1. Preliminaries of Probability Theory

Author: FT Z
Time: 08/07/2024
Last Update: 14/07/2024

Reference

1. Probability: Theory and Examples (Rick Durrett).

Table of Content

1. Probability Spaces
2. Distributions
3. Random Variables
4. Integration
5. Properties of the Integral
6. Expected Value
7. Product Measures, Fubini's Theorem

Go to the bottom of the file

1. Probability Spaces

Go to the table of content.
Go to the bottom of the file

Notations

1. $(\Omega, \mathcal{F}, P)$: Probability space.
2. $\mathcal{R}^d$: The smallest σ -field containing all the open sets in $\mathbb{R}^d$ (Borel algebra). When $d = 1$, we drop superscript.
3. $a_n \downarrow c$: a_n decreases monotonically to c.
4. $a_n \uparrow c$: a_n increases monotonically to c.
5. $\Delta_A F$: $A = (a_1, b_1] \times \cdots \times (a_d, b_d]$, $V = \{a_1, b_1\} \times \cdots \times \{a_d, b_d\}$, $-\infty < a_i < b_i < +\infty$, $\forall v \in V$, let $\text{sgn}(v) = (-1)^{\text{number of } a \text{ in } v}$, $\Delta_A F = \sum\limits_{v \in V} \text{sgn}(v) F(v)$.

Definitions

1. **Probability Space, Probability Measure:**
   1. Suppose $(\Omega, \mathcal{F}, P)$ is a measure space. If $\Omega \in \mathcal{F}$ and $P(\Omega) = 1$, then we call $(\Omega, \mathcal{F}, P)$ a **probability space**, P a **probability measure**.
   2. *Remark: Ω is a set of "outcomes", $\mathcal{F}$ is a set of "events", P is a function that assigns probabilities to events.*
2. **Semialgebra:**
   1. A collection $\mathcal{S}$ of sets is said to be a semialgebra if it satisfy:
      1. It is closed under intersection, i.e. $S, T \in \mathcal{S}$ implies $S \bigcap T \in \mathcal{S}$;
      2. If $S \in \mathcal{S}$, then S^c is a finite disjoint union of sets in $\mathcal{S}$.
3. **Stieltjes Measure Function:**
   1. A **Stieltjes measure function** is defined on $\mathbb{R}$ and has the following properties:
      1. F is nondecreasing;
      2. F is right continuous.

Theorems

1. **Constructing A Measure Using A Stieltjes Measure Function (1 Dimensional Case):**
   1. **Statement:**
      1. Associated with each Stieltjes measure function F there is a unique measure μ on $(\mathbb{R}, \mathcal{R})$ with $\mu((a, b]) = F(b) - F(a)$.
2. **Constructing A Measure Using A Stieltjes Measure Function (High Dimensional Case):**
   1. **Statement:**
      1. Suppose $F : \mathbb{R}^d \to [0, 1]$ satisfy:
         1. F is nondecreasing;
         2. F is right continuous;
         3. If $x_n \downarrow -\infty$ (each coordinate does), then $F(x_n) \downarrow 0$. If $x_n \uparrow +\infty$ (each coordinate does), then $F(x_n) \uparrow 1$.

4. $\forall$ Finite rectangles $A,\ \Delta_A F \geq 0$.

   Then there is a unique probability measure μ on $(\mathbb{R}^d, \mathcal{R}^d)$ so that $\mu(A) = \Delta_A F$ for all finite rectangles.

2. **Example:**
   1. $F(x) = \prod_{i=1}^{d} F_i(x)$, where F_i are Stieltjes measure functions. In this case, $\Delta_A F = \prod_{i=1}^{d} (F_i(b_i) - F_i(a_i))$.

2. Distributions

Go to the table of content.
Go to the bottom of the file

Basic Assumptions

1. $(\Omega, \mathcal{F}, P)$ is a probability space.

Notations

1. $X,\ Y$: Random variable(s).
2. $X \in \mathcal{F}$: X is $\mathcal{F}$ -measurable. (See Definition 1.2)
3. $\mathcal{R}^d$: Borel algebra based on $\mathbb{R}^d$.
4. F: Distribution function of X.
5. F^{-1}: The random variable X. (Even though F may not be 1-1 map, we will call X the inverse of F and denote it by F^{-1})
6. $f_X(x)$: The density function of X.
7. $\mu_1 \perp \mu_2$: μ_1 and μ_2 are measures and they are mutually singular.

