STATISTICS 330 COURSE NOTES

Cyntha A. Struthers
Department of Statistics and Actuarial Science, University of Waterloo

Spring 2024 Edition

Contents

Preface

In order to provide improved versions of these Course Notes for students in subsequent terms, please email corrections, sections that are confusing, or comments/suggestions to castruth@uwaterloo.ca.

1. Preview

The following examples will illustrate the ideas and concepts discussed in these Course Notes. They also indicate how these ideas and concepts are connected to each other.

1.1 Example

The number of service interruptions in a communications system over 200 separate days is summarized in the following frequency table:

Number of interruptions	0	1	2	3	4	5	>5	Total
Observed frequency	64	71	42	18	4	1	0	200

It is believed that a Poisson model will fit these data well. Why might this be a reasonable assumption? (*PROBABILITY MODELS*)

If we let the random variable $X =$ *number of interruptions in a day* and assume that the Poisson model is reasonable then the probability function of X is given by

$$P(X = x) = \frac{\mu^x e^{-\mu}}{x!} \quad \text{for } x = 0, 1, \ldots$$

where μ is a parameter of the model which represents the mean number of service interruptions in a day. (*RANDOM VARIABLES, PROBABILITY FUNCTIONS, EXPECTATION, MODEL PARAMETERS*) Since μ is unknown we might estimate it using the sample mean

$$\bar{x} = \frac{64(0) + 71(1) + \cdots + 1(5)}{200} = \frac{230}{200} = 1.15$$

(*POINT ESTIMATION*) The estimate $\hat{\mu} = \bar{x}$ is the maximum likelihood estimate of μ. It is the value of μ which maximizes the likelihood function. (*MAXIMUM LIKELIHOOD ESTIMATION*) The likelihood function is the probability of the observed data as a function of the unknown parameter(s) in the model. The maximum likelihood estimate is thus the value of μ which maximizes the probability of observing the given data.

In this example the likelihood function is given by

$$\begin{aligned}
L(\mu) &= P(\text{observing 0 interruptions 64 times,} \ldots, > 5 \text{ interruptions 0 times;} \mu) \\
&= \frac{200!}{64!71! \cdots 1!0!} \left(\frac{\mu^0 e^{-\mu}}{0!}\right)^{64} \left(\frac{\mu^1 e^{-\mu}}{1!}\right)^{71} \cdots \left(\frac{\mu^5 e^{-\mu}}{5!}\right)^{1} \left(\sum_{x=6}^{\infty} \frac{\mu^x e^{-\mu}}{x!}\right)^{0} \\
&= c\mu^{64(0)+71(1)+\cdots+1(5)} e^{-(64+71+\cdots+1)\mu} \\
&= c\mu^{-230} e^{-200\mu} \quad \text{for } \mu > 0
\end{aligned}$$

where

$$c = \frac{200!}{64!71! \cdots 1!0!} \left(\frac{1}{0!}\right)^{64} \left(\frac{1}{1!}\right)^{71} \cdots \left(\frac{1}{5!}\right)^{1}$$

The maximum likelihood estimate of μ can be found by solving $\frac{dL}{d\mu} = 0$ or equivalently $\frac{d \log L}{d\mu} = 0$ and verifying that it corresponds to a maximum.

If we want an interval of values for μ which are reasonable given the data then we could construct a confidence interval for μ. (*INTERVAL ESTIMATION*) To construct confidence intervals we need to find the sampling distribution of the estimator. In this example we would need to find the distribution of the estimator

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

where X_i = *number of interruptions in a day* i, $i = 1, 2, \ldots, 200$. (*FUNCTIONS OF RANDOM VARIABLES: cumulative distribution function technique, one-to-one transformations, moment generating function technique*) Since $X_i \sim \text{Poisson}(\mu)$ with $E(X_i) = \mu$ and $Var(X_i) = \mu$ the distribution of $\bar{X}$ for large n is approximately $\text{N}(\mu, \mu/n)$ by the Central Limit Theorem. (*LIMITING DISTRIBUTIONS*)

Suppose the manufacturer of the communications system claimed that the mean number of interruptions was 1. Then we would like to test the hypothesis $H : \mu = 1$. (*TESTS OF HYPOTHESIS*) A test of hypothesis uses a test statistic to measure the evidence based on the observed data against the hypothesis. A test statistic with good properties for testing $H : \mu = \mu_0$ is the likelihood ratio statistic, $-2 \log \left[L(\mu_0) / L(\hat{\mu})\right]$. (*LIKELIHOOD RATIO STATISTIC*) For large n the distribution of the likelihood ratio statistic is approximately $\chi^2(1)$ if the hypothesis $H : \mu = \mu_0$ is true.

1.2 Example

The following are relief times in hours for 20 patients receiving a pain killer:

1.1	1.4	1.3	1.7	1.9	1.8	1.6	2.2	1.7	2.7
4.1	1.8	1.5	1.2	1.4	3.0	1.7	2.3	1.6	2.0

It is believed that the Weibull distribution with probability density function

$$f(x) = \frac{\alpha}{\beta^\alpha} x^{\alpha-1} e^{-(x/\beta)^\alpha} \quad \text{for } x > 0; \quad \alpha > 0, \quad \beta > 0$$

will provide a good fit to the data. (*CONTINUOUS MODELS, PROBABILITY DENSITY FUNCTIONS*) Assuming independent observations the (approximate) likelihood function is

$$L(\alpha, \beta) = \prod_{i=1}^{20} \frac{\alpha}{\beta^{\alpha}} x_i^{\alpha-1} e^{-(x_i/\beta)^{\alpha}} \quad \text{for } \alpha > 0, \quad \beta > 0$$

where x_i is the observed relief time for the ith patient. (*MULTIPARAMETER LIKELIHOODS*) The maximum likelihood estimates $\hat{\alpha}$ and $\hat{\beta}$ are found by simultaneously solving

$$\frac{\partial \log L}{\partial \alpha} = 0 \quad \text{and} \quad \frac{\partial \log L}{\partial \beta} = 0$$

Since an explicit solution to these equations cannot be obtained, a numerical solution must be found using an iterative method. (*NEWTON'S METHOD*) Also, since the maximum likelihood estimators cannot be given explicitly, approximate confidence intervals and tests of hypothesis must be based on the asymptotic distributions of the maximum likelihood estimators. (*LIMITING OR ASYMPTOTIC DISTRIBUTIONS OF MAXIMUM LIKELIHOOD ESTIMATORS*)

2. Univariate Random Variables

In this chapter we review concepts that were introduced in a previous probability course such as STAT 220/230/240 as well as introducing new concepts. In Section 2.1 the concepts of random experiments, sample spaces, probability models, and rules of probability are reviewed. The concepts of a sigma algebra and probability set function are also introduced. In Section 2.2 we define a random variable and its cumulative distribution function. It is important to note that the definition of a cumulative distribution function is the same for all types of random variables. In Section 2.3 we define a discrete random variable and review the named discrete distributions (Hypergeometric, Binomial, Geometric, Negative Binomial, Poisson). In Section 2.4 we define a continuous random variable and review the named continuous distributions (Uniform, Exponential, Normal, and Chi-squared). We also introduce new continuous distributions (Gamma, Two Parameter Exponential, Weibull, Cauchy, Pareto). **A summary of the named discrete and continuous distributions that are used in these Course Notes can be found in Chapter 11**. In Section 2.5 we review the cumulative distribution function technique for finding a function of a random variable and prove a theorem which can be used in the case of a monotone function. In Section 2.6 we review expectations of functions of random variables. In Sections 2.8 – 2.10 we introduce new material related to expectation such as inequalities, variance stabilizing transformations and moment generating functions. **Section 2.11 contains a number of useful calculus results which will be used throughout these Course Notes.**

2.1 Probability

To model real life phenomena for which we cannot predict exactly what will happen we assign numbers, called probabilities, to outcomes of interest which reflect the likelihood of such outcomes. To do this it is useful to introduce the concepts of an experiment and its associated sample space. Consider some phenomenon or process which is repeatable, at least in theory. We call the phenomenon or process a *random experiment* and refer to a single repetition of the experiment as a *trial*. For such an experiment we consider the set of all possible outcomes.

2.1.1 Definition - Sample Space

A *sample space* S is a set of all the distinct outcomes for a random experiment, with the property that in a single trial, one and only one of these outcomes occurs.

To assign probabilities to the events of interest for a given experiment we begin by defining a collection of subsets of a sample space S which is rich enough to define all the events of interest for the experiment. We call such a collection of subsets a *sigma algebra*.

2.1.2 Definition - Sigma Algebra

A collection of subsets of a set S is called a sigma algebra, denoted by $\mathcal{B}$, if it satisfies the following properties:

(1) $\varnothing \in \mathcal{B}$ where $\varnothing$ is the empty set

(2) If $A \in \mathcal{B}$ then $\bar{A} \in \mathcal{B}$

(3) If $A_1, A_2, \ldots \in \mathcal{B}$ then $\bigcup\limits_{i=1}^{\infty} A_i \in \mathcal{B}$

Suppose $A_1, A_2, \ldots$ are subsets of the sample space S which correspond to events of interest for the experiment. To complete the probability model for the experiment we need to assign real numbers $P(A_i)$, $i = 1, 2, \ldots$, where $P(A_i)$ is called the probability of A_i. To develop the theory of probability these probabilities must satisfy certain properties. The following *Axioms of Probability* are a set of axioms which allow a mathematical structure to be developed.

2.1.3 Definition - Probability Set Function

Let $\mathcal{B}$ be a sigma algebra associated with the sample space S. A probability set function is a function P with domain $\mathcal{B}$ that satisfies the following axioms:

$(A1)$ $P(A) \geq 0$ for all $A \in \mathcal{B}$

$(A2)$ $P(S) = 1$

$(A3)$ If $A_1, A_2, \ldots \in \mathcal{B}$ are pairwise mutually exclusive events, that is, $A_i \cap A_j = \varnothing$ for all $i \neq j$, then

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

Note: The probabilities $P(A_1), P(A_2), \ldots$ can be assigned in any way as long they satisfy these three axioms. However, if we wish to model real life phenomena we would assign the probabilities such that they correspond to the relative frequencies of events in a repeatable experiment.

2.1.4 Example

Let $\mathcal{B}$ be a sigma algebra associated with the sample space S and let P be a probability set function with domain $\mathcal{B}$. If $A, B \in \mathcal{B}$ then prove the following:
(a) $P(\varnothing) = 0$
(b) If $A, B \in \mathcal{B}$ and A and B are mutually exclusive events then $P(A \cup B) = P(A) + P(B)$.
(c) $P(\bar{A}) = 1 - P(A)$
(d) If $A \subset B$ then $P(A) \leq P(B)$ **Note:** $A \subset B$ means $a \in A$ implies $a \in B$.

Solution
(a) Let $A_1 = S$ and $A_i = \varnothing$ for $i = 2, 3, \ldots$. Since $\bigcup\limits_{i=1}^{\infty} A_i = S$ then by Definition 2.1.3 $(A3)$ it follows that

$$P(S) = P(S) + \sum_{i=2}^{\infty} P(\varnothing)$$

and by $(A2)$ we have

$$1 = 1 + \sum_{i=2}^{\infty} P(\varnothing)$$

By $(A1)$ the right side is a series of non-negative numbers which must converge to the left side which is 1 which is finite which results in a contradiction unless $P(\varnothing) = 0$ as required.
(b) Let $A_1 = A$, $A_2 = B$, and $A_i = \varnothing$ for $i = 3, 4, \ldots$. Since $\bigcup\limits_{i=1}^{\infty} A_i = A \cup B$ then by $(A3)$

$$P(A \cup B) = P(A) + P(B) + \sum_{i=3}^{\infty} P(\varnothing)$$

and since $P(\varnothing) = 0$ by the result (a) it follows that

$$P(A \cup B) = P(A) + P(B)$$

(c) Since $S = A \cup \bar{A}$ and $A \cap \bar{A} = \varnothing$ then by $(A2)$ and the result proved in (b) it follows that

$$1 = P(S) = P\left(A \cup \bar{A}\right) = P(A) + P\left(\bar{A}\right)$$

or

$$P\left(\bar{A}\right) = 1 - P(A)$$

(d) Since $B = (A \cap B) \cup \left(\bar{A} \cap B\right) = A \cup \left(\bar{A} \cap B\right)$ and $A \cap \left(\bar{A} \cap B\right) = \varnothing$ then by (b) $P(B) = P(A) + P\left(\bar{A} \cap B\right)$. But by $(A1)$, $P\left(\bar{A} \cap B\right) \geq 0$ so it follows that $P(B) \geq P(A)$.

2.1.5 Exercise

Let $\mathcal{B}$ be a sigma algebra associated with the sample space S and let P be a probability set function with domain $\mathcal{B}$. If $A, B \in \mathcal{B}$ then prove the following:
(a) $0 \leq P(A) \leq 1$
(b) $P\left(A \cap \bar{B}\right) = P(A) - P(A \cap B)$
(c) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

For a given experiment we are sometimes interested in the probability of an event given that we know that the event of interest has occurred in a certain subset of S. For example, the experiment might involve people of different ages and we may be interest in an event only for a given age group. This leads us to define conditional probability.

2.1.6 Definition - Conditional Probability

Let $\mathcal{B}$ be a sigma algebra associated with the sample space S and suppose $A, B \in \mathcal{B}$. The conditional probability of event A given event B is

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{provided } P(B) > 0$$

2.1.7 Example

The following table of probabilities are based on data from the 2011 Canadian census. The probabilities are for Canadians aged $25 - 34$.

Highest level of education attained	Employed	Unemployed
No certificate, diploma or degree	0.066	0.010
High school diploma or equivalent	0.185	0.016
Postsecondary certificate, diploma or degree	0.683	0.040

If a person is selected at random what is the probability the person

(a) is employed?

(b) has no certificate, diploma or degree?

(c) is unemployed and has at least a high school diploma or equivalent?

(d) has at least a high school diploma or equivalent given that they are unemployed?

Solution

(a) Let E be the event "employed", A_1 be the event "no certificate, diploma or degree", A_2 be the event "high school diploma or equivalent", and A_3 be the event "postsecondary certificate, diploma or degree".

$$\begin{aligned} P(E) &= P(E \cap A_1) + P(E \cap A_2) + P(E \cap A_3) \\ &= 0.066 + 0.185 + 0.683 \\ &= 0.934 \end{aligned}$$

(b)

$$\begin{aligned} P(A_1) &= P(E \cap A_1) + P(\bar{E} \cap A_1) \\ &= 0.066 + 0.010 \\ &= 0.076 \end{aligned}$$

(c)

$$\begin{aligned} & P(\text{unemployed and have at least a high school diploma or equivalent}) \\ = \; & P(\bar{E} \cap (A_2 \cup A_3)) \\ = \; & P(\bar{E} \cap A_2) + P(\bar{E} \cap A_3) \\ = \; & 0.016 + 0.040 = 0.056 \end{aligned}$$

(d)

$$\begin{aligned} P(A_2 \cup A_3 | \bar{E}) &= \frac{P(\bar{E} \cap (A_2 \cup A_3))}{P(\bar{E})} \\ &= \frac{0.056}{0.066} \\ &= 0.848 \end{aligned}$$

If the occurrence of event B does not affect the probability of the event A, then the events are called independent events.

2.1.8 Definition - Independent Events

Let $\mathcal{B}$ be a sigma algebra associated with the sample space S and suppose $A, B \in \mathcal{B}$. A and B are independent events if

$$P(A \cap B) = P(A) P(B)$$

2.1.9 Example

In Example 2.1.7 are the events, "unemployed" and "no certificate, diploma or degree", independent events?

Solution

The events "unemployed" and "no certificate, diploma or degree" are not independent since

$$\begin{aligned} 0.010 &= P(\bar{E} \cap A_1) \\ &\neq P(\bar{E}) P(A_1) = (0.066)(0.076) \end{aligned}$$

2.2 Random Variables

A probability model for a random experiment is often easier to construct if the outcomes of the experiment are real numbers. When the outcomes are not real numbers, the outcomes can be mapped to numbers using a function called a random variable. When the observed data are numerical values such as the number of interruptions in a day in a communications system or the length of time until relief after taking a pain killer, random variables are still used in constructing probability models.

2.2.1 Definition of a Random Variable

A *random variable* X is a function from a sample space S to the real numbers $\Re$, that is,

$$X : S \to \Re$$

such that $P(X \leq x)$ is defined for all $x \in \Re$.

Note: '$X \leq x$' is an abbreviation for $\{\omega \in S : X(\omega) \leq x\}$ where $\{\omega \in S : X(\omega) \leq x\} \in \mathcal{B}$ and $\mathcal{B}$ is a sigma algebra associated with the sample space S.

2.2.2 Example

Three friends Ali, Benita and Chen are enrolled in STAT 330. Suppose we are interested in whether these friends earn a grade of 70 or more. If we let A represent the event "Ali earns a grade of 70 or more", B represent the event "Benita earns a grade of 70 or more", and C represent the event "Chen earns a grade of 70 or more" then a suitable sample space is

$$S = \left\{ABC, \bar{A}BC, A\bar{B}C, AB\bar{C}, \bar{A}\bar{B}C, A\bar{B}\bar{C}, \bar{A}B\bar{C}, \bar{A}\bar{B}\bar{C}\right\}$$

Suppose we are mostly interested in how many of these friends earn a grade of 70 or more. We can define the random variable $X =$ "number of friends who earn a grade of 70 or more". The range of X is $\{0, 1, 2, 3\}$ with associated mapping

$$\begin{aligned}
X\left(ABC\right) &= 3 \\
X\left(\bar{A}BC\right) &= X\left(A\bar{B}C\right) = X\left(AB\bar{C}\right) = 2 \\
X\left(\bar{A}\bar{B}C\right) &= X\left(A\bar{B}\bar{C}\right) = X\left(\bar{A}B\bar{C}\right) = 1 \\
X\left(\bar{A}\bar{B}\bar{C}\right) &= 0
\end{aligned}$$

An important function associated with random variables is the cumulative distribution function.

2.2.3 Definition - Cumulative Distribution Function

The *cumulative distribution function* (c.d.f.) of a random variable X is defined by

$$F(x) = P(X \leq x) \quad \text{for } x \in \Re$$

Note: The cumulative distribution function is defined for all real numbers.

2.2.4 Properties - Cumulative Distribution Function

(1) F is a non-decreasing function, that is,

$$F(x_1) \leq F(x_2) \ \text{ for all } x_1 < x_2$$

(2)

$$\lim_{x \to -\infty} F(x) = 0 \ \text{ and } \ \lim_{x \to \infty} F(x) = 1$$

(3) F is a right-continuous function, that is,

$$\lim_{x \to a^+} F(x) = F(a)$$

(4) For all $a < b$

$$P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a)$$

(5) For all b

$$P(X = b) = F(b) - \lim_{a \to b^-} F(a)$$

2.2.5 Example

Suppose X is a random variable with cumulative distribution function

$$F(x) = P(X \leq x) = \begin{cases} 0 & x < 0 \\ 0.1 & 0 \leq x < 1 \\ 0.3 & 1 \leq x < 2 \\ 0.6 & 2 \leq x < 3 \\ 1 & x \geq 3 \end{cases}$$

(a) Graph the function $F(x)$.

(b) Determine the probabilities

(i) $P(X \leq 1)$
(ii) $P(X \leq 2)$
(iii) $P(X \leq 2.4)$
(iv) $P(X = 2)$
(v) $P(0 < X \leq 2)$

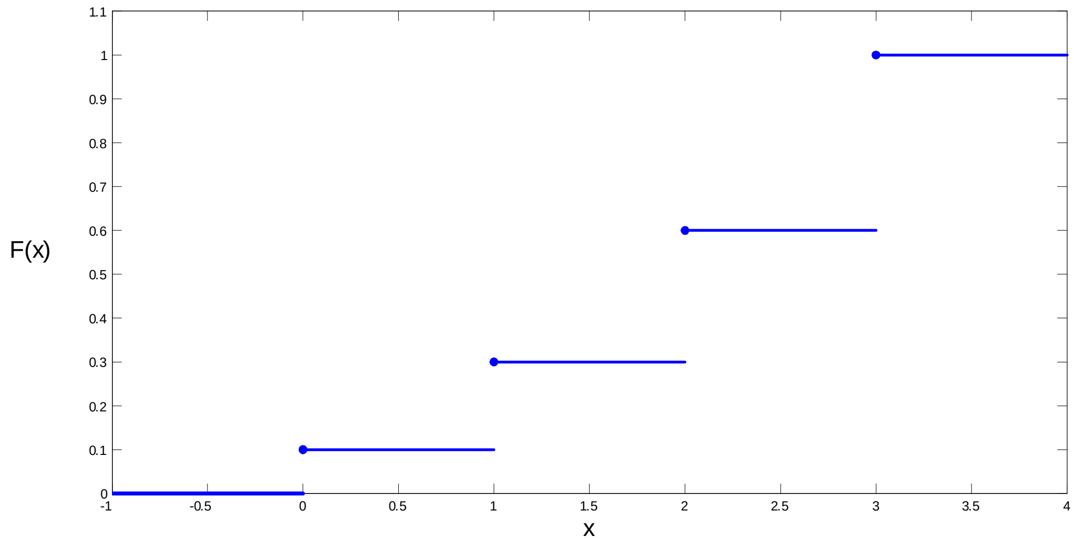


Figure 2.1: Graph of $F(x) = P(X \leq x)$ for Example 2.2.5

(vi) $P(0 \leq X \leq 2)$.

Solution

(a) See Figure 2.1.

(b) (i)

$$P(X \leq 1) = F(1) = 0.3$$

(ii)

$$P(X \leq 2) = F(2) = 0.6$$

(iii)

$$P(X \leq 2.4) = P(X \leq 2) = F(2) = 0.6$$

(iv)

$$P(X = 2) = F(2) - \lim_{x \to 2^-} F(x) = 0.6 - 0.3 = 0.3$$

or

$$P(X = 2) = P(X \leq 2) - P(X \leq 1) = F(2) - F(1) = 0.6 - 0.3 = 0.3$$

(v)

$$P(0 < X \leq 2) = P(X \leq 2) - P(X \leq 0) = F(2) - F(0) = 0.6 - 0.1 = 0.5$$

(vi)

$$P(0 \leq X \leq 2) = P(X \leq 2) - P(X < 0) = F(2) - 0 = 0.6$$

2.2.6 Example

Suppose X is a random variable with cumulative distribution function

$$F(x) = P(X \le x) = \begin{cases} 0 & x \le 0 \\ \frac{2}{5}x^3 & 0 < x \le 1 \\ \frac{1}{5}\left(12x - 3x^2 - 7\right) & 1 < x < 2 \\ 1 & x \ge 2 \end{cases}$$

(a) Graph the function $F(x)$.

(b) Determine the probabilities

(i) $P(X \le 1)$

(ii) $P(X \le 2)$

(iii) $P(X \le 2.4)$

(iv) $P(X = 0.5)$

(v) $P(X = b)$, for $b \in \Re$

(vi) $P(1 < X \le 2.4)$

(vii) $P(1 \le X \le 2.4)$.

Solution

(a) See Figure 2.2

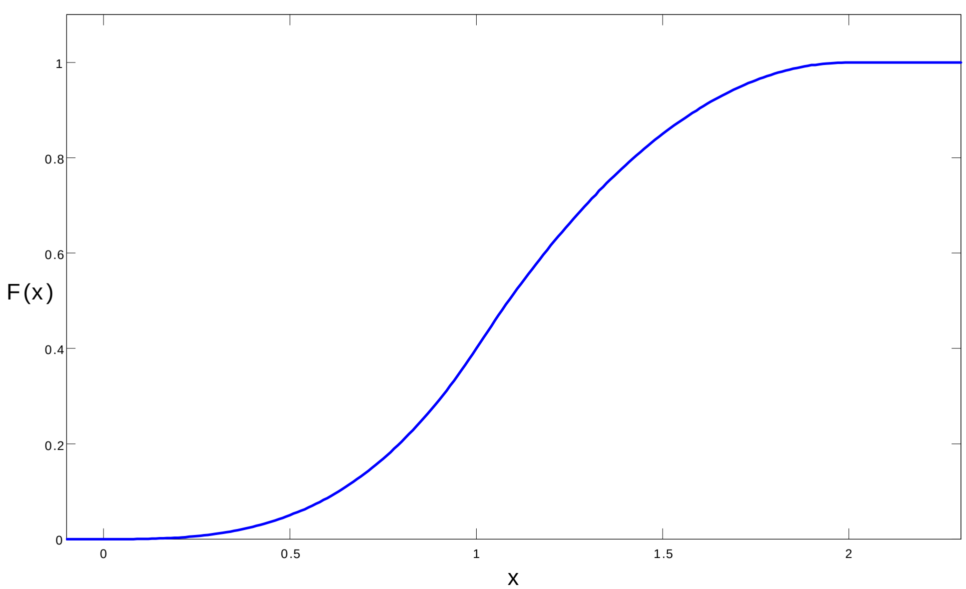


Figure 2.2: Graph of $F(x) = P(X \le x)$ for Example 2.2.6

(*b*) (*i*)

$$\begin{aligned} P(X \le 1) &= F(1) \\ &= \frac{2}{5}(1)^3 = \frac{2}{5} \\ &= 0.4 \end{aligned}$$

(*ii*)

$$P(X \le 2) = F(2) = 1$$

(*iii*)

$$P(X \le 2.4) = F(2.4) = 1$$

(*iv*)

$$P(X = 0.5)$$

$$\begin{aligned} &= F(0.5) - \lim_{x \to 0.5^-} F(x) \\ &= F(0.5) - F(0.5) \\ &= 0 \end{aligned}$$

(*v*)

$$P(X = b)$$

$$\begin{aligned} &= F(b) - \lim_{a \to b^-} F(a) \\ &= F(b) - F(b) \\ &= 0 \quad \text{for all } b \in \Re \end{aligned}$$

(*vi*)

$$\begin{aligned} P(1 < X \le 2.4) &= F(2.4) - F(1) \\ &= 1 - 0.4 \\ &= 0.6 \end{aligned}$$

(*vii*)

$$\begin{aligned} P(1 \le X \le 2.4) &= P(1 < X \le 2.4) \quad \text{since } P(X = 1) = 0 \\ &= 0.6 \end{aligned}$$

2.3 Discrete Random Variables

A set A is countable if the number of elements in the set is finite or the elements of the set can be put into a one-to-one correspondence with the positive integers.

2.3.1 Definition - Discrete Random Variable

A random variable X defined on a sample space S is a *discrete random variable* if there is a countable subset $A \subset \Re$ such that $P(X \in A) = 1$.

2.3.2 Definition - Probability Function

If X is a discrete random variable then the *probability function* (p.f.) of X is given by

$$\begin{aligned} f(x) &= P(X = x) \\ &= F(x) - \lim_{\varepsilon \to 0^+} F(x - \varepsilon) \quad \text{for } x \in \Re \end{aligned}$$

The set $A = \{x : f(x) > 0\}$ is called the support set of X.

2.3.3 Properties - Probability Function

(1)

$$f(x) \geq 0 \quad \text{for } x \in \Re$$

(2)

$$\sum_{x \in A} f(x) = 1$$

2.3.4 Example

In Example 2.2.5 find the support set A, show that X is a discrete random variable and determine its probability function.

Solution

The support set of X is $A = \{0, 1, 2, 3\}$ which is a countable set. Its probability function is

$$f(x) = P(X = x) = \begin{cases} 0.1 & \text{if } x = 0 \\ P(X \leq 1) - P(X \leq 0) = 0.3 - 0.1 = 0.2 & \text{if } x = 1 \\ P(X \leq 2) - P(X \leq 1) = 0.6 - 0.3 = 0.3 & \text{if } x = 2 \\ P(X \leq 3) - P(X \leq 2) = 1 - 0.6 = 0.4 & \text{if } x = 3 \end{cases}$$

or

x	0	1	2	3	Total
$f(x) = P(X = x)$	0.1	0.2	0.3	0.4	1

Since $P(X \in A) = \sum\limits_{x=0}^{3} P(X = x) = 1$, X is a discrete random variable.

In the next example we review four of the named distributions which were introduced in a previous probability course.

2.3.5 Example

Suppose a box containing a red balls and b black balls. For each of the following find the probability function of the random variable X and show that $\sum\limits_{x \in A} f(x) = 1$ where $A = \{x : f(x) > 0\}$ is the support set of X.

(a) $X =$ number of red balls among n balls drawn at random without replacement.

(b) $X =$ number of red balls among n balls drawn at random with replacement.

(c) $X =$ number of black balls selected before obtaining the *first* red ball if sampling is done at random with replacement.

(d) $X =$ number of black balls selected before obtaining the kth red ball if sampling is done at random with replacement.

Solution

(a) If n balls are selected at random without replacement from a box of a red balls and b black balls then the random variable $X =$ number of red balls has a Hypergeometric distribution with probability function

$$f(x) = P(X = x) = \frac{\binom{a}{x}\binom{b}{n-x}}{\binom{a+b}{n}} \quad \text{for } x = \max(0, n-b), \ldots, \min(a, n)$$

By the Hypergeometric identity 2.11.6

$$\begin{aligned}
\sum_{x \in A} f(x) &= \sum_{x=\max(0,n-b)}^{\min(a,n)} \frac{\binom{a}{x}\binom{b}{n-x}}{\binom{a+b}{n}} \\
&= \frac{1}{\binom{a+b}{n}} \sum_{x=0}^{\infty} \binom{a}{x}\binom{b}{n-x} \\
&= \frac{\binom{a+b}{n}}{\binom{a+b}{n}} \\
&= 1
\end{aligned}$$

(b) If n balls are selected at random with replacement from a box of a red balls and b black balls then we have a sequence of Bernoulli trials and the random variable X = number of red balls has a Binomial distribution with probability function

$$f(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \text{for } x = 0, 1, \ldots, n$$

where $p = \frac{a}{a+b}$. By the Binomial series 2.11.3(1)

$$\begin{aligned} \sum_{x \in A} f(x) &= \sum_{x \in 0}^{n} \binom{n}{x} p^x (1-p)^{n-x} \\ &= (p + 1 - p)^n \\ &= 1 \end{aligned}$$

(c) If sampling is done with replacement then we have a sequence of Bernoulli trials and the random variable X = number of black balls selected before obtaining the first red ball has a Geometric distribution with probability function

$$f(x) = P(X = x) = p(1-p)^x \quad \text{for } x = 0, 1, \ldots$$

By the Geometric series 2.11.1

$$\begin{aligned} \sum_{x \in A} f(x) &= \sum_{x \in 0}^{\infty} p(1-p)^x \\ &= \frac{p}{[1-(1-p)]} \\ &= 1 \end{aligned}$$

(d) If sampling is done with replacement then we have a sequence of Bernoulli trials and the random variable X = number of black balls selected before obtaining the kth red ball has a Negative Binomial distribution with probability function

$$f(x) = P(X = x) = \binom{x+k-1}{x} p^k (1-p)^x \quad \text{for } x = 0, 1, \ldots$$

Using the identity 2.11.4(1)

$$\binom{x+k-1}{x} = \binom{-k}{x} (-1)^x$$

the probability function can be written as

$$f(x) = P(X = x) = \binom{-k}{x} p^k (p-1)^x \quad \text{for } x = 0, 1, \ldots$$

By the Binomial series 2.11.3(2)

$$\begin{aligned}
\sum_{x\in A} f(x) &= \sum_{x=0}^{\infty} \binom{-k}{x} p^k (p-1)^x \\
&= p^k \sum_{x=0}^{\infty} \binom{-k}{x} (p-1)^x \\
&= p^k (1+p-1)^{-k} \\
&= 1
\end{aligned}$$

2.3.6 Example

If X is a random variable with probability function

$$f(x) = \frac{\mu^x e^{-\mu}}{x!} \quad \text{for } x = 0, 1, \ldots, \quad \mu > 0 \tag{2.1}$$

show that

$$\sum_{x=0}^{\infty} f(x) = 1$$

Solution

By the Exponential series 2.11.7

$$\begin{aligned}
\sum_{x=0}^{\infty} f(x) &= \sum_{x=0}^{\infty} \frac{\mu^x e^{-\mu}}{x!} \\
&= e^{-\mu} \sum_{x=0}^{\infty} \frac{\mu^x}{x!} \\
&= e^{-\mu} e^{\mu} \\
&= 1
\end{aligned}$$

The probability function (2.1) is called the Poisson probability function.

2.3.7 Exercise

If X is a random variable with probability function

$$f(x) = \frac{-(1-p)^x}{x \log p} \quad \text{for } x = 1, 2, \ldots, \quad 0 < p < 1$$

show that

$$\sum_{x=1}^{\infty} f(x) = 1$$

Hint: Use the Logarithmic series 2.11.8.

Important Note: A summary of the named distributions used in these Course Notes can be found in Chapter 11.

2.4 Continuous Random Variables

2.4.1 Definition - Continuous Random Variable

Suppose X is a random variable with cumulative distribution function F. If F is a continuous function for all $x \in \Re$ and F is differentiable except possibly at countably many points then X is called a *continuous random variable.*

Note: The definition (2.2.3) and properties (2.2.4) of the cumulative distribution function hold for the random variable X regardless of whether X is discrete or continuous.

2.4.2 Example

Suppose X is a random variable with cumulative distribution function

$$F(x) = P(X \leq x) = \begin{cases} 0 & x \leq -1 \\ -\frac{1}{2}(x+1)^2 + x + 1 & -1 < x \leq 0 \\ \frac{1}{2}(x-1)^2 + x & 0 < x < 1 \\ 1 & x \geq 1 \end{cases}$$

Show that X is a continuous random variable.

Solution

The cumulative distribution function F is a continuous function for all $x \in \Re$ since it is a piecewise function composed of continuous functions and

$$\lim_{x \to a} F(x) = F(a)$$

at the break points $a = -1, 0, 1$.

The function F is differentiable for all $x \neq -1, 0, 1$ since it is a piecewise function composed of differentiable functions. Since

$$\begin{aligned} \lim_{h \to 0^-} \frac{F(-1+h) - F(-1)}{h} &= 0 \neq \lim_{h \to 0^+} \frac{F(-1+h) - F(-1)}{h} = 1 \\ \lim_{h \to 0^-} \frac{F(0+h) - F(0)}{h} &= 0 = \lim_{h \to 0^+} \frac{F(0+h) - F(0)}{h} \\ \lim_{h \to 0^-} \frac{F(1+h) - F(1)}{h} &= 1 \neq \lim_{h \to 0^+} \frac{F(1+h) - F(1)}{h} = 0 \end{aligned}$$

F is differentiable at $x = 0$ butnot differentiable at $x = -1, 1$. The set $\{-1, 1\}$ is a countable set.

Since F is a continuous function for all $x \in \Re$ and F is differentiable except for countable many points, therefore X is a continuous random variable.

2.4.3 Definition - Probability Density Function

If X is a continuous random variable with cumulative distribution function $F(x)$ then the *probability density function* (p.d.f.) of X is $f(x) = F'(x)$ if F is differentiable at x. The set $A = \{x : f(x) > 0\}$ is called the support set of X.

Note: At the countably many points at which $F'(a)$ does not exist, $f(a)$ may be assigned any convenient value since the probabilities $P(X \leq x)$ will be unaffected by the choice. We usually choose $f(a) \geq 0$ and most often we choose $f(a) = 0$.

2.4.4 Properties - Probability Density Function

(1) $f(x) \geq 0$ for all $x \in \Re$

(2) $\int\limits_{-\infty}^{\infty} f(x)dx = \lim\limits_{x\to\infty} F(x) - \lim\limits_{x\to-\infty} F(x) = 1$

(3) $f(x) = \lim\limits_{h\to 0} \frac{F(x+h)-F(x)}{h} = \lim\limits_{h\to 0} \frac{P(x \leq X \leq x+h)}{h}$ if this limit exists

(4) $F(x) = \int\limits_{-\infty}^{x} f(t)dt, \quad x \in \Re$

(5) $P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a) = \int\limits_{a}^{b} f(x)dx$

(6) $P(X = b) = F(b) - \lim\limits_{a\to b^-} F(a) = F(b) - F(b) = 0$
(since F is continuous).

2.4.5 Example

Find and sketch the probability density function of the random variable X with the cumulative distribution function in Example 2.2.6.

Solution

By taking the derivative of $F(x)$ we obtain

$$F'(x) = \begin{cases} 0 & x < 0 \\ \frac{d}{dx}\left(\frac{2}{5}x^3\right) = \frac{6}{5}x^2 & 0 < x < 1 \\ \frac{d}{dx}\left(\frac{12x-3x^2-7}{5}\right) = \frac{1}{5}(12-6x) & 1 < x < 2 \\ 0 & x > 2 \end{cases}$$

We can assign any values to $f(0)$, $f(1)$, and $f(2)$. For convenience we choose
$f(0) = f(2) = 0$ and $f(1) = \frac{6}{5}$.

The probability density function is

$$f(x) = F'(x) = \begin{cases} \frac{6}{5}x^2 & 0 < x \leq 1 \\ \frac{1}{5}(12 - 6x) & 1 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

The graph of $f(x)$ is given in Figure 2.3.

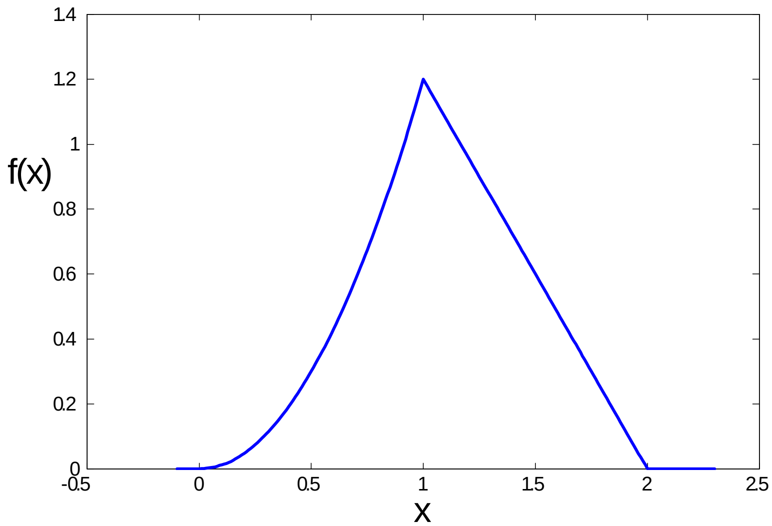


Figure 2.3: Graph of probability density function for Example 2.4.5

Note: See Table 2.1 in Section 2.7 for a summary of the differences between the properties of a discrete and continuous random variable.

2.4.6 Example

Suppose X is a random variable with cumulative distribution function

$$F(x) = \begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & a < x \leq b \\ 1 & x > b \end{cases}$$

where $b > a$.

(a) Sketch $F(x)$ the cumulative distribution function of X.

(b) Find $f(x)$ the probability density function of X and sketch it.

(c) Is it possible for $f(x)$ to take on values greater than one?

Solution

(a) See Figure 2.4 for a sketch of the cumulative distribution function.

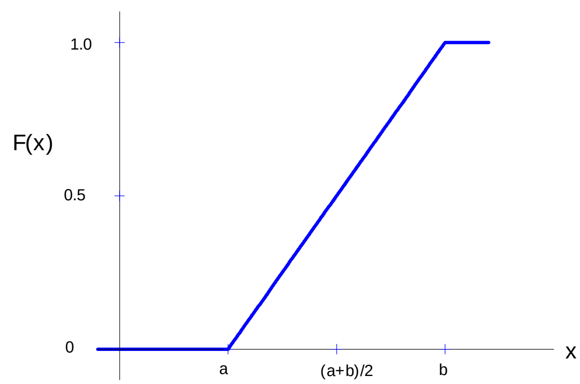


Figure 2.4: Graph of cumulative distribution function for Example 2.4.6

(b) By taking the derivative of $F(x)$ we obtain

$$F'(x) = \begin{cases} 0 & x < a \\ \frac{1}{b-a} & a < x < b \\ 0 & x > b \end{cases}$$

The derivative of $F(x)$ does not exist for $x = a$ or $x = b$. For convenience we define $f(a) = f(b) = \frac{1}{b-a}$ so that

$$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

See Figure 2.5 for a sketch of the probability density function. Note that we could define $f(a)$ and $f(b)$ to be any values and the cumulative distribution function would remain the same since $x = a$ or $x = b$ are countably many points.

The random variable X is said to have a Uniform(a, b) distribution. We write this as $X \sim$ Uniform(a, b).

(c) If $a = 0$ and $b = 0.5$ then $f(x) = 2$ for $0 \leq x \leq 0.5$. This example illustrates that the probability density function is not a probability and that the probability density function can take on values greater than one.

The important restriction for continuous random variables is

$$\int\limits_{-\infty}^{\infty} f(x)\, dx = 1$$

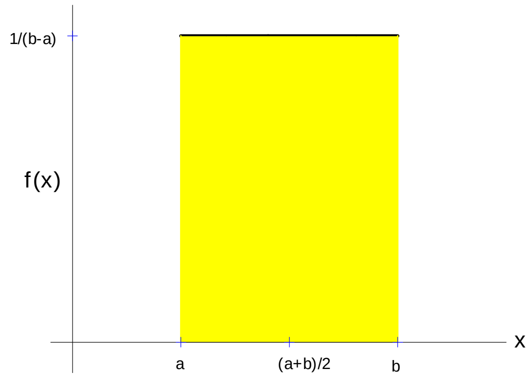


Figure 2.5: Graph of the probability density function of a Uniform(a, b) random variable

2.4.7 Example

Consider the function

$$f(x) = \frac{\theta}{x^{\theta+1}} \quad \text{for } x \geq 1$$

and 0 otherwise. For what values of θ is this function a probability density function?

Solution

Using the result (2.8) from Section 2.11

$$\begin{aligned}
\int_{-\infty}^{\infty} f(x)\,dx &= \int_{1}^{\infty} \frac{\theta}{x^{\theta+1}} dx \\
&= \lim_{b\to\infty} \int_{1}^{b} \theta x^{-\theta-1} dx \\
&= \lim_{b\to\infty} \left[-x^{-\theta}|_1^b \right] \\
&= 1 - \lim_{b\to\infty} \frac{1}{b^{\theta}} \\
&= 1 \quad \text{if } \theta > 0
\end{aligned}$$

Also $f(x) \geq 0$ if $\theta > 0$. Therefore $f(x)$ is a a probability density function for all $\theta > 0$.

X is said to have a Pareto$(1, \theta)$ distribution.

A useful function for evaluating integrals associated with several named random variables is the Gamma function.

2.4.8 Definition - Gamma Function

The *gamma function*, denoted by $\Gamma(\alpha)$ for all $\alpha > 0$, is given by

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy$$

2.4.9 Properties - Gamma Function

(1) $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1) \quad \alpha > 1$

(2) $\Gamma(n) = (n-1)! \quad n = 1, 2, \ldots$

(3) $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

2.4.10 Example

Suppose X is a random variable with probability density function

$$f(x) = \frac{x^{\alpha-1} e^{-x/\beta}}{\Gamma(\alpha)\beta^\alpha} \quad \text{for } x > 0, \quad \alpha > 0, \ \beta > 0$$

and 0 otherwise.

X is said to have a Gamma distribution with parameters α and β and we write $X \sim \text{Gamma}(\alpha, \beta)$.

(a) Verify that

$$\int_{-\infty}^{\infty} f(x) dx = 1$$

(b) What special probability density function is obtained for $\alpha = 1$?

(c) Graph the probability density functions for

(i) $\alpha = 1$, $\beta = 3$
(ii) $\alpha = 2$, $\beta = 1.5$
(iii) $\alpha = 5$, $\beta = 0.6$
(iv) $\alpha = 10$, $\beta = 0.3$

on the same graph.

Note: See Chapter 11 - Summary of Named Distributions. Note that the notation for parameters used for named distributions is not necessarily the same in all textbooks. This is especially true for distributions with two or more parameters.

Solution

(a)

$$\begin{aligned}
\int_{-\infty}^{\infty} f(x)\,dx &= \int_{0}^{\infty} \frac{x^{\alpha-1}e^{-x/\beta}}{\Gamma(\alpha)\beta^{\alpha}}dx \quad \text{let } y = x/\beta \\
&= \frac{1}{\Gamma(\alpha)\beta^{\alpha}} \int_{0}^{\infty} (y\beta)^{\alpha-1} e^{-y}\beta dy \\
&= \frac{1}{\Gamma(\alpha)} \int_{0}^{\infty} y^{\alpha-1}e^{-y}dy \\
&= \frac{\Gamma(\alpha)}{\Gamma(\alpha)} \\
&= 1
\end{aligned}$$

(b) If $\alpha = 1$ the probability density function is

$$f(x) = \frac{1}{\beta}e^{-x/\beta} \quad \text{for } x > 0; \quad \beta > 0$$

and 0 otherwise which is the Exponential(β) distribution.

(c) See Figure 2.6

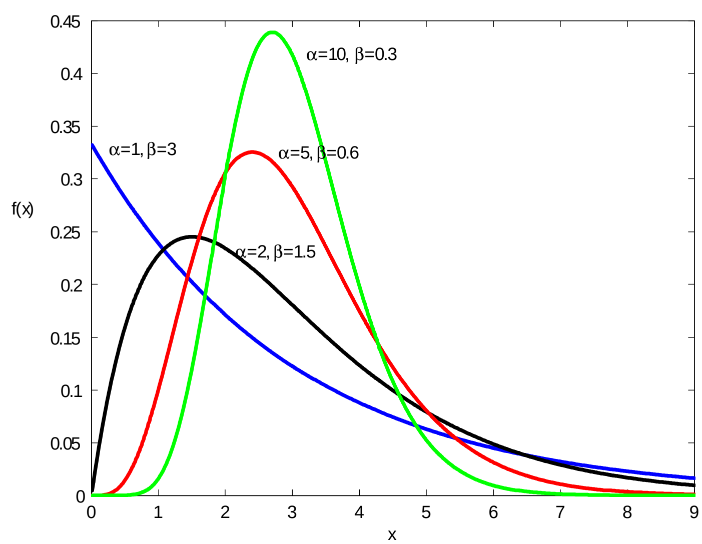


Figure 2.6: Gamma(α, β) probability density functions

2.4.11 Exercise

Suppose X is a random variable with probability density function

$$f(x) = \frac{\alpha}{\beta^\alpha} x^{\alpha-1} e^{-(x/\beta)^\alpha} \quad \text{for } x > 0,\ \alpha > 0,\ \beta > 0$$

and 0 otherwise.

X is said to have a Weibull distribution with parameters α and β and we write $X \sim \text{Weibull}(\alpha, \beta)$.

(a) Verify that

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

(b) What special probability density function is obtained for $\alpha = 1$?

(c) Graph the probability density functions for

(i) $\alpha = 1$, $\beta = 0.5$
(ii) $\alpha = 2$, $\beta = 0.5$
(iii) $\alpha = 2$, $\beta = 1$
(iv) $\alpha = 3$, $\beta = 1$

on the same graph.

2.4.12 Exercise

Suppose X is a random variable with probability density function

$$f(x) = \frac{\beta\alpha^\beta}{x^{\beta+1}} \quad \text{for } x > \alpha,\ \alpha > 0,\ \beta > 0$$

and 0 otherwise.

X is said to have a Pareto distribution with parameters α and β and we write $X \sim \text{Pareto}(\alpha, \beta)$.

(a) Verify that

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

(b) Graph the probability density functions for

(i) $\alpha = 1$, $\beta = 1$
(ii) $\alpha = 1$, $\beta = 2$
(iii) $\alpha = 0.5$, $\beta = 1$
(iv) $\alpha = 0.5$, $\beta = 2$

on the same graph.

2.5 Location and Scale Parameters

In Chapter 6 we will look at methods for constructing confidence intervals for an unknown parameter θ. If the parameter θ is either a location parameter or a scale parameter then a confidence interval is easier to construct.

2.5.1 Definition - Location Parameter

Suppose X is a continuous random variable with probability density function $f(x;\theta)$ where θ is a parameter of the distribution. Let $F_0(x) = F(x;\theta=0)$ and $f_0(x) = f(x;\theta=0)$. The parameter θ is called a *location parameter* of the distribution if

$$F(x;\theta) = F_0(x-\theta) \quad \text{for } \theta \in \Re$$

or equivalently

$$f(x;\theta) = f_0(x-\theta) \quad \text{for } \theta \in \Re$$

2.5.2 Definition - Scale Parameter

Suppose X is a continuous random variable with probability density function $f(x;\theta)$ where θ is a parameter of the distribution. Let $F_1(x) = F(x;\theta=1)$ and $f_1(x) = f(x;\theta=1)$. The parameter θ is called a *scale parameter* of the distribution if

$$F(x;\theta) = F_1\left(\frac{x}{\theta}\right) \quad \text{for } \theta > 0$$

or equivalently

$$f(x;\theta) = \frac{1}{\theta} f_1(\frac{x}{\theta}) \quad \text{for } \theta > 0$$

2.5.3 Example

Suppose X is a continuous random variable with probability density function

$$f(x) = \frac{1}{\beta} e^{-(x-\alpha)/\beta} \quad \text{for } x \geq \alpha,\ \alpha \in \Re,\ \beta > 0$$

and 0 otherwise.

X is said to have a Two Parameter Exponential distribution and we write $X \sim$ Double Exponential(α, β).

(a) If $X \sim$ Two Parameter Exponential$(\theta, 1)$ show that θ is a location parameter for this distribution. Sketch the probability density function for $\theta = -1, 0, 1$ on the same graph.

(b) If $X \sim$ Two Parameter Exponential$(0, \theta)$ show that θ is a scale parameter for this distribution. Sketch the probability density function for $\theta = 0.5, 0, 2$ on the same graph.

Solution

(a) For $X \sim$ Two Parameter Exponential$(\theta, 1)$ the probability density function is

$$f(x;\theta) = e^{-(x-\theta)} \quad \text{for } x \geq \theta,\ \theta \in \Re$$

and 0 otherwise.
Let

$$f_0(x) = f(x;\theta = 0) = e^{-x} \quad \text{for } x > 0$$

and 0 otherwise. Then

$$\begin{aligned} f(x;\theta) &= e^{-(x-\theta)} \\ &= f_0(x-\theta) \quad \text{for all } \theta \in \Re \end{aligned}$$

and therefore θ is a location parameter of this distribution.
See Figure 2.7 for a sketch of the probability density function for $\theta = -1, 0, 1$.

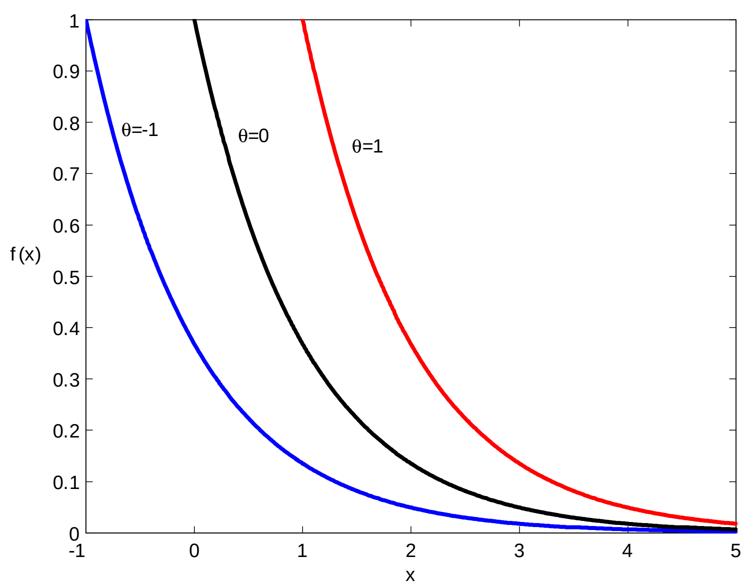


Figure 2.7: Exponential$(\theta, 1)$ probability density function for $\theta = -1, 0, 1$

(b) For $X \sim$ Two Parameter Exponential$(0, \theta)$ the probability density function is

$$f(x;\theta) = \frac{1}{\theta} e^{-x/\theta} \quad \text{for } x > 0,\ \theta > 0$$

and 0 otherwise which is the Exponential(θ) probability density function.

Let

$$f_1(x) = f(x; \theta = 1) = e^{-x} \quad \text{for } x > 0$$

and 0 otherwise. Then

$$f(x; \theta) = \frac{1}{\theta} e^{-x/\theta} = \frac{1}{\theta} f_1\left(\frac{x}{\theta}\right) \quad \text{for all } \theta > 0$$

and therefore θ is a scale parameter of this distribution.

See Figure 2.8 for a sketch of the probability density function for $\theta = 0.5, 1, 2$.

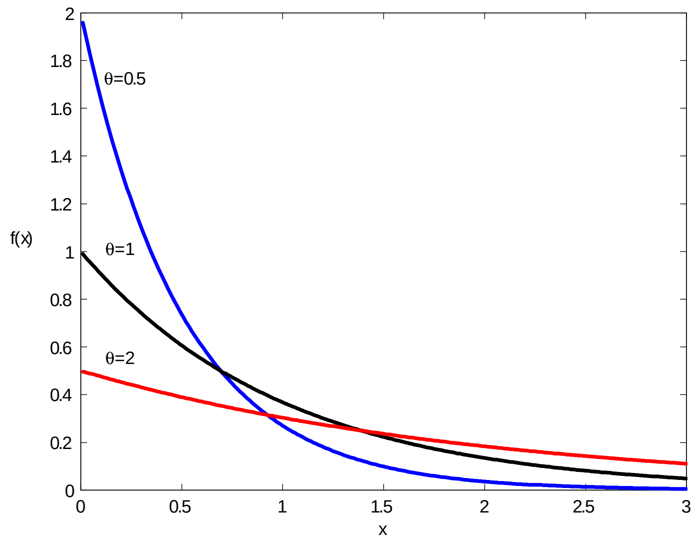


Figure 2.8: Exponential(θ) probability density function for $\theta = 0.5, 1, 2$

2.5.4 Exercise

Suppose X is a continuous random variable with probability density function

$$f(x) = \frac{1}{\beta\pi\{1 + [(x-\mu)/\beta]^2\}} \quad \text{for } x \in \Re,\ \mu \in \Re,\ \beta > 0$$

and 0 otherwise.

X is said to have a two parameter Cauchy distribution and we write
$X \sim \text{Cauchy}(\mu, \beta)$.

(a) If $X \sim \text{Cauchy}(\theta, 1)$ then show that θ is a location parameter for the distribution. Graph the Cauchy$(\theta, 1)$ probability density function for $\theta = -1$, 0 and 1 on the same graph.

(b) If $X \sim \text{Cauchy}(0, \theta)$ then show that θ is a scale parameter for the distribution. Graph the Cauchy$(0, \theta)$ probability density function for $\theta = 0.5$, 1 and 2 on the same graph.

2.6 Functions of a Random Variable

Suppose X is a continuous random variable with probability density function f and cumulative distribution function F and we wish to find the probability density function of the random variable $Y = h(X)$ where h is a real-valued function. In this section we look at techniques for determining the distribution of Y.

2.6.1 Cumulative Distribution Function Technique

A useful technique for determining the distribution of a function of a random variable $Y = h(X)$ is the *cumulative distribution function technique*. This technique involves obtaining an expression for $G(y) = P(Y \le y)$, the cumulative distribution function of Y, in terms of F, the cumulative distribution function of X. The corresponding probability density function g of Y is found by differentiating G. Care must be taken to determine the support set of the random variable Y.

2.6.2 Example

If $Z \sim \mathrm{N}(0,1)$ find the probability density function of $Y = Z^2$. What type of random variable is Y?

Solution

If $Z \sim \mathrm{N}(0,1)$ then the probability density function of Z is

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \quad \text{for } z \in \Re$$

Let $G(y) = P(Y \le y)$ be the cumulative distribution function of $Y = Z^2$. Since the support set of the random variable Z is $\Re$ and $Y = Z^2$ then the support set of Y is $B = \{y : y > 0\}$.

For $y \in B$

$$\begin{aligned} G(y) &= P(Y \le y) \\ &= P\left(Z^2 \le y\right) \\ &= P\left(-\sqrt{y} \le Z \le \sqrt{y}\right) \\ &= \int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz \\ &= 2\int_{0}^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz \quad \text{since } f(z) \text{ is an even function} \end{aligned}$$

For $y \in B$ the probability density function of Y is

$$\begin{aligned} g(y) &= \frac{d}{dy}G(y) = \frac{d}{dy}\left[\int_0^{\sqrt{y}} \frac{2}{\sqrt{2\pi}} e^{-z^2/2} dz\right] = \frac{2}{\sqrt{2\pi}} e^{-\left(\sqrt{y}\right)^2/2} \frac{d}{dy}\left(\sqrt{y}\right) \\ &= \frac{2}{\sqrt{2\pi}} e^{-\left(\sqrt{y}\right)^2/2} \frac{1}{2\sqrt{y}} \qquad (2.2) \\ &= \frac{1}{\sqrt{2\pi y}} e^{-y/2} \qquad (2.3) \end{aligned}$$

by 2.11.10.

We recognize (2.3) as the probability density function of a Chi-squared(1) random variable so $Y = Z^2 \sim \chi^2(1)$.

The above solution provides a proof of the following theorem.

2.6.3 Theorem

If $Z \sim \mathrm{N}(0,1)$ then $Z^2 \sim \chi^2(1)$.

2.6.4 Example

Suppose $X \sim$ Exponential(θ). The cumulative distribution function for X is

$$F(x) = P(X \le x) = \begin{cases} 0 & x \le 0 \\ 1 - e^{-x/\theta} & x > 0 \end{cases}$$

Determine the distribution of the random variable $Y = F(X) = 1 - e^{-X/\theta}$.

Solution

Since $Y = 1 - e^{-X/\theta}$, $X = -\theta \log(1 - Y) = F^{-1}(Y)$ for $X > 0$ and $0 < Y < 1$.

For $0 < y < 1$ the cumulative distribution function of Y is

$$\begin{aligned} G(y) &= P(Y \le y) = P\left(1 - e^{-X/\theta} \le y\right) \\ &= P(X \le -\theta \log(1-y)) \\ &= F(-\theta \log(1-y)) \\ &= 1 - e^{\theta \log(1-y)/\theta} \\ &= 1 - (1-y) \\ &= y \end{aligned}$$

If $U \sim \text{Uniform}(0,1)$ then the cumulative distribution function of U is

$$P(U \leq u) = \begin{cases} 0 & u \leq 0 \\ u & 0 < u < 1 \\ 1 & u \geq 1 \end{cases}$$

Therefore the cumulative distribution function of $Y = F(X) = 1 - e^{-X/\theta}$ is $\text{Uniform}(0,1)$.

This is an example of a result which holds more generally as summarized in the following theorem.

2.6.5 Theorem - Probability Integral Transformation

If X is a continuous random variable with cumulative distribution function F then the random variable

$$Y = F(X) = \int_{-\infty}^{X} f(t)\, dt \tag{2.4}$$

has a $\text{Uniform}(0,1)$ distribution.

Proof

Suppose the continuous random variable X has support set $A = \{x : f(x) > 0\}$. For all $x \in A$, F is an increasing function since F is a cumulative distribution function. Therefore for all $x \in A$ the function F has an inverse function F^{-1}.

For $0 < y < 1$, the cumulative distribution function of $Y = F(X)$ is

$$\begin{aligned} G(y) &= P(Y \leq y) = P(F(X) \leq y) \\ &= P\left(X \leq F^{-1}(y)\right) \\ &= F\left(F^{-1}(y)\right) \\ &= y \end{aligned}$$

which is the cumulative distribution function of a $\text{Uniform}(0,1)$ random variable. Therefore $Y = F(X) \sim \text{Uniform}(0,1)$ as required.

Note: Because of the form of the function (transformation) $Y = F(X)$ in (2.4), this transformation is called the *probability integral transformation*. This result holds for any cumulative distribution function F corresponding to a continuous random variable.

2.6.6 Theorem

Suppose F is a cumulative distribution function for a continuous random variable. If $U \sim$ Uniform$(0,1)$ then the random variable $X = F^{-1}(U)$ also has cumulative distribution function F.

Proof

Suppose that the support set of the random variable $X = F^{-1}(U)$ is A. For $x \in A$, the cumulative distribution function of $X = F^{-1}(U)$ is

$$\begin{aligned} P(X \leq x) &= P\left(F^{-1}(U) \leq x\right) \\ &= P(U \leq F(x)) \\ &= F(x) \end{aligned}$$

since $P(U \leq u) = u$ for $0 < u < 1$ if $U \sim$ Uniform$(0,1)$. Therefore $X = F^{-1}(U)$ has cumulative distribution function F.

Note: The result of the previous theorem is important because it provides a method for generating observations from a continuous distribution. Let u be an observation generated from a Uniform$(0,1)$ distribution using a random number generator. Then by Theorem 2.6.6, $x = F^{-1}(u)$ is an observation from the distribution with cumulative distribution function F.

2.6.7 Example

Explain how Theorem 2.6.6 can be used to generate observations from a Weibull$(\theta,1)$ distribution.

Solution

If X has a Weibull$(\theta,1)$ distribution then the cumulative distribution function is

$$\begin{aligned} F(x;\theta) &= \int_0^x \theta y^{\theta-1} e^{-y^\theta} dy \\ &= 1 - e^{-x^\theta} \quad \text{for } x > 0 \end{aligned}$$

The inverse cumulative distribution function is

$$F^{-1}(u) = [-\log(1-u)]^{1/\theta} \quad \text{for } 0 < u < 1$$

If u is an observation from the Uniform$(0,1)$ distribution then $x = [-\log(1-u)]^{1/\theta}$ is an observation from the Weibull$(\theta,1)$ distribution by Theorem 2.6.6.

If we wish to find the distribution of the random variable $Y = h(X)$ and h is a one-to-one real-valued function then the following theorem can be used.

2.6.8 Theorem - One-to-One Transformation of a Random Variable

Suppose X is a continuous random variable with probability density function f and support set $A = \{x : f(x) > 0\}$. Let $Y = h(X)$ where h is a real-valued function. Let $B = \{y : g(y) > 0\}$ be the support set of the random variable Y. If h is a one-to-one function from A to B and $\frac{d}{dx}h(x)$ is continuous for $x \in A$, then the probability density function of Y is

$$g(y) = f(h^{-1}(y)) \left| \frac{d}{dy} h^{-1}(y) \right| \quad \text{for } y \in B$$

Proof

We prove this theorem using the cumulative distribution function technique.

(1) Suppose h is an increasing function and $\frac{d}{dx}h(x)$ is continuous for $x \in A$. Then $h^{-1}(y)$ is also an increasing function and $\frac{d}{dy}h^{-1}(y) > 0$ for $y \in B$. The cumulative distribution function of $Y = h(X)$ is

$$\begin{aligned} G(y) &= P(Y \le y) = P(h(X) \le y) \\ &= P\left(X \le h^{-1}(y)\right) \quad \text{since } h \text{ is an increasing function} \\ &= F\left(h^{-1}(y)\right) \end{aligned}$$

Therefore

$$\begin{aligned} g(y) &= \frac{d}{dy} G(y) = \frac{d}{dy} F\left(h^{-1}(y)\right) \\ &= F'\left(h^{-1}(y)\right) \frac{d}{dy} h^{-1}(y) \quad \text{by the Chain Rule} \\ &= f\left(h^{-1}(y)\right) \left| \frac{d}{dy} h^{-1}(y) \right| \quad y \in B \text{ since } \frac{d}{dy} h^{-1}(y) > 0 \end{aligned}$$

(2) Suppose h is a decreasing function and $\frac{d}{dx}h(x)$ is continuous for $x \in A$. Then $h^{-1}(y)$ is also a decreasing function and $\frac{d}{dy}h^{-1}(y) < 0$ for $y \in B$. The cumulative distribution function of $Y = h(X)$ is

$$\begin{aligned} G(y) &= P(Y \le y) = P(h(X) \le y) \\ &= P\left(X \ge h^{-1}(y)\right) \quad \text{since } h \text{ is a decreasing function} \\ &= 1 - F\left(h^{-1}(y)\right) \end{aligned}$$

Therefore

$$\begin{aligned} g(y) &= \frac{d}{dy} G(y) = \frac{d}{dy} \left[1 - F\left(h^{-1}(y)\right)\right] \\ &= -F'\left(h^{-1}(y)\right) \frac{d}{dy} h^{-1}(y) \quad \text{by the Chain Rule} \\ &= f\left(h^{-1}(y)\right) \left| \frac{d}{dy} h^{-1}(y) \right| \quad y \in B \text{ since } \frac{d}{dy} h^{-1}(y) < 0 \end{aligned}$$

These two cases give the desired result.

2.6.9 Example

The following two results were used extensively in your previous probability and statistics courses.

(a) If $Z \sim \mathrm{N}(0,1)$ then $Y = \mu + \sigma Z \sim \mathrm{N}\big(\mu, \sigma^2\big)$.

(b) If $X \sim \mathrm{N}(\mu, \sigma^2)$ then $Z = \frac{X-\mu}{\sigma} \sim \mathrm{N}(0,1)$.

Prove these results using Theorem 2.6.8.

Solution

(a) If $Z \sim \mathrm{N}(0,1)$ then the probability density function of Z is

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \quad \text{for } z \in \Re$$

$Y = \mu + \sigma Z = h(Z)$ is an increasing function with inverse function $Z = h^{-1}(Y) = \frac{Y-\mu}{\sigma}$. Since the support set of Z is $A = \Re$, the support set of Y is $B = \Re$.

Since

$$\frac{d}{dy} h^{-1}(y) = \frac{d}{dy}\left(\frac{y-\mu}{\sigma}\right) = \frac{1}{\sigma}$$

then by Theorem 2.6.8 the probability density function of Y is

$$\begin{aligned} g(y) &= f(h^{-1}(y)) \left| \frac{d}{dy} h^{-1}(y) \right| \\ &= \frac{1}{\sqrt{2\pi}} e^{-\left(\frac{y-\mu}{\sigma}\right)^2/2} \frac{1}{\sigma} \quad \text{for } y \in \Re \end{aligned}$$

which is the probability density function of an $\mathrm{N}\big(\mu, \sigma^2\big)$ random variable.

Therefore if $Z \sim \mathrm{N}(0,1)$ then $Y = \mu + \sigma Z \sim \mathrm{N}\big(\mu, \sigma^2\big)$.

(b) If $X \sim \mathrm{N}(\mu, \sigma^2)$ then the probability density function of X is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\left(\frac{x-\mu}{\sigma}\right)^2/2} \quad \text{for } x \in \Re$$

$Z = \frac{X-\mu}{\sigma} = h(X)$ is an increasing function with inverse function $X = h^{-1}(Z) = \mu + \sigma Z$. Since the support set of X is $A = \Re$, the support set of Z is $B = \Re$.

Since

$$\frac{d}{dz} h^{-1}(z) = \frac{d}{dy}(\mu + \rho z) = \sigma$$

then by Theorem 2.6.8 the probability density function of Z is

$$\begin{aligned} g(z) &= f(h^{-1}(z)) \left| \frac{d}{dy} h^{-1}(z) \right| \\ &= \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \quad \text{for } z \in \Re \end{aligned}$$

which is the probability density function of an $\mathrm{N}(0,1)$ random variable.

Therefore if $X \sim \mathrm{N}(\mu, \sigma^2)$ then $Z = \frac{Y-\mu}{\sigma} \sim \mathrm{N}(0,1)$.

2.6.10 Example

Use Theorem 2.6.8 to prove the following relationship between the Pareto distribution and the Exponential distribution.

$$\text{If } X \sim Pareto\,(1,\theta) \text{ then } Y = \log(X) \sim \text{Exponential}\left(\frac{1}{\theta}\right)$$

Solution

If $X \sim \text{Pareto}(1,\theta)$ then the probability density function of X is

$$f(x) = \frac{\theta}{x^{\theta+1}} \quad \text{for } x \geq 1,\ \theta > 0$$

$Y = \log(X) = h(X)$ is an increasing function with inverse function $X = e^Y = h^{-1}(Y)$. Since the support set of X is $A = \{x : x \geq 0\}$, the support set of Y is $B = \{y : y > 0\}$.

Since

$$\frac{d}{dy}h^{-1}(y) = \frac{d}{dy}e^y = e^y$$

then by Theorem 2.6.8 the probability density function of Y is

$$\begin{aligned} g(y) &= f(h^{-1}(y))\left|\frac{d}{dy}h^{-1}(y)\right| \\ &= \frac{\theta}{(e^y)^{\theta+1}}e^y = \theta e^{-\theta y} \quad \text{for } y > 0 \end{aligned}$$

which is the probability density function of an Exponential$\left(\frac{1}{\theta}\right)$ random variable as required.

2.6.11 Exercise

Use Theorem 2.6.8 to prove the following relationship between the Exponential distribution and the Weibull distribution.

$$\text{If } X \sim Exponential(1) \text{ then } Y = \beta X^{1/\alpha} \sim \text{Weibull}(\alpha,\beta) \quad \text{for } y > 0,\ \alpha > 0,\ \beta > 0$$

2.6.12 Exercise

Suppose X is a random variable with probability density function

$$f(x) = \theta x^{\theta-1} \ \text{ for } 0 < x < 1,\ \theta > 0$$

and 0 otherwise.

Use Theorem 2.6.8 to prove that $Y = -\log X \sim \text{Exponential}\left(\frac{1}{\theta}\right)$.

2.7 Expectation

In this section we define the expectation operator E which maps random variables to real numbers. These numbers have an interpretation in terms of long run averages for repeated independent trails of an experiment associated with the random variable. Much of this section is a review of material covered in a previous probability course.

2.7.1 Definition - Expectation

Suppose $h(x)$ is a real-valued function.

If X is a discrete random variable with probability function $f(x)$ and support set A then the *expectation* of the random variable $h(X)$ is defined by

$$E[h(X)] = \sum_{x \in A} h(x) f(x)$$

provided the sum converges absolutely, that is, provided

$$E(|h(X)|) = \sum_{x \in A} |h(x)| f(x) < \infty$$

If X is a continuous random variable with probability density function $f(x)$ then the *expectation* of the random variable $h(X)$ is defined by

$$E[h(X)] = \int_{-\infty}^{\infty} h(x) f(x) dx$$

provided the integral converges absolutely, that is, provided

$$E(|h(X)|) = \int_{-\infty}^{\infty} |h(x)| f(x) dx < \infty$$

If $E(|h(X)|) = \infty$ then we say that $E[h(X)]$ does not exist.

$E[h(X)]$ is also called the *expected value* of the random variable $h(X)$.

2.7.2 Example

Find $E(X)$ if $X \sim$ Geometric(p).

Solution

If $X \sim$ Geometric(p) then

$$f(x) = pq^x \quad \text{for } x = 0, 1, \ldots, \quad q = 1 - p, \quad 0 < p < 1$$

and

$$
\begin{aligned}
E(X) &= \sum_{x\in A} xf(x) = \sum_{x=0}^{\infty} xpq^x = \sum_{x=1}^{\infty} xpq^x \\
&= pq\sum_{x=1}^{\infty} xq^{x-1} \quad \text{which converges if } 0<q<1 \\
&= \frac{pq}{(1-q)^2} \quad \text{by 2.11.2(2)} \\
&= \frac{q}{p} \quad \text{if } 0<q<1
\end{aligned}
$$

2.7.3 Example

Suppose $X \sim \text{Pareto}(1,\theta)$ with probability density function

$$f(x) = \frac{\theta}{x^{\theta+1}} \quad \text{for } x \geq 1,\ \theta > 0$$

and 0 otherwise. Find $E(X)$. For what values of θ does $E(X)$ exist?

Solution

$$
\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} xf(x)dx = \int_{1}^{\infty} x\frac{\theta}{x^{\theta+1}}dx \\
&= \theta\int_{1}^{\infty} \frac{1}{x^{\theta}}dx \quad \text{which converges for } \theta>1 \text{ by 2.8} \\
&= \theta\lim_{b\to\infty}\int_{1}^{b} x^{-\theta}dx = \frac{\theta}{1-\theta}\lim_{b\to\infty} x^{-\theta+1}|_1^b = \frac{\theta}{\theta-1}\left(1-\frac{1}{b^{\theta-1}}\right) \\
&= \frac{\theta}{\theta-1} \quad \text{for } \theta>1
\end{aligned}
$$

Therefore $E(X) = \frac{\theta}{\theta-1}$ and the mean exists only for $\theta > 1$.

2.7.4 Exercise

Suppose X is a nonnegative continuous random variable with cumulative distribution function $F(x)$ and $E(X) < \infty$. Show that

$$E(X) = \int_{0}^{\infty} [1 - F(x)]\, dx$$

Hint: Use integration by parts with $u = [1 - F(x)]$.

2.7.5 Theorem - Expectation is a Linear Operator

Suppose X is a random variable with probability (density) function $f(x)$, a and b are real constants, and $g(x)$ and $h(x)$ are real-valued functions. Then

$$\begin{aligned} E(aX+b) &= aE(X)+b \\ E[ag(X)+bh(X)] &= aE[g(X)]+bE[h(X)] \end{aligned}$$

Proof (Continuous Case)

$$\begin{aligned} E(aX+b) &= \int_{-\infty}^{\infty}(ax+b)f(x)dx \\ &= a\int_{-\infty}^{\infty}xf(x)dx+b\int_{-\infty}^{\infty}f(x)dx \quad \text{by properties of integrals} \\ &= aE(X)+b(1) \quad \text{by Definition 2.7.1 and Property 2.4.4} \\ &= aE(X)+b \end{aligned}$$

$$\begin{aligned} E[ag(X)+bh(X)] &= \int_{-\infty}^{\infty}[ag(x)+bh(x)]f(x)dx \\ &= a\int_{-\infty}^{\infty}g(x)f(x)dx+b\int_{-\infty}^{\infty}h(x)f(x)dx \quad \text{by properties of integrals} \\ &= aE[g(X)]+bE[h(X)] \quad \text{by Definition 2.7.1} \end{aligned}$$

as required.

The following named expectations are used frequently.

2.7.6 Special Expectations

(1) The *mean* of a random variable

$$E(X)=\mu$$

(2) The *kth moment (about the origin)* of a random variable

$$E(X^k)$$

(3) The *kth moment about the mean* of a random variable

$$E\left[(X-\mu)^k\right]$$

(4) The *kth factorial moment* of a random variable

$$E\left(X^{(k)}\right) = E\left[X(X-1)\cdots(X-k+1)\right]$$

(5) The *variance* of a random variable

$$Var(X) = E[(X-\mu)^2] = \sigma^2 \qquad \text{where} \quad \mu = E(X)$$

2.7.7 Theorem - Properties of Variance

$$\begin{aligned}\sigma^2 &= Var(X)\\ &= E(X^2) - \mu^2\\ &= E[X(X-1)] + \mu - \mu^2\end{aligned}$$

$$Var(aX+b) = a^2 Var(X)$$

and

$$E(X^2) = \sigma^2 + \mu^2$$

Proof (Continuous Case)

$$\begin{aligned}Var(X) &= E[(X-\mu)^2] = E\left(X^2 - 2\mu X + \mu^2\right)\\ &= E\left(X^2\right) - 2\mu E\left(X\right) + \mu^2 \quad \text{by Theorem 2.7.5}\\ &= E\left(X^2\right) - 2\mu^2 + \mu^2\\ &= E\left(X^2\right) - \mu^2\end{aligned}$$

Also

$$\begin{aligned}Var(X) &= E\left(X^2\right) - \mu^2 = E\left[X\left(X-1\right) + X\right] - \mu^2\\ &= E\left[X\left(X-1\right)\right] + E\left(X\right) - \mu^2 \quad \text{by Theorem 2.7.5}\\ &= E\left[X\left(X-1\right)\right] + \mu - \mu^2\end{aligned}$$

$$\begin{aligned}Var\left(aX+b\right) &= E\left\{\left[aX + b - (a\mu + b)\right]^2\right\}\\ &= E\left[\left(aX - a\mu\right)^2\right] = E\left[a^2\left(X-\mu\right)^2\right]\\ &= a^2 E[(X-\mu)^2] \quad \text{by Theorem 2.7.5}\\ &= a^2 Var(X) \quad \text{by definition}\end{aligned}$$

Rearranging $\sigma^2 = E(X^2) - \mu^2$ gives

$$E(X^2) = \sigma^2 + \mu^2$$

2.7.8 Example

If $X \sim \text{Binomial}(n, p)$ then show

$$E(X^{(k)}) = n^{(k)} p^k \quad \text{for } k = 1, 2, \ldots$$

and thus find $E(X)$ and $Var(X)$.

Solution

$$\begin{aligned}
E(X^{(k)}) &= \sum_{x=k}^{n} x^{(k)} \binom{n}{x} p^x (1-p)^{n-x} \\
&= \sum_{x=k}^{n} n^{(k)} \binom{n-k}{x-k} p^x (1-p)^{n-x} \quad \text{by 2.11.4(1)} \\
&= n^{(k)} \sum_{y=0}^{n-k} \binom{n-k}{y} p^{y+k} (1-p)^{n-y-k} \qquad y = x - k \\
&= n^{(k)} p^k \sum_{y=0}^{n-k} \binom{n-k}{y} p^y (1-p)^{(n-k)-y} \\
&= n^{(k)} p^k (p + 1 - p)^{n-k} \quad \text{by 2.11.3(1)} \\
&= n^{(k)} p^k \quad \text{for } k = 1, 2, \ldots
\end{aligned}$$

For $k = 1$ we obtain

$$\begin{aligned}
E\left(X^{(1)}\right) &= E(X) \\
&= n^{(1)} p^1 \\
&= np
\end{aligned}$$

For $k = 2$ we obtain

$$\begin{aligned}
E\left(X^{(2)}\right) &= E[X(X-1)] \\
&= n^{(2)} p^2 \\
&= n(n-1) p^2
\end{aligned}$$

Therefore

$$\begin{aligned}
Var(X) &= E[X(X-1)] + \mu - \mu^2 \\
&= n(n-1) p^2 + np - n^2 p^2 \\
&= -np^2 + np \\
&= np(1-p)
\end{aligned}$$

2.7.9 Exercise

Show the following:

(a) If $X \sim$ Poisson(θ) then $E(X^{(k)}) = \theta^k$ for $k = 1, 2, \ldots$

(b) If $X \sim$ Negative Binomial(k, p) then $E(X^{(j)}) = (-k)^{(j)} \left(\frac{p-1}{p}\right)^j$ for $j = 1, 2, \ldots$

(c) If $X \sim$ Gamma(α, β) then $E(X^p) = \beta^p \Gamma(\alpha + p) / \Gamma(\alpha)$ for $p > -\alpha$.

(d) If $X \sim$ Weibull(α, β) then $E(X^k) = \beta^k \Gamma\left(\frac{k}{\alpha} + 1\right)$ for $k = 1, 2, \ldots$

In each case find $E(X)$ and $Var(X)$.

Table 2.1 summarizes the differences between the properties of a discrete and continuous random variable.

Property	Discrete Random Variable	Continuous Random Variable
c.d.f.	$F(x) = P(X \leq x) = \sum\limits_{t \leq x} P(X = t)$ F is a right continuous function for all $x \in \Re$	$F(x) = P(X \leq x) = \int\limits_{-\infty}^{x} f(t)\,dt$ F is a continuous function for all $x \in \Re$
p.f./p.d.f.	$f(x) = P(X = x)$	$f(x) = F'(x) \neq P(X = x) = 0$
Probability of an event	$P(X \in E) = \sum\limits_{x \in E} P(X = x)$ $= \sum\limits_{x \in E} f(x)$	$P(a < X \leq b) = F(b) - F(a)$ $= \int\limits_{a}^{b} f(x)\,dx$
Total Probability	$\sum\limits_{x \in A} P(X = x) = \sum\limits_{x \in A} f(x) = 1$ where A = support set of X	$\int\limits_{-\infty}^{\infty} f(x)\,dx = 1$
Expectation	$E[g(X)] = \sum\limits_{x \in A} g(x) f(x)$ where A = support set of X	$E[g(X)] = \int\limits_{-\infty}^{\infty} g(x) f(x)\,dx$

Table 2.1 Properties of discrete versus continuous random variables

2.8 Inequalities

In Chapter 5 we consider limiting distributions of a sequence of random variables. The following inequalities which involve the moments of a distribution are useful for proving limit theorems.

2.8.1 Markov's Inequality

$$P(|X| \geq c) \leq \frac{E(|X|^k)}{c^k} \quad \text{for all } k,\ c > 0$$

Proof (Continuous Case)

Suppose X is a continuous random variable with probability density function $f(x)$. Let

$$A = \left\{ x : \left|\frac{x}{c}\right|^k \geq 1 \right\} = \{x : |x| \geq c\} \text{ since } c > 0$$

Then

$$\begin{aligned}
\frac{E\left(|X|^k\right)}{c^k} &= E\left(\left|\frac{X}{c}\right|^k\right) = \int_{-\infty}^{\infty} \left|\frac{x}{c}\right|^k f(x)\,dx \\
&= \int_A \left|\frac{x}{c}\right|^k f(x)\,dx + \int_{\bar{A}} \left|\frac{x}{c}\right|^k f(x)\,dx \\
&\geq \int_A \left|\frac{x}{c}\right|^k f(x)\,dx \quad \text{since} \int_{\bar{A}} \left|\frac{x}{c}\right|^k f(x)\,dx \geq 0 \\
&\geq \int_A f(x)\,dx \quad \text{since } \left|\frac{x}{c}\right|^k \geq 1 \text{ for } x \in A \\
&= P(|X| \geq c)
\end{aligned}$$

as required. (The proof of the discrete case follows by replacing integrals with sums.)

