Let $(\mathbb{R}, \sigma(\mathcal{I}^1) = \mathcal{B}(\mathbb{R}))$ be a measurable space. From Theorem 1.4 if we set $S = (0,1]$ and consider $\mathcal{I} = \mathcal{I}^1 \cap S = \{(0,1] \cap A : A \in \mathcal{I}^1\}$ then $\sigma(\mathcal{I}^1 \cap (0,1]) = \mathcal{B}(\mathbb{R}) \cap (0,1]$ is a σ-algebra associated with $(0,1]$. Thus, if we define $\mathcal{B}_{(0,1]} := \sigma(\mathcal{I}^1 \cap (0,1])$, then

$$((0,1], \mathcal{B}_{(0,1]} := \sigma(\mathcal{I}))$$

is a measurable space where $\mathcal{I} = \{(a,b] : 0 \leq a \leq b \leq 1\}$. Define the set function $\lambda : \mathcal{I} \to [0,1]$ such that $\lambda(\emptyset) = 0$ and $\lambda((a,b]) = b - a$.

If λ is σ-additive on $\mathcal{I}$ it is a pre-measure on $\mathcal{I}$ and extends uniquely to $\mathcal{B}_{(0,1]}$.

Theorem 2.9. *λ is σ-additive on $\mathcal{I}$.*

Proof. First, we show that λ is finitely additive on $\mathcal{I}$. Let $(a,b] \in \mathcal{I}$ and $(a,b] = \cup_{i=1}^n (a_i, b_i]$ with $a_1 = a, a_2 = b_1, a_3 = b_2, \ldots, a_n = b_{n-1}, b_n = b$. Then,

$$\begin{aligned}\sum_{i=1}^n \lambda((a_i, b_i]) &= (b_1 - a_1) + (b_2 - a_2) + \cdots + (b_n - a_n) \\ &= (\cancel{a_2} - a) + (\cancel{a_3} - \cancel{a_2}) + \ldots (b - \cancel{a_n}) = b - a \\ &= \lambda((a,b]) = \lambda(\cup_{i=1}^n (a_i, b_i]).\end{aligned}$$

Therefore, λ is finitely additive.

We need to show that for $(a,b] = \cup_{i=1}^\infty (a_i, b_i]$, where $\{(a_i, b_i]\}_{i \in \mathbb{N}}$ is a pairwise disjoint collection we have $b - a = \sum_{i=1}^\infty (b_i - a_i)$.

For any n, let $\{(a_i, b_i]\}_{i=1}^n$ be a pairwise disjoint collection. Then, we can write

$$(a,b] - \cup_{i=1}^n (a_i, b_i] = \cup_{j=1}^m I_j,$$

where the last set is the finite union of pairwise disjoint intervals. Thus, since λ is finitely additive on $\mathcal{I}$

$$\lambda((a,b]) = \sum_{i=1}^n \lambda((a_i, b_i]) + \sum_{i=1}^m \lambda(I_j) \geq \sum_{i=1}^n \lambda((a_i, b_i]).$$

Thus, $\lambda((a,b]) = b - a \geq \lim_{n\to\infty} \sum_{i=1}^n \lambda((a_i, b_i]) = \sum_{i=1}^\infty \lambda((a_i, b_i])$.

Now, for the reverse inequality ($b - a \le \sum_{i=1}^{\infty} \lambda((a_i, b_i]))$ let $0 < \epsilon < b - a$ and note that

$$
\begin{aligned}
(a+\epsilon, b] \subseteq [a+\epsilon, b] \subseteq \cup_{i=1}^{\infty}(a_i, b_i + \frac{1}{2^i}\epsilon) \\
&\subseteq \cup_{i=1}^{n}(a_i, b_i + \frac{1}{2^i}\epsilon) \text{ for some } n \in \mathbb{N} \text{ by the Heine-Borel Theorem} \\
&\subseteq \cup_{i=1}^{n}(a_i, b_i + \frac{1}{2^i}\epsilon].
\end{aligned}
$$

But $\lambda((a_i, b_i] = \lambda((a_i, b_i + \frac{1}{2^i}\epsilon] - \frac{1}{2^i}\epsilon$. Hence,

$$
\begin{aligned}
\lambda((a+\epsilon, b]) &\le \sum_{i=1}^{n} \lambda((a_i, b_i + \frac{1}{2^i}\epsilon]) \text{ by subadditivity} \\
&= \sum_{i=1}^{n} (b_i - a_i + \frac{1}{2^i}\epsilon) \\
b - a - \epsilon &\le \sum_{i=1}^{n} (b_i - a_i) + \epsilon \sum_{i=1}^{n} \frac{1}{2^i} \text{ or} \\
b - a &\le \sum_{i=1}^{n} (b_i - a_i) + \epsilon \left(1 + \sum_{i=1}^{n} \frac{1}{2^i}\right).
\end{aligned}
$$

The last inequality gives $b - a \le \sum_{i=1}^{\infty}(b_i - a_i)$. Hence, combining with the previously obtained reverse inequality we have $b - a = \sum_{i=1}^{\infty}(b_i - a_i)$. ■

Since λ is σ-additive (pre-measure) on $\mathcal{I}$ (a semi-ring), using Carathéodory's theorem, we can state that

$$((0,1], \mathcal{B}_{(0,1]} := \sigma(\mathcal{I}), \lambda^*)$$

is a measure space, where λ^* is the unique extension of λ from $\mathcal{I}$ to $\sigma(\mathcal{I})$. In addition, $0 \le \lambda^*((a,b]) \le 1$, $\lambda^*((0,1]) = 1$. Thus, we have *constructed* a specific probability space.

We will now construct a probability measure on $(\mathbb{R}, \mathcal{B})$.

Definition 2.5. *Let $F : \mathbb{R} \to [0,1]$ be a function with the following properties:*

1. $\lim_{h \downarrow 0} F(x+h) := F(x+) = F(x)$ *for all $x \in \mathbb{R}$ and $h > 0$,*

2. $F(x) \le F(y)$ *if $x < y$,*

3. $\lim_{x\to\infty} F(x) = 1$, $\lim_{x\to-\infty} F(x) = 0$.

F is called a proper distribution function (df). If only conditions 1 and 2 are met, F is called a distribution function or a defective df.

Remark 2.3. *1. Let $F(x-) := \lim_{h\downarrow 0} F(x-h)$ for $h > 0$. The left jump of F at x is defined as $LJ_F(x) = F(x) - F(x-)$ and the right jump of F at x is defined as $RJ_F(x) = F(x+) - F(x)$. The jump of F at x is defined as $J_F(x) = LJ_F(x) + RJ_F(x) = F(x+) - F(x-)$. Since F satisfies condition 2, $RJ_F(x) = 0$ for all $x \in \mathbb{R}$ and $J_F(x) = F(x) - F(x-)$. In addition, since F is nondecreasing $J_F(x) \geq 0$. If $J_F(x) = 0$ then F is continuous at x.*
2. For any two $x \leq y \in \mathbb{R}$ we have that $0 \leq F(y) - F(x) \leq 1$

Definition 2.6. *The left (generalized) inverse of a df F, denoted by $F^{-}(y)$, is defined as*

$$F^{-}(y) := \inf\{x : F(x) \geq y \text{ for } y \in (0,1]\}.$$

Theorem 2.10. *Let $S(y) = \{x : F(x) \geq y\}$ for $y \in (0,1]$. Then,*

1. *$S(y)$ is a closed set.*

2. *$F^{-}(y) > t \iff y > F(t)$ or $F^{-}(y) \leq t \iff y \leq F(t)$.*

Proof. 1. If $s_n \in S(y)$ and $s_n \downarrow s$, by right continuity of F we have $y \leq F(s_n) \downarrow F(s)$. Thus, $y \leq F(s)$ and $s \in S(y)$. If $s_n \in S(y)$ and $s_n \uparrow s$, we have $y \leq F(s_n) \uparrow F(s-) \leq F(s)$. Thus, $y \leq F(s)$ which implies that $s \in S(y)$. Consequently, by a characterization of closed sets, $S(y)$ is closed.
2. Since $S(y)$ is closed, its infimum $F^{-}(y) \in S(y)$ and therefore $F(F^{-}(y)) \geq y$. $t < F^{-}(y) \implies t \notin S(y) \implies F(t) < y$. The reverse implications all apply. ■

