Probability theory

Lecture notes

Instructor: Arijit Chakrabarty

Last updated: 31 May 2025

1 Review of measure theory and integration

1.1 Measure theory

Definition 1. *Let Ω be a non-empty set, and $\mathcal{F}$ a class of subsets of Ω. The class $\mathcal{F}$ is a* **field** *if*

- $\Omega \in \mathcal{F}$;
- $A \in \mathcal{F}$ *implies that* $A^c \in \mathcal{F}$;
- $A, B \in \mathcal{F}$ *implies that* $A \cup B \in \mathcal{F}$.

Definition 2. *A field $\mathcal{F}$ of subsets of Ω is a* σ**-field** *if $A_1, A_2, \ldots \in \mathcal{F}$ implies that $\bigcup_{n=1}^\infty A_n \in \mathcal{F}$.*

Definition 3. *Let $\mathcal{F}$ be a collection of subsets of Ω. A function μ from $\mathcal{F}$ to $[0,\infty]$ is* **countably additive** *if $\mu(A) < \infty$ for some $A \in \mathcal{F}$ and for disjoint $\mathcal{F}$-sets $A_1, A_2, \ldots$ such that $\bigcup_{n=1}^\infty A_n \in \mathcal{F}$, it holds that*

$$\mu\left(\bigcup_{n=1}^\infty A_n\right) = \sum_{n=1}^\infty \mu(A_n)\,.$$

If $\mathcal{F}$ is a σ-field, then a μ satisfying the above is a **measure** *and $(\Omega, \mathcal{F}, \mu)$ is a measure space. If $\mathcal{F}$ is a σ-field and $\mu(\Omega) < \infty$, then μ is a* **finite measure***. If $\mathcal{F}$ is a σ-field and there exist $A_1, A_2, \ldots \in \mathcal{F}$ such that $\Omega = \mathcal{A}_1 \cup A_2 \cup \ldots$ and $\mu(A_n) < \infty$ for all n, then μ is a* σ**-finite measure***. If $\mathcal{F}$ is a σ-field and $\mu(\Omega) = 1$, then μ is a* **probability measure** *and $(\Omega, \mathcal{F}, \mu)$ is a probability space.*

Exercise 1.1. *If $\mathcal{F}$ is a field and μ is countably additive on $\mathcal{F}$, show that $\emptyset \in \mathcal{F}$ and*

$$\mu(\emptyset) = 0\,.$$

Theorem 1.1 (Caratheodory extension theorem)**.** *Suppose that $\mathcal{F}$ is a field on Ω, and μ is a countably additive function on $\mathcal{F}$. Then, there exists a measure μ^* on $(\Omega, \sigma(\mathcal{F}))$ such that*

$$\mu^*(A) = \mu(A) \text{ for all } A \in \mathcal{F}\,.$$

The above theorem, which is ubiquitous in construction of measures, is Theorem 11.2, pg 166, of Billingsley (1995).

Definition 4. *A sequence of sets A_n increases to a set A, denoted by $A_n \uparrow A$ if $A_1 \subset A_2 \subset \ldots$, and $A = \bigcup_{n=1}^{\infty} A_n$. Similarly, $A_n \downarrow A$ is also defined.*

Definition 5. *A family $\mathcal{M}$ of subsets of Ω is a* **monotone class** *if*

- *(closed under monotone union) $A_n \in \mathcal{M}$ and $A_n \uparrow A$ implies that $A \in \mathcal{M}$,*
- *(closed under monotone intersection) $A_n \in \mathcal{M}$ and $A_n \downarrow A$ implies that $A \in \mathcal{M}$.*

The following is Theorem 3.4, pg 43, of Billingsley (1995).

Theorem 1.2 (Monotone class theorem)**.** *If $\mathcal{F}$ is a field, and $\mathcal{M}$ is a monotone class, then $\mathcal{F} \subset \mathcal{M}$ implies that $\sigma(\mathcal{F}) \subset \mathcal{M}$.*

Theorem 1.3 (Uniqueness)**.** *Suppose $\mathcal{F}$ is a field, and μ_1, μ_2 are measures on $(\Omega, \sigma(\mathcal{F}))$ which agree on $\mathcal{F}$ and are σ-finite on $\mathcal{F}$. Then*

$$\mu_1 = \mu_2\,.$$

Proof. Follows from Theorem 1.2. □

Definition 6. *A non-empty collection $\mathcal{S}$ of subsets of Ω is a* **semi-field** *if $A, B \in \mathcal{S}$ implies $A \cap B \in \mathcal{S}$ and $A \in \mathcal{S}$ implies*

$$A^c = A_1 \cup \ldots \cup A_n\,,$$

for some disjoint $A_1, \ldots, A_n \in \mathcal{S}$.

Exercise 1.2. *If $\mathcal{S}$ is a semi-field, show that*

$$\mathcal{F} = \{A_1 \cup \ldots \cup A_n : A_1, \ldots, A_n \in \mathcal{S} \text{ are disjoint}\}$$

is the smallest field generated by $\mathcal{S}$.

The following corollary of Theorems 1.1 and 1.3 will be most useful for us.

Corollary 1.1. *Suppose that $\mathcal{S}$ is a semi-field on Ω, and $\mu : \mathcal{S} \to [0, \infty]$ is a countably additive function. Then, there exists a measure μ^* on $(\Omega, \sigma(\mathcal{S}))$ such that*

$$\mu^*(A) = \mu(A) \text{ for all } A \in \mathcal{S}\,.$$

Furthermore, if μ is σ-finite on $\mathcal{S}$, then μ^* is unique.

Exercise 1.3. *Prove Corollary 1.1.*

Exercise 1.4. *Let μ_1 and μ_2 be measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ defined by*

$$\mu_1(A) = \#(A \cap \mathbb{Q})\,, \mu_2(A) = 2\#(A \cap \mathbb{Q})\,, A \in \mathcal{B}(\mathbb{R})\,.$$

Show that μ_1, μ_2 are σ-finite and agree on $\mathcal{S} = \{(a, b] \cap \mathbb{R} : -\infty \le a \le b \le \infty\}$ which is a semi-field but they do not agree on $(\mathbb{R}, \sigma(\mathcal{S}))$. Corollary 1.1 thus fails if μ_1 and μ_2 are σ-finite, but not on $\mathcal{S}$, and everything else holds.

Definition 7. *For $d \ge 1$, $\mathcal{B}(\mathbb{R}^d)$ is the* **Borel σ-field** *on $\mathbb{R}^d$, that is, the σ-field generated by all open sets in $\mathbb{R}^d$. A measure μ on $\left(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\right)$ is* **Radon** *if*

$$\mu(K) < \infty \text{ for all compact } K \subset \mathbb{R}^d\,.$$

The following, which is Theorem 12.4, pg 176, of Billingsley (1995), is fundamental to probability theory.

Theorem 1.4. *If $F : \mathbb{R} \to \mathbb{R}$ is a non-decreasing right continuous function, then there exists a unique Radon measure μ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that*

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

Furthermore, if $F(\infty) = 1$ and $F(-\infty) = 0$, then μ is a probability measure.

In Theorem 1.4, the following conventions are used:

$$F(\infty) = \lim_{x\to\infty} F(x) \text{ if it exists, and } F(-\infty) = \lim_{x\to-\infty} F(x) \text{ if it exists.}$$

Unless mentioned otherwise, $F(\infty)$ and $F(-\infty)$ will mean the above everywhere.

Exercise 1.5. *Show that a Radon measure μ on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ is regular, that is,*

$$\begin{aligned}\mu(A) &= \inf\left\{\mu(U) : U \text{ open, and } U \supset A\right\} \\ &= \sup\left\{\mu(F) : F \text{ closed, and } F \subset A\right\}\,,\end{aligned}$$

for all $A \in \mathcal{B}(\mathbb{R}^d)$.

Exercise 1.6. *Show that μ_1 as in Exc 1.4 is σ-finite but not regular. Thus not all σ-finite measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ are regular.*

Exercise 1.7. *Suppose that* μ_1, μ_2 *are Radon measures on* $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ *such that*

$$\mu_1(U^c) = \mu_2(U^c) = 0$$

for some open set $U \subset \mathbb{R}^d$. *If*

$$\mu_1(R) \le \mu_2(R)\,,$$

for all rectangles $R = (a_1, b_1] \times \ldots \times (a_d, b_d] \subset U$ *with* $a_1, \ldots, a_d, b_1, \ldots, b_d \in \mathbb{Q}$, *then show that*

$$\mu_1(A) \le \mu_2(A)\,, A \in \mathcal{B}(\mathbb{R}^d)\,.$$

1.2 Integration

Definition 8. *Let* $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$ *and*

$$\mathcal{B}\left(\bar{\mathbb{R}}\right) = \sigma\left(\mathcal{B}(\mathbb{R}) \cup \left\{\{-\infty\}, \{\infty\}\right\}\right)\,.$$

That is, $\mathcal{B}\left(\bar{\mathbb{R}}\right)$ *is the smallest* σ*-field which is a superset of* $\mathcal{B}(\mathbb{R})$ *and contain the singleton sets* $\{-\infty\}, \{\infty\}$. *Given a measurable space* $(\Omega, \mathcal{A})$ *and a function* $f : \Omega \to \bar{\mathbb{R}}$, f *is* ***$\mathcal{A}$-measurable*** *if*

$$f^{-1}A \in \mathcal{A} \text{ for all } A \in \mathcal{B}(\bar{\mathbb{R}})\,,$$

where $f^{-1}A = \{\omega \in \Omega : f(\omega) \in A\}$.

Given a measure space $(\Omega, \mathcal{A}, \mu)$ and a measurable $f : \Omega \to [0, \infty]$, the **integral** of f with respect to μ will be denoted by any of the following:

$$\int f\, d\mu\,, \int f(\omega)\, d\mu(\omega)\,, \int f(\omega)\, \mu(d\omega)\,, \int_\Omega f\, d\mu \text{ etc.}$$

For a measurable $f : \Omega \to \bar{\mathbb{R}}$, its **integral is defined** as

$$\int f\, d\mu = \int f^+\, d\mu - \int f^-\, d\mu\,,$$

whenever either $\int f^+\, d\mu < \infty$ *or* $\int f^-\, d\mu < \infty$, where $x^+ = x \vee 0$ and $x^- = (-x) \vee 0$ for all $x \in \bar{\mathbb{R}}$. The above is defined even when the right hand side is $\pm\infty$; "$\infty - \infty$" is the only case when it is undefined.

We say f is **integrable** if $\int f^+\, d\mu < \infty$ *and* $\int f^-\, d\mu < \infty$, which happens if and only if

$$\int |f|\, d\mu < \infty\,.$$

In other words, f is integrable, f has a finite integral, $|f|$ is integrable, f^+ and f^- are integrable, $f \in L^1$ all mean the same. However, "$\int f\, d\mu$ is defined" is **not the same** as "f is integrable", which deserves emphasis.

Exercise 1.8. *Show that the integral of f, if defined, remains unchanged if the underlying σ-field is changed to anything with respect to which f is measurable.*

Theorem 1.5. *Suppose $g, f, f_1, f_2, \ldots$ are measurable functions from a measure space $(\Omega, \mathcal{A}, \mu)$ to $\bar{\mathbb{R}}$.*

1. *(Monotone convergence theorem) If $0 \le f_n \uparrow f$, then $\int f_n d\mu \uparrow \int f d\mu$.*
2. *(Fatou's lemma) If $f_n \ge 0$, then*
$$\int \left(\liminf_{n\to\infty} f_n \right) d\mu \le \liminf_{n\to\infty} \int f_n \, d\mu \, .$$
3. *(Dominated convergence theorem) If $f_n \to f$, $|f_n| \le g$ and g is integrable, then f is integrable and*
$$\int f_n \, d\mu \to \int g \, d\mu \, .$$

The above are Theorems 16.2, 16.3 and 16.4 on pg 208-209 of Billingsley (1995), respectively.

Definition 9. *Suppose $(\Omega_1, \mathcal{A}_1, \mu)$ is a measure space, $(\Omega_2, \mathcal{A}_2)$ is a measurable space and $T : \Omega_1 \to \Omega_2$ is a measurable map, that is,*
$$T^{-1}A \in \mathcal{A}_1 \text{ for all } A \in \mathcal{A}_2 \, .$$
The **push forward measure** *of μ under T is the measure $\mu \circ T^{-1}$ on $(\Omega_2, \mathcal{A}_2)$ defined by*
$$\mu \circ T^{-1}(A) = \mu\left(T^{-1}A\right) \, , A \in \mathcal{A}_2 \, .$$

Theorem 1.6. *Suppose $(\Omega_1, \mathcal{A}_1, \mu)$ is a measure space, $(\Omega_2, \mathcal{A}_2)$ is a measurable space and $T : \Omega_1 \to \Omega_2$ is a measurable map. Then, for a measurable $f : \Omega_2 \to \bar{\mathbb{R}}$,*
$$\int_{\Omega_1} f\left(T(x)\right) \, d\mu(x) = \int_{\Omega_2} f(y) \, d\left(\mu \circ T^{-1}\right)(y) \, ,$$
whenever the integral on either side is defined.

Exercise 1.9. *Prove the above theorem by first showing it when $f = \mathbf{1}_A$ for some $A \in \mathcal{A}_2$, then for non-negative simple functions f and finally using the monotone convergence theorem.*

Definition 10. *If μ and ν are measures on $(\Omega, \mathcal{A})$, then μ is absolutely continuous with respect to ν, we write $\mu \ll \nu$, if*
$$\nu(A) = 0 \Rightarrow \mu(A) = 0 \, , \text{ for all } A \in \mathcal{A} \, .$$

Theorem 1.7 (Radon-Nikodym)**.** *Suppose* $(\Omega, \mathcal{A})$ *is a measurable space on which* μ, ν *are* σ*-finite measures such that* $\mu \ll \nu$*. Then there exists a measurable* $f : \Omega \to [0, \infty)$ *such that*

$$\int_A f \, d\nu = \mu(A) \, , A \in \mathcal{A} \, . \tag{1.1}$$

If (1.1) *holds with* f *replaced by any other function* g*, then* $g = f$ ν*-a.e.*

The above is Theorem 32.2, page 422, of Billingsley (1995).

Definition 11. *The function* f *satisfying* (1.1) *is the "Radon-Nikodym derivative of* μ *with respect to* ν*", and is denoted by*

$$f = \frac{d\mu}{d\nu} \, .$$

Exercise 1.10. *Suppose* μ, ν *are* σ*-finite measures on* $(\Omega, \mathcal{A})$ *and* $\mu \ll \nu$*. Then, for a measurable* $g : \Omega \to \bar{\mathbb{R}}$*,*

$$\int g \, d\mu = \int g \frac{d\mu}{d\nu} \, d\nu \, ,$$

whenever the integral on either side is defined.

Definition 12. *If* $(\Omega_1, \mathcal{A}_1)$ *and* $(\Omega_2, \mathcal{A}_2)$ *are measurable spaces, then the product* σ*-field* $\mathcal{A}_1 \otimes \mathcal{A}_2$ *is defined by*

$$\mathcal{A}_1 \otimes \mathcal{A}_2 = \sigma \left(A_1 \times A_2 : A_1 \in \mathcal{A}_1, \; A_2 \in \mathcal{A}_2 \right) \, .$$

Exercise 1.11. *Show that*

$$\mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}) = \mathcal{B}\left(\mathbb{R}^2\right) \, .$$

Exercise 1.12. *If* $(\Omega_1, \mathcal{A}_1, \mu_1)$ *and* $(\Omega_2, \mathcal{A}_2, \mu_2)$ *are* σ*-finite measure spaces, then show that there there exists a unique measure* $\mu_1 \otimes \mu_2$ *on* $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$ *satisfying*

$$\mu_1 \otimes \mu_2(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2) \, , A_1 \in \mathcal{A}_1, \; A_2 \in \mathcal{A}_2 \, .$$

Show that $\mu_1 \otimes \mu_2$ *is* σ*-finite.*

Theorem 1.8 (Tonelli)**.** *Suppose* $(\Omega_1, \mathcal{A}_1, \mu_1)$ *and* $(\Omega_2, \mathcal{A}_2, \mu_2)$ *are* σ*-finite measure spaces and* $f : \Omega_1 \times \Omega_2 \to [0, \infty]$ *is* $\mathcal{A}_1 \otimes \mathcal{A}_2$*-measurable. Then, for a fixed* $\omega_1 \in \Omega_1$*,*

$$f(\omega_1, \cdot) \text{ is measurable w.r.t. } \mathcal{A}_2 \, ,$$

$$\int_{\Omega_2} f(\cdot, \omega_2) \, \mu_2(d\omega_2) \text{ if measurable w.r.t. } \mathcal{A}_1 \, ,$$

and likewise with the roles of ω_1 *and* ω_2 *interchanged. Furthermore,*

$$\begin{aligned} \int_{\Omega_1 \times \Omega_2} f \, d(\mu_1 \otimes \mu_2) &= \int_{\Omega_1} \left(\int_{\Omega_2} f(\omega_1, \omega_2) \, \mu_2(d\omega_2) \right) \mu_1(d\omega_1) \\ &= \int_{\Omega_2} \left(\int_{\Omega_1} f(\omega_1, \omega_2) \, \mu_1(d\omega_1) \right) \mu_2(d\omega_2) \, . \end{aligned}$$

Convention

The usual convention for iterated integrals is the following:

$$\int_{\Omega_1}\int_{\Omega_2} f(\omega_1,\omega_2)\,\mu_2(d\omega_2)\mu_1(d\omega_1) = \int_{\Omega_1}\left(\int_{\Omega_2} f(\omega_1,\omega_2)\,\mu_2(d\omega_2)\right)\mu_1(d\omega_1)\,,$$

that is, the left hand side above means the right hand side.

Exercise 1.13. *Suppose* $(\Omega_1,\mathcal{A}_1,\mu_1)$ *and* $(\Omega_2,\mathcal{A}_2,\mu_2)$ *are* σ*-finite measure spaces and* $f:\Omega_1\times\Omega_2\to\bar{\mathbb{R}}$ *is* $\mu_1\otimes\mu_2$*-integrable. Show that*

$$\int_{\Omega_2}|f(\omega_1,\omega_2)|\,\mu_2(d\omega_2)<\infty \text{ for almost every } \omega_1\in\Omega_1\,,$$

and that the similar statement holds for integrals over Ω_1*.*

Theorem 1.9 (Fubini)**.** *Suppose* $(\Omega_1,\mathcal{A}_1,\mu_1)$ *and* $(\Omega_2,\mathcal{A}_2,\mu_2)$ *are* σ*-finite measure spaces and* $f\in L^1\left(\Omega_1\times\Omega_2,\mu_1\otimes\mu_2\right)$*. Then*

$$\int\int f(\omega_1,\omega_2)\,\mu_1(d\omega_1)\mu_2(d\omega_2) = \int\int f(\omega_1,\omega_2)\,\mu_2(d\omega_2)\mu_1(d\omega_1)\,.$$

The theorems of Tonelli and Fubini are subsumed in Theorem 18.3, pg 234, Billingsley (1995).

Definition 13. *The Lebesgue measure* λ *is the measure on* $(\mathbb{R},\mathcal{B}(\mathbb{R}))$ *satisfying*

$$\lambda\big((a,b]\big)=b-a\,,\ \text{for all} -\infty<a\le b<\infty\,,$$

the existence and uniqueness of which is guaranteed by Theorem 1.4 by taking $F(x)=x$*. For a Borel measurable function* $f:(a,b)\to\bar{\mathbb{R}}$*, where* $-\infty\le a<b\le\infty$*, its Lebesgue integral is defined as*

$$\int_a^b f(x)\,dx=\int_{(a,b)} f(x)\,\lambda(dx)\,,$$

whenever the right hand side is defined.

Theorem 1.10 (Fundamental theorem of calculus)**.** *Suppose* $-\infty<a<b<\infty$ *and* $F:[a,b]\to\mathbb{R}$ *is differentiable on* (a,b) *and continuous at* a *and* b*. Then* F'*, the derivative of* F*, is Borel measurable. If*

$$\int_a^b |F'(x)|\,dx<\infty\,, \tag{1.2}$$

then

$$\int_a^b F'(x)\,dx=F(b)-F(a)\,.$$

If (1.2) *holds with* $b=\infty$*, then the above holds as well with* $F(b)$ *replaced by* $\lim_{x\to\infty}F(x)$ *which necessarily exists, and likewise if* $-\infty=a<b\le\infty$*.*

The above is Theorem 7.21, page 149, of Rudin (1987).

Theorem 1.11 (Change of variable)**.** *If $U, V \subset \mathbb{R}$ are open sets, $\psi : U \to V$ is a C^1 bijection whose derivative ψ' never vanishes, then*

$$\int_U f \circ \psi(x) |\psi'(x)| \, dx = \int_V f(y) \, dy \,,$$

for a measurable $f : V \to \bar{\mathbb{R}}$ whenever the integral on either side makes sense.

In other words, for substituting $y = \psi(x)$, dy is to replaced by $|\psi'(x)| dx$. It is being emphasized that if $|\psi'|$ is replaced by ψ then the formula is incorrect. The above is a special case of the change of variables formula in d dimensions, which is Theorem 1.12 below.

Exercise 1.14. *Calculate*

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(x^2+y^2)} \, dx \, dy \,.$$

Soln.: Let

$$\begin{aligned} I &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(x^2+y^2)} \, dx \, dy \\ \text{(Tonelli)} \;\; &= \int_{-\infty}^{\infty} e^{-x^2} \int_{-\infty}^{\infty} e^{-y^2} \, dy \, dx \,. \end{aligned}$$

For a fixed $x \in \mathbb{R} \setminus \{0\}$, put $y = xz$ using Theorem 1.11 to get

$$\int_{-\infty}^{\infty} e^{-y^2} \, dy = |x| \int_{-\infty}^{\infty} e^{-z^2 x^2} \, dz \,.$$

Thus

$$\begin{aligned} I &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |x| e^{-x^2(1+z^2)} \, dz \, dx \\ \text{(Tonelli)} \;\; &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |x| e^{-x^2(1+z^2)} \, dx \, dz \,. \end{aligned}$$

For a fixed $z \in \mathbb{R}$,

$$\begin{aligned} \int_{-\infty}^{\infty} |x| e^{-x^2(1+z^2)} \, dx &= 2 \int_0^{\infty} x e^{-x^2(1+z^2)} \, dx \\ \text{(Theorem 1.11: } y = x^2,\ dy = 2x dx) &= \int_0^{\infty} e^{-y(1+z^2)} \, dz \\ \text{(Theorem 1.10)} &= \left[-\frac{e^{-y(1+z^2)}}{1+z^2} \right]_{y=0}^{y=\infty} \\ &= \frac{1}{1+z^2} \,. \end{aligned}$$

Therefore

$$\begin{aligned} I &= \int_{-\infty}^{\infty} \frac{1}{1+z^2}\, dz \\ \text{(Theorem 1.10)} &= \left[\tan^{-1} z\right]_{-\infty}^{\infty} \\ &= \frac{\pi}{2} - \left(-\frac{\pi}{2}\right) = \pi\,. \end{aligned}$$

This completes the solution of the exercise.

An immediate consequence of the above exercise is

$$\int_{-\infty}^{\infty} e^{-x^2}\, dx = \sqrt{\pi}\,,$$

showing with the help of Theorem 1.11 that

$$\int_{-\infty}^{\infty} e^{-y^2/2}\, dy = \sqrt{2\pi}\,.$$

Definition 14. *The Lebesgue measure λ_d is the d-fold product of the one-dimensional Lebesgue measure λ, that is, λ_d is the unique measure on the space $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ satisfying*

$$\lambda_d\left(A_1 \times \ldots \times A_d\right) = \prod_{i=1}^{d} \lambda(A_i)\,, A_1, \ldots, A_d \in \mathcal{B}(\mathbb{R})\,.$$

For stating the next result, a Jacobian matrix has to be first defined. Consider an open set $U \subset \mathbb{R}^d$ and a function $F : U \to \mathbb{R}^d$. Denote by $f_1, \ldots, f_d$ the coordinate functions of F, that is,

$$F(x) = \big(f_1(x), \ldots, f_d(x)\big)\,, x \in \mathbb{R}^d\,.$$

If the first partial derivatives of F exist, that is, $\partial f_i(x)/\partial x_j$ exists for all $x \in U$ and $1 \le i, j \le d$, then its Jacobian matrix at x, denoted by $J(x)$, is a $d \times d$ matrix defined by

$$J(x) = \left(\frac{\partial f_i(x)}{\partial x_j}\right)_{1 \le i,j \le d}\,, x \in U\,,$$

that is, the (i, j)-th entry of $J(x)$ is $\partial f_i(x)/\partial x_j$. The statement of the theorem is the following, of which Theorem 1.11 is a special case.

Theorem 1.12. *For open subsets U and V of $\mathbb{R}^d$, let $T : U \to V$ be a bijection which is continuously differentiable, that is, the first partial derivatives of T exist and are continuous. Assume that its Jacobian matrix $J(x)$ is non-singular for all $x \in U$. Then for any non-negative measurable function $f : V \to \mathbb{R}$,*

$$\int_U f\big(T(x)\big)\, |\det(J(x))|\, dx = \int_V f(y)\, dy\,,$$

$\det(A)$ *denoting the determinant of A for any square matrix A.*

For the sake of completeness, a proof of the above theorem is provided. The following facts from linear algebra and multivariable analysis are needed.

Fact 1.1. *If $T : \mathbb{R}^d \to \mathbb{R}^d$ is a linear map, then for a compact rectangle*

$$R = [a_1, b_1] \times \ldots \times [a_d, b_d] \subset \mathbb{R}^d \,,$$

with $-\infty < a_i < b_i < \infty$ for $i = 1, \ldots, d$, it holds that

$$\lambda\left(\{T(x) : x \in R\}\right) = |\det(T)| \prod_{i=1}^d (b_i - a_i) = |\det(T)|\lambda(R) \,,$$

where λ is the Lebesgue measure on $\mathbb{R}^d$.

The following is the inverse function theorem.

Fact 1.2. *Let $U \subset \mathbb{R}^d$ be an open set and $T : U \to \mathbb{R}^d$ be continuously differentiable. Denoting by $J(x)$ the Jacobian matrix of T at $x \in U$, assume that $J(x_0)$ is non-singular for some $x_0 \in U$. Then, there exists an open neighbourhood X of x_0 such that T is one-one on X, the set $T(X)$ is open, T^{-1} is continuously differentiable on $T(X)$ and the Jacobian matrix of T^{-1} at y is $\left(J(T^{-1}y)\right)^{-1}$ for all $y \in T(X)$.*

The following is another fact from multivariable analysis which essentially follows from the one-dimensional mean value theorem.

Fact 1.3. *Suppose that $U \subset \mathbb{R}^d$ is open, $R \subset U$ is a closed rectangle and $T : U \to \mathbb{R}^d$ is continuously differentiable such that*

$$|J_{ij}(y) - J_{ij}(x)| \le \alpha \,, x, y \in R \,, 1 \le i, j \le d \,,$$

where $J_{ij}(z)$ is the (i,j)-th entry of the Jacobian matrix $J(z)$ of T at z for all $z \in U$ and $1 \le i, j \le d$. Then,

$$\|T(x) - T(y) - J(x)(x-y)\| \le d\alpha\|x - y\| \,, x, y \in R \,,$$

where $\|\cdot\|$ is the L^∞ norm on $\mathbb{R}^d$ defined by

$$\|x\| = \max_{1 \le i \le d} |x_i| \,, x = (x_1, \ldots, x_d) \in \mathbb{R}^d \,, \tag{1.3}$$

$x, y, T(x), T(y)$ are viewed as $d \times 1$ vectors and hence $J(x)(x-y)$ is also a $d \times 1$ vector.

A proof of the above fact is provided in Subsection 9.1 of the Appendix.

Proof of Theorem 1.12

The proof of Theorem 1.12 will be executed by sequentially showing each step below. Step 4. would complete the proof.

Step 1. For any compact rectangle $R = [a_1, b_1] \times \ldots \times [a_d, b_d] \subset U$ with $-\infty < a_i < b_i < \infty$ for $i = 1, \ldots, d$, and $a_1, \ldots, a_d, b_1, \ldots, b_d \in \mathbb{Q}$,

$$\lambda\big(T(R)\big) \le \int_R |\det(J(x))|\, dx\,. \tag{1.4}$$

Step 2. For all $A \in \mathcal{B}(\mathbb{R}^d)$,

$$\lambda\big(T(A \cap U)\big) \le \int_{A\cap U} |\det(J(x))|\, dx\,. \tag{1.5}$$

Step 3. For any non-negative measurable function $f : V \to \mathbb{R}$,

$$\int_U f\big(T(x)\big)\, |\det(J(x))|\, dx \ge \int_V f(y)\, dy\,. \tag{1.6}$$

Step 4. The inequality in (1.6) is an equality.

The proof of Step 1., which is the main step of the proof, is based on the idea that locally T is like a linear transformation.

Proof of Step 1. Fix a compact rectangle $R = [a_1, b_1] \times \ldots \times [a_d, b_d] \subset U$ where $a_i < b_i$ and $a_1, \ldots, a_d, b_1, \ldots, b_d \in \mathbb{Q}$. Let $\varepsilon > 0$. Since $\det(J(\cdot))$ is a continuous function, it is uniformly continuous on R. Choose $\delta_1 > 0$ such that

$$\big|\det(J(x)) - \det(J(x'))\big| \le \varepsilon \text{ for all } x, x' \in R,\ \|x - x'\| \le \delta_1\,, \tag{1.7}$$

where $\|\cdot\|$ denotes the L^∞ norm as in (1.3) throughout.

Recall that the function $A \to A^{-1}$, from the space of $d \times d$ non-singular matrices to itself, is continuous. Since $J(x)$ is non-singular for all $x \in U$, the map $x \mapsto J(x)^{-1}$ is continuous on U. Thus,

$$f : R \times \{z \in \mathbb{R}^d : \|z\| = 1\} \to \mathbb{R}^d\,,$$

defined by

$$f(x, z) = J(x)^{-1} z\,, (x, z) \in R \times \{z \in \mathbb{R}^d : \|z\| = 1\}\,,$$

is a continuous function defined on a compact set; elements of $\mathbb{R}^d$ are viewed as $d \times 1$ vectors by convention. Therefore,

$$c = \max\Big\{\|f(x, z)\| :\ (x, z) \in R \times \{z \in \mathbb{R}^d : \|z\| = 1\}\Big\} < \infty\,.$$

In other words,

$$\big\|J(x)^{-1} z\big\| \le c\|z\|\,, x \in R,\ z \in \mathbb{R}^d\,. \tag{1.8}$$

Denote by $J_{ij}(x)$ the (i,j)-th entry of $J(x)$ for all $x \in U$ and $1 \le i,j \le d$. Uniform continuity of $J_{ij}(\cdot)$ on R ensures the existence of $\delta_2 > 0$ such that

$$\left|J_{ij}(x) - J_{ij}(x')\right| \le \frac{\varepsilon}{cd} \text{ for all } x, x' \in R,\ \|x - x'\| \le \delta_2\,. \tag{1.9}$$

Let $0 < \delta \le \min\{\delta_1, \delta_2\}$ be such that $\delta^{-1}(b_i - a_i)$ is an integer for every i. Choosing such a δ is possible because $b_i - a_i$ is rational; if p_i, q_i are positive integers with $b_i - a_i = p_i/q_i$, letting

$$\delta = \frac{1}{nq_1 \ldots q_d}\,,$$

works for large n, for example.

Consider the square

$$[a_1 + (i_1 - 1)\delta, a_1 + i_1\delta] \times \ldots \times [a_d + (i_d - 1)\delta, a_d + i_d\delta]\ ,$$

where $i_1, \ldots, i_d$ are positive integers with $i_j \le \delta^{-1}(b_j - a_j)$ for $j = 1, \ldots, d$. Denote the collection of all such squares by $\{Q_1, \ldots, Q_k\}$. In other words, $Q_1, \ldots, Q_k$ are compact squares of side-length δ such that

$$R = Q_1 \cup \ldots \cup Q_k\,,$$

and $\lambda(Q_i \cap Q_j) = 0$ for $1 \le i < j \le k$. Let x_i be the centre of Q_i (the centre of a square or a rectangle is well defined). Recalling that $\|\cdot\|$ is the L^∞ norm, write

$$Q_i = B_{\delta/2}(x_i)\,, i = 1, \ldots, k\,, \tag{1.10}$$

where for $r \ge 0$ and $z = (z_1, \ldots, z_d) \in \mathbb{R}^d$,

$$B_r(z) = \{y \in \mathbb{R}^d : \|y - z\| \le r\} = [z_1 - r, z_1 + r] \times \ldots \times [z_d - r, z_d + r]\,. \tag{1.11}$$

The above is precisely the advantage of working with the L^∞ norm.

For $i = 1, \ldots, k$, fix $x_i \in Q_i$ and define

$$\phi_i(z) = J(x_i)(z - x_i) + T(x_i)\,, z \in \mathbb{R}^d\,.$$

Our first claim is that

$$T\left(Q_i\right) \subset \phi_i\left(Q_i^\varepsilon\right)\,, i = 1, \ldots, k\,, \tag{1.12}$$

where

$$Q_i^\varepsilon = B_{(1+\varepsilon)\delta/2}(x_i)\,, i = 1, \ldots, k\,.$$

Proceeding towards proving (1.12), fix $i \in \{1, \ldots, k\}$, and use Fact 1.3 along with (1.9) to claim that for all $z \in Q_i$,

$$\|T(z) - T(x_i) - J(x_i)(z - x_i)\| \le \frac{\varepsilon}{c}\|z - x_i\|\,.$$

Since the left hand side above equals $\|T(z) - \phi_i(z)\|$, it follows that

$$\|T(z) - \phi_i(z)\| \le \frac{\varepsilon}{c}\|z - x_i\|\,, z \in Q_i\,. \tag{1.13}$$

Therefore, for $z \in Q_i$,

$$\begin{aligned}\left\|\phi_i^{-1} \circ T(z) - z\right\| &= \left\|\phi_i^{-1} \circ T(z) - \phi_i^{-1} \circ \phi_i(z)\right\| \\ &= \left\|J(x_i)^{-1}\left(T(z) - \phi_i(z)\right)\right\| \\ &\le c\,\|T(z) - \phi_i(z)\| \\ &\le \varepsilon\|z - x_i\|\,,\end{aligned}$$

(1.8) and (1.13) implying the inequalities in the penultimate line and the last line, respectively. Thus, for $z \in Q_i$,

$$\left\|\phi_i^{-1} \circ T(z) - x_i\right\| \le \left\|\phi_i^{-1} \circ T(z) - z\right\| + \|z - x_i\| \le (1+\varepsilon)\|z - x_i\|\,.$$

Recall (1.10) to argue that

$$\phi_i^{-1} \circ T(z) \in Q_i^{\varepsilon}\,, z \in Q_i\,,$$

which is equivalent to (1.12).

An immediate implication of (1.12) is that for fixed $i = 1, \ldots, k$,

$$\begin{aligned}\lambda\left(T(Q_i)\right) &\le \lambda\left(\{J(x_i)z + T(x_i) - J(x_i)x_i : z \in Q_i^{\varepsilon}\}\right) \\ &= \lambda\left(\{J(x_i)z : z \in Q_i^{\varepsilon}\}\right) \\ &= |\det(J(x_i))|\lambda(Q_i^{\varepsilon})\,,\end{aligned}$$

the second line following from the translation-invariance of the Lebesgue measure, and Fact 1.1 and the observation that Q_i^{ε} is a rectangle implying the last line. This is the crux of the proof in that it shows how the modulus of the determinant of the Jacobian appears. Further, (1.11) shows Q_i^{ε} is a square of side-length $(1+\varepsilon)\delta$. Therefore,

$$\lambda(Q_i^{\varepsilon}) = (1+\varepsilon)^d\delta^d = (1+\varepsilon)^d\lambda(Q_i)\,,$$

(1.10) implying the second equality. Put everything together to get

$$\lambda\left(T(Q_i)\right) \le |\det(J(x_i))|(1+\varepsilon)^d\lambda(Q_i)\,.$$

Thus,

$$\begin{aligned}\lambda\left(T(R)\right) &= \sum_{i=1}^{k}\lambda(T(Q_i)) \\ &\le (1+\varepsilon)^d\sum_{i=1}^{k}|\det(J(x_i))|\lambda(Q_i) \\ &\le (1+\varepsilon)^d\sum_{i=1}^{k}\left(\varepsilon + \min_{z \in Q_i}|\det(J(z))|\right)\lambda(Q_i) \\ &\le (1+\varepsilon)^d\left(\varepsilon\lambda(R) + \int_R|\det(J(x))|dx\right)\,,\end{aligned}$$

(1.7) and that $\delta \le \delta_1$ implying the penultimate line. Since the above holds for all $\varepsilon > 0$, letting $\varepsilon \downarrow 0$ completes the proof of Step 1. $\square$

While Step 1. was mostly based on analysis and linear algebra, the proof of Step 2. is standard in measure theory and follows from Exc 1.7.

Proof of Step 2. Define measures μ and ν on $\mathbb{R}^d$ by

$$\mu(A) = \lambda(T(A \cap U))\,, A \in \mathcal{B}(\mathbb{R}^d)\,,$$

and

$$\nu(B) = \int_{B\cap U} |\det(J(x))|\; dx\,, B \in \mathcal{B}(\mathbb{R}^d)\,.$$

The claim (1.5) is equivalent to

$$\mu(A) \le \nu(A)\,, A \in \mathcal{B}(\mathbb{R}^d)\,. \tag{1.14}$$

In view of Exc 1.7, it suffices to show that the claim holds for any compact rectangle with rational corners, that is,

$$\mu(R) \le \nu(R)\,, \tag{1.15}$$

if $R = [a_1, b_1] \times \ldots \times [a_d, b_d] \subset U$ for some $a_1, \ldots, a_d, b_1, \ldots, b_d \in \mathbb{Q}$ with $a_i < b_i$, which is precisely what has been shown in Step 1. $\square$

The proof of Step 3., which is also standard, is based on approximating a non-negative measurable function by simple functions from below.

Proof of Step 3. First let $f : V \to \mathbb{R}$ be a non-negative simple function, that is,

$$f = \sum_{i=1}^{k} \alpha_i \mathbf{1}_{A_i}\,,$$

for some $\alpha_1, \ldots, \alpha_k \in [0, \infty]$ and $A_1, \ldots, A_k \in \mathcal{B}(\mathbb{R}^d)$ with $A_i \subset V$ for all i.

Then,

$$\begin{aligned}
\int_V f(y)\,dy &= \sum_{i=1}^k \alpha_i \lambda(A_i) \\
&= \sum_{i=1}^k \alpha_i \lambda(T(T^{-1}A_i)) \\
&\le \sum_{i=1}^k \alpha_i \int_{T^{-1}A_i} |\det(J(x))|\,dx \\
&= \int_U |\det(J(x))| \sum_{i=1}^k \alpha_i \mathbf{1}_{T^{-1}A_i}(x)\,dx \\
&= \int_U |\det(J(x))| \sum_{i=1}^k \alpha_i \mathbf{1}_{A_i}(T(x))\,dx \\
&= \int_U |\det(J(x))|\, f(T(x))\,dx\,,
\end{aligned}$$

the inequality in the third line following from Step 2. Thus,

$$\int_V f(y)\,dy \le \int_U |\det(J(x))|\, f(T(x))\,dx\,. \tag{1.16}$$

For a measurable function $f : V \to [0,\infty)$, there exist non-negative simple functions f_n such that $f_n \uparrow f$. The desired inequality (1.16) holds with f replaced by f_n therein. Letting $n \to \infty$ with the help of MCT, the proof of Step 3. follows. $\square$

Step 4. is a consequence of the inverse function theorem.

Proof of Step 4. Fact 1.2 and the assumption that $J(x)$ is non-singular for all $x \in U$ imply that $T^{-1} : V \to U$ is a continuously differentiable bijection whose Jacobian matrix is $J(T^{-1}y)^{-1}$ for all $y \in V$. Using Step 3. with U, V, T replaced by V, U, T^{-1} implies

$$\int_U g(x)\,dx \le \int_V g \circ T^{-1}(y) \left|\det\left(J(T^{-1}y)^{-1}\right)\right|\,dy\,, \tag{1.17}$$

for any measurable $g : U \to [0,\infty)$.

Fix a measurable $f : V \to [0,\infty)$. Define

$$g(x) = f \circ T(x) |\det(J(x))|\,, x \in U\,.$$

Apply (1.17) to this g to get

$$\begin{aligned}\int_U f\circ T(x)|\det(J(x))|\,dx &\le \int_V g\circ T^{-1}(y)\left|\det\left(J(T^{-1}y)^{-1}\right)\right|\,dy\\ &= \int_V f(y)\left|\det\left(J(T^{-1}y)\right)\right|\left|\det\left(J(T^{-1}y)^{-1}\right)\right|\,dy\\ &= \int_V f(y)\,dy\,.\end{aligned}$$

Compare this with (1.6) obtained in Step 3. to get

$$\int_U f\circ T(x)|\det(J(x))|\,dx = \int_V f(y)\,dy\,.$$

This completes the proof of Step 4. and that of Theorem 1.12 as well. □

2 Random experiments and random variables

A random experiment is an experiment for which there is a set of possible outcomes, of which any one may occur. Though it cannot be predicted which outcome will occur, the "probabilities" of those outcomes are understood from intuition. For example, if a fair coin is tossed, then the possible outcomes are head and tail, each occurring with probability 1/2. No attempts will be made to give a mathematical definition of random experiments.

Given a random experiment, a probability space is associated with it, which naturally captures our intuition about the experiment. In other words, the probability space is the mathematically precise starting point of the study of probability theory. This is best understood from the following few examples. The set of all possible outcomes is called the "sample space" and is usually denoted by Ω.

Example 2.1. *A fair die is rolled n times. The sample space of this experiment is*

$$\Omega = \left\{(x_1,\ldots,x_n) : x_i\in\{1,\ldots,6\},\, i=1,\ldots,n\right\}.$$

The probability space associated with this experiment is $(\Omega,\mathcal{A},P)$ where $\mathcal{A}=\mathcal{P}$ is the power set of Ω and

$$P(A)=\frac{\#A}{\#\Omega}\,, A\subset\Omega\,.$$

Example 2.2. *A fair coin is tossed till the first head is obtained. The sample space is*

$$\Omega=\{H,\,HT,\,HHT,\ldots\}\,.$$

Define $p:\Omega\to[0,1]$ by

$$p\big(T\ldots T(n\ \text{times})H\big)=2^{-n-1}\,, n=0,1,2,\ldots\,,$$

and

$$P(A) = \sum_{\omega \in A} p(\omega)\,, A \subset \Omega\,.$$

Thus $(\Omega, \mathcal{P}(\Omega), P)$ *is the probability space associated with this experiment.*

As in the above examples, if the sample space Ω is countable and $p(\omega) \geq 0$ is the probability of the outcome ω for all $\omega \in \Omega$, where

$$\sum_{\omega \in \Omega} p(\omega) = 1\,,$$

then letting

$$P(A) = \sum_{\omega \in A} p(\omega)\,, A \subset \Omega\,,$$

$\big(\Omega, \mathcal{P}(\Omega), P\big)$ is the natural probability space. The above method, however, fails for an uncountable sample space. For random experiments with such a sample space, measure theory is essential, as illustrated in the next example.

Example 2.3. *A fair coin is tossed infinitely often. That is, for* $n = 1, 2, 3, \ldots$, *there is a* n*-th toss which yields either a head or a tail. The sample space of this experiment is*

$$\Omega = \big\{(\omega_1, \omega_2, \omega_3, \ldots) : \omega_n \in \{H, T\}\,, n = 1, 2, \ldots\big\}\,,$$

which is clearly uncountable. In order to associate a probability measure on a suitably chosen collection of subsets of Ω, *we first need to decide what are the events of interest. In practice, we would be interested in events whose occurrence is decided by the first finitely many tosses, or events that can be constructed from them. Keeping this in mind, we define*

$$E_{\omega_1 \ldots \omega_n} = \{(\omega_1', \omega_2', \ldots) \in \Omega : \omega_i' = \omega_i,\ 1 \leq i \leq n\},$$

for all $n \in \mathbb{N}, \omega_1, \ldots, \omega_n \in \{H, T\}$. *That is,* $E_{\omega_1 \ldots \omega_n}$ *is the event that the first toss yields* ω_1, *the outcome of the second toss is* ω_2, *and so on till the* n*-th toss. Define*

$$\mathcal{S} = \{\emptyset\} \cup \big\{E_{\omega_1 \ldots \omega_n} : n \in \mathbb{N}, \omega_1, \ldots, \omega_n \in \{H, T\}\big\}\,.$$

The following several exercises show that there is a unique probability measure P *on* $(\Omega, \sigma(\mathcal{S}))$ *satisfying*

$$P\left(E_{\omega_1 \ldots \omega_n}\right) = 2^{-n}\,, n = 1, 2, \ldots,\ \omega_1, \ldots, \omega_n \in \{H, T\}\,, \tag{2.1}$$

which is precisely the claim of Exc 2.6.

Exercise 2.1. *If* $\mathcal{S}$ *is a semi-field and* $A, B_1, \ldots, B_n \in \mathcal{S}$, *show that*

$$A \setminus (B_1 \cup \ldots \cup B_n) = C_1 \cup \ldots \cup C_k\,,$$

for some $k \geq 1$ *and disjoint* $C_1, \ldots, C_k \in \mathcal{S}$.

Definition 15. *Given a non-empty set Ω and a non-empty collection $\mathcal{C}$ of subsets of Ω, $\mathcal{C}$ has the "Cantor intersection property" if the following holds. Whenever $C_1, C_2, C_3 \ldots \in \mathcal{C}$ are such that $C_1 \supset C_2 \supset C_3 \supset \ldots$ and $C_n \neq \emptyset$ for all n, it holds that*

$$\bigcap_{n=1}^{\infty} C_n \neq \emptyset\,.$$

For example, the collection of all compact sets of $\mathbb{R}^d$ has the Cantor intersection property.

Exercise 2.2. *Suppose $\mathcal{S}$ is a semi-field having the Cantor intersection property. If $A_1, A_2, \ldots \in \mathcal{S}$ are disjoint and $\bigcup_{n=1}^{\infty} A_n \in \mathcal{S}$, show that $A_n = \emptyset$ for all but finitely many n's.*

Hint.: Assume $A_1, A_2, \ldots \in \mathcal{S}$ are disjoint,

$$A = \bigcup_{n=1}^{\infty} A_n \in \mathcal{S}\,,$$

and $A_n \neq \emptyset$ for infinitely many n's. Use Exc 2.1 to write

$$A \setminus A_1 = C_1 \cup \ldots \cup C_k\,,$$

for some $C_1, \ldots, C_k \in \mathcal{S}$. Then for some i,

$$C_i \not\subset A_2 \cup \ldots \cup A_n \text{ for all finite } n\,,$$

because otherwise A is the union of finitely many A_n's which would imply that all but finitely many of A_n's are empty. Let $B_1 = C_i$. Apply Exc 2.1 to $B_1 \setminus (A_1 \cup A_2)$ to get $B_2 \in \mathcal{S}$ such that $B_2 \subset B_1 \setminus (A_1 \cup A_2)$ and

$$B_2 \not\subset A_3 \cup \ldots \cup A_n \text{ for all finite } n\,.$$

Proceed inductively to obtain $B_{n+1} \in \mathcal{S}$ with

$$B_{n+1} \subset B_n \setminus (A_1 \cup \ldots \cup A_{n+1})\,, \tag{2.2}$$

and

$$B_{n+1} \not\subset A_{n+2} \cup \ldots \cup A_{n+k} \tag{2.3}$$

for any finite k.

