Exercises and solutions for
Fundamental Elements of Probability and Asymptotic Theory

by

Carlos Brunet Martins-Filho
Department of Economics
256 UCB
University of Colorado at Boulder
Boulder, CO 80309-0256 USA
email: carlos.martins@colorado.edu

Chapter 1

Exercises

1. Let $f : \mathbb{N} \times \mathbb{N} \to \mathbb{R}$ be a double sequence with typical value given by $f(m,n)$. Assume that

   (a) for every $n \in \mathbb{N}$, $f(m_1, n) \le f(m_2, n)$ whenever $m_1 \le m_2$,

   (b) for every $m \in \mathbb{N}$, $f(m, n_1) \le f(m, n_2)$ whenever $n_1 \le n_2$.

   Show that $\lim\limits_{n\to\infty}\left(\lim\limits_{m\to\infty} f(m,n)\right) = \lim\limits_{m\to\infty}\left(\lim\limits_{n\to\infty} f(m,n)\right) = \lim\limits_{n\to\infty} f(n,n)$.

   As a corollary, show that if $f(m,n) \ge 0$ then $\sum\limits_{n\in\mathbb{N}}\sum\limits_{m\in\mathbb{N}} f(m,n) = \sum\limits_{m\in\mathbb{N}}\sum\limits_{n\in\mathbb{N}} f(m,n)$.

   Answer: From conditions (a) and (b), $f(1,1) \le f(1,2) \le f(2,2) \le f(2,3) \le f(3,3) \le \cdots$ Hence, $f(m,m) \le f(n,n)$ whenever $m \le n$. The sequence $\{f(n,n)\}_{n\in\mathbb{N}}$ is monotonically increasing, hence it has a limit, which is either finite, if the sequence is bounded above, or infinity, if it is not. Let this limit be denoted by F. By the same reasoning, there exist limits $F_m = \lim\limits_{n\to\infty} f(m,n)$ for each $m \in \mathbb{N}$. Since $f(m,n) \le f(n,n)$, we have that $F_m \le F$ when $m \le n$. Note that $F_{m_1} \le F_{m_2}$ whenever $m_1 \le m_2$, hence $\lim\limits_{m\to\infty} F_m = F'$ exists, and $F' \le F$.

   To complete the proof, we need to show that $F' = F$. If F is finite, for every $\epsilon > 0$ there exists $N(\epsilon)$ such that for all $n \ge N(\epsilon)$, $F - \epsilon \le f(n,n) \le F$. Put $m := N(\epsilon)$, and note that

   $$F_m = \lim_{n\to\infty} f(m,n) \ge f(m,m) := f(N(\epsilon), N(\epsilon)) \ge F - \epsilon.$$

   Hence, $\lim\limits_{n\to\infty} F_m = F \ge F - \epsilon$, which implies that $F \le F'$. Combining the last inequality with $F' \le F$ from the previous paragraph gives $F = F'$. If F is infinite, for any $C > 0$

there exists $N(C)$ such that if $n \geq N(C)$, $f(n,n) \geq C$. If $m = N(C) \leq n$ then $f(m,m) \leq f(m,n)$ and

$$C \leq f(m,m) \leq \lim_{n\to\infty} f(m,n) = F_m,$$

hence it follows that F' must be infinite.

The proof that $\lim_{n\to\infty}\left(\lim_{m\to\infty} f(m,n)\right) = \lim_{n\to\infty} f(n,n)$ follows in exactly the same way by interchanging the indexes m and n due to the symmetry of the equation.

Corollary. Let $g(p,q) = \sum_{m=1}^{p}\sum_{n=1}^{q} f(m,n)$ for $p,q \in \mathbb{N}$. Since, $f(m,n) \geq 0$, $g(p,q)$ satisfies conditions (a) and (b), establishing the result.

2. Let $\mathbb{X}$ be an arbitrary set and consider the collection of all subsets of $\mathbb{X}$ that are countable or have countable complements. Show that this collection is a σ-algebra. Use this fact to obtain the σ-algebra generated by $\mathcal{C} = \{\{x\} : x \in \mathbb{R}\}$.

   Answer: Let $\mathcal{F} = \{A \subseteq \mathbb{X} : \#A \leq \#\mathbb{N} \text{ or } \#A^c \leq \#\mathbb{N}\}$, where $\#$ indicates cardinality. First, note that $\mathbb{X} \in \mathcal{F}$ since $\mathbb{X}^c = \emptyset$, which is countable. Second, if $A \in \mathcal{F}$ then either $A = (A^c)^c$ or A^c are countable. That is, $A^c \in \mathcal{F}$. Third, if $A_n \in \mathcal{F}$ for $n \in \mathbb{N}$ we have two possible cases - A_n are all countable, or at least one of these sets is uncountable, say A_{n_0}. For the first case, $\underset{n\in\mathbb{N}}{\cup} A_n$ is the countable union of countable sets, hence it is countable and consequently in $\mathcal{F}$. For the second case, since A_{n_0} is uncountable and in $\mathcal{F}$, it must be that $A_{n_0}^c$ is countable. Also,

$$\left(\underset{n\in\mathbb{N}}{\cup} A_n\right)^c = \underset{n\in\mathbb{N}}{\cap} A_n^c \subset A_{n_0}^c.$$

   Since subsets of countable sets are countable, $\left(\underset{n\in\mathbb{N}}{\cup} A_n\right)^c$ is countable, and consequently $\underset{n\in\mathbb{N}}{\cup} A_n \in \mathcal{F}$.

   Now, let $\mathcal{F}$ be the σ-algebra defined above. Since $\mathcal{C} \subseteq \mathcal{F}$, $\sigma(\mathcal{C}) \subseteq \mathcal{F}$. Also, if $A \in \mathcal{F}$ either A or A^c is countable. Without loss of generality, suppose A is countable. Then, $A = \underset{x\in C}{\cup}\{x\}$ where C is a countable collection of real numbers. Hence, $A \in \sigma(\mathcal{C})$. Hence, $\mathcal{F} \subseteq \sigma(\mathcal{C})$. Combining the two set containments we have $\sigma(\mathcal{C}) = \mathcal{F}$.

3. Denote by $B(x,r)$ an open ball in $\mathbb{R}^n$ centered at x and with radius r. Show that the Borel sets are generated by the collection $B = \{B_r(x) : x \in \mathbb{R}^n, r > 0\}$.

   Answer: Let $B' = \{B_r(x) : x \in \mathbb{Q}^n, r \in \mathbb{Q}^+\}$. Then, $B' \subset B \subset \mathcal{O}_{\mathbb{R}^n}$ and $\sigma(B') \subset \sigma(B) \subset \sigma(\mathcal{O}_{\mathbb{R}^n})$.

Now, let $S = \bigcup_{B\in B',\, B\subset O} B$. By construction $x \in S \implies x \in O$. Now, suppose $x \in O$. Then, since O is open, there exists $B(x,\epsilon)$ such that $B(x,\epsilon) \subset O$ where ϵ is a rational number. Since $\mathbb{Q}^n$ is a dense subset of $\mathbb{R}^n$, we can find $q \in \mathbb{Q}^n$ such that $\|x-q\| \le \epsilon/2$. Consequently,

$$B(q,\epsilon/2) \subset B(x,\epsilon) \subset O.$$

Hence, $O \subset S$. Thus, every open O can be written as $O = \bigcup_{B\in B',\, B\subset O} B$. Since B' is a collection of balls with rational radius and rational centers, B' is countable. Thus,

$$\mathcal{O}_{\mathbb{R}^n} \subset \sigma(B') \implies \sigma(\mathcal{O}_{\mathbb{R}^n}) \subset \sigma(B').$$

Combining this set containment with $\sigma(B') \subset \sigma(B) \subset \sigma(\mathcal{O}_{\mathbb{R}^n})$ completes the proof.

4. Let $(\Omega,\mathcal{F})$ be a measurable space. Show that: a) if μ_1 and μ_2 are measures on $(\Omega,\mathcal{F})$, then $\mu_c(F) := c_1\mu_1(F) + c_2\mu_2(F)$ for $F \in \mathcal{F}$ and all $c_1, c_2 \ge 0$ is a measure; b) if $\{\mu_i\}_{i\in\mathbb{N}}$ are measures on $(\Omega,\mathcal{F})$ and $\{\alpha_i\}_{i\in\mathbb{N}}$ is a sequence of positive numbers, then $\mu_\infty(F) = \sum_{i\in\mathbb{N}} \alpha_i\mu_i(F)$ for $F \in \mathcal{F}$ is a measure.

   Answer: a) First, note that $\mu_c : \mathcal{F} \to [0,\infty]$ since $c_1, c_2, \mu_1(F), \mu_2(F) \ge 0$ for all $F \in \mathcal{F}$. Second, $\mu_c(\emptyset) = c_1\mu_1(\emptyset) + c_2\mu_2(\emptyset) = 0$ since μ_1 and μ_2 are measures. Third, if $\{F_i\}_{i\in\mathbb{N}} \in \mathcal{F}$ is a pairwise disjoint collection of sets,

$$\begin{aligned}
\mu_c\left(\cup_{i\in\mathbb{N}} F_i\right) &= c_1\mu_1(\cup_{i\in\mathbb{N}} F_i) + c_2\mu_2(\cup_{i\in\mathbb{N}} F_i) \\
&= c_1 \sum_{i\in\mathbb{N}} \mu_1(F_i) + c_2 \sum_{i\in\mathbb{N}} \mu_2(F_i), \text{ since } \mu_1 \text{ and } \mu_2 \text{ are measures} \\
&= \sum_{i\in\mathbb{N}} (c_1\mu_1(F_i) + c_2\mu_2(F_i)) = \sum_{i\in\mathbb{N}} \mu_c(F_i).
\end{aligned}$$

   b) The verification that $\mu_\infty : \mathcal{F} \to [0,\infty]$ and $\mu_\infty(\emptyset) = 0$ follows the same arguments as in item a) when examining μ_c. For σ-additivity, note that if $\{F_j\}_{j\in\mathbb{N}} \in \mathcal{F}$ is a pairwise disjoint collection of sets,

$$\mu_\infty\left(\cup_{j\in\mathbb{N}} F_j\right) = \sum_{i=1}^{\infty} \alpha_i\mu_i\left(\cup_{j\in\mathbb{N}} F_j\right) = \sum_{i=1}^{\infty} \alpha_i \sum_{j=1}^{\infty} \mu_i\left(F_j\right) = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \alpha_i\mu_i\left(F_j\right).$$

   If we are able to interchange the sums in the last term, then we can write

$$\mu_\infty\left(\cup_{j\in\mathbb{N}} F_j\right) = \sum_{j=1}^{\infty}\sum_{i=1}^{\infty} \alpha_i\mu_i\left(F_j\right) = \sum_{j=1}^{\infty} \mu_\infty\left(F_j\right),$$

completing the proof. Now, note that

$$\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\alpha_i\mu_i\left(F_j\right)=\lim_{n\to\infty}\lim_{m\to\infty}\sum_{i=1}^{n}\sum_{j=1}^{m}\alpha_i\mu_i\left(F_j\right)=\sup_{n\in\mathbb{N}}\sup_{m\in\mathbb{N}}\sum_{i=1}^{n}\sum_{j=1}^{m}\alpha_i\mu_i\left(F_j\right)=\sup_{n\in\mathbb{N}}\sup_{m\in\mathbb{N}}S_{nm}$$

since the partial sums are increasing. Now, if $S_{nm}\in\mathbb{R}$, then

$$\sup_{n\in\mathbb{N}}\sup_{m\in\mathbb{N}}S_{nm}=\sup_{m\in\mathbb{N}}\sup_{n\in\mathbb{N}}S_{nm}.$$

Hence, to finish the proof, we require $\mu_i(F_j)<\infty$.

5. Let $(\Omega,\mathcal{F},\mu)$ be a measure space and $\mathcal{G}\subset\mathcal{F}$ be a σ-algebra. In this case, we call $\mathcal{G}$ a sub-σ-algebra of $\mathcal{F}$. Let $\nu:=\mu|_{\mathcal{G}}$ be the restriction of μ to $\mathcal{G}$. That is, $\nu(G)=\mu(G)$ for all $G\in\mathcal{G}$. Is ν a measure? If μ is finite, is ν finite? If μ is a probability, is ν a probability?

   Answer: Since $\emptyset\in\mathcal{G}\subset\mathcal{F}$, $\nu(\emptyset)=\mu(\emptyset)=0$. If $\{A_i\}_{i\in\mathbb{N}}\in\mathcal{G}$ is a pairwise disjoint sequence, we have that $\{A_i\}_{i\in\mathbb{N}}\in\mathcal{F}$. Hence, $\nu(\cup_{i\in\mathbb{N}}A_i)=\mu(\cup_{i\in\mathbb{N}}A_i)=\sum_{i\in\mathbb{N}}\mu(A_i)=\sum_{i\in\mathbb{N}}\nu(A_i)$. Now, μ finite means that $\mu(\Omega)<\infty$. Since $\Omega\in\mathcal{G}$, $\nu(\Omega)=\mu(\Omega)<\infty$. The same holds for $\mu(\Omega)=1$.

6. Show that a measure space $(\Omega,\mathcal{F},\mu)$ is σ-finite if, and only if, there exists $\{F_n\}_{n\in\mathbb{N}}\in\mathcal{F}$ such that $\cup_{n\in\mathbb{N}}F_n=\Omega$ and $\mu(F_n)<\infty$ for all n.

   Answer: $(\Rightarrow)$ By definition, $(\Omega,\mathcal{F},\mu)$ is σ-finite if there exists and increasing sequence $A_1\subset A_2\subset A_3\cdots$ such that $\cup_{n\in\mathbb{N}}A_n=\Omega$ with $\mu(A_n)<\infty$ for all n. Hence, it suffices to let $F_n=A_n$.

   $(\Leftarrow)$ Let $A_n=\cup_{j=1}^{n}F_j$. Then, $A_1\subset A_2\subset\cdots$ and $\cup_{n\in\mathbb{N}}A_n=\cup_{j\in\mathbb{N}}F_j=\Omega$. Also, $\mu(A_n)=\mu(\cup_{j=1}^{n}F_j)\leq\sum_{j=1}^{n}\mu(F_j)<\infty$ since the sum is finite and $\mu(F_j)<\infty$.

7. Let $(\Omega,\mathcal{F},P)$ be a probability space and $\{E_n\}_{n\in\mathbb{N}}\subset\mathcal{F}$. Show that if $\sum_{n=1}^{\infty}P(E_n)<\infty$ then $P\left(\operatorname*{limsup}_{n\to\infty}E_n\right)=0$.

   Answer:

$$\begin{aligned}P\left(\operatorname*{limsup}_{n\to\infty}E_n\right)&=P\left(\lim_{n\to\infty}\cup_{j\geq n}E_j\right)\\&=\lim_{n\to\infty}P\left(\cup_{j\geq n}E_j\right)\text{ by continuity}\\&\leq\operatorname*{limsup}_{n\to\infty}\sum_{j=n}^{\infty}P(E_j)\text{ by subadditivity and definition of limsup.}\end{aligned}$$

Since $\sum_{n=1}^{\infty} P(E_n) < \infty$ it must be that $\sum_{j=n}^{\infty} P(E_j) \to 0$ as $n \to 0$. Consequently, $P\left(\limsup_{n\to\infty} E_n\right) = 0$.

8. Let $\{E_j\}_{j\in J}$ be a collection of pairwise disjoint events. Show that if $P(E_j) > 0$ for each $j \in J$, then J is countable.

