

Chapter 5

Lebesgue's convergence theorems and $\mathcal{L}^p$ spaces

In this chapter we prove the monotone and dominated convergence theorems and apply them to parameter-dependent integrals. Introduce L^p spaces, Lebesgue-Stieltjes measures and compare Riemann and Lebesgue integrals.

5.1 Convergence theorems

We start by proving Markov's Inequality, which will be used in the proof of several theorems that follow.

Theorem 5.1. (*Markov's Inequality*) *Let $(\mathbb{X}, \mathcal{F}, \mu)$ be a measure space and $f \in \mathcal{L}_{\mathbb{R}}$. Then, for all $E \in \mathcal{F}$ and $a > 0$*

$$\mu(\{x : |f(x)| \geq a\} \cap E) \leq \frac{1}{a} \int_E |f| d\mu.$$

Proof. Note that, $aI_{\{x:|f(x)| \geq a\} \cap E} = aI_{\{x:|f(x)| \geq a\}}I_E \leq |f(x)|I_E$ and consequently, integrating both sides, $a\mu(\{x : |f(x)| \geq a\} \cap E) \leq \int_E |f| d\mu$. Therefore,

$$\mu(\{x : |f(x)| \geq a\} \cap E) \leq \frac{1}{a} \int_E |f| d\mu.$$

■

Remark 5.1. Note that if $E = \mathbb{X}$ we have $\mu(\{x : |f(x)| \geq a\}) \leq \frac{1}{a} \int_{\mathbb{X}} |f| d\mu$. When $(\mathbb{X}, \mathcal{F}, \mu) = (\Omega, \mathcal{F}, P)$ a probability space and $f := X$ a random variable, we have

$$P(\{\omega : |X(\omega)| \geq a\}) \leq \frac{1}{a} E_P(|X|).$$

Also, note that if $f(\omega) := (X(\omega) - E_P(X))^2$ and $f \in \mathcal{L}_{\bar{\mathbb{R}}}$ we have

$$P(\{\omega : (X(\omega) - E_P(X))^2 \geq a\}) = P(\{\omega : |X(\omega) - E_P(X)| \geq a^{1/2}\}) \leq \frac{1}{a} E_P((X - E_P(X))^2),$$

and letting $b = a^{1/2}$ we have

$$P(\{\omega : |X(\omega) - E_P(X)| \geq b\}) \leq \frac{1}{b^2} E_P((X - E_P(X))^2) = \frac{1}{b^2} (E_P(X^2) - E_P(X)^2)$$

which is known as Chebyshev's Inequality. Note that $\int_{\Omega} X^2 dP$ must exist as a real number.

Recall that for a measure space $(\mathbb{X}, \mathcal{F}, \mu)$, N is a null set if $N \in \mathcal{F}$ and $\mu(N) = 0$. If a certain property $\mathcal{P}(x)$ that depends on $x \in \mathbb{X}$ holds for all $x \in \mathbb{X}$ except $x \in N_{\mathcal{P}} \subset N$, where N is a null set, we say that the property is true almost everywhere (a.e.) or almost surely (a.s.). Note the set $N_{\mathcal{P}}$ where the property does not hold need not be a measurable set.

Theorem 5.2. Let $(\mathbb{X}, \mathcal{F}, \mu)$ be a measure space and $f : (\mathbb{X}, \mathcal{F}, \mu) \rightarrow (\bar{\mathbb{R}}, \bar{\mathcal{B}})$ be measurable. Then,

$$1. I_N f \in \mathcal{L}_{\bar{\mathbb{R}}} \text{ for all } N \in \mathcal{N}_{\mu} := \{N \in \mathcal{F} : \mu(N) = 0\} \text{ and } \int_N f d\mu = 0$$

$$2. \int_{\mathbb{X}} |f| d\mu = 0 \iff |f| = 0 \text{ a.e.}$$

Proof. 1. For $j \in \mathbb{N}$, let $f_j = \min\{|f|, j\}$ and note $0 \leq f_1 \leq f_2 \leq \dots$ with $\lim_{j \rightarrow \infty} f_j = |f|$.

Hence, by Theorem [4.5](#)

$$\begin{aligned} \int_{\mathbb{X}} I_N |f| d\mu &= \lim_{j \rightarrow \infty} \int_{\mathbb{X}} I_N f_j d\mu = \lim_{j \rightarrow \infty} \int_{\mathbb{X}} I_N \min\{|f|, j\} d\mu \leq \lim_{j \rightarrow \infty} \int_{\mathbb{X}} j I_N d\mu \\ &= \lim_{j \rightarrow \infty} j \int_{\mathbb{X}} I_N d\mu = \lim_{j \rightarrow \infty} j \mu(N) = 0. \end{aligned}$$

Hence, $I_N f$ is integrable. Furthermore, since $|\int_{\mathbb{X}} I_N f d\mu| \leq \int_{\mathbb{X}} I_N |f| d\mu = 0$, we have that $\int_N f d\mu = 0$.

2. $(\Leftarrow) |f| = 0$ a.e. means that $\mu(\{x : |f(x)| \neq 0\}) = 0$. $\int_{\mathbb{X}} |f| d\mu = \int_{\{|f|=0\}} |f| d\mu + \int_{\{|f|\neq 0\}} |f| d\mu = \int_{\{|f|\neq 0\}} |f| d\mu = 0$ by item 1.