Theorem 2.11. *Let $A \subseteq \mathbb{R}$ and $\mathcal{S}_F(A) = \{y \in (0,1] : F^{-}(y) \in A\}$. If $A \in \sigma(\mathcal{I}^1) = \mathcal{B}(\mathbb{R})$, then $\mathcal{S}_F(A) \in \mathcal{B}_{(0,1]} = \sigma(\mathcal{I}^1) \cap (0,1]$.*

Proof. Let $\mathcal{G} = \{A \subseteq \mathbb{R} : \mathcal{S}_F(A) \in \mathcal{B}_{(0,1]}\}$. Note that $\mathcal{G}$ contains intervals of the type $(a, b]$ since

$$\begin{aligned}\mathcal{S}_F((a,b]) &= \{y \in (0,1] : F^-(y) \in (a,b]\} = \{y \in (0,1] : a < F^-(y) \leq b\} \\ &= \{y \in (0,1] : F(a) < y \leq F(b)\} = (F(a), F(b)] \in \mathcal{B}_{(0,1]}.\end{aligned}$$

Note that since $\mathcal{I}^1 \subseteq \mathcal{G}$, if $\mathcal{G}$ is a σ-algebra, $\sigma(\mathcal{I}^1) = \mathcal{B} \subseteq \mathcal{G}$. Hence, $A \in \mathcal{B}$ implies $S_F(A) \in \mathcal{B}_{(0,1]}$. Hence, we now show that $\mathcal{G}$ is a σ-algebra associated with $\mathbb{R}$.

1. $\mathcal{S}_F(\mathbb{R}) = \{y \in (0,1] : F^-(y) \in \mathbb{R}\} = (0,1] \in \mathcal{B}_{(0,1]}$, thus $\mathbb{R} \in \mathcal{G}$.

2. By definition of $\mathcal{S}_F$

$$\begin{aligned}\mathcal{S}_F(A^c) &= \{y \in (0,1] : F^-(y) \in A^c\} \\ &= \{y \in (0,1] : F^-(y) \in A\}^c = (\mathcal{S}_F(A))^c \in \mathcal{B}_{(0,1]}\end{aligned}$$

   where the last inclusion statement follows if $A \in \mathcal{G}$ and the fact that $\mathcal{B}_{(0,1]}$ is a σ-algebra.

3. If $\{A_n\}_{n\in\mathbb{N}} \in \mathcal{G}$ we have by definition of $\mathcal{S}_F$

$$\begin{aligned}\mathcal{S}_F(\underset{n\in\mathbb{N}}{\cup} A_n) &= \{y \in (0,1] : F^-(y) \in \underset{n\in\mathbb{N}}{\cup} A_n\} \\ &= \underset{n\in\mathbb{N}}{\cup}\{y \in (0,1] : F^-(y) \in A_n\} = \underset{n\in\mathbb{N}}{\cup}\mathcal{S}_F(A) \in \mathcal{B}_{(0,1]}\end{aligned}$$

   where the last inclusion statement follows since $A_n \in \mathcal{G}$ and the fact that $\mathcal{B}_{(0,1]}$ is a σ-algebra.

■

Definition 2.7. *Let $A \in \mathcal{B}$ and define $P_F(A) = \lambda(\mathcal{S}_F(A))$ where λ is the Lebesgue measure on $\mathcal{B}_{(0,1]}$.*

Remark 2.4. *It is easy to verify that P_F is a probability measure. First, note that*

$$P_F(\emptyset) = \lambda(\mathcal{S}_F(\emptyset)) = \lambda(\{y \in (0,1] : F^-(y) \in \emptyset\}) = \lambda(\emptyset) = 0.$$

Second, if $\{A_n\}_{n\in\mathbb{N}}$ is a pairwise disjoint collection of sets in $\mathcal{B}$ then

$$P_F(\underset{n\in\mathbb{N}}{\cup} A_n) = \lambda(\mathcal{S}_F(\underset{n\in\mathbb{N}}{\cup} A_n)) = \lambda(\{y \in (0,1] : F^-(y) \in \underset{n\in\mathbb{N}}{\cup} A_n\}) = \lambda(\underset{n\in\mathbb{N}}{\cup} \mathcal{S}_F(A_n))$$
$$= \sum_{n=1}^{\infty} \lambda(\mathcal{S}_F(A_n)) = \sum_{n=1}^{\infty} P_F(A_n).$$

where the next to last equality follows from the fact that λ is a measure and $\{\mathcal{S}_F(A_n)\}_{n\in\mathbb{N}}$ is a pairwise disjoint collection.

Its df can be obtained by noting that

$$P_F((-\infty,x]) = \lambda(\mathcal{S}_F((-\infty,x])) = \lambda(\{y \in (0,1] : F^-(y) \in (-\infty,x]\})$$
$$= \lambda(\{y \in (0,1] : y \le F(x)\}) = \lambda((0,F(x)]) = F(x)$$

This last equality will be used throughout the notes. It is fundamental to our study.

If we take $(\mathbb{X},\mathcal{F},\mu)$ to be a probability space $(\Omega,\mathcal{F},P)$, $(\mathbb{E},\mathcal{E}) := (\mathbb{R},\mathcal{B})$ and $X:\Omega\to\mathbb{R}$ be a random variable, then Theorem 1.18 establishes that

$$P_X(B) = P(X^{-1}(B)) = (P\circ X^{-1})(B) \text{ for all } B \in \mathcal{B} \text{ is a measure on } (\mathbb{R},\mathcal{B}).$$

We call P_X the distribution measure (or distribution law) of X.

Chapter 3

Integration

3.1 Simple functions

In many cases it is convenient to use $-\infty$ or ∞ in calculations. In these cases we work with the extended real line, i.e., $\bar{\mathbb{R}} := \mathbb{R} \cup \{-\infty\} \cup \{\infty\}$. Functions that take values in $\bar{\mathbb{R}}$ are called *numerical* functions. The Borel sets associated with the extended real line are denoted by $\bar{\mathcal{B}} := \mathcal{B}(\bar{\mathbb{R}})$ and are defined as the collection of sets $\bar{B}$ such that $\bar{B} = B \cup S$ where $B \in \mathcal{B}(\mathbb{R})$ and $S \in \{\emptyset, \{-\infty\}, \{\infty\}, \{-\infty, \infty\}\}$. It is easy to verify that $\bar{\mathcal{B}}$ is a σ-algebra and that $\mathcal{B}(\mathbb{R}) = \mathbb{R} \cap \mathcal{B}(\bar{\mathbb{R}})$. In addition, $\bar{\mathcal{B}}$ is generated by a collection of sets of the form $[a, \infty]$ (or $(a, \infty]$, $[-\infty, a]$, $[-\infty, a)$) where $a \in \mathbb{R}$.

Let $(\mathbb{X}, \mathcal{F})$ and $(\mathbb{R}, \mathcal{B})$ be measurable spaces. Since the indicator function of a measurable set is a measurable function, it follows from Theorem 1.19 that if $\{A_j\}_{j=1}^n$ with $n \in \mathbb{N}$ is a pairwise disjoint collection in $\mathcal{F}$ and $a_j \in \mathbb{R}$ for $j = 1, \cdots, n$, the linear combination

$$f(x) = \sum_{i=1}^{n} a_j I_{A_j}(x) \tag{3.1}$$

is a $\mathcal{F}/\mathcal{B}$-measurable function.

Definition 3.1. *A real-valued function on a measurable space* $(\mathbb{X}, \mathcal{F})$ *is said to be simple if it has the representation* (3.1)*. A standard representation of a simple function is given by*

$$f(x) = \sum_{j=0}^{n} a_j I_{A_j}(x) \text{ with } a_0 = 0 \text{ and } A_0 = (\cup_{j=1}^{n} A_j)^c. \tag{3.2}$$

Remark 3.1. 1. *If $f : (\mathbb{X}, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$ is measurable and takes on finitely many values, say $\{a_j\}_{j=1}^n$ then it is a simple function. To see this, note that that B_j is measurable, since $B_j = \{x : f(x) = a_j\} = \{x : f(x) \leq a_j\} - \{x : f(x) < a_j\}$ and f is measurable. Also, note that the collection $\{B_j\}_{j=1}^n$ is pairwise disjoint. Hence,*

$$f(x) = \sum_{j=1}^{n} a_j I_{B_j}(x) = \sum_{j=1}^{n} a_j I_{f=a_j}(x). \tag{3.3}$$

Conversely, if X is simple it takes on finitely many values.

