Math 639: Lecture 1

Measure theory background

Bob Hough

January 24, 2017

Probability spaces

Definition

A *probability space* is a measure space $(\Omega, \mathscr{F}, \text{Prob})$ with Prob a positive measure of mass 1.

- Ω is called the *sample space*, and $\omega \in \Omega$ are called outcomes.
- $\mathscr{F}$, a σ-algebra, is called the *event space*, and $A \in \mathscr{F}$ are called *events*.

Algebras of sets

Definition

A collection of sets $\mathscr{S}$ is a *semialgebra* if

- If $S, T \in \mathscr{S}$ then $S \cap T \in \mathscr{S}$
- If $S \in \mathscr{S}$ then S^c is the finite disjoint union of sets of $\mathscr{S}$.

Example

The empty set together with those sets

$$(a_1, b_1] \times \cdots \times (a_d, b_d] \subset \mathbb{R}^d, \qquad -\infty \leq a_i < b_i \leq \infty$$

form a semialgebra in $\mathbb{R}^d$.

Algebras of sets

Definition

A collection of sets $\mathscr{S}$ is an *algebra* if it is closed under complements and intersections.

Lemma

If $\mathscr{S}$ is a semialgebra, then $\overline{\mathscr{S}}$, given by finite disjoint unions from $\mathscr{S}$, is an algebra.

Definition

A σ-algebra of sets is an algebra which is closed under countable unions.

Borel σ-algebra

Definition

Given a collection of subsets $A_\alpha \subset \Omega$, the *generated σ-algebra* $\sigma(\{A_\alpha\})$ is the smallest σ-algebra containing $\{A_\alpha\}$.

Definition

In the case that Ω has a topology $\mathscr{T}$ of open sets, the *Borel σ-algebra* is $\sigma(\mathscr{T})$.

Borel σ-algebra

Definition

The product of measure spaces $(\Omega_i, \mathscr{F}_i)$, $i = 1, ..., n$ is the set $\Omega = \Omega_1 \times ... \times \Omega_n$ with the σ-algebra $\mathscr{F}_1 \times ... \times \mathscr{F}_n = \sigma\left(\bigcup_{i=1}^n \mathscr{F}_i\right)$.

Exercise

Let $d \geq 1$. With the usual topologies, the Borel σ-algebra $\mathscr{B}_{\mathbb{R}^d}$ is equal to $\mathscr{B}_{\mathbb{R}} \times ... \times \mathscr{B}_{\mathbb{R}}$ (d copies).

Dynkin's $\pi - \lambda$ Theorem

Definition

A π-*system* is a collection $\mathscr{P}$ of sets closed under finite intersections. A λ-*system* is a collection $\mathscr{L}$ of sets satisfying the following

- $\Omega \in \mathscr{L}$
- For any $A, B \in \mathscr{L}$ satisfying $A \subset B$, $B \setminus A \in \mathscr{L}$
- If $A_1 \subset A_2 \subset \dots$ is a sequence from $\mathscr{L}$ and $A = \bigcup_{i=1}^{\infty} A_i$ then $A \in \mathscr{L}$.

Dynkin's $\pi - \lambda$ Theorem

Lemma

Let $\mathscr{L}$ be a λ-system which is closed under intersection. Then $\mathscr{L}$ is a σ-algebra.

Proof.

- If $A \in \mathscr{L}$ then $A^c = \Omega \setminus A \in \mathscr{L}$.
- If $A, B \in \mathscr{L}$ then $A \cup B = (A^c \cap B^c)^c \in \mathscr{L}$.
- Thus, if $\{A_i\}_{i=1}^{\infty}$ is a sequence in $\mathscr{L}$, then for each n, $\bigcup_{i=1}^{n} A_i \in \mathscr{L}$, and hence $\bigcup_{i=1}^{\infty} A_i \in \mathscr{L}$.

□

Dynkin's $\pi - \lambda$ Theorem

Theorem (Dynkin's $\pi - \lambda$ Theorem)

If $\mathscr{P} \subset \mathscr{L}$ with $\mathscr{P}$ a π-system and $\mathscr{L}$ a λ-system then $\sigma(\mathscr{P}) \subset \mathscr{L}$.

Dynkin's $\pi - \lambda$ Theorem

Proof.

Let $\ell(\mathscr{P})$ be the smallest λ-system containing $\mathscr{P}$. We show that $\ell(\mathscr{P})$ is a σ-algebra.

