Keener, Theoretical Statistics
Chapter 1: Probability and Measure

Junhyeok Kil

June 23, 2025

Measures

A measure is the "size" of a set.

Definition

A collection $\mathcal{A}$ of subsets of a set $\mathcal{X}$ is a *σ-algebra* if:

1. $\mathcal{X} \in \mathcal{A}$
2. $A \in \mathcal{A} \Rightarrow A^{\complement} = \mathcal{X} - A \in \mathcal{A}$
3. $A_1, A_2, \cdots \in \mathcal{A} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$

The ordered pair $(\mathcal{X}, \mathcal{A})$ is called a *measurable space*.
$A \in \mathcal{A}$ are called *measurable sets*.

Exercise 4

Let $\mathcal{X} = \{1, 2, 3, 4\}$. Find the smallest σ-field $\mathcal{A}$ of subsets of $\mathcal{X}$ that contains $\{1\}$ and $\{1, 2, 3\}$.
Solution:

$$\{1\}, \{1, 2, 3\} \in \mathcal{A}$$

By properties:

$$\emptyset, \{1, 2, 3, 4\}, \{2, 3, 4\}, \{4\}, \{1, 4\}, \{2, 3\} \in \mathcal{A}$$

Since if $A, B \in \mathcal{A}$ then:

$$A^c, B^c, A \cup B, A^c \cup B^c, (A \cup B)^c = A^c \cap B^c, A \cap B \in \mathcal{A}$$

Thus:

$$\mathcal{A} = \{\emptyset, \{1\}, \{1, 4\}, \{2, 3\}, \{1, 2, 3\}, \{2, 3, 4\}, \{4\}, \{1, 2, 3, 4\}\} \quad \square$$

Measures (continued)

Definition

A function μ on σ-algebra $\mathcal{A}$ of $\mathcal{X}$ is a *measure* if:

1. $\mu : \mathcal{A} \to [0, \infty]$ (non-negativity)
2. If $A_1, A_2, \ldots$ are disjoint, then

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i) \quad \text{(countable additivity)}$$

If at least one set A has finite measure, by countable additivity:

$$\mu(A) = \mu(A \cup \emptyset) = \mu(A) + \mu(\emptyset) \Rightarrow \mu(\emptyset) = 0$$

The triple $(\mathcal{X}, \mathcal{A}, \mu)$ is called a *measure space*.
A *probability measure* is a measure with $\mu(\mathcal{X}) = 1$.

Examples of Measures

Example: Let $\mathcal{X}$ be countable. The *counting measure* is

$$\mu(A) = \#(A) = \#\text{ points in } A.$$

Example: Let $\mathcal{X} = \mathbb{R}^n$. The *Lebesgue measure* is

$$\mu(A) = \int_A \cdots \int_A dx_1 \cdots dx_n.$$

It is not possible to assign a measure to every subset of $\mathbb{R}^n$.

Sigma-Algebra and Domain of Measure

The σ-algebra is the *domain* of a measure μ and is a collection of subsets of $\mathcal{X}$:

$$\mathcal{F} \subseteq 2^{\mathcal{X}}.$$

If $\mathcal{X}$ is countable, we can take $\mathcal{F} = 2^{\mathcal{X}}$.
If $\mathcal{X} = \mathbb{R}^n$, we can take $\mathcal{F}$ to be the Borel set of $\mathbb{R}^n$, which is the smallest σ-algebra that contains all open subsets of $\mathbb{R}^n$.

Integration

Define the indicator function $\mathbf{1}_A$ of a set A as:

$$\mathbf{1}_A(x) = \mathbb{I}(x \in A) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$

Properties of Integrals

Basic properties for integrals:

1. $\int \mathbf{1}_A \, d\mu = \mu(A)$ for all $A \in \mathcal{A}$.
2. If f, g are non-negative, measurable functions and $a, b > 0$,
$$\int (af + bg) \, d\mu = a \int f \, d\mu + b \int g \, d\mu.$$
3. If $f_1 \leq f_2 \leq \cdots$ are non-negative measurable functions and $f(x) = \lim_{n\to\infty} f_n(x)$,
$$\int f \, d\mu = \lim_{n\to\infty} \int f_n \, d\mu.$$

Simple Functions

Using (1) and (2): If $\alpha_1, \ldots, \alpha_m$ are positive constants and $A_1, \ldots, A_m \in \mathcal{A}$,

$$\int \left(\sum_{i=1}^{m} \alpha_i \mathbf{1}_{A_i} \right) d\mu = \sum_{i=1}^{m} \alpha_i \mu(A_i).$$

We call this *simple functions*. Any non-negative measurable function can be approximated by simple functions.