2. *Representation (3.2) is not unique, but a simple function has at least one representation such as (3.2) .*

Theorem 3.1. *Let $f : (\mathbb{X}, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$ and $g : (\mathbb{X}, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$ be simple functions. Then, $f \pm g$, cf for $c > 0$, fg, $f^+ = \max\{f, 0\}$, $f^- = -\min\{f, 0\}$ and $|f|$ are simple functions.*

Proof. Homework. ■

3.2 Integral of simple functions

Definition 3.2. *Let $f : (\mathbb{X}, \mathcal{F}, \mu) \to (\mathbb{R}, \mathcal{B})$ be a non-negative simple function with standard representation (3.2). The integral of f with respect to μ, denoted by $\int_{\mathbb{X}} f d\mu$, is given by*

$$\int_{\mathbb{X}} f d\mu := \sum_{j=0}^{n} a_j \mu(A_j) \in [0, \infty]. \tag{3.4}$$

By definition $a_j \in \mathbb{R}$ for $j = 0, 1, \ldots, n$, but since μ takes values in $[0, \infty]$ we can have $\int_{\mathbb{X}} f d\mu = \infty$. If μ is a finite measure, e.g., a probability measure P, then it must be that $\int_{\mathbb{X}} f d\mu \in \mathbb{R}$. When $\mathbb{X} : \Omega$ an outcome space, $f := X$ is a random variable and $\mu := P$ is a probability measure we write $E_P(X) := \int_{\Omega} X dP$ and call it the expectation of X given probability P. It will be convenient, in the case of simple functions, to write $I_\mu(f) := \int_{\mathbb{X}} f d\mu$.

Remark 3.2. *Since the representation (3.2) is not unique, for uniqueness, the definition of integral requires that it be invariant to the representation used. To see this, suppose that* $f(x) = \sum_{j=0}^{n} a_j I_{A_j}(x) = \sum_{k=0}^{m} b_k I_{B_k}(x)$. *Then,* $\mathbb{X} = \cup_{j=0}^{n} A_j = \cup_{k=0}^{m} B_k$ *and*

$$A_j = \cup_{k=0}^{m}(A_j \cap B_k),\ B_k = \cup_{j=0}^{n}(A_j \cap B_k).$$

Since μ *finitely additive and the sets in the above unions are disjoint we have that*

$$\sum_{j=0}^{n} a_j \mu(A_j) = \sum_{j=0}^{n} a_j \sum_{k=0}^{m} \mu(A_j \cap B_k) = \sum_{j=0}^{n}\sum_{k=0}^{m} a_j \mu(A_j \cap B_k).$$

Similarly,

$$\sum_{k=0}^{m} b_k \mu(B_k) = \sum_{k=0}^{m} b_k \sum_{j=0}^{n} \mu(A_j \cap B_k) = \sum_{j=0}^{n}\sum_{k=0}^{m} b_k \mu(A_j \cap B_k).$$

But $a_j = b_k$ *whenever* $A_j \cap B_k \neq \emptyset$, *and when* $A_j \cap B_k = \emptyset$, $\mu(A_j \cap B_k) = 0$. *Thus,* $a_j \mu(A_j \cap B_k) = b_k \mu(A_j \cap B_k)$ *for all pairs* (j,k), *and* $I_\mu(f)$ *is unique.*

Theorem 3.2. *Let* $f : (\mathbb{X}, \mathcal{F}, \mu) \to (\mathbb{R}, \mathcal{B})$ *and* $g : (\mathbb{X}, \mathcal{F}, \mu) \to (\mathbb{R}, \mathcal{B})$ *be simple non-negative functions. Then,*

a) $\int_{\mathbb{X}} cf d\mu = c \int_{\mathbb{X}} f d\mu$ *for* $c \geq 0$ *and* $\int_{\mathbb{X}} I_E d\mu = \mu(E)$ *for* $E \in \mathcal{F}$.

b) $\int_{\mathbb{X}} (f+g) d\mu = \int_{\mathbb{X}} f d\mu + \int_{\mathbb{X}} g d\mu$,

c) If for $E \in \mathcal{F}$, *we have* $\lambda(E) = \int_{\mathbb{X}} f I_E d\mu$, *then* λ *is a measure on* $\mathcal{F}$.

d) $f \leq g \implies \int_{\mathbb{X}} f d\mu \leq \int_{\mathbb{X}} g d\mu$.

Proof. For a) note that $c \geq 0 \implies cf \geq 0$ with representation $cf(x) = \sum_{j=0}^{n} c a_j I_{A_j}(x)$. Therefore, $\int_{\mathbb{X}} cf d\mu = \sum_{j=0}^{n} c a_j \mu(A_j) = c \sum_{j=0}^{n} a_j \mu(A_j) = c \int_{\mathbb{X}} f d\mu$. For the second part, note that $I_E(x) = I_E(x) + 0 I_{E^c}(x)$. Hence, $\int_{\mathbb{X}} I_E d\mu = \mu(E)$.

For b) let $f(x) = \sum_{j=0}^{n} a_j I_{A_j}(x)$ and $g(x) = \sum_{k=0}^{m} b_k I_{B_k}(x)$. Then, $f(x) + g(x) = \sum_{j=0}^{n}\sum_{k=0}^{m}(a_j + b_k) I_{A_j \cap B_k}(x)$ with $(A_j \cap B_k) \cap (A_{j'} \cap B_{k'}) = \emptyset$ whenever $(j,k) \neq (j',k')$.

Then,

$$\begin{aligned}
\int_{\mathbb{X}} (f+g)\, d\mu &= \sum_{j=0}^{n}\sum_{k=0}^{m}(a_j+b_k)\mu(A_j\cap B_k)\\
&= \sum_{j=0}^{n} a_j \sum_{k=0}^{m}\mu(A_j\cap B_k) + \sum_{k=0}^{m} b_k \sum_{j=0}^{n}\mu(A_j\cap B_k)\\
&= \sum_{j=0}^{n} a_j\mu(A_j) + \sum_{k=0}^{m} b_k\mu(B_k),
\end{aligned}$$

since $\mathbb{X}$ is the union of both $\{A_j\}$ and $\{B_k\}$. Then, by definition $\int_{\mathbb{X}}(f+g)d\mu = \int_{\mathbb{X}} fd\mu + \int_{\mathbb{X}} gd\mu$.

For c) note that $f(x)I_E(x) = \sum_{j=0}^{n} a_j I_{A_j\cap E}(x)$. From b) and a),

$$\lambda(E) = \int_{\mathbb{X}} fI_E d\mu = \sum_{j=0}^{n} a_j \int_{\Omega} I_{A_j\cap E}(x)d\mu = \sum_{j=0}^{n} a_j\mu(A_j\cap E).$$

But $\mu(A_j\cap E)$ is a measure, and we have expressed $\lambda(E)$ as a linear combination of measures on $\mathcal{F}$, hence λ is a measure on $\mathcal{F}$.

For d) write $g = f + (g-f)$. Note that $g-f$ is simple and non-negative since $g \geq f$. Hence, $I_\mu(g) = I_\mu(f) + I_\mu(g-f) \geq I_\mu(f)$. ■

3.3 Integral of non-negative functions

We start with the following fundamental theorem.

