

Theoretical Statistics: Topics for a Core Course

Summary Note

Robert W. Keener

HYUNSUNG KIM
Department of Statistics
Chung-Ang University

Update: September 6, 2021

Contents

Preface	2
1 Probability and Measure	4

Preface

This note contains the solution of the exercises in the textbook, *Theoretical Statistics: Topics for a Core Course*, and it was created by Hyunsung Kim, who is a Ph.D. student. I wrote it when I study a theoretical statistics based on this textbook on my own by solving some exercises and also referred to the solution manual in the textbook.

It contains a few selected exercises in the textbook what I studied, and also note that it may not be the exact solutions. If you want to refer to this note, you should study with doubt about the answer.

Textbook

- Keener, *Theoretical Statistics: Topics for a Core Course*.

Reference

- Durrett, *Probability: Theory and Examples*, 5th edition.
- Royden, *Real Analysis*, 4th edition.

Chapter 1

Probability and Measure

1.1 Measures

Definition 1.1.1 (σ -algebra). A collection of \mathcal{A} of subsets of a set \mathcal{X} is a σ -algebra (or σ -field) if

1. $\mathcal{X} \in \mathcal{A}$ and $\emptyset \in \mathcal{A}$,
2. If $A \in \mathcal{A}$, then $A^c = \mathcal{X} - A \in \mathcal{A}$,
3. If $A_1, A_2, \dots \in \mathcal{A}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$.

Definition 1.1.2 (Measure). A function μ on a σ -algebra \mathcal{A} of \mathcal{X} is a measure if

1. For every $A \in \mathcal{A}$, $0 \leq \mu(A) \leq \infty$; that is, $\mu : \mathcal{A} \rightarrow [0, \infty]$.
2. (Countable Additivity) If A_1, A_2, \dots are disjoint elements of \mathcal{A} , then

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

Example 1.1.3. For a measure μ on a set \mathcal{X} and subsets $A \in \mathcal{X}$,

1. μ is *counting measure* on \mathcal{X} if \mathcal{X} is countable, and define

$$\mu(A) = \#A = \text{number of points in } A,$$

and also its σ -algebra \mathcal{A} be the *power set* of \mathcal{X} , denoted $\mathcal{A} = 2^{\mathcal{X}}$.

2. μ is *Lebesgue measure* on \mathcal{X} if $\mathcal{X} = \mathbb{R}^n$, and define

$$\mu(A) = \int \cdots \int_A dx_1 \cdots dx_n,$$

and for any set A in a σ -algebra \mathcal{A} called the *Borel sets* of $\mathcal{X} = \mathbb{R}^n$, and formally \mathcal{A} is the smallest σ -algebra that contains all open set in \mathbb{R}^n ("rectangles").

3. μ is *probability measure* on \mathcal{X} if $\mu(\mathcal{X}) = 1$.

Remark 1.1.4. If \mathcal{A} is a σ -algebra of subsets of \mathcal{X} , the pair $(\mathcal{X}, \mathcal{A})$ is called a *measurable space*, and if μ is a measure on \mathcal{A} , the triple $(\mathcal{X}, \mathcal{A}, \mu)$ is called a *measure space*.

Proposition 1.1.5 (Continuity property of measures). *If measurable sets B_n , $n \geq 1$, are increasing ($B_1 \subset B_2 \subset \dots$), with $B = \bigcup_{i=1}^{\infty} B_n$, called the limit of the sequence, then*

$$\mu(B) = \lim_{n \rightarrow \infty} \mu(B_n).$$

If a measure μ is a probability measure, it also holds for decreasing sets ($B_1 \supset B_2 \supset \dots$) and its limit $B = \bigcap_{n=1}^{\infty} B_n$.

Proof. 생략 □

Definition 1.1.6 (σ -finite). *There is some notations and definitions about the measure.*

1. A measure μ is *finite* if $\mu(\mathcal{X}) < \infty$.

2. A measure μ is *σ -finite* if $\exists A_1, A_2, \dots \in \mathcal{A}$ with $\mu(A_i) < \infty \forall i = 1, 2, \dots$ and $\bigcup_{i=1}^{\infty} A_i = \mathcal{X}$.

Remark 1.1.7. A probability measure is σ -finite.