Integration of General Functions

Theorem: If f is non-negative and measurable, then there exist non-negative simple functions $f_1 \leq f_2 \leq \cdots$ with

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

Brief Proof

Introduce the positive and negative parts

$$f^+(x) = \max\{f(x), 0\}, \quad f^-(x) = -\min\{f(x), 0\}$$

f^+ and f^- are both non-negative and measurable, and we have

$$f = f^+ - f^-$$

This difference is ambiguous only when the integrals are both infinite. So if either integrals are finite, we define

$$\int f \, d\mu = \int f^+ \, d\mu - \int f^- \, d\mu$$

To avoid $\infty - \infty$, require:

$$\int f^+ \, d\mu + \int f^- \, d\mu = \int |f| \, d\mu < \infty \quad \square$$

When $\int |f| d\mu < \infty$, f is called *integrable*.

Events, Probabilities, and Random Variables

Let P be a probability measure on a measurable space $(\mathcal{E}, \mathcal{B})$, so $(\mathcal{E}, \mathcal{B}, P)$ is a probability space.

Definition

A measurable function $\mathcal{X} : \mathcal{E} \to \mathbb{R}$ is called a *random variable (r.v.)*.

Definition

The probability measure $P_{\mathcal{X}}$ defined by

$$P_{\mathcal{X}}(A) = P\left(\{e \in \mathcal{E} : \mathcal{X}(e) \in A\}\right) \stackrel{\text{def}}{=} P(\mathcal{X} \in A)$$

for Borel sets A is called the *distribution* of $\mathcal{X}$.

Cumulative Distribution Function (CDF)

Definition

The cumulative distribution function (CDF) of $\mathcal{X}$ is defined by

$$F_{\mathcal{X}}(x) = P(\mathcal{X} \leq x) = P\left(\{e \in \mathcal{E} : \mathcal{X}(e) \leq x\}\right) = P_{\mathcal{X}}((-\infty, x])$$

for $x \in \mathbb{R}$.

Null Sets

Definition

Let μ be a measure on $(\mathcal{X}, \mathcal{A})$. A set $\mathcal{N}$ is called *null* (null w.r.t. μ) if

$$\mu(\mathcal{N}) = 0.$$

Almost Everywhere Statements

If a statement holds for $x \in \mathcal{X} \setminus N$ with N null, the statement is said to hold *a.e.* μ.

$$f = 0 \text{ a.e. } \mu \iff \mu\big(\{x \in \mathcal{X} : f(x) \neq 0\}\big) = 0$$

Suppose a statement holds $\iff x \in B$. Then the statement holds a.e. μ

$$\iff \mu(B^c) = 0 \iff \mu(B) = 1.$$

Densities

Definition

Let P and μ be measures on a σ-algebra $\mathcal{A}$ of $\mathcal{X}$. Then P is called *absolutely continuous* w.r.t. μ, written $P \ll \mu$, if

$$P(A) = 0 \text{ whenever } \mu(A) = 0.$$

Theorem (Radon-Nikodym)

If a finite measure P is absolutely continuous w.r.t. a σ-finite measure μ, then there exists a non-negative measurable function f such that

$$P(A) = \int_A f \, d\mu \stackrel{\text{def}}{=} \int_A f \mathbf{1}_A \, d\mu.$$

f is called the *Radon-Nikodym derivative* of P w.r.t. μ, or the *density* of P w.r.t. μ, denoted $f = \frac{dP}{d\mu}$.

PDF and PMF

If $\mathcal{X} \sim P_{\mathcal{X}}$ and $P_{\mathcal{X}}$ is absolutely continuous w.r.t. μ with density $p = \frac{dP_{\mathcal{X}}}{d\mu}$, then we say $\mathcal{X}$ has *density* p w.r.t. μ.
If P is a probability distribution and μ is Lebesgue measure, we call $p(x)$ the *probability density function (PDF)*.
If P is a probability distribution and μ is counting measure, we call $p(x)$ the *probability mass function (PMF)*.

Expectation

Definition

If $\mathcal{X}$ is a random variable on a probability space $(\mathcal{E}, \mathcal{B}, P)$, then the expectation of $\mathcal{X}$ is

$$\mathbb{E}(\mathcal{X}) = \int \mathcal{X}\, dP.$$

This formula is rarely used. Instead, if $\mathcal{X} \sim P_{\mathcal{X}}$, we write

$$\mathbb{E}(\mathcal{X}) = \int x\, dP_{\mathcal{X}}(x).$$

If $Y = f(\mathcal{X})$,

$$\mathbb{E}(Y) = \mathbb{E}(f(\mathcal{X})) = \int f\, dP_{\mathcal{X}}.$$

If $P_{\mathcal{X}}$ has density p w.r.t. μ, then

$$\int f\, dP_{\mathcal{X}} = \int fp\, d\mu.$$

These results can be viewed as change-of-variable results, and proofs are based on properties of integration.