Theorem 3.3. *Let $f(\omega) : (\Omega, \mathcal{F}) \to (\bar{\mathbb{R}}, \bar{\mathcal{B}})$ be a non-negative measurable function. Then, there exists a sequence $\varphi_n(\omega) : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$ of simple non-negative functions such that:*

1. *$\varphi_n(\omega) \leq \varphi_{n+1}(\omega)$, for all $\omega \in \Omega$ and $n \in \mathbb{N}$*

2. *$\lim_{n\to\infty} \varphi_n(\omega) = f(\omega)$, for all $\omega \in \Omega$.*

Proof. 1. For each $n = 1, 2, \ldots$ define the sets

$$E_{k,n} = \begin{cases} \left\{\omega \in \Omega : \frac{k}{2^n} \leq f(\omega) < \frac{k}{2^n} + \frac{1}{2^n}\right\} = f^{-1}([\frac{k}{2^n}, \frac{k}{2^n} + \frac{1}{2^n})) \text{ for } k = 0, 1, \ldots, n2^n - 1 \\ \{\omega \in \Omega : f(\omega) \geq n\} = f^{-1}([n, \infty]) \text{ for } k = n2^n. \end{cases}$$

For each n, the sets $\{E_{k,n} : k = 0, 1, \ldots, n2^n\}$ are disjoint by construction, belong to $\mathcal{F}$ since f is measurable and $\cup_{k=0}^{n2^n} E_{k,n} = \Omega$. Now, let

$$\varphi_n(\omega) = \sum_{k=0}^{n2^n} \frac{k}{2^n} I_{E_{k,n}}(\omega).$$

Fix $\omega \in \Omega$ and for any $n \in \mathbb{N}$ we note that $\omega \in E_{k_0,n}$ for some k_0. By definition

$$\varphi_n(\omega) = \begin{cases} \frac{k_0}{2^n} \text{ if } k_0 = 0, 1, \cdots, n2^n - 1 \\ n \text{ if } k_0 = n2^n. \end{cases}$$

First, let $k_0 \in \{0, 1, \cdots, n2^n - 1\}$ and consider $n+1$. The lower bound on $[\frac{k_0}{2^n}, \frac{k_0}{2^n} + \frac{1}{2^n})$ must coincide with $\frac{k}{2^{n+1}}$, which gives $k = 2k_0$. Thus, $E_{k,n+1} = E_{2k_0,n+1} = f^{-1}\left([\frac{2k_0}{2^{n+1}}, \frac{2k_0}{2^{n+1}} + \frac{1}{2^{n+1}})\right) = f^{-1}\left([\frac{k_0}{2^n}, \frac{k_0}{2^n} + \frac{1}{2^{n+1}})\right)$ and

$$E_{k+1,n+1} = E_{2k_0+1,n+1} = f^{-1}\left([\frac{k_0}{2^n} + \frac{1}{2^{n+1}}, \frac{k_0}{2^n} + \frac{2}{2^{n+1}})\right) = f^{-1}\left([\frac{k_0}{2^n} + \frac{1}{2^{n+1}}, \frac{k_0}{2^n} + \frac{1}{2^n})\right).$$

Consequently, $E_{k_0,n} = E_{k,n+1} \cup E_{k+1,n+1} = E_{2k_0,n+1} \cup E_{2k_0+1,n+1}$. If $\omega \in E_{2k_0,n+1}$ then $\varphi_{n+1}(\omega) = \frac{2k_0}{2^{n+1}}$ and $\varphi_{n+1}(\omega) - \varphi_n(\omega) = \frac{2k_0}{2^{n+1}} - \frac{k_0}{2^n} = 0$. Alternatively, if $\omega \in E_{2k_0+1,n+1}$ then $\varphi_{n+1}(\omega) = \frac{2k_0+1}{2^{n+1}}$ and $\varphi_{n+1}(\omega) - \varphi_n(\omega) = \frac{2k_0+1}{2^{n+1}} - \frac{k_0}{2^n} = \frac{1}{2^{n+1}} > 0$. Consequently, if $\omega \in E_{k_0,n}$ then $\varphi_{n+1}(\omega) - \varphi_n(\omega) \geq 0$.

Second, if $k_0 = n2^n$ then $E_{k_0,n} = f^{-1}([n, \infty])$. Now, if $\omega \in f^{-1}([n+1, \infty])$ then $\varphi_{n+1}(\omega) = n + 1$ and $\varphi_n(\omega) = n$. Consequently, $\varphi_{n+1}(\omega) - \varphi_n(\omega) = 1 > 0$. If $\omega \in f^{-1}([n, n+1])$ then $\varphi_n(\omega) = n$ and $\varphi_{n+1}(\omega) = \frac{k}{2^{n+1}}$ if $\omega \in f^{-1}([\frac{k}{2^{n+1}}, \frac{k}{2^{n+1}} + \frac{1}{2^{n+1}}))$. Setting the lower bound of the interval equal to n gives $k = n2^{n+1}$ and $\varphi_{n+1}(\omega) = n$ if $\omega \in f^{-1}([n, n + \frac{1}{2^{n+1}}))$, giving $\varphi_{n+1}(\omega) - \varphi_n(\omega) = 0$. If $\omega \in f^{-1}([n + \frac{1}{2^{n+1}}, n + \frac{2}{2^{n+1}}))$ then $\varphi_{n+1}(\omega) = \frac{n2^{n+1}+1}{2^{n+1}}$ and consequently $\varphi_{n+1}(\omega) - \varphi_n(\omega) = \frac{1}{2^{n+1}} > 0$. Continuing in this fashion for subsequent sub-intervals of $[n, n+1]$ gives $\varphi_{n+1}(\omega) - \varphi_n(\omega) \geq 0$.

2. From item 1, we have that $\varphi_1(\omega) \leq \varphi_2(\omega) \leq \cdots \leq f(\omega)$ for all $\omega \in \Omega$. Hence, $\lim_{n\to\infty} \varphi_n(\omega) = \sup_{n\in\mathbb{N}} \varphi_n(\omega)$. But $0 \leq f(\omega) - \varphi_n(\omega) \leq \frac{1}{2^n}$ and taking limits as $n \to \infty$ we have $f(\omega) = \lim_{n\to\infty} \varphi_n(\omega) = \sup_{n\in\mathbb{N}} \varphi_n(\omega)$. ■

Definition 3.3. *Let $f : (\mathbb{X}, \mathcal{F}, \mu) \to (\bar{\mathbb{R}}, \bar{\mathcal{B}})$ be a non-negative measurable function. The integral of f with respect to μ is given by*

$$\int_{\mathbb{X}} f d\mu := \sup_{\varphi} \int_{\mathbb{X}} \varphi(x) d\mu := \sup_{\varphi} I_\mu(\varphi) \in [0, \infty], \tag{3.5}$$

where the sup *is taken over all simple functions φ which are non-negative satisfying $\varphi(x) \leq f(x)$ for all $x \in \mathbb{X}$.*

Remark 3.3. *If f is a non-negative simple function $\int_{\mathbb{X}} f d\mu = I_\mu(f)$.*

Theorem 3.4. *(Beppo-Levi Theorem) Let $(\mathbb{X}, \mathcal{F}, \mu)$ be a measure space and $\{f_j\}_{j\in\mathbb{N}}$ be an increasing sequence of non-negative measurable functions $f_j : (\mathbb{X}, \mathcal{F}) \to (\bar{\mathbb{R}}, \bar{\mathcal{B}})$. Then $f = \sup_{j\in\mathbb{N}} f_j$ is a non-negative measurable function and*

$$\int_{\mathbb{X}} f d\mu := \int_{\mathbb{X}} \sup_{j\in\mathbb{N}} f_j d\mu = \sup_{j\in\mathbb{N}} \int_{\mathbb{X}} f_j d\mu.$$

Proof. That f is a non-negative measurable function follows from Theorem 1.20. Note that if g and h are non-negative measurable functions, we have by definition that

$$\int_{\mathbb{X}} g d\mu := \sup_{\varphi} \int_{\mathbb{X}} \varphi d\mu \text{ where } \varphi \leq g, \ \varphi \text{ a simple function.}$$

But since $g \leq h$,

$$\int_{\mathbb{X}} g d\mu \leq \sup_{\varphi} \int_{\mathbb{X}} \varphi d\mu = \int_{\mathbb{X}} h d\mu \text{ where } \varphi \leq h.$$

Now, $f_j \leq f := \sup_{j\in\mathbb{N}} f_j$. By the monotonicity of integrals, which we just established,

$$\int_{\mathbb{X}} f_j d\mu \leq \int_{\mathbb{X}} f d\mu.$$

Taking $\sup_{j\in\mathbb{N}}$ on both sides gives $\sup_{j\in\mathbb{N}} \int_{\mathbb{X}} f_j d\mu \leq \int_{\mathbb{X}} f d\mu$.