   Answer: Let $C_n = \{E_j : P(E_j) > \frac{1}{n} \text{ and } j \in J\}$. By assumption the elements of C_n are disjoint events and

$$P\left(\cup_{j_m} E_{j_m}\right) = \sum_{m=1}^{\infty} P(E_{j_m}) = \infty,$$

   where the last equality follows from the fact that $P(E_{j_m}) > 0$. So, it must be that C_n has finitely many elements. Also, $\{E_j\}_{j\in J} = \cup_{n=1}^{\infty} C_n$, which is countable since it is a countable union of finite sets.

9. Consider the extended real line, i.e., $\bar{\mathbb{R}} := \mathbb{R} \cup \{-\infty\} \cup \{\infty\}$. Let $\bar{\mathcal{B}} := \mathcal{B}(\bar{\mathbb{R}})$ be defined as the collection of sets $\bar{B}$ such that $\bar{B} = B \cup S$ where $B \in \mathcal{B}(\mathbb{R})$ and $S \in \{\emptyset, \{-\infty\}, \{\infty\}, \{-\infty, \infty\}\}$. Show that $\bar{\mathcal{B}}$ is a σ-algebra and that it is generated by a collection of sets of the form $[a, \infty]$ where $a \in \mathbb{R}$.

   Answer: Let's first show that $\bar{\mathcal{B}}$ is a σ-algebra. Since $\bar{B} = B \cup S$ with $B \in \mathcal{B}(\mathbb{R})$, we can choose $B = \mathbb{R}$ and use $S = \{-\infty, \infty\}$ to conclude that $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\} \in \bar{\mathcal{B}}$. Next, note that if $\bar{B} = B \cup S$ we have that $\bar{B}^c = B^c \cap S^c$. But the complement of a set S is an element of $\{\bar{\mathbb{R}}, \mathbb{R}\cup\{\infty\}, \mathbb{R}\cup\{-\infty\}, \mathbb{R}\}$. Hence, either 1) $\bar{B}^c = B^c \cap \bar{\mathbb{R}} = B^c \cup \emptyset \in \bar{\mathcal{B}}$ or, 2) $\bar{B}^c = B^c \cap (\mathbb{R} \cup \{\infty\}) = (B^c \cap \mathbb{R}) \cup \{\infty\}$ where $B^c \cap \mathbb{R} \in \mathcal{B}$ and consequently $\bar{B}^c \in \bar{\mathcal{B}}$ or, 3) $\bar{B}^c = B^c \cap (\mathbb{R} \cup \{-\infty\}) = (B^c \cap \mathbb{R}) \cup \{-\infty\}$ where $B^c \cap \mathbb{R} \in \mathcal{B}$ and consequently $\bar{B}^c \in \bar{\mathcal{B}}$ or, 4) $\bar{B}^c = B^c \cap \mathbb{R} \in \bar{\mathcal{B}}$.

   Lastly, letting $A_i = B_i \cup S$ for $B_i \in \mathcal{B}$ we have that $\cup_{i\in\mathbb{N}} A_i = \cup_{i\in\mathbb{N}}(B_i \cup S) = (\cup_{i\in\mathbb{N}} B_i) \cup S$. Since $\cup_{i\in\mathbb{N}} B_i \in \mathcal{B}$ we have that $\cup_{i\in\mathbb{N}} A_i \in \bar{\mathcal{B}}$.

   If $\bar{\mathcal{B}}$ is a σ-algebra and $\mathcal{C} = \{[a, \infty] : a \in \mathbb{R}\}$, we need to show that $\sigma(\mathcal{C}) = \bar{\mathcal{B}}$.

   First, note that $[a, \infty] = [a, \infty) \cup \{\infty\}$ and we know that $[a, \infty) \in \mathcal{B}$. Thus, $[a, \infty] \in \bar{\mathcal{B}}$ for all $a \in \mathbb{R}$. Then, $\sigma(\mathcal{C}) \subseteq \bar{\mathcal{B}}$.

   Second, observe that for $-\infty < a \leq b < \infty$ we have $[a, b) = [a, \infty] - [b, \infty] = [a, \infty] \cap [b, \infty]^c \in \sigma(\mathcal{C})$ since $\sigma(\mathcal{C})$ contains $[a, \infty]$ and $[b, \infty]^c$ by virtue of being a

σ-algebra. Hence,

$$\mathcal{B} \subseteq \sigma(\mathcal{C}) \subseteq \bar{\mathcal{B}}.$$

Now,

$$\{\infty\} = \cap_{i\in\mathbb{N}}[i,\infty],\ \{-\infty\} = \cap_{i\in\mathbb{N}}[-\infty,-i) = \cap_{i\in\mathbb{N}}[-i,\infty]^c$$

which allows us to conclude that $\{\infty\},\{-\infty\} \in \sigma(\mathcal{C})$. Hence, if $B \in \mathcal{B}$ all sets of the form

$$B, B\cup\{\infty\}, B\cup\{-\infty\}, B\cup\{\infty\}\cup\{-\infty\}$$

are in $\sigma(\mathcal{C})$. Hence, $\bar{\mathcal{B}} \subseteq \sigma(\mathcal{C})$. Combining this set. containment with $\sigma(\mathcal{C}) \subseteq \bar{\mathcal{B}}$ gives the result.

10. If $E_1, E_2, \cdots, E_n$ are independent events, show that the probability that none of them occur is less than or equal to $\exp\left(-\sum_{i=1}^n P(E_i)\right)$.

    Answer: Let $f(x) = \exp(-x)$ and note that for $\lambda \in (0,1)$, by Taylor's Theorem

    $$\exp(-x) = f(x) = f(0) + f^{(1)}(0)x + \frac{1}{2}f^{(2)}(\lambda x)x^2 = 1 - x + \frac{1}{2}\exp(-\lambda x)x^2$$

    Consequently, $1 - x \le \exp(-x)$. Now, we are interested in the event $E = (\cup_{i=1}^n E_i)^c = \cap_{i=1}^n E_i^c$. But since the $E_1, E_2, \cdots, E_n$ are independent, so is the collection $E_1^c, E_2^c, \cdots, E_n^c$. Hence, $P(E) = \prod_{i=1}^n P(E_i^c) = \prod_{i=1}^n (1-P(E_i)) \le \prod_{i=1}^n \exp(-P(E_i)) = \exp\left(-\sum_{i=1}^n P(E_i)\right)$.

11. Let $\{A_n\}_{n\in\mathbb{N}}$ and $\{B_n\}_{n\in\mathbb{N}}$ be events (measurable sets) in a probability space with measure P with $\lim A_n = A$, $\lim B_n = B$, $P(B_n), P(B) > 0$ for all n. Show that $P(A_n|B) \to P(A|B)$, $P(A|B_n) \to P(A|B)$, $P(A_n|B_n) \to P(A|B)$ as $n \to \infty$.

    Answer: Since $P(\cdot|B)$ is a probability measure (proved in the class notes), we have by continuity of probability measures that $P(A_n|B) \to P(A|B)$ if $\lim B_n = B$.

    Now, since $\lim B_n = B$ we have that $A \cap B_n \to A \cap B$. To see this, note that if $A \cap B_n := C_n$ then $D_j = \cup_{n=j}^\infty C_n = A \cap (\cup_{n=1}^\infty B_n)$. Then, $\limsup C_n = \cap_{j=1}^\infty D_j = \cap_{j=1}^\infty (A \cap \cup_{n=1}^\infty B_n) = A \cap B$. Defining $\liminf$ for C_n we can in similar fashion that $\liminf C_n = A \cap B$. Hence, by continuity of probability measures $P(A\cap B_n) \to P(A\cap B)$ and $P(B_n) \to P(B)$. Consequently,

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

    Lastly, since $A_n \cap B_n \to A \cup B$, using the same arguments

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

12. Let $(\mathbb{X}, \bar{\mathcal{F}}, \bar{\mu})$ be the measure space defined in Theorem 1.15 and $\mathcal{C} = \{G \in \mathbb{X} : \exists\, A,\, B \in \mathcal{F} \ni A \subset G \subset B \text{ and } \mu(B - A) = 0\}$. Show that $\bar{\mathcal{F}} = \mathcal{C}$.

    Answer: $G \in \bar{\mathcal{F}} \implies G = A \cup M$ where $A \in \mathcal{F}$ and $M \in \mathcal{S}$. $M \in \mathcal{S} \implies \exists N \in \mathcal{N}_\mu \ni M \subset N$. Then,

    $$A \subset G = A \cup M \subset A \cup N := B \in \mathcal{F}.$$

    Now, $\mu(B - A) = \mu(B \cup A^c) = \mu((A \cup N) - A) \leq \mu(N) = 0$. Thus, $G \in \mathcal{C}$.

    $G \in \mathcal{C} \implies \exists A,\, B \in \mathcal{F} \ni A \subset G \subset B$ and $\mu(B - A) = 0$. Since $A \subset G \subset B$ we have that $G - A \subset B - A$, and since $B - A$ is a μ-null set $G - A \in \mathcal{S}$. Now, $G = A \cup (G - A)$, and since $A \in \mathcal{F}$, $G \in \bar{\mathcal{F}}$.

Chapter 2

Exercises

1. Let μ be a measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that $\mu([-n, n)) < \infty$ for all $n \in \mathbb{N}$. Define,

$$F_\mu(x) := \begin{cases} \mu([0, x)) & \text{if } x > 0, \\ 0 & \text{if } x = 0, \\ -\mu([x, 0)) & \text{if } x < 0. \end{cases}$$

Show that $F_\mu : \mathbb{R} \to \mathbb{R}$ is monotonically increasing and left continuous.

Answer: Given that $\mu([-n, n)) < \infty$, F_μ takes values in $\mathbb{R}$. First, we show that all $x < x'$, $F_\mu(x) \leq F_\mu(x')$. There are three cases to be considered

(a) $(0 \leq x < x')$: if $0 < x < x'$, $F_\mu(x') - F_\mu(x) = \mu([0, x')) - \mu([0, x))$. Since $[0, x') = [0, x) \cup [x, x')$, σ-additivity of μ gives $\mu([0, x')) = \mu([0, x)) + \mu([x, x'))$ or $\mu([x, x')) = \mu([0, x')) - \mu([0, x)) = F_\mu(x') - F_\mu(x) \geq 0$. If $x = 0$, $F_\mu(x') - F_\mu(0) = \mu([0, x')) \geq 0$.

(b) $(x < 0 \leq x')$: If $x' > 0$, $F_\mu(x') - F_\mu(x) = \mu([0, x')) + \mu([x, 0)) \geq 0$. If $x' = 0$, $F_\mu(0) - F_\mu(x) = \mu([x, 0)) \geq 0$.

(c) $(x < x' < 0)$: $F_\mu(x') - F_\mu(x) = -\mu([x', 0)) + \mu([x, 0))$. Since $[x, 0) = [x, x') \cup [x', 0)$, σ-additivity of μ gives $\mu([x, 0)) = \mu([x, x')) + \mu([x', 0))$ or $\mu([x, 0)) - \mu([x', 0)) = F_\mu(x') - F_\mu(x) = \mu([x, x')) \geq 0$.

Second, we must show that $\lim_{n\to\infty} F_\mu(x - h_n) = F_\mu(x)$ for all $x \in \mathbb{R}$. Let $n \in \mathbb{N}$, $h_1 \geq h_2 \geq h_3 \geq \cdots$ with $h_n \downarrow 0$ as $n \to \infty$, and $h_1 > 0$. There are three cases to consider.

(a) $(x > 0)$: Choose $h_1 \in (0, x)$ and define $A_n = [0, x - h_n)$. Then, $A_1 \subset A_2 \subset \cdots$ and $\lim_{n\to\infty} A_n = \bigcup_{n\in\mathbb{N}} A_n = [0, x)$. By continuity of measure from below,

$$\lim_{n\to\infty} F_\mu(x - h_n) = \lim_{n\to\infty} \mu([0, x - h_n)) = \mu([0, x)) = F_\mu(x).$$

(b) $(x = 0)$: Define $A_n = [-h_n, 0)$. Then, $A_1 \supset A_2 \supset \cdots$ and $\lim_{n\to\infty} A_n = \bigcap_{n\in\mathbb{N}} A_n = \emptyset$. By continuity of measures from above, and given that $\mu([-h_1, 0)) < \infty$,

$$\lim_{n\to\infty} F_\mu(-h_n) = \lim_{n\to\infty} \mu([-h_n, 0)) = \mu(\emptyset) = 0 = F_\mu(0).$$

(c) $(x < 0)$: Define $A_n = [x - h_n, 0)$. Then, $A_1 \supset A_2 \supset \cdots$ and $\lim_{n\to\infty} A_n = \cap_{n=1}^{\infty} A_n = [x, 0)$. By continuity of measures from above and given that $\mu([x - h_1, 0)) < \infty$,

$$\lim_{n\to\infty} F_\mu(x - h_n) = \lim_{n\to\infty} -\mu([x - h_n, 0)) = -\mu([x, 0)) = F_\mu(x).$$

2. Let F_μ be defined as in question 1 and let $\nu_{F_\mu}(([a, b)) = F_\mu(b) - F_\mu(a)$ for all $a \leq b$, $a, b \in \mathbb{R}$. Show that ν_{F_μ} extends uniquely to a measure on $\mathcal{B}(\mathbb{R})$ and $\nu_{F_\mu} = \mu$.