- Let $A \in \ell(\mathscr{P})$ and define $L_A = \{B : A \cap B \in \ell(\mathscr{P})\}$.
- We check that L_A is a λ-system.
  - $\Omega \in L_A$ since $A \in \ell(\mathscr{P})$
  - If $B, C \in L_A$ and $B \supset C$, then $A \cap (B - C) = (A \cap B) - (A \cap C) \in \ell(\mathscr{P})$.
  - If $B_1 \subset B_2 \subset ...$ is a sequence from L_A with $B = \bigcup_{i=1}^\infty B_i$ then $B_1 \cap A \subset B_2 \cap A \subset ...$ has $B \cap A = \bigcup_{i=1}^\infty (B_i \cap A)$, and hence $B \cap A \in \ell(\mathscr{P})$ so $B \in L_A$.
- If $A \in \mathscr{P}$ then $L_A = \ell(\mathscr{P})$. Hence, if $B \in \ell(\mathscr{P})$ then $A \cap B \in \ell(\mathscr{P})$. But then this implies $L_B = \ell(\mathscr{P})$. It follows that for all $A, B \in \ell(\mathscr{P})$, $A \cap B \in \ell(\mathscr{P})$.

□

Measures

Definition

A *positive measure on an algebra* $\mathscr{A}$ is a set function μ which satisfies

- $\mu(A) \geq \mu(\emptyset) = 0$ for all $A \in \mathscr{A}$
- If $A_i \in \mathscr{A}$ are disjoint and their union is in $\mathscr{A}$, then

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i).$$

If $\mu(\Omega) = 1$ then μ is a probability measure.

Probability measure properties

A probability measure satisfies the following basic properties.

- (Monotonicity) If $A \subset B$ then $\mathrm{Prob}(A) \leq \mathrm{Prob}(B)$.
- (Sub-additivity) If $A \subset \bigcup_i A_i$ then $\mathrm{Prob}(A) \leq \sum_i \mathrm{Prob}(A_i)$
- (Continuity from below) If $A_1 \subset A_2 \subset ...$ and $A = \bigcup_i A_i$ then $\mathrm{Prob}(A_i) \uparrow \mathrm{Prob}(A)$
- (Continuity from above) If $A_1 \supset A_2 \supset ...$ and $A = \bigcap_i A_i$ then $\mathrm{Prob}(A_i) \downarrow \mathrm{Prob}(A)$.

Atomic measures

Definition

A probability space $(\Omega, \mathscr{F}, \text{Prob})$ is *non-atomic* if $\text{Prob}(A) > 0$ implies that there exists $B \in \mathscr{F}$ satisfying $B \subset A$ and $0 < \text{Prob}(B) < \text{Prob}(A)$.

Outer measures

Definition

An *outer measure* μ^* on a measurable space $(\Omega, \mathscr{F})$ is a set function $\mu^* : \mathscr{F} \to [0, \infty]$ satisfying

- $\mu^*(\emptyset) = 0$ and $\mu^*(A_1) \leq \mu^*(A_2)$ for any $A_1, A_2 \in \mathscr{F}$ with $A_1 \subset A_2$.
- $\mu^*\left(\bigcup_{n=1}^\infty A_n\right) \leq \sum_{n=1}^\infty \mu^*(A_n)$ for any countable collection of sets $\{A_n\} \subset \mathscr{F}$.

Definition

Given an outer measure μ^* on a measurable space $(\Omega, \mathscr{F})$, a set $A \in \mathscr{F}$ is *measurable* (in the sense of Carathéodory) if for each set $E \in \mathscr{F}$,

$$\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c).$$

Outer measures

Theorem

Let μ^* be an outer measure on a measurable space $(\Omega, \mathscr{F})$. The subset $\mathscr{G}$ of μ^*-measurable sets in $\mathscr{F}$ is a σ-algebra, and μ^* restricted to this subset is a measure.

See e.g. Royden pp.54–60.

Lebesgue measure

An outer measure on $(\mathbb{R}, 2^{\mathbb{R}})$ is given by

$$\mu^* (A) = \inf \left\{ \sum_{i=1}^{\infty} b_i - a_i : A \subset \bigcup_{i=1}^{\infty} (a_i, b_i] \right\}.$$

Lebesgue measure is obtained by restricting μ^* to its measurable sets. The σ-algebra so obtained is larger than the Borel σ-algebra.

Carathéodory's Extension Theorem

Theorem

Let μ be a σ-finite measure on an algebra $\mathscr{A}$. Then μ has a unique extension to $\sigma(\mathscr{A})$.