Now, we establish the reverse inequality, i.e., $\sup_{j\in\mathbb{N}} \int_{\mathbb{X}} f_j d\mu \geq \int_{\mathbb{X}} f d\mu$. Let $\varphi(x)$ be a simple function such that $\varphi \leq f$. If we can show that

$$I_\mu(\varphi) = \int_{\mathbb{X}} \varphi d\mu \leq \sup_{j\in\mathbb{N}} \int_{\mathbb{X}} f_j d\mu \tag{3.6}$$

we will have the desired inequality since we can take sup over all simple functions on both sides of (3.6) to give

$$\sup_{\varphi}\int_{\mathbb{X}}\varphi d\mu := \int_{\mathbb{X}} f d\mu \ \leq \sup_{j\in\mathbb{N}}\int_{\mathbb{X}} f_j d\mu.$$

Let φ be a simple nonnegative function such that $\varphi \leq f$. Since $f(x) := \sup_{j\in\mathbb{N}} f_j(x)$, for every $x \in \mathbb{X}$ and $\epsilon \in (0,1)$, there exists $N_{(x,\epsilon)}$ such that

$$f_j(x) \geq \epsilon\varphi(x) \text{ whenever } j \geq N_{(x,\epsilon)}.$$

Now, if $A_j = \{x : f_j(x) \geq \epsilon\varphi(x)\}$ we note that the sets A_j increase as $j \to \infty$ since $f_1 \leq f_2 \cdots$. Furthermore, these sets are measurable by measurability of f_j and φ. By definition of A_j

$$\epsilon I_{A_j}(x)\varphi(x) \leq I_{A_j}(x) f_j(x) \leq f_j(x). \tag{3.7}$$

Since φ is a simple function it has a standard representation $\varphi(x) = \sum_{i=0}^{m} y_i I_{B_i}(x)$ and

$$\epsilon I_{A_j}(x)\sum_{i=0}^{m} y_i I_{B_i}(x) = \epsilon\sum_{i=0}^{m} y_i I_{B_i\cap A_j}(x).$$

Thus, the integral of the simple function in this expression is given by $\epsilon\sum_{i=0}^{m} y_i\mu(B_i\cap A_j)$. By monotonicity of integrals and using (3.7) we have

$$\epsilon\sum_{i=0}^{m} y_i\mu(B_i\cap A_j) \leq \int_{\mathbb{X}} f_j d\mu \leq \sup_{j\in\mathbb{N}}\int_{\mathbb{X}} f_j d\mu.$$

Since $\varphi \leq f$, the collection $\{A_j\}$ grows to $\mathbb{X}$ as $j \to \infty$. Thus, by the fact that μ is continuous from below

$$\mu(B_i\cap A_j) \uparrow \mu(B_i\cap\mathbb{X}) = \mu(B_i) \text{ as } j \to \infty$$

and

$$\epsilon\sum_{i=0}^{m} y_i\mu(B_i) = \epsilon\int_{\mathbb{X}}\varphi d\mu \leq \sup_{j\in\mathbb{N}}\int_{\mathbb{X}} f_j d\mu.$$

Now, just let ϵ be arbitrarily close to 1 to finish the proof. ■

Remark 3.4. *1. If we take $f_j = \varphi_j$ where φ_j are non-negative simple functions and $f = \sup_j \varphi_j$, then*

$$\int_{\mathbb{X}} f d\mu = \sup_{j \in \mathbb{N}} \int_{\mathbb{X}} \varphi_j d\mu.$$

Note that $\sup$ *can be replaced with* $\lim_{j\to\infty}$.

2. If $E \in \mathcal{F}$, then $I_E(x)f(x)$ is a non-negative measurable function if $f \geq 0$. We define

$$\int_E f d\mu = \int_{\mathbb{X}} I_E f d\mu. \tag{3.8}$$

Theorem 3.5. *Let $(\mathbb{X}, \mathcal{F}, \mu)$ be a measure space and $f, g : (\mathbb{X}, \mathcal{F}, \mu) \to (\bar{\mathbb{R}}, \bar{\mathcal{B}})$ be numerical non-negative measurable functions. Then*

1. $\int\limits_{\mathbb{X}} I_A d\mu = \mu(A)$ *for all* $A \in \mathcal{F}$,
2. $\int\limits_{\mathbb{X}} a f d\mu = a \int\limits_{\mathbb{X}} f d\mu$ *for* $a \geq 0$,
3. $\int\limits_{\mathbb{X}} (f+g) d\mu = \int\limits_{\mathbb{X}} f d\mu + \int\limits_{\mathbb{X}} g d\mu$,
4. *If $E, F \in \mathcal{F}$ and $E \subseteq F$, then $\int_E f d\mu \leq \int_F f d\mu$.*

Proof. 1. $\int\limits_{\mathbb{X}} I_A d\mu = I_\mu(I_A) = \mu(A)$. 2. If $a > 0$, let φ_n be an increasing sequence of measurable non-negative simple functions converging to f (such sequence exists by Theorem 3.3). Then, $a\varphi_n(\omega)$ is an increasing sequence converging to af. By Theorem 3.4 and the fact that $I_\mu(a\varphi_n) = aI_\mu(\varphi_n)$

$$\int\limits_{\mathbb{X}} a f d\mu = \lim_{n\to\infty} \int\limits_{\mathbb{X}} a\varphi_n d\mu = a \lim_{n\to\infty} \int\limits_{\mathbb{X}} \varphi_n(\omega) d\mu = a \int\limits_{\mathbb{X}} f d\mu$$

3. Let φ_n, ψ_n be non-negative increasing simple functions converging to f and g. Then

$\varphi_n + \psi_n$ is an increasing sequence converging to $f+g$. Again, by Theorem 3.4

$$\begin{aligned}
\int\limits_{\mathbb{X}} (f+g)d\mu &= \lim_{n\to\infty} \int\limits_{\mathbb{X}} (\varphi_n + \psi_n)d\mu \\
&= \lim_{n\to\infty} \int\limits_{\mathbb{X}} \varphi_n d\mu + \lim_{n\to\infty} \int\limits_{\mathbb{X}} \psi_n d\mu \\
&= \int\limits_{\mathbb{X}} f d\mu + \int\limits_{\mathbb{X}} g d\mu.
\end{aligned}$$

4. $fI_E \leq fI_F$ therefore

$$\int_E f d\mu = \int\limits_{\mathbb{X}} fI_E d\mu \leq \int\limits_{\mathbb{X}} fI_F d\mu = \int_F f d\mu.$$

■

Corollary 3.1. *Let $\{f_j\}_{j\in\mathbb{N}}$ be a sequence of measurable non-negative numerical functions ($f_j : (\mathbb{X}, \mathcal{F}, \mu) \to (\bar{\mathbb{R}}, \bar{\mathcal{B}})$). Then, $\sum_{j=1}^{\infty} f_j$ is measurable and*

$$\int_{\mathbb{X}} \left(\sum_{j=1}^{\infty} f_j \right) d\mu = \sum_{j=1}^{\infty} \int_{\mathbb{X}} f_j d\mu.$$

Proof. Let $S_m = \sum_{j=1}^{m} f_j$, $S = \lim\limits_{m\to\infty} \sum_{j=1}^{m} f_j = \sum_{j=1}^{\infty} f_j$ and note that $0 \leq S_1 \leq S_2 \leq \cdots$. Then, by Theorem 3.5.3 we have that

$$\int_{\mathbb{X}} S_m d\mu = \sum_{j=1}^{m} \int_{\mathbb{X}} f_j d\mu.$$

Taking limits as $m \to \infty$ and using Theorem 3.4, we have

$$\lim_{m\to\infty} \int_{\mathbb{X}} S_m d\mu = \lim_{m\to\infty} \sum_{j=1}^{m} \int_{\mathbb{X}} f_j d\mu = \sum_{j=1}^{\infty} \int_{\mathbb{X}} f_j d\mu = \int_{\mathbb{X}} S d\mu = \int_{\mathbb{X}} \left(\sum_{j=1}^{\infty} f_j \right) d\mu.$$