Definitions

1. **Random Variable:**
   1. A real-valued function X defined on Ω is said to be a **random variable** if for every Borel set $B \subset \mathbb{R}$ we have $X^{-1}(B) = \{\omega | X(\omega) \in B\} \in \mathcal{F}$.
   2. When we want to emphasize $\mathcal{F}$ is a σ -field, we also say X is $\mathcal{F}$ -measurable or write $X \in \mathcal{F}$.
   3. *Remark:*
      1. *X is defined on Ω instead of $\mathcal{F}$.*
2. **Distribution:**
   1. If X is a random variable, then X induces a probability measure on $\mathbb{R}$ called its **distribution** by setting $\mu(A) = P(X \in A)$, where A is a Borel set. (Note that X is a real-valued function)
   2. Using the notation introduced previously, $P(X \in A)$ can also be written as $P(X^{-1}(A))$.
3. **Distribution Function:**
   1. The distribution of X is usually described by giving its **distribution function**, $F(x) = P(X \leq x),\ x \in \mathbb{R}_*$.
4. **(For Random Variables) Equal In Distribution:**
   1. If X and Y induce the same distribution μ on $(\mathbb{R}, \mathcal{R})$, we say X and Y are **equal in distribution**.
5. **Density Function:**
   1. When the distribution function $F(x) = P(X \leq x)$ has the form $F(x) = \int_{-\infty}^{x} f(y)dy$, we say that X has **density function** f.
   2. *Remark:* $P(X = x) = \lim_{\epsilon \to 0} \int_{x-\epsilon}^{x+\epsilon} f(y)dy = 0$.
6. **(Measure) Mutually Singular, Singular With Respect To:**
   1. Two measures μ_1 and μ_2 are said to be **mutually singular** if there exists a set A with $\mu_1(A) = 0$ and $\mu_2(A^c) = 0$.
   2. In this case, we also say μ_1 is **singular with respect to** μ_2 and write $\mu_1 \perp \mu_2$.
7. **Absolutely Continuous Distribution Function On $\mathbb{R}$:**
   1. A distribution function on $\mathbb{R}$ is said to be **absolutely continuous** if it has a density.
8. **Singular Distribution Function On $\mathbb{R}$:**
   1. A distribution function on $\mathbb{R}$ is said to be **singular** if the corresponding measure is singular with respect to Lebesgue measure.
9. **Discrete Probability Measure:**
   1. A probability measure P is said to be **discrete** if there is a countable set S with $P(S^c) = 0$.

Theorems

1. **Properties of Distribution Function:**
   1. **Statement:**
      1. Any distribution function F has the following properties:
         1. F is nondecreasing;
         2. F is right continuous
         3. $\lim_{x \to +\infty} F(x) = 1,\ \lim_{x \to -\infty} F(x) = 0;$

4. If $F(x^-) = \lim_{y \uparrow x} F(y)$, then $F(x^-) = P(X < x)$;
5. $P(X = x) = F(x) - F(x^-)$.

2. **Distribution Function Induces A Random Variable:**
   1. **Statement:**
      1. If F has properties 1, 2, 3 in Theorem 1, then it is the distribution function of some random variable.
   2. **Idea:**
      1. Suppose the events can be represented by numbers, we hope to use the fact that $P(\{\omega|\omega \leq F(x)\}) = F(x)$.
   3. **Proof Steps:**
      1. Let $\Omega = (0,1)$, $\mathcal{F} =$ Borel Algebra, P is Lebesgue measure. If $\omega \in (0,1)$, let $X(\omega) = \sup\{y|F(y) < \omega\}$.
      2. Once we show that $\{\omega|X(\omega) \leq x\} = \{\omega|\omega \leq F(x)\}$, the desired result follows immediately since $P(\{\omega|\omega \leq F(x)\}) = F(x)$.
      3. To check the result in step 2, we observe that if $\omega \leq F(x)$, since F is nondecreasing, then $x \geq \sup\{y|F(y) < \omega\}$, so $X(\omega) \leq x$. Therefore $\{\omega|X(\omega) \leq x\} \supset \{\omega|\omega \leq F(x)\}$
      4. On the other hand, if $\omega > F(x)$, then since F is right-continuous, there exists a $\epsilon > 0$ so that $F(x+\epsilon) < \omega$ and $X(\omega) \leq x + \epsilon > x$, this means $\{\omega|X(\omega) \leq x\}^c \supset \{\omega|\omega \leq F(x)\}^c$. Therefore $\{\omega|X(\omega) \leq x\} \subset \{\omega|\omega \leq F(x)\}$.
3. **An Inequality Related to The Normal Distribution:**
   1. **Statement:**
      1. For $x > 0$, $(x^{-1} - x^{-3})\exp(-\frac{x^2}{2}) \leq \int_x^{+\infty} \exp(-\frac{y^2}{2})dy \leq x^{-1}\exp(-\frac{x^2}{2})$

Examples

1. **A Singular Distribution: Uniform Distribution on the Cantor Set:**
   1. On the Cantor set C, we set $F(x) = 0$ for $x \leq 0$, $F(x) = 1$ for $x \geq 1$, $F(x) = \frac{1}{2}$ for $x \in [\frac{1}{3}, \frac{2}{3}]$, $F(x) = \frac{1}{4}$ for $x \in [\frac{1}{9}, \frac{2}{9}]$, $F(x) = \frac{3}{4}$ for $x \in [\frac{7}{9}, \frac{8}{9}]$, ... There is no f can be density function of F because such an f would be equal to 0 on a set of measure 1. From the definition, it is immediate that the corresponding measure has $\mu(C^c) = 0$, so F is singular with respect to Lebesgue measure.