Proof of uniqueness.

Let μ_1 and μ_2 be two extensions of μ to $\sigma(\mathscr{A})$. Let $A \in \mathscr{A}$ satisfy $\mu(A) < \infty$ and let

$$\mathscr{L} = \{B \in \sigma(\mathscr{A}) : \mu_1(A \cap B) = \mu_2(A \cap B)\}.$$

We show that $\mathscr{L}$ is a λ-system. Since $\mathscr{A} \subset \mathscr{L}$ and $\mathscr{A}$ is a π-system, it then follows that $\mathscr{L} = \sigma(\mathscr{A})$. Uniqueness then follows on taking a sequence $\{A_n\}$ with $A_n \uparrow \Omega$ and $\mu(A_n) < \infty$. □

Carathéodory's Extension Theorem

Proof of uniqueness.

To verify the λ-system property, observe

- $\Omega \in \mathscr{L}$
- If $B, C \in \mathscr{L}$ with $C \subset B$, then

$$\begin{aligned}\mu_1(A \cap (B - C)) &= \mu_1(A \cap B) - \mu_1(A \cap C) \\ &= \mu_2(A \cap B) - \mu_2(A \cap C) = \mu_2(A \cap (B - C)).\end{aligned}$$

- If $B_n \in \mathscr{L}$ and $B_n \uparrow B$ then

$$\mu_1(A \cap B) = \lim_{n \to \infty} \mu_1(A \cap B_n) = \lim_{n \to \infty} \mu_2(A \cap B_n) = \mu_2(A \cap B).$$

$\square$

Proof of existence.

Define set function μ^* on $\sigma(\mathscr{A})$ by

$$\mu^*(E) = \inf\left\{\sum_{i=1}^{\infty} \mu(A_i) : E \subset \bigcup_{i=1}^{\infty} A_i, A_i \in \mathscr{A}\right\}.$$

Evidently $\mu^*(A) = \mu(A)$ for $A \in \mathscr{A}$. Also, $A \in \mathscr{A}$ is measurable, since for $F \in \sigma(\mathscr{A})$ and $\epsilon > 0$ there exists $\{B_i\}_{i=1}^{\infty}$ a sequence from $\mathscr{A}$ satisfying $\sum_i \mu(B_i) \leq \mu^*(F) + \epsilon$. Then

$$\mu(B_i) = \mu^*(B_i \cap A) + \mu^*(B_i \cap A^c)$$

$$\mu^*(F) + \epsilon \geq \sum_i \mu^*(B_i \cap A) + \sum_i \mu^*(B_i \cap A^c) \geq \mu^*(F \cap A) + \mu^*(F^c \cap A).$$

which gives the condition for measurability. □

Proof of existence.

μ^* satisfies the properties of an outer measure, since

- If $E \subset F$ then $\mu^*(E) \leq \mu^*(F)$
- If $F \subset \bigcup_i F_i$ is a countable union, then $\mu^*(F) \leq \sum_i \mu^*(F_i)$.

The restriction of μ^* to its measurable sets gives the required extension of μ. □

Random variables

Definition

A real valued *random variable* on a measure space $(\Omega, \mathscr{F}, \text{Prob})$ is a function $X : \Omega \to \mathbb{R}$ which is $\mathscr{F}$-measurable, that is, for each Borel set $B \subset \mathbb{R}$,

$$X^{-1}(B) = \{\omega : X(\omega) \in B\} \in \mathscr{F}.$$

A *random vector* in $\mathbb{R}^d$ is a measurable map $X : \Omega \to \mathbb{R}^d$.

Given $A \in \mathscr{F}$, the indicator function of A is a random variable,

$$1_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A \end{cases} .$$

Random variables

Theorem

If $X_1, ..., X_n$ *are random variables and* $f : (\mathbb{R}^n, \mathscr{B}_{\mathbb{R}^n}) \to (\mathbb{R}, \mathscr{B})$ *is measurable, then* $f(X_1, ..., X_n)$ *is a random variable.*

Theorem

If $X_1, X_2, ...$ *are random variables then* $X_1 + X_2 + ... + X_n$ *is a random variable, and so are*

$$\inf_n X_n, \quad \sup_n X_n, \quad \limsup_n X_n, \quad \liminf_n X_n.$$

Proof.