■

Theorem 3.6. *(Fatou's Lemma): Let $\{f_j\}_{j\in\mathbb{N}}$ be a sequence of measurable non-negative numerical functions ($f_j : (\mathbb{X}, \mathcal{F}, \mu) \to (\bar{\mathbb{R}}, \bar{\mathcal{B}})$). Then, $f := \liminf\limits_{n\to\infty} f_j$ is measurable and*

$$\int_{\mathbb{X}} f d\mu \leq \liminf_{j\to\infty} \int_{\mathbb{X}} f_j d\mu.$$

Proof. Let $g_n = \inf\{f_n, f_{n+1}, \dots\}$ for $n = 1, 2, \dots$, and note that $g_1 \le f_1$, $g_1 \le f_2, \dots$. Also, $g_2 \le f_2, g_2 \le f_3 \dots$. Thus, $g_n \le f_j$ for all $n \le j$. Furthermore, $g_1 \le g_2 \le \dots$. Now, recall that $f := \liminf_{j\to\infty} f_j := \sup_{n\in\mathbb{N}} \inf_{j\ge n} f_j$ and

$$\lim_{n\to\infty} g_n = \liminf_{j\to\infty} f_j := f.$$

Also, $\int_{\mathbb{X}} g_n d\mu \le \int_{\mathbb{X}} f_j d\mu$ for all $n \le j$ and

$$\int_{\mathbb{X}} g_n d\mu \le \liminf_{j\to\infty} \int_{\mathbb{X}} f_j d\mu.$$

Since the sequence $g_n \uparrow \liminf_{j\to\infty} f_j$, by Theorem 3.4

$$\lim_{n\to\infty} \int_{\mathbb{X}} g_n d\mu = \int_{\mathbb{X}} f d\mu \le \liminf_{j\to\infty} \int_{\mathbb{X}} f_j(\omega) d\mu.$$

■

3.4 Integral of functions

Let $f : (\mathbb{X}, \mathcal{F}, \mu) \to (\bar{\mathbb{R}}, \bar{\mathcal{B}})$ be a measurable numerical function and $f^+ = \max\{f, 0\}$ and $f^- = -\min\{f, 0\}$.

Definition 3.4. *Let $f : (\mathbb{X}, \mathcal{F}, \mu) \to (\bar{\mathbb{R}}, \bar{\mathcal{B}})$ be a measurable numerical function such that $\int_{\mathbb{X}} f^+ d\mu < \infty$ and $\int_{\mathbb{X}} f^- d\mu < \infty$. In this case, we say that f is μ-integrable and we write*

$$\int_{\mathbb{X}} f d\mu := \int_{\mathbb{X}} f^+ d\mu - \int_{\mathbb{X}} f^- d\mu.$$

We note that $\int_{\mathbb{X}} f d\mu \in \mathbb{R}$ and denote by $\mathcal{L}_{\mathbb{R}}$ the set of integrable real functions and $\mathcal{L}_{\bar{\mathbb{R}}}$ the set of integrable numerical functions. A non-negative function f is said to be integrable if, and only if, $\int f d\mu < \infty$. If $(\mathbb{X}, \mathcal{F}, \mu) := (\mathbb{R}^n, \mathcal{B}^n, \lambda^n)$ we call $\int_{\mathbb{R}^n} f d\lambda^n$ the Lebesgue integral.

Theorem 3.7. *Let $f : (\mathbb{X}, \mathcal{F}, \mu) \to (\bar{\mathbb{R}}, \bar{\mathcal{B}})$ be a measurable function. Then,*

$$(1)\, f \in \mathcal{L}_{\bar{\mathbb{R}}} \iff (2)\, |f| \in \mathcal{L}_{\bar{\mathbb{R}}} \iff (3) \text{ there exist } 0 \le g \in \mathcal{L}_{\bar{\mathbb{R}}} \text{ such that } |f| \le g.$$

Proof. ((1) $\implies$ (2)) Since, $|f| = f^+ + f^-$ and since integrability of f implies $\int_{\mathbb{X}} f^+ d\mu < \infty$ and $\int_{\mathbb{X}} f^- d\mu < \infty$ we have $\int_{\mathbb{X}} |f| d\mu = \int_{\mathbb{X}} f^+ d\mu + \int_{\mathbb{X}} f^- d\mu < \infty$.

((2) $\implies$ (3)) Just take $g = |f|$.

((3) $\implies$ (1)) Since $f^+ \leq |f| \leq g$ and $f^- \leq |f| \leq g$ we have by the monotonicity of the integral of non-negative functions and the integrability of g that $f^+, f^- \in \mathcal{L}_{\bar{\mathbb{R}}}$. Hence, $f \in \mathcal{L}_{\bar{\mathbb{R}}}$. ■

Theorem 3.8. *Let $f, g : (\mathbb{X}, \mathcal{F}, \mu) \to (\bar{\mathbb{R}}, \bar{\mathcal{B}})$ be measurable functions such that $f, g \in \mathcal{L}_{\bar{\mathbb{R}}}$ and $a \in \mathbb{R}$. Then,*

1. *$af \in \mathcal{L}_{\bar{\mathbb{R}}}$ and $\int_{\mathbb{X}} af d\mu = a \int_{\mathbb{X}} f d\mu$,*
2. *$(f + g) \in \mathcal{L}_{\bar{\mathbb{R}}}$ and $\int_{\mathbb{X}} (f + g) d\mu = \int_{\mathbb{X}} f d\mu + \int_{\mathbb{X}} g d\mu$,*
3. *$\max\{f, g\}, \min\{f, g\} \in \mathcal{L}_{\bar{\mathbb{R}}}$,*
4. *if $f \leq g$ then $\int_{\mathbb{X}} f d\mu \leq \int_{\mathbb{X}} g d\mu$.*

Proof. Homework. Use Theorems 3.7 and 3.5. ■

Remark 3.5. *Note that*

$$\left| \int_{\mathbb{X}} f d\mu \right| \leq \left| \int_{\mathbb{X}} f^+ d\mu \right| + \left| \int_{\mathbb{X}} f^- d\mu \right| = \int_{\mathbb{X}} f^+ d\mu + \int_{\mathbb{X}} f^- d\mu = \int_{\mathbb{X}} (f^+ + f^-) d\mu = \int_{\mathbb{X}} |f| d\mu.$$

Theorem 3.9. *Let $f : (\mathbb{X}, \mathcal{F}, \mu) \to (\bar{\mathbb{R}}, \bar{\mathcal{B}})$ be a measurable function such that $0 \leq f \in \mathcal{L}_{\bar{\mathbb{R}}}$ and*

$$m(E) = \int_E f d\mu \text{ for all } E \in \mathcal{F}.$$

Then, m is a measure on $\mathcal{F}$.

Proof. Since $f \geq 0$, $m(E) \geq 0$. If $E = \emptyset$, then $f I_E = 0$ and

$$m(\emptyset) = \int_{\emptyset} f d\mu = \int_{\mathbb{X}} f I_{\emptyset} d\mu = \int_{\mathbb{X}} 0 d\mu = 0.$$

Now, let $\{E_j\}_{j\in\mathbb{N}}$ be a disjoint collection of sets in $\mathcal{F}$ such that $\cup_{j=1}^{\infty} E_j = E$ and let $f_n(x) = \sum_{j=1}^{n} f(x) I_{E_j}(x)$. By Theorem 3.5.3 $\int_{\mathbb{X}} f_n d\mu = \sum_{j=1}^{n} \int_{\mathbb{X}} f I_{E_j} d\mu$. Thus, $\int_{\mathbb{X}} f_n d\mu = \sum_{j=1}^{n} m(E_j)$.