3. Random Variables

Go to the table of content.
Go to the bottom of the file

Basic Assumptions

1. "$\mathcal{A}$ generates $\mathcal{S}$" requires $\mathcal{S}$ to be σ -field/ σ -algebra (while only ring or σ -ring).

Notations

1. r.v. : Random variable.
2. X: A map from $(\Omega, \mathcal{F})$ to $(S, \mathcal{S})$. $\mathcal{F}$ and $\mathcal{S}$ are collection of subsets of Ω, S respectively.
3. $\mathcal{R}^d$: Borel algebra based on $\mathbb{R}^d$.
4. $\{X \in B\}$: $B \in \mathcal{S}, \{X \in B\} = \{\omega \in \Omega|X(\omega) \in B\}$.
5. $\sigma(X)$: The sigma-field generated by X.
6. X_∞: $\lim_{n \to +\infty} X_n$.
7. a.s. : Almost surely.

Definitions

1. **Measurable Map, Random Vector, Random Variable:**
   1. A function $X: \Omega \to S$ is said to be a **measurable map** from $(\Omega, \mathcal{F})$ to $(S, \mathcal{S})$ if $\forall B \in \mathcal{S},\ X^{-1}(B) = \{\omega|X(\omega) \in B\} \in \mathcal{F}$.
   2. When $(S, \mathcal{S}) = (\mathbb{R}^d, \mathcal{R}^d)$ and $d > 1$, X is called a **random vector**.
   3. When $(S, \mathcal{S}) = (\mathbb{R}, \mathcal{R})$, X is called a **random variable**. (See also Definition 1.1 in Section 2)
2. **σ -Field Generated by Random Variable:**
   1. The smallest σ -field on Ω that makes X a measurable map, i.e. $\{\{X \in B\}|\forall B \in \mathcal{S}\}$, where $\mathcal{S}$ is a σ -field.
3. **(Random Variables) Converges Almost Surely:**
   1. From Theorem 4, we could know that $\Omega_o = \{\omega|\ \lim_{n \to +\infty} X_n \text{ exists}\} = \{\omega|\limsup_{n \to +\infty} X_n - \liminf_{n \to +\infty} X_n = 0\}$ is a measurable set. If $P(\Omega_o) = 1$, we say that $\{X_n\}_{n=1}^\infty$ **converges almost surely**.
   2. *Remark: This type of convergence is called "convergent almost everywhere" in measure theory.*

Theorems

1. **Methods for Determining the Measurability of A Mapping:**
   1. **Statement:**
      1. If $\forall A \in \mathcal{A},\ \{\omega|X(\omega) \in A\} \in \mathcal{F}$, and $\mathcal{A}$ generates $\mathcal{S}$ (i.e. $\mathcal{S}$ is the smallest σ -field that contains $\mathcal{A}$), then X is measurable.
   2. **Proof Steps:**

1. We need to prove $\forall B \in \mathcal{S}, X^{-1}(B) \in \mathcal{F}$, and we already know $\forall A \in \mathcal{A}, X^{-1}(A) \in \mathcal{F}$, while $\mathcal{A} \subseteq \mathcal{S}$.
2. $\{X \in \bigcup_{i=1}^{\infty} B_i\} = \bigcup_{i=1}^{\infty} \{X \in B_i\}$, $\{X \in B^c\} = \{X \in B\}^c$, so the class of sets $\mathcal{B} = \{B | \{X \in B\} \in \mathcal{F}\}$ is a σ -field.
3. Since $\mathcal{B} \supset \mathcal{A}$ and $\mathcal{A}$ generates $\mathcal{S}$, $\mathcal{B} \supset \mathcal{S}$.