$(\Rightarrow)$ Note that by the fact that μ is a measure

$$\begin{aligned} \mu(\{x : |f(x)| > 0\}) &= \mu\left(\bigcup_{j \in \mathbb{N}} \{|f| \geq 1/j\}\right) \leq \sum_{j \in \mathbb{N}} \mu(\{|f| \geq 1/j\}) \\ &\leq \sum_{j \in \mathbb{N}} j \int_{\mathbb{X}} |f| d\mu = 0 \end{aligned}$$

by Markov's Inequality and the assumption that $\int_{\mathbb{X}} |f| d\mu = 0$. ■

Remark 5.2. 1. If $f, g \geq 0$ are measurable and $f = g$ μ -a.e. then $\int_{\mathbb{X}} f d\mu = \int_{\{x: f(x) \neq g(x)\}} f d\mu +$

$\int_{\{x: f(x) = g(x)\}} f d\mu$. But by Theorem 5.2.1, the first integral in this sum is equal to zero.

Consequently, $\int_{\mathbb{X}} f d\mu = \int_{\{x: f(x) = g(x)\}} f d\mu = \int_{\{x: f(x) = g(x)\}} g d\mu = \int_{\{x: f(x) \neq g(x)\}} g d\mu + \int_{\{x: f(x) = g(x)\}} g d\mu = \int_{\mathbb{X}} g d\mu$.

2. If $f \in \mathcal{L}_{\mathbb{R}}$, g is measurable and $f = g$ μ -a.e. then $g \in \mathcal{L}_{\mathbb{R}}$. To see this, note that $f = g$ μ -a.e. implies $f^+ = g^+$ and $f^- = g^-$ μ -a.e. Using the previous remark on f^+ and f^- we have $\int_{\mathbb{X}} f^+ d\mu = \int_{\mathbb{X}} g^+ d\mu$ and $\int_{\mathbb{X}} f^- d\mu = \int_{\mathbb{X}} g^- d\mu$. Hence, $g \in \mathcal{L}_{\mathbb{R}}$ and $\int_{\mathbb{X}} f d\mu = \int_{\mathbb{X}} g d\mu$.

3. If f is measurable and $0 \leq g \in \mathcal{L}_{\mathbb{R}}$ with $|f| \leq g$ a.e., then

$$f^+ \leq |f| \leq g \text{ a.e. and } f^- \leq |f| \leq g \text{ a.e.}$$

Hence, $\int_{\mathbb{X}} f^+ d\mu \leq \int_{\mathbb{X}} g d\mu$, $\int_{\mathbb{X}} f^- d\mu \leq \int_{\mathbb{X}} g d\mu$ and f is integrable.

Theorem 5.3. Let $f : (\mathbb{X}, \mathcal{F}, \mu) \rightarrow (\bar{\mathbb{R}}, \bar{\mathcal{B}})$ be integrable. Then f is real valued almost everywhere.

Proof. Note that $N := \{x : |f(x)| = \infty\} = \{x : f(x) = \infty\} \cup \{x : f(x) = -\infty\} \in \mathcal{F}$. Let $\bigcap_{n \in \mathbb{N}} \{x : |f(x)| \geq n\} := \bigcap_{n \in \mathbb{N}} B_n$ with $B_1 \supset B_2 \supset \dots$, then $\lim_{n \rightarrow \infty} B_n = \bigcap_{n \in \mathbb{N}} B_n = N$. Also, note that by Markov's Inequality and integrability of f

$$\mu(B_1) = \mu(\{x : |f(x)| \geq 1\}) \leq \int_{\mathbb{X}} |f| d\mu < \infty.$$

Hence, by continuity of measures from above, and Markov's Inequality

$$\mu(N) = \lim_{n \rightarrow \infty} \mu(B_n) = \lim_{n \rightarrow \infty} \mu(\{x : |f(x)| \geq n\}) \leq \lim_{n \rightarrow \infty} \frac{1}{n} \int_{\mathbb{X}} |f| d\mu = 0.$$

■

Theorem 5.4. (*Lebesgue's Monotone Convergence Theorem*) Let $f_n : (\mathbb{X}, \mathcal{F}, \mu) \rightarrow (\bar{\mathbb{R}}, \bar{\mathcal{B}})$ for $n \in \mathbb{N}$ be integrable functions such that $f_1 \leq f_2 \leq \dots$ and $f := \lim_{n \rightarrow \infty} f_n = \sup_{n \in \mathbb{N}} f_n$. Then,

$$f \in \mathcal{L}_{\bar{\mathbb{R}}}(\mu) \iff \sup_{n \in \mathbb{N}} \int_{\mathbb{X}} f_n d\mu < \infty.$$

In this case,

$$\sup_{n \in \mathbb{N}} \int_{\mathbb{X}} f_n d\mu = \int_{\mathbb{X}} \sup_{n \in \mathbb{N}} f_n d\mu = \int_{\mathbb{X}} f d\mu.$$

