Study Notes of Matrix and Tensor

Pei Zhong

March 23, 2024

Contents

Preface

The notes mainly refer to:

- Introduction to Mathematical Statistics 8th Edition
- lecture note
- Study Guide

Chapter 1

Probability and Distributions

1.1 Introduction

Definition 1.1

If an experiment can be repeated under the same conditions it is a random experiment. The set of every possible outcome of an experiment is the sample space, denoted $\mathcal{C}$.

Remark. For an experiment, the sample space is not unique. For example, When talking about the temperature in an area, we can define the sample space as $\mathcal{C} = (-\infty, \infty)$ or $\mathcal{C} = [a, b]$. For a specific random experiment, we can use different sample spaces to describe it. However, it is worth studying how to describe it with an appropriate sample space.

Note/Definition. Notationally, we denote the elements of the sample space with lower case letters such as a, b, c. Subsets of the sample space are *events* and we denote them with upper case letters such as A, B, C.

Definition 1.2

If an experiment is performed N times and a specific event occurs f times, then f is the frequency of the event and f/N is the relative frequency of the event.

1.2 Sets

1.3 The Probability Set Function

We need to define a set function that assigns a probability to the events (subsets of sample space $\mathcal{C}$). We denote the colletion of events as $\mathcal{B}$. If $\mathcal{C}$ is finite set, then we hope to assign a probability to all events (that is, to define a probability set function on the power set of $\mathcal{C}$). More generally, we require that $\mathcal{B}$ (the colletion of events) to satisfy: (1) the sample space $\mathcal{C}$ itself is an event, (2) the complement of every event is again an event, and (3) every countable union of events is again an event. Symbolically, this means (1) $\mathcal{C} \in \mathcal{B}$, (2) if $A \in \mathcal{B}$ then $A^c \in \mathcal{B}$, and (3) if $A_1, A_2, \ldots \in \mathcal{B}$ then $\cup_{n=1}^{\infty} A_n \in \mathcal{B}$. Combining (2) and (3), we see by DeMorgan's Law (for countable unions) that if $A_1, A_2, ... \in \mathcal{B}$ then $\cap_{n=1}^{\infty} A_n \in \mathcal{B}$. So the collection of events $\mathcal{B}$ is closed under complements, countable unions, and countable intersections. Such a collection of sets form a *σ-algebra*.

Definition 1.3

A collection of events $\{A_n | n \in I\}$ (where I is some indexing set) such that $A_i \cap A_j = \emptyset$ is a mutually exclusive collection of events.

Definition 1.4

Let $\mathcal{C}$ be a sample space and let $\mathcal{B}$ be the set of all events (thus, $\mathcal{B}$ is a σ-field). Let P be a real-valued function defined on $\mathcal{B}$. Then P is a probability set function if P satisfies the following three conditions:
(1) $P(A) \geqslant 0$ for $A \in \mathcal{B}$.
(2) $P(\mathcal{C}) = 1$.
(3) If $\{A_n\}$ is a mutually exclusive collection of events, then $P(\cup_{n=1}^{+\infty} A_n) = \sum\limits_{n=1}^{+\infty} P(C_n)$.

Theorem 1.1

For each event $A \in \mathcal{B}$, $P(A) = 1 - P(A^c)$.

Theorem 1.2

The probability of the null set is zero; that is, $P(\emptyset) = 0$.

Theorem 1.3

If A and B are events such that $A \subset B$, then $P(A) \leqslant P(B)$.

Theorem 1.4

For each event $A \in \mathcal{B}$ we have $0 \leqslant P(A) \leqslant 1$.

Theorem 1.5

If A and B are events in $\mathcal{C}$, then $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Theorem 1.6

Let $\{A_n\}$ be a nondecreasing sequence of events (ie. $A_n \subseteq A_{n+1}$). Then

$$\lim_{n\to\infty} P(A_n) = P(\lim_{n\to\infty} A_n) = P(\cup_{n=1}^{\infty} A_n).$$

Let $\{A_n\}$ be a nonincreasing sequence of events (ie. $A_n \supseteq A_{n+1}$). Then

$$\lim_{n\to\infty} P(A_n) = P(\lim_{n\to\infty} A_n) = P(\cap_{n=1}^{\infty} A_n).$$

Theorem 1.7

Let $\{A_n\}$ be an arbitrary sequence of events. Then

$$P(\cup_{n=1}^{\infty} A_n) \leqslant \sum_{n=1}^{\infty} P(A_n).$$

1.4 Conditional Probability and Independence

The idea behind conditional probability is that the initial sample space $\mathcal{C}$ has been replaced with some subset $A \subset \mathcal{C}$.