2.8.2 Chebyshev's Inequality

Suppose X is a random variable with finite mean μ and finite variance σ^2. Then for any $k > 0$

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

2.8.3 Exercise

Use Markov's Inequality to prove Chebyshev's Inequality.

2.9 Variance Stabilizing Transformation

In Chapter 6 we look at methods for constructing a confidence interval for an unknown parameter θ based on data X. To do this it is often useful to find a transformation, $g(X)$, of the data X whose variance is approximately constant with respect to θ.

Suppose X is a random variable with finite mean $E(X) = \theta$. Suppose also that X has finite variance $Var(X) = \sigma^2(\theta)$ and standard deviation $\sqrt{Var(X)} = \sigma(\theta)$ also depending on θ. Let $Y = g(X)$ where g is a differentiable function. By the linear approximation

$$Y = g(X) \approx g(\theta) + g'(\theta)(X - \theta)$$

Therefore

$$E(Y) \approx E\left[g(\theta) + g'(\theta)(X - \theta)\right] = g(\theta)$$

since

$$E\left[g'(\theta)(X - \theta)\right] = g'(\theta)\, E\left[(X - \theta)\right] = 0$$

Also

$$Var(Y) \approx Var\left[g'(\theta)(X - \theta)\right] = \left[g'(\theta)\right]^2 Var(X) = \left[g'(\theta)\,\sigma(\theta)\right]^2 \tag{2.5}$$

If we want $Var(Y) \approx$ constant with respect to θ then we should choose g such that

$$\left[g'(\theta)\right]^2 Var(X) = \left[g'(\theta)\,\sigma(\theta)\right]^2 = \text{ constant}$$

In other words we need to solve the differential equation

$$\frac{dg}{d\theta} = \frac{k}{\sigma(\theta)}$$

where k is a conveniently chosen constant.

2.9.1 Example

If $X \sim$ Poisson(θ) then show that the random variable $Y = g(X) = \sqrt{X}$ has approximately constant variance.

Solution

If $X \sim$ Poisson(θ) then $\sqrt{Var(X)} = \sigma(\theta) = \sqrt{\theta}$. For $g(X) = \sqrt{X}$, $g'(X) = \frac{1}{2}X^{-1/2}$. Therefore by (2.5), the variance of $Y = g(X) = \sqrt{X}$ is approximately

$$\left[g'(\theta)\,\sigma(\theta)\right]^2 = \left[\frac{1}{2}\theta^{-1/2}\sqrt{\theta}\right]^2 = \frac{1}{4}$$

which is a constant.

2.9.2 Exercise

If $X \sim$ Exponential(θ) then show that the random variable $Y = g(X) = \log X$ has approximately constant variance.

2.10 Moment Generating Functions

If we are given the probability (density) function of a random variable X or the cumulative distribution function of the random variable X then we can determine everything there is to know about the distribution of X. There is a third type of function, the *moment generating function*, which also uniquely determines a distribution. The moment generating function is closely related to other transforms used in mathematics, the Laplace and Fourier transforms.

Moment generating functions are a powerful tool for determining the distributions of functions of random variables (Chapter 4), particularly sums, as well as determining the limiting distribution of a sequence of random variables (Chapter 5).

2.10.1 Definition - Moment Generating Function

If X is a random variable then $M(t) = E(e^{tX})$ is called the *moment generating function (m.g.f.) of X* if this expectation exists for all $t \in (-h, h)$ for some $h > 0$.

Important: When determining the moment generating function $M(t)$ of a random variable the values of t for which the expectation exists should always be stated.

2.10.2 Example

(a) Find the moment generating function of the random variable $X \sim \text{Gamma}(\alpha, \beta)$.

(b) Find the moment generating function of the random variable $X \sim \text{Negative Binomial}(k, p)$.

Solution

(a) If $X \sim \text{Gamma}(\alpha, \beta)$ then

$$
\begin{aligned}
M(t) &= \int_{-\infty}^{\infty} e^{tx} f(x)\, dx = \int_{0}^{\infty} e^{tx} \frac{x^{\alpha-1}}{\beta^{\alpha}\Gamma(\alpha)} e^{-x/\beta} dx \\
&= \int_{0}^{\infty} \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-x\left(\frac{1}{\beta}-t\right)} dx \quad \text{which converges for } t < \frac{1}{\beta} \\
&= \frac{1}{\beta^{\alpha}\Gamma(\alpha)} \int_{0}^{\infty} x^{\alpha-1} e^{-x\left(\frac{1-\beta t}{\beta}\right)} dx \quad \text{let } y = \left(\frac{1-\beta t}{\beta}\right) x \\
&= \frac{1}{\beta^{\alpha}\Gamma(\alpha)} \int_{0}^{\infty} \left(\frac{\beta y}{1-\beta t}\right)^{\alpha-1} e^{-y} \left(\frac{\beta}{1-\beta t}\right) dy \\
&= \frac{1}{\Gamma(\alpha)} \left(\frac{1}{1-\beta t}\right)^{\alpha} \int_{0}^{\infty} y^{\alpha-1} e^{-y} dy = \frac{\Gamma(\alpha)}{\Gamma(\alpha)} \left(\frac{1}{1-\beta t}\right)^{\alpha} \\
&= \left(\frac{1}{1-\beta t}\right)^{\alpha} \quad \text{for } t < \frac{1}{\beta}
\end{aligned}
$$

(b) If $X \sim$ Negative Binomial(k, p) then

$$
\begin{aligned}
M(t) &= \sum_{x \in 0}^{\infty} e^{tx} \binom{-k}{x} p^k (-q)^x \quad \text{where } q = 1 - p \\
&= p^k \sum_{x \in 0}^{\infty} \binom{-k}{x} \left[e^t (-q)\right]^x \quad \text{which converges for } \left|-qe^t\right| < 1 \\
&= p^k \left(1 - qe^t\right)^{-k} \quad \text{by 2.11.3(2) for } e^t < q^{-1} \\
&= \left(\frac{p}{1 - qe^t}\right)^k \quad \text{for } t < -\log(q)
\end{aligned}
$$

2.10.3 Exercise

(a) Show that the moment generating function of the random variable $X \sim$ Binomial(n, p) is $M(t) = \left(q + pe^t\right)^n$ for $t \in \Re$.

(b) Show that the moment generating function of the random variable $X \sim$ Poisson(θ) is $M(t) = e^{\theta\left(e^t - 1\right)}$ for $t \in \Re$.

If the moment generating function of random variable X exists then the following theorem gives us a method for determining the distribution of the random variable $Y = aX + b$ which is a linear function of X.

2.10.4 Theorem - Moment Generating Function of a Linear Function

Suppose the random variable X has moment generating function $M_X(t)$ defined for $t \in (-h, h)$ for some $h > 0$. Let $Y = aX + b$ where $a, b \in \mathcal{R}$ and $a \neq 0$. Then the moment generating function of Y is

$$
M_Y(t) = e^{bt} M_X(at) \quad \text{for } |t| < \frac{h}{|a|}
$$

Proof

$$
\begin{aligned}
M_Y(t) &= E\left(e^{tY}\right) = E\left(e^{t(aX+b)}\right) \\
&= e^{bt} E\left(e^{atX}\right) \quad \text{which exists for } |at| < h \\
&= e^{bt} M_X(at) \quad \text{for } |t| < \frac{h}{|a|}
\end{aligned}
$$

as required.

2.10.5 Example

(a) Find the moment generating function of $Z \sim \mathrm{N}(0, 1)$.

(b) Use (a) and the fact that $X = \mu + \sigma Z \sim \mathrm{N}(\mu, \sigma^2)$ to find the moment generating function of a $\mathrm{N}(\mu, \sigma^2)$ random variable.

Solution

(a) The moment generating function of Z is

$$\begin{aligned}
M_Z(t) &= \int_{-\infty}^{\infty} e^{tz} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\left(z^2 - 2tz\right)/2} dz \\
&= e^{t^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(z-t)^2/2} dz \\
&= e^{t^2/2} \quad \text{for } t \in \Re
\end{aligned}$$

by 2.4.4(2) since

$$\frac{1}{\sqrt{2\pi}} e^{-(z-t)/2}$$

is the probability density function of a $N(t, 1)$ random variable.

(b) By Theorem 2.10.4 the moment generating function of $X = \mu + \sigma Z$ is

$$\begin{aligned}
M_X(t) &= e^{\mu t} M_Z(\sigma t) \\
&= e^{\mu t} e^{(\sigma t)^2/2} \\
&= e^{\mu t + \sigma^2 t^2/2} \quad \text{for } t \in \Re
\end{aligned}$$

2.10.6 Exercise

If $X \sim$ Negative Binomial(k, p) then find the moment generating function of $Y = X + k$, $k = 1, 2, \ldots$

2.10.7 Theorem - Moments from Moment Generating Function

Suppose the random variable X has moment generating function $M(t)$ defined for $t \in (-h, h)$ for some $h > 0$. Then $M(0) = 1$ and

$$M^{(k)}(0) = E(X^k) \quad \text{for } k = 1, 2, \ldots$$

where

$$M^{(k)}(t) = \frac{d^k}{dt^k} M(t)$$

is the kth derivative of $M(t)$.

Proof (Continuous Case)

Note that

$$M(0) = E(X^0) = E(1) = 1$$

and also that

$$\frac{d^k}{dt^k} e^{tx} = x^k e^{tx} \quad k = 1, 2, \ldots \tag{2.6}$$

The result (2.6) can be proved by induction.

Now if X is a continuous random variable with moment generating function $M(t)$ defined for $t \in (-h, h)$ for some $h > 0$ then

$$\begin{aligned} M^{(k)}(t) &= \frac{d^k}{dt^k} E\left(e^{tX}\right) \\ &= \frac{d^k}{dt^k} \int_{-\infty}^{\infty} e^{tx} f(x)\, dx \\ &= \int_{-\infty}^{\infty} \frac{d^k}{dt^k} e^{tx} f(x)\, dx \quad k = 1, 2, \ldots \end{aligned}$$

assuming the operations of differentiation and integration can be exchanged. (This interchange of operations cannot always be done but for the moment generating functions of interest in this course the result does hold.)

Using (2.6) we have

$$\begin{aligned} M^{(k)}(t) &= \int_{-\infty}^{\infty} x^k e^{tx} f(x)\, dx \\ &= E\left(X^k e^{tX}\right) \quad t \in (-h, h) \quad \text{for some } h > 0 \end{aligned}$$

Letting $t = 0$ we obtain

$$M^{(k)}(0) = E\left(X^k\right) \quad k = 1, 2, \ldots$$

as required.

2.10.8 Example

If $X \sim \text{Gamma}(\alpha, \beta)$ then $M(t) = (1 - \beta t)^{-\alpha}, \ t < 1/\beta$. Find $E\left(X^k\right)$, $k = 1, 2, \ldots$ using Theorem 2.10.7.

Solution

$$\begin{aligned} \frac{d}{dt} M(t) &= M'(t) = \frac{d}{dt}(1 - \beta t)^{-\alpha} \\ &= \alpha\beta (1 - \beta t)^{-\alpha - 1} \end{aligned}$$

so

$$E(X) = M'(0) = \alpha\beta$$

$$\begin{aligned}
\frac{d^2}{dt^2} M(t) &= M''(t) = \frac{d^2}{dt^2}(1-\beta t)^{-\alpha} \\
&= \alpha(\alpha+1)\beta^2(1-\beta t)^{-\alpha-2}
\end{aligned}$$

so

$$E\left(X^2\right) = M''(0) = \alpha(\alpha+1)\beta^2$$

Continuing in this manner we have

$$\begin{aligned}
\frac{d^k}{dt^k} M(t) &= M^{(k)}(t) = \frac{d^k}{dt^k}(1-\beta t)^{-\alpha} \\
&= \alpha(\alpha+1)\cdots(\alpha+k-1)\beta^k(1-\beta t)^{-\alpha-k} \quad \text{for } k = 1, 2, \ldots
\end{aligned}$$

so

$$\begin{aligned}
E\left(X^k\right) &= M^{(k)}(0) \\
&= \alpha(\alpha+1)\cdots(\alpha+k-1)\beta^k = (\alpha+k-1)^{(k)}\beta^k \quad \text{for } k = 1, 2, \ldots
\end{aligned}$$

2.10.9 Important Idea

Suppose $M^{(k)}(t)$, $k = 1, 2, \ldots$ exists for $t \in (-h, h)$ for some $h > 0$, then $M(t)$ has a Maclaurin series given by

$$\sum_{k=0}^{\infty} \frac{M^{(k)}(0)}{k!} t^k$$

where

$$M^{(0)}(0) = M(0) = 1$$

The coefficient of t^k in this power series is equal to

$$\frac{M^{(k)}(0)}{k!} = \frac{E(X^k)}{k!}$$

Therefore if we can obtain a Maclaurin series for $M(t)$, for example, by using the Binomial series or the Exponential series, then we can find $E(X^k)$ by using

$$E(X^k) = k! \times \text{coefficient of } t^k \text{ in the Maclaurin series for } M(t) \tag{2.7}$$

2.10.10 Example

Suppose $X \sim \text{Gamma}(\alpha, \beta)$. Find $E\left(X^k\right)$ by using the Binomial series expansion for $M(t) = (1-\beta t)^{-\alpha}$, $t < 1/\beta$.

Solution

$$\begin{aligned} M(t) &= (1-\beta t)^{-\alpha} \\ &= \sum_{k=0}^{\infty} \binom{-\alpha}{k} (-\beta t)^k \quad \text{for } |t| < \frac{1}{\beta} \quad \text{by 2.11.3(2)} \\ &= \sum_{k=0}^{\infty} \binom{-\alpha}{k} (-\beta)^k t^k \quad \text{for } |t| < \frac{1}{\beta} \end{aligned}$$

The coefficient of t^k in $M(t)$ is

$$\binom{-\alpha}{k} (-\beta)^k \quad \text{for } k = 1, 2, \ldots$$

Therefore

$$\begin{aligned} E(X^k) &= k! \cdot \text{coefficient of } t^k \text{ in the Maclaurin series for } M(t) \\ &= k! \binom{-\alpha}{k} (-\beta)^k \\ &= k! \frac{(-\alpha)^{(k)}}{k!} (-\beta)^k \\ &= (-\alpha)(-\alpha-1) \cdots (-\alpha-k+2)(-\alpha-k+1)(-\beta)^k \\ &= (\alpha+k-1)(\alpha+k-2) \cdots (\alpha+1)(\alpha)\beta^k \\ &= (\alpha+k-1)^{(k)} \beta^k \quad \text{for } k = 1, 2, \ldots \end{aligned}$$

which is the same result as obtained in Example 2.10.8.

Moment generating functions are particularly useful for finding distributions of sums of independent random variables. The following theorem plays an important role in this technique.

2.10.11 Uniqueness Theorem for Moment Generating Functions

Suppose the random variable X has moment generating function $M_X(t)$ and the random variable Y has moment generating function $M_Y(t)$. Suppose also that $M_X(t) = M_Y(t)$ for all $t \in (-h, h)$ for some $h > 0$. Then X and Y have the same distribution, that is,
$P(X \leq s) = F_X(s) = F_Y(s) = P(Y \leq s)$ for all $s \in \Re$.

Proof

See Problem 18 for the proof of this result in the discrete case.

2.10.12 Example

If $X \sim$ Exponential(1) then find the distribution of $Y = \alpha + \beta X$ where $\beta > 0$ and $\alpha \in \Re$.

Solution

From Example 2.4.10 we know that if $X \sim$ Exponential(1) then $X \sim$ Gamma$(1, 1)$ so

$$M_X(t) = \frac{1}{1-t} \quad \text{for } t < 1$$

By Theorem 2.10.4

$$\begin{aligned} M_Y(t) &= e^{\alpha t} M_X(\beta t) \quad \text{for } \beta t < 1 \\ &= \frac{e^{\alpha t}}{1 - \beta t} \quad \text{for } t < \frac{1}{\beta} \end{aligned}$$

By examining the list of moment generating functions in Chapter 11 we see that this is the moment generating function of a Two Parameter Exponential(α, β) random variable. Therefore by the Uniqueness Theorem for Moment Generating Functions, Y has a Two Parameter Exponential(α, β) distribution.

2.10.13 Example

If $X \sim$ Gamma(α, β), where α is a positive integer, then show $\frac{2X}{\beta} \sim \chi^2(2\alpha)$.

Solution

(a) From Example 2.10.2 the moment generating function of X is

$$M(t) = \left(\frac{1}{1-\beta t}\right)^{\alpha} \quad \text{for } t < \frac{1}{\beta}$$

By Theorem 2.10.4 the moment generating function of $Y = \frac{2X}{\beta}$ is

$$\begin{aligned} M_Y(t) &= M_X\left(\frac{2}{\beta}t\right) \quad \text{for } |t| < \frac{1}{\left|\left(\frac{2}{\beta}\right)(\beta)\right|} \\ &= \left[\frac{1}{1 - \beta\left(\frac{2}{\beta}\right)t}\right]^{\alpha} \quad \text{for } |t| < \frac{1}{2} \\ &= \left(\frac{1}{1-2t}\right)^{\alpha} \quad \text{for } |t| < \frac{1}{2} \end{aligned}$$

By examining the list of moment generating functions in Chapter 11 we see that this is the moment generating function of a $\chi^2(2\alpha)$ random variable if α is a positive integer. Therefore by the Uniqueness Theorem for Moment Generating Functions, Y has a $\chi^2(2\alpha)$ distribution if α is a positive integer.

2.10.14 Exercise

Suppose the random variable X has moment generating function

$$M(t) = e^{t^2/2} \quad \text{for } t \in \Re$$

(a) Use (2.7) and the Exponential series 2.11.7 to find $E(X)$ and $Var(X)$.

(b) Find the moment generating function of $Y = 2X - 1$. What is the distribution of Y?

2.11 Calculus Review

2.11.1 Geometric Series

$$\sum_{x=0}^{\infty} at^x = a + at + at^2 + \cdots = \frac{a}{1-t} \quad \text{for } |t| < 1$$

2.11.2 Useful Results

(1)

$$\sum_{x=0}^{\infty} t^x = \frac{1}{1-t} \quad \text{for } |t| < 1$$

(2)

$$\sum_{x=1}^{\infty} xt^{x-1} = \frac{1}{(1-t)^2} \quad \text{for } |t| < 1$$

2.11.3 Binomial Series

(1) For $n \in \mathcal{Z}^+$ (the positive integers)

$$(a+b)^n = \sum_{x=0}^{n} \binom{n}{x} a^x b^{n-x}$$

where

$$\binom{n}{x} = \frac{n!}{x!\,(n-x)!} = \frac{n^{(x)}}{x!}$$

(2) For $n \in \mathcal{Q}$ (the rational numbers) and $|t| < 1$

$$(1+t)^n = \sum_{x=0}^{\infty} \binom{n}{x} t^x$$

where

$$\binom{n}{x} = \frac{n^{(x)}}{x!} = \frac{n(n-1)\cdots(n-x+1)}{x!}$$

2.11.4 Important Identities

$$(1) \quad x^{(k)}\binom{n}{x} = n^{(k)}\binom{n-k}{x-k}$$

$$(2) \quad \binom{x+k-1}{x} = \binom{x+k-1}{k-1} = (-1)^x\binom{-k}{x}$$

2.11.5 Multinomial Theorem

If n is a positive integer and $a_1, a_2, \ldots, a_k$ are real numbers, then

$$(a_1 + a_2 + \cdots + a_k)^n = \sum\sum\cdots\sum \frac{n!}{x_1!x_2!\cdots x_k!} a_1^{x_1} a_2^{x_2} \cdots a_k^{x_k}$$

where the summation extends over all non-negative integers $x_1, x_2, \ldots, x_k$ with $x_1 + x_2 + \cdots + x_k = n$.

2.11.6 Hypergeometric Identity

$$\sum_{x=0}^{\infty}\binom{a}{x}\binom{b}{n-x} = \binom{a+b}{n}$$

2.11.7 Exponential Series

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \cdots = \sum_{n=0}^{\infty}\frac{x^n}{n!} \quad \text{for } x \in \Re$$

2.11.8 Logarithmic Series

$$\ln(1+x) = \log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots \quad \text{for } -1 < x \leq 1$$

2.11.9 First Fundamental Theorem of Calculus (FTC1)

If f is continuous on $[a, b]$ then the function g defined by

$$g(x) = \int_a^x f(t)\,dt \quad \text{for } a \leq x \leq b$$

is continuous on $[a, b]$, differentiable on (a, b) and $g'(x) = f(x)$.

2.11.10 Fundamental Theorem of Calculus and the Chain Rule

Suppose we want the derivative with respect to x of $G(x)$ where

$$G(x) = \int_a^{h(x)} f(t)\,dt \quad \text{for } a \le x \le b$$

and $h(x)$ is a differentiable function on $[a, b]$. If we define

$$g(u) = \int_a^u f(t)\,dt$$

then $G(x) = g(h(x))$. Then by the Chain Rule

$$\begin{aligned} G'(x) &= g'(h(x))\,h'(x) \\ &= f(h(x))\,h'(x) \quad \text{for } a < x < b \end{aligned}$$

2.11.11 Improper Integrals

(a) If $\int_a^b f(x)\,dx$ exists for every number $b \ge a$ then

$$\int_a^{\infty} f(x)\,dx = \lim_{b \to \infty} \int_a^b f(x)\,dx$$

provided this limit exists. If the limit exists we say the improper integral converges otherwise we say the improper integral diverges.

(b) If $\int_a^b f(x)\,dx$ exists for every number $a \le b$ then

$$\int_{-\infty}^{b} f(x)\,dx = \lim_{a \to -\infty} \int_a^b f(x)\,dx$$

provided this limit exists.

(c) If both $\int_a^{\infty} f(x)\,dx$ and $\int_{-\infty}^{a} f(x)\,dx$ are convergent then we define

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{a} f(x)\,dx + \int_a^{\infty} f(x)\,dx$$

where a is any real number.

2.11.12 Comparison Test for Improper Integrals

Suppose that f and g are continuous functions with $f(x) \geq g(x) \geq 0$ for $x \geq a$.
(a)

$$\text{If } \int_a^\infty f(x)\,dx \text{ is convergent then } \int_a^\infty g(x)\,dx \text{ is convergent.}$$

(b)

$$\text{If } \int_a^\infty g(x)\,dx \text{ is divergent then } \int_a^\infty f(x)\,dx \text{ is divergent.}$$

2.11.13 Useful Result for Comparison Test

$$\int_1^\infty \frac{1}{x^p}dx \quad \text{converges if and only if } p > 1 \tag{2.8}$$

2.11.14 Useful Inequalities

$$\frac{1}{1+y^p} \leq \frac{1}{y^p} \quad \text{for } y \geq 1,\ p > 0$$

$$\frac{1}{1+y^p} \geq \frac{1}{y^p + y^p} = \frac{1}{2y^p} \quad \text{for } y \geq 1,\ p > 0$$

2.11.15 Taylor's Theorem

Suppose f is a real-valued function such that the derivatives $f^{(1)}, f^{(2)}, \ldots, f^{(n+1)}$ all exist on an open interval containing the point $x = a$. Then

$$f(x) = f(a) + f^{(1)}(a)(x-a) + \frac{f^{(2)}(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n + \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1}$$

for some c between a and x.

2.12 Chapter 2 Problems

1. Consider the following functions:

   (a) $f(x) = kx\theta^x \quad$ for $x = 1, 2, \ldots; \quad 0 < \theta < 1$

   (b) $f(x) = k\left[1 + (x/\theta)^2\right]^{-1} \quad$ for $x \in \Re, \quad \theta > 0$

   (c) $f(x) = ke^{-|x-\theta|} \quad$ for $x \in \Re, \quad \theta \in \Re$

   (d) $f(x) = k(1-x)^\theta \quad$ for $0 < x < 1, \quad \theta > 0$

   (e) $f(x) = kx^2e^{-\theta x} \quad$ for $x > 0, \quad \theta > 0$

   (f) $f(x) = kx^{-(\theta+1)} \quad$ for $x \geq 1, \quad \theta > 0$

   (g) $f(x) = ke^{-x/\theta}\left(1 + e^{-x/\theta}\right)^{-2} \quad$ for $x \in \Re, \quad \theta > 0$

   (h) $f(x) = kx^{-3}e^{-1/(\theta x)} \quad$ for $x > 0, \quad \theta > 0$

   (i) $f(x) = k(1+x)^{-\theta} \quad$ for $x > 0, \ \theta > 1$

   (j) $f(x) = k(1-x)x^{\theta-1} \quad$ for $0 < x < 1, \quad \theta > 0$

   In each case:
   (1) Determine k so that $f(x)$ is a probability (density) function and sketch $f(x)$.
   (2) Let X be a random variable with probability (density) function $f(x)$. Find the cumulative distribution function of X.
   (3) Find $E(X)$ and $Var(X)$ using the probability (density) function. Indicate the values of θ for which $E(X)$ and $Var(X)$ exist.
   (4) Find $P(0.5 < X \leq 2)$ and $P(X > 0.5|X \leq 2)$.
   In (a) use $\theta = 0.3$, in (b) use $\theta = 1$, in (c) use $\theta = 0$, in (d) use $\theta = 5$, in (e) use $\theta = 1$, in (f) use $\theta = 1$, in (g) use $\theta = 2$, in (h) use $\theta = 1$, in (i) use $\theta = 2$, in (j) use $\theta = 3$.

2. Determine if θ is a location parameter, a scale parameter, or neither for the distributions in $(b) - (j)$ of Problem 1.

3. (a) If $X \sim$ Weibull$(2, \theta)$ then show θ is a scale parameter for this distribution.

   (b) If $X \sim$ Uniform$(0, \theta)$ then show θ is a scale parameter for this distribution.

4. Suppose X is a continuous random variable with probability density function

$$f(x) = \begin{cases} ke^{-(x-\theta)^2/2} & |x-\theta| \leq c \\ ke^{-c|x-\theta|+c^2/2} & |x-\theta| > c \end{cases}$$

   (a) Show that

$$\frac{1}{k} = \frac{2}{c}e^{-c^2/2} + \sqrt{2\pi}\left[2\Phi(c) - 1\right]$$

   where Φ is the N$(0,1)$ cumulative distribution function.

(b) Find the cumulative distribution function of X, $E(X)$ and $Var(X)$.

(c) Show that θ is a location parameter for this distribution.

(d) On the same graph sketch $f(x)$ for $c = 1$, $\theta = 0$, $f(x)$ for $c = 2$, $\theta = 0$ and the $\mathrm{N}(0, 1)$ probability density function. What do you notice?

5. The Geometric and Exponential distributions both have a property referred to as the memoryless property.

(a) Suppose $X \sim$ Geometric(p). Show that $P(X \geq k + j | X \geq k) = P(X \geq j)$ where k and j are nonnegative integers. Explain why this is called the memoryless property.

(b) Show that if $Y \sim$ Exponential(θ) then $P(Y \geq a + b | Y \geq a) = P(Y \geq b)$ where $a > 0$ and $b > 0$.

6. Suppose that $f_1(x), f_2(x), \ldots, f_k(x)$ are probability density functions with support sets $A_1, A_2, \ldots, A_k$, means $\mu_1, \mu_2, \ldots, \mu_k$, and finite variances $\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2$ respectively. Suppose that $0 < p_1, p_2, \ldots, p_k < 1$ and $\sum\limits_{i=1}^{k} p_i = 1$.

(a) Show that $g(x) = \sum\limits_{i=1}^{k} p_i f_i(x)$ is a probability density function.

(b) Let X be a random variable with probability density function $g(x)$. Find the support set of X, $E(X)$ and $Var(X)$.

7.

(a) If $X \sim$ Gamma(α, β) then find the probability density function of $Y = e^X$.

(b) If $X \sim$ Gamma(α, β) then show $Y = 1/X \sim$ Inverse Gamma(α, β).

(c) If $X \sim$ Gamma(k, β) then show $Y = 2X/\beta \sim \chi^2(2k)$ for $k = 1, 2, \ldots$.

(d) If $X \sim \mathrm{N}(\mu, \sigma^2)$ then find the probability density function of $Y = e^X$.

(e) If $X \sim \mathrm{N}(\mu, \sigma^2)$ then find the probability density function of $Y = X^{-1}$.

(f) If $X \sim$ Uniform($-\frac{\pi}{2}, \frac{\pi}{2}$) then show that $Y = \tan X \sim$ Cauchy(1, 0).

(g) If $X \sim$ Pareto(α, β) then show that $Y = \beta \log(X/\alpha) \sim$ Exponential(1).

(h) If $X \sim$ Weibull($2, \theta$) then show that $Y = X^2 \sim$ Exponential(θ^2).

(i) If $X \sim$ Double Exponential(0, 1) then find the probability density function of $Y = X^2$.

(j) If $X \sim \mathrm{t}(k)$ then show that $Y = X^2 \sim \mathrm{F}(1, k)$.

8. Suppose $T \sim t(n)$.

   (a) Show that $E(T) = 0$ if $n > 1$.

   (b) Show that $Var(T) = n/(n-2)$ if $n > 2$.

9. Suppose $X \sim \text{Beta}(a, b)$.

   (a) Find $E\left(X^k\right)$ for $k = 1, 2, \ldots$. Use this result to find $E(X)$ and $Var(X)$.

   (b) Graph the probability density function for (i) $a = 0.7$, $b = 0.7$, (ii) $a = 1$, $b = 3$, (iii) $a = 2$, $b = 2$, (iv) $a = 2$, $b = 4$, and (v) $a = 3$, $b = 1$ on the same graph.

   (c) What special probability density function is obtained for $a = b = 1$?

10. If $E(|X|^k)$ exists for some integer $k > 1$, then show that $E(|X|^j)$ exists for $j = 1, 2, \ldots, k-1$.

11. If $X \sim \text{Binomial}(n, \theta)$, find the variance stabilizing transformation $g(X)$ such that $Var[g(X)]$ is approximately constant.

12. Prove that for any random variable X,
$$E\left(X^4\right) \geq \frac{1}{4} P\left(X^2 \geq \frac{1}{2}\right)$$

13. For each of the following probability (density) functions derive the moment generating function $M(t)$. State the values for which $M(t)$ exists and use the moment generating function to find the mean and variance.

   (a) $f(x) = \binom{n}{x} p^x (1-p)^{n-x}$ for $x = 0, 1, \ldots, n$; $0 < p < 1$

   (b) $f(x) = \mu^x e^{-\mu}/x!$ for $x = 0, 1, \ldots$; $\mu > 0$

   (c) $f(x) = \frac{1}{\beta} e^{-(x-\theta)/\beta}$ for $x > \theta$; $\theta \in \Re$, $\beta > 0$

   (d) $f(x) = \frac{1}{2} e^{-|x-\theta|}$ for $x \in \Re$; $\theta \in \Re$

   (e) $f(x) = 2x$ for $0 < x < 1$

   (f) $f(x) = \begin{cases} x & 0 \leq x \leq 1 \\ 2 - x & 1 < x \leq 2 \\ 0 & \text{otherwise} \end{cases}$

14. Suppose X is a random variable with moment generating function $M(t) = E(e^{tX})$ which exists for $t \in (-h, h)$ for some $h > 0$. Then $K(t) = \log M(t)$ is called the cumulant generating function of X.

   (a) Show that $E(X) = K'(0)$ and $Var(X) = K''(0)$.

(b) If $X \sim$ Negative Binomial(k, p) then use (a) to find $E(X)$ and $Var(X)$.

15. For each of the following find the Maclaurin series for $M(t)$ using known series. Thus determine all the moments of X if X is a random variable with moment generating function $M(t)$:

    (a) $M(t) = (1-t)^{-3}$ for $|t| < 1$

    (b) $M(t) = (1+t)/(1-t)$ for $|t| < 1$

    (c) $M(t) = e^t/(1-t^2)$ for $|t| < 1$

16. Suppose $Z \sim$ N$(0,1)$ and $Y = |Z|$.

    (a) Show that $M_Y(t) = 2\Phi(t) e^{t^2/2}$ for $t \in \Re$ where $\Phi(t)$ is the cumulative distribution function of a N$(0,1)$ random variable.

    (b) Use (a) to find $E(|Z|)$ and $Var(|Z|)$.

17. Suppose $X \sim \chi^2_{(1)}$ and $Z \sim$ N$(0,1)$. Use the properties of moment generating functions to compute $E\left(X^k\right)$ and $E\left(Z^k\right)$ for $k = 1, 2, \ldots$ How are these two related? Is this what you expected?

18. Suppose X and Y are discrete random variables such that $P(X = j) = p_j$ and $P(Y = j) = q_j$ for $j = 0, 1, \ldots$. Suppose also that $M_X(t) = M_Y(t)$ for $t \in (-h, h)$, $h > 0$. Show that X and Y have the same distribution. (Hint: Compare $M_X(\log s)$ and $M_Y(\log s)$ and recall that if two power series are equal then their coefficients are equal.)

19. Suppose X is a random variable with moment generating function $M(t) = e^t/(1-t^2)$ for $|t| < 1$.

    (a) Find the moment generating function of $Y = (X-1)/2$.

    (b) Use the moment generating function of Y to find $E(Y)$ and $Var(Y)$.

    (c) What is the distribution of Y?

3. Multivariate Random Variables

Models for real phenomena usually involve more than a single random variable. When there are multiple random variables associated with an experiment or process we usually denote them as $X, Y, \ldots$ or as $X_1, X_2, \ldots$. For example, your final mark in a course might involve X_1 = your assignment mark, X_2 = your midterm test mark, and X_3 = your exam mark. We need to extend the ideas introduced in Chapter 2 for univariate random variables to deal with multivariate random variables.

In Section 3.1 we began be defining the joint and marginal cumulative distribution functions since these definitions hold regardless of what type of random variable we have. We define these functions in the case of two random variables X and Y. More than two random variables will be considered in specific examples in later sections. In Section 3.2 we briefly review discrete joint probability functions and marginal probability functions that were introduced in a previous probability course. In Section 3.3 we introduce the ideas needed for two continuous random variables and look at detailed examples since this is new material. In Section 3.4 we define independence for two random variables and show how the Factorization Theorem for Independence can be used. When two random variables are not independent then we are interested in conditional distributions. In Section 3.5 we review the definition of a conditional probability function for discrete random variables and define a conditional probability density function for continuous random variables which is new material. In Section 3.6 we review expectations of functions of discrete random variables. We also define expectations of functions of continuous random variables which is new material except for the case of Normal random variables. In Section 3.7 we define conditional expectations which arise from the conditional distributions discussed in Section 3.5. In Section 3.8 we discuss moment generating functions for two or more random variables, and show how the Factorization Theorem for Moment Generating Functions can be used to prove that random variables are independent. In Section 3.9 we review the Multinomial distribution and its properties. In Section 3.10 we introduce the very important Bivariate Normal distribution and its properties. Section 3.11 contains some useful results related to evaluating double integrals.

3.1 Joint and Marginal Cumulative Distribution Functions

We begin with the definitions and properties of the cumulative distribution functions associated with two random variables.

3.1.1 Definition - Joint Cumulative Distribution Function

Suppose X and Y are random variables defined on a sample space S. The *joint cumulative distribution function of* X *and* Y is given by

$$F(x,y) = P(X \le x, Y \le y) \quad \text{for } (x,y) \in \Re^2$$

3.1.2 Properties - Joint Cumulative Distribution Function

(1) F is non-decreasing in x for fixed y

(2) F is non-decreasing in y for fixed x

(3) $\lim\limits_{x \to -\infty} F(x,y) = 0$ and $\lim\limits_{y \to -\infty} F(x,y) = 0$

(4) $\lim\limits_{(x,y) \to (-\infty,-\infty)} F(x,y) = 0$ and $\lim\limits_{(x,y) \to (\infty,\infty)} F(x,y) = 1$

3.1.3 Definition - Marginal Distribution Function

The *marginal cumulative distribution function of* X is given by

$$\begin{aligned} F_1(x) &= \lim_{y \to \infty} F(x,y) \\ &= P(X \le x) \quad \text{for } x \in \Re \end{aligned}$$

The *marginal cumulative distribution function of* Y is given by

$$\begin{aligned} F_2(y) &= \lim_{x \to \infty} F(x,y) \\ &= P(Y \le y) \quad \text{for } y \in \Re \end{aligned}$$

Note: The definitions and properties of the joint cumulative distribution function and the marginal cumulative distribution functions hold for both (X,Y) discrete random variables and for (X,Y) continuous random variables.

Joint and marginal cumulative distribution functions for discrete random variables are not very convenient for determining probabilities. Joint and marginal probability functions, which are defined in Section 3.2, are more frequently used for discrete random variables. In Section 3.3 we look at specific examples of joint and marginal cumulative distribution functions for continuous random variables. In Chapter 5 we will see the important role of cumulative distribution functions in determining asymptotic distributions.

3.2 Bivariate Discrete Distributions

Suppose X and Y are random variables defined on a sample space S. If there is a countable subset $A \subset \Re^2$ such that $P[(X,Y) \in A] = 1$, then X and Y are discrete random variables.

Probabilities for discrete random variables are most easily handled in terms of joint probability functions.

3.2.1 Definition - Joint Probability Function

Suppose X and Y are discrete random variables.

The *joint probability function* of X and Y is given by

$$f(x,y) = P(X = x, Y = y) \quad \text{for } (x,y) \in \Re^2$$

The set $A = \{(x,y) : f(x,y) > 0\}$ is called the support set of (X,Y).

3.2.2 Properties of Joint Probability Function

(1) $f(x,y) \geq 0$ for $(x,y) \in \Re^2$

(2) $\sum\limits_{(x,y)\in} \sum\limits_{A} f(x,y) = 1$

(3) For any set $R \subset \Re^2$

$$P[(X,Y) \in R] = \sum_{(x,y)\in} \sum_{R} f(x,y)$$

3.2.3 Definition - Marginal Probability Function

Suppose X and Y are discrete random variables with joint probability function $f(x,y)$.

The *marginal probability function of* X is given by

$$\begin{aligned} f_1(x) &= P(X = x) \\ &= \sum_{\textit{all } y} f(x,y) \quad \text{for } x \in \Re \end{aligned}$$

and the *marginal probability function of* Y is given by

$$\begin{aligned} f_2(y) &= P(Y = y) \\ &= \sum_{\textit{all } x} f(x,y) \quad \text{for } y \in \Re \end{aligned}$$

3.2.4 Example

In a fourth year statistics course there are 10 actuarial science students, 9 statistics students and 6 math business students. Five students are selected at random without replacement.

Let X be the number of actuarial science students selected, and let Y be the number of statistics students selected.

Find

(a) the joint probability function of X and Y

(b) the marginal probability function of X

(c) the marginal probability function of Y

(d) $P(X > Y)$

Solution

(a) The joint probability function of X and Y is

$$\begin{aligned} f(x,y) &= P(X=x, Y=y) \\ &= \frac{\binom{10}{x}\binom{9}{y}\binom{6}{5-x-y}}{\binom{25}{5}} \\ \text{for } x &= 0,1,\ldots,5, \;\; y=0,1,\ldots,5, \;\; x+y \leq 5 \end{aligned}$$

(b) The marginal probability function of X is

$$\begin{aligned} f_1(x) &= P(X=x) \\ &= \sum_{y=0}^{\infty} \frac{\binom{10}{x}\binom{9}{y}\binom{6}{5-x-y}}{\binom{25}{5}} \\ &= \frac{\binom{10}{x}\binom{15}{5-x}}{\binom{25}{5}} \sum_{y=0}^{\infty} \frac{\binom{9}{y}\binom{6}{5-x-y}}{\binom{15}{5-x}} \\ &= \frac{\binom{10}{x}\binom{15}{5-x}}{\binom{25}{5}} \quad \text{for } x = 0,1,\ldots,5 \end{aligned}$$

by the Hypergeometric identity 2.11.6. Note that the marginal probability function of X is Hypergeometric$(25, 10, 5)$. This makes sense because, when we are only interested in the number of actuarial science students, we only have two types of objects (actuarial science students and non-actuarial science students) and we are sampling without replacement which gives us the familiar Hypergeometric probability function.

(c) The marginal probability function of Y is

$$\begin{aligned}
f_2(y) &= P(Y=y) \\
&= \sum_{x=0}^{\infty} \frac{\binom{10}{x}\binom{9}{y}\binom{6}{5-x-y}}{\binom{25}{5}} \\
&= \frac{\binom{9}{y}\binom{16}{5-y}}{\binom{25}{5}} \sum_{x=0}^{\infty} \frac{\binom{10}{x}\binom{6}{5-x-y}}{\binom{16}{5-y}} \\
&= \frac{\binom{9}{y}\binom{16}{5-y}}{\binom{25}{5}} \quad \text{for } y = 0, 1, \ldots, 5
\end{aligned}$$

by the Hypergeometric identity 2.11.6. The marginal probability function of Y is Hypergeometric$(25, 9, 5)$.

(d)

$$\begin{aligned}
P(X>Y) &= \sum_{(x,y):\, x>y} \sum f(x,y) \\
&= f(1,0)+f(2,0)+f(3,0)+f(4,0)+f(5,0) \\
&\quad +f(2,1)+f(3,1)+f(4,1) \\
&\quad +f(3,2)
\end{aligned}$$

3.2.5 Exercise

The Hardy-Weinberg law of genetics states that, under certain conditions, the relative frequencies with which three genotypes AA, Aa and aa occur in the population will be θ^2, $2\theta(1-\theta)$ and $(1-\theta)^2$ respectively where $0<\theta<1$. Suppose n members of a very large population are selected at random.

Let X be the number of AA types selected and let Y be the number of Aa types selected. Find

(a) the joint probability function of X and Y

(b) the marginal probability function of X

(c) the marginal probability function of Y

(d) $P(X+Y=t)$ for $t=0,1,\ldots$.

3.3 Bivariate Continuous Distributions

Probabilities for continuous random variables can also be specifed in terms of joint probability density functions.

3.3.1 Definition - Joint Probability Density Function

Suppose that $F(x, y)$ is a continuous function and that

$$f(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y)$$

exists and is a continuous function except possibly along a finite number of curves. Suppose also that

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1$$

Then X and Y are said to be continuous random variables with joint probability density function f. The set $A = \{(x, y) : f(x, y) > 0\}$ is called the support set of (X, Y).

Note: We will arbitrarily define $f(x, y)$ to be equal to 0 when $\frac{\partial^2}{\partial x \partial y} F(x, y)$ does not exist although we could define it to be any real number.

3.3.2 Properties - Joint Probability Density Function

(1)

$$f(x, y) \geq 0 \quad \text{for all } (x, y) \in \Re^2$$

(2)

$$\begin{aligned} P[(X, Y) \in R] &= \iint_R f(x, y) dx dy \quad \text{for } R \subset \Re^2 \\ &= \text{the volume under the surface } z = f(x, y) \\ &\quad \text{and above the region } R \text{ in the } xy-\text{plane} \end{aligned}$$

3.3.3 Example

Suppose X and Y are continuous random variables with joint probability density function

$$f(x, y) = x + y \quad \text{for } 0 < x < 1,\ 0 < y < 1$$

and 0 otherwise. The support set of (X, Y) is $A = \{(x, y) : 0 < x < 1,\ 0 < y < 1\}$.

The joint probability function for $(x, y) \in A$ is graphed in Figure 3.1. We notice that the surface is the portion of the plane $z = x + y$ lying above the region A.

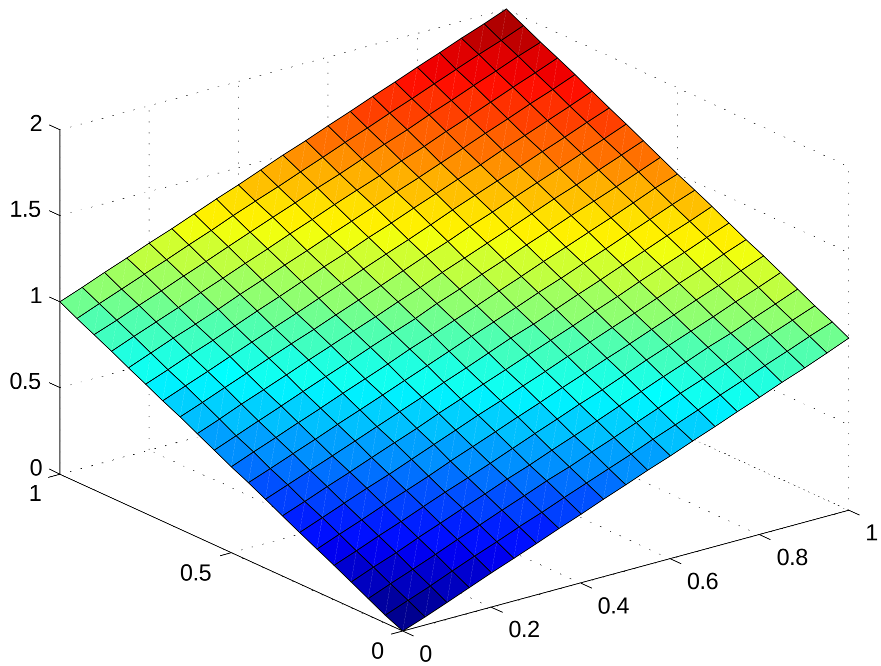


Figure 3.1: Graph of joint probability density function for Example 3.3.3

(a) Show that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = 1$$

(b) Find

(i) $P\left(X \le \frac{1}{3}, Y \le \frac{1}{2}\right)$

(ii) $P(X \le Y)$

(iii) $P\left(X + Y \le \frac{1}{2}\right)$

(iv) $P\left(XY \le \frac{1}{2}\right)$.

Solution

(a) A graph of the support set for (X, Y) is given in Figure 3.2. Such a graph is useful for determining the limits of integration of the double integral.

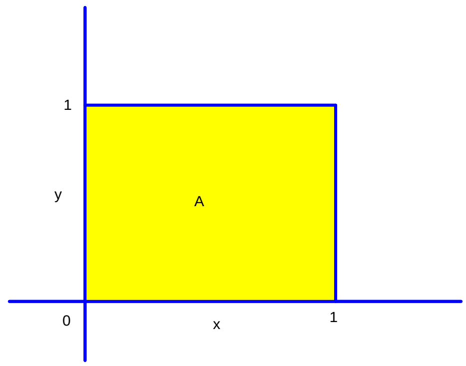


Figure 3.2: Graph of the support set of (X, Y) for Example 3.3.3

$$\begin{aligned}
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy &= \int\int_{(x,y)\,\in A} (x+y)\,dxdy \\
&= \int_0^1\int_0^1 (x+y)\,dxdy \\
&= \int_0^1 \left[\left(\frac{1}{2}x^2 + xy\right)|_0^1\right] dy \\
&= \int_0^1 \left(\frac{1}{2} + y\right) dy \\
&= \left(\frac{1}{2}y + \frac{1}{2}y^2\right)|_0^1 \\
&= \frac{1}{2} + \frac{1}{2} \\
&= 1
\end{aligned}$$

(b) (i) A graph of the region of integration is given in Figure 3.3

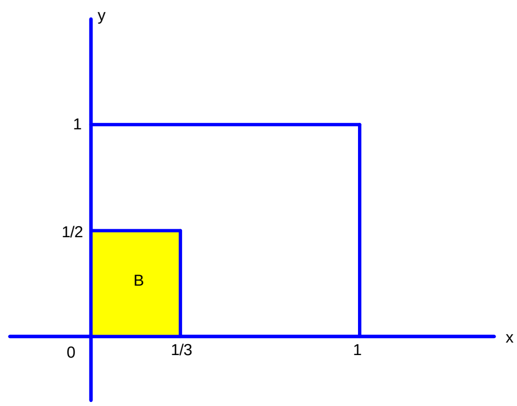


Figure 3.3: Graph of the integration region for Example 3.3.3(b)(i)

$$\begin{aligned}
P\left(X \leq \frac{1}{3}, Y \leq \frac{1}{2}\right) &= \int\limits_{(x,y)}\int\limits_{\in B} (x+y)\, dxdy \\
&= \int\limits_0^{1/2}\int\limits_0^{1/3} (x+y)\, dxdy \\
&= \int\limits_0^{1/2} \left[\left(\frac{1}{2}x^2 + xy\right) |_0^{1/3}\right] dy \\
&= \int\limits_0^{1/2} \left[\frac{1}{2}\left(\frac{1}{3}\right)^2 + \frac{1}{3}y\right] dy \\
&= \left(\frac{1}{18}y + \frac{1}{6}y^2\right) |_0^{1/2} \\
&= \frac{1}{18}\left(\frac{1}{2}\right) + \frac{1}{6}\left(\frac{1}{2}\right)^2 \\
&= \frac{5}{72}
\end{aligned}$$

(*ii*) A graph of the region of integration is given in Figure 3.4. Note that when the region is not rectangular then care must be taken with the limits of integration.

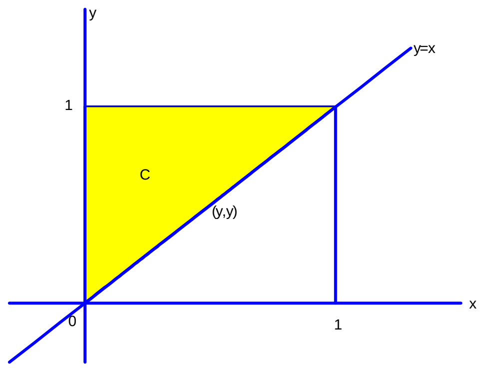


Figure 3.4: Graph of the integration region for Example 3.3.3(b)(ii)

$$\begin{aligned}
P(X &\leq Y) = \int\int_{(x,y)\,\in C} (x+y)\,dxdy \\
&= \int_{y=0}^{1}\int_{x=0}^{y} (x+y)\,dxdy \\
&= \int_{0}^{1}\left[\left(\frac{1}{2}x^2 + xy\right)|_0^y\right] dy \\
&= \int_{0}^{1}\left(\frac{1}{2}y^2 + y^2\right) dy \\
&= \int_{0}^{1}\frac{3}{2}y^2 dy = \frac{1}{2}y^3|_0^1 \\
&= \frac{1}{2}
\end{aligned}$$

Alternatively

$$P(X \leq Y) = \int_{x=0}^{1}\int_{y=x}^{1} (x+y)\,dydx$$

Why does the answer of $1/2$ make sense when you look at Figure 3.1?

(iii) A graph of the region of integration is given in Figure 3.5.

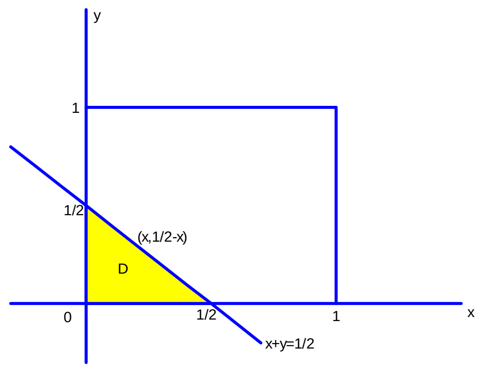


Figure 3.5: Graph of the region of integration for Example 3.3.3(b)(iii)

$$
\begin{aligned}
P\left(X+Y \leq \frac{1}{2}\right) &= \int\int_{(x,y)\,\in D} (x+y)\,dxdy \\
&= \int_{x=0}^{\frac{1}{2}} \int_{y=0}^{\frac{1}{2}-x} (x+y)\,dydx \\
&= \int_0^{\frac{1}{2}} \left[\left(xy+\frac{1}{2}y^2\right)|_0^{\frac{1}{2}-x}\right] dx \\
&= \int_0^{\frac{1}{2}} \left[x\left(\frac{1}{2}-x\right)+\frac{1}{2}\left(\frac{1}{2}-x\right)^2\right] dx \\
&= \int_0^{\frac{1}{2}} \left(-\frac{x^2}{2}+\frac{1}{8}\right) dx = \left(-\frac{x^3}{6}+\frac{1}{8}x\right)|_0^{1/2} \\
&= \frac{2}{48}
\end{aligned}
$$

Alternatively

$$P\left(X+Y \leq \frac{1}{2}\right) = \int_{y=0}^{\frac{1}{2}} \int_{x=0}^{\frac{1}{2}-y} (x+y)\,dxdy$$

Why does this small probability make sense when you look at Figure 3.1?

(*iv*) A graph of the region of integration E is given in Figure 3.6. In this example the integration can be done more easily by integrating over the region F.

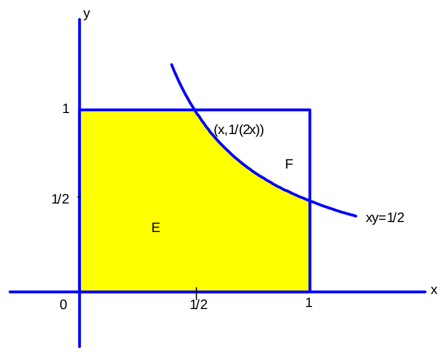


Figure 3.6: Graph of the region of integration for Example 3.3.3(b)(iv)

$$\begin{aligned}
P\left(XY \leq \frac{1}{2}\right) &= \int\limits_{(x,y)\ \in E}\int (x+y)\,dxdy \\
&= 1 - \int\limits_{(x,y)\ \in F}\int (x+y)\,dxdy \\
&= 1 - \int\limits_{x=\frac{1}{2}}^{1}\int\limits_{y=\frac{1}{2x}}^{1} (x+y)\,dydx \\
&= 1 - \int\limits_{\frac{1}{2}}^{1}\left[\left(xy+\frac{1}{2}y^2\right)|_{\frac{1}{2x}}^{1}\right]dx \\
&= 1 - \int\limits_{\frac{1}{2}}^{1}\left\{x+\frac{1}{2}-\left[x\left(\frac{1}{2x}\right)+\frac{1}{2}\left(\frac{1}{2x}\right)^2\right]\right\}dx \\
&= 1 - \int\limits_{\frac{1}{2}}^{1}\left(x-\frac{1}{8x^2}\right)dx \\
&= 1 - \left[\left(\frac{1}{2}x^2+\frac{1}{8x}\right)|_{\frac{1}{2}}^{1}\right] \\
&= \frac{3}{4}
\end{aligned}$$

3.3.4 Definition of Marginal Probability Density Function

Suppose X and Y are continuous random variables with joint probability density function $f(x,y)$.

The *marginal probability density function of* X is given by

$$f_1(x) = \int_{-\infty}^{\infty} f(x,y)dy \quad \text{for } x \in \Re$$

and the *marginal probability density function of* Y is given by

$$f_2(y) = \int_{-\infty}^{\infty} f(x,y)dx \quad \text{for } y \in \Re$$

3.3.5 Example

For the joint probability density function in Example 3.3.3 determine:

(a) the marginal probability density function of X and the marginal probability density function of Y

(b) the joint cumulative distribution function of X and Y

(c) the marginal cumulative distribution function of X and the marginal cumulative distribution function of Y

Solution

(a) The marginal probability density function of X is

$$f_1(x) = \int_{-\infty}^{\infty} f(x,y)dy = \int_0^1 (x+y)\,dy = \left(xy + \frac{1}{2}y^2\right)|_0^1 = x + \frac{1}{2} \quad \text{for } 0 < x < 1$$

and 0 otherwise.

Since both the joint probability density function $f(x,y)$ and the support set A are symmetric in x and y then by symmetry the marginal probability density function of Y is

$$f_2(y) = y + \frac{1}{2} \quad \text{for } 0 < y < 1$$

and 0 otherwise.

(b) Since

$$\begin{aligned} P(X \le x, Y \le y) &= \int_0^y \int_0^x (s+t)\,ds dt = \int_0^y \left[\left(\frac{1}{2}s^2 + st\right)|_0^x\right] dt = \int_0^y \left(\frac{1}{2}x^2 + xt\right) dt \\ &= \left(\frac{1}{2}x^2 t + \frac{1}{2}xt^2\right)|_0^y \\ &= \frac{1}{2}\left(x^2 y + xy^2\right) \quad \text{for } 0 < x < 1,\ 0 < y < 1 \end{aligned}$$

$$\begin{aligned}
P(X \leq x, Y \leq y) &= \int_0^x \int_0^1 (s+t)\, dt ds = \frac{1}{2}\left(x^2 + x\right) \quad \text{for } 0 < x < 1,\, y \geq 1 \\
P(X \leq x, Y \leq y) &= \int_0^y \int_0^1 (s+t)\, ds dt = \frac{1}{2}\left(y^2 + y\right) \quad \text{for } x \geq 1,\, 0 < y < 1
\end{aligned}$$

the joint cumulative distribution function of X and Y is

$$F(x,y) = P(X \leq x, Y \leq y) = \begin{cases} 0 & x \leq 0 \text{ or } y \leq 0 \\ \frac{1}{2}\left(x^2 y + x y^2\right) & 0 < x < 1,\, 0 < y < 1 \\ \frac{1}{2}\left(x^2 + x\right) & 0 < x < 1,\, y \geq 1 \\ \frac{1}{2}\left(y^2 + y\right) & x \geq 1,\, 0 < y < 1 \\ 1 & x \geq 1,\, y \geq 1 \end{cases}$$

(c) Since the support set of (X,Y) is $A = \{(x,y) : 0 < x < 1,\ 0 < y < 1\}$ then

$$\begin{aligned}
F_1(x) &= P(X \leq x) \\
&= \lim_{y\to\infty} F(x,y) = \lim_{y\to\infty} P(X \leq x, Y \leq y) \\
&= F(x,1) \\
&= \frac{1}{2}\left(x^2 + x\right) \quad \text{for } 0 < x < 1
\end{aligned}$$

Alternatively

$$\begin{aligned}
F_1(x) &= P(X \leq x) = \int_{-\infty}^x f_1(s)\, ds = \int_0^x \left(s + \frac{1}{2}\right) ds \\
&= \frac{1}{2}\left(x^2 + x\right) \quad \text{for } 0 < x < 1
\end{aligned}$$

In either case the marginal cumulative distribution function of X is

$$F_1(x) = \begin{cases} 0 & x \leq 0 \\ \frac{1}{2}\left(x^2 + x\right) & 0 < x < 1 \\ 1 & x \geq 1 \end{cases}$$

By symmetry the marginal cumulative distribution function of Y is

$$F_2(y) = \begin{cases} 0 & y \leq 0 \\ \frac{1}{2}\left(y^2 + y\right) & 0 < y < 1 \\ 1 & y \geq 1 \end{cases}$$

3.3.6 Exercise

Suppose X and Y are continuous random variables with joint probability density function

$$f(x, y) = \frac{k}{(1 + x + y)^3} \quad \text{for } x \geq 0,\ y \geq 0$$

and 0 otherwise.

(a) Determine k and sketch $f(x, y)$.

(b) Find

(i) $P(X \leq 1, Y \leq 2)$
(ii) $P(X \leq Y)$
(iii) $P(X + Y \leq 1)$

(c) Determine the marginal probability density function of X and the marginal probability density function of Y.

(d) Determine the joint cumulative distribution function of X and Y.

(e) Determine the marginal cumulative distribution function of X and the marginal cumulative distribution function of Y.

3.3.7 Exercise

Suppose X and Y are continuous random variables with joint probability density function

$$f(x, y) = ke^{-x-y} \quad \text{for } y \geq x \geq 0$$

and 0 otherwise.

(a) Determine k and sketch $f(x, y)$.

(b) Find

(i) $P(X \leq 1, Y \leq 2)$
(ii) $P(X \leq Y)$
(iii) $P(X + Y \leq 1)$

(c) Determine the marginal probability density function of X and the marginal probability density function of Y.

(d) Determine the joint cumulative distribution function of X and Y.

(e) Determine the marginal cumulative distribution function of X and the marginal cumulative distribution function of Y.

3.4 Independent Random Variables

Suppose we are modeling a phenomenon involving two random variables. For example suppose X is your final mark in this course and Y is the time you spent doing practice problems. We would be interested in whether the distribution of one random variable affects the distribution of the other. The following definition defines this idea precisely. The definition should remind you of the definition of independent events (see Definition 2.1.8).

3.4.1 Definition - Independent Random Variables

Two random variables X and Y are called *independent random variables* if and only if

$$P(X \in A \text{ and } Y \in B) = P(X \in A)P(Y \in B)$$

for all sets A and B of real numbers.

Definition 3.4.1 is not very convenient for determining the independence of two random variables. The following theorem shows how to use the marginal and joint cumulative distribution functions or the marginal and joint probability (density) functions to determine if two random variables are independent.

3.4.2 Theorem - Independent Random Variables

(1) Suppose X and Y are random variables with joint cumulative distribution function $F(x, y)$. Suppose also that $F_1(x)$ is the marginal cumulative distribution function of X and $F_2(y)$ is the marginal cumulative distribution function of Y. Then X and Y are independent random variables if and only if

$$F(x, y) = F_1(x) F_2(y) \quad \text{for all } (x, y) \in \Re^2$$

(2) Suppose X and Y are random variables with joint probability (density) function $f(x, y)$. Suppose also that $f_1(x)$ is the marginal probability (density) function of X with support set $A_1 = \{x : f_1(x) > 0\}$ and $f_2(y)$ is the marginal probability (density) function of Y with support set $A_2 = \{y : f_2(y) > 0\}$. Then X and Y are independent random variables if and only if

$$f(x, y) = f_1(x) f_2(y) \quad \text{for all } (x, y) \in A_1 \times A_2$$

where $A_1 \times A_2 = \{(x, y) : x \in A_1, y \in A_2\}$.

Proof

(1) For given (x, y), let $A_x = \{s : s \leq x\}$ and let $B_y = \{t : t \leq y\}$. Then by Definition 3.4.1 X and Y are independent random variables if and only if

$$P(X \in A_x \text{ and } Y \in B_y) = P(X \in A_x) P(Y \in B_y)$$

for all $(x, y) \in \Re^2$.

But

$$P(X \in A_x \text{ and } Y \in B_y) = P(X \leq x, Y \leq y) = F(x, y)$$

$$P(X \in A_x) = P(X \leq x) = F_1(x)$$

and

$$P(Y \in B_y) = F_2(y)$$

Therefore X and Y are independent random variables if and only if

$$F(x, y) = F_1(x) F_2(y) \quad \text{for all } (x, y) \in \Re^2$$

as required.

(2) **(Continuous Case)** From (1) we have X and Y are independent random variables if and only if

$$F(x, y) = F_1(x) F_2(y) \tag{3.1}$$

for all $(x, y) \in \Re^2$. Now $\frac{\partial}{\partial x} F_1(x)$ exists for $x \in A_1$ and $\frac{\partial}{\partial y} F_2(y)$ exists for $y \in A_2$. Taking the partial derivative $\frac{\partial^2}{\partial x \partial y}$ of both sides of (3.1) where the partial derivative exists implies that X and Y are independent random variables if and only if

$$\frac{\partial^2}{\partial x \partial y} F(x, y) = \frac{\partial}{\partial x} F_1(x) \frac{\partial}{\partial y} F_2(y) \quad \text{for all } (x, y) \in A_1 \times A_2$$

or

$$f(x, y) = f_1(x) f_2(y) \quad \text{for all } (x, y) \in A_1 \times A_2$$

as required.

Note: The discrete case can be proved using an argument similar to the one used for (1).

3.4.3 Example

(a) In Example 3.2.4 determine if X and Y are independent random variables.

(b) In Example 3.3.3 determine if X and Y are independent random variables.

Solution

(a) Since the total number of students is fixed, a larger number of actuarial science students would imply a smaller number of statistics students and we would guess that the random variables are not independent. To show this we only need to find one pair of values (x, y)

for which $P(X=x,Y=y) \neq P(X=x)P(Y=y)$. Since

$$\begin{aligned}
P(X=0,Y=0) &= \frac{\binom{10}{0}\binom{9}{0}\binom{6}{5}}{\binom{25}{5}} = \frac{\binom{6}{5}}{\binom{25}{5}} \\
P(X=0) &= \frac{\binom{10}{0}\binom{15}{5}}{\binom{25}{5}} = \frac{\binom{15}{5}}{\binom{25}{5}} \\
P(Y=0) &= \frac{\binom{9}{0}\binom{16}{5}}{\binom{25}{5}} = \frac{\binom{16}{5}}{\binom{25}{5}}
\end{aligned}$$

and

$$P(X=0,Y=0) = \frac{\binom{6}{5}}{\binom{25}{5}} \neq P(X=0)P(Y=0) = \frac{\binom{15}{5}}{\binom{25}{5}}\frac{\binom{16}{5}}{\binom{25}{5}}$$

therefore by Theorem 3.4.2, X and Y are not independent random variables.

(b) Since

$$f(x,y) = x+y \quad \text{for } 0<x<1, 0<y<1$$

$$f_1(x) = x + \frac{1}{2} \quad \text{for } 0<x<1, \quad f_2(y) = y + \frac{1}{2} \quad \text{for } 0<y<1$$

it would appear that X and Y are not independent random variables. To show this we only need to find one pair of values (x,y) for which $f(x,y) \neq f_1(x)f_2(y)$. Since

$$f\left(\frac{2}{3},\frac{1}{3}\right) = 1 \neq f_1\left(\frac{2}{3}\right)f_2\left(\frac{1}{3}\right) = \frac{7}{6}\left(\frac{5}{6}\right)$$

therefore by Theorem 3.4.2, X and Y are not independent random variables.

3.4.4 Exercise

In Exercises 3.2.5, 3.3.6 and 3.3.7 determine if X and Y independent random variables.

In the previous examples we determined whether the random variables were independent using the joint probability (density) function and the marginal probability (density) functions. The following very useful theorem does not require us to determine the marginal probability (density) functions.

3.4.5 Factorization Theorem for Independence

Suppose X and Y are random variables with joint probability (density) function $f(x, y)$. Suppose also that A is the support set of (X, Y), A_1 is the support set of X, and A_2 is the support set of Y. Then X and Y are independent random variables if and only if there exist non-negative functions $g(x)$ and $h(y)$ such that

$$f(x, y) = g(x)\, h(y) \quad \text{for all } (x, y) \in A_1 \times A_2$$

Notes:

(1) If the Factorization Theorem for Independence holds then the marginal probability (density) function of X will be proportional to g and the marginal probability (density) function of Y will be proportional to h.

(2) Whenever the support set A is not rectangular the random variables will not be independent. The reason for this is that when the support set is not rectangular it will always be possible to find a point (x, y) such that $x \in A_1$ with $f_1(x) > 0$, and $y \in A_2$ with $f_2(y) > 0$ so that $f_1(x) f_2(y) > 0$, but $(x, y) \notin A$ so $f(x, y) = 0$. This means there is a point (x, y) such that $f(x, y) \neq f_1(x) f_2(y)$ and therefore X and Y are not independent random variables.

(3) The above definitions and theorems can easily be extended to the random vector $(X_1, X_2, \ldots, X_n)$.

Proof (Continuous Case)

If X and Y are independent random variables then by Theorem 3.4.2

$$f(x, y) = f_1(x)\, f_2(y) \quad \text{for all } (x, y) \in A_1 \times A_2$$

Letting $g(x) = f_1(x)$ and $h(y) = f_2(y)$ proves there exist $g(x)$ and $h(y)$ such that

$$f(x, y) = g(x)\, h(y) \quad \text{for all } (x, y) \in A_1 \times A_2$$

If there exist non-negative functions $g(x)$ and $h(y)$ such that

$$f(x, y) = g(x)\, h(y) \quad \text{for all } (x, y) \in A_1 \times A_2$$

then

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y) dy = \int_{y \in A_2} g(x)\, h(y)\, dy = cg(x) \quad \text{for } x \in A_1$$

and

$$f_2(y) = \int_{-\infty}^{\infty} f(x, y) dx = \int_{x \in A_1} g(x)\, h(y)\, dy = kh(y) \quad \text{for } y \in A_2$$

Now

$$\begin{aligned}
1 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy \\
&= \int_{y\in A_2}\int_{x\in A_1} f_1(x)\, f_2(y)\, dxdy \\
&= \int_{y\in A_2}\int_{x\in A_1} cg(x)\, kh(y)\, dxdy \\
&= ck
\end{aligned}$$

Since $ck = 1$

$$\begin{aligned}
f(x,y) &= g(x)\, h(y) = ckg(x)\, h(y) \\
&= cg(x)\, kh(y) \\
&= f_1(x)\, f_2(y) \quad \text{for all } (x,y) \in A_1 \times A_2
\end{aligned}$$

and by Theorem 3.4.2 X and Y are independent random variables.

3.4.6 Example

Suppose X and Y are discrete random variables with joint probability function

$$f(x,y) = \frac{\theta^{x+y}e^{-2\theta}}{x!y!} \quad \text{for } x = 0, 1\ldots, \quad y = 0, 1, \ldots$$

(a) Determine if X and Y independent are random variables.

(b) Determine the marginal probability function of X and the marginal probability function of Y.

Solution

(a) The support set of (X, Y) is $A = \{(x,y) : x = 0, 1, \ldots;\ y = 0, 1, \ldots\}$ which is rectangular. The support set of X is $A_1 = \{x : x = 0, 1, \ldots\}$, and the support set of Y is $A_2 = \{y : y = 0, 1, \ldots\}$.

Let

$$g(x) = \frac{\theta^x e^{-\theta}}{x!} \quad \text{and} \quad h(y) = \frac{\theta^y e^{-\theta}}{y!}$$

Then $f(x,y) = g(x)h(y)$ for all $(x,y) \in A_1 \times A_2$. Therefore by the Factorization Theorem for Independence X and Y are independent random variables.

(b) By inspection we can see that $g(x)$ is the probability function for a Poisson(θ) random variable. Therefore the marginal probability function of X is

$$f_1(x) = \frac{\theta^x e^{-\theta}}{x!} \quad \text{for } x = 0, 1\ldots$$

Similarly the marginal probability function of Y is

$$f_2(y) = \frac{\theta^y e^{-\theta}}{y!} \quad \text{for } y = 0, 1 \ldots$$

and $Y \sim \text{Poisson}(\theta)$.

3.4.7 Example

Suppose X and Y are continuous random variables with joint probability density function

$$f(x,y) = \frac{3}{2} y \left(1 - x^2\right) \quad \text{for } -1 < x < 1,\ 0 < y < 1$$

and 0 otherwise.

(a) Determine if X and Y independent are random variables.

(b) Determine the marginal probability function of X and the marginal probability function of Y.

Solution

(a) The support set of (X,Y) is $A = \{(x,y) : -1 < x < 1,\ 0 < y < 1\}$ which is rectangular. The support set of X is $A_1 = \{x : -1 < x < 1\}$, and the support set of Y is $A_2 = \{y : 0 < y < 1\}$.

Let

$$g(x) = 1 - x^2 \text{ and } h(y) = \frac{3}{2} y$$

Then $f(x,y) = g(x)h(y)$ for all $(x,y) \in A_1 \times A_2$. Therefore by the Factorization Theorem for Independence X and Y are independent random variables.

(b) Since the marginal probability density function of Y is proportional to $h(y)$ we know $f_2(y) = kh(y)$ for $0 < y < 1$ where k is determined by

$$1 = k \int_0^1 \frac{3}{2} y dy = \frac{3k}{4} y^2 |_0^1 = \frac{3k}{4}$$

Therefore $k = \frac{4}{3}$ and

$$f_2(y) = 2y \quad \text{for } 0 < y < 1$$

and 0 otherwise.

Since X and Y are independent random variables $f(x,y) = f_1(x) f_2(y)$ or $f_1(x) = f(x,y)/f_2(y)$ for $x \in A_1$. Therefore the marginal probability density function of X is

$$\begin{aligned} f_1(x) &= \frac{f(x,y)}{f_2(y)} = \frac{\frac{3}{2} y \left(1 - x^2\right)}{2y} \\ &= \frac{3}{4} \left(1 - x^2\right) \quad \text{for } -1 < x < 1 \end{aligned}$$

and 0 otherwise.

3.4.8 Example

Suppose X and Y are continuous random variables with joint probability density function

$$f(x,y) = \frac{2}{\pi} \quad \text{for } 0 < x < \sqrt{1-y^2}, \quad -1 < y < 1$$

and 0 otherwise.

(a) Determine if X and Y independent are random variables.

(b) Determine the marginal probability function of X and the marginal probability function of Y.

Solution

(a) The support set of (X,Y) which is

$$A = \left\{(x,y) : 0 < x < \sqrt{1-y^2}, \ -1 < y < 1\right\}$$

is graphed in Figure 3.7.

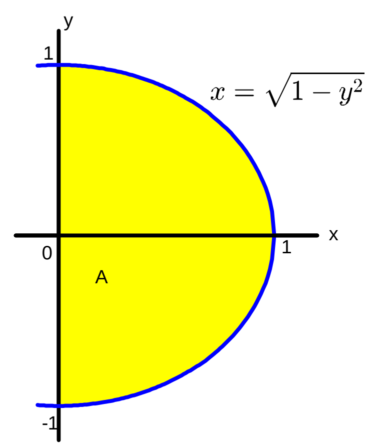


Figure 3.7: Graph of the support set of (X,Y) for Example 3.4.8

The region A can also be described as

$$A = \left\{(x,y) : 0 < x < 1, \ -\sqrt{1-x^2} < y < \sqrt{1-x^2}\right\}$$

The region A is half of a unit circle which has area equal to $\pi/2$. Since the probability density function is constant or uniform on this region we can see that $f(x,y)$ must equal $2/\pi$ on the region A since the volume of the solid must be equal to 1.

The support set for X is

$$A_1 = \{x : 0 < x < 1\}$$

and the support set for Y is

$$A_2 = \{y : -1 < y < 1\}$$

If we choose the point $(0.9, 0.9) \in A_1 \times A_2$, $f(0.9, 0.9) = 0$ but $f_1(0.9) > 0$ and $f_2(0.9) > 0$ so $f(0.9, 0.9) \neq f_1(0.9) f_2(0.9)$ and therefore X and Y are not independent random variables.

(b) When the support set is not rectangular care must be taken to determine the marginal probability functions.

To find the marginal probability density function of X we use the description of the support set in which the range of X does not depend on y which is

$$A = \left\{(x, y) : 0 < x < 1, -\sqrt{1 - x^2} < y < \sqrt{1 - x^2}\right\}$$

The marginal probability density function of X is

$$\begin{aligned} f_1(x) &= \int_{-\infty}^{\infty} f(x, y) dy \\ &= \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{2}{\pi} dy \\ &= \frac{4}{\pi}\sqrt{1 - x^2} \quad \text{for } 0 < x < 1 \end{aligned}$$

and 0 otherwise.

To find the marginal probability density function of Y which use the description of the support set in which the range of Y does not depend on x which is

$$A = \left\{(x, y) : 0 < x < \sqrt{1 - y^2}, -1 < y < 1\right\}$$

The marginal probability density function of Y is

$$\begin{aligned} f_2(y) &= \int_{-\infty}^{\infty} f(x, y) dx \\ &= \int_{0}^{\sqrt{1-y^2}} \frac{2}{\pi} dx \\ &= \frac{2}{\pi}\sqrt{1 - y^2} \quad \text{for } -1 < y < 1 \end{aligned}$$

and 0 otherwise.

3.5 Conditional Distributions

In Section 2.1 we defined the conditional probability of event A given event B as

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{provided } P(B) > 0$$

The concept of conditional probability can also be extended to random variables.

3.5.1 Definition - Conditional Probability (Density) Function

Suppose X and Y are random variables with joint probability (density) function $f(x, y)$, and marginal probability (density) functions $f_1(x)$ and $f_2(y)$ respectively. Suppose also that the support set of (X, Y) is $A = \{(x, y) : f(x, y) > 0\}$.

The *conditional probability (density) function of* X *given* $Y = y$ is given by

$$f_1(x|y) = \frac{f(x, y)}{f_2(y)} \tag{3.2}$$

for $(x, y) \in A$ provided $f_2(y) \neq 0$.

The *conditional probability (density) function of* Y *given* $X = x$ is given by

$$f_2(y|x) = \frac{f(x, y)}{f_1(x)} \tag{3.3}$$

for $(x, y) \in A$ provided $f_1(x) \neq 0$.

Notes:

(1) If X and Y are discrete random variables then

$$\begin{aligned} f_1(x|y) &= P(X = x|Y = y) \\ &= \frac{P(X = x, Y = y)}{P(Y = y)} \\ &= \frac{f(x, y)}{f_2(y)} \end{aligned}$$

and

$$\begin{aligned} \sum_x f_1(x|y) &= \sum_x \frac{f(x, y)}{f_2(y)} \\ &= \frac{1}{f_2(y)} \sum_x f(x, y) \\ &= \frac{f_2(y)}{f_2(y)} \\ &= 1 \end{aligned}$$

Similarly for $f_2(y|x)$.

(2) If X and Y are continuous random variables

$$\begin{aligned}
\int_{-\infty}^{\infty} f_1(x|y)dx &= \int_{-\infty}^{\infty} \frac{f(x,y)}{f_2(y)}dx \\
&= \frac{1}{f_2(y)} \int_{-\infty}^{\infty} f(x,y)dx \\
&= \frac{f_2(y)}{f_2(y)} \\
&= 1
\end{aligned}$$

Similarly for $f_2(y|x)$.

(3) If X is a continuous random variable then $f_1(x) \neq P(X = x)$ and $P(X = x) = 0$ for all x. Therefore to justify the definition of the conditional probability density function of Y given $X = x$ when X and Y are continuous random variables we consider $P(Y \le y|X = x)$ as a limit

$$\begin{aligned}
&P(Y \le y|X = x) = \lim_{h \to 0} P(Y \le y|x \le X \le x + h) \\
&= \lim_{h\to 0} \frac{\int\limits_{x}^{x+h} \int\limits_{-\infty}^{y} f(u,v)\, dv du}{\int\limits_{x}^{x+h} f_1(u)\, du} \\
&= \lim_{h\to 0} \frac{\frac{d}{dh} \int\limits_{x}^{x+h} \int\limits_{-\infty}^{y} f(u,v)\, dv du}{\frac{d}{dh} \int\limits_{x}^{x+h} f_1(u)\, du} \quad \text{by L'Hôpital's Rule} \\
&= \lim_{h\to 0} \frac{\int\limits_{-\infty}^{y} f(x+h,v)\, dv}{f_1(x+h)} \quad \text{by the Fundamental Theorem of Calculus} \\
&= \frac{\lim_{h\to 0} \int\limits_{-\infty}^{y} f(x+h,v)\, dv}{\lim_{h\to 0} f_1(x+h)} \\
&= \frac{\int\limits_{-\infty}^{y} f(x,v)\, dv}{f_1(x)}
\end{aligned}$$

assuming that the limits exist and that integration and the limit operation can be interchanged. If we differentiate the last term with respect to y using the Fundamental Theorem of Calculus we have

$$\frac{d}{dy} P(Y \le y|X = x) = \frac{f(x,y)}{f_1(x)}$$

which gives us a justification for using

$$f_2(y|x) = \frac{f(x,y)}{f_1(x)}$$

as the conditional probability density function of Y given $X = x$.

(4) For a given value of x, call it x^*, we can think of obtaining the conditional probability density function of Y given $X = x^*$ geometrically in the following way. Think of the curve of intersection which is obtained by cutting through the surface $z = f(x,y)$ with the plane $x = x^*$ which is parallel to the yz plane. The curve of intersection is $z = f(x^*, y)$ which is a curve lying in the plane $x = x^*$. The area under the curve $z = f(x^*, y)$ and lying above the xy plane is not necessarily equal to 1 and therefore $z = f(x^*, y)$ is not a proper probability density function. However if we consider the curve $z = f(x^*, y) / f_1(x^*)$, which is just a rescaled version of the curve $z = f(x^*, y)$, then the area lying under the curve $z = f(x^*, y) / f_1(x^*)$ and lying above the xy plane is equal to 1. This is the probability density function we want.

3.5.2 Example

In Example 3.4.8 determine the conditional probability density function of X given $Y = y$ and the conditional probability density function of Y given $X = x$.

Solution

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

$$\begin{aligned} f_1(x|y) &= \frac{f(x,y)}{f_2(y)} \\ &= \frac{\frac{2}{\pi}}{\frac{2}{\pi}\sqrt{1-y^2}} \\ &= \frac{1}{\sqrt{1-y^2}} \quad \text{for } 0 < x < \sqrt{1-y^2}, \ -1 < y < 1 \end{aligned}$$

Note that for each $y \in (-1, 1)$, the conditional probability density function of X given $Y = y$ is Uniform$\left(0, \sqrt{1-y^2}\right)$. This makes sense because the joint probability density function is constant on its support set.

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

$$\begin{aligned} f_2(y|x) &= \frac{f(x,y)}{f_1(x)} \\ &= \frac{\frac{2}{\pi}}{\frac{4}{\pi}\sqrt{1-x^2}} \\ &= \frac{1}{2\sqrt{1-x^2}} \quad \text{for } -\sqrt{1-x^2} < y < \sqrt{1-x^2}, \ 0 < x < 1 \end{aligned}$$

Note that for each $x \in (0,1)$, the conditional probability density function of Y given $X = x$ is Uniform$\left(-\sqrt{1-x^2}, \sqrt{1-x^2}\right)$. This again makes sense because the joint probability density function is constant on its support set.

