Study Notes of Mathematical Statistics

Pei Zhong

May 28, 2024

Contents

1 Probability and Distributions **3**

1.1 Introduction . . . 3
1.2 Sets . . . 3
1.3 The Probability Set Function . . . 3
1.4 Conditional Probability and Independence . . . 5
1.5 Random variables . . . 5
1.6 Discrete Random Variables . . . 6
1.7 Continuous Random Variables . . . 6
1.8 Expectation of a Random Variable . . . 6
1.9 Some Special Expectations . . . 6
1.9.1 The Moment Generating Function . . . 6
1.10 Homework . . . 7
1.11 Reference . . . 7

2 Multivariate Distributions **8**

2.1 Distributions of Two Random Variables . . . 8
2.2 Transformations: Bivariate Random Variables . . . 9
2.3 Conditional Distributions and Expectations . . . 10
2.4 Independent Random Variables . . . 12
2.5 The Correlation Coefficient . . . 12
2.6 Homework . . . 12
2.7 Reference . . . 14

3 Some Special Distributions **15**

3.1 The Binomial and Related Distributions . . . 15
3.2 The Poisson Distribution . . . 15
3.3 The Γ,χ^2, and β Distributions . . . 15
3.4 The Normal Distribution . . . 15
3.5 The Multivariate Normal Distribution . . . 15
3.6 $t-$ and $F-$Distributions . . . 15

3.7 Mixture Distributions 15
3.8 Homework 15
3.9 Reference 16

4 Some Elementary Statistical Inferences 17
4.1 Introduction 17
4.1.1 Types of statistics 17
4.1.2 Statistical inference definition 17
4.1.3 Types of statistical inference 18
4.2 Point Estimators 18
4.3 Confidence intervals 19
4.3.1 Confidence interval for mean μ 19
4.4 Testing hypotheses 20
4.5 Homework 20
4.6 Reference 21

5 Consistency and Limiting Distributions 23
5.1 Homework 23
5.2 Reference 24

6 Maximum Likelihood Methods 25
6.1 Rao-Cramér Lower Bound and Efficiency 25
6.2 Homework 25
6.3 Reference 26

7 Sufficiency 27
7.1 Reference 27

Preface

The notes mainly refer to:

- Introduction to Mathematical Statistics 8th Edition
- lecture note
- Study Guide
- Introduction to mathematical statistics exercise solution

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 $A_1, A_2, \ldots \in \mathcal{B}$ 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 $\mathcal{D} = \{x | x = X(c) \text{ for some } c \in \mathcal{C}\}$. 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: 1.9.7

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. $\square$

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

We now consider transformations of random vectors, say $Y = g(X_1, X_2)$. We desire to find the cumulative distribution function of Y . We give several examples, but state no new theorems.

Let (X_1, X_2) have a jointly continuous distribution with pdf $f_{X_1,X_2}(x_1, x_2)$ and support set $\mathcal{S}$. Consider the transformed random vector $(Y_1, Y_2) = T(X_1, X_2)$ where T is a one-to-one continuous transformation. Let $\mathcal{T} = T(S)$ denote the support of (Y_1, Y_2). The transformation is depicted in Figure 2.1. Rewrite the transformation in terms of its components as $(Y_1, Y_2) = T(X_1, X_2) = (u_1(X_1, X_2), u_2(X_1, X_2))$, where the functions $y_1 = u_1(x_1, x_2)$ and $y_2 = u_2(x1, x2)$ define T . Since the transformation is one-to-one, the inverse transformation T^{-1} exists. We write it as $x_1 = w_1(y_1, y_2), x_2 = w_2(y_1, y_2)$. Finally, we need the Jacobian of the transformation which is the determinant of order 2 given by