Answer: Recall that $\mathcal{S} = \{[a, b) : a \leq b,\ a, b \in \mathbb{R}\}$ is a semi-ring (if $a = b$, $[a, a) = \emptyset$). Given F_μ, we define $\nu_{F_\mu} : \mathcal{S} \to [0, \infty)$ as $\nu_{F_\mu}([a, b)) = F_\mu(b) - F_\mu(a)$ for all $a \leq b$. Since F_μ is monotonically increasing, $F_\mu(b) - F_\mu(a) \geq 0$ and $\nu_{F_\mu}([a, a) = \emptyset) = F_\mu(a) - F_\mu(a) = 0$. Also, ν_{F_μ} is finitely additive since for $a < c < b$, we have that $[a, b) = [a, c) \cup [c, b)$ and $\nu_{F_\mu}([a, b)) = F_\mu(b) - F_\mu(a) = F_\mu(c) - F_\mu(a) + F_\mu(b) - F_\mu(c) = \nu_{F_\mu}([a, c)) + \nu_{F_\mu}([c, b))$. We now show that ν_{F_μ} is σ-additive, i.e., for $[a_n, b_n)$, $n \in \mathbb{N}$ a disjoint collection such that $[a, b) = \cup_{n\in\mathbb{N}} [a_n, b_n)$, we have $\nu_{F_\mu}([a, b)) = \sum_{n\in\mathbb{N}} \nu_{F_\mu}([a_n, b_n))$. Fix ϵ_n, $\epsilon > 0$ and note that $(a_n - \epsilon_n, b_n) \supset [a_n, b_n)$. Hence, $\cup_{n\in\mathbb{N}} (a_n - \epsilon_n, b_n) \supset \cup_{n\in\mathbb{N}} [a_n, b_n) = [a, b) \supset [a, b - \epsilon]$. Since $\cup_{n\in\mathbb{N}} (a_n - \epsilon_n, b_n)$ is an open cover for the compact set $[a, b - \epsilon]$, by the Heine-Borel Theorem, there exists $N \in \mathbb{N}$ such that

$$\cup_{n=1}^{N} [a_n - \epsilon_n, b_n) \supset \cup_{n=1}^{N} (a_n - \epsilon_n, b_n) \supset [a, b - \epsilon] \supset [a, b - \epsilon). \tag{2.1}$$

Now, since $\cup_{n\in\mathbb{N}} [a_n, b_n) = [a, b)$ we have $\cup_{n=1}^{N} [a_n, b_n) \subset [a, b)$ and

$$\nu_{F_\mu}([a, b)) \geq \nu_{F_\mu}\left(\cup_{n=1}^{N} [a_n, b_n)\right) = \sum_{n=1}^{N} \nu_{F_\mu}([a_n, b_n)) \text{ by finite additivity.}$$

Hence, we have

$$
\begin{aligned}
0 &\le \nu_{F_\mu}([a,b)) - \sum_{n=1}^{N} \nu_{F_\mu}([a_n,b_n)) \\
&= \nu_{F_\mu}([a,b-\epsilon)) + \nu_{F_\mu}([b-\epsilon,b)) - \sum_{n=1}^{N} \big(\nu_{F_\mu}([a_n-\epsilon_n,b_n)) - \nu_{F_\mu}([a_n-\epsilon_n,a_n))\big) \\
&= \nu_{F_\mu}([a,b-\epsilon)) - \sum_{n=1}^{N} \nu_{F_\mu}([a_n-\epsilon_n,b_n)) \text{ this term} < 0 \text{ by (2.1)} \\
&+ \nu_{F_\mu}([b-\epsilon,b)) + \sum_{n=1}^{N} \nu_{F_\mu}([a_n-\epsilon_n,a_n)) \\
&\le \nu_{F_\mu}([b-\epsilon,b)) + \sum_{n=1}^{N} \nu_{F_\mu}([a_n-\epsilon_n,a_n)) = F_\mu(b) - F_\mu(b-\epsilon) + \sum_{n=1}^{N}(F_\mu(a_n) - F_\mu(a_n-\epsilon_n)).
\end{aligned}
$$

By left-continuity of F_μ, we can choose ϵ such that $F_\mu(b) - F_\mu(b-\epsilon) < \eta/2$ and ϵ_n such that $F_\mu(a_n) - F_\mu(a_n-\epsilon_n) < 2^{-n}\,\eta/2$. Hence,

$$
0 \le \nu_{F_\mu}([a,b)) - \sum_{n=1}^{N} \nu_{F_\mu}([a_n,b_n)) \le \frac{\eta}{2}\left(1 + \sum_{n=1}^{N} 2^{-n}\right).
$$

Letting $N \to \infty$ we have that $\nu_{F_\mu}([a,b)) = \sum_{n=1}^{\infty} \nu_{F_\mu}([a_n,b_n))$.

Since ν_{F_μ} is a pre-measure on a semi-ring, by Carathéodory's Theorem, it has an extension to $\sigma(\mathcal{S}) = \mathcal{B}(\mathbb{R})$. Furthermore, since for $n \in \mathbb{N}$, $[-n,n) \uparrow \mathbb{R}$ and $\nu_{F_\mu}([-n,n)) = F_\mu(n) - F_\mu(-n) = \mu([0,n)) + \mu([-n,0))) < \infty$, this extension is unique.

To verify that $\nu_{F_\mu} = \mu$, it suffices to verify that $\nu_{F_\mu} = \mu$ on $\mathcal{S}$, since ν_{F_μ} extends uniquely to $\mathcal{B}(\mathbb{R})$. There are three cases:

Case 1 ($0 \le a < b$): $\nu_{F_\mu}([a,b)) = F_\mu(b) - F_\mu(a) = \mu([0,b)) - \mu([0,a)) = \mu([0,a)) + \mu([a,b)) - \mu([0,a)) = \mu([a,b))$, since $[0,b) = [0,a) \cup [a,b)$,

Case 2 ($a < 0 < b$): $\nu_{F_\mu}([a,b)) = F_\mu(b) - F_\mu(a) = \mu([0,b)) + \mu([a,0)) = \mu([a,b))$, since $[a,b) = [a,0) \cup [0,b)$,

Case 3 ($a < b \le 0$): $\nu_{F_\mu}([a,b)) = F_\mu(b) - F_\mu(a) = -\mu([b,0)) + \mu([a,0)) = \mu([a,b))$, since $[a,b) = [a,0) - [b,0)$, which completes the proof.

3. If F is a distribution function, show that it can have an infinite number of jump discontinuities, but at most countably many.

Answer: A jump of F, denoted by $J_F(x)$ exists if $J_F(x) = F(x) - \lim_{h\to 0} F(x-h) > 0$ for $h > 0$. This happens if and only if $P(\{x\}) > 0$. Now, the collection of events $E_x := \{\{x\} : P(\{x\}) > 0\}$ is disjoint and all have positive probability. We now show that this collection is countable. Let $C_n = \{E_x : P(E_x) > \frac{1}{n}$ and $x \in \mathbb{R}\}$. The elements of C_n are disjoint events and

$$P\left(\cup_{x_m} E_{x_m}\right) = \sum_{m=1}^{\infty} P(E_{x_m}) = \infty,$$

where the last equality follows from the fact that $P(E_{x_m}) > 0$. So, it must be that C_n has finitely many elements. Also, $\{E_x\}_{x\in\mathbb{R}} = \cup_{n=1}^{\infty} C_n$, which is countable since it is a countable union of finite sets.

4. Show that $\lambda^1((a,b)) = b - a$ for all $a, b \in \mathbb{R}$, $a \leq b$. State and prove the same for λ^n.

   Answer: Let $a < b$ and note that $[a + \frac{1}{k}, b) \uparrow (a,b)$ as $k \to \infty$. Thus, by continuity of measures,

$$\lambda((a,b)) = \lim_{k\to\infty} \lambda([a + 1/k, b) = \lim_{k\to\infty} (b - a - 1/k) = b - a.$$

   Since $\lambda([a,b)) = b - a$, this proves that $\lambda(\{a\}) = 0$.

5. Consider the measure space $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \lambda^n)$. Show that for every $B \in \mathcal{B}(\mathbb{R}^n)$ and $x \in \mathbb{R}^n$, $x+B \in \mathcal{B}(\mathbb{R}^n)$ and that $\lambda^n(x+B) = \lambda^n(B)$. Note: $x+B := \{z : z = x+b,\, b \in B\}$.

   Answer: First, we need to show that $x + B \in \mathcal{B}(\mathbb{R}^n)$ for all $x \in \mathbb{R}^n$ and for all $B \in \mathcal{B}(\mathbb{R}^n)$. Let $\mathcal{A}_x = \{B \in \mathcal{B}(\mathbb{R}^n) : x + B \in \mathcal{B}(\mathbb{R}^n)\}$ and note that $\mathcal{A}_x \subset \mathcal{B}(\mathbb{R}^n)$. Also, $\mathcal{A}_x$ is a σ-algebra associated with $\mathbb{R}^n$, since:

   (a) $\mathbb{R}^n \in \mathcal{A}_x$ given that $x + b \in \mathbb{R}^n$ for all $b \in \mathbb{R}^n$ and $\mathbb{R}^n \in \mathcal{B}(\mathbb{R}^n)$,

   (b) $B \in \mathcal{A}_x \implies x+B \in \mathcal{B}(\mathbb{R}^n) \implies (x+B)^c \in \mathcal{B}(\mathbb{R}^n)$. But since $(x+B)^c = x+B^c$ and $B^c \in \mathcal{B}(\mathbb{R}^n)$, $B^c \in \mathcal{A}_x$.

   (c) $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{A}_x \implies x + A_n \in \mathcal{B}(\mathbb{R}^n)$ for all $n \in \mathbb{N}$. Since $\mathcal{B}(\mathbb{R}^n)$ is a σ-algebra $\bigcup_{n\in\mathbb{N}}(x + A_n) = x + \bigcup_{n\in\mathbb{N}} A_n \in \mathcal{B}(\mathbb{R}^n)$. But since $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{B}(\mathbb{R}^n)$, $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{A}_x$.

   Now, let $R^{n,h} = \times_{i=1}^n [l_i, u_i) \in \mathcal{I}^{n,h} \subset \mathcal{B}(\mathbb{R}^n)$ and note that $x + R^{n,h} \in \mathcal{I}^{n,h} \subset \mathcal{B}(\mathbb{R}^n)$. Hence, $R^{n,h} \in \mathcal{A}_x \implies x + R^{n,h} \in \mathcal{A}_x$. Hence,

$$\mathcal{B}(\mathbb{R}^n) = \sigma(\mathcal{I}^{n,h}) \subset \mathcal{A}_x \subset \mathcal{B}(\mathbb{R}^n),$$

which implies that $x + B \in \mathcal{B}(\mathbb{R}^n)$ for all $x \in \mathbb{R}^n$ and for all $B \in \mathcal{B}(\mathbb{R}^n)$.

Now, set $v(B) = \lambda^n(x+B)$. If $B = \emptyset$, $v(\emptyset) = \lambda^n(x+\emptyset) = \lambda^n(\emptyset) = 0$. Also, for a pairwise disjoint sequence $\{A_n\}_{n\in\mathbb{N}}$, $v\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \lambda^n\left(x + \bigcup_{n\in\mathbb{N}} A_n\right) = \lambda^n\left(\bigcup_{n\in\mathbb{N}}(x + A_n)\right) = \sum_{n\in\mathbb{N}} \lambda^n(x + A_n) = \sum_{n\in\mathbb{N}} v(A_n)$. Hence, v is a measure and

$$v(R^{n,h}) = \lambda^n(x + R^{n,h}) = \prod_{i=1}^{n}(u_i + x_i - (l_i + x_i)) = \prod_{i=1}^{n}(u_i - l_i) = \lambda^n(R^{n,h}).$$

Hence, $v(R^{n,h}) = \lambda^n(R^{n,h})$ for every $R^{n,h} \in \mathcal{I}^{n,h}$. Since $\mathcal{I}^{n,h}$ is a π-system, generates $\mathcal{B}(\mathbb{R}^n)$ and admits an exhausting sequence $[-k, k) \uparrow \mathbb{R}^n$ with $\lambda^n([-k, k)^n) = (2k)^n < \infty$, we have by Carathéodory Theorem that $\lambda^n = v$ on $\mathcal{B}(\mathbb{R}^n)$.

Chapter 3

Exercises

1. Suppose $(\Omega, \mathcal{F})$ and $(\mathbb{Y}, \mathcal{G})$ are measure spaces and $f : \Omega \to \mathbb{Y}$. Show that: a) $I_{f^{-1}(A)}(\omega) = (I_A \circ f)(\omega)$ for all ω; b) f is measurable if, and only if, $\sigma(\{f^{-1}(A) : A \in \mathcal{G}\}) \subset \mathcal{F}$.

   Answer: a) For any subset $A \subset Y$, we have $f^{-1}(A) = \{\omega : f(\omega) \in A\}$. Then,

   $$I_{f^{-1}(A)}(\omega) = I_{\{\omega: f(\omega) \in A\}}(\omega) = I_A(f(\omega)) = (I_A \circ f)(\omega).$$

   b) Since f is measurable, $f^{-1}(\mathcal{G}) \subset \mathcal{F}$. By monotonicity of σ-algebras, $\sigma(f^{-1}(\mathcal{G})) = \sigma(\{f^{-1}(A) : A \in \mathcal{G}\}) \subset \mathcal{F}$. Now, $\sigma(f^{-1}(\mathcal{G})) = f^{-1}(\sigma(\mathcal{G})) = f^{-1}(\mathcal{G}) \subset \mathcal{F}$. The last set containment implies measurability.

2. Show that for any function $f : \mathbb{X} \to \mathbb{Y}$ and any collection of subsets $\mathcal{G}$ of $\mathbb{Y}$, $f^{-1}(\sigma(\mathcal{G})) = \sigma(f^{-1}(\mathcal{G}))$

   Answer: $f^{-1}(\sigma(\mathcal{G}))$ is a σ-algebra associated with $\mathbb{X}$. Since $\mathcal{G} \subset \sigma(\mathcal{G})$, $f^{-1}(\mathcal{G}) \subset f^{-1}(\sigma(\mathcal{G}))$ and consequently $\sigma(f^{-1}(\mathcal{G})) \subset f^{-1}(\sigma(\mathcal{G}))$.

   Now, as in Theorem 3.1, $\mathcal{U} = \{U \in 2^{\mathbb{Y}} : f^{-1}(U) \in \sigma(f^{-1}(\mathcal{G}))\}$ is a σ-algebra. By definition of $\mathcal{U}$

   $$f^{-1}(\mathcal{U}) \subset \sigma(f^{-1}(\mathcal{G})).$$

   Also, $\mathcal{G} \subset \mathcal{U}$ since $f^{-1}(\mathcal{G}) \subset f^{-1}(\mathcal{U}) \subset \sigma(f^{-1}(\mathcal{G}))$. Since $\mathcal{U}$ is a σ-algebra we have that $\sigma(\mathcal{G}) \subset \mathcal{U}$. So,

   $$f^{-1}(\sigma(\mathcal{G})) \subset f^{-1}(\mathcal{U}) \subset \sigma(f^{-1}(\mathcal{C})).$$

   The last set containment combined with the reverse obtained on the last paragraph completes the proof.

3. Let $i \in I$ where I is an arbitrary index set. Consider $f_i : (\mathbb{X}, \mathcal{F}) \to (\mathbb{X}_i, \mathcal{F}_i)$.

   (a) Show that for all i, the smallest σ-algebra associated with $\mathbb{X}$ that makes f_i measurable is given by $f_i^{-1}(\mathcal{F}_i)$.

   (b) Show that $\sigma\left(\bigcup_{i \in I} f_i^{-1}(\mathcal{F}_i)\right)$ is the smallest σ-algebra associated with $\mathbb{X}$ that makes all f_i simultaneously measurable.

   Answer: a) f_i is measurable if $f_i^{-1}(\mathcal{F}_i) \subset \mathcal{F}$. But by monotonicity of $\sigma(\cdot)$ we have $\sigma(f_i^{-1}(\mathcal{F}_i)) = f_i^{-1}(\mathcal{F}_i) \subset \mathcal{F}$ since $f_i^{-1}(\mathcal{F}_i)$ is a σ-algebra. b) $f_i^{-1}(\mathcal{F}_i) \subset \mathcal{F}$ for all $i \in I$ because f_i is measurable. But any sub-σ-algebra of $\mathcal{F}$ that makes all f_i measurable functions must contain all $f_i^{-1}(\mathcal{F}_i)$, i.e., $\bigcup_{i \in I} f_i^{-1}(\mathcal{F}_i)$. However, unions of σ-algebras are not necessarily σ-algebras. Hence, we consider $\sigma\left(\bigcup_{i \in I} f_i^{-1}(\mathcal{F}_i)\right)$, the smallest σ-algebra that makes all f_i simultaneously measurable.

4. Let $X : (\Omega, \mathcal{F}, P) \to (S, \mathcal{B}_S)$ where $S \subset \mathbb{R}^k$ and $\mathcal{B}_S = \{B \cap S : B \in \mathcal{B}^k\}$ be a random vector with $k \in \mathbb{N}$, and $g : (S, \mathcal{B}_S) \to (T, \mathcal{B}_T)$ be measurable where $T \subset \mathbb{R}^p$ with $p \in \mathbb{N}$. If $Y = g(X)$, show that

   (a) $\sigma(Y) := Y^{-1}(\mathcal{B}_T) \subset \sigma(X) := X^{-1}(\mathcal{B}_S)$,

   (b) if $k = p$ and g is bijective, $\sigma(Y) = \sigma(X)$.