Definition 1.5

Let B and A be events with $P(A) > 0$. Then the conditional probability of B given A as $P(B|A) = \frac{P(A\cap B)}{P(A)}$.

Note/Definition. If A and B are events where $P(A) > 0$ then $P(A \cap B) = P(A)P(B|A)$ by Definition 1.5. This is called the multiplication rule also.

Definition 1.6

Let A and B be two events. Then A and B are Independent is $P(A \cap B) = P(A)P(B)$.

1.5 Random variables

Definition 1.7

Consider a random experiment with a sample space $\mathcal{C}$. A function X which assigns to each $c \in \mathcal{C}$ one and only one real number $X(c) = x$ is a random variable. The space (or range) of X is the set of real numbers $\mathcal{D} = \{x|x = X(c)$ for some $c \in \mathcal{C}\}$. If $\mathcal{D}$ is a countable set then X is a discrete random variable and if $\mathcal{D}$ is an interval of real numbers then X is a continuous random variable.

Definition 1.8

Let X be a random variable. Then its cumulative distribution function (cdf) $F : \mathbb{R} \to [0, 1]$ is defined as follows:

$$F(x) = P(X \leqslant x).$$

Theorem 1.8

1.6 Discrete Random Variables

1.7 Continuous Random Variables

1.8 Expectation of a Random Variable

1.9 Some Special Expectations

1.9.1 The Moment Generating Function

Recall ethe McLaurin series

$$f(\alpha) = e^{\alpha} = \sum_{m=0}^{\infty} \frac{\alpha^m}{m!},$$

if we write the random variable

$$e^{tX} = \sum_{m=0}^{\infty} \frac{t^m}{m!} X^m,$$

then its expectation value defines something called the moment generating function (mgf)

$$M(t) = E(e^{tX}) = \sum_{m=0}^{\infty} \frac{t^m}{m!} E(X^m).$$

If we take the mth derivative of the mgf, evaluated at $t = 0$, we get the mth ($m \geqslant 1$) moment:

$$M^m(0) = E(X^m).$$

For this to work, the mgf has to be defined in a neighborhood of the origin, i.e., for $-h < t < h$ where $h > 0$ is some positive number.

Definition 1.9

Let X be a random variable such that for some $h > 0$, the expectation of e^{tX} exists for $-h < t < h$. The moment generating function (or mgf) of X is the function $M(t) = E(e^{tX})$ for $-h < t < h$.

Remark. When a moment generating function exists, we must have for $t = 0$ that $M(0) = E(1) = 1$.

1.10 Homework

Exercise 1.1

Show that the moment generating function of the random variable X having the pdf $f(x) = \frac{1}{3}$, $-1 < x < 2$, zero elsewhere, is

$$M(t) = \begin{cases} \frac{e^{2t}-e^{-t}}{3t} & t \neq 0 \\ 1 & t = 0. \end{cases}$$

Solve For $t \neq 0$,

$$M(t) = E(e^{tX}) = \int_{-\infty}^{+\infty} e^{tx} f(x) dx = \int_{-1}^{2} \frac{1}{3} e^{tx} dx = \frac{1}{3} \frac{e^{tx}}{t} |_{x=-1}^{x=2} = \frac{e^{2t} - e^{-t}}{3t}.$$

And $M(0) = 1$ when a moment generating function exists and so the result follows. □

1.11 Reference

- lecture note
- Probability and Distributions
- Sample space is unique?
- proof of 1.3

Chapter 2

Multivariate Distributions

2.1 Distributions of Two Random Variables

Definition 2.1

Given a random experiment with a sample space $\mathcal{C}$, consider two random variables X_1 and X_2 which assign to each element c of $\mathcal{C}$ one and only one ordered pair of numbers (X_1, X_2) is a random vector. The space of (X_1, X_2) is the set of ordered pairs $\mathcal{D} = \{(x_1, x_2)|x_1 = X_1(c), x_2 = X_2(c), x \in \mathcal{C}\}$.