Thus $B_1, B_2, \ldots \in \mathcal{S}$ with $A \supset B_1 \supset B_2 \supset \ldots$, and (2.2) implies that for $n = 2, 3, \ldots$,

$$B_n \subset B_{n-1} \setminus (A_1 \cup \ldots \cup A_n) \subset A \setminus (A_1 \cup \ldots \cup A_n) = \bigcup_{i=n+1}^{\infty} A_i\,.$$

Besides $B_n \neq \emptyset$ for $n = 1, 2, \ldots$, for else (2.3) would be contradicted.

The Cantor intersection property of $\mathcal{S}$ implies

$$\emptyset \neq \bigcap_{n=1}^{\infty} B_n \subset \bigcap_{n=1}^{\infty} \bigcup_{k=n+1}^{\infty} A_k \,,$$

which is a contradiction because $A_1, A_2, \ldots$ are disjoint.

Exercise 2.3. *Let Ω and $\mathcal{S}$ be as in Example 2.3. Show that $\mathcal{S}$ is a semi-field having the Cantor intersection property.*

Exercise 2.4. *Let Ω and $\mathcal{S}$ be as in Example 2.3. Define $P : \mathcal{S} \to [0, 1]$ by*

$$P(E) = 0 \text{ if } E = \emptyset \,,$$

and by (2.1) *otherwise. Show that P is finitely additive on $\mathcal{S}$, that is, if $A_1, \ldots, A_n \in \mathcal{S}$ are disjoint such that $A_1 \cup \ldots \cup A_n \in \mathcal{S}$, then*

$$P\left(\bigcup_{i=1}^{n} A_i\right) = P(A_1) + \ldots + P(A_n) \,.$$

Hint.: Prove this by induction on n. This is a tautology for $n = 1$ and easy to prove for $n = 2$. Assuming it for n, prove it for $n + 1$ by proceeding along the following lines. Suppose $A_1, \ldots, A_{n+1} \in \mathcal{S}$ are disjoint (and non-empty WLOG) and their union is in $\mathcal{S}$. Thus,

$$A_i = E_{\omega_1^i \omega_2^i \ldots \omega_{k_i}^i} \text{ for some } k_i \geq 1, \text{ and } \omega^i, \ldots, \omega_{k_i}^i \in \{H, T\}\,, i = 1, \ldots, n+1 \,.$$

Assume WLOG that $k_1 = k_1 \vee \ldots \vee k_{n+1}$. Let u be the opposite of $\omega_{k_1}^1$, that is $u = H$ if $\omega_{k_1}^1 = T$ and vice versa. Argue that for some i,

$$\left(\omega_1^1, \ldots, \omega_{k_1-1}^1, u, u, u, \ldots\right) \in A_i \,.$$

For this i, show that $k_i = k_1$ and

$$\left(\omega_1^i, \ldots, \omega_{k_i}^i\right) = \left(\omega_1^1, \ldots, \omega_{k_1-1}^1, u\right) \,.$$

Thus $A_1 \cup A_i = E_{\omega_1^1 \ldots \omega_{k_1-1}^1}$, which allows the induction hypothesis to be used.

Exercise 2.5. *Let Ω and $\mathcal{S}$ be as in Example 2.3. Use Exc 2.2 and 2.3 to show if $A_1, A_2, \ldots \in \mathcal{S}$ are disjoint such that*

$$\bigcup_{n=1}^{\infty} A_n \in \mathcal{S} \,,$$

then all but finitely many of A_n's are empty. Hence argue that P defined in Exc 2.4 is countably additive on $\mathcal{S}$.

Exercise 2.6. *Let Ω and $\mathcal{S}$ be as in Example 2.3. Use the corollary of Theorems 1.1 and 1.3 to show that there exists a unique probability measure P on $(\Omega, \sigma(\mathcal{S}))$ satisfying* (2.1).

Now that the natural association of a probability space with a random experiment is understood, we shall define a random variable and its C.D.F.

Definition 16. *Given a probability space $(\Omega, \mathcal{A}, P)$,* random variable *X defined on it is a measurable function $X : \Omega \to \bar{\mathbb{R}}$ such that*

$$P\left(X^{-1}\{-\infty, \infty\}\right) = 0\,. \tag{2.4}$$

A measurable function X for which (2.4) *fails is an* improper random variable. *Given a random variable X, its* cumulative distribution function *(C.D.F.) is a function $F : \mathbb{R} \to [0,1]$ defined by*

$$F(x) = P\left(X^{-1}(-\infty, x]\right)\,, x \in \mathbb{R}\,.$$

It should be noted that an improper random variable is not a random variable. The following theorem gives necessary and sufficient conditions for a function to be a C.D.F.

Theorem 2.1. *If F is the C.D.F. of a random variable X, then*

1. *F is non-decreasing,*
2. *F is right continuous,*
3. *$F(-\infty) = 0$,*
4. *and $F(\infty) = 1$.*

Conversely, if $F : \mathbb{R} \to [0,1]$ is a function satisfying 1.–4. above, then there exists a random variable X defined on some probability space whose C.D.F. is F.

Above and elsewhere, the convention adopted is

$$F(\infty) = \lim_{x\to\infty} F(x) \text{ whenever it exists}\,,$$

and

$$F(-\infty) = \lim_{x\to-\infty} F(x) \text{ whenever it exists}\,.$$

Proof of Theorem 2.1. The proof of 1.–4. is easy and is left as an exercise when F is a C.D.F. Conversely, for a $F : \mathbb{R} \to [0,1]$ satisfying 1.–4., Theorem 1.4 guarantees the existence of a unique probability measure P on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ satisfying

$$P\left((a,b]\right) = F(b) - F(a)\,, -\infty < a < b < \infty\,.$$

Letting $\Omega = \mathbb{R}$, $\mathcal{A} = \mathcal{B}(\mathbb{R})$ and $X : \mathbb{R} \to \mathbb{R}$ to be the identity function, it is easy to see that F is the C.D.F. of X which is a random variable on $(\Omega, \mathcal{A}, P)$. This completes the proof. $\square$

Henceforth, $(\Omega, \mathcal{A}, P)$ will be the probability space underlying any random variable talked about, unless explicitly mentioned otherwise. Theorem 2.1 guarantees that such a probability space exists whenever a few conditions are satisfied.

Definition 17. *Given a possibly improper random variable X, its* distribution*, usually denoted by $P(X \in \cdot)$ or $P \circ X^{-1}(\cdot)$, is the push forward measure of P on $(\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}))$ under X, that is,*

$$P(X \in B) = P \circ X^{-1}(B) = P(\{\omega \in \Omega : X(\omega) \in B\}) \, , B \in \mathcal{B}(\bar{\mathbb{R}}) \, .$$

For a Borel function $f : \bar{\mathbb{R}} \to \bar{\mathbb{R}}$, we denote by

$$\int_{\bar{\mathbb{R}}} f(x) P(X \in dx) \text{ or } \int_{\bar{\mathbb{R}}} f(x) P \circ X^{-1}(dx) \, ,$$

the integral of f with respect to the distribution of X whenever it is defined.

Exercise 2.7. *For a random variable X, show that its distribution is the unique measure μ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ satisfying*

$$\mu\left((a, b] \cap \mathbb{R}\right) = F(b) - F(a) \, , -\infty \le a \le b \le \infty \, ,$$

were F is the C.D.F. of X.

Definition 18. *For a possibly improper random variable X, its* expectation *is defined as*

$$\mathrm{E}(X) = \int_{\Omega} X(\omega) P(d\omega) \, ,$$

*whenever the integral on the right hand side makes sense. The word "*mean*" is an often used synonym of "expectation".*

Note that $\mathrm{E}(X)$ is defined if either of $\mathrm{E}(X^+)$ and $\mathrm{E}(X^-)$ is finite whereas $\mathrm{E}(X)$ is finite when both are finite which happens if and only if $\mathrm{E}(|X|) < \infty$.

The following is a formula relating expectation with C.D.F.

Theorem 2.2. *For a possibly improper random variable X whose expectation is defined*

$$\mathrm{E}(X) = \int_0^\infty P(X > x) \, dx - \int_{-\infty}^0 P(X \le x) \, dx \, . \tag{2.5}$$

The following exercise is needed for the proof.

Exercise 2.8. *For a possibly improper random variable $X \ge 0$, show that*

$$\{(\omega, x) \in \Omega \times \mathbb{R} : 0 \le X(\omega) < x\} \in \mathcal{A} \otimes \mathcal{B}(\mathbb{R}) \, .$$

Hint.: First show this for a simple function X.

Proof of Theorem 2.2. We first show this when $X \geq 0$. The definition of expectation implies

$$\begin{aligned}
\mathrm{E}(X) &= \int_\Omega X(\omega) P(d\omega) \\
&= \int_\Omega \left(\int_0^\infty \mathbf{1}_{[0 \leq x < X(\omega)]} dx \right) P(d\omega) \\
&= \int_{\{(\omega,x)\in\Omega\times\mathbb{R}: 0 \leq X(\omega) < x\}} P(d\omega) \otimes dx \,,
\end{aligned}$$

the last line following from Tonelli and Exc 2.8. Use Tonelli again to write

$$\begin{aligned}
\mathrm{E}(X) &= \int_0^\infty \int_\Omega \mathbf{1}_{[0 \leq x < X(\omega)]} P(d\omega)\, dx \\
&= \int_0^\infty P(X > x) \,.
\end{aligned}$$

For X which is not necessarily non-negative,

$$\mathrm{E}\left(X^+\right) = \int_0^\infty P(X^+ > x)\, dx = \int_0^\infty P(X > x)\, dx \,,$$

the second equality follows from the observation that for $x \geq 0$, $X > x \iff X^+ > x$. Similarly,

$$\begin{aligned}
\mathrm{E}\left(X^-\right) &= \int_0^\infty P(X^- > x)\, dx \\
&= \int_{-\infty}^0 P(X^- > -x)\, dx \\
\left(X^- > -x \iff X < x \text{ for } x \leq 0\right) &= \int_{-\infty}^0 P(X < -x)\, dx \\
&= \int_{-\infty}^0 P(X \leq -x)\, dx \,,
\end{aligned}$$

the last line following from the fact that $P(X \leq \cdot)$ and $P(X < \cdot)$ differs on a set which is at most countable and hence has Lebesgue measure zero. Recalling that $X = X^+ - X^-$ and either of $\mathrm{E}(X^+)$ and $\mathrm{E}(X^-)$ is finite, the proof follows. $\square$

The following theorem relates the expectation with its distribution.

Theorem 2.3. *For a possibly improper random variable X and a Borel function $f : \bar{\mathbb{R}} \to \bar{\mathbb{R}}$,*

$$\mathrm{E}(f(X)) = \int_{\bar{\mathbb{R}}} f(x) P(X \in dx) \,,$$

whenever either side makes sense. In particular, if $\mathrm{E}(X)$ is defined then

$$\mathrm{E}(X) = \int_{\bar{\mathbb{R}}} x P(X \in dx) \,.$$

Proof. Follows from Theorem 1.6. $\square$

The proof of the next result follows directly from the definition.

Theorem 2.4 (Linearity of expectation)**.** *If X and Y have finite expectations, then so does $\alpha X + \beta Y$ for $\alpha, \beta \in \mathbb{R}$ and then*

$$\mathrm{E}\left(\alpha X + \beta Y\right) = \alpha \mathrm{E}(X) + \beta \mathrm{E}(Y)\,.$$

Proof. Follows from the inequality $|\alpha X + \beta Y| \le |\alpha||X| + |\beta||Y|$ and the fact that integral on a measure space is monotone and linear. $\square$

If Definition 18 were replaced by any other definition, for example, by (2.5), then proving the above theorem would have become extremely difficult.

Definition 19. *For a random variable X with a finite mean μ, its* variance *is defined as*

$$\mathrm{Var}(X) = \mathrm{E}\left[(X-\mu)^2\right]\,.$$

The standard deviation *of X is $\sqrt{\mathrm{Var}(X)}$. The convention is to declare* $\mathrm{Var}(X) = \infty$ *if* $\mathrm{E}(X^2) = \infty$.

Theorem 2.5. *The variance of X is defined and finite if and only if*

$$\mathrm{E}(X^2) < \infty,$$

in which case,

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - \left(\mathrm{E}(X)\right)^2\,. \tag{2.6}$$

For the proof, the following fact is needed.

Fact 2.1 (Cauchy-Schwarz inequality)**.** *If f, g are measurable functions from a measure space $(\Omega, \mathcal{A}, \mu)$ to $\mathbb{R}$, then*

$$\int |fg|\,d\mu \le \left(\int f^2\,d\mu\right)^{1/2}\left(\int g^2\,d\mu\right)^{1/2}\,.$$

Proof of Theorem 2.5. The Cauchy-Schwarz inequality implies that

$$\mathrm{E}(|XY|) \le \sqrt{\mathrm{E}(X^2)\mathrm{E}(Y^2)}\,,$$

for random variables X and Y. Take Y to be identically 1 and square both sides to get

$$\left(\mathrm{E}(|X|)\right)^2 \le \mathrm{E}(X^2)\,. \tag{2.7}$$

If $\mathrm{E}(X^2) < \infty$, then (2.7) shows X has a finite expectation. Let $\mu = \mathrm{E}(X)$. Write

$$(X-\mu)^2 = X^2 - 2\mu X + \mu^2\,.$$

Since X^2 and X have a finite expectation, the linearity of expectation implies so does the left hand side and

$$\mathrm{E}\left[(X-\mu)^2\right] = \mathrm{E}(X^2) - 2\mu\mathrm{E}(X) + \mu^2 = \mathrm{E}(X^2) - \mu^2\,,$$

which proves the "if" part and (2.6).

Conversely, if $\mathrm{Var}(X)$ is defined and finite, that is, if $\mathrm{E}(X) = \mu$ is finite and so is

$$\mathrm{E}\left[(X-\mu)^2\right]\,,$$

then writing

$$X^2 = (X-\mu)^2 + 2\mu X - \mu^2\,,$$

it follows that $\mathrm{E}(X^2) < \infty$. This shows the "only if" part and thus completes the proof. $\square$

The formula (2.6) is used almost always for calculating variance. A word of caution: (2.7) should not be misinterpreted as

$$\left(\int |f|\,d\mu\right)^2 \le \int f^2\,d\mu\,,$$

when μ is not a probability measure. The above is clearly false, for example, if μ is the Lebesgue measure on $\mathbb{R}$ and $f(x) = x^{-1}\mathbf{1}(x>1)$ because then the right hand side is finite whereas the left hand side is not.

Definition 20. *A random variable X is* discrete *if there exists a countable set $C \subset \mathbb{R}$ such that $P(X \in C) = 1$. The* probability mass function *of a discrete random variable X is the function $f : \mathbb{R} \to [0,1]$ defined by*

$$f(x) = P(X = x)\,, x \in \mathbb{R}\,.$$

Theorem 2.6. *If X is a discrete random variable, then for any measurable $f : \mathbb{R} \to \bar{\mathbb{R}}$,*

$$\mathrm{E}\left(f(X)\right) = \sum_{x\in\mathbb{R}} f(x)P(X=x)\,, \tag{2.8}$$

whenever the left hand side is defined, where the sum on the right hand side is to be interpreted as the sum over those x for which $P(X = x) > 0$. In particular, if X has an expectation, then

$$\mathrm{E}(X) = \sum_{x\in\mathbb{R}} xP(X=x)\,.$$

Proof. Let $C = \{x \in \mathbb{R} : P(X = x) > 0\}$. Since X is discrete, $P(X \in C) = 1$. Let μ be the counting measure on C. The fact that for $A, E \in \mathcal{A}$ with

$P(E) = 1$, $P(A \cap E) = P(A)$, implies for $B \in \mathcal{B}(\mathbb{R})$,

$$\begin{aligned} P(X \in B) &= P(X \in B \cap C) \\ &= \sum_{x \in B \cap C} P(X = x) \\ &= \int_B P(X = x)\mu(dx)\,. \end{aligned}$$

In other words,

$$\frac{P(X \in dx)}{\mu(dx)} = P(x = x)\,,$$

that is, $P(X = \cdot)$ is the Radon-Nikodym derivative of $P(X \in \cdot)$ with respect to μ. Exc 1.10 shows that for a measurable $f : \mathbb{R} \to \bar{\mathbb{R}}$,

$$\int_{\mathbb{R}} f(x)P(X \in dx) = \int_{\mathbb{R}} f(x)P(X = x)\mu(dx)\,,$$

whenever either side is defined. The left hand side equals $\mathrm{E}(f(X))$ by Theorem 2.3, and the right hand side is simply $\sum_{x \in C} f(x)P(X = x)$. Since $P(X = x) = 0$ for $x \notin C$, this completes the proof of (2.8). The second claim being a special case of (2.8), the proof follows. $\square$

Example 2.4. *A coin with chances of head p is tossed infinitely often. Proceeding like in Example 2.3, construct the probability space for this experiment. That is, if Ω and $\mathcal{S}$ are as therein, show that $P : \mathcal{S} \to [0, 1]$, defined by*

$$P\left(E_{\omega_1 \ldots \omega_n}\right) = \prod_{i=1}^{n} \left[p\mathbf{1}(\omega_i = H) + q\mathbf{1}(\omega_i = T)\right]\,, n \in \mathbb{N}\,, \omega_1, \ldots, \omega_n \in \{H, T\}\,,$$

where $q = 1 - p$ and $P(\emptyset) = 0$, is countably additive. Use Theorem 1.1 to complete the construction.

Let X be the number of heads obtained in the first n tosses. Show that the PMF of X is

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

The convention followed here and elsewhere is that left hand side is to be interpreted as zero for those k for which it has not been defined, that is, for $k \in \mathbb{R} \setminus \{0, 1, \ldots, n\}$ in this case. The distribution of X is called Binomial(n, p). Check that its mean and variance are np and npq, respectively. If $n = 1$, then this has another name, which is, Bernoulli(p). In other words, a Bernoulli(p) random variable takes the values 0 and 1 with probabilities q and p, respectively.

Let Y be the number of tosses needed to get the first head. The PMF of Y is

$$P(Y = n) = q^{n-1}p\,, n \in \mathbb{N}\,.$$

The distribution of Y is called Geometric(p). Check that its mean and variance are p^{-1} and $p^{-2}q$, respectively.

For a fixed $k = 1, 2, 3, \ldots$, let Z be the number of tosses needed to get the k-th head. The PMF of Z is

$$P(Z = n) = \binom{n-1}{k-1} q^{n-k}p^k\,, n \in \{k, k+1, k+2, \ldots\}\,.$$

Since the right hand side is the $(n-k+1)$-th term in the expansion of $p^k(1-q)^{-k}$, the distribution of Z is called Negative Binomial(k, p). Check that its mean and variance are kp^{-1} and $kp^{-2}q$, respectively.

Exercise 2.9. *Suppose X_n is a Bin(n, p_n) random variable defined on a probability space $(\Omega_n, \mathcal{A}_n, P_n)$. If*

$$\lim_{n\to\infty} np_n = \lambda \in (0, \infty)\,,$$

show that for $k = 0, 1, 2, \ldots$,

$$\lim_{n\to\infty} P_n(X_n = k) = e^{-\lambda}\frac{\lambda^k}{k!}\,.$$

Exercise 2.10. *Show using Theorem 2.1 that there exists a discrete random variable X defined on some probability space with*

$$P(X = n) = e^{-\lambda}\frac{\lambda^n}{n!}\,, n \in \{0, 1, 2, \ldots\}\,.$$

The distribution of X is called Poisson(λ). Show that its mean and variance both equal λ.

Definition 21. *A random variable X is* continuous *if*

$$P(X = x) = 0 \text{ for all } x \in \mathbb{R}\,.$$

Exercise 2.11. *If X is a random variable with C.D.F. F, show that for all $x \in \mathbb{R}$,*

$$P(X = x) = F(x) - F(x-)\,.$$

Hence argue that X is a continuous random variable if and only if F is a continuous function.

Definition 22. *A Borel function $f : \mathbb{R} \to [0, \infty)$ is the* density *of a random variable X if*

$$P(X \in B) = \int_B f(x)\,dx\,, \text{ for all } B \in \mathcal{B}(\mathbb{R})\,.$$

Exercise 2.12. *1. Show that f is the density of X is equivalent to*

$$\frac{P(X \in dx)}{dx} = f(x)\,, x \in \mathbb{R}\,,$$

that is, f is the Radon-Nikodym derivative of $P(X \in \cdot)$ with respect to the Lebesgue measure.
2. Prove that if f and g are densities of X, then $f = g$ a.e. In other words, a density is unique upto a set of Lebesgue measure zero.
3. Prove that a non-negative Borel function f on $\mathbb{R}$ is a density of X having C.D.F. F if and only if

$$\int_{-\infty}^{x} f(t)\,dt = F(x)\,, x \in \mathbb{R}\,.$$

4. Show that a random variable is continuous if it has a density.

Definition 23. *A function $F : \mathbb{R} \to \mathbb{R}$ is* absolutely continuous *if given $\varepsilon > 0$ there exists $\delta > 0$ such that*

$$\sum_{i=1}^{n} |F(y_i) - F(x_i)| \le \varepsilon\,,$$

whenever $x_1 \le y_1 \le x_2 \le y_2 \le \ldots \le x_n \le y_n$ are such that

$$\sum_{i=1}^{n} (y_i - x_i) \le \delta\,.$$

Theorem 2.7. *A random variable has a density if and only if its C.D.F. is absolutely continuous.*

The proof uses the following exercise.

Exercise 2.13. *If h is an integrable function on a measure space $(\Omega, \mathcal{A}, \mu)$, then given $\varepsilon > 0$ there exists $\delta > 0$ such that*

$$\int_A |h|\,d\mu \le \varepsilon\,,$$

for all $A \in \mathcal{A}$ with $\mu(A) \le \delta$.

Proof of Theorem 2.7. We start with proving the "if" part. Let X be a random variable with an absolutely continuous C.D.F. F. In view of Theorem 1.7, which is the Radon-Nikodym theorem, it suffices to show

$$P(X \in \cdot) \ll \lambda\,, \tag{2.9}$$

λ being the Lebesgue measure. To prove this, fix any $B \in \mathcal{B}(\mathbb{R})$ with $\lambda(B) = 0$. We shall prove $P(X \in B) = 0$ by showing for any $\varepsilon > 0$

$$P(X \in B) \le \varepsilon\,. \tag{2.10}$$

Fix $\varepsilon > 0$. Absolute continuity of F implies there exists $\delta > 0$ such that

$$\sum_{i=1}^n |F(y_i) - F(x_i)| \le \varepsilon \,,$$

whenever $x_1 \le y_1 \le x_2 \le y_2 \le \ldots \le x_n \le y_n$ are such that

$$\sum_{i=1}^n (y_i - x_i) \le \delta \,.$$

Since $\lambda(B) = 0$ and the Lebesgue measure is regular, see Exc 1.5, there exists an open set $U \subset \mathbb{R}$ with $U \supset B$ and $\lambda(B) \le \delta$. An open subset of $\mathbb{R}$ is the union of countably many disjoint open intervals, that is,

$$U = \bigcup_{i\ge1} (x_i, y_i) \,,$$

for some $x_1, y_1, x_2, y_2, \ldots$ satisfying $x_i < y_i$ and $(x_i, y_i) \cup (x_j, y_j) = \emptyset$ for all $i \neq j$. Therefore,

$$\sum_{i=1}^n (y_i - x_i) = \lambda\left(\bigcup_{i=1}^n (x_i, y_i)\right) \le \lambda(U) \le \delta \,,$$

showing that

$$\begin{aligned} \varepsilon &\ge \sum_{i=1}^n [F(y_i) - F(x_i)] \\ \text{(absolute continuity implies continuity)} &= \sum_{i=1}^n [F(y_i-) - F(x_i)] \\ &= \sum_{i=1}^n P(x_i < X < y_i) \\ &= P\left(X \in \bigcup_{i=1}^n (x_i, y_i)\right) . \end{aligned}$$

Since

$$P\left(X \in \bigcup_{i=1}^n (x_i, y_i)\right) \uparrow P(X \in U) \ge P(X \in B) \,,$$

(2.10) follows. Arbitrariness of ε shows (2.9) which by an appeal to Theorem 1.7 proves the existence of the density of X which is nothing but the Radon-Nikodym derivative of $P \circ X^{-1}$ with respect to Lebesgue.

Conversely, suppose that X has a density f. That is, $f : \mathbb{R} \to [0, \infty)$ is Borel and satisfies

$$\int_{-\infty}^{\infty} f(x)\, dx = 1 \,.$$

Fix $\varepsilon > 0$. Use Exc 2.13 to choose $\delta > 0$ such that

$$\int_B f(x)\,dx \le \varepsilon\,,$$

for all $B \in \mathcal{B}(\mathbb{R})$ with $\lambda(B) \le \delta$. For $n \ge 1$, fix $x_1 \le y_1 \le \ldots \le x_n \le y_n$ with

$$\sum_{i=1}^n (y_i - x_i) \le \delta\,.$$

Letting $B = (x_1, y_1] \cup \ldots \cup (x_n, y_n]$, the above simply means $\lambda(B) \le \delta$. The choice of δ implies

$$\varepsilon \ge \int_B f(x)\,dx = P(X \in B) = \sum_{i=1}^n \left[F(y_i) - F(x_i)\right]\,,$$

F being the C.D.F. of X. Thus F is absolutely continuous. This proves the "only if" part and thereby completes the proof. □

The following result is most useful in getting the density of a random variable when it exists.

Theorem 2.8. *A random variable X with C.D.F. F has a density if and only if*

$$\int_{-\infty}^{\infty} f(x)\,dx = 1\,,$$

where

$$f(x) = \begin{cases} \frac{d}{dx}F(x)\,, & \text{if } F \text{ is differentiable at } x\,, \\ 0\,, & \text{otherwise}\,. \end{cases} \tag{2.11}$$

In that case, f is the density of X.

The proof uses the following facts.

Fact 2.2 (Theorem 31.2, pg 404, Billingsley (1995))**.** *A non-decreasing function $F : [a, b] \to \mathbb{R}$ is differentiable a.e. on (a, b). If f is as in (2.11), then f is Borel measurable, non-negative and satisfies*

$$\int_a^b f(x)\,dx \le F(b) - F(a)\,.$$

Fact 2.3 (Theorem 31.3, pg 406, Billingsley (1995))**.** *If $f : [a, b] \to [0, \infty)$ is integrable and*

$$F(x) = \int_a^x f(t)\,dt\,, x \in [a, b]\,,$$

then F is differentiable a.e. on (a, b) and

$$\frac{d}{dx}F(x) = f(x) \text{ for almost all } x \in (a, b)\,.$$

Proof of Theorem 2.8. We start with proving the "if" part. Assume

$$\int_{-\infty}^{\infty} f(x)\,dx = 1\,, \tag{2.12}$$

f being as in (2.11). Fact 2.2 implies

$$\int_a^b f(x)\,dx \le F(b) - F(a)\,, -\infty < a < b < \infty\,.$$

Keeping b fixed and letting $a \to -\infty$, MCT shows that the left hand side goes to the corresponding integral from $-\infty$ to b. Since F is a C.D.F., $F(-\infty) = 0$, showing that

$$\int_{-\infty}^{b} f(x)\,dx \le F(b)\,. \tag{2.13}$$

A similar argument shows that for a fixed,

$$\int_a^{\infty} f(x)\,dx \le 1 - F(a)\,.$$

Thus

$$\begin{aligned} F(a) &\le 1 - \int_a^{\infty} f(x)\,dx \\ \big(\text{(using (2.12))}\big) &= \int_{-\infty}^{a} f(x)\,dx \\ &\le F(a)\,, \end{aligned}$$

(2.13) implying the last line by putting $b = a$. Thus

$$\int_{-\infty}^{a} f(x)\,dx = F(a) \text{ for all } a \in \mathbb{R}\,.$$

Exc 2.12.3 shows f is the density of X. This proves the "if" part.

For the "only if" part, assume X has a density g. That is,

$$\int_{-\infty}^{x} g(t)\,dt = P(X \le x) = F(x)\,, x \in \mathbb{R}\,.$$

Thus, for $-\infty < a < b < \infty$,

$$F(x) - F(a) = \int_a^x g(t)\,dt\,, x \in [a,b]\,.$$

Fact 2.3 shows that the left hand side is differentiable a.e. on (a,b) and

$$\frac{d}{dx}F(x) = g(x) \text{ for a.e. } x \in (a,b)\,.$$

Since this is true for all a, b with $-\infty < a < b < \infty$, the above equality holds a.e. on $\mathbb{R}$. A comparison with (2.11) shows $f = g$ a.e. Therefore

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{\infty} g(x)\,dx = 1\,,$$

proving the "only if" part. This completes the proof. □

Theorems 2.7 and 2.8 and their proofs essentially show the following result.

Theorem 2.9. *If X is a random variable with C.D.F. F, then the following are equivalent.*

1. *A density of X exists.*
2. *The function F is absolutely continuous.*
3. *The distribution of X is absolutely continuous with respect to Lebesgue, that is,*
$$P(X \in \cdot) \ll \lambda(\cdot)\,.$$
4. *Given $\varepsilon > 0$ there exists $\delta > 0$ such that*
$$P(X \in B) \le \varepsilon \text{ whenever } B \in \mathcal{B}(\mathbb{R}) \text{ and } \lambda(B) \le \delta\,.$$
5. *If f is the derivative of F wherever it exists and zero elsewhere, then*
$$\int_{-\infty}^{\infty} f(x)\,dx = 1\,.$$

If any of these hold, then f defined above is the density of X.

Proof. Exc. □

The equivalence of 2. and 5. in the above theorem is important from the point of view of analysis as well.

Theorem 2.10. *If X has density f, then for any measurable $g : \mathbb{R} \to \bar{\mathbb{R}}$,*

$$\mathrm{E}\,(g(X)) = \int_{-\infty}^{\infty} g(x) f(x)\,dx\,,$$

whenever either side makes sense. In particular,

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} x f(x)\,dx\,,$$

if either side is defined.

Proof. Similar to the proof of Theorem 2.6 once it is observed that

$$f(x) = \frac{P(X \in dx)}{dx}\,.$$

□

Example 2.5. *The distribution of X is Uniform(a,b) for $-\infty < a < b < \infty$ if its C.D.F. is*

$$F(x) = \begin{cases} 0\,, & x < a\,, \\ \frac{x-a}{b-a}\,, & a \le x \le b\,, \\ 1\,, & x > b\,. \end{cases}$$

Check that the density of X is

$$f(x) = \frac{1}{b-a}\,, a \le x \le b\,,$$

with the usual convention of interpreting it as zero wherever it is not defined. Show that the mean and variance of Uniform(a,b) are $(a+b)/2$ and $(b-a)^2/12$, respectively.

Example 2.6. *A random variable X follows standard normal or standard Gaussian if its density is*

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,, x \in \mathbb{R}\,.$$

It is easy to see that $\mathrm{E}(X) = 0$ *and*

$$\begin{aligned} \mathrm{E}(X^2) &= \int_{-\infty}^{\infty} x^2 f(x)\,dx \\ &= \sqrt{\frac{2}{\pi}} \int_0^{\infty} x^2 e^{-x^2/2}\,dx \\ &= \sqrt{\frac{2}{\pi}} \int_0^{\infty} x \left(x e^{-x^2/2}\right)\,dx\,. \end{aligned}$$

Integrating by parts with the help of the observation that

$$\frac{d}{dx}\left(-e^{-x^2/2}\right) = x e^{-x^2/2}\,,$$

we get

$$\begin{aligned} \mathrm{E}(X^2) &= \sqrt{\frac{2}{\pi}} \left[x \left(-e^{-x^2/2}\right)\Big|_0^{\infty} - \int_0^{\infty} \left(-e^{-x^2/2}\right)\,dx \right] \\ &= \sqrt{\frac{2}{\pi}} \int_0^{\infty} e^{-x^2/2}\,dx \\ &= 1\,. \end{aligned}$$

Thus the mean and variance of the standard normal distribution are 0 *and* 1*, respectively.*

The distribution with density

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right), x \in \mathbb{R},$$

is called Normal(μ, σ^2) *for* $\mu \in \mathbb{R}$ *and* $\sigma > 0$*. Show that* f *defined as above is indeed a density, and that the mean and variance of Normal*(μ, σ^2) *are* μ *and* σ^2*, respectively.*

Definition 24. *Let* $(\Omega, \mathcal{A}, P)$ *be a probability space and* $A, B \in \mathcal{A}$ *with* $P(A) > 0$*. The* conditional probability *of* B *given* A *is defined as*

$$P(B|A) = \frac{P(B \cap A)}{P(A)}.$$

Example 2.7. *Suppose we want a non-negative random variable* X *which has the "memoryless" property, that is,* $P(X > t) > 0$ *for all* $t \geq 0$ *and*

$$P(X > s + t | X > t) = P(X > s), s, t \geq 0.$$

This is the same as

$$P(X > s + t) = P(X > s)P(X > t), s, t \geq 0,$$

that is,

$$1 - F(s + t) = \big(1 - F(s)\big)\big(1 - F(t)\big), s, t \geq 0.$$

Letting $G = \log(1 - F)$*, the condition is*

$$G(s + t) = G(s) + G(t), s, t \geq 0.$$

It follows that

$$G(r) = -\lambda r, r \in [0, \infty) \cap \mathbb{Q},$$

where $\lambda = -G(1) = -\log\big(1 - F(1)\big) \geq 0$*. Right continuity of* G *implies*

$$G(x) = \lambda x, x \geq 0,$$

that is,

$$F(x) = 1 - e^{-\lambda x}, x \geq 0.$$

Since $F(\infty) = 1$*, it is necessary that* $\lambda > 0$*. As* $X \geq 0$*,* $F(x) = 0$ *for* $x < 0$*.*

For $\lambda > 0$*,* X *follows Exponential*(λ) *if* F *defined by*

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

is its C.D.F. Show that the density of Exponential(λ) *is*

$$f(x) = \lambda e^{-\lambda x}, x > 0,$$

and its mean and variance are λ^{-1} *and* λ^{-2}*, respectively.*

Definition 25. *For $\alpha > 0$, define*

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}\, dx\,.$$

Exercise 2.14. *Show that*

1. $\Gamma(\alpha) < \infty$ *for all* $\alpha > 0$,
2. $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)\,, \alpha > 0\,,$
3. *and*
$$\Gamma(n) = (n-1)!\,, n \in \mathbb{N}\,.$$

Example 2.8. *For $\alpha > 0$, Gamma(α) is the distribution with density*

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

Check that its mean and variance both equal α.

Definition 26. *For $\alpha, \beta > 0$, define*

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,. \tag{2.14}$$

Exercise 2.15. 1. *Show that the RHS of (2.14) is finite for $\alpha, \beta > 0$.*

2. *Show that*
$$B\left(\frac{1}{2}, \frac{1}{2}\right) = \pi\,,$$
by substituting $\theta = \sin^{-1}\sqrt{x}$ in the RHS of (2.14).

Example 2.9. *For $\alpha, \beta > 0$, X follows Beta(α, β) if its density is*

$$f(x) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}\,, 0 < x < 1\,.$$

When $\alpha = \beta = 1/2$, show that X has C.D.F.

$$F(x) = \begin{cases} 0\,, & x < 0\,, \\ \frac{2}{\pi}\sin^{-1}\sqrt{x}\,, & 0 \le x \le 1\,, \\ 1\,, & x > 1\,. \end{cases}$$

This is why the Beta$(1/2, 1/2)$ distribution is also called the "arc-sine law".

3 Independence

In this chapter the concept of independence is studied, which is a fundamental concept in probability theory. Unless mentioned otherwise, $(\Omega, \mathcal{A}, P)$ is the probability space underlying everything we talk about. For example, all random variables are defined on this space and any collection of sets we talk about is a subset of $\mathcal{A}$, unless the contrary is explicitly stated.

Definition 27. *A collection of σ-fields $\mathcal{A}_1, \ldots, \mathcal{A}_n$ are* independent *if*

$$P(A_1 \cap \ldots \cap A_n) = \prod_{i=1}^{n} P(A_i)\,,$$

for all $A_1 \in \mathcal{A}_1, \ldots, A_n \in \mathcal{A}_n$.

The following is an important observation which connects the usually given definition of independence of events with the above.

Exercise 3.1. *If $\mathcal{A}_1, \ldots, \mathcal{A}_n$ are independent σ-fields and $A_1, \ldots, A_n$ belong to $\mathcal{A}_1, \ldots, \mathcal{A}_n$, respectively, show that*

$$P(A_{i_1} \cap \ldots \cap A_{i_k}) = P(A_{i_1}) \ldots P(A_{i_k})\,,$$

for all $1 \le i_1 < \ldots < i_k \le n$.

The above when $n = 2$ should be compared with Definition 24.

Theorem 3.1. *If $\mathcal{S}_1, \ldots, \mathcal{S}_n$ are semi-fields such that*

$$P\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} P(A_i)\,,\ \textit{for all } A_1 \in \mathcal{S}_1, \ldots, A_n \in \mathcal{S}_n\,,$$

then $\sigma(\mathcal{S}_1), \ldots, \sigma(\mathcal{S}_n)$ are independent.

Proof. The first step is to show

$$P\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} P(A_i)\,,\ \text{for all } A_1 \in \sigma(\mathcal{S}_1), A_2 \in \mathcal{S}_2, \ldots, A_n \in \mathcal{S}_n\,. \quad (3.1)$$

To that end, fix $A_2 \in \mathcal{S}_1, \ldots, A_n \in \mathcal{S}_n$ and define $\mu_1, \mu_2 : \mathcal{A} \to [0, \infty)$ by

$$\mu_1(A) = P(A) \prod_{i=2}^{n} P(A_i)\,,$$
$$\mu_2(A) = P(A \cap A_2 \cap \ldots \cap A_n)\,,$$

for all $A \in \mathcal{A}$. Thus μ_1 and μ_2 are finite measures on $(\Omega, \mathcal{A})$, which agree on $\mathcal{S}_1$ by the hypothesis of the theorem. As $\mathcal{S}_1$ is a semi-field, Corollary 1.1 implies that μ_1 and μ_2 agree on $\sigma(\mathcal{S}_1)$. In other words, (3.1) holds.

We shall now show inductively that for $i = 1, \ldots, n$,

$$P\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} P(A_i)\,, \tag{3.2}$$

for all $A_1 \in \sigma(\mathcal{S}_1), \ldots, A_i \in \sigma(\mathcal{S}_i), A_{i+1} \in \mathcal{S}_{i+1}, \ldots, A_n \in \mathcal{S}_n$; (3.1) shows this holds for $i = 1$. As the induction hypothesis, assume (3.2) for some $i \in \{1, \ldots, n-1\}$. Fix $A_1, \ldots, A_i, A_{i+2}, \ldots, A_n$ in $\sigma(\mathcal{S}_1), \ldots, \sigma(\mathcal{S}_i), \mathcal{S}_{i+2}, \ldots, \mathcal{S}_n$, respectively. As before, define finite measures ν_1, ν_2 on $(\Omega, \mathcal{A})$ by

$$\nu_1(A) = P(A) \prod_{1 \le j \le n,\, j \ne i+1} P(A_j)\,,$$

$$\nu_2(A) = P\left(A \cap \bigcap_{1 \le j \le n,\, j \ne i+1} A_j\right)\,,$$

for all $A \in \mathcal{A}$. The induction hypothesis implies that ν_1 and ν_2 agree on $\mathcal{S}_{i+1}$ and hence they do so on $\sigma(\mathcal{S}_{i+1})$ by Corollary 1.1. As this holds for all $A_1, \ldots, A_i, A_{i+2}, \ldots, A_n$ in $\sigma(\mathcal{S}_1), \ldots, \sigma(\mathcal{S}_i), \mathcal{S}_{i+2}, \ldots, \mathcal{S}_n$, respectively, (3.2) follows for $i + 1$. Mathematical induction shows (3.2) for $i = n$, which completes the proof. $\square$

Definition 28. *Random variables $X_1, \ldots, X_n$, all of which by convention are defined on $(\Omega, \mathcal{A}, P)$, are* independent *if $\sigma(X_1), \ldots, \sigma(X_n)$ are independent σ-fields, where $\sigma(X)$ is the smallest σ-field with respect to which X is measurable for any $X : \Omega \to \bar{\mathbb{R}}$, that is,*

$$\sigma(X) = \left\{X^{-1}B : B \in \mathcal{B}(\bar{\mathbb{R}})\right\}\,.$$

Exercise 3.2. *Show that discrete random variables $X_1, \ldots, X_n$ are independent if and only if*

$$P(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} P(X_i = x_i)\,,\ \textit{for all } x_1, \ldots, x_n \in \mathbb{R}\,.$$

Theorem 3.2. *Random variables $X_1, \ldots, X_n$ are independent if and only if*

$$P(X_1 \le x_1, \ldots, X_n \le x_n) = \prod_{i=1}^{n} P(X_i \le x_i)\,, x_1, \ldots, x_n \in \mathbb{R}\,. \tag{3.3}$$

The proof uses the following exercise.

Exercise 3.3. *For random variables $X_1, \ldots, X_n$ and $-\infty \le a_i \le b_i \le \infty$ for $i = 1, \ldots, n$,*

$$\begin{aligned} &P\left(a_i < X_i \le b_i\,, i = 1, \ldots, n\right) \\ &= \sum_{(x_1,\ldots,x_n) \in \{a_1,b_1\} \times \ldots \times \{a_n,b_n\}} (-1)^{\#\{1 \le i \le n : x_i = a_i\}} P(X_1 \le x_1, \ldots, X_n \le x_n)\,. \end{aligned}$$

Soln.: As usual, denote by $\mathbf{1}_A$ or $\mathbf{1}(A)$ the indicator of A, that is, it is one or zero depending on whether A occurs or not, respectively. Write

$$P\left(a_i < X_i \le b_i\,, i = 1, \ldots, n\right) = \mathrm{E}\left(\mathbf{1}\left(a_i < X_i \le b_i\,, i = 1, \ldots, n\right)\right) . \quad (3.4)$$

Observe that

$$\begin{aligned}
&\mathbf{1}\left(a_i < X_i \le b_i\,, i = 1, \ldots, n\right) \\
&= \prod_{i=1}^n \mathbf{1}(a_i < X_i \le b_i) \\
&= \prod_{i=1}^n \left[\mathbf{1}(X_i \le b_i) - \mathbf{1}(X_i \le a_i)\right] \\
&= \prod_{i=1}^n \sum_{x_i \in \{a_i, b_i\}} (-1)^{\mathbf{1}(x_i = a_i)} \mathbf{1}(X_i \le x_i) \\
&= \sum_{x_1 \in \{a_1, b_1\}} \cdots \sum_{x_n \in \{a_n, b_n\}} \prod_{i=1}^n (-1)^{\mathbf{1}(x_i = a_i)} \mathbf{1}(X_i \le x_i) \\
&= \sum_{(x_1, \ldots, x_n) \in \{a_1, b_1\} \times \ldots \times \{a_n, b_n\}} (-1)^{\#\{1 \le i \le n : x_i = a_i\}} \mathbf{1}(X_1 \le x_1, \ldots, X_n \le x_n) .
\end{aligned}$$

Taking expectation on both sides and using (3.4), the solution follows.

Proof of Theorem 3.2. The "only if" part is trivial because (3.3) follows from the observation that $X_1^{-1}(-\infty, x_1], \ldots, X_n^{-1}(-\infty, x_n]$ belong to $\sigma(X_1)$, $\ldots, \sigma(X_n)$, respectively. For the "if" part, assume (3.3). The first observation is that (3.3) holds for $x_1, \ldots, x_n \in \bar{\mathbb{R}}$ because if $x_i = -\infty$ for one or more i, then both sides are zero and if $x_i \to \infty$, then both sides of (3.3) increase to the respective quantities obtained by putting $x_i = \infty$ for those i's. Thus, (3.3) can be assumed to hold for all $x_1, \ldots, x_n \in \bar{\mathbb{R}}$ without loss of generality.

Let $\mathcal{S}_i = \{X_i^{-1}(a_i, b_i] : -\infty \le a_i \le b_i \le \infty\}$ for $i = 1, \ldots, n$. For

$A_1 \in \mathcal{S}_1, \ldots, A_n \in \mathcal{S}_n$, that is, $A_i = X_i^{-1}(a_i, b_i]$ for some $\infty \le a_i \le b_i \le \infty$,

$$
\begin{aligned}
&P(A_1 \cap \ldots \cap A_n) \\
&= P(a_1 < X_1 \le b_1, \ldots, a_n < X_n \le b_n) \\
&= \sum_{(x_1,\ldots,x_n)\in\{a_1,b_1\}\times\ldots\times\{a_n,b_n\}} (-1)^{\#\{1\le i\le n: x_i=a_i\}} P(X_1 \le x_1, \ldots, X_n \le x_n) \\
&= \sum_{(x_1,\ldots,x_n)\in\{a_1,b_1\}\times\ldots\times\{a_n,b_n\}} (-1)^{\#\{1\le i\le n: x_i=a_i\}} \prod_{i=1}^n P(X_i \le x_i) \\
&= \sum_{(x_1,\ldots,x_n)\in\{a_1,b_1\}\times\ldots\times\{a_n,b_n\}} \prod_{i=1}^n (-1)^{\mathbf{1}(x_i=a_i)} P(X_i \le x_i) \\
&= \prod_{i=1}^n \sum_{x_i\in\{a_i,b_i\}} (-1)^{\mathbf{1}(x_i=a_i)} P(X_i \le x_i) \\
&= \prod_{i=1}^n \left(P(X_i \le b_i) - P(X_i \le a_i)\right) \\
&= \prod_{i=1}^n P(a_i < X_i \le b_i) \\
&= P(A_1)\ldots P(A_n)\,,
\end{aligned}
$$

Exc 3.3 implying the third line and the fourth line following from (3.3) which holds for all $x_1, \ldots, x_n \in \bar{\mathbb{R}}$. Theorem 3.1 shows $\sigma(\mathcal{S}_1), \ldots, \sigma(\mathcal{S}_n)$ are independent, which is the same as independence of $X_1, \ldots, X_n$. $\square$

Exercise 3.4. *Let $\Omega = (0,1]$, $\mathcal{A}$ be the collection of Borel subsets of $(0,1]$ and P be the restriction of Lebesgue measure to $(0,1]$.*

1. *For all $\omega \in \Omega$, show that there exist unique $X_1(\omega), X_2(\omega), \ldots \in \{0,1,2\}$ such that*
$$\omega = \sum_{n=1}^\infty 3^{-n} X_n(\omega)\,,$$
and $\{n : X_n(\omega)$ equals either 1 or $2\}$ is an infinite set. In other words, the ternary expansion of ω is being considered and in case where multiple expansions are possible, the non-terminating one, that is, the one which has infinitely many 2's is being taken.