3.5.3 Exercise

In Exercise 3.2.5 show that the conditional probability function of Y given $X = x$ is

$$\text{Binomial}\left(n-x,\ \frac{2\theta(1-\theta)}{1-\theta^2}\right)$$

Why does this make sense?

3.5.4 Exercise

In Example 3.3.3 and Exercises 3.3.6 and 3.3.7 determine the conditional probability density function of X given $Y = y$ and the conditional probability density function of Y given $X = x$. Be sure to check that

$$\int_{-\infty}^{\infty} f_1(x|y)dx = 1 \quad \text{and} \quad \int_{-\infty}^{\infty} f_2(y|x)dy = 1$$

When choosing a model for bivariate data it is sometimes easier to specify a conditional probability (density) function and a marginal probability (density) function. The joint probability (density) function can then be determined using the Product Rule which is obtained by rewriting (3.2) and (3.3).

3.5.5 Product Rule

Suppose X and Y are random variables with joint probability (density) function $f(x,y)$, marginal probability (density) functions $f_1(x)$ and $f_2(y)$ respectively and conditional probability (density) function's $f_1(x|y)$ and $f_2(y|x)$. Then

$$\begin{aligned} f(x,y) &= f_1(x|y)f_2(y) \\ &= f_2(y|x)f_1(x) \end{aligned}$$

3.5.6 Example

In modeling survival in a certain insect population it is assumed that the number of eggs laid by a single female follows a Poisson(μ) distribution. It is also assumed that each egg has probability p of surviving independently of any other egg. Determine the probability function of the number of eggs that survive.

Solution

Let Y = number of eggs laid and let X = number of eggs that survive. Then $Y \sim$ Poisson(μ) and $X|Y = y \sim$ Binomial(y, p). We want to determine the marginal probability function of X.

By the Product Rule the joint probability function of X and Y is

$$\begin{aligned} f(x, y) &= f_1(x|y) f_2(y) \\ &= \binom{y}{x} p^x (1-p)^{y-x} \frac{\mu^y e^{-\mu}}{y!} \\ &= \frac{p^x e^{-\mu}}{x!} \frac{(1-p)^{y-x} \mu^y}{(y-x)!} \end{aligned}$$

with support set is

$$A = \{(x, y) : x = 0, 1, \ldots, y;\ y = 0, 1, \ldots\}$$

which can also be written as

$$A = \{(x, y) : y = x, x+1, \ldots;\ x = 0, 1, \ldots\} \tag{3.4}$$

The marginal probability function of X can be obtained using

$$f_1(x) = \sum_{all\ y} f(x, y)$$

Since we are summing over y we need to use the second description of the support set given in (3.4). So

$$\begin{aligned} f_1(x) &= \sum_{y=x}^{\infty} \frac{p^x e^{-\mu}}{x!} \frac{(1-p)^{y-x} \mu^y}{(y-x)!} \\ &= \frac{p^x e^{-\mu} \mu^x}{x!} \sum_{y=x}^{\infty} \frac{(1-p)^{y-x} \mu^{y-x}}{(y-x)!} \quad \text{let } u = y - x \\ &= \frac{p^x e^{-\mu} \mu^x}{x!} \sum_{u=0}^{\infty} \frac{[\mu(1-p)]^u}{u!} \\ &= \frac{p^x e^{-\mu} \mu^x}{x!} e^{\mu(1-p)} \quad \text{by the Exponential series 2.11.7} \\ &= \frac{(p\mu)^x e^{-p\mu}}{x!} \quad \text{for } x = 0, 1, \ldots \end{aligned}$$

which we recognize as a Poisson$(p\mu)$ probability function.

3.5.7 Example

Determine the marginal probability function of X if $Y \sim$ Gamma$(\alpha, \frac{1}{\theta})$ and the conditional distribution of X given $Y = y$ is Weibull$(p, y^{-1/p})$.

Solution

Since $Y \sim$ Gamma$(\alpha, \frac{1}{\theta})$

$$f_2(y) = \frac{\theta^{\alpha} y^{\alpha-1} e^{-\theta y}}{\Gamma(\alpha)} \quad \text{for } y > 0$$

and 0 otherwise.

Since the conditional distribution of X given $Y = y$ is Weibull$(p, y^{-1/p})$

$$f_1(x|y) = pyx^{p-1}e^{-yx^p} \quad \text{for } x > 0$$

By the Product Rule the joint probability density function of X and Y is

$$\begin{aligned} f(x, y) &= f_1(x|y) f_2(y) \\ &= pyx^{p-1}e^{-yx^p} \frac{\theta^{\alpha} y^{\alpha-1} e^{-\theta y}}{\Gamma(\alpha)} \\ &= \frac{p\theta^{\alpha} x^{p-1}}{\Gamma(\alpha)} y^{\alpha} e^{-y(\theta + x^p)} \end{aligned}$$

The support set is

$$A = \{(x, y) : x > 0;\ y > 0\}$$

which is a rectangular region.

The marginal probability function of X is

$$\begin{aligned} f_1(x) &= \int_{-\infty}^{\infty} f(x, y)\, dy \\ &= \frac{p\theta^{\alpha} x^{p-1}}{\Gamma(\alpha)} \int_0^{\infty} y^{\alpha} e^{-y(\theta+x^p)} dy \quad \text{let } u = y(\theta + x^p) \\ &= \frac{p\theta^{\alpha} x^{p-1}}{\Gamma(\alpha)} \int_0^{\infty} \left(\frac{u}{\theta + x^p}\right)^{\alpha} e^{-u} \left(\frac{1}{\theta + x^p}\right) du \\ &= \frac{p\theta^{\alpha} x^{p-1}}{\Gamma(\alpha)} \left(\frac{1}{\theta + x^p}\right)^{\alpha+1} \int_0^{\infty} u^{\alpha} e^{-u} du \\ &= \frac{p\theta^{\alpha} x^{p-1}}{\Gamma(\alpha)} \left(\frac{1}{\theta + x^p}\right)^{\alpha+1} \Gamma(\alpha + 1) \quad \text{by 2.4.8} \\ &= \frac{\alpha p\theta^{\alpha} x^{p-1}}{(\theta + x^p)^{\alpha+1}} \quad \text{since } \Gamma(\alpha + 1) = \alpha\Gamma(\alpha) \end{aligned}$$

for $x > 0$ and 0 otherwise. This distribution is a member of the Burr family of distributions which is frequently used by actuaries for modeling household income, crop prices, insurance risk, and many other financial variables.

The following theorem gives us one more method for determining whether two random variables are independent.

3.5.8 Theorem

Suppose X and Y are random variables with marginal probability (density) functions $f_1(x)$ and $f_2(y)$ respectively and conditional probability (density) functions $f_1(x|y)$ and $f_2(y|x)$. Suppose also that A_1 is the support set of X, and A_2 is the support set of Y. Then X and Y are independent random variables if and only if either of the following holds

$$f_1(x|y) = f_1(x) \quad \text{for all } x \in A_1$$

or

$$f_2(y|x) = f_2(y) \quad \text{for all } y \in A_2$$

3.5.9 Example

Suppose the conditional distribution of X given $Y = y$ is

$$f_1(x|y) = \frac{e^{-x}}{1 - e^{-y}} \quad \text{for } 0 < x < y$$

and 0 otherwise. Are X and Y independent random variables?

Solution

Since the conditional distribution of X given $Y = y$ depends on y then $f_1(x|y) = f_1(x)$ cannot hold for all x in the support set of X and therefore X and Y are not independent random variables.

3.6 Joint Expectations

As with univariate random variables we define the expectation operator for bivariate random variables. The discrete case is a review of material you would have seen in a previous probability course.

3.6.1 Definition - Joint Expectation

Suppose $h(x, y)$ is a real-valued function.

If X and Y are discrete random variables with joint probability function $f(x, y)$ and support set A then

$$E[h(X,Y)] = \sum_{(x,y)} \sum_{\in A} h(x,y) f(x,y)$$

provided the joint sum converges absolutely.

If X and Y are continuous random variables with joint probability density function $f(x,y)$ then

$$E[h(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,y)f(x,y)dxdy$$

provided the joint integral converges absolutely.

3.6.2 Theorem

Suppose X and Y are random variables with joint probability (density) function $f(x,y)$, a and b are real constants, and $g(x,y)$ and $h(x,y)$ are real-valued functions. Then

$$E[ag(X,Y)+bh(X,Y)] = aE[g(X,Y)] + bE[h(X,Y)]$$

Proof (Continuous Case)

$$\begin{aligned}
E[ag(X,Y)+bh(X,Y)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [ag(x,y)+bh(x,y)]\, f(x,y)dxdy \\
&= a\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)dxdy + b\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,y)f(x,y)dxdy \\
&\quad \text{by properties of double integrals} \\
&= aE[g(X,Y)] + bE[h(X,Y)] \quad \text{by Definition 3.6.1}
\end{aligned}$$

3.6.3 Corollary

(1)

$$E(aX+bY) = aE(X) + bE(Y) = a\mu_X + b\mu_Y$$

where $\mu_X = E(X)$ and $\mu_Y = E(Y)$.

(2) If $X_1, X_2, \ldots, X_n$ are random variables and $a_1, a_2, \ldots, a_n$ are real constants then

$$E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i E(X_i) = \sum_{i=1}^{n} a_i \mu_i$$

where $\mu_i = E(X_i)$.

(3) If $X_1, X_2, \ldots, X_n$ are random variables with $E(X_i) = \mu$, $i = 1, 2, \ldots, n$ then

$$E\left(\bar{X}\right) = \frac{1}{n}\sum_{i=1}^{n} E\left(X_i\right) = \frac{1}{n}\sum_{i=1}^{n} \mu = \frac{n\mu}{n} = \mu$$

Proof of (1) (Continuous Case)

$$\begin{aligned}
E(aX+bY) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(ax+by)\,f(x,y)dxdy \\
&= a\int_{-\infty}^{\infty} x\left[\int_{-\infty}^{\infty} f(x,y)dy\right]dx + b\int_{-\infty}^{\infty} y\left[\int_{-\infty}^{\infty} f(x,y)dx\right]dy \\
&= a\int_{-\infty}^{\infty} xf_1(x)\,dx + b\int_{-\infty}^{\infty} yf_2(y)\,dy \\
&= aE(X)+bE(Y) = a\mu_X + b\mu_Y
\end{aligned}$$

3.6.4 Theorem - Expectation and Independence

(1) If X and Y are independent random variables and $g(x)$ and $h(y)$ are real valued functions then

$$E\left[g\left(X\right)h\left(Y\right)\right] = E\left[g\left(X\right)\right]E\left[h\left(Y\right)\right]$$

(2) More generally if $X_1, X_2, \ldots, X_n$ are independent random variables and $h_1, h_2, \ldots, h_n$ are real valued functions then

$$E\left[\prod_{i=1}^{n} h_i(X_i)\right] = \prod_{i=1}^{n} E\left[h_i(X_i)\right]$$

Proof of (1) (Continuous Case)

Since X and Y are independent random variables then by Theorem 3.4.2

$$f(x,y) = f_1(x)\,f_2(y) \quad \text{for all } (x,y)\in A$$

where A is the support set of (X,Y). Therefore

$$\begin{aligned}
E\left[g(X)h(Y)\right] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)f_1(x)\,f_2(y)\,dxdy \\
&= \int_{-\infty}^{\infty} h(y)f_2(y)\left[\int_{-\infty}^{\infty} g(x)f_1(x)\,dx\right]dy \\
&= E\left[g(X)\right]\int_{-\infty}^{\infty} h(y)f_2(y)\,dy \\
&= E\left[g(X)\right]E\left[h(Y)\right]
\end{aligned}$$

3.6.5 Definition - Covariance

The *covariance* of random variables X and Y is defined by

$$Cov(X,Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

If $Cov(X,Y) = 0$ then X and Y are called *uncorrelated* random variables.

3.6.6 Theorem - Covariance and Independence

If X and Y are random variables then

$$Cov(X,Y) = E(XY) - \mu_X \mu_Y$$

If X and Y are independent random variables then $Cov(X,Y) = 0$.

Proof

$$\begin{aligned}
Cov(X,Y) &= E\left[(X - \mu_X)(Y - \mu_Y)\right] \\
&= E\left(XY - \mu_X Y - X\mu_Y + \mu_X \mu_Y\right) \\
&= E(XY) - \mu_X E(Y) - \mu_Y E(X) + \mu_X \mu_Y \\
&= E(XY) - E(X)E(Y) - E(Y)E(X) + E(X)E(Y) \\
&= E(XY) - E(X)E(Y)
\end{aligned}$$

Now if X and Y are independent random variables then by Theorem 3.6.4 $E(XY) = E(X)E(Y)$ and therefore $Cov(X,Y) = 0$.

3.6.7 Theorem - Variance of a Linear Combination

(1) Suppose X and Y are random variables and a and b are real constants then

$$\begin{aligned}
Var(aX + bY) &= a^2 Var(X) + b^2 Var(Y) + 2abCov(X,Y) \\
&= a^2 \sigma_X^2 + b^2 \sigma_Y^2 + 2abCov(X,Y)
\end{aligned}$$

(2) Suppose $X_1, X_2, \ldots, X_n$ are random variables with $Var(X_i) = \sigma_i^2$ and $a_1, a_2, \ldots, a_n$ are real constants then

$$Var\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 \sigma_i^2 + 2\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} a_i a_j Cov\left(X_i, X_j\right)$$

(3) If $X_1, X_2, \ldots, X_n$ are independents random variables and $a_1, a_2, \ldots, a_n$ are real constants then

$$Var\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 \sigma_i^2$$

(4) If $X_1, X_2, \ldots, X_n$ are independent random variables with $Var(X_i) = \sigma^2$, $i = 1, 2, \ldots, n$ then

$$\begin{aligned} Var\left(\bar{X}\right) &= Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \left(\frac{1}{n}\right)^2 \sum_{i=1}^{n} Var\left(X_i\right) \\ &= \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{n\sigma^2}{n^2} \\ &= \frac{\sigma^2}{n} \end{aligned}$$

Proof of (1)

$$\begin{aligned} Var\left(aX + bY\right) &= E\left[\left(aX + bY - a\mu_X - b\mu_Y\right)^2\right] \\ &= E\left\{\left[a\left(X - \mu_X\right) + b\left(Y - \mu_Y\right)\right]^2\right\} \\ &= E\left[a^2\left(X - \mu_X\right)^2 + b^2\left(Y - \mu_Y\right)^2 + 2ab\left(X - \mu_X\right)\left(Y - \mu_Y\right)\right] \\ &= a^2 E\left[\left(X - \mu_X\right)^2\right] + b^2 E\left[\left(Y - \mu_Y\right)^2\right] + 2abE\left[\left(X - \mu_X\right)\left(Y - \mu_Y\right)\right] \\ &= a^2\sigma_X^2 + b^2\sigma_Y^2 + 2abCov\left(X, Y\right) \end{aligned}$$

3.6.8 Definition - Correlation Coefficient

The *correlation coefficient* of random variables X and Y is defined by

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

3.6.9 Example

For the joint probability density function in Example 3.3.3 find $\rho(X, Y)$.

Solution

$$\begin{aligned} E\left(XY\right) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x, y)dxdy = \int_{0}^{1}\int_{0}^{1} xy\left(x + y\right)dxdy \\ &= \int_{0}^{1}\int_{0}^{1}\left(x^2y + xy^2\right)dxdy = \int_{0}^{1}\left(\frac{1}{3}x^3y + \frac{1}{2}x^2y^2\right)|_0^1 dy \\ &= \int_{0}^{1}\left(\frac{1}{3}y + \frac{1}{2}y^2\right)dy = \left(\frac{1}{6}y^2 + \frac{1}{6}y^3\right)|_0^1 = \frac{1}{6} + \frac{1}{6} \\ &= \frac{1}{3} \end{aligned}$$

$$
\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} x f_1(x) dx = \int_0^1 x\left(x+\frac{1}{2}\right) dx = \int_0^1 \left(x^2+\frac{1}{2}x\right) dx \\
&= \left(\frac{1}{3}x^3+\frac{1}{4}x^2\right)|_0^1 \\
&= \frac{1}{3}+\frac{1}{4}=\frac{7}{12}
\end{aligned}
$$

$$
\begin{aligned}
E\left(X^2\right) &= \int_{-\infty}^{\infty} x^2 f_1(x) dx = \int_0^1 x^2\left(x+\frac{1}{2}\right) dx = \int_0^1 \left(x^3+\frac{1}{2}x^2\right) dx \\
&= \left(\frac{1}{4}x^4+\frac{1}{6}x^3\right)|_0^1 \\
&= \frac{1}{4}+\frac{1}{6}=\frac{3+2}{12}=\frac{5}{12}
\end{aligned}
$$

$$
\begin{aligned}
Var(X) &= E\left(X^2\right)-[E(X)]^2=\frac{5}{12}-\left(\frac{7}{12}\right)^2 \\
&= \frac{60-49}{144}=\frac{11}{144}
\end{aligned}
$$

By symmetry

$$E(Y)=\frac{7}{12} \text{ and } Var(Y)=\frac{11}{144}$$

Therefore

$$Cov(X,Y)=E(XY)-E(X)E(Y)=\frac{1}{3}-\left(\frac{7}{12}\right)\left(\frac{7}{12}\right)=\frac{48-49}{144}=\frac{-1}{144}$$

and

$$
\begin{aligned}
\rho(X,Y) &= \frac{Cov(X,Y)}{\sigma_X\sigma_Y} \\
&= \frac{\frac{-1}{144}}{\sqrt{\left(\frac{11}{144}\right)\left(\frac{11}{144}\right)}} \\
&= \frac{-1}{144}\left(\frac{144}{11}\right) \\
&= -\frac{1}{11}
\end{aligned}
$$

3.6.10 Exercise

For the joint probability density function in Exercise 3.3.7 find $\rho(X,Y)$.

3.6.11 Theorem

If $\rho(X,Y)$ is the correlation coefficient of random variables X and Y then

$$-1 \leq \rho(X,Y) \leq 1$$

$\rho(X,Y) = 1$ if and only if $Y = aX + b$ for some $a > 0$ and $\rho(X,Y) = -1$ if and only if $Y = aX + b$ for some $a < 0$.

Proof

Let $S = X + tY$, where $t \in \Re$. Then $E(S) = \mu_S$ and

$$\begin{aligned} Var(S) &= E\left[(S-\mu_S)^2\right] \\ &= E\{[(X+tY) - (\mu_X + t\mu_Y)]^2\} \\ &= E\{[(X-\mu_X) + t(Y-\mu_Y)]^2\} \\ &= E\left[(X-\mu_X)^2 + 2t(X-\mu_X)(Y-\mu_Y) + t^2(Y-\mu_Y)^2\right] \\ &= t^2\sigma_Y^2 + 2Cov(X,Y)t + \sigma_X^2 \end{aligned}$$

Now $Var(S) \geq 0$ for any $t \in \Re$ implies that the quadratic equation $Var(S) = t^2\sigma_Y^2 + 2Cov(X,Y)t + \sigma_X^2$ in the variable t must have at most one real root. To have at most one real root the discrimant of this quadratic equation must be less than or equal to zero. Therefore

$$[2Cov(X,Y)]^2 - 4\sigma_X^2\sigma_Y^2 \leq 0$$

or

$$[Cov(X,Y)]^2 \leq \sigma_X^2\sigma_Y^2$$

or

$$|\rho(X,Y)| = \left|\frac{Cov(X,Y)}{\sigma_X\sigma_Y}\right| \leq 1$$

and therefore

$$-1 \leq \rho(X,Y) \leq 1$$

To see that $\rho(X,Y) = \pm 1$ corresponds to a linear relationship between X and Y, note that $\rho(X,Y) = \pm 1$ implies

$$|Cov(X,Y)| = \sigma_X\sigma_Y$$

and therefore

$$[2Cov(X,Y)]^2 - 4\sigma_X^2\sigma_Y^2 = 0$$

which corresponds to a zero discriminant in the quadratic equation. This means that there exists one real number t^* for which

$$Var(S) = Var(X + t^*Y) = 0$$

But $Var(X + t^*Y) = 0$ implies $X + t^*Y$ must equal a constant, that is, $X + t^*Y = c$. Thus X and Y satisfy a linear relationship.

3.7 Conditional Expectation

Since conditional probability (density) functions are also probability (density) function, expectations can be defined in terms of these conditional probability (density) functions as in the following definition.

3.7.1 Definition - Conditional Expectation

The *conditional expectation of* $g(X)$ *given* $Y = y$ is given by

$$E\left[g(X)|y\right] = \sum_{x} g(x) f_1(x|y)$$

if Y is a discrete random variable and

$$E[g(X)|y] = \int_{-\infty}^{\infty} g(x) f_1(x|y) dx$$

if Y is a continuous random variable provided the sum/integral converges absolutely.

The conditional expectation of $h(Y)$ given $X = x$ is defined in a similar manner.

3.7.2 Special Cases

(1) The *conditional mean of* X *given* $Y = y$ is denoted by $E\left(X|y\right)$.

(2) The *conditional variance of* X *given* $Y = y$ is denoted by $Var\left(X|y\right)$ and is given by

$$\begin{aligned} Var\left(X|y\right) &= E\left\{\left[X - E\left(X|y\right)\right]^2 |y\right\} \\ &= E\left(X^2|y\right) - \left[E\left(X|y\right)\right]^2 \end{aligned}$$

3.7.3 Example

For the joint probability density function in Example 3.4.8 find $E\left(Y|x\right)$ the conditional mean of Y given $X = x$, and $Var\left(Y|x\right)$ the conditional variance of Y given $X = x$.

Solution

From Example 3.5.2 we have

$$f_2(y|x) = \frac{1}{2\sqrt{1-x^2}} \quad \text{for } -\sqrt{1-x^2} < y < \sqrt{1-x^2},\ 0 < x < 1$$

Therefore

$$\begin{aligned}
E(Y|x) &= \int_{-\infty}^{\infty} y f_2(y|x)dy \\
&= \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} y \frac{1}{2\sqrt{1-x^2}} dy \\
&= \frac{1}{2\sqrt{1-x^2}} \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} y dy \\
&= \frac{1}{\sqrt{1-x^2}} \left(y^2 |_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \right) \\
&= 0
\end{aligned}$$

Since $E(Y|x) = 0$

$$\begin{aligned}
Var(Y|x) &= E\left(Y^2|x\right) = \int_{-\infty}^{\infty} y^2 f_2(y|x)dy \\
&= \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} y^2 \frac{1}{2\sqrt{1-x^2}} dy \\
&= \frac{1}{2\sqrt{1-x^2}} \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} y^2 dy \\
&= \frac{1}{6\sqrt{1-x^2}} \left(y^3 |_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \right) \\
&= \frac{1}{3}\left(1 - x^2\right)
\end{aligned}$$

Recall that the conditional distribution of Y given $X = x$ is Uniform$\left(\sqrt{1-x^2}, -\sqrt{1-x^2}\right)$. The results above can be verified by noting that if $U \sim$ Uniform(a, b) then $E(U) = \frac{a+b}{2}$ and $Var(U) = \frac{(b-a)^2}{12}$.

3.7.4 Exercise

In Exercises 3.5.3 and 3.3.7 find $E(Y|x)$, $Var(Y|x)$, $E(X|y)$ and $Var(X|y)$.

3.7.5 Theorem

If X and Y are independent random variables then $E[g(X)|y] = E[g(X)]$ and $E[h(Y)|x] = E[h(Y)]$.

Proof (Continuous Case)

$$\begin{aligned}
E[g(X)|y] &= \int_{-\infty}^{\infty} g(x) f_1(x|y) dx \\
&= \int_{-\infty}^{\infty} g(x) f_1(x) dx \quad \text{by Theorem 3.5.8} \\
&= E\left[g\left(X\right)\right]
\end{aligned}$$

as required.

$E\left[h\left(Y\right)|x\right] = E\left[h\left(Y\right)\right]$ follows in a similar manner.

3.7.6 Definition

$E\left[g\left(X\right)|Y\right]$ is the function of the random variable Y whose value is $E\left[g\left(X\right)|y\right]$ when $Y = y$. This means of course that $E\left[g\left(X\right)|Y\right]$ is a random variable.

3.7.7 Example

In Example 3.5.6 the joint model was specified by $Y \sim$ Poisson(μ) and $X|Y = y \sim$ Binomial(y, p) and we showed that $X \sim$ Poisson$(p\mu)$. Determine $E\left(X|Y = y\right)$, $E\left(X|Y\right)$, $E[E(X|Y)]$, and $E\left(X\right)$. What do you notice about $E[E(X|Y)]$ and $E\left(X\right)$.

Solution

Since $X|Y = y \sim$ Binomial(y, p)

$$E\left(X|Y = y\right) = py$$

and

$$E\left(X|Y\right) = pY$$

which is a random variable. Since $Y \sim$ Poisson(μ) and $E\left(Y\right) = \mu$

$$E\left[E\left(X|Y\right)\right] = E\left(pY\right) = pE\left(Y\right) = p\mu$$

Now since $X \sim$ Poisson$(p\mu)$ then $E\left(X\right) = p\mu$.

We notice that

$$E\left[E\left(X|Y\right)\right] = E\left(X\right)$$

The following theorem indicates that this result holds generally.

3.7.8 Theorem

Suppose X and Y are random variables then

$$E\left\{E\left[g\left(X\right)|Y\right]\right\} = E\left[g\left(X\right)\right]$$

Proof (Continuous Case)

$$\begin{aligned} E\left\{E\left[g\left(X\right)|Y\right]\right\} &= E\left[\int_{-\infty}^{\infty} g\left(x\right) f_1\left(x|y\right) dx\right] \\ &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} g\left(x\right) f_1\left(x|y\right) dx\right] f_2\left(y\right) dy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g\left(x\right) f_1\left(x|y\right) f_2\left(y\right) dxdy \\ &= \int_{-\infty}^{\infty} g\left(x\right)\left[\int_{-\infty}^{\infty} f\left(x,y\right) dy\right] dx \\ &= \int_{-\infty}^{\infty} g\left(x\right) f_1\left(x\right) dx \\ &= E\left[g\left(X\right)\right] \end{aligned}$$

3.7.9 Corollary - Law of Total Expectation

Suppose X and Y are random variables then

$$E[E(X|Y)] = E(X)$$

Proof

Let $g\left(X\right) = X$ in Theorem 3.7.8 and the result follows.

3.7.10 Theorem - Law of Total Variance

Suppose X and Y are random variables then

$$Var(X) = E[Var(X|Y)] + Var[E(X|Y)]$$

Proof

$$\begin{aligned} Var\left(X\right) &= E\left(X^2\right) - \left[E\left(X\right)\right]^2 \\ &= E\left[E\left(X^2|Y\right)\right] - \left\{E\left[E\left(X|Y\right)\right]\right\}^2 \quad \text{by Theorem 3.7.8} \\ &= E\left[E\left(X^2|Y\right)\right] - E\left\{\left[E\left(X|Y\right)\right]^2\right\} + E\left\{\left[E(X|Y)\right]^2\right\} - \left\{E\left[E(X|Y)\right]\right\}^2 \\ &= E[Var(X|Y)] + Var[E(X|Y)] \end{aligned}$$

When the joint model is specified in terms of a conditional distribution $X|Y=y$ and a marginal distribution for Y then Theorems 3.7.8 and 3.7.10 give a method for finding expectations for functions of X without having to determine the marginal distribution for X.

3.7.11 Example

Suppose $P \sim$ Uniform$(0, 0.1)$ and $Y|P=p \sim$ Binomial$(10, p)$. Find $E(Y)$ and $Var(Y)$.

Solution

Since $P \sim$ Uniform$(0, 0.1)$

$$E(P) = \frac{0+0.1}{2} = \frac{1}{20}, \quad Var(P) = \frac{(0.1-0)^2}{12} = \frac{1}{1200}$$

and

$$\begin{aligned} E\left(P^2\right) &= Var(P) + [E(P)]^2 = \frac{1}{1200} + \left(\frac{1}{20}\right)^2 \\ &= \frac{1}{1200} + \frac{1}{400} = \frac{1+3}{1200} = \frac{4}{1200} = \frac{1}{300} \end{aligned}$$

Since $Y|P=p \sim$ Binomial$(10, p)$

$$E(Y|p) = 10p, \quad E(Y|P) = 10P$$

and

$$Var(Y|p) = 10p(1-p), \quad Var(Y|P) = 10P(1-P) = 10\left(P - P^2\right)$$

Therefore

$$E(Y) = E[E(Y|P)] = E(10P) = 10E(P) = 10\left(\frac{1}{20}\right) = \frac{1}{2}$$

and

$$\begin{aligned} Var(Y) &= E[Var(Y|P)] + Var[E(Y|P)] \\ &= E\left[10\left(P-P^2\right)\right] + Var(10P) \\ &= 10\left[E(P) - E\left(P^2\right)\right] + 100Var(P) \\ &= 10\left(\frac{1}{20} - \frac{1}{300}\right) + 100\left(\frac{1}{1200}\right) = \frac{11}{20} \end{aligned}$$

3.7.12 Exercise

In Example 3.5.7 find $E(X)$ and $Var(X)$ using Corollary 3.7.9 and Theorem 3.7.10.

3.7.13 Exercise

Suppose $P \sim$ Beta(a, b) and $Y|P=p \sim$ Geometric(p). Find $E(Y)$ and $Var(Y)$.

3.8 Joint Moment Generating Functions

Moment generating functions can also be defined for bivariate and multivariate random variables. As mentioned previously, moment generating functions are a powerful tool for determining the distributions of functions of random variables (Chapter 4), particularly sums, as well as determining the limiting distribution of a sequence of random variables (Chapter 5).

3.8.1 Definition - Joint Moment Generating Function

If X and Y are random variables then

$$M(t_1, t_2) = E\left(e^{t_1 X + t_2 Y}\right)$$

is called the *joint moment generating function of* X *and* Y if this expectation exists (joint sum/integral converges absolutely) for all $t_1 \in (-h_1, h_1)$ for some $h_1 > 0$, and all $t_2 \in (-h_2, h_2)$ for some $h_2 > 0$.

More generally if $X_1, X_2, \ldots, X_n$ are random variables then

$$M(t_1, t_2, \ldots, t_n) = E\left[\exp\left(\sum_{i=1}^{n} t_i X_i\right)\right]$$

is called the *joint moment generating function of* $X_1, X_2, \ldots, X_n$ if this expectation exists for all $t_i \in (-h_i, h_i)$ for some $h_i > 0$, $i = 1, 2, \ldots, n$.

If the joint moment generating function is known that it is straightforward to obtain the moment generating functions of the marginal distributions.

3.8.2 Important Note

If $M(t_1, t_2)$ exists for all $t_1 \in (-h_1, h_1)$ for some $h_1 > 0$, and all $t_2 \in (-h_2, h_2)$ for some $h_2 > 0$, then the moment generating function of X is given by

$$M_X(t) = E(e^{tX}) = M(t, 0) \quad \text{for } t \in (-h_1, h_1)$$

and the moment generating function of Y is given by

$$M_Y(t) = E(e^{tY}) = M(0, t) \quad \text{for } t \in (-h_2, h_2)$$

3.8.3 Example

Suppose X and Y are continuous random variables with joint probability density function

$$f(x, y) = e^{-y} \quad \text{for } 0 < x < y < \infty$$

and 0 otherwise.

(a) Find the joint moment generating function of X and Y.

(b) What is the moment generating function of X and what is the marginal distribution of X?

(c) What is the moment generating function of Y and what is the marginal distribution of Y?

Solution

(a) The joint moment generating function is

$$\begin{aligned}
M(t_1,t_2) &= E\left(e^{t_1X+t_2Y}\right) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{t_1x+t_2y} f(x,y)\,dxdy \\
&= \int_{y=0}^{\infty}\int_{x=0}^{y} e^{t_1x+t_2y}e^{-y}dxdy \\
&= \int_0^{\infty} e^{t_2y-y}\left(\int_0^{y} e^{t_1x}dx\right)dy \\
&= \int_0^{\infty} e^{t_2y-y}\left(\frac{1}{t_1}e^{t_1x}|_0^y\right)dy \\
&= \frac{1}{t_1}\int_0^{\infty} e^{t_2y-y}\left(e^{t_1y}-1\right)dy \\
&= \frac{1}{t_1}\int_0^{\infty}\left(e^{-(1-t_1-t_2)y}-e^{-(1-t_2)y}\right)dy
\end{aligned}$$

which converges for $t_1+t_2<1$ and $t_2<1$

Therefore

$$\begin{aligned}
M(t_1,t_2) &= \frac{1}{t_1}\lim_{b\to\infty}\left[\frac{-1}{1-t_1-t_2}e^{-(1-t_1-t_2)y}|_0^b+\frac{1}{1-t_2}e^{-(1-t_2)y}|_0^b\right] \\
&= \frac{1}{t_1}\lim_{b\to\infty}\left\{\frac{-1}{1-t_1-t_2}\left[e^{-(1-t_1-t_2)b}-1\right]+\frac{1}{1-t_2}\left[e^{-(1-t_2)b}-1\right]\right\} \\
&= \frac{1}{t_1}\left[\frac{1}{1-t_1-t_2}-\frac{1}{1-t_2}\right] \\
&= \frac{1}{t_1}\frac{(1-t_2)-(1-t_1-t_2)}{(1-t_1-t_2)(1-t_2)} \\
&= \frac{1}{(1-t_1-t_2)(1-t_2)} \quad \text{for } t_1+t_2<1 \text{ and } t_2<1
\end{aligned}$$

(b) The moment generating function of X is

$$\begin{aligned} M_X(t) &= E(e^{tX}) \\ &= M(t,0) \\ &= \frac{1}{(1-t-0)(1-0)} \\ &= \frac{1}{1-t} \quad \text{for } t<1 \end{aligned}$$

By examining the list of moment generating functions in Chapter 11 we see that this is the moment generating function of a Exponential(1) random variable. Therefore by the Uniqueness Theorem for Moment Generating Functions, X has a Exponential(1) distribution.

(c) The moment generating function of Y is

$$\begin{aligned} M_Y(t) &= E(e^{tY}) \\ &= M(0,t) \\ &= \frac{1}{(1-0-t)(1-t)} \\ &= \frac{1}{(1-t)^2} \quad \text{for } t<1 \end{aligned}$$

By examining the list of moment generating functions in Chapter 11 we see that this is the moment generating function of a Gamma$(2,1)$ random variable. Therefore by the Uniqueness Theorem for Moment Generating Functions, Y has a Gamma$(2,1)$ distribution.

3.8.4 Example

Suppose X and Y are continuous random variables with joint probability density function

$$f(x,y) = e^{-x-y} \quad \text{for } x>0,\ y>0$$

and 0 otherwise.

(a) Find the joint moment generating function of X and Y.

(b) What is the moment generating function of X and what is the marginal distribution of X?

(c) What is the moment generating function of Y and what is the marginal distribution of Y?

Solution

(a) The joint moment generating function is

$$\begin{aligned}
M(t_1, t_2) &= E\left(e^{t_1 X + t_2 Y}\right) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{t_1 x + t_2 y} f(x, y)\, dx dy \\
&= \int_{0}^{\infty}\int_{0}^{\infty} e^{t_1 x + t_2 y} e^{-x-y} dx dy \\
&= \left(\int_{0}^{\infty} e^{-y(1-t_2)} dy\right)\left(\int_{0}^{\infty} e^{-x(1-t_1)} dx\right) \quad \text{which converges for } t_1 < 1,\, t_2 < 1 \\
&= \lim_{b\to\infty}\left(\frac{-e^{-y(1-t_2)}}{(1-t_2)}|_0^b\right)\lim_{b\to\infty}\left(\frac{-e^{-x(1-t_1)}}{(1-t_1)}|_0^b\right) \\
&= \left(\frac{1}{1-t_1}\right)\left(\frac{1}{1-t_2}\right) \quad \text{for } t_1 < 1,\, t_2 < 1
\end{aligned}$$

(b) The moment generating function of X is

$$\begin{aligned}
M_X(t) &= E(e^{tX}) = M(t, 0) \\
&= \frac{1}{(1-t)(1-0)} \\
&= \frac{1}{1-t} \quad \text{for } t < 1
\end{aligned}$$

By examining the list of moment generating functions in Chapter 11 we see that this is the moment generating function of a Exponential(1) random variable. Therefore by the Uniqueness Theorem for Moment Generating Functions, X has a Exponential(1) distribution.

(c) The moment generating function of Y is

$$\begin{aligned}
M_Y(t) &= E(e^{tY}) = M(0, t) \\
&= \frac{1}{(1-0)(1-t)} \\
&= \frac{1}{1-t} \quad \text{for } t < 1
\end{aligned}$$

By examining the list of moment generating functions in Chapter 11 we see that this is the moment generating function of a Exponential(1) random variable. Therefore by the Uniqueness Theorem for Moment Generating Functions, Y has a Exponential(1) distribution.

3.8.5 Theorem

If X and Y are random variables with joint moment generating function $M(t_1, t_2)$ which exists for all $t_1 \in (-h_1, h_1)$ for some $h_1 > 0$, and all $t_2 \in (-h_2, h_2)$ for some $h_2 > 0$ then

$$E(X^j Y^k) = \frac{\partial^{j+k}}{\partial t_1^j \partial t_2^k} M(t_1, t_2)|_{(t_1,t_2)=(0,0)}$$

Proof

See Problem $11(a)$.

3.8.6 Independence Theorem for Moment Generating Functions

Suppose X and Y are random variables with joint moment generating function $M(t_1, t_2)$ which exists for all $t_1 \in (-h_1, h_1)$ for some $h_1 > 0$, and all $t_2 \in (-h_2, h_2)$ for some $h_2 > 0$. Then X and Y are independent random variables if and only if

$$M(t_1, t_2) = M_X(t_1) M_Y(t_2)$$

for all $t_1 \in (-h_1, h_1)$ and $t_2 \in (-h_2, h_2)$ where $M_X(t_1) = M(t_1, 0)$ and $M_Y(t_2) = M(0, t_2)$.

Proof

See Problem $11(b)$.

3.8.7 Example

Use Theorem 3.8.6 to determine if X and Y are independent random variables in Examples 3.8.3 and 3.8.4.

Solution

For Example 3.8.3

$$\begin{aligned}
M(t_1, t_2) &= \frac{1}{(1 - t_1 - t_2)(1 - t_2)} \quad \text{for } t_1 + t_2 < 1 \text{ and } t_2 < 1 \\
M_X(t_1) &= \frac{1}{1 - t_1} \quad \text{for } t_1 < 1 \\
M_Y(t_2) &= \frac{1}{(1 - t_2)^2} \quad \text{for } t_2 < 1
\end{aligned}$$

Since

$$M\left(\frac{1}{4}, \frac{1}{4}\right) = \frac{1}{\left(1 - \frac{1}{4} - \frac{1}{4}\right)\left(1 - \frac{1}{4}\right)} = \frac{8}{3} \neq M_X\left(\frac{1}{4}\right) M_Y\left(\frac{1}{4}\right) = \frac{1}{\left(1 - \frac{1}{4}\right)} \frac{1}{\left(1 - \frac{1}{4}\right)^2} = \left(\frac{4}{3}\right)^3$$

then by Theorem 3.8.6 X and Y are not independent random variables.

For Example 3.8.4

$$\begin{aligned}
M(t_1,t_2) &= \left(\frac{1}{1-t_1}\right)\left(\frac{1}{1-t_2}\right) \quad \text{for } t_1<1,\ t_1<1 \\
M_X(t_1) &= \frac{1}{1-t_1} \quad \text{for } t_1<1 \\
M_Y(t_2) &= \frac{1}{1-t_2} \quad \text{for } t_2<1
\end{aligned}$$

Since

$$M(t_1,t_2) = \left(\frac{1}{1-t_1}\right)\left(\frac{1}{1-t_2}\right) = M_X(t_1)\,M_Y(t_2) \quad \text{for all } t_1<1,\ t_1<1$$

then by Theorem 3.8.6 X and Y are independent random variables.

3.8.8 Example

Suppose $X_1, X_2, \ldots, X_n$ are independent and identically distributed random variables each with moment generating function $M(t)$, $t \in (-h,h)$ for some $h>0$. Find $M(t_1,t_2,\ldots,t_n)$ the joint moment generating function of $X_1, X_2, \ldots, X_n$. Find the moment generating function of $T = \sum\limits_{i=1}^{n} X_i$.

Solution

Since the X_i's are independent random variables each with moment generating function $M(t)$, $t \in (-h,h)$ for some $h>0$, the joint moment generating function of $X_1, X_2, \ldots, X_n$ is

$$\begin{aligned}
M(t_1,t_2,\ldots,t_n) &= E\left[\exp\left(\sum_{i=1}^{n} t_i X_i\right)\right] \\
&= E\left(\prod_{i=1}^{n} e^{t_i X_i}\right) \\
&= \prod_{i=1}^{n} E\left(e^{t_i X_i}\right) \\
&= \prod_{i=1}^{n} M(t_i) \quad \text{for } t_i \in (-h,h),\ i=1,2,\ldots,n \text{ for some } h>0
\end{aligned}$$

The moment generating function of $T = \sum\limits_{i=1}^{n} X_i$ is

$$\begin{aligned}
M_T(t) &= E\left(e^{tT}\right) \\
&= E\left[\exp\left(\sum_{i=1}^{n} tX_i\right)\right] \\
&= M(t,t,\ldots,t) \\
&= \prod_{i=1}^{n} M(t) \\
&= [M(t)]^n \quad \text{for } t \in (-h,h) \text{ for some } h>0
\end{aligned}$$

3.9 Multinomial Distribution

The discrete multivariate distribution which is the most widely used is the Multinomial distribution which was introduced in a previous probability course.

We give its joint probability function and its important properties here.

3.9.1 Definition - Multinomial Distribution

Suppose $(X_1, X_2, \ldots, X_k)$ are discrete random variables with joint probability function

$$f(x_1, x_2, \ldots, x_k) = \frac{n!}{x_1! x_2! \cdots x_k!} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}$$

$$\text{for } \; x_i = 0, 1, \ldots, n, \; i = 1, 2, \ldots, k, \; \sum_{i=1}^{k} x_i = n; \quad 0 \le p_i \le 1, i = 1, 2, \ldots, k, \; \sum_{i=1}^{k} p_i = 1$$

Then $(X_1, X_2, \ldots, X_k)$ is said to have a *Multinomial distribution*.
We write $(X_1, X_2, \ldots, X_k) \sim \text{Multinomial}(n;\; p_1, p_2, \ldots, p_k)$.

Notes: (1) Since $\sum_{i=1}^{k} X_i = n$, the Multinomial distribution is actually a joint distribution for $k-1$ random variables which can be written as

$$f(x_1, x_2, \ldots, x_{k-1}) = \frac{n!}{x_1! x_2! \cdots x_{k-1}! \left(n - \sum_{i=1}^{k-1} x_i\right)!} p_1^{x_1} p_2^{x_2} \cdots p_{k-1}^{x_{k-1}} \left(1 - \sum_{i=1}^{k-1} p_i\right)^{n - \sum_{i=1}^{k-1} x_i}$$

$$\text{for } \; x_i = 0, 1, \ldots, n, \; i = 1, 2, \ldots, k-1, \; \sum_{i=1}^{k-1} x_i \le n; \quad 0 \le p_i \le 1, i = 1, 2, \ldots, k-1, \; \sum_{i=1}^{k-1} p_i \le 1$$

(2) If $k = 2$ we obtain the familiar Binomial distribution

$$\begin{aligned} f(x_1) &= \frac{n!}{x_1! (n - x_1)!} p_1^{x_1} (1 - p_1)^{n - x_1} \\ &= \binom{n}{x_1} p_1^{x_1} (1 - p_1)^{n - x_1} \end{aligned}$$

$$\text{for } x_1 = 0, 1, \ldots, n, \; 0 \le p_1 \le 1$$

(2) If $k = 3$ we obtain the Trinomial distribution

$$f(x_1, x_2) = \frac{n!}{x_1! x_2! (n - x_1 - x_2)!} p_1^{x_1} p_2^{x_2} (1 - p_1 - p_2)^{n - x_1 - x_2}$$

$$\text{for } x_i = 0, 1, \ldots, n, \; i = 1, 2, \; x_1 + x_2 \le n \; \text{ and } \; 0 \le p_i \le 1, i = 1, 2, \; p_1 + p_2 \le 1$$

3.9.2 Theorem - Properties of the Multinomial Distribution

If $(X_1, X_2, \ldots, X_k) \sim$ Multinomial$(n; p_1, p_2, \ldots, p_k)$, then

(1) $(X_1, X_2, \ldots, X_{k-1})$ has joint moment generating function

$$\begin{aligned} M(t_1, t_2, \ldots, t_{k-1}) &= E\left(e^{t_1 X_1 + t_2 X_2 + \cdots + t_{k-1} X_{k-1}}\right) \\ &= \left(p_1 e^{t_1} + p_2 e^{t_2} \cdots + p_{k-1} e^{t_{k-1}} + p_k\right)^n \end{aligned} \tag{3.5}$$

for $(t_1, t_2, \ldots, t_{k-1}) \in \Re^{k-1}$.

(2) Any subset of $X_1, X_2, \ldots, X_k$ also has a Multinomial distribution. In particular

$$X_i \sim \text{ Binomial}(n,\ p_i) \quad \text{for } i = 1, 2, \ldots, k$$

(3) If $T = X_i + X_j$, $i \neq j$, then

$$T \sim \text{ Binomial}\left(n, p_i + p_j\right)$$

(4)

$$Cov\left(X_i, X_j\right) = -np_i p_j \quad \text{for } i \neq j$$

(5) The conditional distribution of any subset of $(X_1, X_2, \ldots, X_k)$ given the remaining of the coordinates is a Multinomial distribution. In particular the conditional probability function of X_i given $X_j = x_j$, $i \neq j$, is

$$X_i | X_j = x_j \sim \text{ Binomial}\left(n - x_j, \frac{p_i}{1 - p_j}\right)$$

(6) The conditional distribution of X_i given $T = X_i + X_j = t$, $i \neq j$, is

$$X_i | X_i + X_j = t \sim \text{ Binomial}\left(t, \frac{p_i}{p_i + p_j}\right)$$

3.9.3 Example

Suppose $(X_1, X_2, \ldots, X_k) \sim$ Multinomial$(n; p_1, p_2, \ldots, p_k)$

(a) Prove $(X_1, X_2, \ldots, X_{k-1})$ has joint moment generating function

$$M(t_1, t_2, \ldots, t_{k-1}) = \left(p_1 e^{t_1} + p_2 e^{t_2} \cdots + p_{k-1} e^{t_{k-1}} + p_k\right)^n$$

for $(t_1, t_2, \ldots, t_{k-1}) \in \Re^{k-1}$.

(b) Prove $(X_1, X_2, X_3) \sim$ Multinomial$(n; p_1, p_1, 1 - p_1 - p_2)$.

(c) Prove $X_i \sim$ Binomial(n, p_i) for $i = 1, 2, \ldots, k$.

(d) Prove $T = X_1 + X_2 \sim$ Binomial$(n, p_1 + p_2)$.

Solution

(a) Let $A = \left\{ (x_1, x_2, \ldots, x_k) : x_i = 0, 1, \ldots, n, \; i = 1, 2, \ldots, k, \; \sum\limits_{i=1}^{k} x_i = n \right\}$ then

$$M\left(t_1, t_2, \ldots, t_{k-1}\right) = E\left(e^{t_1 X_1 + t_2 X_2 + \cdots + t_{k-1} X_{k-1}}\right)$$

$$\begin{aligned}
&= \sum_{(x_1, x_2, \ldots, x_k)} \sum_{\in A} \cdots \sum e^{t_1 x_1 + t_2 x_2 + \cdots + t_{k-1} x_{k-1}} \frac{n!}{x_1! x_2! \cdots x_k!} p_1^{x_1} p_2^{x_2} \cdots p_{k-1}^{x_k - 1} p_k^{x_k} \\
&= \sum_{(x_1, x_2, \ldots, x_k)} \sum_{\in A} \cdots \sum \frac{n!}{x_1! x_2! \cdots x_k!} \left(p_1 e^{t_1}\right)^{x_1} \left(p_2 e^{t_2}\right)^{x_2} \cdots \left(p_{k-1} e^{t_{k-1}}\right)^{x_{k-1}} p_k^{x_k} \\
&= \left(p_1 e^{t_1} + p_2 e^{t_2} \cdots + p_{k-1} e^{t_{k-1}} + p_k\right)^n \quad \text{for } \left(t_1, t_2, \ldots, t_{k-1}\right) \in \Re^k
\end{aligned}$$

by the Multinomial Theorem 2.11.5.

(b) The joint moment generating function of (X_1, X_2, X_3) is

$$M\left(t_1, t_2, 0, \ldots, 0\right) = \left[p_1 e^{t_1} + p_2 e^{t_2} + (1 - p_1 - p_2)\right]^n \quad \text{for } \left(t_1, t_2\right) \in \Re^2$$

which is of the form 3.5 so by the Uniqueness Theorem for Moment Generating Functions, $(X_1, X_2, X_3) \sim$ Multinomial$(n; p_1, p_1, 1 - p_1 - p_2)$.

(c) The moment generating function of X_i is

$$M\left(0, 0, \ldots, t, 0, \ldots, 0\right) = \left[p_i e^{t_i} + (1 - p_i)\right]^n \quad \text{for } t_i \in \Re$$

for $i = 1, 2, \ldots, k$ which is the moment generating function of a Binomial(n, p_i) random variable. By the Uniqueness Theorem for Moment Generating Functions, $X_i \sim$ Binomial(n, p_i) for $i = 1, 2, \ldots, k$.

(d) The moment generating function of $T = X_1 + X_2$ is

$$\begin{aligned}
M_T\left(t\right) &= E\left(e^{tT}\right) = E\left(e^{t(X_1 + X_2)}\right) \\
&= E\left(e^{tX_1 + tX_2}\right) \\
&= M\left(t, t, 0, 0, \ldots, 0\right) \\
&= \left[p_1 e^t + p_2 e^t + (1 - p_1 - p_2)\right]^n \\
&= \left[\left(p_1 + p_2\right) e^t + (1 - p_1 - p_2)\right]^n \quad \text{for } t \in \Re
\end{aligned}$$

which is the moment generating function of a Binomial$(n, p_1 + p_2)$ random variable. By the Uniqueness Theorem for Moment Generating Functions, $T \sim$ Binomial$(n, p_1 + p_2)$.

3.9.4 Exercise

Prove property (3) in Theorem 3.9.2.

3.10 Bivariate Normal Distribution

The best known bivariate continuous distribution is the Bivariate Normal distribution. We give its joint probability density function written in vector notation so we can easily introduce the multivariate version of this distribution called the Multivariate Normal distribution in Chapter 7.

3.10.1 Definition - Bivariate Normal Distribution (BVN)

Suppose X_1 and X_2 are random variables with joint probability density function

$$f(x_1, x_2) = \frac{1}{2\pi|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})^T\right\} \quad \text{for } (x_1, x_2) \in \Re^2$$

where

$$\mathbf{x} = \begin{bmatrix} x_1 & x_2 \end{bmatrix}, \qquad \boldsymbol{\mu} = \begin{bmatrix} \mu_1 & \mu_2 \end{bmatrix}, \qquad \Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

and Σ is an nonsingular matrix. Then $\mathbf{X} = (X_1, X_2)$ is said to have a *bivariate normal distribution*. We write $\mathbf{X} \sim \text{BVN}(\boldsymbol{\mu}, \Sigma)$.

The Bivariate Normal distribution has many special properties.

3.10.2 Theorem - Properties of the BVN Distribution

If $\mathbf{X} \sim \text{BVN}(\boldsymbol{\mu}, \Sigma)$, then

(1) X_1, X_2 has joint moment generating function

$$\begin{aligned} M(t_1, t_2) &= E\left(e^{t_1X_1 + t_2X_2}\right) = E\left[\exp\left(\mathbf{X}\mathbf{t}^T\right)\right] \\ &= \exp\left(\boldsymbol{\mu}\mathbf{t}^T + \frac{1}{2}\mathbf{t}\Sigma\mathbf{t}^T\right) \quad \text{for all } \mathbf{t} = (t_1, t_2) \in \Re^2 \end{aligned}$$

(2) $X_1 \sim \text{N}(\mu_1, \sigma_1^2)$ and $X_2 \sim \text{N}(\mu_2, \sigma_2^2)$.

(3) $Cov(X_1, X_2) = \rho\sigma_1\sigma_2$ and $Cor(X_1, X_2) = \rho$ where $-1 \leq \rho \leq 1$.

(4) X_1 and X_2 are independent random variables if and only if $\rho = 0$.

(5) If $\mathbf{c} = (c_1, c_2)$ is a nonzero vector of constants then

$$c_1X_1 + c_2X_2 \sim \text{N}\left(\boldsymbol{\mu}\mathbf{c}^T, \mathbf{c}\Sigma\mathbf{c}^T\right)$$

(6) If A is a 2×2 nonsingular matrix and $\mathbf{b}$ is a 1×2 vector then

$$\mathbf{X}A + \mathbf{b} \sim \text{BVN}\left(\boldsymbol{\mu}A + \mathbf{b}, A^T\Sigma A\right)$$

(7)

$$X_2|X_1 = x_1 \sim \text{N}\left(\mu_2 + \rho\sigma_2(x_1 - \mu_1)/\sigma_1,\ \sigma_2^2(1-\rho^2)\right)$$

and

$$X_1|X_2 = x_2 \sim \mathrm{N}\left(\mu_1 + \rho\sigma_1(x_2 - \mu_2)/\sigma_2,\ \sigma_1^2(1-\rho^2)\right)$$

(8) $(\mathbf{X} - \boldsymbol{\mu})\Sigma^{-1}(\mathbf{X} - \boldsymbol{\mu})^T \sim \chi^2(2)$

Proof

For proofs of properties (1) – (4) and (6) – (7) see Problem 13.

(5) The moment generating function of $c_1X_1 + c_2X_2$ is

$$\begin{aligned}
& E\left(e^{t(c_1X_1+c_2X_2)}\right) \\
= \;& E\left(e^{(c_1t)X_1+(c_2t)X_2}\right) \\
= \;& \exp\left(\begin{bmatrix} \mu_1 & \mu_2 \end{bmatrix}\begin{bmatrix} c_1t \\ c_2t \end{bmatrix} + \frac{1}{2}\begin{bmatrix} c_1t & c_2t \end{bmatrix}\Sigma\begin{bmatrix} c_1t \\ c_2t \end{bmatrix}\right) \\
= \;& \exp\left(\begin{bmatrix} \mu_1 & \mu_2 \end{bmatrix}\begin{bmatrix} c_1 \\ c_2 \end{bmatrix}t + \frac{1}{2}\begin{bmatrix} c_1 & c_2 \end{bmatrix}\Sigma\begin{bmatrix} c_1 \\ c_2 \end{bmatrix}t^2\right) \\
= \;& \exp\left[\left(\boldsymbol{\mu}\mathbf{c}^T\right)t + \frac{1}{2}\left(\mathbf{c}\Sigma\mathbf{c}^T\right)t^2\right] \quad \text{for } t \in \Re \text{ where } \mathbf{c} = (c_1, c_2)
\end{aligned}$$

which is the moment generating function of a $\mathrm{N}\left(\boldsymbol{\mu}\mathbf{c}^T, \mathbf{c}\Sigma\mathbf{c}^T\right)$ random variable. Therefore by the Uniqueness Theorem for Moment Generating Functions, $c_1X_1 + c_2X_2 \sim \mathrm{N}\left(\boldsymbol{\mu}\mathbf{c}^T, \mathbf{c}\Sigma\mathbf{c}^T\right)$.

The BVN joint probability density function is graphed in Figures 3.8 - 3.10.

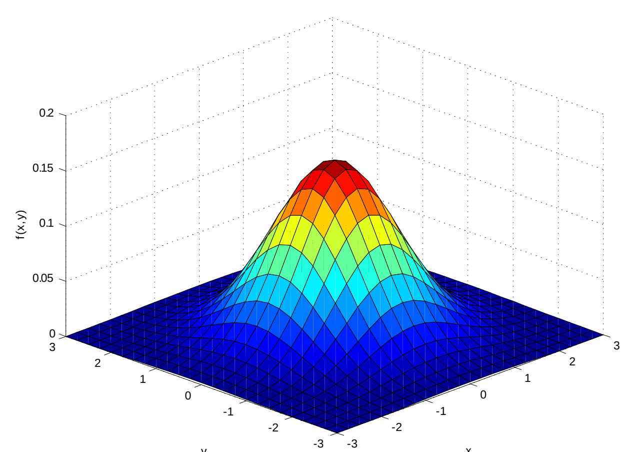


Figure 3.8: Graph of BVN p.d.f. with $\mu = \begin{bmatrix} 0 & 0 \end{bmatrix}$ and $\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

The graphs all have the same mean vector $\mu = [0\ 0]$ but different variance/covariance matrices Σ. The axes all have the same scale.

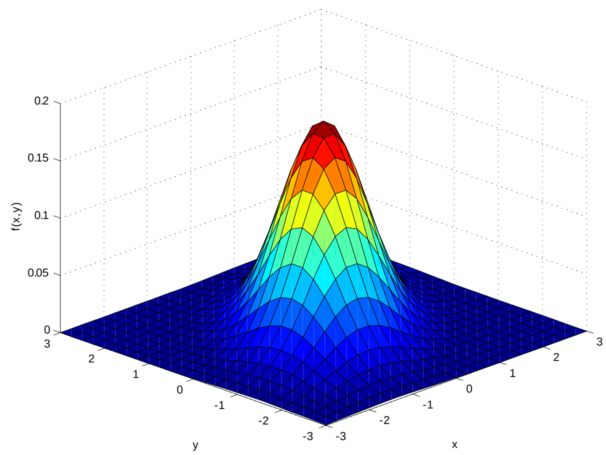


Figure 3.9: Graph of BVN p.d.f. with $\mu = \begin{bmatrix} 0 & 0 \end{bmatrix}$ and $\Sigma = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$

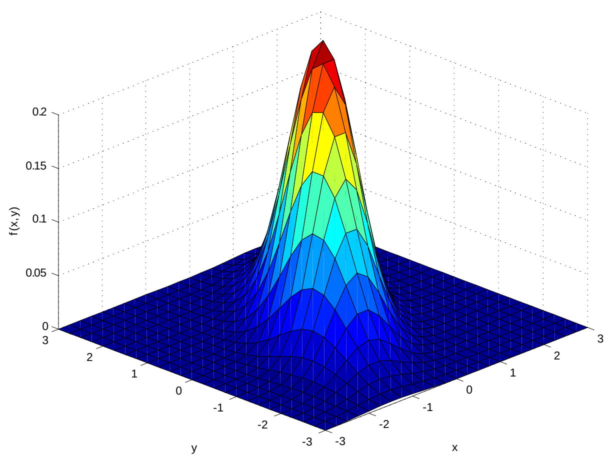


Figure 3.10: Graph of BVN p.d.f. with $\mu = \begin{bmatrix} 0 & 0 \end{bmatrix}$ and $\Sigma = \begin{bmatrix} 0.6 & 0.5 \\ 0.5 & 1 \end{bmatrix}$

3.11 Calculus Review

Consider the region R in the xy-plane in Figure 3.11.

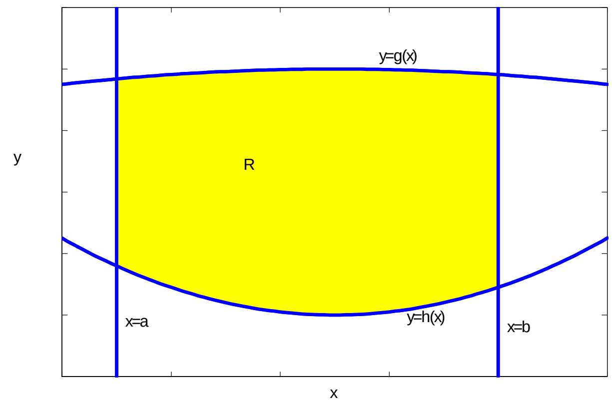


Figure 3.11: Region 1

Suppose $f(x,y) \geq 0$ for all $(x,y) \in \Re^2$. The graph of $z = f(x,y)$ is a surface in 3-space lying above or touching the xy-plane. The volume of the solid bounded by the surface $z = f(x,y)$ and the xy-plane above the region R is given by

$$\text{Volume} = \int\limits_{x=a}^{b} \int\limits_{y=g(x)}^{h(x)} f(x,y)dydx$$

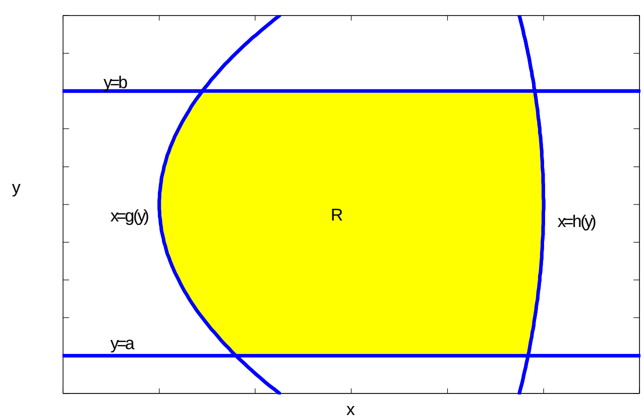


Figure 3.12: Region 2

If R is the region in Figure 3.12 then the volume is given by

$$\text{Volume} = \int_{y=c}^{d} \int_{x=g(y)}^{h(y)} f(x,y)dxdy$$

Give an expression for the volume of the solid bounded by the surface $z = f(x,y)$ and the xy-plane above the region $R = R_1 \cup R_2$ in Figure 3.13.

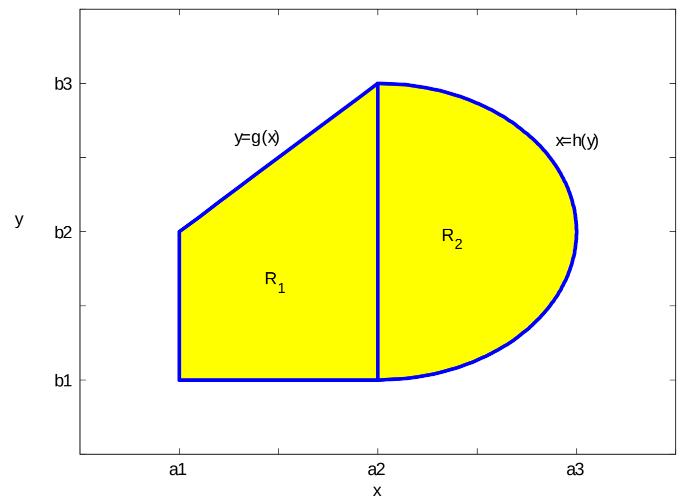


Figure 3.13: Region 3

3.12 Chapter 3 Problems

1. Suppose X and Y are discrete random variables with joint probability function

$$f(x,y) = kq^2p^{x+y} \quad \text{for } x = 0,1,\ldots, \quad y = 0,1,\ldots; \quad 0 < p < 1, \quad q = 1-p$$

   (a) Determine the value of k.

   (b) Find the marginal probability function of X and the marginal probability function of Y. Are X and Y independent random variables?

   (c) Find $P(X = x|X + Y = t)$.

2. Suppose X and Y are discrete random variables with joint probability function

$$f(x,y) = \frac{e^{-2}}{x!(y-x)!} \quad \text{for } x = 0,1,\ldots,y; \quad y = 0,1,\ldots$$

   (a) Find the marginal probability function of X and the marginal probability function of Y.

   (b) Are X and Y independent random variables?

3. Suppose X and Y are continuous random variables with joint probability density function

$$f(x,y) = k(x^2 + y) \quad \text{for } 0 < y < 1 - x^2, \quad -1 < x < 1$$

   (a) Determine k.

   (b) Find the marginal probability density function of X and the marginal probability density function of Y.

   (c) Are X and Y independent random variables?

   (d) Find $P(Y \leq X + 1)$.

4. Suppose X and Y are continuous random variables with joint probability density function

$$f(x,y) = kx^2y \quad \text{for } x^2 < y < 1$$

   (a) Determine k.

   (b) Find the marginal probability density function of X and the marginal probability density function of Y.

   (c) Are X and Y independent random variables?

   (d) Find $P(X \geq Y)$.

   (e) Find the conditional probability density function of X given $Y = y$ and the conditional probability density function of Y given $X = x$.

5. Suppose X and Y are continuous random variables with joint probability density function
$$f(x, y) = kxe^{-y} \quad \text{for } 0 < x < 1, \quad 0 < y < \infty$$
   (a) Determine k.
   (b) Find the marginal probability density function of X and the marginal probability density function of Y.
   (c) Are X and Y independent random variables?
   (d) Find $P(X + Y \leq t)$.

6. Suppose each of the following functions is a joint probability density function for continuous random variables X and Y.
   (a) $f(x, y) = k \quad \text{for } 0 < x < y < 1$
   (b) $f(x, y) = kx \quad \text{for } 0 < x^2 < y < 1$
   (c) $f(x, y) = kxy \quad \text{for } 0 < y < x < 1$
   (d) $f(x, y) = k(x + y) \quad \text{for } 0 < x < y < 1$
   (e) $f(x, y) = \frac{k}{x} \quad \text{for } 0 < y < x < 1$
   (f) $f(x, y) = kx^2y \quad \text{for } 0 < x < 1, 0 < y < 1, 0 < x + y < 1$
   (g) $f(x, y) = ke^{-x-2y} \quad \text{for } 0 < y < x < \infty$

   In each case:
   (i) Determine k.
   (ii) Find the marginal probability density function of X and the marginal probability density function of Y.
   (iii) Find the conditional probability density function of X given $Y = y$ and the conditional probability density function of Y given $X = x$.
   (iv) Find $E(X|y)$ and $E(Y|x)$.

7. Suppose $X \sim$ Uniform$(0, 1)$ and the conditional probability density function of Y given $X = x$ is
$$f_2(y|x) = \frac{1}{1 - x} \quad \text{for } 0 < x < y < 1$$
   Determine:
   (a) the joint probability density function of X and Y
   (b) the marginal probability density function of Y
   (c) the conditional probability density function of X given $Y = y$.

8. Suppose X and Y are continuous random variables. Suppose also that the marginal probability density function of X is

$$f_1(x) = \frac{1}{3}(1+4x) \quad \text{for } 0 < x < 1$$

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

$$f_2(y|x) = \frac{2y+4x}{1+4x} \quad \text{for } 0 < x < 1,\ 0 < y < 1$$

Determine:

   (a) the joint probability density function of X and Y
   (b) the marginal probability density function of Y
   (c) the conditional probability density function of X given $Y = y$.

9. Suppose that $\theta \sim \text{Beta}(a, b)$ and $Y|\theta \sim \text{Binomial}(n, \theta)$. Find $E(Y)$ and $Var(Y)$.

10. Assume that Y denotes the number of bacteria in a cubic centimeter of liquid and that $Y|\mu \sim \text{Poisson}(\mu)$. Further assume that μ varies from location to location and $\mu \sim \text{Gamma}(\alpha, \beta)$.

   (a) Find $E(Y)$ and $Var(Y)$.
   (b) If α is a positive integer then show that the marginal probability function of Y is Negative Binomial.

11. Suppose X and Y are random variables with joint moment generating function $M(t_1, t_2)$ which exists for all $|t_1| < h_1$ and $|t_2| < h_2$ for some $h_1, h_2 > 0$.

   (a) Show that

$$E(X^j Y^k) = \frac{\partial^{j+k}}{\partial t_1^j \partial t_2^k} M(t_1, t_2)|_{(t_1,t_2)=(0,0)}$$

   (b) Prove that X and Y are independent random variables if and only if $M(t_1, t_2) = M_X(t_1) M_Y(t_2)$.
   (c) If $(X_1, X_2, X_3) \sim \text{Multinomial}(n, p_1, p_2, p_3)$ find $Cov(X_1, X_2)$.

12. Suppose X and Y are discrete random variables with joint probability function

$$f(x, y) = \frac{e^{-2}}{x!(y-x)!} \quad \text{for } x = 0, 1, \ldots, y; \quad y = 0, 1, \ldots$$

   (a) Find the joint moment generating function of X and Y.
   (b) Find $Cov(X, Y)$.

13. Suppose $X = (X_1, X_2) \sim \text{BVN}(\boldsymbol{\mu}, \Sigma)$.

    (a) Let $t = (t_1, t_2)$. Use matrix multiplication to verify that

    $$\begin{aligned} & (\mathbf{x} - \boldsymbol{\mu})\,\Sigma^{-1}\,(\mathbf{x} - \boldsymbol{\mu})^T - 2\mathbf{x}\mathbf{t}^T \\ = \; & [\mathbf{x} - (\boldsymbol{\mu} + \mathbf{t}\Sigma)]\,\Sigma^{-1}\,[\mathbf{x} - (\boldsymbol{\mu} + \mathbf{t}\Sigma)]^T - 2\boldsymbol{\mu}\mathbf{t}^T - \mathbf{t}\Sigma\mathbf{t}^T \end{aligned}$$

    Use this identity to show that the joint moment generating function of X_1 and X_2 is

    $$\begin{aligned} M(t_1, t_2) &= E\left(e^{t_1 X_1 + t_2 X_2}\right) = E\left[\exp\left(\mathbf{X}\mathbf{t}^T\right)\right] \\ &= \exp\left(\boldsymbol{\mu}\mathbf{t}^T + \frac{1}{2}\mathbf{t}\Sigma\mathbf{t}^T\right) \quad \text{for all } \mathbf{t} = (t_1, t_2) \in \Re^2 \end{aligned}$$

    (b) Use moment generating functions to show $X_1 \sim \text{N}(\mu_1, \sigma_1^2)$ and $X_2 \sim \text{N}(\mu_2, \sigma_2^2)$.

    (c) Use moment generating functions to show $Cov(X_1, X_2) = \rho\sigma_1\sigma_2$. Hint: Use the result in Problem $11(a)$.

    (d) Use moment generating functions to show that X_1 and X_2 are independent random variables if and only if $\rho = 0$.

    (e) Let A be a 2×2 nonsingular matrix and $\mathbf{b}$ be a 1×2 vector. Use the moment generating function to show that

    $$\mathbf{X}A + \mathbf{b} \sim \text{BVN}\left(\boldsymbol{\mu}A + \mathbf{b}, A^T \Sigma A\right)$$

    (f) Verify that

    $$(\mathbf{x} - \boldsymbol{\mu})\,\Sigma^{-1}\,(\mathbf{x} - \boldsymbol{\mu})^T - \left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2 = \frac{1}{\sigma_2^2\,(1 - \rho^2)}\left\{x_2 - \left[\mu_2 + \frac{\rho\sigma_2}{\sigma_1}\,(x_1 - \mu_1)\right]\right\}^2$$

    and thus show that the conditional distribution of X_2 given $X_1 = x_1$ is $\text{N}(\mu_2 + \rho\sigma_2(x_1 - \mu_1)/\sigma_1,\ \sigma_2^2(1 - \rho^2))$. Note that by symmetry the conditional distribution of X_1 given $X_2 = x_2$ is $\text{N}(\mu_1 + \rho\sigma_1(x_2 - \mu_2)/\sigma_2,\ \sigma_1^2(1 - \rho^2))$.

14. Suppose X and Y are continuous random variables with joint probability density function

    $$f(x, y) = 2e^{-x-y} \quad \text{for } 0 < x < y < \infty$$

    (a) Find the joint moment generating function of X and Y.

    (b) Determine the marginal distributions of X and Y.

    (c) Find $Cov(X, Y)$.

4. Functions of Two or More Random Variables

In this chapter we look at techniques for determining the distributions of functions of two or more random variables. These techniques are extremely important for determining the distributions of estimators such as maximum likelihood estimators, the distributions of pivotal quantities for constructing confidence intervals, and the distributions of test statistics for testing hypotheses.

In Section 4.1 we extend the cumulative distribution function technique introduced in Section 2.6 to a function of two or more random variables. In Section 4.2 we look at a method for determining the distribution of a one-to-one transformation of two or more random variables which is an extension of the result in Theorem 2.6.8. In particular we show how the t distribution, which you would have used in your previous statistics course, arises as the ratio of two independent Chi-squared random variables each divided by their degrees of freedom. In Section 4.3 we see how moment generating functions can be used for determining the distribution of a sum of random variables which is an extension of Theorem 2.10.4. In particular we prove that a linear combination of independent Normal random variables has a Normal distribution. This is a result which was used extensively in previous probability and statistics courses.

4.1 Cumulative Distribution Function Technique

Suppose $X_1, X_2, \ldots, X_n$ are continuous random variables with joint probability density function $f(x_1, x_2, \ldots, x_n)$. The probability density function of $Y = h(X_1, X_2, \ldots, X_n)$ can be determined using the cumulative distribution function technique that was used in Section 2.6 for the case $n = 1$.

4.1.1 Example

Suppose X and Y are continuous random variables with joint probability density function

$$f(x, y) = 3y \quad \text{for } 0 < x < y < 1$$

and 0 otherwise. Determine the probability density function of $T = XY$.

Solution

The support set of (X, Y) is $A = \{(x, y) : 0 < x < y < 1\}$ which is the union of the regions E and F shown in Figure 4.1

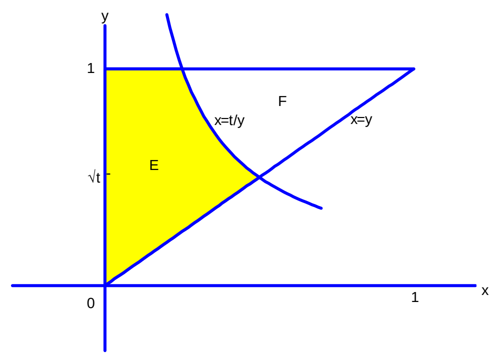


Figure 4.1: Support set for Example 4.1.1

For $0 < t < 1$

$$G(t) = P(T \le t) = P(XY \le t) = \int\int_{(x,y)\,\in\, E} 3y dx dy$$

Due to the shape of the region E, the double integral over the region E would have to be written as the sum of two double integrals. It is easier to find $G(t)$ using

$$\begin{aligned}
G(t) &= \int\int_{(x,y)\,\in\, E} 3y dx dy \\
&= 1 - \int\int_{(x,y)\,\in\, F} 3y dx dy \\
&= 1 - \int_{y=\sqrt{t}}^{1} \int_{x=t/y}^{y} 3y dx dy = 1 - \int_{\sqrt{t}}^{1} 3y \left(x|_{t/y}^{y} \right) dy \\
&= 1 - \int_{\sqrt{t}}^{1} 3y \left(y - \frac{t}{y} \right) dy = 1 - \int_{\sqrt{t}}^{1} \left(3y^2 - 3t \right) dy \\
&= 1 - \left(y^3 - 3ty \right)|_{\sqrt{t}}^{1} = 1 - \left(1 - 3t - t^{3/2} - 3t^{3/2} \right) \\
&= 3t - 2t^{3/2} \quad \text{for } 0 < t < 1
\end{aligned}$$

The cumulative distribution function for T is

$$G(t) = \begin{cases} 0 & t \leq 0 \\ 3t - 2t^{3/2} & 0 < t < 1 \\ 1 & t \geq 1 \end{cases}$$

Now a cumulative distribution function must be a continuous function for all real values. Therefore as a check we note that

$$\lim_{t \to 0^+} \left(3t - 2t^{3/2}\right) = 0 = G(0)$$

and

$$\lim_{t \to 1^-} \left(3t - 2t^{3/2}\right) = 1 = G(1)$$

so indeed $G(t)$ is a continuous function for all $t \in \Re$.

Since $\frac{d}{dt}G(t) = 0$ for $t < 0$ and $t > 0$, and

$$\begin{aligned} \frac{d}{dt}G(t) &= \frac{d}{dt}\left(3t - 2t^{3/2}\right) \\ &= 3 - 3t^{1/2} \quad \text{for } 0 < t < 1 \end{aligned}$$

the probability density function of T is

$$g(t) = 3 - 3t^{1/2} \quad \text{for } 0 < t < 1$$

and 0 otherwise.

4.1.2 Exercise

Suppose X and Y are continuous random variables with joint probability density function

$$f(x, y) = 3y \quad \text{for } 0 \leq x \leq y \leq 1$$

and 0 otherwise. Find the probability density function of $S = Y/X$.

4.1.3 Example

Suppose $X_1, X_2, \ldots, X_n$ are independent and identically distributed continuous random variables each with probability density function $f(x)$ and cumulative distribution function $F(x)$. Find the probability density function of $U = \max(X_1, X_2, \ldots, X_n) = X_{(n)}$ and $V = \min(X_1, X_2, \ldots, X_n) = X_{(1)}$.

Solution

For $u \in \Re$, the cumulative distribution function of U is

$$\begin{aligned} G(u) &= P(U \leq u) = P[\max(X_1, X_2, \ldots, X_n) \leq u] \\ &= P(X_1 \leq u, X_2 \leq u, \ldots, X_n \leq u) \\ &= P(X_1 \leq u) P(X_2 \leq u) \ldots P(X_n \leq u) \\ &\quad \text{since } X_1, X_2, \ldots, X_n \text{ are independent random variables} \\ &= \prod_{i=1}^{n} P(X_i \leq u) \\ &= \prod_{i=1}^{n} F(u) \quad \text{since } X_1, X_2, \ldots, X_n \text{ are identically distributed} \\ &= [F(u)]^n \end{aligned}$$

Suppose A is the support set of X_i, $i = 1, 2, \ldots, n$. The probability density function of U is

$$\begin{aligned} g(u) &= \frac{d}{du} G(u) = \frac{d}{du} [F(u)]^n \\ &= n [F(u)]^{n-1} f(u) \quad \text{for } u \in A \end{aligned}$$

and 0 otherwise.

For $v \in \Re$, the cumulative distribution function of V is

$$\begin{aligned} H(v) &= P(V \leq v) = P[\min(X_1, X_2, \ldots, X_n) \leq v] \\ &= 1 - P[\min(X_1, X_2, \ldots, X_n) > v] \\ &= 1 - P(X_1 > v, X_2 > v, \ldots, X_n > v) \\ &= 1 - P(X_1 > v) P(X_2 > v) \ldots P(X_n > v) \\ &\quad \text{since } X_1, X_2, \ldots, X_n \text{ are independent random variables} \\ &= 1 - \prod_{i=1}^{n} P(X_i > v) \\ &= 1 - \prod_{i=1}^{n} [1 - F(v)] \quad \text{since } X_1, X_2, \ldots, X_n \text{ are identically distributed} \\ &= 1 - [1 - F(v)]^n \end{aligned}$$

Suppose A is the support set of X_i, $i = 1, 2, \ldots, n$. The probability density function of V is

$$\begin{aligned} h(v) &= \frac{d}{dv} H(v) = \frac{d}{dv} \{1 - [1 - F(v)]^n\} \\ &= n [1 - F(v)]^{n-1} f(v) \quad \text{for } v \in A \end{aligned}$$

and 0 otherwise.

4.2 One-to-One Transformations

In this section we look at how to determine the joint distribution of a one-to-one transformation of two or more random variables. We concentrate on the bivariate case for ease of presentation. The method does extend to more than two random variables. See Problems 12 and 13 at the end of this chapter for examples of one-to-one transformations of three random variables.

We begin with some notation and a theorem which gives sufficient conditions for determining whether a transformation is one-to-one in the bivariate case followed by the theorem which gives the joint probability density function for the two new random variables.

Suppose the transformation S defined by

$$\begin{aligned} u &= h_1(x,y) \\ v &= h_2(x,y) \end{aligned}$$

is a one-to-one transformation for all $(x,y) \in R_{XY}$ and that S maps the region R_{XY} into the region R_{UV} in the uv plane. Since $S : (x,y) \to (u,v)$ is a one-to-one transformation there exists a inverse transformation T defined by

$$\begin{aligned} x &= w_1(u,v) \\ y &= w_2(u,v) \end{aligned}$$

such that $T = S^{-1} : (u,v) \to (x,y)$ for all $(u,v) \in R_{UV}$. The Jacobian of the transformation T is

$$\frac{\partial(x,y)}{\partial(u,v)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \left[\frac{\partial(u,v)}{\partial(x,y)}\right]^{-1}$$

where $\frac{\partial(u,v)}{\partial(x,y)}$ is the Jacobian of the transformation S.

4.2.1 Inverse Mapping Theorem

Consider the transformation S defined by

$$\begin{aligned} u &= h_1(x,y) \\ v &= h_2(x,y) \end{aligned}$$

Suppose the partial derivatives $\frac{\partial u}{\partial x}$, $\frac{\partial u}{\partial y}$, $\frac{\partial v}{\partial x}$ and $\frac{\partial v}{\partial y}$ are continuous functions in the neighbourhood of the point (a,b). Suppose also that $\frac{\partial(u,v)}{\partial(x,y)} \neq 0$ at the point (a,b). Then there is a neighbourhood of the point (a,b) in which S has an inverse.

Note: These are sufficient but not necessary conditions for the inverse to exist.