   Answer: (a) $E \in Y^{-1}(\mathcal{B}_T) \implies E = Y^{-1}(B_T)$ for some $B_T \in \mathcal{B}_T$. Now,

$$\begin{aligned} E &= \{\omega : Y(\omega) \in B_T\} = \{\omega : g(X(\omega)) \in B_T\} = \{\omega : X(\omega) \in g^{-1}(B_T)\} \\ &= X^{-1}(g^{-1}(B_T)). \end{aligned}$$

   Since g is measurable, $g^{-1}(B_T) \in \mathcal{B}_S$ and since X is a random vector $X^{-1}(g^{-1}(B_T)) \in \sigma(X) := X^{-1}(\mathcal{B}_S)$. Hence, $\sigma(Y) \subset \sigma(X)$.

   (b) First, observe that since g is bijective, it must be that $k = p$ and $S = T$. For any $B_T \in \mathcal{B}_T$,

$$\begin{aligned} g^{-1}(B_T) &= g^{-1}(g(B)) \text{ for some } B \subset S \\ &= B \in \mathcal{B}_S \text{ since } g^{-1} \text{ is an inverse function and } g \text{ is measurable.} \end{aligned}$$

   Hence, any $B_T \in \mathcal{B}_T$ is such that $B_T = g(B)$ where $B \in \mathcal{B}_S$. Similarly, due to the existence of the inverse g^{-1}, for any $B_S \in \mathcal{B}_S$, $B_S = g^{-1}(B)$ where $B \in \mathcal{B}_T$. Hence, if

$\mathcal{C} := \{g^{-1}(B) : B \in \mathcal{B}_T\}$ then $\mathcal{B}_S \subset \mathcal{C}$. But measurability of g assures that $\mathcal{C} \subset \mathcal{B}_S$ Hence, $X^{-1}(\mathcal{B}_S) := \sigma(X) = X^{-1}(\mathcal{C}) = \{X^{-1}(g^{-1}(B)) : B \in \mathcal{B}_T\} = \sigma(Y)$.

Chapter 4

Exercises

1. Prove Theorem 4.2.

   Answer: Let $f = \sum_{i=0}^{I} y_i I_{A_i}$ and $f = \sum_{j=0}^{J} y_j I_{B_j}$ be standard representations of f and g. Then,

   $$f \pm g = \sum_{i=0}^{I} \sum_{j=0}^{J} (y_i \pm z_j) I_{A_i \cap B_j}$$

   and

   $$fg = \sum_{i=0}^{I} \sum_{j=0}^{J} (y_i z_j) I_{A_i \cap B_j}$$

   with $(A_i \cap B_j) \cap (A_{i'} \cap B_{j'}) = \emptyset$ whenever $(i, j) \neq (i', j')$. After relabeling and merging the double sums into single sums we have the result. The case for cf is obvious. f simple implies f^+ and f^- are simple by definition, and since $|f| = f^+ + f^-$, $|f|$ is simple.

2. Show that if f is a non-negative measurable simple function, its integral, as defined in Definition 4.3 is equal to $I_\mu(f)$.
   Answer: Since f is simple and $f \leq f$, f is one of the simple functions (denoted by ϕ) appearing in Definition 21 of the class notes. Hence, $\int f d\mu \geq I_\mu(f)$. Also, if ϕ is a simple function such that $\phi \leq f$, by monotonicity of the integral of simple functions we have $I_\mu(\phi) \leq I_\mu(f)$, hence

   $$\sup_\phi I_\mu(\phi) := \int f d\mu \leq I_\mu(f).$$

   Combining the two inequalities we have $\int f d\mu = I_\mu(f)$.

3. Let $(\mathbb{X}, \mathcal{F})$ be a measurable space and $\{\mu_n\}_{n\in\mathbb{N}}$ be a sequence of measures defined on it. Noting that $\mu = \sum_{n\in\mathbb{N}} \mu_n$ is also a measure on $(\mathbb{X}, \mathcal{F})$ (you don't have to prove this), show that

$$\int_{\mathbb{X}} f d\mu = \sum_{n\in\mathbb{N}} \int_{\mathbb{X}} f d\mu_n$$

for f non-negative and measurable.

Answer: First, let $f = I_F \geq 0$ for $F \in \mathcal{F}$. Then, f is measurable and

$$\int_{\mathbb{X}} f d\mu = \int_{\mathbb{X}} I_F d\mu = \mu(F) = \sum_{n\in\mathbb{N}} \mu_n(F) = \sum_{n\in\mathbb{N}} \int_{\mathbb{X}} I_F d\mu_n = \sum_{n\in\mathbb{N}} \int_{\mathbb{X}} f d\mu_n.$$

Hence, the result holds for indicator functions. Now, consider a simple non-negative function $f = \sum_{j=0}^{m} a_j I_{A_j}$ where $a_j \geq 0$ and $A_j \in \mathcal{F}$. Then,

$$\begin{aligned}\int_{\mathbb{X}} f d\mu &= \int_{\mathbb{X}} \sum_{j=0}^{m} a_j I_{A_j} d\mu = \sum_{j=0}^{m} a_j \int_{\mathbb{X}} I_{A_j} d\mu = \sum_{j=0}^{m} a_j \mu(A_j) = \sum_{j=0}^{m} a_j \sum_{n\in\mathbb{N}} \mu_n(A_j) \\ &= \sum_{n\in\mathbb{N}} \sum_{j=0}^{m} a_j \mu_n(A_j) = \sum_{n\in\mathbb{N}} \int_{\mathbb{X}} f d\mu_n.\end{aligned}$$

Hence, the result holds for simple non-negative functions. Lastly, let f be non-negative and measurable. By Theorem 3.3 in the class notes, there exists a sequence $\{\phi_n\}_{n\in\mathbb{N}}$ of non-negative, non-decreasing, measurable simple function such that $\sup_{n\in\mathbb{N}} \phi_n = f$. By Beppo-Levi's Theorem

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

Hence,

$$
\begin{aligned}
\int_{\mathbb{X}} f d\mu &= \sup_{n\in\mathbb{N}} \int_{\mathbb{X}} \phi_n d\mu = \sup_{n\in\mathbb{N}} \sum_{j=1}^{\infty} \int_{\mathbb{X}} \phi_n d\mu_j \\
&= \sup_{n\in\mathbb{N}} \sup_{m\in\mathbb{N}} \sum_{j=1}^{m} \int_{\mathbb{X}} \phi_n d\mu_j \text{ since } \textstyle\int_{\mathbb{X}} \phi_n d\mu_j \text{ is nondecreasing.} \\
&= \sup_{m\in\mathbb{N}} \sup_{n\in\mathbb{N}} \sum_{j=1}^{m} \int_{\mathbb{X}} \phi_n d\mu_j = \sup_{m\in\mathbb{N}} \lim_{n\to\infty} \sum_{j=1}^{m} \int_{\mathbb{X}} \phi_n d\mu_j \\
&= \sup_{m\in\mathbb{N}} \sum_{j=1}^{m} \lim_{n\to\infty} \int_{\mathbb{X}} \phi_n d\mu_j \\
&= \sup_{m\in\mathbb{N}} \sum_{j=1}^{m} \int_{\mathbb{X}} \lim_{n\to\infty} \phi_n d\mu_j \text{ by Beppo-Levi's Theorem} \\
&= \sup_{m\in\mathbb{N}} \sum_{j=1}^{m} \int_{\mathbb{X}} f d\mu_j = \sum_{j\in\mathbb{N}} \int_{\mathbb{X}} f d\mu_j .
\end{aligned}
$$

4. Let $(\mathbb{X}, \mathcal{F}, \mu)$ be a measure space and $f : (\mathbb{X}, \mathcal{F}, \mu) \to (\mathbb{R}, \mathcal{B})$ be measurable and non-negative. For every $F \in \mathcal{F}$ consider $\int I_F f d\mu$. Is this a measure?
Answer: Let $v(F) = \int I_F f d\mu$. Then v is a $[0,\infty]$-valued set function defined for $F \in \mathcal{F}$. Then,

   (a) $I_\emptyset = 0$ and clearly $v(\emptyset) = 0$.

   (b) Let $F = \cup_{i\in\mathbb{N}} F_i$ be a union of pairwise disjoint sets in $\mathcal{F}$. Then, $\sum_{i=1}^{\infty} I_{F_i} = I_F$ and

$$
\begin{aligned}
v(F) &= \int \left(\sum_{i=1}^{\infty} I_{F_i}\right) f d\mu = \int \left(\sum_{i=1}^{\infty} I_{F_i} f\right) d\mu \\
&= \sum_{i=1}^{\infty} \int I_{F_i} f d\mu = \sum_{i=1}^{\infty} v(F_i)
\end{aligned}
$$

.

5. Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\{F_n\}_{n\in\mathbb{N}} \subset \mathcal{F}$.

   (a) Prove that $I_{\liminf\limits_{n\to\infty} F_n} = \liminf\limits_{n\to\infty} I_{F_n}$ and $I_{\limsup\limits_{n\to\infty} F_n} = \limsup\limits_{n\to\infty} I_{F_n}$.

   (b) Prove that $P\left(\liminf\limits_{n\to\infty} F_n\right) \leq \liminf\limits_{n\to\infty} P(F_n)$.

(c) Prove that $\limsup_{n\to\infty} P(F_n) \leq P\left(\limsup_{n\to\infty} F_n\right)$.

Answer: Part (a) is straightforward by noting that $I_{\cap F_n} = \inf I_{F_n}$ and $I_{\cup F_n} = \sup I_{A_n}$. (b) Part (a) combined with Fatou's Lemma gives,

$$P(\liminf F_n) = \int I_{\liminf F_n} dP = \int \liminf I_{F_n} dP \leq \liminf \int I_{F_n} dP.$$

(c) Again, by Fatou's Lemma (the reverse) we have,

$$P(\limsup F_n) = \int I_{\limsup F_n} dP = \int \limsup I_{F_n} dP \geq \limsup \int I_{F_n} dP.$$

Chapter 5

Exercises

1. Prove Theorem 4.2.

   Answer: Let $f = \sum_{i=0}^{I} y_i I_{A_i}$ and $f = \sum_{j=0}^{J} y_j I_{B_j}$ be standard representations of f and g. Then,

$$f \pm g = \sum_{i=0}^{I} \sum_{j=0}^{J} (y_i \pm z_j) I_{A_i \cap B_j}$$

   and

$$fg = \sum_{i=0}^{I} \sum_{j=0}^{J} (y_i z_j) I_{A_i \cap B_j}$$

   with $(A_i \cap B_j) \cap (A_{i'} \cap B_{j'}) = \emptyset$ whenever $(i, j) \neq (i', j')$. After relabeling and merging the double sums into single sums we have the result. The case for cf is obvious. f simple implies f^+ and f^- are simple by definition, and since $|f| = f^+ + f^-$, $|f|$ is simple.

2. Prove Theorem 4.10.

   Answer: Since $f = f^+ - f^-$ and f^+ and f^- are nonnegative, use Theorems 4.6 and 4.8 in your notes.

3. Use Markov's inequality to prove the following for $a > 0$ and $g : (0, \infty) \to (0, \infty)$ that is increasing:

$$P(|X(\omega)| \geq a) \leq \frac{1}{g(a)} \int g(|X|) dP$$

   Answer: Since g is increasing, $\{\omega : |X(\omega)| \geq a\} = \{\omega : g(|X(\omega)|) \geq g(a)\}$. Hence, since g is positive

$$g(a) I_{\{\omega:|X(\omega)| \geq a\}} = g(a) I_{\{\omega:g(|X(\omega)|) \geq g(a)\}} \leq g(|X(\omega)|).$$

Integrating both sides we have $g(a)P(\{\omega : |X(\omega)| \geq a\}) \leq \int g(|X(\omega)|)dP$. This completes the proof as $g(a) > 0$.

4. Let X be a random variable defined in the probability space $(\Omega, \mathcal{F}, P)$ with $E(X^2) < \infty$. Consider a function $f : \mathbb{R} \to \mathbb{R}$. What restrictions are needed on f to guarantee that $f(X)$ is a random variable with $E(f(X)^2) < \infty$?

   Answer: Recall that if $X : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$, we say that X is a random variable (measurable real valued function) if, and only if, for all $B \in \mathcal{B}_{\mathbb{R}}$ we have $X^{-1}(B) \in \mathcal{F}$. Hence, if $h(\omega) := f(X(\omega)) = (f \circ X)(\omega) : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ we require that for all $B \in \mathcal{B}_{\mathbb{R}}$ we have $h^{-1}(B) = (f \circ X)^{-1}(B) = X^{-1}(f^{-1}(B)) \in \mathcal{F}$. That is, $f^{-1}(B) \in \mathcal{B}_{\mathbb{R}}$.

   Since X is a random variable (measurable) and given that $f^{-1}(B) \in \mathcal{B}_{\mathbb{R}}$ for all $B \in \mathcal{B}_{\mathbb{R}}$, $f(X)$ is a random variable (measurable). Since the f^2 is a continuous function of f, f^2 is also a random variable (measurable). Hence, we can consider the integrability (or not) of $f(X)^2$, i.e., whether or not $E(f(X)^2) < \infty$. We give two general restrictions on f that give $E(f(X)^2) < \infty$. First, suppose that $\sup_{\omega \in \Omega} |h(\omega)| = \sup_{\omega \in \Omega} |(f \circ X)(\omega)| < C$. Then,

   $$\left| \int f^2 dP \right| \leq \int h^2 dP \leq C^2 \int dP = C^2.$$

   Second, suppose that $h^2 \leq X^2$ for all $\omega \in \Omega$. Then, $\int h^2 dP \leq \int X^2 dP < \infty$.

   Note that, in general, it is not true that $E(f(X)^2) < \infty$ even if $E(X^2) < \infty$. For example, suppose that $X \sim U[0,1]$. Then, $E(X^2) = 1/3$. Now, let $Y := f(X) = \tan\left(\pi(X - \frac{1}{2})\right)$ and we can easily obtain that the probability density of Y is

   $$f_Y(y) = \left| \frac{d}{dy} f^{-1}(y) \right| = \left| \frac{d}{dy} \left(\frac{1}{2} + \frac{1}{\pi} \arctan(y) \right) \right| = \frac{1}{\pi} \frac{1}{1+y^2},\ y \in \mathbb{R}.$$

   But this is the Cauchy density and $\int y^2 f_Y(y) dy$ does not exist.

5. Let $X : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B})$ be a random variable. Show that if $V(X) := E\left((X - E(X))\right)^2 = 0$ then X is a constant with probability 1.

   Answer: From your notes, if $\int_\Omega X^2 dP = 0$ then $X^2 = 0$ almost everywhere. If N is a null set $\int_\Omega X^2 dP = \int_N X^2 dP + \int_{N^c} X^2 dP = \int_N X^2 dP + \int_{N^c} 0 dP = 0$. Thus, $P(X^2 = x) = 0$ for $x \neq 0$ and $P(X^2 = 0) = 1$. But this is equivalent to $P(X = 0) = 1$. Hence, $V(X) = E\left((X - E(X))\right)^2 = 0$ implies $P(X - E(X) = 0) = P(X = E(X)) = 1$.

6. Consider the following statement: *f is continuous almost everywhere if, and only if, it is almost everywhere equal to an everywhere continuous function.* Is this true or false? Explain, with precise mathematical arguments.

   Answer: False. Consider the function $I_{\mathbb{Q}}(x)$, where $x \in \mathbb{R}$. This function is nowhere continuous in $\mathbb{R}$, but it is equal to 0 almost everywhere, an everywhere continuous function. Alternatively, the function $I_{[0,\infty)}(x)$ is continuous everywhere except at $\{0\}$, a set of measure zero. So, it is continuous almost everywhere. However, there is no everywhere continuous function in $\mathbb{R}$ that is equal $I_{[0,\infty)}(x)$ almost everywhere.