3. *Remark:*
    1. *Utilizing the generated algebraic construct is the smallest one, we want to construct a set with target property, then prove this set is superset of target set, hence prove the statement. See also* Proof Steps of "The Uniqueness Theorem for Outer Extension" in Chapter 1, Section 3.
    2. *From these proof steps we could know that if $\mathcal{S}$ is a σ -field, then $\{\{X \in B\} | B \in \mathcal{S}\}$ is a σ -field. It is the smallest σ -field on Ω that makes X a measurable map.*

2. **Measurability of Composite Mappings:**
    1. **Statement:**
        1. If $X : (\Omega, \mathcal{F}) \to (S, \mathcal{S})$ and $f : (S, \mathcal{S}) \to (T, \mathcal{T})$ are measurable maps, then $f(X)$ is a measurable map from $(\Omega, \mathcal{F})$ to $(T, \mathcal{T})$.
3. **A Measurable Mapping of A Random Variable is Still A Random Variable:**
    1. **Statement:**
        1. If $X_1, \ldots, X_n$ are random variables and $f : (\mathbb{R}^n, \mathcal{R}^n) \to (\mathbb{R}, \mathcal{R})$ is measurable, then $f(X_1, \ldots, X_n)$ is a random variable.
4. **Different Limits of A Sequence of Random Variables are Still Random Variables:**
    1. **Statement:**
        1. If $\{X_i\}_{i=1}^{\infty}$ is a sequence of random variables, then $\inf_n X_n$, $\sup_n X_n$, $\limsup_n X_n$, $\liminf_n X_n$ are also random variables.
    2. **Proof Steps:**
        1. Refer to "The Limit of A Sequence of Measurable Functions" in Chapter 0, Section 5.
        2. Note that
            1. $\{\inf_n X_n < a\} = \bigcup_{n=1}^{\infty} \{X_n < a\} \in \mathcal{F}$
            2. $\{\sup_n X_n > a\} = \bigcup_{n=1}^{\infty} \{X_n > a\} \in \mathcal{F}$
            3. $\liminf_{n\to\infty} X_n = \sup_n \left(\inf_{m \geq n} X_m \right)$
            4. $\limsup_{n\to\infty} X_n = \inf_n \left(\sup_{m \geq n} X_m \right)$
    3. *Remark:*
        1. To see the meaning of $\liminf_{n\to\infty} X_n, \limsup_{n\to\infty} X_n$, refer to Notation 6 and Notation 7 in Section 5, Chapter 0.

Examples

1. **Example for $\mathcal{A}$ and $\mathcal{S}$ in Theorem 1:**
    1. If $(S, \mathcal{S}) = (\mathbb{R}, \mathcal{R})$, then $\mathcal{A} = \{(-\infty, x] | x \in \mathbb{R}\}$ or $\{(-\infty, x) | x \in \mathbb{Q}\}$.
    2. If $(S, \mathcal{S}) = (\mathbb{R}^d, \mathcal{R}^d)$, then $\mathcal{A} = \{(a_1, b_1) \times \cdots \times (a_d, b_d) | -\infty < a_i < b_i < +\infty\}$.

Exercises

1. **Statement:** Show that if $\mathcal{A}$ generates $\mathcal{S}$, then $X^{-1}(\mathcal{A}) = \{\{X \in A\} | \forall A \in \mathcal{A}\}$ generates $\sigma(X) = \{\{X \in B\} | \forall B \in \mathcal{S}\}$.
    1. **Proof Steps:**
        1. From the property of generated σ -ring, we know that $\sigma(X) \subseteq \sigma(X^{-1}(\mathcal{A}))$.
        2. For all $E \in \sigma(X^{-1}(\mathcal{A}))$, there must exist a sequence of sets in $\mathcal{A}$: $\{A_n\}_{n=1}^{\infty}$ satisfy $\bigcup_{n=1}^{\infty} X^{-1}(A_n) = E$. Since $\bigcup_{n=1}^{\infty} X^{-1}(A_n) = X^{-1}(\bigcup_{n=1}^{\infty} A_n)$, $\bigcup_{n=1}^{\infty} A_n \in \mathcal{S}$, $E \in \sigma(X)$. Therefore $\sigma(X^{-1}(\mathcal{A})) \subseteq \sigma(X)$.

4. Integration

Go to the table of content.
Go to the bottom of the file

Basic Assumptions

1. μ is a σ -finite measure on $(\Omega, \mathcal{F})$.
2. $(\Omega, \mathcal{F}, \mu)$ is a measure space.

Notations

1. $a \wedge b$: $a, b \in \mathbb{R}, a \wedge b = \min(a, b)$.
2. $\mathcal{R}^d$: Borel algebra based on $\mathbb{R}^d$.
3. Notation of integrals for special cases:
    1. When $(\Omega, \mathcal{F}, \mu) = (\mathbb{R}^d, \mathcal{R}^d, m)$, we write $\int f(x)dx$ for $\int f dm$;

2. When $(\Omega, \mathcal{F}, \mu) = (\mathbb{R}, \mathcal{R}, m)$ and $E = [a, b]$, we write $\int_a^b f(x)dx$ for $\int_E fdm$;
3. When $(\Omega, \mathcal{F}, \mu) = (\mathbb{R}, \mathcal{R}, \mu)$ with $\mu((a, b]) = G(b) - G(a)$ for $a < b$, we write $\int f(x)dG(x)$ for $\int fd\mu$;
4. When Ω is a countable set, $\mathcal{F}$ is collection of subsets of Ω, and μ is counting measure, we write $\sum\limits_{i\in\Omega} f(i)$ for $\int fd\mu$.