2. *For $n \ge 1$ and $i_1, \ldots, i_n \in \{0,1,2\}$, show that for $\omega \in \Omega$,*
$$X_1(\omega) = i_1, \ldots, X_n(\omega) = i_n \iff \sum_{j=1}^n 3^{-j} i_j < \omega \le 3^{-n} + \sum_{j=1}^n 3^{-j} i_j\,.$$
Hence prove that $X_1, \ldots, X_n$ are independent and each takes values $0, 1, 2$ with probability $1/3$ for each.

3. *Prove that* $\inf\{n \geq 1 : X_n = 2\}$ *is a proper random variable, that is, it is finite almost surely ("almost surely" or "a.s." simply means "with probability 1"). In fact, observe that it is a Geometric*$(1/3)$ *random variable.*

4. *Show that the set of* $\omega \in \Omega$ *which have multiple ternary expansions is countable.*

5. *Use the above two claims to argue that*

$$C = \Big\{\omega \in (0,1] : \omega \text{ has a unique ternary expansion and} \qquad (3.5)$$
$$\omega = \sum_{n=1}^{\infty} 3^{-n} x_n \text{ for some } x_1, x_2, \ldots \in \{0,1\}\Big\}$$

is a Borel set of zero Lebesgue measure.

Hint *for* (3.5). *Observe that*

$$C = \left(\bigcap_{n=1}^{\infty} \{\omega : X_n(\omega) \neq 2\}\right) \cap \{\omega : \omega \text{ has a unique ternary expansion}\}.$$

Exercise 3.5. *Let* $(\Omega, \mathcal{A}, P)$ *be as in Example 2.3. Define* $X : \Omega \to \mathbb{R}$ *by*

$$X(\omega) = \sum_{n=1}^{\infty} 3^{-n} \mathbf{1}(\omega_n = H)\,, \omega = (\omega_1, \omega_2, \ldots) \in \Omega\,.$$

1. *Show that* X *is a random variable, that is, it is a measurable function.*

2. *Prove that* X *is a one-one function. Hence show that* X *is a continuous random variable.*

3. *If* C *is as in* (3.5), *show that* $P(X \in C) = 1$.

4. *Prove that* X *cannot have a density.*
Hint. *Show that if* X *has a density, then* $P(X \in C)$ *would be zero because Lebesgue measure of* C *is zero.*

5. *Argue using Theorem 2.7 that the C.D.F. of* X *is continuous but not absolutely continuous.*

Definition 29. *A possibly infinite collection* $\{\mathcal{A}_i : i \in I\}$ *is* independent *if for all* $n \geq 2$ *and distinct* $i_1, \ldots, i_n \in I$*, the* σ*-fields* $\mathcal{A}_{i_1}, \ldots, \mathcal{A}_{i_n}$ *are independent.*

For a collection $\{\mathcal{A}_i : i \in I\}$ of σ-fields, denote

$$\bigvee_{i \in I} \mathcal{A}_i = \sigma\left(\bigcup_{i \in I} \mathcal{A}_i\right).$$

Theorem 3.3. *If $\{\mathcal{A}_i : i \in I\}$ is an independent collection of σ-fields and $I_1, I_2, \ldots, I_k$ are disjoint non-empty subsets of I, then*

$$\bigvee_{i \in I_1} \mathcal{A}_i \ , \ldots , \bigvee_{i \in I_1} \mathcal{A}_i$$

are independent.

Proof. The first step of the proof is to show the claim when $I_1, \ldots, I_k$ are finite sets. Suppose

$$I_i = \{n_{i1}, \ldots, n_{ik_i}\} \ , i = 1, \ldots, k \, .$$

Define

$$\mathcal{S}_i = \{A_1 \cap \ldots \cap A_{k_i} : A_j \in \mathcal{A}_{ij} \text{ for } j = 1, \ldots, k_i\} \ , i = 1, \ldots, k \, .$$

Obviously, $\mathcal{S}_i$ is a semi-field because it is closed under finite intersections. Further, if $A_j \in \mathcal{A}_{ij}$ for $j = 1, \ldots, k_i$,

$$(A_1 \cap \ldots \cap A_{k_i})^c = B_1 \cup \ldots \cup B_{k_i} \, ,$$

where

$$B_1 = A_1^c \, , \ B_2 = A_1 \cap A_2^c \, , \ \ldots , \ B_{k_i} = A_1 \cap \ldots \cap A_{k_i - 1} \cap A_{k_i}^c \, .$$

Since $B_1, \ldots, B_{k_i}$ are disjoint $\mathcal{S}_i$-sets, $\mathcal{S}_i$ is a semi-field. Since $\{\mathcal{A}_{ij} : j = 1, \ldots, k_i, \ i = 1, \ldots, k\}$ are independent, it follows that

$$P(A_1 \cap \ldots \cap A_k) = P(A_1) \ldots P(A_k) \text{ for all } A_1 \in \mathcal{S}_1, \ldots, A_k \in \mathcal{S}_k \, .$$

Theorem 3.1 implies independence of $\sigma(\mathcal{S}_1), \ldots, \sigma(S_k)$, which are same as

$$\bigvee_{i \in I_1} \mathcal{A}_i \ , \ldots , \bigvee_{i \in I_1} \mathcal{A}_i \, ,$$

respectively.

Now let $I_1, \ldots, I_k$ be disjoint non-empty subsets of I. Define

$$\mathcal{F}_i = \bigcup_{J \subset I_i, \, J \text{ finite}} \ \bigvee_{j \in J} \mathcal{A}_j \, , i = 1, \ldots, k \, .$$

Clearly, $\mathcal{F}_i$ is a field for $i = 1, \ldots, k$ and the first step implies that

$$P(A_1 \cap \ldots \cap A_k) = P(A_1) \ldots P(A_k) \, , A_1 \in \mathcal{F}_1, \ldots, A_k \in \mathcal{F}_k \, .$$

Once again, Theorem 3.1 implies independence of $\sigma(\mathcal{F}_1), \ldots, \sigma(\mathcal{F}_k)$ and completes the proof. $\square$

An immediate corollary of the above theorem is the following.

Corollary 3.1. *If $\{A_i : i \in I\}$ is an independent collection of σ-fields and for each $\alpha \in \Theta$, $\emptyset \neq I_\alpha \subset I$ is such that*

$$I_\alpha \cap I_\beta = \emptyset\,, \text{ for all } \alpha, \beta \in \Theta\,, \alpha \neq \beta\,,$$

then $\{\bigvee_{i \in I_\alpha} \mathcal{A}_i : \alpha \in \Theta\}$ is independent.

Exercise 3.6. 1. *If $\mathcal{A}_i$ is a σ-field for all $i \in I$, show that $A \in \bigvee_{i \in I} \mathcal{A}_i$ if and only if*

$$A \in \bigvee_{i \in I_0} \mathcal{A}_i\,,$$

for some countable $I_0 \subset I$.

2. *If X_i is a random-variable for all $i \in I$, show that $A \in \sigma(X_i : i \in I)$ if and only if*

$$A \in \sigma(X_i : i \in I_0)$$

for some countable $I_0 \subset I$.

3. *If $A_i \in \mathcal{A}$ for all $i \in I$, show that $A \in \sigma(A_i : i \in I)$ if and only if*

$$A \in \sigma(A_i : i \in I_0)$$

for some countable $I_0 \subset I$.

Definition 30. *A possibly infinite collection of random variables $\{X_i : i \in I\}$ is* independent *if $\{\sigma(X_i) : i \in I\}$ is an independent collection of σ-fields. Two random variables Y and Z are* identically distributed*, that is, $Y \stackrel{d}{=} Z$, if*

$$P(Y \in B) = P(Z \in B)\,, B \in \mathcal{B}(\mathbb{R})\,.$$

The collection $\{X_i : i \in I\}$ is independent and identically distributed *or* i.i.d. *if it is independent and*

$$X_i \stackrel{d}{=} X_j\,, i, j \in I\,.$$

Exercise 3.7. *If $\{\mathcal{A}_i : i \in I\}$ is an independent collection of σ-fields and X_i is $\mathcal{A}_i$-measurable for each $i \in I$, show that $\{X_i : i \in I\}$ is independent.*

Theorem 3.4. *If X and Y are independent random variables with finite mean, then XY has a finite mean and*

$$\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)\,.$$

Proof. As the first step, we show that for simple non-negative independent random variables X and Y,

$$\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)\,.$$

Since X and Y are simple and non-negative,

$$X = \sum_{i=1}^{m} \alpha_i \mathbf{1}_{A_i}\,, \quad Y = \sum_{i=1}^{n} \beta_i \mathbf{1}_{B_i}\,,$$

for some $\alpha_1, \ldots, \alpha_m, \beta_1, \ldots, \beta_n \ge 0$, $A_1, \ldots, A_m \in \sigma(X)$ and $B_1, \ldots, B_n \in \sigma(Y)$. Thus,

$$\begin{aligned}\mathrm{E}(XY) &= \sum_{i=1}^{m}\sum_{j=1}^{n} \alpha_i\beta_j P(A_i \cap B_j)\\ &= \sum_{i=1}^{m}\sum_{j=1}^{n} \alpha_i\beta_j P(A_i)P(B_j)\\ &= \mathrm{E}(X)\mathrm{E}(Y)\,,\end{aligned}$$

the independence of $\sigma(X)$ and $\sigma(Y)$ being used in the second line.

The second step is to show that for non-negative independent random variables X and Y, $\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$. There exist $\sigma(X)$-measurable simple random variables s_n such that $0 \le s_n \uparrow X$ and $\sigma(Y)$-measurable simple random variables t_n such that $0 \le t_n \uparrow Y$. As s_n and t_n are independent by Exc 3.7, the first step implies

$$\mathrm{E}(s_n t_n) = \mathrm{E}(s_n)\mathrm{E}(t_n)\,.$$

Observing that $0 \le s_n t_n \uparrow XY$, letting $n \to \infty$ and using MCT, it follows that $\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$.

Finally suppose X and Y are independent and integrable. Then $|X|$ and $|Y|$ are independent. The second step shows

$$\mathrm{E}\left(|X|\,|Y|\right) = \mathrm{E}(|X|)\mathrm{E}(|Y|) < \infty\,.$$

Thus, XY is integrable. Splitting $X = X^+ - X^-$ and likewise for Y, the proof follows. □

Definition 31. *For random variables X and Y such that X, Y, XY are integrable, the* covariance *of X and Y is*

$$\mathrm{Cov}(X,Y) = \mathrm{E}\left[\big(X - \mathrm{E}(X)\big)\big(Y - \mathrm{E}(Y)\big)\right]\,.$$

Theorem 3.5. *If X and Y are random variables with finite variance, then* $\mathrm{Cov}(X,Y)$ *is defined and*

$$|\mathrm{Cov}(X,Y)| \le \sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}\,.$$

Proof. Follows from Cauchy-Schwarz. □

Theorem 3.6. *1. If* $\mathrm{Cov}(X, Y)$ *is defined, then*

$$\mathrm{Cov}(X, Y) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)\,.$$

2. If X *and* Y *are independent and integrable, then* $\mathrm{Cov}(X, Y)$ *exists and equals zero.*

3. If X *has a finite variance,* $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$*.*

4. If $\mathrm{Cov}(X, Y)$ *is defined, then*

$$\mathrm{Cov}(\alpha X + \gamma, \beta Y + \delta) = \alpha\beta\mathrm{Cov}(X, Y)\,, \alpha, \beta, \delta, \gamma \in \mathbb{R}\,.$$

5. If $X_1, \ldots, X_n$ *have finite variances, then*

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2 \sum_{1 \le i < j \le n} \mathrm{Cov}(X_i, X_j)\,.$$

Proof. Exc. □

Definition 32. *For random variables* X *and* Y *whose variances are finite and positive, their* correlation *is*

$$\mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}\,.$$

Exercise 3.8. *For a random variable* X*, show that* $\mathrm{Var}(X) = 0$ *if and only if* X *is a degenerate random variable, that is, for some* $c \in \mathbb{R}$*,* $X = c$ *a.s.*

Theorem 3.7. *Suppose* X *and* Y *are non-degenerate random variables with finite variances.*

1. For $\alpha, \beta, \gamma, \delta \in \mathbb{R}$ *with* $\alpha, \beta \neq 0$*,*

$$\mathrm{Corr}\left(\alpha X + \gamma, \beta Y + \delta\right) = \mathrm{sgn}(\alpha\beta)\mathrm{Corr}(X, Y)\,.$$

2. A correlation coefficient always lies between -1 *and* 1*, that is,*

$$|\mathrm{Corr}(X, Y)| \le 1\,.$$

3. There exist $a, b, c \in \mathbb{R}$ *with* $a, b \neq 0$ *such that* $aX + bY = c$ *a.s. if and only if*

$$\mathrm{Corr}(X, Y) = \pm 1\,.$$

Proof. 1. A trivial consequence of Theorem 3.6.4.
2. Follows from Theorem 3.5.
3. For the "only if" part, suppose $aX + bY = c$ a.s. for some $a, b \neq 0$. Then

$$\begin{aligned}
\mathrm{Corr}(X,Y) &= \mathrm{Corr}\left(X, \frac{c}{b} - \frac{a}{b}X\right) \\
&= \mathrm{sgn}\left(-\frac{a}{b}\right)\mathrm{Corr}(X,X) \\
&= \mathrm{sgn}\left(-\frac{a}{b}\right),
\end{aligned}$$

1. being used in the second line and the last line follows from the observation that $\mathrm{Corr}(X,X) = 1$ which is a restatement of Theorem 3.6.3. Thus $aX + bY = c$ a.s. for some $a, b \neq 0$ implies $\mathrm{Corr}(X,Y) = \pm 1$.

Conversely, suppose $\mathrm{Corr}(X,Y) = \pm 1$. Define

$$X' = \frac{X - \mathrm{E}(X)}{\sqrt{\mathrm{Var}(X)}}, \quad Y' = \frac{Y - \mathrm{E}(Y)}{\sqrt{\mathrm{Var}(Y)}}.$$

Then $\mathrm{E}(X'^2) = \mathrm{E}(Y'^2) = 1$ and $\mathrm{E}(X'Y') = \mathrm{Corr}(X,Y)$. If $\mathrm{Corr}(X,Y) = 1$, then

$$\mathrm{E}\left[(X' - Y')^2\right] = \mathrm{E}(X'^2) + \mathrm{E}(Y'^2) - 2\mathrm{E}(X'Y') = 2 - 2\mathrm{Corr}(X,Y) = 0\,,$$

showing that $X' = Y'$ a.s. In other words, $\mathrm{Corr}(X,Y) = 1$ implies

$$\frac{X}{\sqrt{\mathrm{Var}(X)}} - \frac{Y}{\mathrm{Var}(Y)} = \frac{\mathrm{E}(X)}{\sqrt{\mathrm{Var}(X)}} - \frac{\mathrm{E}(Y)}{\mathrm{Var}(Y)} \text{ a.s.}$$

A similar calculation shows that $\mathrm{Corr}(X,Y) = -1$ implies

$$\frac{X}{\sqrt{\mathrm{Var}(X)}} + \frac{Y}{\mathrm{Var}(Y)} = \frac{\mathrm{E}(X)}{\sqrt{\mathrm{Var}(X)}} + \frac{\mathrm{E}(Y)}{\mathrm{Var}(Y)} \text{ a.s.}$$

This proves the "if" part and thus completes the proof. □

Now we proceed towards showing that given a countable collection of CDFs, there exist independent random variables with those CDFs. The first step in that direction is the following result, which is an alternate way of proving the second part of Theorem 2.1.

Theorem 3.8. *Let F be a C.D.F. and define*

$$F^{\leftarrow}(y) = \inf\{x \in \mathbb{R} : F(x) \geq y\}\,, 0 < y < 1\,. \tag{3.6}$$

If $U \sim$ Uniform$(0,1)$, then $F^{\leftarrow}(U)$ has C.D.F. F.

Proof. It suffices to prove that for $x_0 \in \mathbb{R}$ and $0 < y_0 < 1$,

$$F^{\leftarrow}(y_0) \le x_0 \iff y_0 \le F(x_0)\,, \tag{3.7}$$

because then it would follow from the fact that $0 < U < 1$ a.s. that for $x \in \mathbb{R}$,

$$P\left(F^{\leftarrow}(U) \le x\right) = P\left(U \le F(x)\right) = F(x)\,.$$

Proceeding towards proving (3.7), first assume $y_0 \le F(x_0)$. In other words,

$$x_0 \in \{x \in \mathbb{R} : F(x) \ge y_0\}\,.$$

Thus, $x_0 \ge \inf\{x \in \mathbb{R} : F(x) \ge y_0\} = F^{\leftarrow}(y_0)$, proving the "$\Leftarrow$" part, that is, the "if" part.

For the reverse implication of (3.7), we shall show that $y_0 > F(x_0) \Rightarrow F^{\leftarrow}(y_0) > x_0$. Assume $y_0 > F(x_0)$. Right continuity of F implies there exists $x_1 > x_0$ with $F(x_1) < y_0$. As F is non-decreasing,

$$F\big((-\infty, x_1]\big) = \big(0, F(x_1)\big] \subset (0, y_0)\,.$$

In other words, $F(x) < y_0$ for all $x \le x_1$, which is equivalent to

$$\{x : F(x) \ge y_0\} \subset (x_1, \infty)\,.$$

Thus, $\inf\{x : F(x) \ge y_0\} \ge x_1 > x_0$. That is, $F^{\leftarrow}(y_0) > x_0$, as desired. This proves the "$\Rightarrow$" implication, that is, the "only if" part of (3.7), which completes the proof. □

The next step in the same direction gives an alternate way of generating an Uniform$(0,1)$ random variable.

Theorem 3.9. *If $X_1, X_2, \ldots$ are i.i.d. from Bernoulli$(1/2)$, that is, they take values 0 and 1 with probability $1/2$ each, then*

$$U = \sum_{n=1}^{\infty} 2^{-n} X_n$$

follows Uniform$(0,1)$.

Proof. For $n = 1, 2, 3, \ldots$, define

$$U_n = \sum_{i=1}^{n} 2^{-i} X_i\,.$$

We shall first show by induction on n that

$$P\left(U_n = 2^{-n} i\right) = 2^{-n}\,, i = 0, 1, \ldots, 2^n - 1\,. \tag{3.8}$$

Since $U_1 = X_1/2$, that is, U_1 takes values 0 and $1/2$ each with probability $1/2$, (3.8) trivially holds for $n = 1$. Assume (3.8) for some n as the induction hypothesis. Notice

$$U_{n+1} - U_n = 2^{-n-1} X_{n+1} \,. \tag{3.9}$$

Since U_n is measurable with respect to $\sigma(X_1, \ldots, X_n)$ which is independent of $\sigma(X_{n+1})$ by Corollary 3.1, U_n is independent of $U_{n+1} - U_n$. Another implication of (3.9) and the fact $U_n \in \{2^{-n}i : i = 0, 1, \ldots, 2^n - 1\}$ is that

$$U_{n+1} \in \begin{cases} \{2^{-n}i : i = 0, 1, \ldots, 2^n - 1\} \,, & \text{if } U_{n+1} - U_n = 0 \,, \\ \{2^{-n}i + 2^{-n-1} : i = 0, 1, \ldots, 2^n - 1\} \,, & \text{otherwise.} \end{cases} \tag{3.10}$$

Since $\{2^{-n}i : i = 0, 1, \ldots, 2^n - 1\} \cap \{2^{-n}i + 2^{-n-1} : i = 0, 1, \ldots, 2^n - 1\} = \emptyset$, for $i = 0, 1, \ldots, 2^n - 1$, (3.10) implies

$$\begin{aligned} P(U_{n+1} = 2^{-n}i) &= P(U_{n+1} = 2^{-n}i, U_{n+1} - U_n = 0) \\ &= P(U_n = 2^{-n}i, U_{n+1} - U_n = 0) \\ \text{(independence of } U_n, U_{n+1} - U_n) &= P(U_n = 2^{-n}i) P(U_{n+1} - U_n = 0) \\ &= \frac{1}{2} P(U_n = 2^{-n}i) \\ &= 2^{-n-1} \,, \end{aligned}$$

the penultimate line follows from (3.9), whereas (3.8) for n implies the last line. A similar calculation with (3.10) shows

$$P\left(U_{n+1} = 2^{-n}i + 2^{-n-1}\right) = 2^{-n-1} \,, i = 0, 1, \ldots, 2^n - 1 \,.$$

Observing that

$$\begin{aligned} &\{2^{-n}i : i = 0, 1, \ldots, 2^n - 1\} \cup \{2^{-n}i + 2^{-n-1} : i = 0, 1, \ldots, 2^n - 1\} \\ &= \{2^{-n-1}i : i = 0, 1, \ldots, 2^{n+1} - 1\} \,, \end{aligned}$$

(3.8) follows for $n + 1$. Mathematical induction shows (3.8) for all n.

Note that $U_n \uparrow U$ and hence for any $x \in \mathbb{R}$, $[U_n \le x] \downarrow [U \le x]$. Thus, for $x \in (0, 1)$,

$$\begin{aligned} P(U \le x) &= \lim_{n \to \infty} P(U_n \le x) \\ &= \lim_{n \to \infty} 2^{-n} \left([2^n x] + 1\right) \\ &= x \,, \end{aligned}$$

(3.8) implying the second line, $[z]$ denoting the largest integer less than or equal to z. Monotonicity of the CDF implies $P(U \le 0) = 0$ and $P(U \le 1) = 1$. Hence,

$$P(U \le x) = \begin{cases} 0, & x \le 0 \,, \\ x, & 0 < x < 1 \,, \\ 1, & x \ge 1 \,, \end{cases}$$

that is, U follows Uniform$(0, 1)$. This completes the proof. $\square$

Now we are in a position to prove the existence of a countable collection of independent random variables.

Theorem 3.10. *Given CDFs $F_1, F_2, \ldots$, there exist independent random variables $X_1, X_2, \ldots$, defined on some probability space $(\Omega, \mathcal{A}, P)$, whose CDFs are $F_1, F_2, \ldots$, respectively.*

Proof. Let $(\Omega, \mathcal{A}, P)$ be the probability space associated with infinite tosses of a fair coin as in Example 2.3. Define for $n = 1, 2, \ldots$,

$$Y_n(\omega) = \mathbf{1}(\omega_n = \mathrm{H}), \omega = (\omega_1, \omega_2, \omega_3, \ldots) \in \Omega .$$

Clearly, $Y_1, Y_2, \ldots$ are i.i.d. from Bernoulli(1/2). Fix a bijection $\phi : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ and define

$$Z_{ij} = Y_{\phi(i,j)}, i, j \in \mathbb{N} .$$

It is immediate that $\{Z_{ij} : i, j \in \mathbb{N}\}$ are i.i.d. from Bernoulli(1/2).

Define

$$U_i = \sum_{j=1}^{\infty} 2^{-j} Z_{ij}, i \in \mathbb{N} . \tag{3.11}$$

For $i = 1, 2, \ldots$, U_i is measurable with respect to $\mathcal{A}_i = \sigma(Z_{i1}, Z_{i2}, \ldots)$. Corollary 3.1 shows $\mathcal{A}_1, \mathcal{A}_2, \ldots$ are independent σ-fields. Hence, $U_1, U_2, \ldots$ are independent. Theorem 3.9 in view of (3.11) shows U_i follows Uniform$(0, 1)$ for every i. That is, $U_1, U_2, \ldots$ are i.i.d. from Uniform$(0, 1)$.

Finally, set $X_i = F_i^{\leftarrow}(U_i)$, where $F_i^{\leftarrow}$ is as in (3.6). Theorem 3.8 shows X_i has CDF F_i. Since $U_1, U_2, \ldots$ are independent, so are $X_1, X_2, \ldots$, and hence the proof follows. □

Remark 1. *The construction* (3.11) *along with Theorems 3.8 and 3.9 gives an alternative proof of the second part of Theorem 2.1 which completely bypasses Theorem 1.4.*

Definition 33. *Let* $\mathbb{R}^{\mathbb{N}} = \{(x_1, x_2, x_3, \ldots) : x_n \in \mathbb{R}, n = 1, 2, \ldots\}$ *and*

$$\mathcal{B}\left(\mathbb{R}^{\mathbb{N}}\right) = \sigma\left\{\left(\prod_{i=1}^{n} A_i\right) \times \mathbb{R} \times \mathbb{R} \times \ldots : n \geq 1, A_1, \ldots, A_n \in \mathcal{B}(\mathbb{R})\right\} .$$

Theorem 3.11. *If $P_1, P_2, \ldots$ are probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, then there exists a unique probability measure P on $\left(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}})\right)$ such that*

$$P\left(A_1 \times \ldots \times A_n \times \mathbb{R} \times \mathbb{R} \times \ldots\right) = \prod_{i=1}^{n} P_i(A_i), A_1, \ldots, A_n \in \mathcal{B}(\mathbb{R}), n \geq 1 . \tag{3.12}$$

Proof. For $n = 1, 2, \ldots$, let

$$F_n(x) = P_n\left((-\infty, x]\right), x \in \mathbb{R}.$$

Theorem 3.10 shows the existence of independent random variables $X_1, X_2, X_3, \ldots$ having CDFs $F_1, F_2, \ldots$, respectively, defined on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Exc 2.7 shows that $\mathbb{P}(X_i \in A) = P_i(A)$ for $i = 1, 2, \ldots$ and $A \in \mathcal{B}(\mathbb{R})$.

Define $T : \Omega \to \mathbb{R}^{\mathbb{N}}$ by

$$T(\omega) = (X_1(\omega), X_2(\omega), \ldots), \omega \in \Omega.$$

The map T is measurable, that is, $T^{-1}B \in \mathcal{A}$ for all $B \in \mathcal{B}(\mathbb{R}^{\mathbb{N}})$, because for all $n \geq 1$ and $A_1, \ldots, A_n \in \mathcal{B}(\mathbb{R})$,

$$T^{-1}(A_1 \times \ldots \times A_n \times \mathbb{R} \times \mathbb{R} \times \ldots) = \bigcap_{i=1}^{n} X_i^{-1} A_i \in \mathcal{A}.$$

Let $P = \mathbb{P} \circ T^{-1}$. Then, for $A_1, \ldots, A_n$ as above,

$$\begin{aligned} P\left(A_1 \times \ldots \times A_n \times \mathbb{R} \times \mathbb{R} \times \ldots\right) &= \mathbb{P}\left(T^{-1}(A_1 \times \ldots \times A_n \times \mathbb{R} \times \mathbb{R} \times \ldots)\right) \\ &= \mathbb{P}\left(\bigcap_{i=1}^{n} X_i^{-1} A_i\right) \\ &= \prod_{i=1}^{n} \mathbb{P}(X_i \in A_i) \\ &= \prod_{i=1}^{n} P_i(A_i), \end{aligned}$$

the independence of $X_1, X_2, \ldots$ implying the penultimate line. Thus, P satisfies (3.12). Uniqueness follows from the observation that

$$\{A_1 \times \ldots \times A_n \times \mathbb{R} \times \mathbb{R} \times \ldots : n \geq 1, A_1, \ldots, A_n \in \mathcal{B}(\mathbb{R})\}$$

is a semi-field. This completes the proof. □

Definition 34. *The probability measure P on $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}}))$ satisfying* (3.12) *is the* product measure *of $P_1, P_2, \ldots$ and is denoted by*

$$P = \bigotimes_{n=1}^{\infty} P_n.$$

The above infinite product could be defined because $P_1, P_2, \ldots$ are all probability measures.

We conclude this chapter by pointing out that measure theory is indispensable for a rigorous treatment of probability theory, which is now amply clear. The following are a few instances where usage of measure theory was necessary.

1. For studying the simple random experiment of infinite tosses of a fair coin, as in Example 2.3.

2. For defining expectation of a general random variable, that is, one which is neither discrete nor has a density.

3. Showing linearity of expectation is very difficult, even for random variables having a density, without the measure theoretic definition.

4. Answering the question of when a random variable has a density is impossible without the Radon-Nikodym theorem.

5. Last but not the least, the study of independence in this chapter became much easier thanks to measure theory.

4 Several random variables

Definition 35. *For random variables $X_1, \ldots, X_d$, which by convention are defined on the same probability space $(\Omega, \mathcal{A}, P)$, the* joint CDF *of $(X_1, \ldots$ $\ldots, X_d)$ is a function $F : \mathbb{R}^d \to [0,1]$ defined by*

$$F(x_1, \ldots, x_d) = P(X_1 \le x_1, \ldots, X_d \le x_d)\,, x_1, \ldots, x_d \in \mathbb{R}\,.$$

A joint CDF will often be referred to simply by 'CDF'. For this chapter, we introduce the following notations:

$$\mathcal{H} = \{(a_1, b_1] \times \ldots \times (a_d, b_d] : -\infty < a_i < b_i < \infty \text{ for } i = 1, \ldots, d\}\ ,$$
$$\Delta_R F = \sum_{(x_1,\ldots,x_d) \in \{a_1,b_1\} \times \ldots \times \{a_d,b_d\}} (-1)^{\#\{i : x_i = a_i\}} F(x_1, \ldots, x_d)\,, \tag{4.1}$$

for a function $F : \mathbb{R}^d \to \mathbb{R}$ and $R = (a_1, b_1] \times \ldots \times (a_d, b_d] \in \mathcal{H}$. A restatement of Exc 3.3 in the above notations is that

$$P\big((X_1, \ldots, X_d) \in R\big) = \Delta_R F\,, R \in \mathcal{H}\,, \tag{4.2}$$

if F is the CDF of $(X_1, \ldots, X_d)$. The following theorem can be proved along similar lines as the direct part of Theorem 2.1, with the help of (4.2).

Theorem 4.1. *If F is the joint CDF of d random variables, then*

1. *$\Delta_R F \ge 0$ for all $R \in \mathcal{H}$,*

2. *F is continuous from above, that is,*

$$\lim_{y_1 \downarrow x_1, \ldots, y_d \downarrow x_d} F(y_1, \ldots, y_d) = F(x_1, \ldots, x_d) \text{ for all } x_1, \ldots, x_d \in \mathbb{R}\,,$$

3. for any $k = 1, \ldots, d$ and $1 \leq i_i < \ldots < i_k \leq d$,

$$\lim_{i_1 \to -\infty, \ldots, i_k \to -\infty} F(x_1, \ldots, x_d) = 0\,,$$

where $x_j \in \mathbb{R}$ is fixed for all $j \in \{1, \ldots, d\} \setminus \{i_1, \ldots, i_k\}$,

4. and

$$\lim_{x_1 \to \infty, \ldots, x_d \to \infty} F(x_1, \ldots, x_d) = 1\,.$$

Proof. Exercise. $\square$

The following measure-theoretic fact is a d-dimensional generalization of Theorem 1.4.

Fact 4.1. *If $F : \mathbb{R}^d \to \mathbb{R}$ is a function which is continuous from above and satisfies $\Delta_R F \geq 0$ for all $R \in \mathcal{H}$, then there exists a unique Radon measure μ on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ such that*

$$\mu(R) = \Delta_R F\,,\ \textit{for all } R \in \mathcal{H}\,.$$

Neither the above fact nor the following theorem, which is built on it and gives a converse of Theorem 4.1, will be used much in the course. Nonetheless, a proof of the above fact is given in Subsection 9.2 of the Appendix.

Theorem 4.2. *If $F : \mathbb{R}^d \to [0,1]$ satisfies 1.-4. of Theorem 4.1, then there exist random variables $X_1, \ldots, X_d$ defined on some probability space such that F is the joint CDF of $(X_1, \ldots, X_d)$.*

Proof. Since F satisfies 1. and 2., the preceding fact implies there exists a Radon measure μ on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ such that

$$\mu(R) = \Delta_R F\,,\ R \in \mathcal{H}\,.$$

Our first goal is to show μ is a probability measure. To that end, rewrite the above for $R = (-m, n]^d$, where $m, n \in \mathbb{N}$, as

$$\mu\left((-m, n]^d\right) = \sum_{(x_1, \ldots, x_d) \in \{-m, n\}^d} (-1)^{\#\{i : x_i = -m\}} F(x_1, \ldots, x_d)\,, m, n \in \mathbb{N}\,.$$

In the right hand side above, if n is fixed and $m \to \infty$, then every term except $F(n, \ldots, n)$ goes to zero by 3. of Theorem 4.1, which F satisfies by hypothesis. As μ is a measure, the left hand side increases to $\mu((-\infty, n]^d)$ as $m \to \infty$. Therefore,

$$\mu\left((-\infty, n]^d\right) = F(n, n, \ldots, n)\,, n \geq 1\,. \tag{4.3}$$

Let $n \to \infty$ and use 4. to conclude $\mu(\mathbb{R}^d) = 1$. In other words, μ is a probability measure.

Let $\Omega = \mathbb{R}^d$, $\mathcal{A} = \mathcal{B}(\mathbb{R}^d)$ and $P = \mu$. Define $X_i : \mathbb{R}^d \to \mathbb{R}$ by

$$X_i(x_1, \ldots, x_d) = x_i \, , (x_1, \ldots, x_d) \in \mathbb{R}^d \, ,$$

for $i = 1, \ldots, d$. Then $X_1, \ldots, X_d$ are random variables on the probability space $(\Omega, \mathcal{A}, P)$. Arguments leading to (4.3) can be slightly tweaked to show that

$$P\left((-\infty, x_1] \times \ldots \times (-\infty, x_d]\right) = F(x_1, \ldots, x_d) \, , x_1, \ldots, x_d \in \mathbb{R} \, .$$

As the above is the same as saying F is the CDF of $(X_1, \ldots, X_d)$, the proof follows. $\square$

Definition 36. *For random variables $X_1, \ldots, X_d$, the* joint distribution *of $(X_1, \ldots, X_d)$ is the measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ given by $P \circ (X_1, \ldots, X_d)^{-1}$, that is, the measure pushed forward to $\mathbb{R}^d$ by $(X_1, \ldots, X_d)$. For a measurable function $f : \mathbb{R}^d \to \mathbb{R}$, the integral of f with respect to the measure $P \circ (X_1, \ldots, X_d)^{-1}$, if defined, is denoted by*

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \ldots, x_d) P(X_1 \in dx_1, \ldots, X_d \in dx_d) \, .$$

For random variables $Y_1, \ldots, Y_d$, $(X_1, \ldots, X_d) \stackrel{d}{=} (Y_1, \ldots, Y_d)$ means

$$P\left((X_1, \ldots, X_d) \in B\right) = P\left((Y_1, \ldots, Y_d) \in B\right) \text{ for all } B \in \mathcal{B}(\mathbb{R}^d) \, .$$

The following result is similar to its one-dimensional analogue.

Theorem 4.3. *1. For random variables $X_1, \ldots, X_d$ and a Borel measurable $f : \mathbb{R}^d \to \mathbb{R}$,*

$$\mathrm{E}\left(f(X_1, \ldots, X_d)\right) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \ldots, x_d) P(X_1 \in dx_1, \ldots, X_d \in dx_d) \, ,$$

whenever either side is defined.
2. For random variables $X_1, \ldots, X_d, Y_1, \ldots, Y_d$,

$$(X_1, \ldots, X_d) \stackrel{d}{=} (Y_1, \ldots, Y_d)$$

if and only if the CDFs of $(X_1, \ldots, X_d)$ and $(Y_1, \ldots, Y_d)$ are the same.

Proof. 1. Follows from Theorem 1.6.
2. Follows from (4.2) and the observation that

$$\left\{ \left(\prod_{i=1}^{d} (a_i, b_i] \right) \cap \mathbb{R}^d : -\infty \le a_i \le b_i \le \infty \, , i = 1, \ldots, d \right\}$$

is a semi-field and for every set in the above class, there exists sets in $\mathcal{H}$ increasing to that. $\square$

Definition 37. *For discrete random variables* $X_1, \ldots, X_d$*, the* joint PMF *of* $(X_1, \ldots, X_d)$ *is the function* $p : \mathbb{R}^d \to [0,1]$ *defined by*

$$p(x_1, \ldots, x_d) = P(X_1 = x_1, \ldots, X_d = x_d)\,, x_1, \ldots, x_d \in \mathbb{R}\,.$$

A Borel function $f : \mathbb{R}^d \to [0, \infty)$ *is the* joint density *of* $(X_1, \ldots, X_d)$ *if*

$$P\left[(X_1, \ldots, X_d) \in B\right] = \int_B f(x)\, dx\,, B \in \mathcal{B}(\mathbb{R}^d)\,.$$

Theorem 4.4. *A Borel map* $f : \mathbb{R}^d \to [0, \infty)$ *is the density of* $(X_1, \ldots, X_d)$ *if and only if*

$$F(x_1, \ldots, x_d) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_d} f(z_1, \ldots, z_d)\ dz_d \ldots dz_1\,,$$

for all $x_1, \ldots, x_d \in \mathbb{R}$*, where* F *is the CDF of* $(X_1, \ldots, X_d)$*.*

Proof. Follows from Theorem 4.3.2. □

Theorem 4.5. *If* $X_1, \ldots, X_d$ *are discrete random variables, then for* $k = 1, \ldots, d-1$*,*

$$P(X_1 = x_1, \ldots, X_k = x_k) = \sum_{x_{k+1} \in \mathbb{R}} \cdots \sum_{x_d \in \mathbb{R}} P(X_1 = x_1, \ldots, X_d = x_d)\,,$$

for all $x_1, \ldots, x_k \in \mathbb{R}$*. If* f *is the joint density of* $(X_1, \ldots, X_d)$*, then for* $k = 1, \ldots, d-1$*,* $g : \mathbb{R}^k \to \mathbb{R}$ *defined by*

$$g(x_1, \ldots, x_k) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \ldots, x_d)\, dx_{k+1} \ldots dx_d\,, x_1, \ldots, x_k \in \mathbb{R}\,,$$

is the density of $(X_1, \ldots, X_k)$*.*

Proof. We prove the second claim; the proof of the first one is similar. Let f be the density of $(X_1, \ldots, X_d)$. Then for $B \in \mathcal{B}(\mathbb{R}^k)$,

$$\begin{aligned}
&P\left[(X_1, \ldots, X_k) \in B\right] \\
&= P\left[(X_1, \ldots, X_d) \in B \times \mathbb{R}^{d-k}\right] \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \mathbf{1}\left[(x_1, \ldots, x_k) \in B\right] f(x_1, \ldots, x_d) dx_d \ldots dx_1 \\
&= \int_{\mathbb{R}^k} \mathbf{1}\left((x_1, \ldots, x_k) \in B\right) \left(\int_{\mathbb{R}^{d-k}} f(x_1, \ldots, x_d) dx_{k+1} \ldots dx_d \right)\ dx_1 \ldots dx_k \\
&= \int_{\mathbb{R}^k} \mathbf{1}\left((x_1, \ldots, x_k) \in B\right) g(x_1, \ldots, x_k)\ dx_1 \ldots dx_k\,.
\end{aligned}$$

As this is true for all $B \in \mathcal{B}(\mathbb{R}^k)$, g is the density of $(X_1, \ldots, X_k)$, as claimed. □

Theorem 4.6. *For discrete independent random variables, the joint PMF is the product of the marginal PMFs, that is, if $X_1, \ldots, X_d$ are discrete and independent, then*

$$P(X_1 = x_1, \ldots, X_d = x_d) = \prod_{i=1}^{d} P(X_i = x_i)\,, (x_1, \ldots, x_d) \in \mathbb{R}^d\,.$$

If $X_1, \ldots, X_d$ are independent with respective densities $f_1, \ldots, f_d$, then f, defined by,

$$f(x_1, \ldots, x_d) = \prod_{i=1}^{d} f_i(x_i)\,, x_1, \ldots, x_d \in \mathbb{R}\,,$$

is the density of $(X_1, \ldots, X_d)$.

Proof. The first claim follows immediately from the definition of independence. For the second claim, using independence, write for $x_1, \ldots, x_d \in \mathbb{R}$,

$$\begin{aligned} P(X_1 \le x_1, \ldots, X_d \le x_d) &= \prod_{i=1}^{d} P(X_i \le x_i) \\ &= \prod_{i=1}^{d} \int_{-\infty}^{x_i} f_i(z_i)\, dz_i \\ &= \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_d} f(z_1, \ldots, z_d) dz_d \ldots dz_1\,. \end{aligned}$$

Theorem 4.4 completes the proof. □

The following is a converse of the above, and in fact, slightly stronger than that.

Theorem 4.7. *If $X_1, \ldots, X_d$ are discrete random variables for which there exist $c \in \mathbb{R}$ and functions $p_1, \ldots, p_d : \mathbb{R} \to \mathbb{R}$ such that*

$$P(X_1 = x_1, \ldots, X_d = x_d) = c \prod_{i=1}^{d} p_i(x_i)\,, x_1, \ldots, x_d \in \mathbb{R}\,,$$

then $X_1, \ldots, X_d$ are independent. Furthermore, if

$$\sum_{x \in \mathbb{R}} p_i(x) = 1\,, i = 1, \ldots, d\,,$$

then $c = 1$ and $p_1, \ldots, p_d$ are the respective marginal PMFs of $X_1, \ldots, X_d$.

If $(X_1, \ldots, X_d)$ has a joint density f which can be written as

$$f(x_1, \ldots, x_d) = c \prod_{i=1}^{d} f_i(x_i)\,, (x_1, \ldots, x_d) \in \mathbb{R}^d\,, \tag{4.4}$$

for measurable functions $f_1, \ldots, f_d : \mathbb{R} \to \mathbb{R}$, then $X_1, \ldots, X_d$ are independent. If, in addition, $f_1, \ldots, f_d$ integrate to 1, then $c = 1$ and $f_1, \ldots, f_d$ are the marginal densities of $X_1, \ldots, X_d$, respectively.

Proof. The second claim will be proved as the proof of the first claim is similar. Non-negativity of f and (4.4) show

$$|c| \prod_{i=1}^{d} |f_i(x_i)| = f(x_1, \ldots, x_d)\,, x_1, \ldots, x_d \in \mathbb{R}\,. \tag{4.5}$$

Thus,

$$\begin{aligned} |c| \prod_{i=1}^{d} \int_{-\infty}^{\infty} |f_i(x_i)|\, dx_i &= \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} |c| \prod_{i=1}^{d} |f_i(x_i)|\; dx_1 \ldots dx_d \\ &= 1\,, \end{aligned}$$

(4.5) and that f is a density implying the second line. Hence,

$$\int_{-\infty}^{\infty} |f_i(x)|\, dx < \infty\,, i = 1, \ldots, d\,.$$

This allows integrating both sides of (4.4) over $(x_1, \ldots, x_d) \in \mathbb{R}^d$ which yields

$$c\,\alpha_1 \ldots \alpha_d = 1\,, \tag{4.6}$$

where

$$\alpha_i = \int_{\infty}^{\infty} f_i(x)\, dx\,, i = 1, \ldots, d\,.$$

Theorem 4.5 shows that for $i = 1, \ldots, d$, the density g_i of X_i can be obtained by fixing x_i and integrating the right hand side of (4.4) over all other variables. That is,

$$g_i(x_i) = c \left(\prod_{j \in \{1, \ldots, d\} \setminus \{i\}} \alpha_j \right) f_i(x_i)\,, x_i \in \mathbb{R}\,;$$

(4.6) implies

$$g_i(x) = \alpha_i^{-1} f_i(x)\,, x \in \mathbb{R}\,.$$

Use this to rewrite (4.4) as

$$\begin{aligned} f(x_1, \ldots, x_d) &= c \prod_{i=1}^{d} \alpha_i g_i(x_i) \\ &= \prod_{i=1}^{d} g_i(x_i)\,, \end{aligned}$$

for all $x_1, \ldots, x_d \in \mathbb{R}$, the last line following from (4.6). Therefore, for $B_1, \ldots$ $\ldots, B_d \in \mathcal{B}(\mathbb{R})$,

$$\begin{aligned}
P(X_1 \in B_1, \ldots, X_d \in B_d) &= \int_{B_1} \cdots \int_{B_d} f(x_1, \ldots, x_d)\ dx_d \ldots dx_1 \\
&= \int_{B_1} \cdots \int_{B_d} \prod_{i=1}^{d} g_i(x_i)\ dx_d \ldots dx_1 \\
&= \prod_{i=1}^{d} \int_{B_i} g_i(x_i)\ dx_i \\
&= \prod_{i=1}^{d} P(X_i \in B_i)\,,
\end{aligned}$$

the equality in the last line holds because $g_1, \ldots, g_d$ are the respective densities of $X_1, \ldots, X_d$. Thus, $X_1, \ldots, X_d$ are independent.

If

$$\int_{-\infty}^{\infty} f_i(x)\, dx = 1\,, i = 1, \ldots, d\,,$$

then $\alpha_i = 1$ for all i, showing $c = 1$ by (4.6) and that $g_i = f_i$. In this case, $f_1, \ldots, f_d$ are thus the respective marginal densities of $X_1, \ldots, X_d$. This completes the proof. □

The following exercise is a variant of the above theorem.

Exercise 4.1. *Suppose f is the density of $(X_1, \ldots, X_m, Y_1, \ldots, Y_n)$. If $(X_1, \ldots \ldots, X_m)$ and $(Y_1, \ldots, Y_n)$ are independent, then show that*

$$f(x_1, \ldots, x_m, y_1, \ldots, y_n) = f_X(x_1, \ldots, x_m) f_Y(y_1, \ldots, y_n)\,,$$

for almost all $(x_1, \ldots, x_m, y_1, \ldots, y_n) \in \mathbb{R}^{m+n}$, where f_X and f_Y are the densities of $(X_1, \ldots, X_m)$ and $(Y_1, \ldots, Y_n)$, respectively. Conversely, if there exist measurable $g : \mathbb{R}^m \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}$ and $c \in \mathbb{R}$ such that

$$f(x_1, \ldots, x_m, y_1, \ldots, y_n) = c g(x_1, \ldots, x_m) h(y_1, \ldots, y_n)\,,$$

for almost all $(x_1, \ldots, x_m, y_1, \ldots, y_n) \in \mathbb{R}^{m+n}$, then show that $(X_1, \ldots, X_m)$ and $(Y_1, \ldots, Y_n)$ are independent. Besides, if

$$\int_{\mathbb{R}^m} g(x)\, dx = 1 = \int_{\mathbb{R}^n} h(x)\, dx\,,$$

then show that $c = 1$, that g, h are non-negative a.e., and that $g \vee 0$ and $h \vee 0$ are the respective densities of $(X_1, \ldots, X_m)$ and $(Y_1, \ldots, Y_n)$.

