ST5215 Advanced Statistical Theory, Lecture 2

HUANG Dongming

National University of Singapore

13 Aug 2020

Review

Last time

- σ-field (Borel filed $\mathcal{B}$ on $\mathcal{R}$, product σ-field)

Review

Last time

- σ-field (Borel filed $\mathcal{B}$ on $\mathcal{R}$, product σ-field)
- Measure (counting measure, Lebesgue measure, σ-finite)

Review

Last time

- σ-field (Borel filed $\mathcal{B}$ on $\mathcal{R}$, product σ-field)
- Measure (counting measure, Lebesgue measure, σ-finite)
- Measurable functions (indicator functions, simple functions)

Review

Last time

- σ-field (Borel filed $\mathcal{B}$ on $\mathcal{R}$, product σ-field)
- Measure (counting measure, Lebesgue measure, σ-finite)
- Measurable functions (indicator functions, simple functions)
- Lebesgue Integration (3 steps)

Review

Last time

- σ-field (Borel filed $\mathcal{B}$ on $\mathcal{R}$, product σ-field)
- Measure (counting measure, Lebesgue measure, σ-finite)
- Measurable functions (indicator functions, simple functions)
- Lebesgue Integration (3 steps)

Some clarifications

- Cartesian product of two sets $A \times B = \{(a, b) : a \in A, b \in B\}$

Review

Last time

- σ-field (Borel filed $\mathcal{B}$ on $\mathcal{R}$, product σ-field)
- Measure (counting measure, Lebesgue measure, σ-finite)
- Measurable functions (indicator functions, simple functions)
- Lebesgue Integration (3 steps)

Some clarifications

- Cartesian product of two sets $A \times B = \{(a, b) : a \in A, b \in B\}$
- "Homework" mentioned in slides

Review

Last time

- σ-field (Borel filed $\mathcal{B}$ on $\mathcal{R}$, product σ-field)
- Measure (counting measure, Lebesgue measure, σ-finite)
- Measurable functions (indicator functions, simple functions)
- Lebesgue Integration (3 steps)

Some clarifications

- Cartesian product of two sets $A \times B = \{(a, b) : a \in A, b \in B\}$
- "Homework" mentioned in slides
- P9 of the old slide has some typos

Properties of measures

Let $(\Omega, \mathcal{F}, \nu)$ be a measure space. Assume all sets below are $\mathcal{F}$-measurable

- (Monotonicity) If $A \subset B$, then $\nu(A) \leq \nu(B)$.
- (Subadditivity) For any sequence of sets A_n's, $\nu(\cup_{n=1}^{\infty} A_n) \leq \sum_{n=1}^{\infty} \nu(A_n)$.
- (Continuity) If A_n's is an increasing sequence of sets, and $\lim_{n\to\infty} A_n := \cup_{n=1}^{\infty} A_n$, then $\nu(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} \nu(A_n)$.
- If A_n's is a decreasing sequence of sets and $\lim_{n\to\infty} A_n := \cap_{n=1}^{\infty} A_n$, and if $\nu(A_1) < \infty$, then $\nu(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} \nu(A_n)$.

Sets under mappings

For any function $f : \Omega \mapsto \Lambda$ and any index set I, we have

- $f(\bigcup_{\alpha \in I} A_\alpha) = \bigcup_{\alpha \in I} f(A_\alpha)$, where $A_\alpha \subset \Omega$

Sets under mappings

For any function $f : \Omega \mapsto \Lambda$ and any index set I, we have

- $f(\bigcup_{\alpha \in I} A_\alpha) = \bigcup_{\alpha \in I} f(A_\alpha)$, where $A_\alpha \subset \Omega$
- $f^{-1}(\bigcup_{\alpha \in I} B_\alpha) = \bigcup_{\alpha \in I} f^{-1}(B_\alpha)$, where $B_\alpha \subset \Lambda$

Sets under mappings

For any function $f : \Omega \mapsto \Lambda$ and any index set I, we have

- $f(\bigcup_{\alpha\in I} A_\alpha) = \bigcup_{\alpha\in I} f(A_\alpha)$, where $A_\alpha \subset \Omega$
- $f^{-1}(\bigcup_{\alpha\in I} B_\alpha) = \bigcup_{\alpha\in I} f^{-1}(B_\alpha)$, where $B_\alpha \subset \Lambda$
- $f^{-1}(B^c) = (f^{-1}(B))^c$, where $B \subset \Lambda$

Null set and a.e.

- If A is measurable and $\nu(A) = 0$, then A is called a *null set*.

Null set and a.e.

- If A is measurable and $\nu(A) = 0$, then A is called a *null set*.
- a.e. (almost everywhere) and a.s. statements: A statement holds ν-a.e. (or simply a.e.) if there is a null set A, the statement holds for all ω in A^c

Null set and a.e.

- If A is measurable and $\nu(A) = 0$, then A is called a *null set*.
- a.e. (almost everywhere) and a.s. statements: A statement holds ν-a.e. (or simply a.e.) if there is a null set A, the statement holds for all ω in A^c
- if ν is a probability, a.e. is often written as a.s. (almost surely)

Null set and a.e.

- If A is measurable and $\nu(A) = 0$, then A is called a *null set*.
- a.e. (almost everywhere) and a.s. statements: A statement holds ν-a.e. (or simply a.e.) if there is a null set A, the statement holds for all ω in A^c
- if ν is a probability, a.e. is often written as a.s. (almost surely) Sometimes, we says the statement holds with probability 1 (w.p. 1)

Null set and a.e.

- If A is measurable and $\nu(A) = 0$, then A is called a *null set*.
- a.e. (almost everywhere) and a.s. statements: A statement holds ν-a.e. (or simply a.e.) if there is a null set A, the statement holds for all ω in A^c
- if ν is a probability, a.e. is often written as a.s. (almost surely) Sometimes, we says the statement holds with probability 1 (w.p. 1)

Examples

- Let $f(x) = x^2$, then $f(x) > 0$ m-a.e. (recall: m denotes the Lebesgue measure on $\mathcal{R}$): $f(x) \leq 0$ iff $x = 0$, and $m(\{0\}) = 0$

Null set and a.e.

- If A is measurable and $\nu(A) = 0$, then A is called a *null set*.

- a.e. (almost everywhere) and a.s. statements: A statement holds ν-a.e. (or simply a.e.) if there is a null set A, the statement holds for all ω in A^c

- if ν is a probability, a.e. is often written as a.s. (almost surely) Sometimes, we says the statement holds with probability 1 (w.p. 1)

Examples

- Let $f(x) = x^2$, then $f(x) > 0$ m-a.e. (recall: m denotes the Lebesgue measure on $\mathcal{R}$): $f(x) \leq 0$ iff $x = 0$, and $m(\{0\}) = 0$
- Let $f_n(x) = 1/(1 + nx^2)$. Then $f_n \to 0$ m-a.e. : $f_n(x) \nrightarrow 0$ iff $x = 0$
- If $f = 0$ ν-a.e. then $\int f \, \mathrm{d}\nu = 0$: For any simple function g such that $0 \leq g \leq f_+$, easy to see $\int g \, \mathrm{d}\nu = 0$.

Properties

Assume the expectation of random variables below exists

1. Linearity: $\mathbb{E}(aX + bY) = a\mathbb{E}X + b\mathbb{E}Y$
2. $\mathbb{E}X$ is finite if and only if $\mathbb{E}|X|$ is finite
3. If $X \geq 0$ a.s., then $\mathbb{E}X \geq 0$
   - If $X \leq Y$ a.s., then $\mathbb{E}X \leq \mathbb{E}Y$
   - If $X = Y$ a.s., then $\mathbb{E}X = \mathbb{E}Y$
4. $|\mathbb{E}X| \leq \mathbb{E}|X|$
5. If $X \geq 0$ a.s., and $\mathbb{E}X = 0$, then $X = 0$ a.s.

Properties

Assume the expectation of random variables below exists

1. Linearity: $\mathbb{E}(aX + bY) = a\mathbb{E}X + b\mathbb{E}Y$
2. $\mathbb{E}X$ is finite if and only if $\mathbb{E}|X|$ is finite
3. If $X \geq 0$ a.s., then $\mathbb{E}X \geq 0$
   - If $X \leq Y$ a.s., then $\mathbb{E}X \leq \mathbb{E}Y$
   - If $X = Y$ a.s., then $\mathbb{E}X = \mathbb{E}Y$
4. $|\mathbb{E}X| \leq \mathbb{E}|X|$
5. If $X \geq 0$ a.s., and $\mathbb{E}X = 0$, then $X = 0$ a.s.

Standard method to prove (1) in 3 steps: simple functions, nonnegative functions, and then general functions. W.L.O.G. (with out loss of generality) assume $a, b > 0$

Properties

Assume the expectation of random variables below exists

1. Linearity: $\mathbb{E}(aX + bY) = a\mathbb{E}X + b\mathbb{E}Y$
2. $\mathbb{E}X$ is finite if and only if $\mathbb{E}|X|$ is finite
3. If $X \geq 0$ a.s., then $\mathbb{E}X \geq 0$
   - If $X \leq Y$ a.s., then $\mathbb{E}X \leq \mathbb{E}Y$
   - If $X = Y$ a.s., then $\mathbb{E}X = \mathbb{E}Y$
4. $|\mathbb{E}X| \leq \mathbb{E}|X|$
5. If $X \geq 0$ a.s., and $\mathbb{E}X = 0$, then $X = 0$ a.s.

Standard method to prove (1) in 3 steps: simple functions, nonnegative functions, and then general functions. W.L.O.G. (with out loss of generality) assume $a, b > 0$
(2 – 4) follows from the definition of integration.

Properties

Assume the expectation of random variables below exists

1. Linearity: $\mathbb{E}(aX + bY) = a\mathbb{E}X + b\mathbb{E}Y$
2. $\mathbb{E}X$ is finite if and only if $\mathbb{E}|X|$ is finite
3. If $X \geq 0$ a.s., then $\mathbb{E}X \geq 0$
   - If $X \leq Y$ a.s., then $\mathbb{E}X \leq \mathbb{E}Y$
   - If $X = Y$ a.s., then $\mathbb{E}X = \mathbb{E}Y$
4. $|\mathbb{E}X| \leq \mathbb{E}|X|$
5. If $X \geq 0$ a.s., and $\mathbb{E}X = 0$, then $X = 0$ a.s.

Standard method to prove (1) in 3 steps: simple functions, nonnegative functions, and then general functions. W.L.O.G. (with out loss of generality) assume $a, b > 0$

(2 – 4) follows from the definition of integration.

(5): Let $A = \{X > 0\}$ and $A_n = \{X > 1/n\}$.