Expectation: Continuous and Discrete Case

If $\mathcal{X}$ is an absolutely continuous r.v. with density p,

$$\mathbb{E}(\mathcal{X}) = \int x \, dP_{\mathcal{X}}(x) = \int xp(x) \, dx$$

and

$$\mathbb{E}(f(\mathcal{X})) = \int f(x)p(x) \, dx. \quad (1)$$

If $\mathcal{X}$ is discrete with $P(\mathcal{X} \in \mathcal{X}_0) = 1$ for a countable set $\mathcal{X}_0$, μ is counting measure on $\mathcal{X}_0$, and p is the mass function $p(x) = P(\mathcal{X} = x)$, then

$$\mathbb{E}(\mathcal{X}) = \int x \, dP_{\mathcal{X}}(x) = \int xp(x) \, d\mu(x) = \sum_{x \in \mathcal{X}_0} xp(x)$$

and

$$\mathbb{E}(f(\mathcal{X})) = \sum_{x \in \mathcal{X}_0} f(x)p(x). \quad (2)$$

Variance

Definition

The variance of r.v. $\mathcal{X}$ with finite expectation is defined as

$$\mathrm{Var}(\mathcal{X}) = \mathbb{E}\big(\mathcal{X} - \mathbb{E}(\mathcal{X})\big)^2$$

$$= \mathbb{E}(\mathcal{X}^2 - 2\mathcal{X}\mathbb{E}(\mathcal{X}) + (\mathbb{E}(\mathcal{X}))^2) = \mathbb{E}(\mathcal{X}^2) - (\mathbb{E}(\mathcal{X}))^2.$$

If $\mathcal{X}$ is absolutely continuous with density p, by (1)

$$\mathrm{Var}(\mathcal{X}) = \int (x - \mathbb{E}(\mathcal{X}))^2 p(x)\, dx.$$

If $\mathcal{X}$ is discrete with mass function p, by (2)

$$\mathrm{Var}(\mathcal{X}) = \sum_{x \in \mathcal{X}_0} (x - \mathbb{E}(\mathcal{X}))^2 p(x).$$

Covariance and Correlation

Definition

The covariance between r.v. $\mathcal{X}$ and $\mathcal{Y}$ with finite expectation is

$$\mathrm{Cov}(\mathcal{X}, \mathcal{Y}) = \mathbb{E}\big((\mathcal{X} - \mathbb{E}(\mathcal{X}))(\mathcal{Y} - \mathbb{E}(\mathcal{Y}))\big)$$

whenever the expectation exists.

If $\mathcal{Y} = \mathcal{X}$:

$$\mathrm{Cov}(\mathcal{X}, \mathcal{X}) = \mathrm{Var}(\mathcal{X})$$

By linearity,

$$\mathrm{Cov}(\mathcal{X}, \mathcal{Y}) = \mathbb{E}(\mathcal{X}\mathcal{Y}) - \mathbb{E}(\mathcal{X})\mathbb{E}(\mathcal{Y})$$

Correlation

Because covariances are influenced by the measurement scale, a more natural measure is the correlation:

$$\mathrm{Cor}(\mathcal{X}, \mathcal{Y}) = \frac{\mathrm{Cov}(\mathcal{X}, \mathcal{Y})}{\sqrt{\mathrm{Var}(\mathcal{X})\,\mathrm{Var}(\mathcal{Y})}}$$

Product Measures and Independence

Let $(\mathcal{X}, \mathcal{A}, \mu)$ and $(\mathcal{Y}, \mathcal{B}, \nu)$ be measure spaces. Then there exists a unique measure $\mu \times \nu$, called the *product measure*, on $(\mathcal{X} \times \mathcal{Y}, \mathcal{A} \otimes \mathcal{B})$ such that

$$(\mu \times \nu)(A \times B) = \mu(A)\nu(B)$$

for all $A \in \mathcal{A}$ and $B \in \mathcal{B}$.
The σ-algebra $\mathcal{A} \otimes \mathcal{B}$ is the smallest σ-algebra containing all sets $A \times B$ with $A \in \mathcal{A}, B \in \mathcal{B}$.

Examples and Fubini's Theorem

- If μ, ν are Lebesgue measures on $\mathbb{R}^n$ and $\mathbb{R}^m$, then $\mu \times \nu$ is Lebesgue measure on $\mathbb{R}^{n+m}$.
- If μ, ν are counting measures on countable sets $\mathcal{X}_0, \mathcal{Y}_0$, then $\mu \times \nu$ is counting measure on $\mathcal{X}_0 \times \mathcal{Y}_0$.