Theorem 4.8. *Suppose $X = (X_1, \ldots, X_d)$ is a random vector with $P(X \in U) = 1$ for some open set $U \subset \mathbb{R}^d$. Let $\psi : U \to V$ be a bijection for some open set $V \subset \mathbb{R}^d$. Let $T : V \to U$ be the inverse of ψ. Assume T is continuously differentiable and its Jacobian matrix $J(y)$ at $y \in V$, defined by*

$$J(y) = \frac{\partial T(y)}{\partial y}\,,$$

is non-singular for all $y \in V$. Then the joint density of $Y = (Y_1, \ldots, Y_d) = \psi(X)$ is

$$g(y) = \begin{cases} f \circ T(y) |\det(J(y))|\,, & y \in V\,, \\ 0\,, & y \notin V\,. \end{cases}$$

Proof. Since $(Y_1, \ldots, Y_d) \in V$ a.s., for $B \in \mathcal{B}(\mathbb{R}^d)$,

$$\begin{aligned} P\big((Y_1, \ldots, Y_d) \in B\big) &= P\big((Y_1, \ldots, Y_d) \in B \cap V\big) \\ &= P\big((X_1, \ldots, X_d) \in T(B \cap V)\big) \\ &= \int_{T(B\cap V)} f(x)\, dx \\ &= \int_{B\cap V} f \circ T(y) |\det(J(y))|\, dy \\ &= \int_B g(y)\, dy\,, \end{aligned}$$

the penultimate line following from Theorem 1.12. Hence the proof follows. □

Example 4.1. *Let $X \sim Gamma(\alpha)$ and $Y \sim Gamma(\beta)$ independently of each other. We want to find the distribution of $W = X/(X + Y)$.*

Theorem 4.8 is the only tool at our disposal, which is valid for one-one functions from an open subset of $\mathbb{R}^2$ to $\mathbb{R}^2$. Therefore, we define an auxiliary random variable $Z = X + Y$. Thus, $(W, Z) = \psi(X, Y)$ where $\psi : U \to V$ is a bijection defined by

$$\psi(x, y) = \left(\frac{x}{x + y}, x + y\right)\,, (x, y) \in U\,,$$

and $U = (0, \infty)^2$ and $V = (0, 1) \times (0, \infty)$ are open sets. The inverse of ψ is $T : V \to U$ defined by

$$T(w, z) = (wz, z - wz)\,, (w, z) \in V\,.$$

The Jacobian matrix of T is

$$J(w, z) = \begin{bmatrix} z & w \\ -z & 1 - w \end{bmatrix}\,,$$

showing $|\det J(w,z)| = z$ *for* $(w,z) \in V$. *The joint density of* (X,Y) *is*

$$f(x,y) = \frac{1}{\Gamma(\alpha)\Gamma(\beta)} e^{-x-y} x^{\alpha-1} y^{\beta-1}\,, (x,y) \in U\,.$$

Theorem 4.8 shows that the joint density g *of* (W,Z) *at* $(w,z) \in V$ *is*

$$\begin{aligned} g(w,z) &= f \circ T(w,z) |\det J(w,z)| \\ &= \frac{1}{\Gamma(\alpha)\Gamma(\beta)} e^{-z} (wz)^{\alpha-1} (z-wz)^{\beta-1} z \\ &= \frac{1}{\Gamma(\alpha)\Gamma(\beta)} w^{\alpha-1}(1-w)^{\beta-1} e^{-z} z^{\alpha+\beta-1}\,, \end{aligned}$$

and $g(w,z) = 0$ *for* $(w,z) \notin V$. *In other words,*

$$g(w,z) = c\, h_1(w) h_2(z)\,, (w,z) \in \mathbb{R}^2\,,$$

where

$$h_1(w) = \frac{1}{B(\alpha,\beta)} w^{\alpha-1}(1-w)^{\beta-1} \mathbf{1}(0 < w < 1)\,, w \in \mathbb{R}\,,$$

$$h_2(z) = \frac{1}{\Gamma(\alpha+\beta)} e^{-z} z^{\alpha+\beta-1} \mathbf{1}(z > 0)\,, z \in \mathbb{R}\,,$$

and

$$c = \frac{B(\alpha,\beta)\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\,.$$

Since h_1 *and* h_2 *are densities of Beta*(α,β) *and Gamma*$(\alpha+\beta)$, *respectively, Theorem 4.7 shows that* $c = 1$ *and* W *and* Z *are independent with respective densities* h_1 *and* h_2. *In particular, this means* $X/(X+Y)$ *follows Beta*(α,β). *Furthermore,* $c = 1$ *reconfirms that*

$$B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}\,, \alpha,\beta > 0\,.$$

In the next result and elsewhere, a vector $x \in \mathbb{R}^n$, for $n = 2, 3, \ldots$, is to be thought of as an $n \times 1$ column vector, unless mentioned otherwise.

Theorem 4.9. *If* $X = (X_1, \ldots, X_d)$ *has density* f, A *is a* $d \times d$ *non-singular matrix and*

$$Y = AX + \mu\,,$$

for some fixed $\mu \in \mathbb{R}^d$, *then* $Y = (Y_1, \ldots, Y_d)$ *has density*

$$g(y) = \frac{1}{|\det(A)|} f\left(A^{-1}(y-\mu)\right)\,, y \in \mathbb{R}^d\,.$$

Proof. Follows immediately from Theorem 4.8 by observing that $Y = \psi(X)$ where $\psi : \mathbb{R}^d \to \mathbb{R}^d$ is a bijection defined by

$$\psi(x) = Ax + \mu \,, x \in \mathbb{R}^d \,,$$

that the inverse of ψ is T defined by

$$T(y) = A^{-1}(y - \mu) \,, y \in \mathbb{R}^d \,,$$

and that the Jacobian matrix of T is A^{-1}. □

The next result is a striking application of Theorem 4.8 which is very useful in statistics.

Theorem 4.10. *If $X_1, \ldots, X_n$ are i.i.d. from standard normal for $n \geq 2$, and*

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \,, \text{ and } S = \sum_{i=1}^{n}(X_i - \bar{X})^2 \,,$$

then $\bar{X}$ and S are independent.

Proof. Let P be an $n \times n$ orthogonal matrix whose first row is

$$\left[\frac{1}{\sqrt{n}} \cdots \frac{1}{\sqrt{n}}\right] \,;$$

such a P exists because the above is a vector of norm 1. Let

$$X = (X_1, \ldots, X_n),$$

which is to be thought of as a column vector by convention. Define

$$Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = PX \,.$$

The density of X is

$$f(x) = (2\pi)^{-n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{n} x_i^2\right) = (2\pi)^{-n/2} \exp\left(-\frac{1}{2}x^T x\right) \,,$$

for all $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$. Theorem 4.9 shows that the density of Y is

$$\begin{aligned} g(y) &= \frac{1}{|\det(P^T)|} f\left(P^T y\right) \\ \big(\text{since } \det(P^T) = \pm 1\big) &= (2\pi)^{-n/2} \exp\left(-\frac{1}{2}(P^T y)^T (P^T y)\right) \\ &= (2\pi)^{-n/2} \exp\left(-\frac{1}{2}y^T P P^T y\right) \\ &= (2\pi)^{-n/2} \exp\left(-\frac{1}{2}y^T y\right) \,, \end{aligned}$$

for $y \in \mathbb{R}^n$, the last line following from the fact that P is an orthogonal matrix. Thus,

$$g(y) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-y_i^2/2}\,, y = (y_1, \ldots, y_n) \in \mathbb{R}^n\,. \tag{4.7}$$

In other words, the joint density $g(y_1, \ldots, y_n)$ of $(Y_1, \ldots, Y_n)$ can be factorized as the product of the standard normal density evaluated at $y_1, \ldots, y_n$. Theorem 4.7 shows $Y_1, \ldots, Y_n$ are i.i.d. from standard normal.

The choice of the first row of P implies

$$Y_1 = \sqrt{n}\bar{X}\,. \tag{4.8}$$

Once again, P is an orthogonal matrix implies

$$\sum_{i=1}^{n} X_i^2 = \sum_{i=1}^{n} Y_i^2\,. \tag{4.9}$$

Write

$$\begin{aligned} S &= \sum_{i=1}^{n} \left(X_i^2 - 2X_i\bar{X} + (\bar{X})^2\right) \\ &= \sum_{i=1}^{n} X_i^2 - n(\bar{X})^2 \\ &= \sum_{i=1}^{n} Y_i^2 - Y_1^2 \\ &= \sum_{i=2}^{n} Y_i^2\,, \end{aligned}$$

(4.8) and (4.9) implying the penultimate line. As S is a function of $Y_2, \ldots, Y_n$ and $\bar{X}$ is a function of Y_1, the independence of $\bar{X}$ and S follows, which completes the proof. $\square$

Definition 38. *If $Z_1, \ldots, Z_n$ are i.i.d. from standard normal, the distribution of $\sum_{i=1}^n Z_i^2$ is called χ_n^2.*

Exercise 4.2. *Show that S, which is as in Theorem 4.10, has the χ_{n-1}^2 distribution.*

Exercise 4.3. *Let $X_1, \ldots, X_n$ be i.i.d. from standard normal, and*

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}.$$

Fix $\mu \in \mathbb{R}^n$ and let Σ be a $n \times n$ real symmetric positive definite (p.d.) matrix, that is, $\Sigma^T = \Sigma$ and $x^T \Sigma x > 0$ for all $x \in \mathbb{R}^n \setminus \{0\}$. Let $\Sigma^{1/2}$ be the p.d. square root of Σ, that is, $\Sigma^{1/2}$ is the unique p.d. matrix whose square is Σ. Define

$$Y = \mu + \Sigma^{1/2} X \,. \tag{4.10}$$

Show that the density of $Y = (Y_1, \ldots, Y_n)$ is

$$g(y) = \frac{1}{(2\pi)^{n/2} \det(\Sigma^{1/2})} \exp\left(-\frac{1}{2}(y-\mu)^T \Sigma^{-1}(y-\mu)\right), y \in \mathbb{R}^n \,.$$

Soln.: Follows from Theorem 4.9.

The density obtained in the above exercise is the density of the so-called multivariate normal distribution which is formally defined below. The next several results are devoted to understanding the properties of this distribution. Observing that $\det(\Sigma^{1/2}) = \sqrt{\det(\Sigma)}$, the following definition makes perfect sense.

Definition 39. *If $X = (X_1, \ldots, X_n)$ has density*

$$f(x) = \frac{1}{(2\pi)^{n/2} \sqrt{\det(\Sigma)}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right), x \in \mathbb{R}^n \,,$$

for some $\mu \in \mathbb{R}^n$ and $n \times n$ p.d. matrix Σ, then X follows n-dimensional multivariate normal distribution *with parameters μ and Σ, which is written as*

$$X \sim N_n(\mu, \Sigma) \,.$$

The interpretation of μ and Σ in the distribution $N_n(\mu, \Sigma)$ will be clear after a couple of results. The following theorem is essentially the converse of Exc 4.3.

Theorem 4.11. *If $X \sim N_n(\mu, \Sigma)$ and $(Y_1, \ldots, Y_n) = Y = \Sigma^{-1/2}(X - \mu)$, then $Y_1, \ldots, Y_n$ are i.i.d. from standard normal.*

Proof. The density of X is

$$f(x) = \frac{1}{(2\pi)^{n/2} \sqrt{\det(\Sigma)}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right), x \in \mathbb{R}^n$$

Writing

$$Y = \Sigma^{-1/2} X - \Sigma^{-1/2} \mu \,,$$

Theorem 4.9 with $A = \Sigma^{-1/2}$ and μ replaced by $-\Sigma^{-1/2}\mu$ therein implies that the density of Y is

$$\begin{aligned}
g(y) &= \frac{1}{\det(\Sigma^{-1/2})} f\left(\Sigma^{1/2}\left(y + \Sigma^{-1/2}\mu\right)\right) \\
&= \frac{1}{\det(\Sigma^{-1/2})} f\left(\Sigma^{1/2} y + \mu\right) \\
&= (2\pi)^{-n/2} \exp\left(-\frac{1}{2}\left(\Sigma^{1/2}y\right)^T \Sigma^{-1}\left(\Sigma^{1/2}y\right)\right) \\
&= (2\pi)^{-n/2} \exp\left(-\frac{1}{2} y^T y\right),
\end{aligned}$$

$\det(\Sigma^{-1/2}) = (\det(\Sigma))^{-1/2}$ being used in the penultimate line, and the last line follows from the fact that $\Sigma^{1/2}$ is symmetric and

$$\Sigma^{1/2}\Sigma^{-1}\Sigma^{1/2} = I\,.$$

Like in (4.7) and the subsequent argument, Theorem 4.7 shows $Y_1, \ldots, Y_n$ are i.i.d. from standard normal, which completes the proof. □

The next theorem shows that if $X \sim N_n(\mu, \Sigma)$, then μ and Σ are the "mean vector" and the "covariance matrix" of X, respectively.

Theorem 4.12. *If* $X \sim N_n(\mu, \Sigma)$ *where*

$$\mu = (\mu_1, \ldots, \mu_n) \text{ and } \Sigma = ((\sigma_{ij}))_{1\le i,j\le n}\,,$$

then

$$\begin{aligned}
\mathrm{E}(X_i) &= \mu_i\,, i = 1, \ldots, n\,, \\
\mathrm{Cov}(X_i, X_j) &= \sigma_{ij}\,, 1 \le i, j \le n\,.
\end{aligned}$$

In particular, $\mathrm{Var}(X_i) = \sigma_{ii}$ *for* $i = 1, \ldots, n$.

Proof. Let $(Y_1, \ldots, Y_n) = Y = \Sigma^{-1/2}(X - \mu)$; $Y_1, \ldots, Y_n$ are i.i.d. from standard normal by Theorem 4.11. Rewrite the above as

$$X = \mu + \Sigma^{1/2} Y\,,$$

or

$$X_i = \mu_i + \sum_{j=1}^n \theta_{ij} Y_j\,, i = 1, \ldots, n\,,$$

where $\Sigma^{1/2} = ((\theta_{ij}))_{1\le i,j\le n}$. Since $Y_1, \ldots, Y_n$ are zero mean random variables, it immediately follows $\mathrm{E}(X_i) = \mu_i$ for $i = 1, \ldots, n$. Theorem 3.6 shows that for fixed $1 \leq i, j \leq n$,

$$\begin{aligned}
\mathrm{Cov}(X_i, X_j) &= \mathrm{Cov}\left(\sum_{k=1}^{n} \theta_{ik} Y_k, \sum_{l=1}^{n} \theta_{jl} Y_l\right) \\
&= \sum_{k=1}^{n} \sum_{l=1}^{n} \theta_{ik} \theta_{jl} \mathrm{Cov}\left(Y_k, Y_l\right) \\
&= \sum_{k=1}^{n} \theta_{ik} \theta_{jk} \, ,
\end{aligned}$$

the last line following from the fact that $Y_1, \ldots, Y_n$ are independent and each has variance one. Recalling that θ_{ik} is the (i,k)-th entry of $\Sigma^{1/2}$ which is a symmetric matrix, write

$$\begin{aligned}
\sum_{k=1}^{n} \theta_{ik} \theta_{jk} &= \sum_{k=1}^{n} \theta_{ik} \theta_{kj} \\
&= (i,j)\text{-th entry of } \Sigma^{1/2} \Sigma^{1/2} \\
&= (i,j)\text{-th entry of } \Sigma \\
&= \sigma_{ij} \, .
\end{aligned}$$

It thus follows that

$$\mathrm{Cov}(X_i, X_j) = \sigma_{ij} \, , 1 \leq i, j \leq n \, .$$

Taking $i = j$ implies $\mathrm{Var}(X_i) = \sigma_{ii}$ and completes the proof. □

Exercise 4.4. *If $Y_1, \ldots, Y_n$ are random variables, each having variance one, such that*

$$\mathrm{Cov}(Y_i, Y_j) = 0 \, , 1 \leq i < j \leq n \, ,$$

A is an $m \times n$ matrix and

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_m \end{bmatrix} = A \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} ,$$

show that the covariance matrix of X is AA^T.

The next theorem is consistent with the above exercise .

Theorem 4.13. *If $X \sim N_n(0, I)$, where I is the $n \times n$ identity matrix, then for any $m \times n$ matrix B with* $\mathrm{Rank}(B) = m$ *and $\mu \in \mathbb{R}^m$,*

$$BX + \mu \sim N_m\left(\mu, BB^T\right) \, .$$

Proof. Let us first deal with the case $m = n$. In this case, B is a non-singular matrix as $\text{Rank}(B) = m$. The density of X is

$$f(x) = (2\pi)^{-n/2} \exp\left(-\frac{1}{2}x^T x\right), x \in \mathbb{R}^n .$$

Theorem 4.9 implies that the density of $Y = BX + \mu$ is

$$\begin{aligned} g(y) &= \frac{1}{|\det(B)|} f\left(B^{-1}(y-\mu)\right) \\ &= \frac{1}{|\det(B)|(2\pi)^{n/2}} \exp\left(-\frac{1}{2}(B^{-1}(y-\mu))^T (B^{-1}(y-\mu))\right) \\ &= \frac{1}{|\det(B)|(2\pi)^{n/2}} \exp\left(-\frac{1}{2}(y-\mu)^T (BB^T)^{-1}(y-\mu)\right) \\ &= \frac{1}{(2\pi)^{n/2}\sqrt{\det(\Sigma)}} \exp\left(-\frac{1}{2}(y-\mu)^T \Sigma^{-1}(y-\mu)\right), \end{aligned}$$

where $\Sigma = BB^T$ is a p.d. matrix because B is non-singular and therefore $\sqrt{\det(\Sigma)} = |\det(B)|$. This shows the stated claim when $m = n$.

Now assume $1 \le m < n$. Since $\text{Rank}(B) = m$, the rows of B form a basis of the row space of B. Let C be a matrix whose rows form a basis of the orthogonal complement of the row space of B. In other words, C is an $(n-m) \times n$ matrix with $\text{Rank}(C) = n - m$ and

$$BC^T = 0 . \tag{4.11}$$

Define $Y_1, \ldots, Y_n$ by

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = AX + \tilde{\mu} ,$$

where $A = \left[\begin{smallmatrix} B \\ C \end{smallmatrix}\right]$ is an $n \times n$ non-singular matrix and $\tilde{\mu}$ is an $n \times 1$ vector defined by $\tilde{\mu} = \left[\begin{smallmatrix} \mu \\ 0 \end{smallmatrix}\right]$. Using the result for the case $m = n$, which has already been shown, it follows that

$$(Y_1, \ldots, Y_n) \sim N_n\left(\tilde{\mu}, AA^T\right) . \tag{4.12}$$

Using (4.11) to write

$$AA^T = \begin{bmatrix} BB^T & 0 \\ 0 & CC^T \end{bmatrix} , \tag{4.13}$$

it is immediate that

$$\left(AA^T\right)^{-1} = \begin{bmatrix} \left(BB^T\right)^{-1} & 0 \\ 0 & \left(CC^T\right)^{-1} \end{bmatrix} .$$

Thus for $y^{(1)} \in \mathbb{R}^m$ and $y^{(2)} \in \mathbb{R}^{n-m}$, which are column vectors by convention, and letting

$$y = \begin{bmatrix} y^{(1)} \\ y^{(2)} \end{bmatrix}, \tag{4.14}$$

we get

$$\begin{aligned} &(y - \tilde{\mu})^T \left(AA^T\right)^{-1} (y - \tilde{\mu}) \\ &= (y^{(1)} - \mu)^T \left(BB^T\right)^{-1} (y^{(1)} - \mu) + y^{(2)} \left(CC^T\right)^{-1} y^{(2)}. \end{aligned} \tag{4.15}$$

Recall (4.12) to write the density of $(Y_1, \ldots, Y_n)$ as

$$\begin{aligned} g(y) &= \frac{1}{(2\pi)^{n/2}\sqrt{\det(AA^T)}} \exp\left(-\frac{1}{2}(y - \tilde{\mu})^T \left(AA^T\right)^{-1} (y - \tilde{\mu})\right) \\ &= \frac{1}{(2\pi)^{n/2}\sqrt{\det(AA^T)}} \exp\left(-\frac{1}{2}(y^{(1)} - \mu)^T \left(BB^T\right)^{-1} (y^{(1)} - \mu)\right) \\ &\qquad \times \exp\left(-\frac{1}{2}y^{(2)} \left(CC^T\right)^{-1} y^{(2)}\right), \end{aligned}$$

by (4.15), where y is partitioned as in (4.14). Use (4.13) to write

$$\det(AA^T) = \det(BB^T)\det(CC^T),$$

which allows simplifying g to

$$\begin{aligned} g(y) &= \frac{1}{(2\pi)^{m/2}\sqrt{\det(BB^T)}} \exp\left(-\frac{1}{2}(y^{(1)} - \mu)^T \left(BB^T\right)^{-1} (y^{(1)} - \mu)\right) \\ &\qquad \times \frac{1}{(2\pi)^{(n-m)/2}\sqrt{\det(CC^T)}} \exp\left(-\frac{1}{2}y^{(2)} \left(CC^T\right)^{-1} y^{(2)}\right) \\ &= g_1\left(y^{(1)}\right) g_2\left(y^{(2)}\right), \end{aligned}$$

where g_1 and g_2 are the densities of $N_m(\mu, BB^T)$ and $N_{n-m}(0, CC^T)$, respectively. Exc 4.1 shows that $(Y_1, \ldots, Y_m)$ and $(Y_{m+1}, \ldots, Y_n)$ are independent from $N_m(\mu, BB^T)$ and $N_{n-m}(0, CC^T)$, respectively. Since

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_m \end{bmatrix} = BX + \mu,$$

a restatement of the former is that

$$BX + \mu \sim N_m(\mu, BB^T),$$

which completes the proof. $\square$

Theorem 4.14. *If $X \sim N_n(\mu, \Sigma)$ and B is an $m \times n$ matrix with* $\text{Rank}(B) = m$, *then*
$$BX \sim N_m\left(B\mu, B\Sigma B^T\right) .$$

Proof. Let $(Y_1, \ldots, Y_n) = Y = \Sigma^{-1/2}(X - \mu)$. Theorem 4.11 shows $Y_1, \ldots$ $\ldots, Y_n$ are i.i.d. from standard normal, that is, $Y \sim N_n(0, I)$.

Write

$$BX = B(X-\mu) + B\mu = B\Sigma^{1/2}\Sigma^{-1/2}(X-\mu) + B\mu = AY + B\mu ,$$

where $A = B\Sigma^{1/2}$. Since A is an $m \times n$ matrix with $\text{Rank}(A) = m$ because $\Sigma^{1/2}$ is non-singular, Theorem 4.13 shows

$$AY + B\mu \sim N_m\left(B\mu, AA^T\right) .$$

Observing that
$$AA^T = B\Sigma B^T ,$$
the proof follows. □

An immediate corollary of the above theorem is the following.

Corollary 4.1. *If $X_1, \ldots, X_n$ are independent and $X_i \sim N(\mu_i, \sigma_i^2)$ for $i = 1, \ldots, n$, then*
$$\sum_{i=1}^n X_i \sim N\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right) .$$

Proof. Follows from Theorem 4.14 by taking B to be the $1 \times n$ matrix $[1 \ldots 1]$ and observing that
$$(X_1, \ldots, X_n) \sim N_n(\mu, \Sigma) ,$$
where $\mu = (\mu_1, \ldots, \mu_n)$, $\Sigma = ((\sigma_{ij}))_{1 \le i,j \le n}$ and

$$\sigma_{ij} = \begin{cases} \sigma_i^2 , & i = j , \\ 0 , & i \neq j . \end{cases}$$

□

Exercise 4.5. *If (X, Y) follows bivariate normal, that is,*

$$(X, Y) \sim N_2(\mu, \Sigma)$$

for some $\mu \in \mathbb{R}^2$ and 2×2 p.d. matrix Σ, show that

$$\text{Corr}(X, Y) = 0 \iff X, Y \text{ are independent.}$$

Show that the above equivalence fails if each of X and Y follows normal but (X, Y) is not necessarily bivariate normal.

Exercise 4.6. *If $X \sim N_n(0, \Sigma)$, show that*

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i^2\right) = 2\,\mathrm{Tr}\left(\Sigma^2\right)\,,$$

where $\mathrm{Tr}(\cdot)$ denotes the trace of a square matrix.
Hint. *The spectral theorem for real symmetric matrices implies $\Sigma = PDP^T$ for some orthogonal matrix P and diagonal matrix D. Define $Y = P^T X$ and show that*

$$Y^T Y = X^T X\,.$$

Exercise 4.7. *Suppose $(X_1, \ldots, X_n) \sim N_n(\mu, \Sigma)$ where $\mu = (\mu_1, \ldots, \mu_n)$ and $\Sigma = ((\sigma_{ij}))_{1 \le i,j \le n}$.*

1. *Show that $X_i \sim N(\mu_i, \sigma_{ii})$ for $i = 1, \ldots, n$.*
2. *If $i_1, \ldots, i_k \in \{1, \ldots, n\}$ are distinct, show that*

$$(X_{i_1}, \ldots, X_{i_k}) \sim N_k\left((\mu_{i_j})_{1 \le j \le k}, ((\sigma_{i_j, i_{j'}}))_{1 \le j,j' \le k}\right)\,.$$

3. *For $A, B \subset \{1, \ldots, n\}$ with $A \neq \emptyset$, $B \neq \emptyset$ and $A \cap B = \emptyset$, show that*

$$(X_i : i \in A) \text{ and } (X_i : i \in B) \text{ are independent}$$

if and only if

$$\mathrm{Cov}(X_i, X_j) = 0 \text{ for all } i \in A, j \in B\,.$$

Exercise 4.8. *Suppose X and Y are i.i.d. from standard normal. Let $\rho \in (-1, 1)$ and set*

$$Z = \rho X + \sqrt{1 - \rho^2}\, Y\,.$$

Show that

$$(X, Z) \sim N_2\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right)\,.$$

Exercise 4.9. *If $X_1, \ldots, X_n$ are i.i.d. from standard normal, P is an $m \times n$ matrix with $1 \le m < n$ and $PP^T = I_m$, and*

$$(Y_1, \ldots, Y_m) = Y = PX\,,$$

show that

$$\sum_{i=1}^{m} Y_i^2 \sim \chi^2_m\,,$$

$$\sum_{i=1}^{n} X_i^2 - \sum_{i=1}^{m} Y_i^2 \sim \chi^2_{n-m}\,,$$

and

$$\sum_{i=1}^{m} Y_i^2\,,\ \sum_{i=1}^{n} X_i^2 - \sum_{i=1}^{m} Y_i^2 \text{ are independent.}$$

The last topic to be studied in this chapter is order statistics. We start with defining the same.

Definition 40. *The* ascending sort map *is a map $T : \mathbb{R}^n \to \mathbb{R}^n$ which sorts the entries of a vector in ascending order, that is, $T(x_1, \ldots, x_n) = (y_1, \ldots, y_n)$ means $y_1 \le \ldots \le y_n$ and $(y_1, \ldots, y_n)$ is a permutation of $(x_1, \ldots, x_n)$ for all $(x_1, \ldots, x_n) \in \mathbb{R}^n$. For random variables $X_1, \ldots, X_n$, their* order statistics *$X_{(1)}, \ldots, X_{(n)}$ are defined by*

$$\left(X_{(1)}, \ldots, X_{(n)}\right) = T(X_1, \ldots, X_n) .$$

The ascending order map is a continuous function from $\mathbb{R}^n$ to $\mathbb{R}^n$ and thus Borel measurable. Therefore, if $X_1, \ldots, X_n$ are random variables, which by convention are defined on the same probability space, their order statistics are random variables as well.

Theorem 4.15. *If $X_1, \ldots, X_n$ are i.i.d. from some density f, then the density of their order statistics $(X_{(1)}, \ldots, X_{(n)})$ is*

$$g(x_1, \ldots, x_n) = \begin{cases} n! f(x_1) \ldots f(x_n) , & x_1 < x_2 < \ldots < x_n , \\ 0 , & \textit{otherwise} . \end{cases}$$

The proof uses the following exercise which is a special case of Exc 1.7.

Exercise 4.10. *Suppose P, Q are finite measures on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ such that for some open set U,*

$$P(U^c) = 0 = Q(U^c) .$$

If $P(R) = Q(R)$ for all $R = (a_1, b_1] \times \ldots \times (a_n, b_n] \subset U$ with $-\infty < a_i < b_i < \infty$, $i = 1, \ldots, n$, show that P and Q agree on $\mathcal{B}(\mathbb{R}^n)$.

Proof of Theorem 4.15. Letting $U = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : x_1 < \ldots < x_n\}$, the proof would follow from the above exercise once it is shown that

$$P\left((X_{(1)}, \ldots, X_{(n)}) \in U^c\right) = 0 = \int_{U^c} g(x_1, \ldots, x_n)\, dx_1 \ldots dx_n , \tag{4.16}$$

and

$$P\left((X_{(1)}, \ldots, X_{(n)}) \in R\right) = \int_R g(x_1, \ldots, x_n)\, dx_1 \ldots dx_n , \tag{4.17}$$

for all

$$R = (a_1, b_1] \times \ldots \times (a_n, b_n] \subset U \text{ with } -\infty < a_i < b_i < \infty , i = 1, \ldots, n . \tag{4.18}$$

Since $X_1, \ldots, X_n$ are i.i.d. from a density f, $(X_1, \ldots, X_n)$ has a density h given by

$$h(x_1, \ldots, x_n) = f(x_1) \ldots f(x_n) , x_1, \ldots, x_n \in \mathbb{R} .$$

Since $\{(x_1, \ldots, x_n) : x_i = x_j \text{ for some } 1 \le i < j \le n\}$ is a finite union of lower dimension subspaces of $\mathbb{R}^n$, it is a set of zero Lebesgue measure and hence the integral of h on that set is zero. In other words,

$$P(X_i = X_j \text{ for some } 1 \le i < j \le n) = 0 .$$

Consequently,

$$P(X_{(i)} = X_{(j)} \text{ for some } 1 \le i < j \le n) = 0 .$$

This along with the obvious fact that $X_{(1)} \le \ldots \le X_{(n)}$ shows that

$$P\left((X_{(1)}, \ldots, X_{(n)}) \in U^c\right) = 0 .$$

That

$$\int_{U^c} g(x_1, \ldots, x_n)\, dx_1 \ldots dx_n = 0$$

follows tautologically from the definition of g. That is, (4.16) holds.

For (4.17), fix R as in (4.18). An immediate consequence of $R \subset U$ is that $a_1 < b_1 \le a_2 < b_2 \le \ldots \le a_n < b_n$ and hence

$$(a_i, b_i] \cap (a_j, b_j] = \emptyset\,, 1 \le i < j \le n . \tag{4.19}$$

Therefore,

$$\begin{aligned}
&P\left((X_{(1)}, \ldots, X_{(n)}) \in R\right) \\
&= P\left(a_i < X_{(i)} \le b_i\,, i = 1, \ldots, n\right) \\
&= P\left(\bigcup_{\pi \text{ permutation of } \{1,\ldots,n\}} \left[a_i < X_{\pi(i)} \le b_i\,, i = 1, \ldots, n\right]\right) \\
&= \sum_{\pi \text{ permutation of } \{1,\ldots,n\}} P\left(a_i < X_{\pi(i)} \le b_i\,, i = 1, \ldots, n\right) ,
\end{aligned}$$

the penultimate line following from the fact that $X_{(1)}(\omega), \ldots, X_{(n)}(\omega)$ is a permutation of $X_1(\omega), \ldots, X_n(\omega)$ for every $\omega \in \Omega$, and the last line follows from the observation that for distinct permutations π and π',

$$\left[a_i < X_{\pi(i)} \le b_i\,, i = 1, \ldots, n\right] \cap \left[a_i < X_{\pi'(i)} \le b_i\,, i = 1, \ldots, n\right] = \emptyset\,,$$

which is another consequence of (4.19).

For a fixed permutation π, the independence of $X_{\pi(1)}, \ldots, X_{\pi(n)}$ implies

$$\begin{aligned}
P\left((X_{(1)}, \ldots, X_{(n)}) \in R\right) &= \sum_{\pi \text{ permutation of } \{1,\ldots,n\}} \prod_{i=1}^n P\left(a_i < X_{\pi(i)} \le b_i\right) \\
\left(X_{\pi(i)} \stackrel{d}{=} X_1\,, i = 1, \ldots, n\right) &= \sum_{\pi \text{ permutation of } \{1,\ldots,n\}} \prod_{i=1}^n \int_{a_i}^{b_i} f(x)\,dx \\
&= n! \prod_{i=1}^n \int_{a_i}^{b_i} f(x)\,dx \\
&= n! \int_{a_1}^{b_1} \ldots \int_{a_n}^{b_n} f(x_1) \ldots f(x_n)\,dx_n \ldots dx_1 \\
&= \int_R g(x_1, \ldots, x_n)\,dx_1 \ldots dx_n\,.
\end{aligned}$$

Thus (4.17) holds for every R as in (4.18). This completes the proof. $\square$

Exercise 4.11. *If $(X_1, \ldots, X_n)$ is a random vector in $\mathbb{R}^n$ with joint density f and $(X_{(1)}, \ldots, X_{(n)})$ is its order statistic, show that the density of the latter is*

$$g(x_1, \ldots, x_n) = \begin{cases} \sum_\pi f\left(x_{\pi(1)}, \ldots, x_{\pi(n)}\right), & x_1 < \ldots < x_n, \\ 0, & \textit{otherwise}, \end{cases}$$

where the sum is over all permutations of $\{1, \ldots, n\}$.

5 Conditional expectation

Following the usual convention, $(\Omega, \mathcal{A}, P)$ is the probability space underlying everything we talk about, unless specifically mentioned otherwise. As in Definition 24, for $A, B \in \mathcal{A}$ with $P(A) > 0$, the conditional probability of B given that A has occurred is

$$P(B|A) = \frac{P(B \cap A)}{P(A)}\,.$$

Suppose now that $0 < P(A) < 1$ and we want to define the conditional probability of B given that we know whether A has occurred or not. If A has occurred, then the above is the natural definition of the said conditional probability, whereas it is

$$\frac{P(B \cap A^c)}{P(A^c)}$$

if A has not occurred, that is, A^c has occurred. In other words,

$$\frac{P(B \cap A)}{P(A)}\mathbf{1}_A + \frac{P(B \cap A^c)}{P(A^c)}\mathbf{1}_{A^c} \tag{5.1}$$

is a natural definition of the conditional probability of B given the knowledge of whether A has occurred or not. The said conditional probability is thus a random variable depending on $\mathbf{1}_A$ and $\mathbf{1}_{A^c}$ alone. To generalize this, if $A_1, A_2, A_3, \ldots$ are mutually and exhaustive events of positive probability, that is,

$$A_i \cap A_j = \emptyset \text{ for all } i \neq j\,, \quad \bigcup_{i=1}^{\infty} A_i = \Omega\,, \quad \text{ and } P(A_i) > 0\,, i = 1, 2, \ldots\,, \tag{5.2}$$

then a similar reasoning says that the conditional probability of B given the knowledge which one of $A_1, A_2, \ldots$ has occurred is

$$Z = \sum_{i=1}^{\infty} \frac{P(B \cap A_i)}{P(A_i)} \mathbf{1}_{A_i}\,,$$

which should be thought of as a generalization of (5.1).

Thus, Z is a random variable depending on $\mathbf{1}_{A_1}, \mathbf{1}_{A_2}, \ldots,$ alone. Furthermore, if $E \in \sigma(\{A_1, A_2, \ldots\})$, then (5.2) shows

$$E = \bigcup_{i \in N} A_i \text{ for some } N \subset \mathbb{N}\,.$$

Therefore,

$$\begin{aligned} P(B \cap E) &= \sum_{i \in N} P(B \cap A_i) \\ \left(\int_{A_i} Z\, dP = \frac{P(B \cap A_i)}{P(A_i)} \int_{A_i} dP = P(B \cap A_i)\right) &= \sum_{i \in N} \int_{A_i} Z\, dP \\ &= \int_E Z\, dP\,. \end{aligned}$$

The conditional probability of B, given the knowledge of which one of $A_1, A_2, \ldots$ has occurred, is thus a random variable Z that is $\sigma(\{A_1, A_2, \ldots\})$-measurable and satisfies

$$\int_E Z\, dP = P(B \cap E) \text{ for all } E \in \sigma(\{A_1, A_2, \ldots\})\,.$$

Since the right hand side is the same as $\int_E \mathbf{1}_B\, dP$, and interpreting the conditional probability of B as the "conditional expectation" of $\mathbf{1}_B$, a natural candidate for the latter is thus a $\sigma(\{A_1, A_2, \ldots\})$-measurable random variable Z satisfying

$$\int_E Z\, dP = \int_E \mathbf{1}_B\, dP \text{ for all } E \in \sigma(\{A_1, A_2, \ldots\})\,.$$

Reasoning along similar lines, for an integrable random variable X and a σ-field $\mathcal{F} \subset \mathcal{A}$, the conditional expectation of X given $\mathcal{F}$ should be defined as an $\mathcal{F}$-measurable random variable Z which satisfies

$$\int_E Z\,dP = \int_E X\,dP \text{ for all } E \in \mathcal{F}\,.$$

The following theorem guarantees the existence of such Z and its uniqueness upto zero probability sets.

Theorem 5.1. *For an integrable random variable X and a σ-field $\mathcal{F} \subset \mathcal{A}$, there exists an integrable random variable Z which is $\mathcal{F}$-measurable and satisfies*

$$\int_E Z\,dP = \int_E X\,dP \textit{ for all } E \in \mathcal{F}\,. \tag{5.3}$$

If Z' is another $\mathcal{F}$-measurable and integrable random variable such that the above holds with Z replaced by Z', then $Z' = Z$ a.s.

Proof. Write $X = X^+ - X^-$ where $X^+ = X \vee 0$ and $X^- = (-X) \vee 0$. Since X is integrable, so are X^+ and X^-. Define measures μ_+ and μ_- on $(\Omega, \mathcal{F})$ by

$$\mu^+(E) = \int_E X^+\,dP\,, \mu^-(E) = \int_E X^-\,dP \text{ for all } E \in \mathcal{F}\,.$$

Thus μ^+ and μ^- are finite measures on $(\Omega, \mathcal{F})$ and each of them is absolutely continuous with respect to P. Theorem 1.7, which is the Radon-Nikodym theorem, implies there exist $\mathcal{F}$-measurable functions Z_1 and Z_2 from Ω to $[0, \infty)$ satisfying

$$\int_E Z_1\,dP = \mu_+(E)\,, \text{ and } \int_E Z_2\,dP = \mu_-(E)\,, \text{ for all } E \in \mathcal{F}\,.$$

Letting $Z = Z_1 - Z_2$, (5.3) clearly holds.

If Z' is another $\mathcal{F}$-measurable and integrable random variable such that (5.3) holds with Z replaced by Z', then it would follow that

$$\int_E (Z - Z')\,dP = 0\,, E \in \mathcal{F}\,.$$

Since $Z - Z'$ is $\mathcal{F}$-measurable, the above implies $Z - Z' = 0$ a.s. This completes the proof. □

Definition 41. *For an integrable random variable X and a σ-field $\mathcal{F} \subset \mathcal{A}$, the* conditional expectation *of X given $\mathcal{F}$, denoted by* $\mathrm{E}(X|\mathcal{F})$*, is an integrable random variable Z which is $\mathcal{F}$-measurable and satisfies* (5.3)*. Theorem 5.1 guarantees the existence of such Z and its uniqueness upto sets of zero probability.*

While the above definition is motivated by (5.1), the same can be arrived at by another route of reasoning. Recall that for any $X \in L^2(\Omega)$,

$$\mathrm{E}(X) = \arg\min_{\alpha \in \mathbb{R}} \mathrm{E}\left[(X-\alpha)^2\right] \,,$$

that is, $\mathrm{E}(X)$ is the unique $\alpha \in \mathbb{R}$ at which the right hand side is minimized. When no additional information is available, minimization over the set of real numbers makes sense. However, if for a σ-field $\mathcal{F} \subset \mathcal{A}$, we know for each $E \in \mathcal{F}$ whether it has occurred or not, then the class should be expanded to all $\mathcal{F}$-measurable functions, because any $\mathcal{F}$-measurable function is now "known". The following result makes this idea precise.

Theorem 5.2. *If* $\mathrm{E}(X^2) < \infty$*, and* $\mathcal{F} \subset \mathcal{A}$ *is a* σ*-field, then almost surely,*

$$\mathrm{E}(X|\mathcal{F}) = \arg\min \left\{ \int (X-Y)^2 \, dP : Y \text{ is measurable with respect to } \mathcal{F} \right\} .$$

Proof. Recall that $L^2(\Omega, \mathcal{A}, P)$ is a Hilbert space of which $L^2(\Omega, \mathcal{F}, P)$ is a subspace. Further, $L^2(\Omega, \mathcal{F}, P)$ is a complete metric space. As a consequence, $L^2(\Omega, \mathcal{F}, P)$ is a closed subspace of $L^2(\Omega, \mathcal{A}, P)$. Since $X \in L^2(\Omega, \mathcal{A}, P)$, it has a projection onto $L^2(\Omega, \mathcal{F}, P)$, which we call Z. In other words,

$$Z = \arg\min \left\{ \int (X-Y)^2 \, dP : Y \in L^2(\Omega, \mathcal{F}, P) \right\} . \tag{5.4}$$

Our first task is to show that the above Z actually minimizes the L^2 distance from X over all $\mathcal{F}$-measurable functions. Indeed, for a random variable Y with $\mathrm{E}(Y^2) = \infty$, it holds that

$$\int (X-Y)^2 \, dP = \infty$$

because otherwise $Y = X - (X - Y)$ would be in L^2 as X is in L^2. Since $Z \in L^2(\Omega, \mathcal{F}, P)$,

$$\int (X-Z)^2 \, dP < \infty = \int (X-Y)^2 \, dP \,, \text{ if } \mathrm{E}(Y^2) = \infty \,.$$

Thus,

$$Z = \arg\min \left\{ \int (X-Y)^2 \, dP : Y \text{ is measurable with respect to } \mathcal{F} \right\} .$$

To complete the proof, all that remains to show is $Z = \mathrm{E}(X|\mathcal{F})$ a.s. Since Z is $\mathcal{F}$-measurable, this would follow once (5.3) is shown to hold for this Z.

Results in functional analysis show that Z as in (5.4) is an orthogonal projection onto $L^2(\Omega, \mathcal{F}, P)$. That is, $X - Z$ belongs to the orthogonal complement of $L^2(\Omega, \mathcal{F}, P)$. Since $\mathbf{1}_E \in L^2(\Omega, \mathcal{F}, P)$ for any $E \in \mathcal{F}$, it thus follows that

$$\int (X - Z)\mathbf{1}_E \, dP = 0 \,,$$

which is the same as (5.3). Hence the proof follows. □

The inadequacy of the statement of Theorem 5.2 as a definition of conditional expectation is that it works only for $L^2(\Omega)$, whereas Definition 41 is valid on $L^1(\Omega)$, which is a superset of $L^2(\Omega)$ because P is a finite measure. Therefore, the conditional expectation of any integrable random variable will be as in Definition 41.

The following theorem is in line with our intuition that given a σ-field $\mathcal{F}$, any $\mathcal{F}$-measurable random variable is a known quantity, and hence should come outside the conditional expectation just like a constant comes out of an expectation. Throughout this chapter $\mathcal{F}$ is a σ-field with $\mathcal{F} \subset \mathcal{A}$, unless mentioned otherwise.

Theorem 5.3. *If X and Y are random variables such that Y and XY are integrable and X is $\mathcal{F}$-measurable, then*

$$\mathrm{E}\,(XY|\mathcal{F}) = X\,\mathrm{E}(Y|\mathcal{F}) \ \ a.s.$$

Proof. Let us first assume $Y \geq 0$ and μ be a finite measure on $(\Omega, \mathcal{A})$ defined by

$$\mu(E) = \int_E Y \, dP \,, E \in \mathcal{A} \,.$$

As shown in the proof of Theorem 5.1, $Z = \mathrm{E}(Y|\mathcal{F})$ is simply the Radon-Nikodym derivative of μ with respect to P on $(\Omega, \mathcal{F})$. Exc 1.10 and the fact that X is $\mathcal{F}$-measurable show that

$$\begin{aligned} \int_{(\Omega,\mathcal{F})} |X| Z \, dP &= \int_{(\Omega,\mathcal{F})} |X| \, d\mu \\ \text{(by Exc 1.8)} &= \int_{(\Omega,\mathcal{A})} |X| \, d\mu \\ &= \int_{(\Omega,\mathcal{A})} |X| Y \, dP < \infty \,, \end{aligned}$$

the equality in the last line is implied by the fact that Y is the Radon-Nikodym derivative of μ with respect to P on $(\Omega, \mathcal{A})$ and the inequality follows from the hypothesis that XY is integrable. Thus XZ is P-integrable.

Thus, for all $E \in \mathcal{F}$, $XZ\mathbf{1}_E$ is P-integrable, showing by a similar argument that

$$\int_{(\Omega,\mathcal{F})} X\mathbf{1}_E Z \, dP = \int_{(\Omega,\mathcal{F})} X\mathbf{1}_E \, d\mu \,, \tag{5.5}$$

because $X\mathbf{1}_E$ is $\mathcal{F}$-measurable. A similar argument shows

$$\int_{(\Omega,\mathcal{A})} X\mathbf{1}_E Y\, dP = \int_{(\Omega,\mathcal{A})} X\mathbf{1}_E\, d\mu\,. \tag{5.6}$$

Once again, the right hand sides of (5.5) and (5.6) are equal by Exc 1.8. This shows

$$\int_E XZ\, dP = \int_E XY\, dP\,, E \in \mathcal{F}\,.$$

Since XZ is $\mathcal{F}$-measurable, it follows that $XZ = \mathrm{E}(XY|\mathcal{F})$ which completes the proof for the case $Y \geq 0$. For the general case, the proof follows from similar arguments by splitting $Y = Y^+ - Y^-$. □

Theorem 5.4. *If X and Y are integrable random variables, then the following hold.*

1. *The random variable $X + Y$ is integrable and*

$$\mathrm{E}(X + Y|\mathcal{F}) = \mathrm{E}(X|\mathcal{F}) + \mathrm{E}(Y|\mathcal{F}) \text{ a.s.}$$

2. *For $\alpha \in \mathbb{R}$,*

$$\mathrm{E}(\alpha X|\mathcal{F}) = \alpha \mathrm{E}(X|\mathcal{F}) \text{ a.s.}$$

3. *If $X \geq 0$ a.s., then*

$$\mathrm{E}(X|\mathcal{F}) \geq 0 \text{ a.s.}$$

4. *If $X \leq Y$ a.s., then*

$$\mathrm{E}(X|\mathcal{F}) \leq \mathrm{E}(Y|\mathcal{F}) \text{ a.s.}$$

Proof.
1. Follows from the definition of conditional expectation.
2. A special case of Theorem 5.3, though it follows from the definition as well.
3. Follows from the definition of conditional expectation.
4. Implied by 1.-3. above by the following arguments:

$$\begin{aligned} &\mathrm{E}(Y|\mathcal{F}) - \mathrm{E}(X|\mathcal{F}) \\ \text{(by 1. and 2.)} &= \mathrm{E}(Y - X|\mathcal{F}) \\ \text{(by 3.)} &\geq 0 \text{ a.s.} \end{aligned}$$

□

The following is the so-called tower property and is in line with the intuition that conditional expectation given $\mathcal{F}$ of an L^2 random variable is projection onto $L^2(\Omega, \mathcal{F}, P)$ as in Theorem 5.2.