Properties

Assume the expectation of random variables below exists

1. Linearity: $\mathbb{E}(aX + bY) = a\mathbb{E}X + b\mathbb{E}Y$
2. $\mathbb{E}X$ is finite if and only if $\mathbb{E}|X|$ is finite
3. If $X \geq 0$ a.s., then $\mathbb{E}X \geq 0$
   - If $X \leq Y$ a.s., then $\mathbb{E}X \leq \mathbb{E}Y$
   - If $X = Y$ a.s., then $\mathbb{E}X = \mathbb{E}Y$
4. $|\mathbb{E}X| \leq \mathbb{E}|X|$
5. If $X \geq 0$ a.s., and $\mathbb{E}X = 0$, then $X = 0$ a.s.

Standard method to prove (1) in 3 steps: simple functions, nonnegative functions, and then general functions. W.L.O.G. (with out loss of generality) assume $a, b > 0$
(2 – 4) follows from the definition of integration.
(5): Let $A = \{X > 0\}$ and $A_n = \{X > 1/n\}$. Since $X \geq XI_{A_n} \geq \frac{1}{n} I_{A_n}$,

Properties

Assume the expectation of random variables below exists

1. Linearity: $\mathbb{E}(aX + bY) = a\mathbb{E}X + b\mathbb{E}Y$
2. $\mathbb{E}X$ is finite if and only if $\mathbb{E}|X|$ is finite
3. If $X \geq 0$ a.s., then $\mathbb{E}X \geq 0$
   - If $X \leq Y$ a.s., then $\mathbb{E}X \leq \mathbb{E}Y$
   - If $X = Y$ a.s., then $\mathbb{E}X = \mathbb{E}Y$
4. $|\mathbb{E}X| \leq \mathbb{E}|X|$
5. If $X \geq 0$ a.s., and $\mathbb{E}X = 0$, then $X = 0$ a.s.

Standard method to prove (1) in 3 steps: simple functions, nonnegative functions, and then general functions. W.L.O.G. (with out loss of generality) assume $a, b > 0$
$(2-4)$ follows from the definition of integration.
(5): Let $A = \{X > 0\}$ and $A_n = \{X > 1/n\}$. Since $X \geq XI_{A_n} \geq \frac{1}{n} I_{A_n}$,we have $0 = \mathbb{E}X \geq \frac{1}{n}\mathbb{E}I_{A_n}$,

Properties

Assume the expectation of random variables below exists

1. Linearity: $\mathbb{E}(aX + bY) = a\mathbb{E}X + b\mathbb{E}Y$
2. $\mathbb{E}X$ is finite if and only if $\mathbb{E}|X|$ is finite
3. If $X \geq 0$ a.s., then $\mathbb{E}X \geq 0$
   - If $X \leq Y$ a.s., then $\mathbb{E}X \leq \mathbb{E}Y$
   - If $X = Y$ a.s., then $\mathbb{E}X = \mathbb{E}Y$
4. $|\mathbb{E}X| \leq \mathbb{E}|X|$
5. If $X \geq 0$ a.s., and $\mathbb{E}X = 0$, then $X = 0$ a.s.

Standard method to prove (1) in 3 steps: simple functions, nonnegative functions, and then general functions. W.L.O.G. (with out loss of generality) assume $a, b > 0$
$(2-4)$ follows from the definition of integration.
(5): Let $A = \{X > 0\}$ and $A_n = \{X > 1/n\}$. Since $X \geq XI_{A_n} \geq \frac{1}{n} I_{A_n}$,we have $0 = \mathbb{E}X \geq \frac{1}{n}\mathbb{E}I_{A_n}$,so $\mathbb{P}(A_n) = 0$.

Properties

Assume the expectation of random variables below exists

1. Linearity: $\mathbb{E}(aX + bY) = a\mathbb{E}X + b\mathbb{E}Y$
2. $\mathbb{E}X$ is finite if and only if $\mathbb{E}|X|$ is finite
3. If $X \geq 0$ a.s., then $\mathbb{E}X \geq 0$
   - If $X \leq Y$ a.s., then $\mathbb{E}X \leq \mathbb{E}Y$
   - If $X = Y$ a.s., then $\mathbb{E}X = \mathbb{E}Y$
4. $|\mathbb{E}X| \leq \mathbb{E}|X|$
5. If $X \geq 0$ a.s., and $\mathbb{E}X = 0$, then $X = 0$ a.s.

Standard method to prove (1) in 3 steps: simple functions, nonnegative functions, and then general functions. W.L.O.G. (with out loss of generality) assume $a, b > 0$
(2 – 4) follows from the definition of integration.
(5): Let $A = \{X > 0\}$ and $A_n = \{X > 1/n\}$. Since $X \geq XI_{A_n} \geq \frac{1}{n} I_{A_n}$,we have $0 = \mathbb{E}X \geq \frac{1}{n}\mathbb{E}I_{A_n}$,so $\mathbb{P}(A_n) = 0$.Since $A_n \uparrow A$, we have $\mathbb{P}(A) = 0$.

Convergence Theorems for Integration

Theorem

Let $\{f_n\}_{n=1}^{\infty}$ *be a sequence of Borel functions on* $(\Omega, \mathcal{F}, \nu)$.

Convergence Theorems for Integration

Theorem

Let $\{f_n\}_{n=1}^{\infty}$ *be a sequence of Borel functions on* $(\Omega, \mathcal{F}, \nu)$*.*

- *Monotone convergence theorem: If* $0 \leq f_1 \leq \cdots$ *and* $\lim_n f_n = f$ *a.e., then*

$$\int \lim_n f_n \, \mathrm{d}\nu = \lim_n \int f_n \, \mathrm{d}\nu. \tag{1}$$

Convergence Theorems for Integration

Theorem

Let $\{f_n\}_{n=1}^{\infty}$ *be a sequence of Borel functions on* $(\Omega, \mathcal{F}, \nu)$*.*

- *Monotone convergence theorem: If* $0 \le f_1 \le \cdots$ *and* $\lim_n f_n = f$ *a.e., then*

$$\int \lim_n f_n \, \mathrm{d}\nu = \lim_n \int f_n \, \mathrm{d}\nu. \tag{1}$$

- *Fatou's lemma: If* $f_n \ge 0$*, then*

$$\int \liminf_n \; f_n \, \mathrm{d}\nu \le \liminf_n \int f_n \, \mathrm{d}\nu. \tag{2}$$

Convergence Theorems for Integration

Theorem

Let $\{f_n\}_{n=1}^{\infty}$ *be a sequence of Borel functions on* $(\Omega, \mathcal{F}, \nu)$.

- *Monotone convergence theorem: If* $0 \le f_1 \le \cdots$ *and* $\lim_n f_n = f$ *a.e., then*
$$\int \lim_n f_n \, \mathrm{d}\nu = \lim_n \int f_n \, \mathrm{d}\nu. \tag{1}$$
- *Fatou's lemma: If* $f_n \ge 0$*, then*
$$\int \liminf_n \; f_n \, \mathrm{d}\nu \le \liminf_n \int f_n \, \mathrm{d}\nu. \tag{2}$$
- *Dominated convergence theorem: If* $\lim_{n\to\infty} f_n = f$ *a.e. and there exists an integrable function* g *such that* $|f_n| \le g$ *a.e., then*
$$\int \lim_n f_n \, \mathrm{d}v = \lim_n \int f_n \, \mathrm{d}\nu. \tag{3}$$

Without the dominating integrable function

On the measure space $(\mathcal{R}, \mathcal{B}, m)$

- Let $f_n(x) = I_{(n,n+1)}(x)$.
- Let $f(x) = 0$
- $f_n \to f$ a.e.

Without the dominating integrable function

On the measure space $(\mathcal{R}, \mathcal{B}, m)$

- Let $f_n(x) = I_{(n,n+1)}(x)$.
- Let $f(x) = 0$
- $f_n \to f$ a.e.
- but $\int f_n \, \mathrm{d}m = 1$ and $\int f \, \mathrm{d}m = 0$,

Without the dominating integrable function

On the measure space $(\mathcal{R}, \mathcal{B}, m)$

- Let $f_n(x) = I_{(n,n+1)}(x)$.
- Let $f(x) = 0$
- $f_n \to f$ a.e.
- but $\int f_n \, \mathrm{d}m = 1$ and $\int f \, \mathrm{d}m = 0$, so $\lim_n \int f_n \, \mathrm{d}m \neq \int f \, \mathrm{d}m$.

Change of variables

- Let $(\Omega, \mathcal{F}, \nu)$ be a measure space, and $(\Lambda, \mathcal{G})$ be a measurable space

Change of variables

- Let $(\Omega, \mathcal{F}, \nu)$ be a measure space, and $(\Lambda, \mathcal{G})$ be a measurable space
- Let f be measurable from $(\Omega, \mathcal{F})$ to $(\Lambda, \mathcal{G})$.

Change of variables

- Let $(\Omega, \mathcal{F}, \nu)$ be a measure space, and $(\Lambda, \mathcal{G})$ be a measurable space
- Let f be measurable from $(\Omega, \mathcal{F})$ to $(\Lambda, \mathcal{G})$.
- Then f induces a measure on Λ, denoted by $\nu \circ f^{-1}$ and defined by

$$\nu \circ f^{-1}(B) = \nu(f^{-1}(B)), \qquad \forall B \in \mathcal{G}. \tag{4}$$

Change of variables

- Let $(\Omega, \mathcal{F}, \nu)$ be a measure space, and $(\Lambda, \mathcal{G})$ be a measurable space
- Let f be measurable from $(\Omega, \mathcal{F})$ to $(\Lambda, \mathcal{G})$.
- Then f induces a measure on Λ, denoted by $\nu \circ f^{-1}$ and defined by

$$\nu \circ f^{-1}(B) = \nu(f^{-1}(B)), \qquad \forall B \in \mathcal{G}. \tag{4}$$

Theorem (Change of variables)

Suppose g is a Borel function on $(\Lambda, \mathcal{G})$. The integral of Borel function $g \circ f$ is computed via

$$\int_\Omega g \circ f \, \mathrm{d}\nu = \int_\Lambda g \, \mathrm{d}(\nu \circ f^{-1}). \tag{5}$$

If either integral exists, then so does the other, and the two are the same.

If $\nu = P$ is a probability measure and $\Lambda = \mathcal{R}$ and $f = X$ is a random variable, then $P \circ X^{-1}$ is often denoted by P_X.

If $\nu = P$ is a probability measure and $\Lambda = \mathcal{R}$ and $f = X$ is a random variable, then $P \circ X^{-1}$ is often denoted by P_X.

- P_X is called the law (or the distribution) of X

If $\nu = P$ is a probability measure and $\Lambda = \mathcal{R}$ and $f = X$ is a random variable, then $P \circ X^{-1}$ is often denoted by P_X.

- P_X is called the law (or the distribution) of X
- The cumulative distribution function (CDF, c.d.f.) of X is denoted by F_X and defined by $F_X(x) = P(X \leq x)$.

If $\nu = P$ is a probability measure and $\Lambda = \mathcal{R}$ and $f = X$ is a random variable, then $P \circ X^{-1}$ is often denoted by P_X.