Theorem (Fubini)

If $f \geq 0$, then

$$\int f \, d(\mu \times \nu) = \int \left[\int f(x,y) \, d\nu(y) \right] d\mu(x)$$

$$= \int \left[\int f(x,y) \, d\mu(x) \right] d\nu(y)$$

Independence of Random Vectors

Definition (Independence)

Two random vectors $\mathcal{X} \in \mathbb{R}^n$, $\mathcal{Y} \in \mathbb{R}^m$ are independent if

$$P(\mathcal{X} \in A, \mathcal{Y} \in B) = P(\mathcal{X} \in A)P(\mathcal{Y} \in B)$$

for all Borel sets A, B.

If $\mathcal{Z} = \begin{pmatrix} \mathcal{X} \\ \mathcal{Y} \end{pmatrix}$, then

$$\mathcal{Z} \in A \times B \iff \mathcal{X} \in A \text{ and } \mathcal{Y} \in B$$

Joint Distributions and Densities

The above condition for independence can be expressed in terms of distributions as:

$$P_{\mathcal{Z}}(A \times B) = P_{\mathcal{X}}(A) P_{\mathcal{Y}}(B)$$

This shows that the distribution of $\mathcal{Z}$ is the product measure:

$$P_{\mathcal{Z}} = P_{\mathcal{X}} \times P_{\mathcal{Y}}$$

The density of $\mathcal{Z}$ can be given by the product of the densities of $\mathcal{X}$ and $\mathcal{Y}$:

$$p_{\mathcal{Z}}(x, y) = p_{\mathcal{X}}(x) p_{\mathcal{Y}}(y)$$

Here $P_{\mathcal{Z}}$ is called the *joint distribution* of $\mathcal{X}$ and $\mathcal{Y}$.

Functions of Independent Random Vectors

If $\mathcal{X}_1, \ldots, \mathcal{X}_n$ are independent random vectors and $f_1, \ldots, f_n$ are measurable functions, then

$$f_1(\mathcal{X}_1), \ldots, f_n(\mathcal{X}_n) \text{ are independent.}$$

Joint Density of Independent Variables

If $\mathcal{X}_i$ has density $p_{\mathcal{X}_i}$ w.r.t. μ_i for $i = 1, \ldots, n$, then $\mathcal{X}_\infty, \ldots \mathcal{X}_n$ have joint density p given by

$$p(x_1, \ldots, x_n) = p_{\mathcal{X}_1}(x_1) \times \cdots \times p_{\mathcal{X}_n}(x_n)$$

w.r.t. $\mu = \mu_1 \times \cdots \times \mu_n$.

Independent and Identically Distributed (i.i.d.)

If $\mathcal{X}_1, \ldots, \mathcal{X}_n$ are independent and all have the same distribution $\mathcal{X}_i \sim Q$, then

$$\mathcal{X}_1, \ldots, \mathcal{X}_n \text{ are i.i.d.}$$

and the collection $\mathcal{X}_1, \ldots, \mathcal{X}_n$ is called a *random sample* from Q.

Conditional Distributions

Definition

If X and Y are r.v.s, the conditional distribution of Y given $X = x$ is

$$Q_x(B) = P(Y \in B \mid X = x) = \frac{P(Y \in B, X = x)}{P(X = x)}$$

for Borel sets B (when X is discrete).

Q_x is a probability measure on Y for each x, forming a *stochastic transition kernel*. If X and Y are independent, $Q_x = P_Y$ for all x.

Conditional Expectation

Integration w.r.t. conditional distribution gives the conditional expectation:

$$\mathbb{E}[f(X, Y) \mid X = x] = \int f(x, y)\, dQ_x(y)$$

If X, Y are discrete:

$$\mathbb{E}[f(X, Y) \mid X = x] = \sum_y f(x, y) q_x(y)$$

where

$$q_x(y) = P(Y = y \mid X = x)$$

Law of Total Expectation (Smoothing)

$$\mathbb{E}f(X, Y) = \mathbb{E}\big(\mathbb{E}[f(X, Y) \mid X]\big)$$

Examples:

$$\mathbb{E}Y = \mathbb{E}\big(\mathbb{E}[Y \mid X]\big)$$

$$P(B) = \mathbb{E}\big(P(B \mid X)\big)$$

These identities are called the *law of total probability*, *tower property*, or *smoothing*.

Conditional of Conditional

Conditional expectations and probabilities satisfy:

$$\mathbb{E}(Y \mid X) = \mathbb{E}\big(\mathbb{E}[Y \mid X, W] \mid X\big)$$

$$P(B \mid X) = \mathbb{E}\big(P(B \mid X, W) \mid X\big)$$