Proof. Since $f_n \in \mathcal{L}_{\bar{\mathbb{R}}}$ and $f_1 \leq f_2 \leq \dots$ we have that $0 \leq f_n - f_1 \in \mathcal{L}_{\bar{\mathbb{R}}}$ forms an increasing sequence of nonnegative measurable functions.¹ Hence, by Theorem 4.5

$$0 \leq \sup_{n \in \mathbb{N}} \int_{\mathbb{X}} f_n d\mu - \int_{\mathbb{X}} f_1 d\mu = \sup_{n \in \mathbb{N}} \int_{\mathbb{X}} (f_n - f_1) d\mu = \int_{\mathbb{X}} \sup_{n \in \mathbb{N}} (f_n - f_1) d\mu = \int_{\mathbb{X}} (\sup_{n \in \mathbb{N}} f_n - f_1) d\mu. \quad (5.1)$$

Now, suppose $f \in \mathcal{L}_{\bar{\mathbb{R}}}$ and note that from the left side of equation (5.1)

$$\begin{aligned} \sup_{n \in \mathbb{N}} \int_{\mathbb{X}} f_n d\mu - \int_{\mathbb{X}} f_1 d\mu &= \int_{\mathbb{X}} (f - f_1) d\mu, \text{ or} \\ \sup_{n \in \mathbb{N}} \int_{\mathbb{X}} f_n d\mu &= \int_{\mathbb{X}} f_1 d\mu + \int_{\mathbb{X}} (f - f_1) d\mu \\ &= \int_{\mathbb{X}} f_1 d\mu + \int_{\mathbb{X}} f d\mu - \int_{\mathbb{X}} f_1 d\mu = \int_{\mathbb{X}} f d\mu < \infty. \end{aligned}$$

¹ $f_n - f_1$ may not be defined (e.g. $\infty - \infty$). But since f_n is integrable replacing f_n by $f_n I_{\{x : |f(x)| = \infty\}^c}$ does not impact any of the integrals given Theorem 5.3

If $\sup_{n \in \mathbb{N}} \int_{\mathbb{X}} f_n d\mu < \infty$, then from equation (5.1) we have $\int_{\mathbb{X}} (f - f_1) d\mu < \infty$ and since f_1 is integrable $f = (f - f_1) + f_1$ is integrable. Therefore,

$$\int f d\mu = \int (f - f_1) d\mu + \int f_1 d\mu = \sup_{n \in \mathbb{N}} \int_{\mathbb{X}} f_n d\mu < \infty.$$

■

Theorem 5.5. (*Lebesgue's Dominated Convergence Theorem*) Let $(\mathbb{X}, \mathcal{F}, \mu)$ be a measure space and $\{f_n\}_{n \in \mathbb{N}}$ be a sequence of integrable functions such that for each n , $|f_n| \leq g$ μ -a.e. where g is some integrable nonnegative function. If $\lim_{n \rightarrow \infty} f_n(x) = f(x)$ exists μ -a.e. in $\bar{\mathbb{R}}$, and f is measurable, then f is integrable and

$$\lim_{n \rightarrow \infty} \int_{\mathbb{X}} f_n d\mu = \int_{\mathbb{X}} \lim_{n \rightarrow \infty} f_n d\mu = \int_{\mathbb{X}} f d\mu.$$

Proof. We start by observing that since the f_n and g are measurable, the set

$$\mathcal{N} = \{x : \lim_{n \rightarrow \infty} f_n(x) \text{ does not exist}\} \cup \left(\bigcup_{n \in \mathbb{N}} \{x : |f_n(x)| > g(x)\} \right)$$

is measurable and $\mu(\mathcal{N}) = 0$. Thus, the functions $f I_{\mathcal{N}^c}$ and $f_n I_{\mathcal{N}^c}$ satisfy the conditions on the theorem at every $x \in \mathbb{X} - \mathcal{N}$ with majorant $g I_{\mathcal{N}^c}$. And their integrals coincide with those of f and f_n . Hence, we proceed as if the conditions on the theorem held everywhere on $\mathbb{X}$.

By the pointwise limit of the sequence f_n , for any $\epsilon > 0$ there exists $N_{(\epsilon, x)} \in \mathbb{N}$ such that for all $n > N_{(\epsilon, x)}$

$$\begin{aligned} |f| &= |f - f_n + f_n| \leq |f_n| + |f - f_n| \\ &\leq g + |f - f_n| \text{ by } |f_n| < g \\ &\leq g + \epsilon, \end{aligned}$$

which implies that $|f| \leq g$. Therefore, $\int_{\mathbb{X}} f d\mu < \infty$ provided $g \in \mathcal{L}_{\bar{\mathbb{R}}}(\mu)$. Also, $|f_n| \leq g \iff$

$-g \leq f_n \leq g$. Hence, $f_n + g \geq 0$. By Fatou's Lemma,

$$\begin{aligned} \int \liminf_{n \rightarrow \infty} (f_n + g) d\mu &= \int (f + g) d\mu \leq \liminf_{n \rightarrow \infty} \int (f_n + g) d\mu \\ &= \liminf_{n \rightarrow \infty} \int f_n d\mu + \int g d\mu. \end{aligned}$$