- P_X is called the law (or the distribution) of X
- The cumulative distribution function (CDF, c.d.f.) of X is denoted by F_X and defined by $F_X(x) = P(X \leq x)$.
- Using the theorem, we have $\mathbb{E}X = \int_\Omega X(\omega)\, \mathrm{d}P(\omega) = \int_{\mathcal{R}} x\, \mathrm{d}P_X(x)$

Fubini's Theorem

Let ν_i be a σ-finite measure on $(\Omega_i, \mathcal{F}_i)$, $i = 1, 2$, and let f be a Borel function on $(\Omega_1 \times \Omega_2, \sigma(\mathcal{F}_1 \times \mathcal{F}_2))$.

Fubini's Theorem

Let ν_i be a σ-finite measure on $(\Omega_i, \mathcal{F}_i)$, $i = 1, 2$, and let f be a Borel function on $(\Omega_1 \times \Omega_2, \sigma(\mathcal{F}_1 \times \mathcal{F}_2))$.
Suppose that either $f \geq 0$ or $\int |f| \, \mathrm{d}(\nu_1 \times \nu_2) < \infty$.

Fubini's Theorem

Let ν_i be a σ-finite measure on $(\Omega_i, \mathcal{F}_i)$, $i = 1, 2$, and let f be a Borel function on $(\Omega_1 \times \Omega_2, \sigma(\mathcal{F}_1 \times \mathcal{F}_2))$.
Suppose that either $f \geq 0$ or $\int |f| \, \mathrm{d}(\nu_1 \times \nu_2) < \infty$. Then

- $$g(\omega_2) = \int_{\Omega_1} f(\omega_1, \omega_2) \, \mathrm{d}\nu_1(\omega_1) \tag{6}$$

  exists ν_2-a.e., g is a Borel function on $(\Omega_2, \mathcal{F}_2)$, and the integral of g exists.

Fubini's Theorem

Let ν_i be a σ-finite measure on $(\Omega_i, \mathcal{F}_i)$, $i = 1, 2$, and let f be a Borel function on $(\Omega_1 \times \Omega_2, \sigma(\mathcal{F}_1 \times \mathcal{F}_2))$.
Suppose that either $f \geq 0$ or $\int |f| \, \mathrm{d}(\nu_1 \times \nu_2) < \infty$. Then

- $$g(\omega_2) = \int_{\Omega_1} f(\omega_1, \omega_2) \, \mathrm{d}\nu_1(\omega_1) \tag{6}$$

  exists ν_2-a.e., g is a Borel function on $(\Omega_2, \mathcal{F}_2)$, and the integral of g exists.

- $$\begin{aligned} \int_{\Omega_1 \times \Omega_2} f \, \mathrm{d}(\nu_1 \times \nu_2) &= \int_{\Omega_2} \left[\int_{\Omega_1} f(\omega_1, \omega_2) \, \mathrm{d}\nu_1(\omega_1) \right] \, \mathrm{d}\nu_2(\omega_2) \\ &= \int_{\Omega_1} \left[\int_{\Omega_2} f(\omega_1, \omega_2) \, \mathrm{d}\nu_2(\omega_2) \right] \, \mathrm{d}\nu_1(\omega_1). \end{aligned}$$

Fubini's Theorem

Example

Let $\Omega_1 = \Omega_2 = \{1, 2, \ldots\}$,and $\nu_1 = \nu_2$ be the counting measure. A function a on $\Omega_1 \times \Omega_2$ defines a double sequence, and $a(i, j)$ is often denoted by a_{ij}.

Fubini's Theorem

Example

Let $\Omega_1 = \Omega_2 = \{1, 2, \ldots\}$,and $\nu_1 = \nu_2$ be the counting measure. A function a on $\Omega_1 \times \Omega_2$ defines a double sequence, and $a(i,j)$ is often denoted by a_{ij}.

If either $a_{ij} \geq 0$ for all i, j or $\int |a| \, \mathrm{d}(\nu_1 \times \nu_2) = \sum_{ij} |a_{ij}| < \infty$, then

$$\sum_{ij} a_{ij} = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} a_{ij} = \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} a_{ij}. \tag{7}$$

Exercise

Assume X is a non-negative r.v. on $(\Omega, \mathcal{F}, P)$.

Exercise

Assume X is a non-negative r.v. on $(\Omega, \mathcal{F}, P)$.
Show: $\mathbb{E}X = \int_0^\infty P(X > t)\, \mathrm{d}m(t)$.

Exercise

Assume X is a non-negative r.v. on $(\Omega, \mathcal{F}, P)$.
Show: $\mathbb{E}X = \int_0^\infty P(X > t)\, \mathrm{d}m(t)$.
Proof:

- $X = \int_0^\infty I_{X>t}\, \mathrm{d}m(t)$

Exercise

Assume X is a non-negative r.v. on $(\Omega, \mathcal{F}, P)$.
Show: $\mathbb{E}X = \int_0^\infty P(X > t)\, \mathrm{d}m(t)$.
Proof:

- $X = \int_0^\infty I_{X>t}\, \mathrm{d}m(t)$
- By Fubini's theorem

$$\begin{aligned}
\mathbb{E}X &= \int_\Omega \left[\int_0^\infty I_{X>t}\, \mathrm{d}m(t)\right]\, \mathrm{d}P(\omega) \\
&= \int_0^\infty \left[\int_\Omega I_{X>t}\, \mathrm{d}P(\omega)\right]\, \mathrm{d}m(t) \\
&= \int_0^\infty P(X > t)\, \mathrm{d}m(t)
\end{aligned}$$

Absolutely continuity

Let λ and ν be two measures on a measurable space $(\Omega, \mathcal{F})$. We say λ is *absolutely continuous* w.r.t. ν and write $\lambda \ll \nu$ iff

$$\nu(A) = 0 \qquad \text{implies} \qquad \lambda(A) = 0. \tag{8}$$

Absolutely continuity

Let λ and ν be two measures on a measurable space $(\Omega, \mathcal{F})$. We say λ is *absolutely continuous* w.r.t. ν and write $\lambda \ll \nu$ iff

$$\nu(A) = 0 \qquad \text{implies} \qquad \lambda(A) = 0. \tag{8}$$

Exercise: Show that if the measure λ defined by

$$\lambda(A) := \int_A f \, \mathrm{d}\nu, A \in \mathcal{F} \tag{9}$$

for a non-negative Borel function f, then $\lambda \ll \nu$.

Absolutely continuity

Let λ and ν be two measures on a measurable space $(\Omega, \mathcal{F})$. We say λ is *absolutely continuous* w.r.t. ν and write $\lambda \ll \nu$ iff

$$\nu(A) = 0 \qquad \text{implies} \qquad \lambda(A) = 0. \tag{8}$$

Exercise: Show that if the measure λ defined by

$$\lambda(A) := \int_A f \, \mathrm{d}\nu, A \in \mathcal{F} \tag{9}$$

for a non-negative Borel function f, then $\lambda \ll \nu$.
Proof: It is easy to check λ is a measure.

Absolutely continuity

Let λ and ν be two measures on a measurable space $(\Omega, \mathcal{F})$. We say λ is *absolutely continuous* w.r.t. ν and write $\lambda \ll \nu$ iff

$$\nu(A) = 0 \qquad \text{implies} \qquad \lambda(A) = 0. \tag{8}$$

Exercise: Show that if the measure λ defined by

$$\lambda(A) := \int_A f \, \mathrm{d}\nu, A \in \mathcal{F} \tag{9}$$

for a non-negative Borel function f, then $\lambda \ll \nu$.
Proof: It is easy to check λ is a measure.
Suppose $A \in \mathcal{F}$ and $\nu(A) = 0$.

Absolutely continuity

Let λ and ν be two measures on a measurable space $(\Omega, \mathcal{F})$. We say λ is *absolutely continuous* w.r.t. ν and write $\lambda \ll \nu$ iff

$$\nu(A) = 0 \qquad \text{implies} \qquad \lambda(A) = 0. \tag{8}$$

Exercise: Show that if the measure λ defined by

$$\lambda(A) := \int_A f \, \mathrm{d}\nu, A \in \mathcal{F} \tag{9}$$

for a non-negative Borel function f, then $\lambda \ll \nu$.
Proof: It is easy to check λ is a measure.
Suppose $A \in \mathcal{F}$ and $\nu(A) = 0$. Since $I_A f = 0 \;\; \nu-$a.e.,

Absolutely continuity

Let λ and ν be two measures on a measurable space $(\Omega, \mathcal{F})$. We say λ is *absolutely continuous* w.r.t. ν and write $\lambda \ll \nu$ iff

$$\nu(A) = 0 \qquad \text{implies} \qquad \lambda(A) = 0. \tag{8}$$

Exercise: Show that if the measure λ defined by

$$\lambda(A) := \int_A f \, \mathrm{d}\nu, A \in \mathcal{F} \tag{9}$$

for a non-negative Borel function f, then $\lambda \ll \nu$.
Proof: It is easy to check λ is a measure.
Suppose $A \in \mathcal{F}$ and $\nu(A) = 0$. Since $I_A f = 0 \;\; \nu-$a.e.,we have $\int_A f \, \mathrm{d}\nu = \int I_A f \, \mathrm{d}\nu = 0$, i.e., $\lambda(A) = 0$. $\square$

Absolutely continuity

Let λ and ν be two measures on a measurable space $(\Omega, \mathcal{F})$. We say λ is *absolutely continuous* w.r.t. ν and write $\lambda \ll \nu$ iff

$$\nu(A) = 0 \qquad \text{implies} \qquad \lambda(A) = 0. \tag{8}$$

Exercise: Show that if the measure λ defined by

$$\lambda(A) := \int_A f \, \mathrm{d}\nu, A \in \mathcal{F} \tag{9}$$

for a non-negative Borel function f, then $\lambda \ll \nu$.
Proof: It is easy to check λ is a measure.
Suppose $A \in \mathcal{F}$ and $\nu(A) = 0$. Since $I_A f = 0$ $\nu-$a.e.,we have
$\int_A f \, \mathrm{d}\nu = \int I_A f \, \mathrm{d}\nu = 0$, i.e., $\lambda(A) = 0$. $\square$

Conversely, if $\lambda \ll \nu$, can we find a Borel function f s.t. $\lambda(A) = \int_A f \, \mathrm{d}\nu$?