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix}$$

Note that J plays the role of dx/dy in the univariate case. We assume that these first-order partial derivatives are continuous and that the Jacobian J is not identically equal to zero in T. Let B be any region in T and let $A = T^{-1}(B)$ as shown in Figure 2.1. Because the transformation T is one-to-one, $P[(X_1, X_2) \in A] = P[T(X_1, X_2) \in T(A)] = P[(Y_1, Y_2) \in B]$. Then we have

$$\begin{aligned} P[(X_1, X_2) \in A] &= \iint_A f_{X_1,X_2}(x_1, x_2) dx_1 dx_2 \\ &= \iint_{T(A)} f_{X_1,X_2}[T^{-1}(y_1, y_2)]|J| dy_1 dy_2 \\ &= \iint_B f_{X_1,X_2}[w_1(y_1, y_2), w_2(y_1, y_2)]|J| dy_1 dy_2 \\ &= \iint_B f_{X_1,X_2}[w_1(y_1, y_2), w_2(y_1, y_2)]|J| dy_1 dy_2. \end{aligned}$$

Since B is arbitrary, the last integrand must be the joint pdf of (Y_1, Y_2). That is the pdf of (Y_1, Y_2) is

$$f_{Y_1,Y_2}(y_1, y_2) = \begin{cases} f_{X_1,X_2}[w_1(y_1, y_2), w_2(y_1, y_2)]|J| & (y_1, y_2) \in \mathcal{T} \\ 0 & \text{elsewhere.} \end{cases}$$

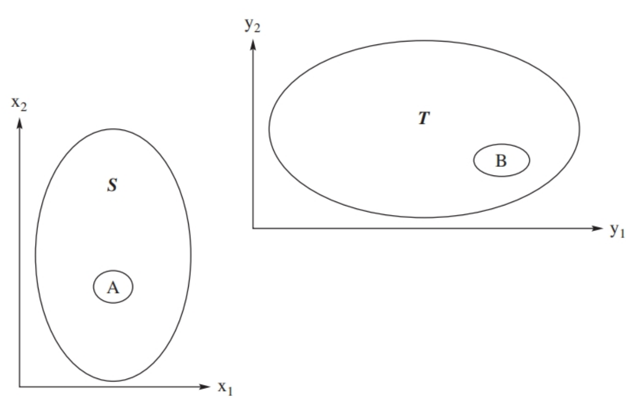

Figure 2.1

2.3 Conditional Distributions and Expectations

We now introduce a parameter ρ of the joint distribution of (X, Y) which quantifies the dependence between X and Y (so that $\rho = 0$ when X and Y are independent). We assume the existence of all expectation under discussion.

Definition 2.5

Let (X, Y) have a joint distribution. Denote the means of X and Y respectively by μ_1 and μ_2 and their respective variances by σ_1^2 and σ_2^2. The covariance of (X, Y) is

$$\text{cov}(X, Y) = E[(X - \mu_1)(Y - \mu_2)].$$

Remark. Since the expectation operator is linear, then

$$\begin{aligned}\text{cov}(X, Y) &= E[XY - \mu_2 X - \mu_1 Y + \mu_1\mu_2] = E[XY] - \mu_2 E[X] - \mu_1 E[Y] + \mu_1\mu_2 \\ &= E[XY] - \mu_1\mu_2 - \mu_1\mu_2 + \mu_1\mu_2 = E[XY] - \mu_1\mu_2.\end{aligned}$$

Definition 2.6

If each of σ_1 and σ_2 is positive then the correlation coefficient between X and Y is

$$\rho = \frac{E[(X - \mu_1)(Y - \mu_2)]}{\sigma_1\sigma_2} = \frac{\text{cov}(X, Y)}{\sigma_1\sigma_2}.$$

Remark. We can relate these parameters as

$$\begin{aligned}E[XY] &= \mu_1\mu_2 + \text{cov}(X, Y) \\ &= \mu_1\mu_2 + \rho\sigma_1\sigma_2.\end{aligned}$$

Theorem 2.1

For all jointly distributed random variables (X, Y) whose correlation coefficient ρ exists (so that $\sigma_1 > 0$ and $\sigma_2 > 0$ by the definition of ρ), we have $-1 \leqslant \rho \leqslant 1$.