1.2 Integration

Example 1.2.1. The integral of "nice" function f against a measure μ is written as $\int f d\mu$ or $\int f(x) d\mu(x)$.

1. If μ is *counting measure* on \mathcal{X} , then the integral of f against μ is

$$\int f d\mu = \sum_{x \in \mathcal{X}} f(x).$$

2. If μ is *Lebesgue measure* on \mathbb{R}^n , then the integral of f against μ is

$$\int f d\mu = \int \cdots \int f(x_1, \dots, x_n) dx_1 \dots dx_n.$$

Definition 1.2.2 (Measurable). *If $(\mathcal{X}, \mathcal{A})$ is a measurable space and f is a real-valued function on \mathcal{X} , then f is *measurable* if*

$$f^{-1}(B) \stackrel{\text{def}}{=} \{x \in \mathcal{X} : f(x) \in B\} \in \mathcal{A},$$

for every Borel set B .

Definition 1.2.3 (Indicator function). *The indicator function 1_A of a set A is defined as*

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

Proposition 1.2.4. *The integral of the indicator function 1_A has following properties:*

1. For any set $A \in \mathcal{A}$, $\int 1_A d\mu = \mu(A)$.
2. If f and g are non-negative measurable functions, and if a and b are positive constants,

$$\int (af + bg) d\mu = a \int f d\mu + b \int g d\mu.$$

3. If $f_1 \leq f_2 \leq \dots$ are non-negative measurable functions, and if $f(x) = \lim_{n \rightarrow \infty} f_n(x)$, then

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

Proof. 생략 □

Definition 1.2.5 (Simple function). *A function f is simple if*

$$f = \sum_{i=1}^m a_i 1_{A_i},$$

where a_1, \dots, a_m are positive constants, and A_1, \dots, A_m are sets in \mathcal{A}

Remark 1.2.6. By using 1 and 2 in Proposition 1.2.4, the integral of simple function is obtained as

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

Theorem 1.2.7. *If f is non-negative and measurable, then there exist non-negative simple functions $f_1 \leq f_2 \leq \dots$ with $f = \lim_{n \rightarrow \infty} f_n$.*

Proof. 생략 □

Definition 1.2.8 (Integrable). *A measurable function f is integrable if $\int |f| d\mu < \infty$.*

1.3 Events, Probabilities, and Random Variables

Definition 1.3.1 (Probability space). *If \mathcal{E} is a sample space, \mathcal{B} is a σ -algebra of subsets of \mathcal{X} , and P is a probability measure, then the triple $(\mathcal{X}, \mathcal{A}, P)$ is called a probability space. Also $B \in \mathcal{B}$ are called events, points $e \in \mathcal{E}$ are called outcomes, and $P(B)$ is called the probability of B .*

Definition 1.3.2 (Random variable). *A measurable function $X : \mathcal{E} \rightarrow \mathbb{R}$ is called a random variable.*

Definition 1.3.3 (Distribution). *The probability measure P_X is defined by*

$$P_X(A) = P(\{e \in \mathcal{E} : X(e) \in A\}) \stackrel{\text{def}}{=} P(X \in A),$$

for Borel sets A is called distribution of X , and if we let $P_X := Q$, then it is denoted as $X \sim Q$, means that a random variable X has distribution Q .

Definition 1.3.4 (Cumulative distribution function). *The cumulative distribution function of X is defined by*

$$F_X(x) = P(X \leq x) = P(\{e \in \mathcal{E} : X(e) \leq x\}) = P_X((-\infty, x]),$$

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

1.4 Null Sets

Definition 1.4.1 (Null set). *For a measure μ on the measurable space $(\mathcal{X}, \mathcal{A})$, a set N is called null with respect to μ if*

$$\mu(N) = 0.$$

Definition 1.4.2 (Almost everywhere). *If a statement holds for $x \in \mathcal{X} - N$ with N null, the statement is said to hold almost everywhere (a.e.) or a.e. μ .*

Example 1.4.3. $f = 0$ a.e. $\mu \iff \mu(\{x \in \mathcal{X} : f(x) \neq 0\}) = 0$.