Absolutely continuity

Let λ and ν be two measures on a measurable space $(\Omega, \mathcal{F})$. We say λ is *absolutely continuous* w.r.t. ν and write $\lambda \ll \nu$ iff

$$\nu(A) = 0 \qquad \text{implies} \qquad \lambda(A) = 0. \tag{8}$$

Exercise: Show that if the measure λ defined by

$$\lambda(A) := \int_A f \,\mathrm{d}\nu, A \in \mathcal{F} \tag{9}$$

for a non-negative Borel function f, then $\lambda \ll \nu$.
Proof: It is easy to check λ is a measure.
Suppose $A \in \mathcal{F}$ and $\nu(A) = 0$. Since $I_A f = 0 \;\; \nu-$a.e.,we have $\int_A f \,\mathrm{d}\nu = \int I_A f \,\mathrm{d}\nu = 0$, i.e., $\lambda(A) = 0$. $\square$

Conversely, if $\lambda \ll \nu$, can we find a Borel function f s.t. $\lambda(A) = \int_A f \,\mathrm{d}\nu$?
Yes, if ν is σ-finite.

Theorem (Radon-Nikodym)

Let ν and λ be two measures on $(\Omega, \mathcal{F})$ and ν be σ-finite. If $\lambda \ll \nu$, then there exists a nonnegative Borel function f on Ω such that

$$\lambda(A) = \int_A f \, \mathrm{d}\nu, \qquad A \in \mathcal{F}. \tag{10}$$

In addition, f is unique ν-a.e., i.e., if another g also satisfies this equation then $f = g$ ν-a.e.

Theorem (Radon-Nikodym)

Let ν and λ be two measures on $(\Omega, \mathcal{F})$ and ν be σ-finite. If $\lambda \ll \nu$, then there exists a nonnegative Borel function f on Ω such that

$$\lambda(A) = \int_A f \, \mathrm{d}\nu, \qquad A \in \mathcal{F}. \tag{10}$$

In addition, f is unique ν-a.e., i.e., if another g also satisfies this equation then $f = g$ ν-a.e.

- The function f above is called the *Radon-Nikodym derivative* or density of λ w.r.t. ν, and denoted by $\frac{\mathrm{d}\lambda}{\mathrm{d}\nu}$.

Theorem (Radon-Nikodym)

Let ν and λ be two measures on $(\Omega, \mathcal{F})$ and ν be σ-finite. If $\lambda \ll \nu$, then there exists a nonnegative Borel function f on Ω such that

$$\lambda(A) = \int_A f \, \mathrm{d}\nu, \qquad A \in \mathcal{F}. \tag{10}$$

In addition, f is unique ν-a.e., i.e., if another g also satisfies this equation then $f = g$ ν-a.e.

- The function f above is called the *Radon-Nikodym derivative* or density of λ w.r.t. ν, and denoted by $\frac{\mathrm{d}\lambda}{\mathrm{d}\nu}$.
- The proof of existence is beyond this course.

Theorem (Radon-Nikodym)

Let ν and λ be two measures on $(\Omega, \mathcal{F})$ and ν be σ-finite. If $\lambda \ll \nu$, then there exists a nonnegative Borel function f on Ω such that

$$\lambda(A) = \int_A f \, \mathrm{d}\nu, \qquad A \in \mathcal{F}. \tag{10}$$

In addition, f is unique ν-a.e., i.e., if another g also satisfies this equation then $f = g$ ν-a.e.

- The function f above is called the *Radon-Nikodym derivative* or density of λ w.r.t. ν, and denoted by $\frac{\mathrm{d}\lambda}{\mathrm{d}\nu}$.
- The proof of existence is beyond this course.
- The uniqueness can be proved: Let $A = \{f \geq g\}$.

Theorem (Radon-Nikodym)

Let ν and λ be two measures on $(\Omega, \mathcal{F})$ and ν be σ-finite. If $\lambda \ll \nu$, then there exists a nonnegative Borel function f on Ω such that

$$\lambda(A) = \int_A f \, \mathrm{d}\nu, \qquad A \in \mathcal{F}. \tag{10}$$

In addition, f is unique ν-a.e., i.e., if another g also satisfies this equation then $f = g$ ν-a.e.

- The function f above is called the *Radon-Nikodym derivative* or density of λ w.r.t. ν, and denoted by $\frac{\mathrm{d}\lambda}{\mathrm{d}\nu}$.
- The proof of existence is beyond this course.
- The uniqueness can be proved: Let $A = \{f \geq g\}$. Note that $I_A \cdot (f - g)$ is nonnegative and $\int I_A \cdot (f - g) \, \mathrm{d}\nu = 0$,

Theorem (Radon-Nikodym)

Let ν and λ be two measures on $(\Omega, \mathcal{F})$ and ν be σ-finite. If $\lambda \ll \nu$, then there exists a nonnegative Borel function f on Ω such that

$$\lambda(A) = \int_A f \, \mathrm{d}\nu, \qquad A \in \mathcal{F}. \tag{10}$$

In addition, f is unique ν-a.e., i.e., if another g also satisfies this equation then $f = g$ ν-a.e.

- The function f above is called the *Radon-Nikodym derivative* or density of λ w.r.t. ν, and denoted by $\frac{\mathrm{d}\lambda}{\mathrm{d}\nu}$.
- The proof of existence is beyond this course.
- The uniqueness can be proved: Let $A = \{f \geq g\}$. Note that $I_A \cdot (f - g)$ is nonnegative and $\int I_A \cdot (f - g) \, \mathrm{d}\nu = 0$, so $I_A \cdot (f - g) = 0$ ν-a.e.

Theorem (Radon-Nikodym)

Let ν and λ be two measures on $(\Omega, \mathcal{F})$ and ν be σ-finite. If $\lambda \ll \nu$, then there exists a nonnegative Borel function f on Ω such that

$$\lambda(A) = \int_A f \, \mathrm{d}\nu, \qquad A \in \mathcal{F}. \tag{10}$$

In addition, f is unique ν-a.e., i.e., if another g also satisfies this equation then $f = g$ ν-a.e.

- The function f above is called the *Radon-Nikodym derivative* or density of λ w.r.t. ν, and denoted by $\frac{\mathrm{d}\lambda}{\mathrm{d}\nu}$.
- The proof of existence is beyond this course.
- The uniqueness can be proved: Let $A = \{f \geq g\}$. Note that $I_A \cdot (f - g)$ is nonnegative and $\int I_A \cdot (f - g) \, \mathrm{d}\nu = 0$, so $I_A \cdot (f - g) = 0$ ν-a.e.
  Similarly, we have $I_{f<g} \cdot (g - f) = 0$ ν-a.e..

Theorem (Radon-Nikodym)

Let ν and λ be two measures on $(\Omega, \mathcal{F})$ and ν be σ-finite. If $\lambda \ll \nu$, then there exists a nonnegative Borel function f on Ω such that

$$\lambda(A) = \int_A f \, \mathrm{d}\nu, \qquad A \in \mathcal{F}. \tag{10}$$

In addition, f is unique ν-a.e., i.e., if another g also satisfies this equation then $f = g$ ν-a.e.

- The function f above is called the *Radon-Nikodym derivative* or density of λ w.r.t. ν, and denoted by $\frac{\mathrm{d}\lambda}{\mathrm{d}\nu}$.
- The proof of existence is beyond this course.
- The uniqueness can be proved: Let $A = \{f \geq g\}$. Note that $I_A \cdot (f - g)$ is nonnegative and $\int I_A \cdot (f - g) \, \mathrm{d}\nu = 0$, so $I_A \cdot (f - g) = 0$ ν-a.e.
  Similarly, we have $I_{f<g} \cdot (g - f) = 0$ ν-a.e..
  Subtracting yields $f - g = 0$ ν-a.e.

Probability density function (PDF)

Example (Probability density)

Let $\lambda = P$ be a probability measure on $\mathcal{R}$ corresponding to a CDF F, and $\nu = m$ is the Lebesgue measure.

Probability density function (PDF)

Example (Probability density)

Let $\lambda = P$ be a probability measure on $\mathcal{R}$ corresponding to a CDF F, and $\nu = m$ is the Lebesgue measure. If $P \ll m$, its Radon-Nikodym derivative w.r.t. m, $\frac{\mathrm{d}P}{\mathrm{d}m}$, is called the (Lebesgue) PDF of F.

Probability density function (PDF)

Example (Probability density)

Let $\lambda = P$ be a probability measure on $\mathcal{R}$ corresponding to a CDF F, and $\nu = m$ is the Lebesgue measure. If $P \ll m$, its Radon-Nikodym derivative w.r.t. m, $\frac{\mathrm{d}P}{\mathrm{d}m}$, is called the (Lebesgue) PDF of F.
In particular, suppose F has a derivative f in the usual sense of calculus:

$$P((-\infty, x]) = F(x) = \int_{-\infty}^{x} f(y)\,\mathrm{d}m(y) = \int_{-\infty}^{x} f(y)\,\mathrm{d}y. \tag{11}$$

In this case, one can show that the Radon-Nikodym derivative is the same as the usual derivative in calculus, that is, $\frac{\mathrm{d}P}{\mathrm{d}m} = f$, a.e.

Probability density function (PDF)

Example (Probability density)

Let $\lambda = P$ be a probability measure on $\mathcal{R}$ corresponding to a CDF F, and $\nu = m$ is the Lebesgue measure. If $P \ll m$, its Radon-Nikodym derivative w.r.t. m, $\frac{\mathrm{d}P}{\mathrm{d}m}$, is called the (Lebesgue) PDF of F.
In particular, suppose F has a derivative f in the usual sense of calculus:

$$P((-\infty, x]) = F(x) = \int_{-\infty}^{x} f(y)\, \mathrm{d}m(y) = \int_{-\infty}^{x} f(y)\, \mathrm{d}y. \tag{11}$$

In this case, one can show that the Radon-Nikodym derivative is the same as the usual derivative in calculus, that is, $\frac{\mathrm{d}P}{\mathrm{d}m} = f$, a.e.

For any general σ-finite ν, if P is a probability measure on $\mathcal{R}$ corresponding to a CDF F and $P \ll \nu$, then $\frac{\mathrm{d}P}{\mathrm{d}\nu}$ is called the PDF of F w.r.t. ν.

Example (Discrete CDF and PDF)

Let $a_1 < a_2 < \cdots$ be a sequence of real numbers and X a random variable that $X \in \Lambda = \{a_1, a_2, \ldots\}$. Let $p_n = P(X = a_n)$.

Example (Discrete CDF and PDF)

Let $a_1 < a_2 < \cdots$ be a sequence of real numbers and X a random variable that $X \in \Lambda = \{a_1, a_2, \ldots\}$. Let $p_n = P(X = a_n)$. Then the CDF of X is

$$F(x) = \begin{cases} \sum_{i=1}^{n} p_i & a_n \le x < a_{n+1}, \\ 0 & -\infty < x < a_1. \end{cases} \tag{12}$$

Example (Discrete CDF and PDF)

Let $a_1 < a_2 < \cdots$ be a sequence of real numbers and X a random variable that $X \in \Lambda = \{a_1, a_2, \ldots\}$. Let $p_n = P(X = a_n)$. Then the CDF of X is

$$F(x) = \begin{cases} \sum_{i=1}^{n} p_i & a_n \leq x < a_{n+1}, \\ 0 & -\infty < x < a_1. \end{cases} \tag{12}$$

This CDF is a stepwise function, and called a discrete CDF.