Exercise, or see Durrett, pp. 14–15. □

Distributions

Definition

The *distribution* of a random variable X on $\mathbb{R}$ is the probability measure μ on $(\mathbb{R}, \mathscr{B})$ defined by

$$\mu(A) = \text{Prob}(X \in A).$$

The *distribution function* of X is

$$F(x) = \text{Prob}(X \leq x).$$

Distributions

Theorem

Any distribution function F has the following properties:

1. *F is nondecreasing.*
2. $\lim_{x\to\infty} F(x) = 1, \lim_{x\to-\infty} F(x) = 0$.
3. *F is right continuous, that is,* $\lim_{y\downarrow x} F(y) = F(x)$.
4. *If* $F(x-) = \lim_{y\uparrow x} F(y)$ *then* $F(x-) = \mathrm{Prob}(X < x)$.
5. $\mathrm{Prob}(X = x) = F(x) - F(x-)$.

Furthermore, any function satisfying the first three items is the distribution function of a random variable.

Proof.

All of the forward claims are straightforward.
For the reverse claim, let $\Omega = (0,1)$, $\mathscr{F} = \mathscr{B}$ and set Prob to be Lebesgue measure. Define

$$X(\omega) = \sup\{y : F(y) < \omega\}.$$

Then

$$\{\omega : X(\omega) \leq x\} = \{\omega : \omega \leq F(x)\},$$

which follows by the right-continuity of F. Hence $\mathrm{Prob}(X \leq x) = F(x)$. □

Distributions

Definition

If X and Y induce the same distribution μ on $(\mathbb{R}, \mathscr{B})$, we say X and Y are *equal in distribution*. We write $X =_d Y$.

Definition

When the distribution function $F(x) = \text{Prob}(X \leq x)$ has the form

$$F(x) = \int_{-\infty}^{x} f(y)dy$$

we say that X has *density function* f.

Example distributions

- Uniform distribution on (0,1). Density $f(x) = 1$ for $x \in (0,1)$ and 0 otherwise.
- Exponential distribution with rate λ. Density $f(x) = \lambda e^{-\lambda x}$ for $x > 0$, 0 otherwise.
- Standard normal distribution. Density $f(x) = \frac{\exp\left(-\frac{x^2}{2}\right)}{\sqrt{2\pi}}$.

Example distributions

- Uniform distribution on the Cantor set. Define distribution function F by $F(x) = 0$ 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 $x \in [\frac{7}{9}, \frac{8}{9}]$,....
- Point mass at 0. The distribution function has $F(x) = 0$ for $x < 0$, $F(x) = 1$ for $x \geq 0$.
- Lognormal distribution. Let X be a standard Gaussian variable. $\exp(X)$ is lognormal.
- Chi-square distribution. Let X be a standard Gaussian variable. X^2 has a chi-squared distribution.

Example distributions on $\mathbb{Z}$

- Bernoulli distribution, parameter p. $\text{Prob}(X = 1) = p$, $\text{Prob}(X = 0) = 1 - p$.
- Poisson distribution, parameter λ. X is supported on $\mathbb{Z}$ and $\text{Prob}(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}$.
- Geometric distribution, success probability $p \in (0, 1)$. X is supported on $\mathbb{Z}$ and $\text{Prob}(X = k) = p(1 - p)^{k-1}$, for $k = 1, 2, \ldots$.

Integration

The Lebesgue integral against a σ-finite measure is defined as usual for

1. Simple functions
2. Bounded functions
3. Nonnegative functions
4. General functions

Integral inequalities

Theorem (Jensen's inequality)

Let ϕ be convex on $\mathbb{R}$. If μ is a probability measure, and f and $\phi(f)$ are integrable then

$$\phi\left(\int f d\mu\right) \leq \int \phi(f) d\mu.$$

Proof.

Let $c = \int f d\mu$ and let $\ell(x) = ax + b$ be a linear function which satisfies $\ell(c) = \phi(c)$ and $\phi(x) \geq \ell(x)$. Thus

$$\int \phi(f) d\mu \geq \int (af + b) d\mu = \phi\left(\int f d\mu\right).$$

$\square$

Hölder's inequality

Theorem

If $p, q \in (1, \infty)$ *with* $\frac{1}{p} + \frac{1}{q} = 1$, *then*

$$\int |fg| d\mu \leq \|f\|_p \|g\|_q.$$

Hölder's inequality

Proof.

We may assume that $\|f\|_p > 0$ and $\|g\|_q > 0$, since otherwise both sides vanish. Dividing both sides by $\|f\|_p \|g\|_q$, we may assume that $\|f\|_p = \|g\|_q = 1$.