7. Adapt the proof of Lebesgue's Dominated Convergence Theorem in your notes to show that any sequence $\{f_n\}_{n\in\mathbb{N}}$ of measurable functions such that $\lim_{n\to\infty} f_n(x) = f(x)$ and $|f_n| \leq g$ for some g with g^p nonnegative and integrable satisfies

$$\lim_{n\to\infty} \int |f_n - f|^p d\mu = 0.$$

   Answer: (3 points) First, note that $|f_n - f|^p \leq (|f_n| + |f|)^p$. Since $|f_n - f| \to 0$ we have that $|f_n| \to |f|$. Consequently, for all $\epsilon > 0$ there exists $N_\epsilon \in \mathbb{N}$ such that for $n \geq N_\epsilon$ we have

$$|f_n| - \epsilon \leq |f| \leq |f_n| + \epsilon \leq g + \epsilon$$

   since $|f_n| < g$. Consequently, $|f| \leq g$, $|f|^p \leq g^p$ and $|f_n - f|^p \leq 2^p g^p$ where g^p is nonnegative and integrable. Now, letting $\phi_n = |f_n - f|^p$ we have that $\lim\limits_{n\to\infty} \phi_n = 0$ and by Lebesgue's dominated convergence theorem in the class notes

$$\lim_{n\to\infty} \int_{\mathbb{X}} \phi_n d\mu = \int_{\mathbb{X}} \lim_{n\to\infty} \phi_n d\mu = 0.$$

8. Let λ be the one-dimensional Lebesgue measure for the Borel sets of $\mathbb{R}$. Show that for every integrable function f, the function

$$g(x) = \int_{(0,x)} f(t) d\lambda, \text{ for } x > 0$$

   is continuous.

Answer: Consider a sequence $\{y_n\}_{n\in\mathbb{N}}$ with $0 < x < y_n$ such that $\lim_{n\to\infty} y_n = x$. Then,

$$\begin{aligned}
g(y_n) - g(x) &= \int_{(0,y_n)} f d\lambda - \int_{(0,x)} f d\lambda = \int_{(0,\infty)} I_{(0,y_n)} f d\lambda - \int_{(0,\infty)} I_{(0,x)} f d\lambda \\
&= \int_{(0,\infty)} (I_{(0,y_n)} - I_{(0,x)}) f d\lambda = \int_{(0,\infty)} I_{(x,y_n)} f d\lambda \\
|g(y_n) - g(x)| &\leq \int_{(0,\infty)} I_{[x,y_n)} |f| d\lambda.
\end{aligned}$$

Now, $I_{[x,y_n)}|f| \leq |f|$ and $\int_{(0,\infty)} |f| d\lambda < \infty$ since f is integrable. Also, $\lim_{n\to\infty} I_{[x,y_n)} f = 0$ almost everywhere (ae). Thus, by dominated convergence in the class notes

$$\begin{aligned}
\lim_{n\to\infty} |g(y_n) - g(x)| &\leq \lim_{n\to\infty} \int_{(0,\infty)} I_{(x,y_n)} |f| d\lambda \\
&= \int_{(0,\infty)} \lim_{n\to\infty} I_{(x,y_n)} |f| d\lambda = 0.
\end{aligned}$$

By repeating the argument for $y_n \uparrow x$ we obtain continuity of g at x.

9. Show that if X is a random variable with $E(|X|^p) < \infty$ then $|X|$ is almost everywhere real valued.

   Answer: Let $N = \{\omega : |X(\omega)| = \infty\} = \{\omega : |X(\omega)|^p = \infty\}$. Then $N = \cap_{n\in\mathbb{N}}\{\omega : |X(\omega)|^p \geq n\}$. Then,

$$\begin{aligned}
P(N) &= P\left(\cap_{n\in\mathbb{N}}\{\omega : |X(\omega)|^p \geq n\}\right) \\
&= \lim_{n\to\infty} P\left(\{\omega : |X(\omega)|^p \geq n\}\right) \text{ by continuity of probability measures} \\
&\leq \lim_{n\to\infty} \frac{1}{k} \int_\Omega |X|^p dP \text{ by Markov's Inequality} \\
&= 0 \text{ since } \textstyle\int_\Omega |X|^p dP \text{ is finite.}
\end{aligned}$$

10. Suppose $X : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B})$ is a random variable with $E(|X|) < \infty$. Let $N \in \mathcal{F}$ be such that $P(N) = 0$ and define

$$Y(\omega) = \begin{cases} X(\omega) & \text{if } \omega \notin N \\ c & \text{if } \omega \in N \end{cases},$$

   where $c \in \mathbb{R}$. Is Y integrable? Is $E(X) = E(Y)$?

   Answer: Yes, for both questions. We can change an integrable random variables at any set of measure zero without changing the integral.

Chapter 6

Exercises

1. Let $X_1, X_2 \in \mathcal{L}^2$ and define $\text{Cov}(X_1, X_2) = E([X_1 - E(X_1)][X_2 - E(X_2)])$. Show that $\text{Cov}(X_1, X_2) = E(X_1 X_2) - E(X_1)E(X_2)$ and that if X_1 and X_2 are independent $\text{Cov}(X_1, X_2) = 0$.

   Answer: From the definition of $\text{Cov}(X_1, X_2)$ and linearity of expectations

   $\text{Cov}(X_1, X_2) = E(X_1X_2 - X_1E(X_2) - X_2E(X_1) + E(X_1)E(X_2)) = E(X_1X_2) - E(X_1)E(X_2).$

   Independence of X_1 and X_2 implies that $E(X_1X_2) = E(X_1)E(X_2)$. Hence, $\text{Cov}(X_1, X_2) = 0$.

2. Let $\{X_n\}_{n \in \mathbb{N}}$ be a sequence of random variables that are independent and share the same continuous distribution. Let p be a permutation of $\{1, \cdots, n\}$ for $n \in \mathbb{N}$. Show that $(X_1, \cdots, X_n)$ and $(X_{p(1)}, \cdots, X_{p(n)})$ have the same distribution.

   Answer: Since the random variables are independent and have the same distribution, say F,

   $$P(X_1 \leq x_1, \cdots, X_n \leq x_n) = \prod_{i=1}^{n} F(x_i)$$

   for all $x_i \in \mathbb{R}$. If $\{p(i)\}_{i=1}^n$ is a permutation of $\{1, \cdots, n\}$, then

   $$P(X_{p(1)} \leq x_1, \cdots, X_{p(n)} \leq x_n) = \prod_{i=1}^{n} F(x_i).$$

   Hence, $(X_1, \cdots, X_n)$ and $X_{p(1)}, \cdots, X_{p(n)}$ have the same distribution.

3. Let I be a finite index set and consider the collection of σ-algebras $\{\mathcal{B}_i\}_{i\in I}$. Show that this collection is independent if, and only if, for every choice of non-negative $\mathcal{B}_i$-measurable random variable X_i, we have $E\left(\prod_{i\in I} X_i\right) = \prod_{i\in I} E(X_i)$.

   Answer: If $E\left(\prod_{i\in I} X_i\right) = \prod_{i\in I} E(X_i)$ whenever $X_i \in \mathcal{B}_i$, then for any $A_i \in \mathcal{B}_i$, take $X_i = I_{A_i}$ and

$$E\left(\prod_{i\in I} X_i\right) = P\left(\bigcap_{i\in I} A_i\right) = \prod_{i\in I} P(A_i) = \prod_{i\in I} E(X_i)$$

   and consequently $\{A_i\}_{i\in I}$ are independent and $\{\mathcal{B}_i\}_{i\in I}$ are independent σ-algebras.

   Now, suppose $\{\mathcal{B}_i\}_{i\in I}$ are independent σ-algebras. For $A_i \in \mathcal{B}_i$ and $X_i = I_{A_i}$ we have

$$E\left(\prod_{i\in I} X_i\right) = \prod_{i\in I} E(X_i).$$

   Now, if $\{X_i\}_{i\in I}$ are simple, then write $X_i = \sum_j x_{ij} I_{A_{ij}}$ for $A_{ij} \in \mathcal{B}_i$. Then, we have

$$\begin{aligned} E\left(\prod_{i\in I} X_i\right) &= E\left(\prod_{i\in I}\sum_{j(i)} x_{ij(i)} I_{A_{ij(i)}}\right) = E\left(\sum_{j(i),i\in I}\prod_{i\in I} x_{ij(i)} I_{\underset{i\in I}{\cap} A_{ij(i)}}\right) \\ &= \sum_{j(i),i\in I}\prod_{i\in I} x_{ij(i)} P\left(\underset{i\in I}{\cap} A_{ij(i)}\right) = \sum_{j(i),i\in I}\prod_{i\in I} x_{ij(i)} P(A_{ij(i)}) \\ &= \prod_{i\in I}\sum_{j(i)} x_{ij(i)} P(A_{ij(i)}) = \prod_{i\in I} E(X_i). \end{aligned}$$

   If X_i is a non-negative $\mathcal{B}_i$-measurable function, there exists $X_i^{(n)}$ such that $X_i^{(n)}$ is simple and $0 < X_i^{(n)} \uparrow X_i$. Then, it follows that $\prod_{i\in I} X_i^{(n)} \uparrow \prod_{i\in I} X_i$ and by the monotone convergence theorem $E\left(\prod_{i\in I} X_i^{(n)}\right) \uparrow E\left(\prod_{i\in I} X_i\right)$ and from the previous argument , the left side is $\prod_{i\in I} E\left(X_i^{(n)}\right) \uparrow \prod_{i\in I} E\left(X_i\right)$ again using the monotone convergence theorem.

4. If E is an event that is independent of the π-system P and $E \in \sigma(P)$, then $P(E)$ is either 0 or 1.

   Answer: Set $\mathcal{C}_1 = E$ and $\mathcal{C}_2 = P$ and it follows that $\mathcal{C}_1$ is independent of $\mathcal{C}_2$. This implies that $\sigma(\mathcal{C}_1)$ is independent of $\sigma(\mathcal{C}_2)$. Therefore, E is independent of E and $P(E) = 0$ or 1.

5. Let $\{A_i\}_{i=1}^n$ be independent events. Show that $P\left(\cup_{i=1}^n A_i\right) = 1 - \prod_{i=1}^n P(A_i^c)$.

   Answer: By De Morgan's Law $\left(\bigcup_{i=1}^n A_i\right)^c = \bigcap_{i=1}^n A_i^c$. Hence,

$$P\left(\left(\bigcup_{i=1}^n A_i\right)^c\right) = 1 - P\left(\bigcup_{i=1}^n A_i\right) = P\left(\bigcap_{i=1}^n A_i^c\right)$$

   which implies

$$P\left(\bigcup_{i=1}^n A_i\right) = 1 - P\left(\bigcap_{i=1}^n A_i^c\right) = 1 - \bigcap_{i=1}^n P(A_i^c)$$

   where the last equality follows from the fact that if $\{A_i\}_{i=1}^n$ are independent events, so are $\{A_i^c\}_{i=1}^n$.

6. We have proved that if X and Y are independent, then $f(X)$ and $g(Y)$ are independent if f and g are measurable. Is it possible to have X and Y be dependent and $f(X)$ and $g(Y)$ be independent? If so, give an example, if not, prove.

   Answer: Yes, it is possible. Consider two independent random variables X_1 and X_2 and another random variable W that is independent of X_1 and X_2 and takes on the values 1 and -1 with probability $1/2$ each. Now, define two new random variables $X = WX_1$ and $Y = WX_2$. X and Y are functionally connected and cannot be independent. However, $X^2 = X_1^2$ and $Y^2 = X_2^2$, which are independent since X_1 and X_2 are independent.

Chapter 7

Exercises

1. Let $\{X_n\}_{n\in\mathbb{N}} \subset \mathcal{L}^p$ for $p \in [1,\infty)$ be a sequence of non-negative functions. Show that

$$\|\sum_{n=1}^{\infty} X_n\|_p \leq \sum_{n=1}^{\infty} \|X_n\|_p.$$

Answer: Let $S_n = \sum_{n=1}^{N} X_n$. Since $X_n \geq 0$ for all n, $0 \leq S_1 \leq S_2 \leq \cdots$. Given that $|S_N|^p \leq 2^p \sum_{n=1}^{N} |X_n|^p$. we have $\int_\Omega |S_N|^p dP \leq 2^p \sum_{n=1}^{N} \int_\Omega |X_n|^p dP < \infty$. Consequently, $S_N \in \mathcal{L}^p$. By Minkowski's inequality

$$\|S_N\|_p \leq \sum_{n=1}^{N} \|X_n\|_p \leq \sum_{n=1}^{\infty} \|X_n\|_p, \tag{7.1}$$

which implies $\|S_N\|_p^p \leq (\sum_{n=1}^{\infty} \|X_n\|_p)^p$. By Beppo-Levi's Theorem

$$\begin{aligned}\sup_{n\in\mathbb{N}}\|S_N\|_p^p &= \sup_{n\in\mathbb{N}} \int_\Omega S_N^p dP = \int_\Omega \sup_{n\in\mathbb{N}} S_N^p dP = \int_\Omega \sup_{n\in\mathbb{N}} \left(\sum_{n=1}^{N} X_n\right)^p dP \\ &= \int_\Omega \left(\sup_{n\in\mathbb{N}} \sum_{n=1}^{N} X_n\right)^p dP = \|\sum_{n=1}^{\infty} X_n\|_p^p. \end{aligned}\tag{7.2}$$

Hence, by inequality (7.1) and (7.2) we have

$$\sup_{n\in\mathbb{N}}\|S_N\|_p^p = \|\sum_{n=1}^{\infty} X_n\|_p^p \leq \left(\sum_{n=1}^{\infty} \|X_n\|_p\right)^p.$$

Consequently, $\| \sum_{n=1}^{\infty} X_n\|_p \leq \sum_{n=1}^{\infty} \|X_n\|_p$.

2. Show that if $\sum_{n\in\mathbb{N}} x_n$ converges absolutely, then it converges.

Answer: Suppose $N_1,\, N_2 \in \mathbb{N},\, N_1 < N_2$ and $\sum_{n\in\mathbb{N}} x_n$ converges absolutely. Note that $\sum_{n=1}^{N_2} |x_n| - \sum_{n=1}^{N_1} |x_n| = \sum_{n=N_1+1}^{N_2} |x_n|$. If $N_1 \to \infty$, then $\sum_{n=1}^{N_2} |x_n| - \sum_{n=1}^{N_1} |x_n| \to 0$, as every convergent sequence is Cauchy. Also, since

$$|x_{N_1+1} + x_{N_2+1} + \cdots x_{N_2}| \leq \sum_{n=N_1+1}^{N_2} |x_n|,$$

$$|\sum_{n=1}^{N_2} x_n - \sum_{n=1}^{N_1} x_n| = |x_{N_1+1} + x_{N_2+1} + \cdots x_{N_2}| \leq \sum_{n=N_1+1}^{N_2} |x_n| \to 0.$$

Since $\mathbb{R}$ is complete, $\lim_{N\to\infty} \sum_{n=1}^{N} x_n$ converges.