Theorem 5.5 (Tower property)**.** *If $\mathcal{F} \subset \mathcal{G} \subset \mathcal{A}$ and $\mathcal{F}, \mathcal{G}$ are σ-fields, then for any integrable X,*

$$\mathrm{E}\left(\mathrm{E}(X|\mathcal{G})\middle|\mathcal{F}\right) = \mathrm{E}(X|\mathcal{F}) \ \textit{a.s.}$$

Proof. Let $Y = \mathrm{E}(X|\mathcal{G})$ and $Z = \mathrm{E}(X|\mathcal{F})$. Then for any $A \in \mathcal{F}$,

$$\begin{aligned} \int_A Z\,dP &= \int_A X\,dP \\ (Y = \mathrm{E}(X|\mathcal{G}) \text{ and } A \in \mathcal{F} \subset \mathcal{G}) &= \int_A Y\,dP\,. \end{aligned}$$

Since Z is $\mathcal{F}$-measurable and the above holds for all $A \in \mathcal{F}$, we get

$$Z = \mathrm{E}(Y|\mathcal{F}) \text{ a.s.}\,,$$

which is precisely the claim of the theorem. □

The following special case of the tower property deserves special mention.

Corollary 5.1. *For an integrable X and a σ-field $\mathcal{G} \subset \mathcal{A}$,*

$$\mathrm{E}\big(\mathrm{E}\,(X|\mathcal{G})\big) = \mathrm{E}(X)\,.$$

Proof. Follows from Theorem 5.5 by taking $\mathcal{F} = \{\emptyset, \Omega\}$ and observing that for any integrable random variable Z,

$$\mathrm{E}(Z|\mathcal{F}) = \mathrm{E}(Z)\,.$$

□

Theorem 5.6. *If X is an integrable random variable and $\mathcal{G} \subset \mathcal{A}$ is a σ-field which is independent of $\sigma(X) \vee \mathcal{F}$, then*

$$\mathrm{E}\,(X|\mathcal{F} \vee \mathcal{G}) = \mathrm{E}(X|\mathcal{F}) \ \textit{a.s.}$$

Proof. Let $Z = \mathrm{E}(X|\mathcal{F})$. Since Z is integrable and $\mathcal{F} \vee \mathcal{G}$-measurable, all that needs to be shown is

$$\int_E Z\,dP = \int_E X\,dP \text{ for all } E \in \mathcal{F} \vee \mathcal{G}\,. \tag{5.7}$$

Let $\mathcal{S} = \{A \cap B : A \in \mathcal{F},\, B \in \mathcal{G}\}$; (5.7) will first be shown to hold for all $E \in \mathcal{S}$. Fix $E \in \mathcal{S}$, that is, $E = A \cap B$ for some $A \in \mathcal{F}$ and $B \in \mathcal{G}$. Then

$$\begin{aligned} \mathrm{E}\,(X\mathbf{1}_E) &= \mathrm{E}\,((X\mathbf{1}_A)\mathbf{1}_B) \\ (X\mathbf{1}_A \text{ is } \sigma(X) \vee \mathcal{F}\text{-measurable and } B \in \mathcal{G}) &= \mathrm{E}(X\mathbf{1}_A)\mathrm{E}(\mathbf{1}_B) \\ (Z = \mathrm{E}(X|\mathcal{F}),\, A \in \mathcal{F}) &= \mathrm{E}(Z\mathbf{1}_A)\mathrm{E}(\mathbf{1}_B) \\ &= \mathrm{E}\,(Z\mathbf{1}_E)\ , \end{aligned}$$

the last line again following from the independence of $\mathcal{F}$ and $\mathcal{G}$ and that $Z\mathbf{1}_A$ and $\mathbf{1}_B$ are measurable with respect to them, respectively. Thus, (5.7) holds for all $E \in \mathcal{S}$.

Since $\mathcal{S}$ is a semi-field, (5.7) can easily be shown to hold for all E in the field generated by $\mathcal{S}$. Finally, standard arguments using Theorem 1.2, which is the monotone class theorem, completes the proof. $\square$

Corollary 5.2. *If X is integrable and $\mathcal{G}$ is a σ-field independent of $\sigma(X)$, then*

$$\mathrm{E}(X|\mathcal{G}) = \mathrm{E}(X) \ \textit{a.s.}$$

Proof. Follows from Theorem 5.7 by taking $\mathcal{F} = \{\emptyset, \Omega\}$. $\square$

The above corollary is diametrically opposite to

$$\mathrm{E}(X|\mathcal{G}) = X\,, \text{ if } X \text{ is measurable w.r.t. } \mathcal{G}\,.$$

Definition 42. *If X and Y are random variables and the former is integrable, define*

$$\mathrm{E}(X|Y) = \mathrm{E}\big(X|\sigma(Y)\big)\,.$$

Theorem 5.7. *Suppose X and Y are independent and $f : \mathbb{R}^2 \to \mathbb{R}$ is a Borel function such that*

$$\mathrm{E}\left(|f(X,Y)|\right) < \infty\,.$$

Then

$$Y \in \left\{ y \in \mathbb{R} : \int_{\mathbb{R}} |f(x,y)| P(X \in dx) < \infty \right\} \ \textit{a.s.} \tag{5.8}$$

Further,

$$\mathrm{E}\left(f(X,Y)|Y\right) = g(Y)\,,$$

where

$$g(y) = \begin{cases} \int_{\mathbb{R}} f(x,y) P(X \in dx)\,, & \textit{if } \int_{\mathbb{R}} |f(x,y)| P(X \in dx) < \infty\,, \\ 0\,, & \textit{otherwise}\,. \end{cases}$$

Proof. Denote by μ_X and μ_Y the respective distributions of X and Y, that is, $\mu_X(B) = P(X \in B)$ for all $B \in \mathcal{B}(\mathbb{R})$, and likewise for μ_Y. Independence of X and Y implies

$$P\left((X,Y) \in B\right) = \mu_X \otimes \mu_Y(B) \text{ for all } B \in \mathcal{B}(\mathbb{R}^2)\,.$$

Thus,

$$\begin{aligned} \int_{\mathbb{R}} \int_{\mathbb{R}} |f(x,y)| \mu_X(dx) \mu_Y(dy) &= \mathrm{E}\left(|f(X,Y)|\right) \\ &< \infty\,. \end{aligned}$$

Tonelli's theorem implies

$$\mu_Y\left(y \in \mathbb{R} : \int_{\mathbb{R}} |f(x,y)|\mu_X(dx) < \infty\right) = 1\,,$$

which is exactly the same as (5.8).

Fubini's theorem implies for any $A \in \sigma(Y)$, that is, for $A = Y^{-1}B$ for some $B \in \mathcal{B}(\mathbb{R})$,

$$\begin{aligned}
\mathrm{E}\left(f(X,Y)\mathbf{1}_A\right) &= \int_{\mathbb{R}}\int_{\mathbb{R}} f(x,y)\mathbf{1}_B(y)\mu_X(dx)\mu_Y(dy) \\
&= \int_B \left(\int_{\mathbb{R}} f(x,y)\mu_X(dx)\right)\mu_Y(dy) \\
\text{(by (5.8))} &= \int_B g(y)\mu_Y(dy) \\
\left(\text{because } A = Y^{-1}B\right) &= \mathrm{E}\left(g(Y)\mathbf{1}_A\right)\,.
\end{aligned}$$

As this holds for all $A \in \sigma(Y)$, the proof follows. $\square$

Exercise 5.1. *If X and Y are independent from standard normal and standard uniform, respectively, calculate* $\mathrm{E}(Ye^{XY})$.

Soln.: First note that

$$0 \le Ye^{XY} \le e^{XY} \le e^{|XY|} \le e^{|X|}\,.$$

Since $\mathrm{E}(e^{|X|}) < \infty$ because X follows normal, Ye^{XY} is integrable. The tower property implies

$$\begin{aligned}
\mathrm{E}\left(Ye^{XY}\right) &= \mathrm{E}\left(\mathrm{E}\left(Ye^{XY}|Y\right)\right) \\
\text{(Theorem 5.3)} &= \mathrm{E}\left(Y\mathrm{E}(e^{XY}|Y)\right) \\
\left(\text{the above theorem and that } \mathrm{E}(e^{Xy}) = e^{y^2/2},\ y \in \mathbb{R}\right) &= \mathrm{E}\left(Ye^{Y^2/2}\right) \\
&= \int_0^1 ye^{y^2/2}\,dy \\
\left(z = y^2/2,\ dz = y\,dy\right) &= \int_0^{1/2} e^z\,dz \\
&= e^{1/2} - 1\,.
\end{aligned}$$

Exercise 5.2. *For independent X and Y and $0 < p < \infty$, show that*

$$\mathrm{E}\left(|X+Y|^p\right) < \infty \iff \mathrm{E}(|X|^p) < \infty \textit{ and } \mathrm{E}(|Y|^p) < \infty\,.$$

Soln.: The "$\Leftarrow$" part follows trivially from the observation

$$|X+Y|^p \le 2^p\left(|X|^p + |Y|^p\right)\,,$$

and doesn't really need independence.

For the "$\Rightarrow$" part, assume

$$\mathrm{E}\left(|X+Y|^p\right) < \infty\,.$$

Independence of X, Y and (5.8) imply

$$Y \in \{y \in \mathbb{R} : \mathrm{E}\left(|X+y|^p\right) < \infty\} \text{ a.s.}$$

In particular, the above set is non-empty. Thus, there exists $y \in \mathbb{R}$ such that

$$\mathrm{E}\left(|X+y|^p\right) < \infty\,.$$

Since

$$|X|^p \le 2^p \left(|X+y|^p + |y|^p\right)\,,$$

we get $\mathrm{E}(|X|^p) < \infty$. This also shows $\mathrm{E}(|Y|^p) < \infty$ and thus proves the "$\Rightarrow$" part.

Exercise 5.3. *A random variable X is infinitely divisible if for all fixed $n = 1, 2, \ldots,$ there exist i.i.d. random variables $X_{n1}, \ldots, X_{nn}$ defined on some probability space such that*

$$X \stackrel{d}{=} X_{n1} + \ldots + X_{nn}\,. \tag{5.9}$$

If X is an infinitely divisible random variable with mean zero and variance one, show that

$$\mathrm{E}(X^4) \ge 3\,.$$

Hint. *Using the above exercise, show that if* $\mathrm{E}(X^4) < \infty$*, then*

$$\mathrm{E}(X^4) = n\mathrm{E}\left(X_{n1}^4\right) + 3n(n-1)\left(\mathrm{E}\left(X_{n1}^2\right)\right)^2\,.$$

Exercise 5.4. *If X and Y are independent and either of them is a continuous random variable, show that*

$$X \ne Y \text{ a.s.}$$

Exercise 5.5. *If*

$$(X, Y) \sim N_2\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix}\right)\,,$$

show that $\mathrm{E}(Y|X) = \rho X$.

6 Various modes of convergence and the laws of large numbers

In this chapter, we shall understand the "frequentists' interpretation" of probability. Recall that "the probability of Heads for a coin is p" means that in large number of tosses of the coin, the observed proportion of Heads is close to p. To make this precise, we need notions of convergence for random variables. As usual, all random variables are defined on $(\Omega, \mathcal{A}, P)$, unless mentioned otherwise.

Definition 43. *For random variables X_n and X, if it holds that*

$$P\left(\left\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\}\right) = 1\,,$$

then as $n \to \infty$, X_n converges to X almost surely, or

$$X_n \to X \text{ a.s.}$$

Exercise 6.1. *For random variables $X, X_1, X_2, \ldots$, show that*

$$\left\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\} = \left[\liminf_{n\to\infty} X_n \ge X \ge \limsup_{n\to\infty} X_n\right] \in \mathcal{A}\,.$$

The following exercise shows that a.s. convergence is as good as convergence for every sample point, for practical purposes.

Exercise 6.2. *If $X_n \to X$ a.s., show that there exist random variables X_1', $X_2', \ldots$ such that $X_n' \to X$ and*

$$P\left(X_n' = X_n \text{ for all } n\right) = 1\,.$$

Theorem 6.1. *If $X_n \to X$ a.s. and $|X_n| \le Y$ for some Y which has finite expectation, then show that X has a finite expectation, and*

$$\lim_{n\to\infty} \mathrm{E}(X_n) = \mathrm{E}(X)\,.$$

Proof. Exercise. □

Definition 44. *A sequence of random variables (X_n) converges in probability to X or $X_n \xrightarrow{P} X$ if for every $\varepsilon > 0$,*

$$\lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0\,.$$

Exercise 6.3. *If $X_n \xrightarrow{P} X$ and $X_n \xrightarrow{P} X'$, then show that $X = X'$, a.s.*

Theorem 6.2. *If $X_n \to X$ a.s., then $X_n \xrightarrow{P} X$.*

Proof. Assume that $X_n \to X$ a.s. Fix $\varepsilon > 0$. Clearly,

$$[X_n \to X] \subset [\mathbf{1}(|X_n - X| > \varepsilon) \to 0],$$

and hence

$$\mathbf{1}(|X_n - X| > \varepsilon) \to 0 \text{ a.s.}$$

Theorem 6.1 shows

$$\lim_{n\to\infty} \mathrm{E}\left(\mathbf{1}(|X_n - X| > \varepsilon)\right) = 0,$$

that is,

$$P(|X_n - X| > \varepsilon) \to 0.$$

Since this holds for all $\varepsilon > 0$, it follows that $X_n \xrightarrow{P} X$ which completes the proof. $\square$

Example 6.1. *Let* $\Omega = (0,1]$, $\mathcal{A} = \mathcal{B}((0,1])$ *and* P *be the restriction of Lebesgue measure to* $(0,1]$. *Define for all* $\omega \in \Omega$,

$$\begin{aligned}
X_1(\omega) &= \mathbf{1}\left(0 < \omega \le \frac{1}{2}\right), \\
X_2(\omega) &= \mathbf{1}\left(\frac{1}{2} < \omega \le 1\right), \\
X_3(\omega) &= \mathbf{1}\left(0 < \omega \le \frac{1}{4}\right), \\
X_4(\omega) &= \mathbf{1}\left(\frac{1}{4} < \omega \le \frac{1}{2}\right), \\
X_5(\omega) &= \mathbf{1}\left(\frac{1}{2} < \omega \le \frac{3}{4}\right), \\
X_6(\omega) &= \mathbf{1}\left(\frac{3}{4} < \omega \le 1\right), \\
&\vdots \quad .
\end{aligned}$$

Then, $X_n \xrightarrow{P} 0$ *but*

$$P\left(\lim_{n\to\infty} X_n = 0\right) = 0.$$

The above example shows that convergence in probability is a strictly weaker notion of convergence than almost sure convergence.

Theorem 6.3 (Markov inequality)**.** *For a non-negative random variable* Z *and* $a > 0$,

$$P(Z \ge a) \le \frac{1}{a}\mathrm{E}(Z).$$

Proof. Since $Z \geq 0$, it holds that

$$\begin{aligned} \mathrm{E}(Z) &\geq \mathrm{E}\left(Z\mathbf{1}(Z \geq a)\right) \\ &\geq \mathrm{E}\left(a\mathbf{1}(Z \geq a)\right) \\ &= aP(Z \geq a)\,, \end{aligned}$$

from which the proof follows. □

Definition 45. *For $1 \leq p < \infty$ and $X, X_1, X_2, \ldots \in L^p(\Omega)$, $X_n \to X$ in L^p if*

$$\lim_{n\to\infty} \mathrm{E}\left(|X_n - X|^p\right) = 0\,.$$

Theorem 6.4. *If $X_n \to X$ in L_p for some $1 \leq p < \infty$, then the following hold.*

1. *For $1 \leq q \leq p$, $X_n \to X$ in L^q.*
2. *As $n \to \infty$, $X_n \xrightarrow{P} X$.*

Proof. 1. Follows from the fact that for any random variable Z,

$$\|Z\|_q \leq \|Z\|_p\,,$$

as long as $0 < q \leq p$, the solution of which follows by applying Jensen to get

$$\phi\left(\mathrm{E}(|Z|^q)\right) \leq \mathrm{E}\left(\phi(|Z|^q)\right)\,,$$

where $\phi(x) = |x|^{p/q}$ is convex.
2. For $\varepsilon > 0$,

$$P\left(|X_n - X| > \varepsilon\right) = P\left(|X_n - X|^p > \varepsilon^p\right) \leq \varepsilon^{-p}\mathrm{E}\left(|X_n - X|^p\right)\,,$$

by the Markov inequality. Letting $n \to \infty$ completes the proof. □

Exercise 6.4. *If $X_n \to Y$ in L^p for some $1 \leq p < \infty$ and $X_n \to Z$ a.s., show that*

$$Y = Z \text{ a.s.}$$

Exercise 6.5.

1. *In Example 6.1, show that for all $p \in [1, \infty)$, $X_n \to 0$ in L^p.*
2. *Show that for $p \in [1, \infty)$, convergence in L^p neither implies nor is implied by a.s. convergence.*

Theorem 6.5 (Weak law of large numbers (WLLN) for finite variance)**.** *If $X_1, X_2, \ldots$ are i.i.d. random variables with mean μ and finite variance, then*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to \mu$$

as $n \to \infty$ in L^2 and hence in probability.

Proof. Since

$$\mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \mu\,, n = 1, 2, \ldots\,,$$

it follows that

$$\begin{aligned}
\mathrm{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right)^2\right] &= \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\
(X_1, \ldots, X_n \text{ are independent}) &= \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}(X_i) \\
\left(X_1 \stackrel{d}{=} \ldots \stackrel{d}{=} X_n\right) &= \frac{1}{n}\mathrm{Var}(X_1) \\
&\to 0\,, n \to \infty\,.
\end{aligned}$$

Thus

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to \mu \text{ as } n \to \infty\,,$$

in L^2 and hence in probability. This completes the proof. □

Exercise 6.6. *If a coin with probability of Heads p is tossed infinitely many times, and X_n denotes the proportion of Heads observed in the first n tosses, then show that as $n \to \infty$,*

$$X_n \xrightarrow{P} p\,.$$

Theorem 6.6 (Borel-Cantelli lemma)**.** *If $A_1, A_2, \ldots$ are events such that*

$$\sum_{n=1}^{\infty} P(A_n) < \infty\,,$$

then

$$P\left(\{\omega \in \Omega : \omega \in A_n \textit{ for infinitely many } n\}\right) = 0\,.$$

Proof. Let

$$B_n = A_n \cup A_{n+1} \cup \ldots,\; n \geq 1\,,$$

and

$$B_\infty = \bigcap_{n=1}^{\infty} B_n\,.$$

Clearly,

$$B_\infty = \{\omega \in \Omega : \omega \in A_n \text{ for infinitely many } n\}\;.$$

Furthermore, since $B_n \downarrow B_\infty$, it follows that

$$P(B_\infty) = \lim_{n\to\infty} P(B_n) \leq \lim_{n\to\infty} \sum_{k=n}^{\infty} P(A_k) = 0\,,$$

which completes the proof. □

Theorem 6.7. *If $X_n \xrightarrow{P} X$, then X_n has a subsequence X_{n_k} such that*

$$X_{n_k} \to X \text{ a.s.},$$

as $k \to \infty$.

Proof. Since $X_n \xrightarrow{P} X$, there exists n_1 such that

$$P(|X_{n_1} - X| > 1) \le \frac{1}{2}.$$

There exists N_2 such that

$$P\left(|X_n - X| > \frac{1}{2}\right) \le 2^{-2} \text{ for all } n \ge N_2.$$

Define $n_2 = N_2 \vee (n_1 + 1)$. Proceeding similarly, we get positive integers $n_1 < n_2 < n_3 < \ldots$ such that

$$P\left(|X_{n_k} - X| > \frac{1}{k}\right) \le 2^{-k} \text{ for all } k.$$

Hence,

$$\sum_{k=1}^{\infty} P\left(|X_{n_k} - X| > \frac{1}{k}\right) < \infty.$$

Borel-Cantelli Lemma implies that

$$P\left(|X_{n_k} - X| > \frac{1}{k} \text{ for infinitely many } k\right) = 0.$$

Thus,

$$X_{n_k} \to X \text{ a.s.},$$

as $k \to \infty$. This completes the proof. $\square$

Exercise 6.7. *If $X_n \xrightarrow{P} X$ and $|X_n| \le Y$ for some Y with $\mathrm{E}(Y) < \infty$, then prove that*

$$\lim_{n\to\infty} \mathrm{E}(X_n) = \mathrm{E}(X).$$

Exercise 6.8. *Prove or disprove the following claim. If X_n and X are random variables such that any subsequence $\{X_{n_k} : k \ge 1\}$ of X_n has a further subsequence $\{X_{n_{k_l}} : l \ge 1\}$ such that*

$$X_{n_{k_l}} \to X \text{ a.s.},$$

then $X_n \to X$ a.s.

Exercise 6.9. *Show that the following are equivalent for random variables X_n and X.*

1. *As $n \to \infty$, $X_n \xrightarrow{P} X$.*

2. *Every subsequence $\{X_{n_k} : k \geq 1\}$ of $\{X_n : n \geq 1\}$ has a further subsequence $\{X_{n_{k_l}} : l \geq 1\}$ such that as $l \to \infty$,*

$$X_{n_{k_l}} \to X \ \text{a.s.}$$

3. *Every subsequence $\{X_{n_k} : k \geq 1\}$ of $\{X_n : n \geq 1\}$ has a further subsequence $\{X_{n_{k_l}} : l \geq 1\}$ such that as $l \to \infty$,*

$$X_{n_{k_l}} \xrightarrow{P} X\,.$$

Theorem 6.8. *If $X, X_1, X_2, \ldots$ are random variables such that*

$$\sum_{n=1}^{\infty} P\left(|X_n - X| > \varepsilon\right) < \infty \ \text{for all } \varepsilon > 0\,,$$

then $X_n \to X$ a.s.

Proof. Let

$$Z = \limsup_{n\to\infty} |X_n - X|\,,$$

which is a possibly improper random variable. The proof would follow if it can be shown that $Z = 0$ a.s., which is the same as

$$P(Z > \varepsilon) = 0 \text{ for all } \varepsilon > 0\,. \tag{6.1}$$

For $\varepsilon > 0$,

$$P(Z > \varepsilon) = P\left(|X_n - X| > \varepsilon \text{ for infinitely many } n\right)\,.$$

The hypothesis in conjunction with the Borel-Cantelli lemma shows that the right hand side is zero. Thus (6.1) follows, which completes the proof. □

Theorem 6.9 (Strong law of large numbers (SLLN) for finite fourth moment)**.** *If $X_1, X_2, \ldots$ are i.i.d. random variables with finite fourth moment, show that as $n \to \infty$,*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to \mathrm{E}(X_1) \ \text{a.s. and in } L^4\,.$$

Proof. Without loss of generality, assume that $\mathrm{E}(X_1) = 0$. It suffices to show that

$$\sum_{n=1}^{\infty} \mathrm{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^4\right] < \infty\,, \tag{6.2}$$

for the following reasons. Markov's inequality would show for $\varepsilon > 0$ and $n = 1, 2, \ldots$,

$$P\left(\left|\sum_{i=1}^n X_i\right| > \varepsilon\right) \le \varepsilon^{-4}\mathrm{E}\left[\left(\frac{1}{n}\sum_{i=1}^n X_i\right)^4\right].$$

From (6.2) it would follow that

$$\sum_{n=1}^\infty P\left(\left|\sum_{i=1}^n X_i\right| > \varepsilon\right) < \infty\,,$$

which in conjunction with Theorem 6.8 would prove

$$\frac{1}{n}\sum_{i=1}^n X_i \to 0 \text{ a.s.}$$

Besides, (6.2) would show that

$$\lim_{n\to\infty} \mathrm{E}\left[\left(\frac{1}{n}\sum_{i=1}^n X_i\right)^4\right] = 0\,,$$

which is the same as

$$\frac{1}{n}\sum_{i=1}^n X_i \to 0 \text{ in } L^4\,, n \to \infty\,.$$

Thus, showing (6.2) suffices.

For proving (6.2), write for a fixed n,

$$\begin{aligned}\mathrm{E}\left[\left(\frac{1}{n}\sum_{i=1}^n X_i\right)^4\right] &= n^{-4}\mathrm{E}\left(\sum_{i,j,k,l=1}^n X_iX_jX_kX_l\right) \\ &= n^{-4}\sum_{i,j,k,l=1}^n \mathrm{E}(X_iX_jX_kX_l)\,.\end{aligned} \tag{6.3}$$

Since $X_1, \ldots, X_n$ are i.i.d. and zero mean,

$$\mathrm{E}(X_iX_jX_kX_l) \neq 0$$

implies either $i = j = k = l$, in which case $\mathrm{E}(X_iX_jX_kX_l) = \mathrm{E}(X_1^4)$, or exactly one of the following holds:

$$i = j \neq k = l,\ i = k \neq j = l \text{ or } i = l \neq j = k\,.$$

In each of the above 3 cases, $\mathrm{E}(X_i X_j X_k X_l) = (\mathrm{E}(X_1^2))^2$. Thus,

$$\mathrm{E}\left[\left(\frac{1}{n}\sum_{i=1}^n X_i\right)^4\right] = n^{-4}\left[n\mathrm{E}(X_1^4) + 3n(n-1)\left(\mathrm{E}\left(X_1^2\right)\right)^2\right] \tag{6.4}$$
$$\leq n^{-2}\left[\mathrm{E}(X_1^4) + 3\left(\mathrm{E}\left(X_1^2\right)\right)^2\right].$$

Thus (6.2) follows, which completes the proof. □

The next result, which is the most general strong law of large numbers, assumes only finite mean and no higher moment. The following example shows that such random variables exist.

Example 6.2. *Suppose X is a random variable with density*

$$f(x) = \frac{1}{cx^2(\log x)^2}\mathbf{1}(x > e)\,, x \in \mathbb{R}\,,$$

where

$$c = \int_e^\infty \frac{dx}{x^2(\log x)^2}\,.$$

Thus $X \geq 0$ and

$$\begin{aligned} \mathrm{E}(X) &= \frac{1}{c}\int_e^\infty \frac{dx}{x(\log x)^2} \\ \left(y = \log x,\, dy = \frac{dx}{x}\right) &= \frac{1}{c}\int_1^\infty \frac{dy}{y^2} \\ &= \frac{1}{c} < \infty\,. \end{aligned}$$

For any $\varepsilon > 0$,

$$\begin{aligned} \mathrm{E}(X^{1+\varepsilon}) &= \frac{1}{c}\int_e^\infty \frac{dx}{x^{1-\varepsilon}(\log x)^2} \\ &= \infty \end{aligned}$$

because

$$\lim_{x\to\infty}\frac{x}{x^{1-\varepsilon}(\log x)^2} = \infty\,,$$

and

$$\int_e^\infty \frac{dx}{x} = \infty\,.$$

Thus $\mathrm{E}(X) < \infty = \mathrm{E}(X^{1+\varepsilon})$ for all $\varepsilon > 0$. That is, X has finite mean but any higher moment is infinite.

Theorem 6.10 (SLLN)**.** *For i.i.d. random variables $X_1, X_2, \ldots$ with finite mean μ,*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to \mu \ a.s.\,,$$

as $n \to \infty$.

For proving the SLLN, the following inequality will be used.

Theorem 6.11 (Kolmogorov maximal inequality)**.** *Let $X_1, \ldots, X_n$ be independent random variables with finite variance. Then, for any $\alpha > 0$,*

$$P\left(\max_{1\le k\le n} |S_k - \mathrm{E}(S_k)| \ge \alpha\right) \le \alpha^{-2}\mathrm{Var}(S_n)\,,$$

where

$$S_k = \sum_{i=1}^{k} X_i,\ 1 \le k \le n\,.$$

The following inequality obtained by putting $n = 1$ above is known as Chebyshev's inequality in probability theory. This follows directly from Markov's inequality as well.

Corollary 6.1 (Chebyshev's inequality)**.** *If X has mean μ and finite variance σ^2, then*

$$P\left(|X - \mu| \ge \alpha\right) \le \alpha^{-2}\sigma^2\,, \alpha > 0\,.$$

Proof of Theorem 6.11. WLOG, assume that $X_1, \ldots, X_n$ are zero mean. We start with the observation that

$$\left[\max_{1\le k\le n} |S_k| \ge \alpha\right] = \bigcup_{k=1}^{n} A_k\,,$$

where

$$A_k = \left[|S_k| \ge \alpha > |S_j| \text{ for all } 1 \le j \le k-1\right],\ k = 1, \ldots, n\,.$$

Since $A_1, \ldots, A_n$ are disjoint, it follows that

$$\begin{aligned}
\mathrm{Var}(S_n) &= \mathrm{E}(S_n^2) \\
&= \sum_{k=1}^{n} \mathrm{E}\left(S_n^2 \mathbf{1}_{A_k}\right) \\
&= \sum_{k=1}^{n} \left[\mathrm{E}\left((S_n - S_k)^2 \mathbf{1}_{A_k}\right) + \mathrm{E}\left(S_k^2 \mathbf{1}_{A_k}\right) + 2\mathrm{E}\left((S_n - S_k)S_k \mathbf{1}_{A_k}\right)\right] \\
&\ge \sum_{k=1}^{n} \left[\mathrm{E}\left(S_k^2 \mathbf{1}_{A_k}\right) + 2\mathrm{E}\left((S_n - S_k)S_k \mathbf{1}_{A_k}\right)\right]\,.
\end{aligned}$$

Since $S_n - S_k$ and $S_k \mathbf{1}_{A_k}$ are independent and the former has zero mean, it follows that

$$\mathrm{E}\left((S_n - S_k)S_k \mathbf{1}_{A_k}\right) = 0\,,$$

and hence

$$\begin{aligned}
\mathrm{Var}(S_n) &\geq \sum_{k=1}^{n} \mathrm{E}\left(S_k^2 \mathbf{1}_{A_k}\right) \\
&\geq \sum_{k=1}^{n} \mathrm{E}\left(\alpha^2 \mathbf{1}_{A_k}\right) \\
&= \alpha^2 P\left(\max_{1 \leq k \leq n} |S_k| \geq \alpha\right).
\end{aligned}$$

This completes the proof. □

Proof of Theorem 6.10. WLOG, assume that $E(X_1) = 0$. For $n \geq 1$, define

$$S_n := \sum_{i=1}^{n} X_i\,,$$

$$X_n' := X_n \mathbf{1}(|X_n| \leq n)\,,$$

and

$$S_n' := \sum_{k=1}^{n} X_k'\,.$$

Notice that

$$\begin{aligned}
\sum_{n=1}^{\infty} P\left(X_n \neq X_n'\right) &= \sum_{n=1}^{\infty} P(|X_1| > n) \\
&\leq \sum_{n=1}^{\infty} \int_{n-1}^{n} P(|X_1| > s)ds \qquad (6.5) \\
&= \int_0^{\infty} P(|X_1| > s)ds \\
&= E(|X_1|) \\
&< \infty\,,
\end{aligned}$$

(6.5) following from the observation that $P(|X_1| > n) \leq P(|X_1| > s)$ for $s \leq n$. From the Borel-Cantelli lemma, it follows that

$$\limsup_{n\to\infty} \left|S_n - S_n'\right| < \infty$$

almost surely, and hence

$$n^{-1}S_n - n^{-1}S_n' \to 0$$

almost surely. So it suffices to show that

$$n^{-1}S'_n \to 0 \text{ a.s.} \tag{6.6}$$

Notice that by DCT,

$$E(X'_n) = E\left[X_1\mathbf{1}(|X_1| \le n)\right] \to E(X_1) = 0\,,$$

and hence,

$$\lim_{n\to\infty} n^{-1}\mathrm{E}(S'_n) = 0\,.$$

Therefore, (6.6) will follow if we can show that

$$n^{-1}\left[S'_n - E(S'_n)\right] \to 0 \text{ a.s.}$$

For $r \ge 1$, set

$$Z_r := \max_{2^{r-1}\le k<2^r} |S'_k - E(S'_k)|\,.$$

Since

$$\frac{1}{k}|S'_k - E(S'_k)| \le 2^{-(r-1)}Z_r \text{ for all } 2^{r-1} \le k \le 2^r\,,$$

it suffices to show that

$$2^{-r}Z_r \to 0 \text{ a.s.}$$

The above will follows from Theorem 6.8 if it can be shown that

$$\sum_{r=1}^{\infty} P\left[|Z_r| > 2^r\varepsilon\right] < \infty\,.$$

for any $\varepsilon > 0$. Kolmogorov's inequality implies that

$$\begin{aligned}
\sum_{r=1}^{\infty} P\left[|Z_r| > 2^r\varepsilon\right] &\le \sum_{r=1}^{\infty} P\left(\max_{1\le k\le 2^r}\left|S'_k - \mathrm{E}(S'_k)\right| > 2^r\varepsilon\right)\\
&\le \sum_{r=1}^{\infty} \varepsilon^{-2}4^{-r}\mathrm{Var}(S'_{2^r})\\
&= \varepsilon^{-2}\sum_{r=1}^{\infty} 4^{-r}\sum_{j=1}^{2^r}\mathrm{Var}(X'_j)\\
&= \varepsilon^{-2}\sum_{j=1}^{\infty}\mathrm{Var}(X'_j)\sum_{r=\lceil\log_2 j\rceil}^{\infty} 4^{-r}\\
&\le K\sum_{j=1}^{\infty} j^{-2}\mathrm{Var}(X'_j)\,,
\end{aligned}$$

the last line following from the calculation that

$$\begin{aligned}\sum_{r=\lceil \log_2 j\rceil}^{\infty} 4^{-r} &= \frac{4}{3}4^{-\lceil \log_2 j\rceil} \\ &\le \frac{4}{3}4^{-\log_2 j} \\ &= \frac{4}{3}j^{-2}.\end{aligned}$$

Thus, in order to complete the proof, all that needs to be shown is that

$$\sum_{j=1}^{\infty} j^{-2}\mathrm{Var}(X_j') < \infty.$$

To that end, observe that

$$\begin{aligned}\sum_{j=1}^{\infty} j^{-2}\mathrm{Var}(X_j') &\le \sum_{j=1}^{\infty} j^{-2}E(X_j'^2) \\ &= \sum_{j=1}^{\infty} j^{-2}E(X_1^2\mathbf{1}(|X_1| \le j)) \\ &= \sum_{k=1}^{\infty} E(X_1^2\mathbf{1}(k-1 < |X_1| \le k))\sum_{j=k}^{\infty} j^{-2} \\ &\le \sum_{k=1}^{\infty} E(X_1^2\mathbf{1}(k-1 < |X_1| \le k))2/k \qquad (6.7) \\ &\le 2\sum_{k=1}^{\infty} E(|X_1|\mathbf{1}(k-1 < |X_1| \le k)) \\ &= 2E|X_1| < \infty,\end{aligned}$$

(6.7) following from the fact that for $k \ge 2$,

$$\sum_{j=k}^{\infty} j^{-2} \le \sum_{j=k}^{\infty}\int_{j-1}^{j} x^{-2}\,dx = \frac{1}{k-1} \le \frac{2}{k},$$

and

$$\sum_{j=1}^{\infty} j^{-2} \le 1 + \int_1^{\infty} x^{-2}\,dx = 2,$$

which together imply

$$\sum_{j=k}^{\infty} j^{-2} \le \frac{2}{k}, k \in \mathbb{N}.$$

Hence, the proof follows. $\square$

Exercise 6.10. *Let $X_1, X_2, \ldots$ be i.i.d. taking values 1 and -1, with respective probabilities p and $1-p$. Define*

$$S_n = \sum_{i=1}^{n} X_i \,, n = 0, 1, 2, \ldots ;$$

$(S_n : n \geq 0)$ is called a random walk. Show that

$$P(S_n = k \text{ for some } n \geq 0) = 1 \text{ for all } k = 1, 2, \ldots \text{ if } p > \frac{1}{2} \,,$$

and

$$P(S_n = k \text{ for some } n \geq 0) = 1 \text{ for all } k = -1, -2, \ldots \text{ if } p < \frac{1}{2} \,.$$

Theorem 6.12 (Kolmogorov's zero-one law)**.** *If $\mathcal{A}_1, \mathcal{A}_2, \ldots$ are independent σ-fields and*

$$\mathcal{T} = \bigcap_{n=1}^{\infty} \bigvee_{k=n}^{\infty} \mathcal{A}_k \,,$$

then $P(A)$ equals either 0 or 1 for all $A \in \mathcal{T}$.

Proof. For a fixed $n = 1, 2, \ldots$, $\mathcal{T} \subset \mathcal{A}_{n+1} \vee \mathcal{A}_{n+2} \vee \ldots$ and hence $\mathcal{T}$ is independent of $\mathcal{A}_1 \vee \ldots \vee \mathcal{A}_n$. Thus,

$$P(A \cap B) = P(A)P(B) \text{ for all } A \in \mathcal{T},\, B \in \mathcal{A}_1 \vee \ldots \vee \mathcal{A}_n \,.$$

The above holds for all $n = 1, 2, \ldots$, showing

$$P(A \cap B) = P(A)P(B) \text{ for all } A \in \mathcal{T},\, B \in \bigcup_{n=1}^{\infty} (\mathcal{A}_1 \vee \ldots \vee \mathcal{A}_n) \,.$$

Since $\bigcup_{n=1}^{\infty}(\mathcal{A}_1 \vee \ldots \vee \mathcal{A}_n)$ is a field, Theorem 3.1 shows that $\mathcal{T}$ is independent of the sigma-field generated by $\bigcup_{n=1}^{\infty}(\mathcal{A}_1 \vee \ldots \vee \mathcal{A}_n)$. Observing that

$$\sigma\left(\bigcup_{n=1}^{\infty} (\mathcal{A}_1 \vee \ldots \vee \mathcal{A}_n)\right) = \bigvee_{n=1}^{\infty} \mathcal{A}_n \supset \mathcal{T} \,,$$

$\mathcal{T}$ is thus independent of itself. In other words,

$$P(A \cap A) = P(A)P(A) \text{ for all } A \in \mathcal{T} \,,$$

showing that $P(A)$ has to be either 0 or 1, which completes the proof. $\square$

Exercise 6.11. *If $X_1, X_2, \ldots$ are independent random variables and*

$$S_n = X_1 + \ldots + X_n \,, n = 1, 2, \ldots \,,$$

show that

$$P\left(\limsup_{n\to\infty} S_n = \infty\right) = 0 \text{ or } 1\,,$$

and that there exists $a \in [-\infty, \infty]$ *such that*

$$P\left(\liminf_{n\to\infty} \frac{1}{n} S_n = a\right) = 1\,.$$

Convince yourself that the above is not necessarily true with $\frac{1}{n}S_n$ *replaced by* S_n*.*

Theorem 6.13 (Second Borel-Cantelli lemma)**.** *If* $A_1, A_2, \ldots$ *are independent events such that*

$$\sum_{n=1}^{\infty} P(A_n) = \infty\,,$$

then

$$P\left(A_n \text{ occurs for infinitely many } n\right) = 1\,.$$

Proof. Since

$$E = [A_n \text{ occurs for infinitely many } n] = \bigcap_{n=1}^{\infty} \bigvee_{k=n}^{\infty} \{A_k, A_k^c, \emptyset, \Omega\}\,,$$

Kolmogorov's zero-one law shows $P(E)$ is either 0 or 1. Thus it suffices to show

$$P(E) > 0\,. \tag{6.8}$$

Recall that for any $n \geq 1$,

$$\begin{aligned}
P\left(\bigcup_{i=1}^{n} A_i\right) &= 1 - P\left(\bigcap_{i=1}^{n} A_i^c\right) \\
&= 1 - \prod_{i=1}^{n} (1 - P(A_i)) \\
\left(1 - x \leq e^{-x} \text{ for all } x \in \mathbb{R}\right) &\geq 1 - \prod_{i=1}^{n} e^{-P(A_i)} \\
&= 1 - \exp\left(-\sum_{i=1}^{n} P(A_i)\right) \\
&\to 1\,,
\end{aligned}$$

as $n \to \infty$ because $\sum_{i=1}^{\infty} P(A_i) = \infty$.

Let $\alpha_1, \alpha_2, \ldots \in (0, 1)$ be such that

$$\prod_{i=1}^{\infty} \alpha_i > 0\,.$$

For example, $\alpha_i = e^{-1/i^2}$ for $i = 1, 2, \ldots$ satisfies the above. The above calculations show there exists n_1 such that

$$P\left(\bigcup_{i=1}^{n_1} A_i\right) \geq \alpha_1 .$$

Since $\sum_{i=n_1+1}^{\infty} P(A_i) = \infty$, a similar calculation shows there exists $n_2 > n_1$ such that

$$P\left(\bigcup_{i=n_1+1}^{n_2} A_i\right) \geq \alpha_2 .$$

Proceeding inductively, get integers $0 = n_0 < n_1 < n_2 < \ldots$ such that

$$P\left(\bigcup_{i=n_{k-1}+1}^{n_k} A_i\right) \geq \alpha_k \, , k \in \mathbb{N} .$$

Clearly,

$$E \supset \bigcap_{k=1}^{\infty} \bigcup_{i=n_{k-1}+1}^{n_k} A_i .$$

Therefore,

$$\begin{aligned} P(E) &\geq P\left(\bigcap_{k=1}^{\infty} \bigcup_{i=n_{k-1}+1}^{n_k} A_i\right) \\ &= \prod_{k=1}^{\infty} P\left(\bigcup_{i=n_{k-1}+1}^{n_k} A_i\right) \\ &\geq \prod_{k=1}^{\infty} \alpha_k > 0 . \end{aligned}$$

Thus (6.8) holds, from which the proof follows. □

An immediate consequence of the second Borel-Cantelli lemma is the following, which should be compared with Theorem 6.8.

Exercise 6.12. *If $X_1, X_2, X_3, \ldots$ are independent random variables, then show that*

$$X_n \to X \ \textit{a.s.} \iff \sum_{n=1}^{\infty} P\left(|X_n - X| > \varepsilon\right) < \infty \ \textit{for all } \varepsilon > 0 .$$

Show that if the above holds, then X is a degenerate random variable.

Exercise 6.13. *Suppose $X_1, X_2, \ldots$ are random variables such that*

$$\sum_{n=1}^{\infty} \mathrm{E}(X_n^2) < \infty\,.$$

If $Y_1, Y_2, \ldots$ are such that $\sigma(X_n : n \ge 1), \sigma(Y_1), \sigma(Y_2), \ldots$ are independent and Y_n takes values 1 and -1, each with probability $1/2$ for $n = 1, 2, \ldots$, show that

$$\sum_{i=1}^{n} X_i Y_i \to Z\,,\ \text{as } n \to \infty\,,$$

in L^2, for some $Z \in L^2(\Omega)$.

Soln.: Let

$$Z_n = \sum_{i=1}^{n} X_i Y_i\,, n \ge 1\,.$$

Since $L^2(\Omega)$ is a complete metric space, it suffices to show that $\{Z_n : n \ge 1\}$ is a Cauchy sequence in $L^2(\Omega)$. For $1 \le m < n$,

$$\begin{aligned}
\mathrm{E}\left[(Z_n - Z_m)^2\right] &= \mathrm{E}\left[\left(\sum_{i=m+1}^{n} X_i Y_i\right)^2\right] \\
\left(Y_i^2 = 1\right) &= \sum_{i=m+1}^{n} \mathrm{E}(X_i^2) + 2 \sum_{m+1 \le i < j \le n} \mathrm{E}\left(X_i X_j Y_i Y_j\right)\,.
\end{aligned}$$

For $m+1 \le i < j \le n$, independence of $\sigma(X_1, X_2, \ldots), \sigma(Y_i), \sigma(Y_j)$ shows

$$\mathrm{E}\left(X_i X_j Y_i Y_j\right) = \mathrm{E}(X_i X_j)\mathrm{E}(Y_i)\mathrm{E}(Y_j) = 0\,.$$

Thus,

$$\mathrm{E}\left[(Z_n - Z_m)^2\right] = \sum_{i=m+1}^{n} \mathrm{E}(X_i^2)\,.$$

Given $\varepsilon > 0$, choosing N such that

$$\sum_{i=N+1}^{\infty} \mathrm{E}(X_i^2) \le \varepsilon\,,$$

which is possible from the given hypothesis, it holds that for $N \le m < n$,

$$\mathrm{E}\left[(Z_n - Z_m)^2\right] = \sum_{i=m+1}^{n} \mathrm{E}(X_i^2) \le \sum_{i=N+1}^{\infty} \mathrm{E}(X_i^2) \le \varepsilon\,,$$

showing $\{Z_n : n \ge 1\}$ is a Cauchy sequence in $L^2(\Omega)$.

Exercise 6.14. *Suppose $X_1, X_2, \ldots$ are i.i.d. from standard exponential, and $X_{(n,1)}, \ldots, X_{(n,n)}$ are the order statistics of $X_1, \ldots, X_n$. Then show that*

$$X_{(n,[n/2])} \to \log 2 \text{ a.s., as } n \to \infty\,.$$

Exercise 6.15. 1. *If $X_1, X_2, \ldots$ are independent and $\alpha_n \to \infty$ such that*

$$\frac{1}{\alpha_n}\sum_{i=1}^n X_i \xrightarrow{P} X \text{ as } n \to \infty\,,$$

show that X is a degenerate random variable.

2. *Hence or otherwise, prove that if $X_1, X_2, \ldots$ are i.i.d. from standard normal, then there* ***does not exist*** *a random variable Z such that*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i \xrightarrow{P} Z\,, n \to \infty\,.$$

7 Characteristic function and moment generating function

By convention, $(\Omega, \mathcal{A}, P)$ is the underlying probability space.

Definition 46. *A function $Z : \Omega \to \mathbb{C}$ which satisfies $Z^{-1}A \in \mathcal{A}$ for all $A \in \mathcal{B}(\mathbb{C})$ is a* complex-valued random variable. *A $\mathbb{C}$-valued random variable Z is* integrable *if*

$$\int_\Omega |Z|\, dP < \infty\,,$$

and in that case its expectation is defined as

$$\mathrm{E}(Z) = \int_\Omega Z\, dP\,.$$

Throughout this chapter, $\Re(z)$ and $\Im(z)$ will denote the real and imaginary parts of z for $z \in \mathbb{C}$ and $\iota = \sqrt{-1}$. That is, for $z = x + \iota y$ where $x, y \in \mathbb{R}$,

$$\Re(z) = x\,, \Im(z) = y\,.$$

Exercise 7.1. 1. *For a function $Z : \Omega \to \mathbb{C}$, show that*

$$\sigma(Z) := \{Z^{-1}A : A \in \mathcal{B}(\mathbb{C})\} = \sigma\left(\Re(Z)\right) \vee \sigma\left(\Im(Z)\right)\,.$$

Hence prove that Z is $\mathcal{B}(\mathbb{C})$-measurable if and only if $\Re(Z)$ and $\Im(Z)$ are $\mathcal{B}(\mathbb{R})$-measurable.