Note that $f_1 \le f_2 \le \dots$ and converges to $f I_E$. Hence, by Theorem 3.4

$$m(E) = \int_{\mathbb{X}} f I_E d\mu = \lim_{n\to\infty} \int_{\mathbb{X}} f_n d\mu = \lim_{n\to\infty} \sum_{j=1}^{n} m(E_j) = \sum_{j=1}^{\infty} \lambda(E_j).$$

■

Remark 3.6. *m is called the measure with density function f with respect to μ and denoted by $m = f\mu$. If m has a density with respect to μ it is traditional in mathematics to write $dm/d\mu$ for the the density function. We note that with a little more work we can recognize f as the Radon-Nikodým derivative of m with respect to the measure μ.*

3.5 Lebesgue's convergence theorems

Theorem 3.10 (Lebesgue's Monotone Convergence Theorem)**.** *Let $f_n : (\mathbb{X}, \mathcal{F}, \mu) \to (\bar{\mathbb{R}}, \bar{\mathcal{B}})$ for $n \in \mathbb{N}$ be integrable functions such that $f_1 \le f_2 \le \cdots$ and $f := lim_{n\to\infty} f_n := \sup_{n\in\mathbb{N}} f_n$. Then,*

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

In this case,

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

Proof. Since $f_n \in \mathcal{L}(\mu)$ and $f_1 \le f_2 \le \cdots$ we have that $0 \le f_n - f_1 \in \mathcal{L}(\mu)$ forms an increasing sequence of nonnegative measurable functions. Hence, by Theorem 3.4

$$0 \le \sup_{n\in\mathbb{N}} \int_{\mathbb{X}} (f_n - f_1) d\mu = \int_{\mathbb{X}} \sup_{n\in\mathbb{N}} (f_n - f_1) d\mu. \tag{3.9}$$

Now, suppose $f \in \mathcal{L}(\mu)$ and note that from equation (3.9)

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

If $\sup_{n\in\mathbb{N}} \int_{\mathbb{X}} f_n d\mu < \infty$, then from equation (3.9) 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,

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

■

In the context of a measure space $(\mathbb{X}, \mathcal{F}, \mu)$, we will call N 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 $x \in N_{\mathcal{P}} \subseteq N$, where N is a null set, we say that the property is true almost everywhere (ae) or almost surely (as). Note the set $N_{\mathcal{P}}$ where the property does not hold need not be a measurable set.

Theorem 3.11. *(Markov's Inequality) Let $(\mathbb{X}, \mathcal{F}, \mu)$ be a measure space and $f \in \mathcal{L}_{\bar{\mathbb{R}}}(\mu)$. Then, for all $E \in \mathcal{F}$ and $a > 0$*

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

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

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

■

Note that if $E = \mathbb{X}$ we have $\mu(\{|f| \geq a\}) \leq \frac{1}{a}\int_{\mathbb{X}} |f| d\mu$. When $(\mathbb{X}, \mathcal{F}, \mu) = (\Omega, \mathcal{F}, P)$ a probability space and $f := X$ a random variable, the last result is commonly stated as

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

Theorem 3.12. *Let* $(\mathbb{X}, \mathcal{F}, \mu)$ *be a measure space and* $f \in \mathcal{L}_{\bar{\mathbb{R}}}(\mu)$. *Then,*

1. *if* N *is a null set* $\int_N f d\mu = 0$,

2. $\int_{\mathbb{X}} |f| d\mu = 0 \iff |f| = 0$ *ae.*

Proof. 1. Let $f_j = \min\{|f|, j\}$ and note $0 \leq f_1 \leq f_2 \leq \cdots$ with $\lim_{j\to\infty} f_j = |f|$. Hence, by Theorem 3.4

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

2. $(\Leftarrow)$ $\int_{\mathbb{X}} |f| d\mu = \int_{\{|f|=0\}} |f| d\mu + \int_{\{|f|\neq 0\}} |f| d\mu = \int_{\{|f|\neq 0\}} |f| d\mu = 0$ by item 1.

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

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

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

■

A direct consequence of this Theorem is that functions that are equal almost everywhere have the same integral.

Remark 3.7. *1. If* $f, g \geq 0$ *are measurable and* $f = g$ *ae 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 3.12.1, the first integral 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 $f \in \mathcal{L}_{\bar{\mathbb{R}}}$ *and* $f = g$ *ae then* $f^{+,-} = g^{+,-}$ *ae. Using the previous remark on* $f^{+,-}$ *we have* $\int_{\mathbb{X}} f^{+,-} d\mu = \int_{\mathbb{X}} g^{+,-} d\mu$. *Hence,*$\int_{\mathbb{X}} f d\mu = \int_{\mathbb{X}} g d\mu$.

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

$$f^{+,-} \leq |f| \leq g \ ae \ .$$

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

Theorem 3.13. *(Lebesgue 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* $|f_n| \leq g$ *for all* n *and some integrable nonnegative function* g. *If* $\lim_{n\to\infty} f_n(x) = f(x)$ *exists ae, then* f *is integrable and*

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

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

$$N = \{x : \lim_{n\to\infty} f_n \text{ does not exist}\}$$

is measurable and $\mu(N) = 0$. Thus, we proceed by taking $N = \emptyset$ as it does not contribute to the integrals.

For any $\epsilon > 0$ there exists $N_{(\epsilon,x)}$ such that for all $n > N_{(\epsilon,x)}$

$$\begin{aligned} |f| &= |f - f_n + f_n| \leq |f_n| + |f - f_n| \\ &\leq g + |f - f_n| \text{ by the bound assumption on the theorem statement} \\ &\leq g + \epsilon. \end{aligned}$$

Therefore, $\int f d\mu < \infty$ provided $g \in \mathcal{L}(\mu)$. Also, $|f_n| \leq g \iff -g \leq f_n \leq g$. Hence, $f_n + g \geq 0$. By Fatou's Lemma,

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

Therefore,

$$\int f d\mu \leq \liminf_{n\to\infty} \int f_n d\mu. \tag{3.10}$$

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

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

The first inequality together with the last equality imply that

$$\int f d\mu \geq \limsup_{n\to\infty} \int f_n d\mu. \tag{3.11}$$

Combining (3.10) and (3.11) completes the proof. ■

We now consider a measurable function that is indexed by a parameter $\theta \in (a, b)$ for $a < b$. As such, we define $f(x, \theta) : (\mathbb{X}, \mathcal{F}, \mu) \times (a, b) \to (\mathbb{R}, \mathcal{B})$ where f is measurable for all $\theta \in (a, b)$.

Theorem 3.14. *Let $f(x, \theta) : (\mathbb{X}, \mathcal{F}, \mu) \times (a, b) \to (\mathbb{R}, \mathcal{B})$ where f is measurable and $f \in \mathcal{L}(\mu)$ for all $\theta \in (a, b)$. Also, assume that $f(x, \theta)$ is continuous for every $x \in \mathbb{X}$ and $|f(x, \theta)| \leq g(x)$ for all $(x, t) \in \mathbb{X} \times (a, b)$ and some nonnegative integrable function g. Then, the function $h : (a, b) \to \mathbb{R}$ 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 $f(x,\theta)$. It suffices to show that for any sequence $\theta_n \in (a,b)$ such that $\theta_n \to \theta$ we have $h(\theta_n) \to h(\theta)$. By continuity of $f(x,\theta)$ for every x we have $f(x,\theta_n) \to f(x,\theta)$ and $|f(x,\theta_n)| \leq g(x)$ for all $x \in \mathbb{X}$. By Lebesgue's Dominated Convergence Theorem,

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

■

Theorem 3.15. *Let $f(x,\theta) : (\mathbb{X}, \mathcal{F}, \mu) \times (a,b) \to (\mathbb{R}, \mathcal{B})$ where f is measurable and $f \in \mathcal{L}(\mu)$ for all $\theta \in (a,b)$. Also, assume that $f(x,\theta)$ is differentiable on (a,b) for every $x \in \mathbb{X}$ and $|\frac{d}{d\theta} f(x,\theta)| \leq g(x)$ for all $(t,x) \in (a,b) \times \mathbb{X}$ and some nonnegative integrable function g. Then, the function $h : (a,b) \to \mathbb{R}$ 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. Recall that $\theta, \theta_n \in (a,b)$ with $\theta_n \to \theta$ and $\theta_n \neq \theta$.