Remark 1.4.4 (Almost sure). For the probability space, the statement holds a.e. if and only if $\mu(B^c) = 0$ or $\mu(B) = 1$. Especially, it can be denoted that the statement holds *almost surely* (a.s.) or *with probability one*.

Proposition 1.4.5. *There are a few useful facts about integration:*

1. If $f = 0$ (a.e. μ), then $\int f d\mu = 0$.
2. If $f \geq 0$ and $\int f d\mu = 0$, then $f = 0$ (a.e. μ).
3. If $f = g$ (a.e. μ), then $\int f d\mu = \int g d\mu$ for at least of the integral exists.
4. If $\int 1_{(c,x)} f d\mu = 0$ for all $x > c$, then $f(x) = 0$ for a.e. $x > c$. The constant c here can be $-\infty$.
5. If f and g are integrable and $f > g$, then $\int f d\mu > \int g d\mu$. (From 2)

Proof. 생략

□

1.5 Densities

Definition 1.5.1 (Absolutely continuity). *Let P and μ be measures on a σ -field \mathcal{A} of \mathcal{X} . Then P is called absolutely continuous with respect to μ , written $P \ll \mu$, if $P(A) = 0$ whenever $\mu(A) = 0$.*

Theorem 1.5.2 (Radon-Nikodym). *If a finite measure P is absolutely continuous with respect to a σ -finite measure μ , then there exists a non-negative measurable function f such that*

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

Remark 1.5.3. The function f in Theorem 1.5.2 is called Radon-Nikodym derivative of P with respect to μ , or the *density* of P with respect to μ , denoted

$$f = \frac{dP}{d\mu}.$$

Remark 1.5.4. If $X \sim P_X$ and P_X is absolutely continuous with respect to μ with density $p = dP_X/d\mu$, it is called that X has density p with respect to μ .

Example 1.5.5 (Absolutely continuous random variables). If a random variable X has density p with respect to Lebesgue measure on \mathbb{R} , then X or its distribution P_X is called *absolutely continuous* with density p . Then, from the Radon-Nikodym theorem,

$$F_X(x) = P(X \leq x) = P_X((-\infty, x]) = \int_{-\infty}^x p(u)du$$

And also by fundamental theorem of calculus, $p(x) = F'_X(x)$.

Example 1.5.6 (Discrete random variables). Let \mathcal{X}_0 be a countable subset of \mathbb{R} . The measure μ defined by

$$\mu(B) = \#(\mathcal{X}_0 \cap B)$$

for Borel sets B is (also) called *counting measure on \mathcal{X}_0* . Then, the integration is defined as

$$\int f d\mu = \sum_{x \in \mathcal{X}_0} f(x).$$

If a random variable X satisfies

$$P(X \in \mathcal{X}_0) = P_X(\mathcal{X}_0) = 1,$$

then X is called a *discrete random variable*. And also P_X is absolutely continuous with respect to μ . (See the textbook page 8.)

The density p of P_X with respect to μ satisfies

$$P(X \in A) = P_X(A) = \int_A p d\mu = \sum_{x \in \mathcal{X}_0} p(x)1_A(x).$$

If $A = \{y\}$ with $y \in \mathcal{X}_0$, then

$$P(X = y) = \sum_{x \in \mathcal{X}_0} p(x)1_{\{y\}}(x) = p(y).$$

This density p is called the *mass function* for X .

1.6 Expectation

Definition 1.6.1 (Expectation). *If X is a random variable on a probability space $(\mathcal{E}, \mathcal{B}, P)$, then the expectation or expected value of X is defined as*

$$EX = \int X dP.$$

If $X \sim P_X$, then

$$EX = \int x dP_X(x).$$

Also, if $Y = f(X)$, then

$$EY = Ef(X) = \int f dP_X.$$

Remark 1.6.2. If P_X has density p with respect to μ , then

$$\int f dP_X = \int f p d\mu.$$

Example 1.6.3. If X is an absolutely continuous random variable with density p , then

$$EX = \int x dP_X(x) = \int x p(x) dx$$

and

$$Ef(X) = \int f(x) p(x) dx.$$

Example 1.6.4. If X is a discrete with $P(X \in \mathcal{X}_0) = 1$ for a countable set \mathcal{X}_0 , if μ is counting measure on \mathcal{X}_0 , and if p is the mass function given by $p(x) = P(X = x)$, then

$$EX = \int x dP_X(x) = \int x p(x) d\mu(x) = \sum_{x \in \mathcal{X}_0} x p(x)$$

and

$$Ef(X) = \sum_{x \in \mathcal{X}_0} f(x) p(x).$$

Proposition 1.6.5. *If X and Y are random variables with finite expectations, the followings are satisfied.*