2. *For a $\mathbb{C}$-valued integrable random variable Z, show that*

$$|\mathrm{E}(Z)| \le \mathrm{E}(|Z|)\,.$$

3. *Show that a $\mathbb{C}$-valued random variable Z is integrable if and only if $\Re(Z)$ and $\Im(Z)$ are integrable, and in that case,*

$$\mathrm{E}(Z) = \mathrm{E}\left(\Re(Z)\right) + \iota \mathrm{E}(\Im(Z))\,.$$

4. *For integrable $\mathbb{C}$-valued random variables Z_1, Z_2 and $\alpha, \beta \in \mathbb{C}$, show that $\alpha Z_1 + \beta Z_2$ is integrable and*

$$\mathrm{E}(\alpha Z_1 + \beta Z_2) = \alpha \mathrm{E}(Z_1) + \beta \mathrm{E}(Z_2)\,.$$

5. *If Z_1 and Z_2 are integrable $\mathbb{C}$ valued random variables which are independent, that is, $\sigma(Z_1)$ and $\sigma(Z_2)$ are independent, then show that $Z_1 Z_2$ is integrable and*

$$\mathrm{E}(Z_1 Z_2) = \mathrm{E}(Z_1)\mathrm{E}(Z_2)\,.$$

Definition 47. *For a probability measure μ on $\mathbb{R}$, its* characteristic function *(CHF) is a function $\phi : \mathbb{R} \to \mathbb{C}$ defined by*

$$\phi(t) := \int_{\mathbb{R}} e^{\iota t x} \mu(dx),\, t \in \mathbb{R}\,,$$

where $\iota := \sqrt{-1}$. The characteristic function *of a random variable X is the CHF of the measure $P \circ X^{-1}$ on $\mathbb{R}$, that is,*

$$\phi_X(t) := \int_{\mathbb{R}} e^{\iota t x} P(X \in dx),\, t \in \mathbb{R}\,.$$

Exercise 7.2. *For a random variable X with CHF ϕ_X, show that*

$$\phi_X(t) := E\left[e^{\iota t X}\right] = E[\cos tX] + \iota E[\sin tX],\, t \in \mathbb{R}\,.$$

Theorem 7.1. *Let ϕ_X be the characteristic function of a random variable X. Then*

1. *$\phi_X(0) = 1$ and $|\phi_X(t)| \le 1$ for all t,*

2. *ϕ_X is uniformly continuous,*

3. *$aX + b$ has the characteristic function ϕ_{aX+b} given by*

$$\phi_{aX+b}(t) = e^{ibt}\phi_X(at),\, t \in \mathbb{R}\,.$$

Proof. Proof of 1. follows immediately from Exc 7.1.2, which implies

$$\left|\mathrm{E}(e^{\iota t X})\right| \le \mathrm{E}\left(|e^{\iota t X}|\right) = 1\,.$$

For 2., notice that for any $t, h \in \mathbb{R}$,

$$\begin{aligned}
|\phi_X(t+h) - \phi_X(t)| &= \left| E\left[e^{i(t+h)X} - e^{itX} \right] \right| \\
\text{(Exc 7.1.2)} \quad &\leq \mathrm{E}\left| e^{i(t+h)X} - e^{itX} \right| \\
&= \mathrm{E}\left| e^{ihX} - 1 \right| .
\end{aligned}$$

By the DCT, it follows that

$$\lim_{h\to 0} E\left[\left| e^{ihX} - 1 \right| \right] = 0 \,,$$

which completes the proof of uniform continuity.
Finally, 3. follows immediately from Exc 7.1.4 which implies

$$\mathrm{E}\left(e^{\iota t(aX+b)} \right) = e^{\iota bt} \mathrm{E}\left(e^{\iota a t X} \right) \,, t \in \mathbb{R} \,.$$

□

Theorem 7.2. *If $\lambda \neq 0$ and ϕ_X is the characteristic function of X, then the following three statements are equivalent.*

1. *$\phi_X(\lambda) = 1$.*
2. *ϕ_X has period λ.*
3. $$P\left[X \in \frac{2\pi}{\lambda} \mathbb{Z} \right] = 1 \,.$$

Proof. 1.⇒3. Assume that $\phi_X(\lambda) = 1$. This means that

$$E[\cos \lambda X] = 1 \,.$$

Therefore, $\cos \lambda X = 1$ almost surely, which shows 3.
3.⇒2. If 3. holds, then

$$e^{\iota \lambda X} = 1 \text{ a.s}$$

Thus, for a fixed $t \in \mathbb{R}$,

$$e^{\iota(t+\lambda)X} = e^{\iota t X} \text{ a.s.}$$

Taking expectation of both sides shows $\phi_X(t+\lambda) = \phi_X(t)$, from which 2. follows.
2.⇒1. Trivial because $\phi_X(0) = 1$. □

Exercise 7.3. *If the CHF ϕ_X of X satisfies*

$$|\phi_X(\lambda)| = 1 \textit{ for some } \lambda \neq 0 \,,$$

show that there exists $a \in \mathbb{R}$ such that

$$P\left(X - a \in \frac{2\pi}{\lambda} \mathbb{Z} \right) = 1 \,.$$

Definition 48. *For a probability measure μ on $\mathbb{R}$, its* moment generating function *(MGF) is a function $\psi : \mathbb{R} \to (0, \infty]$ defined by*

$$\psi(t) = \int_{-\infty}^{\infty} e^{tx} \mu(dx) \,.$$

The moment generating function of a random variable X is ψ_X defined by

$$\psi_X(t) = \mathrm{E}\left(e^{tX}\right) , t \in \mathbb{R} \,.$$

While the characteristic function (CHF) takes values in $\{z \in \mathbb{C} : |z| \le 1\}$, the moment generating function at t is possibly $+\infty$ for all $t \ne 0$. However, like the CHF, the MGF also maps 0 to 1.

Theorem 7.3. *Let μ be a probability measure on $\mathbb{R}$ with MGF ψ. If*

$$\alpha = \inf\{t \in \mathbb{R} : \psi(t) < \infty\} \text{ and } \beta = \sup\{t \in \mathbb{R} : \psi(t) < \infty\} \,,$$

then $\{t \in \mathbb{R} : \psi(t) < \infty\} \supset (\alpha, \beta)$. If $\alpha < \beta$ then e^{zx} is μ-integrable for z with $\alpha < \Re(z) < \beta$ and $f : \{z \in \mathbb{C} : \alpha < \Re(z) < \beta\} \to \mathbb{C}$ defined by

$$f(z) = \int_{-\infty}^{\infty} e^{zx} \mu(dx) \,, \tag{7.1}$$

is a holomorphic function. If, in addition, $\alpha < 0 < \beta$, then all moments of μ are finite, that is,

$$\int_{-\infty}^{\infty} |x|^n \mu(dx) < \infty \,, n = 1, 2, \dots \,,$$

and

$$f(z) = \sum_{n=0}^{\infty} \frac{z^n}{n!} \int_{-\infty}^{\infty} x^n \mu(dx) \,, z \in \mathbb{C}, |z| < (-\alpha) \wedge \beta \,. \tag{7.2}$$

Proof. Monotonicity and positivity of the exponential function on $\mathbb{R}$ implies that for $t_1 < t_2 < t_3$,

$$e^{t_2 x} \le e^{t_1 x} + e^{t_3 x} \,, x \in \mathbb{R} \,, \tag{7.3}$$

showing that

$$\psi(t_2) \le \psi(t_1) + \psi(t_3) \,.$$

Thus, $\{t \in \mathbb{R} : \psi(t) < \infty\}$ is a convex subset of $\mathbb{R}$ which contains (α, β).

If $\alpha < \Re(z) < \beta$ for some $z \in \mathbb{C}$, then the above shows $\psi(\Re(z)) < \infty$ and thus

$$\int_{-\infty}^{\infty} |e^{zx}| \, \mu(dx) = \int_{-\infty}^{\infty} e^{x\Re(z)} \, \mu(dx) = \psi(\Re(z)) < \infty \,.$$

DCT for complex-valued function shows that f defined by (7.1) is continuous. A standard application of Fubini and Morera's theorem in conjunction with (7.3) proves f is holomorphic.

For the final claim, assume $\alpha < 0 < \beta$. For $0 < t < (-\alpha) \wedge \beta$ and $n \geq 1$,

$$|x|^n \leq t^{-n} n! \frac{|tx|^n}{n!} \leq n! t^{-n} e^{|tx|} \leq n! t^{-n} \left(e^{tx} + e^{-tx}\right), x \in \mathbb{R}.$$

Thus,

$$\int_{\mathbb{R}} |x|^n \mu(dx) \leq n! t^{-n} \int_{-\infty}^{\infty} \left(e^{tx} + e^{-tx}\right) \mu(dx) = n! t^{-n} \left(\psi(t) + \psi(-t)\right) < \infty.$$

Further, for $z \in \mathbb{C}$ with $|z| \leq t$,

$$f(z) = \int_{\mathbb{R}} \lim_{n\to\infty} \sum_{i=0}^{n} \frac{1}{i!} (zx)^i \, \mu(dx).$$

Since for all n and $x \in \mathbb{R}$,

$$\left| \sum_{i=0}^{n} \frac{1}{i!} (zx)^i \right| \leq \sum_{i=0}^{n} \frac{1}{i!} |zx|^i \leq e^{|zx|} \leq e^{t|x|},$$

and

$$\int_{\mathbb{R}} e^{t|x|} \mu(dx) \leq \psi(t) + \psi(-t) < \infty,$$

DCT shows that

$$\begin{aligned} \int_{\mathbb{R}} \lim_{n\to\infty} \sum_{i=0}^{n} \frac{1}{i!} (zx)^i \, \mu(dx) &= \lim_{n\to\infty} \int_{\mathbb{R}} \sum_{i=0}^{n} \frac{1}{i!} (zx)^i \, \mu(dx) \\ &= \lim_{n\to\infty} \sum_{i=0}^{n} \frac{1}{i!} z^i \int_{\mathbb{R}} x^i \, \mu(dx) \\ &= \sum_{i=0}^{\infty} \frac{1}{i!} z^i \int_{\mathbb{R}} x^i \mu(dx). \end{aligned}$$

Since this holds for all $z \in \mathbb{C}$ with $|z| \leq t$ and t is arbitrary in $(0, (-\alpha) \wedge \beta)$, (7.2) follows. □

Corollary 7.1. *If μ is a probability measure such that $\alpha < 0 < \beta$, where α, β are as in Theorem 7.3, then the MGF ψ of μ satisfies*

$$\psi(t) = \sum_{n=0}^{\infty} \frac{t^n}{n!} \int_{-\infty}^{\infty} x^n \mu(dx), t \in \mathbb{R}, |t| < (-\alpha) \wedge \beta,$$

and the CHF ϕ of μ satisfies

$$\phi(t) = \sum_{n=0}^{\infty} \frac{(\iota t)^n}{n!} \int_{-\infty}^{\infty} x^n \mu(dx), t \in \mathbb{R}, |t| < (-\alpha) \wedge \beta.$$

Remark 2. *Only DCT and no complex analysis is used in the proof of* (7.2)*, and therefore for Corollary 7.1.*

Exercise 7.4. *Show that the characteristic function* ϕ *of standard normal is*

$$\phi(t) := e^{-t^2/2},\ t \in \mathbb{R}\,. \tag{7.4}$$

1. *Show that*
$$\int_{-\infty}^{\infty} e^{tx-x^2/2}\,dx = \sqrt{2\pi}\,e^{t^2/2}\,, t \in \mathbb{R}\,.$$

2. *Interpret the above as that the MGF of standard normal is*
$$\psi(t) = e^{t^2/2}\,, t \in \mathbb{R}\,, \tag{7.5}$$
whose analytic continuation to $\mathbb{C}$ *is*
$$\tilde{\psi}(z) = e^{z^2/2}\,, z \in \mathbb{C}\,.$$
Use Theorem 7.3 to arrive at (7.4)*. This line of argument essentially justifies replacing* t *by* ιt *in* (7.5)*.*

3. *Derive that*
$$\int_{-\infty}^{\infty} e^{\iota tx-x^2/2}\,dx = \sqrt{2\pi}\,e^{-t^2/2}\,, t \in \mathbb{R}\,,$$
which is the same as (7.4)*.*

Theorem 7.4 (Inversion theorem)**.** *If the probability measure* μ *has characteristic function* ϕ*, and* $a < b$ *are such that* $\mu\{a, b\} = 0$*, then*

$$\mu(a, b] = \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-\iota ta} - e^{-\iota tb}}{\iota t} \phi(t)dt\,.$$

Lemma 7.1. *If*

$$S(T) := \int_0^T \frac{\sin x}{x} dx,\ T \geq 0\,,$$

then

$$\lim_{T\to\infty} S(T) = \frac{\pi}{2}\,.$$

Proof of Theorem 7.4. Fix a, b satisfying the hypotheses. Before giving the actual proof, let us start with a sketch of the proof; every step in the sketch will eventually be justified. For $T > 0$,

$$\begin{aligned}
&\int_{-T}^{T} \frac{e^{-\iota ta} - e^{-\iota tb}}{\iota t} \phi(t)\, dt \\
&= \int_{-T}^{T} \frac{e^{-\iota ta} - e^{-\iota tb}}{\iota t} \int_{-\infty}^{\infty} e^{\iota tx}\, \mu(dx)\, dt \\
&= \int_{-\infty}^{\infty} \int_{-T}^{T} \frac{e^{\iota t(x-a)} - e^{\iota t(x-b)}}{\iota t}\, dt\, \mu(dx) && (7.6)\\
&= \int_{-\infty}^{\infty} \int_{-T}^{T} t^{-1} \left(\sin(t(x-a)) - \sin(t(x-b))\right)\, dt\, \mu(dx) && (7.7)\\
&= \int_{-\infty}^{\infty} 2\big(\operatorname{sgn}(x-a) S(T|x-a|) - \operatorname{sgn}(x-b) S(T|x-b|)\big)\, \mu(dx)\,, && (7.8)
\end{aligned}$$

S being as in Lemma 7.1, provided (7.6)-(7.8) can be justified. The said lemma implies that

$$\begin{aligned}
&\lim_{T\to\infty} \big(\operatorname{sgn}(x-a) S(T|x-a|) - \operatorname{sgn}(x-b) S(T|x-b|)\big) \\
&= \frac{\pi}{2} \left(\operatorname{sgn}(x-a) - \operatorname{sgn}(x-b)\right) \\
&= \begin{cases} \pi\,, & a < x < b\,, \\ 0\,, & x < a \text{ or } x > b\,, \\ \frac{\pi}{2}\,, & x = a \text{ or } x = b\,. \end{cases}
\end{aligned}$$

Thus,

$$\begin{aligned}
&\lim_{T\to\infty} \int_{-T}^{T} \frac{e^{-\iota ta} - e^{-\iota tb}}{\iota t} \phi(t)\, dt \\
&= \lim_{T\to\infty} \int_{-\infty}^{\infty} 2\big(\operatorname{sgn}(x-a) S(T|x-a|) - \operatorname{sgn}(x-b) S(T|x-b|)\big)\, \mu(dx) \\
&= \int_{-\infty}^{\infty} \left(2\pi \mathbf{1}(a < x < b) + \pi \mathbf{1}(x \in \{a, b\})\right)\, \mu(dx) && (7.9)\\
&= 2\pi\mu\big((a, b]\big)\,,
\end{aligned}$$

because $\mu(\{a, b\}) = 0$ by assumption, provided the interchange of integral and limit in (7.9) can be justified. The claim would thus follow once (7.6)-(7.9) are justified.

For justifying the interchange of integrals in (7.6), notice that

$$\begin{aligned}
\left| \frac{e^{\iota t(x-a)} - e^{\iota t(x-b)}}{\iota t} \right| &= \left| \frac{e^{-\iota ta} - e^{-\iota tb}}{\iota t} \right| \\
&= \left| \int_a^b e^{-\iota tx} dx \right| \\
&\le b - a\,. && (7.10)
\end{aligned}$$

Therefore,

$$\int_{-\infty}^{\infty}\int_{-T}^{T}\left|\frac{e^{\iota t(x-a)}-e^{\iota t(x-b)}}{\iota t}\right| dt\,\mu(dx) \le (b-a)\int_{-\infty}^{\infty}\int_{-T}^{T} 1\,dt\,\mu(dx)$$
$$\text{(Tonelli)} = 2T(b-a) < \infty\,.$$

Thus, (7.6) follows from Fubini.

The equality in (7.7) follows immediately by observing that for fixed x,

$$\int_{-T}^{T}\frac{e^{\iota t(x-a)}-e^{\iota t(x-b)}}{\iota t}\,dt = \int_{-T}^{T} t^{-1}\left(\sin(t(x-a))-\sin(t(x-b))\right)dt$$
$$-\iota\int_{-T}^{T} t^{-1}\left(\cos(t(x-a))-\cos(t(x-b))\right)dt\,,$$

and that $t\mapsto t^{-1}\left(\cos(t(x-a))-\cos(t(x-b))\right)$ is an odd function.

For justifying (7.8), first fix $x\neq a$ and write

$$\begin{aligned}\int_{-T}^{T}\frac{\sin(t(x-a))}{t}\,dt &= (x-a)\int_{-T}^{T}\frac{\sin(t(x-a))}{t(x-a)}\,dt\\ (y=t(x-a),\ dy=|x-a|\,dt) &= \frac{x-a}{|x-a|}\int_{-T|x-a|}^{T|x-a|}\frac{\sin y}{y}\,dy\\ &= \operatorname{sgn}(x-a)\int_{-T|x-a|}^{T|x-a|}\frac{\sin y}{y}\,dy\\ &= 2\operatorname{sgn}(x-a)S\left(T|x-a|\right)\,,\end{aligned}$$

as $y\mapsto y^{-1}\sin y$ is an even function. The identity

$$\int_{-T}^{T}\frac{\sin(t(x-a))}{t}\,dt = 2\operatorname{sgn}(x-a)S\left(T|x-a|\right)$$

holds for $x=a$ as well because in that case both sides vanish. The above holds with a replaced by b, which establishes (7.8).

Finally, (7.9) is justified by the observation that

$$K=\sup_{t\ge 0}|S(t)|<\infty\,, \tag{7.11}$$

which follows from Lemma 7.1 and the fact that $S(\cdot)$ is a continuous function. Since

$$|\operatorname{sgn}(x-a)S(T|x-a|)-\operatorname{sgn}(x-b)S(T|x-b|)|\le 2K\,,$$

and μ is a finite measure, DCT justifies the interchange of limit and integral in (7.9). This completes the proof. $\square$

The following is an immediate corollary of Theorem 7.4.

Corollary 7.2. *If μ_1 and μ_2 are probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ with respective CHFs ϕ_1 and ϕ_2, then*

$$\phi_1(t) = \phi_2(t) \text{ for all } t \in \mathbb{R} \iff \mu_1 = \mu_2 \,.$$

Theorem 7.5. *Suppose μ_1 and μ_2 are probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ with respective MGFs ψ_1 and ψ_2. If there exists $\theta > 0$ such that*

$$\psi_1(t) = \psi_2(t) < \infty \,, t \in [-\theta, \theta] \,,$$

then $\mu_1 = \mu_2$.

Proof. For $i = 1, 2$, define $f_i : \{z \in \mathbb{C} : |\Re(z)| < \theta\} \to \mathbb{C}$ by

$$f_i(z) = \int_{\mathbb{R}} e^{zx} \mu_i(dx) \,,$$

which is possible because the MGFs of μ_1 and μ_2 are finite on $[-\theta, \theta]$. Theorem 7.3 shows that f_1 and f_2 are holomorphic. Since ψ_i is the restriction of f_i to $(-\theta, \theta)$, the assumption implies f_1 and f_2 agree on an uncountable set. Thus, $f_1 = f_2$. As $\{\iota t : t \in \mathbb{R}\}$ is contained in the domains of f_1 and f_2, it follows that

$$f_1(\iota t) = f_2(\iota t) \,, t \in \mathbb{R} \,.$$

The above is the same as saying the CHFs of μ_1 and μ_2 are identical. Corollary 7.2 completes the proof. ☐

Theorem 7.6 (Inversion theorem for densities)**.** *If the characteristic function ϕ of a probability measure μ is integrable on $\mathbb{R}$, that is,*

$$\int_{-\infty}^{\infty} |\phi(t)| dt < \infty \,,$$

then f defined by

$$f(x) := \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\iota t x} \phi(t) dt \,,$$

is a density of μ.

Proof. Let F be the CDF of μ, that is,

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

Step 1. The function F is continuous.

Proof of Step 1. Suffices to show that for all $x \in \mathbb{R}$, $\mu\{x\} = 0$. Fix $x \in \mathbb{R}$. Let $a < x \le b$ be such that $\mu\{a, b\} = 0$. By the preceding result, it follows that

$$\mu(a, b] = \lim_{T \to \infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-\iota t a} - e^{-\iota t b}}{\iota t} \phi(t) dt \,.$$

Notice that for all $T \geq 0$,

$$\left|\int_{-T}^{T} \frac{e^{-\iota ta} - e^{-\iota tb}}{\iota t}\phi(t)dt\right| \leq \int_{-T}^{T} \left|\frac{e^{-\iota ta} - e^{-\iota tb}}{\iota t}\right| |\phi(t)|dt$$
$$\leq (b-a)\int_{-\infty}^{\infty} |\phi(t)|dt\,,$$

(7.10) implying the inequality in the last line. Hence

$$\mu(a,b] \leq (b-a)\int_{-\infty}^{\infty} |\phi(t)|dt\,.$$

Let (a_n) and (b_n) be such that $a_n < x \leq b_n$, $\mu\{a_n, b_n\} = 0$, $a_n, b_n \to x$. Then

$$\mu\{x\} \leq \mu(a_n, b_n] \leq (b_n - a_n)\int_{-\infty}^{\infty} |\phi(t)|dt\,,$$

and the RHS converges to zero as $n \to \infty$. This completes the proof of Step 1. □

Step 2. The function F is differentiable, and

$$F'(x) = f(x)\,.$$

Proof of Step 2. Fix $x \in \mathbb{R}$ and $h \neq 0$. Then, by Step 1 and the preceding result, it follows that

$$\frac{F(x+h) - F(x)}{h} = \frac{1}{h}\mu\left((x, x+h]\right) = \lim_{T\to\infty} \frac{1}{2\pi}\int_{-T}^{T} \frac{e^{-\iota tx} - e^{-\iota t(x+h)}}{\iota th}\phi(t)dt\,.$$

Since

$$\left|\frac{e^{-\iota tx} - e^{-\iota t(x+h)}}{\iota th}\phi(t)\right| = \frac{1}{|th|}|\phi(t)|\left|e^{-\iota th} - 1\right| \leq |\phi(t)|\,, \tag{7.12}$$

the inequality following from (7.10) by putting $b = h$ and $a = 0$. As $|\phi(t)|$ is integrable on $\mathbb{R}$, by DCT, it follows that

$$\frac{F(x+h) - F(x)}{h} = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{e^{-\iota tx} - e^{-\iota t(x+h)}}{\iota th}\phi(t)dt\,. \tag{7.13}$$

Since,

$$\lim_{h\to 0} \frac{e^{-\iota t(x+h)} - e^{-\iota tx}}{h} = \frac{d}{dx}e^{-\iota tx} = -\iota te^{-\iota tx}\,,$$

it follows that the integrand in (7.13) converges to $e^{-\iota tx}\phi(t)$ as $h \to 0$. By (7.12), the modulus of the integrand in (7.13) is bounded above by $|\phi(t)|$. DCT allows the limit as $h \to 0$ to be interchanged with the integral in the RHS of (7.13), which completes the proof of Step 2. □

Arguments similar to those in the proof of Theorem 7.1.2 show that f is a continuous function. By Step 2, it follows that for all real $a < b$,

$$\mu(a, b] = F(b) - F(a) = \int_a^b f(x) dx \, ,$$

the second equality following by the fundamental theorem of calculus. This completes the proof. $\square$

The following is an immediate corollary of Theorem 7.6.

Corollary 7.3. *Suppose μ is a probability measure on $\mathbb{R}$ which has a continuous density f. If the CHF ϕ of μ is integrable on $\mathbb{R}$ with respect to the Lebesgue measure, then*

$$\int_{-\infty}^{\infty} e^{-\iota t x} \phi(t) \, dt = 2\pi f(x) \text{ for all } x \in \mathbb{R} \, .$$

Exercise 7.5. *Use Corollary 7.3 to give yet another proof of the fact*

$$\int_{-\infty}^{\infty} e^{-x^2/2} \, dx = \sqrt{2\pi} \, .$$

Example 7.1. *Let*

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

and μ be the probability measure whose density is f. The CHF of μ can be calculated and shown to be

$$\phi(t) = \frac{1}{1 + t^2} \, , t \in \mathbb{R} \, .$$

Since f is continuous, the above corollary implies

$$\int_{-\infty}^{\infty} e^{-\iota t x} \frac{dt}{1 + t^2} = 2\pi f(x) = \pi e^{-|x|} \, , x \in \mathbb{R} \, .$$

Replacing t by $-t$ using the symmetry of ϕ and dividing throughout by π, the above is the same as saying that the CHF ξ of the Cauchy distribution is

$$\xi(x) = e^{-|x|} \, , x \in \mathbb{R} \, .$$

Definition 49. *The CHF of a probability measure μ on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ is a function $\phi : \mathbb{R}^d \to \mathbb{C}$ defined by*

$$\phi(t_1, \ldots, t_d) = \int_{\mathbb{R}^d} \exp\left(\iota \sum_{j=1}^d t_j x_j\right) \mu(dx_1, \ldots, dx_d) \, , (t_1, \ldots, t_d) \in \mathbb{R}^d \, .$$

For an $\mathbb{R}^d$-valued random variable X, its CHF ϕ_X is

$$\phi_X(t) = \mathrm{E}\left(e^{\iota \langle t, X \rangle}\right) \, , t \in \mathbb{R}^d \, .$$

Theorem 7.7 (Inversion theorem on $\mathbb{R}^d$)**.** *If ϕ is the CHF of probability measure μ on $\mathbb{R}^d$, then for $\Delta = [a_1, b_1] \times \ldots \times [a_d, b_d]$ where $a_j < b_j$ for $j = 1, \ldots, d$, and $\mu(\partial\Delta) = 0$ where $\partial\Delta$ is the boundary of Δ,*

$$\mu(\Delta) \\ = (2\pi)^{-d} \lim_{T\to\infty} \int_{[-T,T]^d} \phi(t_1, \ldots, t_d) \left(\prod_{j=1}^{d} \frac{e^{-\iota t_j a_j} - e^{-\iota t_j b_j}}{\iota t_j} \right) dt_1 \ldots dt_d \,.$$

Proof. We shall proceed along the lines of the proof of Theorem 7.4. Let $\Delta = [a_1, b_1] \times \ldots \times [a_d, b_d]$ satisfy the hypotheses. For $T > 0$,

$$\int_{[-T,T]^d} \phi(t_1, \ldots, t_d) \left(\prod_{j=1}^{d} \frac{e^{-\iota t_j a_j} - e^{-\iota t_j b_j}}{\iota t_j} \right) dt_1 \ldots dt_d \tag{7.14}$$

$$= \int_{[-T,T]^d} \left(\prod_{j=1}^{d} \frac{e^{-\iota t_j a_j} - e^{-\iota t_j b_j}}{\iota t_j} \right) \left(\int_{\mathbb{R}^d} \exp\left(\iota \sum_{j=1}^{d} t_j x_j \right) \mu(dx_1, \ldots, dx_d) \right) dt_1 \ldots dt_d$$

$$= \int_{[-T,T]^d} \int_{\mathbb{R}^d} \left(\prod_{j=1}^{d} \frac{e^{\iota t_j (x_j - a_j)} - e^{\iota t_j (x_j - b_j)}}{\iota t_j} \right) \mu(dx_1, \ldots, dx_d)\, dt_1, \ldots, dt_d \,. \tag{7.15}$$

Recall (7.10) to write

$$\int_{[-T,T]^d} \int_{\mathbb{R}^d} \left| \prod_{j=1}^{d} \frac{e^{\iota t_j (x_j - a_j)} - e^{\iota t_j (x_j - b_j)}}{\iota t_j} \right| \mu(dx_1, \ldots, dx_d)\, dt_1, \ldots, dt_d$$

$$\le \int_{[-T,T]^d} \int_{\mathbb{R}^d} \prod_{j=1}^{d} (b_j - a_j)\, \mu(dx_1, \ldots, dx_d)\, dt_1, \ldots, dt_d$$

$$= (2T)^d \prod_{j=1}^{d} (b_j - a_j) < \infty \,.$$

Fubini implies that the quantity in (7.15) equals

$$\int_{\mathbb{R}^d}\int_{[-T,T]^d}\left(\prod_{j=1}^d \frac{e^{\iota t_j(x_j-a_j)}-e^{\iota t_j(x_j-b_j)}}{\iota t_j}\right) dt_1,\ldots,dt_d\,\mu(dx_1,\ldots,dx_d)$$

$$=\int_{\mathbb{R}^d}\left(\prod_{j=1}^d\int_{-T}^T \frac{e^{\iota t_j(x_j-a_j)}-e^{\iota t_j(x_j-b_j)}}{\iota t_j}\,dt_j\right)\mu(dx_1,\ldots,dx_d)$$

$$=\int_{\mathbb{R}^d}\left(\prod_{j=1}^d 2\left(\operatorname{sgn}(x_j-a_j)S(T|x_j-a_j|)-\operatorname{sgn}(x_j-b_j)S(T|x_j-b_j|)\right)\right)$$
$$\mu(dx_1,\ldots,dx_d)\,,$$

(7.8) implying the last equality.

Denote for $x_1,\ldots,x_d\in\mathbb{R}$,

$$\psi_T(x_1,\ldots,x_d)$$
$$=\prod_{j=1}^d 2\left(\operatorname{sgn}(x_j-a_j)S(T|x_j-a_j|)-\operatorname{sgn}(x_j-b_j)S(T|x_j-b_j|)\right)\,,$$

and use Lemma 7.1 to argue

$$\lim_{T\to\infty}\psi_T(x_1,\ldots,x_d)=\pi^d\prod_{j=1}^d\left(\operatorname{sgn}(x_j-a_j)-\operatorname{sgn}(x_j-b_j)\right)$$
$$=:\psi_\infty(x_1,\ldots,x_d)\,.$$

For $(x_1,\ldots,x_d)$ in the interior of Δ, that is, if $a_j<x_j<b_j$ for all j, then

$$\operatorname{sgn}(x_j-a_j)-\operatorname{sgn}(x_j-b_j)=2\,,j=1,\ldots,d\,,$$

and hence

$$\psi_\infty(x_1,\ldots,x_d)=(2\pi)^d\,.$$

On the other hand, if $(x_1,\ldots,x_d)\in\Delta^c$, then there exists j for which either $x_j<a_j$ or $x_j>b_j$ and hence for that j,

$$\operatorname{sgn}(x_j-a_j)-\operatorname{sgn}(x_j-b_j)=0\,.$$

In other words,

$$\psi_\infty(x_1,\ldots,x_d)=0\,,(x_1,\ldots,x_d)\in\Delta^c\,.$$

Thus,

$$\psi_\infty(x_1,\ldots,x_d)=(2\pi)^d\mathbf{1}\left((x_1,\ldots,x_d)\in\Delta\right)\,,(x_1,\ldots,x_d)\in(\partial\Delta)^c\,.\quad(7.16)$$

Letting K as in (7.11), it is immediate that

$$|\psi_T(x_1,\ldots,x_d)| \le (4K)^d\,,$$

which along with DCT implies that as $T\to\infty$,

$$\begin{aligned}\int_{\mathbb{R}^d}\psi_T(x_1,\ldots,x_d)\mu(dx_1,\ldots,dx_d) \to \int_{\mathbb{R}^d}\psi_\infty(x_1,\ldots,x_d)\mu(dx_1,\ldots,x_d) \\ = (2\pi)^d\mu(\Delta)\,,\end{aligned}$$

(7.16) and that $\mu(\partial\Delta)=0$ imply the second line. The left hand side of the first line above is the same as the quantity in (7.14). That is, we have shown

$$\lim_{T\to\infty}\int_{[-T,T]^d}\phi(t_1,\ldots,t_d)\left(\prod_{j=1}^d \frac{e^{-\iota t_j a_j}-e^{-\iota t_j b_j}}{\iota t_j}\right)dt_1\ldots dt_d = (2\pi)^d\mu(\Delta)\,,$$

which is precisely the claimed formula. □

Theorem 7.8 (Uniqueness theorem on $\mathbb{R}^d$)**.** *If μ_1 and μ_2 are probability measures on $\mathbb{R}^d$ with identical CHFs, then $\mu_1=\mu_2$.*

Proof. Let μ_1 and μ_2 be probability measures on $\mathbb{R}^d$ with identical CHFs. To show $\mu_1=\mu_2$, in view of Theorem 4.3.2, it suffices to prove that

$$F_1(x)=F_2(x)\,, x\in\mathbb{R}^d\,, \tag{7.17}$$

where F_1,F_2 are the respective CDFs of μ_1,μ_2, that is,

$$F_i(x)=\mu_i\left((-\infty,x_1]\times\ldots\times(-\infty,x_d]\right)\,, i=1,2,\ x=(x_1,\ldots,x_d)\in\mathbb{R}^d\,.$$

An immediate consequence of Theorem 7.7 is that for a compact rectangle $\Delta\subset\mathbb{R}^d$,

$$\mu_1(\Delta)=\mu_2(\Delta)\text{ if }\mu_1\left(\partial\Delta\right)=0=\mu_2\left(\partial\Delta\right)\,, \tag{7.18}$$

$\partial\Delta$ being the boundary of Δ. Define

$$C_{ij}=\left\{x\in\mathbb{R}:\mu_j\left(\{(x_1,\ldots,x_d)\in\mathbb{R}^d:x_i=x\}\right)>0\right\},$$

for all $i=1,\ldots,d$, $j=1,2$. Since C_{ij} is countable, so is C defined by

$$C=\bigcup_{i=1}^d\bigcup_{j=1}^2 C_{ij}\,.$$

Thus C^c is dense in $\mathbb{R}$ and for $\Delta=[a_1,b_1]\times\ldots\times[a_d,b_d]$,

$$\mu_1\left(\partial\Delta\right)=0=\mu_2\left(\partial\Delta\right)\ \text{ if } a_1,b_1,\ldots,a_d,b_d\in C^c \text{ and } a_i<b_i\,, i=1,\ldots,d\,. \tag{7.19}$$

Proceeding towards (7.17), fix $x_1, \ldots, x_d \in C^c$. Since C^c is dense, there exist $a_n \in C^c$ such that $\wedge_{i=1}^d x_i > a_1 > a_2 > \ldots$ and $a_n \to -\infty$. It follows from (7.18) and (7.19) that

$$\mu_1 ([a_n, x_1] \times \ldots \times [a_n, x_d]) = \mu_2 ([a_n, x_1] \times \ldots \times [a_n, x_d]) \ , n \geq 1 \, .$$

Letting $n \to \infty$, (7.17) follows for $x = (x_1, \ldots, x_d)$ if $x_1, \ldots, x_d \in C^c$. Using the facts that C^c is dense once again, and that F is continuous from above, (7.17) follows for all $x \in \mathbb{R}^d$, which completes the proof. $\square$

The following result is an immediate consequence of Theorem 7.8.

Theorem 7.9 (Cramér-Wold device). *For $\mathbb{R}^d$-valued random variables X and Y,*

$$X \stackrel{d}{=} Y \iff \langle \lambda, X \rangle \stackrel{d}{=} \langle \lambda, Y \rangle \text{ for all } \lambda \in \mathbb{R}^d \, .$$

Proof. Follows from Theorem 7.8. $\square$

Remark 3. *No elementary proof of the Cramér-Wold device, without using CHFs which essentially belong to the domain of Fourier analysis, is known.*

Example 7.2. *Let $Z_1, \ldots, Z_d$ be i.i.d. from standard normal and $Z = (Z_1, \ldots, Z_d)$. Fix a $d \times d$ symmetric non-negative definite (n.n.d.) matrix Σ and $\mu \in \mathbb{R}^d$ and define*

$$X = \mu + \Sigma^{1/2} Z \, , \tag{7.20}$$

where elements of $\mathbb{R}^d$ are to be interpreted as column vectors by convention. Let us calculate the CHF of X. Fix $\lambda \in \mathbb{R}^d$ and write

$$\lambda^T X = \lambda^T \mu + \theta^T Z \, ,$$

where

$$\theta = \Sigma^{1/2} \lambda \, .$$

Recall that $\theta^T Z$ follows $N(0, \|\theta\|^2)$, where $\|\cdot\|$ is the L^2-norm, if $\|\theta\| > 0$; $\theta^T Z$ is degenerate at zero otherwise. Assuming for a moment that $t = \|\theta\| > 0$,

$$\begin{aligned}
\mathrm{E}\left(e^{\iota \theta^T Z}\right) &= \mathrm{E}\left(\exp\left(\iota t \frac{\theta^T Z}{\|\theta\|}\right)\right) \\
\left(\text{because } \frac{\theta^T Z}{\|\theta\|} \sim N(0,1)\right) &= e^{-t^2/2} \\
&= e^{-\|\theta\|^2/2} \\
&= \exp\left(-\frac{1}{2}\theta^T \theta\right) \\
&= \exp\left(-\frac{1}{2}\lambda^T \Sigma \lambda\right) .
\end{aligned}$$

If $\|\theta\| = 0$, that is, θ is the zero vector, then also

$$\mathrm{E}\left(e^{\iota\theta^T Z}\right) = \exp\left(-\frac{1}{2}\lambda^T\Sigma\lambda\right),$$

because both sides equal 1 *in this case. Thus,*

$$\mathrm{E}\left(e^{\iota\lambda^T X}\right) = e^{\iota\lambda^T\mu}\mathrm{E}\left(e^{\iota\theta^T Z}\right) = \exp\left(\iota\lambda^T\mu - \frac{1}{2}\lambda^T\Sigma\lambda\right).$$

In other words, the CHF ϕ_X of X is

$$\phi_X(\lambda) = \exp\left(\iota\lambda^T\mu - \frac{1}{2}\lambda^T\Sigma\lambda\right), \lambda \in \mathbb{R}^d.$$

Definition 50. *An $\mathbb{R}^d$-valued random variable X follows $N_d(\mu, \Sigma)$ for $\mu \in \mathbb{R}^d$ and a $d \times d$ symmetric n.n.d. matrix Σ, if the CHF of X is*

$$\phi_X(\lambda) = \exp\left(\iota\lambda^T\mu - \frac{1}{2}\lambda^T\Sigma\lambda\right), \lambda \in \mathbb{R}^d.$$

The above definition is consistent with Definition 39 in the following sense. If Σ is p.d. and $X \sim N_d(\mu, \Sigma)$ according to Definition 50, then the density of X is f as in Definition 39. Indeed, (7.20) should be compared with (4.10) to see this immediately.

Remark 4. *The $N_d(\mu, \Sigma)$ is called a "singular normal distribution" if Σ is n.n.d. but not p.d. It should be noted that a singular normal distribution in one dimension is a degenerate distribution.*

Exercise 7.6. *Show that a $N_d(\mu, \Sigma)$ distribution has a density if and only if Σ is p.d.*

Theorem 7.10. *For an $\mathbb{R}^d$-valued random variable X, $\mu \in \mathbb{R}^d$ and a $d \times d$ n.n.d. matrix Σ,*

$$X \sim N_d(\mu, \Sigma) \iff \langle\lambda, X\rangle \sim N\left(\lambda^T\mu, \lambda^T\Sigma\lambda\right) \text{ for all } \lambda \in \mathbb{R}^d.$$

Proof. For the "$\Rightarrow$ part", assume $X \sim N_d(\mu, \Sigma)$ and fix $\lambda \in \mathbb{R}^d$. Then for $t \in \mathbb{R}$,

$$\begin{aligned}
\mathrm{E}\left(e^{\iota t\langle\lambda, X\rangle}\right) &= \mathrm{E}\left(e^{\iota\langle t\lambda, X\rangle}\right) \\
&= \exp\left(\iota(t\lambda)^T\mu - \frac{1}{2}(t\lambda)^T\Sigma(t\lambda)\right) \\
&= e^{\iota t\theta - \sigma^2 t^2/2},
\end{aligned}$$

where $\theta = \lambda^T\mu$ and $\sigma^2 = \lambda^T\Sigma\lambda$. As the above is true for all $t \in \mathbb{R}$, Definition 50 shows that $\langle\lambda, X\rangle \sim N(\theta, \sigma^2)$. This proves the "$\Rightarrow$ part".

For the reverse implication, assume that

$$\langle \lambda, X \rangle \sim N\left(\lambda^T \mu, \lambda^T \Sigma \lambda\right) \text{ for all } \lambda \in \mathbb{R}^d .$$

Let $Y \sim N_d(\mu, \Sigma)$. The already proven "$\Rightarrow$ part" shows that

$$\langle \lambda, Y \rangle \sim N\left(\lambda^T \mu, \lambda^T \Sigma \lambda\right) \text{ for all } \lambda \in \mathbb{R}^d .$$

Thus $\langle \lambda, X \rangle \stackrel{d}{=} \langle \lambda, Y \rangle$ for all $\lambda \in \mathbb{R}^d$. The Cramér-Wold device shows

$$X \stackrel{d}{=} Y ,$$

from which the "$\Leftarrow$ part" follows. This completes the proof. $\square$

Exercise 7.7. *If X is a $\mathbb{R}^d$-valued random vector such that for all $\lambda \in \mathbb{R}^d$, $\lambda^T X$ follows one-dimensional normal, show that $X \sim N_d(\mu, \Sigma)$, where μ and Σ are the mean vector and the variance-covariance matrix of X, respectively.*

Exercise 7.8. *For a random variable X with CHF ϕ, show that the following are equivalent.*

1. *The distribution of X is symmetric, that is, $X \stackrel{d}{=} -X$.*
2. *For all $t \in \mathbb{R}$, $\Im(\phi(t)) = 0$, that is, ϕ is a real function.*
3. *The function ϕ is even, that is, $\phi(-t) = \phi(t)$ for all $t \in \mathbb{R}$.*

Exercise 7.9. *If $X_1, X_2, \ldots$ are i.i.d. from the Cauchy distribution, show that there does not exist a random variable Z such that*

$$\frac{1}{n} \sum_{i=1}^{n} X_i \xrightarrow{P} Z \, , n \to \infty .$$

Exercise 7.10. *Suppose X and Y are independent random variables.*

1. *Show that $X + Y$ has a density if either X or Y has a density.*
   Hint. *If f is the density of X, use Theorem 5.7 to show that for fixed $z \in \mathbb{R}$,*
   $$P\left(X + Y \le z | \sigma(Y)\right) = \int_{-\infty}^{z-Y} f(x)\, dx .$$
   The "conditional probability" of an event given a σ-field is the same as the conditional expectation of the indicator of that event.
2. *Show that $X+Y$ has a bounded continuous density if the CHF of either X or Y is integrable on $\mathbb{R}$.*

8 Weak convergence and the central limit theorem

Definition 51. *For probability measures $\mu, \mu_1, \mu_2, \dots$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, μ_n* converges weakly *to μ, or $\mu_n \Rightarrow \mu$ if*

$$\lim_{n\to\infty} \mu_n\left((-\infty, x]\right) = \mu\left((-\infty, x]\right) ,$$

for all $x \in \mathbb{R}$ with $\mu(\{x\}) = 0$. For $\mathbb{R}$-valued random variables $X, X_1, X_2, \dots$, we say X_n converges to X in law, in distribution or weakly*, and denote it by $X_n \Rightarrow X$, if*

$$P \circ X_n^{-1} \Rightarrow P \circ X^{-1} .$$

Henceforth, $\mathcal{B}(\mathbb{R})$ or $\mathcal{B}(\mathbb{R}^d)$ will be the underlying σ-field, depending on the context, unless specifically mentioned otherwise.

Exercise 8.1. *If $X, X_1, X_2, \dots$ are random variables with respective CDFs $F, F_1, F_2, \dots$, show that $X_n \Rightarrow X$ if and only if*

$$\lim_{n\to\infty} F_n(x) = F(x) \text{ for every continuity point } x \text{ of } F .$$

Example 8.1. *If $X_n \sim$ Binomial(n, p_n) where $p_n \in (0,1)$ are such that*

$$\lim_{n\to\infty} np_n = \lambda \in (0, \infty) ,$$

then $X_n \Rightarrow X$ where $X \sim$ Poisson(λ).

Exercise 8.2. *If $X_n \Rightarrow Y$ and $X_n \Rightarrow Z$, show that $Y \stackrel{d}{=} Z$.*

Theorem 8.1. *If $X_n \xrightarrow{P} X$, then $X_n \Rightarrow X$.*

Proof. Fix $x \in \mathbb{R}$ such that $P(X = x) = 0$. It suffices to show for all such x,

$$\lim_{n\to\infty} P(X_n \le x) = P(X \le x) . \tag{8.1}$$

Fix $\varepsilon > 0$. Since $P(X = x) = 0$, there exists $w < x < y$ such that

$$P(X \in [w, y]) \le \varepsilon .$$

Clearly,

$$[X > y] \cap [|X_n - X| \le y - x] \subset [X_n > x] .$$

Take complements of both sides to get

$$[X \le y] \cup [X_n - X| > y - x] \supset [X_n \le x] .$$

Thus,

$$\begin{aligned} P(X_n \le x) &\le P\left([X \le y] \cup [X_n - X| > y - x]\right) \\ &\le P(X \le y) + P(|X_n - X| > y - x) . \end{aligned}$$

Let $n \to \infty$ and use the fact $X_n \xrightarrow{P} X$ to argue

$$\limsup_{n\to\infty} P(X_n \le x) \le P(X \le y) \le P(X \le x) + \varepsilon\,,$$

the right inequality following from the choice of w and y. Since ε is arbitrary, we get

$$\limsup_{n\to\infty} P(X_n \le x) \le P(X \le x)\,.$$

Since

$$[X_n > x] \cap [|X_n - X| \le x - w] \subset [X > w]\,,$$

proceeding along similar lines would yield

$$\liminf_{n\to\infty} P(X_n \le x) \ge P(X \le x)\,,$$

from which (8.1) would follow and would complete the proof. $\square$

Exercise 8.3. *If X follows standard normal and $X_n = -X$, show that*

$$X_n \Rightarrow X \text{ but } X_n \overset{P}{\not\longrightarrow} X\,.$$

Exercise 8.4. *If X is a degenerate random variable, show that*

$$X_n \Rightarrow X \iff X_n \xrightarrow{P} X\,.$$

Theorem 8.2. *For probability measures $\mu_1, \mu_2, \ldots, \mu_\infty$ on $\mathbb{R}$, $\mu_n \Rightarrow \mu_\infty$ if and only if*

$$\lim_{n\to\infty} \int f\, d\mu_n = \int f\, d\mu_\infty\,, \tag{8.2}$$

for every bounded continuous function $f : \mathbb{R} \to \mathbb{R}$.

Proof. For the "if part", assume (8.2) and fix $x \in \mathbb{R}$ with $\mu_\infty(\{x\}) = 0$. Fix $\varepsilon > 0$ and let $w < x < y$ be such that $\mu_\infty([w, y]) \le \varepsilon$. Let $f : \mathbb{R} \to \mathbb{R}$ be the function which is 0 on $[x, \infty)$, 1 on $(-\infty, w]$ and is the line segment joining $(w, 1)$ and $(x, 0)$ on $[w, x]$. Thus f is bounded and continuous and

$$\mathbf{1}_{(-\infty,w]} \le f \le \mathbf{1}_{(-\infty,x]}\,. \tag{8.3}$$

The right inequality above implies that for $n = 1, 2, \ldots$,

$$\mu_n\left((-\infty, x]\right) \ge \int f\, d\mu_n\,.$$

Letting $n \to \infty$,

$$\begin{aligned}
\liminf_{n\to\infty} \mu_n\left((-\infty, x]\right) &\ge \liminf_{n\to\infty} \int f\, d\mu_n \\
\text{(by (8.2))} &= \int f\, d\mu_\infty \\
\text{(the left inequality of (8.3))} &\ge \mu_\infty\left((-\infty, w]\right) \\
&\ge \mu_\infty\left((-\infty, x]\right) - \varepsilon\,,
\end{aligned}$$

the choice of w implying the last line. Arbitrariness of ε shows

$$\liminf_{n\to\infty} \mu_n\left((-\infty,x]\right) \geq \mu_\infty\left((-\infty,x]\right) .$$

A similar argument with $g:\mathbb{R}\to\mathbb{R}$ which is 0 on $[y,\infty)$, 1 on $(-\infty,x]$ and is the line segment joining $(x,1)$ and $(y,0)$ on $[x,y]$ yields the desired upper bound and thus proves $\mu_n \Rightarrow \mu_\infty$. This proves the "if part".