Therefore,

$$\int f d\mu \leq \liminf_{n \rightarrow \infty} \int f_n d\mu. \quad (5.2)$$

Also, $g - f_n \geq 0$ and again by Fatou's Lemma,

$$\begin{aligned} 0 &\leq \int \liminf_{n \rightarrow \infty} (g - f_n) d\mu = \int g d\mu - \int f d\mu \\ &\leq \liminf_{n \rightarrow \infty} \int (g - f_n) d\mu \\ &= \int g d\mu + \liminf_{n \rightarrow \infty} - \int f_n d\mu \\ &= \int g d\mu - \limsup_{n \rightarrow \infty} \int f_n d\mu. \end{aligned}$$

The second inequality together with the last equality imply that

$$\int f d\mu \geq \limsup_{n \rightarrow \infty} \int f_n d\mu. \quad (5.3)$$

Combining (5.2) and (5.3) completes the proof. ■

We now consider a measurable function that is indexed by a parameter $\theta \in S$, where S is an open, nonempty subset of $\mathbb{R}^p$ for $p \in \mathbb{N}$. As such, we define $f(x, \theta) : (\mathbb{X}, \mathcal{F}, \mu) \times S \rightarrow (\mathbb{R}, \mathcal{B})$ where f is measurable for any $\theta \in S$.

Theorem 5.6. *Let $f : (\mathbb{X}, \mathcal{F}, \mu) \times S \rightarrow (\mathbb{R}, \mathcal{B})$ where f is $\mathcal{F} - \mathcal{B}$ measurable for every $\theta \in S$ and $f \in \mathcal{L}_{\mathbb{R}}$. Also, assume that for any $x \in \mathbb{X}$, $f(x, \theta)$ is continuous on S and $|f(x, \theta)| \leq g(x)$ for every $(x, \theta) \in \mathbb{X} \times S$ and some nonnegative integrable function g . Then, the function $h : S \rightarrow \mathbb{R}$ given by*

$$h(\theta) := \int_{\mathbb{X}} f(x, \theta) d\mu$$

is continuous.

Proof. The function h is well defined because of integrability of $f(x, \theta)$ for any $\theta \in S$. It suffices to show that for any sequence $\{\theta_n\}_{n \in \mathbb{N}} \subset S$ such that $\theta_n \rightarrow \theta \in S$ we have $h(\theta_n) \rightarrow h(\theta)$. By continuity of $f(x, \theta)$ on S for every x , we have $f(x, \theta_n) \rightarrow f(x, \theta)$ and $|f(x, \theta_n)| \leq g(x)$. By Lebesgue's Dominated Convergence Theorem,

$$\lim_{n \rightarrow \infty} h(\theta_n) = \int_{\mathbb{X}} \lim_{n \rightarrow \infty} f(x, \theta_n) d\mu = \int_{\mathbb{X}} f(x, \theta) d\mu = h(\theta),$$

establishing continuity of h on S . ■

Theorem 5.7. Let $f : (\mathbb{X}, \mathcal{F}, \mu) \times (a, b) \rightarrow (\mathbb{R}, \mathcal{B})$ where f is measurable and $f \in \mathcal{L}_{\mathbb{R}}$ for all $\theta \in (a, b)$. Also, assume that for any $x \in \mathbb{X}$, $f(x, \theta)$ is differentiable on (a, b) and $|\frac{d}{d\theta} f(x, \theta)| \leq g(x)$ for all $(x, \theta) \in \mathbb{X} \times (a, b)$ and some nonnegative integrable function g . Then, the function $h : (a, b) \rightarrow \mathbb{R}$ given by

$$h(\theta) := \int_{\mathbb{X}} f(x, \theta) d\mu$$

is differentiable and its derivative is given by

$$\frac{d}{d\theta} h(\theta) = \int_{\mathbb{X}} \frac{d}{d\theta} f(x, \theta) d\mu.$$

Proof. For $\theta, \theta_n \in (a, b)$ with $\theta_n \rightarrow \theta$ as $n \rightarrow \infty$ and $\theta_n \neq \theta$ for all n

$$\frac{d}{d\theta} f(x, \theta) = \lim_{n \rightarrow \infty} \frac{f(x, \theta_n) - f(x, \theta)}{\theta_n - \theta}$$

for all $x \in \mathbb{X}$ and consequently $\frac{d}{d\theta} f(x, \theta)$ is measurable. By the Mean Value Theorem, there exists $\theta_{c,n} = c\theta_n + (1 - c)\theta$ with $c \in (0, 1)$ such that $f(x, \theta_n) - f(x, \theta) = \frac{d}{d\theta} f(x, \theta_{c,n})(\theta_n - \theta)$.

Consequently,

$$\left| \frac{f(x, \theta_n) - f(x, \theta)}{\theta_n - \theta} \right| = \left| \frac{d}{d\theta} f(x, \theta_{c,n}) \right| \leq g(x),$$

and since g is integrable and f is measurable by assumption, we have by Theorem [4.8](#) that

$\frac{f(x, \theta_n) - f(x, \theta)}{\theta_n - \theta}$ is integrable. Thus,

$$\frac{h(\theta_n) - h(\theta)}{\theta_n - \theta} = \int_{\mathbb{X}} \frac{f(x, \theta_n) - f(x, \theta)}{\theta_n - \theta} d\mu.$$