Theorems

1. **Approximation of A Sequence of Integrals:**
   1. **Statement:**
      1. Suppose $f \geq 0$, let $E_n \uparrow \Omega$ have $\mu(E_n) < +\infty$. Then $\int_{E_n}(f \wedge n)d\mu \uparrow \int_\Omega fd\mu$ as $n \uparrow \infty$. The definition of f can be seen in Definition 2.2.1 in Chapter 0 Section 7.
   2. **Proof Steps:**
      1. The left-hand side increases as n does. Since $h = (f \wedge n)1_{E_n}$ is a probability in the sup, each term is smaller than the integral on the right.
      2. To prove that the limit is $\int_\Omega fd\mu$, observe that if $0 \leq h \leq f,\ h \leq M$, and $\mu(\{x|h(x) > 0\}) < +\infty$, then for $n \geq M$ using $h \leq M$, $\int_{E_n}(f \wedge n)d\mu \geq \int_{E_n} hd\mu = \int_\Omega hd\mu - \int_{E_n^c} hd\mu$.
      3. Now $0 \leq \int_{E_n^c} hd\mu \leq M\mu(E_n^c \bigcap \{x|h(x) > 0\}) \to 0$ as $n \to +\infty$, so $\liminf\limits_{n\to+\infty} \int_{E_n}(f \wedge n)d\mu \geq \int_\Omega hd\mu$, which proves the desired result by the definition of the integral of f.

5. Properties of the Integral

Go to the table of content.
Go to the bottom of the file

Basic Assumptions

1. $\int fd\mu$ represents the integral over the entire domain.

Notations

1. $\|f\|_p$: $(\int |f|^p d\mu)^{\frac{1}{p}}$ for $1 \leq p < +\infty$.

Theorems

1. **Jensen's Inequality:**
   1. **Statement:**
      1. Suppose φ is convex, if μ is a probability measure, and f and $\varphi(f)$ are integrable, then $\varphi(\int fd\mu) \leq \int \varphi(f)d\mu$.
   2. **Proof Steps:**
      1. Let $c = \int fd\mu$ and let $l(x) = ax + b$ be a linear function that has $l(c) = \varphi(c)$ and $\varphi(x) \geq l(x)$.
      2. To see such a function exists, recall that convexity implies $\lim\limits_{h\downarrow 0} \frac{\varphi(c)-\varphi(c-h)}{h} \leq \lim\limits_{h\downarrow 0} \frac{\varphi(c+h)-\varphi(c)}{h}$, the limit exist since the sequence are monotone. If we let a be any number between the two limits and let $l(x) = a(x - c) + \varphi(c)$, then l has the desired properties.
      3. $\int \varphi(f)d\mu \geq \int(af + b)d\mu = a\int fd\mu + b = l(\int fd\mu) = \varphi(\int fd\mu)$.
2. **Holder's Inequality:**
   1. **Statement:**
      1. If $p, q \in (1, +\infty)$ with $\frac{1}{p} + \frac{1}{q} = 1$, then $\int |fg|d\mu \leq \|f\|_p\|g\|_q$.
   2. **Proof Steps:**
      1. If $\|f\|_p = 0$ or $\|g\|_q = 0$, then $|fg| = 0,\ a.e.$, so it suffices to prove the result when $\|f\|_p$ and $\|g\|_q > 0$ or by dividing both sides by $\|f\|_p\|g\|_q$, when $\|f\|_p = \|g\|_q = 1$.
      2. Fix $y \geq 0$ and let $\varphi(x) = \frac{x^p}{p} + \frac{y^q}{q} - xy$ for $x \geq 0$, $\varphi'(x) = x^{p-1} - y$, $\varphi''(x) = (p-1)x^{p-2}$, so φ has a minimum at $x_o = y^{\frac{1}{p-1}}$.
      3. Since $q = \frac{p}{p-1}$ and $x_o^p = y^{\frac{p}{p-1}} = y^q$, so $\varphi(x_o) = y^q(\frac{1}{p} + \frac{1}{q}) - y^{\frac{1}{p-1}}y = 0$.
      4. Since x_o is the minimum, it follows that $xy \leq \frac{x^p}{p} + \frac{y^q}{q}$.
      5. Letting $x = |f|, y = |g|$, and integrating $\int |fg|d\mu \leq \frac{1}{p} + \frac{1}{q} = 1 = \|f\|_p\|g\|_q$.