For fixed $y \geq 0$,

$$\phi(x) = \frac{x^p}{p} + \frac{y^q}{q} - xy$$

has a minimum in $x \geq 0$ at $x_0 = y^{\frac{1}{p-1}}$ and $x_0^p = y^{\frac{p}{p-1}} = y^q$, so $\phi(x_0) = 0$. Thus $xy \leq \frac{x^p}{p} + \frac{y^q}{q}$ in $x, y \geq 0$. The claim follows by setting $x = |f|$, $y = |g|$ and integrating. □

Bounded convergence theorem

Definition

We say that $f_n \to f$ in measure if, for any $\epsilon > 0$,

$$\mu(\{x : |f_n(x) - f(x)| > \epsilon\}) \to 0$$

as $n \to \infty$.

Theorem (Bounded convergence theorem)

Let E be a set with $\mu(E) < \infty$. Suppose f_n vanishes on E^c, $|f_n(x)| \leq M$, and $f_n \to f$ in measure. Then

$$\int f d\mu = \lim_{n\to\infty} \int f_n d\mu.$$

Bounded convergence theorem

Proof.

Let $\epsilon > 0$, $G_n = \{x : |f_n(x) - f(x)| < \epsilon\}$ and $B_n = E - G_n$. Thus

$$\begin{aligned}\left|\int f d\mu - \int f_n d\mu\right| &\leq \int |f - f_n| d\mu \\ &= \int_{G_n} |f - f_n| d\mu + \int_{B_n} |f - f_n| d\mu \\ &\leq \epsilon\mu(E) + 2M\mu(B_n).\end{aligned}$$

Since $f_n \to f$ in measure, $\mu(B_n) \to 0$. The proof follows on letting $\epsilon \downarrow 0$. □

Fatou's lemma

Lemma (Fatou's lemma)

If $f_n \geq 0$ *then*

$$\liminf_{n\to\infty} \int f_n d\mu \geq \int \left(\liminf_{n\to\infty} f_n\right) d\mu.$$

Fatou's lemma

Proof.

Let $g_n(x) = \inf_{m \geq n} f_m(x)$, and note that

$$g_n(x) \uparrow g(x) = \liminf_{n \to \infty} f_n(x).$$

It suffices to verify that $\lim_{n\to\infty} \int g_n d\mu \geq \int g d\mu$. To do so, let $E_m \uparrow \Omega$ be sets of finite measure. For each fixed m, as $n \to \infty$,

$$\int g_n d\mu \geq \int_{E_m} g_n \wedge m d\mu \to \int_{E_m} g \wedge m d\mu.$$

Letting $m \to \infty$ proves the result. □

Monotone convergence theorem

Theorem (Monotone convergence theorem)

If $f_n \geq 0$ *and* $f_n \uparrow f$ *then*

$$\int f_n d\mu \uparrow \int f d\mu.$$

Monotone convergence theorem

Proof.

By Fatou's lemma, $\lim_{n\to\infty} \int f_n d\mu \geq \int f d\mu$. The reverse inequality is immediate. □

Dominated convergence theorem

Theorem (Dominated convergence theorem)

If $f_n \to f$ a.e., $|f_n| \leq g$ for all n and g is integrable, then $\int f_n d\mu \to \int f d\mu$.

Dominated convergence theorem

Proof.

Since $f_n + g \geq 0$, Fatou's lemma gives

$$\liminf_{n\to\infty} \int (f_n + g) d\mu \geq \int (f + g) d\mu.$$

Thus $\liminf_{n\to\infty} \int f_n d\mu \geq \int f d\mu$. To prove the limit, replace f_n with $-f_n$. □

Expected value

Definition

Let X be a random variable on $(\Omega, \mathscr{F}, \mathrm{Prob})$, and write $X = X^+ + X^-$ in a positive and negative part.

The *expected value* of X^+ is $\mathrm{E}[X^+] = \int X^+ dP$, similarly X^-. If either $\mathrm{E}[X^+]$ or $\mathrm{E}[X^-]$ is finite we say $\mathrm{E}[X]$ exists and its value is

$$\mathrm{E}[X] = \mathrm{E}[X^+] + \mathrm{E}[X^-].$$

$\mathrm{E}[X]$ is also called the mean, μ.