Hence, by the Lebesgue's Dominated Convergence Theorem

$$\lim_{n \rightarrow \infty} \frac{h(\theta_n) - h(\theta)}{\theta_n - \theta} = \frac{d}{d\theta} h(\theta) = \int_{\mathbb{X}} \lim_{n \rightarrow \infty} \frac{f(x, \theta_n) - f(x, \theta)}{\theta_n - \theta} d\mu = \int_{\mathbb{X}} \frac{d}{d\theta} f(x, \theta) d\mu.$$

■

5.2 $\mathcal{L}^p$ spaces

Definition 5.1. The collection of measurable functions $f : (\mathbb{X}, \mathcal{F}, \mu) \rightarrow (\mathbb{R}, \mathcal{B})$ such that $\int_{\mathbb{X}} |f|^p d\mu < \infty$ for $p \in [1, \infty)$ is denoted by $\mathcal{L}_{\mathbb{R}}^p$, $\mathcal{L}_{\mathbb{R}}^p(\mu)$ or $\mathcal{L}_{\mathbb{R}}^p(\mathbb{X}, \mathcal{F}, \mu)$.

Let $f, g \in \mathcal{L}_{\mathbb{R}}^p(\mathbb{X}, \mathcal{F}, \mu)$ and define $s : (\mathbb{X}, \mathcal{F}, \mu) \rightarrow (\mathbb{R}, \mathcal{B})$ as $s(x) = f(x) + g(x)$ for all $x \in \mathbb{X}$. Then, s is measurable and $|s(x)| \leq |f(x)| + |g(x)| \leq 2 \max\{|f(x)|, |g(x)|\}$ and

$$|s(x)|^p \leq 2^p \max\{|f(x)|, |g(x)|\}^p = 2^p \max\{|f(x)|^p, |g(x)|^p\} \leq 2^p(|f(x)|^p + |g(x)|^p).$$

Consequently, $\int_{\mathbb{X}} |s|^p d\mu \leq 2^p(\int_{\mathbb{X}} |f|^p d\mu + \int_{\mathbb{X}} |g|^p d\mu) < \infty$. Also, if $a \in \mathbb{R}$ and $m : (\mathbb{X}, \mathcal{F}, \mu) \rightarrow (\mathbb{R}, \mathcal{B})$ is defined as $m(x) = af(x)$ for all $x \in \mathbb{X}$, m is measurable and we have $|m(x)|^p = |a|^p |f(x)|^p$ and $\int_{\mathbb{X}} |m|^p d\mu = |a|^p \int_{\mathbb{X}} |f|^p d\mu < \infty$. Lastly, if we take $\mathbf{0}(x) = 0$ for all $x \in \mathbb{X}$ to be the null vector in $\mathcal{L}_{\mathbb{R}}^p(\mathbb{X}, \mathcal{F}, \mu)$, then $\mathcal{L}_{\mathbb{R}}^p(\mathbb{X}, \mathcal{F}, \mu)$ is a vector space.

For every $f \in \mathcal{L}_{\mathbb{R}}^p(\mathbb{X}, \mathcal{F}, \mu)$ we define $\|f\|_p = (\int_{\mathbb{X}} |f|^p d\mu)^{1/p}$ and prove the following inequality.

Theorem 5.8. (Hölder's Inequality) If $1 < p < \infty$, $p^{-1} + q^{-1} = 1$, $f \in \mathcal{L}_{\mathbb{R}}^p$, $g \in \mathcal{L}_{\mathbb{R}}^q$, then $fg \in \mathcal{L}_{\mathbb{R}}$ and $\int_{\mathbb{X}} |fg| d\mu \leq \|f\|_p \|g\|_q$.

Proof. If $\|f\|_p = 0$ then, by Theorem [5.2](#) $|f|^p = 0$ a.e. and $|f| = 0$ a.e. Hence, $|fg| = 0$ a.e. and fg is measurable. Hence, $\int |fg| d\mu = 0$ and the inequality holds. Likewise for $\|g\|_q = 0$. So, assume $\|f\|_p, \|g\|_q \neq 0$. Let $x = f/\|f\|_p$, $y = g/\|g\|_q$ and note that $\|x\|_p = 1$ and $\|y\|_q = 1$. It suffices to prove $\int |xy| d\mu \leq 1$.

Now, note that for any $a, b > 0$ and $0 < \alpha < 1$,

$$a^\alpha b^{1-\alpha} \leq \alpha a + (1-\alpha)b. \quad (5.4)$$

To see this, divide by b to obtain $(\frac{a}{b})^\alpha \leq \alpha \frac{a}{b} + (1-\alpha)$. It suffices to show $u^\alpha \leq \alpha u + (1-\alpha)$, for $u > 0$.

The inequality holds for $u = 1$. Now, $\frac{d}{du}u^\alpha = \alpha u^{\alpha-1} = \alpha \frac{1}{u^{1-\alpha}}$. Since $\alpha \in (0, 1)$ we have that $u^{1-\alpha} < 1$ if $u < 1$. Consequently, in this case, $u^{\alpha-1} > 1$ and $\frac{d}{du}u^\alpha > \alpha$. Also, using the same arguments, if $u > 1$ we have that $\frac{d}{du}u^\alpha < \alpha$. By the Mean Value Theorem, for some $\lambda \in (0, 1)$ and $u > 1$

$$u^\alpha - 1 = \alpha(\lambda u + (1-\lambda))^{\alpha-1}(u-1) < \alpha(u-1) \implies u^\alpha < 1 - \alpha + \alpha u$$

Also, for $u < 1$

$$u^\alpha - 1 = \alpha(\lambda u + (1-\lambda))^{\alpha-1}(u-1) < -\alpha(1-u) \implies u^\alpha < 1 + \alpha u - \alpha.$$

Thus, $u^\alpha \leq \alpha u + (1-\alpha)$ for $u > 0$. Note that if $a = 0$ or $b = 0$ the inequality (5.4) holds trivially.