1. *For a and b are nonzero constants,*

$$E(aX + bY) = aEX + bEY.$$

2. *If $X \leq y$ (a.e. P), then $EX \leq EY$, and equality holds if $X = Y$ (a.e. P).*

Proof. 생략 □

Definition 1.6.6 (Variance). *The variance of a random variable X with finite expectation is defined as*

$$\text{Var}(X) = E(X - EX)^2 = EX^2 - (EX)^2.$$

1. If X is absolutely continuous with density p ,

$$\text{Var}(X) = \int (x - EX)^2 p(x) dx.$$

2. If X is discrete with mass function p ,

$$\text{Var}(X) = \sum_{x \in \mathcal{X}_0} (x - EX)^2 p(x).$$

Definition 1.6.7 (Covariance). The covariance between two random variables X and Y with finite expectation is defined as

$$\text{Cov}(X, Y) = E(X - EX)(Y - EY) = EXY - (EX)(EY).$$

Definition 1.6.8 (Correlation). The correlation between two random variables X and Y with finite expectation is defined as

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

and it always lie in $[-1, 1]$.

1.7 Random Vectors

Definition 1.7.1 (Random vector). If X_1, \dots, X_n are random variables, then the function $X : \mathcal{E} \rightarrow \mathbb{R}^n$ defined by

$$X(e) = \begin{pmatrix} X_1(e) \\ \vdots \\ X_n(e) \end{pmatrix}, \quad e \in \mathcal{E},$$

is called a random vector.

Remark 1.7.2. Similarly with the random variables, the notations of random vectors is directly extended by follows:

1. The distribution of P_X of X is defined by

$$P_X(B) = P(X \in B) \stackrel{\text{def}}{=} P(\{e \in \mathcal{E} : X(e) \in B\})$$

for Borel sets $B \in \mathbb{R}^n$, and denote $X \sim P_X$.

2. The random vector X or its distribution P_X is called *absolutely continuous* with density p if P_X is absolutely continuous with respect to Lebesgue measure on \mathbb{R}^n . And the probability is defined as

$$P(X \in B) = \int \cdots \int_B p(x) dx.$$

3. The random vector X is *discrete* if $P(X \in \mathcal{X}_0) = 1$ for some countable set $\mathcal{X}_0 \subset \mathbb{R}^n$. If $p(x) = P(X = x)$, then P_X has density p with respect to counting measure on \mathcal{X}_0 and

$$P(X \in B) = \sum_{x \in \mathcal{X} \cap B} p(x).$$

4. The expectation of random vector X is defined as

$$EX = \begin{pmatrix} EX_1 \\ \vdots \\ EX_n \end{pmatrix}.$$

If $T : \mathbb{R}^n \rightarrow \mathbb{R}$ is a measurable function, then $T(X)$ is a random variable, and with finite integral,

$$ET(X) = \int T dP_X.$$

If P_X has a density p with respect to a dominating measure μ , this integral can be expressed as $\int T p d\mu$, which becomes

$$\sum_{x \in \mathcal{X}_0} T(x)p(x) \text{ or } \int \cdots \int T(x)p(x)dx,$$

where discrete and absolutely continuous cases with μ counting or Lebesgue measure, respectively.

1.8 Covariance Matrices

Definition 1.8.1 (Random matrix). *A matrix W is called a random matrix if the entries W_{ij} are random variables. If W is a random matrix, then EW is the matrix of expectations of the entries,*

$$(EW)_{ij} = EW_{ij}.$$

Remark 1.8.2. If v is a constant vector, A , B , and C are constant matrices, X is a random vector, and W is a random matrix, then

$$E[v + AX] = v + AE(X),$$

$$E[A + BWC] = A + B(EW)C.$$

Definition 1.8.3 (Covariance matrix).