6. Expected Value

Go to the table of content.
Go to the bottom of the file

Basic Assumptions

1. $\int fd\mu$ represents the integral over the entire domain.
2. $(\Omega, \mathcal{F}, P)$ is probability space.
3. X is a random variable on $(\Omega, \mathcal{F}, P)$
4. Unless specified, $n \to \infty$ means n goes to $+\infty$.

Notations

1. $EX, E(X)$: Expected value of X.
2. $E(X; A)$: $\int_A XdP$.
3. $\mathcal{R}^d$: Borel algebra based on $\mathbb{R}^d$.
4. $a \wedge b : \min(a, b)$

Definitions

1. **Expected Value/ Mean of Random Variables:**
   1. If $X \geq 0$, we define its **expected value** to be $EX = \int XdP$ (equals to ∞ is allowed).
   2. For general case, $EX = EX^+ - EX^-$.
   3. EX is often called the **mean** of X.
   4. *Remark: The element of* $\{X \in A\}$ *is not the value of* X *but the event in* Ω*:* $\{X \in A\} = \{\omega \in \Omega | X(\omega) \in A\}$.
2. **Expected Value Exists:**
   1. For general case, we declare that EX **exists** whenever the subtraction makes sense, i.e., $EX^+ < +\infty$ or $EX^- < +\infty$.
3. **(kth) Moment, Mean, Variance of Random Variables:**
   1. If k is a positive integer, then EX^k is called the **kth moment** of X.
   2. The first moment EX is usually called the **mean**.
   3. If $EX^2 < \infty$, then the **variance** of X is defined to be $Var(X) = E(X - EX)^2$.
   4. *Remark:* $\text{Var}(X) = EX^2 - (EX)^2$.

Theorems

1. **Basic Properties of Expectation:**
   1. **Statement:**
      1. Suppose $X, Y \geq 0$ or $E|X|, E|Y| < +\infty$,
         1. $E(X + Y) = EX + EY$;
         2. $E(aX + b) = aEX + b$ for any real numbers a, b;
         3. If $X \geq Y$, then $EX \geq EY$.
      2. $EX < \infty$ implies $X < \infty\ a.s.$
2. **Jensen's Inequality:**
   1. **Statement:**
      1. Suppose φ is convex, then $E(\varphi(X)) \geq \varphi(EX)$ provided both expectations exist, i.e., $E|X|$ and $E|\varphi(X)| < +\infty$.
   2. **Corollary:**
      1. $|EX| \leq E|X|$, $(EX)^2 \leq E(X^2)$.
3. **Holder's Inequality:**
   1. **Statement:**
      1. If $p, q \in [1, \infty]$ with $\frac{1}{p} + \frac{1}{q} = 1$, then $E|XY| \leq \|X\|_p \|Y\|_q$. Here $\|X\|_r = (E|X|^r)^{\frac{1}{r}}$ for $r \in [1, \infty)$; $\|X\|_\infty = \inf\{M | P(|X| > M) = 0\}$.
4. **Chebyshev's Inequality:**
   1. **Statement:**
      1. Suppose $\varphi : \mathbb{R} \to \mathbb{R}$ has $\varphi \geq 0$, let $A \in \mathcal{R}$ and $i_A = \inf\{\varphi(y) | y \in A\}$, then $i_A P(X \in A) \leq E(\varphi(X); X \in A) \leq E\varphi(X)$.
   2. **Proof Steps:**
      1. The definition of i_A and the fact that $\varphi \geq 0$ imply that $i_A 1_{(X \in A)} \leq \varphi(X) 1_{(X \in A)} \leq \varphi(X)$.
      2. Taking expected values.
   3. *Remark:*
      1. *Some authors call this result "**Markov's inequality**".*
      2. *They use the name Chebyshev's inequality for the special case in which* $\varphi(x) = x^2, A = \{x | \, |x| \geq a\}$*:* $a^2 P(|X| \geq a) \leq EX^2$.
5. **Interchange Limits and Integrals (Theorem on Convergence of Integrals, Expected Value Version):**
   1. **Statement:**
      1. **Fatou's Lemma:**
         1. If $X_n \geq 0$,then $\liminf_{n \to \infty} EX_n \geq E(\liminf_{n \to \infty} X_n)$.
      2. **Monotone Convergence Theorem:**
         1. If $0 \leq X_n \uparrow X$, then $EX_n \uparrow EX$.
      3. **Dominated Convergence Theorem:**
         1. If $X_n \to X\ a.s.$, $|X_n| \leq Y$ for all n, and $EY < +\infty$, then $EX_n \to EX$.
      4. **Bounded Convergence Theorem:**
         1. The special case of dominated convergence theorem in which Y is constant.