4.2.2 Theorem - One-to-One Bivariate Transformations

Let X and Y be continuous random variables with joint probability density function $f(x,y)$ and let $R_{XY} = \{(x,y) : f(x,y) > 0\}$ be the support set of (X,Y). Suppose the transformation S defined by

$$\begin{aligned} U &= h_1(X,Y) \\ V &= h_2(X,Y) \end{aligned}$$

is a one-to-one transformation with inverse transformation

$$\begin{aligned} X &= w_1(U,V) \\ Y &= w_2(U,V) \end{aligned}$$

Suppose also that S maps R_{XY} into R_{UV}. Then $g(u,v)$, the joint joint probability density function of U and V, is given by

$$g(u,v) = f(w_1(u,v), w_2(u,v)) \left| \frac{\partial(x,y)}{\partial(u,v)} \right|$$

for all $(u,v) \in R_{UV}$. (Compare Theorem 2.6.8 for univariate random variables.)

4.2.3 Proof

We want to find $g(u,v)$, the joint probability density function of the random variables U and V. Suppose S^{-1} maps the region $B \subset R_{UV}$ into the region $A \subset R_{XY}$ then

$$\begin{aligned} P[(U,V) &\in B] \\ &= \iint\limits_B g(u,v)dudv \qquad (4.1) \\ &= P\left[(X,Y) \in A\right] \\ &= \iint\limits_A f(x,y)dxdy \\ &= \iint\limits_B f(w_1(u,v), w_2(u,v)) \left| \frac{\partial(x,y)}{\partial(u,v)} \right| dudv \qquad (4.2) \end{aligned}$$

where the last line follows by the Change of Variable Theorem. Since this is true for all $B \subset R_{UV}$ we have, by comparing (4.1) and (4.2), that the joint probability density function of U and V is given by

$$g(u,v) = f(w_1(u,v), w_2(u,v)) \left| \frac{\partial(x,y)}{\partial(u,v)} \right|$$

for all $(u,v) \in R_{UV}$.

In the following example we see how Theorem 4.2.2 can be used to show that the sum of two independent Exponential(1) random variables is a Gamma random variable.

4.2.4 Example

Suppose $X \sim$ Exponential(1) and $Y \sim$ Exponential(1) independently. Find the joint probability density function of $U = X + Y$ and $V = X$. Show that $U \sim$ Gamma$(2, 1)$.

Solution

Since $X \sim$ Exponential(1) and $Y \sim$ Exponential(1) independently, the joint probability density function of X and Y is

$$\begin{aligned} f(x, y) &= f_1(x) f_2(y) = e^{-x} e^{-y} \\ &= e^{-x-y} \end{aligned}$$

with support set $R_{XY} = \{(x, y) : x > 0, y > 0\}$ which is shown in Figure 4.2.

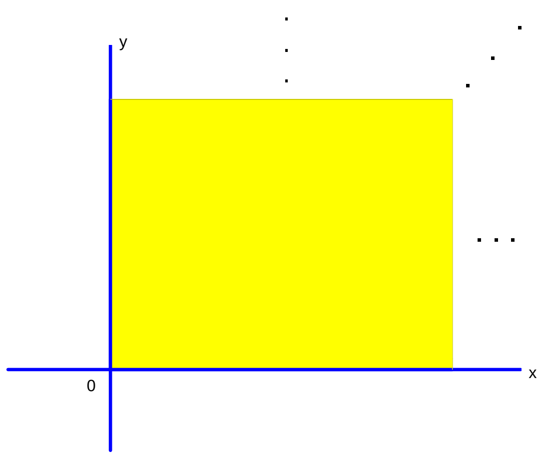


Figure 4.2: Support set R_{XY} for Example 4.2.6

The transformation

$$S: \; U = X + Y, \quad V = X$$

has inverse transformation

$$X = V, \quad Y = U - V$$

Under S the boundaries of R_{XY} are mapped as

$$\begin{aligned} (k, 0) &\rightarrow (k, k) \quad \text{for } k \geq 0 \\ (0, k) &\rightarrow (k, 0) \quad \text{for } k \geq 0 \end{aligned}$$

and the point $(1, 2)$ is mapped to the point $(3, 1)$. Thus S maps R_{XY} into

$$R_{UV} = \{(u, v) : 0 < v < u\}$$

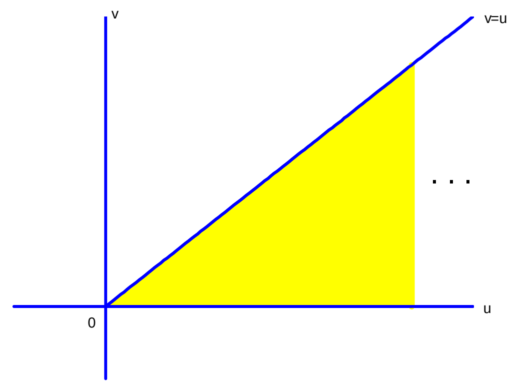


Figure 4.3: Support set R_{UV} for Example 4.2.4

as shown in Figure 4.3.

The Jacobian of the inverse transformation is

$$\frac{\partial(x,y)}{\partial(u,v)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} 0 & 1 \\ 1 & \frac{\partial y}{\partial v} \end{vmatrix} = -1$$

Note that the transformation S is a linear transformation and so we would expect the Jacobian of the transformation to be a constant.

The joint probability density function of U and V is given by

$$\begin{aligned} g(u,v) &= f(w_1(u,v), w_2(u,v)) \left| \frac{\partial(x,y)}{\partial(u,v)} \right| \\ &= f(v, u-v)\,|-1| \\ &= e^{-u} \quad \text{for } (u,v) \in R_{UV} \end{aligned}$$

and 0 otherwise.

To find the marginal probability density functions for U we note that the support set R_{UV} is not rectangular and the range of integration for v will depend on u. The marginal probability density function of U is

$$\begin{aligned} g_1(u) &= \int_{-\infty}^{\infty} g(u,v)\,dv = e^{-u} \int_{v=0}^{u} dv \\ &= ue^{-u} \quad \text{for } u > 0 \end{aligned}$$

and 0 otherwise which is the probability density function of a Gamma$(2,1)$ random variable. Therefore $U \sim$ Gamma$(2,1)$.

In the following exercise we see how the sum and difference of two independent Exponential(1) random variables give a Gamma random variable and a Double Exponential random variable respectively.

4.2.5 Exercise

Suppose $X \sim$ Exponential(1) and $Y \sim$ Exponential(1) independently. Find the joint probability density function of $U = X + Y$ and $V = X - Y$. Show that $U \sim$ Gamma$(2, 1)$ and $V \sim$ Double Exponential$(0, 1)$.

In the following example we see how the Gamma and Beta distributions are related.

4.2.6 Example

Suppose $X \sim$ Gamma$(a, 1)$ and $Y \sim$ Gamma$(b, 1)$ independently. Find the joint probability density function of $U = X + Y$ and $V = \frac{X}{X+Y}$. Show that $U \sim$ Gamma$(a + b, 1)$ and $V \sim$ Beta(a, b) independently. Find $E(V)$ by finding $E\left(\frac{X}{X+Y}\right)$.

Solution

Since $X \sim$ Gamma$(a, 1)$ and $Y \sim$ Gamma$(b, 1)$ independently, the joint probability density function of X and Y is

$$\begin{aligned} f(x, y) &= f_1(x) f_2(y) = \frac{x^{a-1} e^{-x}}{\Gamma(a)} \frac{y^{b-1} e^{-y}}{\Gamma(b)} \\ &= \frac{x^{a-1} y^{b-1} e^{-x-y}}{\Gamma(a)\Gamma(b)} \end{aligned}$$

with support set $R_{XY} = \{(x, y) : x > 0, y > 0\}$ which is the same support set as shown in Figure 4.2.

The transformation

$$S: \; U = X + Y, \quad V = \frac{X}{X+Y}$$

has inverse transformation

$$X = UV, \quad Y = U(1 - V)$$

Under S the boundaries of R_{XY} are mapped as

$$\begin{aligned} (k, 0) &\to (k, 1) \quad \text{for } k \geq 0 \\ (0, k) &\to (k, 0) \quad \text{for } k \geq 0 \end{aligned}$$

and the point $(1, 2)$ is mapped to the point $\left(3, \frac{1}{3}\right)$. Thus S maps R_{XY} into

$$R_{UV} = \{(u, v) : u > 0, \; 0 < v < 1\}$$

as shown in Figure 4.4.

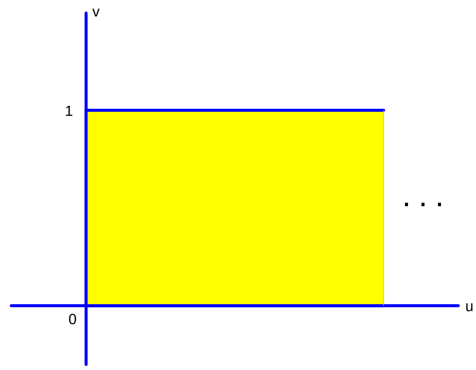


Figure 4.4: Support set R_{UV} for Example 4.2.6

The Jacobian of the inverse transformation is

$$\frac{\partial(x,y)}{\partial(u,v)} = \begin{vmatrix} v & u \\ 1-v & -u \end{vmatrix} = -uv - u + uv = -u$$

The joint probability density function of U and V is given by

$$\begin{aligned} g(u,v) &= f(uv, u(1-v))\,|-u| \\ &= \frac{(uv)^{a-1}[u(1-v)]^{b-1} e^{-u}\,|-u|}{\Gamma(a)\Gamma(b)} \\ &= u^{a+b-1}e^{-u}\frac{v^{a-1}(1-v)^{b-1}}{\Gamma(a)\Gamma(b)} \quad \text{for } (u,v) \in R_{UV} \end{aligned}$$

and 0 otherwise.

To find the marginal probability density functions for U and V we note that the support set of U is $B_1 = \{u : u > 0\}$ and the support set of V is $B_2 = \{v : 0 < v < 1\}$. Since

$$g(u,v) = \underbrace{u^{a+b-1}e^{-u}}_{h_1(u)} \underbrace{\frac{v^{a-1}(1-v)^{b-1}}{\Gamma(a)\Gamma(b)}}_{h_2(v)}$$

for all $(u,v) \in B_1 \times B_2$ then, by the Factorization Theorem for Independence, U and V are independent random variables. Also by the Factorization Theorem for Independence the probability density function of U must be proportional to $h_1(u)$. By writing

$$g(u,v) = \left[\frac{u^{a+b-1}e^{-u}}{\Gamma(a+b)}\right]\left[\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}v^{a-1}(1-v)^{b-1}\right]$$

we note that the function in the first square bracket is the probability density function of a Gamma$(a+b,1)$ random variable and therefore $U \sim$ Gamma$(a+b,1)$. It follows that the function in the second square bracket must be the probability density function of V which is a Beta(a, b) probability density function. Therefore $U \sim$ Gamma$(a+b, 1)$ independently of $V \sim$ Beta(a, b).

In Chapter 2, Problem 9 the moments of a Beta random variable were found by integration. Here is a rather clever way of finding $E(V)$ using the mean of a Gamma random variable. In Exercise 2.7.9 it was shown that the mean of a Gamma(α, β) random variable is $\alpha\beta$.

Now

$$E(UV) = E\left[(X+Y)\frac{X}{(X+Y)}\right] = E(X) = (a)(1) = a$$

since $X \sim$ Gamma$(a, 1)$. But U and V are independent random variables so

$$a = E(UV) = E(U)E(V)$$

But since $U \sim$ Gamma$(a+b, 1)$ we know $E(U) = a+b$ so

$$a = E(U)E(V) = (a+b)E(V)$$

Solving for $E(V)$ gives

$$E(V) = \frac{a}{a+b}$$

Higher moments can be found in a similar manner using the higher moments of a Gamma random variable.

4.2.7 Exercise

Suppose $X \sim$ Beta(a, b) and $Y \sim$ Beta$(a+b, c)$ independently. Find the joint probability density function of $U = XY$ and $V = X$. Show that $U \sim$ Beta$(a, b+c)$.

In the following example we see how a rather unusual transformation can be used to transform two independent Uniform$(0,1)$ random variables into two independent N$(0,1)$ random variables. This transformation is referred to as the Box-Muller Transformation after the two statisticians George E. P. Box and Mervin Edgar Muller who published this result in 1958.

4.2.8 Example - Box-Muller Transformation

Suppose $X \sim$ Uniform$(0,1)$ and $Y \sim$ Uniform$(0,1)$ independently. Find the joint probability density function of

$$\begin{aligned} U &= (-2\log X)^{1/2}\cos(2\pi Y) \\ V &= (-2\log X)^{1/2}\sin(2\pi Y) \end{aligned}$$

Show that $U \sim$ N$(0,1)$ and $V \sim$ N$(0,1)$ independently. Explain how you could use this result to generate independent observations from a N$(0,1)$ distribution.

Solution

Since $X \sim \text{Uniform}(0,1)$ and $Y \sim \text{Uniform}(0,1)$ independently, the joint probability density function of X and Y is

$$\begin{aligned} f(x,y) &= f_1(x)\, f_2(y) = (1)(1) \\ &= 1 \end{aligned}$$

with support set $R_{XY} = \{(x,y) : 0 < x < 1, 0 < y < 1\}$.

Consider the transformation

$$S: \ U = (-2\log X)^{1/2}\cos(2\pi Y), \quad V = (-2\log X)^{1/2}\sin(2\pi Y)$$

To determine the support set of (U,V) we note that $0 < y < 1$ implies $-1 < \cos(2\pi y) < 1$. Also $0 < x < 1$ implies $0 < (-2\log X)^{1/2} < \infty$. Therefore $u = (-2\log x)^{1/2}\cos(2\pi y)$ takes on values in the interval $(-\infty,\infty)$. By a similar argument $v = (-2\log x)^{1/2}\sin(2\pi y)$ also takes on values in the interval $(-\infty,\infty)$. Therefore the support set of (U,V) is $R_{UV} = \Re^2$.

The inverse of the transformation S can be determined. In particular we note that since

$$\begin{aligned} U^2 + V^2 &= \left[(-2\log X)^{1/2}\cos(2\pi Y)\right]^2 + \left[(-2\log X)^{1/2}\sin(2\pi Y)\right]^2 \\ &= (-2\log X)\left[\cos^2(2\pi Y) + \sin^2(2\pi Y)\right] \\ &= -2\log X \end{aligned}$$

and

$$\frac{V}{U} = \frac{\sin(2\pi Y)}{\cos(2\pi Y)} = \tan(2\pi Y)$$

the inverse transformation is

$$X = e^{\frac{-1}{2}\left(U^2+V^2\right)}, \quad Y = \frac{1}{2\pi}\arctan\left(\frac{V}{U}\right)$$

To determine the Jacobian of the inverse transformation it is simpler to use the result

$$\frac{\partial(x,y)}{\partial(u,v)} = \left[\frac{\partial(u,v)}{\partial(x,y)}\right]^{-1} = \left[\left|\begin{array}{cc} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{array}\right|\right]^{-1}$$

Since

$$\begin{aligned} &\left|\begin{array}{cc} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{array}\right| \\ &= \left|\begin{array}{cc} -\frac{1}{x}(-2\log x)^{-1/2}\cos(2\pi y) & -2\pi(-2\log x)^{-1/2}\sin(2\pi y) \\ -\frac{1}{x}(-2\log x)^{-1/2}\sin(2\pi y) & 2\pi(-2\log x)^{-1/2}\cos(2\pi y) \end{array}\right| \\ &= -\frac{2\pi}{x}\left[\cos^2(2\pi y) + \sin^2(2\pi y)\right] \\ &= -\frac{2\pi}{x} \end{aligned}$$

Therefore

$$
\begin{aligned}
\frac{\partial(x,y)}{\partial(u,v)} &= \left[\frac{\partial(u,v)}{\partial(x,y)}\right]^{-1} = \left(-\frac{2\pi}{x}\right)^{-1} \\
&= -\frac{x}{2\pi} \\
&= -\frac{1}{2\pi} e^{\frac{-1}{2}\left(u^2+v^2\right)}
\end{aligned}
$$

The joint probability density function of U and V is

$$
\begin{aligned}
g\left(u,v\right) &= f(w_1(u,v), w_2(u,v)) \left|\frac{\partial(x,y)}{\partial(u,v)}\right| \\
&= (1)\left|-\frac{1}{2\pi} e^{\frac{-1}{2}\left(u^2+v^2\right)}\right| \\
&= \frac{1}{2\pi} e^{\frac{-1}{2}\left(u^2+v^2\right)} \quad \text{for } \ (u,v) \in \Re^2
\end{aligned}
$$

The support set of U is $\Re$ and the support set of V is $\Re$. Since $g\left(u,v\right)$ can be written as

$$
g\left(u,v\right) = \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}\right)\left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}v^2}\right)
$$

for all $(u,v) \in \Re \times \Re = \Re^2$, therefore by the Factorization Theorem for Independence, U and V are independent random variables. We also note that the joint probability density function is the product of two $\mathrm{N}(0,1)$ probability density functions. Therefore $U \sim \mathrm{N}(0,1)$ and $V \sim \mathrm{N}(0,1)$ independently.

Let x and y be two independent $\text{Uniform}(0,1)$ observations which have been generated using a random number generator. Then from the result above we have that

$$
\begin{aligned}
u &= \left(-2\log x\right)^{1/2} \cos\left(2\pi y\right) \\
v &= \left(-2\log x\right)^{1/2} \sin\left(2\pi y\right)
\end{aligned}
$$

are two independent $\mathrm{N}(0,1)$ observations.

The result in the following theorem is one that was used (without proof) in a previous statistics course such as STAT 221/231/241 to construct confidence intervals and test hypotheses regarding the mean in a $\mathrm{N}\left(\mu, \sigma^2\right)$ model when the variance σ^2 is unknown.

4.2.9 Theorem - t Distribution

If $X \sim \chi^2(n)$ independently of $Z \sim \mathrm{N}(0,1)$ then

$$
T = \frac{Z}{\sqrt{X/n}} \sim \mathrm{t}(n)
$$

Proof

The transformation $T = \frac{Z}{\sqrt{X/n}}$ is not a one-to-one transformation. However if we add the variable $U = X$ to *complete the transformation* and consider the transformation

$$S: \ T = \frac{Z}{\sqrt{X/n}}, \quad U = X$$

then this transformation has an inverse transformation given by

$$X = U, \quad Z = T\left(\frac{U}{n}\right)^{1/2}$$

Since $X \sim \chi^2(n)$ independently of $Z \sim \mathrm{N}(0,1)$ the joint probability density function of X and Z is

$$\begin{aligned} f(x,z) &= f_1(x) f_2(z) = \frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-x/2} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \\ &= \frac{1}{2^{(n+1)/2}\Gamma(n/2)\sqrt{\pi}} x^{n/2-1} e^{-x/2} e^{-z^2/2} \end{aligned}$$

with support set $R_{XZ} = \{(x,z) : x > 0, z \in \Re\}$. The transformation S maps R_{XZ} into $R_{TU} = \{(t,u) : t \in \Re, u > 0\}$.

The Jacobian of the inverse transformation is

$$\frac{\partial(x,z)}{\partial(t,u)} = \begin{vmatrix} \frac{\partial x}{\partial t} & \frac{\partial x}{\partial u} \\ \frac{\partial z}{\partial t} & \frac{\partial z}{\partial u} \end{vmatrix} = \begin{vmatrix} 0 & 1 \\ \left(\frac{u}{n}\right)^{1/2} & \frac{\partial z}{\partial u} \end{vmatrix} = -\left(\frac{u}{n}\right)^{1/2}$$

The joint probability density function of U and V is given by

$$\begin{aligned} g(t,u) &= f\left(t\left(\frac{u}{n}\right)^{1/2}, u\right) \left| -\left(\frac{u}{n}\right)^{1/2} \right| \\ &= \frac{1}{2^{(n+1)/2}\Gamma(n/2)\sqrt{\pi}} u^{n/2-1} e^{-u/2} e^{-t^2 u/(2n)} \left(\frac{u}{n}\right)^{1/2} \\ &= \frac{1}{2^{(n+1)/2}\Gamma(n/2)\sqrt{\pi n}} u^{(n+1)/2-1} e^{-u\left(1+t^2/n\right)/2} \quad \text{for } (t,u) \in R_{TU} \end{aligned}$$

and 0 otherwise.

To determine the distribution of T we need to find the marginal probability density function for T.

$$\begin{aligned} g_1(t) &= \int_{-\infty}^{\infty} g(t,u)\, du \\ &= \frac{1}{2^{(n+1)/2}\Gamma(n/2)\sqrt{\pi n}} \int_0^{\infty} u^{(n+1)/2-1} e^{-u\left(1+t^2/n\right)/2} du \end{aligned}$$

Let $y = \frac{u}{2}\left(1+\frac{t^2}{n}\right)$ so that $u = 2y\left(1+\frac{t^2}{n}\right)^{-1}$ and $du = 2\left(1+\frac{t^2}{n}\right)^{-1} dy$. Note that when $u=0$ then $y=0$, and when $u\to\infty$ then $y\to\infty$. Therefore

$$\begin{aligned}
g_1(t) &= \frac{1}{2^{(n+1)/2}\Gamma(n/2)\sqrt{\pi n}} \int_0^\infty \left[2y\left(1+\frac{t^2}{n}\right)^{-1}\right]^{(n+1)/2-1} e^{-y}\left[2\left(1+\frac{t^2}{n}\right)^{-1}\right] dy \\
&= \frac{1}{2^{(n+1)/2}\Gamma(n/2)\sqrt{\pi n}} 2^{(n+1)/2}\left[\left(1+\frac{t^2}{n}\right)^{-1}\right]^{(n+1)/2} \int_0^\infty y^{(n+1)/2-1}e^{-y}dy \\
&= \frac{1}{\Gamma(n/2)\sqrt{\pi n}}\left(1+\frac{t^2}{n}\right)^{-(n+1)/2}\Gamma\left(\frac{n+1}{2}\right) \\
&= \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma(n/2)\sqrt{\pi n}}\left(1+\frac{t^2}{n}\right)^{-(n+1)/2} \qquad \text{for } t\in\Re
\end{aligned}$$

which is the probability density function of a random variable with a $\mathrm{t}(n)$ distribution. Therefore

$$T = \frac{Z}{\sqrt{X/n}} \sim t(n)$$

as required.

4.2.10 Example

Use Theorem 4.2.9 to find $E(T)$ and $Var(T)$ if $T \sim \mathrm{t}(n)$.

Solution

If $X \sim \chi^2(n)$ independently of $Z \sim \mathrm{N}(0,1)$ then we know from the previous theorem that

$$T = \frac{Z}{\sqrt{X/n}} \sim \mathrm{t}(n)$$

Now

$$E(T) = E\left(\frac{Z}{\sqrt{X/n}}\right) = \sqrt{n}E(Z)\,E\left(X^{-1/2}\right)$$

since X and Z are independent random variables. Since $E(Z)=0$ it follows that $E(T)=0$

as long as $E\left(X^{-1/2}\right)$ exists. Since $X \sim \chi^2(n)$

$$
\begin{aligned}
E\left(X^k\right) &= \int_0^\infty x^k \frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-x/2} dx \\
&= \frac{1}{2^{n/2}\Gamma(n/2)} \int_0^\infty x^{k+n/2-1} e^{-x/2} dx \quad \text{let } y = x/2 \\
&= \frac{1}{2^{n/2}\Gamma(n/2)} \int_0^\infty (2y)^{k+n/2-1} e^{-y} (2)\, dy \\
&= \frac{2^{k+n/2}}{2^{n/2}\Gamma(n/2)} \int_0^\infty y^{k+n/2-1} e^{-y} dy \\
&= \frac{2^k \Gamma(n/2+k)}{\Gamma(n/2)} \qquad (4.3)
\end{aligned}
$$

which exists for $n/2 + k > 0$. If $k = -1/2$ then the integral exists for $n/2 > 1/2$ or $n > 1$. Therefore

$$E(T) = 0 \quad \text{for } n > 1$$

Now

$$
\begin{aligned}
Var(T) &= E\left(T^2\right) - [E(T)]^2 \\
&= E\left(T^2\right) \quad \text{since } E(T) = 0
\end{aligned}
$$

and

$$E\left(T^2\right) = E\left(\frac{Z^2}{X/n}\right) = nE\left(Z^2\right) E\left(X^{-1}\right)$$

Since $Z \sim \mathrm{N}(0,1)$ then

$$
\begin{aligned}
E\left(Z^2\right) &= Var(Z) + [E(Z)]^2 = 1 + 0^2 \\
&= 1
\end{aligned}
$$

Also by (4.3)

$$
\begin{aligned}
E\left(X^{-1}\right) &= \frac{2^{-1}\Gamma(n/2-1)}{\Gamma(n/2)} = \frac{1}{2(n/2-1)} \\
&= \frac{1}{n-2}
\end{aligned}
$$

which exists for $n > 2$. Therefore

$$
\begin{aligned}
Var(T) &= E\left(T^2\right) = nE\left(Z^2\right) E\left(X^{-1}\right) \\
&= n(1)\left(\frac{1}{n-2}\right) \\
&= \frac{n}{n-2} \quad \text{for } n > 2
\end{aligned}
$$

The following theorem concerns the F distribution which is used in testing hypotheses about the parameters in a multiple linear regression model.

4.2.11 Theorem - F Distribution

If $X \sim \chi^2(n)$ independently of $Y \sim \chi^2(m)$ then

$$U = \frac{X/n}{Y/m} \sim \mathrm{F}(n, m)$$

4.2.12 Exercise

(a) Prove Theorem 4.2.11. Hint: Complete the transformation with $V = Y$.

(b) Find $E(U)$ and $Var(U)$ and note for what values of n and m that they exist.

Hint: Use the technique and results of Example 4.2.10.

4.3 Moment Generating Function Technique

The moment generating function technique is particularly useful in determining the distribution of a sum of two or more independent random variables if the moment generating functions of the random variables exist.

4.3.1 Theorem

Suppose $X_1, X_2, \ldots, X_n$ are independent random variables and X_i has moment generating function $M_i(t)$ which exists for $t \in (-h, h)$ for some $h > 0$. The moment generating function of $Y = \sum_{i=1}^{n} X_i$ is given by

$$M_Y(t) = \prod_{i=1}^{n} M_i(t)$$

for $t \in (-h, h)$.

If the X_i's are independent and identically distributed random variables each with moment generating function $M(t)$ then $Y = \sum_{i=1}^{n} X_i$ has moment generating function

$$M_Y(t) = [M(t)]^n$$

for $t \in (-h, h)$.

Proof

The moment generating function of $Y = \sum_{i=1}^{n} X_i$ is

$$\begin{aligned}
M_Y(t) &= E\left(e^{tY}\right) \\
&= E\left[\exp\left(t\sum_{i=1}^{n} X_i\right)\right] \\
&= \prod_{i=1}^{n} E\left(e^{tX_i}\right) \quad \text{since } X_1, X_2, \ldots, X_n \text{ are independent random variables} \\
&= \prod_{i=1}^{n} M_i(t) \quad \text{for } t \in (-h, h)
\end{aligned}$$

If $X_1, X_2, \ldots, X_n$ are identically distributed each with moment generating function $M(t)$ then

$$\begin{aligned}
M_Y(t) &= \prod_{i=1}^{n} M(t) \\
&= [M(t)]^n \quad \text{for } t \in (-h, h)
\end{aligned}$$

as required.

Note: This theorem in conjunction with the Uniqueness Theorem for Moment Generating Functions can be used to find the distribution of Y.

Here is a summary of results about sums of random variables for the named distributions.

4.3.2 Special Results

(1) If $X_i \sim$ Binomial(n_i, p), $i = 1, 2, \ldots, n$ independently, then

$$\sum_{i=1}^{n} X_i \sim \text{ Binomial}\left(\sum_{i=1}^{n} n_i,\ p\right)$$

(2) If $X_i \sim$ Poisson(μ_i), $i = 1, 2, \ldots, n$ independently, then

$$\sum_{i=1}^{n} X_i \sim \text{ Poisson}\left(\sum_{i=1}^{n} \mu_i\right)$$

(3) If $X_i \sim$ Negative Binomial(k_i, p), $i = 1, 2, \ldots, n$ independently, then

$$\sum_{i=1}^{n} X_i \sim \text{ Negative Binomial}\left(\sum_{i=1}^{n} k_i,\ p\right)$$

(4) If $X_i \sim$ Exponential(β), $i = 1, 2, \ldots, n$ independently, then

$$\sum_{i=1}^{n} X_i \sim \text{ Gamma}(n, \beta)$$

(5) If $X_i \sim$ Gamma(α_i, β), $i = 1, 2, \ldots, n$ independently, then

$$\sum_{i=1}^{n} X_i \sim \text{ Gamma}\left(\sum_{i=1}^{n} \alpha_i,\ \beta\right)$$

(6) If $X_i \sim \chi^2(k_i)$, $i = 1, 2, \ldots, n$ independently, then

$$\sum_{i=1}^{n} X_i \sim \chi^2\left(\sum_{i=1}^{n} k_i\right)$$

(7) If $X_i \sim \mathrm{N}(\mu, \sigma^2)$, $i = 1, 2, \ldots, n$ independently, then

$$\sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2(n)$$

Proof

(1) Suppose $X_i \sim \text{Binomial}(n_i, p)$, $i = 1, 2, \ldots, n$ independently. The moment generating function of X_i is

$$M_i(t) = \left(pe^t + q\right)^{n_i} \quad \text{for } t \in \Re$$

for $i = 1, 2, \ldots, n$. By Theorem 4.3.1 the moment generating function of $Y = \sum_{i=1}^{n} X_i$ is

$$\begin{aligned} M_Y(t) &= \prod_{i=1}^{n} M_i(t) \\ &= \prod_{i=1}^{n} \left(pe^t + q\right)^{n_i} \\ &= \left(pe^t + q\right)^{\sum_{i=1}^{n} n_i} \quad \text{for } t \in \Re \end{aligned}$$

which is the moment generating function of a $\text{Binomial}\left(\sum_{i=1}^{n} n_i,\ p\right)$ random variable. Therefore by the Uniqueness Theorem for Moment Generating Functions $\sum_{i=1}^{n} X_i \sim \text{Binomial}\left(\sum_{i=1}^{n} n_i,\ p\right)$.

(2) Suppose $X_i \sim \text{Poisson}(\mu_i)$, $i = 1, 2, \ldots, n$ independently. The moment generating function of X_i is

$$M_i(t) = e^{\mu_i\left(e^t - 1\right)} \quad \text{for } t \in \Re$$

for $i = 1, 2, \ldots, n$. By Theorem 4.3.1 the moment generating function of $Y = \sum_{i=1}^{n} X_i$ is

$$\begin{aligned} M_Y(t) &= \prod_{i=1}^{n} M_i(t) \\ &= \prod_{i=1}^{n} e^{\mu_i\left(e^t - 1\right)} \\ &= e^{\left(\sum_{i=1}^{n} \mu_i\right)\left(e^t - 1\right)} \quad \text{for } t \in \Re \end{aligned}$$

which is the moment generating function of a $\text{Poisson}\left(\sum_{i=1}^{n} \mu_i\right)$ random variable. Therefore by the Uniqueness Theorem for Moment Generating Functions $\sum_{i=1}^{n} X_i \sim \text{Poisson}\left(\sum_{i=1}^{n} \mu_i\right)$.

(3) Suppose $X_i \sim$ Negative Binomial$(k_i,\ p)$, $i = 1, 2, \ldots, n$ independently. The moment generating function of X_i is

$$M_i(t) = \left(\frac{p}{1 - qe^t}\right)^{k_i} \quad \text{for } t < -\log(q)$$

for $i = 1, 2, \ldots, n$. By Theorem 4.3.1 the moment generating function of $Y = \sum\limits_{i=1}^{n} X_i$ is

$$\begin{aligned} M_Y(t) &= \prod_{i=1}^{n} M_i(t) \\ &= \prod_{i=1}^{n} \left(\frac{p}{1 - qe^t}\right)^{k_i} \\ &= \left(\frac{p}{1 - qe^t}\right)^{\sum\limits_{i=1}^{n} k_i} \quad \text{for } t < -\log(q) \end{aligned}$$

which is the moment generating function of a Negative Binomial$\left(\sum\limits_{i=1}^{n} k_i,\ p\right)$ random variable. Therefore by the Uniqueness Theorem for Moment Generating Functions
$\sum\limits_{i=1}^{n} X_i \sim$ Negative Binomial$\left(\sum\limits_{i=1}^{n} k_i,\ p\right)$.

(4) Suppose $X_i \sim$ Exponential(β), $i = 1, 2, \ldots, n$ independently. The moment generating function of each X_i is

$$M(t) = \frac{1}{1 - \beta t} \quad \text{for } t < \frac{1}{\beta}$$

for $i = 1, 2, \ldots, n$. By Theorem 4.3.1 the moment generating function of $Y = \sum\limits_{i=1}^{n} X_i$ is

$$M_Y(t) = [M(t)]^n = \left(\frac{1}{1 - \beta t}\right)^n = \quad \text{for } t < \frac{1}{\beta}$$

which is the moment generating function of a Gamma(n, β) random variable. Therefore by the Uniqueness Theorem for Moment Generating Functions $\sum\limits_{i=1}^{n} X_i \sim$ Gamma(n, β).

(5) Suppose $X_i \sim$ Gamma(α_i, β), $i = 1, 2, \ldots, n$ independently. The moment generating function of X_i is

$$M_i(t) = \left(\frac{1}{1 - \beta t}\right)^{\alpha_i} \quad \text{for } t < \frac{1}{\beta}$$

for $i = 1, 2, \ldots, n$. By Theorem 4.3.1 the moment generating function of $Y = \sum\limits_{i=1}^{n} X_i$ is

$$\begin{aligned} M_Y(t) &= \prod_{i=1}^{n} M_i(t) = \prod_{i=1}^{n} \left(\frac{1}{1 - \beta t}\right)^{\alpha_i} \\ &= \left(\frac{1}{1 - \beta t}\right)^{\sum\limits_{i=1}^{n} \alpha_i} \quad \text{for } t < \frac{1}{\beta} \end{aligned}$$

which is the moment generating function of a Gamma$\left(\sum\limits_{i=1}^{n}\alpha_i,\ \beta\right)$ random variable. Therefore by the Uniqueness Theorem for Moment Generating Functions $\sum\limits_{i=1}^{n} X_i \sim \text{Gamma}\left(\sum\limits_{i=1}^{n}\alpha_i,\ \beta\right)$.

(6) Suppose $X_i \sim \chi^2(k_i)$, $i = 1, 2, \ldots, n$ independently. The moment generating function of X_i is

$$M_i(t) = \left(\frac{1}{1-2t}\right)^{k_i} \quad \text{for } t < \frac{1}{2}$$

for $i = 1, 2, \ldots, n$. By Theorem 4.3.1 the moment generating function of $Y = \sum\limits_{i=1}^{n} X_i$ is

$$\begin{aligned} M_Y(t) &= \prod_{i=1}^{n} M_i(t) \\ &= \prod_{i=1}^{n}\left(\frac{1}{1-2t}\right)^{k_i} \\ &= \left(\frac{1}{1-2t}\right)^{\sum\limits_{i=1}^{n} k_i} \quad \text{for } t < \frac{1}{2} \end{aligned}$$

which is the moment generating function of a $\chi^2\left(\sum\limits_{i=1}^{n} k_i\right)$ random variable. Therefore by the Uniqueness Theorem for Moment Generating Functions $\sum\limits_{i=1}^{n} X_i \sim \chi^2\left(\sum\limits_{i=1}^{n} k_i\right)$.

(7) Suppose $X_i \sim \text{N}(\mu, \sigma^2)$, $i = 1, 2, \ldots, n$ independently. Then by Example 2.6.9 and Theorem 2.6.3

$$\left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2(1) \quad \text{for } i = 1, 2, \ldots, n$$

and by (6)

$$\sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2(n)$$

4.3.3 Exercise

Suppose $X_1, X_2, \ldots, X_n$ are independent and identically distributed random variables with moment generating function $M(t)$, $E(X_i) = \mu$, and $Var(X_i) = \sigma^2 < \infty$. Give an expression for the moment generating function of $Z = \sqrt{n}\left(\bar{X} - \mu\right)/\sigma$ where $\bar{X} = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$ in terms of $M(t)$.

The following theorem is one that was used in your previous probability and statistics courses without proof. The method of moment generating functions now allows us to easily proof this result.

4.3.4 Theorem - Linear Combination of Independent Normal Random Variables

If $X_i \sim \mathrm{N}(\mu_i, \sigma_i^2)$, $i = 1, 2, \ldots, n$ independently, then

$$\sum_{i=1}^{n} a_i X_i \sim \mathrm{N}\left(\sum_{i=1}^{n} a_i \mu_i,\ \sum_{i=1}^{n} a_i^2 \sigma_i^2\right)$$

Proof

Suppose $X_i \sim \mathrm{N}(\mu_i, \sigma_i^2)$, $i = 1, 2, \ldots, n$ independently. The moment generating function of X_i is

$$M_i(t) = e^{\mu_i t + \sigma_i^2 t^2/2} \quad \text{for } t \in \Re$$

for $i = 1, 2, \ldots, n$. The moment generating function of $Y = \sum_{i=1}^{n} a_i X_i$ is

$$\begin{aligned}
M_Y(t) &= E\left(e^{tY}\right) \\
&= E\left[\exp\left(t \sum_{i=1}^{n} a_i X_i\right)\right] \\
&= \prod_{i=1}^{n} E\left(e^{(a_i t) X_i}\right) \quad \text{since } X_1, X_2, \ldots, X_n \text{ are independent random variables} \\
&= \prod_{i=1}^{n} M_i(a_i t) \\
&= \prod_{i=1}^{n} e^{\mu_i a_i t + \sigma_i^2 a_i^2 t^2/2} \\
&= \exp\left[\left(\sum_{i=1}^{n} a_i \mu_i\right) t + \left(\sum_{i=1}^{n} a_i^2 \sigma_i^2\right) t^2/2\right] \quad \text{for } t \in \Re
\end{aligned}$$

which is the moment generating function of a $\mathrm{N}\left(\sum_{i=1}^{n} a_i \mu_i,\ \sum_{i=1}^{n} a_i^2 \sigma_i^2\right)$ random variable. Therefore by the Uniqueness Theorem for Moment Generating Functions

$$\sum_{i=1}^{n} a_i X_i \sim \mathrm{N}\left(\sum_{i=1}^{n} a_i \mu_i,\ \sum_{i=1}^{n} a_i^2 \sigma_i^2\right)$$

4.3.5 Corollary

If $X_i \sim \mathrm{N}(\mu, \sigma^2)$, $i = 1, 2, \ldots, n$ independently then

$$\sum_{i=1}^{n} X_i \sim \mathrm{N}\left(n\mu, n\sigma^2\right)$$

and

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \sim \mathrm{N}\left(\mu,\ \frac{\sigma^2}{n}\right)$$

Proof

To prove that

$$\sum_{i=1}^{n} X_i \sim \mathrm{N}\left(n\mu, n\sigma^2\right)$$

let $a_i = 1$, $\mu_i = \mu$, and $\sigma_i^2 = \sigma^2$ in Theorem 4.3.4 to obtain

$$\sum_{i=1}^{n} X_i \sim \mathrm{N}\left(\sum_{i=1}^{n} \mu,\ \sum_{i=1}^{n} \sigma^2\right)$$

or

$$\sum_{i=1}^{n} X_i \sim \mathrm{N}\left(n\mu, n\sigma^2\right)$$

To prove that

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim \mathrm{N}\left(\mu,\ \frac{\sigma^2}{n}\right)$$

we note that

$$\bar{X} = \sum_{i=1}^{n}\left(\frac{1}{n}\right) X_i$$

Let $a_i = \frac{1}{n}$, $\mu_i = \mu$, and $\sigma_i^2 = \sigma^2$ in Theorem 4.3.4 to obtain

$$\bar{X} = \sum_{i=1}^{n}\left(\frac{1}{n}\right) X_i \sim \mathrm{N}\left(\sum_{i=1}^{n}\left(\frac{1}{n}\right)\mu,\ \sum_{i=1}^{n}\left(\frac{1}{n}\right)^2\sigma^2\right)$$

or

$$\bar{X} \sim \mathrm{N}\left(\mu,\ \frac{\sigma^2}{n}\right)$$

The following identity will be used in proving Theorem 4.3.8.

4.3.6 Useful Identity

$$\sum_{i=1}^{n} (X_i - \mu)^2 = \sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 + n\left(\bar{X} - \mu\right)^2$$

4.3.7 Exercise

Prove the identity 4.3.6

As mentioned previously the t distribution is used to construct confidence intervals and test hypotheses regarding the mean in a $\mathrm{N}(\mu, \sigma^2)$ model. We are now able to prove the theorem on which these results are based.

4.3.8 Theorem

If $X_i \sim \mathrm{N}(\mu, \sigma^2)$, $i = 1, 2, \ldots, n$ independently then

$$\bar{X} \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

independently of

$$\frac{(n-1)\,S^2}{\sigma^2} = \frac{\sum\limits_{i=1}^{n} \left(X_i - \bar{X}\right)^2}{\sigma^2} \sim \chi^2(n-1)$$

where

$$S^2 = \frac{\sum\limits_{i=1}^{n} \left(X_i - \bar{X}\right)^2}{n-1}$$

Proof

For a proof that $\bar{X}$ and S^2 are independent random variables please see Problem 16.

By identity 4.3.6

$$\sum_{i=1}^{n} (X_i - \mu)^2 = \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2 + n\left(\bar{X} - \mu\right)^2$$

Dividing both sides by σ^2 gives

$$\underbrace{\sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2}_{Y} = \underbrace{\frac{(n-1)\,S^2}{\sigma^2}}_{U} + \underbrace{\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right)^2}_{V}$$

Since $\bar{X}$ and S^2 are independent random variables, it follows that U and V are independent random variables.

By 4.3.2(7)

$$Y = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2(n)$$

with moment generating function

$$M_Y(t) = (1 - 2t)^{-n/2} \quad \text{for } t < \frac{1}{2} \tag{4.4}$$

$\bar{X} \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right)$ was proved in Corollary 4.3.5. By Example 2.6.9

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathrm{N}(0, 1)$$

and by Theorem 2.6.3

$$V = \left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right)^2 \sim \chi^2(1)$$

with moment generating function

$$M_V(t) = (1-2t)^{-1/2} \quad \text{for } t < \frac{1}{2} \tag{4.5}$$

Since U and V are independent random variables and $Y = U + V$ then

$$M_Y(t) = E\left(e^{tY}\right) = E\left(e^{t(U+V)}\right) = E\left(e^{tU}\right)E\left(e^{tV}\right) = M_U(t)\,M_V(t) \tag{4.6}$$

Substituting (4.4) and (4.5) into (4.6) gives

$$(1-2t)^{-n/2} = M_U(t)\,(1-2t)^{-1/2} \quad \text{for } t < \frac{1}{2}$$

or

$$M_U(t) = (1-2t)^{-(n-1)/2} \quad \text{for } t < \frac{1}{2}$$

which is the moment generating function of a $\chi^2(n-1)$ random variable. Therefore by the Uniqueness Theorem for Moment Generating Functions

$$U = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

4.3.9 Theorem

If $X_i \sim \mathrm{N}(\mu, \sigma^2)$, $i = 1, 2, \ldots, n$ independently then

$$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim \mathrm{t}(n-1)$$

Proof

$$\frac{\bar{X}-\mu}{S/\sqrt{n}} = \frac{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{\frac{(n-1)S^2}{\sigma^2}}{n-1}}} = \frac{Z}{\sqrt{\frac{U}{n-1}}}$$

where

$$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim \mathrm{N}(0,1)$$

independently of

$$U = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

Therefore by Theorem 4.2.9

$$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim \mathrm{t}(n-1)$$

The following theorem is useful for testing the equality of variances in a two sample Normal model.

4.3.10 Theorem

Suppose $X_1, X_2, \ldots, X_n$ are independent $\mathrm{N}\big(\mu_1, \sigma_1^2\big)$ random variables, and independently $Y_1, Y_2, \ldots, Y_m$ are independent $\mathrm{N}\big(\mu_2, \sigma_2^2\big)$ random variables. Let

$$S_1^2 = \frac{\sum\limits_{i=1}^{n} \left(X_i - \bar{X}\right)^2}{n-1} \quad \text{and} \quad S_2^2 = \frac{\sum\limits_{i=1}^{m} \left(Y_i - \bar{Y}\right)^2}{m-1}$$

Then

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim \mathrm{F}\left(n-1, m-1\right)$$

4.3.11 Exercise

Prove Theorem 4.3.10. Hint: Use Theorems 4.3.8 and 4.2.11.

4.4 Chapter 4 Problems

1. Show that if X and Y are independent random variables then $U = h(X)$ and $V = g(Y)$ are also independent random variables where h and g are real-valued functions.

2. Suppose X and Y are continuous random variables with joint probability density function

$$f(x, y) = 24xy \quad \text{for } 0 < x + y < 1, \quad 0 < x < 1, \quad 0 < y < 1$$

and 0 otherwise.

   (a) Find the joint probability density function of $U = X + Y$ and $V = X$. Be sure to specify the support set of (U, V).

   (b) Find the marginal probability density function of U and the marginal probability density function of V. Be sure to specify their support sets.

3. Suppose X and Y are continuous random variables with joint probability density function

$$f(x, y) = e^{-y} \quad \text{for } 0 < x < y < \infty$$

and 0 otherwise.

   (a) Find the joint probability density function of $U = X + Y$ and $V = X$. Be sure to specify the support set of (U, V).

   (b) Find the marginal probability density function of U and the marginal probability density function of V. Be sure to specify their support sets.

4. Suppose X and Y are nonnegative continuous random variables with joint probability density function $f(x, y)$. Show that the probability density function of $U = X + Y$ is given by

$$g(u) = \int_0^\infty f(v, u - v)\, dv$$

Hint: Consider the transformation $U = X + Y$ and $V = X$.

5. Suppose X and Y are continuous random variables with joint probability density function

$$f(x, y) = 2(x + y) \quad \text{for } 0 < x < y < 1$$

and 0 otherwise.

   (a) Find the joint probability density function of $U = X$ and $V = XY$. Be sure to specify the support set of (U, V).

(b) Are U and V independent random variables?

(c) Find the marginal probability density function's of U and V. Be sure to specify their support sets.

6. Suppose X and Y are continuous random variables with joint probability density function

$$f(x,y) = 4xy \quad \text{for } 0 < x < 1, \quad 0 < y < 1$$

and 0 otherwise.

(a) Find the probability density function of $T = X + Y$ using the cumulative distribution function technique.

(b) Find the joint probability density function of $S = X$ and $T = X + Y$. Find the marginal probability density function of T and compare your answer to the one you obtained in (a).

(c) Find the joint probability density function of $U = X^2$ and $V = XY$. Be sure to specify the support set of (U, V).

(d) Find the marginal probability density function's of U and V.

(e) Find $E(V^3)$. (Hint: Are X and Y independent random variables?)

7. Suppose X and Y are continuous random variables with joint probability density function

$$f(x,y) = 4xy \quad \text{for } 0 < x < 1, \quad 0 < y < 1$$

and 0 otherwise.

(a) Find the joint probability density function of $U = X/Y$ and $V = XY$. Be sure to specify the support set of (U, V).

(b) Are U and V independent random variables?

(c) Find the marginal probability density function's of U and V. Be sure to specify their support sets.

8. Suppose X and Y are independent Uniform$(0, \theta)$ random variables. Find the probability density function of $U = X - Y$.
(Hint: Complete the transformation with $V = X + Y$.)

9. Suppose $Z_1 \sim \mathrm{N}(0,1)$ and $Z_2 \sim \mathrm{N}(0,1)$ independently. Let

$$X_1 = \mu_1 + \sigma_1 Z_1, \quad X_2 = \mu_2 + \sigma_2[\rho Z_1 + \left(1 - \rho^2\right)^{1/2} Z_2]$$

where $-\infty < \mu_1,\ \mu_2 < \infty$, $\sigma_1,\ \sigma_2 > 0$ and $-1 < \rho < 1$.

(a) Show that $(X_1, X_2)^T \sim \mathrm{BVN}(\mu, \Sigma)$.

(b) Show that $(X-\mu)^T\Sigma^{-1}(X-\mu) \sim \chi^2(2)$. Hint: Show $(X-\mu)^T\Sigma^{-1}(X-\mu) = Z^TZ$ where $Z = (Z_1, Z_2)^T$.

10. Suppose $X \sim \mathrm{N}(\mu, \sigma^2)$ and $Y \sim \mathrm{N}(\mu, \sigma^2)$ independently. Let $U = X + Y$ and $V = X - Y$.

    (a) Find the joint moment generating function of U and V.

    (b) Use (a) to show that U and V are independent random variables.

11. Let X and Y be independent $\mathrm{N}(0,1)$ random variables and let $U = X/Y$.

    (a) Show that $U \sim \mathrm{Cauchy}(1,0)$.

    (b) Show that the Cauchy$(1,0)$ probability density function is the same as the $\mathrm{t}(1)$ probability density function

12. Let X_1, X_2, X_3 be independent Exponential(1) random variables. Let the random variables Y_1, Y_2, Y_3 be defined by

$$Y_1 = \frac{X_1}{X_1 + X_2}, \quad Y_2 = \frac{X_1 + X_2}{X_1 + X_2 + X_3}, \quad Y_3 = X_1 + X_2 + X_3$$

    Show that Y_1, Y_2, Y_3 are independent random variables and find their marginal probability density function's.

13. Let X_1, X_2, X_3 be independent $\mathrm{N}(0,1)$ random variables. Let the random variables Y_1, Y_2, Y_3 be defined by

$$\begin{aligned} X_1 &= Y_1 \cos Y_2 \sin Y_3, \quad X_2 = Y_1 \sin Y_2 \sin Y_3, \quad X_3 = Y_1 \cos Y_3 \\ \text{for } 0 &< y_1 < \infty, \quad 0 < y_2 < 2\pi, \quad 0 < y_3 < \pi \end{aligned}$$

    Show that Y_1, Y_2, Y_3 are independent random variables and find their marginal probability density function's.

14. Suppose $X_1, X_2, \ldots, X_n$ is a random sample from the Poisson(μ) distribution. Find the conditional probability function of $X_1, X_2, \ldots, X_n$ given $T = \sum\limits_{i=1}^{n} X_i = t$.

15. Suppose $X \sim \chi^2(n)$, $X + Y \sim \chi^2(m)$, $m > n$ and X and Y independent random variables. Use the properties of moment generating functions to show that $Y \sim \chi^2(m-n)$.

16. Suppose $X_1, X_2, \ldots, X_n$ is a random sample from the $\mathrm{N}(\mu, \sigma^2)$ distribution. In this problem we wish to show that $\bar{X}$ and $\sum_{i=1}^{n}(X_i - \bar{X})^2$ are independent random variables. Note that this implies that $\bar{X}$ and $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ are also independent random variables.
Let $U = (U_1, U_2, \ldots, U_n)$ where $U_i = X_i - \bar{X}$, $i = 1, 2, \ldots, n$ and let

$$M(s_1, s_2, \ldots, s_n, s) = E[\exp(\sum_{i=1}^{n} s_i U_i + s\bar{X})]$$

be the joint moment generating function of U and $\bar{X}$.

(a) Let $t_i = s_i - \bar{s} + s/n$, $i = 1, 2, \ldots, n$ where $\bar{s} = \frac{1}{n}\sum_{i=1}^{n} s_i$. Show that

$$E[\exp(\sum_{i=1}^{n} s_i U_i + s\bar{X})] = E[\exp(\sum_{i=1}^{n} t_i X_i)] = \exp\left[\mu \sum_{i=1}^{n} t_i + \sigma^2 \sum_{i=1}^{n} t_i^2/2\right]$$

Hint: Since $X_i \sim \mathrm{N}(\mu, \sigma^2)$

$$E[\exp(t_i X_i)] = \exp\left[\mu t_i + \sigma^2 t_i^2/2\right]$$

(b) Verify that $\sum_{i=1}^{n} t_i = s$ and $\sum_{i=1}^{n} t_i^2 = \sum_{i=1}^{n}(s_i - \bar{s})^2 + s^2/n$.

(c) Use (a) and (b) to show that

$$M(s_1, s_2, \ldots, s_n, s) = \exp[\mu s + (\sigma^2/n)(s^2/2)]\exp[\sigma^2 \sum_{i=1}^{n}(s_i - \bar{s})^2/2]$$

(d) Show that the random variable $\bar{X}$ is independent of the random vector U and thus $\bar{X}$ and $\sum_{i=1}^{n}(X_i - \bar{X})^2$ are independent. Hint: $M_{\bar{X}}(s) = M(0, 0, \ldots, 0, s)$ and $M_U(s_1, s_2, \ldots, s_n) = M(s_1, s_2, \ldots, s_n, 0)$.

5. Limiting or Asymptotic Distributions

In a previous probability course the Poisson approximation to the Binomial distribution

$$\binom{n}{x} p^x (1-p)^{n-x} \approx \frac{(np)^x e^{-np}}{x!} \quad \text{for } x = 0, 1, \ldots, n$$

if n is large and p is small was used.

As well the Normal approximation to the Binomial distribution

$$\begin{aligned} P(X_n \leq x) &= \binom{n}{x} p^x (1-p)^{n-x} \quad \text{for } x = 0, 1, \ldots, n \\ &\approx P\left(Z \leq \frac{x - np}{\sqrt{np(1-p)}}\right) \quad \text{where } Z \sim \mathrm{N}(0,1) \end{aligned}$$

if n is large and p is close to $1/2$ (a special case of the very important Central Limit Theorem) was used. These are examples of what we will call limiting or asymptotic distributions.

In this chapter we consider a sequence of random variables $X_1, X_2, \ldots, X_n, \ldots$ and look at the definitions and theorems related to determining the limiting distribution of such a sequence. In Section 5.1 we define convergence in distribution and look at several examples to illustrate its meaning. In Section 5.2 we define convergence in probability and examine its relationship to convergence in distribution. In Section 5.3 we look at the Weak Law of Large Numbers which is an important theorem when examining the behaviour of estimators of unknown parameters (Chapter 6). In Section 5.4 we use the moment generating function to find limiting distributions including a proof of the Central Limit Theorem. The Central Limit Theorem was used in STAT 221/231/241 to construct an approximate confidence interval for an unknown parameter. In Section 5.5 additional limit theorems for finding limiting distributions are introduced. These additional theorems allow us to determine new limiting distributions by combining the limiting distributions which have been determined from definitions, the Weak Law of Large Numbers, and/or the Central Limit Theorem.

5.1 Convergence in Distribution

In calculus you studied sequences of real numbers $a_1, a_2, \ldots, a_n, \ldots$ and learned theorems which allowed you to evaluate limits such as $\lim\limits_{n\to\infty} a_n$. In this course we are interested in a sequence of random variables $X_1, X_2, \ldots, X_n, \ldots$ and what happens to the distribution of X_n as $n \to \infty$. We do this be examining what happens to $F_n(x) = P(X_n \leq x)$, the cumulative distribution function of X_n, as $n \to \infty$. Note that for a fixed value of x, the sequence $F_1(x), F_2(x), \ldots, F_n(x) \ldots$ is a sequence of real numbers. In general we will obtain a different sequence of real numbers for each different value of x. Since we have a sequence of real numbers we will be able to use limit theorems you have used in your previous calculus courses to evaluate $\lim\limits_{n\to\infty} F_n(x)$. We will need to take care in determining how $F_n(x)$ behaves as $n \to \infty$ for all real values of x. To formalize these ideas we give the following definition for *convergence in distribution of a sequence of random variables*.

5.1.1 Definition - Convergence in Distribution

Let $X_1, X_2, \ldots, X_n, \ldots$ be a sequence of random variables and let $F_1(x), F_2(x), \ldots, F_n(x), \ldots$ be the corresponding sequence of cumulative distribution functions, that is, X_n has cumulative distribution function $F_n(x) = P(X_n \leq x)$. Let X be a random variable with cumulative distribution function $F(x) = P(X \leq x)$. We say X_n *converges in distribution* to X and write

$$X_n \to_D X$$

if

$$\lim_{n\to\infty} F_n(x) = F(x)$$

at all points x at which $F(x)$ is continuous. We call F the *limiting* or *asymptotic distribution of* X_n.

Note:

(1) Although we say the random variable X_n converges in distribution to the random variable X, the definition of convergence in distribution is defined in terms of the pointwise convergence of the corresponding sequence of cumulative distribution functions.

(2) This definition holds for both discrete and continuous random variables.

(3) One way to think about convergence in distribution is that, if $X_n \to_D X$, then for large n

$$F_n(x) = P(X_n \leq x) \approx F(x) = P(X \leq x)$$

if x is a point of continuity of $F(x)$. How good the approximation is will depend on the values of n and x.

The following theorem and corollary will be useful in determining limiting distributions.

5.1.2 Theorem - e Limit

If b and c are real constants and $\lim_{n\to\infty} \psi(n) = 0$ then

$$\lim_{n\to\infty} \left[1 + \frac{b}{n} + \frac{\psi(n)}{n}\right]^{cn} = e^{bc}$$

5.1.3 Corollary

If b and c are real constants then

$$\lim_{n\to\infty} \left(1 + \frac{b}{n}\right)^{cn} = e^{bc}$$

5.1.4 Example

Let $Y_i \sim$ Exponential(1) , $i = 1, 2, \ldots$ independently. Consider the sequence of random variables $X_1, X_2, \ldots, X_n, \ldots$ where $X_n = \max(Y_1, Y_2, \ldots, Y_n) - \log n$. Find the limiting distribution of X_n.

Solution

Since $Y_i \sim$ Exponential(1)

$$P(Y_i \leq y) = \begin{cases} 0 & y \leq 0 \\ 1 - e^{-y} & y > 0 \end{cases}$$

for $i = 1, 2, \ldots$. Since the Y_i's are independent random variables

$$\begin{aligned}
F_n(x) &= P(X_n \leq x) = P(\max(Y_1, Y_2, \ldots, Y_n) - \log n \leq x) \\
&= P(\max(Y_1, Y_2, \ldots, Y_n) \leq x + \log n) \\
&= P(Y_1 \leq x + \log n, Y_2 \leq x + \log n, \ldots, Y_n \leq x + \log n) \\
&= \prod_{i=1}^{n} P(Y_i \leq x + \log n) \\
&= \prod_{i=1}^{n} \left(1 - e^{-(x+\log n)}\right) \quad \text{for } x + \log n > 0 \\
&= \left(1 - \frac{e^{-x}}{n}\right)^n \quad \text{for } x > -\log n
\end{aligned}$$

As $n \to \infty$, $-\log n \to -\infty$ so

$$\begin{aligned}
\lim_{n\to\infty} F_n(x) &= \lim_{n\to\infty} \left(1 + \frac{(-e^{-x})}{n}\right)^n \\
&= e^{-e^{-x}} \quad \text{for } x \in \Re
\end{aligned}$$

by 5.1.3.

Consider the function

$$F(x) = e^{-e^{-x}} \quad \text{for } x \in \Re$$

Since $F'(x) = e^{-x}e^{-e^{-x}} > 0$ for all $x \in \Re$, therefore $F(x)$ is a continuous, increasing function for $x \in \Re$. Also $\lim_{x\to-\infty} F(x) = \lim_{x\to-\infty} e^{-e^{-x}} = 0$, and $\lim_{x\to\infty} F(x) = \lim_{x\to\infty} e^{-e^{-x}} = 1$. Therefore $F(x)$ is a cumulative distribution function for a continuous random variable.

Let X be a random variable with cumulative distribution function $F(x)$. Since

$$\lim_{n\to\infty} F_n(x) = F(x)$$

for all $x \in \Re$, that is at all points x at which $F(x)$ is continuous, therefore

$$X_n \to_D X$$

In Figure 5.1 you can see how quickly the curves $F_n(x)$ approach the limiting curve $F(x)$.

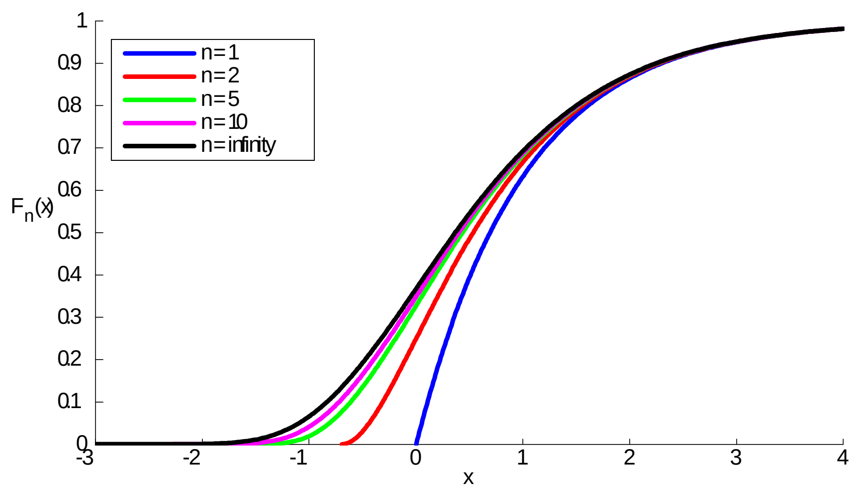


Figure 5.1: Graphs of $F_n(x) = \left(1 - \frac{e^{-x}}{n}\right)^n$ for $n = 1, 2, 5, 10, \infty$

5.1.5 Example

Let $Y_i \sim \text{Uniform}(0, \theta)$, $i = 1, 2, \ldots$ independently. Consider the sequence of random variables $X_1, X_2, \ldots, X_n, \ldots$ where $X_n = \max(Y_1, Y_2, \ldots, Y_n)$. Find the limiting distribution of X_n.

Solution

Since $Y_i \sim \text{Uniform}(0, \theta)$

$$P(Y_i \le y) = \begin{cases} 0 & y \le 0 \\ \frac{y}{\theta} & 0 < y < \theta \\ 1 & y \ge \theta \end{cases}$$

for $i = 1, 2, \ldots$. Since the Y_i's are independent random variables

$$\begin{aligned}
F_n(x) &= P(X_n \le x) = P(\max(Y_1, Y_2, \ldots, Y_n) \le x) \\
&= P(Y_1 \le x, Y_2 \le x, \ldots, Y_n \le x) = \prod_{i=1}^{n} P(Y_i \le x) \\
&= \begin{cases} \prod\limits_{i=1}^{n} 0 & x \le 0 \\ \prod\limits_{i=1}^{n} \frac{x}{\theta} & 0 < x < \theta \\ \prod\limits_{i=1}^{n} 1 & x \ge \theta \end{cases} \\
&= \begin{cases} 0 & x \le 0 \\ \left(\frac{x}{\theta}\right)^n & 0 < x < \theta \\ 1 & x \ge \theta \end{cases}
\end{aligned}$$

Therefore

$$\lim_{n\to\infty} F_n(x) = \begin{cases} 0 & x < \theta \\ 1 & x \ge \theta \end{cases} = F(x)$$

In Figure 5.2 you can see how quickly the curves $F_n(x)$ approach the limiting curve $F(x)$.

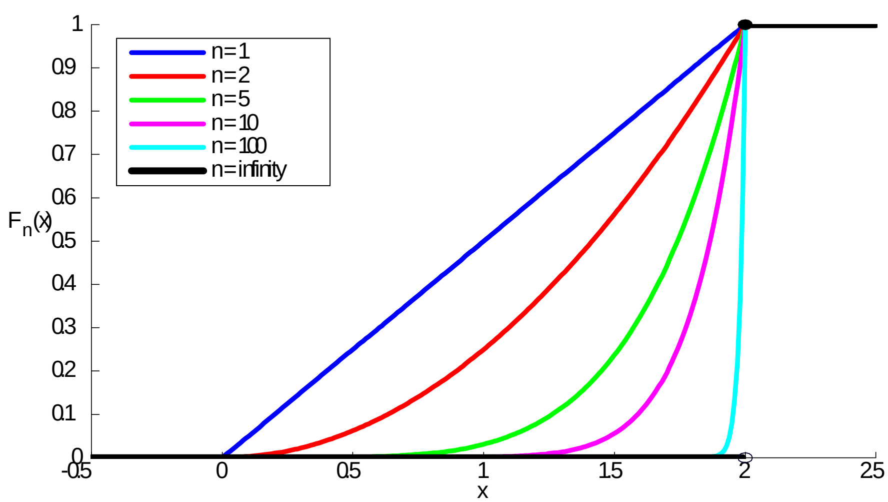


Figure 5.2: Graphs of $F_n(x) = \left(\frac{x}{\theta}\right)^n$ for $\theta = 2$ and $n = 1, 2, 5, 10, 100, \infty$

It is straightforward to check that $F(x)$ is a cumulative distribution function for the discrete random variable X with probability function

$$f(x) = \begin{cases} 1 & y = \theta \\ 0 & \text{otherwise} \end{cases}$$

Therefore

$$X_n \to_D X$$

Since X only takes on one value with probability one, X is called a degenerate random variable. When X_n converges in distribution to a degenerate random variable we also call this *convergence in probability to a constant* as defined in the next section.

5.1.6 Comment

Suppose $X_1, X_2, \ldots, X_n, \ldots$is a sequence of random variables such that $X_n \to_D X$. Then for large n we can use the approximation

$$P(X_n \leq x) \approx P(X \leq x)$$

If X is degenerate at b then $P(X = b) = 1$ and this approximation is not very useful. However, if the limiting distribution is degenerate then we could use this result in another way. In Example 5.1.5 we showed that if $Y_i \sim \text{Uniform}(0, \theta)$, $i = 1, 2, \ldots, n$ independently then $X_n = \max(Y_1, Y_2, \ldots, Y_n)$ converges in distribution to a degenerate random variable X with $P(X = \theta) = 1$. This result is rather useful since, if we have observed data $y_1, y_2, \ldots, y_n$ from a Uniform$(0, \theta)$ distribution and θ is unknown, then this suggests using $y_{(n)} = \max(y_1, y_2, \ldots, y_n)$ as an estimate of θ if n is reasonably large. We will discuss this idea in more detail in Chapter 6.

5.2 Convergence in Probability

Definition 5.1.1 is useful for finding the limiting distribution of a sequence of random variables $X_1, X_2, \ldots, X_n, \ldots$ when the sequence of corresponding cumulative distribution functions $F_1, F_2, \ldots, F_n, \ldots$ can be obtained. In other cases we may use the following definition to determine the limiting distribution.

5.2.1 Definition - Convergence in Probability

A sequence of random variables $X_1, X_2, \ldots, X_n, \ldots$ *converges in probability* to a random variable X if, for all $\varepsilon > 0$,

$$\lim_{n\to\infty} P(|X_n - X| \geq \varepsilon) = 0$$

or equivalently

$$\lim_{n\to\infty} P(|X_n - X| < \varepsilon) = 1$$

We write

$$X_n \to_p X$$

Convergence in probability is a stronger former of convergence than convergence in distribution in the sense that convergence in probability implies convergence in distribution as stated in the following theorem. However, if X_n converges in distribution to X, then X_n may or may not converge in probability to X.

5.2.2 Theorem - Convergence in Probability Implies Convergence in Distribution

If $X_n \rightarrow_p X$ then $X_n \rightarrow_D X$.

In Example 5.1.5 the limiting distribution was degenerate. When the limiting distribution is degenerate we say X_n converges in probability to a constant. The following definition, which follows from Definition 5.2.1, can be used for proving convergence in probability.

5.2.3 Definition - Convergence in Probability to a Constant

A sequence of random variables $X_1, X_2, \ldots, X_n, \ldots$ *converges in probability* to a constant b if, for all $\varepsilon > 0$,

$$\lim_{n\to\infty} P\left(|X_n - b| \geq \varepsilon\right) = 0$$

or equivalently

$$\lim_{n\to\infty} P\left(|X_n - b| < \varepsilon\right) = 1$$

We write

$$X_n \rightarrow_p b$$

5.2.4 Example

Suppose $X_1, X_2, \ldots, X_n, \ldots$ is a sequence of random variables with $E(X_n) = \mu_n$ and $Var(X_n) = \sigma_n^2$. If $\lim\limits_{n\to\infty} \mu_n = a$ and $\lim\limits_{n\to\infty} \sigma_n^2 = 0$ then show $X_n \rightarrow_p a$.

Solution

To show $X_n \rightarrow_p a$ we need to show that for all $\varepsilon > 0$

$$\lim_{n\to\infty} P\left(|X_n - a| < \varepsilon\right) = 1$$

or equivalently

$$\lim_{n\to\infty} P\left(|X_n - a| \geq \varepsilon\right) = 0$$

Recall Markov's Inequality. For all $k,\ c > 0$

$$P\left(|X| \geq c\right) \leq \frac{E\left(\left|X^k\right|\right)}{c^k}$$

Therefore by Markov's Inequality with $k = 2$ and $c = \varepsilon$ we have

$$0 \leq \lim_{n\to\infty} P\left(|X_n - a| \geq \varepsilon\right) \leq \lim_{n\to\infty} \frac{E\left(|X_n - a|^2\right)}{\varepsilon^2} \tag{5.1}$$

Now

$$\begin{aligned} E\left(|X_n - a|^2\right) &= E\left[(X_n - a)^2\right] \\ &= E\left[(X_n - \mu_n)^2\right] + 2\left(\mu_n - a\right) E\left(X_n - \mu_n\right) + \left(\mu_n - a\right)^2 \end{aligned}$$

Since

$$\lim_{n\to\infty} E\left([X_n - \mu_n]^2\right) = \lim_{n\to\infty} Var\left(X_n\right) = \lim_{n\to\infty} \sigma_n^2 = 0$$

and

$$\lim_{n\to\infty} (\mu_n - a) = 0 \quad \text{since} \quad \lim_{n\to\infty} \mu_n = a$$

therefore

$$\lim_{n\to\infty} E\left(|X_n - a|^2\right) = \lim_{n\to\infty} E\left[(X_n - a)^2\right] = 0$$

which also implies

$$\lim_{n\to\infty} \frac{E\left(|X_n - a|^2\right)}{\varepsilon^2} = 0 \tag{5.2}$$

Thus by (5.1), (5.2) and the Squeeze Theorem

$$\lim_{n\to\infty} P\left(|X_n - a| \geq \varepsilon\right) = 0$$

for all $\varepsilon > 0$ as required.

The proof in Example 5.2.4 used Definition 5.2.3 to prove convergence in probability. The reason for this is that the distribution of the X_i's was not specified. Only conditions on $E(X_n)$ and $Var(X_n)$ were specified. This means that the result in Example 5.2.4 holds for any sequence of random variables $X_1, X_2, \ldots, X_n, \ldots$ satisfying the given conditions.

If the sequence of corresponding cumulative distribution functions $F_1, F_2, \ldots, F_n, \ldots$ can be obtained then the following theorem can also be used to prove convergence in probability to a constant.

5.2.5 Theorem

Suppose $X_1, X_2, \ldots, X_n, \ldots$ is a sequence of random variables such that X_n has cumulative distribution function $F_n(x)$. If

$$\lim_{n\to\infty} F_n(x) = \lim_{n\to\infty} P\left(X_n \leq x\right) = \begin{cases} 0 & x < b \\ 1 & x > b \end{cases}$$

then $X_n \to_p b$.

Note: We do not need to worry about whether $\lim\limits_{n\to\infty} F_n(b)$ exists since $x = b$ is a point of discontinuity of the limiting distribution (see Definition 5.1.1).

5.2.6 Example

Let $Y_i \sim$ Exponential$(\theta, 1)$, $i = 1, 2, \ldots$ independently. Consider the sequence of random variables $X_1, X_2, \ldots, X_n, \ldots$ where $X_n = \min(Y_1, Y_2, \ldots, Y_n)$. Show that $X_n \to_p \theta$.

Solution

Since $Y_i \sim$ Exponential$(\theta, 1)$

$$P(Y_i > y) = \begin{cases} e^{-(y-\theta)} & y > \theta \\ 1 & y \leq \theta \end{cases}$$

for $i = 1, 2, \ldots$. Since $Y_1, Y_2, \ldots, Y_n$ are independent random variables

$$\begin{aligned} F_n(x) &= 1 - P(X_n > x) = 1 - P(\min(Y_1, Y_2, \ldots, Y_n) > x) \\ &= 1 - P(Y_1 > x, Y_2 > x, \ldots, Y_n > x) = 1 - \prod_{i=1}^{n} P(Y_i > x) \\ &= \begin{cases} 1 - \prod\limits_{i=1}^{n} 1 & x \leq \theta \\ 1 - \prod\limits_{i=1}^{n} e^{-(x-\theta)} & x > \theta \end{cases} \\ &= \begin{cases} 0 & x \leq \theta \\ 1 - e^{-n(x-\theta)} & x > \theta \end{cases} \end{aligned}$$

Therefore

$$\lim_{n \to \infty} F_n(x) = \begin{cases} 0 & x \leq \theta \\ 1 & x > \theta \end{cases}$$

which we note is not a cumulative distribution function since the function is not right-continuous at $x = \theta$. However

$$\lim_{n \to \infty} F_n(x) = \begin{cases} 0 & x < \theta \\ 1 & x > \theta \end{cases}$$

and therefore by Theorem 5.2.5, $X_n \to_p \theta$. In Figure 5.3 you can see how quickly the limit is approached.

5.3 Weak Law of Large Numbers

In this section we look at a very important result which we will use in Chapter 6 to show that maximum likelihood estimators have good properties. This result is called the Weak Law of Large Numbers. Needless to say there is another law called the Strong Law of Large Numbers but we will not consider this law here.

Also in this section we will look at some simulations to illustrate the theoretical result in the Weak Law of Large Numbers.

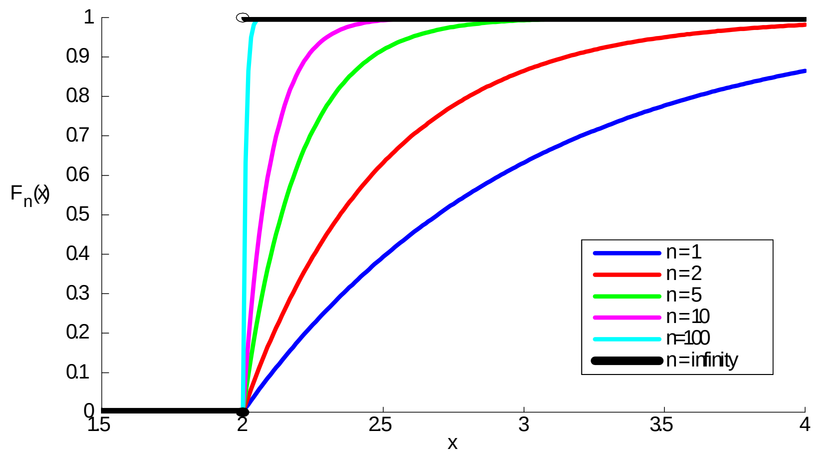


Figure 5.3: Graphs of $F_n(x) = 1 - e^{-n(x-\theta)}$ for $\theta = 2$

5.3.1 Weak Law of Large Numbers

Suppose $X_1, X_2, \ldots$ are independent and identically distributed random variables with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2 < \infty$. Consider the sequence of random $\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_n, \ldots$ where

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Then

$$\bar{X}_n \to_p \mu$$

Proof

Using Definition 5.2.3 we need to show

$$\lim_{n\to\infty} P\left(\left|\bar{X}_n - \mu\right| \geq \varepsilon\right) = 0 \quad \text{for all } \varepsilon > 0$$

We apply Chebyshev's Theorem (see 2.8.2) to the random variable $\bar{X}_n$ where $E\left(\bar{X}_n\right) = \mu$ (see Corollary 3.6.3(3)) and $Var\left(\bar{X}_n\right) = \sigma^2/n$ (see Theorem 3.6.7(4)) to obtain

$$P\left(\left|\bar{X}_n - \mu\right| \geq \frac{k\sigma}{\sqrt{n}}\right) \leq \frac{1}{k^2} \quad \text{for all } k > 0$$

Let $k = \frac{\sqrt{n}\varepsilon}{\sigma}$. Then

$$0 \leq P\left(\left|\bar{X}_n - \mu\right| \geq \varepsilon\right) \leq \frac{\sigma^2}{n\varepsilon} \quad \text{for all } \varepsilon > 0$$

Since

$$\lim_{n\to\infty} \frac{\sigma^2}{n\varepsilon} = 0$$

therefore by the Squeeze Theorem

$$\lim_{n\to\infty} P\left(\left|\bar{X}_n - \mu\right| \geq \varepsilon\right) = 0 \quad \text{for all } \varepsilon > 0$$

as required.

Notes:

(1) The proof of the Weak Law of Large Numbers does not actually require that the random variables be identically distributed, only that they all have the same mean and variance. As well the proof does not require knowing the distribution of these random variables.

(2) In words the Weak Law of Large Numbers says that the sample mean $\bar{X}_n$ approaches the population mean μ as $n \to \infty$.

5.3.2 Example

If $X \sim$ Pareto$(1, \theta)$ then X has probability density function

$$f(x) = \frac{\theta}{x^{\theta+1}} \quad \text{for } x \geq 1$$

and 0 otherwise. X has cumulative distribution function

$$F(x) = \begin{cases} 0 & \text{if } x < 1 \\ 1 - \frac{1}{x^\theta} & \text{for } x \geq 1 \end{cases}$$

and inverse cumulative distribution function

$$F^{-1}(x) = (1-x)^{-1/\theta} \quad \text{for } 0 < x < 1$$

Also

$$E(X) = \begin{cases} \infty & \text{if } 0 < \theta \leq 1 \\ \frac{\theta}{\theta-1} & \text{if } \theta > 1 \end{cases}$$

and

$$Var(X) = \frac{\theta}{(\theta-1)^2(\theta-2)}$$

Suppose $X_1, X_2, \ldots, X_n$ are independent and identically distributed Pareto$(1, \theta)$ random variables. By the Weak Law of Large Numbers

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \to_p E(X) = \frac{\theta}{\theta-1} \quad \text{for } \theta > 1$$

If $U_i \sim$ Uniform$(0,1)$, $i = 1, 2, \ldots, n$ independently then by Theorem 2.6.6

$$\begin{aligned} X_i &= F^{-1}(U_i) \\ &= (1-U_i)^{-1/\theta} \sim \text{Pareto}(1, \theta) \end{aligned}$$

$i = 1, 2, \ldots, n$ independently. If we generate Uniform$(0,1)$ observations $u_1, u_2, \ldots, u_n$ using a random number generator and then let $x_i = (1-u_i)^{-1/\theta}$ then $x_1, x_2, \ldots, x_n$ are observations from the Pareto$(1, \theta)$ distribution.

The points (i, x_i), $i = 1, 2, \ldots, 500$ for one simulation of 500 observations from a Pareto$(1, 5)$ distribution are plotted in Figure 5.4.

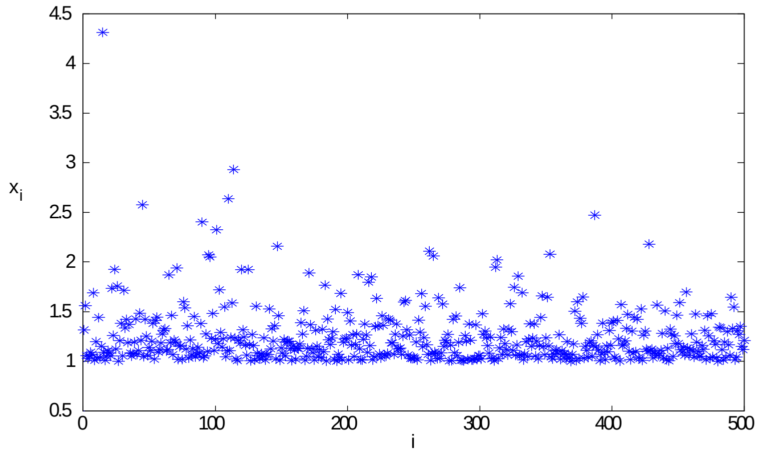


Figure 5.4: 500 observations from a Pareto$(1, 5)$ distribution

Figure 5.5 shows a plot of the points $(n, \bar{x}_n)$, $n = 1, 2, \ldots, 500$ where $\bar{x}_n = \frac{1}{n}\sum\limits_{i=1}^{n} x_i$ is the sample mean. We note that the sample mean $\bar{x}_n$ is approaching the population mean $\mu = E(X) = \frac{5}{5-1} = 1.25$ as n increases.

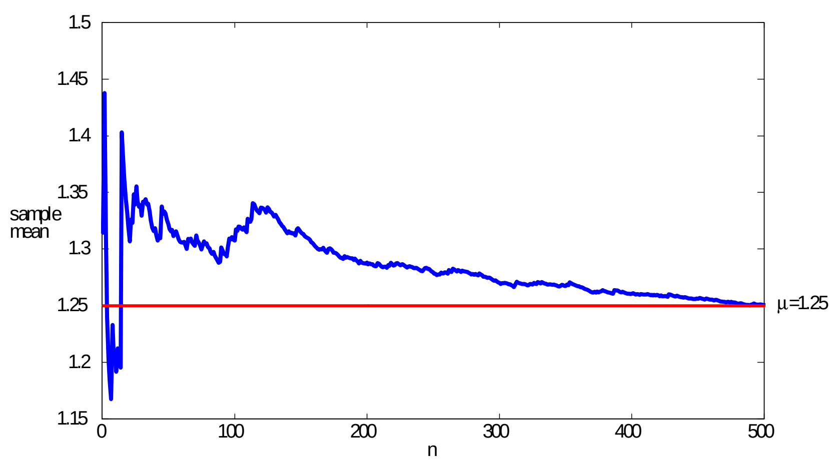


Figure 5.5: Graph of $\bar{x}_n$ versus n for 500 Pareto$(1, 5)$ observations

If we generate a further 1000 values of x_i, and plot (i, x_i), $i = 1, 2, \ldots, 1500$ we obtain the graph in Figure 5.6.