Example (Discrete CDF and PDF)

Let $a_1 < a_2 < \cdots$ be a sequence of real numbers and X a random variable that $X \in \Lambda = \{a_1, a_2, \ldots\}$. Let $p_n = P(X = a_n)$. Then the CDF of X is

$$F(x) = \begin{cases} \sum_{i=1}^n p_i & a_n \le x < a_{n+1}, \\ 0 & -\infty < x < a_1. \end{cases} \tag{12}$$

This CDF is a stepwise function, and called a discrete CDF.
The corresponding probability measure on Λ is given by

$$P_X(A) = \sum_{i: a_i \in A} p_i, \quad A \in \mathcal{F} = \text{the power set of } \Lambda. \tag{13}$$

Let ν be the counting measure on $\mathcal{F}$ and define f on Λ by $f(a_i) = p_i$.

Example (Discrete CDF and PDF)

Let $a_1 < a_2 < \cdots$ be a sequence of real numbers and X a random variable that $X \in \Lambda = \{a_1, a_2, \ldots\}$. Let $p_n = P(X = a_n)$. Then the CDF of X is

$$F(x) = \begin{cases} \sum_{i=1}^{n} p_i & a_n \leq x < a_{n+1}, \\ 0 & -\infty < x < a_1. \end{cases} \tag{12}$$

This CDF is a stepwise function, and called a discrete CDF.
The corresponding probability measure on Λ is given by

$$P_X(A) = \sum_{i:a_i \in A} p_i, \quad A \in \mathcal{F} = \text{the power set of } \Lambda. \tag{13}$$

Let ν be the counting measure on $\mathcal{F}$ and define f on Λ by $f(a_i) = p_i$.
Then $P_X(A) = \int_A f \, \mathrm{d}\nu$. So f is the PDF of F w.r.t. the counting measure.

Example (Discrete CDF and PDF)

Let $a_1 < a_2 < \cdots$ be a sequence of real numbers and X a random variable that $X \in \Lambda = \{a_1, a_2, \ldots\}$. Let $p_n = P(X = a_n)$. Then the CDF of X is

$$F(x) = \begin{cases} \sum_{i=1}^{n} p_i & a_n \leq x < a_{n+1}, \\ 0 & -\infty < x < a_1. \end{cases} \tag{12}$$

This CDF is a stepwise function, and called a discrete CDF.
The corresponding probability measure on Λ is given by

$$P_X(A) = \sum_{i: a_i \in A} p_i, \quad A \in \mathcal{F} = \text{the power set of } \Lambda. \tag{13}$$

Let ν be the counting measure on $\mathcal{F}$ and define f on Λ by $f(a_i) = p_i$. Then $P_X(A) = \int_A f \, \mathrm{d}\nu$. So f is the PDF of F w.r.t. the counting measure.

Any discrete CDF has a PDF w.r.t. the counting measure, and such PDF is called discrete PDF (or PMF, probability mass function).

Calculus with Radon-Nikodym derivatives

Let ν be a σ-finite measure on a measurable space $(\Omega, \mathcal{F})$. Suppose all other measures in (1-3) are also defined on $(\Omega, \mathcal{F})$

(1) If $\lambda \ll \nu$ and $f \geq 0$, then

$$\int f \, \mathrm{d}\lambda = \int f \frac{\mathrm{d}\lambda}{\mathrm{d}\nu} \, \mathrm{d}\nu. \tag{14}$$

(2) If $\lambda_i \ll \nu$, then $\lambda_1 + \lambda_2 \ll \nu$ and

$$\frac{\mathrm{d}(\lambda_1 + \lambda_2)}{\mathrm{d}\nu} = \frac{\mathrm{d}\lambda_1}{\mathrm{d}\nu} + \frac{\mathrm{d}\lambda_2}{\mathrm{d}\nu} \qquad \nu\text{-a.e.} \tag{15}$$

Properties of Radon-Nikodym derivatives (Cont.)

(3) Chain rule: If λ is σ-finite and $\tau \ll \lambda \ll \nu$, then

$$\frac{\mathrm{d}\tau}{\mathrm{d}\nu} = \frac{\mathrm{d}\tau}{\mathrm{d}\lambda}\frac{\mathrm{d}\lambda}{\mathrm{d}\nu} \qquad \nu\text{-a.e.} \tag{16}$$

In particular, if $\lambda \ll \nu$ and $\nu \ll \lambda$,then let $\tau = \nu$ in the above, and we have

$$\frac{\mathrm{d}\lambda}{\mathrm{d}\nu} = \left(\frac{\mathrm{d}\nu}{\mathrm{d}\lambda}\right)^{-1} \qquad \nu \text{ or } \lambda\text{-a.e.} \tag{17}$$

(4) Let $(\Omega_i, \mathcal{F}_i, \nu_i)$ be a measure space and ν_i be σ-finite, $i = 1, 2$. Let λ_i be a measure on $(\Omega_i, \mathcal{F}_i)$ and $\lambda_i \ll \nu_i$, $i = 1, 2$. Then
$\lambda_1 \times \lambda_2 \ll \nu_1 \times \nu_2$
and

$$\frac{\mathrm{d}(\lambda_1 \times \lambda_2)}{\mathrm{d}(\nu_1 \times \nu_2)}(\omega_1, \omega_2) = \frac{\mathrm{d}\lambda_1}{\mathrm{d}\nu_1}(\omega_1) \cdot \frac{\mathrm{d}\lambda_2}{\mathrm{d}\nu_2}(\omega_2), \quad (\nu_1 \times \nu_2) - \text{a.e.} \tag{18}$$

Moments

Definitions of moments of a random variable X:

- If $\mathbb{E}|X|^p < \infty$ for some real number p, then $\mathbb{E}|X|^p$ is called the pth *absolute moment* of X or its law P_X

Moments

Definitions of moments of a random variable X:

- If $\mathbb{E}|X|^p < \infty$ for some real number p, then $\mathbb{E}|X|^p$ is called the pth *absolute moment* of X or its law P_X
- If $\mathbb{E}X^k$ is finite, where k is a positive integer, then $\mathbb{E}X^k$ is called the kth *moment* of X or P_X
  - when $k = 1$, it is the expectation of X

Moments

Definitions of moments of a random variable X:

- If $\mathbb{E}|X|^p < \infty$ for some real number p, then $\mathbb{E}|X|^p$ is called the pth *absolute moment* of X or its law P_X
- If $\mathbb{E}X^k$ is finite, where k is a positive integer, then $\mathbb{E}X^k$ is called the kth *moment* of X or P_X
    - when $k = 1$, it is the expectation of X
- If $\mu = \mathbb{E}X$ and $\mathbb{E}(X - \mu)^k$ are finite for a positive integer k, then $\mathbb{E}(X - \mu)^k$ is called the kth *central moment* of X or P_X

Moments

Definitions of moments of a random variable X:

- If $\mathbb{E}|X|^p < \infty$ for some real number p, then $\mathbb{E}|X|^p$ is called the pth *absolute moment* of X or its law P_X
- If $\mathbb{E}X^k$ is finite, where k is a positive integer, then $\mathbb{E}X^k$ is called the kth *moment* of X or P_X
  - when $k = 1$, it is the expectation of X
- If $\mu = \mathbb{E}X$ and $\mathbb{E}(X - \mu)^k$ are finite for a positive integer k, then $\mathbb{E}(X - \mu)^k$ is called the kth *central moment* of X or P_X
  - When $k = 2$, it is called the *variance* of X or P_X, and denoted by $\mathrm{Var}(X)$ or σ_X^2
  - $\sqrt{\mathrm{Var}(X)}$ is called the standard deviation of X, often denoted by σ_X

We have similar definitions for a random vector $X \in \mathcal{R}^d$ or a random matrix $X \in \mathcal{R}^{d_1 \times d_2}$

- $(a_1, a_2, \ldots, a_d)$ denotes a row vector, and $(a_1, a_2, \ldots, a_d)^\top$ denotes its transport, which is a column vector

We have similar definitions for a random vector $X \in \mathcal{R}^d$ or a random matrix $X \in \mathcal{R}^{d_1 \times d_2}$

- $(a_1, a_2, \ldots, a_d)$ denotes a row vector, and $(a_1, a_2, \ldots, a_d)^\top$ denotes its transport, which is a column vector
- For a random vector $X = (X_1, \ldots, X_d)^\top$, we use $\mathbb{E}X$ to denote $(\mathbb{E}X_1, \ldots, \mathbb{E}X_d)^\top$

We have similar definitions for a random vector $X \in \mathcal{R}^d$ or a random matrix $X \in \mathcal{R}^{d_1 \times d_2}$

- $(a_1, a_2, \ldots, a_d)$ denotes a row vector, and $(a_1, a_2, \ldots, a_d)^\top$ denotes its transport, which is a column vector
- For a random vector $X = (X_1, \ldots, X_d)^\top$, we use $\mathbb{E}X$ to denote $(\mathbb{E}X_1, \ldots, \mathbb{E}X_d)^\top$
- Similarly, for a random matrix

$$X = \begin{pmatrix} X_{11} & X_{21} & \cdots & X_{1d} \\ X_{21} & X_{22} & \cdots & X_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ X_{d1} & X_{d2} & \cdots & X_{dd} \end{pmatrix} \tag{19}$$

we denote

$$\mathbb{E}X = \begin{pmatrix} \mathbb{E}X_{11} & \mathbb{E}X_{21} & \cdots & \mathbb{E}X_{1d} \\ \mathbb{E}X_{21} & \mathbb{E}X_{22} & \cdots & \mathbb{E}X_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{E}X_{d1} & \mathbb{E}X_{d2} & \cdots & \mathbb{E}X_{dd} \end{pmatrix} \tag{20}$$

- For a random vector $X \in \mathcal{R}^d$, $\mathrm{Var}(X) = \mathbb{E}\{(X - \mathbb{E}X)(X - \mathbb{E}X)^\top$ is called the *covariance matrix* of X

- For a random vector $X \in \mathcal{R}^d$, $\mathrm{Var}(X) = \mathbb{E}\{(X - \mathbb{E}X)(X - \mathbb{E}X)^\top$ is called the *covariance matrix* of X
  $\mathrm{Var}(X)$ is a matrix, whose (i, j)th element is $\mathbb{E}\{(X_i - \mathbb{E}X_i)(X_j - \mathbb{E}X_j)\}$.