6. **Another Result on Integration to the Limit:**
    1. **Statement:**
        1. Suppose $X_n \to X\ a.s.$ Let g, h be continuous functions with $g \geq 0$ and $g(x) \to \infty$ as $|x| \to \infty$; $\frac{|h(x)|}{g(x)} \to 0$ as $|x| \to \infty$; $Eg(X_n) \leq K < +\infty$ for all n. Then $Eh(X_n) \to Eh(X)$.
    2. **Idea:**
        1. We could take up a method similar to "3ϵ " principle.
        $A \to B \iff |A - B| < \epsilon \Leftarrow |A - A_1| < \frac{\epsilon}{3}, |A_1 - B_1| < \frac{\epsilon}{3}, |B_1 - B| < \frac{\epsilon}{3} \iff |A_1 - A| < \frac{\epsilon}{3}, |A_1 - B_1| < \frac{\epsilon}{3}, |B_1 - B| < \frac{\epsilon}{3}$.
    3. **Proof Steps:**
        1. By subtracting a constant from h, we can suppose without loss of generality that $h(0) = 0$.
        2. Pick M large so that $P(|X| = M) = 0$ and $g(x) > 0$ when $|x| \geq M$.
        3. Given a random variable Y, let $\overline{Y} = Y\mathbb{1}(|Y| \leq M)$. Since $X_n \to X, P(|X| = M) = 0$, then $\overline{X_n} \to \overline{X},\ a.s.$ Since $\overline{X_n}$ is bounded, h is continuous, $h(\overline{X_n})$ is also bounded, and it follows from the bounded convergence theorem that $Eh(\overline{X_n}) \to Eh(\overline{X})$.
        4. To control the error, we use the following: $|Eh(\overline{Y}) - Eh(Y)| \leq E|h(\overline{Y}) - h(Y)| \leq E(|h(Y)|; |Y| > M) \leq \epsilon_M Eg(Y)$ where $\epsilon_M = \sup\left\{\frac{|h(x)|}{g(x)} \middle| |x| \geq M\right\}$. Taking $Y = X_n$ and using $Eg(X_n) \leq K < +\infty$, it follows that $|Eh(\overline{X_n}) - Eh(X_n)| \leq K\epsilon_M$.
        5. To estimate $|Eh(\overline{X}) - Eh(X)|$, we observe that $g \geq 0$ and g is continuous, so Fatou's lemma implies $Eg(X) \leq \liminf_{n\to\infty} Eg(X_n) \leq K$. Taking $Y = X$, we have $|Eh(\overline{X}) - Eh(X)| \leq K\epsilon_M$.
        6. The triangle inequality implies $|Eh(X_n) - Eh(X)| \leq |Eh(X_n) - Eh(\overline{X_n})| + |Eh(\overline{X_n}) - Eh(\overline{X})| + |Eh(\overline{X}) - Eh(x)|$. Taking limits, we have $\limsup_{n\to\infty} |Eh(X_n) - Eh(X)| \leq 2K\epsilon_M$.
7. **Change of Variables Formula:**
    1. **Statement:**
        1. Let X be a random element of $(S, \mathcal{S})$ with distribution μ, i.e., $\mu(A) = P(X \in A)$. If f is a measurable function from $(S, \mathcal{S})$ to $(\mathbb{R}, \mathcal{R})$ so that $f \geq 0$ or $E|f(X)| < \infty$, then $Ef(X) = \int_S f(y)\mu(dy)$.
    2. **Idea:**
        1. We will prove this result by verifying it in four increasingly more general special cases that similar to the way that the integral was defined.
    3. **Proof Steps:**
        1. Case 1: For indicator functions: If $B \in \mathcal{S}$ and $f = \mathbb{1}_B$, then $E\mathbb{1}_B(X) = P(X \in B) = \mu(B) = \int_S \mathbb{1}_B(y)\mu(dy)$. (Recall that μ is the measure in $(S, \mathcal{S})$).
        2. Case 2: For simple functions: Let $f(x) = \sum_{m=1}^{n} c_m \mathbb{1}_{B_m}$ where $c_m \in \mathbb{R}, B_m \in \mathcal{S}$. The linearity of expected value, the result of Case 1, and the linearity of integration imply $Ef(X) = \sum_{m=1}^{n} c_m E\mathbb{1}_{B_m}(X) = \sum_{m=1}^{n} c_m \int_S \mathbb{1}_{B_m}(y)\mu(dy) = \int_S f(y)\mu(dy)$.
        3. Case 3: For nonnegative functions: Now if $f \geq 0$ and we let $f_n(x) = (\frac{[2^n f(x)]}{2^n}) \wedge n$, then the f_n are simple and $f_n \uparrow f$, so using the result for simple functions and the monotone convergence theorem:
        $Ef(X) = \lim_n Ef_n(X) = \lim_n \int_S f_n(y)\mu(dy) = \int_S f(y)\mu(dy)$.
        4. Case 4: For integrable functions: The general case now follows by writing $f(x) = f^+(x) - f^-(x)$. The condition $E|f(x)| < \infty$ guarantees that $Ef^+(X)$ and $Ef^-(X)$ are finite. So
        $Ef(X) = Ef^+(X) - Ef^-(X) = \int_S f(y)^+\mu(dy) - \int_S f^-(y)\mu(dy) = \int_S f(y)\mu(dy)$.
    4. *Remark:*
        1. *We should note the method described in Idea, since it will be used several times. This is because the definitions of the integral of a simple function and a nonnegative function are different from the integral of a general function.*
        2. *We also need to pay attention to the way we construct the simple function to approximate the original function.*
        3. *We usually apply this statement with $S = \mathbb{R}^d$ for doing calculus. Therefore we can compute expected values of functions of random variables by performing integrals on the real line.*