For the "only if part", assume $\mu_n \Rightarrow \mu_\infty$. Let $f:\mathbb{R}\to\mathbb{R}$ be bounded continuous and fix $\varepsilon>0$. Fix $a,b\in\mathbb{R}$ with $a<b$ such that $\mu_\infty(\{a,b\})=0$ and

$$\mu_\infty\left([a,b]^c\right) \leq \varepsilon .$$

Continuity of f implies it is uniformly continuous on $[a,b]$. Thus there exists $\delta>0$ satisfying

$$|f(x)-f(y)| \leq \varepsilon \text{ for all } x,y\in[a,b]\,,\text{ with } |x-y|\leq\delta .$$

Choose $x_0=a<x_1<\ldots<x_k=b$ satisfying $\mu(\{x_0,\ldots,x_k\})=0$ and

$$x_i-x_{i-1}\leq\delta\,, i=1,\ldots,k\,;$$

this is possible because a and b have been chosen to be continuity points of μ_∞. Thus, for $n=1,2,\ldots,\infty$,

$$\begin{aligned}
&\left|\int_{(a,b]} f(x)\mu_n(dx) - \sum_{i=1}^k f(x_i)\mu_n\left((x_{i-1},x_i]\right)\right| \qquad (8.4)\\
&= \left|\sum_{i=1}^k \int_{(x_{i-1},x_i]} \left[f(x)-f(x_i)\right]\mu_n(dx)\right|\\
&\leq \sum_{i=1}^k \int_{(x_{i-1},x_i]} \left|f(x)-f(x_i)\right|\mu_n(dx)\\
&\leq \sum_{i=1}^k \mu_n\left((x_{i-1},x_i]\right) \max_{x\in[x_{i-1},x_i]} |f(x)-f(x_i)|\\
&\leq \varepsilon\sum_{i=1}^k \mu_n\left((x_{i-1},x_i]\right)\\
&= \varepsilon\mu_n\left((a,b]\right) \leq \varepsilon ,
\end{aligned}$$

the inequality in the penultimate line following from the choice of δ and that $x_i-x_{i-1}\leq\delta\,, i=1,\ldots,k$.

Thus, for $n = 1, 2, \ldots$,

$$\begin{aligned}
&\left| \int_{(a,b]} f(x)\, \mu_n(dx) - \int_{(a,b]} f(x)\, \mu_\infty(dx) \right| \\
&\le \left| \int_{(a,b]} f(x) \mu_n(dx) - \sum_{i=1}^k f(x_i) \mu_n\left((x_{i-1}, x_i]\right) \right| \\
&\quad + \left| \int_{(a,b]} f(x) \mu_\infty(dx) - \sum_{i=1}^k f(x_i) \mu_\infty\left((x_{i-1}, x_i]\right) \right| \\
&\quad + \left| \sum_{i=1}^k f(x_i) \mu_n\left((x_{i-1}, x_i]\right) - \sum_{i=1}^k f(x_i) \mu_\infty\left((x_{i-1}, x_i]\right) \right| \\
&\le 2\varepsilon + \left| \sum_{i=1}^k f(x_i) \mu_n\left((x_{i-1}, x_i]\right) - \sum_{i=1}^k f(x_i) \mu_\infty\left((x_{i-1}, x_i]\right) \right| .
\end{aligned}$$

Since $x_0, \ldots, x_k$ are continuity points of μ_∞ which is the weak limit of μ_n,

$$\lim_{n\to\infty} \mu_n\left((x_{i-1}, x_i]\right) = \mu_\infty\left((x_{i-1}, x_i]\right) ,$$

and hence

$$\lim_{n\to\infty} \left| \sum_{i=1}^k f(x_i) \mu_n\left((x_{i-1}, x_i]\right) - \sum_{i=1}^k f(x_i) \mu_\infty\left((x_{i-1}, x_i]\right) \right| = 0 . \tag{8.5}$$

Therefore,

$$\limsup_{n\to\infty} \left| \int_{(a,b]} f(x)\, \mu_n(dx) - \int_{(a,b]} f(x)\, \mu_\infty(dx) \right| \le 2\varepsilon . \tag{8.6}$$

Let $K = \sup_x |f(x)|$ which is finite because f is bounded. Thus,

$$\left| \int_{(a,b]^c} f(x)\, \mu_\infty(dx) \right| \le K \mu_\infty\left((a,b]^c\right) \le K\varepsilon ,$$

and

$$\begin{aligned}
\limsup_{n\to\infty} \left| \int_{(a,b]^c} f(x)\, \mu_n(dx) \right| &\le K \limsup_{n\to\infty} \mu_n\left((a,b]^c\right) \\
(\mu_\infty(\{a,b\}) = 0) &= K \mu_\infty\left((a,b]^c\right) \\
&\le K\varepsilon .
\end{aligned}$$

Combine these with (8.6) to get

$$\limsup_{n\to\infty} \left| \int f\, d\mu_n - \int f\, d\mu_\infty \right| \le 2(K+1)\varepsilon .$$

Since ε is arbitrary, (8.2) follows. This proves the “only if part” and therefore completes the proof. $\square$

Theorem 8.3 (Lévy Continuity theorem)**.** *Let μ_n, μ be probability measures on $\mathbb{R}$ with characteristic functions ϕ_n, ϕ. Then, $\mu_n \Rightarrow \mu$ if and only if*

$$\lim_{n\to\infty} \phi_n(t) = \phi(t) \textit{ for all } t \in \mathbb{R}\,.$$

Proof. The "only if" part follows trivially from Theorem 8.2. For the "if" part, assume that

$$\lim_{n\to\infty} \phi_n(t) = \phi(t) \text{ for all } t \in \mathbb{R}\,.$$

Step 1. Let F_n be the c.d.f. of μ_n. There exist integers $1 \le n_1 < n_2 < \ldots$ such that

$$\lim_{k\to\infty} F_{n_k}(r) \text{ exists for all } r \in \mathbb{Q}\,.$$

Proof of Step 1. Follows immediately from Cantor's diagonal argument because $\mathbb{Q}$ is countable and for all $r \in \mathbb{Q}$, $\{F_n(r) : r \in \mathbb{Q}\}$ is a bounded sequence. □

Step 2. Denote

$$H(r) := \lim_{k\to\infty} F_{n_k}(r),\ r \in \mathbb{Q}\,,$$

and

$$G(x) := \inf\{H(r) : r > x, r \in \mathbb{Q}\}\,.$$

Then, G is a non-decreasing right continuous function.

Proof of Step 2. Non-decreasing is immediate. For right continuity, fix $x \in \mathbb{R}$ and $\varepsilon > 0$. Clearly, there exists $r \in (x,\infty) \cap \mathbb{Q}$ such that

$$H(r) \le G(x) + \varepsilon\,.$$

Clearly,

$$G((x+r)/2) \le H(r) \le G(x) + \varepsilon\,.$$

Thus, G is right continuous at x. □

Step 3. For every continuity point x of G,

$$\lim_{k\to\infty} F_{n_k}(x) = G(x)\,.$$

Proof of Step 3. Fix a continuity point x of G and $\varepsilon > 0$. Therefore, there exist $w < x < y$ such that

$$G(x) - \varepsilon \le G(w) \le G(y) \le G(x) + \varepsilon\,.$$

Let r_1, r_2 be rationals such that $w < r_1 < x < r_2 < y$. Then,

$$\begin{aligned}
G(x) - \varepsilon &\le G(w) \\
&\le H(r_1) \\
&= \lim_{k\to\infty} F_{n_k}(r_1) \\
&\le \liminf_{k\to\infty} F_{n_k}(x) \\
&\le \limsup_{k\to\infty} F_{n_k}(x) \\
&\le \lim_{k\to\infty} F_{n_k}(r_2) \\
&= H(r_2) \\
&\le G(y) \qquad (8.7) \\
&\le G(x) + \varepsilon\,,
\end{aligned}$$

the inequality in (8.7) following from the fact that for all $r \in (y,\infty)\cap\mathbb{Q}$,

$$H(r) \ge H(r_2)\,,$$

that is, $H(r_2)$ is a lower bound of the set of which $G(y)$ is the infimum. Letting $\varepsilon \downarrow 0$ completes the proof of Step 3. $\square$

Step 4. Given $\varepsilon > 0$, there exists a such that

$$\limsup_{n\to\infty} \mu_n\{x : |x| \ge a\} \le \varepsilon\,.$$

Proof of Step 4. Observe that for $u > 0$,

$$\begin{aligned}
\frac{1}{u}\int_{-u}^{u}(1-\phi_n(t))dt &= \frac{1}{u}\int_{-u}^{u}\int_{\mathbb{R}}\left(1-e^{\iota tx}\right)\mu_n(dx)dt \\
&= \int_{\mathbb{R}}\frac{1}{u}\int_{-u}^{u}\left(1-e^{\iota tx}\right)dt\mu_n(dx) \\
\left(\text{Interpreting } \frac{\sin 0}{0} = 1\right) \quad &= 2\int_{\mathbb{R}}\left(1-\frac{\sin ux}{ux}\right)\mu_n(dx) \\
\left(\frac{\sin z}{z} \le 1\,, z\in\mathbb{R}\right) \quad &\ge 2\int_{[|x|\ge 2/u]}\left(1-\frac{\sin ux}{ux}\right)\mu_n(dx) \\
\left(\frac{\sin ux}{ux} \le \frac{|\sin ux|}{|ux|} \le \frac{1}{|ux|}\right) \quad &\ge 2\int_{[|x|\ge 2/u]}\left(1-\frac{1}{|ux|}\right)\mu_n(dx) \\
&\ge 2\int_{[|x|\ge 2/u]}\frac{1}{2}\mu_n(dx) \\
&= \mu_n\left[|x|\ge 2/u\right]\,.
\end{aligned}$$

By DCT, it follows that for all fixed $u > 0$,

$$\lim_{n\to\infty}\frac{1}{u}\int_{-u}^{u}(1-\phi_n(t))dt = \frac{1}{u}\int_{-u}^{u}(1-\phi(t))dt \le \frac{1}{u}\int_{-u}^{u}|1-\phi(t)|\,dt\,.$$

Fix $\varepsilon > 0$. Since ϕ is a characteristic function, it is continuous at 0, and hence there exists $u > 0$ such that

$$|\phi(t) - 1| \leq \varepsilon/2, \; |t| \leq u\,.$$

Letting $a = 2/u$, putting everything together,

$$\begin{aligned}
\limsup_{n\to\infty} \mu_n\{x : |x| \geq a\} &\leq \lim_{n\to\infty} \frac{1}{u}\int_{-u}^{u} (1 - \phi_n(t))dt \\
&\leq \frac{1}{u}\int_{-u}^{u} |1 - \phi(t)|\, dt \\
&\leq \varepsilon\,.
\end{aligned}$$

Thus Step 4 follows. □

Step 5. As $x \to \infty$, $G(x) \to 1$ and as $x \to -\infty$, $G(x) \to 0$.

Proof of Step 5. Since G is non-decreasing, $G(-\infty)$ and $G(\infty)$ exist. Fix $\varepsilon > 0$. Use Step 4 to get $a > 0$ such that

$$\limsup_{n\to\infty} \mu_n\left((-a, a)^c\right) \leq \varepsilon\,.$$

Let $x \leq -a$ be a continuity point of G. Since G is non-decreasing,

$$\begin{aligned}
G(-\infty) &\leq G(x) \\
\text{(By Step 3)} &= \lim_{k\to\infty} F_{n_k}(x) \\
&\leq \limsup_{k\to\infty} F_{n_k}(-a) \\
&= \limsup_{k\to\infty} \mu_{n_k}\left((-\infty, -a]\right) \\
&\leq \limsup_{n\to\infty} \mu_n\left((-a, a)^c\right) \\
&\leq \varepsilon\,.
\end{aligned}$$

Since ε is arbitrary and G is non-negative, it follows that $G(-\infty) = 0$. A similar argument shows that if $y \geq -a$ is a continuity point of G, then $G(y) \geq 1 - \varepsilon$, and hence $G(\infty) = 1$. This proves Step 5. □

Step 6. As $k \to \infty$, $\mu_{n_k} \Longrightarrow \mu$.

Proof of Step 6. Steps 2 and 5 in conjunction with Theorem 1.4 imply there exists a probability measure ν on $\mathbb{R}$ such that

$$\nu(-\infty, x] = G(x), \; x \in \mathbb{R}\,.$$

Step 3 implies that

$$\mu_{n_k} \Longrightarrow \nu\,.$$

By the already proven "only if" part, it follows that

$$\lim_{k\to\infty} \phi_{n_k}(t) = \int e^{\iota tx} \nu(dx) \text{ for all } t \in \mathbb{R}\,.$$

This in conjunction with the hypothesis

$$\lim_{n\to\infty} \phi_n(t) = \phi(t)\,, t \in \mathbb{R}\,,$$

shows

$$\phi(t) = \int e^{\iota tx} \nu(dx) \text{ for all } t \in \mathbb{R}\,.$$

Since ϕ is the CHF of μ, Corollary 7.2 implies

$$\mu = \nu\,.$$

This completes the proof. □

Step 7. As $n \to \infty$, $\mu_n \Longrightarrow \mu$.

Proof of Step 7. Let μ_{m_k} be any subsequence of μ_n. Steps 1 - 6 show that μ_{m_k} has a further subsequence $\mu_{m_{k_l}}$ such that

$$\mu_{m_{k_l}} \Longrightarrow \mu \text{ as } l \to \infty\,.$$

Since this is true for all subsequences μ_{m_k}, it follows that $\mu_n \Longrightarrow \mu$ as $n \to \infty$. □

Step 7 clearly completes the proof of the "only if" part, and thereby proves the theorem. □

Exercise 8.5. *Suppose that $\mu_1, \mu_2, \ldots$ are probability measures on $\mathbb{R}$ with CHFs $\phi_1, \phi_2, \ldots$, respectively. Assume*

$$\lim_{n\to\infty} \phi_n(t) = \phi(t)\,, t \in \mathbb{R}\,.$$

Show that there exists a probability measure μ whose CHF is ϕ if and only if ϕ is continuous at zero and in that case $\mu_n \Rightarrow \mu$.

Theorem 8.4 (Central limit theorem (CLT) on $\mathbb{R}$ for i.i.d.)**.** *Let $X_1, X_2, \ldots$ be i.i.d. random variables with mean μ and variance $\sigma^2 \in (0, \infty)$. Then, as $n \to \infty$,*

$$\frac{\sum_{j=1}^n X_j - n\mu}{n^{1/2}\sigma} \Longrightarrow Z\,,$$

where Z follows standard normal.

Lemma 8.1. *For all $\theta \in \mathbb{R}$,*

$$\left| e^{\iota\theta} - \left(1 + \iota\theta - \frac{1}{2}\theta^2\right) \right| \le 2\min(\theta^2, |\theta|^3)\,.$$

Proof. Notice that

$$\left|e^{\iota\theta} - \left(1 + \iota\theta - \frac{1}{2}\theta^2\right)\right| \;\leq\; \left|\cos\theta - \left(1 - \frac{1}{2}\theta^2\right)\right| + |\sin\theta - \theta| \;.$$

Denote

$$\begin{aligned} R_1 &:= \cos\theta - \left(1 - \frac{1}{2}\theta^2\right) , \\ R_2 &:= \sin\theta - \theta \,. \end{aligned}$$

By Taylor's theorem, there exists ξ, ξ' such that

$$\cos\theta = 1 - \frac{\theta^2}{2} + \frac{\theta^3}{6}\sin\xi \tag{8.8}$$

$$= 1 - \frac{\theta^2}{2}\cos\xi' \,. \tag{8.9}$$

Equations (8.8) and (8.9) respectively show that

$$\begin{aligned} |R_1| &\leq \frac{|\theta|^3}{6} \leq |\theta|^3 , \\ |R_1| &\leq \frac{\theta^2}{2}(1 + |\cos\xi'|) \leq \theta^2 \,. \end{aligned}$$

Therefore,

$$|R_1| \leq \min(\theta^2, |\theta|^3) \,.$$

Applying Taylor to $\sin\theta$ shows the existence of η, η' satisfying

$$\begin{aligned} \sin\theta &= \theta - \frac{\theta^3}{6}\cos\eta \\ &= \theta - \frac{\theta^2}{2}\sin\eta' \,. \end{aligned}$$

Thus,

$$|R_2| \leq \min(\theta^2, |\theta|^3) ,$$

and this completes the proof. □

Lemma 8.2. *For $y, z \in \mathbb{C}$ with $|y| \vee |z| \leq 1$, and $n \in \mathbb{N}$,*

$$|y^n - z^n| \leq n|y - z| \,.$$

Proof. The observation

$$\begin{aligned} |y^n - z^n| &= \left|(y - z)\sum_{j=0}^{n-1} y^{n-1-j}z^j\right| \\ &= |y - z|\left|\sum_{j=0}^{n-1} y^{n-1-j}z^j\right| \\ &\leq n|y - z| , \end{aligned}$$

completes the proof. □

Proof of Theorem 8.4. WLOG, we assume that $\mu = 0$ and $\sigma = 1$. Then, what needs to be shown is that

$$n^{-1/2}\sum_{j=1}^n X_j \Longrightarrow N(0,1)\,.$$

Let ϕ be the characteristic function of X_1. In view of the Lévy continuity theorem, what needs to be shown is that

$$\lim_{n\to\infty}\phi(t/\sqrt{n})^n = e^{-t^2/2} \text{ for all } t \in \mathbb{R}\,. \tag{8.10}$$

Fix $t \in \mathbb{R}$, and notice that

$$\begin{aligned}
&\left|\mathrm{E}[e^{\iota t X_1/\sqrt{n}}] - \mathrm{E}\left[1 + \frac{\iota t}{\sqrt{n}}X_1 - \frac{t^2}{2n}X_1^2\right]\right| \\
\leq\ & \mathrm{E}\left|e^{\iota t X_1/\sqrt{n}} - \left[1 + \frac{\iota t}{\sqrt{n}}X_1 - \frac{t^2}{2n}X_1^2\right]\right| \\
\text{(by Lemma 8.1)}\quad \leq\ & 2\mathrm{E}\left(\min(t^2X_1^2/n, |t|^3|X_1|^3/n^{3/2})\right)\,,
\end{aligned}$$

that is,

$$\left|\phi(t/\sqrt{n}) - \left(1 - \frac{t^2}{2n}\right)\right| \leq 2\mathrm{E}\left(\min(t^2X_1^2/n, |t|^3|X_1|^3/n^{3/2})\right)\,.$$

By Lemma 8.2, it follows that for $n > t^2$,

$$\begin{aligned}
\left|\phi(t/\sqrt{n})^n - \left(1 - \frac{t^2}{2n}\right)^n\right| &\leq n\left|\phi(t/\sqrt{n}) - \left(1 - \frac{t^2}{2n}\right)\right| \\
&\leq 2\mathrm{E}\left(\min(t^2X_1^2, |t|^3|X_1|^3/n^{1/2})\right)\,.
\end{aligned}$$

By DCT, the extreme RHS goes to 0 as $n \to \infty$. Thus, (8.10) follows, and completes the proof. □

We now proceed towards the multivariate CLT, that is, CLT on $\mathbb{R}^d$, for which, weak convergence on $\mathbb{R}^d$ is to be defined.

Definition 52. *For probability measures $\mu, \mu_1, \mu_2, \ldots$ on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, $\mu_n \Rightarrow \mu$ if*

$$\lim_{n\to\infty}\int_{\mathbb{R}^d} f\,d\mu_n = \lim_{n\to\infty}\int_{\mathbb{R}^d} f\,d\mu\,,$$

for all bounded continuous $f : \mathbb{R}^d \to \mathbb{R}$. If $X, X_1, X_2, \ldots$ are $\mathbb{R}^d$-valued random variables, then $X_n \Rightarrow X$ if $P \circ X_n^{-1} \Rightarrow P \circ X$.

Theorem 8.2 shows the above definition is consistent with Definition 51 in the case $d = 1$. The advantage of the above definition is that the weak limit, if exists, can easily be shown to be unique, as claimed in the following exercise. For a probability measure on $\mathbb{R}^d$, the underlying σ-field $\mathcal{B}(\mathbb{R}^d)$ will not be mentioned henceforth.

Exercise 8.6. *If μ_n, μ, ν are probability measures on $\mathbb{R}^d$ such that $\mu_n \Rightarrow \mu$ and $\mu_n \Rightarrow \nu$, show that $\mu = \nu$.*
Hint. *Use Theorem 7.8.*

Another convenience of Definition 52 is that the following result now becomes automatic.

Theorem 8.5 (Continuous mapping theorem)**.** *Suppose $X_1, X_2, \ldots, X_\infty$ are $\mathbb{R}^{d_1}$-valued random variables and $X_n \Rightarrow X_\infty$. If $g : \mathbb{R}^{d_1} \to \mathbb{R}^{d_2}$ is a continuous function, then*

$$g(X_n) \Rightarrow g(X_\infty)\,,$$

as $\mathbb{R}^{d_2}$-valued random variables.

Proof. According to the definition, it suffices to check that for any bounded continuous $f : \mathbb{R}^{d_2} \to \mathbb{R}$,

$$\lim_{n\to\infty} \mathrm{E}\left(f \circ g(X_n)\right) = \mathrm{E}\left(f \circ g(X_\infty)\right)\,.$$

Fix such f. Since f, g are continuous, so is $f \circ g$. As f is bounded, so is $f \circ g$. Thus $f \circ g : \mathbb{R}^{d_1} \to \mathbb{R}$ is bounded continuous and the definition of weak convergence implies the above. This completes the proof. □

The following result shows, among other things, that if weak convergence on $\mathbb{R}^d$ were defined by CDFs as in Definition 51, then that would have been equivalent to Definition 52.

Theorem 8.6 (Portmanteau theorem)**.** *For probability measures $\mu_1, \mu_2, \ldots$ $\ldots, \mu_\infty$ on $\mathbb{R}^d$ with respective CDFs $F_1, F_2, \ldots, F_\infty$, the following are equivalent.*

1. *As $n \to \infty$, $\mu_n \Rightarrow \mu_\infty$.*
2. *For any closed set $F \subset \mathbb{R}^d$,*
$$\limsup_{n\to\infty} \mu_n(F) \le \mu_\infty(F)\,.$$
3. *For any open set $U \subset \mathbb{R}^d$,*
$$\liminf_{n\to\infty} \mu_n(U) \ge \mu_\infty(U)\,.$$
4. *For $A \in \mathcal{B}(\mathbb{R}^d)$ with $\mu_\infty(\partial A) = 0$, where ∂A is the boundary of A,*
$$\lim_{n\to\infty} \mu_n(A) = \mu_\infty(A)\,.$$
5. *For all $x \in \mathbb{R}^d$ at which F_∞ is continuous,*
$$\lim_{n\to\infty} F_n(x) = F_\infty(x)\,.$$

The proof uses the following exercise.

Exercise 8.7. *If μ is a probability measure on $\mathbb{R}^d$ with CDF F, show that for $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$,*

$$F \text{ is continuous at } x \iff \mu(\partial E_x) = 0\,,$$

where $E_x = (-\infty, x_1] \times \ldots \times (-\infty, x_d]$. Equivalently, show that if F is the CDF of an $\mathbb{R}^d$-valued random variable $(X_1, \ldots, X_d)$, then for every $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$,

$$F \text{ is continuous at } x \iff P\left(\bigvee_{i=1}^d (X_i - x_i) = 0\right) = 0\,.$$

Proof of Portmanteau theorem. Since $\mu_1, \mu_2, \ldots, \mu_\infty$ are probability measures, it follows trivially that

$$2 \iff 3\,.$$

Thus it suffices to show 1⇒2⇒4⇒5⇒1.

Proof of 1⇒2. Assume 1, that is, $\mu_n \Rightarrow \mu_\infty$. Let $\|\cdot\|$ be any norm on $\mathbb{R}^d$ and define

$$d(F, x) = \inf\{\|x - y\| : y \in F\}\,, x \in \mathbb{R}^d\,.$$

Fix $\varepsilon > 0$ and define $f : \mathbb{R}^d \to \mathbb{R}$ by

$$f_\varepsilon(x) = 1 - \left(1 \wedge \varepsilon^{-1} d(F, x)\right)\,, x \in \mathbb{R}^d\,.$$

Since $d(F, \cdot)$ is a continuous function, so is f. Further, $0 \le f_\varepsilon \le 1$,

$$f_\varepsilon(x) = 1\,, \text{ if } x \in F\,,$$

and

$$f_\varepsilon(x) = 0\,, \text{ if } d(F, x) \ge \varepsilon\,.$$

In other words,

$$\mathbf{1}_F(x) \le f_\varepsilon(x) \le \mathbf{1}\left(d(F, x) < \varepsilon\right)\,.$$

Thus,

$$\begin{aligned}
\limsup_{n\to\infty} \mu_n(F) &= \limsup_{n\to\infty} \int \mathbf{1}_F(x) \mu_n(dx) \\
&\le \lim_{n\to\infty} \int f_\varepsilon(x) \mu_n(dx) \\
(\text{as } \mu_n \Rightarrow \mu_\infty) &= \int f_\varepsilon(x) \mu_\infty(dx) \\
&\le \mu_\infty\left(\{x \in \mathbb{R}^d : d(F, x) < \varepsilon\}\right)\,.
\end{aligned}$$

As $\varepsilon \downarrow 0$,

$$\{x \in \mathbb{R}^d : d(F, x) < \varepsilon\} \downarrow \{x \in \mathbb{R}^d : d(F, x) = 0\} = F\,,$$

the set theoretic equality following from the fact that F is closed. Thus,

$$\lim_{\varepsilon \downarrow 0} \mu_\infty \Big(\{x \in \mathbb{R}^d : d(F, x) < \varepsilon\} \Big) = \mu_\infty(F)\,.$$

Therefore,

$$\limsup_{n\to\infty} \mu_n(F) \le \mu_\infty(F)\,.$$

Hence it follows that 1⇒2. □

Proof of 2⇒4. From the equivalence of 2 and 3, which is indeed a tautology, assume that

$$\limsup_{n\to\infty} \mu_n(F) \le \mu_\infty(F)\,, F \subset \mathbb{R}^d \text{ closed}\,, \tag{8.11}$$

and

$$\liminf_{n\to\infty} \mu_n(U) \ge \mu_\infty(U)\,, U \subset \mathbb{R}^d \text{ open}\,. \tag{8.12}$$

Fix $A \in \mathcal{B}(\mathbb{R}^d)$ such that $\mu_\infty(\partial A) = 0$, that is,

$$\mu_\infty(\bar{A}) = \mu_\infty(A^\circ) = \mu_\infty(A)\,, \tag{8.13}$$

where $\bar{A}$ and A° are the closure and interior of A, respectively. Invoke (8.12) with $U = A^\circ$ to get

$$\begin{aligned}
\mu_\infty(A^\circ) &\le \liminf_{n\to\infty} \mu_n(A^\circ) \\
(\text{as } A^\circ \subset A) &\le \liminf_{n\to\infty} \mu_n(A) \\
&\le \limsup_{n\to\infty} \mu_n(A) \\
&\le \limsup_{n\to\infty} \mu_n(\bar{A}) \\
&\le \mu_\infty(\bar{A})\,,
\end{aligned}$$

(8.11) implying the last line. This in conjunction with (8.13) shows

$$\lim_{n\to\infty} \mu_n(A) = \mu_\infty(A)\,.$$

Thus, 2⇒4. □

Proof of 4⇒5. Assume 4. Let $x = (x_1, \ldots, x_d)$ be a continuity point of F_∞. Exc 8.7 shows that

$$\mu_\infty(\partial E_x) = 0\,,$$

where $E_x = (-\infty, x_1] \times \ldots \times (-\infty, x_d]$. The hypothesis 4 which has been assumed shows that

$$\lim_{n\to\infty} \mu_n(E_x) = \mu_\infty(E_x)\,,$$

which is exactly the same as

$$\lim_{n\to\infty} F_n(x) = F_\infty(x)\,.$$

Thus 4⇒5. $\square$

Proof of 5⇒1. Assume 5. Let

$$\mathcal{C} = \{x \in \mathbb{R}^d : F_\infty \text{ is countinuous at } x\}\,.$$

The assumption 5 immediately implies

$$\lim_{n\to\infty} \Delta_R F_n = \Delta_R F_\infty\,, R = \prod_{i=1}^d (a_i, b_i] \text{ if } \{a_1, b_1\} \times \ldots \times \{a_d, b_d\} \subset \mathcal{C}\,, \tag{8.14}$$

where $\Delta_R F$ is as in (4.1). Recall Exc 3.3, a restatement of which is that

$$\mu_n(R) = \Delta_R F_n\,, n = 1, \ldots, \infty\,. \tag{8.15}$$

Let

$$C_i = \left\{z \in \mathbb{R} : \mu_\infty\left(\{(x_1, \ldots, x_d) \in \mathbb{R}^d : x_i = z\}\right) > 0\right\}\,, i = 1, \ldots, d\,.$$

Clearly, $C_1, \ldots, C_d$ are countable sets and hence

$$D = (C_1 \cup \ldots \cup C_d)^c$$

is dense in $\mathbb{R}$. It is immediate that for $(z_1, \ldots, z_d) \in D^d$,

$$\mu_\infty\left(\{(x_1, \ldots, x_d) \in \mathbb{R}^d : x_i = z_i \text{ for some } i = 1, \ldots, d\}\right) = 0\,.$$

Thus for such $(z_1, \ldots, z_d)$,

$$\mu_\infty\Big(\partial\big((-\infty, z_1] \times \ldots \times (-\infty, z_d]\big)\Big) = 0\,.$$

In view of Exc 8.7, this means $D^d \subset \mathcal{C}$. Combine this with (8.14) and (8.15) to get

$$\lim_{n\to\infty} \mu_n(R) = \mu_\infty(R)\,, R = \prod_{i=1}^d (a_i, b_i]\,, a_1, b_1, \ldots, a_d, b_d \in D\,. \tag{8.16}$$

Fix a bounded continuous $f : \mathbb{R}^d \to \mathbb{R}$ and $\varepsilon > 0$. Fix $a, b \in D$, which is dense in $\mathbb{R}$, such that

$$\mu_\infty\left((a, b]^d\right) \geq 1 - \varepsilon\,. \tag{8.17}$$

As $[a,b]^d$ is a compact set, f is uniformly continuous there. Hence there exists $\delta > 0$ such that

$$|f(x) - f(y)| \le \varepsilon\,, \text{ for all } x, y \in [a,b]^d, \|x - y\| \le \delta\,, \tag{8.18}$$

where $\|(z_1, \ldots, z_d)\| = |z_1| \vee \ldots \vee |z_d|$ is the max norm on $\mathbb{R}^d$. Let $a = x_0 < x_1 < \ldots < x_k = b$ be such that $x_0, \ldots, x_k \in D$ and

$$x_i - x_{i-1} \le \delta\,, i = 1, \ldots, k\,.$$

Set

$$\mathcal{H} = \left\{ \prod_{j=1}^{d} (x_{i_j - 1}, x_{i_j}] : 1 \le i_1, \ldots, i_d \le k \right\}.$$

A consequence of (8.16) is that

$$\lim_{n\to\infty} \mu_n(R) = \mu_\infty(R)\,, R \in \mathcal{H}\,.$$

For $R \in \mathcal{H}$, (8.18) and that $\|\cdot\|$ has been chosen to be the max-norm imply

$$|f(y) - f(z)| \le \varepsilon\,, \text{ for all } y, z \in R\,.$$

Since the k^d many rectangles in $\mathcal{H}$ are disjoint and their union is $(a,b]^d$, we get

$$\int_{(a,b]^d} f\, d\mu_n = \sum_{R\in\mathcal{H}} \int_R f\, d\mu_n\,, n = 1, 2, \ldots, \infty\,.$$

Proceeding like in (8.4)-(8.5) with the help of the above three claims, the analogue of (8.6) can be shown, which is

$$\limsup_{n\to\infty} \left| \int_{(a,b]^d} f\, d\mu_n - \int_{(a,b]^d} f\, d\mu_\infty \right| \le 2\varepsilon\,.$$

Finally, (8.16) also implies

$$\lim_{n\to\infty} \mu_n\left((a,b]^d\right) = \mu_\infty\left((a,b]^d\right)\,,$$

which in conjunction with (8.17) shows

$$\limsup_{n\to\infty} \left| \int_{\left((a,b]^d\right)^c} f\, d\mu_n \right| \le K\varepsilon\,,$$

where $K = \sup_{x\in\mathbb{R}^d} |f(x)|$ which is finite because f is bounded. Trivially,

$$\left| \int_{\left((a,b]^d\right)^c} f\, d\mu_\infty \right| \le K\varepsilon\,,$$

Thus

$$\limsup_{n\to\infty}\left|\int_{\mathbb{R}^d} f\,d\mu_n - \int_{\mathbb{R}^d} f\,d\mu_\infty\right| \le 2(K+1)\varepsilon\,.$$

Since ε is arbitrary, it follows that

$$\lim_{n\to\infty}\int_{\mathbb{R}^d} f\,d\mu_n = \int_{\mathbb{R}^d} f\,d\mu_\infty\,.$$

This being true for any bounded continuous f, $\mu_n \Rightarrow \mu_\infty$. Thus, 5⇒1. □

The proof of Portmanteau theorem is now complete. □

Definition 53. *A sequence $\{\mu_n : n = 1, 2, \ldots\}$ of probability measures on $\mathbb{R}^d$ is* tight *if given $\varepsilon > 0$ there exists a compact set $K \subset \mathbb{R}^d$ such that*

$$\liminf_{n\to\infty}\mu_n(K) \ge 1 - \varepsilon\,.$$

The following result connects tightness with weak convergence.

Theorem 8.7. *If $\{\mu_n : n = 1, 2, \ldots\}$ is a tight sequence of probability measures on $\mathbb{R}^d$, then there exists a subsequence $\{\mu_{n_k} : k = 1, 2, \ldots\}$ and a probability measure μ on $\mathbb{R}^d$ such that*

$$\mu_{n_k} \Rightarrow \mu\,, k \to \infty\,.$$

Proof. We shall proceed like in the proof of Lévy continuity theorem. Let F_n be the CDF of μ_n, that is,

$$F_n\left(x\right) = \mu_n\left(\left(-\infty, x_1\right] \times \ldots \times \left(-\infty, x_d\right]\right)\,, n \in \mathbb{N},\, x = (x_1, \ldots, x_d) \in \mathbb{R}^d\,.$$

As $\mathbb{Q}^d$ is a countable set and for every $r \in \mathbb{Q}^d$, $\{F_n(r) : n = 1, 2, \ldots\}$ is a bounded sequence of real numbers, there exist $1 \le n_1 < n_2 < \ldots$ such that

$$\lim_{k\to\infty} F_{n_k}(r) \text{ exists for all } r \in \mathbb{Q}^d\,.$$

Define

$$G(r) = \lim_{k\to\infty} F_{n_k}(r)\,, r \in \mathbb{Q}^d\,,$$

and $F : \mathbb{R}^d \to [0, 1]$ by

$$F(x_1, \ldots, x_d) = \inf\left\{G(r_1, \ldots, r_d) : r_1 > x_1, \ldots, r_d > x_d,\, r_1, \ldots, r_d \in \mathbb{Q}\right\}\,.$$

We shall show that F is a CDF, that is, it satisfies the assumptions of Theorem 4.2, from which it would follows that F induces a probability measure μ on $\mathbb{R}^d$. It will be shown that $\mu_{n_k} \Rightarrow \mu$. This is achieved in the following few steps.

Step 1. The function F is continuous from above, that is,

$$\lim_{y_1\downarrow x_1, \ldots, y_d\downarrow x_d} F(y_1, \ldots, y_d) = F(x_1, \ldots, x_d)\,, \text{ for all } x_1, \ldots, x_d \in \mathbb{R}\,.$$

Proof of Step 1. The definition of F implies that for fixed $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ and $\varepsilon > 0$, there exist rationals $r_1 > x_1, \ldots, r_d > x_d$ such that

$$G(r_1, \ldots, r_d) \le F(x_1, \ldots, x_d) + \varepsilon\,.$$

Define

$$\delta = \frac{1}{2} \bigwedge_{i=1}^{d} (r_i - x_i)\,.$$

For $y \in [x_1, x_1 + \delta] \times \ldots \times [x_d, x_d + \delta]$, once again, the definition of F implies that

$$\begin{aligned} F(x) &\le F(y) \\ &\le G(r_1, \ldots, r_d) \\ &\le F(x) + \varepsilon\,. \end{aligned}$$

Thus Step 1 follows. □

Step 2. For $R = (a_1, b_1] \times \ldots \times (a_d, b_d]$ where $-\infty < a_i < b_i < \infty$ for $i = 1, \ldots, d$,

$$\Delta_R F \ge 0\,.$$

Proof of Step 2. Fix R as above and $\varepsilon > 0$. Let $x = (x_1, \ldots, x_d) \in E = \{a_1, b_1\} \times \ldots \times \{a_d, b_d\}$. There exist rationals $r_1 > x_1, \ldots, r_d > x_d$ such that

$$G(r_1, \ldots, r_d) \le F(x_1, \ldots, x_d) + \varepsilon 2^{-d}\,.$$

Set

$$\delta_x = \bigwedge_{i=1}^{d} (r_i - x_i)\,.$$

Since G is non-decreasing by definition, it thus follows that

$$F(x) \le G(s_1, \ldots, s_d) \le G(r_1, \ldots, r_d) \le F(x) + \varepsilon$$

for all $s_1, \ldots, s_d \in \mathbb{Q}$ with $x_i \le s_i \le x_i + \delta_x\,, i = 1, \ldots, d\,.$

Taking

$$\delta = \left(\bigwedge_{x \in E} \delta_x \right) \wedge \left(\frac{1}{2} \bigwedge_{i=1}^{d} (b_i - a_i) \right)\,,$$

it thus follows that for all $x = (x_1, \ldots, x_d) \in E$,

$$|F(x) - G(s)| \le \varepsilon 2^{-d}\,, s \in \mathbb{Q}^d \cap ([x_1, x_1 + \delta] \times \ldots \times [x_d, x_d + \delta])\,. \quad (8.19)$$

Choose $t_i \in [a_i, a_i + \delta] \cap \mathbb{Q}$ and $u_i \in [b_i, b_i + \delta] \cap \mathbb{Q}$ for $i = 1, \ldots, d$. Since $\delta < b_i - a_i$ for all i, $t_i \le a_i + \delta < b_i \le u_i$ for all i. Letting

$$R' = (t_1, u_1] \times \ldots \times (t_d, u_d]\,,$$

(8.19) implies

$$|\Delta_R F - \Delta_{R'} G| \leq \varepsilon \,.$$

Since

$$\Delta_{R'} G = \lim_{k\to\infty} \Delta_{R'} F_{n_k} \geq 0 \,,$$

it follows that

$$\Delta_R F \geq -\varepsilon \,.$$

As ε is arbitrary, Step 2 follows. □

Step 3. As $x_1 \to \infty, \ldots, x_d \to \infty$, $F(x_1, \ldots, x_d) \to 1$. On the other hand, as $\bigwedge_{i=1}^d x_i \to -\infty$, $F(x_1, \ldots, x_d) \to 0$.

Proof of Step 3. This is the only step in which tightness of $\{\mu_n\}$ is used. Since $0 \leq F(x) \leq 1$ for all $x \in \mathbb{R}^d$, it suffices to show that for $\varepsilon > 0$ there exists $a, b \in \mathbb{R}$ such that for $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$,

$$F(x) \geq 1 - \varepsilon \text{ if } \bigwedge_{i=1}^d x_i > b \,, \tag{8.20}$$

and

$$F(x) \leq \varepsilon \text{ if } \bigwedge_{i=1}^d x_i < a \,. \tag{8.21}$$

Fix $\varepsilon > 0$. Tightness implies there exists a compact set K such that

$$\liminf_{n\to\infty} \mu_n(K) \geq 1 - \varepsilon \,.$$

Since K is compact and hence bounded, there exist $a, b \in \mathbb{Q}$ with $a < b$ and $K \subset (a, b]^d$. Thus

$$\begin{aligned} G(b, \ldots, b) &= \lim_{k\to\infty} F_{n_k}(b, \ldots, b) \\ &= \lim_{k\to\infty} \mu_{n_k}\left((-\infty, b]^d\right) \\ &\geq \liminf_{n\to\infty} \mu_n(K) \\ &\geq 1 - \varepsilon \,. \end{aligned}$$

The definition of F and that G is non-decreasing imply that

$$F(z) \geq G(v) \text{ if } z = (z_1, \ldots, z_d) \in \mathbb{R}^d, \text{ and } v \in \mathbb{Q}^d \cap \prod_{i=1}^d (-\infty, z_i] \,. \tag{8.22}$$

Thus,

$$F(x_1, \ldots, x_d) \geq G(b, \ldots, b) \text{ for all } x_1 \geq b, \ldots, x_d \geq b \,,$$

showing that (8.20) holds. Fix $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ with $x_j < a$ for some fixed j. Let $r \in \mathbb{Q}$ be such that $r > (x_1 \vee \ldots \vee x_d)$ and define $y = (y_1, \ldots, y_d)$ where

$$y_i = \begin{cases} r, & i \neq j\,, \\ a, & i = j\,. \end{cases}$$

Thus $y \in \mathbb{Q}^d$ and $y_i > x_i$ for all i, which shows

$$\begin{aligned} F(x) &\leq G(y) \\ &= \lim_{k\to\infty} \mu_{n_k}\left((-\infty, y_1] \times \ldots \times (-\infty, y_d]\right) \\ &\leq \limsup_{n\to\infty} \mu_n\left((-\infty, y_1] \times \ldots \times (-\infty, y_d]\right) \\ &\leq \limsup_{n\to\infty} \mu_n(K^c)\,, \end{aligned}$$

the last line following from the argument that $y_j = a$ and $K \subset (a, b]^d$ show

$$\left((-\infty, y_1] \times \ldots \times (-\infty, y_d]\right) \cap K = \emptyset\,,$$

and hence

$$(-\infty, y_1] \times \ldots \times (-\infty, y_d] \subset K^c\,.$$

Finally,

$$\limsup_{n\to\infty} \mu_n(K^c) = 1 - \liminf_{n\to\infty} \mu_n(K) \leq \varepsilon\,,$$

which establishes (8.21). This proves Step 3. □

Steps 1-3 in conjunction with Theorem 4.2 show that there exists a probability measure μ on $\mathbb{R}^d$ satisfying

$$\mu\left((-\infty, x_1] \times \ldots \times (-\infty, x_d]\right) = F(x)\,, x = (x_1, \ldots, x_d) \in \mathbb{R}^d\,.$$

To complete the proof by showing $\mu_{n_k} \Rightarrow \mu$, $k \to \infty$, in view of the Portmanteau theorem, it suffices to prove that

$$\lim_{k\to\infty} F_{n_k}(x) = F(x)\,, \tag{8.23}$$

for every continuity point x of F. Fix such $x = (x_1, \ldots, x_d)$ and $\varepsilon > 0$. By continuity, there exist $w_i < x_i < y_i$ for $i = 1, \ldots, d$ such that

$$F(w_1, \ldots, w_d) \geq F(x_1, \ldots, x_d) - \varepsilon\,,$$

and

$$F(y_1, \ldots, y_d) \leq F(x_1, \ldots, x_d) + \varepsilon\,.$$

Let $r_1, \ldots, r_d, s_1, \ldots, s_d \in \mathbb{Q}$ be such that $w_i < r_i < x_i < s_i < y_i$ for $i = 1, \ldots, d$. Thus,

$$
\begin{aligned}
F(x) - \varepsilon &\le F(w_1, \ldots, w_d) \\
\text{(definition of } F) &\le G(r_1, \ldots, r_d) \\
&= \lim_{k\to\infty} F_{n_k}(r_1, \ldots, r_d) \\
&\le \liminf_{k\to\infty} F_{n_k}(x) \\
&\le \limsup_{k\to\infty} F_{n_k}(x) \\
&\le \lim_{k\to\infty} F_{n_k}(s_1, \ldots, s_d) \\
&= G(s_1, \ldots, s_d) \\
\text{(by (8.22))} &\le F(y_1, \ldots, y_d) \\
&\le F(x) + \varepsilon \,.
\end{aligned}
$$

Since ε is arbitrary, (8.23) follows, which completes the proof of Theorem 8.7. □

The following is a generalization of Theorem 7.9, and hence this also is called the Cramér-Wold device.

Theorem 8.8 (Cramér-Wold device for weak convergence)**.** *For $\mathbb{R}^d$-valued random variables $X_1, X_2, \ldots X_\infty$, $X_n \Rightarrow X_\infty$ if and only if*

$$\langle \lambda, X_n \rangle \Rightarrow \langle \lambda, X_\infty \rangle \text{ for all } \lambda \in \mathbb{R}^d \,. \tag{8.24}$$

The proof uses the following exercise from real analysis.

Exercise 8.8. *1. If $F, F_1, F_2, \ldots$ are functions from $\mathbb{R}^d$ to $[0,1]$, show that*

$$\lim_{n\to\infty} F_n(x) = F(x)$$

for every continuity point x of F if and only if every subsequence $\{F_{n_k}\}$ of $\{F_n\}$ has a further subsequence $\{F_{n_{k_l}}\}$ such that

$$\lim_{l\to\infty} F_{n_{k_l}}(x) = F(x)$$

for every continuity point x of F.
2. Hence or otherwise, prove that for probability measures $\mu, \mu_1, \mu_2, \ldots$ on $\mathbb{R}^d$, $\mu_n \Rightarrow \mu$ if and only if every subsequence $\{\mu_{n_k}\}$ of $\{\mu_n\}$ has a further subsequence $\{\mu_{n_{k_l}}\}$ such that

$$\mu_{n_{k_l}} \Rightarrow \mu \,, l \to \infty \,.$$

Proof of Theorem 8.8. The "only if" part follows trivially from the continuous mapping theorem because for a fixed $\lambda \in \mathbb{R}^d$, $x \mapsto \langle \lambda, x \rangle$ is a continuous map from $\mathbb{R}^d$ to $\mathbb{R}$.