Theorem 2.2

If X and Y are independent random variables then $\text{cov}(X, Y) = 0$ and hence $\rho = 0$.

Theorem 2.3

Suppose (X, Y) have a joint distribution with the variances of X and Y finite and positive. Denote the means and variances of X and Y by μ_1, μ_2 and σ_1^2, σ_2^2, respectively, and let ρ be the correlation coefficient between X and Y. If $E[Y|X]$ is linear in X then

$$\begin{gathered}E[Y|X] = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(X - \mu_1) \text{ and} \\ E[Var(Y|X)] = \sigma_2^2(1 - \rho^2).\end{gathered}$$

2.4 Independent Random Variables

2.5 The Correlation Coefficient

2.6 Homework

Exercise 2.1: 2.3.6

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

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

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

Exercise 2.2: 2.6.9

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

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

and

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

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

(b) Since

$$
\begin{aligned}
P(X_1 < X_2 < X_3, X_3 < 1) &= \int_0^1 \{\int_0^{x_3} \{\int_0^{x_2} e^{-(x_1+x_2+x_3)} dx_1\} dx_2\} dx_3 \\
&= \int_0^1 \{\int_0^{x_3} -e^{-(x_1+x_2+x_3)}|_{x_1=0}^{x_1=x_2} dx_2\} dx_3 \\
&= \int_0^1 \{\int_0^{x_3} -e^{-(2x_2+x_3)} + e^{-(x_2+x_3)} dx_2\} dx_3 \\
&= \int_0^1 \frac{1}{2} e^{-(2x_2+x_3)} - e^{-(x_2+x_3)}|_{x_2=0}^{x_2=x_3} dx_3 \\
&= \int_0^1 \frac{1}{2} e^{-x_3} - e^{-2x_3} + \frac{1}{2} e^{-3x_3} dx_3 \\
&= -\frac{1}{2} e^{-x_3} + \frac{1}{2} e^{-2x_3} - \frac{1}{6} e^{-3x_3}|_{x_3=0}^{x_3=1} \\
&= -\frac{1}{2} e^{-1} + \frac{1}{2} e^{-2} - \frac{1}{6} e^{-3} + \frac{1}{6}
\end{aligned}
$$

and

$$
P(X_3 < 1) = \int_0^1 e^{-x} dx = -e^{-x}|_{x=0}^{x=1} = -e^{-1} + 1,
$$

$P(X_1 < X_2 < X_3 | X_3 < 1) = \frac{P(X_1<X_2<X_3, X_3<1)}{P(X_3<1)} = \frac{1-3e^{-1}+3e^{-2}-e^{-3}}{6(1-e^{-1})}$. □

2.7 Reference

- chapter 2
- 2.1
- 2.3
- ch2 solution

Chapter 3

Some Special Distributions

3.1 The Binomial and Related Distributions

3.2 The Poisson Distribution

3.3 The Γ,χ^2, and β Distributions

3.4 The Normal Distribution

3.5 The Multivariate Normal Distribution

3.6 $t-$ and $F-$Distributions

3.7 Mixture Distributions

3.8 Homework

Exercise 3.1: 3.2.17

Let X_1 and X_2 be two independent random variables. Suppose that X_1 and $Y = X_1 + X_2$ have Possion Distributions with means μ_1 and $\mu > \mu_1$, respectively. Find the distribution of X_2.

Exercise 3.2: 3.4.21

Let $f(x)$ and $F(x)$ be the pdf and the cdf, respectively, of a distribution of the continuous type such that $f'(x)$ exists for all x. Let the mean of the truncated distribution that has pdf $g(y) = f(y)/F(b)$, $-\infty < y < b$, zero elsewhere, be equal to $-f(b)/F(b)$ for all real b. Prove that $f(x)$ is a pdf of a standard normal distribution.