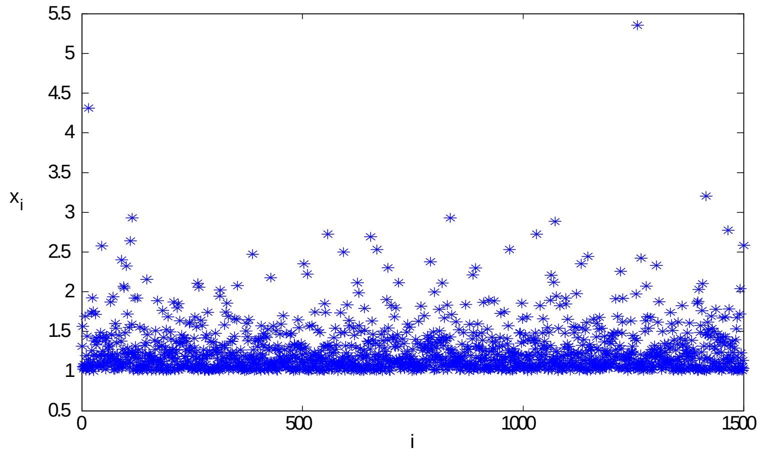


Figure 5.6: 1500 observations from a Pareto$(1, 5)$ distribution

The corresponding plot of $(n, \bar{x}_n)$, $n = 1, 2, \ldots, 1500$ in shown in Figure 5.7. We note that the sample mean $\bar{x}_n$ stays very close to the population mean $\mu = 1.25$ for $n > 500$.

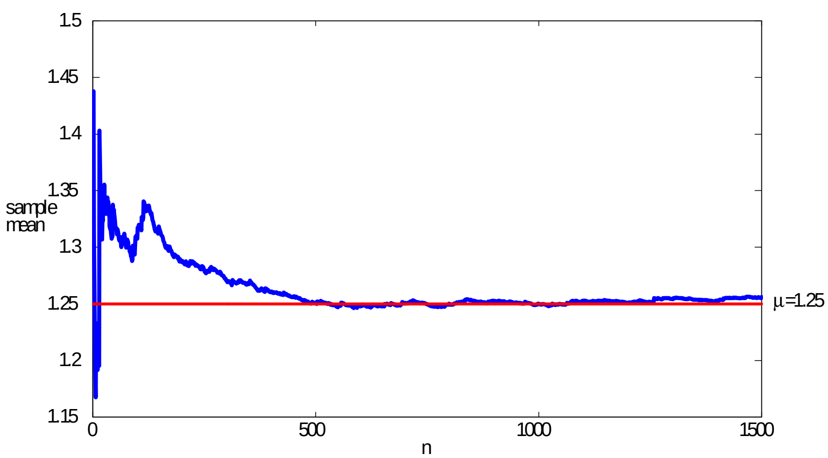


Figure 5.7: Graph of $\bar{x}_n$ versus n for 1500 Pareto$(1, 5)$ observations

Note that these figures correspond to only one set of simulated data. If we generated another set of data using a random number generator the actual data points would change. However what would stay the same is that the sample mean for the new data set would still approach the mean value $E(X) = 1.25$ as n increases.

The points (i, x_i), $i = 1, 2, \ldots, 500$ for one simulation of 500 observations from a Pareto$(1, 0.5)$ distribution are plotted in Figure 5.8. For this distribution

$$E(X) = \int_1^\infty x \frac{\theta}{x^{\theta+1}} dx = \theta \int_1^\infty \frac{1}{x^\theta} dx \quad \text{which diverges to } \infty$$

Note in Figure 5.8 that the are some very large observations. In particular there is one observation which is close to $40,000$.

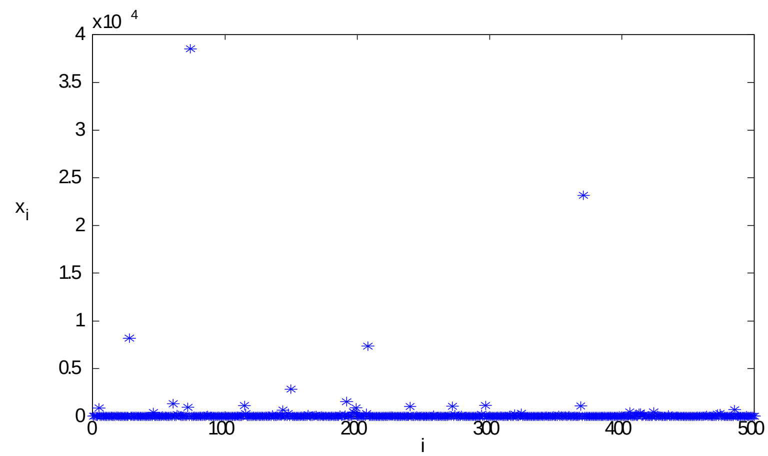


Figure 5.8: 500 observations from a Pareto$(1, 0.5)$ distribution

The corresponding plot of $(n, \bar{x}_n)$, $n = 1, 2, \ldots, 500$ for these data is given in Figure 5.9. Note that the mean $\bar{x}_n$ does not appear to be approaching a fixed value.

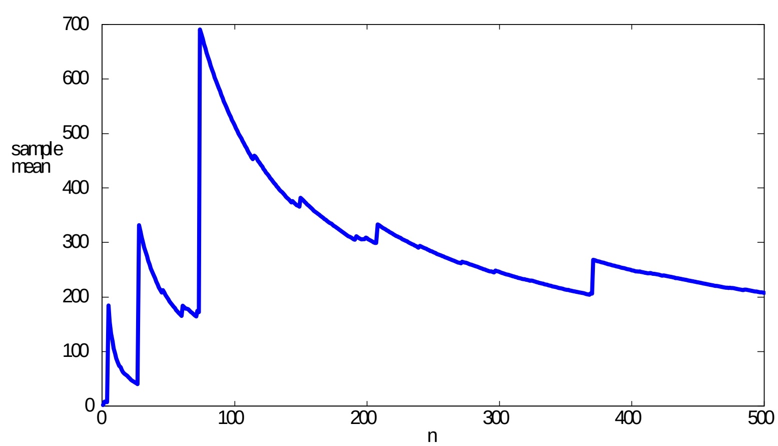


Figure 5.9: Graph of $\bar{x}_n$ versus n for 500 Pareto$(1, 0.5)$ observations

In Figure 5.10 the points $(n, \bar{x}_n)$ for a set of $50,000$ observations generated from a Pareto$(1, 0.5)$ distribution are plotted. Note that the mean $\bar{x}_n$ does not approach a fixed value and in general is getting larger as n gets large. This is consistent with $E(X)$ diverging to ∞.

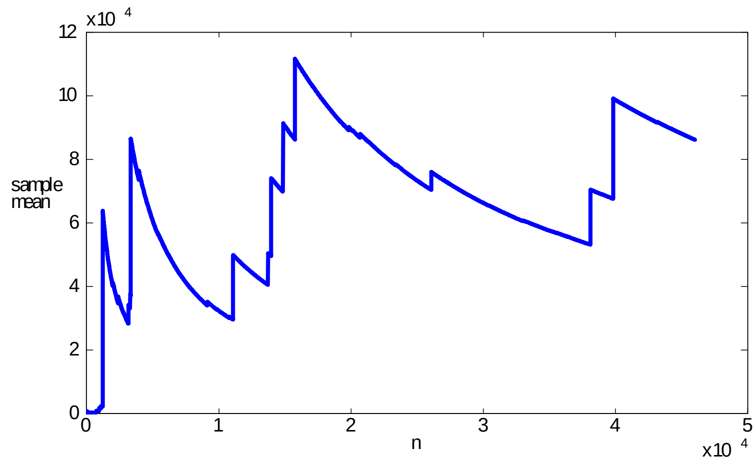


Figure 5.10: Graph of $\bar{x}_n$ versus n for 50000 Pareto$(1, 0.5)$ observations

5.3.3 Example

If $X \sim$ Cauchy$(0, 1)$ then

$$f(x) = \frac{1}{\pi(1 + x^2)} \quad \text{for } x \in \Re$$

The probability density function, shown in Figure 5.11, is symmetric about the y axis and the median of the distribution is equal to 0.

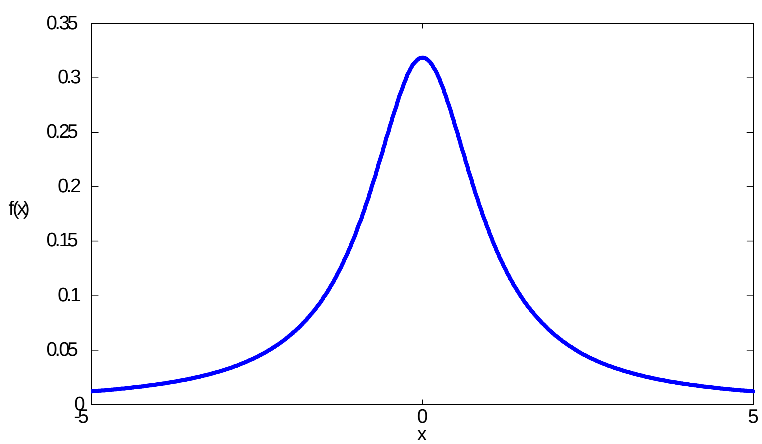


Figure 5.11: Cauchy$(0, 1)$ probability density function

The mean of a Cauchy$(0,1)$ random variable does not exist since

$$\frac{1}{\pi}\int_{-\infty}^{0} \frac{x}{1+x^2}dx \quad \text{diverges to } -\infty \tag{5.3}$$

and

$$\frac{1}{\pi}\int_{0}^{\infty} \frac{x}{1+x^2}dx \quad \text{diverges to } \infty \tag{5.4}$$

The cumulative distribution function of a Cauchy$(0,1)$ random variable is

$$F(x) = \frac{1}{\pi}\arctan x + \frac{1}{2} \quad \text{for } x \in \Re$$

and the inverse cumulative distribution function is

$$F^{-1}(x) = \tan\left[\pi\left(x - \frac{1}{2}\right)\right] \quad \text{for } 0 < x < 1$$

If we generate Uniform$(0,1)$ observations $u_1, u_2, \ldots, u_N$ using a random number generator and then let $x_i = \tan\left[\pi\left(u_i - \frac{1}{2}\right)\right]$, then the x_i's are observations from the Cauchy$(0,1)$ distribution.

The points $(n, \bar{x}_n)$ for a set of $500,000$ observations generated from a Cauchy$(0,1)$ distribution are plotted in Figure 5.12. Note that $\bar{x}_n$ does not approach a fixed value. However, unlike the Pareto$(1, 0.5)$ example in which $\bar{x}_n$ was getting larger as n got larger, we see in Figure 5.12 that $\bar{x}_n$ drifts back and forth around the line $y = 0$. This behaviour, which is consistent with (5.3)and (5.4), continues even if more observations are generated.

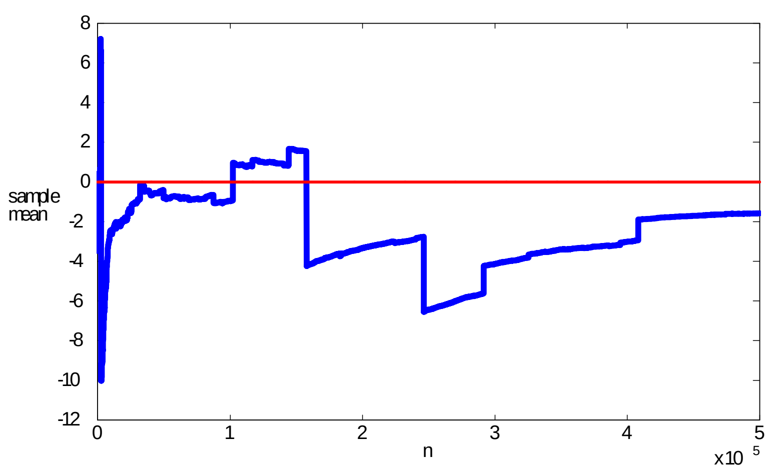


Figure 5.12: Graph of $\bar{x}_n$ versus n for 50000 Cauchy$(0,1)$ observations

5.4 Moment Generating Function Technique for Limiting Distributions

We now look at the moment generating function technique for determining a limiting distribution. Suppose we have a sequence of random variables $X_1, X_2, \ldots, X_n, \ldots$ and $M_1(t), M_1(t), \ldots, M_n(t), \ldots$ is the corresponding sequence of moment generating functions. For a fixed value of t, the sequence $M_1(t), M_1(t), \ldots, M_n(t), \ldots$ is a sequence of real numbers. In general we will obtain a different sequence of real numbers for each different value of t. Since we have a sequence of real numbers we will be able to use limit theorems you have used in your previous calculus courses to evaluate $\lim\limits_{n\to\infty} M_n(t)$. We will need to take care in determining how $M_n(t)$ behaves as $n \to \infty$ for an interval of values of t containing the value 0. Of course this technique only works if the the moment generating function exists and is tractable.

5.4.1 Limit Theorem for Moment Generating Functions

Let $X_1, X_2, \ldots, X_n, \ldots$ be a sequence of random variables such that X_n has moment generating function $M_n(t)$. Let X be a random variable with moment generating function $M(t)$. If there exists an $h > 0$ such that

$$\lim_{n\to\infty} M_n(t) = M(t) \quad \text{for all } t \in (-h, h)$$

then

$$X_n \to_D X$$

Note:

(1) The sequence of random variables $X_1, X_2, \ldots, X_n, \ldots$ converges if the corresponding sequence of moment generating functions $M_1(t), M_1(t), \ldots, M_n(t), \ldots$ converges pointwise.

(2) This definition holds for both discrete and continuous random variables.

Recall from Definition 5.1.1 that

$$X_n \to_D X$$

if

$$\lim_{n\to\infty} F_n(x) = F(x)$$

at all points x at which $F(x)$ is continuous. If X is a discrete random variable then the cumulative distribution function is a right continuous function. The values of x of main interest for a discrete random variable are exactly the points at which $F(x)$ is discontinuous. The following theorem indicates that $\lim\limits_{n\to\infty} F_n(x) = F(x)$ holds for the values of x at which $F(x)$ is discontinuous if X_n and X are non-negative integer-valued random variables. The named discrete distributions Bernoulli, Binomial, Geometric, Negative Binomial, and Poisson are all non-negative integer-valued random variables.

5.4.2 Theorem

Suppose X_n and X are non-negative integer-valued random variables. If $X_n \to_D X$ then $\lim\limits_{n\to\infty} P(X_n \le x) = P(X \le x)$ holds for all x and in particular

$$\lim_{n\to\infty} P(X_n = x) = P(X = x) \quad \text{for } x = 0, 1, \ldots.$$

5.4.3 Example

Consider the sequence of random variables $X_1, X_2, \ldots, X_k, \ldots$ where $X_k \sim$ Negative Binomial(k, p). Use Theorem 5.4.1 to determine the limiting distribution of X_k as $k \to \infty$, $p \to 1$ such that $kq/p = \mu$ remains constant where $q = 1 - p$. Use this limiting distribution and Theorem 5.4.2 to give an approximation for $P(X_k = x)$.

Solution

If $X_k \sim$ Negative Binomial(k, p) then

$$M_k(t) = E\left(e^{tX_k}\right) = \left(\frac{p}{1 - qe^t}\right)^k \quad \text{for } t < -\log q \tag{5.5}$$

If $\mu = kq/p$ then

$$p = \frac{k}{\mu + k} \quad \text{and} \quad q = \frac{\mu}{\mu + k} \tag{5.6}$$

Substituting 5.6 into 5.5 and simplifying gives

$$\begin{aligned} M_k(t) &= \left(\frac{\frac{k}{\mu+k}}{1 - \frac{\mu}{\mu+k}e^t}\right)^k = \left(\frac{1}{\frac{\mu+k-\mu e^t}{k}}\right)^k \\ &= \left[\frac{1}{1 - \frac{\mu(e^t-1)}{k}}\right]^k \\ &= \left[1 - \frac{\mu\left(e^t - 1\right)}{k}\right]^{-k} \quad \text{for } t < -\log\left(\frac{\mu}{\mu + k}\right) \end{aligned}$$

Now

$$\lim_{k\to\infty} \left[1 - \frac{\mu\left(e^t - 1\right)}{k}\right]^{-k} = e^{\mu\left(e^t - 1\right)} \quad \text{for } t < \infty$$

by Corollary 5.1.3. Since $M(t) = e^{\mu\left(e^t-1\right)}$ for $t \in \Re$ is the moment generating function of a Poisson(μ) random variable then by Theorem 5.4.1, $X_k \to_D X \sim$ Poisson(μ).

By Theorem 5.4.2

$$P(X_k = x) = \binom{-k}{x} p^k (-q)^x \approx \frac{\left(\frac{kq}{p}\right)^x e^{-kq/p}}{x!} \quad \text{for } x = 0, 1, \ldots$$

5.4.4 Exercise - Poisson Approximation to the Binomial Distribution

Consider the sequence of random variables $X_1, X_2, \ldots, X_n, \ldots$ where $X_n \sim \text{Binomial}(n, p)$. Use Theorem 5.4.1 to determine the limiting distribution of X_n as $n \to \infty$, $p \to 0$ such that $np = \mu$ remains constant. Use this limiting distribution and Theorem 5.4.2 to give an approximation for $P(X_n = x)$.

In your previous probability and statistics courses you would have used the Central Limit Theorem (without proof!) for approximating Binomial and Poisson probabilities as well as constructing approximate confidence intervals. We now give a proof of this theorem.

5.4.5 Central Limit Theorem

Suppose $X_1, X_2, \ldots$ are independent and identically distributed random variables with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2 < \infty$. Consider the sequence of random variables $Z_1, Z_2, \ldots, Z_n, \ldots$ where $Z_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}$ and $\bar{X}_n = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$. Then

$$Z_n = \frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{\sigma} \to_D Z \sim \text{N}(0, 1)$$

Proof

We can write Z_n as

$$Z_n = \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu)$$

Suppose that for $i = 1, 2, \ldots$, X_i has moment generating function $M_X(t)$, $t \in (-h, h)$ for some $h > 0$. Then for $i = 1, 2, \ldots$, $(X_i - \mu)$ has moment generating function $M(t) = e^{-\mu t} M_X(t)$, $t \in (-h, h)$ for some $h > 0$. Note that

$$M(0) = 1, \quad M'(0) = E(X_i - \mu) = E(X_i) - \mu = 0$$

and

$$M''(0) = E\left[(X_i - \mu)^2\right] = Var(X_i) = \sigma^2$$

Also by Taylor's Theorem (see 2.11.15) for $n = 2$ we have

$$\begin{aligned} M(t) &= M(0) + M'(0)\,t + \frac{1}{2}M''(c)\,t^2 \\ &= 1 + \frac{1}{2}M''(c)\,t^2 \qquad (5.7) \end{aligned}$$

for some c between 0 and t.

Since $X_1, X_2, \ldots, X_n$ are independent and identically distributed, the moment generating function of Z_n is

$$\begin{aligned} M_n(t) &= E\left(e^{tZ_n}\right) \\ &= E\left[\exp\left(\frac{t}{\sigma\sqrt{n}}\sum_{i=1}^{n}(X_i-\mu)\right)\right] \\ &= \left[M\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n \quad \text{for } \left|\frac{t}{\sigma\sqrt{n}}\right| < h \end{aligned} \tag{5.8}$$

Using (5.7) in (5.8) gives

$$\begin{aligned} M_n(t) &= \left[1+\frac{1}{2}M''(c_n)\left(\frac{t}{\sigma\sqrt{n}}\right)^2\right]^n \\ &= \left\{1+\frac{1}{2}\left(\frac{t}{\sigma\sqrt{n}}\right)^2\left[M''(c_n)-M''(0)\right]+\frac{1}{2}\left(\frac{t}{\sigma\sqrt{n}}\right)^2 M''(0)\right\}^n \\ &\quad \text{for some } c_n \text{ between } 0 \text{ and } \frac{t}{\sigma\sqrt{n}} \end{aligned}$$

But $M''(0)=\sigma^2$ so

$$\begin{aligned} M_n(t) &= \left\{1+\frac{\left(\frac{1}{2}t^2\right)}{n}+\left(\frac{t^2}{2\sigma^2}\right)\frac{\left[M''(c_n)-M''(0)\right]}{n}\right\}^n \\ &\quad \text{for some } c_n \text{ between } 0 \text{ and } \frac{t}{\sigma\sqrt{n}} \end{aligned}$$

Since c_n is between 0 and $\frac{t}{\sigma\sqrt{n}}$, $c_n \to 0$ as $n \to \infty$. Since $M''(t)$ is continuous on $(-h, h)$

$$\lim_{n\to\infty} M''(c_n) = M''\left(\lim_{n\to\infty} c_n\right) = M''(0) = \sigma^2$$

and

$$\lim_{n\to\infty}\left(\frac{t^2}{2\sigma^2}\right)\frac{\left[M''(c_n)-M''(0)\right]}{n} = 0$$

Therefore by Theorem 5.1.2, with

$$\psi(n) = \left(\frac{t^2}{2\sigma^2}\right)\left[M''(c_n)-M''(0)\right]$$

we have

$$\begin{aligned} \lim_{n\to\infty} M_n(t) &= \lim_{n\to\infty}\left\{1+\frac{\left(\frac{1}{2}t^2\right)}{n}+\left(\frac{t^2}{2\sigma^2}\right)\frac{\left[M''(c_n)-M''(0)\right]}{n}\right\}^n \\ &= e^{\frac{1}{2}t^2} \quad \text{for } |t| < \infty \end{aligned}$$

which is the moment generating function of a $\mathrm{N}(0,1)$ random variable. Therefore by Theorem 5.4.1

$$Z_n \to_D Z \sim \mathrm{N}(0,1)$$

as required.

Note: Although this proof assumes that the moment generating function of X_i, $i = 1, 2, \ldots$ exists, it does not make any assumptions about the form of the distribution of the X_i's. There are other more general proofs of the Central Limit Theorem which only assume the existence of the variance σ^2 (which implies the existence of the mean μ).

5.4.6 Example - Normal Approximation to the χ^2 Distribution

Suppose $Y_n \sim \chi^2(n)$, $n = 1, 2, \ldots$. Consider the sequence of random variables $Z_1, Z_2, \ldots, Z_n, \ldots$ where $Z_n = (Y_n - n)/\sqrt{2n}$. Show that

$$Z_n = \frac{Y_n - n}{\sqrt{2n}} \to_D Z \sim \mathrm{N}(0, 1)$$

Solution

Let $X_i \sim \chi^2(1)$, $i = 1, 2, \ldots$ independently. Since $X_1, X_2, \ldots$ are independent and identically distributed random variables with $E(X_i) = 1$ and $Var(X_i) = 2$, then by the Central Limit Theorem

$$\frac{\sqrt{n}\left(\bar{X}_n - 1\right)}{\sqrt{2}} \to_D Z \sim \mathrm{N}(0, 1)$$

But $\bar{X}_n = \frac{1}{n} \sum\limits_{i=1}^{n} X_i$ so

$$\frac{\sqrt{n}\left(\frac{1}{n} \sum\limits_{i=1}^{n} X_i - 1\right)}{\sqrt{2}} = \frac{S_n - n}{\sqrt{2n}}$$

where $S_n = \sum\limits_{i=1}^{n} X_i$. Therefore

$$\frac{S_n - n}{\sqrt{2n}} \to_D Z \sim \mathrm{N}(0, 1)$$

Now by 4.3.2(6), $S_n \sim \chi^2(n)$ and therefore Y_n and S_n have the same distribution. It follows that

$$Z_n = \frac{Y_n - n}{\sqrt{2n}} \to_D Z \sim \mathrm{N}(0, 1)$$

5.4.7 Exercise - Normal Approximation to the Binomial Distribution

Suppose $Y_n \sim \mathrm{Binomial}(n, p)$, $n = 1, 2, \ldots$ Consider the sequence of random variables $Z_1, Z_2, \ldots, Z_n, \ldots$ where $Z_n = (Y_n - np)/\sqrt{np(1-p)}$. Show that

$$Z_n = \frac{Y_n - np}{\sqrt{np(1-p)}} \to_D Z \sim \mathrm{N}(0, 1)$$

Hint: Let $X_i \sim \mathrm{Binomial}(1, p)$, $i = 1, 2, \ldots, n$ independently.

5.5 Additional Limit Theorems

Suppose we know the limiting distribution of one or more sequences of random variables by using the definitions and/or theorems in the previous sections of this chapter. The theorems in this section allow us to more easily determine the limiting distribution of a function of these sequences.

5.5.1 Limit Theorems

(1) If $X_n \to_p a$ and g is continuous at $x = a$ then $g(X_n) \to_p g(a)$.

(2) If $X_n \to_p a$, $Y_n \to_p b$ and $g(x, y)$ is continuous at (a, b) then $g(X_n, Y_n) \to_p g(a, b)$.

(3) (Slutsky's Theorem) If $X_n \to_p a$, $Y_n \to_D Y$ and $g(a, y)$ is continuous for all $y \in$ support set of Y then $g(X_n, Y_n) \to_D g(a, Y)$.

Proof of (1)

Since g is continuous at $x = a$ then for every $\varepsilon > 0$ there exists a $\delta > 0$ such that $|x - a| < \delta$ implies $|g(x) - g(a)| < \varepsilon$. By Example 2.1.4($d$) this implies that

$$P(|g(X_n) - g(a)| < \varepsilon) \geq P(|X - a| < \delta)$$

But $X_n \to_p a$ it follows that for every $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$\lim_{n\to\infty} P(|g(X_n) - g(a)| < \varepsilon) \geq \lim_{n\to\infty} P(|X - a| < \delta) = 1$$

But

$$\lim_{n\to\infty} P(|g(X_n) - g(a)| < \varepsilon) \leq 1$$

so by the Squeeze Theorem

$$\lim_{n\to\infty} P(|g(X_n) - g(a)| < \varepsilon) = 1$$

and therefore $g(X_n) \to_p g(a)$.

5.5.2 Example

If $X_n \to_p a > 0$, $Y_n \to_p b \neq 0$ and $Z_n \to_D Z \sim \mathrm{N}(0, 1)$ then find the limiting distributions of each of the following:

(b) $\sqrt{X_n}$

(b) $X_n + Y_n$

(c) $Y_n + Z_n$

(d) $X_n Z_n$

(e) Z_n^2

Solution

(a) Let $g(x) = \sqrt{x}$ which is a continuous function for all $x \in \Re^+$. Since $X_n \to_p a$ then by 5.5.1(1), $\sqrt{X_n} = g(X_n) \to_p g(a) = \sqrt{a}$ or $\sqrt{X_n} \to_p \sqrt{a}$.

(b) Let $g(x, y) = x + y$ which is a continuous function for all $(x, y) \in \Re^2$. Since $X_n \to_p a$ and $Y_n \to_p b$ then by 5.5.1(2), $X_n + Y_n = g(X_n, Y_n) \to_p g(a, b) = a + b$ or $X_n + Y_n \to_p a + b$.

(c) Let $g(y, z) = y + z$ which is a continuous function for all $(y, z) \in \Re^2$. Since $Y_n \to_p b$ and $Z_n \to_D Z \sim \mathrm{N}(0, 1)$ then by 5.5.1(3), $Y_n + Z_n = g(Y_n, Z_n) \to_D g(b, z) = b + Z$ or $Y_n + Z_n \to_D b + Z$ where $Z \sim \mathrm{N}(0, 1)$. Since $b + Z \sim \mathrm{N}(b, 1)$, therefore $Y_n + Z_n \to_D b + Z \sim \mathrm{N}(b, 1)$

(d) Let $g(x, z) = xz$ which is a continuous function for all $(x, z) \in \Re^2$. Since $X_n \to_p a$ and $Z_n \to_D Z \sim \mathrm{N}(0, 1)$ then by Slutsky's Theorem, $X_n Z_n = g(X_n, Z_n) \to_D g(a, z) = aZ$ or $X_n Z_n \to_D aZ$ where $Z \sim \mathrm{N}(0, 1)$. Since $aZ \sim \mathrm{N}\big(0, a^2\big)$, therefore $X_n Z_n \to_D aZ \sim \mathrm{N}\big(0, a^2\big)$

(e) Let $g(x, z) = z^2$ which is a continuous function for all $(x, z) \in \Re^2$. Since $Z_n \to_D Z \sim \mathrm{N}(0, 1)$ then by Slutsky's Theorem, $Z_n^2 = g(X_n, Z_n) \to_D g(a, z) = Z^2$ or $Z_n^2 \to_D Z^2$ where $Z \sim \mathrm{N}(0, 1)$. Since $Z^2 \sim \chi^2(1)$, therefore $Z_n^2 \to_D Z^2 \sim \chi^2(1)$

5.5.3 Exercise

If $X_n \to_p a > 0$, $Y_n \to_p b \neq 0$ and $Z_n \to_D Z \sim \mathrm{N}(0, 1)$ then find the limiting distributions of each of the following:

(a) X_n^2

(b) $X_n Y_n$

(c) X_n / Y_n

(d) $X_n - 2Z_n$

(e) $1/Z_n$

In Example 5.5.2 we identified the function g in each case. As with other limit theorems we tend not to explicitly identify the function g once we have a good idea of how the theorems work as illustrated in the next example.

5.5.4 Example

Suppose $X_i \sim \text{Poisson}(\mu)$, $i = 1, 2, \ldots$ independently. Consider the sequence of random variables $Z_1, Z_2, \ldots, Z_n, \ldots$ where

$$Z_n = \frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{\sqrt{\bar{X}_n}}$$

Find the limiting distribution of Z_n.

Solution

Since $X_1, X_2, \ldots$ are independent and identically distributed random variables with $E(X_i) = \mu$ and $Var(X_i) = \mu$, then by the Central Limit Theorem

$$W_n = \frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{\sqrt{\mu}} \to_D Z \sim \mathrm{N}(0,1) \tag{5.9}$$

and by the Weak Law of Large Numbers

$$\bar{X}_n \to_p \mu \tag{5.10}$$

By (5.10) and 5.5.1(1)

$$U_n = \sqrt{\frac{\bar{X}_n}{\mu}} \to_p \sqrt{\frac{\mu}{\mu}} = 1 \tag{5.11}$$

Now

$$\begin{aligned} Z_n &= \frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{\sqrt{\bar{X}_n}} \\ &= \frac{\frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{\sqrt{\mu}}}{\sqrt{\frac{\bar{X}_n}{\mu}}} \\ &= \frac{W_n}{U_n} \end{aligned}$$

By (5.9), (5.11), and Slutsky's Theorem

$$Z_n = \frac{W_n}{U_n} \to_D \frac{Z}{1} = Z \sim \mathrm{N}(0,1)$$

5.5.5 Example

Suppose $X_i \sim$ Uniform$(0,1)$, $i = 1, 2, \ldots$ independently. Consider the sequence of random variables $U_1, U_2, \ldots, U_n, \ldots$ where $U_n = \max(X_1, X_2, \ldots, X_n)$. Show that

(a) $U_n \to_p 1$

(b) $e^{U_n} \to_p e$

(c) $\sin(1 - U_n) \to_p 0$

(d) $V_n = n(1 - U_n) \to_D V \sim$ Exponential(1)

(e) $1 - e^{-V_n} \to_D 1 - e^{-V} \sim$ Uniform$(0,1)$

(f) $(U_n + 1)^2 [n(1 - U_n)] \to_D 4V \sim$ Exponential(4)

Solution

(a) Since $X_1, X_2, \ldots, X_n$ are Uniform$(0,1)$ random variables then for $i = 1, 2, \ldots$

$$P(X_i \leq x) = \begin{cases} 0 & x \leq 0 \\ x & 0 < x < 1 \\ 1 & x \geq 1 \end{cases}$$

Since $X_1, X_2, \ldots, X_n$ are independent random variables

$$\begin{aligned} F_n(x) &= P(U_n \leq u) \\ &= P(\max(X_1, X_2, \ldots, X_n) \leq u) \\ &= \prod_{i=1}^{n} P(X_i \leq u) \end{aligned}$$

$$= \begin{cases} 0 & u \leq 0 \\ u^n & 0 < u < 1 \\ 1 & u \geq 1 \end{cases}$$

Therefore

$$\lim_{n\to\infty} F_n(u) = \begin{cases} 0 & u < 1 \\ 1 & u \geq 1 \end{cases}$$

and by Theorem 5.2.5

$$U_n = \max(X_1, X_2, \ldots, X_n) \to_p 1 \tag{5.12}$$

(b) By (5.12) and 5.5.1(1)

$$e^{U_n} \to_p e^1 = e$$

(c) By (5.12) and 5.5.1(1)

$$\sin(1 - U_n) \to_p \sin(1 - 1) = \sin(0) = 0$$

(d) The cumulative distribution function of $V_n = n(1 - U_n)$ is

$$\begin{aligned} G_n(x) &= P(V_n \leq v) \\ &= P(n(1 - \max(X_1, X_2, \ldots, X_n)) \leq v) \\ &= P\left(\max(X_1, X_2, \ldots, X_n) \geq 1 - \frac{v}{n}\right) \\ &= 1 - P\left(\max(X_1, X_2, \ldots, X_n) \leq 1 - \frac{v}{n}\right) \\ &= 1 - \prod_{i=1}^{n} P\left(X_i \leq 1 - \frac{v}{n}\right) \\ &= \begin{cases} 0 & v \leq 0 \\ 1 - \left(1 - \frac{v}{n}\right)^n & v > 0 \end{cases} \end{aligned}$$

Therefore

$$\lim_{n\to\infty} G_n(v) = \begin{cases} 0 & v \leq 0 \\ 1 - e^{-v} & v > 0 \end{cases}$$

which is the cumulative distribution function of an Exponential(1) random variable. Therefore by Definition (5.1.1)

$$V_n = n(1 - U_n) \to_D V \sim \text{Exponential}(1) \tag{5.13}$$

(e) By (5.13) and Slutsky's Theorem

$$1 - e^{-V_n} \to_D 1 - e^{-V} \quad \text{where} \quad V \sim \text{Exponential}(1)$$

The probability density function of $W = 1 - e^{-V}$ is

$$\begin{aligned} g(w) &= e^{\log(1-w)} \left| \frac{d}{dw} [-\log(1-w)] \right| \\ &= 1 \quad \text{for } 0 < w < 1 \end{aligned}$$

which is the probability of a Uniform$(0,1)$ random variable. Therefore

$$1 - e^{-V_n} \to_D 1 - e^{-V} \sim \text{Uniform}(0,1)$$

(f) By (5.12) and 5.5.1(1)

$$(U_n + 1)^2 \to_p (1+1)^2 = 4 \tag{5.14}$$

By (5.13), (5.14), and Slutsky's Theorem

$$(U_n + 1)^2 [n(1 - U_n)] \to_D 4V$$

where $V \sim$ Exponential(1). If $V \sim$ Exponential(1) then $4V \sim$ Exponential(4) so

$$(U_n + 1)^2 [n(1 - U_n)] \to_D 4V \sim \text{Exponential}(4)$$

5.5.6 Delta Method

Let $X_1, X_2, \ldots, X_n, \ldots$ be a sequence of random variables such that

$$n^b(X_n - a) \to_D X \tag{5.15}$$

for some $b > 0$. Suppose the function $g(x)$ is differentiable at a and $g'(a) \neq 0$. Then

$$n^b[g(X_n) - g(a)] \to_D g'(a)X$$

Proof

By Taylor's Theorem (2.11.15) we have

$$g(X_n) = g(a) + g'(c_n)(X_n - a)$$

or

$$g(X_n) - g(a) = g'(c_n)(X_n - a) \tag{5.16}$$

where c_n is between a and X_n.

From (5.15) it follows that $X_n \to_p a$. Since c_n is between X_n and a, therefore $c_n \to_p a$ and by 5.5.1(1)

$$g'(c_n) \to_p g'(a) \tag{5.17}$$

Multiplying (5.16) by n^b gives

$$n^b[g(X_n) - g(a)] = g'(c_n)\, n^b(X_n - a) \tag{5.18}$$

Therefore by (5.15), (5.17), (5.18) and Slutsky's Theorem

$$n^b[g(X_n) - g(a)] \to_D g'(a)X$$

5.5.7 Example

Suppose $X_i \sim$ Exponential(θ), $i = 1, 2, \ldots$ independently. Find the limiting distributions of each of the following:

(a) $\bar{X}_n$

(b) $U_n = \sqrt{n}\left(\bar{X}_n - \theta\right)$

(c) $Z_n = \frac{\sqrt{n}\left(\bar{X}_n - \theta\right)}{\bar{X}_n}$

(d) $V_n = \sqrt{n}\left[\log(\bar{X}_n) - \log\theta\right]$

Solution

(a) Since $X_1, X_2, \ldots$ are independent and identically distributed random variables with $E(X_i) = \theta$ and $Var(X_i) = \theta^2$, then by the Weak Law of Large Numbers

$$\bar{X}_n \to_p \theta \tag{5.19}$$

(b) Since $X_1, X_2, \ldots$ are independent and identically distributed random variables with $E(X_i) = \theta$ and $Var(X_i) = \theta^2$, then by the Central Limit Theorem

$$W_n = \frac{\sqrt{n}\left(\bar{X}_n - \theta\right)}{\theta} \to_D Z \sim \mathrm{N}(0, 1) \tag{5.20}$$

Therefore by Slutsky's Theorem

$$U_n = \sqrt{n}\left(\bar{X}_n - \theta\right) \to_D \theta Z \sim \mathrm{N}\left(0, \theta^2\right) \tag{5.21}$$

(c) Z_n can be written as

$$Z_n = \frac{\sqrt{n}\left(\bar{X}_n - \theta\right)}{\bar{X}_n} = \frac{\frac{\sqrt{n}\left(\bar{X}_n - \theta\right)}{\theta}}{\frac{\bar{X}_n}{\theta}}$$

By (5.19) and 5.5.1(1),

$$\frac{\bar{X}_n}{\theta} \rightarrow_p \frac{\theta}{\theta} = 1 \tag{5.22}$$

By (5.20), (5.22), and Slutsky's Theorem

$$Z_n \rightarrow_D \frac{Z}{1} = Z \sim \mathrm{N}(0,1)$$

(d) Let $g(x) = \log x$, $a = \theta$, and $b = 1/2$. Then $g'(x) = \frac{1}{x}$ and $g'(a) = g'(\theta) = \frac{1}{\theta}$. By (5.21) and the Delta Method

$$n^{1/2}\left[\log(\bar{X}_n) - \log\theta\right] \rightarrow_D \frac{1}{\theta}(\theta Z) = Z \sim \mathrm{N}(0,1)$$

5.5.8 Exercise

Suppose $X_i \sim \text{Poisson}(\mu)$, $i = 1, 2, \ldots$ independently. Show that

$$\begin{aligned} U_n &= \sqrt{n}\left(\bar{X}_n - \mu\right) \rightarrow_D U \sim \mathrm{N}(0,\mu) \\ \text{and } V_n &= \sqrt{n}\left(\sqrt{\bar{X}_n} - \sqrt{\mu}\right) \rightarrow_D V \sim \mathrm{N}\left(0, \frac{1}{4}\right) \end{aligned}$$

5.5.9 Theorem

Let $X_1, X_2, \ldots, X_n, \ldots$ be a sequence of random variables such that

$$\sqrt{n}(X_n - a) \rightarrow_D X \sim \mathrm{N}\left(0, \sigma^2\right) \tag{5.23}$$

Suppose the function $g(x)$ is differentiable at a. Then

$$\sqrt{n}[g(X_n) - g(a)] \rightarrow_D W \sim \mathrm{N}\left(0, \left[g'(a)\right]^2\sigma^2\right)$$

provided $g'(a) \neq 0$.

Proof

Suppose $g(x)$ is a differentiable function at a and $g'(a) \neq 0$. Let $b = 1/2$. Then by (5.23) and the Delta Method it follows that

$$\sqrt{n}[g(X_n) - g(a)] \rightarrow_D W \sim \mathrm{N}\left(0, \left[g'(a)\right]^2\sigma^2\right)$$

5.6 Chapter 5 Problems

1. Suppose $Y_i \sim$ Exponential$(\theta, 1)$, $i = 1, 2, \ldots$ independently. Find the limiting distributions of

   (a) $X_n = \min(Y_1, Y_2, \ldots, Y_n)$

   (b) $U_n = X_n/\theta$

   (c) $V_n = n(X_n - \theta)$

   (d) $W_n = n^2(X_n - \theta)$

2. Suppose $X_1, X_2, \ldots, X_n$ are independent and identically distributed continuous random variables with cumulative distribution function $F(x)$ and probability density function $f(x)$. Let $Y_n = \max(X_1, X_2, \ldots, X_n)$.

   Show that
   $$Z_n = n[1 - F(Y_n)] \to_D Z \sim \text{Exponential}(1)$$

3. Suppose $X_i \sim$ Poisson(μ), $i = 1, 2, \ldots$ independently. Find $M_n(t)$ the moment generating function of
   $$Y_n = \sqrt{n}(\bar{X}_n - \mu)$$
   Show that
   $$\lim_{n\to\infty} \log M_n(t) = \frac{1}{2}\mu t^2$$
   What is the limiting distribution of Y_n?

4. Suppose $X_i \sim$ Exponential(θ), $i = 1, 2, \ldots$ independently. Show that the moment generating function of
   $$Z_n = \left(\sum_{i=1}^{n} X_i - n\theta\right) / \sqrt{n}$$
   is
   $$M_n(t) = \left[e^{\theta t/\sqrt{n}}\left(1 - \theta t/\sqrt{n}\right)\right]^{-n}$$
   Find $\lim\limits_{n\to\infty} M_n(t)$ and thus determine the limiting distribution of Z_n.

5. If $Z \sim$ N$(0, 1)$ and $W_n \sim \chi^2(n)$ independently then we know
   $$T_n = \frac{Z}{\sqrt{W_n/n}} \sim t(n)$$
   Show that
   $$T_n \to_D Y \sim \text{N}(0, 1)$$

6. Suppose $X_1, X_2, \ldots$ are independent and identically distributed random variables with $E(X_i) = \mu$, $Var(X_i) = \sigma^2 < \infty$, and $E(X_i^4) < \infty$. Let

$$\begin{aligned}\bar{X}_n &= \frac{1}{n}\sum_{i=1}^{n} X_i \\ S_n^2 &= \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2\end{aligned}$$

and

$$T_n = \frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{S_n}$$

Show that

$$S_n \to_p \sigma$$

and

$$T_n \to_D Z \sim \mathrm{N}(0,1)$$

7. Let $X_n \sim$ Binomial(n, θ). Find the limiting distributions of

   (a) $T_n = \frac{X_n}{n}$
   (b) $U_n = \frac{X_n}{n}\left(1 - \frac{X_n}{n}\right)$
   (c) $W_n = \sqrt{n}\left(\frac{X_n}{n} - \theta\right)$
   (d) $Z_n = \frac{W_n}{\sqrt{U_n}}$
   (e) $V_n = \sqrt{n}\left(\arcsin\sqrt{\frac{X_n}{n}} - \arcsin\sqrt{\theta}\right)$
   (f) Compare the variances of the limiting distributions of W_n, Z_n and V_n and comment.

8. Suppose $X_i \sim$ Geometric(θ), $i = 1, 2, \ldots$ independently. Let

$$Y_n = \sum_{i=1}^{n} X_i$$

   Find the limiting distributions of

   (a) $\bar{X}_n = \frac{Y_n}{n}$
   (b) $W_n = \sqrt{n}\left(\bar{X}_n - \frac{(1-\theta)}{\theta}\right)$
   (c) $V_n = \frac{1}{1+\bar{X}_n}$
   (d) $Z_n = \frac{\sqrt{n}(V_n - \theta)}{\sqrt{V_n^2(1-V_n)}}$

9. Suppose $X_i \sim \text{Gamma}(2, \theta)$, $i = 1, 2, \ldots$ independently. Let

$$Y_n = \sum_{i=1}^{n} X_i$$

Find the limiting distributions of

(a) $\bar{X}_n = \frac{Y_n}{n}$ and $\frac{\sqrt{2}\theta}{\bar{X}_n/\sqrt{2}}$

(b) $W_n = \frac{\sqrt{n}\left(\bar{X}_n - 2\theta\right)}{\sqrt{2\theta^2}}$ and $V_n = \sqrt{n}\left(\bar{X}_n - 2\theta\right)$

(c) $Z_n = \frac{\sqrt{n}\left(\bar{X}_n - 2\theta\right)}{\bar{X}_n/\sqrt{2}}$

(d) $U_n = \sqrt{n}\left[\log(\bar{X}_n) - \log(2\theta)\right]$

(e) Compare the variances of the limiting distributions of Z_n and U_n.

10. Suppose $X_1, X_2, \ldots, X_n, \ldots$ is a sequence of random variables with

$$E(X_n) = \mu$$

and

$$Var(X_n) = \frac{a}{n^p} \quad \text{for } p > 0$$

Show that

$$X_n \to_p \mu$$

6. Maximum Likelihood Estimation - One Parameter

In this chapter we look at the method of maximum likelihood to obtain both point and interval estimates of one unknown parameter. Some of this material was introduced in a previous statistics course such as STAT 221/231/241.

In Section 6.2 we review the definitions needed for the method of maximum likelihood estimation, the derivations of the maximum likelihood estimates for the unknown parameter in the Binomial, Poisson and Exponential models, and the important invariance property of maximum likelihood estimates. You will notice that we pay more attention to verifying that the maximum likelihood estimate does correspond to a maximum using the first derivative test. Example 6.2.9 is new and illustrates how the maximum likelihood estimate is found when the support set of the random variable depends on the unknown parameter.

In Section 6.3 we define the score function, the information function, and the expected information function. These functions play an important role in the distribution of the maximum likelihood estimator. These functions are also used in Newton's Method which is a method for determining the maximum likelihood estimate in cases where there is no explicit solution. Although the maximum likelihood estimates in nearly all the examples you saw previously could be found explicitly, this is not true in general.

In Section 6.4 we review likelihood intervals. Likelihood intervals provide a way to summarize the uncertainty in an estimate. In Section 6.5 we give a theorem on the limiting distribution of the maximum likelihood estimator. This important theorem tells us why maximum likelihood estimators are good estimators.

In Section 6.6 we review how to find a confidence interval using a pivotal quantity. Confidence intervals also give us a way to summarize the uncertainty in an estimate. We also give a theorem on how to obtain a pivotal quantity using the maximum likelihood estimator if the parameter θ is either a scale or location parameter. In Section 6.7 we review how to find an approximate confidence interval using an asymptotic pivotal quantity. We then show how to use asymptotic pivotal quantities based on the limiting distribution of the maximum likelihood estimator to construct approximate confidence intervals.

6.1 Introduction

Suppose the random variable $\mathbf{X}$ (possibly a vector of random variables) has probability function/probability density function $f(\mathbf{x};\theta)$. Suppose also that θ is unknown and $\theta \in \Omega$ where Ω is the parameter space or the set of possible values of θ. Let $\mathbf{X}$ be the potential data that is to be collected. In your previous statistics course you learned how numerical and graphical summaries as well as goodness of fit tests could be used to check whether the assumed model for an observed set of data $\mathbf{x}$ was reasonable. In this course we will assume that the fit of the model has been checked and that the main focus now is to use the model and the data to determine point and interval estimates of θ.

6.1.1 Definition - Statistic

A *statistic*, $T = T(\mathbf{X})$, is a function of the data $\mathbf{X}$ which does not depend on any unknown parameters.

6.1.2 Example

Suppose $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ is a random sample, that is, $X_1, X_2, \ldots, X_n$ are independent and identically distributed random variables, from a distribution with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$ where μ and σ^2 are unknown.

The sample mean $\bar{X} = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$, the sample variance $S^2 = \frac{1}{n-1}\sum\limits_{i=1}^{n}\left(X_i - \bar{X}\right)^2$, and the sample minimum $X_{(1)} = \min\left(X_1, X_2, \ldots, X_n\right)$ are statistics.

The random variable $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ is not a statistic since it is not only a function of the data $\mathbf{X}$ but also a function of the unknown parameters μ and σ^2.

6.1.3 Definition - Estimator and Estimate

A statistic $T = T(\mathbf{X})$ that is used to estimate $\tau(\theta)$, a function of θ, is called an *estimator* of $\tau(\theta)$ and an observed value of the statistic $t = t(\mathbf{x})$ is called an *estimate* of $\tau(\theta)$.

6.1.4 Example

Suppose $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ are independent and identically distributed random variables from a distribution with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$, $i = 1, 2, \ldots, n$. Suppose $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is an observed random sample from this distribution.

The random variable $\bar{X}$ is an estimator of μ. The number $\bar{x}$ is an estimate of μ.

The random variable S is an estimator of σ. The number s is an estimate of σ.

6.2 Maximum Likelihood Method

Suppose $\mathbf{X}$ is discrete random variable with probability function $P(\mathbf{X} = \mathbf{x}; \theta) = f(\mathbf{x}; \theta)$, $\theta \in \Omega$ where the scalar parameter θ is unknown. Suppose also that $\mathbf{x}$ is an observed value of the random variable $\mathbf{X}$. Then the probability of observing this value is, $P(\mathbf{X} = \mathbf{x}; \theta) = f(\mathbf{x}; \theta)$. With the observed value of $\mathbf{x}$ substituted into $f(\mathbf{x}; \theta)$ we have a function of the parameter θ only, referred to as the *likelihood function* and denoted $L(\theta; \mathbf{x})$ or $L(\theta)$. In the absence of any other information, it seems logical that we should estimate the parameter θ using a value most compatible with the data. For example we might choose the value of θ which maximizes the probability of the observed data or equivalently the value of θ which maximizes the likelihood function.

6.2.1 Definition - Likelihood Function: Discrete Case

Suppose $\mathbf{X}$ is a discrete random variable with probability function $f(\mathbf{x}; \theta)$, where θ is a scalar and $\theta \in \Omega$ and $\mathbf{x}$ is an observed value of $\mathbf{X}$. The *likelihood function* for θ based on the observed data $\mathbf{x}$ is

$$\begin{aligned} L(\theta) &= L(\theta; \mathbf{x}) \\ &= P(\text{observing the data } \mathbf{x}; \theta) \\ &= P(\mathbf{X} = \mathbf{x}; \theta) \\ &= f(\mathbf{x}; \theta) \quad \text{for } \theta \in \Omega \end{aligned}$$

If $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ is a random sample from a distribution with probability function $f(x; \theta)$ and $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ are the observed data then the likelihood function for θ based on the observed data $\mathbf{x}$ is

$$\begin{aligned} L(\theta) &= L(\theta; \mathbf{x}) \\ &= P(\text{observing the data } \mathbf{x}; \theta) \\ &= P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n; \theta) \\ &= \prod_{i=1}^{n} f(x_i; \theta) \quad \text{for } \theta \in \Omega \end{aligned}$$

6.2.2 Definition - Maximum Likelihood Estimate and Estimator

The value of θ that maximizes the likelihood function $L(\theta)$ is called the *maximum likelihood estimate*. The maximum likelihood estimate is a function of the observed data $\mathbf{x}$ and we write $\hat{\theta} = \hat{\theta}(\mathbf{x})$. The corresponding *maximum likelihood estimator*, which is a random variable, is denoted by $\tilde{\theta} = \tilde{\theta}(\mathbf{X})$.

The shape of the likelihood function and the value of θ at which it is maximized are not affected if $L(\theta)$ is multiplied by a constant. Indeed it is not the absolute value of the likelihood function that is important but the relative values at two different values of the parameter, for example, $L(\theta_1)/L(\theta_2)$. This ratio can be interpreted as how much more or less consistent the data are with the parameter θ_1 as compared to θ_2. The ratio $L(\theta_1)/L(\theta_2)$ is also unaffected if $L(\theta)$ is multiplied by a constant. In view of this the likelihood may be defined as $P(\mathbf{X} = \mathbf{x}; \theta)$ or as any constant multiple of it.

To find the maximum likelihood estimate we usually solve the equation $\frac{d}{d\theta}L(\theta) = 0$. If the value of θ which maximizes $L(\theta)$ occurs at an endpoint of Ω, of course this does not provide the value at which the likelihood is maximized. In Example 6.2.9 we see that solving $\frac{d}{d\theta}L(\theta) = 0$ does not give the maximum likelihood estimate. Most often however we will find $\hat{\theta}$ by solving $\frac{d}{d\theta}L(\theta) = 0$.

Since the log function (Note: log = ln) is an increasing function, the value of θ which maximizes the likelihood $L(\theta)$ also maximizes $\log L(\theta)$, the logarithm of the likelihood function. Since it is usually simpler to find the derivative of the sum of n terms rather than the product, it is often easier to determine the maximum likelihood estimate of θ by solving $\frac{d}{d\theta}\log L(\theta) = 0$.

It is important to verify that $\hat{\theta}$ is the value of θ which maximizes $L(\theta)$ or equivalently $l(\theta)$. This can be done using the first derivative test. Recall that the second derivative test checks for a local extremum.

6.2.3 Definition - Log Likelihood Function

The *log likelihood function* is defined as

$$l(\theta) = l(\theta; \mathbf{x}) = \log L(\theta) \quad \text{for } \theta \in \Omega$$

where $\mathbf{x}$ are the observed data and log is the natural logarithmic function.

6.2.4 Example

Suppose in a sequence of n Bernoulli trials the probability of success is equal to θ and we have observed x successes. Find the likelihood function, the log likelihood function, the maximum likelihood estimate of θ and the maximum likelihood estimator of θ.

Solution

The likelihood function for θ based on x successes in n trials is

$$\begin{aligned} L(\theta) &= P(X = x; \theta) \\ &= \binom{n}{x}\theta^x (1-\theta)^{n-x} \quad \text{for } 0 \le \theta \le 1 \end{aligned}$$

or more simply

$$L(\theta) = \theta^x (1-\theta)^{n-x} \quad \text{for } 0 \le \theta \le 1$$

Suppose $x \neq 0$ and $x \neq n$. The log likelihood function is

$$l(\theta) = x \log \theta + (n - x) \log (1 - \theta) \quad \text{for } 0 < \theta < 1$$

with derivative

$$\begin{aligned} \frac{d}{d\theta} l(\theta) &= \frac{x}{\theta} - \frac{n - x}{1 - \theta} \\ &= \frac{x (1 - \theta) - \theta (n - x)}{\theta (1 - \theta)} \\ &= \frac{x - n\theta}{\theta (1 - \theta)} \quad \text{for } 0 < \theta < 1 \end{aligned}$$

The solution to $\frac{d}{d\theta} l(\theta) = 0$ is $\theta = x/n$ which is the sample proportion. Since $\frac{d}{d\theta} l(\theta) > 0$ if $0 < \theta < x/n$ and $\frac{d}{d\theta} l(\theta) < 0$ if $x/n < \theta < 1$ then, by the first derivative test, $l(\theta)$ has a absolute maximum at $\theta = x/n$.

If $x = 0$ then

$$L(\theta) = (1 - \theta)^n \quad \text{for } 0 \leq \theta \leq 1$$

which is a decreasing function of θ on the interval $[0, 1]$. $L(\theta)$ is maximized at the endpoint $\theta = 0$ or $\theta = 0/n$.

If $x = n$ then

$$L(\theta) = \theta^n \quad \text{for } 0 \leq \theta \leq 1$$

which is a increasing function of θ on the interval $[0, 1]$. $L(\theta)$ is maximized at the endpoint $\theta = 1$ or $\theta = n/n$.

In all cases the value of θ which maximizes the likelihood function is the sample proportion $\theta = x/n$. Therefore the maximum likelihood estimate of θ is $\hat{\theta} = x/n$ and the maximum likelihood estimator is $\tilde{\theta} = x/n$.

6.2.5 Example

Suppose $x_1, x_2, \ldots, x_n$ is an observed random sample from the Poisson(θ) distribution. Find the likelihood function, the log likelihood function, the maximum likelihood estimate of θ and the maximum likelihood estimator of θ.

Solution

The likelihood function is

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{n} f(x_i; \theta) = \prod_{i=1}^{n} P(X_i = x_i; \theta) \\ &= \prod_{i=1}^{n} \frac{\theta^{x_i} e^{-\theta}}{x_i!} \\ &= \left(\prod_{i=1}^{n} \frac{1}{x_i!} \right) \theta^{\sum\limits_{i=1}^{n} x_i} e^{-n\theta} \quad \text{for } \theta \geq 0 \end{aligned}$$

or more simply

$$L(\theta) = \theta^{n\bar{x}} e^{-n\theta} \quad \text{for } \theta \geq 0$$

Suppose $\bar{x} \neq 0$. The log likelihood function is

$$l(\theta) = n(\bar{x}\log\theta - \theta) \quad \text{for } \theta > 0$$

with derivative

$$\begin{aligned} \frac{d}{d\theta} l(\theta) &= n\left(\frac{\bar{x}}{\theta} - 1\right) \\ &= \frac{n}{\theta}(\bar{x} - \theta) \quad \text{for } \theta > 0 \end{aligned}$$

The solution to $\frac{d}{d\theta} l(\theta) = 0$ is $\theta = \bar{x}$ which is the sample mean. Since $\frac{d}{d\theta} l(\theta) > 0$ if $0 < \theta < \bar{x}$ and $\frac{d}{d\theta} l(\theta) < 0$ if $\theta > \bar{x}$ then, by the first derivative test, $l(\theta)$ has a absolute maximum at $\theta = \bar{x}$.

If $\bar{x} = 0$ then

$$L(\theta) = e^{-n\theta} \quad \text{for } \theta \geq 0$$

which is a decreasing function of θ on the interval $[0, \infty)$. $L(\theta)$ is maximized at the endpoint $\theta = 0$ or $\theta = 0 = \bar{x}$.

In all cases the value of θ which maximizes the likelihood function is the sample mean $\theta = \bar{x}$. Therefore the maximum likelihood estimate of θ is $\hat{\theta} = \bar{x}$ and the maximum likelihood estimator is $\tilde{\theta} = \bar{X}$.

6.2.6 Likelihood Functions for Continuous Models

Suppose X is a continuous random variable with probability density function $f(x;\theta)$. For continuous continuous random variable, $P(X = x;\theta)$ is unsuitable as a definition of the likelihood function since this probability always equals zero.

For continuous data we usually observe only the value of X rounded to some degree of precision, for example, data on waiting times is rounded to the closest second or data on heights is rounded to the closest centimeter. The actual observation is really a discrete random variable. For example, suppose we observe X correct to one decimal place. Then

$$P(\text{we observe } 1.1\ ;\theta) = \int_{1.05}^{1.15} f(x;\theta)dx \approx (0.1)f(1.1;\theta)$$

assuming the function $f(x;\theta)$ is reasonably smooth over the interval. More generally, suppose $x_1, x_2, \ldots, x_n$ are the observations from a random sample from the distribution with probability density function $f(x;\theta)$ which have been rounded to the nearest Δ which is assumed to be small. Then

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n;\theta) \approx \prod_{i=1}^{n} \Delta f(x_i;\theta) = \Delta^n \prod_{i=1}^{n} f(x_i;\theta)$$

If we assume that the precision Δ does not depend on the unknown parameter θ, then the term Δ^n can be ignored. This argument leads us to adopt the following definition of the likelihood function for a random sample from a continuous distribution.

6.2.7 Definition - Likelihood Function: Continuous Case

If $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ are the observed values of a random sample from a distribution with probability density function $f(x;\theta)$, then the likelihood function is defined as

$$L(\theta) = L(\theta;\mathbf{x}) = \prod_{i=1}^{n} f(x_i;\theta) \quad \text{for } \theta \in \Omega$$

6.2.8 Example

Suppose $x_1, x_2, \ldots, x_n$ is an observed random sample from the Exponential(θ) distribution. Find the likelihood function, the log likelihood function, the maximum likelihood estimate of θ and the maximum likelihood estimator of θ.

Solution

The likelihood function is

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{n} f(x_i;\theta) = \prod_{i=1}^{n} \frac{1}{\theta} e^{-x_i/\theta} \\ &= \frac{1}{\theta^n} \exp\left(-\sum_{i=1}^{n} x_i/\theta\right) \\ &= \theta^{-n} e^{-n\bar{x}/\theta} \quad \text{for } \theta > 0 \end{aligned}$$

The log likelihood function is

$$l(\theta) = -n\left(\log\theta + \frac{\bar{x}}{\theta}\right) \quad \text{for } \theta > 0$$

with derivative

$$\begin{aligned} \frac{d}{d\theta} l(\theta) &= -n\left(\frac{1}{\theta} - \frac{\bar{x}}{\theta^2}\right) \\ &= \frac{n}{\theta^2}(\bar{x} - \theta) \end{aligned}$$

Now $\frac{d}{d\theta} l(\theta) = 0$ for $\theta = \bar{x}$. Since $\frac{d}{d\theta} l(\theta) > 0$ if $0 < \theta < \bar{x}$ and $\frac{d}{d\theta} l(\theta) < 0$ if $\theta > \bar{x}$ then, by the first derivative test, $l(\theta)$ has a absolute maximum at $\theta = \bar{x}$. Therefore the maximum likelihood estimate of θ is $\hat{\theta} = \bar{x}$ and the maximum likelihood estimator is $\tilde{\theta} = \bar{X}$.

6.2.9 Example

Suppose $x_1, x_2, \ldots, x_n$ is an observed random sample from the Uniform$(0,\theta)$ distribution. Find the likelihood function, the maximum likelihood estimate of θ and the maximum likelihood estimator of θ.

Solution

The probability density function of a Uniform$(0, \theta)$ random variable is

$$f(x;\theta) = \frac{1}{\theta} \quad \text{for } 0 \le x \le \theta$$

and zero otherwise. The support set of the random variable X is $[0, \theta]$ which depends on the unknown parameter θ. In such examples care must be taken in determining the maximum likelihood estimate of θ.

The likelihood function is

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{n} f(x_i;\theta) \\ &= \prod_{i=1}^{n} \frac{1}{\theta} \quad \text{if } 0 \le x_i \le \theta,\ i = 1, 2, \ldots, n, \text{ and } \theta > 0 \\ &= \frac{1}{\theta^n} \quad \text{if } 0 \le x_i \le \theta,\ i = 1, 2, \ldots, n, \text{ and } \theta > 0 \end{aligned}$$

To determine the value of θ which maximizes $L(\theta)$ we note that $L(\theta)$ can be written as

$$L(\theta) = \begin{cases} 0 & \text{if } 0 < \theta < x_{(n)} \\ \frac{1}{\theta^n} & \text{if } \theta \ge x_{(n)} \end{cases}$$

where $x_{(n)} = \max(x_1, x_2, \ldots, x_n)$ is the maximum of the sample. To see this remember that in order to observe the sample $x_1, x_2, \ldots, x_n$ the value of θ must be larger than all the observed x_i's.

$L(\theta)$ is a decreasing function of θ on the interval $[x_{(n)}, \infty)$. Therefore $L(\theta)$ is maximized at $\theta = x_{(n)}$. The maximum likelihood estimate of θ is $\hat{\theta} = x_{(n)}$ and the maximum likelihood estimator is $\tilde{\theta} = X_{(n)}$.

Note: In this example there is no solution to $\frac{d}{d\theta} l(\theta) = \frac{d}{d\theta}(-n \log \theta) = 0$ and the maximum likelihood estimate of θ is not found by solving $\frac{d}{d\theta} l(\theta) = 0$.

One of the reasons the method of maximum likelihood is so widely used is the invariance property of the maximum likelihood estimate under one-to-one transformations.

6.2.10 Theorem - Invariance of the Maximum Likelihood Estimate

If $\hat{\theta}$ is the maximum likelihood estimate of θ then $g(\hat{\theta})$ is the maximum likelihood estimate of $g(\theta)$.

Note: The invariance property of the maximum likelihood estimate means that if we know the maximum likelihood estimate of θ then we know the maximum likelihood estimate of any function of θ.

6.2.11 Example

In Example 6.2.8 find the maximum likelihood estimate of the median of the distribution and the maximum likelihood estimate of $Var(\tilde{\theta})$.

Solution

If X has an Exponential(θ) distribution then the median m is found by solving

$$0.5 = \int_0^m \frac{1}{\theta} e^{-x/\theta} dx$$

to obtain

$$m = -\theta \log(0.5)$$

By the Invariance of the Maximum Likelihood Estimate the maximum likelihood estimate of m is $\hat{m} = -\hat{\theta} \log(0.5) = \bar{x} \log(0.5)$.

Since X_i has an Exponential(θ) distribution with $Var(X_i) = \theta^2$, $i = 1, 2, \ldots, n$ independently, the variance of the maximum likelihood estimator $\tilde{\theta} = \bar{X}$ is

$$Var\left(\tilde{\theta}\right) = Var(\bar{X}) = \frac{\theta^2}{n}$$

By the Invariance of the Maximum Likelihood Estimate the maximum likelihood estimate of $Var(\tilde{\theta})$ is $(\hat{\theta})^2/n = (\bar{x})^2/n$.

6.3 Score and Information Functions

The derivative of the log likelihood function plays an important role in the method of maximum likelihood. This function is often called the *score function*.

6.3.1 Definition - Score Function

The *score function* is defined as

$$S(\theta) = S(\theta; \mathbf{x}) = \frac{d}{d\theta} l(\theta) = \frac{d}{d\theta} \log L(\theta) \quad \text{for } \theta \in \Omega$$

where $\mathbf{x}$ are the observed data.

Another function which plays an important role in the method of maximum likelihood is the *information function*.

6.3.2 Definition - Information Function

The *information function* is defined as

$$I(\theta) = I(\theta; \mathbf{x}) = -\frac{d^2}{d\theta^2} l(\theta) = -\frac{d^2}{d\theta^2} \log L(\theta) \quad \text{for } \theta \in \Omega$$

where $\mathbf{x}$ are the observed data. $I(\hat{\theta})$ is called the observed information.

In Section 6.7 we will see how the observed information $I(\hat{\theta})$ can be used to construct an approximate confidence interval for the unknown parameter θ. $I(\hat{\theta})$ also tells us about the concavity of the log likelihood function $l(\theta)$.

6.3.3 Example

Find the observed information for Example 6.2.5. Suppose the maximum likelihood estimate of θ was $\hat{\theta} = 2$. Compare $I(\hat{\theta}) = I(2)$ if $n = 10$ and $n = 25$. Plot the function $r(\theta) = l(\theta) - l(\hat{\theta})$ for $n = 10$ and $n = 25$ on the same graph.

Solution

From Example 6.2.5, the score function is

$$S(\theta) = \frac{d}{d\theta} l(\theta) = \frac{n}{\theta}(\bar{x} - \theta) \quad \text{for } \theta > 0$$

Therefore the information function is

$$\begin{aligned} -\frac{d^2}{d\theta^2} l(\theta) &= -\frac{d}{d\theta}\left[n\left(\frac{\bar{x}}{\theta} - 1\right)\right] = -\left[-\frac{n\bar{x}}{\theta^2}\right] \\ &= \frac{n\bar{x}}{\theta^2} \end{aligned}$$

and the observed information is

$$\begin{aligned} I(\hat{\theta}) &= \frac{n\bar{x}}{\hat{\theta}^2} = \frac{n\hat{\theta}}{\hat{\theta}^2} \\ &= \frac{n}{\hat{\theta}} \end{aligned}$$

If $n = 10$ then $I(\hat{\theta}) = I(2) = n/\hat{\theta} = 10/2 = 5$. If $n = 25$ then $I(\hat{\theta}) = I(2) = 25/2 = 12.5$. See Figure 6.1. The function $r(\theta) = l(\theta) - l(\hat{\theta})$ is more concave and symmetric for $n = 25$ than for $n = 10$. As the number of observations increases we have more "information" about the unknown parameter θ.

Although we view the likelihood, log likelihood, score and information functions as functions of θ, they are also functions of the observed data $\mathbf{x}$. When it is important to emphasize the dependence on the data $\mathbf{x}$ we will write $L(\theta; \mathbf{x})$, $S(\theta; \mathbf{x})$, and $I(\theta; \mathbf{x})$. When we wish to determine the sampling distribution of the corresponding random variables we will write $L(\theta; \mathbf{X})$, $S(\theta; \mathbf{X})$, and $I(\theta; \mathbf{X})$.

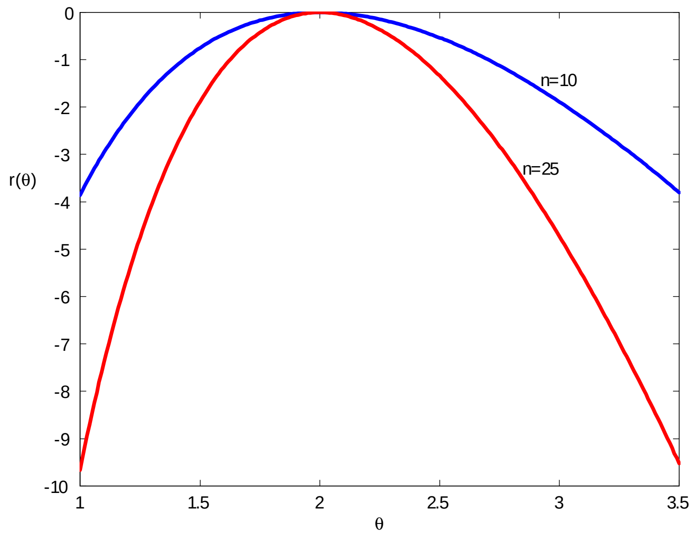


Figure 6.1: Poisson Log Likelihoods for $n = 10$ and $n = 25$

Here is one more function which plays an important role in the method of maximum likelihood.

6.3.4 Definition - Expected Information Function

If θ is a scalar then the *expected information function* is given by

$$\begin{aligned} J(\theta) &= E\left[I(\theta;\mathbf{X})\right] \\ &= E\left[-\frac{d^2}{d\theta^2} l(\theta;\mathbf{X})\right] \quad \text{for } \theta \in \Omega \end{aligned}$$

where $\mathbf{X}$ is the potential data.

Note:

If $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ is a random sample from $f(x;\theta)$ then

$$\begin{aligned} J(\theta) &= E\left[-\frac{d^2}{d\theta^2} l(\theta;\mathbf{X})\right] \\ &= nE\left[-\frac{d^2}{d\theta^2} \log f(X;\theta)\right] \quad \text{for } \theta \in \Omega \end{aligned}$$

where X has probability density function $f(x;\theta)$.

6.3.5 Example

For each of the following find the observed information $I(\hat{\theta})$ and the expected information $J(\theta)$. Compare $I(\hat{\theta})$ and $J(\hat{\theta})$. Determine the mean and variance of the maximum likelihood estimator $\tilde{\theta}$. Compare the expected information with the variance of the maximum likelihood estimator.

(a) Example 6.2.4 (Binomial)

(b) Example 6.2.5 (Poisson)

(c) Example 6.2.8 (Exponential)

Solution

(a) From Example 6.2.4, the score function based on x successes in n Bernoulli trials is

$$S(\theta) = \frac{d}{d\theta} l(\theta) = \frac{x - n\theta}{\theta(1-\theta)} \quad \text{for } 0 < \theta < 1$$

Therefore the information function is

$$\begin{aligned} I(\theta) &= -\frac{d^2}{d\theta^2} l(\theta) = -\frac{d}{d\theta}\left(\frac{x}{\theta} - \frac{n-x}{1-\theta}\right) = -\left[\frac{-x}{\theta^2} - \frac{n-x}{(1-\theta)^2}\right] \\ &= \frac{x}{\theta^2} + \frac{n-x}{(1-\theta)^2} \quad \text{for } 0 < \theta < 1 \end{aligned}$$

Since the maximum likelihood estimate is $\hat{\theta} = \frac{x}{n}$ the observed information is

$$\begin{aligned} I(\hat{\theta}) &= \frac{x}{\hat{\theta}^2} + \frac{n-x}{(1-\hat{\theta})^2} = \frac{x}{\left(\frac{x}{n}\right)^2} + \frac{n-x}{\left(1-\frac{x}{n}\right)^2} \\ &= n\left[\frac{\frac{x}{n}}{\left(\frac{x}{n}\right)^2} + \frac{1-\frac{x}{n}}{\left(1-\frac{x}{n}\right)^2}\right] = n\left[\frac{1}{\frac{x}{n}} + \frac{1}{1-\frac{x}{n}}\right] \\ &= \frac{n}{\left(\frac{x}{n}\right)\left(1-\frac{x}{n}\right)} \\ &= \frac{n}{\hat{\theta}(1-\hat{\theta})} \end{aligned}$$

If X has a Binomial(n, θ) distribution then $E(X) = n\theta$ and $Var(X) = n\theta(1-\theta)$. Therefore the expected information is

$$\begin{aligned} J(\theta) &= E[I(\theta; X)] = E\left[\frac{X}{\theta^2} + \frac{n-X}{(1-\theta)^2}\right] = \frac{n\theta}{\theta^2} + \frac{n(1-\theta)}{(1-\theta)^2} \\ &= \frac{n[(1-\theta)+\theta]}{\theta(1-\theta)} \\ &= \frac{n}{\theta(1-\theta)} \quad \text{for } 0 < \theta < 1 \end{aligned}$$

We note that $I(\hat{\theta}) = J(\hat{\theta})$.

The maximum likelihood estimator is $\tilde{\theta} = X/n$ with

$$E(\tilde{\theta}) = E\left(\frac{X}{n}\right) = \frac{1}{n}E\left(X\right) = \frac{n\theta}{n} = \theta$$

and

$$\begin{aligned} Var(\tilde{\theta}) &= Var\left(\frac{X}{n}\right) = \frac{1}{n^2}Var\left(X\right) = \frac{n\theta\left(1-\theta\right)}{n^2} = \frac{\theta\left(1-\theta\right)}{n} \\ &= \frac{1}{J\left(\theta\right)} \end{aligned}$$

(b) From Example 6.3.3 the information function based on Poisson data $x_1, x_2, \ldots, x_n$ is

$$I(\theta) = \frac{n\bar{x}}{\theta^2} \quad \text{for } \theta > 0$$

Since the maximum likelihood estimate is $\hat{\theta} = \bar{x}$ the observed information is

$$I(\hat{\theta}) = \frac{n\bar{x}}{\hat{\theta}^2} = \frac{n}{\hat{\theta}}$$

We note that $I(\hat{\theta}) = J(\hat{\theta})$.

Since X_i has a Poisson(θ) distribution with $E\left(X_i\right) = Var\left(X_i\right) = \theta$ then $E\left(\bar{X}\right) = \theta$ and $Var\left(\bar{X}\right) = \theta/n$. Therefore the expected information is

$$\begin{aligned} J(\theta) &= E\left[I(\theta; X_1, X_2, \ldots, X_n)\right] = E\left(\frac{n\bar{X}}{\theta^2}\right) = \frac{n}{\theta^2}\left(\theta\right) \\ &= \frac{n}{\theta} \quad \text{for } \theta > 0 \end{aligned}$$

The maximum likelihood estimator is $\tilde{\theta} = \bar{X}$ with

$$E(\tilde{\theta}) = E\left(\bar{X}\right) = \theta$$

and

$$Var(\tilde{\theta}) = Var\left(\bar{X}\right) = \frac{\theta}{n} = \frac{1}{J\left(\theta\right)}$$

(c) From Example 6.2.8 the score function based on Exponential data $x_1, x_2, \ldots, x_n$ is

$$S(\theta) = \frac{d}{d\theta}l\left(\theta\right) = -n\left(\frac{1}{\theta} - \frac{\bar{x}}{\theta^2}\right) \quad \text{for } \theta > 0$$

Therefore the information function is

$$\begin{aligned} I\left(\theta\right) &= -\frac{d^2}{d\theta^2}l\left(\theta\right) = n\frac{d}{d\theta}\left(\frac{1}{\theta} - \frac{\bar{x}}{\theta^2}\right) \\ &= n\left(\frac{-1}{\theta^2} + \frac{2\bar{x}}{\theta^3}\right) \quad \text{for } \theta > 0 \end{aligned}$$

Since the maximum likelihood estimate is $\hat{\theta} = \bar{x}$ the observed information is

$$I(\hat{\theta}) = n\left(\frac{-1}{\hat{\theta}^2} + \frac{2\hat{\theta}}{\hat{\theta}^3}\right) = \frac{n}{(\hat{\theta})^2}$$

Since X_i has a Exponential(θ) distribution with $E(X_i) = \theta$ and $Var(X_i) = \theta^2$ then $E(\bar{X}) = \theta$ and $Var(\bar{X}) = \theta^2/n$. Therefore the expected information is

$$\begin{aligned} J(\theta) &= E\left[I(\theta; X_1, X_2, \ldots, X_n)\right] = nE\left(\frac{-1}{\theta^2} + \frac{2\bar{X}}{\theta^3}\right) = n\left(\frac{-1}{\theta^2} + \frac{2\theta}{\theta^2}\right) \\ &= \frac{n}{\theta^2} \quad \text{for } \theta > 0 \end{aligned}$$

We note that $I(\hat{\theta}) = J(\hat{\theta})$.

The maximum likelihood estimator is $\tilde{\theta} = \bar{X}$ with

$$E(\tilde{\theta}) = E\left(\bar{X}\right) = \theta$$

and

$$Var(\tilde{\theta}) = Var\left(\bar{X}\right) = \frac{\theta^2}{n} = \frac{1}{J(\theta)}$$

In all three examples we have $I(\hat{\theta}) = J(\hat{\theta})$, $E(\tilde{\theta}) = \theta$, and $Var(\tilde{\theta}) = [J(\theta)]^{-1}$.

In the three previous examples, we observed that $E(\tilde{\theta}) = \theta$ and therefore $\tilde{\theta}$ was an unbiased estimator of θ. This is not always true for maximum likelihood estimators as we see in the next example. However, maximum likelihood estimators usually have other good properties. Suppose $\tilde{\theta}_n = \tilde{\theta}_n(X_1, X_2, \ldots, X_n)$ is the maximum likelihood estimator based on a sample of size n. If $\lim\limits_{n\to\infty} E(\tilde{\theta}_n) = \theta$ then $\tilde{\theta}_n$ is an *asymptotically unbiased estimator* of θ. If $\tilde{\theta}_n \to_p \theta$ then $\tilde{\theta}_n$ is called a *consistent estimator* of θ.

6.3.6 Example

Suppose $x_1, x_2, \ldots, x_n$ is an observed random sample from the distribution with probability density function

$$f(x;\theta) = \theta x^{\theta-1} \quad \text{for } 0 \le x \le 1 \quad \text{for } \theta > 0 \tag{6.1}$$

(a) Find the score function, the maximum likelihood estimator, the information function, the observed information, and the expected information.

(b) Show that $T = -\sum\limits_{i=1}^{n} \log X_i \sim \text{Gamma}\left(n, \frac{1}{\theta}\right)$

(c) Use (b) and 2.7.9 to show that $\tilde{\theta}$ is not an unbiased estimator of θ. Show however that $\tilde{\theta}$ is an aysmptotically unbiased estimator of θ.

(d) Show that $\tilde{\theta}$ is a consistent estimator of θ.

(e) Use (b) and 2.7.9 to find $Var(\tilde{\theta})$. Compare $Var(\tilde{\theta})$ with the expected information.

Solution

(a) The likelihood function is

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{n} f(x_i;\theta) = \prod_{i=1}^{n} \theta x_i^{\theta-1} \\ &= \theta^n \left(\prod_{i=1}^{n} x_i\right)^{\theta-1} \quad \text{for } \theta > 0 \end{aligned}$$

or more simply

$$L(\theta) = \theta^n \left(\prod_{i=1}^{n} x_i\right)^{\theta} \quad \text{for } \theta > 0$$

The log likelihood function is

$$\begin{aligned} l(\theta) &= n\log\theta + \theta\sum_{i=1}^{n}\log x_i \\ &= n\log\theta - \theta t \quad \text{for } \theta > 0 \end{aligned}$$

where $t = -\sum_{i=1}^{n}\log x_i$. The score function is

$$\begin{aligned} S(\theta) &= \frac{d}{d\theta} l(\theta) \\ &= \frac{n}{\theta} - t \\ &= \frac{1}{\theta}(n - \theta t) \quad \text{for } \theta > 0 \end{aligned}$$

Now $\frac{d}{d\theta} l(\theta) = 0$ for $\theta = n/t$. Since $\frac{d}{d\theta} l(\theta) > 0$ if $0 < \theta < n/t$ and $\frac{d}{d\theta} l(\theta) < 0$ if $\theta > n/t$ then, by the first derivative test, $l(\theta)$ has a absolute maximum at $\theta = n/t$. Therefore the maximum likelihood estimate of θ is $\hat{\theta} = n/t$ and the maximum likelihood estimator is $\tilde{\theta} = n/T$ where $T = -\sum_{i=1}^{n}\log X_i$.