Conversely, assume (8.24). Denote

$$X_n = (X_{n1}, \ldots, X_{nd}) \, , n = 1, \ldots, \infty \, .$$

For fixed $i \in \{1, \ldots, d\}$, letting λ be the vector whose i-th coordinate is 1 and rest are 0, (8.24) implies

$$X_{ni} \Rightarrow X_{\infty i} \, , i = 1, \ldots, d \, . \tag{8.25}$$

We shall first show that $\{P \circ X_n^{-1} : n = 1, 2, \ldots\}$ is tight, that is, given $\varepsilon > 0$, a compact $K \subset \mathbb{R}^d$ will be obtained satisfying

$$\liminf_{n \to \infty} P(X_n \in K) \geq 1 - \varepsilon \, . \tag{8.26}$$

Fix $\varepsilon > 0$. Let $0 < \alpha < \infty$ be such that

$$P(|X_{\infty i}| < \alpha) \geq 1 - \frac{\varepsilon}{d} \, , i = 1, \ldots, d \, .$$

Use 3 of the Portmanteau theorem with $d = 1$, $U = (-\alpha, \alpha)$ and (8.25) to get

$$\liminf_{n \to \infty} P(|X_{ni}| < \alpha) \geq P(|X_{\infty i}| < \alpha) \geq 1 - \frac{\varepsilon}{d} \, ,$$

a consequence of which is

$$\limsup_{n \to \infty} P(|X_{ni}| \geq \alpha) \leq \frac{\varepsilon}{d} \, , i = 1, \ldots, d \, . \tag{8.27}$$

Let $K = [-\alpha, \alpha]^d$. Thus,

$$\begin{aligned}
\liminf_{n \to \infty} P(X_n \in K) &= 1 - \limsup_{n \to \infty} P(X_n \in K^c) \\
&= 1 - \limsup_{n \to \infty} P\left(\bigcup_{i=1}^d [|X_{ni}| > \alpha] \right) \\
&\geq 1 - \limsup_{n \to \infty} \sum_{i=1}^d P\left(|X_{ni}| > \alpha\right) \\
&\geq 1 - \sum_{i=1}^d \limsup_{n \to \infty} P\left(|X_{ni}| > \alpha\right) \\
&\geq 1 - \sum_{i=1}^d \limsup_{n \to \infty} P\left(|X_{ni}| \geq \alpha\right) \\
&\geq 1 - \sum_{i=1}^d \frac{\varepsilon}{d} = 1 - \varepsilon \, ,
\end{aligned}$$

(8.27) implying the inequality in the last line. Thus, (8.26) holds. In other words, $\{P \circ X_n^{-1} : n = 1, 2, \ldots\}$ is tight.

By Exc 8.8.2, it suffices to show that every subsequence $\{X_{n_k}\}$ of $\{X_n\}$ has a further subsequence converging weakly to X_∞. Fix a subsequence $\{X_{n_k}\}$. Since $\{P \circ X_n^{-1} : n = 1, 2, \ldots\}$ is tight, so is $\{P \circ X_{n_k}^{-1} : k = 1, 2, \ldots\}$. Theorem 8.7 implies $\{X_{n_k}\}$ has a subsequence $\{X_{n_{k_l}} : l = 1, 2, \ldots\}$ such that

$$X_{n_{k_l}} \Rightarrow Y \,, l \to \infty \,,$$

for some $\mathbb{R}^d$-valued random variable Y. The already proven "only if" part of this theorem implies

$$\langle \lambda, X_{n_{k_l}} \rangle \Rightarrow \langle \lambda, Y \rangle \,, l \to \infty \,, \lambda \in \mathbb{R}^d \,.$$

Comparing this with the hypothesis (8.24) yields

$$\langle \lambda, X_\infty \rangle \stackrel{d}{=} \langle \lambda, Y \rangle \,, \lambda \in \mathbb{R}^d \,.$$

Theorem 7.9 implies

$$X_\infty \stackrel{d}{=} Y \,.$$

Therefore,

$$X_{n_{k_l}} \Rightarrow X_\infty \,, l \to \infty \,.$$

This gives us the desired further subsequence of $\{X_{n_k}\}$ which converges weakly to X_∞. Hence the proof follows. ☐

The CLT in $\mathbb{R}^d$ now becomes a trivial consequence of the above theorem.

Theorem 8.9 (CLT in $\mathbb{R}^d$)**.** *Suppose $X_1, X_2, \ldots$ are i.i.d. random variables taking values in $\mathbb{R}^d$ such that each coordinate of X_1 has mean zero and finite variance. Then*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \Rightarrow Z \,, n \to \infty$$

where $Z \sim N_d(0, \Sigma)$ and Σ is the covariance matrix of X_1.

Proof. In view of Theorem 8.8, it suffices to prove that for all $\lambda \in \mathbb{R}^d$,

$$\left\langle \lambda, \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \right\rangle \Rightarrow \langle \lambda, Z \rangle \,, n \to \infty \,. \tag{8.28}$$

To that end fix $\lambda \in \mathbb{R}^d$, write

$$\left\langle \lambda, \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \right\rangle = \frac{1}{\sqrt{n}} \sum_{i=1}^n \langle \lambda, X_i \rangle \,,$$

and notice that $\langle \lambda, X_1 \rangle, \langle \lambda, X_2 \rangle, \ldots$ are i.i.d. Denoting $X_1 = (X_{11}, \ldots, X_{1d})$, the assumption that $X_{11}, \ldots, X_{1d}$ are zero mean implies

$$\mathrm{E}\left(\langle \lambda, X_1 \rangle\right) = 0\,.$$

Further, if σ_{ij} is the (i,j)-th entry of Σ, that is,

$$\sigma_{ij} = \mathrm{Cov}(X_{1i}, X_{ij})\,,$$

then writing $\lambda = [\lambda_1 \ldots \lambda_d]^T$,

$$\begin{aligned}
\mathrm{Var}\left(\langle \lambda, X_1 \rangle\right) &= \mathrm{Var}\left(\sum_{i=1}^d \lambda_i X_{1i}\right) \\
&= \sum_{i=1}^d \sum_{j=1}^d \lambda_i \lambda_j \sigma_{ij} \\
&= \lambda^T \Sigma \lambda\,.
\end{aligned}$$

The CLT on $\mathbb{R}$, which is Theorem 8.4, implies

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \langle \lambda, X_i \rangle \Rightarrow Y_\lambda\,, n \to \infty\,,$$

where $Y_\lambda \sim N(0, \lambda^T \Sigma \lambda)$. Theorem 7.10 shows that

$$Y_\lambda \stackrel{d}{=} \langle \lambda, Z \rangle\,.$$

In other words, (8.28) holds, from which the proof follows. $\square$

The last theorem of this course is Lindeberg's CLT, which is a generalization of Theorem 8.4 in that the assumption of identical distribution therein is relaxed.

Theorem 8.10 (Lindeberg's CLT)**.** *Suppose for all $n \in \mathbb{N}$, $X_{n1}, \ldots, X_{nn}$ are independent $\mathbb{R}$-valued random variables satisfying the following:*

$$\mathrm{E}(X_{ni}) = 0\,, i = 1, \ldots, n,\ n = 1, 2, \ldots\,,$$

$$\lim_{n \to \infty} \sum_{i=1}^n \mathrm{E}\left(X_{ni}^2\right) = \sigma^2 < \infty\,,$$

and

$$\lim_{n \to \infty} \sum_{i=1}^n \mathrm{E}\left(X_{ni}^2 \mathbf{1}(|X_{ni}| > \varepsilon)\right) = 0\,,\ \textit{for every } \varepsilon > 0\,. \tag{8.29}$$

Then, as $n \to \infty$,

$$\sum_{i=1}^n X_{ni} \Rightarrow Z\,,$$

where $Z \sim N(0, \sigma^2)$.

The assumption (8.29) is called Lindeberg's condition. The family $\{X_{ni} : 1 \le i \le n,\, n = 1, 2, \ldots\}$ is called a triangular array, which is why, Theorem 8.10 is also known as CLT for triangular arrays. Theorem 8.4 follows from Theorem 8.10 as claimed in the following exercise.

Exercise 8.9. *Suppose $X_1, X_2, \ldots$ are i.i.d. zero mean random variables with finite variance σ^2. Define*

$$X_{ni} = \frac{1}{\sqrt{n}} X_i \,, 1 \le i \le n,\, n \ge 1 \,.$$

Show that $\{X_{ni} : 1 \le i \le n,\, n = 1, 2, \ldots\}$ satisfies the assumptions of Theorem 8.10 and hence argue that Theorem 8.4 is a special case of that.

Theorem 8.10 can be proven along the lines of the proof of Theorem 8.4, that is, with the help of the Lévy continuity theorem. For pedagogical reasons, we shall prove it using Lindeberg's principle which completely bypasses the Fourier analytic method, that is, the use of characteristic functions. The following two exercises, for example, can be easily solved using the Lévy continuity theorem, though the solutions hinted at don't use it.

Exercise 8.10. *If $X_n \sim N(0, \sigma_n^2)$ and $0 \le \sigma_n \to \sigma < \infty$, show that $X_n \Rightarrow X$ where $X \sim N(0, \sigma^2)$.*
Hint. *If $Z \sim N(0,1)$, then $X_n \stackrel{d}{=} \sigma_n Z \to \sigma Z$.*

Exercise 8.11. *Suppose $X, X_1, X_2, \ldots$ are random variables such that for all thrice differentiable bounded $f : \mathbb{R} \to \mathbb{R}$ whose first three derivatives are bounded, it holds that*

$$\lim_{n\to\infty} \mathrm{E}\left(f(X_n)\right) = \mathrm{E}\left(f(X)\right) \,.$$

Show that $X_n \Rightarrow X$.
Hint. *Let*

$$f(x) = \begin{cases} 1, & x \le 0 \,, \\ (1 - x^4)^4, & 0 < x < 1 \,, \\ 0, & x \ge 1 \,. \end{cases}$$

Observe that for $w < y$,

$$\mathbf{1}_{(-\infty, w]}(x) \le f\left(\frac{x - w}{y - w}\right) \le \mathbf{1}_{(-\infty, y]}(x) \text{ for all } x \in \mathbb{R} \,.$$

Proof of Theorem 8.10. Using Exc 8.11, it suffices to show that

$$\lim_{n\to\infty} \mathrm{E}\left(f(S_n)\right) = \mathrm{E}\left(f(Z)\right) \,, \tag{8.30}$$

for all thrice differentiable $f : \mathbb{R} \to \mathbb{R}$ such that f and its first three derivatives are bounded, where

$$S_n = \sum_{i=1}^n X_{ni}\,, n \geq 1.$$

Fix such f.

Let $(Z_1, Z_2, \ldots)$ be a collection of i.i.d. standard normal random variables which is independent of the triangular array $\{X_{ni} : 1 \leq i \leq n, n \geq 1\}$. Set

$$\sigma_{ni} = \sqrt{\mathrm{E}(X_{ni}^2)}\,, 1 \leq i \leq n\,, n = 1, 2, \ldots\,,$$

and

$$\sigma_n = \sqrt{\sum_{i=1}^n \sigma_{ni}^2}\,, n \geq 1\,.$$

Since

$$\sum_{i=1}^n \sigma_{ni} Z_i \sim N(0, \sigma_n^2)\,, n = 1, 2, \ldots\,, \tag{8.31}$$

and $\sigma_n^2 \to \sigma^2$, Exc 8.10 shows

$$\lim_{n\to\infty} \mathrm{E}\left(f\left(\sum_{i=1}^n \sigma_{ni} Z_i\right)\right) = \mathrm{E}\left(f(Z)\right)\,.$$

Thus, (8.30) would follow once it is shown that

$$\lim_{n\to\infty} \mathrm{E}\left(f(S_n) - f\left(\sum_{i=1}^n \sigma_{ni} Z_i\right)\right) = 0\,. \tag{8.32}$$

Fix $n \in \{1, 2, \ldots\}$ and write

$$f(S_n) - f\left(\sum_{i=1}^n \sigma_{ni} Z_i\right) = \sum_{i=1}^n (f(Y_{i-1}) - f(Y_i))\,,$$

where

$$Y_i = \sum_{j=i+1}^n X_{nj} + \sum_{j=1}^i \sigma_{nj} Z_j\,, i = 0, 1, \ldots, n\,,$$

with the usual interpretation of the sum as zero if the lower limit exceeds the upper limit. Thus,

$$\left|\mathrm{E}\left(f(S_n) - f\left(\sum_{i=1}^n \sigma_{ni} Z_i\right)\right)\right| \leq \sum_{i=1}^n \left|\mathrm{E}(f(Y_{i-1}) - f(Y_i))\right|\,. \tag{8.33}$$

Fix $i \in \{1, \ldots, n\}$ and write

$$Y_i = W + \sigma_{ni} Z_i \,,$$

and

$$Y_{i-1} = W + X_{ni} \,,$$

where

$$W = \sum_{j=i+1}^{n} X_{nj} + \sum_{j=1}^{i-1} \sigma_{nj} Z_j \,.$$

It is immediate that W, X_{ni}, Z_i are independent. Taylor's theorem implies

$$f(Y_{i-1}) = f(W) + X_{ni} f'(W) + \frac{1}{2} X_{ni}^2 f''(\xi_1) \tag{8.34}$$

$$= f(W) + X_{ni} f'(W) + \frac{1}{2} X_{ni}^2 f''(W) + \frac{1}{6} X_{ni}^3 f'''(\xi_2) \,, \tag{8.35}$$

for some ξ_1 and ξ_2 between W and Y_{i-1}, where f', f'', f''' are the first three derivatives of f, respectively. Let

$$K = \sup_{x \in \mathbb{R}} \left(|f(x)| \vee |f'(x)| \vee |f''(x)| \vee |f'''(x)| \right) \,,$$

which is finite by assumption. A consequence of (8.34) is that

$$\begin{aligned}
&\left| f(Y_{i-1}) - \left(f(W) + X_{ni} f'(W) + \frac{1}{2} X_{ni}^2 f''(W) \right) \right| \\
&= \frac{1}{2} X_{ni}^2 \left| f''(\xi_1) - f''(W) \right| \\
&\le \frac{1}{2} X_{ni}^2 \left(|f''(\xi_1)| + |f''(W)| \right) \\
&\le K X_{ni}^2 \,.
\end{aligned}$$

Similarly, (8.35) shows

$$\left| f(Y_{i-1}) - \left(f(W) + X_{ni} f'(W) + \frac{1}{2} X_{ni}^2 f''(W) \right) \right| \le \frac{1}{6} K |X_{ni}|^3 \le K |X_{ni}|^3 \,.$$

Thus,

$$\left| f(Y_{i-1}) - \left(f(W) + X_{ni} f'(W) + \frac{1}{2} X_{ni}^2 f''(W) \right) \right| \le K (X_{ni}^2 \wedge |X_{ni}|^3) \,.$$

Therefore,

$$\begin{aligned}
K \mathrm{E} \left(X_{ni}^2 \wedge |X_{ni}|^3 \right) &\ge \mathrm{E} \left| f(Y_{i-1}) - \left(f(W) + X_{ni} f'(W) + \frac{1}{2} X_{ni}^2 f''(W) \right) \right| \\
&\ge \left| \mathrm{E}(f(Y_{i-1})) - \mathrm{E} \left(f(W) + X_{ni} f'(W) + \frac{1}{2} X_{ni}^2 f''(W) \right) \right| \\
&= \left| \mathrm{E}(f(Y_{i-1})) - \mathrm{E}(f(W)) - \frac{1}{2} \sigma_{ni}^2 \mathrm{E}(f''(W)) \right| \,,
\end{aligned}$$

the last line following from the independence of W and X_{ni} and that the mean and variance of X_{ni} are zero and σ_{ni}^2, respectively. A similar calculation shows

$$\left|\mathrm{E}(f(Y_i)) - \mathrm{E}(f(W)) - \frac{1}{2}\sigma_{ni}^2 \mathrm{E}(f''(W))\right| \le K\mathrm{E}\left(|\sigma_{ni} Z_i|^3\right) = C\sigma_{ni}^3\,,$$

where $C = K\mathrm{E}(|Z_1|^3)$. Combine the two inequalities obtained to get

$$|\mathrm{E}\left(f(Y_{i-1}) - f(Y_i)\right)| \le K\mathrm{E}\left(X_{ni}^2 \wedge |X_{ni}|^3\right) + C\sigma_{ni}^3\,.$$

Summing the above inequality over $i = 1, \ldots, n$ and using (8.33), we get

$$\left|\mathrm{E}\left(f(S_n) - f\left(\sum_{i=1}^n \sigma_{ni} Z_i\right)\right)\right| \le C\sum_{i=1}^n \sigma_{ni}^3 + K\sum_{i=1}^n \mathrm{E}\left(X_{ni}^2 \wedge |X_{ni}|^3\right)\,.$$

Thus, (8.32) would follow, which would complete the proof, once the following are shown:

$$\lim_{n\to\infty}\sum_{i=1}^n \sigma_{ni}^3 = 0\,, \tag{8.36}$$

$$\text{and}\ \lim_{n\to\infty}\sum_{i=1}^n \mathrm{E}\left(X_{ni}^2 \wedge |X_{ni}|^3\right) = 0\,. \tag{8.37}$$

For (8.36), write

$$\begin{aligned}\sum_{i=1}^n \sigma_{ni}^3 &\le \sigma_n^2 \max_{1\le i\le n} \sigma_{ni}\\ &= \sigma_n^2 \sqrt{\max_{1\le i\le n} \sigma_{ni}^2}\,.\end{aligned}$$

Since $\sigma_n^2 \to \sigma^2 < \infty$, (8.36) would follow if it can be shown that

$$\lim_{n\to\infty}\max_{1\le i\le n} \sigma_{ni}^2 = 0\,.$$

Fix $\varepsilon > 0$ and write

$$\sigma_{ni}^2 = \mathrm{E}(X_{ni}^2 \mathbf{1}(|X_{ni}| \le \varepsilon)) + \mathrm{E}(X_{ni}^2 \mathbf{1}(|X_{ni}| > \varepsilon)) \le \varepsilon^2 + \mathrm{E}(X_{ni}^2 \mathbf{1}(|X_{ni}| > \varepsilon))\,.$$

Hence

$$\begin{aligned}\max_{1\le i\le n} \sigma_{ni}^2 &\le \varepsilon^2 + \max_{1\le i\le n} \mathrm{E}(X_{ni}^2 \mathbf{1}(|X_{ni}| > \varepsilon))\\ &\le \varepsilon^2 + \sum_{i=1}^n \mathrm{E}(X_{ni}^2 \mathbf{1}(|X_{ni}| > \varepsilon))\,.\end{aligned}$$

Invoke (8.29) to argue

$$\limsup_{n\to\infty} \max_{1\le i\le n} \sigma_{ni}^2 \le \varepsilon^2 \,.$$

Since ε is arbitrary,

$$\lim_{n\to\infty} \max_{1\le i\le n} \sigma_{ni}^2 = 0 \,,$$

which shows (8.36).

Finally, for (8.37), fix $\varepsilon > 0$ and write

$$\begin{aligned}
\sum_{i=1}^n \mathrm{E}\left(X_{ni}^2 \wedge |X_{ni}|^3\right) &\le \sum_{i=1}^n \mathrm{E}\left(|X_{ni}|^3 \mathbf{1}(|X_{ni}| \le \varepsilon\right) + \sum_{i=1}^n \mathrm{E}\left(X_{ni}^2 \mathbf{1}(|X_{ni}| > \varepsilon)\right) \\
&\le \varepsilon \sum_{i=1}^n \mathrm{E}(X_{ni}^2) + \sum_{i=1}^n \mathrm{E}\left(X_{ni}^2 \mathbf{1}(|X_{ni}| > \varepsilon)\right) \\
&= \varepsilon \sigma_n^2 + \sum_{i=1}^n \mathrm{E}\left(X_{ni}^2 \mathbf{1}(|X_{ni}| > \varepsilon)\right) \,.
\end{aligned}$$

Let $n \to \infty$ and use (8.29) to get

$$\limsup_{n\to\infty} \sum_{i=1}^n \mathrm{E}\left(X_{ni}^2 \wedge |X_{ni}|^3\right) \le \varepsilon \sigma^2 \,.$$

Since ε is arbitrary, (8.37) follows. This in conjunction with (8.36) shows (8.32), which completes the proof. □

Remark 5. *The above proof is transparent in that it displays the property of normal that has been used. Indeed,* (8.31) *does use the fact that the sum of independent normal random variables also follows normal.*

Exercise 8.12. *If $X_1, X_2, \ldots$ are i.i.d. and $P(X_1 = 0) < 1$, show that there does not exist a random variable Z such that*

$$\sum_{i=1}^n X_i \Rightarrow Z \,, n \to \infty \,.$$

Exercise 8.13. *Show that a sequence of probability measure $\{\mu_n\}$ on $\mathbb{R}^d$ is tight if and only if given any subsequence of $\{\mu_n\}$, there exists a further subsequence which converges to a probability measure μ on $\mathbb{R}^d$. This is a special case of Prohorov's theorem*

Exercise 8.14. *Show that the Lindeberg condition* (8.29) *is implied by the Lyapunov condition*

$$\lim_{n\to\infty} \sum_{i=1}^n \mathrm{E}\left(|X_{ni}|^{2+\delta}\right) = 0 \textit{ for some } \delta > 0 \,.$$

Exercise 8.15. *If $X_n \sim$ Binomial(n, p_n) where p_n are such that*

$$\lim_{n\to\infty} np_n(1-p_n) = \infty\,,$$

show that

$$\frac{X_n - np_n}{\sqrt{np_n(1-p_n)}} \Rightarrow Z\,,$$

where $Z \sim N(0,1)$.

Exercise 8.16. *Suppose X is as in Exc 5.3, that is, it is infinitely divisible,* $\mathrm{E}(X) = 0$ *and* $\mathrm{Var}(X) = 1$*. Show that*

$$\mathrm{E}(X^4) = 3 \iff X \sim N(0,1)\,.$$

Hint. *If $X_{n1}, \ldots, X_{nn}$ are as in* (5.9) *and* $\mathrm{E}(X^4) = 3$*, show that*

$$\mathrm{E}(X_{n1}^4) = \frac{3}{n^2}\,.$$

Use Exc 8.14.

Exercise 8.17. *A coin with probability of head $p \in (0,1)$ is tossed infinitely many times. Let X_n be the number of the toss on which the n-th head is obtained. Show that*

$$n^{-1/2}\left(X_n - \frac{n}{p}\right) \Rightarrow Z\,,$$

where $Z \sim N(0, \sigma^2)$ for some σ^2. Calculate σ^2.

Exercise 8.18. *There are k boxes numbered $1, \ldots, k$ and an infinite supply of balls. The balls are thrown, one by one, randomly into one of the boxes. Let $X_{n1}, \ldots, X_{nk}$ denote the number of balls in Boxes $1, \ldots, k$, respectively, after the first n balls are thrown. Show that*

$$n^{-1/2}\left(X_{n1} - \frac{n}{k}, \ldots, X_{nk} - \frac{n}{k}\right) \Rightarrow (Z_1, \ldots, Z_k)\,,$$

where $(Z_1, \ldots, Z_k) \sim N_k(0, \Sigma)$ for some $k \times k$ matrix Σ. Calculate Σ.

Exercise 8.19. *If $X_n \sim$ Binomial(n, p_n) and*

$$\lim_{n\to\infty} np_n = \lambda \in (0,\infty)\,,$$

use the Lévy continuity theorem to show that $X_n \Rightarrow Z$ where Z follows Poisson(λ).

Exercise 8.20. *Suppose that $X_1, X_2, \ldots$ are i.i.d. random variables with density*

$$f(x) = e^{-1}x^{-2}\mathbf{1}(x > e^{-1}),\; x \in \mathbb{R}\,.$$

Show that as $n \to \infty$,

$$(X_1 \ldots X_n)^{1/\sqrt{n}} \Rightarrow Z\,,$$

where Z follows the log-normal distribution, that is, $\log Z$ *follows standard normal.*

Exercise 8.21. *If $X_1, X_2 \ldots$ are i.i.d. with zero mean and finite positive variance, show that there does not exist a random variable Z such that*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \xrightarrow{P} Z \, , n \to \infty \, .$$

Exercise 8.22. *Suppose X_n are random variables with all moments finite such that*

$$\lim_{n\to\infty} \mathrm{E}(X_n^k) = m_k \in \mathbb{R} \, , k \in \{1, 2, \ldots\} \, .$$

If there exists a unique probability measure μ on $\mathbb{R}$ such that

$$\int_{\mathbb{R}} x^k \mu(dx) = m_k \, , k = 1, 2, \ldots \, ,$$

show that $X_n \Rightarrow X$ where $P \circ X^{-1} = \mu$.
Hint. *First show $\{P \circ X_n^{-1} : n = 1, 2, \ldots\}$ is tight.*

Exercise 8.23. *Suppose $X_1, X_2, \ldots$ are i.i.d. from the density*

$$f(x) = x^{-2} \, , x \geq 1 \, ,$$

then show that there exists a random variable Z such that

$$n^{-1} \max_{1 \leq i \leq n} X_i \Rightarrow Z \, .$$

Find the distribution of Z.

Exercise 8.24. *Suppose $X, X_1, X_2, \ldots$ are random variables with CDFs F, $F_1, F_2, \ldots$, respectively. If $X_n \Rightarrow X$ and F is continuous, show that*

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} |F_n(x) - F(x)| = 0 \, .$$

Exercise 8.25. *Suppose $\mu, \mu_1, \mu_2, \ldots$ are probability measures on $\mathbb{R}$ having densities $f, f_1, f_2, \ldots$ with respect to the Lebesgue measure (λ). If*

$$f_n(x) \to f(x) \text{ for a.e.}(\lambda)\, x \, ,$$

show that $f_n \to f$ in $L^1(\mathbb{R}, \lambda)$ and hence

$$\lim_{n\to\infty} \mu_n(B) = \mu(B) \, , B \in \mathcal{B}(\mathbb{R}) \, .$$

Hint. *Write*

$$|f_n - f| = f_n + f - 2(f_n \wedge f) \, .$$

9 Appendix

9.1 Proof of Fact 1.3

Proof of Fact 1.3. Since $\|\cdot\|$ is the L^∞ norm, it suffices to show that the absolute value of each entry of the $d \times 1$ vector $T(x) - T(y) - J(x)(x-y)$ is at most. $d\alpha\|x-y\|$. In other words, it suffices to show that if $f : U \to \mathbb{R}$ is continuously differentiable, and

$$|f_i(y) - f_i(x)| \le \alpha\,, x, y \in R,\ i = 1, \dots, d\,,$$

where

$$f_i(x) = \frac{\partial f(x)}{\partial x_i}\,, x \in U,\ i = 1, \dots, d\,,$$

then

$$\left| f(x) - f(y) - \sum_{i=1}^d f_i(x)(x_i - y_i) \right| \le d\alpha\|x-y\|\,, x, y \in R\,.$$

Let f be a function satisfying the hypotheses. Let $x^0 = x$, $x^d = y$, and for $1 \le i \le d-1$,

$$x^i = (y_1, \dots, y_i, x_{i+1}, \dots, x_d)\ .$$

Since R is a rectangle, $x^1, \dots, x^{d-1} \in R$. For a fixed $i = 1, \dots, d$, x^{i-1} and x^i have all entries identical except the i-th one, which are x_i and y_i respectively. The one-dimensional mean value theorem implies there exists ξ_i between x_i and y_i such that

$$f\left(x^{i-1}\right) - f\left(x^i\right) = (x_i - y_i) f_i(y_1, \dots, y_{i-1}, \xi_i, x_{i+1}, \dots, x_d)\,.$$

Since $\tilde{\xi}_i = (y_1, \dots, y_{i-1}, \xi_i, x_{i+1}, \dots, x_d) \in R$ because R is a rectangle, the hypotheses on f imply

$$\left| f_i(\tilde{\xi}_i) - f_i(x) \right| \le \alpha\,, i = 1, \dots, d\,.$$

Therefore,

$$
\begin{aligned}
&\left| f(x) - f(y) - \sum_{i=1}^d f_i(x)(x_i - y_i) \right| \\
&= \left| f\left(x^0\right) - f\left(x^d\right) - \sum_{i=1}^d f_i(x)(x_i - y_i) \right| \\
&= \left| \sum_{i=1}^d \left[f\left(x^{i-1}\right) - f\left(x^i\right) \right] - \sum_{i=1}^d f_i(x)(x_i - y_i) \right| \\
&= \left| \sum_{i=1}^d \left(f_i(\tilde{\xi}_i) - f_i(x) \right) (x_i - y_i) \right| \\
&\le \sum_{i=1}^d \left| f_i(\tilde{\xi}_i) - f_i(x) \right| |x_i - y_i| \\
&\le d\alpha \max_{1 \le i \le d} |x_i - y_i| \\
&= d\alpha \|x - y\| \,.
\end{aligned}
$$

This completes the proof. $\square$

9.2 Proof of Fact 4.1

Proof of Fact 4.1. Let $F : \mathbb{R}^d \to \mathbb{R}$ satisfy the assumptions, that is,

$$
\lim_{y_1 \downarrow x_1, \ldots, y_d \downarrow x_d} F(y_1, \ldots, y_d) = F(x_1, \ldots, x_d) \text{ for all } (x_1, \ldots, x_d) \in \mathbb{R}^d \,, \tag{9.1}
$$

and

$$
\Delta_R F \ge 0 \text{ for all } R \in \mathcal{H} \,, \tag{9.2}
$$

where

$$
\begin{aligned}
\mathcal{H} &= \{(a_1, b_1] \times \ldots \times (a_d, b_d] : -\infty < a_i < b_i < \infty \text{ for } i = 1, \ldots, d\} \,, \\
\Delta_R F &= \sum_{(x_1, \ldots, x_d) \in \{a_1, b_1\} \times \ldots \times \{a_d, b_d\}} (-1)^{\#\{i : x_i = a_i\}} F(x_1, \ldots, x_d) \,,
\end{aligned} \tag{9.3}
$$

for all $R = (a_1, b_1] \times \ldots \times (a_d, b_d] \in \mathcal{H}$.
Step 1. The function $R \mapsto \Delta_R F$ is a finitely additive set function on $\mathcal{H}$, that is, for disjoint $R_1, \ldots, R_n \in \mathcal{H}$ such that $R = R_1 \cup \ldots \cup R_n \in \mathcal{H}$,

$$
\Delta_R F = \sum_{i=1}^n \Delta_{R_i} F \,.
$$

Proof of Step 1. For $R = (a_1, b_1] \times (a_d, b_d] \in \mathcal{H}$, and $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$, define

$$\operatorname{sgn}(x, R) = \begin{cases} (-1)^{\#\{i : x_i = a_i\}}, & x \in \{a_1, b_1\} \times \ldots \times \{a_d, b_d\}, \\ 0, & \text{otherwise}. \end{cases}$$

That is, $\operatorname{sgn}(x, R)$ is zero unless x is a vertex of R.

Rewrite (9.3) as

$$\Delta_R F = \sum_{x = (x_1, \ldots, x_d) \in \{a_1, b_1\} \times \ldots \times \{a_d, b_d\}} \operatorname{sgn}(x, R) F(x).$$

Suppose $R = (a_1, b_1] \times \ldots \times (a_d, b_d] \in \mathcal{H}$ and for some $n_1, \ldots, n_d \in \mathbb{N}$,

$$a_i = a_{i,0} < a_{i,1} < \ldots < a_{i,n_i} = b_i, \, i = 1, \ldots, d.$$

Let

$$R_{k_1, \ldots, k_d} = \prod_{i=1}^{d} (a_{i,k_i - 1}, a_{i,k_i}], \, 1 \le k_1 \le n_1, \ldots, 1 \le k_d \le n_d. \tag{9.4}$$

We shall first show that

$$\sum_{k_1=1}^{n_1} \cdots \sum_{k_d=1}^{n_d} \Delta_{R_{k_1, \ldots, k_d}} F = \Delta_R F. \tag{9.5}$$

The LHS above equals

$$\sum_{x \in A} F(x) \sum_{k_1=1}^{n_1} \cdots \sum_{k_d=1}^{n_d} \operatorname{sgn}\left(x, R_{k_1, \ldots, k_d}\right), \tag{9.6}$$

where $A = \prod_{i=1}^{d} \{a_{i,0}, a_{i,1}, \ldots, a_{i,n_i}\}$. Let $A_0 = \{a_1, b_1\} \times \ldots \times \{a_d, b_d\}$ and observe that for $x \in A_0$, there exists unique $k_1, \ldots, k_d$ such that

$$\operatorname{sgn}\left(x, R_{k_1, \ldots, k_d}\right) \neq 0,$$

and for this $k_1, \ldots, k_d$,

$$\operatorname{sgn}\left(x, R_{k_1, \ldots, k_d}\right) = \operatorname{sgn}(x, R).$$

Thus, the quantity in (9.6) equals

$$\sum_{x \in A_0} \operatorname{sgn}(x, R) F(x) + \sum_{x \in A \setminus A_0} F(x) \sum_{k_1=1}^{n_1} \cdots \sum_{k_d=1}^{n_d} \operatorname{sgn}\left(x, R_{k_1, \ldots, k_d}\right).$$

Since the first term above is the same as $\Delta_R F$, (9.5) would follow once it is shown that

$$\sum_{k_1=1}^{n_1} \cdots \sum_{k_d=1}^{n_d} \operatorname{sgn}(x, R_{k_1,\ldots,k_d}) = 0\,, x \in A \setminus A_0\,. \tag{9.7}$$

Fix $x = (x_1, \ldots, x_d) \in A \setminus A_0$. Then there exists $i \in \{1, \ldots, d\}$ such that

$$x_i = a_{i,u_i} \text{ for some } 1 \le u_i \le n_i - 1\,.$$

Thus for $1 \le k_1 \le n_1, \ldots, 1 \le k_d \le n_d$, x is not a vertex of $R_{k_1,\ldots,k_d}$ by (9.4), unless k_i equals either u_i or $u_i + 1$, that is,

$$\operatorname{sgn}(x, R_{k_1,\ldots,k_d}) = 0 \text{ if } k_i \notin \{u_i, u_i + 1\}\,.$$

Further,

$$\operatorname{sgn}\left(x, R_{k_1,\ldots,k_{i-1},u_i,k_{i+1},\ldots,k_d}\right) = -\operatorname{sgn}\left(x, R_{k_1,\ldots,k_{i-1},u_i+1,k_{i+1},\ldots,k_d}\right)\,.$$

Thus (9.7) follows which proves (9.5).

To complete the proof of Step 1, let $R_1, \ldots, R_n \in \mathcal{H}$ be disjoint such that $R = R_1 \cup \ldots \cup R_n \in \mathcal{H}$. Let $R = (a_1, b_1] \times \ldots \times (a_d, b_d] \in \mathcal{H}$ and

$$a_i = a_{i,0} < a_{i,1} < \ldots < a_{i,n_i} = b_i\,, i = 1, \ldots, d\,,$$

be such that vertices of $R_1, \ldots, R_n$ belong to $\prod_{i=1}^d \{a_{i,0}, a_{i,1}, \ldots, a_{i,n_i}\}$. If $R_{k_1,\ldots,k_d}$ is as in (9.4), then

$$\text{either } R_{k_1,\ldots,k_d} \subset R_i \text{ or } R_{k_1,\ldots,k_d} \cap R_i = \emptyset\,,$$

for $1 \le k_1 \le n_1, \ldots, 1 \le k_d \le n_d$ and $i = 1, \ldots, n$. Use (9.5) to write

$$\begin{aligned}
\Delta_R F &= \sum_{1 \le k_1 \le n_1,\ldots,1 \le k_d \le n_d} \Delta_{R_{k_1,\ldots,k_d}} F \\
&= \sum_{i=1}^n \sum_{1 \le k_1 \le n_1,\ldots,1 \le k_d \le n_d :\, R_{k_1,\ldots,k_d} \subset R_i} \Delta_{R_{k_1,\ldots,k_d}} F \\
&= \sum_{i=1}^n \Delta_{R_i} F\,,
\end{aligned}$$

(9.5) being used again in the last line. This completes the proof of Step 1. □

Step 2. If $R_1, R_2 \in \mathcal{H}$ and $R_1 \subset R_2$, then $\Delta_{R_1} F \le \Delta_{R_2} F$.

Proof of Step 2. Follows from (9.2) and Step 1 by observing that $R_2 \setminus R_1 = S_1 \cup \ldots \cup S_n$ for some disjoint $S_1, \ldots, S_n \in \mathcal{H}$. □

Step 3. If $R = (a_1, b_1] \times \ldots \times (a_d, b_d] \in \mathcal{H}$ and for $\varepsilon > 0$, $R_\varepsilon = (a_1, b_1+\varepsilon] \times \ldots \times (a_d, b_d+\varepsilon]$, then

$$\lim_{\varepsilon \downarrow 0} \Delta_{R_\varepsilon} F = \Delta_R F\,.$$

Proof of Step 3. Follows from (9.1). □

Step 4. If $R = (a_1, b_1] \times \ldots \times (a_d, b_d] \in \mathcal{H}$,

$$\lim_{a_1' \downarrow a_1, \ldots, a_d' \downarrow a_d} \Delta_{(a_1', b_1] \times \ldots \times (a_d', b_d]} F = \Delta_R F\,.$$

Proof of Step 4. Follows from (9.1). □

For the next several steps, fix $n = (n_1, \ldots, n_d) \in \mathbb{Z}^d$ and let

$$\Omega_n = (n_1 - 1, n_1] \times \ldots \times (n_d - 1, n_d]\,,$$

and

$$\mathcal{S}_n = \{\emptyset\} \cup \{R \in \mathcal{H} : R \subset \Omega_n\}\,.$$

Step 5. The collection $\mathcal{S}_n$ is a semi-field on Ω_n and $\mu_n : \mathcal{S}_n \to [0, \infty)$ defined by

$$\mu_n(R) = \Delta_R F\,, \emptyset \neq R \in \mathcal{S}_n\,,$$

and $\mu_n(\emptyset) = 0$ is a finitely additive set function.

Proof of Step 5. That $\mathcal{S}_n$ is a semi-field is immediate. Finite additivity of μ_n follows from Step 1. □

Step 6. Let $\mathcal{F}_n = \{A_1 \cup \ldots \cup A_k : A_1, \ldots, A_k \in \mathcal{S}_n \text{ are disjoint}\}$. Then $\mathcal{F}_n$ is a field on Ω_n. Extend μ_n to $\mathcal{F}_n$ by

$$\mu_n(A_1 \cup \ldots \cup A_k) = \sum_{i=1}^k \mu_n(A_i)\,, A_1, \ldots, A_k \in \mathcal{S}_n \text{ are disjoint}\,.$$

Then μ_n is well defined on $\mathcal{F}_n$, that is, different representations yield the same definition, is finitely additive on $\mathcal{F}_n$, monotone on $\mathcal{F}_n$, that is, $\mu_n(A) \leq \mu_n(B)$ for $A, B \in \mathcal{F}_n$ with $A \subset B$ and finitely sub-additive on $\mathcal{F}_n$, that is,

$$\mu_n(A_1 \cup \ldots \cup A_k) \leq \sum_{i=1}^k \mu_n(A_i)\,, A_1, \ldots, A_k \in \mathcal{F}_n\,.$$

Proof of Step 6. That $\mathcal{F}_n$ is a field follows from Step 5 which says $\mathcal{S}_n$ is a semi-field. If $A_1, \ldots, A_k \in \mathcal{S}_n$ are disjoint and so are $\mathcal{B}_1, \ldots, B_l \in \mathcal{S}_n$ such that

$$A_1 \cup \ldots \cup A_k = B_1 \cup \ldots \cup B_l\,,$$

then Step 5 shows

$$\sum_{i=1}^{k} \mu_n(A_i) = \sum_{i=1}^{k} \sum_{j=1}^{l} \mu_n(A_i \cap B_j) = \sum_{j=1}^{l} \mu_n(B_j)\,.$$

Thus, μ_n is well defined on $\mathcal{F}_n$ in that the definition is not dependent on the representation. A similar argument shows μ_n is finitely additive on $\mathcal{F}_n$. If $A, B \in \mathcal{F}_n$ and $A \subset B$, then finite additivity shows

$$\mu_n(B) = \mu_n(A) + \mu_n(B \setminus A) \geq \mu_n(A)\,,$$

showing μ_n is monotone on $\mathcal{F}_n$. Finally for $A, B \in \mathcal{F}_n$, finite additivity shows

$$\mu_n(A \cup B) = \mu_n(A) + \mu_n(B \setminus A) \leq \mu_n(A) + \mu_n(B)\,,$$

the inequality following from monotonicity of μ_n. Induction shows μ_n is finitely sub-additive on $\mathcal{F}_n$. This completes the proof of Step 6. $\square$

Step 7. The set function μ_n is countably additive on $\mathcal{S}_n$.

Proof of Step 7. Let $R_1, R_2, \ldots \in \mathcal{S}_n$ be disjoint such that

$$R = R_1 \cup R_2 \cup \ldots \in \mathcal{S}_n\,.$$

For $k = 1, 2, 3, \ldots$, finite additivity of μ_n on $\mathcal{F}_n$ shown in Step 6 implies

$$\sum_{i=1}^{k} \mu_n(R_i) = \mu_n\left(\bigcup_{i=1}^{k} R_i\right) \leq \mu_n(R)\,,$$

the inequality following from monotonicity of μ_n. Thus, countable additivity would follow once it is shown that

$$\mu_n(R) \leq \sum_{i=1}^{\infty} \mu_n(R_i)\,. \tag{9.8}$$

Let $R = (a_1, b_1] \times \ldots \times (a_d, b_d]$ and for $i = 1, 2, \ldots$,

$$R_i = (a_{i,1}, b_{i,1}] \times \ldots \times (a_{i,d}, b_{i,d}]\,.$$

Fix $\delta > 0$. Use Step 3 to get $\varepsilon_i > 0$ such that $\Delta_{\tilde{R}_i} F \leq \Delta_{R_i} F + 2^{-i}\delta$ where

$$\tilde{R}_i = (a_{i,1}, b_{i,1} + \varepsilon_i] \times \ldots \times (a_{i,d}, b_{i,d} + \varepsilon_i]\,.$$

Fix $a_i' \in (a_i, b_i)$ for $i = 1, \ldots, d$. Since

$$[a_1', b_1] \times \ldots \times [a_d', b_d] \subset R = \bigcup_{i=1}^{\infty} R_i \subset \bigcup_{i=1}^{\infty} (a_{i,1}, b_{i,1} + \varepsilon_i) \times \ldots \times (a_{i,d}, b_{i,d} + \varepsilon_i)\,,$$

the Heine-Borel theorem implies

$$[a_1', b_1] \times \ldots \times [a_d', b_d] \subset \bigcup_{i=1}^{k} (a_{i,1}, b_{i,1} + \varepsilon_i) \times \ldots \times (a_{i,d}, b_{i,d} + \varepsilon_i)$$

for some finite k. Letting $R' = (a_1', b_1] \times \ldots \times (a_d', b_d]$, it follows that

$$R' \subset \Omega_n \cap \left(\tilde{R}_1 \cup \ldots \cup \tilde{R}_k \right) .$$

Monotonicity and finite sub-additivity of μ_n shown in Step 6 implies

$$\begin{aligned}
\mu_n(R') &\le \sum_{i=1}^{k} \mu_n(\tilde{R}_i \cap \Omega_n) \\
&= \sum_{i=1}^{k} \Delta_{\tilde{R}_i \cap \Omega_n} F \\
\text{(Step 2)} &\le \sum_{i=1}^{k} \Delta_{\tilde{R}_i} F \\
\text{(choice of } \varepsilon_i) &\le \sum_{i=1}^{\infty} \left(\Delta_{R_i} F + 2^{-i} \delta \right) \\
&= \delta + \sum_{i=1}^{\infty} \mu_n(R_i) .
\end{aligned}$$

Since δ is arbitrary, it follows that

$$\mu_n(R') \le \sum_{i=1}^{\infty} \mu_n(R_i) .$$

Letting $a_1' \downarrow a_1, \ldots, a_d' \downarrow a_d$ and using Step 4, (9.8) follows. This completes the proof of Step 7. □

Step 8. The set function μ_n can be extended to a measure on $(\Omega_n, \sigma(\mathcal{S}_n))$.

Proof of Step 8. Follows from Step 7 and Corollary 1.1 of the Caratheodory extension theorem. □

Step 9. If

$$\mu(A) = \sum_{n \in \mathbb{Z}^d} \mu_n(A \cap \Omega_n) \, , A \in \mathcal{B}(\mathbb{R}^d) \, ,$$

then μ is a Radon measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ satisfying

$$\mu(R) = \Delta_R F \, , R \in \mathcal{H} \, . \tag{9.9}$$

Proof of Step 9. As μ_n is a measure on $(\Omega_n, \sigma(\mathcal{S}_n))$ for each $n \in \mathbb{Z}^d$ by Step 8 and $(\Omega_n : n \in \mathbb{Z}^d)$ is a partition of $\mathbb{R}^d$, μ defined above is a measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. For $R \in \mathcal{H}$, as R is bounded and non-empty, there exist $n_1, \ldots, n_k \in \mathbb{Z}^d$ such that $R \cap \Omega_{n_i} \neq \emptyset$ for $i = 1, \ldots, k$ and $R \subset \Omega_{n_1} \cup \ldots \cup \Omega_{n_k}$. Thus,

$$\begin{aligned}
\mu(R) &= \sum_{i=1}^{k} \mu_{n_i}(R \cap \Omega_{n_i}) \\
(\emptyset \neq R \cap \Omega_{n_i} \in \mathcal{S}_{n_i}) &= \sum_{i=1}^{k} \Delta_{R \cap \Omega_{n_i}} F \\
(\text{Step 1}) &= \Delta_R F\,,
\end{aligned}$$

showing (9.9). To see that μ is Radon, for any compact set $K \subset \mathbb{R}^d$, there exists $n \in \mathbb{N}$ such that $R = (-n, n]^d \supset K$. Thus

$$\mu(K) \leq \mu(R) = \Delta_R F\,,$$

by (9.9). This shows μ is a Radon measure and completes the proof of Step 9. □

Step 10. The measure μ is the only measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ satisfying (9.9).

Proof of Step 10. Suppose μ' is a measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ such that (9.9) holds with μ replaced by μ'. Then μ and μ' agree on $\mathcal{H}$, and hence on

$$\mathcal{S} = \left\{ \mathbb{R}^d \cap \prod_{i=1}^{d} (a_i, b_i] : -\infty \leq a_i \leq b_i \leq \infty \right\}\,,$$

because for every set in $\mathcal{S}$ there exist sets in $\mathcal{H}$ increasing to the former. Further, μ and μ' are σ-finite on $\mathcal{H}$ and hence on $\mathcal{S}$ which is a semi-field that generates $\mathcal{B}(\mathbb{R}^d)$. Corollary 1.1 shows μ and μ' agree on $\mathcal{B}(\mathbb{R}^d)$, as claimed in Step 10. □

Steps 9 and 10 complete the proof of the fact. □

Remark 6. *A function F satisfying* (9.1) *and* (9.2) *is not necessarily monotonic. For example, $F : \mathbb{R}^2 \to \mathbb{R}$ defined by*

$$F(x, y) = xy\,,$$

satisfies (9.1) *and* (9.2), *and in fact induces the Lebesgue measure on $\mathbb{R}^2$, though F is not monotonic because*

$$F(0, 0) = 0 < F(1, 1) = F(-1, -1) = 1\,.$$

That is, $x_1 \leq x_2$ and $y_1 \leq y_2$ implies neither $F(x_1, y_1) \leq F(x_2, y_2)$ nor $F(x_1, y_1) \geq F(x_2, y_2)$.

References

Billingsley, P. (1995). *Probability and Measure*. Wiley, New York, 3rd edition.

Rudin, W. (1987). *Real and Complex Analysis*. McGraw Hill Book Company, third edition.