3. Prove Theorem 7.9.

   Answer: Just $X_n \xrightarrow{\mathcal{L}^1} X$ on item 2. For $\epsilon > 0$ let $A_n = \{|X_n - X| > \epsilon\}$

$$\begin{aligned} E|X_n - X| &= E(|X_n - X| I_{|X_n - X| < \epsilon} + |X_n - X| I_{|X_n - X| \geq \epsilon}) \\ &\leq \epsilon + E(|X_n| I_{A_n}) + E(|X| I_{A_n}). \end{aligned}$$

   $P(A_n) \to 0$ as $n \to \infty$, hence $E(|X_n| I_{A_n}) \to 0$ by Theorem 7.6. Similarly, $E(|X| I_{A_n}) \to 0$, which gives the result.

4. Let $\{g_n\}_{n=1,2,\cdots}$ be a sequence of real valued functions that converge uniformly to g on an open set S, containing x, and g is continuous at x. Show that if $\{X_n\}_{n=1,2,\cdots}$ is a sequence of random variables taking values in S such that $X_n \xrightarrow{p} X$, then

$$g_n(X_n) \xrightarrow{p} g(X).$$

   Note: Recall that a sequence of real valued functions $\{g_n\}_{n=1,2,\cdots}$ converges uniformly to g on a set S if, for every $\epsilon > 0$ there exists $N_\epsilon \in \mathbb{N}$ (depending only on ϵ) such that for all $n > N_\epsilon$, $|g_n(x) - g(x)| < \epsilon$ for every $x \in S$.

   Answer: Let $\epsilon, \delta > 0$ and define the following subsets of the sample space: $S_1^n = \{\omega : |g_n(X_n) - g(X)| < \epsilon\}$, $S_2^n = \{\omega : |g_n(X_n) - g(X_n)| < \epsilon/2\}$, $S_3^n = \{\omega : |g(X_n) - g(X)| < \epsilon/2\}$, $S_4^n = \{\omega : X_n \in S\}$. By the triangle inequality, $S_1^n \supseteq S_2^n \cap S_3^n$. By continuity of g at X and openness of S, there exists γ_ϵ such that whenever $|X_n - X| < \gamma_\epsilon$, $|g(X_n) - g(X)| < \epsilon/2$ and $X_n \in S$. Letting, $S_5^n = \{\omega : |X_n - X| < \gamma_\epsilon\}$, we see that $S_5^n \subseteq S_3^n \cap S_4^n$. Since $X_n \xrightarrow{p} X$ and uniform convergence of g_n, there exists $N_{\delta,\epsilon}$ such that whenever $n > N_{\delta,\epsilon}$, $|g_n(X) - g(X)| < \epsilon/2$ for all $X \in S$ and $P(S_5^n) > 1 - \delta$. Thus, $n > N_{\delta,\epsilon}$ implies $S_4^n \subseteq S_2^n$. Consequently, $n > N_{\delta,\epsilon}$ implies $S_1^n \supseteq S_2^n \cap S_3^n \supseteq S_4^n \cap S_3^n \supseteq S_5^n$. Thus, $P(S_1^n) \geq P(S_5^n) > 1 - \delta$.

5. Show that $X_n \overset{as}{\to} X$ is equivalent to $P\left(\{\omega : \sup_{j \geq n} |X_j - X| \geq \epsilon\}\right) \to 0$ for all $\epsilon > 0$ as $n \to \infty$.

   Answer: For any $\epsilon > 0$ and $k \in \mathbb{N}$ let $A_k(\epsilon) = \{\omega : |X_k(\omega) - X(\omega)| > \epsilon\}$. If for all $n \in \mathbb{N}$ we have that $P\left(\cup_{k>n} A_k(\epsilon)\right) > 0$ then it must be that $X_n \overset{as}{\nrightarrow} X$. Consequently,

$$\begin{aligned} X_n \overset{as}{\to} X \quad &\Leftrightarrow \quad \lim_{n\to\infty} P\left(\cup_{n<k} A_k(\epsilon)\right) = 0 \\ &\Leftrightarrow \quad P\left(\{\omega : \sup_{j\geq n} |X_j - X| > \epsilon\}\right) \to 0 \text{ as } n \to \infty. \end{aligned}$$

6. Prove item 1 of Remark 7.1.
   Answer: For $\epsilon > 0$ we have that

$$\{\omega : |X_n + Y_n - X - Y| > \epsilon\} \subseteq \{\omega : |X_n - X| > \epsilon/2\} \cup \{\omega : |Y_n - Y| > \epsilon/2\}$$

   The probability of the events on the union on right-hand side go to zero as $n \to \infty$. By monotonicity of probability measures we have the results.

7. Let $n \in \mathbb{N}$ and $h_n > 0$ such that $h_n \to 0$ as $n \to \infty$. Show that if $\sum_{n=1}^{\infty} P(\{\omega : |X_n - X| \geq h_n\}) < \infty$ then $X_n \xrightarrow{p} X$.

   Answer: From question 5,

$$X_n \overset{as}{\to} X \quad \Leftrightarrow \quad \lim_{n\to\infty} P\left(\cup_{n<k} A_k(h_n)\right) = 0.$$

   But $P\left(\cup_{n<k} A_k(h_n)\right) \leq \sum_{k\geq n} P(A_k(\epsilon))$ and if $\sum_{n=1}^{\infty} P(\{\omega : |X_n - X| \geq h_n\}) < \infty$ then it must be that $\lim_{n\to\infty} \sum_{k\geq n} P(A_k(\epsilon)) = 0$. Since convergence almost surely implies convergence in probability, the proof is complete.

8. Show that if $Y_n \xrightarrow{d} Y$ then $Y_n = O_p(1)$.

Answer: Without loss of generality let $a > 0$. Provided that a and $-a$ are continuity points of F_Y, we can write that,$P(|Y_n| > a) \to P(|Y| > a)$ as $n \to \infty$. Hence, for every $\epsilon > 0$ there exists N_ϵ such that,

$$|P(|Y_n| > a) - P(|Y| > a)| < \epsilon \text{ for all } n \geq N_\epsilon$$

or

$$P(|Y| > a) - \epsilon < P(|Y_n| > a) < P(|Y| > a) + \epsilon.$$

We can choose a such that $P(|Y| > a) < \delta$ for any $\delta > 0$. Thus, $P(|Y_n| > a) < \delta + \epsilon$ for all $n \geq N_\epsilon$.

9. Let $g : S \subseteq \mathbb{R}$ be continuous on S, and X_t and X_s be random variables defined on $(\Omega, \mathcal{F}, P)$ taking values in S. Show that: a) if X_t is independent of X_s, then $g(X_t)$ is independent of $g(X_s)$; b) if X_t and X_s are identically distributed, then $g(X_t)$ and $g(X_s)$ are identically distributed.

   Answer: Let $Y_t = g(X_t)$ and $Y_s = g(X_s)$. g continuous assures that both Y_t and Y_s are random variables.

   a) $F_{Y_t,Y_s}(a, b) = P(S = \{\omega : Y_t \leq a \text{ and } Y_s \leq b\})$. Let $S_t = \{X_t(\omega) : Y_t(\omega) \leq a\}, S_s = \{X_s(\omega) : Y_s(\omega) \leq b\}$. Since, $S = S_t \cap S_s$ and by independence $P(S) = P(S_t)P(S_s)$ which implies $F_{Y_t,Y_s}(a, b) = F_{Y_t}(a)F_{Y_s}(b)$.

   b) $F_{Y_t}(a) = P(S_t) = P(\{X_s(\omega) : Y_s(\omega) \leq a\}) = F_{Y_s}(a)$.

10. Let $\{X_n\}$ be a sequence of independent random variables that converges in probability to a limit X. Show that X is almost surely a constant.

    Answer: Recall that if X is almost surely a constant, say c, $P(\{\omega : X(\omega) \neq c\}) = 0$. Then, the distribution function F associated with X is given by

$$F(x) = \begin{cases} 0, & \text{if } x < c \\ 1, & \text{if } x \geq c \end{cases} .$$

If X is not a constant, there exists a c and $0 < \epsilon < 1/2$ such that $P(X < c) > 2\epsilon$ and $P(X \leq c + \epsilon) < 1 - 2\epsilon$ or $P(X > c + \epsilon) > 2\epsilon$. Since $X_n \xrightarrow{p} X$ than $X_n \xrightarrow{d} X$. Consequently, for n sufficiently large and c a point of continuity of F we have

$$F(c) - \epsilon < F_n(c) < F(c) + \epsilon$$

which implies that $\epsilon < F_n(c)$. Also, $1 - F_n(c + \epsilon) > 1 - F(c + \epsilon) - \epsilon$ which implies $P(X_n > c + \epsilon) > P(X > c + \epsilon) - \epsilon > \epsilon$. Since $X_n \xrightarrow{p} X$, for n sufficiently large $P(\{\omega : |X_r - X_s| > \epsilon\}) < \epsilon^3$. Since $\{\omega : |X_r - X_s| > \epsilon\} = \{\omega : X_r - X_s > \epsilon\} \cup \{\omega : X_r - X_s < -\epsilon\}$ we note that if $X_r < c$ and $X_s > c + \epsilon$ then $X_r - X_s > \epsilon$ is equivalent to $X_r - X_s < -\epsilon$. Consequently,

$$P(\{\omega : |X_r - X_s| > \epsilon\}) \leq P(\{\omega : X_r < c \text{ and } X_s > c + \epsilon\}).$$

But since X_r and X_s are independent $P(\{\omega : X_r < c \text{ and } X_s > c + \epsilon\}) = P(\{\omega : X_r < c\})P(\{\omega : X_s > c + \epsilon\}) > \epsilon^2$. Hence,

$$\epsilon^3 > P(\{\omega : |X_r - X_s| > \epsilon\}) > \epsilon^2,$$

a contradiction.

11. Suppose $\frac{X_n - \mu}{\sigma_n} \xrightarrow{d} Z$ where the non-random sequence $\sigma_n \to 0$ as $n \to \infty$, and g is a function which is differentiable at μ. Then, show that $\frac{g(X_n) - g(\mu)}{g^{(1)}(\mu)\sigma_n} \xrightarrow{d} Z$.

    Answer: From question 2, if $Z_n \xrightarrow{d} Z$ then $Z_n = O_p(1)$. Let $Z_n = \frac{X_n - \mu}{\sigma_n}$ and write $X_n = \mu + \sigma_n Z_n = \mu + O_p(\sigma_n)$. By Taylor's Theorem

    $$\frac{1}{\sigma_n} g(X_n) - g(\mu) = g^{(1)}(\mu)\frac{(X_n - \mu)}{\sigma_n} + o_p(1).$$

    Since $\frac{X_n - \mu}{\sigma_n} \xrightarrow{d} Z$, we have the result.

12. Show that if $\{X_n\}_{n \in \mathbb{N}}$ and X are random variables defined on the same probability space and $r > s \geq 1$ and $X_n \xrightarrow{\mathcal{L}_r} X$, then $X_n \xrightarrow{\mathcal{L}_s} X$.

    Answer: For arbitrary W let $Z = |W|^s$, $Y = 1$ and $p = r/s$. Then, by Hölder's Inequality

    $$E|ZY| \leq \|Z\|_p \|Y\|_{p/(p-1)}.$$

Substituting Z and Y gives $E(|W|^s) \leq E(|W|^{sp})^{1/p} = E(|W|^{s\frac{r}{s}})^{s/r}$. Raising both sides to $1/s$ gives

$$E(|W|^s)^{1/s} \leq E(|W|^r)^{1/r}.$$

Setting $W = X_n - X$ and taking limits as $n \to \infty$ gives the result.

Chapter 8

Exercises

1. Let U and V be two points in an n-dimensional unit cube, i.e., $[0,1]^n$ and X_n be the Euclidean distance between these two points which are chosen independently and uniformly. Show that $\frac{X_n}{\sqrt{n}} \xrightarrow{p} \frac{1}{\sqrt{6}}$.

   Answer: Let $U' = \begin{pmatrix} U_1 & \cdots & U_n \end{pmatrix}$ and $V' = \begin{pmatrix} V_1 & \cdots & V_n \end{pmatrix}$. Then, $X_n = \left(\sum_{i=1}^n (U_i - V_i)^2\right)^{1/2}$ and we can write

$$\frac{1}{n}E(X_n^2) = \frac{1}{n}\sum_{i=1}^n E((U_i - V_i)^2) = \int_0^1 \int_0^1 (u-v)^2 dudv = 1/6$$

   where the last equality follows from routine integration. Then, since $E(|(U-V)^2|) = E((U-V)^2) < \infty$, by Kolmogorov's Law of Large Numbers

$$\frac{1}{n}X_n^2 = \frac{1}{n}\sum_{i=1}^n (U_i - V_i)^2 \xrightarrow{p} 1/6.$$

   Since, $f(x) = x^{1/2}$ is a continuous function $[0,\infty)$, by Slutsky Theorem if $\frac{1}{n}X_n^2 \xrightarrow{p} 1/6$ then $f\left(\frac{1}{n}X_n^2\right) \xrightarrow{p} f(1/6)$. Consequently,

$$\frac{1}{\sqrt{n}}X_n \xrightarrow{p} 1/\sqrt{6}.$$

2. Show that if $\{X_j\}_{j\in\mathbb{N}}$ be a sequence of random variables with $E(X_j) = 0$ and $\sum_{j=1}^\infty \frac{1}{a_j^p} E(|X_j|^p) < \infty$ for some $p \geq 1$ and a sequence of positive constants $\{a_j\}_{j\in\mathbb{N}}$. Then,

$$\sum_{j=1}^\infty P(|X_j| > a_j) < \infty \text{ and } \sum_{j=1}^\infty \frac{1}{a_j}|E(X_j I_{\{\omega:|X_j|\leq a_j\}})| < \infty.$$

Furthermore, for any $r \geq p$,

$$\sum_{j=1}^{\infty} \frac{1}{a_j^r} E(|X_j|^r I_{\{\omega:|X_j|\leq a_j\}}) < \infty.$$

Use this result to prove Theorem 8.4 in your class notes with convergence in probability.