The information function is

$$\begin{aligned} I(\theta) &= -\frac{d^2}{d\theta^2} l(\theta) \\ &= \frac{n}{\theta^2} \quad \text{for } \theta > 0 \end{aligned}$$

and the observed information is

$$I(\hat{\theta}) = \frac{n}{\hat{\theta}^2}$$

The expected information is

$$\begin{aligned} J(\theta) &= E\left[I(\theta; X_1, X_2, \ldots, X_n)\right] = E\left(\frac{n}{\theta^2}\right) \\ &= \frac{n}{\theta^2} \quad \text{for } \theta > 0 \end{aligned}$$

(b) From Exercise 2.6.12 we have that if X_i has the probability density function (6.1) then

$$Y_i = -\log X_i \sim Exponential\left(\frac{1}{\theta}\right) \quad \text{for } i = 1, 2, \ldots, n \tag{6.2}$$

From 4.3.2(4) have that

$$T = -\sum_{i=1}^{n} \log X_i = \sum_{i=1}^{n} Y_i \sim \text{Gamma}\left(n, \frac{1}{\theta}\right)$$

(c) Since $T \sim \text{Gamma}\left(n, \frac{1}{\theta}\right)$ then from 2.7.9 we have

$$E\left(T^p\right) = \frac{\Gamma(n+p)}{\theta^p \Gamma(n)} \quad \text{for } p > -n$$

Assuming $n > 1$ we have

$$E\left(T^{-1}\right) = \frac{\Gamma(n-1)}{\theta^{-1}\Gamma(n)} = \frac{\theta}{n-1}$$

so

$$\begin{aligned} E\left(\tilde{\theta}\right) &= E\left(\frac{n}{T}\right) = nE\left(\frac{1}{T}\right) \\ &= nE\left(T^{-1}\right) \\ &= n\left(\frac{\theta}{n-1}\right) \\ &= \frac{\theta}{\left(1 - \frac{1}{n}\right)} \neq \theta \end{aligned}$$

and therefore $\tilde{\theta}$ is not an unbiased estimator of θ.

Now $\tilde{\theta}$ is an estimator based on a sample of size n. Since

$$\lim_{n\to\infty} E\left(\tilde{\theta}\right) = \lim_{n\to\infty} \frac{\theta}{\left(1 - \frac{1}{n}\right)} = \theta$$

therefore $\tilde{\theta}$ is an asymptotically unbiased estimator of θ.

(d) By (6.2), $Y_1, Y_2, \ldots, Y_n$ are independent and identically distributed random variables with $E(Y_i) = \frac{1}{\theta}$ and $Var(Y_i) = \frac{1}{\theta^2} < \infty$. Therefore by the Weak Law of Large Numbers

$$\frac{T}{n} = \frac{1}{n}\sum_{i=1}^{n} Y_i \to_p \frac{1}{\theta}$$

and by the Limit Theorems

$$\tilde{\theta} = \frac{n}{T} \to_p \theta$$

Thus $\tilde{\theta}$ is a consistent estimator of θ.

(e) To determine $Var(\tilde{\theta})$ we note that

$$\begin{aligned}
Var(\tilde{\theta}) &= Var\left(\frac{n}{T}\right) \\
&= n^2 Var\left(\frac{1}{T}\right) \\
&= n^2\left\{E\left[\left(\frac{1}{T}\right)^2\right] - \left[E\left(\frac{1}{T}\right)\right]^2\right\} \\
&= n^2\left\{E\left(T^{-2}\right) - \left[E\left(T^{-1}\right)\right]^2\right\}
\end{aligned}$$

Since

$$E\left(T^{-2}\right) = \frac{\Gamma(n-2)}{\theta^{-2}\Gamma(n)} = \frac{\theta^2}{(n-1)(n-2)}$$

then

$$\begin{aligned}
Var(\tilde{\theta}) &= n^2\left\{E\left(T^{-2}\right) - \left[E\left(T^{-1}\right)\right]^2\right\} \\
&= n^2\left[\frac{\theta^2}{(n-1)(n-2)} - \left(\frac{\theta}{n-1}\right)^2\right] \\
&= n^2\theta^2\left[\frac{(n-1)-(n-2)}{(n-1)^2(n-2)}\right] \\
&= \frac{\theta^2}{\left(1-\frac{1}{n}\right)^2(n-2)}
\end{aligned}$$

We note that $Var\left(\tilde{\theta}\right) \neq \frac{\theta^2}{n} = \frac{1}{J(\theta)}$, however for large n, $Var\left(\tilde{\theta}\right) \approx \frac{1}{J(\theta)}$.

6.3.7 Example

Suppose $x_1, x_2, \ldots, x_n$ is an observed random sample from the Weibull$(\theta, 1)$ distribution with probability density function

$$f(x;\theta) = \theta x^{\theta-1}e^{-x^\theta} \quad \text{for } x > 0,\ \theta > 0$$

Find the score function and the information function. How would you find the maximum likelihood estimate of θ?

Solution

The likelihood function is

$$\begin{aligned}
L(\theta) &= \prod_{i=1}^{n} f(x_i;\theta) = \prod_{i=1}^{n} \theta x_i^{\theta-1}e^{-x_i^\theta} \\
&= \theta^n\left(\prod_{i=1}^{n} x_i\right)^{\theta-1} \exp\left(-\sum_{i=1}^{n} x_i^\theta\right) \quad \text{for } \theta > 0
\end{aligned}$$

or more simply

$$L(\theta) = \theta^n \left(\prod_{i=1}^{n} x_i \right)^{\theta} \exp\left(-\sum_{i=1}^{n} x_i^{\theta} \right) \quad \text{for } \theta > 0$$

The log likelihood function is

$$l(\theta) = n \log \theta + \theta \sum_{i=1}^{n} \log x_i - \sum_{i=1}^{n} x_i^{\theta} \quad \text{for } \theta > 0$$

The score function is

$$\begin{aligned} S(\theta) &= \frac{d}{d\theta} l(\theta) \\ &= \frac{n}{\theta} + t - \sum_{i=1}^{n} x_i^{\theta} \log x_i \quad \text{for } \theta > 0 \end{aligned}$$

where $t = \sum_{i=1}^{n} \log x_i$.

Notice that $S(\theta) = 0$ cannot be solved explicitly. The maximum likelihood estimate can only be determined numerically for a given sample of data $x_1, x_2, \ldots, x_n$. Since

$$\frac{d}{d\theta} S(\theta) = -\left[\frac{n}{\theta^2} + \sum_{i=1}^{n} x_i^{\theta} (\log x_i)^2 \right] \quad \text{for } \theta > 0$$

is negative for all values of $\theta > 0$ then we know that the function $S(\theta)$ is always decreasing and therefore there is only one solution to $S(\theta) = 0$. The solution to $S(\theta) = 0$ gives the maximum likelihood estimate.

The information function is

$$\begin{aligned} I(\theta) &= -\frac{d^2}{d\theta^2} l(\theta) \\ &= \frac{n}{\theta^2} + \sum_{i=1}^{n} x_i^{\theta} (\log x_i)^2 \quad \text{for } \theta > 0 \end{aligned}$$

To illustrate how to find the maximum likelihood estimate for a given sample of data, we randomly generate 20 observations from the Weibull$(\theta, 1)$ distribution. To do this we use the result of Example 2.6.7 in which we showed that if u is an observation from the Uniform$(0, 1)$ distribution then $x = [-\log(1-u)]^{1/\theta}$ is an observation from the Weibull$(\theta, 1)$ distribution.

The following R code generates the data, plots the likelihood function, finds $\hat{\theta}$ by solving $S(\theta) = 0$ using the R function `uniroot`, and determines $S(\hat{\theta})$ and the observed information $I(\hat{\theta})$.

```
# randomly generate 20 observations from a Weibull(theta,1)
# using a random theta value between 0.5 and 1.5
set.seed(20086689)  # set the seed so results can be reproduced
truetheta<-runif(1,min=0.5,max=1.5)
# data are sorted and rounded to two decimal places for easier display
x<-sort(round((-log(1-runif(20)))^(1/truetheta),2))
x
#
# function for calculating Weibull likelihood for data x and theta=th
WBLF<-function(th,x)
{n<-length(x)
L<-th^n*prod(x)^th*exp(-sum(x^th))
return(L)}
#
# function for calculating Weibull score for data x and theta=th
WBSF<-function(th,x)
{n<-length(x)
t<-sum(log(x))
S<-(n/th)+t-sum(log(x)*x^th)
return(S)}
#
# function for calculating Weibull information for data x and theta=th
WBIF<-function(th,x)
{n<-length(x)
I<-(n/th^2)+sum(log(x)^2*x^th)
return(I)}
#
# plot the Weibull likelihood function
th<-seq(0.25,0.75,0.01)
L<-sapply(th,WBLF,x)
plot(th,R,"l",xlab=expression(theta),
ylab=expression(paste("L(",theta,")")),lwd=3)
#
# find thetahat using uniroot function
thetahat<-uniroot(function(th) WBSF(th,x),lower=0.4,upper=0.6)$root
cat("thetahat = ",thetahat) # display value of thetahat
# display value of Score function at thetahat
cat("S(thetahat) = ",WBSF(thetahat,x))
# calculate observed information
cat("Observed Information = ",WBIF(thetahat,x))
```

The generated data are

0.01	0.01	0.05	0.07	0.10	0.11	0.23	0.28	0.44	0.46
0.64	1.07	1.16	1.25	2.40	3.03	3.65	5.90	6.60	30.07

The likelihood function is graphed in Figure 6.2. The solution to $S(\theta) = 0$ determined by `uniroot` was $\hat{\theta} = 0.4951607$. $S(\hat{\theta}) = -2.468574 \times 10^{-5}$ which is close to zero and the observed information was $I(\hat{\theta}) = 181.8069$ which is positive and indicates a local maximum. We have already shown in this example that the solution to $S(\theta) = 0$ is the unique maximum likelihood estimate.

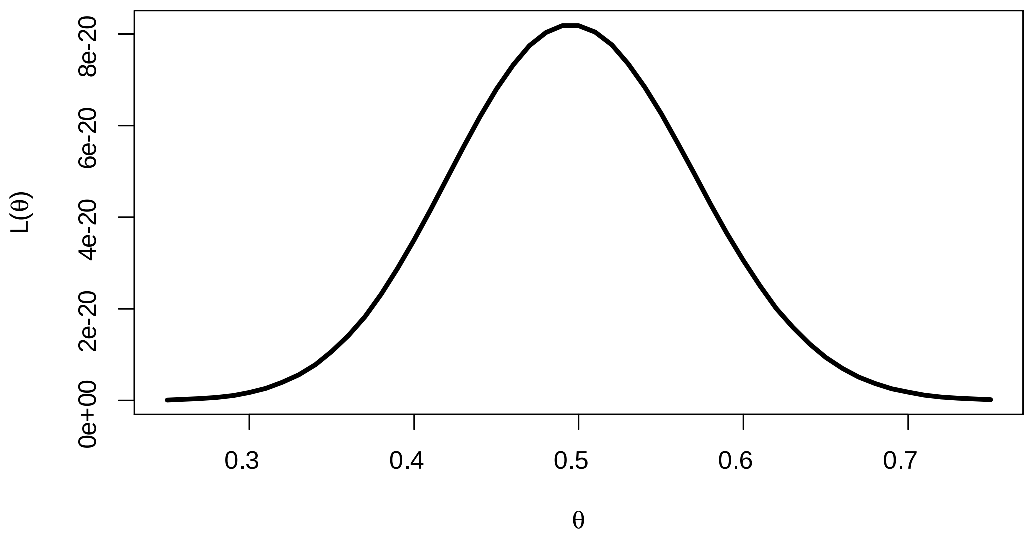


Figure 6.2: Likelihood function for Example 6.3.7

Note that the interval $[0.25, 0.75]$ used to graph the likelihood function was determined by trial and error. The values `lower=0.4` and `upper=0.6` used for `uniroot` were determined from the graph of the likelihood function. From the graph is it easy to see that the value of $\hat{\theta}$ lies in the interval $[0.4, 0.6]$.

Newton's Method, which is a numerical method for finding the roots of an equation usually discussed in first year calculus, is a method which can be used for finding the maximum likelihood estimate. Newton's Method usually works quite well for finding maximum likelihood estimates because the likelihood function is often very quadratic in shape.

6.3.8 Newton's Method

Let $\theta^{(0)}$ be an initial estimate of θ. The estimate $\theta^{(i)}$ can be updated using

$$\theta^{(i+1)} = \theta^{(i)} + \frac{S(\theta^{(i)})}{I(\theta^{(i)})} \quad \text{for } i = 0, 1, \ldots$$

Notes:

(1) The initial estimate, $\theta^{(0)}$, may be determined by graphing $L(\theta)$ or $l(\theta)$.

(2) The algorithm is usually run until the value of $\theta^{(i)}$ no longer changes to a reasonable number of decimal places. When the algorithm is stopped it is always important to check that the value of θ obtained does indeed maximize $L(\theta)$.

(3) This algorithm is also called the Newton-Raphson Method.

(4) $I(\theta)$ can be replaced by $J(\theta)$ for a similar algorithm which is called the Method of Scoring.

(5) If the support set of X depends on θ (e.g. Uniform$(0, \theta)$) then $\hat{\theta}$ is not found by solving $S(\theta) = 0$.

6.3.9 Example

Use Newton's Method to find the maximum likelihood in Example 6.3.7.

Solution

Here is R code for Newton's Method for the Weibull Example

```
# Newton's Method for Weibull Example
NewtonWB<-function(th,x)
{thold<-th
thnew<-th+0.1
while (abs(thold-thnew)>0.00001)
{thold<-thnew
thnew<-thold+WBSF(thold,x)/WBIF(thold,x)
print(thnew)}
return(thnew)}
#
thetahat<-NewtonWB(0.2,x)
```

Newton's Method converges after four iterations and the value of thetahat returned is $\hat{\theta} = 0.4951605$ which is the same value to six decimal places as was obtained above using the `uniroot` function.

6.4 Likelihood Intervals

In your previous statistics course, likelihood intervals were introduced as one approach to constructing an interval estimate for the unknown parameter θ.

6.4.1 Definition - Relative Likelihood Function

The *relative likelihood function* $R(\theta)$ is defined by

$$R(\theta) = R(\theta; \mathbf{x}) = \frac{L(\theta)}{L(\hat{\theta})} \quad \text{for } \theta \in \Omega$$

where $\mathbf{x}$ are the observed data.

The relative likelihood function takes on values between 0 and 1 and can be used to rank parameter values according to their plausibilities in light of the observed data. If $R(\theta_1) = 0.1$, for example, then θ_1 is rather an implausible parameter value because the data are ten times more probable when $\theta = \hat{\theta}$ than they are when $\theta = \theta_1$. However, if $R(\theta_1) = 0.5$, then θ_1 is a fairly plausible value because it gives the data 50% of the maximum possible probability under the model.

6.4.2 Definition - Likelihood Interval

The set of θ values for which $R(\theta) \geq p$ is called a $100p\%$ *likelihood interval* for θ.

Values of θ inside a 10% likelihood interval are referred to as plausible values in light of the observed data. Values of θ outside a 10% likelihood interval are referred to as implausible values given the observed data. Values of θ inside a 50% likelihood interval are very plausible and values of θ outside a 1% likelihood interval are very implausible in light of the data.

6.4.3 Definition - Log Relative Function

The *log relative likelihood function* is the natural logarithm of the relative likelihood function:

$$r(\theta) = r(\theta; \mathbf{x}) = \log[R(\theta)] \quad \text{for } \theta \in \Omega$$

where $\mathbf{x}$ are the observed data.

Likelihood regions or intervals may be determined from a graph of $R(\theta)$ or $r(\theta)$. Alternatively, they can be found by solving $R(\theta) - p = 0$ or $r(\theta) - \log p = 0$. In most cases this must be done numerically.

6.4.4 Example

Plot the relative likelihood function for θ in Example 6.3.7. Find 10% and 50% likelihood intervals for θ.

Solution

Here is R code to plot the relative likelihood function for the Weibull Example with lines for determining 10% and 50% likelihood intervals for θ as well as code to determine these intervals using `uniroot`.

The R function `WBRLF` uses the R function `WBLF` from Example 6.3.7.

```
# function for calculating Weibull relative likelihood function
WBRLF<-function(th,thetahat,x)
{R<-WBLF(th,x)/WBLF(thetahat,x)
return(R)}
#
# plot the Weibull relative likelihood function
th<-seq(0.25,0.75,0.01)
R<-sapply(th,WBRLF,thetahat,x)
plot(th,R,"l",xlab=expression(theta),
   ylab=expression(paste("R(",theta,")")),lwd=3)
# add lines to determine 10% and 50% likelihood intervals
abline(a=0.10,b=0,col="red",lwd=2)
abline(a=0.50,b=0,col="blue",lwd=2)
#
# use uniroot to determine endpoints of 10%, 15%, and 50% likelihood intervals
uniroot(function(th) WBRLF(th,thetahat,x)-0.1,lower=0.3,upper=0.4)$root
uniroot(function(th) WBRLF(th,thetahat,x)-0.1,lower=0.6,upper=0.7)$root
uniroot(function(th) WBRLF(th,thetahat,x)-0.5,lower=0.35,upper=0.45)$root
uniroot(function(th) WBRLF(th,thetahat,x)-0.5,lower=0.55,upper=0.65)$root
uniroot(function(th) WBRLF(th,thetahat,x)-0.15,lower=0.3,upper=0.4)$root
uniroot(function(th) WBRLF(th,thetahat,x)-0.15,lower=0.6,upper=0.7)$root
```

The graph of the relative likelihood function is given in Figure 6.3.

The upper and lower values used for `uniroot` were determined using this graph.

The 10% likelihood interval for θ is $[0.34, 0.65]$.

The 50% likelihood interval for θ is $[0.41, 0.58]$.

For later reference the 15% likelihood interval for θ is $[0.3550, 0.6401]$.

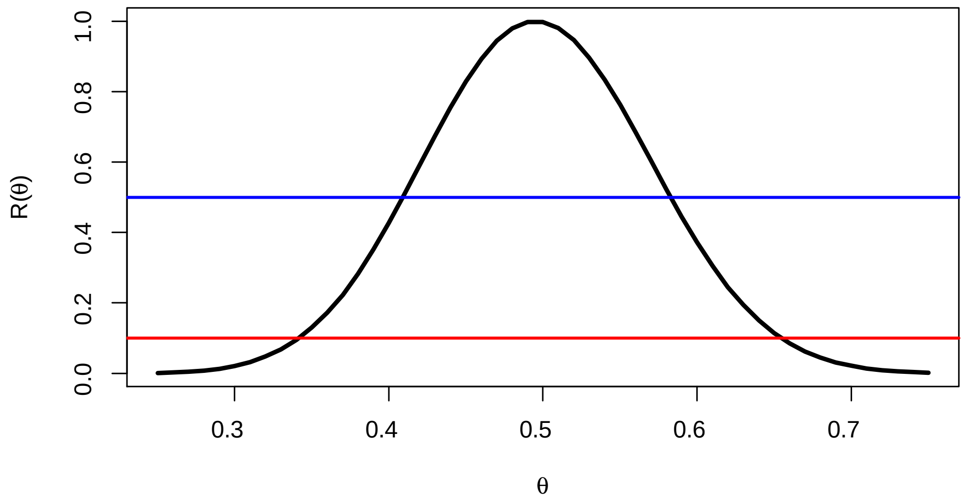


Figure 6.3: Relative Likelihood function for Example 6.3.7

6.4.5 Example

Suppose $x_1, x_2, \ldots, x_n$ is an observed random sample from the Uniform$(0, \theta)$ distribution. Plot the relative likelihood function for θ if $n = 20$ and $x_{(20)} = 1.5$. Find 10% and 50% likelihood intervals for θ.

Solution

From Example 6.2.9 the likelihood function for $n = 10$ and $\hat{\theta} = x_{(10)} = 0.5$ is

$$L(\theta) = \begin{cases} 0 & \text{if } 0 < \theta < 0.5 \\ \frac{1}{\theta^{10}} & \text{if } \theta \geq 0.5 \end{cases}$$

The relative likelihood function is

$$R(\theta) = \begin{cases} 0 & \text{if } 0 < \theta < 0.5 \\ \left(\frac{0.5}{\theta}\right)^{10} & \text{if } \theta \geq 0.5 \end{cases}$$

is graphed in Figure 6.4 along with lines for determining 10% and 50% likelihood intervals.

To determine the value of θ at which the horizontal line $R = p$ intersects the graph of $R(\theta)$ we solve $\left(\frac{0.5}{\theta}\right)^{10} = p$ to obtain $\theta = 0.5p^{-1/10}$. Since $R(\theta) = 0$ if $0 < \theta < 0.5$ then a $100p\%$ likelihood interval for θ is of the form $\left[0.5, 0.5p^{-1/10}\right]$.

A 10% likelihood interval is

$$\left[0.5, 0.5\,(0.1)^{-1/10}\right] = [0.5, 0.629]$$

A 50% likelihood interval is

$$\left[0.5, 0.5\,(0.5)^{-1/10}\right] = [0.5, 0.536]$$

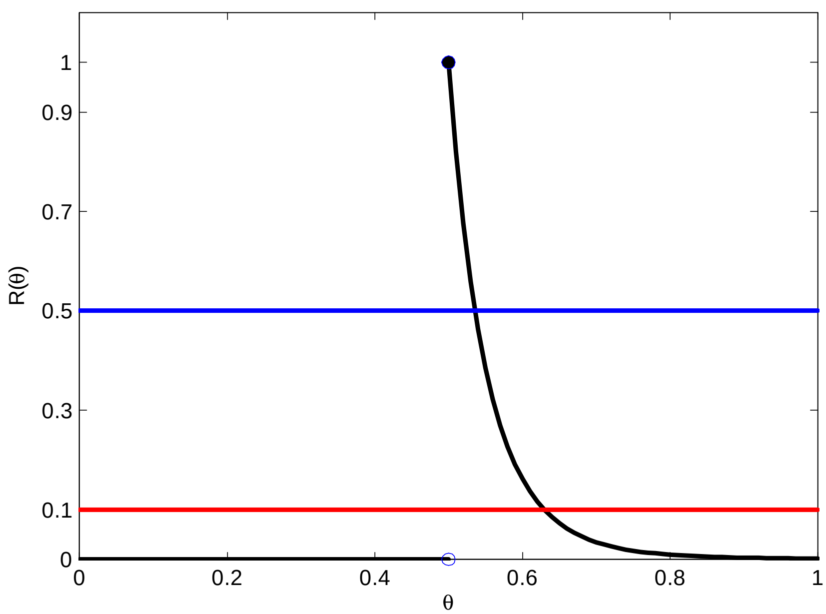


Figure 6.4: Relative likelihood function for Example 6.4.5

More generally for an observed random sample $x_1, x_2, \ldots, x_n$ from the Uniform$(0, \theta)$ distribution a $100p\%$ likelihood interval for θ will be of the form $\left[x_{(n)}, x_{(n)} p^{-1/n}\right]$.

6.4.6 Exercise

Suppose $x_1, x_2, \ldots, x_n$ is an observed random sample from the Two Parameter Exponential$(1, \theta)$ distribution. Plot the relative likelihood function for θ if $n = 12$ and $x_{(1)} = 2$. Find 10% and 50% likelihood intervals for θ.

6.5 Limiting Distribution of Maximum Likelihood Estimator

If there is no explicit expression for the maximum likelihood estimator $\tilde{\theta}$, as in Example 6.3.7, then its sampling distribution can only be obtained by simulation. This makes it difficult to determine the properties of the maximum likelihood estimator. In particular, to determine how good an estimator is, we look at its mean and variance. The mean indicates whether the estimator is close, on average, to the true value of θ, while the variance indicates the uncertainty in the estimator. The larger the variance the more uncertainty in our estimation. To determine how good an estimator is we also look at how the mean and variance behave as the sample size $n \to \infty$.

We saw in Example 6.3.5 that $E(\tilde{\theta}) = \theta$ and $Var(\tilde{\theta}) = [J(\theta)]^{-1} \to 0$ as $n \to \infty$ for the Binomial, Poisson, Exponential models. For each of these models the maximum likelihood estimator is an unbiased estimator of θ. In Example 6.3.6 we saw that $E(\tilde{\theta}) \to \theta$ and $Var(\tilde{\theta}) \to [J(\theta)]^{-1} \to 0$ as $n \to \infty$. The maximum likelihood estimator $\tilde{\theta}$ is an asymptotically unbiased estimator. In Example 6.3.7 we are not able to determine $E(\tilde{\theta})$ and $Var(\tilde{\theta})$.

The following theorem gives the limiting distribution of the maximum likelihood estimator in general under certain restrictions.

6.5.1 Theorem - Limiting Distribution of Maximum Likelihood Estimator

Suppose $\mathbf{X}_n = (X_1, X_2, \ldots, X_n)$ be a random sample from the probability (density) function $f(x;\theta)$ for $\theta \in \Omega$. Let $\tilde{\theta}_n = \tilde{\theta}_n(\mathbf{X}_n)$ be the maximum likelihood estimator of θ based on $\mathbf{X}_n$.

Then under certain (regularity) conditions

$$\tilde{\theta}_n \to_p \theta \tag{6.3}$$

$$(\tilde{\theta}_n - \theta)[J(\theta)]^{1/2} \to_D Z \sim \mathrm{N}(0,1) \tag{6.4}$$

$$-2\log R(\theta;\mathbf{X}_n) = -2\log\left[\frac{L(\theta;\mathbf{X}_n)}{L(\tilde{\theta}_n;\mathbf{X}_n)}\right] \to_D W \sim \chi^2(1) \tag{6.5}$$

for each $\theta \in \Omega$.

The proof of this result which depends on applying Taylor's Theorem to the score function is beyond the scope of this course. The regularity conditions are a bit complicated but essentially they are a set of conditions which ensure that the error term in Taylor's Theorem goes to zero as $n \to \infty$. One of the conditions is that the support set of $f(x;\theta)$ does not depend on θ. Therefore, for example, this theorem cannot be applied to the maximum likelihood estimator in the case of a random sample from the Uniform$(0,\theta)$ distribution.

This is actually not a problem since the distribution of the maximum likelihood estimator in this case can be determined exactly.

Since (6.3) holds $\tilde{\theta}_n$ is called a *consistent estimator* of θ.

Theorem 6.5.1 implies that for sufficiently large n, $\tilde{\theta}_n$ has an approximately $\mathrm{N}(\theta, 1/J(\theta))$ distribution. Therefore for large n

$$E(\tilde{\theta}_n) \approx \theta$$

and the maximum likelihood estimator is an asymptotically unbiased estimator of θ.

Since $\tilde{\theta}_n$ has an approximately $\mathrm{N}(\theta, 1/J(\theta))$ distribution this also means that for sufficiently large n

$$Var(\tilde{\theta}_n) \approx \frac{1}{J(\theta)}$$

$1/J(\theta)$ is called the *asymptotic variance* of $\tilde{\theta}_n$. Of course $J(\theta)$ is unknown because θ is unknown. By (6.3), (6.4), and Slutsky's Theorem

$$(\tilde{\theta}_n - \theta)\sqrt{J(\tilde{\theta}_n)} \to_D Z \sim \mathrm{N}(0,1) \tag{6.6}$$

which implies that the asymptotic variance of $\tilde{\theta}_n$ can be estimated using $1/J(\hat{\theta}_n)$. Therefore for sufficiently large n we have

$$Var(\tilde{\theta}_n) \approx \frac{1}{J(\hat{\theta}_n)}$$

By the Weak Law of Large Numbers

$$\frac{1}{n} I(\theta; \mathbf{X}_n) = -\frac{1}{n}\sum_{i=1}^{n} \frac{d^2}{d\theta^2} l(\theta; X_i) \to_p E\left[-\frac{d^2}{d\theta^2} l(\theta; X_i)\right] \tag{6.7}$$

Therefore by (6.3), (6.4), (6.7) and the Limit Theorems it follows that

$$(\tilde{\theta}_n - \theta)\sqrt{I(\tilde{\theta}_n; \mathbf{X}_n)} \to_D Z \sim \mathrm{N}(0,1) \tag{6.8}$$

which implies the asymptotic variance of $\tilde{\theta}_n$ can be also be estimated using $1/I(\hat{\theta}_n)$ where $I(\hat{\theta}_n)$ is the observed information. Therefore for sufficiently large n we have

$$Var(\tilde{\theta}_n) \approx \frac{1}{I(\hat{\theta}_n)}$$

Results (6.6), (6.8) and (6.5) can be used to construct approximate confidence intervals for θ.

In Chapter 8 we will see how result (6.5) can be used in a test of hypothesis.

Although we will not prove Theorem 6.5.1 in general we can prove the results in a particular case. The following example illustrates how techniques and theorems from previous chapters can be used together to obtained the results of interest. It is also a good review of several ideas covered thus far in these Course Notes.

6.5.2 Example

(a) Suppose $X \sim$ Weibull$(2, \theta)$. Show that $E\left(X^2\right) = \theta^2$ and $Var\left(X^2\right) = \theta^4$.

(b) Suppose $X_1, X_2, \ldots, X_n$ is a random sample from the Weibull$(2, \theta)$ distribution. Find the maximum likelihood estimator $\tilde{\theta}$ of θ, the information function $I(\theta)$, the observed information $I(\hat{\theta})$, and the expected information $J(\theta)$.

(c) Show that

$$\tilde{\theta}_n \rightarrow_p \theta$$

(d) Show that

$$(\tilde{\theta}_n - \theta)[J(\theta)]^{1/2} \rightarrow_D Z \sim \mathrm{N}(0,1)$$

Solution

(a) From Exercise 2.7.9 we have

$$E\left(X^k\right) = \theta^k \Gamma\left(\frac{k}{2} + 1\right) \quad \text{for } k = 1, 2, \ldots \tag{6.9}$$

if $X \sim$ Weibull$(2, \theta)$. Therefore

$$\begin{aligned} E\left(X^2\right) &= \theta^2 \Gamma\left(\frac{2}{2} + 1\right) = \theta^2 \\ E\left(X^4\right) &= \theta^4 \Gamma\left(\frac{4}{2} + 1\right) = 2\theta^4 \end{aligned}$$

and

$$Var\left(X^2\right) = E\left[\left(X^2\right)^2\right] - \left[E\left(X^2\right)\right]^2 = 2\theta^4 - \left(\theta^2\right)^2 = \theta^4$$

(b) The likelihood function is

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{n} f(x_i; \theta) = \prod_{i=1}^{n} \frac{2x_i e^{-(x_i/\theta)^2}}{\theta^2} \\ &= \theta^{-2n} \left(\prod_{i=1}^{n} 2x_i\right) \exp\left(-\frac{1}{\theta^2} \sum_{i=1}^{n} x_i^2\right) \quad \text{for } \theta > 0 \end{aligned}$$

or more simply

$$L(\theta) = \theta^{-2n} \exp\left(-\frac{t}{\theta^2}\right) \quad \text{for } \theta > 0$$

where $t = \sum\limits_{i=1}^{n} x_i^2$. The log likelihood function is

$$l(\theta) = -2n \log \theta - \frac{t}{\theta^2} \quad \text{for } \theta > 0$$

The score function is

$$S(\theta) = \frac{d}{d\theta} l(\theta) = \frac{-2n}{\theta} + \frac{2t}{\theta^3} = \frac{2}{\theta^2}\left(t - n\theta^2\right) \quad \text{for } \theta > 0$$

Now $\frac{d}{d\theta}l(\theta) = 0$ for $\theta = \left(\frac{t}{n}\right)^{1/2}$. Since $\frac{d}{d\theta}l(\theta) > 0$ if $0 < \theta < \left(\frac{t}{n}\right)^{1/2}$ and $\frac{d}{d\theta}l(\theta) < 0$ if $\theta > \left(\frac{t}{n}\right)^{1/2}$ then, by the first derivative test, $l(\theta)$ has a absolute maximum at $\theta = \left(\frac{t}{n}\right)^{1/2}$. Therefore the maximum likelihood estimate of θ is $\hat{\theta} = \left(\frac{t}{n}\right)^{1/2}$. The maximum likelihood estimator is

$$\tilde{\theta} = \left(\frac{T}{n}\right)^{1/2} \quad \text{where} \quad T = \sum_{i=1}^{n} X_i^2$$

The information function is

$$\begin{aligned} I(\theta) &= -\frac{d^2}{d\theta^2}l(\theta) \\ &= \frac{6T}{\theta^4} - \frac{2n}{\theta^2} \quad \text{for } \theta > 0 \end{aligned}$$

and the observed information is

$$\begin{aligned} I(\hat{\theta}) &= \frac{6t}{\hat{\theta}^4} - \frac{2n}{\hat{\theta}^2} = \frac{6(n\hat{\theta}^2)}{\hat{\theta}^4} - \frac{2n}{\hat{\theta}^2} \\ &= \frac{4n}{\hat{\theta}^2} \end{aligned}$$

To find the expected information we note that, from (a), $E\left(X_i^2\right) = \theta^2$ and thus

$$E(T) = E\left(\sum_{i=1}^{n} X_i^2\right) = n\theta^2$$

Therefore the expected information is

$$\begin{aligned} J(\theta) &= E\left(\frac{6T}{\theta^4} - \frac{2n}{\theta^2}\right) = \frac{6E(T)}{\theta^4} - \frac{2n}{\theta^2} = \frac{6n\theta^2}{\theta^4} - \frac{2n}{\theta^2} \\ &= \frac{4n}{\theta^2} \quad \text{for } \theta > 0 \end{aligned}$$

(c) To show that $\tilde{\theta}_n \to_p \theta$ we need to show that

$$\tilde{\theta} = \left(\frac{T}{n}\right)^{1/2} = \left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right)^{1/2} \to_p \theta$$

Since $X_1^2, X_2^2, \ldots, X_n^2$ are independent and identically distributed random variables with $E\left(X_i^2\right) = \theta^2$ and $Var\left(X_i^2\right) = \theta^4$ for $i = 1, 2, \ldots, n$ then by the Weak Law of Large Numbers

$$\frac{T}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \to_p \theta^2$$

and by 5.5.1(1)

$$\tilde{\theta} = \left(\frac{T}{n}\right)^{1/2} \to_p \theta \tag{6.10}$$

as required.

(d) To show that

$$(\tilde{\theta}_n - \theta)[J(\theta)]^{1/2} \to_D Z \sim \mathrm{N}(0,1)$$

we need to show that

$$\begin{aligned} & [J(\theta)]^{1/2}(\tilde{\theta}_n - \theta) \\ = & \frac{2\sqrt{n}}{\theta}\left[\left(\frac{T}{n}\right)^{1/2} - \theta\right] \to_D Z \sim \mathrm{N}(0,1) \end{aligned}$$

Since $X_1^2, X_2^2, \ldots, X_n^2$ are independent and identically distributed random variables with $E\left(X_i^2\right) = \theta^2$ and $Var\left(X_i^2\right) = \theta^4$ for $i = 1, 2, \ldots, n$ then by the Central Limit Theorem

$$\frac{\sqrt{n}\left(\frac{T}{n} - \theta^2\right)}{\theta^2} \to_D Z \sim \mathrm{N}(0,1) \tag{6.11}$$

Let $g(x) = \sqrt{x}$ and $a = \theta^2$. Then $\frac{d}{dx}g(x) = \frac{1}{2\sqrt{x}}$ and $g'(a) = \frac{1}{2\theta}$. By (6.11) and the Delta Method we have

$$\frac{\sqrt{n}\left(\sqrt{\frac{T}{n}} - \sqrt{\theta^2}\right)}{\theta^2} \to_D \frac{1}{2\theta}Z \sim \mathrm{N}\left(0, \frac{1}{4\theta^2}\right)$$

or

$$(2\theta)\frac{\sqrt{n}\left(\sqrt{\frac{T}{n}} - \theta\right)}{\theta^2} \tag{6.12}$$

$$= \frac{2\sqrt{n}}{\theta}\left[\left(\frac{T}{n}\right)^{1/2} - \theta\right] \to_D Z \sim \mathrm{N}(0,1) \tag{6.13}$$

as required.

6.5.3 Example

Suppose $X_1, X_2, \ldots, X_n$ is a random sample from the Uniform$(0,\theta)$ distribution. Since the support set of X_i depends on θ Theorem 6.5.1 does not hold. Show however that the maximum likelihood estimator $\tilde{\theta}_n = X_{(n)}$ is a consistent estimator of θ.

Solution

In Example 5.1.5 we showed that $\tilde{\theta}_n = \max(X_1, X_2, \ldots, X_n) \to_p \theta$ and therefore $\tilde{\theta}_n$ is a consistent estimator of θ.

6.6 Confidence Intervals

In your previous statistics course a confidence interval was used to summarize the available information about an unknown parameter. Confidence intervals allow us to quantify the uncertainty in the unknown parameter.

6.6.1 Definition - Confidence Interval

Suppose $\mathbf{X}$ is a random variable (possibly a vector) whose distribution depends on θ, and $A(\mathbf{X})$ and $B(\mathbf{X})$ are statistics. If

$$P[A(\mathbf{X}) \leq \theta \leq B(\mathbf{X})] = p \quad \text{for } 0 < p < 1$$

then $[a(\mathbf{x}), b(\mathbf{x})]$ is called a $100p\%$ *confidence interval* for θ where $\mathbf{x}$ are the observed data.

Confidence intervals can be constructed in a straightforward manner if a pivotal quantity exists.

6.6.2 Definition - Pivotal Quantity

Suppose $\mathbf{X}$ is a random variable (possibly a vector) whose distribution depends on θ. The random variable $Q(\mathbf{X}; \theta)$ is called a *pivotal quantity* if the distribution of Q does not depend on θ.

Pivotal quantities can be used for constructing confidence intervals in the following way. Since the distribution of $Q(\mathbf{X}; \theta)$ is known we can write down a probability statement of the form

$$P\left(q_1 \leq Q(\mathbf{X}; \theta) \leq q_2\right) = p$$

where q_1 and q_2 do not depend on θ. If Q is a monotone function of θ then this statement can be rewritten as

$$P\left[A(\mathbf{X}) \leq \theta \leq B(\mathbf{X})\right] = p$$

and the interval $[a(\mathbf{x}), b(\mathbf{x})]$ is a $100p\%$ confidence interval.

6.6.3 Example

Suppose $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ is a random sample from the Exponential(θ) distribution. Determine the distribution of

$$Q(\mathbf{X}; \theta) = \frac{2 \sum\limits_{i=1}^{n} X_i}{\theta}$$

and thus show $Q(\mathbf{X}; \theta)$ is a pivotal quantity. Show how $Q(\mathbf{X}; \theta)$ can be used to construct a $100p\%$ equal tail confidence interval for θ.

Solution

Since $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ is a random sample from the Exponential(θ) distribution then by 4.3.2(4)

$$Y = \sum_{i=1}^{n} X_i \sim \text{ Gamma}\,(n, \theta)$$

and by Chapter 2, Problem 7(c)

$$Q(\mathbf{X}; \theta) = \frac{2Y}{\theta} = \frac{2\sum\limits_{i=1}^{n} X_i}{\theta} \sim \chi^2\,(2n)$$

and therefore $Q(\mathbf{X}; \theta)$ is a pivotal quantity.

Find values a and b such that

$$P\,(W \leq a) = \frac{1-p}{2} \quad \text{and} \quad P\,(W \geq b) = \frac{1-p}{2}$$

where $W \sim \chi^2\,(2n)$.

Since $P\,(a \leq W \leq b) = p$ then

$$P\left(a \leq \frac{2\sum\limits_{i=1}^{n} X_i}{\theta} \leq b\right) = p$$

or

$$P\left(\frac{2\sum\limits_{i=1}^{n} X_i}{b} \leq \theta \leq \frac{2\sum\limits_{i=1}^{n} X_i}{a}\right) = p$$

Therefore

$$\left[\frac{2\sum\limits_{i=1}^{n} x_i}{b}, \frac{2\sum\limits_{i=1}^{n} x_i}{a}\right]$$

is a $100p\%$ equal tail confidence interval for θ.

The following theorem gives the pivotal quantity in the case in which θ is either a location or scale parameter.

6.6.4 Theorem

Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ be a random sample from $f\,(x; \theta)$ and let $\tilde{\theta} = \tilde{\theta}(\mathbf{X})$ be the maximum likelihood estimator of the scalar parameter θ based on $\mathbf{X}$.

(1) If θ is a location parameter of the distribution then $Q(\mathbf{X}; \theta) = \tilde{\theta} - \theta$ is a pivotal quantity.

(2) If θ is a scale parameter of the distribution then $Q(\mathbf{X}; \theta) = \tilde{\theta}/\theta$ is a pivotal quantity.

6.6.5 Example

Suppose $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ is a random sample from the Weibull$(2, \theta)$ distribution. Use Theorem 6.6.4 to find a pivotal quantity $Q(\mathbf{X}; \theta)$. Show how the pivotal quantity can be used to construct a $100p\%$ equal tail confidence interval for θ.

Solution

From Chapter 2, Problem 3(a) we know that θ is a scale parameter for the Weibull$(2, \theta)$ distribution. From Example 6.5.2 the maximum likelihood estimator of θ is

$$\tilde{\theta} = \left(\frac{T}{n}\right)^{1/2} = \left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right)^{1/2}$$

Therefore by Theorem 6.6.4

$$\begin{aligned} Q(\mathbf{X}; \theta) &= \frac{\tilde{\theta}}{\theta} \\ &= \frac{1}{\theta}\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right)^{1/2} \end{aligned}$$

is a pivotal quantity. To construct a confidence interval for θ we need to determine the distribution of $Q(\mathbf{X}; \theta)$. This looks difficult at first until we notice that

$$\left(\frac{\tilde{\theta}}{\theta}\right)^2 = \frac{1}{n\theta^2}\sum_{i=1}^{n} X_i^2$$

This form suggests looking for the distribution of $\sum\limits_{i=1}^{n} X_i^2$ which is a sum of independent and identically distributed random variables $X_1^2, X_2^2, \ldots, X_n^2$.

From Chapter 2, Problem 7(h) we have that if $X_i \sim$ Weibull$(2, \theta)$, $i = 1, 2, \ldots, n$ then

$$X_i^2 \sim \text{Exponential}(\theta^2) \quad \text{for } i = 1, 2, \ldots, n$$

Therefore $\sum\limits_{i=1}^{n} X_i^2$ is a sum of independent Exponential(θ^2) random variables. By 4.3.2(4)

$$\sum_{i=1}^{n} X_i^2 \sim \text{Gamma}\left(n, \theta^2\right)$$

and by Chapter 2, Problem 7(h)

$$Q_1(\mathbf{X}; \theta) = \frac{2\sum\limits_{i=1}^{n} X_i^2}{\theta^2} \sim \chi^2\left(2n\right)$$

and therefore $Q_1(\mathbf{X}; \theta)$ is a pivotal quantity.

Now

$$Q_1(\mathbf{X};\theta) = 2n\left(\frac{\tilde{\theta}}{\theta}\right)^2 = 2n\left[Q(\mathbf{X};\theta)\right]^2$$

is a one-to-one function of $Q(\mathbf{X};\theta)$. To see that $Q_1(\mathbf{X};\theta)$ and $Q(\mathbf{X};\theta)$ generate the same confidence intervals for θ we note that

$$\begin{aligned}
& P\left(a \leq Q_1(\mathbf{X};\theta) \leq b\right) \\
= \; & P\left(a \leq 2n\left(\frac{\tilde{\theta}}{\theta}\right)^2 \leq b\right) \\
= \; & P\left[\left(\frac{a}{2n}\right)^{1/2} \leq \frac{\tilde{\theta}}{\theta} \leq \left(\frac{b}{2n}\right)^{1/2}\right] \\
= \; & P\left[\left(\frac{a}{2n}\right)^{1/2} \leq Q(\mathbf{X};\theta) \leq \left(\frac{b}{2n}\right)^{1/2}\right]
\end{aligned}$$

To construct a $100p\%$ equal tail confidence interval we choose a and b such that

$$P\left(W \leq a\right) = \frac{1-p}{2} \;\; \text{and} \;\; P\left(W \geq b\right) = \frac{1-p}{2}$$

where $W \sim \chi^2\left(2n\right)$. Since

$$\begin{aligned}
p &= P\left[\left(\frac{a}{2n}\right)^{1/2} \leq \frac{\tilde{\theta}}{\theta} \leq \left(\frac{b}{2n}\right)^{1/2}\right] \\
&= P\left[\tilde{\theta}\left(\frac{2n}{b}\right)^{1/2} \leq \theta \leq \tilde{\theta}\left(\frac{2n}{a}\right)^{1/2}\right]
\end{aligned}$$

a $100p\%$ equal tail confidence interval is

$$\left[\hat{\theta}\left(\frac{2n}{b}\right)^{1/2}, \; \hat{\theta}\left(\frac{2n}{a}\right)^{1/2}\right]$$

where $\hat{\theta} = \left(\frac{1}{n}\sum\limits_{i=1}^{n} x_i^2\right)^{1/2}$

6.6.6 Example

Suppose $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ is a random sample from the Uniform$(0, \theta)$ distribution. Use Theorem 6.6.4 to find a pivotal quantity $Q(\mathbf{X};\theta)$. Show how the pivotal quantity can be used to construct a $100p\%$ confidence interval for θ of the form $[\hat{\theta}, \, a\hat{\theta}]$.

Solution

From Chapter 2, Problem 3(b) we know that θ is a scale parameter for the Uniform$(0, \theta)$ distribution. From Example 6.2.9 the maximum likelihood estimator of θ is

$$\tilde{\theta} = X_{(n)}$$

Therefore by Theorem 6.6.4

$$Q(\mathbf{X}; \theta) = \frac{\tilde{\theta}}{\theta} = \frac{X_{(n)}}{\theta}$$

is a pivotal quantity. To construct a confidence interval for θ we need to determine the distribution of $Q(\mathbf{X}; \theta)$.

$$\begin{aligned} P\left(Q\left(\mathbf{X}; \theta\right) \leq q\right) &= P\left(\frac{\tilde{\theta}}{\theta} \leq q\right) \\ &= P\left(X_{(n)} \leq q\theta\right) \\ &= \prod_{i=1}^{n} P\left(X_i \leq q\theta\right) \\ &= \prod_{i=1}^{n} q \quad \text{since } P\left(X_i \leq x\right) = \frac{x}{\theta} \quad \text{for } 0 \leq x \leq \theta \\ &= q^n \quad \text{for } 0 \leq q \leq 1 \end{aligned}$$

To construct a $100p\%$ confidence interval for θ of the form $[\hat{\theta}, a\hat{\theta}]$ we need to choose a such that

$$\begin{aligned} p &= P\left(\tilde{\theta} \leq \theta \leq a\tilde{\theta}\right) \\ &= P\left(1 \leq \frac{\theta}{\tilde{\theta}} \leq a\right) = P\left(\frac{1}{a} \leq \frac{\tilde{\theta}}{\theta} \leq 1\right) \\ &= P\left(\frac{1}{a} \leq Q\left(\mathbf{X}; \theta\right) \leq 1\right) \\ &= P\left(Q\left(\mathbf{X}; \theta\right) \leq 1\right) - P\left(Q\left(\mathbf{X}; \theta\right) \leq \frac{1}{a}\right) \\ &= 1 - P\left(Q\left(\mathbf{X}; \theta\right) \leq \frac{1}{a}\right) \\ &= 1 - a^{-n} \end{aligned}$$

or $a = (1 - p)^{-1/n}$. The $100p\%$ confidence interval for θ is

$$\left[\hat{\theta}, (1 - p)^{-1/n}\, \hat{\theta}\right]$$

where $\hat{\theta} = x_{(n)}$.

6.7 Approximate Confidence Intervals

A pivotal quantity does not always exist. For example there is no pivotal quantity for the Binomial or the Poisson distributions. In these cases we use an asymptotic pivotal quantity to construct approximate confidence intervals.

6.7.1 Definition - Asymptotic Pivotal Quantity

Suppose $\mathbf{X}_n = (X_1, X_2, \ldots, X_n)$ is a random variable whose distribution depends on θ. The random variable $Q(\mathbf{X}_n;\theta)$ is called an *asymptotic pivotal quantity* if the limiting distribution of $Q(\mathbf{X}_n;\theta)$ as $n \to \infty$ does not depend on θ.

In your previous statistics course, approximate confidence intervals for the Binomial and Poisson distribution were justified using a Central Limit Theorem argument. We are now able to clearly justify the asymptotic pivotal quantity using the theorems of Chapter 5.

6.7.2 Example

Suppose $\mathbf{X}_n = (X_1, X_2, \ldots, X_n)$ is a random sample from the Poisson(θ) distribution. Show that

$$Q(\mathbf{X}_n;\theta) = \frac{\sqrt{n}\left(\bar{X}_n - \theta\right)}{\sqrt{\bar{X}_n}}$$

is an asymptotic pivotal quantity. Show how $Q(\mathbf{X}_n;\theta)$ can be used to construct an approximate $100p\%$ equal tail confidence interval for θ.

Solution

In Example 5.5.4, the Weak Law of Large Numbers, the Central Limit Theorem and Slutsky's Theorem were all used to prove

$$Q(\mathbf{X}_n;\theta) = \frac{\sqrt{n}\left(\bar{X}_n - \theta\right)}{\sqrt{\bar{X}_n}} \to_D Z \sim \mathrm{N}(0,1) \tag{6.14}$$

and thus $Q(\mathbf{X}_n;\theta)$ is an asymptotic pivotal quantity.

Let a be the value such that $P\left(Z \le a\right) = \left(1+p\right)/2$ where $Z \sim \mathrm{N}(0,1)$. Then by (6.14) we have

$$\begin{aligned} p &\approx P\left(-a \le \frac{\sqrt{n}\left(\bar{X}_n - \theta\right)}{\sqrt{\bar{X}_n}} \le a\right) \\ &= P\left(\bar{X}_n - a\sqrt{\frac{\bar{X}_n}{n}} \le \theta \le \bar{X}_n - a\sqrt{\frac{\bar{X}_n}{n}}\right) \end{aligned}$$

and an approximate $100p\%$ equal tail confidence interval for θ is

$$\left[\bar{x}_n - a\sqrt{\frac{\bar{x}_n}{n}}, \bar{x}_n + a\sqrt{\frac{\bar{x}_n}{n}}\right]$$

or

$$\left[\hat{\theta}_n - a\sqrt{\frac{\hat{\theta}_n}{n}}, \hat{\theta}_n + a\sqrt{\frac{\hat{\theta}_n}{n}}\right] \tag{6.15}$$

since $\hat{\theta}_n = \bar{x}_n$.

6.7.3 Exercise

Suppose $X_n \sim \text{Binomial}(n, \theta)$. Show that

$$Q(X_n; \theta) = \frac{\sqrt{n}\left(\frac{X_n}{n} - \theta\right)}{\sqrt{\frac{X_n}{n}\left(1 - \frac{X_n}{n}\right)}}$$

is an asymptotic pivotal quantity. Show that an approximate $100p\%$ equal tail confidence interval for θ based on $Q(X_n; \theta)$ is given by

$$\left[\hat{\theta}_n - a\sqrt{\frac{\hat{\theta}_n(1-\hat{\theta}_n)}{n}}, \hat{\theta}_n + a\sqrt{\frac{\hat{\theta}_n(1-\hat{\theta}_n)}{n}}\right] \tag{6.16}$$

where $\hat{\theta}_n = \frac{x_n}{n}$.

6.7.4 Approximate Confidence Intervals and the Limiting Distribution of the Maximum Likelihood Estimator

The limiting distribution of the maximum likelihood estimator $\tilde{\theta}_n = \tilde{\theta}_n(X_1, X_2, \ldots, X_n)$ can also be used to construct approximate confidence intervals. This is particularly useful in cases in which the maximum likelihood estimate cannot be found explicitly.

Since

$$\left[J(\tilde{\theta}_n)\right]^{1/2}(\tilde{\theta}_n - \theta) \to_D Z \sim \text{N}(0,1)$$

then $\left[J(\tilde{\theta}_n)\right]^{1/2}(\tilde{\theta}_n - \theta)$ is an asymptotic pivotal quantity. An approximate $100p\%$ confidence interval based on this asymptotic pivotal quantity is given by

$$\hat{\theta}_n \pm a\sqrt{\frac{1}{J(\hat{\theta}_n)}} = \left[\hat{\theta}_n - a\sqrt{\frac{1}{J(\hat{\theta}_n)}},\ \hat{\theta}_n + a\sqrt{\frac{1}{J(\hat{\theta}_n)}}\right] \tag{6.17}$$

where $P(Z \le a) = \frac{1+p}{2}$ and $Z \sim \text{N}(0,1)$.

Similarly since

$$[I(\tilde{\theta}_n; \mathbf{X})]^{1/2}(\tilde{\theta}_n - \theta) \to_D Z \sim \text{N}(0,1)$$

then $[I(\tilde{\theta}_n; \mathbf{X})]^{1/2}(\tilde{\theta}_n - \theta)$ is an asymptotic pivotal quantity. An approximate $100p\%$ confidence interval based on this asymptotic pivotal quantity is given by

$$\hat{\theta}_n \pm a\sqrt{\frac{1}{I(\hat{\theta}_n)}} = \left[\hat{\theta}_n - a\sqrt{\frac{1}{I(\hat{\theta}_n)}},\ \hat{\theta}_n + a\sqrt{\frac{1}{I(\hat{\theta}_n)}}\right] \tag{6.18}$$

where $I(\hat{\theta}_n)$ is the observed information.

Notes:

(1) One drawback of these intervals is that we don't know how large n needs to be to obtain a good approximation.

(2) These approximate confidence intervals are both symmetric about $\hat{\theta}_n$ which may not be a reasonable summary of the plausible values of θ in light of the observed data. See likelihood intervals below.

(3) It is possible to obtain approximate confidence intervals based on a given data set which contain values which are not valid for θ. For example, an interval may contain negative values although θ must be positive.

6.7.5 Example

Use the results from Example 6.3.5 to determine the approximate $100p\%$ confidence intervals based on (6.17) and (6.18) in the case of Binomial data and Poisson data. Compare these intervals with the intervals in (6.16) and (6.15).

Solution

From Example 6.3.5F we have that for Binomial data

$$I(\hat{\theta}_n) = J(\hat{\theta}_n) = \frac{n}{\hat{\theta}_n(1-\hat{\theta}_n)}$$

so (6.17) and (6.18) both give the intervals

$$\left[\hat{\theta}_n - a\sqrt{\frac{\hat{\theta}_n(1-\hat{\theta}_n)}{n}}, \hat{\theta}_n + a\sqrt{\frac{\hat{\theta}_n(1-\hat{\theta}_n)}{n}}\right]$$

which is the same interval as in (6.16).

From Example 6.3.5F we have that for Poisson data

$$I(\hat{\theta}_n) = J(\hat{\theta}_n) = \frac{n}{\hat{\theta}_n}$$

so (6.17) and (6.18) both give the intervals

$$\left[\hat{\theta}_n - a\sqrt{\frac{\hat{\theta}_n}{n}}, \hat{\theta}_n + a\sqrt{\frac{\hat{\theta}_n}{n}}\right]$$

which is the same interval as in (6.15).

6.7.6 Example

For Example 6.3.7 construct an approximate 95% confidence interval based on (6.18). Compare this with the 15% likelihood interval determined in Example 6.4.4.

Solution

From Example 6.3.7 we have $\hat{\theta} = 0.4951605$ and $I(\hat{\theta}) = 181.8069$. Therefore an approximate 95% confidence interval based on (6.18) is

$$\begin{aligned}
& \hat{\theta} \pm 1.96\sqrt{\frac{1}{I(\hat{\theta})}} \\
= \; & 0.4951605 \pm 1.96\sqrt{\frac{1}{181.8069}} \\
= \; & 0.4951605 \pm 0.145362 \\
= \; & [0.3498, 0.6405]
\end{aligned}$$

From Example 6.4.4 the 15% likelihood interval is $[0.3550, 0.6401]$ which is very similar. We expect this to happen since the relative likelihood function in Figure 6.3 is very symmetric.

6.7.7 Approximate Confidence Intervals and Likelihood Intervals

In your previous statistics course you learned that likelihood intervals are also approximate confidence intervals.

6.7.8 Theorem

If a is a value such that $p = 2P(Z \leq a) - 1$ where $Z \sim N(0,1)$, then the likelihood interval $\left\{\theta : R(\theta) \geq e^{-a^2/2}\right\}$ is an approximate $100p\%$ confidence interval.

Proof

By Theorem 6.5.1

$$-2\log R(\theta;\mathbf{X}_n) = 2\log\left[\frac{L(\theta;\mathbf{X}_n)}{L(\tilde{\theta}_n;\mathbf{X}_n)}\right] \to_D W \sim \chi^2(1) \tag{6.19}$$

where $\mathbf{X}_n = (X_1, X_2, \ldots, X_n)$. The confidence coefficient corresponding to the interval $\left\{\theta : R(\theta) \geq e^{-a^2/2}\right\}$ is

$$\begin{aligned}
P\left[\frac{L(\theta;\mathbf{X}_n)}{L(\tilde{\theta}_n;\mathbf{X}_n)} \geq e^{-a^2/2}\right] &= P\left[-2\log R(\theta;\mathbf{X}_n) \leq a^2\right] \\
&\approx P\left(W \leq a^2\right) \quad \text{where} \quad W \sim \chi^2(1) \quad \text{by 6.19} \\
&= 2P(Z \leq a) - 1 \quad \text{where} \quad Z \sim N(0,1) \\
&= p
\end{aligned}$$

as required.

6.7.9 Example

Since

$$0.95 = 2P(Z \leq 1.96) - 1 \quad \text{where} \quad Z \sim N(0,1)$$

and

$$e^{-(1.96)^2/2} = e^{-1.9208} \approx 0.1465 \approx 0.15$$

therefore a 15% likelihood interval for θ is also an approximate 95% confidence interval for θ.

6.7.10 Exercise

(a) Show that a 1% likelihood interval is an approximate 99.8% confidence interval.

(b) Show that a 50% likelihood interval is an approximate 76% confidence interval.

Note that while the confidence intervals given by (6.17) or (6.18) are symmetric about the point estimate $\hat{\theta}_n$, this is not true in general for likelihood intervals.

6.7.11 Example

For Example 6.3.7 compare a 15% likelihood interval with the approximate 95% confidence interval in Example 6.7.6.

Solution

From Example 6.3.7 the 15% likelihood interval is

$$[0.3550, 0.6401]$$

and from Example 6.7.6 the approximate 95% confidence interval is

$$[0.3498, 0.6405]$$

These intervals are very close and agree to 2 decimal places. The reason for this is because the likelihood function (see Figure 6.3) is very symmetric about the maximum likelihood estimate. The approximate intervals (6.17) or (6.18) will be close to the corresponding likelihood interval whenever the likelihood function is reasonably symmetric about the maximum likelihood estimate.

6.7.12 Exercise

Suppose $x_1, x_2, \ldots, x_n$ is an observed random sample from the Logistic$(\theta, 1)$ distribution with probability density function

$$f(x;\theta) = \frac{e^{-(x-\theta)}}{\left[1 + e^{-(x-\theta)}\right]^2} \quad \text{for } x \in \Re, \ \theta \in \Re$$

(a) Find the likelihood function, the score function, and the information function. How would you find the maximum likelihood estimate of θ?

(b) Show that if u is an observation from the Uniform$(0,1)$ distribution then

$$x = \theta - \log\left(\frac{1}{u} - 1\right)$$

is an observation from the Logistic$(\theta, 1)$ distribution.

(c) Use the following R code to randomly generate 30 observations from a Logistic$(\theta, 1)$ distribution.

```
# randomly generate 30 observations from a Logistic(theta,1)
# using a random theta value between 2 and 3
set.seed(21086689) # set the seed so results can be reproduced
truetheta<-runif(1,min=2,max=3)
# data are sorted and rounded to two decimal places for easier display
x<-sort(round((truetheta-log(1/runif(30)-1)),2))
x
```

(d) Use R to plot the likelihood function for θ based on these data.

(e) Use Newton's Method and R to find $\hat{\theta}$.

(f) What are the values of $S(\hat{\theta})$ and $I(\hat{\theta})$?

(g) Use R to plot the relative likelihood function for θ based on these data.

(h) Compare the 15% likelihood interval with the approximate 95% confidence interval (6.18).

6.8 Chapter 6 Problems

1. Suppose $X \sim \text{Binomial}(n, \theta)$. Plot the log relative likelihood function for θ if $x = 3$ is observed for $n = 100$. On the same graph plot the log relative likelihood function for θ if $x = 6$ is observed for $n = 200$. Compare the graphs as well as the 10% likelihood interval and 50% likelihood interval for θ.

2. Suppose $x_1, x_2, \ldots, x_n$ is an observed random sample from the Discrete Uniform$(1, \theta)$ distribution. Find the likelihood function, the maximum likelihood estimate of θ and the maximum likelihood estimator of θ. If $n = 20$ and $x_{20} = 33$, find a 15% likelihood interval for θ.

3. Suppose $x_1, x_2, \ldots, x_n$ is an observed random sample from the Geometric(θ) distribution.

   (a) Find the score function and the maximum likelihood estimator of θ.

   (b) Find the observed information and the expected information.

   (c) Find the maximum likelihood estimator of $E(X_i)$.

   (d) If $n = 20$ and $\sum\limits_{i=1}^{20} x_i = 40$ then find the maximum likelihood estimate of θ and a 15% likelihood interval for θ. Is $\theta = 0.5$ a plausible value of θ? Why?

4. Suppose $(X_1, X_2, X_3) \sim \text{Multinomial}(n, \theta^2, 2\theta(1-\theta), (1-\theta)^2)$. Find the maximum likelihood estimator of θ, the observed information and the expected information.

5. Suppose $x_1, x_2, \ldots, x_n$ is an observed random sample from the Pareto$(1, \theta)$ distribution.

   (a) Find the score function and the maximum likelihood estimator of θ.

   (b) Find the observed information and the expected information.

   (c) Find the maximum likelihood estimator of $E(X_i)$.

   (d) If $n = 20$ and $\sum\limits_{i=1}^{20} \log x_i = 10$ find the maximum likelihood estimate of θ and a 15% likelihood interval for θ. Is $\theta = 0.1$ a plausible value of θ? Why?

   (e) Show that
$$Q(\mathbf{X}; \theta) = 2\theta \sum_{i=1}^{n} \log X_i$$
   is a pivotal quantity. (Hint: What is the distribution of $\log X$ if $X \sim \text{Pareto}(1, \theta)$?) Use this pivotal quantity to determine a 95% equal tail confidence interval for the data in (d). Compare this interval with the 15% likelihood interval.

6. The following model is proposed for the distribution of family size in a large population:

$$\begin{aligned} P\,(k \text{ children in family; } \theta) &= \theta^k \quad \text{for } k = 1, 2, \ldots \\ P\,(0 \text{ children in family; } \theta) &= \frac{1 - 2\theta}{1 - \theta} \end{aligned}$$

The parameter θ is unknown and $0 < \theta < \frac{1}{2}$. Fifty families were chosen at random from the population. The observed numbers of children are given in the following table:

No. of children	0	1	2	3	4	Total
Frequency observed	17	22	7	3	1	50

(a) Find the likelihood, log likelihood, score and information functions for θ.

(b) Find the maximum likelihood estimate of θ and the observed information.

(c) Find a 15% likelihood interval for θ.

(d) A large study done 20 years earlier indicated that $\theta = 0.45$. Is this value plausible for these data?

(e) Calculate estimated expected frequencies. Does the model give a reasonable fit to the data?

7. Suppose $x_1, x_2, \ldots, x_n$ is a observed random sample from the Two Parameter Exponential$(\theta, 1)$ distribution. Show that θ is a location parameter and $\tilde{\theta} = X_{(1)}$ is the maximum likelihood estimator of θ. Show that

$$P(\tilde{\theta} - \theta \leq q) = 1 - e^{-nq} \quad \text{for } q \geq 0$$

and thus show that

$$\left[\hat{\theta} + \frac{1}{n}\log(1-p)\,,\ \hat{\theta}\right]$$

and

$$\left[\hat{\theta} + \frac{1}{n}\log\left(\frac{1-p}{2}\right),\ \hat{\theta} + \frac{1}{n}\log\left(\frac{1+p}{2}\right)\right]$$

are both $100p\%$ confidence intervals for θ. Which confidence interval seems more reasonable?

8. Suppose $X_1, X_2, \ldots, X_n$ is a random sample from the Gamma$\left(\frac{1}{2}, \frac{1}{\theta}\right)$ distribution.

(a) Find $\tilde{\theta}_n$, the maximum likelihood estimator of θ.

(b) Justify the statement $\tilde{\theta}_n \to_p \theta$.

(c) Find the maximum likelihood estimator of $Var(X_i)$.

(d) Use moment generating functions to show that $Q = 2\theta \sum_{i=1}^{n} X_i \sim \chi^2(n)$. If $n = 20$ and $\sum_{i=1}^{20} x_i = 6$, use the pivotal quantity Q to construct an exact 95% equal tail confidence interval for θ. Is $\theta = 0.7$ a plausible value of θ?

(e) Verify that $(\tilde{\theta}_n - \theta)\left[J(\tilde{\theta}_n)\right]^{1/2} \to_D Z \sim \mathrm{N}(0,1)$. Use this asymptotic pivotal quantity to construct an approximate 95% confidence interval for θ. Compare this interval with the exact confidence interval from (d) and a 15% likelihood interval for θ. What do the approximate confidence interval and the likelihood interval indicate about the plausibility of the value $\theta = 0.7$?

9. The number of coliform bacteria X in a 10 cubic centimeter sample of water from a section of lake near a beach has a Poisson(μ) distribution.

(a) If a random sample of n specimen samples is taken and $X_1, X_2, \ldots, X_n$ are the respective numbers of observed bacteria, find the likelihood function, the maximum likelihood estimator and the expected information for μ.

(b) If $n = 20$ and $\sum_{i=1}^{20} x_i = 40$, obtain an approximate 95% confidence interval for μ and a 15% likelihood interval for μ. Compare the intervals.

(c) Suppose there is a fast, simple test which can detect whether there are bacteria present, but not the exact number. If Y is the number of samples out of n which have bacteria, show that

$$P(Y = y) = \binom{n}{y}(1 - e^{-\mu})^y (e^{-\mu})^{n-y} \quad \text{for } y = 0, 1, \ldots, n$$

(d) If $n = 20$ and we found $y = 17$ of the samples contained bacteria, use the likelihood function from part (c) to get an approximate 95% confidence interval for μ. Hint: Let $\theta = 1 - e^{-\mu}$ and use the likelihood function for θ to get an approximate confidence interval for θ and then transform this to an approximate confidence interval for $\mu = -\log(1 - \theta)$.

7. Maximum Likelihood Estimation - Multiparameter

In this chapter we look at the method of maximum likelihood to obtain both point and interval estimates for the case in which the unknown parameter is a vector of unknown parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$. In your previous statistics course you would have seen the $\mathrm{N}(\mu, \sigma^2)$ model with two unknown parameters $\boldsymbol{\theta} = (\mu, \sigma^2)$ and the simple linear regression model $\mathrm{N}(\alpha + \beta x, \sigma^2)$ with three unknown parameters $\boldsymbol{\theta} = (\alpha, \beta, \sigma^2)$.

Although the case of k parameters is a natural extension of the one parameter case, the k parameter case is more challenging. For example, the maximum likelihood estimates are usually found be solving k nonlinear equations in k unknowns $\theta_1, \theta_2, \ldots, \theta_k$. In most cases there are no explicit solutions and the maximum likelihood estimates must be found using a numerical method such as Newton's Method. Another challenging issue is how to summarize the uncertainty in the k estimates. For one parameter it is straightforward to summarize the uncertainty using a likelihood interval or a confidence interval. For k parameters these intervals become regions in $\Re^k$ which are difficult to visualize and interpret.

In Section 7.1 we give all the definitions related to finding the maximum likelihood estimates for k unknown parameters. These definitions are analogous to the definitions which were given in Chapter 6 for one unknown parameter. We also give the extension of the invariance property of maximum likelihood estimates and Newton's Method for k variables. In Section 7.2 we define likelihood regions and show how to find them for the case $k = 2$.

In Section 7.3 we introduce the Multivariate Normal distribution which is the natural extension of the Bivariate Normal distribution discussed in Section 3.10. We also give the limiting distribution of the maximum likelihood estimator of $\boldsymbol{\theta}$ which is a natural extension of Theorem 6.5.1.

In Section 7.4 we show how to obtain approximate confidence regions for $\boldsymbol{\theta}$ based on the limiting distribution of the maximum likelihood estimator of $\boldsymbol{\theta}$ and show how to find them for the case $k = 2$. We also show how to find approximate confidence intervals for individual parameters and indicate that these intervals must be used with care.

7.1 Likelihood and Related Functions

7.1.1 Definition - Likelihood Function

Suppose $\mathbf{X}$ is a (vector) random variable with probability (density) function $f(x;\boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k) \in \Omega$ and Ω is the parameter space or set of possible values of $\boldsymbol{\theta}$. Suppose also that $\mathbf{x}$ is an observed value of $\mathbf{X}$. The *likelihood function* for $\boldsymbol{\theta}$ based on the observed data $\mathbf{x}$ is

$$L(\boldsymbol{\theta}) = L(\boldsymbol{\theta};\mathbf{x}) = f(\mathbf{x};\boldsymbol{\theta}) \quad \text{for } \boldsymbol{\theta} \in \Omega \tag{7.1}$$

If $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ is a random sample from a distribution with probability function $f(x;\boldsymbol{\theta})$ and $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ are the observed data then the likelihood function for θ based on the observed data $x_1, x_2, \ldots, x_n$ is

$$\begin{aligned} L(\boldsymbol{\theta}) &= L(\boldsymbol{\theta};\mathbf{x}) \\ &= \prod_{i=1}^{n} f(x_i;\boldsymbol{\theta}) \quad \text{for } \boldsymbol{\theta} \in \Omega \end{aligned}$$

Note: If $\mathbf{X}$ is a discrete random variable then $L(\boldsymbol{\theta}) = P\,(\text{observing the data } \mathbf{x};\boldsymbol{\theta})$. If $\mathbf{X}$ is a continuous random variable then an argument similar to the one in 6.2.6 can be made to justify the use of (7.1).

7.1.2 Definition - Maximum Likelihood Estimate and Estimator

The value of $\boldsymbol{\theta}$ that maximizes the likelihood function $L(\boldsymbol{\theta})$ is called the *maximum likelihood estimate*. The maximum likelihood estimate is a function of the observed data $\mathbf{x}$ and we write $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}\,(\mathbf{x})$. The corresponding *maximum likelihood estimator* which is a random vector is denoted by $\tilde{\boldsymbol{\theta}} = \tilde{\boldsymbol{\theta}}(\mathbf{X})$.

As in the case of $k = 1$ it is frequently easier to work with the log likelihood function which is maximized at the same value of $\boldsymbol{\theta}$ as the likelihood function.

7.1.3 Definition - Log Likelihood Function

The *log likelihood function* is defined as

$$l(\boldsymbol{\theta}) = l\,(\boldsymbol{\theta};\mathbf{x}) = \log L(\boldsymbol{\theta}) \quad \text{for } \boldsymbol{\theta} \in \Omega$$

where $\mathbf{x}$ are the observed data and log is the natural logarithmic function.

The maximum likelihood estimate of $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k)$ is usually found by solving $\frac{\partial l(\boldsymbol{\theta})}{\partial \theta_j} = 0$, $j = 1, 2, \ldots, k$ simultaneously. See Chapter 7, Problem 1 for an example in which the maximum likelihood estimate is not found in this way.

7.1.4 Definition - Score Vector

If $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$ then the *score vector (function)* is a $1 \times k$ vector of functions defined as

$$S(\boldsymbol{\theta}) = S(\boldsymbol{\theta}; \mathbf{x}) = \left[\frac{\partial l(\boldsymbol{\theta})}{\partial \theta_1}, \frac{\partial l(\boldsymbol{\theta})}{\partial \theta_2}, \ldots, \frac{\partial l(\boldsymbol{\theta})}{\partial \theta_k}\right] \quad \text{for } \boldsymbol{\theta} \in \Omega$$

where $\mathbf{x}$ are the observed data.

We will see that, as in the case of one parameter, the information matrice will provides information about the variance of the maximum likelihood estimator.

7.1.5 Definition - Information Matrix

If $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$ then the *information matrix (function)* $I(\boldsymbol{\theta}) = I(\boldsymbol{\theta}; \mathbf{x})$ is a $k \times k$ symmetric matrix of functions whose (i, j) entry is given by

$$-\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \quad \text{for } \boldsymbol{\theta} \in \Omega$$

where $\mathbf{x}$ are the observed data. $I(\hat{\boldsymbol{\theta}})$ is called the *observed information matrix*.

7.1.6 Definition - Expected Information Matrix

If $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$ then the *expected information matrix (function)* $J(\boldsymbol{\theta})$ is a $k \times k$ symmetric matrix of functions whose (i, j) entry is given by

$$E\left[-\frac{\partial^2 l(\boldsymbol{\theta}; \mathbf{X})}{\partial \theta_i \partial \theta_j}\right] \quad \text{for } \boldsymbol{\theta} \in \Omega$$

The invariance property of the maximum likelihood estimator also holds in the multiparameter case.

7.1.7 Theorem - Invariance of the Maximum Likelihood Estimate

If $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k)$ is the maximum likelihood estimate of $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$ then $g(\hat{\boldsymbol{\theta}})$ is the maximum likelihood estimate of $g(\boldsymbol{\theta})$.

7.1.8 Example

Suppose $x_1, x_2, \ldots, x_n$ is a observed random sample from the $\mathrm{N}(\mu, \sigma^2)$ distribution. Find the score vector, the information matrix, the expected information matrix and the maximum likelihood estimator of $\boldsymbol{\theta} = (\mu, \sigma^2)$. Find the observed information matrix $I(\hat{\mu}, \hat{\sigma}^2)$ and thus verify that $(\hat{\mu}, \hat{\sigma}^2)$ is the maximum likelihood estimator of (μ, σ^2). What is the maximum likelihood estimator of the parameter $\tau = \tau(\mu, \sigma^2) = \mu/\sigma$ which is called the coefficient of variation?

Solution

The likelihood function is

$$\begin{aligned} L\left(\mu,\sigma^2\right) &= \prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}\exp\left[\frac{-1}{2\sigma^2}\left(x_i-\mu\right)^2\right] \\ &= \left(2\pi\right)^{-n/2}\left(\sigma^2\right)^{-n/2}\exp\left[\frac{-1}{2\sigma^2}\sum_{i=1}^{n}\left(x_i-\mu\right)^2\right] \quad \text{for } \mu\in\Re,\ \sigma^2>0 \end{aligned}$$

or more simply

$$L\left(\mu,\sigma^2\right) = \left(\sigma^2\right)^{-n/2}\exp\left[\frac{-1}{2\sigma^2}\sum_{i=1}^{n}\left(x_i-\mu\right)^2\right] \quad \text{for } \mu\in\Re,\ \sigma^2>0$$

The log likelihood function is

$$\begin{aligned} l\left(\mu,\sigma^2\right) &= -\frac{n}{2}\log\left(\sigma^2\right)-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(x_i-\mu\right)^2 \\ &= -\frac{n}{2}\log\left(\sigma^2\right)-\frac{1}{2}\left(\sigma^2\right)^{-1}\left[\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2+n\left(\bar{x}-\mu\right)^2\right] \\ &= -\frac{n}{2}\log\left(\sigma^2\right)-\frac{1}{2}\left(\sigma^2\right)^{-1}\left[\left(n-1\right)s^2+n\left(\bar{x}-\mu\right)^2\right] \quad \text{for } \mu\in\Re,\ \sigma^2>0 \end{aligned}$$

where

$$s^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2$$

Now

$$\frac{\partial l}{\partial \mu}=\frac{n}{\sigma^2}\left(\bar{x}-\mu\right)=n\left(\sigma^2\right)^{-1}\left(\bar{x}-\mu\right)$$

and

$$\frac{\partial l}{\partial \sigma^2}=-\frac{n}{2}\left(\sigma^2\right)^{-1}+\frac{1}{2}\left(\sigma^2\right)^{-2}\left[\left(n-1\right)s^2+n\left(\bar{x}-\mu\right)^2\right]$$

The equations

$$\frac{\partial l}{\partial \mu}=0, \qquad \frac{\partial l}{\partial \sigma^2}=0$$

are solved simultaneously for

$$\mu=\bar{x} \quad \text{and} \quad \sigma^2=\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2=\frac{(n-1)}{n}s^2$$

Since

$$\begin{aligned} -\frac{\partial^2 l}{\partial \mu^2} &= \frac{n}{\sigma^2}, \qquad -\frac{\partial^2 l}{\partial \sigma^2 \partial \mu}=\frac{n\left(\bar{x}-\mu\right)}{\sigma^4} \\ -\frac{\partial^2 l}{\partial\left(\sigma^2\right)^2} &= -\frac{n}{2}\frac{1}{\sigma^4}+\frac{1}{\sigma^6}\left[\left(n-1\right)s^2+n\left(\bar{x}-\mu\right)^2\right] \end{aligned}$$

the information matrix is

$$I\left(\mu,\sigma^2\right)=\begin{bmatrix} \frac{n}{\sigma^2} & \frac{n(\bar{x}-\mu)}{\sigma^4} \\ \frac{n(\bar{x}-\mu)}{\sigma^4} & -\frac{n}{2}\frac{1}{\sigma^4}+\frac{1}{\sigma^6}\left[(n-1)\,s^2+n\left(\bar{x}-\mu\right)^2\right] \end{bmatrix}$$

Since

$$I_{11}\left(\hat{\mu},\hat{\sigma}^2\right)=\frac{n}{\hat{\sigma}^2}>0 \quad\text{and}\quad \det I\left(\hat{\mu},\hat{\sigma}^2\right)=\frac{n^2}{2\hat{\sigma}^6}>0$$

then by the second derivative test the maximum likelihood estimates of of μ and σ^2 are

$$\hat{\mu}=\bar{x} \quad\text{and}\quad \hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2=\frac{(n-1)}{n}s^2$$

and the maximum likelihood estimators are

$$\tilde{\mu}=\bar{X} \quad\text{and}\quad \tilde{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2=\frac{(n-1)}{n}S^2$$

The observed information is

$$I\left(\hat{\mu},\hat{\sigma}^2\right)=\begin{bmatrix} \frac{n}{\hat{\sigma}^2} & 0 \\ 0 & \frac{n}{2\hat{\sigma}^4} \end{bmatrix}$$

Now

$$E\left(\frac{n}{\sigma^2}\right)=\frac{n}{\sigma^2}, \qquad E\left[\frac{n\left(\bar{X}-\mu\right)}{\sigma^4}\right]=0$$

Also

$$\begin{aligned}
&E\left\{-\frac{n}{2}\frac{1}{\sigma^4}+\frac{1}{\sigma^6}\left[(n-1)\,S^2+n\left(\bar{X}-\mu\right)^2\right]\right\}\\
=\;& -\frac{n}{2}\frac{1}{\sigma^4}+\frac{1}{\sigma^6}\left\{(n-1)\,E(S^2)+nE\left[\left(\bar{X}-\mu\right)^2\right]\right\}\\
=\;& -\frac{n}{2}\frac{1}{\sigma^4}+\frac{1}{\sigma^6}\left\{(n-1)\,\sigma^2+\sigma^2\right\}\\
=\;& \frac{n}{2\sigma^4}
\end{aligned}$$

since

$$E\left(\bar{X}-\mu\right)=0,\; E\left[\left(\bar{X}-\mu\right)^2\right]=Var\left(\bar{X}\right)=\frac{\sigma^2}{n} \quad\text{and}\quad E\left(S^2\right)=\sigma^2$$

Therefore the expected information matrix is

$$J\left(\mu,\sigma^2\right)=\begin{bmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{bmatrix}$$

and the inverse of the expected information matrix is

$$\left[J\left(\mu,\sigma^2\right)\right]^{-1}=\begin{bmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{2\sigma^4}{n} \end{bmatrix}$$

Note that

$$\begin{aligned} Var\left(\bar{X}\right) &= \frac{\sigma^2}{n} \\ Var\left(\hat{\sigma}^2\right) &= Var\left[\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2\right] = \frac{2(n-1)\sigma^4}{n^2} \approx \frac{2\sigma^4}{n} \end{aligned}$$

and

$$Cov(\bar{X}, \hat{\sigma}^2) = \frac{1}{n} Cov(\bar{X}, \ \sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2) = 0$$

since $\bar{X}$ and $\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$ are independent random variables.

By the invariance property of maximum likelihood estimators the maximum likelihood estimator of $\tau = \mu/\sigma$ is $\hat{\tau} = \hat{\mu}/\hat{\sigma}$.

Recall from your previous statistics course that inferences for μ and σ^2 are made using the independent pivotal quantities

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim \ t\left(n-1\right) \quad \text{and} \quad \frac{(n-1)S^2}{\sigma^2} \sim \chi^2\left(n-1\right)$$

See Figure 7.1 for a graph of $R\left(\mu, \sigma^2\right)$ for $n = 50$, $\hat{\mu} = 5$ and $\hat{\sigma}^2 = 4$.