Definition 2.2

Let $\mathcal{D}$ be the space associated with the random vectors (X_1,X_2). For $A \subset \mathcal{D}$ we call A an event. The cumulative distribution function (cdf) for (X_1,X_2) is

$$F_{X_1,X_2}(x_1, x_2) = P(\{X_1 \leqslant x_1\} \cap \{X_2 \leqslant x_2\}) \tag{2.1}$$

for $(x_1, x_2) \in \mathbb{R}^2$. This is the *joint cumulative distribution function* of (X_1, X_2). If F_{X_1,X_2} is continuous then random variable (X_1, X_2) is said to be continuous.

Definition 2.3

A random vector (X_1, X_2) is a discrete random vector if its space $\mathcal{D}$ is finite or countable. (Hence X_1 and X_2 both must be discrete.) The joint probability mass function of (X_1, X_2) is $p_{X_1,X_2}(x_1, x_2) = P(X_1 = x_1, X_2 = x_2)$ for all $(x_1, x_2) \in \mathcal{D}$.

Definition 2.4

If for random vector (X_1, X_2) with cumulative distribution function F_{X_1,X_2}, there is a function $f_{X_1,X_2} : \mathbb{R}^2 \to \mathbb{R}$ such that

$$F_{X_1,X_2}(x_1, x_2) = \int_{-\infty}^{x_1} \int_{\infty}^{x_2} f_{X_1,X_2}(w_1, w_2) dw_1 dw_2.$$

Then f_{X_1,X_2} is the joint probability density function (pdf) of (X_1, X_2). The support of (X_1, X_2) is the set of all points (x_1, x_2) for which $f_{X_1,X_2}(x_1, x_2) > 0$, denoted $\mathcal{S}$.

Remark. In this course, continuous random vectors will have joint probability density functions that determine the cumulative distribution function. By the Fundamental Theorem of Calculus (applied twice)

$$\frac{\partial^2 F_{X_1,X_2}(x_1, x_2)}{\partial x_1 \partial x_2} = f_{X_1,X_2}(x_1, x_2).$$

For event $A \in \mathcal{D}$, we have

$$P((X_1, X_2) \in A) = \int \int_A f_{X_1,X_2}(x_1, x_2) dx_1 dx_2.$$

Remark. We can find the distribution of random variable X_1 and X_2 (called marginal distribution) based on the joint distribution of (X_1, X_2). We have

$$\{X \leqslant x_1\} = \{X_1 \leqslant x_1\} \cap \{-\infty < X_2 < \infty\},$$

so with F_{x_1}, the cumulative distribution function of X_1 we get for $x_1 \in \mathbb{R}$

$$\begin{aligned} F_{X_1}(x_1) &= P(X \leqslant x_1) = P(X_1 \leqslant x_1, -\infty < X_2 < \infty) \\ &= \lim_{x_2 \to \infty} F_{X_1,X_2}(x_1, x_2). \end{aligned}$$

We can similarly find the marginal distribution F_{X_2} in terms of F_{X_1,X_2}. In the continuous case,

$$\begin{aligned} f_{X_1}(x_1) &= \int_{-\infty}^{\infty} f_{X_1,X_2}(x_1, x_2) dx_2, \\ f_{X_2}(x_2) &= \int_{-\infty}^{\infty} f_{X_1,X_2}(x_1, x_2) dx_1. \end{aligned}$$

2.2 Transformations: Bivariate Random Variables

2.3 Conditional Distributions and Expectations

2.4 Independent Random Variables

2.5 The Correlation Coefficient

2.6 Homework

Exercise 2.1

Let the joint pdf of X and Y be given by

$$f(x,y) = \begin{cases} \frac{2}{(1+x+y)^3} & 0 < x < \infty, 0 < y < \infty \\ 0 & \text{elsewhere.} \end{cases}$$

(a) Compute the marginal pdf of X and the conditional pdf of Y, given $X = x$.
(b) For a fixed $X = x$, compute $E(1 + x + Y|x)$ and use the result to compute $E(Y|x)$.

Solve (a) By the definition of marginal probability density function:

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f(x,y)dy = \int_0^{\infty} \frac{2}{(1+x+y)^3}dy \overset{t=1+x+y}{=} \int_{1+x}^{\infty} \frac{2}{t^3}dt \\ &= -t^{-2}|_{t=1+x}^{t=\infty} = 0 - (-(1+x)^{-2}) = \frac{1}{(1+x)^2}, \text{ for } 0 < x < \infty. \\ f_Y(y) &= \int_{-\infty}^{\infty} f(x,y)dx = \int_0^{\infty} \frac{2}{(1+x+y)^3}dx \\ &= \frac{1}{(1+y)^2}, \text{ for } 0 < y < \infty. \end{aligned}$$

Hence, $f_X(x) = \begin{cases} \frac{1}{(1+x)^2} & 0 < x < \infty \\ 0 & \text{elsewhere} \end{cases}$ and $f_Y(y) = \begin{cases} \frac{1}{(1+y)^2} & 0 < y < \infty \\ 0 & \text{elsewhere} \end{cases}$.