Now, let $\alpha = 1/p$, $a(\omega) = |x(\omega)|^p$, $b(\omega) = |y(\omega)|^q$ and $1-\alpha = 1/q$. Then,

$$(|x(\omega)|^p)^{1/p}(|y(\omega)|^q)^{1/q} \leq \alpha |x(\omega)|^p + (1-\alpha)|y(\omega)|^q, \text{ or}$$

$$|x(\omega)y(\omega)| \leq \alpha |x(\omega)|^p + (1-\alpha)|y(\omega)|^q.$$

Thus, integrating both sides of the inequality we obtain $\int |xy| d\mu \leq \alpha \|x\|_p^p + (1-\alpha)\|y\|_q^q = 1$.

■

Theorem 5.9. (*Minkowski Inequality*) For $1 \leq p < \infty$, if f and g are in $\mathcal{L}^p$ we have $\|f + g\|_p \leq \|f\|_p + \|g\|_p$.

Proof. Note that $\|f + g\|_p^p = \int |f + g|^p d\mu$ and by the triangle inequality

$$\begin{aligned} \|f + g\|_p^p &\leq \int (|f||f + g|^{p-1} + |g||f + g|^{p-1}) d\mu \\ &= \int |f||f + g|^{p-1} d\mu + \int |g||f + g|^{p-1} d\mu, \text{ and if } p = 1 \text{ the proof is complete.} \end{aligned}$$

If $p > 1$, we can apply Hölder's Inequality since $|f + g|^{(p-1)q} = |f + g|^p$ and $1/p + 1/q = 1$ and

$$\|f + g\|_p^p \leq \|f\|_p \|f + g\|_q^{p-1} + \|g\|_p \|f + g\|_q^{p-1}.$$

Thus,

$$\|f + g\|_p^p \leq \|f\|_p \|f + g\|_q^{p/q} + \|g\|_p \|f + g\|_q^{p/q} = (\|f\|_p + \|g\|_p) \|f + g\|_q^{p/q}. \quad (5.5)$$

Now,

$$\begin{aligned} \|f + g\|_q^{p/q} &= \left(\int (|f + g|^{p/q})^q d\mu \right)^{1/q} = \left(\int |f + g|^p d\mu \right)^{1/q} \\ &= \left(\int |f + g|^p d\mu \right)^{\frac{p-1}{p}} = \|f + g\|_p^{p-1} \end{aligned} \quad (5.6)$$

Substituting (5.6) in (5.5) and dividing by $\|f + g\|_p^{p-1}$ on both sides of (5.5) when $\|f + g\|_p^{p-1} \neq 0$ we obtain $\|f + g\|_p^{p-(p-1)} = \|f + g\|_p \leq \|f\|_p + \|g\|_p$. If $\|f + g\|_p^{p-1} = 0$, there is nothing to prove. ■

Remark 5.3. 1. The Minkowski Inequality and the fact that for $a \in \mathbb{R}$, $\|af\|_p = |a|\|f\|_p$ and $\|f\|_p \geq 0$ shows that $\|\cdot\|_p$ has almost all of the properties of a norm. The exception is that $\|f\|_p = 0$ does not imply that $f(x) = 0$ for all $x \in \mathbb{X}$. It only implies that $f(x) = 0$ a.e.

2. $f, g \in \mathcal{L}_\mathbb{R}^p(\mathbb{X}, \mathcal{F}, \mu)$ are taken to be equivalent if they differ at most on a μ -null set, i.e., $f \sim g$ if $\mu(\{x : f(x) \neq g(x)\}) = 0$. $\sim$ is an equivalence relation (reflexive, symmetric) and transitive in that $\{x : f(x) \neq h(x)\} \subset \{x : f(x) \neq g(x)\} \cup \{x : g(x) \neq h(x)\}$. Then, for every $f \in \mathcal{L}_\mathbb{R}^p(\mathbb{X}, \mathcal{F}, \mu)$ we can define an equivalence class denoted by $[f]_p$. The space of all equivalence classes $[f]_p$ of functions $f \in \mathcal{L}_\mathbb{R}^p$ is denoted by $L_\mathbb{R}^p$ with norm $\|[f]_p\|_p := \inf\{\|g\|_p : g \in \mathcal{L}_\mathbb{R}^p \text{ and } g \sim f\} = \|f\|_p$. $(L^p, \|[f]_p\|_p)$ is a norm vector space and in what follows we will dispense with these technicalities and identify $[f]_p$ with f .

A commonly encountered case for Hölder's Inequality, treated in the next theorem, has $p = 2$ and $X, Y : (\Omega, \mathcal{F}, P) \rightarrow (\mathbb{R}, \mathcal{B})$ being random variables such that $X, Y \in \mathcal{L}^2_{\mathbb{R}}(\Omega, \mathcal{F}, P)$.