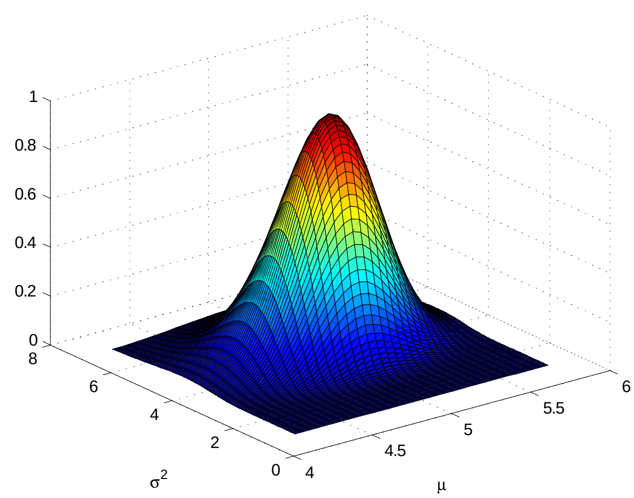


Figure 7.1: Normal Relative Likelihood Function for $n = 50$, $\hat{\mu} = 5$ and $\hat{\sigma}^2 = 4$

7.1.9 Exercise

Suppose $Y_i \sim \mathrm{N}(\alpha+\beta x_i, \sigma^2)$, $i = 1, 2, \ldots, n$ independently where the x_i are known constants. Show that the maximum likelihood estimators of α, β and σ^2 are given by

$$\begin{aligned}
\tilde{\alpha} &= \bar{Y} - \hat{\beta}\bar{x} \\
\tilde{\beta} &= \frac{\sum\limits_{i=1}^{n}(x_i - \bar{x})\left(Y_i - \bar{Y}\right)}{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2} \\
\tilde{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n}(Y_i - \tilde{\alpha} - \tilde{\beta}x_i)^2
\end{aligned}$$

Note: $\tilde{\alpha}$ and $\tilde{\beta}$ are also the least squares estimators of α and β.

7.1.10 Example

Suppose $x_1, x_2, \ldots, x_n$ is an observed random sample from the Beta(a, b) distribution with probability density function

$$f(x; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1} \quad \text{for } 0 < x < 1,\ a > 0,\ b > 0$$

Find the likelihood function, the score vector, and the information matrix and the expected information matrix. How would you find the maximum likelihood estimates of a and b?

Solution

The likelihood function is

$$\begin{aligned}
L(a, b) &= \prod_{i=1}^{n} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x_i^{a-1}(1-x_i)^{b-1} \quad \text{for } a > 0,\ b > 0 \\
&= \left[\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\right]^n \left[\prod_{i=1}^{n} x_i\right]^{a-1} \left[\prod_{i=1}^{n}(1-x_i)\right]^{b-1}
\end{aligned}$$

or more simply

$$L(a, b) = \left[\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\right]^n \left[\prod_{i=1}^{n} x_i\right]^{a} \left[\prod_{i=1}^{n}(1-x_i)\right]^{b} \quad \text{for } a > 0,\ b > 0$$

The log likelihood function is

$$l(a, b) = n\left[\log\Gamma(a+b) - \log\Gamma(a) - \log\Gamma(b) + at_1 + bt_2\right] \quad \text{for } a > 0,\ b > 0$$

where

$$t_1 = \frac{1}{n}\sum_{i=1}^{n}\log x_i \quad \text{and} \quad t_2 = \frac{1}{n}\sum_{i=1}^{n}\log(1-x_i)$$

Let

$$\Psi(z) = \frac{d \log \Gamma(z)}{dz} = \frac{\Gamma'(z)}{\Gamma(z)}$$

which is called the digamma function.

The score vector is

$$\begin{aligned} S(a,b) &= \begin{bmatrix} \frac{\partial l}{\partial a} & \frac{\partial l}{\partial b} \end{bmatrix} \\ &= n \begin{bmatrix} \Psi(a+b) - \Psi(a) + t_1 & \Psi(a+b) - \Psi(b) + t_2 \end{bmatrix} \end{aligned}$$

for $a > 0$, $b > 0$. $S(a,b) = (0,0)$ must be solved numerically to find the maximum likelihood estimates of a and b.
Let

$$\Psi'(z) = \frac{d}{dz}\Psi(z)$$

which is called the trigamma function.

The information matrix is

$$I(a,b) = n \begin{bmatrix} \Psi'(a) - \Psi'(a+b) & -\Psi'(a+b) \\ -\Psi'(a+b) & \Psi'(b) - \Psi'(a+b) \end{bmatrix}$$

for $a > 0$, $b > 0$, which is also expected information matrix.

7.1.11 Exercise

Suppose $x_1, x_2, \ldots, x_n$ is an observed random sample from the Gamma(α, β) distribution. Find the likelihood function, the score vector, and the information matrix and the expected information matrix. How would you find the maximum likelihood estimates of α and β?

Often $S(\theta_1, \theta_2, \ldots, \theta_k) = (0, 0, \ldots, 0)$ must be solved numerically using a method such as Newton's Method.

7.1.12 Newton's Method

Let $\boldsymbol{\theta}^{(0)}$ be an initial estimate of $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$. The estimate $\boldsymbol{\theta}^{(i)}$ can be updated using

$$\boldsymbol{\theta}^{(i+1)} = \boldsymbol{\theta}^{(i)} + \left[S(\boldsymbol{\theta}^{(i)})\right]\left[I(\boldsymbol{\theta}^{(i)})\right]^{-1} \quad \text{for } i = 0, 1, \ldots$$

Note: The initial estimate, $\boldsymbol{\theta}^{(0)}$, may be determined by calculating $L(\boldsymbol{\theta})$ for a grid of values to determine the region in which $L(\boldsymbol{\theta})$ obtains a maximum.

7.1.13 Example

Use the following R code to randomly generate 35 observations from a Beta(a, b) distribution

```
# randomly generate 35 observations from a Beta(a,b)
set.seed(32086689)    # set the seed so results can be reproduced
# use randomly generated a and b values
truea<-runif(1,min=2,max=3)
trueb<-runif(1,min=1,max=4)
# data are sorted and rounded to two decimal places for easier display
x<-sort(round(rbeta(35,truea,trueb),2))
x
```

Use Newton's Method and R to find $(\hat{a}, \hat{b})$.

What are the values of $S(\hat{a}, \hat{b})$ and $I(\hat{a}, \hat{b})$?

Solution

The generated data are

0.08	0.19	0.21	0.25	0.28	0.29	0.29	0.30	0.30	0.32
0.34	0.36	0.39	0.45	0.45	0.47	0.48	0.49	0.54	0.54
0.55	0.55	0.56	0.56	0.61	0.63	0.64	0.65	0.69	0.69
0.73	0.77	0.79	0.81	0.85					

The maximum likelihood estimates of a and b can be found using Newton's Method given by

$$\begin{bmatrix} a^{(i+1)} \\ b^{(i+1)} \end{bmatrix} = \begin{bmatrix} a^{(i)} \\ b^{(i)} \end{bmatrix} + S(a^{(i)}, b^{(i)}) \left[I(a^{(i)}, b^{(i)}) \right]^{-1}$$

for $i = 0, 1, ...$ until convergence.

Here is R code for Newton's Method for the Beta Example.

```
# function for calculating Beta score for a and b and data x
BESF<-function(a,b,x)
{S<-length(x)*c(digamma(a+b)-digamma(a)+mean(log(x)),
    digamma(a+b)-digamma(b)+mean(log(1-x)))
return(S)}
#)
# function for calculating Beta information for a and b)
BEIF<-function(a,b)
{I<-length(x)*cbind(c(trigamma(a)-trigamma(a+b),-trigamma(a+b)),
    c(-trigamma(a+b),trigamma(b)-trigamma(a+b)))
return(I)}
```

```
# Newton's Method for Beta Example
NewtonBE<-function(a,b,x)
{thold<-c(a,b)
thnew<-thold+0.1
while (sum(abs(thold-thnew))>0.0000001)
{thold<-thnew
thnew<-thold+BESF(thold[1],thold[2],x)%*%solve(BEIF(thold[1],thold[2]))
print(thnew)}
return(thnew)}
thetahat<-NewtonBE(2,2,x)
```

The maximum likelihood estimates are $\hat{a} = 2.824775$ and $\hat{b} = 2.97317$. The score vector evaluated at $(\hat{a}, \hat{b})$ is

$$S(\hat{a}, \hat{b}) = \begin{bmatrix} 3.108624 \times 10^{-14} & 7.771561 \times 10^{-15} \end{bmatrix}$$

which indicates we have obtained a local extrema. The observed information matrix is

$$I(\hat{a}, \hat{b}) = \begin{bmatrix} 8.249382 & -6.586959 \\ -6.586959 & 7.381967 \end{bmatrix}$$

Note that since

$$\det[I(\hat{a}, \hat{b})] = (8.249382)\,(7.381967) - (6.586959)^2 > 0$$

and

$$[I(\hat{a}, \hat{b})]_{11} = 8.249382 > 0$$

then by the second derivative test we have found the maximum likelihood estimates.

7.1.14 Exercise

Use the following R code to randomly generate 30 observations from a Gamma(α, β) distribution

```
# randomly generate 35 observations from a Gamma(a,b)
set.seed(32067489)    # set the seed so results can be reproduced
# use randomly generated a and b values
truea<-runif(1,min=1,max=3)
trueb<-runif(1,min=3,max=5)
# data are sorted and rounded to two decimal places for easier display
x<-sort(round(rgamma(30,truea,scale=trueb),2))
x
```

Use Newton's Method and R to find $(\hat{\alpha}, \hat{\beta})$. What are the values of $S(\hat{\alpha}, \hat{\beta})$ and $I(\hat{\alpha}, \hat{\beta})$?

7.2 Likelihood Regions

For one unknown parameter likelihood intervals provide a way of summarizing the uncertainty in the maximum likelihood estimate by providing an intervals of values which are plausible given the observed data. For k unknown parameter summarizing the uncertainty is more challenging. We begin with the definition of a likelihood region which is the natural extension of a likelihood interval.

7.2.1 Definition - Likelihood Regions

The set of $\boldsymbol{\theta}$ values for which $R(\boldsymbol{\theta}) \geq p$ is called a $100p\%$ *likelihood region* for $\boldsymbol{\theta}$.

A $100p\%$ likelihood region for two unknown parameters $\boldsymbol{\theta} = (\theta_1, \theta_2)$ is given by $\{(\theta_1, \theta_2)\,;\ R(\theta_1, \theta_2) \geq p\}$. These regions will be elliptical in shape. To show this we note that for (θ_1, θ_2) sufficiently close to $(\hat{\theta}_1, \hat{\theta}_2)$ we have

$$\begin{aligned} L(\theta_1, \theta_2) &\approx L(\hat{\theta}_1, \hat{\theta}_2) + (\hat{\theta}_1, \hat{\theta}_2) \begin{bmatrix} \hat{\theta}_1 - \theta_1 \\ \hat{\theta}_2 - \theta_2 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} \hat{\theta}_1 - \theta_1 & \hat{\theta}_2 - \theta_2 \end{bmatrix} I(\hat{\theta}_1, \hat{\theta}_2) \begin{bmatrix} \hat{\theta}_1 - \theta_1 \\ \hat{\theta}_2 - \theta_2 \end{bmatrix} \\ &= L(\hat{\theta}_1, \hat{\theta}_2) + \frac{1}{2} \begin{bmatrix} \hat{\theta}_1 - \theta_1 & \hat{\theta}_2 - \theta_2 \end{bmatrix} I(\hat{\theta}_1, \hat{\theta}_2) \begin{bmatrix} \hat{\theta}_1 - \theta_1 \\ \hat{\theta}_2 - \theta_2 \end{bmatrix} \qquad \text{since } S(\hat{\theta}_1, \hat{\theta}_2) = 0 \end{aligned}$$

Therefore

$$\begin{aligned} R(\theta_1, \theta_2) &= \frac{L(\theta_1, \theta_2)}{L(\hat{\theta}_1, \hat{\theta}_2)} \\ &\approx 1 - \frac{1}{2L(\hat{\theta}_1, \hat{\theta}_2)} \begin{bmatrix} \hat{\theta}_1 - \theta_1 & \hat{\theta}_2 - \theta_2 \end{bmatrix} \begin{bmatrix} \hat{I}_{11} & \hat{I}_{12} \\ \hat{I}_{12} & \hat{I}_{22} \end{bmatrix} \begin{bmatrix} \hat{\theta}_1 - \theta_1 \\ \hat{\theta}_2 - \theta_2 \end{bmatrix} \\ &= 1 - \left[2L(\hat{\theta}_1, \hat{\theta}_2)\right]^{-1} \left[(\theta_1 - \hat{\theta}_1)^2 \hat{I}_{11} + 2\left(\theta_1 - \hat{\theta}_1\right)(\theta_2 - \hat{\theta}_2)\hat{I}_{12} + (\theta_2 - \hat{\theta}_2)^2 \hat{I}_{22}\right] \end{aligned}$$

The set of points (θ_1, θ_2) which satisfy $R(\theta_1, \theta_2) = p$ is approximately the set of points (θ_1, θ_2) which satisfy

$$(\theta_1 - \hat{\theta}_1)^2 \hat{I}_{11} + 2(\theta_1 - \hat{\theta}_1)(\theta_2 - \hat{\theta}_2)\hat{I}_{12} + (\theta_2 - \hat{\theta}_2)^2 \hat{I}_{22} = 2(1-p)\, L(\hat{\theta}_1, \hat{\theta}_2)$$

which we recognize as the points on an ellipse centred at $(\hat{\theta}_1, \hat{\theta}_2)$. Therefore a $100p\%$ likelihood region for two unknown parameters $\boldsymbol{\theta} = (\theta_1, \theta_2)$ will be the set of points on and inside a region which will be approximately elliptical in shape.

A similar argument can be made to show that the likelihood regions for three unknown parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$ will be approximate ellipsoids in $\Re^3$.

7.2.2 Example

(a) Use R to graph 1%, 5%, 10%, 50%, and 90% likelihood regions for the parameters (a, b) in Example 7.1.13. Comment on the shapes of the regions.

(b) Is the value $(2.5, 3.5)$ a plausible value of (a, b)?

Solution

(a) The following R code generates the required likelihood regions.

```
# function for calculating Beta relative likelihood function
# for parameters a,b and data x
BERLF<-function(a,b,that,x)
{t1<-prod(x)
t2<-prod(1-x)
n<-length(x)
ah<-that[1]
bh<-that[2]
L<-<-((gamma(a+b)*gamma(ah)*gamma(bh))/
   (gamma(a)*gamma(b)*gamma(ah+bh)))^n*t1^(a-ah)*t2^(b-bh)
return(L)}
#
a<-seq(0.5,5.5,0.01)
b<-seq(0.5,6,0.01)
R<-outer(a,b,FUN = BERLF,thetahat,x)
contour(a,b,R,levels=c(0.01,0.05,0.10,0.50,0.9),xlab="a",ylab="b",lwd=2)
```

The 1%, 5%, 10%, 50%, and 90% likelihood regions for (a, b) are shown in Figure 7.2.

The likelihood contours are approximate ellipses but they are not symmetric about the maximum likelihood estimates $(\hat{a}, \hat{b}) = (2.824775, 2.97317)$. The likelihood regions are more stretched for larger values of a and b. The ellipses are also skewed relative to the ab coordinate axes. The skewness of the likelihood contours relative to the ab coordinate axes is determined by the value of $\hat{I}_{12}$. If the value of $\hat{I}_{12}$ is close to zero the skewness will be small.

(b) Since $R(2.5, 3.5) = 0.082$ the point $(2.5, 3.5)$ lies outside a 10% likelihood region so it is not a very plausible value of (a, b).

Note however that $a = 2.5$ is a plausible value of a for some values of b, for example, $a = 2.5$, $b = 2.5$ lies inside a 50% likelihood region so $(2.5, 2.5)$ is a plausible value of (a, b). We see that when there is more than one parameter then we need to determine whether a set of values are jointly plausible given the observed data.

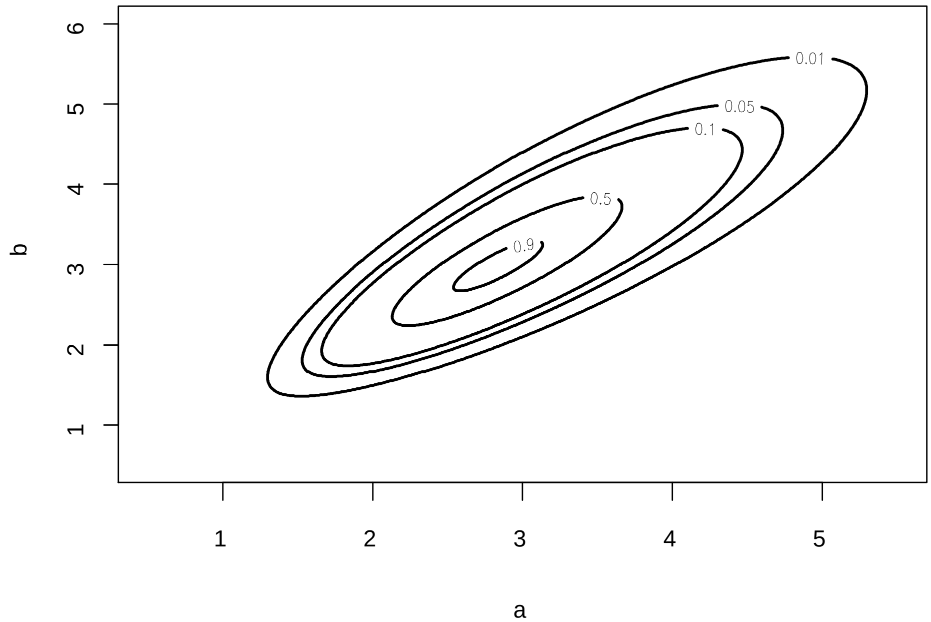


Figure 7.2: Likelihood regions for Beta example

7.2.3 Exercise

(a) Use R to graph 1%, 5%, 10%, 50%, and 90% likelihood regions for the parameters (α, β) in Exercise 7.1.14. Comment on the shapes of the regions.

(b) Is the value $(3, 2.7)$ a plausible value of (α, β)?

(c) Use the R code in Exercise 7.1.14 to generate 100 observations from the Gamma(α, β) distribution.

(d) Use R to graph 1%, 5%, 10%, 50%, and 90% likelihood regions for (α, β) for the data generated in (c). Comment on the shapes of these regions as compared to the regions in (a).

7.3 Limiting Distribution of Maximum Likelihood Estimator

To discuss the asymptotic properties of the maximum likelihood estimator in the multiparameter case we need to define convergence in probability and convergence in distribution of a sequence of random vectors.

7.3.1 Definition - Convergence of a Sequence of Random Vectors

Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n, \ldots$ be a sequence of random vectors where $\mathbf{X}_n = (X_{1n}, X_{2n}, \ldots, X_{kn})$. Let $\mathbf{X} = (X_1, X_2, \ldots, X_k)$ be a random vector and $\mathbf{x} = (x_1, x_2, \ldots, x_k)$.

(1) If $X_{in} \to_p X_i$ for $i = 1, 2, \ldots, k$, then

$$\mathbf{X}_n \to_p \mathbf{X}$$

(2) Let $F_n(\mathbf{x}) = P(X_{1n} \leq x_1, X_{2n} \leq x_2, \ldots, X_{kn} \leq x_k)$ be the cumulative distribution function of $\mathbf{X}_n$. Let $F(\mathbf{x}) = P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_k \leq x_k)$ be the cumulative distribution function of $\mathbf{X}$. If

$$\lim_{n\to\infty} F_n(\mathbf{x}) = F(\mathbf{x})$$

at all points of continuity of $F(\mathbf{x})$ then

$$\mathbf{X}_n \to_D \mathbf{X} = (X_1, X_2, \ldots, X_k)$$

To discuss the asymptotic properties of the maximum likelihood estimator in the multiparameter case we also need the definition and properties of the Multivariate Normal Distribution. The Multivariate Normal distribution is the natural extension of the Bivariate Normal distribution which was discussed in Section 3.10.

7.3.2 Definition - Multivariate Normal Distribution (MVN)

Let $\mathbf{X} = (X_1, X_2, \ldots, X_k)$ be a $1 \times k$ random vector with $E(X_i) = \mu_i$ and $Cov(X_i, X_j) = \sigma_{ij}$, $i, j = 1, 2, \ldots, k$. (Note: $Cov(X_i, X_i) = \sigma_{ii} = Var(X_i) = \sigma_i^2$.) Let $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_k)$ be the mean vector and Σ be the $k \times k$ symmetric covariance matrix whose (i, j) entry is σ_{ij}. Suppose also that the inverse matrix of Σ, Σ^{-1}, exists. If the joint probability density function of $(X_1, X_2, \ldots, X_k)$ is given by

$$f(x_1, x_2, \ldots, x_k) = \frac{1}{(2\pi)^{k/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})^T\right] \quad \text{for } \mathbf{x} \in \Re^k$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_k)$ then $\mathbf{X}$ is said to have a *Multivariate Normal distribution*. We write $\mathbf{X} \sim \text{MVN}(\boldsymbol{\mu}, \Sigma)$.

The following theorem gives some important properties of the Multivariate Normal distribution. These properties are a natural extension of the properties of the Bivariate Normal distribution found in Theorem 3.10.2.

7.3.3 Theorem - Properties of MVN Distribution

Suppose $\mathbf{X} = (X_1, X_2, \ldots, X_k) \sim \text{MVN}(\boldsymbol{\mu}, \Sigma)$. Then

(1) X has joint moment generating function

$$M(\mathbf{t}) = \exp\left(\boldsymbol{\mu}\mathbf{t}^T + \frac{1}{2}\mathbf{t}\Sigma\mathbf{t}^T\right) \quad \text{for } \mathbf{t} = (t_1, t_2, \ldots, t_k) \in \Re^k$$

(2) Any subset of $X_1, X_2, \ldots, X_k$ also has a MVN distribution and in particular $X_i \sim \text{N}\big(\mu_i, \sigma_i^2\big)$, $i = 1, 2, \ldots, k$.

(3)

$$(\mathbf{X} - \boldsymbol{\mu})\Sigma^{-1}(\mathbf{X} - \boldsymbol{\mu})^T \sim \chi^2(k)$$

(4) Let $\mathbf{c} = (c_1, c_2, \ldots, c_k)$ be a nonzero vector of constants then

$$\mathbf{X}\mathbf{c}^T = \sum_{i=1}^{k} c_i X_i \sim \text{N}\left(\boldsymbol{\mu}\mathbf{c}^T,\ \mathbf{c}\Sigma\mathbf{c}^T\right)$$

(5) Let A be a $k \times p$ vector of constants of rank $p \leq k$ then

$$\mathbf{X}A \sim \text{N}(\boldsymbol{\mu}A,\ A^T\Sigma A)$$

(6) The conditional distribution of any subset of $(X_1, X_2, \ldots, X_k)$ given the rest of the coordinates is a MVN distribution. In particular the conditional probability density function of X_i given $X_j = x_j$, $i \neq j$, is

$$X_i | X_j = x_j \sim \text{N}(\mu_i + \rho_{ij}\sigma_i(x_j - \mu_j)/\sigma_j,\ (1 - \rho_{ij}^2)\sigma_i^2)$$

The following theorem gives the asymptotic distribution of the maximum likelihood estimator in the multiparameter case. This theorem looks very similar to Theorem 6.5.1 with the scalar quantities replaced by the appropriate vectors and matrices.

7.3.4 Theorem - Limiting Distribution of the Maximum Likelihood Estimator

Suppose $\mathbf{X}_n = (X_1, X_2, \ldots, X_n)$ is a random sample from $f(x; \boldsymbol{\theta})$ where $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k) \in \Omega$ and the dimension of Ω is k. Let $\tilde{\boldsymbol{\theta}}_n = \tilde{\boldsymbol{\theta}}_n(X_1, X_2, \ldots, X_n)$ be the maximum likelihood estimator of $\boldsymbol{\theta}$ based on $\mathbf{X}_n$. Let $\mathbf{0}_k$ be a $1 \times k$ vector of zeros, $\mathbf{I}_k$ be the $k \times k$ identity matrix, and $[J(\boldsymbol{\theta})]^{1/2}$ be a matrix such that $[J(\boldsymbol{\theta})]^{1/2}[J(\boldsymbol{\theta})]^{1/2} = J(\boldsymbol{\theta})$. Then under certain (regularity) conditions

$$\tilde{\boldsymbol{\theta}}_n \to_p \boldsymbol{\theta} \tag{7.2}$$

$$(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta})[J(\boldsymbol{\theta})]^{1/2} \to_D \mathbf{Z} \sim \text{MVN}(\mathbf{0}_k, \mathbf{I}_k) \tag{7.3}$$

$$-2\log R(\boldsymbol{\theta}; \mathbf{X}_n) = 2[l(\tilde{\boldsymbol{\theta}}_n; \mathbf{X}_n) - l(\boldsymbol{\theta}; \mathbf{X}_n)] \to_D W \sim \chi^2(k) \tag{7.4}$$

for each $\boldsymbol{\theta} \in \Omega$.

Since $\tilde{\boldsymbol{\theta}}_n \to_p \boldsymbol{\theta}$, $\tilde{\boldsymbol{\theta}}_n$ is a consistent estimator of $\boldsymbol{\theta}$.

Theorem 7.3.4 implies that for sufficiently large n, $\tilde{\boldsymbol{\theta}}_n$ has an approximately MVN$(\boldsymbol{\theta}, [J(\theta_0)]^{-1})$ distribution. Therefore for sufficiently large n

$$E(\tilde{\boldsymbol{\theta}}_n) \approx \boldsymbol{\theta}$$

and therefore $\tilde{\boldsymbol{\theta}}_n$ is an asymptotically unbiased estimator of $\boldsymbol{\theta}$. Also

$$Var(\tilde{\boldsymbol{\theta}}_n) \approx [J(\boldsymbol{\theta})]^{-1}$$

where $[J(\boldsymbol{\theta})]^{-1}$ is the inverse matrix of of the matrix $J(\boldsymbol{\theta})$. (Since $J(\boldsymbol{\theta})$ is a $k \times k$ symmetric matrix, $[J(\boldsymbol{\theta})]^{-1}$ also a $k \times k$ symmetric matrix.) $[J(\boldsymbol{\theta})]^{-1}$ is called the *asymptotic variance/covariance* matrix of $\tilde{\boldsymbol{\theta}}_n$. Of course $J(\boldsymbol{\theta})$ is unknown because $\boldsymbol{\theta}$ is unknown. But (7.2), (7.3) and Slutsky's Theorem imply that

$$(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta})[J(\tilde{\boldsymbol{\theta}}_n)]^{1/2} \to_D Z \sim \text{MVN}(\mathbf{0}_k, \mathbf{I}_k)$$

Therefore for sufficiently large n we have

$$Var(\tilde{\boldsymbol{\theta}}_n) \approx \left[J(\hat{\boldsymbol{\theta}}_n)\right]^{-1}$$

where $\left[J(\hat{\boldsymbol{\theta}}_n)\right]^{-1}$ is the inverse matrix of $J(\hat{\boldsymbol{\theta}}_n)$.

It is also possible to show that

$$(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta})[I(\tilde{\boldsymbol{\theta}}_n; \mathbf{X})]^{1/2} \to_D \text{MVN}(\mathbf{0}_k, \mathbf{I}_k)$$

so that for sufficiently large n we also have

$$Var(\tilde{\boldsymbol{\theta}}_n) \approx \left[I(\hat{\boldsymbol{\theta}}_n)\right]^{-1}$$

where $\left[I(\hat{\boldsymbol{\theta}}_n)\right]^{-1}$ is the inverse matrix of the observed information matrix $I(\hat{\boldsymbol{\theta}}_n)$.

These results can be used to construct approximate confidence regions for $\boldsymbol{\theta}$ as shown in the next section.

Note: These results do not hold if the support set of $\mathbf{X}$ depends on $\boldsymbol{\theta}$.

7.4 Approximate Confidence Regions

7.4.1 Definition - Confidence Region

A $100p\%$ *confidence region* for $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$ based on $\mathbf{X}$ is a region $R(\mathbf{X}) \subset \Re^k$ which satisfies

$$P[\boldsymbol{\theta} \in R(\mathbf{X})] = p$$

Exact confidence regions can only be obtained in a very few special cases such as Normal linear models. More generally we must rely on approximate confidence regions based on the results of Theorem 7.3.3.

7.4.2 Asymptotic Pivotal Quantities and Approximate Confidence Regions

The limiting distribution of $\tilde{\boldsymbol{\theta}}_n$ can be used to obtain approximate confidence regions for $\boldsymbol{\theta}$. Since

$$(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta})[J(\boldsymbol{\theta})]^{1/2} \to_D Z \sim \text{MVN}(\mathbf{0}_k, \mathbf{I}_k)$$

it follows from Theorem 7.3.3(3) and Limit Theorems that

$$(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta})J(\tilde{\boldsymbol{\theta}}_n)(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta})^T \to_D W \sim \chi^2(k)$$

An approximate $100p\%$ confidence region for $\boldsymbol{\theta}$ based on the asymptotic pivotal quantity $(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta})J(\tilde{\boldsymbol{\theta}}_n)(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta})^T$ is the set of all $\boldsymbol{\theta}$ vectors in the set

$$\{\boldsymbol{\theta} : (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})J(\hat{\boldsymbol{\theta}}_n)(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})^T \le c\}$$

where c is the value such that $P(W \le c) = p$ and $W \sim \chi^2(k)$.

Similarly since

$$(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta})[I(\tilde{\boldsymbol{\theta}}_n; \mathbf{X}_n)]^{1/2} \to_D Z \sim \text{MVN}(\mathbf{0}_k, \mathbf{I}_k)$$

it follows from Theorem 7.3.3(3) and Limit Theorems that

$$(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta})I(\tilde{\boldsymbol{\theta}}_n; \mathbf{X}_n)(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta})^T \to_D W \sim \chi^2(k)$$

An approximate $100p\%$ confidence region for θ based on the asymptotic pivotal quantity $(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta})I(\tilde{\boldsymbol{\theta}}_n; \mathbf{X}_n)(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta})^T$ is the set of all $\boldsymbol{\theta}$ vectors in the set

$$\{\boldsymbol{\theta} : (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})I(\hat{\boldsymbol{\theta}}_n)(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})^T \le c\}$$

where $I(\hat{\boldsymbol{\theta}}_n)$ is the observed information.

Finally since

$$-2\log R(\boldsymbol{\theta}; \mathbf{X}_n) \to_D W \sim \chi^2(k)$$

an approximate $100p\%$ confidence region for θ based on this asymptotic pivotal quantity is the set of all $\boldsymbol{\theta}$ vectors satisfying

$$\{\boldsymbol{\theta} : -2\log R(\boldsymbol{\theta}; \mathbf{x}) \le c\}$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ are the observed data. Since

$$\begin{aligned} & \{\boldsymbol{\theta} : -2\log R(\boldsymbol{\theta}; \mathbf{x}) \le c\} \\ = & \left\{\boldsymbol{\theta} : R(\boldsymbol{\theta}; \mathbf{x}) \ge e^{-c/2}\right\} \end{aligned}$$

we recognize that this interval is actually a likelihood region.

7.4.3 Example

Use R and the results from Examples 7.1.10 and 7.1.13 to graph approximate 90%, 95%, and 99% confidence regions for (a, b). Compare these approximate confidence regions with the likelihood regions in Example 7.2.2.

Solution

From Example 7.1.10 that for a random sample from the Beta(a, b) distribution the information matrix and the expected information matrix are given by

$$I(a,b) = n \begin{bmatrix} \Psi'(a) - \Psi'(a+b) & -\Psi'(a+b) \\ -\Psi'(a+b) & \Psi'(b) - \Psi'(a+b) \end{bmatrix} = J(a,b)$$

Since

$$\begin{bmatrix} \tilde{a} - a & \tilde{b} - b \end{bmatrix} J(\tilde{a}, \tilde{b}) \begin{bmatrix} \tilde{a} - a \\ \tilde{b} - b \end{bmatrix} \to_D W \sim \chi^2(2)$$

an approximate $100p\%$ confidence region for (a, b) is given by

$$\{(a,b) : \begin{bmatrix} \hat{a} - a & \hat{b} - b \end{bmatrix} J(\hat{a}, \hat{b}) \begin{bmatrix} \hat{a} - a \\ \hat{b} - b \end{bmatrix} \leq c\}$$

where $P(W \leq c) = p$. Since $\chi^2(2) = \text{Gamma}(1, 2) = \text{Exponential}(2)$, c can be determined using

$$p = P(W \leq c) = \int_0^c \frac{1}{2} e^{-x/2} dx = 1 - e^{-c/2}$$

which gives

$$c = -2\log(1-p)$$

For $p = 0.95$, $c = -2\log(0.05) = 5.99$, an approximate 95% confidence region is given by

$$\{(a,b) : \begin{bmatrix} \hat{a} - a & \hat{b} - b \end{bmatrix} J(\hat{a}, \hat{b}) \begin{bmatrix} \hat{a} - a \\ \hat{b} - b \end{bmatrix} \leq 5.99\}$$

If we let

$$J(\hat{a}, \hat{b}) = \begin{bmatrix} \hat{J}_{11} & \hat{J}_{12} \\ \hat{J}_{12} & \hat{J}_{22} \end{bmatrix}$$

then the approximate confidence region can be written as

$$\{(a,b) : (\hat{a} - a)^2 \hat{J}_{11} + 2(\hat{a} - a)(\hat{b} - b)\hat{J}_{12} + (\hat{b} - b)^2 \hat{J}_{22} \leq 5.99\}$$

We note that the approximate confidence region is the set of points on and inside the ellipse

$$(\hat{a} - a)^2 \hat{J}_{11} + 2(\hat{a} - a)(\hat{b} - b)\hat{J}_{12} + (\hat{b} - b)^2 \hat{J}_{22} = 5.99$$

which is centred at $(\hat{a}, \hat{b})$.

For the data in Example 7.1.13, $\hat{a} = 2.824775$, $\hat{b} = 2.97317$ and

$$I(\hat{a}, \hat{b}) = J(\hat{a}, \hat{b}) = \begin{bmatrix} 8.249382 & -6.586959 \\ -6.586959 & 7.381967 \end{bmatrix}$$

Approximate 90% ($-2\log(0.1) = 4.61$), 95% ($-2\log(0.05) = 5.99$), and 99% ($-2\log(0.01) = 9.21$) confidence regions are shown in Figure 7.3.

The following R code generates the required approximate confidence regions.

```
# function for calculating values for determining confidence regions
ConfRegion<-function(a,b,th,info)
{c<-(th[1]-a)^2*info[1,1]+2*(th[1]-a)*
     (th[2]-b)*info[1,2]+(th[2]-b)^2*info[2,2]
return(c)}
#
# graph approximate confidence regions
a<-seq(1,5.5,0.01)
b<-seq(1,6,0.01)
c<-outer(a,b,FUN = ConfRegion,thetahat,Ithetahat)
contour(a,b,c,levels=c(4.61,5.99,9.21),xlab="a",ylab="b",lwd=2)
#
```

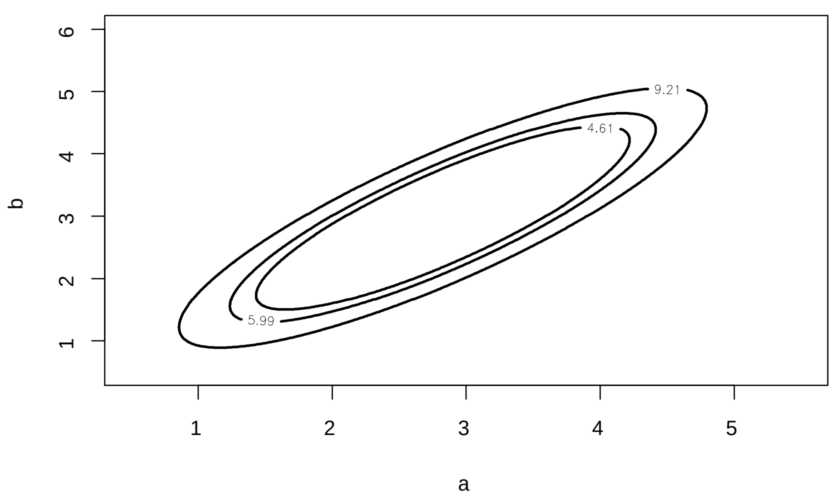


Figure 7.3: Approximate confidence regions for Beta(a, b) example

A 10% likelihood region for (a,b) is given by $\{(a,b) : R(a,b;\mathbf{x}) \geq 0.1\}$. Since

$$-2\log R(a,b;\mathbf{X}_n) \to_D W \sim \chi^2(2) = \text{ Exponential}(2)$$

we have

$$\begin{aligned}
P[R(a,b;\mathbf{X}) \geq 0.1] &= P[-2\log R(a,b;\mathbf{X}) \leq -2\log(0.1)] \\
&\approx P(W \leq -2\log(0.1)) \\
&= 1 - e^{-[-2\log(0.1)]/2} \\
&= 1 - 0.1 = 0.9
\end{aligned}$$

and therefore a 10% likelihood region corresponds to an approximate 90% confidence region. Similarly 1% and 5% likelihood regions correspond to approximate 99% and 95% confidence regions respectively.

If we compare the likelihood regions in Figure 7.2 with the approximate confidence regions shown in Figure 7.3 we notice that the confidence regions are exact ellipses centred at the maximum likelihood estimates whereas the likelihood regions are only approximate ellipses not centered at the maximum likelihood estimates. We notice that there are values inside an approximate 99% confidence region but which are outside a 1% likelihood region. The point $(a,b) = (1, 1.5)$ is an example. There were only 35 observations in this data set. The differences between the likelihood regions and the approximate confidence regions indicate that the Normal approximation might not be good. In this example the likelihood regions provide a better summary of the uncertainty in the estimates.

7.4.4 Exercise

Use R and the results from Exercises 7.1.11 and 7.1.14 to graph approximate 90%, 95%, and 99% confidence regions for (a,b). Compare these approximate confidence regions with the likelihood regions in Exercise 7.2.3.

Since likelihood regions and approximate confidence regions cannot be graphed or easily interpreted for more than two parameters, we often construct approximate confidence intervals for individual parameters. Such confidence intervals are often referred to as *marginal confidence intervals*. These confidence intervals must be used with care as we will see in Example 7.4.6.

Approximate confidence intervals can also be constructed for a linear combination of parameters. An illustration is given in Example 7.4.6.

7.4.5 Approximate Marginal Confidence Intervals

Let θ_i be the ith entry in the vector $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$. Since

$$(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta})[J(\boldsymbol{\theta})]^{1/2} \to_D Z \sim \text{MVN}(\mathbf{0}_k, \mathbf{I}_k)$$

it follows that an approximate $100p\%$ marginal confidence interval for θ_i is given by

$$[\hat{\theta}_i - a\sqrt{\hat{v}_{ii}},\ \hat{\theta}_i + a\sqrt{\hat{v}_{ii}}]$$

where $\hat{\theta}_i$ is the ith entry in the vector $\hat{\boldsymbol{\theta}}_n$, $\hat{v}_{ii}$ is the (i,i) entry of the matrix $[J(\hat{\boldsymbol{\theta}}_n)]^{-1}$, and a is the value such that $P(Z \leq a) = \frac{1+p}{2}$ where $Z \sim \text{N}(0,1)$.

Similarly since

$$(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta})[I(\tilde{\boldsymbol{\theta}}_n; \mathbf{X}_n)]^{1/2} \to_D Z \sim \text{MVN}(\mathbf{0}_k, \mathbf{I}_k)$$

it follows that an approximate $100p\%$ confidence interval for θ_i is given by

$$[\hat{\theta}_i - a\sqrt{\hat{v}_{ii}},\ \hat{\theta}_i + a\sqrt{\hat{v}_{ii}}]$$

where $\hat{v}_{ii}$ is now the (i,i) entry of the matrix $[I(\hat{\boldsymbol{\theta}}_n)]^{-1}$.

7.4.6 Example

Using the results from Examples 7.1.10 and 7.1.13 determine approximate 95% marginal confidence intervals for a, b, and an approximate confidence interval for $a+b$.

Solution

Let

$$\left[J(\hat{a},\hat{b})\right]^{-1} = \begin{bmatrix} \hat{v}_{11} & \hat{v}_{12} \\ \hat{v}_{12} & \hat{v}_{22} \end{bmatrix}$$

Since

$$\begin{bmatrix} \tilde{a}-a & \tilde{b}-b \end{bmatrix} [J(\tilde{a},\tilde{b})]^{1/2} \to_D Z \sim \text{BVN}\left(\begin{bmatrix} 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right)$$

then for large n, $Var(\tilde{a}) \approx \hat{v}_{11}$, $Var(\tilde{b}) \approx \hat{v}_{22}$ and $Cov(\tilde{a},\tilde{b}) \approx \hat{v}_{12}$. Therefore an approximate 95% confidence interval for a is given by

$$[\hat{a} - 1.96\sqrt{\hat{v}_{11}},\ \hat{a} + 1.96\sqrt{\hat{v}_{11}}]$$

and an approximate 95% confidence interval for b is given by

$$[\hat{b} - 1.96\sqrt{\hat{v}_{22}},\ \hat{b} + 1.96\sqrt{\hat{v}_{22}}]$$

For the data in Example 7.1.13, $\hat{a} = 2.824775$, $\hat{b} = 2.97317$ and

$$\begin{aligned}
\left[I(\hat{a},\hat{b})\right]^{-1} &= \left[J(\hat{a},\hat{b})\right]^{-1} \\
&= \begin{bmatrix} 8.249382 & -6.586959 \\ -6.586959 & 7.381967 \end{bmatrix}^{-1} \\
&= \begin{bmatrix} 0.4216186 & 0.3762120 \\ 0.3762120 & 0.4711608 \end{bmatrix}
\end{aligned}$$

An approximate 95% marginal confidence interval for a is

$$[2.824775 + 1.96\sqrt{0.4216186},\ 2.824775 - 1.96\sqrt{0.4216186}] = [1.998403, 3.651148]$$

and an approximate 95% confidence interval for b is

$$[2.97317 - 1.96\sqrt{0.4711608},\ 2.97317 + 1.96\sqrt{0.4711608}] = [2.049695, 3.896645]$$

Note that $a = 2.1$ is in the approximate 95% marginal confidence interval for a and $b = 3.8$ is in the approximate 95% marginal confidence interval for b and yet the point $(2.1, 3.8)$ is not in the approximate 95% joint confidence region for (a, b). Clearly these marginal confidence intervals for a and b must be used with care.

To obtain an approximate 95% marginal confidence interval for $a + b$ we note that

$$\begin{aligned}
Var(\tilde{a} + \tilde{b}) &= Var(\tilde{a}) + Var(\tilde{b}) + 2Cov(\tilde{a}, \tilde{b}) \\
&\approx \hat{v}_{11} + \hat{v}_{22} + 2\hat{v}_{12} = \hat{v}
\end{aligned}$$

so that an approximate 95% confidence interval for $a + b$ is given by

$$[\hat{a} + \hat{b} - 1.96\sqrt{\hat{v}},\ \hat{a} + \hat{b} + 1.96\sqrt{\hat{v}}]$$

For the data in Example 7.1.13

$$\begin{aligned}
\hat{a} + \hat{b} &= 2.824775 + 2.824775 = 5.797945 \\
\hat{v} &= \hat{v}_{11} + \hat{v}_{22} + 2\hat{v}_{12} = 0.4216186 + 0.4711608 + 2(0.3762120) = 1.645203
\end{aligned}$$

and an approximate 95% marginal confidence interval for $a + b$ is

$$[5.797945 + 1.96\sqrt{1.645203},\ 5.797945 - 1.96\sqrt{1.645203}] = [2.573347, 9.022544]$$

7.4.7 Exercise

Using the results from Exercises 7.1.11 and 7.1.14 determine approximate 95% marginal confidence intervals for a, b, and an approximate confidence interval for $a + b$.

7.5 Chapter 7 Problems

1. Suppose $x_1, x_2, \ldots, x_n$ is a observed random sample from the distribution with cumulative distribution function

$$F(x; \theta_1, \theta_2) = 1 - \left(\frac{\theta_1}{x}\right)^{\theta_2} \quad \text{for } x \geq \theta_1, \quad \theta_1 > 0, \quad \theta_2 > 0$$

   Find the maximum likelihood estimates and the maximum likelihood estimators of θ_1 and θ_2.

2. Suppose $(X_1, X_2, X_3) \sim \text{Multinomial}(n, \theta_1, \theta_2, \theta_3)$. Verify that the maximum likelihood estimators of θ_1 and θ_2 are $\tilde{\theta}_1 = X_1/n$ and $\tilde{\theta}_2 = X_2/n$. Find the expected information for θ_1 and θ_2.

3. Suppose $x_{11}, x_{12}, \ldots, x_{1n_1}$ is an observed random sample from the $\text{N}(\mu_1, \sigma^2)$ distribution and independently $x_{21}, x_{22}, \ldots, x_{2n_2}$ is an observed random sample from the $\text{N}(\mu_2, \sigma^2)$ distribution. Find the maximum likelihood estimators of μ_1, μ_2, and σ^2.

4. In a large population of males ages $40 - 50$, the proportion who are regular smokers is α where $0 \leq \alpha \leq 1$ and the proportion who have hypertension (high blood pressure) is β where $0 \leq \beta \leq 1$. Suppose that n men are selected at random from this population and the observed data are

Category	SH	$S\bar{H}$	$\bar{S}H$	$\bar{S}\bar{H}$
Frequency	x_{11}	x_{12}	x_{21}	x_{22}

   where S is the event the male is a smoker and H is the event the male has hypertension.

   (a) Assuming the events S and H are independent determine the likelihood function, the score vector, the maximum likelihood estimates, and the information matrix for α and β.

   (b) Determine the expected information matrix and its inverse matrix. What do you notice regarding the diagonal entries of the inverse matrix?

5. Suppose $x_1, x_2, \ldots, x_n$ is an observed random sample from the $\text{Logistic}(\mu, \beta)$ distribution.

   (a) Find the likelihood function, the score vector, and the information matrix for μ and β. How would you find the maximum likelihood estimates of μ and β?

   (b) Show that if u is an observation from the $\text{Uniform}(0, 1)$ distribution then

$$x = \mu - \beta \log\left(\frac{1}{u} - 1\right)$$

   is an observation from the $\text{Logistic}(\mu, \beta)$ distribution.

(c) Use the following R code to randomly generate 30 observations from a Logistic(μ, β) distribution.

```
# randomly generate 30 observations from a Logistic(mu,beta)
# using a random mu and beta values
set.seed(21086689) # set the seed so results can be reproduced
truemu<-runif(1,min=2,max=3)
truebeta<-runif(1,min=3,max=4)
# data are sorted and rounded to two decimal places for easier display
x<-sort(round((truemu-truebeta*log(1/runif(30)-1)),2))
x
```

(d) Use Newton's Method and R to find $(\hat{\mu}, \hat{\beta})$. Determine $S(\hat{\mu}, \hat{\beta})$ and $I(\hat{\mu}, \hat{\beta})$.

(e) Use R to graph 1%, 5%, 10%, 50%, and 90% likelihood regions for (μ, β).

(f) Use R to graph approximate 90%, 95%, and 99% confidence regions for (μ, β). Compare these approximate confidence regions with the likelihood regions in (e).

(g) Determine approximate 95% marginal confidence intervals for μ, β, and an approximate confidence interval for $\mu + \beta$.

6. Suppose $x_1, x_2, \ldots, x_n$ is an observed random sample from the Weibull(α, β) distribution.

(a) Find the likelihood function, the score vector, and the information matrix for α and β. How would you find the maximum likelihood estimates of α and β?

(b) Show that if u is an observation from the Uniform$(0, 1)$ distribution then

$$x = \beta \left[-\log(1-u)\right]^{1/\alpha}$$

is an observation from the Weibull(α, β) distribution.

(c) Use the following R code to randomly generate 40 observations from a Weibull(α, β) distribution.

```
# randomly generate 40 observations from a Weibull(alpha,beta)
# using random values for alpha and beta
set.seed(21086689) # set the seed so results can be reproduced
truealpha<-runif(1,min=2,max=3)
truebeta<-runif(1,min=3,max=4)
# data are sorted and rounded to two decimal places for easier display
x<-sort(round(truebeta*(-log(1-runif(40)))^(1/truealpha),2))
x
```

(d) Use Newton's Method and R to find $(\hat{\alpha}, \hat{\beta})$. Determine $S(\hat{\alpha}, \hat{\beta})$ and $I(\hat{\alpha}, \hat{\beta})$.

(e) Use R to graph 1%, 5%, 10%, 50%, and 90% likelihood regions for (α, β).

(f) Use R to graph approximate 90%, 95%, and 99% confidence regions for (α, β). Compare these approximate confidence regions with the likelihood regions in (e).

(g) Determine approximate 95% marginal confidence intervals for α, β, and an approximate confidence interval for $\alpha + \beta$.

7. Suppose $Y_i \sim \text{Binomial}(1, p_i)$, $i = 1, 2, \ldots, n$ independently where $p_i = \left(1 + e^{-\alpha - \beta x_i}\right)^{-1}$ and the x_i are known constants.

(a) Determine the likelihood function, the score vector, and the expected information matrix for α and β.

(b) Explain how you would use Newton's method to find the maximum likelihood estimates of α and β.

8. Suppose $x_1, x_2, \ldots, x_n$ is an observed random sample from the Three Parameter Burr distribution with probability density function

$$f(x; \alpha, \beta, \gamma) = \frac{\alpha\gamma\,(x/\beta)^{\alpha-1}}{\beta\left[1 + (x/\beta)^{\alpha}\right]^{\gamma+1}} \quad \text{for } x > 0,\ \alpha > 0,\ \beta > 0,\ \gamma > 0$$

(a) Find the likelihood function, the score vector, and the information matrix for α, β, and γ. How would you find the maximum likelihood estimates of α, β, and γ?

(b) Show that if u is an observation from the Uniform$(0,1)$ distribution then

$$x = \beta\left[(1-u)^{-1/\gamma} - 1\right]^{1/\alpha}$$

is an observation from the Three Parameter Burr distribution.

(c) Use the following R code to randomly generate 60 observations from the Three Parameter Burr distribution.

```
# randomly generate 60 observations from the 3 Parameter Burr
# distribution using random values for alpha, beta and gamma
set.seed(21086689) # set the seed so results can be reproduced
truea<-runif(1,min=2,max=3)
trueb<-runif(1,min=3,max=4)
truec<-runif(1,min=3,max=4)
# data are sorted and rounded to 2 decimal places for easier display
x<-sort(round(trueb*((1-runif(60))^(-1/truec)-1)^(1/truea),2))
x
```

(d) Use Newton's Method and R to find $(\hat{\alpha}, \hat{\beta}, \hat{\gamma})$. Determine $S(\hat{\alpha}, \hat{\beta}, \hat{\gamma})$ and $I(\hat{\alpha}, \hat{\beta}, \hat{\gamma})$. Use the second derivative test to verify that $(\hat{\alpha}, \hat{\beta}, \hat{\gamma})$ are the maximum likelihood estimates.

(e) Determine approximate 95% marginal confidence intervals for α, β, and γ, and an approximate confidence interval for $\alpha + \beta + \gamma$.

8. Hypothesis Testing

Point estimation is a useful statistical procedure for estimating unknown parameters in a model based on observed data. Interval estimation is a useful statistical procedure for quantifying the uncertainty in these estimates. Hypothesis testing is another important statistical procedure which is used for deciding whether a given statement is supported by the observed data.

In Section 8.1 we review the definitions and steps of a test of hypothesis. Much of this material was introduced in a previous statistics course such as STAT 221/231/241. In Section 8.2 we look at how the likelihood function can be use to construct a test of hypothesis when the model is completely specified by the hypothesis of interest. The material in this section is mostly a review of material covered in a previous statistics course. In Section 8.3 we look at how the likelihood function function can be use to construct a test of hypothesis when the model is not completely specified by the hypothesis of interest. The material in this section is mostly new material.

8.1 Test of Hypothesis

In order to analyse a set of data $\mathbf{x}$ we often assume a model $f(\mathbf{x};\boldsymbol{\theta})$ where $\boldsymbol{\theta} \in \Omega$ and Ω is the parameter space or set of possible values of $\boldsymbol{\theta}$. A *test of hypothesis* is a statistical procedure used for evaluating the strength of the evidence provided by the observed data against an hypothesis. An hypothesis is a statement about the model. In many cases the hypothesis can be formulated in terms of the parameter $\boldsymbol{\theta}$ as

$$H_0 : \boldsymbol{\theta} \in \Omega_0$$

where Ω_0 is some subset of Ω. H_0 is called the null hypothesis. When conducting a test of hypothesis there is usually another statement of interest which is the statement which reflects what might be true if H_0 is not supported by the observed data. This statement is called the alternative hypothesis and is denoted H_A or H_1. In many cases H_A may simply take the form

$$H_A : \boldsymbol{\theta} \notin \Omega_0$$

In constructing a test of hypothesis it is useful to distinguish between simple and composite hypotheses.

8.1.1 Definition - Simple and Composite Hypotheses

If the hypothesis completely specifies the model including any parameters in the model then the hypothesis is simple otherwise the hypothesis is composite.

8.1.2 Example

For each of the following indicate whether the null hypothesis is simple or composite. Specify Ω and Ω_0 and determine the dimension of each.

(a) It is assumed that the observed data $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ represent a random sample from a Poisson(θ) distribution. The hypothesis of interest is $H_0 : \theta = \theta_0$ where θ_0 is a specified value of θ.

(b) It is assumed that the observed data $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ represent a random sample from a Gamma(α, β) distribution. The hypothesis of interest is $H_0 : \alpha = \alpha_0$ where α_0 is a specified value of α.

(c) It is assumed that the observed data $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ represent a random sample from an Exponential(θ_1) distribution and independently the observed data $\mathbf{y} = (y_1, y_2, \ldots, y_m)$ represent a random sample from an Exponential(θ_2) distribution. The hypothesis of interest is $H_0 : \theta_1 = \theta_2$.

(d) It is assumed that the observed data $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ represent a random sample from a N(μ_1, σ_1^2) distribution and independently the observed data $\mathbf{y} = (y_1, y_2, \ldots, y_m)$ represent a random sample from a N(μ_2, σ_2^2) distribution. The hypothesis of interest is $H_0 : \mu_1 = \mu_2,\ \sigma_1^2 = \sigma_2^2$.

Solution

(a) This is a simple hypothesis since the model and the unknown parameter are completely specified. $\Omega = \{\theta : \theta > 0\}$ which has dimension 1 and $\Omega_0 = \{\theta_0\}$ which has dimension 0.

(b) This is a composite hypothesis since α is not specified by H_0. $\Omega = \{(\alpha, \beta) : \alpha > 0,\ \beta > 0\}$ which has dimension 2 and $\Omega_0 = \{(\alpha_0, \beta) : \beta > 0\}$ which has dimension 1.

(c) This is a composite hypothesis since θ_1 and θ_2 are not specified by H_0.
$\Omega = \{(\theta_1, \theta_2) : \theta_1 > 0,\ \theta_2 > 0\}$ which has dimension 2 and
$\Omega_0 = \{(\theta_1, \theta_2) : \theta_1 = \theta_2,\ \theta_1 > 0,\ \theta_2 > 0\}$ which has dimension 1.

(d) This is a composite hypothesis since μ_1, σ_1^2, μ_2, and σ_2^2 are not specified by H_0.
$\Omega = \left\{ \left(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2\right) : \mu_1 \in \Re,\ \sigma_1^2 > 0,\ \mu_2 \in \Re,\ \sigma_2^2 > 0 \right\}$ which has dimension 4
and $\Omega_0 = \left\{ \left(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2\right) : \mu_1 = \mu_2,\ \sigma_1^2 = \sigma_2^2,\ \mu_1 \in \Re,\ \sigma_1^2 > 0,\ \mu_2 \in \Re,\ \sigma_2^2 > 0 \right\}$ which has dimension 2.

To measure the evidence against H_0 based on the observed data we use a test statistic or discrepancy measure.

8.1.3 Definition - Test Statistic or Discrepancy Measure

A test statistic or discrepancy measure D is a function of the data $\mathbf{X}$ that is constructed to measure the degree of "agreement" between the data $\mathbf{X}$ and the null hypothesis H_0.

A test statistic is usually chosen so that a small observed value of the test statistic indicates close agreement between the observed data and the null hypothesis H_0 while a large observed value of the test statistic indicates poor agreement. The test statistic is chosen before the data are examined and the choice reflects the type of departure from the null hypothesis H_0 that we wish to detect as specified by the alternative hypothesis H_A. A general method for constructing test statistics can be based on the likelihood function as we will see in the next two sections.

8.1.4 Example

For Example 8.1.2(a) suggest a test statistic which could be used if the alternative hypothesis is $H_A : \theta \neq \theta_0$. Suggest a test statistic which could be used if the alternative hypothesis is $H_A : \theta > \theta_0$ and if the alternative hypothesis is $H_A : \theta < \theta_0$.

Solution

If $H_0 : \theta = \theta_0$ is true then $E\left(\bar{X}\right) = \theta_0$. If the alternative hypothesis is $H_A : \theta \neq \theta_0$ then a reasonable test statistic which could be used is $D = \left|\bar{X} - \theta_0\right|$.

If the alternative hypothesis is $H_A : \theta > \theta_0$ then a reasonable test statistic which could be used is $D = \bar{X} - \theta_0$.

If the alternative hypothesis is $H_A : \theta < \theta_0$ then a reasonable test statistic which could be used is $D = \theta_0 - \bar{X}$.

After the data have been collected the observed value of the test statistic is calculated. Assuming the null hypothesis H_0 is true we compute the probability of observing a value of the test statistic at least as great as that observed. This probability is called the *p-value* of the data in relation to the null hypothesis H_0.

8.1.5 Definition - *p-value*

Suppose we use the test statistic $D = D\left(\mathbf{X}\right)$ to test the null hypothesis H_0. Suppose also that $d = D\left(\mathbf{x}\right)$ is the observed value of D. The *p-value* or *observed significance level* of the test of hypothesis H_0 using test statistic D is

$$p\text{-}value = P(D \geq d; H_0)$$

The *p-value* is the probability of observing such poor agreement using test statistic D between the null hypothesis H_0 and the data if the null hypothesis H_0 is true. If the *p-value* is very small, then such poor agreement would occur very rarely if the null hypothesis H_0 is true, and we interpret this to mean that the observed data are providing evidence against the null hypothesis H_0. The smaller the *p-value* the stronger the evidence against the null hypothesis H_0 based on the observed data. A large *p-value* does not mean that the null hypothesis H_0 is true but only indicates a lack of evidence against the null hypothesis H_0 based on the observed data and using the test statistic D.

The following table gives a rough guideline for interpreting *p-values*. **These are only guidelines**. The interpretation of *p-values* must always be made in the context of a given study.

Table 10.1: Guidelines for interpreting *p-values*

p-value	Interpretation
p-value > 0.10	No evidence against H_0 based on the observed data.
$0.05 <$ *p-value* ≤ 0.10	Weak evidence against H_0 based on the observed data.
$0.01 <$ *p-value* ≤ 0.05	Evidence against H_0 based on the observed data.
$0.001 <$ *p-value* ≤ 0.01	Strong evidence against H_0 based on the observed data.
p-value ≤ 0.001	Very strong evidence against H_0 based on the observed data.

8.1.6 Example

For Example 8.1.4 suppose $\bar{x} = 5.7$, $n = 25$ and $\theta_0 = 5$. Determine the *p-value* for both $H_A : \theta \neq \theta_0$ and $H_A : \theta > \theta_0$. Give a conclusion in each case.

Solution

For $\bar{x} = 5.7$, $n = 25$, $\theta_0 = 5$, and $H_A : \theta \neq 5$ the observed value of the test statistic is $d = |5.7 - 5| = 0.7$, and

$$\begin{aligned}
\textit{p-value} &= P\left(\left|\bar{X} - 5\right| \geq 0.7; H_0 : \theta = 5\right) \\
&= P\left(|T - 125| \geq 17.5\right) \quad \text{where } T = \sum_{i=1}^{25} X_i \sim \text{Poisson}\,(125) \\
&= P\left(T \leq 107.5\right) + P\left(T \geq 142.5\right) \\
&= P\left(T \leq 107\right) + P\left(T \geq 143\right) \\
&= 0.05605429 + 0.06113746 \\
&= 0.1171917
\end{aligned}$$

calculated using R. Since *p-value* > 0.1 there is no evidence against $H_0 : \theta = 5$ based on the data.

For $\bar{x} = 5.7$, $n = 25$, $\theta_0 = 5$, and $H_A : \theta > 5$ the observed value of the test statistic is

$d = 5.7 - 5 = 0.7$, and

$$\begin{aligned} p\text{-}value &= P\left(\bar{X} - 5 \geq 0.7; H_0 : \theta = 5\right) \\ &= P\left(T - 125 \geq 17.5\right) \quad \text{where } T = \sum_{i=1}^{25} X_i \sim \text{Poisson}\,(125) \\ &= P\left(T \geq 143\right) \\ &= 0.06113746 \end{aligned}$$

Since $0.05 <$ *p-value* ≤ 0.1 there is weak evidence against $H_0 : \theta = 5$ based on the data.

8.1.7 Exercise

Suppose in a Binomial experiment 42 successes have been observed in 100 trials and the hypothesis of interest is $H_0 : \theta = 0.5$.

(a) If the alternative hypothesis is $H_A : \theta \neq 0.5$, suggest a suitable test statistic, calculate the *p-value* and give a conclusion.

(b) If the alternative hypothesis is $H_A : \theta < 0.5$, suggest a suitable test statistic, calculate the *p-value* and give a conclusion.

8.2 Likelihood Ratio Tests for Simple Hypotheses

In Examples 8.1.4 and 8.1.7 it was reasonably straightforward to suggest a test statistic which made sense. In this section we consider a general method for constructing a test statistic which has good properties in the case of a simple hypothesis. The test statistic we use is the *likelihood ratio test statistic* which was introduced in your previous statistics course.

Suppose $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ is a random sample from $f(x;\theta)$ where $\boldsymbol{\theta} \in \Omega$ and the dimension of Ω is k. Suppose also that the hypothesis of interest is $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ where the elements of $\boldsymbol{\theta}_0$ are completely specified. H_0 can also be written as $H_0 : \boldsymbol{\theta} \in \Omega_0$ where Ω_0 consists of the single point $\boldsymbol{\theta}_0$. The dimension of Ω_0 is zero. H_0 is a simple hypothesis since the model and all the parameters are completely specified. The likelihood ratio test statistic for this simple hypothesis is

$$\begin{aligned} \Lambda\left(\mathbf{X}; \boldsymbol{\theta}_0\right) &= -2 \log R\left(\boldsymbol{\theta}_0; \mathbf{X}\right) \\ &= -2 \log \left[\frac{L\left(\boldsymbol{\theta}_0; \mathbf{X}\right)}{L(\tilde{\boldsymbol{\theta}}; \mathbf{X})}\right] \\ &= 2\left[l(\tilde{\boldsymbol{\theta}}; \mathbf{X}) - l\left(\boldsymbol{\theta}_0; \mathbf{X}\right)\right] \end{aligned}$$

where $\tilde{\boldsymbol{\theta}} = \tilde{\boldsymbol{\theta}}\left(\mathbf{X}\right)$ is the maximum likelihood estimator of $\boldsymbol{\theta}$. Note that this test statistic implicitly assumes that the alternative hypothesis is $H_A : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ or $H_A : \boldsymbol{\theta} \notin \Omega_0$.

Let the observed value of the likelihood ratio test statistic be

$$\lambda(\mathbf{x};\boldsymbol{\theta}_0) = 2\left[l(\hat{\boldsymbol{\theta}};\mathbf{x}) - l\left(\boldsymbol{\theta}_0;\mathbf{x}\right)\right]$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ are the observed data. The *p-value* is

$$p\text{-}value = P\left[\Lambda\left(\mathbf{X};\boldsymbol{\theta}_0\right) \geq \lambda(\mathbf{x};\boldsymbol{\theta}_0); H_0\right]$$

Note that the *p-value* is calculated assuming $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ is true. In general this *p-value* is difficult to determine exactly since the distribution of the random variable $\Lambda\left(\mathbf{X};\boldsymbol{\theta}_0\right)$ is usually intractable. We use the result from Theorem 7.3.4 which says that under certain (regularity) conditions

$$-2\log R(\boldsymbol{\theta};\mathbf{X}_n) = 2[l(\tilde{\boldsymbol{\theta}}_n;\mathbf{X}_n) - l(\boldsymbol{\theta};\mathbf{X}_n)] \rightarrow_D W \sim \chi^2(k) \tag{8.1}$$

for each $\boldsymbol{\theta} \in \Omega$ where $\mathbf{X}_n = (X_1, X_2, \ldots, X_n)$ and $\tilde{\boldsymbol{\theta}}_n = \tilde{\boldsymbol{\theta}}_n\left(X_1, X_2, \ldots, X_n\right)$.

Therefore based on the asymptotic result (8.1) and assuming $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ is true, the *p-value* for testing $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ using the likelihood ratio test statistic can be approximated using

$$p\text{-}value \approx P\left[W \geq \lambda(\mathbf{x};\boldsymbol{\theta}_0)\right] \quad \text{where } W \sim \chi^2(k)$$

8.2.1 Example

Suppose $X_1, X_2, \ldots, X_n$ is a random sample from the $\mathrm{N}\big(\mu, \sigma^2\big)$ distribution where σ^2 is known. Show that, in this special case, the likelihood ratio test statistic for testing $H_0 : \mu = \mu_0$ has exactly a $\chi^2(1)$ distribution.

Solution

From Example 7.1.8 we have that the likelihood function of μ is

$$L(\mu) = \sigma^{-n}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \bar{x})^2\right\}\exp\left[-\frac{n(\bar{x}-\mu)^2}{2\sigma^2}\right] \quad \text{for } \mu \in \Re$$

or more simply

$$L(\mu) = \exp\left[-\frac{n(\bar{x}-\mu)^2}{2\sigma^2}\right] \quad \text{for } \mu \in \Re$$

The corresponding log likelihood function is

$$l(\mu) = -\frac{n(\bar{x}-\mu)^2}{2\sigma^2} \quad \text{for } \mu \in \Re$$

Solving

$$\frac{dl}{d\mu} = \frac{n(\bar{x}-\mu)}{\sigma^2} = 0$$

gives $\mu = \bar{x}$. Since $l(\mu)$ is a quadratic function which is concave down we know that $\hat{\mu} = \bar{x}$ is the maximum likelihood estimate. The corresponding maximum likelihood estimator of μ is

$$\tilde{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Since $L(\hat{\mu}) = 1$, the relative likelihood function is

$$\begin{aligned} R(\mu) &= \frac{L(\mu)}{L(\hat{\mu})} \\ &= \exp\left[-\frac{n(\bar{x}-\mu)^2}{2\sigma^2}\right] \quad \text{for } \mu \in \Re \end{aligned}$$

The likelihood ratio test statistic for testing the hypothesis $H_0 : \mu = \mu_0$ is

$$\begin{aligned} \Lambda(\mu_0; \mathbf{X}) &= -2\log\left[\frac{L(\mu_0; \mathbf{X})}{L(\tilde{\mu}; \mathbf{X})}\right] \\ &= -2\log\left\{\exp\left[-\frac{n(\bar{X}-\mu_0)^2}{2\sigma^2}\right]\right\} \quad \text{since } \tilde{\mu} = \bar{X} \\ &= \frac{n(\bar{X}-\mu_0)^2}{\sigma^2} \\ &= \left(\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}\right)^2 \end{aligned}$$

If $H_0 : \mu = \mu_0$ is true then $\bar{X} \sim \mathrm{N}(\mu_0, \sigma^2)$ and

$$\left(\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}\right)^2 \sim \chi^2(1)$$

as required.

8.2.2 Example

Suppose $X_1, X_2, \ldots, X_n$ is a random sample from the Poisson(θ) distribution.
(a) Find the likelihood ratio test statistic for testing $H_0 : \theta = \theta_0$. Verify that the likelihood ratio statistic takes on large values if $\tilde{\theta} > \theta_0$ or $\tilde{\theta} < \theta_0$.
(b) Suppose $\bar{x} = 6$ and $n = 25$. Use the likelihood ratio test statistic to test $H_0 : \theta = 5$. Compare this with the test in Example 8.1.6.

Solution

(a) From Example 6.2.5 we have the likelihood function

$$L(\theta) = \theta^{n\bar{x}} e^{-n\theta} \quad \text{for } \theta \geq 0$$

and maximum likelihood estimate $\hat{\theta} = \bar{x}$. The relative likelihood function can be written as

$$R(\theta) = \frac{L(\theta)}{L(\hat{\theta})} = \left(\frac{\theta}{\hat{\theta}}\right)^{n\hat{\theta}} e^{n(\hat{\theta}-\theta)} \quad \text{for } \theta \geq 0$$

The likelihood ratio test statistic for $H_0 : \theta = \theta_0$ is

$$\begin{aligned}
\Lambda(\theta_0; \mathbf{X}) &= -2\log R(\theta_0; \mathbf{X}) \\
&= -2\log\left[\left(\frac{\theta_0}{\tilde{\theta}}\right)^{n\tilde{\theta}} e^{n(\tilde{\theta}-\theta_0)}\right] \\
&= 2n\left[-\tilde{\theta}\log\left(\frac{\theta_0}{\tilde{\theta}}\right) + \left(\theta_0 - \tilde{\theta}\right)\right] \\
&= 2n\tilde{\theta}\left[\left(\frac{\theta_0}{\tilde{\theta}} - 1\right) - \log\left(\frac{\theta_0}{\tilde{\theta}}\right)\right] \qquad (8.2)
\end{aligned}$$

To verify that the likelihood ratio statistic takes on large values if $\tilde{\theta} > \theta_0$ or $\tilde{\theta} < \theta_0$ or equivalently if $\frac{\theta_0}{\tilde{\theta}} < 1$ or $\frac{\theta_0}{\tilde{\theta}} > 1$, consider the function

$$g(t) = a\left[(t-1) - \log(t)\right] \quad \text{for } t > 0 \text{ and } a > 0 \qquad (8.3)$$

We note that $g(t) \to \infty$ as $t \to 0^+$ and $t \to \infty$. Now

$$g'(t) = a\left(1 - \frac{1}{t}\right) = a\left(\frac{t-1}{t}\right) \quad \text{for } t > 0 \text{ and } a > 0$$

Since $g'(t) < 0$ for $0 < t < 1$, and $g'(t) > 0$ for $t > 1$ we can conclude that the function $g(t)$ is a decreasing function for $0 < t < 1$ and an increasing function for $t > 1$ with an absolute minimum at $t = 1$. Since $g(1) = 0$, $g(t)$ is positive for all $t > 0$ and $t \neq 0$.

Therefore if we let $t = \frac{\theta_0}{\tilde{\theta}}$ in (8.2) then we see that $\Lambda(\theta_0; \mathbf{X})$ will be large for small values of $t = \frac{\theta_0}{\tilde{\theta}} < 1$ or large values of $t = \frac{\theta_0}{\tilde{\theta}} > 1$.

(b) If $\bar{x} = 6$, $n = 25$, and $H_0 : \theta = 5$ then the observed value of the likelihood ratio test statistic is

$$\begin{aligned}
\lambda(5; \mathbf{x}) &= -2\log R(\theta_0; \mathbf{X}) \\
&= -2\log\left[\left(\frac{5}{6}\right)^{25(5.6)} e^{25(6-5)}\right] \\
&= 4.6965
\end{aligned}$$

The parameter space is $\Omega = \{\theta : \theta > 0\}$ which has dimension 1 and thus $k = 1$. The approximate *p-value* is

$$\begin{aligned}
p\text{-}value &\approx P(W \geq 4.6965) \quad \text{where } W \sim \chi^2(1) \\
&= 2\left[1 - P\left(Z \leq \sqrt{4.6965}\right)\right] \quad \text{where } Z \sim \mathrm{N}(0,1) \\
&= 0.0302
\end{aligned}$$

calculated using R. Since $0.01 < p\text{-}value \leq 0.05$ there is evidence against $H_0 : \theta = 5$ based on the data. Compared with the answer in Example 8.1.6 for $H_A : \theta \neq 5$ we note that the *p-values* are slightly different by the conclusion is the same.

8.2.3 Example

Suppose $X_1, X_2, \ldots, X_n$ is a random sample from the Exponential(θ) distribution.
(a) Find the likelihood ratio test statistic for testing $H_0 : \theta = \theta_0$. Verify that the likelihood ratio statistic takes on large values if $\tilde{\theta} > \theta_0$ or $\tilde{\theta} < \theta_0$.
(b) Suppose $\bar{x} = 6$ and $n = 25$. Use the likelihood ratio test statistic to test $H_0 : \theta = 5$.
(c) From Example 6.6.3 we have

$$Q(\mathbf{X}; \theta) = \frac{2\sum\limits_{i=1}^{n} X_i}{\theta} = \frac{2n\tilde{\theta}}{\theta} \sim \chi^2(2n)$$

is a pivotal quantity. Explain how this pivotal quantity could be used to test $H_0 : \theta = \theta_0$ if (i) $H_A : \theta < \theta_0$, (ii) $H_A : \theta > \theta_0$, and (iii) $H_A : \theta \neq \theta_0$.
(d) Suppose $\bar{x} = 6$ and $n = 25$. Use the test statistic from (c) for $H_A : \theta \neq \theta_0$ to test $H_0 : \theta = 5$. Compare the answer with the answer in (b).

Solution

(a) From Example 6.2.8 we have the likelihood function

$$L(\theta) = \theta^{-n} e^{-n\bar{x}/\theta} \quad \text{for } \theta > 0$$

and maximum likelihood estimate $\hat{\theta} = \bar{x}$. The relative likelihood function can be written as

$$R(\theta) = \frac{L(\theta)}{L(\hat{\theta})} = \left(\frac{\hat{\theta}}{\theta}\right)^n e^{n(1-\hat{\theta}/\theta)} \quad \text{for } \theta \geq 0$$

The likelihood ratio test statistic for $H_0 : \theta = \theta_0$ is

$$\begin{aligned}
\Lambda(\theta_0; \mathbf{X}) &= -2\log R(\theta_0; \mathbf{X}) \\
&= -2\log\left[\left(\frac{\tilde{\theta}}{\theta}\right)^n e^{n(1-\tilde{\theta}/\theta)}\right] \\
&= 2n\left[\left(\frac{\tilde{\theta}}{\theta_0} - 1\right) - \log\left(\frac{\tilde{\theta}}{\theta_0}\right)\right]
\end{aligned}$$

To verify that the likelihood ratio statistic takes on large values if $\tilde{\theta} > \theta_0$ or $\tilde{\theta} < \theta_0$ or equivalently if $\frac{\tilde{\theta}}{\theta_0} < 1$ and $\frac{\tilde{\theta}}{\theta_0} > 1$ we note that $\Lambda(\theta_0; \mathbf{X})$ is of the form 8.3 so an argument similar to Example 8.2.2(a) can be used with $t = \frac{\tilde{\theta}}{\theta_0}$.
(b) If $\bar{x} = 6$, $n = 25$, and $H_0 : \theta = 5$ then the observed value of the likelihood ratio test statistic is

$$\begin{aligned}
\lambda(5; \mathbf{x}) &= -2\log R(\theta_0; \mathbf{X}) \\
&= -2\log\left[\left(\frac{6}{5}\right)^{25} e^{25(1-6)/5}\right] \\
&= 0.8839222
\end{aligned}$$

The parameter space is $\Omega = \{\theta : \theta > 0\}$ which has dimension 1 and thus $k = 1$. The approximate *p-value* is

$$\begin{aligned} p\text{-}value &\approx P\left(W \geq 0.8839222\right) \quad \text{where } W \sim \chi^2\left(1\right) \\ &= 2\left[1 - P\left(Z \leq \sqrt{0.8839222}\right)\right] \quad \text{where } Z \sim \mathrm{N}\left(0, 1\right) \\ &= 0.3471 \end{aligned}$$

calculated using R. Since $p\text{-}value > 0.1$ there is no evidence against $H_0 : \theta = 5$ based on the data.

(c) (i) If $H_A : \theta > \theta_0$ we could let $D = \frac{\tilde{\theta}}{\theta_0}$. If $H_0 : \theta = \theta_0$ is true then since $E(\tilde{\theta}) = \theta_0$ we would expect observed values of $D = \frac{\tilde{\theta}}{\theta_0}$ to be close to 1. However if $H_A : \theta > \theta_0$ is true then $E(\tilde{\theta}) = \theta > \theta_0$ and we would expect observed values of $D = \frac{\tilde{\theta}}{\theta_0}$ to be larger than 1 and therefore large values of D provide evidence against $H_0 : \theta = \theta_0$. The corresponding *p-value* would be

$$\begin{aligned} p\text{-}value &= P\left(\frac{\tilde{\theta}}{\theta_0} \geq \frac{\hat{\theta}}{\theta_0}; H_0\right) \\ &= P\left(W \geq \frac{2n\hat{\theta}}{\theta_0}\right) \quad \text{where } W \sim \chi^2\left(2n\right) \end{aligned}$$

(ii) If $H_A : \theta < \theta_0$ we could still let $D = \frac{\tilde{\theta}}{\theta_0}$. If $H_0 : \theta = \theta_0$ is true then since $E(\tilde{\theta}) = \theta_0$ we would expect observed values of $D = \frac{\tilde{\theta}}{\theta_0}$ to be close to 1. However if $H_A : \theta < \theta_0$ is true then $E(\tilde{\theta}) = \theta < \theta_0$ and we would expect observed values of $D = \frac{\tilde{\theta}}{\theta_0}$ to be smaller than 1 and therefore small values of D provide evidence against $H_0 : \theta = \theta_0$. The corresponding *p-value* would be

$$\begin{aligned} p\text{-}value &= P\left(\frac{\tilde{\theta}}{\theta_0} \leq \frac{\hat{\theta}}{\theta_0}; H_0\right) \\ &= P\left(W \leq \frac{2n\hat{\theta}}{\theta_0}\right) \quad \text{where } W \sim \chi^2\left(2n\right) \end{aligned}$$

(iii) If $H_A : \theta \neq \theta_0$ we could still let $D = \frac{\tilde{\theta}}{\theta_0}$. If $H_0 : \theta = \theta_0$ is true then since $E(\tilde{\theta}) = \theta_0$ we would expect observed values of $D = \frac{\tilde{\theta}}{\theta_0}$ to be close to 1. However if $H_A : \theta \neq \theta_0$ is true then $E(\tilde{\theta}) = \theta \neq \theta_0$ and we would expect observed values of $D = \frac{\tilde{\theta}}{\theta_0}$ to be either larger or smaller than 1 and therefore both large and small values of D provide evidence against $H_0 : \theta = \theta_0$. If a large (small) value of D is observed it is not simple to determine exactly which small (large) values should also be considered. Since we are not that concerned about the exact *p-value*, the *p-value* is usually calculated more simply as

$$p\text{-}value = \min\left(2P\left(W \leq \frac{2n\hat{\theta}}{\theta_0}\right), 2P\left(W \geq \frac{2n\hat{\theta}}{\theta_0}\right)\right) \quad \text{where } W \sim \chi^2\left(2n\right)$$

(d) If $\bar{x} = 6$, $n = 25$, and $H_0 : \theta = 5$ then the observed value of the $D = \frac{\tilde{\theta}}{\theta_0}$ is $d = \frac{6}{5}$ with

$$\begin{aligned} p\text{-}value &= \min\left(2P\left(W \leq \frac{(50)\,6}{5}\right), 2P\left(W \geq \frac{(50)\,6}{5}\right)\right) \quad \text{where } W \sim \chi^2(50) \\ &= \min(1.6855, 0.3145) \\ &= 0.3145 \end{aligned}$$

calculated using R. Since *p-value* > 0.1 there is no evidence against $H_0 : \theta = 5$ based on the data.

We notice the test statistic $D = \frac{\tilde{\theta}}{\theta_0}$ and $\Lambda(\theta_0; \mathbf{X}) = 2n\left[\left(\frac{\tilde{\theta}}{\theta_0} - 1\right) - \log\left(\frac{\tilde{\theta}}{\theta_0}\right)\right]$ are both functions of $\frac{\tilde{\theta}}{\theta_0}$. For this example the *p-values* are similar and the conclusions are the same.

8.2.4 Example

The following table gives the observed frequencies of the six faces in 100 rolls of a die:

Face: j	1	2	3	4	5	6	Total
Observed Frequency: x_j	16	15	14	20	22	13	100

Are these observations consistent with the hypothesis that the die is fair?