- For a random vector $X \in \mathcal{R}^d$, $\mathrm{Var}(X) = \mathbb{E}\{(X - \mathbb{E}X)(X - \mathbb{E}X)^\top$ is called the *covariance matrix* of X
  $\mathrm{Var}(X)$ is a matrix, whose (i, j)th element is $\mathbb{E}\{(X_i - \mathbb{E}X_i)(X_j - \mathbb{E}X_j)\}$.
- For two random variables X and Y, the quantity $\mathbb{E}\{(X - \mathbb{E}X)(X - \mathbb{E}Y)\}$, denoted by $\mathrm{Cov}(X, Y)$, is called the *covariance* of X and Y
  - If $\mathrm{Cov}(X, Y) = 0$, then we say X and Y are *uncorrelated*
  - The standardized covariance, $\mathrm{Cov}(X, Y)/(\sigma_X \sigma_Y)$, is called the *correlation* of X and Y, and denoted by $\mathrm{Corr}(X, Y)$

Basic properties

Assume all the moments below exist

- Symmetry of covariance: $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$

Basic properties

Assume all the moments below exist

- Symmetry of covariance: $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$
- $\mathrm{Cov}(\sum_i a_i X_i, \sum_j b_j Y_j) = \sum_i \sum_j a_i b_j \mathrm{Cov}(X_i, Y_j)$

Basic properties

Assume all the moments below exist

- Symmetry of covariance: $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$
- $\mathrm{Cov}(\sum_i a_i X_i, \sum_j b_j Y_j) = \sum_i \sum_j a_i b_j \mathrm{Cov}(X_i, Y_j)$
- If X is a random matrix, then $\mathbb{E}(AXB) = A\mathbb{E}(X)B$ for non-random matrices A and B

Basic properties

Assume all the moments below exist

- Symmetry of covariance: $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$
- $\mathrm{Cov}(\sum_i a_i X_i, \sum_j b_j Y_j) = \sum_i \sum_j a_i b_j \mathrm{Cov}(X_i, Y_j)$
- If X is a random matrix, then $\mathbb{E}(AXB) = A\mathbb{E}(X)B$ for non-random matrices A and B
- For some non-random vector $a \in \mathcal{R}^d$, we have
  - $\mathbb{E}(a^\top X) = a^\top \mathbb{E}X$
  - $\mathrm{Var}(a^\top X) = a^\top \mathrm{Var}(X) a$

Basic properties

Assume all the moments below exist

- Symmetry of covariance: $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$
- $\mathrm{Cov}(\sum_i a_i X_i, \sum_j b_j Y_j) = \sum_i \sum_j a_i b_j \mathrm{Cov}(X_i, Y_j)$
- If X is a random matrix, then $\mathbb{E}(AXB) = A\mathbb{E}(X)B$ for non-random matrices A and B
- For some non-random vector $a \in \mathcal{R}^d$, we have
  - $\mathbb{E}(a^\top X) = a^\top \mathbb{E}X$
  - $\mathrm{Var}(a^\top X) = a^\top \mathrm{Var}(X) a$

Basic properties (Cont.)

For a random vector X, $\mathrm{Var}(X)$ is a *symmetric positive semi-definite* (SPSD) matrix

- a matrix M is symmetric if $M = M^\top$
- a $d \times d$ square matrix M is positive semi-definite (PSD) if for any $v \in \mathcal{R}^d$, $v^\top M v \geq 0$.

Basic properties (Cont.)

For a random vector X, $\mathrm{Var}(X)$ is a *symmetric positive semi-definite* (SPSD) matrix

- a matrix M is symmetric if $M = M^\top$
- a $d \times d$ square matrix M is positive semi-definite (PSD) if for any $v \in \mathcal{R}^d$, $v^\top M v \geq 0$.

Proof: Let $M = \mathrm{Var}(X)$.

- symmetry: $M_{ij} = \mathrm{Cov}(X_i, X_j) = \mathrm{Cov}(X_j, X_i) = M_{ji}$.

Basic properties (Cont.)

For a random vector X, $\mathrm{Var}(X)$ is a *symmetric positive semi-definite* (SPSD) matrix

- a matrix M is symmetric if $M = M^\top$
- a $d \times d$ square matrix M is positive semi-definite (PSD) if for any $v \in \mathcal{R}^d$, $v^\top M v \geq 0$.

Proof: Let $M = \mathrm{Var}(X)$.

- symmetry: $M_{ij} = \mathrm{Cov}(X_i, X_j) = \mathrm{Cov}(X_j, X_i) = M_{ji}$.
- positive semi-definite:

$$\begin{aligned} v^\top M v &= v^\top \mathbb{E}\{(X - \mathbb{E}X)(X - \mathbb{E}X)^\top\} v \\ &= \mathbb{E}\{v^\top (X - \mathbb{E}X)(X - \mathbb{E}X)^\top v\}. \end{aligned}$$

  Let $Y = v^\top (X - \mathbb{E}X)$. Note that Y is a scalar. Then $v^\top M v = \mathbb{E}(YY^\top) = \mathbb{E}(Y^2) \geq 0$.

Cauchy-Schwarz inequality

$\mathrm{Cov}(X, Y)^2 \leq \mathrm{Var}(X)\mathrm{Var}(Y)$ and $[\mathbb{E}(XY)]^2 \leq \mathbb{E}(X^2)\mathbb{E}(Y^2)$

Cauchy-Schwarz inequality

$\mathrm{Cov}(X, Y)^2 \le \mathrm{Var}(X)\mathrm{Var}(Y)$ and $[\mathbb{E}(XY)]^2 \le \mathbb{E}(X^2)\mathbb{E}(Y^2)$

Proof:

- Since the variance of $(X, Y)^\top$ is SPSD, its determinant is nonnegative

Cauchy-Schwarz inequality

$\mathrm{Cov}(X, Y)^2 \leq \mathrm{Var}(X)\mathrm{Var}(Y)$ and $[\mathbb{E}(XY)]^2 \leq \mathbb{E}(X^2)\mathbb{E}(Y^2)$

Proof:

- Since the variance of $(X, Y)^\top$ is SPSD, its determinant is nonnegative
- $\det\left(\mathrm{Var}((X, Y)^\top)\right) = \mathrm{Var}(X)\mathrm{Var}(Y) - \mathrm{Cov}(X, Y)^2$

Cauchy-Schwarz inequality

$\mathrm{Cov}(X, Y)^2 \leq \mathrm{Var}(X)\mathrm{Var}(Y)$ and $[\mathbb{E}(XY)]^2 \leq \mathbb{E}(X^2)\mathbb{E}(Y^2)$

Proof:

- Since the variance of $(X, Y)^\top$ is SPSD, its determinant is nonnegative
- $\det\left(\mathrm{Var}((X, Y)^\top)\right) = \mathrm{Var}(X)\mathrm{Var}(Y) - \mathrm{Cov}(X, Y)^2$
- Apply the same argument to the 2nd absolute moment to obtain the other inequality

Jensen's Inequality

- A set $A \subset \mathcal{R}^d$ is *convex* if for any $x, y \in A$, $tx + (1-t)y \in A$.

Jensen's Inequality

- A set $A \subset \mathcal{R}^d$ is *convex* if for any $x, y \in A$, $tx + (1-t)y \in A$.
- A real function f is *convex* on A if

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y) \tag{21}$$

holds for any $x, y \in A$ and $t \in [0,1]$. If the inequality holds with $\leq$ replaced by $<$, then f is *strictly convex*.

Jensen's Inequality

- A set $A \subset \mathcal{R}^d$ is *convex* if for any $x, y \in A$, $tx + (1-t)y \in A$.
- A real function f is *convex* on A if

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y) \tag{21}$$

holds for any $x, y \in A$ and $t \in [0, 1]$. If the inequality holds with $\leq$ replaced by $<$, then f is *strictly convex*.

Theorem

For a random vector and a convex function φ,

$$\varphi(\mathbb{E}X) \leq \mathbb{E}\varphi(X). \tag{22}$$

If φ is strictly convex and $\varphi(X)$ is not a constant, then $\varphi(\mathbb{E}X) < \mathbb{E}\varphi(X)$.

- If φ is twice differentiable, then the convexity of φ is implied by the positive semi-definiteness of its Hessian (or second derivative if φ is univariate) φ''.

Proof:

- A well known fact: For any $y \in A$, there exists a vector $c_y \in \mathcal{R}^d$ such that

$$\varphi(x) \geq \varphi(y) + < c_y, x - y >, \forall x \in A. \tag{23}$$

Proof:

- A well known fact: For any $y \in A$, there exists a vector $c_y \in \mathcal{R}^d$ such that

$$\varphi(x) \geq \varphi(y) + < c_y, x - y >, \forall x \in A. \tag{23}$$

- Set $y = \mathbb{E}X$. Substitute $x = X$, and then take expectation on both sides.

$\square$

Proof:

- A well known fact: For any $y \in A$, there exists a vector $c_y \in \mathcal{R}^d$ such that

$$\varphi(x) \geq \varphi(y) + < c_y, x - y >, \forall x \in A. \tag{23}$$

- Set $y = \mathbb{E}X$. Substitute $x = X$, and then take expectation on both sides.

$\square$

Examples: Suppose X is a nonconstant positive r.v., then

- $(\mathbb{E}X)^{-1} < \mathbb{E}(X^{-1})$

Proof:

- A well known fact: For any $y \in A$, there exists a vector $c_y \in \mathcal{R}^d$ such that
$$\varphi(x) \geq \varphi(y) + < c_y, x - y >, \forall x \in A. \tag{23}$$

- Set $y = \mathbb{E}X$. Substitute $x = X$, and then take expectation on both sides.

$\square$

Examples: Suppose X is a nonconstant positive r.v., then

- $(\mathbb{E}X)^{-1} < \mathbb{E}(X^{-1})$ (since $f(x) = x^{-1}$ is convex)

Proof:

- A well known fact: For any $y \in A$, there exists a vector $c_y \in \mathcal{R}^d$ such that

$$\varphi(x) \geq \varphi(y) + < c_y, x - y >, \forall x \in A. \tag{23}$$

- Set $y = \mathbb{E}X$. Substitute $x = X$, and then take expectation on both sides.

$\square$

Examples: Suppose X is a nonconstant positive r.v., then

- $(\mathbb{E}X)^{-1} < \mathbb{E}(X^{-1})$ (since $f(x) = x^{-1}$ is convex)
- $\mathbb{E}(\log X) < \log(\mathbb{E}X)$

Proof:

- A well known fact: For any $y \in A$, there exists a vector $c_y \in \mathcal{R}^d$ such that
$$\varphi(x) \geq \varphi(y) + < c_y, x - y >, \forall x \in A. \tag{23}$$

- Set $y = \mathbb{E}X$. Substitute $x = X$, and then take expectation on both sides.