Theorem 5.10. *Let $X, Y : (\Omega, \mathcal{F}, P) \rightarrow (\mathbb{R}, \mathcal{B})$ be random variables such that $X, Y \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$.*

1. $XY \in \mathcal{L}(\Omega, \mathcal{F}, P)$ and $|\int_{\Omega} XY dP| \leq \int_{\Omega} |XY| dP \leq (\int_{\Omega} X^2 dP)^{1/2} (\int_{\Omega} Y^2 dP)^{1/2} = \|X\|_2 \|Y\|_2$,
2. $X \in \mathcal{L}(\Omega, \mathcal{F}, P)$ and $(\int_{\Omega} X dP)^2 \leq \int_{\Omega} X^2 dP$,
3. $\mathcal{L}^2(\Omega, \mathcal{F}, P)$ is a vector space.

Proof. 1. This is just a special case of Hölder's Inequality with $p = q = 2$.

2. Let $X \in \mathcal{L}^2$ and note that $I_{\Omega} \in \mathcal{L}^2$ with $\int_{\Omega} I_{\Omega} dP = \int_{\Omega} dP$. Then,

$$\left| \int_{\Omega} X I_{\Omega} dP \right| \leq \int_{\Omega} |X I_{\Omega}| dP \leq \left(\int_{\Omega} X^2 dP \right)^{1/2} \left(\int_{\Omega} dP \right)^{1/2}.$$

Since $\int_{\Omega} dP = 1$, we have

$$\left| \int_{\Omega} X dP \right| \leq \left(\int_{\Omega} X^2 dP \right)^{1/2} \text{ or } \left(\int_{\Omega} X dP \right)^2 \leq \int_{\Omega} X^2 dP.$$

3. follows from the comments after Definition [5.1](#) ■

Remark 5.4. *If $X \in \mathcal{L}^2$ we define $V_P(X) = \int_{\Omega} (X - E_P(X))^2 dP = \int_{\Omega} X^2 dP - (\int_{\Omega} X dP)^2$ and call it the variance of X (under P).*

Theorem 5.11. *Let X be a random variable defined on the probability space $(\Omega, \mathcal{F}, P)$ taking values in $(\mathbb{R}, \mathcal{B})$ and $h : (\mathbb{R}, \mathcal{B}) \rightarrow (\mathbb{R}, \mathcal{B})$ be measurable.*

1. $f := h \circ X$ is integrable in $(\Omega, \mathcal{F}, P)$ if, and only if, h is integrable in $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_X)$, where P_X is as defined in Example [3.2](#).
2. $E_P(h(X)) := \int_{\Omega} f dP = \int_{\mathbb{R}} h dP_X$.

Proof. First, let $h(x) = \sum_{j=0}^m y_j I_{B_j}$ be a non-negative simple function with $B_j = \{x \in \mathbb{R} : h(x) = y_j\} \in \mathcal{B}$ and observe that $h \circ X = \sum_{j=0}^m y_j I_{X^{-1}(B_j)} = \sum_{j=0}^m y_j I_{A_j}$, $A_j \in \mathcal{F}$. Consequently,

$$\begin{aligned} I_P(f) &= \int_{\Omega} f dP = \sum_{j=0}^m y_j P(A_j) = \sum_{j=0}^m y_j P(X^{-1}(B_j)) = \sum_{j=0}^m y_j (P \circ X^{-1})(B_j) \\ &= \sum_{j=0}^m y_j P_X(B_j) = \int_{\mathbb{R}} h dP_X = I_{P_X}(h). \end{aligned}$$

Second, let $h \geq 0$, not necessarily simple. Then, by Theorem 4.4 there exists a nondecreasing sequence of nonnegative simple functions ϕ_n such that $\phi_n \rightarrow h$ as $n \rightarrow \infty$. Hence, if we define $f_n(\omega) = \phi_n(X(\omega)) = (\phi_n \circ X)(\omega)$, it is a sequence of increasing simple functions that converges to f .

$$\begin{aligned} \int_{\Omega} f dP &= \int_{\Omega} (h \circ X) dP = \int_{\Omega} \lim_{n \rightarrow \infty} (\phi_n \circ X) dP \\ &= \lim_{n \rightarrow \infty} \int_{\Omega} (\phi_n \circ X) dP \text{ by Beppo Levi's Theorem} \\ &= \lim_{n \rightarrow \infty} \int_{\mathbb{R}} \phi_n dP_X \text{ by the first part of the argument for simple functions} \\ &= \int_{\mathbb{R}} h dP_X, \text{ by Beppo Levi's Theorem.} \end{aligned}$$

This proves 2. for simple and nonnegative h . If h takes values in $\mathbb{R}$, consider $|h|$ and let ϕ_n be a sequence of increasing nonnegative simple function such that $\phi_n \rightarrow |h|$ as $n \rightarrow \infty$. Then, we have from above that

$$\int_{\Omega} |f| dP = \int_{\mathbb{R}} |h| dP_X.$$

But from Remark 4.5, if $|h|$ is integrable in $(\mathbb{R}, \mathcal{B}, P_X)$ then h is integrable in $(\mathbb{R}, \mathcal{B}, P_X)$, establishing 1. Now, for arbitrary h we can prove the rest of part 2 by applying the same arguments to h^+ and h^- and using the fact that $h = h^+ - h^-$. ■

Clearly, taking $h(x) = x$ in the previous theorem gives $E_P(X) := \int_{\Omega} X dP = \int_{\mathbb{R}} x dP_X(x)$ where in the last integral we emphasize that the “variable” in integration is taking values in

$\mathbb{R}$. In this proof, there is no requirement that $P(\Omega) = 1$. Hence, we can take $(\Omega, \mathcal{F}, P)$ to be an arbitrary measure space.