Answer: Note that

$$P\left(\{\omega : |X_j| > a_j\}\right) = 1 - P(\{\omega : |X_j| \leq a_j\}) = \int_{\Omega} \left(1 - I_{\{\omega:|X_j|\leq a_j\}}\right) dP.$$

If $\omega \in \{\omega : |X_j| \leq a_j\}$, then $P\left(\{\omega : |X_j| > a_j\}\right) = 0$. If $|X_j| > a_j$, then $|X_j|^p > a_j^p$ and $|X_j|^p/a_j^p > 1$. Hence,

$$P\left(\{\omega : |X_j| > a_j\}\right) < \int_{\Omega} |X_j|^p/a_j^p dP = \frac{1}{a_j^p} E\left(|X_j|^p\right)$$

and

$$\sum_{j=1}^{\infty} P\left(\{\omega : |X_j| > a_j\}\right) < \sum_{j=1}^{\infty} \frac{1}{a_j^p} E\left(|X_j|^p\right) < \infty.$$

Now,

$$\begin{aligned}
\frac{1}{a_j}|E(X_j I_{\{\omega:|X_j|\leq a_j\}})| &= \frac{1}{a_j}|E(X_j) - E(X_j I_{\{\omega:|X_j|\leq a_j\}})|, \text{ since } E(X_j) = 0. \\
&\leq \frac{1}{a_j} E\left(|X_t|(1 - I_{\{\omega:|X_j|\leq a_j\}}\right) \\
&\leq \frac{1}{a_j^p} E\left(|X_j|^p(1 - I_{\{\omega:|X_j|\leq a_j\}}\right) \text{ since } \tfrac{|X_j|^p}{a_j^p} \geq \tfrac{|X_j|}{a_j} \text{ if } p \geq 1. \\
&\leq \frac{1}{a_j^p} E\left(|X_j|^p\right).
\end{aligned}$$

Hence,

$$\sum_{j=1}^{\infty} \frac{1}{a_j}|E(X_j I_{\{\omega:|X_j|\leq a_j\}})| < \sum_{j=1}^{\infty} \frac{1}{a_j^p} E\left(|X_j|^p\right) < \infty.$$

Lastly, if $|X_j| \leq a_j$ we have that $\frac{1}{a_j}|X_j| \leq 1$. Then, for $r \geq p \geq 1$

$$\frac{1}{a_j^r}|X_j|^r I_{\{\omega:|X_j|\leq a_j\}} \leq \frac{1}{a_j^p}|X_j|^p I_{\{\omega:|X_j|\leq a_j\}} \leq \frac{1}{a_j}|X_j| I_{\{\omega:|X_j|\leq a_j\}}$$

and

$$E\left(\frac{1}{a_j^r}|X_j|^r I_{\{\omega:|X_j|\leq a_j\}}\right) \leq E\left(\frac{1}{a_j^p}|X_j|^p I_{\{\omega:|X_j|\leq a_j\}}\right).$$

Hence,

$$\sum_{j=1}^{\infty} E\left(\frac{1}{a_j^r}|X_j|^r I_{\{\omega:|X_j|\le a_j\}}\right) < \infty.$$

In Theorem 8.4, the sequence of random variables $\{X_j\}_{j\in\mathbb{N}}$ is independent and has expectation μ_j. Hence, if $W_j := X_j - \mu_j$, we have $E(W_j) = 0$. Furthermore, in Theorem 8.4 it is assumed that for some $\delta > 0$

$$\sum_{j=1}^{\infty} \frac{E(|W_j|^{1+\delta})}{j^{1+\delta}} < \infty.$$

Now, note that for any $n \in \mathbb{N}$ we have $\sum_{j=1}^{n} \frac{E(|W_j|^{1+\delta})}{n^{1+\delta}} \le \sum_{j=1}^{n} \frac{E(|W_j|^{1+\delta})}{j^{1+\delta}}$ and

$$\lim_{n\to\infty} \sum_{j=1}^{n} \frac{E(|W_j|^{1+\delta})}{n^{1+\delta}} \le \lim_{n\to\infty} \sum_{j=1}^{n} \frac{E(|W_j|^{1+\delta})}{j^{1+\delta}} < \infty.$$

Now, in the first part of this answer, take $a_j = n$ for all j and for any $r > 1+\delta$. Then, we have

$$\sum_{j=1}^{\infty} P(|W_j| > n) < \infty \text{ and } \sum_{j=1}^{\infty} \frac{1}{n^r} E(|W_j|^r I_{\{\omega:|W_j|\le n\}}) < \infty.$$

Hence, taking $r = 2$ the conditions on Theorem 8.2 are met and we have

$$\frac{1}{n}\sum_{j=1}^{n} W_j - \frac{1}{n}\sum_{i=1}^{n} E\left(W_j I_{\{\omega:|W_j|\le n\}}\right) = \frac{1}{n}\sum_{j=1}^{n}(X_j - \mu_j) - \frac{1}{n}\sum_{i=1}^{n} E\left(W_j I_{\{\omega:|W_j|\le n\}}\right) = o_p(1).$$

But since $E(W_j) = 0$, we have $E\left(W_j I_{\{\omega:|W_j|\le n\}}\right) \to 0$ as $n \to \infty$. Thus, $\frac{1}{n}\sum_{j=1}^{n}(X_j - \mu_j) = o_p(1)$.

3. Let $\{X_i\}_{i=2,3,\cdots}$ be a sequence of independent random variables such that

$$P(X_i = i) = P(X_i = -i) = \frac{1}{2i\log i},\ P(X_i = 0) = 1 - \frac{1}{i\log i}$$

Show that $\frac{1}{n}\sum_{i=2}^{n} X_i \xrightarrow{p} 0$.

Answer: Let $S_n = \sum_{i=2}^{n} X_i$ and note that $E(X_i) = 0$. Hence, by independence

$$E(S_n^2) = \sum_{i=2}^{n} E(X_i^2) = \sum_{i=2}^{n} \frac{i}{\log i} \le \frac{n^2}{\log n}.$$

Hence, $V(S_n/n) = \frac{1}{n^2}V(S_n) = \frac{1}{n^2}E(S_n^2) \le \frac{1}{n^2}\frac{n^2}{\log n} = \frac{1}{\log n} \to 0$ as $n \to \infty$. Consequently, $\frac{1}{n}S_n \xrightarrow{p} 0$ by Chebyshev's inequality.

Chapter 9

Exercises

1. Assess the veracity of the following statement: "Since knowledge of X implies knowledge of $f(X)$, conditioning on X is the same as conditioning on $f(X)$. Hence, $E(Y|f(X)) = E(Y|X)$." Explain using mathematical arguments.

   Answer: The statement is false. Recall that conditioning on a random variable X means conditioning on the sub-σ-algebra generated by X, i.e., $X^{-1}(\mathcal{B})$. Hence, conditioning on $f(X)$ means conditioning on the sub-σ-algebra generated by $f(X)$, i.e., $X^{-1}(f^{-1}(\mathcal{B}))$ which is generally different from $X^{-1}(\mathcal{B})$. Take, for example, the following random vector: $(Y, X) : \Omega \to \mathbb{R}^2$ with $(Y(\omega), X(\omega)) = (1, -1)$ if $\omega \in E_1$ and $(Y(\omega), X(\omega)) = (2, -1)$ if $\omega \in E_2$, $(Y(\omega), X(\omega)) = (1, 1)$ if $\omega \in E_3$ and $(Y(\omega), X(\omega)) = (2, 1)$ if $\omega \in E_4$, with $P(E_j) = 1/6$ for $j = 1, 2$, $P(E_3) = 3/6$, $P(E_4) = 1/6$ and $\Omega = \cup_{j=1}^{4} E_j$ and $E_i \cup E_j = \emptyset$ for $i \neq j$. Now, let $f(X) = X^2$. Then,

   $$E(Y|X) = \begin{cases} 1.5 & \text{if } X = -1 \\ 5/4 & \text{if } X = 1 \end{cases} \quad \text{and } E(Y|X^2) = 8/6.$$

2. Let X and Y be independent random variables defined in the same probability space. Show that if $E(|Y|) < \infty$ then

   $$P\left(E(Y|X) = E(Y)\right) = 1.$$

   Answer: Let $\mathcal{F}_X$ be the σ-algebra generated by X. Let $E \in \mathcal{F}_X$ and note that there exists B such that $E = \{\omega : X(\omega) \in B\}$.

   $$\int_A Y dP = \int_\Omega Y I_A dP = \int_\Omega Y I_{X \in B} dP = E(Y I_{X \in B}) = E(Y)E(I_{X \in B})$$

where the last equality follows by independence. Now,

$$E(Y)E(I_{X\in B}) = E(Y)\int_\Omega I_{X\in B}dP = E(Y)\int_\Omega I_A dP = \int_A E(Y)dP.$$

Consequently, since A is arbitrary in $\mathcal{F}_X$

$$\int_A Y dP = \int_A E(Y)dP \text{ or } \int_A (Y - E(Y))dP = 0$$

By definition of conditional expectation we have that $E(Y|X) = E(Y)$ since A is arbitrary in $\mathcal{F}_X$.

3. Let $(\Omega, \mathcal{F}, P)$ be a probability space. The set of random variables $X : \Omega \to \mathbb{R}$ such that $\int_\Omega X^2 dP < \infty$ is denoted by $L^2(\Omega, \mathcal{F}, P)$. On this set $\|X\| = \left(\int_\Omega X^2 dP\right)^{1/2}$ is a norm and $< X, Y >= \int_\Omega XY dP$ is an inner product. If $\mathcal{G}$ is a σ-algebra and $\mathcal{G} \subset \mathcal{F}$, the conditional expectation of X with respect to $\mathcal{G}$, denoted by $E(X|\mathcal{G})$ is the orthogonal projection of X onto the closed subspace $L^2(\Omega, \mathcal{G}, P)$ of $L^2(\Omega, \mathcal{F}, P)$. Prove the following results:

   (a) For $X, Y \in L^2(\Omega, \mathcal{F}, P)$ we have $< E(X|\mathcal{G}), Y >=< E(Y|\mathcal{G})), X >=< E(X|\mathcal{G}), E(Y|\mathcal{G}) >$.

   (b) If $X = Y$ almost everywhere then $E(X|\mathcal{G}) = E(Y|\mathcal{G})$ almost everywhere.

   (c) For $X \in L^2(\Omega, \mathcal{G}, P)$ we have $E(X|\mathcal{G}) = X$.

   (d) If $\mathcal{H} \subset \mathcal{G}$ is a σ-algebra, then $E(E(X|\mathcal{G})|\mathcal{H}) = E(X|\mathcal{H})$.

   (e) If $Y \in L^2(\Omega, \mathcal{G}, P)$ and there exists a constant $C > 0$ such that $P(|Y| \geq C) = 0$, we have that $E(YX|\mathcal{G}) = YE(X|\mathcal{G})$.

   (f) If $\{Y_n\}_{n\in\mathbb{N}}$, $X \in L^2(\Omega, \mathcal{F}, P)$ and $\|Y_n - X\| \to 0$ as $n \to \infty$, then $E(Y_n|\mathcal{G}) \xrightarrow{p} E(X|\mathcal{G})$ as $n \to \infty$.

   Answer: (a) By definition of conditional expectation, for all measurable $s \in L^2(\Omega, \mathcal{G}, P)$,

$$E\left([X - E(X|\mathcal{G})]s\right) = 0 \iff E(Xs) = E\left(E(X|\mathcal{G})s\right). \tag{9.1}$$

   Since $E(Y|\mathcal{G}) \in L^2(\Omega, \mathcal{G}, P)$, we have $E(XE(Y|\mathcal{G})) = E\left(E(X|\mathcal{G})E(Y|\mathcal{G})\right)$. But by definition of the inner product the last equality is $< E(Y|\mathcal{G})), X >=< E(X|\mathcal{G}), E(Y|\mathcal{G}) >$. Similarly, changing X for Y in equation (9.1) we obtain $E(Ys) = E\left(E(Y|\mathcal{G})s\right)$. Letting, $s = E(X|\mathcal{G})$ we get $E(YE(X|\mathcal{G})) = E\left(E(Y|\mathcal{G})E(X|\mathcal{G})\right)$ and $E(YE(X|\mathcal{G})) = E(XE(Y|\mathcal{G}))$, which is equivalent to $< E(X|\mathcal{G}), Y >=< E(Y|\mathcal{G})), X >$.

(b) Let $y = E(Y|\mathcal{G})$ and $x = E(X|\mathcal{G})$. Then,

$$\begin{aligned}
(y-x)^2 &= (y-Y+Y-x)(y-x) = (y-Y)(y-x) + (Y-x)(y-x) \\
&= (y-Y)(y-x) + (Y-X)(y-x) + (X-x)(y-x)
\end{aligned}$$

But from item 1, $E(y-x)^2 := \|y-x\|^2 = E(Y-X)(y-x) \leq E|(Y-X)(y-x)| \leq \|Y-X\|\|y-x\|$, which gives $\|y-x\| \leq \|Y-X\|$. Lastly, if $X = Y$ almost everywhere, then $\|Y-X\| = 0$ and $x = y$ almost everywhere.

(c) Since $X \in L^2(\Omega, \mathcal{G}, P)$, it follows from the projection theorem that $E(X|\mathcal{G}) = X$.

(d) From item (a), we have $< E(E(X|\mathcal{G})|\mathcal{H}), Y >=< E(X|\mathcal{G}), E(Y|\mathcal{H}) >=< X, E(E(Y|\mathcal{H})|\mathcal{G}) >$. Since $E(Y|\mathcal{H}) \in L^2(\Omega, \mathcal{G}, P)$, we have that by item (c) $< X, E(E(Y|\mathcal{H})|\mathcal{G}) >=< X, E(Y|\mathcal{H}) >=< E(X|\mathcal{H}), Y >$. Hence, $E(E(X|\mathcal{G})|\mathcal{H}) = E(X|\mathcal{H})$ almost everywhere.

(e) Since $L^2(\Omega, \mathcal{G}, P)$ is a closed linear subspace of $L^2(\Omega, \mathcal{F}, P)$ and $E(\cdot|\mathcal{G})$ is a linear projector, any $X \in L^2(\Omega, \mathcal{F}, P)$ can be written as

$$X = E(X|\mathcal{G}) + (X - E(X|\mathcal{G})) \tag{9.2}$$

where $(X - E(X|\mathcal{G}))$ is orthogonal to any element of $L^2(\Omega, \mathcal{G}, P)$. Hence, (9.2) gives

$$XY = E(X|\mathcal{G})Y + (X - E(X|\mathcal{G}))Y. \tag{9.3}$$

Now, note that for any $s \in L^2(\Omega, \mathcal{G}, P)$ and $Y \in L^2(\Omega, \mathcal{G}, P)$ bounded almost everywhere, as assumed in the question, we have $sY \in L^2(\Omega, \mathcal{G}, P)$. Hence, $E((X - E(X|\mathcal{G}))sY) = 0$ and using (9.3) we have

$$E(sXY) = E(sE(X|\mathcal{G})Y) \iff E([XY - E(X|\mathcal{G})Y]s) = 0,$$

and the conclusion that $E(XY|\mathcal{G}) = E(X|\mathcal{G})Y$.

(f) From item (b)

$$\|E(Y_n|\mathcal{G}) - E(Z|\mathcal{G})\| \leq \|Y_n - Z\|.$$

Taking limits on both sides as $n \to \infty$ we obtain $\|E(Y_n|\mathcal{G}) - E(Z|\mathcal{G})\| \to 0$, since $\|Y_n - Z\| \to 0$ by assumption. That is, $E(Y_n|\mathcal{G})$ converges in quadratic mean to $E(Z|\mathcal{G})$. But by Chebyshev's inequality, convergence in quadratic mean implies convergence in probability. Hence, $E(Y_n|\mathcal{G}) \xrightarrow{p} E(Z|\mathcal{G})$.

4. Let $X, Y \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$ be random variables and assume that $E(Y|X) = aX$ where $a \in \mathbb{R}$.

   (a) Show that if $E(X^2) > 0$, $a = E(XY)/E(X^2)$.

   (b) If $\{(Y_i\ X_i)^T\}_{i=1}^n$ is a sequence of independent random vectors with components having the same distribution as $(Y\ X)^T$, show that

$$\frac{1}{n}\sum_{i=1}^n X_i^2 \xrightarrow{p} E(X^2) \text{ and } \frac{1}{n}\sum_{i=1}^n Y_i X_i \xrightarrow{p} E(XY).$$

   (c) Let $a_n = \left(\frac{1}{n}\sum_{i=1}^n X_i^2\right)^{-1} \frac{1}{n}\sum_{i=1}^n Y_i X_i$. Does $a_n \xrightarrow{p} a$? Can a_n be defined for all n? Explain.

Answer: (a) Note that $E(Y|X) = \underset{a}{\text{argmin}} \int_\Omega (Y - aX)^2 dP$. Now,

$$\int_\Omega (Y - aX)^2 dP = \int_\Omega Y^2 dP + a^2 \int_\Omega X^2 dP - 2a \int_\Omega XY dP,$$

$$\frac{d}{da}\int_\Omega (Y - aX)^2 dP = 2a \int_\Omega X^2 dP - 2\int_\Omega XY dP \text{ and } \frac{d^2}{da^2}\int_\Omega (Y - aX)^2 dP = 2\int_\Omega X^2 dP > 0.$$

Hence, setting the first derivative equal to zero gives, $E(Y|X)\int_\Omega X^2 dP = \int_\Omega XY dP \iff E(Y|X) = \frac{E(XY)}{E(X^2)}$.