Exercise 3.3: 3.5.9

Say the correlation coefficient between the heights of husbands and wives is 0.70 and the mean male height is 5 feet 10 inches with standard deviation 2 inches, and the mean female height is 5 feet 4 inches with standard deviation $1\frac{1}{2}$ inches. Assuming a bivariate normal distribution, what is the best guess of the height of a woman whose husband's height is 6 feet? Find a 95% prediction interval for her height.

3.9 Reference

- ex3.2.17
- ex3.4.21
- ex3.5.9

Chapter 4

Some Elementary Statistical Inferences

4.1 Introduction

Statistics is a branch of Mathematics, that deals with the collection, analysis, interpretation, and the presentation of the numerical data. In other words, it is defined as the collection of quantitative data. The main purpose of Statistics is to make an accurate conclusion using a limited sample about a greater population.

4.1.1 Types of statistics

Statistics can be classified into two different categories. The two different types of Statistics are:

- Descriptive Statistics
- Inferential Statistics

In Statistics, descriptive statistics describe the data, whereas inferential statistics help you make predictions from the data. In inferential statistics, the data are taken from the sample and allows you to generalize the population. In general, inference means "guess", which means making inference about something. So, statistical inference means, making inference about the population. To take a conclusion about the population, it uses various statistical analysis techniques. In this article, one of the types of statistics called inferential statistics is explained in detail. Now, you are going to learn the proper definition of statistical inference, types, solutions, and examples.

4.1.2 Statistical inference definition

Statistical inference is the process of analysing the result and making conclusions from data subject to random variation. It is also called inferential statistics. Hypothesis testing and confidence intervals are the applications of the statistical inference. Statistical inference is a method of making decisions about the parameters of a population, based on random sampling. It helps to assess the relationship between the dependent and independent variables. The purpose of statistical inference to estimate the uncertainty or sample to sample variation. It allows us to provide a probable range of values for the true values of something in the population. The components used for making statistical inference are:

- Sample Size
- Variability in the sample
- Size of the observed differences

4.1.3 Types of statistical inference

There are different types of statistical inferences that are extensively used for making conclusions. They are:

- One sample hypothesis testing
- Confidence Interval
- Pearson Correlation
- Bi-variate regression
- Multi-variate regression
- Chi-square statistics and contingency table
- ANOVA or T-test

4.2 Point Estimators

Until now we have studied Probability, proceeding as follows: we assumed parameters of all distributions to be known and, based on this, computed probabilities of various outcomes (in a random experiment). In this chapter we make the essential transition to Statistics, which is concerned with the exact opposite: the random experiment is performed (usually many times) and the individual outcomes recorded; based on these, we want to estimate values of the distribution parameters (one or more).

Definition 4.1

If the sample $\boldsymbol{X}_1, \boldsymbol{X}_2, ..., \boldsymbol{X}_n$ are iid, then they constitute a random independent sample (RIS) of size n from the population $\mathbf{X}$.

Definition 4.2

Let $T = T(\boldsymbol{X}_1, \boldsymbol{X}_2, ..., \boldsymbol{X}_n)$ be a function of the sample $\boldsymbol{X}_1, \boldsymbol{X}_2, ..., \boldsymbol{X}_n$. Then T is called a statistic.

Remark. Once the sample is drawn, then $t = T(\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_n)$ is called the realization of T, where $\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_n$ is the value of the sample.

Example 4.1

How should we estimate the mean μ of a Normal distribution $N(\mu, \sigma)$, based on a RIS of size n? We would probably take $\overline{X}$ (the sample mean) to be a 'reasonable' estimator of μ [note that this name applies to the random variable $\overline{X}$, with all its potential (would-be) values; as soon as the experiment is completed and a particular value of $\overline{X}$ recorded, this value (i.e. a specific number) is called an estimate of μ].