$\square$

Examples: Suppose X is a nonconstant positive r.v., then

- $(\mathbb{E}X)^{-1} < \mathbb{E}(X^{-1})$ (since $f(x) = x^{-1}$ is convex)
- $\mathbb{E}(\log X) < \log(\mathbb{E}X)$ (since $f(x) = \log(x)$ is concave, i.e. $-\log(x)$ is convex)

Theorem (Chebyshev)

Let X be a random variable and φ a nonnegative and nondecreasing function on $[0, \infty)$ and $\varphi(-t) = \varphi(t)$ for all real t. Then, for each constant $t \geq 0$,

$$\varphi(t)P(|X| \geq t) \leq \int_{\{|X| \geq t\}} \varphi(X) \, \mathrm{d}P \leq \mathbb{E}\varphi(X). \tag{24}$$

Theorem (Chebyshev)

Let X be a random variable and φ a nonnegative and nondecreasing function on $[0, \infty)$ and $\varphi(-t) = \varphi(t)$ for all real t. Then, for each constant $t \geq 0$,

$$\varphi(t)P(|X| \geq t) \leq \int_{\{|X| \geq t\}} \varphi(X) \, \mathrm{d}P \leq \mathbb{E}\varphi(X). \tag{24}$$

Examples:

- $\varphi(t) = |t|$, we have Markov's inequality

$$P(|X| \geq t) \leq \frac{\mathbb{E}\varphi(X)}{t}, \quad \forall t > 0. \tag{25}$$

Theorem (Chebyshev)

Let X be a random variable and φ a nonnegative and nondecreasing function on $[0, \infty)$ and $\varphi(-t) = \varphi(t)$ for all real t. Then, for each constant $t \geq 0$,

$$\varphi(t)P(|X| \geq t) \leq \int_{\{|X| \geq t\}} \varphi(X) \, \mathrm{d}P \leq \mathbb{E}\varphi(X). \tag{24}$$

Examples:

- $\varphi(t) = |t|$, we have Markov's inequality

$$P(|X| \geq t) \leq \frac{\mathbb{E}\varphi(X)}{t}, \quad \forall t > 0. \tag{25}$$

- $\varphi(t) = t^2$ and X is replaced with $X - \mu$ where $\mu = \mathbb{E}X$, we obtain the classic Chebyshev' inequality:

$$P(|X - \mu| \geq t^2) \leq \frac{\sigma_X^2}{t^2}. \tag{26}$$

Exercises

1. Let $\mathcal{C}$ be a collection of subsets of Ω and $\Gamma = \{\mathcal{F} : \mathcal{F} \text{ is a } \sigma\text{-field on } \Omega \text{ and } \mathcal{C} \subset \mathcal{F}\}$. Show that $\Gamma \neq \emptyset$ and $\sigma(\mathcal{C}) = \bigcap_{\mathcal{F} \in \Gamma} \mathcal{F}$.
2. Let $(\Omega, \mathcal{F})$ be a measurable space. $f : \Omega \mapsto \mathcal{R}$ is Borel if and only if $f^{-1}(a, \infty) \in \mathcal{F}$ for all $a \in \mathcal{R}$.
3. Show that a monotone function from $\mathcal{R}$ to $\mathcal{R}$ is Borel
4. Let f be a Borel function on $\mathcal{R}^2$. Define a function g from $\mathcal{R}$ to $\mathcal{R}$ as $g(x) = f(x, y_0)$, where y_0 is a fixed point in $\mathcal{R}$. Show that g is Borel.

Exercise 1

Let $\mathcal{C}$ be a collection of subsets of Ω and
$\Gamma = \{\mathcal{F} : \mathcal{F} \text{ is a } \sigma\text{-field on } \Omega \text{ and } \mathcal{C} \subset \mathcal{F}\}$.
Show that $\Gamma \neq \emptyset$ and $\sigma(\mathcal{C}) = \bigcap_{\mathcal{F} \in \Gamma} \mathcal{F}$.

Exercise 1

Let $\mathcal{C}$ be a collection of subsets of Ω and
$\Gamma = \{\mathcal{F} : \mathcal{F}$ is a σ-field on Ω and $\mathcal{C} \subset \mathcal{F}\}$.
Show that $\Gamma \neq \emptyset$ and $\sigma(\mathcal{C}) = \underset{\mathcal{F}\in\Gamma}{\cap} \mathcal{F}$.

Proof:

- Since $\mathcal{C} \subset 2^{\Omega} \in \Gamma$, Γ is not empty.

Exercise 1

Let $\mathcal{C}$ be a collection of subsets of Ω and
$\Gamma = \{\mathcal{F} : \mathcal{F} \text{ is a } \sigma\text{-field on } \Omega \text{ and } \mathcal{C} \subset \mathcal{F}\}$.
Show that $\Gamma \neq \emptyset$ and $\sigma(\mathcal{C}) = \bigcap_{\mathcal{F} \in \Gamma} \mathcal{F}$.

Proof:

- Since $\mathcal{C} \subset 2^{\Omega} \in \Gamma$, Γ is not empty.
- It is straightforward to show that $\bigcap_{\mathcal{F} \in \Gamma} \mathcal{F}$ is a σ-field by checking the 3 conditions.

Exercise 1

Let $\mathcal{C}$ be a collection of subsets of Ω and
$\Gamma = \{\mathcal{F} : \mathcal{F} \text{ is a } \sigma\text{-field on } \Omega \text{ and } \mathcal{C} \subset \mathcal{F}\}$.
Show that $\Gamma \neq \emptyset$ and $\sigma(\mathcal{C}) = \bigcap_{\mathcal{F} \in \Gamma} \mathcal{F}$.

Proof:

- Since $\mathcal{C} \subset 2^{\Omega} \in \Gamma$, Γ is not empty.
- It is straightforward to show that $\bigcap_{\mathcal{F} \in \Gamma} \mathcal{F}$ is a σ-field by checking the 3 conditions.
  For instance, if $A \in \bigcap_{\mathcal{F} \in \Gamma} \mathcal{F}$ then for any $\mathcal{F} \in \Gamma$, $A \in \mathcal{F}$, and thus $A^c \in \mathcal{F}$. So $A^c \in \bigcap_{\mathcal{F} \in \Gamma} \mathcal{F}$.

Exercise 1

Let $\mathcal{C}$ be a collection of subsets of Ω and
$\Gamma = \{\mathcal{F} : \mathcal{F} \text{ is a } \sigma\text{-field on } \Omega \text{ and } \mathcal{C} \subset \mathcal{F}\}$.
Show that $\Gamma \neq \emptyset$ and $\sigma(\mathcal{C}) = \bigcap_{\mathcal{F} \in \Gamma} \mathcal{F}$.

Proof:

- Since $\mathcal{C} \subset 2^{\Omega} \in \Gamma$, Γ is not empty.
- It is straightforward to show that $\bigcap_{\mathcal{F} \in \Gamma} \mathcal{F}$ is a σ-field by checking the 3 conditions.
  For instance, if $A \in \bigcap_{\mathcal{F} \in \Gamma} \mathcal{F}$ then for any $\mathcal{F} \in \Gamma$, $A \in \mathcal{F}$, and thus $A^c \in \mathcal{F}$. So $A^c \in \bigcap_{\mathcal{F} \in \Gamma} \mathcal{F}$.
- If $\mathcal{E}$ is a σ-field and $\mathcal{C} \subset \mathcal{E}$, then $\mathcal{E} \in \Gamma$.

Exercise 1

Let $\mathcal{C}$ be a collection of subsets of Ω and
$\Gamma = \{\mathcal{F} : \mathcal{F}$ is a σ-field on Ω and $\mathcal{C} \subset \mathcal{F}\}$.
Show that $\Gamma \neq \emptyset$ and $\sigma(\mathcal{C}) = \bigcap_{\mathcal{F}\in\Gamma} \mathcal{F}$.

Proof:

- Since $\mathcal{C} \subset 2^{\Omega} \in \Gamma$, Γ is not empty.
- It is straightforward to show that $\bigcap_{\mathcal{F}\in\Gamma} \mathcal{F}$ is a σ-field by checking the 3 conditions.
  For instance, if $A \in \bigcap_{\mathcal{F}\in\Gamma} \mathcal{F}$ then for any $\mathcal{F} \in \Gamma$, $A \in \mathcal{F}$, and thus $A^c \in \mathcal{F}$. So $A^c \in \bigcap_{\mathcal{F}\in\Gamma} \mathcal{F}$.
- If $\mathcal{E}$ is a σ-field and $\mathcal{C} \subset \mathcal{E}$, then $\mathcal{E} \in \Gamma$. So $\bigcap_{\mathcal{F}\in\Gamma} \mathcal{F} \subset \mathcal{E}$.

Exercises 2

Let $(\Omega, \mathcal{F})$ be a measurable space. $f : \Omega \mapsto \mathcal{R}$ is Borel if and only if $f^{-1}(a, \infty) \in \mathcal{F}$ for all $a \in \mathcal{R}$.

Exercises 2

Let $(\Omega, \mathcal{F})$ be a measurable space. $f : \Omega \mapsto \mathcal{R}$ is Borel if and only if $f^{-1}(a, \infty) \in \mathcal{F}$ for all $a \in \mathcal{R}$.
Proof: "Only if" is trivial. We only prove the "if" part.

Exercises 2

Let $(\Omega, \mathcal{F})$ be a measurable space. $f : \Omega \mapsto \mathcal{R}$ is Borel if and only if $f^{-1}(a, \infty) \in \mathcal{F}$ for all $a \in \mathcal{R}$.

Proof: "Only if" is trivial. We only prove the "if" part.

1. Let $\mathcal{G} = \{A \subset \mathcal{R} : f^{-1}(A) \in \mathcal{F}\}$. Only need to show $\mathcal{B} \subset \mathcal{G}$.

Exercises 2

Let $(\Omega, \mathcal{F})$ be a measurable space. $f : \Omega \mapsto \mathcal{R}$ is Borel if and only if $f^{-1}(a, \infty) \in \mathcal{F}$ for all $a \in \mathcal{R}$.

Proof: "Only if" is trivial. We only prove the "if" part.

1. Let $\mathcal{G} = \{A \subset \mathcal{R} : f^{-1}(A) \in \mathcal{F}\}$. Only need to show $\mathcal{B} \subset \mathcal{G}$.
2. Check $\mathcal{G}$ is a σ-field

Exercises 2

Let $(\Omega, \mathcal{F})$ be a measurable space. $f : \Omega \mapsto \mathcal{R}$ is Borel if and only if $f^{-1}(a, \infty) \in \mathcal{F}$ for all $a \in \mathcal{R}$.

Proof: "Only if" is trivial. We only prove the "if" part.

1. Let $\mathcal{G} = \{A \subset \mathcal{R} : f^{-1}(A) \in \mathcal{F}\}$. Only need to show $\mathcal{B} \subset \mathcal{G}$.
2. Check $\mathcal{G}$ is a σ-field
3. Let $\mathcal{K} = \{(a, \infty) : a \in \mathcal{R}\}$. Show $\sigma(\mathcal{K}) = \mathcal{B}$.

Exercises 2

Let $(\Omega, \mathcal{F})$ be a measurable space. $f : \Omega \mapsto \mathcal{R}$ is Borel if and only if $f^{-1}(a, \infty) \in \mathcal{F}$ for all $a \in \mathcal{R}$.

Proof: "Only if" is trivial. We only prove the "if" part.