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

for all $x \in \mathbb{X}$ and consequently $\frac{df}{d\theta}(x,\theta)$ is measurable. By the Mean Value Theorem, $f(x,\theta_n) - f(x,\theta) = \frac{df}{d\theta}(x,\theta_{z,n})(\theta_n - \theta)$ with $\theta_{z,n} = \lambda\theta_n + (1-\lambda)\theta$, $\lambda \in (0,1)$, $\theta_{z,n} \in (a,b)$. Consequently,

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

so that $\left| \frac{f(x,\theta_n) - f(x,\theta)}{\theta_n - \theta} \right|$ is integrable. Thus,

$$\frac{h(\theta_n) - h(\theta)}{\theta_n - \theta_0} = \int_{\mathbb{X}} \frac{f(x,\theta_n) - f(x,\theta)}{\theta_n - \theta} d\mu.$$

Hence, by the Lebesgue's Dominated Convergence Theorem

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

■

3.6 $\mathcal{L}^p$ spaces

Definition 3.5. *The collection of measurable functions $f:(\mathbb{X},\mathcal{F},\mu)\to(\mathbb{R},\mathcal{B})$ such that $\int_{\mathbb{X}}|f|^p d\mu<\infty$ for $p\in[1,\infty)$ is denoted by $\mathcal{L}^p(\mu)$ or $\mathcal{L}^p(\mathbb{X},\mathcal{F},\mu)$.*

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

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

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

If $f\in\mathcal{L}^p(\mathbb{X},\mathcal{F},\mu)$ we define the function $\|\cdot\|_p:\mathcal{L}^p(\mathbb{X},\mathcal{F},\mu)\to[0,\infty)$ as $\|f\|_p=\left(\int_{\mathbb{X}}|f|^p d\mu\right)^{1/p}$ and prove Hölder's Inequality.

Theorem 3.16. *(Hölder's Inequality) If $1<p<\infty$, $p^{-1}+q^{-1}=1$, $f\in\mathcal{L}^p$, $g\in\mathcal{L}^q$, then $fg\in\mathcal{L}$ and $\int|fg|d\mu\le\|f\|_p\|g\|_q$.*

Proof. If $\|f\|_p=0$ then, by Theorem 3.12 $|f|=0$ ae, so $|fg|=0$ ae. 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 $\int|xy|d\mu\le 1$.

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

$$a^\alpha b^{1-\alpha}\le\alpha a+(1-\alpha)b.$$

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 $\lambda \in (0,1)$

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

Also,

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

Thus, $u^\alpha \leq \alpha u + (1-\alpha)$ for $u > 0$.

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

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

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

■

Theorem 3.17. *(Minkowski-Riez 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. By the triangle inequality

$$\begin{aligned}\|f+g\|_p^p = \int |f+g||f+g|^{p-1}d\mu &\leq \int \left(|f||f+g|^{p-1} + |g||f+g|^{p-1}\right) d\mu\\ &= \int |f||f+g|^{p-1}d\mu + \int |g||f+g|^{p-1}d\mu, \text{ and if } p = 1 \text{ the proof is complete.}\\ &\text{If } p > 1, \text{ by Hölder's Inequality}\\ &\leq \|f\|_p\||f+g|^{p-1}\|_q + \|g\|_p\||f+g|^{p-1}\|_q,\end{aligned}$$

where $1/p + 1/q = 1$ which implies $1/q = 1 - 1/p \implies q = \frac{p}{p-1}$. Thus,

$$\|f+g\|_p^p \leq \|f\|_p \||f+g|^{p/q}\|_q + \|g\|_p \||f+g|^{p/q}\|_q = (\|f\|_p + \|g\|_p)\||f+g|^{p/q}\|_q. \qquad (3.12)$$

Now,

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

Using this in inequality (3.12) we obtain $\|f+g\|_p^{p-(p-1)} = \|f+g\|_p \leq \|f\|_p + \|g\|_p$. ■

Remark 3.8. *1. The Minkowski-Riez Inequality and the fact that $a \in \mathbb{R}$, $\|af\|_p = |a|\|f\|_p$ 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 $f(x) = 0$ almost everywhere.*

2. $f, g \in \mathcal{L}^p(\mathbb{X}, \mathcal{F}, \mu)$ are taken to be equivalent if they differ at most on a set of μ-measure zero (null set), i.e., $f \sim g$ if $\{x : f(x) \neq g(x)\}$ is a null set. Then, for every $f \in \mathcal{L}^p(\mathbb{X}, \mathcal{F}, \mu)$ we can define an equivalence class (reflexive, symmetric and transitive) of $\mathcal{L}^p$ functions induced by f, which will be denoted by $[f]_p$. The space of all equivalence classes $[f]_p$ of functions $f \in \mathcal{L}^p$ is denoted by L^p with norm $\|[f]_p\|_p := \inf\{\|g\|_p : g \in \mathcal{L}^p$ and $g \sim f\}$. $(L^p, \|f_{[p]}\|_p)$ 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, treated in the next theorem, has $p = 2$ and $X, Y : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B})$ being random variables such that $X, Y \in \mathcal{L}^2_{\mathbb{R}}(\Omega, \mathcal{F}, P)$.

Theorem 3.18. *Let $X, Y : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B})$ 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 $|\int_\Omega XY dP| \leq \left(\int_\Omega X^2 dP\right)^{1/2} \left(\int_\Omega Y^2 dP\right)^{1/2}$, *2.* *If* $X \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$ *then* $X \in \mathcal{L}(\Omega, \mathcal{F}, P)$ *and* $\left(\int_\Omega X dP\right)^2 \leq \int_\Omega X^2 dP$*,*

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 $p = q = 2$. 3. follows from the comments after Definition 3.5. 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,

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

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

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

■

Remark 3.9. *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 3.19. *Let* X *be a random variable defined on the probability space* $(\Omega, \mathcal{F}, P)$ *taking values in* $(\mathbb{R}, \mathcal{B})$ *and* $h : (\mathbb{R}, \mathcal{B}) \to (\mathbb{R}, \mathcal{B})$ *be measurable.*

1. $f := h \circ X$ *is integrable in* $(\Omega, \mathcal{F}, P)$ *if, and only if,* h *is integrable in* $(\mathbb{R}, \mathcal{B}, P_X)$*.*

2. $E(h(X)) := \int_\Omega f dP = \int_\mathbb{R} h dP_X$*.*

Proof. First, let h be a non-negative simple function. Then we have that $f(\omega) = \sum_{j=1}^m y_j I_{A_j}(\omega)$ where $A_j \in \mathcal{F}$. Consequently,

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

Second, let $h \geq 0$. Then, by Theorem 3.3 there exists a sequence of increasing non-negative simple function ϕ_n such that $\phi_n \to h$ as $n \to \infty$. Hence, if we define $f_n(\omega) = \phi_n(X(\omega)) = (\phi_n \circ X)(\omega)$, it is a sequence of increasing simple function that converges to f.

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

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

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

But from Remark 3.5, 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 $\mathbb{R}$.

Definition 3.6. *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 non-negative Borel measurable function f_X that satisfies*

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

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

Theorem 3.20. *f_X is a density $\iff$ $\int_{\mathbb{R}} f_X d\lambda = 1$, f_X is unique almost everywhere.*

Proof. ($\implies$) f_X a density implies $F_X(a) = P_X((-\infty, a]) = \int_{(-\infty,a]} f_X d\lambda$. $\lim_{a\to\infty} P_X((-\infty, a]) = 1 = \lim_{a\to\infty} \int_{(-\infty,a]} f_X d\lambda$, where the first equality follows from definition 2.5 and continuity of probability measures.

($\impliedby$) Suppose f_X is a non-negative Borel measurable function such that $\int f_X d\lambda = 1$. For all $A \in \mathcal{B}$, we put

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

By Theorem 3.9, P_X is a measure on $\mathcal{B}$ with $P_X(\mathbb{R}) = 1$, by assumption. Taking $A = (-\infty, a]$,

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

and f_X is a density for F_X.

Now, suppose g_X is another density for F_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 $A_1 \subseteq A_2 \subseteq \ldots$. $\lim_{n\to\infty} A_n = \cup_{n=1}^{\infty} A_n = A = \{x : g_X(x) > f_X(x)\}$ and $\lambda(A) = \lim_{n\to\infty} \lambda(A_n) = 0$. Similarly, we have $\lambda(B) = 0$ for $B = \{x : g_X(x) < f_X(x)\}$. So, $\lambda(\{x : g_X = f_X\}) = 1$. ■

Theorem 3.21. *Let $X : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B})$ be a random variable with density f_X and $h : (\mathbb{R}, \mathcal{B}) \to (\mathbb{R}, \mathcal{B})$ 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. Homework. ■