There is a few related issues we have to sort out:

- How do we know that $\overline{X}$ is a 'good' estimator of μ, i.e. is there some sensible set of criteria which would enable us to judge the quality of individual estimators?
- Using these criteria, can we then find the best estimator of a parameter, at least in some restricted sense?
- Would not it be better to use, instead of a single number [the so called *point estimate*, which can never precisely agree with the exact value of the unknown parameter, and is thus in this sense always wrong], an interval of values which may have a good chance of containing the correct answer?

The rest of this section tackles the first two issues. We start with

4.3 Confidence intervals

The last section considered the issue of so called *point estimates* (good, better and best), but one can easily see that, even for the best of these, a statement which claims a parameter, say μ, to be close to 8.3, is not very informative, unless we can specify what 'close' means. This is the purpose of a confidence interval, which requires quoting the estimate together with specific limits, e.g. 8.3 ± 0.1 (or $8.2 \leftrightarrow 8.4$, using an interval form).

The limits are established to meet a certain (usually 95%) level of confidence (not a probability, since the statement does not involve any randomness – we are either 100% right, or 100% wrong!). The level of confidence ($1 - \alpha$ in general) corresponds to the original, a-priori probability (i.e. before the sample is even taken) of the procedure to get it right (the probability is, as always, in the random sampling). To be able to calculate this probability exactly, we must know what distribution we are sampling from. So, until further notice, we will assume that the distribution is Normal.

4.3.1 Confidence interval for mean μ

Theorem 4.1: Sums of Independent Normal Random Variables

If $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ are mutually independent normal random variables with means $\mu_1, \mu_2, \ldots, \mu_n$ and variances $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$, then the linear combination:

$$\boldsymbol{Y} = \sum_{i=1}^{n} c_i \boldsymbol{X}_i$$

follows the normal distribution:

$$N(\sum_{i=1}^{n} c_i\mu_i, \sum_{i=1}^{n} c_i^2\sigma_i^2)$$

Corollary 4.1

If $\boldsymbol{X}_1, \boldsymbol{X}_2, ..., \boldsymbol{X}_n$ are observations of a random sample of size n from a $N(\mu, \sigma^2)$ population.

- $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i$ is the sample mean of the n observations, and
- $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{X}_i - \overline{\boldsymbol{X}})^2$ is the sample variance of the n observations.

Then:
(1) $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$;
(2) $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$
(3) $\overline{X}$ and S^2 are independent

We first assume that, even though μ is to be estimated (being unknown), we still know the exact (population) value of σ (based on past experience). We know that

4.4 Testing hypotheses

Suppose now that, instead of trying to estimate

4.5 Homework

Exercise 4.1: 4.5.8

Let us say the life of a tire in miles, say X, is normally distributed with mean θ and standard deviation 5000. Past experience indicates that $\theta = 30,000$. The manufacturer claims that the tires made by a new process have mean $\theta > 30,000$. It is possible that $\theta = 35,000$. Check his claim by testing $H_0 : \theta = 30,000$ against $H_1 : \theta > 30,000$. We observe n independent values of X, say $x_1, ..., x_n$, and we reject H_0 (thus accept H_1) if and only if $\overline{x} \geqslant c$. Determine n and c so that the power function $\gamma(\theta)$ of the test has the values $\gamma(30,000) = 0.01$ and $\gamma(35,000) = 0.98$.

Exercise 4.2: 4.5.11

Let $Y_1 < Y_2 < Y_3 < Y_4$ be the order statistics of a random sample of size $n = 4$ from a distribution with pdf $f(x;\theta) = 1/\theta, 0 < x < \theta$, zero elsewhere, where $0 < \theta$. The hypothesis $H_0 : \theta = 1$ is rejected and $H_1 : \theta > 1$ is accepted if the observed $Y_4 \geqslant c$.
(a) Find the constant c so that the significance level is $\alpha = 0.05$.

(b) Determine the power function of the test.