(b) Since $X_i^2 = \begin{pmatrix} 0 & 1 \end{pmatrix}\begin{pmatrix} Y_i \\ X_i \end{pmatrix}\begin{pmatrix} Y_i \\ X_i \end{pmatrix}^T \begin{pmatrix} 0 & 1 \end{pmatrix}^T$ and $X_i Y_i = \begin{pmatrix} 1 & 0 \end{pmatrix}\begin{pmatrix} Y_i \\ X_i \end{pmatrix}\begin{pmatrix} Y_i \\ X_i \end{pmatrix}^T \begin{pmatrix} 0 & 1 \end{pmatrix}$ they are measurable function of $\begin{pmatrix} Y_i \\ X_i \end{pmatrix}$. Hence, $\{X_i^2\}_{i\in\mathbb{N}}$ and $\{X_iY_i\}_{i\in\mathbb{N}}$ are IID sequences. Since, $E(X_i^2) = E(X^2)$ and $E(X_iY_i) = E(XY)$ by the law of large numbers for IID random variables

$$\frac{1}{n}\sum_{i=1}^n X_i^2 \xrightarrow{p} E(X^2) > 0 \text{ and } \frac{1}{n}\sum_{i=1}^n Y_i X_i \xrightarrow{p} E(XY).$$

(c) To define a_n we need $\frac{1}{n}\sum_{i=1}^n X_i^2 > 0$ which is not assured from the assumptions. What can be said is that $\frac{1}{n}\sum_{i=1}^n X_i^2 \xrightarrow{p} E(X^2) > 0$. Hence, a_n exists in probability as $n \to \infty$.

5. Prove the following:

(a) If $Y \in \mathcal{L}(\Omega, \mathcal{F}, P)$ and $\mathcal{G} \subset \mathcal{F}$ is a σ-algebra, show that $|E(Y|\mathcal{G})| \leq E(|Y||\mathcal{G})$.

Answer: $|Y| = Y^+ - Y^-$ where Y^+, $Y^- \geq 0$. By linearity of conditional expectation

$$E(|Y||\mathcal{G}) = E(Y^+|\mathcal{G}) + E(Y^-|\mathcal{G})$$

and from Theorem 7.9 $E(Y^+|\mathcal{G}) \geq 0$, $E(Y^-|\mathcal{G}) \geq 0$. Hence,

$$\begin{aligned} |E(Y|\mathcal{G})| &= |E(Y^+|\mathcal{G}) - E(Y^-|\mathcal{G})| \leq |E(Y^+|\mathcal{G})| + |E(Y^-|\mathcal{G})| \\ &= E(Y^+|\mathcal{G}) + E(Y^-|\mathcal{G}) = E(|Y||\mathcal{G}) \end{aligned}$$

(b) Let c be a scalar constant and suppose $X = c$ almost surely. Show that $E(X|\mathcal{G}) = c$ almost surely.

Answer: It suffices to show that $\int_\Omega |c - E(X|\mathcal{G})|dP = 0$. Now,

$$\int_\Omega |c - E(X|\mathcal{G})|dP \quad = \quad \int_{c \geq E(X|\mathcal{G})} (c - E(X|\mathcal{G}))dP + \int_{c < E(X|\mathcal{G})} (E(X|\mathcal{G}) - c)dP.$$

Now, $\int_{c \geq E(X|\mathcal{G})} (c - E(X|\mathcal{G}))dP = \int_\Omega (c - E(X|\mathcal{G})) I_{\{c \geq E(X|\mathcal{G})\}} dP$. Now, since $E(X|\mathcal{G}) \in \mathcal{L}(\Omega, \mathcal{G}, P)$, $I_{\{c \geq E(X|\mathcal{G})\}}$ is $\mathcal{G}$-measurable. Hence, by the definition of conditional expectation

$$\int_{c \geq E(X|\mathcal{G})} (c - E(X|\mathcal{G}))dP = 0.$$

Similarly, $\int_{c < E(X|\mathcal{G})} (E(X|\mathcal{G}) - c)dP = 0$. Hence, $c = E(X|\mathcal{G})$ almost surely.

(c) If $Y \in \mathcal{L}(\Omega, \mathcal{F}, P)$ and $\mathcal{G} \subset \mathcal{F}$ is a σ-algebra, show that for $a > 0$

$$P\left(\{\omega : |Y(\omega)| \geq a\}|\mathcal{G}\right) \leq \frac{1}{a} E(|Y(\omega)||\mathcal{G}).$$

What is the definition of $P\left(\{\omega : |Y(\omega)| \geq a\}|\mathcal{G}\right)$? Is this a legitimate probability measure?

Answer: Note that $aI_{\{\omega : |Y(\omega)| \geq a\}} \leq |Y(\omega)|$ and

$$aE\left(I_{\{\omega:|Y(\omega)|\geq a\}}|\mathcal{G}\right) \leq E(|Y(\omega)||\mathcal{G}) \iff E\left(I_{\{\omega:|Y(\omega)|\geq a\}}|\mathcal{G}\right) \leq \frac{1}{a} E(|Y(\omega)||\mathcal{G}).$$

If we define $E\left(I_{\{\omega:|Y(\omega)|\geq a\}}|\mathcal{G}\right) := P\left(\{\omega : |Y(\omega)| \geq a\}|\mathcal{G}\right)$ we have

$$P\left(\{\omega : |Y(\omega)| \geq a\}|\mathcal{G}\right) \leq \frac{1}{a} E(|Y(\omega)||\mathcal{G}).$$

Now, to verify that $P(\cdot|\mathcal{G})$ is a legitimate probability measure note that, $E\left(I_{\Omega}|\mathcal{G}\right) = E\left(1|\mathcal{G}\right) = 1 = P\left(\Omega|\mathcal{G}\right)$ almost surely. Also, if $\{E_j\}_{j\in\mathbb{N}}$ is a countable collection of disjoint events $I_{\cup_{j\in\mathbb{N}}E_j} = \sum_{j\in\mathbb{N}} I_{E_j}$ and

$$P\left(\cup_{j\in\mathbb{N}}E_j|\mathcal{G}\right) = E\left(I_{\cup_{j\in\mathbb{N}}E_j}|\mathcal{G}\right) = E\left(\sum_{j\in\mathbb{N}} I_{E_j}|\mathcal{G}\right) = \sum_{j\in\mathbb{N}} E\left(I_{E_j}|\mathcal{G}\right) = \sum_{j\in\mathbb{N}} P\left(E_j|\mathcal{G}\right).$$

6. Let Y and X be random variables such that $Y, X \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$ and define $\varepsilon = Y - E(Y|X)$.

   (a) Show that $E(\varepsilon|X) = 0$ and $E(\varepsilon) = 0$.

   (b) Let $V(Y|X) = E(Y^2|X) - E(Y|X)^2$. Show that $V(Y|X) = V(\varepsilon|X)$, $V(\varepsilon) = E(V(Y|X))$;

   (c) $Cov(\varepsilon, h(X)) = 0$ for any function of X whose expectation exists.

   (d) Assume that $E(Y|X) = \alpha + \beta X$ where $\alpha, \beta \in \mathbb{R}$. Let $E(Y) = \mu_Y$, $E(X) = \mu_X$, $V(Y) = \sigma_Y^2$, $V(X) = \sigma_X^2$ and $\rho = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$. Show that,

   $$E(Y|X) = \mu_Y + \rho\sigma_Y \frac{X - \mu_X}{\sigma_X} \text{ and } E(V(Y|X)) = (1-\rho^2)\sigma_Y^2.$$

Answers:

(a) $E(\varepsilon|X) = E(Y - E(Y|X)|X) = E(Y|X) - E(Y|X) = 0$. By the law on iterated expectations $E(\varepsilon) = E(E(\varepsilon|X)) = 0$.

(b) $V(Y|X) = E((Y - E(Y|X))^2|X) = E(\varepsilon^2|X) = V(\varepsilon|X)$ since $E(\varepsilon|X) = 0$. Also, since $E(\varepsilon) = 0$ we have that $V(\varepsilon) = E(\varepsilon^2) = E(E(\varepsilon^2|X)) = E(V(\varepsilon|X)) = E(V(Y|X))$.

(c) $Cov(\varepsilon, h(X)) = E(\varepsilon h(X)) - E(\varepsilon)E(h(X)) = E(\varepsilon h(X))$ since $E(\varepsilon) = 0$. But by definition of conditional expectation

$$E(\varepsilon h(X)) = E(h(X)E(\varepsilon|X)) = 0 \text{ since } E(\varepsilon|X) = 0.$$

(d) First note that

$$\mu_Y = E(E(Y|X)) = E(\alpha + \beta X) = \alpha + \beta\mu_X. \tag{9.4}$$

Now, by definition of conditional expectation

$$E(XY) = E(X(\alpha + \beta X)) = \alpha\mu_X + \beta E(X^2) = \alpha\mu_X + \beta(\sigma_X^2 + \mu_X^2).$$

Also, $E(XY) = Cov(X,Y) + \mu_X\mu_Y = \rho\sigma_X\sigma_Y + \mu_X\mu_Y$. Then, we have

$$\alpha\mu_X + \beta(\sigma_X^2 + \mu_X^2) = \rho\sigma_X\sigma_Y + \mu_X\mu_Y. \tag{9.5}$$

Equations (9.4) and (9.5) form a system with two unknowns (α, β). The solution is given by,

$$\beta = \frac{\rho\sigma_Y}{\sigma_X} \text{ and } \alpha = \mu_Y - \mu_X\frac{\rho\sigma_Y}{\sigma_X}.$$

Substituting α and β into $E(Y|X) = \alpha + \beta X$ gives the desired result.

Lastly,

$$\begin{aligned}
\sigma_Y^2 := V(Y) = E\left(Y - E(Y)\right)^2 &= E\left(Y - E(Y|X) + E(Y|X) - E(Y)\right)^2 \\
&= E\left((Y - E(Y|X))^2\right) + E\left((E(Y|X) - E(Y))^2\right) \\
&+ 2E\left((Y - E(Y|X))(E(Y|X) - E(Y))\right) \\
&= E\left((Y - E(Y|X))^2\right) + V\left(E(Y|X)\right) + 2E\left(\varepsilon(E(Y|X) - E(Y))\right) \\
&= E(V(Y|X)) + V\left(E(Y|X)\right).
\end{aligned}$$

Consequently,

$$\begin{aligned}
E(V(Y|X)) &= \sigma_Y^2 - V\left(\mu_Y + \rho\sigma_Y\frac{X - \mu_X}{\sigma_X}\right) \\
&= \sigma_Y^2 - \rho^2\sigma_Y^2 = \sigma_Y^2(1 - \rho^2)
\end{aligned}$$

Chapter 10

Exercises

1. Suppose $\{X_i\}_{i=1,2,\cdots}$ is a sequence of independent and identically distributed random variables and $Y_i(x) = I_{\{\omega : X_i \leq x\}}$, where I_A is the indicator function of the set A. Now define

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} Y_i(x)$$

for fixed x. Obtain the asymptotic distribution of $\sqrt{n}(F_n(x) - F(x))$. You can use a Central Limit Theorem, but otherwise show all your work.

Answer: (3 points) First, note that $E(Y_i(x)) = P(\{\omega : X_i \leq x\}) = F(x)$ and $V(Y_i(x)) = F(x) - F(x)^2 = F(x)(1 - F(x))$.

$$\sqrt{n}(F_n(x) - F(x)) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}(Y_i(x) - E(Y_i(x)))\right).$$

Now, since the sequence is $\{Y_i(x)\}$ is IID, this is so because I_A is measurable, by Lévy's CLT

$$\frac{\frac{1}{n}\sum_{i=1}^{n}(Y_i(x) - E(Y_i(x)))}{\sqrt{\frac{F(x)(1-F(x))}{n}}} = \sqrt{n}\frac{\frac{1}{n}\sum_{i=1}^{n}(Y_i(x) - E(Y_i(x)))}{\sqrt{F(x)(1 - F(x))}} = \frac{\sqrt{n}(F_n(x) - F(x))}{\sqrt{F(x)(1 - F(x))}} \xrightarrow{d} Z \sim N(0,1)$$

2. Let $\{X_n\}_{n=1,2,\cdots}$ and $\{Y_n\}_{n=1,2,\cdots}$ be sequences of random variables defined on the same probability space. Suppose $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} Y$ and assume X_n and Y_n are independent for all n and X and Y are independent. Show that $X_n + Y_n \xrightarrow{d} X + Y$. Hint: use the characteristic function for a sum of independent random variables.

Answer: The characteristic function of $X_n + Y_n$ is given by

$$\phi_{X_n+Y_n}(t) = E(\exp it(X_n + Y_n)) = E(\exp it(X_n)\exp it(Y_n)) = E(\exp it(X_n))E(\exp it(Y_n)) = \phi_{X_n}(t)\phi_{Y_n}($$

where the next to last equality follows by independence of X_n and Y_n. Since, $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} Y$, it must be that $\phi_{X_n}(t) \rightarrow \phi_X(t)$ and $\phi_{Y_n}(t) \rightarrow \phi_Y(t)$. So,

$$\phi_{X_n+Y_n}(t) = \phi_{X_n}(t)\phi_{Y_n}(t) \rightarrow \phi_X(t)\phi_Y(t) = \phi_{X+Y}(t),$$

where the last equality follows from independence of X and Y. Thus, $X_n+Y_n \xrightarrow{d} X+Y$.

3. Let $\{X_i\}_{i=1,2,\cdots}$ be a sequence of independent and identically random variables with $E(X_i) = 1$ and $\sigma^2_{X_i} = \sigma^2 < \infty$. Show that if $S_n = \sum_{i=1}^n X_i$

$$\frac{2}{\sigma}\left(S_n^{1/2} - n^{1/2}\right) \xrightarrow{d} Z \sim N(0,1).$$

Answer: Note that,

$$\begin{aligned} \frac{2}{\sigma}(S_n - n) &= \frac{2}{\sigma}\left(S_n^{1/2} - n^{1/2}\right)\left(S_n^{1/2} + n^{1/2}\right) \\ &= \frac{2}{\sigma}\left(S_n^{1/2} - n^{1/2}\right) n^{1/2}\left((S_n/n)^{1/2} + 1\right) \end{aligned}$$

So that,

$$\frac{2}{\sigma}\sqrt{n}\left((S_n/n) - 1\right) = \frac{2}{\sigma}\left(S_n^{1/2} - n^{1/2}\right)\left((S_n/n)^{1/2} + 1\right)$$

and

$$\left((S_n/n)^{1/2} + 1\right)^{-1}\frac{2}{\sigma}\sqrt{n}\left((S_n/n) - 1\right) = \frac{2}{\sigma}\left(S_n^{1/2} - n^{1/2}\right).$$

Since, $\{X_i\}_{i=1,2,\cdots}$ is a sequence of independent and identically random variables with $E(X_i) = 1$, by Slutsky Theorem $\left((S_n/n)^{1/2} + 1\right)^{-1} \xrightarrow{p} 2^{-1}$ and since $\sigma^2_{X_i} = \sigma^2 < \infty$, by Lévy's CLT $\frac{1}{\sigma}\sqrt{n}\left((S_n/n) - 1\right) \xrightarrow{d} N(0,1)$. Hence, $\frac{2}{\sigma}\left(S_n^{1/2} - n^{1/2}\right) \xrightarrow{d} Z \sim N(0,1)$.