The conditional probability density function of Y given $X = x$ is

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{\frac{2}{(1+x+y)^3}}{\frac{1}{(1+x)^2}} = \frac{2(1+x)^2}{(1+x+y)^3}, \text{ for } 0 < x < \infty.$$

Hence, $f_{Y|X}(y|x) = \begin{cases} \frac{2(1+x)^2}{(1+x+y)^3} & 0 < y < \infty \\ 0 & \text{elsewhere.} \end{cases}$

(b) The conditional expectation of $g(Y) = 1 + X + Y$ given $X = x$ is

$$\begin{aligned}
E(1 + x + Y|x) &= \int_{-\infty}^{\infty} g(y) f_{Y|X}(y|x) dy \\
&= \int_0^{\infty} (1 + x + y) \frac{2(1+x)^2}{(1+x+y)^2} dy \\
&\overset{t=1+x+y}{=} \int_{1+x}^{\infty} \frac{2(1+x)^2}{t^2} dt = -\frac{2(1+x)^2}{t}\Big|_{t=1+x}^{t=\infty} = 2(1+x).
\end{aligned}$$

Since $E(1 + x + Y|x) = 1 + x + E(Y|x)$, $E(Y|x) = 2(1 + x) - (1 + x) = (1 + x)$. □

Exercise 2.2

Let X_1, X_2, X_3 be iid with common pdf $f(x) = \exp(-x)$, $0 < x < \infty$, zero elsewhere. Evaluste:
(a) $P(X_1 < X_2 | X_1 < 2X_2)$.
(b) $P(X_1 < X_2 < X_3 | X_3 < 1)$.

Solve The joint common pdf of X_1, X_2 is

$$f_{X_1,X_2}(x_1, x_2) = \begin{cases} e^{-(x_1+x_2)} & 0 < x_1 < \infty, 0 < x_2 < \infty \\ 0 & \text{elsewhere} \end{cases}$$

The joint common pdf of X_1, X_2, X_3 is

$$f_{X_1,X_2,X_3}(x_1, x_2, x_3) = \begin{cases} e^{-(x_1+x_2+x_3)} & 0 < x_1 < \infty, 0 < x_2 < \infty, 0 < x_3 < \infty \\ 0 & \text{elsewhere} \end{cases}$$

(a) Since

$$\begin{aligned}
P(X_1 < X_2, X_1 < 2X_2) &= \int_0^{\infty} dx_1 \int_{x_1}^{\infty} e^{-(x_1+x_2)} dx_2 = \int_0^{\infty} -e^{-x_1} e^{-x_2} \Big|_{x_2=x_1}^{x_2=\infty} dx_1 \\
&= \int_0^{\infty} 0 - (-e^{-2x_1}) dx_1 \\
&= -\frac{1}{2} e^{-2x_1} \Big|_{x_1=0}^{x_1=\infty} \\
&= \frac{1}{2}
\end{aligned}$$

and

$$\begin{aligned}P(X_1 < 2X_2) &= \int_0^\infty dx_1 \int_{\frac{x_1}{2}}^\infty e^{-(x_1+x_2)} dx_2 = \int_0^\infty -e^{-x_1}e^{-x_2}|_{x_2=\frac{x_1}{2}}^{x_2=\infty} dx_1 \\ &= \int_0^\infty 0 - (-e^{-x_1}e^{-\frac{x_1}{2}})dx_1 \\ &= -\frac{2}{3}e^{-\frac{3}{2}x_1}|_{x_1=0}^{x_1=\infty} \\ &= \frac{2}{3},\end{aligned}$$

$P(X_1 < X_2 | X_1 < 2X_2) = \frac{P(X_1<X_2, X_1<2X_2)}{P(X_1<2X_2)} = \frac{\frac{1}{2}}{\frac{2}{3}} = \frac{3}{4}$. □

2.7 Reference

- chapter 2
- 2.1
- 2.3