Exercise 4.3: 4.6.5

On page 373 Rasmussen (1992) discussed a paired design. A baseball coach paired 20 members of his team by their speed; i.e., each member of the pair has about the same speed. Then for each, he randomly chose one member of the pair and told him that if could beat his best time in circling the bases he would give him an award (call this response the time of the "self" member). For the other member of the pair the coach's instruction was an award if he could beat the time of the other member of the pair (call this response the time of the "rival" member). Each member of the pair knew who his rival was. The data are given below, but are also in the file `selfrival.rda`. Let μ_d be the true difference in times (rival minus self) for a pair. The hypotheses of interest are $H_0 : \mu_d = 0$ versus $H_1 : \mu_d < 0$. The data are in order by pairs, so do not mix the order.

```
self:  16.20 16.78 17.38 17.59 17.37 17.49 18.18 18.16 18.36 18.53 15.92
16.58 17.57 16.75 17.28 17.32 17.51 17.58 18.26 17.87
rival:  15.95 16.15 17.05 16.99 17.34 17.53 17.34 17.51 18.10 18.19 16.04
16.80 17.24 16.81 17.11 17.22 17.33 17.82 18.19 17.88
```

(a) Obtain comparison boxplots of the data. Comment on the comparison plots. Are there any outliers?

(b) Compute the paired t-test and obtain the p-value. Are the data significant at the 5% level of significance?

(c) Obtain a point estimate of μ_d and a 95% confidence interval for it.

(d) Conclude in terms of the problem.

4.6 Reference

- Statistical Inference
- Sampling and Statistics
- Chapter 3 RANDOM SAMPLING
- Chapter 5 ESTIMATING DISTRIBUTION PARAMETERS
- Chapter 6 CONFIDENCE INTERVALS
- Chapter 7 TESTING HYPOTHESES
- Sampling distribution of a single normal population
- Sampling distribution of two normal populations
- Sampling Distribution of Sample Mean
- Power of a Statistical Test
- Chapter 4: Some Elementary Statistical Inferences

- ex4.5.8
- ex4.5.11
- ex4.6.5

Chapter 5

Consistency and Limiting Distributions

5.1 Homework

Exercise 5.1: 5.1.7

Let $X_1, \ldots, X_n$ be iid random variables with common pdf

$$f(x) = \begin{cases} e^{-(x-\theta)} & x > \theta, -\infty < \theta < \infty \\ 0 & \text{elsewhere.} \end{cases}$$

This pdf is called the shifted exponential. Let $Y_n = \min\{X_1, \ldots, X_n\}$. Prove that $Y_n \to \theta$ in probability by first obtaining the cdf of Y_n.

Exercise 5.2: 5.1.9

For Exercise 5.1.7, obtain the mean of Y_n. Is Y_n an unbiased estimator of θ? Obtain an unbiased estimator of θ based on Y_n.

Exercise 5.3: 5.2.17

Let $\overline{X}_n$ denote the mean of a random sample of size n from a distribution that has pdf $f(x) = e^{-x}, 0 < x < \infty$, zero elsewhere.
(a) Show that the mgf $M_{Y_n}(t)$ of $Y_n = \sqrt{n}(\overline{X}_n - 1)$ is

$$M_{Y_n}(t) = [e^{t/\sqrt{n}} - (t/\sqrt{n})e^{t/\sqrt{n}}]^{-n}, \quad t < \sqrt{n}.$$

(b) Find the Limiting distribution of Y_n as $n \to \infty$.

Exercise 5.4: 5.3.8

Let Y be $b(n, 0.55)$. Find the smallest value of n such that (approximately) $P(\frac{Y}{n} > \frac{1}{2}) \geqslant 0.95$.

Exercise 5.5: 5.3.14

Assume that $X_1, \ldots, X_n$ is a random sample from a $\Gamma(1, \beta)$ distribution. Determine the asymptotic distribution of $\sqrt{n}(\overline{X} - \beta)$. Then find a transformation $g(\overline{X})$ whose asymptotic variance is free of β.