Definition 5.2. *The density of a probability measure P_X associated with a random variable X defined on a probability space $(\Omega, \mathcal{F}, P)$ is a nonnegative Borel measurable function f_X that satisfies for all $a \in \mathbb{R}$*

$$P_X((-\infty, a]) = \int_{(-\infty, a]} f_X d\lambda = \int_{\mathbb{R}} I_{(-\infty, a]} f_X d\lambda$$

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

Theorem 5.12. (a) *If f_X is a density of P_X , then $\int_{\mathbb{R}} f_X d\lambda = 1$ and $P_X(A) = \int_A f_X d\lambda$ for every $A \in \mathcal{B}(\mathbb{R})$. (b) Conversely, if $f \geq 0$ is Borel measurable with $\int_{\mathbb{R}} f d\lambda = 1$, then $\nu_f(A) := \int_A f d\lambda$ defines a probability measure on $\mathcal{B}$ whose density is f . (c) If f_X and g_X are densities of the same P_X , then $f_X = g_X$ λ -a.e.*

Proof. (a) f_X a density implies that for all $n \in \mathbb{N}$, $P_X((-\infty, n]) = \int_{(-\infty, n]} f_X d\lambda = \int_{\mathbb{R}} f_X I_{(-\infty, n]} d\lambda$. By Lebesgue's Monotone Convergence Theorem $\lim_{n \rightarrow \infty} P_X((-\infty, n]) = 1 = \int_{\mathbb{R}} \lim_{n \rightarrow \infty} f_X I_{(-\infty, n]} d\lambda = \int_{\mathbb{R}} f_X d\lambda$.

Now, let $m(A) := \int_A f_X d\lambda$ be a measure on $\mathcal{B}$. m and P_X are finite and coincide on the collection $\mathcal{P} = \{(-\infty, a] : a \in \mathbb{R}\}$. $\mathcal{P}$ is a π -system that generates $\mathcal{B}$ and $\lim_{a \uparrow \infty} (-\infty, a] = \mathbb{R}$. Hence, by Theorem 2.4, m and P_X coincide on $\mathcal{B}$.

(b) Suppose f is a nonnegative Borel measurable function such that $\int_{\mathbb{R}} f d\lambda = 1$. For all $A \in \mathcal{B}$, we put

$$\nu_f(A) = \int_A f d\lambda = \int_{\mathbb{R}} I_A f d\lambda.$$

By Theorem 4.11, ν_f is a measure on $\mathcal{B}$ with $\nu_f(\mathbb{R}) = 1$, by assumption. Taking $A = (-\infty, a]$,

$$\nu_f((-\infty, a]) = \int_{(-\infty, a]} f d\lambda$$

and f is a density for ν_f .

(c) If g_X is a density for P_X , then, $P_X(A) = \int_A g_X d\lambda = \int_{\mathbb{R}} g_X I_A d\lambda$. Let $A_n = \{x : g_X(x) \geq f_X(x) + 1/n\}$. For all $n \in \mathbb{N}$, $\int_{A_n} g_X d\lambda \geq \int_{A_n} (f_X + \frac{1}{n}) d\lambda = \int_{A_n} f_X d\lambda + \frac{1}{n} \lambda(A_n)$. Since $\int_{A_n} f_X d\lambda = \int_{A_n} g_X d\lambda$ it must be that $\lambda(A_n) = 0$.

Note that $A_1 \subset A_2 \subset \dots$. $\lim_{n \rightarrow \infty} A_n = \cup_{n=1}^{\infty} A_n = A = \{x : g_X(x) > f_X(x)\}$ and $\lambda(A) = \lim_{n \rightarrow \infty} \lambda(A_n) = 0$. Similarly, we have $\lambda(B) = 0$ for $B = \{x : g_X(x) < f_X(x)\}$. So, $\lambda(\{x : g_X \neq f_X\}) \leq \lambda(A) + \lambda(B) = 0$. ■

Theorem 5.13. *Let $X : (\Omega, \mathcal{F}, P) \rightarrow (\mathbb{R}, \mathcal{B})$ be a random variable with density f_X and $h : (\mathbb{R}, \mathcal{B}) \rightarrow (\mathbb{R}, \mathcal{B})$ be a measurable function such that $\int_{\Omega} |h \circ X| dP < \infty$, i.e., $f = h \circ X$ is integrable. Then,*

$$\int_{\Omega} (h \circ X) dP = \int_{\mathbb{R}} h dP_X = \int_{\mathbb{R}} h(x) f_X(x) d\lambda(x)$$

Proof. Exercise. Apply Theorem 5.11. ■