1. Let $\mathcal{G} = \{A \subset \mathcal{R} : f^{-1}(A) \in \mathcal{F}\}$. Only need to show $\mathcal{B} \subset \mathcal{G}$.
2. Check $\mathcal{G}$ is a σ-field
3. Let $\mathcal{K} = \{(a, \infty) : a \in \mathcal{R}\}$. Show $\sigma(\mathcal{K}) = \mathcal{B}$.
4. Since $\mathcal{K} \subset \mathcal{G}$, $\sigma(\mathcal{K}) \subset \mathcal{G}$, i.e. $\mathcal{B} \subset \mathcal{G}$.

Exercises 2

Let $(\Omega, \mathcal{F})$ be a measurable space. $f : \Omega \mapsto \mathcal{R}$ is Borel if and only if $f^{-1}(a, \infty) \in \mathcal{F}$ for all $a \in \mathcal{R}$.

Proof: "Only if" is trivial. We only prove the "if" part.

1. Let $\mathcal{G} = \{A \subset \mathcal{R} : f^{-1}(A) \in \mathcal{F}\}$. Only need to show $\mathcal{B} \subset \mathcal{G}$.
2. Check $\mathcal{G}$ is a σ-field
3. Let $\mathcal{K} = \{(a, \infty) : a \in \mathcal{R}\}$. Show $\sigma(\mathcal{K}) = \mathcal{B}$.
4. Since $\mathcal{K} \subset \mathcal{G}$, $\sigma(\mathcal{K}) \subset \mathcal{G}$, i.e. $\mathcal{B} \subset \mathcal{G}$.

We need a theorem (see page 6 of *Real Analysis* by E. Stein and R. Shakarchi):

Every open set of $\mathcal{R}$ can be written as a countable union of disjoint open intervals.

Exercise 3

Show that a monotone function f from $\mathcal{R}$ to $\mathcal{R}$ is Borel.

Exercise 3

Show that a monotone function f from $\mathcal{R}$ to $\mathcal{R}$ is Borel.
Proof: Use Exercise 2.

Exercise 3

Show that a monotone function f from $\mathcal{R}$ to $\mathcal{R}$ is Borel.
Proof: Use Exercise 2.
For any $a \in \mathcal{R}$, $f^{-1}(a, \infty)$ is either (b, ∞) or $[b, \infty)$ for some b, so it is a Borel set. So f is Borel.

Exercise 4

Let f be a Borel function on $\mathcal{R}^2$. Define a function g from $\mathcal{R}$ to $\mathcal{R}$ as $g(x) = f(x, y_0)$, where y_0 is a fixed point in $\mathcal{R}$. Show that g is Borel.

Exercise 4

Let f be a Borel function on $\mathcal{R}^2$. Define a function g from $\mathcal{R}$ to $\mathcal{R}$ as $g(x) = f(x, y_0)$, where y_0 is a fixed point in $\mathcal{R}$. Show that g is Borel.
Proof:

- For any $B \in \mathcal{B}$, we want to show $g^{-1}(B) \in \mathcal{B}$.

Exercise 4

Let f be a Borel function on $\mathcal{R}^2$. Define a function g from $\mathcal{R}$ to $\mathcal{R}$ as $g(x) = f(x, y_0)$, where y_0 is a fixed point in $\mathcal{R}$. Show that g is Borel.
Proof:

- For any $B \in \mathcal{B}$, we want to show $g^{-1}(B) \in \mathcal{B}$.
- $g^{-1}(B) = \{x : f(x, y_0) \in B\} = \{x : (x, y_0) \in f^{-1}(B)\}$

Exercise 4

Let f be a Borel function on $\mathcal{R}^2$. Define a function g from $\mathcal{R}$ to $\mathcal{R}$ as $g(x) = f(x, y_0)$, where y_0 is a fixed point in $\mathcal{R}$. Show that g is Borel.
Proof:

- For any $B \in \mathcal{B}$, we want to show $g^{-1}(B) \in \mathcal{B}$.
- $g^{-1}(B) = \{x : f(x, y_0) \in B\} = \{x : (x, y_0) \in f^{-1}(B)\}$
- Note that $f^{-1}(B) \subset \mathcal{R}^2$. If $f^{-1}(B)$ makes the above set a Borel set, then it has a "good property".

Exercise 4

Let f be a Borel function on $\mathcal{R}^2$. Define a function g from $\mathcal{R}$ to $\mathcal{R}$ as $g(x) = f(x, y_0)$, where y_0 is a fixed point in $\mathcal{R}$. Show that g is Borel.
Proof:

- For any $B \in \mathcal{B}$, we want to show $g^{-1}(B) \in \mathcal{B}$.
- $g^{-1}(B) = \{x : f(x, y_0) \in B\} = \{x : (x, y_0) \in f^{-1}(B)\}$
- Note that $f^{-1}(B) \subset \mathcal{R}^2$. If $f^{-1}(B)$ makes the above set a Borel set, then it has a "good property".
- Define the collection of sets of "good property":

$$\mathcal{G} = \left\{C \subset \mathcal{R}^2 : \{x : (x, y_0) \in C\} \in \mathcal{B}\right\} \tag{27}$$

Exercise 4

Let f be a Borel function on $\mathcal{R}^2$. Define a function g from $\mathcal{R}$ to $\mathcal{R}$ as $g(x) = f(x, y_0)$, where y_0 is a fixed point in $\mathcal{R}$. Show that g is Borel.
Proof:

- For any $B \in \mathcal{B}$, we want to show $g^{-1}(B) \in \mathcal{B}$.
- $g^{-1}(B) = \{x : f(x, y_0) \in B\} = \{x : (x, y_0) \in f^{-1}(B)\}$
- Note that $f^{-1}(B) \subset \mathcal{R}^2$. If $f^{-1}(B)$ makes the above set a Borel set, then it has a "good property".
- Define the collection of sets of "good property":

$$\mathcal{G} = \left\{C \subset \mathcal{R}^2 : \{x : (x, y_0) \in C\} \in \mathcal{B}\right\} \tag{27}$$

- We want to show $f^{-1}(B) \in \mathcal{G}$.

Exercise 4

Let f be a Borel function on $\mathcal{R}^2$. Define a function g from $\mathcal{R}$ to $\mathcal{R}$ as $g(x) = f(x, y_0)$, where y_0 is a fixed point in $\mathcal{R}$. Show that g is Borel.
Proof:

- For any $B \in \mathcal{B}$, we want to show $g^{-1}(B) \in \mathcal{B}$.
- $g^{-1}(B) = \{x : f(x, y_0) \in B\} = \{x : (x, y_0) \in f^{-1}(B)\}$
- Note that $f^{-1}(B) \subset \mathcal{R}^2$. If $f^{-1}(B)$ makes the above set a Borel set, then it has a "good property".
- Define the collection of sets of "good property":

$$\mathcal{G} = \left\{ C \subset \mathcal{R}^2 : \{x : (x, y_0) \in C\} \in \mathcal{B} \right\} \tag{27}$$

- We want to show $f^{-1}(B) \in \mathcal{G}$.
- Since f is Borel, $f^{-1}(B) \in \mathcal{B}^2$. We only need to show $\mathcal{B}^2 \subset \mathcal{G}$.

Exercise 4 (Cont.)

Let f be a Borel function on $\mathcal{R}^2$. Define a function g from $\mathcal{R}$ to $\mathcal{R}$ as $g(x) = f(x, y_0)$, where y_0 is a fixed point in $\mathcal{R}$. Show that g is Borel.

It remains to show $\mathcal{B}^2 \subset \mathcal{G}$:

- Check $\mathcal{G}$ is a σ-field

Exercise 4 (Cont.)

Let f be a Borel function on $\mathcal{R}^2$. Define a function g from $\mathcal{R}$ to $\mathcal{R}$ as $g(x) = f(x, y_0)$, where y_0 is a fixed point in $\mathcal{R}$. Show that g is Borel.

It remains to show $\mathcal{B}^2 \subset \mathcal{G}$:

- Check $\mathcal{G}$ is a σ-field
  - $\emptyset \in \mathcal{G}$.
  - If $C \in \mathcal{G}$, then $\{x : (x, y_0) \in C^c\} = \{x : (x, y_0) \in C\}^c \in \mathcal{B}$, i.e., $C^c \in \mathcal{G}$.
  - If $C_i \in \mathcal{G}$, $i = 1, 2, \ldots$, then

$$\{x : (x, y_0) \in \cup_i C_i\} = \cup_i \{x : (x, y_0) \in C_i\} \in \mathcal{B},$$

    i.e., $\cup_i C_i \in \mathcal{G}$

Exercise 4 (Cont.)

Let f be a Borel function on $\mathcal{R}^2$. Define a function g from $\mathcal{R}$ to $\mathcal{R}$ as $g(x) = f(x, y_0)$, where y_0 is a fixed point in $\mathcal{R}$. Show that g is Borel.

It remains to show $\mathcal{B}^2 \subset \mathcal{G}$:

- Check $\mathcal{G}$ is a σ-field
  - $\emptyset \in \mathcal{G}$.
  - If $C \in \mathcal{G}$, then $\{x : (x, y_0) \in C^c\} = \{x : (x, y_0) \in C\}^c \in \mathcal{B}$, i.e., $C^c \in \mathcal{G}$.
  - If $C_i \in \mathcal{G}$, $i = 1, 2, \ldots$, then

    $$\{x : (x, y_0) \in \cup_i C_i\} = \cup_i \{x : (x, y_0) \in C_i\} \in \mathcal{B},$$

    i.e., $\cup_i C_i \in \mathcal{G}$
- Check that any closed cube $[a, b] \times [c, d] \in \mathcal{G}$

Exercise 4 (Cont.)

Let f be a Borel function on $\mathcal{R}^2$. Define a function g from $\mathcal{R}$ to $\mathcal{R}$ as $g(x) = f(x, y_0)$, where y_0 is a fixed point in $\mathcal{R}$. Show that g is Borel.

It remains to show $\mathcal{B}^2 \subset \mathcal{G}$:

- Check $\mathcal{G}$ is a σ-field
  - $\emptyset \in \mathcal{G}$.
  - If $C \in \mathcal{G}$, then $\{x : (x, y_0) \in C^c\} = \{x : (x, y_0) \in C\}^c \in \mathcal{B}$, i.e., $C^c \in \mathcal{G}$.
  - If $C_i \in \mathcal{G}$, $i = 1, 2, \ldots$, then

$$\{x : (x, y_0) \in \cup_i C_i\} = \cup_i \{x : (x, y_0) \in C_i\} \in \mathcal{B},$$

    i.e., $\cup_i C_i \in \mathcal{G}$
- Check that any closed cube $[a, b] \times [c, d] \in \mathcal{G}$
- $\mathcal{O}^2 \subset \mathcal{G}$. We need a theorem: Every open subset of $\mathcal{R}^d$ can be written as a countable union of closed cubes. (see p7 of *Real Analysis*)