Theorem

Suppose $X_n \to X$ a.s. Let g and h be continuous functions on $\mathbb{R}$ satisfying

- $g \geq 0$ *and* $g(x) \to \infty$ *as* $|x| \to \infty$
- $|h(x)|/g(x) \to 0$ *as* $|x| \to \infty$
- *There exists* $K \geq 0$ *such that* $\mathrm{E}[g(X_n)] \leq K$ *for all* n.

Then $\mathrm{E}[h(X_n)] \to \mathrm{E}[h(X)]$ *as* $n \to \infty$.

A common application of this theorem takes $h(x) = x$ and $g(x) = |x|^p$ for some $p > 1$.

Expected value

Proof.

The proof method is an example of *truncation*.

- Assume w.l.o.g. that $h(0) = 0$.
- Let $M > 0$ be such that $\mathrm{Prob}(X = M) = 0$ and $g(x) > 0$ for $|x| > M$.
- Define $\overline{Y} = Y\mathbf{1}_{(|Y| \leq M)}$. By bounded convergence, $\mathrm{E}[h(\overline{X}_n)] \to \mathrm{E}[h(\overline{X})]$.

□

Expected value

Proof.

- Use

$$\left|\mathrm{E}[h(\overline{Y})] - \mathrm{E}[h(Y)]\right| \le \mathrm{E}[|h(\overline{Y}) - h(Y)|]$$
$$= \mathrm{E}[|h(Y)|\mathbf{1}_{(|Y|>M)}] \le \epsilon_M \,\mathrm{E}[g(Y)]$$

  where $\epsilon_M = \sup\{\frac{|h(x)|}{g(x)} : |x| > M\}$.

- Thus $|\,\mathrm{E}[h(\overline{X_n})] - E[h(X_n)]| \le K\epsilon_M$. Also,

$$\mathrm{E}[g(X)] \le \liminf_{n\to\infty} \mathrm{E}[g(X_n)] \le K$$

  so $|\,\mathrm{E}[h(\overline{X})] - \mathrm{E}[h(X)]| \le K\epsilon_M$.

□

Expected value

Proof.

- It follows from the triangle inequality that

$$|\operatorname{E}[h(X_n)] - \operatorname{E}[h(X)]| \le 2K\epsilon_M + |\operatorname{E}[h(\overline{X}_n)] - \operatorname{E}[h(\overline{X})]|.$$

  Letting first n, then m tend to infinity proves the claim.

□

Change of variable formula

Theorem

Let X be a random element of $(S, \mathscr{S})$ with distribution μ, that is, $\mu(A) = \mathrm{Prob}(X \in A)$. If f is measurable from $(S, \mathscr{S}) \to (\mathbb{R}, \mathscr{B})$ and is such that $f \geq 0$ or $E[|f(X)|] < \infty$, then

$$\mathrm{E}[f(X)] = \int_S f(y)\mu(dy).$$

Change of variable formula

Proof.

- If $B \in \mathscr{S}$ and $f = \mathbf{1}_B$ then

$$\mathrm{E}[\mathbf{1}_B(X)] = \mathrm{Prob}(X \in B) = \mu(B) = \int_S \mathbf{1}_B(y)\mu(dy).$$

- The equality thus holds for simple functions by linearity.
- The equality holds for non-negative functions f by taking a sequence of simple functions $f_n \uparrow f$ and applying monotone convergence.
- The equality holds for general f by linearity again.

□

Variance

Definition

Let X be a random variable which is square integrable. The variance of X is

$$\mathrm{Var}(X) = \mathrm{E}\left[X^2\right] - E\left[X\right]^2$$

and the standard deviation is $\sigma = \mathrm{Var}(X)^{\frac{1}{2}}$.

Markov's inequality

Theorem

Let $X \geq 0$ be a non-negative random variable with finite mean μ. Then for all $\lambda \geq 1$,

$$\mathrm{Prob}(X > \lambda\mu) \leq \frac{1}{\lambda}.$$

Proof.

The result holds if $\mu = 0$, so assume otherwise. Write

$$\lambda\mu\,\mathrm{Prob}(X > \lambda\mu) \leq \mathrm{E}\left[X 1_{(X > \lambda\mu)}\right] \leq \mathrm{E}[X] = \mu$$

to conclude. □

Chebyshev's inequality

Theorem

Let X be a square-integrable random variable with mean μ and standard deviation σ. Then for all $\lambda \geq 1$,

$$\text{Prob}(|X - \mu| > \lambda\sigma) \leq \frac{1}{\lambda^2}.$$

Proof.

Apply Markov's inequality to $(X - \mu)^2$. □