5.2 Reference

- Chapter 5: Consistency and Limiting Distributions
- ex5.1.7
- ex5.1.9
- ex5.2.17
- ex5.3.8
- ex5.3.14

Chapter 6

Maximum Likelihood Methods

6.1 Rao-Cramér Lower Bound and Efficiency

6.2 Homework

Exercise 6.1: 6.1.4

Suppose $X_1, \ldots, X_n$ are iid with pdf $f(x;\theta) = \frac{2x}{\theta^2}, 0 < x \leqslant \theta$, zero elsewhere. Note this is a nonregular case. Find:
(a) The mle $\hat{\theta}$ for θ.
(b) The constant c so that $E(c\hat{\theta}) = \theta$.
(c) The mle for the median of the distribution. Show that it is a consistent estimator.

Exercise 6.2: 6.2.8

Let X be $N(0,\theta), 0 < \theta < \infty$.
(a) Find the Fisher information $I(\theta)$.
(b) If $X_1, X_2, \ldots, X_n$ is a random sample from this distribution, show that the mle of θ is an efficient estimator of θ.
(c) What is the asymptotic distribution of $\sqrt{n}(\hat{\theta} - \theta)$?

Exercise 6.3: 6.2.11

Let $\overline{X}$ be the mean of a random sample of size n from a $N(\theta, \sigma^2)$ distribution, $-\infty < \theta < \infty, \sigma^2 > 0$. Assume that σ^2 is known. Show that $\overline{X}^2 - \frac{\sigma^2}{n}$ is an unbiased estimator of θ^2 and find its efficiency.

Exercise 6.4: 6.4.3

Let $X_1, X_2, \ldots, X_n$ be iid, each with the distribution having pdf $f(x;\theta_1,\theta_2) = (1/\theta_2)e^{-(x-\theta_1)/\theta_2}, \theta_1 \leqslant x < \infty, -\infty < \theta_2 < \infty$, zero elsewhere. Find the maximum likelihood estimators of θ_1 and θ_2.

6.3 Reference

- Maximum Likelihood Methods
- ex6.1.4
- ex6.2.8

Chapter 7

Sufficiency

Exercise 7.1: 7.1.2

Let $X_1, X_2, \ldots, X_n$ denote a random sample from a normal distribution with mean zero and variance θ, $0 < \theta < \infty$. Show that $\sum_{i=1}^{n} X_i^2/n$ is an unbiased estimator of θ and has variance $2\theta^2/n$.

Exercise 7.2: 7.1.6

Let $X_1, X_2, \ldots, X_n$ denote a random sample from a Poisson distribution with parameter θ, $0 < \theta < \infty$. Let $Y = \sum_{i=1}^{n} X_i$ and let $\mathcal{L}[\theta, \delta(y)] = [\theta - \delta(y)]^2$. If we restrict our considerations to decision functions of the form $\delta(y) = b + y/n$, where b does not depend on y, show that $R(\theta, \delta) = b^2 + \theta/n$. What decision function of this form yields a uniformly smaller risk than every other decision function of this form? With this solution, say δ, and $0 < \theta < \infty$, determine $\max_\theta R(\theta, \delta)$ if it exists.

Exercise 7.3: 7.3.3

If X_1,X_2 is a random sample of size 2 from a distribution having pdf $f(x; \theta) = (1/\theta)e^{-x/\theta}$, $0 < x < \infty$,$0 < \theta < \infty$,zero elsewhere, find the joint pdf of the sufficient statistic $Y_1 = X_1 + X_2$ for θ and $Y_2 = X_2$. Show that Y_2 is an unbiased estimator of θ with variance θ^2. Find $E(Y_2|y_1) = \phi(y_1)$ and the variance of $\phi(Y_1)$.

7.1 Reference

- Sufficiency
- 7.1.2
- 7.1.6