7. Product Measures, Fubini's Theorem

Go to the table of content.
Go to the bottom of the file

Basic Assumptions

1. Let $(X, \mathcal{A}, \mu_1)$ and $(Y, \mathcal{B}, \mu_2)$ be two σ -finite measure spaces. Let $\Omega = X \times Y = \{(x, y) | x \in X, y \in Y\}$, $\mathcal{S} = \{A \times B | A \in \mathcal{A}, B \in \mathcal{B}\}$. Let $\mathcal{F} = \mathcal{A} \times \mathcal{B}$ be the σ -algebra generated by $\mathcal{S}$.

Notations

1. $\mathcal{A} \times \mathcal{B}$: $\sigma(\{A \times B | A \in \mathcal{A}, B \in \mathcal{B}\})$.

Theorems

1. **The Uniqueness of Product Measure:**

1. **Statement:**
   1. There is a unique measure μ on $\mathcal{F}$ with $\mu(A \times B) = \mu_1(A)\mu_2(B)$.
2. **Proof Steps:**
   1. By Theorem 7.1 in Chapter 0, Section 3, it is enough to show that μ is σ -finite, namely if $A \times B = \bigcup_{i=1}^{\infty}(A_i \times B_i)$ is a finite or countable disjoint union, then $\mu(A \times B) = \sum_{i=1}^{\infty} \mu(A_i \times B_i)$.
   2. For each $x \in A$, let $I(x) = \{i | x \in A_i\}, \ B = \bigcup_{i \in I(x)} B_i$ is a disjoint union, so $\mathbb{1}_A(x)\mu_2(B) = \sum_{i \in I(x)} \mathbb{1}_{A_i}(x)\mu_2(B_i)$.
   3. Integrating with respect to μ_1 and using monotone convergence theorem gives $\mu_1(A)\mu_2(B) = \sum_{i=1}^{\infty} \mu_1(A_i)\mu_2(B_i)$. (If $g_m \geq 0$, then $\int \sum_{m=0}^{\infty} g_m d\mu = \sum_{m=0}^{\infty} \int g_m d\mu$)
3. **Corollary:**
   1. Using induction, it follows that if $(\Omega_i, \mathcal{F}_i, \mu_i), \ i = 1, \ldots, n$, are σ -finite measure spaces and $\Omega = \Omega_1 \times \cdots \Omega_n$, there is a unique measure μ on the σ -algebra $\mathcal{F}$ generated by sets of the form $A_1 \times \cdots \times A_n, \ A_i \in \mathcal{F}_i$, which has $\mu(A_1 \times \cdots \times A_n) = \prod_{m=1}^{n} \mu_m(A_m)$.
4. *Remark:*
   1. *This theorem is a special case of* Theorem 5.1.3 in Chapter 0, Section 9. *The proof can be seen on the page 297 of* Reference 1 for Chapter 0.
   2. μ *is often denoted by* $\mu_1 \times \mu_2$.
   3. *In* Corollary 1, *when* $(\Omega_i, \mathcal{F}_i, \mu_i) = (\mathbb{R}, \mathcal{R}, \lambda)$ *for all* i, *the result is Lebesgue measure on the Borel subsets of* n *dimensional Euclidean space on* $\mathbb{R}^n$.

2. **Fubini's Theorem:**
   1. **Statement:**
      1. (Based on the notation of Basic Assumption 1 and Theorem 1) If $f \geq 0$ or $\int |f| d\mu < \infty$, then $\int_X \int_Y f(x,y)\mu_2(dy)\mu_1(dx) = \int_{X \times Y} f d\mu = \int_Y \int_X f(x,y)\mu_1(dx)\mu_2(dy)$.

End

Go to the table of content