Solution

The model for these data is $(X_1, X_2, \ldots, X_6) \sim \text{Multinomial}(100, \theta_1, \theta_2, \ldots, \theta_6)$ and the hypothesis of interest is $H_0 : \theta_1 = \theta_2 = \cdots = \theta_6 = \frac{1}{6}$. Since the model and parameters are completely specified this is a simple hypothesis. Since $\sum_{j=1}^{6} \theta_j = 1$ there are really only $k = 5$ parameters. The relative likelihood function for $(\theta_1, \theta_2, \ldots, \theta_5)$ is

$$L(\theta_1, \theta_2, \ldots, \theta_5) = \frac{n!}{x_1! x_2! \cdots x_5! x_6!} \theta_1^{x_1} \theta_2^{x_2} \cdots \theta_5^{x_5} (1 - \theta_1 - \theta_2 - \cdots - \theta_5)^{x_6}$$

or more simply

$$L(\theta_1, \theta_2, \ldots, \theta_5) = \theta_1^{x_1} \theta_2^{x_2} \cdots \theta_5^{x_5} (1 - \theta_1 - \theta_2 - \cdots - \theta_5)^{x_6}$$

for $0 \leq \theta_j \leq 1$ for $j = 1, 2, \ldots, 5$ and $\sum_{j=1}^{5} \theta_j \leq 1$. The log likelihood function is

$$l(\theta_1, \theta_2, \ldots, \theta_5) = \sum_{j=1}^{5} x_j \log \theta_j + x_6 \log(1 - \theta_1 - \theta_2 - \cdots - \theta_5)$$

Now

$$\begin{aligned} \frac{\partial l}{\partial \theta_j} &= \frac{x_j}{\theta_j} - \frac{n - x_1 - x_2 - \cdots - x_5}{1 - \theta_1 - \theta_2 - \cdots - \theta_5} \quad \text{for } j = 1, 2, \ldots, 5 \\ &= \frac{x_j(1 - \theta_1 - \theta_2 - \cdots - \theta_5) - \theta_j(n - x_1 - x_2 - \cdots - x_5)}{\theta_j(1 - \theta_1 - \theta_2 - \cdots - \theta_5)} \end{aligned}$$

since $x_6 = n - x_1 - x_2 - \cdots - x_5$. We could solve $\frac{\partial l}{\partial \theta_j} = 0$ for $j = 1, 2, \ldots, 5$ simultaneously. In the Binomial case we know $\hat{\theta} = \frac{x}{n}$. It seems reasonable that the maximum likelihood estimate of θ_j is $\hat{\theta}_j = \frac{x_j}{n}$ for $j = 1, 2, \ldots, 5$. To verify this is true we substitute $\theta_j = \frac{x_j}{n}$ for $j = 1, 2, \ldots, 5$ into $\frac{\partial l}{\partial \theta_j}$ to obtain

$$x_j - \frac{x_j}{n} \sum_{i \neq j} x_i - x_j + \frac{x_j}{n} \sum_{i \neq j} x_i = 0$$

Therefore the maximum likelihood estimator of θ_j is $\tilde{\theta}_j = \frac{X_j}{n}$ for $j = 1, 2, \ldots, 5$. Note also that by the invariance property of maximum likelihood estimators $\tilde{\theta}_6 = 1 - \sum\limits_{j=1}^{5} \tilde{\theta}_j = \frac{X_6}{n}$. Therefore we can write

$$l(\tilde{\theta}_1, \tilde{\theta}_2, \tilde{\theta}_3, \tilde{\theta}_4, \tilde{\theta}_5; \mathbf{X}) = \sum_{i=1}^{6} X_j \log\left(\frac{X_j}{n}\right)$$

Since the null hypothesis is $H_0 : \theta_1 = \theta_2 = \cdots = \theta_6 = \frac{1}{6}$

$$l\left(\boldsymbol{\theta}_0; \mathbf{X}\right) = \sum_{i=1}^{6} X_j \log\left(\frac{1}{6}\right)$$

so the likelihood ratio test statistic is

$$\begin{aligned}
\Lambda\left(\mathbf{X}; \boldsymbol{\theta}_0\right) &= 2\left[l(\tilde{\boldsymbol{\theta}}; \mathbf{X}) - l\left(\boldsymbol{\theta}_0; \mathbf{X}\right)\right] \\
&= 2\left[\sum_{i=1}^{6} X_j \log\left(\frac{X_j}{n}\right) - \sum_{i=1}^{6} X_j \log\left(\frac{1}{6}\right)\right] \\
&= 2\sum_{i=1}^{6} X_j \log\left(\frac{X_j}{E_j}\right)
\end{aligned}$$

where $E_j = n/6$ is the expected frequency for outcome j. This test statistic is the likelihood ratio Goodness of Fit test statistic introduced in your previous statistics course.

For these data the observed value of the likelihood ratio test statistic is

$$\begin{aligned}
\lambda(\mathbf{x}; \boldsymbol{\theta}_0) &= 2\sum_{i=1}^{6} x_j \log\left(\frac{x_j}{100/6}\right) \\
&= 2\left[16 \log\left(\frac{16}{100/6}\right) + 15 \log\left(\frac{15}{100/6}\right) + \cdots + 13 \log\left(\frac{13}{100/6}\right)\right] \\
&= 3.699649
\end{aligned}$$

The approximate *p-value* is

$$\begin{aligned}
p\text{-}value &\approx P\left(W \geq 3.699649\right) \quad \text{where } W \sim \chi^2(5) \\
&= 0.5934162
\end{aligned}$$

calculated using R. Since *p-value*> 0.1 there is no evidence based on the data against the hypothesis of a fair die.

Note:

(1) In this example the data $(X_1, X_2, \ldots, X_6)$ are not a random sample. The conditions for (8.1) hold by thinking of the experiment as a sequence of n independent trials with 6 outcomes on each trial.

(2) You may recall from your previous statistics course that the χ^2 approximation is reasonable if the expected frequency E_j in each category is at least 5.

8.2.5 Exercise

In a long-term study of heart disease in a large group of men, it was noted that 63 men who had no previous record of heart problems died suddenly of heart attacks. The following table gives the number of such deaths recorded on each day of the week:

Day of Week	Mon.	Tues.	Wed.	Thurs.	Fri.	Sat.	Sun.
No. of Deaths	22	7	6	13	5	4	6

Test the hypothesis of interest that the deaths are equally likely to occur on any day of the week.

8.3 Likelihood Ratio Tests for Composite Hypotheses

Suppose $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ is a random sample from $f(x; \boldsymbol{\theta})$ where $\boldsymbol{\theta} \in \Omega$ and Ω has dimension k. Suppose we wish to test $H_0 : \boldsymbol{\theta} \in \Omega_0$ where Ω_0 is a subset of Ω of dimension q where $0 < q < k$. The hypothesis H_0 is a composite hypothesis since all the values of the unknown parameters are not specified. For testing composite hypotheses we use the *likelihood ratio test statistic*

$$\begin{aligned}\Lambda(\mathbf{X};\Omega_0) &= -2\log\left[\frac{\max\limits_{\boldsymbol{\theta}\in\Omega_0} L(\boldsymbol{\theta};\mathbf{X})}{\max\limits_{\boldsymbol{\theta}\in\Omega} L(\boldsymbol{\theta};\mathbf{X})}\right] \\ &= 2\left[l(\tilde{\boldsymbol{\theta}};\mathbf{X}) - \max_{\boldsymbol{\theta}\in\Omega_0} l(\boldsymbol{\theta};\mathbf{X})\right]\end{aligned}$$

where $\tilde{\boldsymbol{\theta}} = \tilde{\boldsymbol{\theta}}(X_1, X_2, \ldots, X_n)$ is the maximum likelihood estimator of $\boldsymbol{\theta}$. Note that this test statistic implicitly assumes that the alternative hypothesis is $H_A : \boldsymbol{\theta} \notin \Omega_0$.

Let the observed value of the likelihood ratio test statistic be

$$\lambda(\mathbf{x};\Omega_0) = 2\left[l(\hat{\boldsymbol{\theta}};\mathbf{x}) - \max_{\boldsymbol{\theta}\in\Omega_0} l(\boldsymbol{\theta};\mathbf{x})\right]$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ are the observed data. The *p-value* is

$$p\text{-}value = P\left[\Lambda(\mathbf{X};\Omega_0) \geq \lambda(\mathbf{x};\Omega_0); H_0\right]$$

Note that the *p-value* is calculated assuming $H_0 : \boldsymbol{\theta} \in \Omega_0$ is true. In general this *p-value* is difficult to determine exactly since the distribution of the random variable $\Lambda(\mathbf{X};\Omega_0)$ is usually intractable. Under certain (regularity) conditions it can be shown that, assuming the hypothesis $H_0 : \boldsymbol{\theta} \in \Omega_0$ is true,

$$\Lambda(\mathbf{X};\Omega_0) = 2\left[l(\tilde{\boldsymbol{\theta}}_n; \mathbf{X}_n) - \max_{\boldsymbol{\theta}\in\Omega_0} l(\boldsymbol{\theta}; \mathbf{X}_n)\right] \to_D W \sim \chi^2(k-q)$$

where $\mathbf{X}_n = (X_1, X_2, \ldots, X_n)$ and $\tilde{\boldsymbol{\theta}}_n = \tilde{\boldsymbol{\theta}}_n\,(X_1, X_2, \ldots, X_n)$. The approximate *p-value* is given by

$$p\text{-}value \approx P\left[W \geq \lambda(\mathbf{x};\Omega_0)\right]$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ are the observed data.

Note:

The number of degrees of freedom is the difference between the dimension of Ω and the dimension of Ω_0. The degrees of freedom can also be determined as the number of parameters estimated under the model with no restrictions minus the number of parameters estimated under the model with restrictions imposed by the null hypothesis H_0.

8.3.1 Example

(a) Suppose $X_1, X_2, \ldots, X_n$ is a random sample from the Gamma(α, β) distribution. Find the likelihood ratio test statistic for testing $H_0 : \alpha = \alpha_0$ where α is unknown. Indicate how to find the approximate *p-value*.

(b) For the data in Example 7.1.14 test the hypothesis $H_0 : \alpha = 2$.

Solution

(a) From Example 8.1.2(b) we have $\Omega = \{(\alpha, \beta) : \alpha > 0,\ \beta > 0\}$ which has dimension $k = 2$ and $\Omega_0 = \{(\alpha_0, \beta) : \beta > 0\}$ which has dimension $q = 1$ and the hypothesis is composite.

From Exercise 7.1.11 the likelihood function is

$$L(\alpha, \beta) = [\Gamma(\alpha)\,\beta^\alpha]^{-n}\left(\prod_{i=1}^n x_i\right)^\alpha \exp\left(-\frac{1}{\beta}\sum_{i=1}^n x_i\right) \quad \text{for } \alpha > 0,\ \beta > 0$$

where and the log likelihood function is

$$\begin{aligned} l(\alpha, \beta) &= \log L(\alpha, \beta) \\ &= -n\log\Gamma(\alpha) - n\alpha\log\beta + \alpha\sum_{i=1}^n \log x_i - \frac{1}{\beta}\sum_{i=1}^n x_i \quad \text{for } \alpha > 0,\ \beta > 0 \end{aligned}$$

The maximum likelihood estimators cannot be found explicitly so we write

$$l(\tilde{\alpha}, \tilde{\beta}; \mathbf{X}) = -n\log\Gamma(\tilde{\alpha}) - n\tilde{\alpha}\log\tilde{\beta} + \tilde{\alpha}\sum_{i=1}^n \log X_i - \frac{1}{\tilde{\beta}}\sum_{i=1}^n X_i$$

If $\alpha = \alpha_0$ then the log likelihood function is

$$l(\alpha_0, \beta) = -n\log\Gamma(\alpha_0) - n\alpha_0\log\beta + \alpha_0\sum_{i=1}^n \log x_i - \frac{\sum_{i=1}^n x_i}{\beta} \quad \text{for } \beta > 0$$

which is only a function of β. To determine $\max\limits_{(\alpha,\beta)\in\Omega_0} l(\alpha,\beta;\mathbf{X})$ we note that

$$\frac{d}{d\beta}l\left(\alpha_0,\beta\right)=\frac{-n\alpha_0}{\beta}+\frac{\sum\limits_{i=1}^{n}x_i}{\beta^2}$$

and $\frac{d}{d\beta}l\left(\alpha_0,\beta\right)=0$ for $\beta=\frac{1}{n\alpha_0}\sum\limits_{i=1}^{n}x_i=\frac{\bar{x}}{\alpha_0}$ and therefore

$$\max_{(\alpha,\beta)\in\Omega_0} l(\alpha,\beta;\mathbf{X})=-n\log\Gamma\left(\alpha_0\right)-n\alpha_0\log\left(\frac{\bar{X}}{\alpha_0}\right)+\alpha_0\sum_{i=1}^{n}\log X_i-n\alpha_0$$

The likelihood ratio test statistic is

$$\begin{aligned}
&\Lambda(\mathbf{X};\Omega_0)\\
=\;&2\left[l(\tilde{\alpha},\tilde{\beta};\mathbf{X})-\max_{(\alpha,\beta)\in\Omega_0} l(\alpha,\beta;\mathbf{X})\right]\\
=\;&2\Big[-n\log\Gamma\left(\tilde{\alpha}\right)-n\tilde{\alpha}\log\tilde{\beta}+\tilde{\alpha}\sum_{i=1}^{n}\log X_i-\frac{1}{\tilde{\beta}}\sum_{i=1}^{n}X_i\\
&+n\log\Gamma\left(\alpha_0\right)+n\alpha_0\log\left(\frac{\bar{X}}{\alpha_0}\right)-\alpha_0\sum_{i=1}^{n}\log X_i+n\alpha_0\Big]\\
=\;&2n\left\{\log\left[\frac{\Gamma\left(\alpha_0\right)}{\Gamma\left(\tilde{\alpha}\right)}\right]+\frac{\left(\tilde{\alpha}-\alpha_0\right)}{n}\sum_{i=1}^{n}\log X_i+\alpha_0\log\left(\frac{\bar{X}}{\alpha_0}\right)-\tilde{\alpha}\log\tilde{\beta}+\alpha_0-\frac{\bar{X}}{\tilde{\beta}}\right\}
\end{aligned}$$

with corresponding observed value

$$\begin{aligned}
&\lambda(\mathbf{x};\Omega_0)\\
=\;&2n\left\{\log\left[\frac{\Gamma\left(\alpha_0\right)}{\Gamma\left(\hat{\alpha}\right)}\right]+\frac{\left(\hat{\alpha}-\alpha_0\right)}{n}\sum_{i=1}^{n}\log x_i+\alpha_0\log\left(\frac{\bar{x}}{\alpha_0}\right)-\hat{\alpha}\log\hat{\beta}+\alpha_0-\frac{\bar{x}}{\hat{\beta}}\right\}
\end{aligned}$$

Since $k-q=2-1=1$

$$\begin{aligned}
p\text{-}value &\approx P\left[W\geq\lambda(\mathbf{x};\Omega_0)\right]\quad\text{where } W\sim\chi^2(1)\\
&= 2\left[1-P\left(Z\leq\sqrt{\lambda(\mathbf{x};\Omega_0)}\right)\right]\quad\text{where } Z\sim\mathrm{N}\left(0,1\right)
\end{aligned}$$

The degrees of freedom can also be determined by noticing that, under the full model two parameters (α and β) were estimated, and under the null hypothesis $H_0:\alpha=\alpha_0$ only one parameter (β) was estimated. Therefore $2-1=1$ are the degrees of freedom.

(b) For $H_0:\alpha=2$ and the data in Example 7.1.14 we have $n=30$, $\bar{x}=6.824333$, $\frac{1}{n}\sum\limits_{i=1}^{n}\log x_i=1.794204$, $\hat{\alpha}=4.118407$, $\hat{\beta}=1.657032$. The observed value of the likelihood ratio test statistic is $\lambda(\mathbf{x};\Omega_0)=6.886146$ with

$$\begin{aligned}
p\text{-}value &\approx P\left(W\geq 6.886146\right)\quad\text{where } W\sim\chi^2(1)\\
&= 2\left[1-P\left(Z\leq\sqrt{6.886146}\right)\right]\quad\text{where } Z\sim\mathrm{N}\left(0,1\right)\\
&= 0.008686636
\end{aligned}$$

calculated using R. Since $p\text{-}value\leq 0.01$ there is strong evidence against $H_0:\alpha=2$ based on the data.

8.3.2 Example

(a) Suppose $X_1, X_2, \ldots, X_n$ is a random sample from the Exponential(θ_1) distribution and independently $Y_1, Y_2, \ldots, Y_m$ is a random sample from the Exponential(θ_2) distribution. Find the likelihood ratio test statistic for testing $H_0 : \theta_1 = \theta_2$. Indicate how to find the approximate *p-value.*

(b) Find the approximate *p-value* if the observed data are $n = 10$, $\sum\limits_{i=1}^{10} x_i = 22$, $m = 15$, $\sum\limits_{i=1}^{15} y_i = 40$. What would you conclude?

Solution

(a) From Example 8.1.2(c) we have $\Omega = \{(\theta_1, \theta_2) : \theta_1 > 0,\ \theta_2 > 0\}$ which has dimension $k = 2$ and $\Omega_0 = \{(\theta_1, \theta_2) : \theta_1 = \theta_2,\ \theta_1 > 0,\ \theta_2 > 0\}$ which has dimension $q = 1$ and the hypothesis is composite.

From Example 6.2.8 the likelihood function for an observed random sample $x_1, x_2, \ldots, x_n$ from an Exponential(θ_1) distribution is

$$L_1(\theta_1) = \theta_1^{-n} e^{-n\bar{x}/\theta_1} \quad \text{for } \theta_1 > 0$$

with maximum likelihood estimate $\hat{\theta}_1 = \bar{x}$.

Similarly the likelihood function for an observed random sample $y_1, y_2, \ldots, y_m$ from an Exponential(θ_2) distribution is

$$L_2(\theta_2) = \theta_2^{-m} e^{-m\bar{y}/\theta_2} \quad \text{for } \theta_2 > 0$$

with maximum likelihood estimate $\hat{\theta}_2 = \bar{y}$. Since the samples are independent the likelihood function for (θ_1, θ_2) is

$$L(\theta_1, \theta_2) = L_1(\theta_1) L_2(\theta_2) \quad \text{for } \theta_1 > 0, \theta_2 > 0$$

and the log likelihood function

$$l(\theta_1, \theta_2) = -n \log \theta_1 - \frac{n\bar{x}}{\theta_1} - m \log \theta_2 - \frac{m\bar{y}}{\theta_2} \quad \text{for } \theta_1 > 0, \theta_2 > 0$$

The independence of the samples also implies the maximum likelihood estimators are still $\tilde{\theta}_1 = \bar{X}$ and $\tilde{\theta}_2 = \bar{Y}$. Therefore

$$l(\tilde{\theta}_1, \tilde{\theta}_2; \mathbf{X}, \mathbf{Y}) = -n \log \bar{X} - m \log \bar{Y} - (n + m)$$

If $\theta_1 = \theta_2 = \theta$ then the log likelihood function is

$$l(\theta) = -(n + m) \log \theta - \frac{(n\bar{x} + m\bar{y})}{\theta} \quad \text{for } \theta > 0$$

which is only a function of θ. To determine $\max\limits_{(\theta_1,\theta_2)\in\Omega_0} l(\theta_1,\theta_2;\mathbf{X},\mathbf{Y})$ we note that

$$\frac{d}{d\theta}l(\theta) = \frac{-(n+m)}{\theta} + \frac{(n\bar{x}+m\bar{y})}{\theta^2}$$

and $\frac{d}{d\theta}l(\theta) = 0$ for $\theta = \frac{n\bar{x}+m\bar{y}}{n+m}$ and therefore

$$\max_{(\theta_1,\theta_2)\in\Omega_0} l(\theta_1,\theta_2;\mathbf{X},\mathbf{Y}) = -(n+m)\log\left(\frac{n\bar{X}+m\bar{Y}}{n+m}\right) - (n+m)$$

The likelihood ratio test statistic is

$$\begin{aligned}
&\Lambda(\mathbf{X},\mathbf{Y};\Omega_0) \\
= \;& 2\left[l(\tilde{\theta}_1,\tilde{\theta}_2;\mathbf{X},\mathbf{Y}) - \max_{(\theta_1,\theta_2)\in\Omega_0} l(\theta_1,\theta_2;\mathbf{X},\mathbf{Y})\right] \\
= \;& 2\left[-n\log\bar{X} - m\log\bar{Y} - (n+m) + (n+m)\log\left(\frac{n\bar{X}+m\bar{Y}}{n+m}\right) + (n+m)\right] \\
= \;& 2\left[(n+m)\log\left(\frac{n\bar{X}+m\bar{Y}}{n+m}\right) - n\log\bar{X} - m\log\bar{Y}\right]
\end{aligned}$$

with corresponding observed value

$$\lambda(\mathbf{x},\mathbf{y};\Omega_0) = 2\left[(n+m)\log\left(\frac{n\bar{x}+m\bar{y}}{n+m}\right) - n\log\bar{x} - m\log\bar{y}\right]$$

Since $k-q = 2-1 = 1$

$$\begin{aligned}
p\text{-}value &\approx P\left[W \geq \lambda(\mathbf{x},\mathbf{y};\Omega_0)\right] \quad \text{where } W \sim \chi^2(1) \\
&= 2\left[1 - P\left(Z \leq \sqrt{\lambda(\mathbf{x},\mathbf{y};\Omega_0)}\right)\right] \quad \text{where } Z \sim \mathrm{N}(0,1)
\end{aligned}$$

(b) For $n = 10$, $\sum\limits_{i=1}^{10} x_i = 22$, $m = 15$, $\sum\limits_{i=1}^{15} y_i = 40$ the observed value of the likelihood ratio test statistic is $\lambda(\mathbf{x},\mathbf{y};\Omega_0) = 0.2189032$ and

$$\begin{aligned}
p\text{-}value &\approx P(W \geq 0.2189032) \quad \text{where } W \sim \chi^2(1) \\
&= 2\left[1 - P\left(Z \leq \sqrt{0.2189032}\right)\right] \quad \text{where } Z \sim \mathrm{N}(0,1) \\
&= 0.6398769
\end{aligned}$$

calculated using R. Since $p\text{-}value > 0.5$ there is no evidence against $H_0: \theta_1 = \theta_2$ based on the observed data.

8.3.3 Exercise

(a) Suppose $X_1, X_2, \ldots, X_n$ is a random sample from the Poisson(θ_1) distribution and independently $Y_1, Y_2, \ldots, Y_m$ is a random sample from the Poisson(θ_2) distribution. Find the likelihood ratio test statistic for testing $H_0 : \theta_1 = \theta_2$. Indicate how to find the approximate *p-value*.

(b) Find the approximate *p-value* if the observed data are $n = 10$, $\sum\limits_{i=1}^{10} x_i = 22$, $m = 15$, $\sum\limits_{i=1}^{15} y_i = 40$. What would you conclude?

8.3.4 Example

In a large population of males ages $40 - 50$, the proportion who are regular smokers is α where $0 \leq \alpha \leq 1$ and the proportion who have hypertension (high blood pressure) is β where $0 \leq \beta \leq 1$. Suppose that n men are selected at random from this population and the observed data are

Category	SH	$S\bar{H}$	$\bar{S}H$	$\bar{S}\bar{H}$
Frequency	x_{11}	x_{12}	x_{21}	x_{22}

where S is the event the male is a smoker and H is the event the male has hypertension. Find the likelihood ratio test statistic for testing H_0 : events S and H are independent. Indicate how to find the approximate *p-value*.

Solution

The model for these data is $(X_1, X_2, \ldots, X_6) \sim \text{Multinomial}(100, \theta_{11}, \theta_{12}, \theta_{21}, \theta_{22})$ with parameter space

$$\Omega = \left\{ (\theta_{11}, \theta_{12}, \theta_{21}, \theta_{22}) : 0 \leq \theta_{ij} \leq 1 \text{ for } i, j = 1, 2 \ \sum_{j=1}^{2} \sum_{i=1}^{2} \theta_{ij} \leq 1 \right\}$$

which has dimension $k = 3$.

Let $P(S) = \alpha$ and $P(H) = \beta$ then the hypothesis of interest can be written as $H_0 : \theta_{11} = \alpha\beta$, $\theta_{12} = \alpha(1-\beta)$, $\theta_{21} = (1-\alpha)\beta$, $\theta_{22} = (1-\alpha)(1-\beta)$ and

$$\begin{aligned} \Omega_0 &= \{(\theta_{11}, \theta_{12}, \theta_{21}, \theta_{22}) : \theta_{11} = \alpha\beta,\ \theta_{12} = \alpha(1-\beta), \\ \theta_{21} &= (1-\alpha)\beta,\ \theta_{22} = (1-\alpha)(1-\beta),\ 0 \leq \alpha \leq 1,\ 0 \leq \beta \leq 1\} \end{aligned}$$

which has dimension $q = 2$ and the hypothesis is composite.

From Example 8.2.4 we can see that the relative likelihood function for $(\theta_{11}, \theta_{12}, \theta_{21}, \theta_{22})$ is

$$L(\theta_{11}, \theta_{12}, \theta_{21}, \theta_{22}) = \theta_{11}^{x_{11}} \theta_{12}^{x_{12}} \theta_{21}^{x_{21}} \theta_{22}^{x_{22}}$$

The log likelihood function is

$$l(\theta_{11}, \theta_{12}, \theta_{21}, \theta_{22}) = \sum_{j=1}^{2} \sum_{i=1}^{2} x_{ij} \log \theta_{ij}$$

and the maximum likelihood estimate of θ_{ij} is $\hat{\theta}_{ij} = \frac{x_{ij}}{n}$ for $i, j = 1, 2$. Therefore

$$l(\tilde{\theta}_{11}, \tilde{\theta}_{12}, \tilde{\theta}_{21}, \tilde{\theta}_{22}; \mathbf{X}) = \sum_{j=1}^{2} \sum_{i=1}^{2} X_{ij} \log\left(\frac{X_{ij}}{n}\right)$$

If the events S and H are independent events then from Chapter 7, Problem 4 we have that the likelihood function is

$$\begin{aligned} L(\alpha, \beta) &= (\alpha\beta)^{x_{11}} \left[\alpha\left(1-\beta\right)\right]^{x_{12}} \left[\left(1-\alpha\right)\beta\right]^{x_{21}} \left[\left(1-\alpha\right)\left(1-\beta\right)\right]^{x_{22}} \\ &= \alpha^{x_{11}+x_{12}} \left(1-\alpha\right)^{x_{21}+x_{22}} \beta^{x_{11}+x_{21}} \left(1-\beta\right)^{x_{12}+x_{22}} \quad \text{for } 0 \le \alpha \le 1,\, 0 \le \beta \le 1 \end{aligned}$$

The log likelihood is

$$\begin{aligned} l(\alpha, \beta) &= (x_{11}+x_{12}) \log\alpha + (x_{21}+x_{22}) \log(1-\alpha) \\ &\quad + (x_{11}+x_{21}) \log\beta + (x_{12}+x_{22}) \log(1-\beta) \\ \text{for } 0 &< \alpha < 1,\, 0 < \beta < 1 \end{aligned}$$

and the maximum likelihood estimates are

$$\hat{\alpha} = \frac{x_{11}+x_{12}}{n} \quad \text{and} \quad \hat{\beta} = \frac{x_{11}+x_{21}}{n}$$

therefore

$$\begin{aligned} &\max_{(\theta_{11},\theta_{12},\theta_{21},\theta_{22})\in\Omega_0} l(\theta_{11}, \theta_{12}, \theta_{21}, \theta_{22}; \mathbf{X}) \\ = \; &(X_{11}+X_{12}) \log\left(\frac{X_{11}+X_{12}}{n}\right) + (X_{21}+X_{22}) \log\left(\frac{X_{21}+X_{22}}{n}\right) \\ &+ (X_{11}+X_{21}) \log\left(\frac{X_{11}+X_{21}}{n}\right) + (X_{12}+X_{22}) \log\left(\frac{X_{12}+X_{22}}{n}\right) \end{aligned}$$

The likelihood ratio test statistic can be written as

$$\begin{aligned} &\Lambda(\mathbf{X};\Omega_0) \\ = \; &2\left[l(\tilde{\theta}_{11}, \tilde{\theta}_{12}, \tilde{\theta}_{21}, \tilde{\theta}_{22}; \mathbf{X}) - \max_{(\theta_{11},\theta_{12},\theta_{21},\theta_{22})\in\Omega_0} l(\theta_{11}, \theta_{12}, \theta_{21}, \theta_{22}; \mathbf{X})\right] \\ = \; &2 \sum_{j=1}^{2} \sum_{i=1}^{2} X_{ij} \log\left(\frac{X_{ij}}{E_{ij}}\right) \end{aligned}$$

where $E_{ij} = \frac{R_i C_j}{n}$, $R_i = X_{i1} + X_{i2}$, $C_j = X_{1j} + X_{2j}$ for $i, j = 1, 2$. E_{ij} is the expected frequency if the hypothesis of independence is true.

The corresponding observed value is

$$\lambda(\mathbf{x};\Omega_0) = 2 \sum_{j=1}^{2} \sum_{i=1}^{2} x_{ij} \log\left(\frac{x_{ij}}{e_{ij}}\right)$$

where $e_{ij} = \frac{r_i c_j}{n}$, $r_i = x_{i1} + x_{i2}$, $c_j = x_{1j} + x_{2j}$ for $i, j = 1, 2$.

Since $k - q = 3 - 2 = 1$

$$\begin{aligned} p\textit{-value} &\approx P\left[W \geq \lambda(\mathbf{x};\Omega_0)\right] \quad \text{where } W \sim \chi^2(1) \\ &= 2\left[1 - P\left(Z \leq \sqrt{\lambda(\mathbf{x};\Omega_0)}\right)\right] \quad \text{where } Z \sim \mathrm{N}(0,1) \end{aligned}$$

This of course is the usual test of independence in a two-way table which was discussed in your previous statistics course.

8.4 Chapter 8 Problems

1. Suppose $X_1, X_2, \ldots, X_n$ is a random sample from the distribution with probability density function

$$f(x;\theta) = \theta x^{\theta-1} \quad \text{for } 0 \le x \le 1 \quad \text{for } \theta > 0$$

   (a) Find the likelihood ratio test statistic for testing $H_0 : \theta = \theta_0$.

   (b) If $n = 20$ and $\sum\limits_{i=1}^{20} \log x_i = -25$ find the approximate *p-value* for testing $H_0 : \theta = 1$ using the asymptotic distribution of the likelihood ratio statistic. What would you conclude?

2. Suppose $X_1, X_2, \ldots, X_n$ is a random sample from the Pareto$(1, \theta)$ distribution.

   (a) Find the likelihood ratio test statistic for testing $H_0 : \theta = \theta_0$. Indicate how to find the approximate *p-value*.

   (b) For the data $n = 25$ and $\sum\limits_{i=1}^{25} \log x_i = 40$ find the approximate *p-value* for testing $H_0 : \theta_0 = 1$. What would you conclude?

3. Suppose $X_1, X_2, \ldots, X_n$ is a random sample from the distribution with probability density function

$$f(x;\theta) = \frac{x}{\theta^2} e^{-\frac{1}{2}(x/\theta)^2} \quad \text{for } x > 0,\ \theta > 0$$

   (a) Find the likelihood ratio test statistic for testing $H_0 : \theta = \theta_0$. Indicate how to find the approximate *p-value*.

   (b) If $n = 20$ and $\sum\limits_{i=1}^{20} x_i^2 = 10$ find the approximate *p-value* for testing $H : \theta = 0.1$. What would you conclude?

4. Suppose $X_1, X_2, \ldots, X_n$ is a random sample from the Weibull$(2, \theta_1)$ distribution and independently $Y_1, Y_2, \ldots, Y_m$ is a random sample from the Weibull$(2, \theta_2)$ distribution. Find the likelihood ratio test statistic for testing $H_0 : \theta_1 = \theta_2$. Indicate how to find the approximate *p-value*.

5. Suppose $(X_1, X_2, X_3) \sim$ Multinomial$(n, \theta_1, \theta_2, \theta_3)$.

   (a) Find the likelihood ratio test statistic for testing $H_0 : \theta_1 = \theta_2 = \theta_3$. Indicate how to find the approximate *p-value*.

   (b) Find the likelihood ratio test statistic for testing $H_0 : \theta_1 = \alpha^2$, $\theta_2 = 2\alpha(1-\alpha)$, $\theta_3 = (1-\alpha)^2$. Indicate how to find the approximate *p-value*.

6. Suppose $X_1, X_2, \ldots, X_n$ is a random sample from the $\mathrm{N}(\mu_1, \sigma_1^2)$ distribution and independently $Y_1, Y_2, \ldots, Y_m$ is a random sample from the $\mathrm{N}(\mu_2, \sigma_2^2)$ distribution. Find the likelihood ratio test statistic for testing $H_0 : \mu_1 = \mu_2,\ \sigma_1^2 = \sigma_2^2$. Indicate how to find the approximate *p-value*.

9. Solutions to Chapter Exercises

9.1 Chapter 2

Exercise 2.1.5

(a) $P(A) \geq 0$ follows from Definition 2.1.3(A1). From Example 2.1.4(c) we have $P(\bar{A}) = 1 - P(A)$. But from Definition 2.1.3(A1) $P(\bar{A}) \geq 0$ and therefore $P(A) \leq 1$.

(b) Since $A = (A \cap B) \cup (A \cap \bar{B})$ and $(A \cap B) \cap (A \cap \bar{B}) = \varnothing$ then by Example 2.1.4(b)

$$P(A) = P(A \cap B) + P(A \cap \bar{B})$$

or

$$P(A \cap \bar{B}) = P(A) - P(A \cap B)$$

as required.

(c) Since

$$A \cup B = (A \cap \bar{B}) \cup (A \cap B) \cup (\bar{A} \cap B)$$

is the union of three mutually exclusive events then by Definition 2.1.3(A3) and Example 2.1.4(a) we have

$$P(A \cup B) = P(A \cap \bar{B}) + P(A \cap B) + P(\bar{A} \cap B) \tag{9.1}$$

By the result proved in (b)

$$P(A \cap \bar{B}) = P(A) - P(A \cap B) \tag{9.2}$$

and similarly

$$P(\bar{A} \cap B) = P(B) - P(A \cap B) \tag{9.3}$$

Substituting (9.2) and (9.3) into (9.1) gives

$$\begin{aligned} P(A \cup B) &= P(A) - P(A \cap B) + P(A \cap B) + P(B) - P(A \cap B) \\ &= P(A) + P(B) - P(A \cap B) \end{aligned}$$

as required.

Exercise 2.3.7

By 2.11.8

$$\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots \quad \text{for } -1 < x \le 1$$

Let $x = p - 1$ to obtain

$$\log p = (p-1) - \frac{(p-1)^2}{2} + \frac{(p-1)^3}{3} - \tag{9.4}$$

which holds for $0 < p \le 2$ and therefore also hold for $0 < p < 1$. Now (9.4) can be written as

$$\begin{aligned} \log p &= -(1-p) - \frac{(1-p)^2}{2} - \frac{(1-p)^3}{3} - \\ &= \sum_{x=1}^{\infty} \frac{-(1-p)^x}{x} \quad \text{for } 0 < p < 1 \end{aligned}$$

Therefore

$$\begin{aligned} \sum_{x=1}^{\infty} f(x) &= \sum_{x=1}^{\infty} \frac{-(1-p)^x}{x \log p} \\ &= \frac{1}{\log p} \sum_{x=1}^{\infty} \frac{-(1-p)^x}{x} \\ &= \frac{\log p}{\log p} = 1 \end{aligned}$$

which holds for $0 < p < 1$.

Exercise 2.4.11

(a)

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{-0}^{\infty} \frac{\alpha x^{\alpha-1}}{\beta^\alpha} e^{-(x/\beta)^\alpha} dx$$

Let $y = (x/\beta)^\alpha$. Then $dy = \frac{\alpha x^{\alpha-1}}{\beta^\alpha} dx$. When $x = 0$, $y = 0$ and as $x \to \infty$, $y \to \infty$. Therefore

$$\begin{aligned} \int_{-\infty}^{\infty} f(x)\,dx &= \int_{-0}^{\infty} \frac{\alpha x^{\alpha-1}}{\beta^\alpha} e^{-(x/\beta)^\alpha} dx \\ &= \int_{-0}^{\infty} e^{-y} dy = \Gamma(1) = 0! = 1 \end{aligned}$$

(b) If $\alpha = 1$ then

$$f(x) = \frac{1}{\beta} e^{-x/\beta} \quad \text{for } x > 0$$

which is the probability density function of an Exponential(β) random variable.

(c) See Figure 9.1

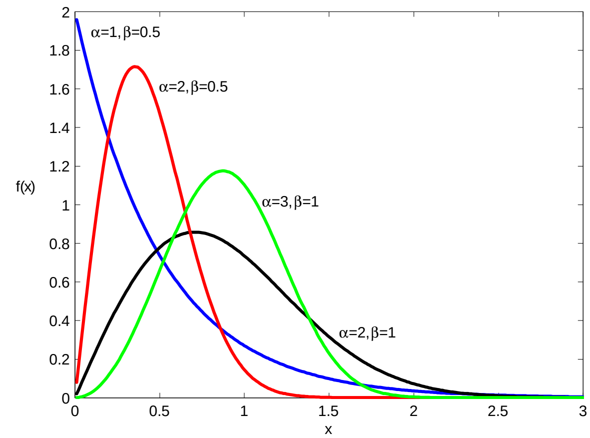


Figure 9.1: Graphs of Weibull probability density function

Exercise 2.4.12

(a)

$$\begin{aligned}
\int_{-\infty}^{\infty} f(x)\,dx &= \int_{\alpha}^{\infty} \frac{\beta\alpha^{\beta}}{x^{\beta+1}}dx = \lim_{b\to\infty}\int_{1}^{b} \beta\alpha^{\beta}x^{-\beta-1}dx \\
&= \lim_{b\to\infty}\left[-\alpha^{\beta}x^{-\beta}\big|_{\alpha}^{b}\right] = 1-\alpha^{\beta}\lim_{b\to\infty}\left[b^{-\beta}\right] \\
&= 1-\alpha^{\beta}\lim_{b\to\infty}\frac{1}{b^{\beta}} = 1 \quad \text{since } \beta > 0
\end{aligned}$$

(b) See Figure 9.2

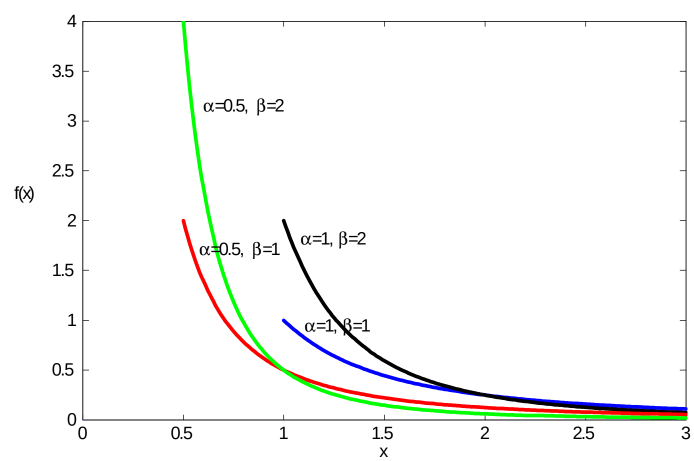


Figure 9.2: Graphs of Pareto probability density function

Exercise 2.5.4

(a) For $X \sim$ Cauchy$(\theta, 1)$ the probability density function is

$$f(x;\theta) = \frac{1}{\pi\left[1+(x-\theta)^2\right]} \quad \text{for } x \in \Re, \quad \theta \in \Re$$

and 0 otherwise. See Figure 9.3 for a sketch of the probability density function for $\theta = -1, 0, 1$.

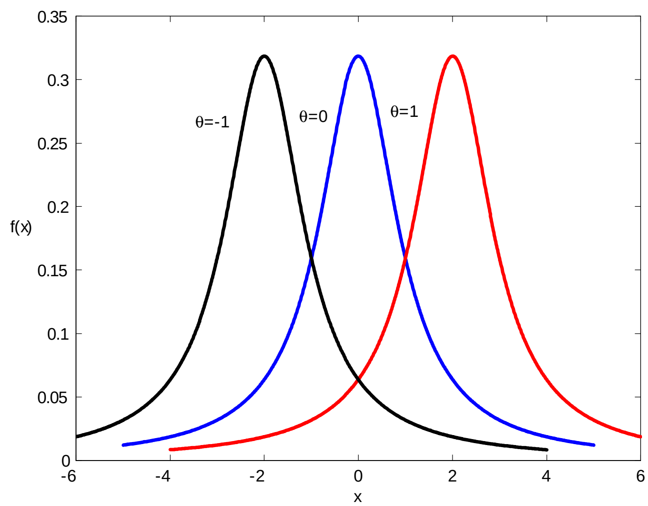


Figure 9.3: Cauchy$(\theta, 1)$ probability density functions for $\theta = -1, 0, 1$

Let

$$f_0(x) = f(x;\theta = 0) = \frac{1}{\pi(1+x^2)} \quad \text{for } x \in \Re$$

and 0 otherwise. Then

$$f(x;\theta) = \frac{1}{\pi\left[1+(x-\theta)^2\right]} = f_0(x-\theta) \quad \text{for } x \in \Re, \quad \theta \in \Re$$

and therefore θ is a location parameter of this distribution.

(b) For $X \sim$ Cauchy$(0, \theta)$ the probability density function is

$$f(x;\theta) = \frac{1}{\theta\pi\left[1+(x/\theta)^2\right]} \quad \text{for } x \in \Re, \quad \theta > 0$$

and 0 otherwise. See Figure 9.4 for a sketch of the probability density function for $\theta = 0.5, 1, 2$.

Let

$$f_1(x) = f(x;\theta = 1) = \frac{1}{\pi(1+x^2)} \quad \text{for } x \in \Re$$

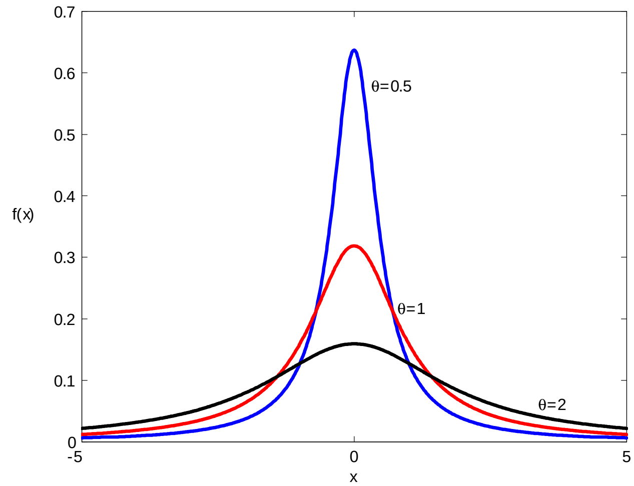


Figure 9.4: Cauchy$(0,\theta)$ probability density function for $\theta = 0.5, 1, 2$

and 0 otherwise. Then

$$f(x;\theta) = \frac{1}{\theta\pi\left[1+(x/\theta)^2\right]} = \frac{1}{\theta} f_1\left(\frac{x}{\theta}\right) \quad \text{for } x \in \Re, \quad \theta > 0$$

and therefore θ is a scale parameter of this distribution.

Exercise 2.6.11

If $X \sim$ Exponential(1) then the probability density function of X is

$$f(x) = e^{-x} \quad \text{for } x \geq 0$$

$Y = \beta X^{1/\alpha} = h(X)$ for $\alpha > 0$, $\beta > 0$ is an increasing function with inverse function $X = (Y/\beta)^{\alpha} = h^{-1}(Y)$. Since the support set of X is $A = \{x : x \geq 0\}$, the support set of Y is $B = \{y : y \geq 0\}$.

Since

$$\begin{aligned} \frac{d}{dy}h^{-1}(y) &= \frac{d}{dy}\left(\frac{y^{\alpha}}{\beta^{\alpha}}\right) \\ &= \frac{\alpha y^{\alpha-1}}{\beta^{\alpha}} \end{aligned}$$

then by Theorem 2.6.8 the probability density function of Y is

$$\begin{aligned} g(y) &= f(h^{-1}(y))\left|\frac{d}{dy}h^{-1}(y)\right| \\ &= \frac{\alpha y^{\alpha-1}}{\beta^{\alpha}}e^{-(y/\beta)^{\alpha}} \quad \text{for } y \geq 0 \end{aligned}$$

which is the probability density function of a Weibull(α, β) random variable as required.

Exercise 2.6.12

X is a random variable with probability density function

$$f(x) = \theta x^{\theta-1} \quad \text{for } 0 < x < 1,\ \theta > 0$$

and 0 otherwise. $Y = -\log X$ is a decreasing function with inverse function $X = e^{-Y} = h^{-1}(Y)$. Since the support set of X is $A = \{x : 0 < x < 1\}$, the support set of Y is $B = \{y : y > 0\}$.

Since

$$\begin{aligned} \frac{d}{dy}h^{-1}(y) &= \frac{d}{dy}\left(e^{-y}\right) \\ &= -e^{-y} \end{aligned}$$

then by Theorem 2.6.8 the probability density function of Y is

$$\begin{aligned} g(y) &= f(h^{-1}(y))\left|\frac{d}{dy}h^{-1}(y)\right| \\ &= \theta\left(e^{-y}\right)^{\theta-1}\left|-e^{-y}\right| \\ &= \theta e^{-\theta y} \quad \text{for } y \geq 0 \end{aligned}$$

which is the probability density function of a Exponential$\left(\frac{1}{\theta}\right)$ random variable as required.

Exercise 2.7.4

Using integration by parts with $u = 1 - F(x)$ and $dv = dx$ gives $du = -f(x)\,dx$, $v = x$ and

$$\begin{aligned}\int_0^\infty [1 - F(x)]\,dx &= x[1 - F(x)]|_0^\infty + \int_0^\infty xf(x)\,dx \\ &= \lim_{x\to\infty} x\int_x^\infty f(t)\,dt + E(X)\end{aligned}$$

The desired result holds if

$$\lim_{x\to\infty} x\int_{-x}^\infty f(t)\,dt = 0 \tag{9.5}$$

Since

$$0 \le x\int_x^\infty f(t)\,dt \le \int_x^\infty tf(t)\,dt$$

then (9.5) holds by the Squeeze Theorem if

$$\lim_{x\to\infty} \int_x^\infty tf(t)\,dt = 0$$

Since $E(X) = \int\limits_0^\infty xf(x)\,dx$ exists then $G(x) = \int\limits_0^x tf(t)\,dt$ exists for all $x > 0$ and $\lim\limits_{x\to\infty} G(x) = E(X)$.

By the First Fundamental Theorem of Calculus 2.11.9 and the definition of an improper integral 2.11.11

$$\int_x^\infty tf(t)\,dt = \lim_{b\to\infty} G(b) - G(x) = E(X) - G(x)$$

Therefore

$$\begin{aligned}\lim_{x\to\infty}\int_x^\infty tf(t)\,dt &= \lim_{x\to\infty}[E(X) - G(x)] = E(X) - \lim_{x\to\infty} G(x) \\ &= E(X) - E(X) \\ &= 0\end{aligned}$$

and (9.5) holds.

Exercise 2.7.9

(a) If $X \sim$ Poisson(θ) then

$$\begin{aligned}
E(X^{(k)}) &= \sum_{x=0}^{\infty} x^{(k)} \frac{\theta^x e^{-\theta}}{x!} \\
&= e^{-\theta}\theta^k \sum_{x=k}^{\infty} \frac{\theta^{x-k}}{(x-k)!} \quad \text{let } y = x-k \\
&= e^{-\theta}\theta^k \sum_{y=0}^{\infty} \frac{\theta y}{y!} \\
&= e^{-\theta}\theta^k e^{\theta} \quad \text{by 2.11.7} \\
&= \theta^k \quad \text{for } k = 1, 2, \ldots
\end{aligned}$$

Letting $k = 1$ and $k = 2$ we have

$$E\left(X^{(1)}\right) = E\left(X\right) = \theta$$

and

$$E\left(X^{(2)}\right) = E\left[X\left(X-1\right)\right] = \theta^2$$

so

$$\begin{aligned}
Var\left(X\right) &= E\left[X\left(X-1\right)\right] + E\left(X\right) - \left[E\left(X\right)\right]^2 \\
&= \theta^2 + \theta - \theta^2 = \theta
\end{aligned}$$

(b) If $X \sim$ Negative Binomial(k, p) then

$$\begin{aligned}
E(X^{(j)}) &= \sum_{x=0}^{\infty} x^{(j)} \binom{-k}{x} p^k \left(p-1\right)^x \\
&= p^k \left(p-1\right)^j \left(-k\right)^{(j)} \sum_{x=j}^{\infty} \binom{-k-j}{x-j} \left(p-1\right)^{x-j} \quad \text{by 2.11.4(1)} \\
&= p^k \left(p-1\right)^j \left(-k\right)^{(j)} \sum_{x=j}^{\infty} \binom{-k-j}{x-j} \left(p-1\right)^{x-j} \quad \text{let } y = x-j \\
&= p^k \left(p-1\right)^j \left(-k\right)^{(j)} \sum_{y=0}^{\infty} \binom{-k-j}{y} \left(p-1\right)^{y} \\
&= p^k \left(p-1\right)^j \left(-k\right)^{(j)} \left(1+p-1\right)^{-k-j} \quad \text{by 2.11.3(2)} \\
&= \left(-k\right)^{(j)} \left(\frac{p-1}{p}\right)^j \quad \text{for } j = 1, 2, \ldots
\end{aligned}$$

Letting $k = 1$ and $k = 2$ we have

$$E\left(X^{(1)}\right) = E\left(X\right) = \left(-k\right)^{(1)} \left(\frac{p-1}{p}\right)^1 = \frac{k\left(1-p\right)}{p}$$

and

$$E\left(X^{(2)}\right) = E\left[X\left(X-1\right)\right] = \left(-k\right)^{(2)} \left(\frac{p-1}{p}\right)^2 = k\left(k+1\right)\left(\frac{1-p}{p}\right)^2$$

so

$$
\begin{aligned}
Var(X) &= E[X(X-1)] + E(X) - [E(X)]^2 \\
&= k(k+1)\left(\frac{1-p}{p}\right)^2 + \frac{k(1-p)}{p} - \left[\frac{k(1-p)}{p}\right]^2 \\
&= \frac{k(1-p)^2}{p^2} + \frac{k(1-p)}{p} = \frac{k(1-p)}{p^2}(1-p+p) \\
&= \frac{k(1-p)}{p^2}
\end{aligned}
$$

(c) If $X \sim$ Gamma(α, β) then

$$
\begin{aligned}
E(X^p) &= \int_0^\infty x^p \frac{x^{\alpha-1}e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)} dx = \int_0^\infty \frac{x^{\alpha+p-1}e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)} dx \quad \text{let } y = \frac{x}{\beta} \\
&= \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^\infty (\beta y)^{\alpha+p-1} e^{-y} \beta dy \\
&= \frac{\beta^{\alpha+p}}{\beta^\alpha \Gamma(\alpha)} \int_0^\infty y^{\alpha+p-1} e^{-y} dy \quad \text{which converges for } \alpha + p > 0 \\
&= \beta^p \frac{\Gamma(\alpha+p)}{\Gamma(\alpha)} \quad \text{for } p > -\alpha
\end{aligned}
$$

Letting $k = 1$ and $k = 2$ we have

$$
\begin{aligned}
E(X) &= \beta \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} = \beta \frac{\alpha \Gamma(\alpha)}{\Gamma(\alpha)} \\
&= \alpha\beta
\end{aligned}
$$

and

$$
\begin{aligned}
E\left(X^2\right) &= E\left(X^2\right) = \beta^2 \frac{\Gamma(\alpha+2)}{\Gamma(\alpha)} = \beta^2 \frac{\alpha(\alpha+1)\Gamma(\alpha)}{\Gamma(\alpha)} \\
&= \alpha(\alpha+1)\beta^2
\end{aligned}
$$

so

$$
\begin{aligned}
Var(X) &= E\left(X^2\right) - [E(X)]^2 = \alpha(\alpha+1)\beta^2 - (\alpha\beta)^2 \\
&= \alpha\beta^2
\end{aligned}
$$

(c) If $X \sim$ Weibull(α, β) then

$$\begin{aligned}
E(X^k) &= \int_0^\infty x^k \frac{\alpha x^{\alpha-1} e^{-(x/\beta)^\alpha}}{\beta^\alpha} dx \quad \text{let } y = \left(\frac{x}{\beta}\right)^\alpha \\
&= \int_0^\infty \left(\beta y^{1/\alpha}\right)^k e^{-y} dy = \beta^k \int_0^\infty y^{k/\alpha} e^{-y} dy \\
&= \beta^k \Gamma\left(\frac{k}{\alpha} + 1\right) \quad \text{for } k = 1, 2, \ldots
\end{aligned}$$

Letting $k = 1$ and $k = 2$ we have

$$E(X) = \beta\Gamma\left(\frac{1}{\alpha} + 1\right)$$

and

$$E\left(X^2\right) = E\left(X^2\right) = \beta^2\Gamma\left(\frac{2}{\alpha} + 1\right)$$

so

$$\begin{aligned}
Var(X) &= E\left(X^2\right) - [E(X)]^2 = \beta^2\Gamma\left(\frac{2}{\alpha} + 1\right) - \left[\beta\Gamma\left(\frac{1}{\alpha} + 1\right)\right]^2 \\
&= \beta^2 \left\{\Gamma\left(\frac{2}{\alpha} + 1\right) - \left[\Gamma\left(\frac{1}{\alpha} + 1\right)\right]^2\right\}
\end{aligned}$$

Exercise 2.8.3

From Markov's inequality we know

$$P(|Y| \geq c) \leq \frac{E\left(|Y|^k\right)}{c^k} \quad \text{for all } k, c > 0 \tag{9.6}$$

Since we are given that X is a random variable with finite mean μ and finite variance σ^2 then

$$\sigma^2 = E\left[(X - \mu)^2\right] = E\left(|X - \mu|^2\right)$$

Substituting $Y = X - \mu$, $k = 1$, and $c = k\sigma$ into (9.6) we obtain

$$P(|X - \mu| \geq k\sigma) \leq \frac{E\left(|X - \mu|^2\right)}{(k\sigma)^2} = \frac{\sigma^2}{(k\sigma)^2} = \frac{1}{k^2}$$

or

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

for all $k > 0$ as required.

Exercise 2.9.2

For a Exponential(θ) random variable $\sqrt{Var\,(X)} = \sigma\,(\theta) = \theta$. For $g\,(X) = \log X$, $g'\,(X) = \frac{1}{X}$. Therefore by (2.5), the variance of $Y = g\,(X) = \log X$ is approximately

$$\begin{aligned}\left[g'\,(\theta)\,\sigma\,(\theta)\right]^2 &= \left[\frac{1}{\theta}\,(\theta)\right]^2 \\ &= 1\end{aligned}$$

which is a constant.

Exercise 2.10.3

(a) If $X \sim$ Binomial(n, p) then

$$\begin{aligned}M\,(t) &= \sum_{x\in 0}^{\infty} e^{tx}\binom{n}{x}p^x\,(1-p)^{n-x} \\ &= \sum_{x\in 0}^{\infty}\binom{n}{x}\left(e^t p\right)^x (1-p)^{n-x} \quad \text{which converges for } t \in \Re \\ &= \left(pe^t + 1 - p\right)^{-k} \quad \text{by 2.11.3(1)}\end{aligned}$$

(b) If $X \sim$ Poisson(θ) then

$$\begin{aligned}M\,(t) &= \sum_{x=0}^{\infty} e^{tx}\frac{\theta^x e^{-\theta}}{x!} \\ &= e^{-\theta}\sum_{x=0}^{\infty}\frac{\left(\theta e^t\right)^x}{x!} \quad \text{which converges for } t \in \Re \\ &= e^{-\theta}e^{\theta e^t} \quad \text{by 2.11.7} \\ &= e^{-\theta+\theta e^t} \quad \text{for } t \in \Re\end{aligned}$$

Exercise 2.10.6

If $X \sim$ Negative Binomial(k, p) then

$$M_X\,(t) = \left(\frac{p}{1-qe^t}\right)^k \quad \text{for } t < -\log q$$

By Theorem 2.10.4 the moment generating function of $Y = X + k$ is

$$\begin{aligned}M_Y(t) &= e^{kt}M_X(t) \\ &= \left(\frac{pe^t}{1-qe^t}\right)^k \quad \text{for } t < -\log q\end{aligned}$$

Exercise 2.10.14

(a) By the Exponential series 2.11.7

$$\begin{aligned} M(t) &= e^{t^2/2} = 1 + \frac{\left(t^2/2\right)}{1!} + \frac{\left(t^2/2\right)^2}{2!} + \cdots \\ &= 1 + \frac{1}{2}t^2 + \frac{1}{2!2^2}t^4 \quad \text{for } t \in \Re \end{aligned}$$

Since $E(X^k) = k! \cdot$coefficient of t^k in the Maclaurin series for $M(t)$ we have

$$E(X) = 1!(0) = 0 \quad \text{and} \quad E\left(X^2\right) = 2!\left(\frac{1}{2}\right) = 1$$

and so

$$Var(X) = E\left(X^2\right) - [E(X)]^2 = 1$$

(b) By Theorem 2.10.4 the moment generating function of $Y = 2X - 1$ is

$$\begin{aligned} M_Y(t) &= e^{-t}M_X(2t) \quad \text{for } t \in \Re \\ &= e^{-t}e^{(2t)^2/2} \\ &= e^{-t+(2t)^2/2} \quad \text{for } t \in \Re \end{aligned}$$

By examining the list of moment generating functions in Chapter 11 we see that this is the moment generating function of a $\mathrm{N}(-1, 4)$ random variable. Therefore by the Uniqueness Theorem for Moment Generating Functions, Y has a $\mathrm{N}(-1, 4)$ distribution.

9.2 Chapter 3

Exercise 3.2.5

(a) The joint probability function of X and Y is

$$\begin{aligned} f(x,y) &= P(X=x, Y=y) \\ &= \frac{n!}{x!y!\,(n-x-y)!}\left(\theta^2\right)^x \left[2\theta\,(1-\theta)\right]^y \left[(1-\theta)^2\right]^{n-x-y} \\ \text{for } x &= 0,1,\ldots,n;\ y=0,1,\ldots,n;\ x+y\le n \end{aligned}$$

which is a Multinomial$\left(n;\theta^2, 2\theta\,(1-\theta)\,,(1-\theta)^2\right)$ distribution or the trinomial distribution.

(b) The marginal probability function of X is

$$\begin{aligned} f_1(x) &= P(X=x) = \sum_{y=0}^{n-x} \frac{n!}{x!y!\,(n-x-y)!}\left(\theta^2\right)^x \left[2\theta\,(1-\theta)\right]^y \left[(1-\theta)^2\right]^{n-x-y} \\ &= \frac{n!}{x!\,(n-x)!}\left(\theta^2\right)^x \sum_{y=0}^{\infty} \frac{(n-x)!}{y!\,(n-x-y)!}\left[2\theta\,(1-\theta)\right]^y \left[(1-\theta)^2\right]^{n-x-y} \\ &= \binom{n}{x}\left(\theta^2\right)^x \left[2\theta\,(1-\theta)+(1-\theta)^2\right]^{n-x} \quad \text{by the Binomial Series 2.11.3(1)} \\ &= \binom{n}{x}\left(\theta^2\right)^x \left(1-\theta^2\right)^{n-x} \quad \text{for } x=0,1,\ldots,n \end{aligned}$$

and so $X \sim$ Binomial$\left(n,\theta^2\right)$.

(c) In a similar manner to (b) the marginal probability function of Y can be shown to be Binomial$(n, 2\theta\,(1-\theta))$ since $P\,(Aa) = 2\theta\,(1-\theta)$.

(d)

$$\begin{aligned} P\,(X+Y=t) &= \sum_{(x,y):}\sum_{x+y=t} f(x,y) = \sum_{x=0}^{t} f(x,t-x) \\ &= \sum_{x=0}^{t} \frac{n!}{x!\,(t-x)!\,(n-t)!}\left(\theta^2\right)^x \left[2\theta\,(1-\theta)\right]^{t-x} \left[(1-\theta)^2\right]^{n-t} \\ &= \binom{n}{t}\left[(1-\theta)^2\right]^{n-t} \sum_{x=0}^{t}\binom{t}{x}\left(\theta^2\right)^x \left[2\theta\,(1-\theta)\right]^{t-x} \\ &= \binom{n}{t}\left[(1-\theta)^2\right]^{n-t}\left[\theta^2+2\theta\,(1-\theta)\right]^t \quad \text{by the Binomial Series 2.11.3(1)} \\ &= \binom{n}{t}\left[\theta^2+2\theta\,(1-\theta)\right]^t\left[(1-\theta)^2\right]^{n-t} \quad \text{for } t=0,1,\ldots,n \end{aligned}$$

Thus $X+Y \sim$Binomial$\left(n,\theta^2+2\theta\,(1-\theta)\right)$ which makes sense since $X+Y$ is counting the number of times an AA or Aa type occurs in n trials and the probability of success is $P\,(AA)+P\,(Aa) = \theta^2+2\theta\,(1-\theta)$.

Exercise 3.3.6

(a)

$$
\begin{aligned}
1 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = k\int_{0}^{\infty}\int_{0}^{\infty}\frac{1}{(1+x+y)^3}dxdy \\
&= \frac{k}{2}\int_{0}^{\infty}\left[\lim_{a\to\infty}\left(\frac{-1}{(1+x+y)^2}|_0^a\right)\right]dy \\
&= \frac{k}{2}\int_{0}^{\infty}\left(\lim_{a\to\infty}\frac{-1}{(1+a+y)^2}+\frac{1}{(1+y)^2}\right)dy \\
&= \frac{k}{2}\int_{0}^{\infty}\frac{1}{(1+y)^2}dy \\
&= \frac{k}{2}\lim_{a\to\infty}\left(\frac{-1}{(1+y)}|_0^a\right) \\
&= \frac{k}{2}\left(\lim_{a\to\infty}\frac{-1}{(1+a)}+1\right) \\
&= \frac{k}{2}
\end{aligned}
$$

Therefore $k = 2$. A graph of the joint probability density function is given in Figure 9.5.

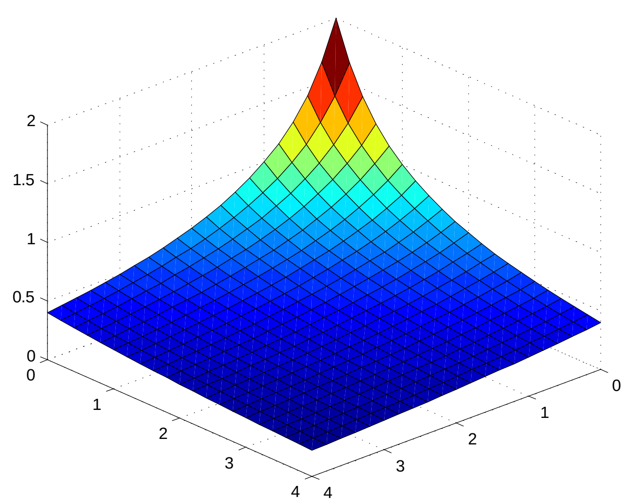


Figure 9.5: Graph of joint probability density function for Exercise 3.3.6

(b) (i)

$$
\begin{aligned}
P(X \le 1, Y \le 2) &= \int_0^2 \int_0^1 \frac{2}{(1+x+y)^3} dx dy = \int_0^2 \left(\frac{-1}{(1+x+y)^2} |_0^1 \right) dy \\
&= \int_0^2 \left[\frac{-1}{(2+y)^2} + \frac{1}{(1+y)^2} \right] dy = \left(\frac{1}{(2+y)} - \frac{1}{(1+y)} \right) |_0^2 \\
&= \frac{1}{4} - \frac{1}{3} - \frac{1}{2} + 1 = \frac{1}{12}(3 - 4 - 6 + 12) = \frac{5}{12}
\end{aligned}
$$

(ii) Since $f(x,y)$ and the support set $A = \{(x,y) : x \ge 0,\ y \ge 0\}$ are both symmetric in x and y, $P(X \le Y) = 0.5$

(iii)

$$
\begin{aligned}
P(X + Y \le 1) &= \int_0^1 \int_0^{1-y} \frac{2}{(1+x+y)^3} dx dy = \int_0^1 \left(\frac{-1}{(1+x+y)^2} |_0^{1-y} \right) dy \\
&= \int_0^1 \left[-\frac{1}{(2)^2} + \frac{1}{(1+y)^2} \right] dy = \left(-\frac{1}{4} y |_0^1 - \frac{1}{(1+y)} |_0^1 \right) \\
&= \left(-\frac{1}{4} + 0 - \frac{1}{2} + 1 \right) = \frac{1}{4}
\end{aligned}
$$

(c) Since

$$
\begin{aligned}
\int_{-\infty}^{\infty} f(x,y) dy &= \int_0^{\infty} \frac{2}{(1+x+y)^3} dy = \lim_{a \to \infty} \left(\frac{-1}{(1+x+y)^2} |_0^a \right) \\
&= \frac{1}{(1+x)^2} \quad \text{for } x \ge 0
\end{aligned}
$$

the marginal probability density function of X is

$$
f_1(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{(1+x)^2} & x \ge 0 \end{cases}
$$

By symmetry the marginal probability density function of Y is

$$
f_2(y) = \begin{cases} 0 & y < 0 \\ \frac{1}{(1+y)^2} & y \ge 0 \end{cases}
$$

(d) Since

$$\begin{aligned}
P(X \leq x, Y \leq y) &= \int_0^y \int_0^x \frac{2}{(1+s+t)^3} ds dt = \int_0^y \left(\frac{-1}{(1+s+t)^2} |_0^x \right) dt \\
&= \int_0^y \left[\frac{-1}{(1+x+t)^2} + \frac{1}{(1+t)^2} \right] dt \\
&= \left(\frac{1}{1+x+t} - \frac{1}{1+t} \right) |_0^y \\
&= \frac{1}{1+x+y} - \frac{1}{1+y} - \frac{1}{1+x} + 1 \quad \text{for } x \geq 0,\ y \geq 0
\end{aligned}$$

the joint cumulative distribution function of X and Y is

$$F(x,y) = P(X \leq x, Y \leq y) = \begin{cases} 0 & x < 0 \text{ or } y < 0 \\ 1 + \frac{1}{1+x+y} - \frac{1}{1+y} - \frac{1}{1+x} & x \geq 0,\ y \geq 0 \end{cases}$$

(e) Since

$$\begin{aligned}
\lim_{y \to \infty} F(x,y) &= \lim_{y \to \infty} \left(1 + \frac{1}{1+x+y} - \frac{1}{1+y} - \frac{1}{1+x} \right) \\
&= 1 - \frac{1}{1+x} \quad \text{for } x \geq 0
\end{aligned}$$

the marginal cumulative distribution function of X is

$$F_1(x) = P(X \leq x) = \begin{cases} 0 & x < 0 \\ 1 - \frac{1}{1+x} & x \geq 0 \end{cases}$$

Check:

$$\begin{aligned}
\frac{d}{dx} F_1(x) &= \frac{d}{dx} \left(1 - \frac{1}{1+x} \right) \\
&= \frac{1}{(1+x)^2} \\
&= f_1(x) \quad \text{for } x \geq 0
\end{aligned}$$

By symmetry the marginal cumulative distribution function of Y is

$$F_2(y) = P(Y \leq y) = \begin{cases} 0 & y < 0 \\ 1 - \frac{1}{1+y} & y \geq 0 \end{cases}$$

Exercise 3.3.7

(a) Since the support set is $A = \{(x, y) : y > x \geq 0\}$

$$\begin{aligned}
1 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = k\int_{0}^{\infty}\int_{0}^{y} e^{-x-y}dxdy \\
&= k\int_{0}^{\infty}\left(-e^{-x-y}|_0^y\right)dy \\
&= k\int_{0}^{\infty}\left(-e^{-2y}+e^{-y}\right)dy \\
&= k\left[\lim_{a\to\infty}\left(\frac{1}{2}e^{-2y}-e^{-y}\right)|_0^a\right] \\
&= k\left[\lim_{a\to\infty}\left(\frac{1}{2}e^{-2a}-e^{-a}\right)-\frac{1}{2}+1\right] \\
&= \frac{k}{2}
\end{aligned}$$

and therefore $k = 2$. A graph of the joint probability density function is given in Figure 9.6

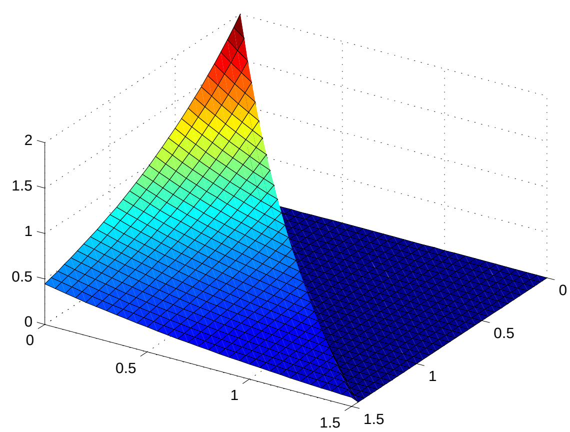


Figure 9.6: Graph of joint probability density function for Exercise 3.3.7

(b) (i) The region of integration is shown in Figure 9.7.

$$\begin{aligned}
P(X \le 1, Y \le 2) &= 2\int_{x=0}^{1}\int_{y=x}^{2} e^{-x-y}dydx = 2\int_{x=0}^{1} e^{-x}\left(-e^{-y}|_x^2\right)dx \\
&= 2\int_{x=0}^{1} e^{-x}\left(-e^{-2}+e^{-x}\right)dx = 2\int_{x=0}^{1}\left(-e^{-2}e^{-x}+e^{-2x}\right)dx \\
&= 2\left(e^{-2}e^{-x}-\frac{1}{2}e^{-2x}\right)|_0^1 = 2\left(e^{-3}-\frac{1}{2}e^{-2}-e^{-2}+\frac{1}{2}\right) \\
&= 1+2e^{-3}-3e^{-2}
\end{aligned}$$

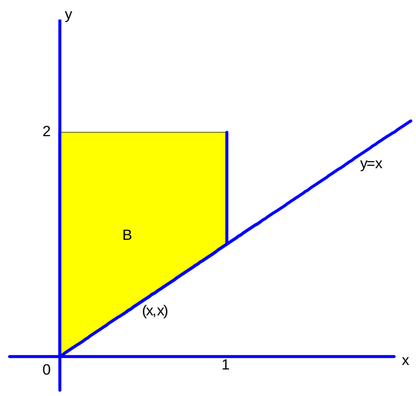


Figure 9.7: Region of integration for Exercise 3.3.7(b)(i)

(ii) Since the support set $A = \{(x,y) : 0 < x < y < \infty\}$ contains only values for which $x < y$ then $P(X \le Y) = 1$.

(iii) The region of integration is shown in Figure 9.8

$$\begin{aligned}
P(X+Y \le 1) &= 2\int_{x=0}^{1/2}\int_{y=x}^{1-x} e^{-x-y}dydx = 2\int_{x=0}^{1} e^{-x}\left(-e^{-y}|_x^{1-x}\right)dx \\
&= 2\int_{x=0}^{1} e^{-x}\left(-e^{x-1}+e^{-x}\right)dx = 2\int_{x=0}^{1}\left(-e^{-1}+e^{-2x}\right)dx \\
&= 2\left(-e^{-1}x-\frac{1}{2}e^{-2x}\right)|_0^{1/2} = 2\left[-e^{-1}\left(\frac{1}{2}\right)-\frac{1}{2}e^{-1}+0+\frac{1}{2}\right] \\
&= 1-2e^{-1}
\end{aligned}$$

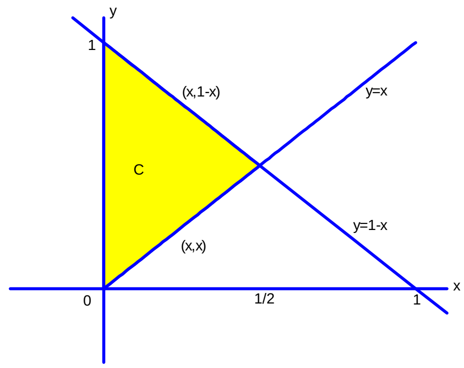


Figure 9.8: Region of integration for Exercise 3.3.7(b)(iii)

(c) The marginal probability density function of X is

$$f_1(x) = \int_{-\infty}^{\infty} f(x,y)dy = 2\int_{x}^{\infty} e^{-x-y}dy = 2e^{-x}\lim_{a\to\infty}\left(-e^{-y}|_x^a\right) = 2e^{-2x} \quad \text{for } x>0$$

and 0 otherwise which we recognize is an Exponential(1/2) probability density function. The marginal probability density function of Y is

$$\int_{-\infty}^{\infty} f(x,y)dy = 2\int_{0}^{y} e^{-x-y}dx = 2e^{-y}\left(-e^{-x}|_0^y\right) = 2e^{-y}\left(1-e^{-y}\right) \quad \text{for } y>0$$

and 0 otherwise.

(d) Since

$$\begin{aligned} P(X\le x, Y\le y) &= 2\int_{s=0}^{x}\int_{t=s}^{y} e^{-s-t}dtds = 2\int_0^x e^{-s}\left(-e^{-t}|_s^y\right)ds \\ &= 2\int_0^x e^{-s}\left(-e^{-y}+e^{-s}\right)ds \\ &= 2\int_0^x \left(-e^{-y}e^{-s}+e^{-2s}\right)ds = 2\left(e^{-y}e^{-s}-\frac{1}{2}e^{-2s}\right)|_0^x \\ &= 2e^{-x-y}-e^{-2x}-2e^{-y}+1 \quad \text{for } y\ge x\ge 0 \end{aligned}$$

$$\begin{aligned} P(X \leq x, Y \leq y) &= 2\int_{s=0}^{y}\int_{t=s}^{y} e^{-s-t}dtds \\ &= e^{-2y} - 2e^{-y} + 1 \quad \text{for } x > y > 0 \end{aligned}$$

and

$$P(X \leq x, Y \leq y) = 0 \quad \text{for } x \leq 0 \text{ or } y \leq 0$$

therefore the joint cumulative distribution function of X and Y is

$$F(x,y) = P(X \leq x, Y \leq y) = \begin{cases} 2e^{-x-y} - e^{-2x} - 2e^{-y} + 1 & y \geq x \geq 0 \\ e^{-2y} - 2e^{-y} + 1 & x > y > 0 \\ 0 & x \leq 0 \text{ or } y \leq 0 \end{cases}$$

(e) Since the support set is $A = \{(x,y) : y \geq x \geq 0\}$ and

$$\begin{aligned} \lim_{y\to\infty} F(x,y) &= \lim_{y\to\infty}\left(2e^{-x-y} - e^{-2x} - 2e^{-y} + 1\right) \\ &= 1 - e^{-2x} \quad \text{for } x > 0 \end{aligned}$$

marginal cumulative distribution function of X is

$$F_1(x) = P(X \leq x) = \begin{cases} 0 & x \leq 0 \\ 1 - e^{-2x} & x > 0 \end{cases}$$

which we also recognize as the cumulative distribution function of an Exponential(1/2) random variable.

Since

$$\begin{aligned} \lim_{x\to\infty} F(x,y) &= \lim_{x\to y}\left(2e^{-x-y} - e^{-2x} - 2e^{-y} + 1\right) = 2e^{-2y} - e^{-2y} - 2e^{-y} + 1 \\ &= e^{-2y} - 2e^{-y} + 1 \quad \text{for } y > 0 \end{aligned}$$

the marginal cumulative distribution function of Y is

$$F_2(y) = P(Y \leq y) = \begin{cases} 0 & y \leq 0 \\ 1 + e^{-2y} - 2e^{-y} & y \geq 0 \end{cases}$$

Check:

$$\begin{aligned} P(Y \leq y) &= \int_0^y f_2(t)\,dt = 2\int_0^y \left(e^{-t} - e^{-2t}\right)dt = 2\left(-e^{-t} + \frac{1}{2}e^{-2t}\right)\Big|_0^y \\ &= 2\left(-e^{-y} + \frac{1}{2}e^{-2y} + 1 - \frac{1}{2}\right) = 1 + e^{-2y} - 2e^{-y} \quad \text{for } y > 0 \end{aligned}$$

or

$$\frac{d}{dy}F_2(y) = \frac{d}{dy}\left(e^{-2y} - 2e^{-y} + 1\right) = -2e^{-2y} + 2e^{-y} = f_2(y) \quad \text{for } y > 0$$

Exercise 3.4.4

From the solution to Exercise 3.2.5 we have

$$\begin{aligned} f(x,y) &= \frac{n!}{x!y!\,(n-x-y)!}\left(\theta^2\right)^x \left[2\theta\left(1-\theta\right)\right]^y \left[\left(1-\theta\right)^2\right]^{n-x-y} \\ \text{for } x &= 0,1,\ldots,n;\ y=0,1,\ldots,n;\ x+y\le n \end{aligned}$$

$$f_1\left(x\right) = \binom{n}{x}\left(\theta^2\right)^x\left(1-\theta^2\right)^{n-x} \quad \text{for } x=0,1,\ldots,n$$

and

$$f_2\left(y\right) = \binom{n}{y}\left[2\theta\left(1-\theta\right)\right]^y\left[1-2\theta\left(1-\theta\right)\right]^{n-y} \quad \text{for } y=0,1,\ldots,n$$

Since

$$f\left(0,0\right) = \left(1-\theta\right)^{2n} \neq f_1\left(0\right) f_2\left(0\right) = \left(1-\theta^2\right)^n\left[1-2\theta\left(1-\theta\right)\right]^n$$

therefore X and Y are not independent random variables.

From the solution to Exercise 3.3.6 we have

$$f\left(x,y\right) = \frac{2}{\left(1+x+y\right)^3} \quad \text{for } x\ge 0,\ y\ge 0$$

$$f_1\left(x\right) = \frac{1}{\left(1+x\right)^2} \quad \text{for } x\ge 0$$

and

$$f_2\left(y\right) = \frac{1}{\left(1+y\right)^2} \quad \text{for } y\ge 0$$

Since

$$f\left(0,0\right) = 2 \neq f_1\left(0\right) f_2\left(0\right) = \left(1\right)\left(1\right)$$

therefore X and Y are not independent random variables.

From the solution to Exercise 3.3.7 we have

$$f\left(x,y\right) = 2e^{-x-y} \quad \text{for } 0<x<y<\infty$$

$$f_1\left(x\right) = 2e^{-2x} \quad \text{for } x>0$$

and

$$f_2\left(y\right) = 2e^{-y}\left(1-e^{-y}\right) \quad \text{for } y>0$$

Since

$$f\left(1,2\right) = 2e^{-3} \neq f_1\left(1\right) f_2\left(2\right) = \left(2e^{-1}\right)\left[2e^{-2}\left(1-e^{-2}\right)\right]$$

therefore X and Y are not independent random variables.

Exercise 3.5.3

$$
\begin{aligned}
P(Y=y|X=x) &= \frac{P(X=x,Y=y)}{P(X=x)} \\
&= \frac{\frac{n!}{x!y!(n-x-y)!}\left(\theta^2\right)^x\left[2\theta(1-\theta)\right]^y\left[(1-\theta)^2\right]^{n-x-y}}{\frac{n!}{x!(n-x)!}\left(\theta^2\right)^x\left(1-\theta^2\right)^{n-x}} \\
&= \frac{(n-x)!}{y!\,(n-x-y)!}\,\frac{\left[2\theta(1-\theta)\right]^y\left[(1-\theta)^2\right]^{n-x-y}}{\left(1-\theta^2\right)^y\left(1-\theta^2\right)^{n-x-y}} \\
&= \binom{n-x}{y}\left[\frac{2\theta(1-\theta)}{\left(1-\theta^2\right)}\right]^y\left[\frac{(1-\theta)^2}{\left(1-\theta^2\right)}\right]^{(n-x)-y} \\
&= \binom{n-x}{y}\left[\frac{2\theta(1-\theta)}{\left(1-\theta^2\right)}\right]^y\left[\frac{1-2\theta+\theta^2}{\left(1-\theta^2\right)}\right]^{(n-x)-y} \\
&= \binom{n-x}{y}\left[\frac{2\theta(1-\theta)}{\left(1-\theta^2\right)}\right]^y\left[\frac{1-\theta^2-2\theta+2\theta^2}{\left(1-\theta^2\right)}\right]^{(n-x)-y} \\
&= \binom{n-x}{y}\left[\frac{2\theta(1-\theta)}{\left(1-\theta^2\right)}\right]^y\left[1-\frac{2\theta(1-\theta)}{\left(1-\theta^2\right)}\right]^{(n-x)-y}
\end{aligned}
$$

for $y=0,1,\ldots,(n-x)$ which is the probability function of a Binomial$\left(n-x, \frac{2\theta(1-\theta)}{\left(1-\theta^2\right)}\right)$ as required.

We are given that there are x genotypes of type AA. Therefore there are only $n-x$ members (trials) whose genotype must be determined. The genotype can only be of type Aa or type aa. In a population with only these two types the proportion of type Aa would be

$$\frac{2\theta(1-\theta)}{2\theta(1-\theta)+(1-\theta)^2} = \frac{2\theta(1-\theta)}{\left(1-\theta^2\right)}$$

and the proportion of type aa would be

$$\frac{(1-\theta)^2}{2\theta(1-\theta)+(1-\theta)^2} = \frac{(1-\theta)^2}{\left(1-\theta^2\right)} = 1-\frac{2\theta(1-\theta)}{\left(1-\theta^2\right)}$$

Since we have $(n-x)$ independent trials with probability of Success (type Aa) equal to $\frac{2\theta(1-\theta)}{\left(1-\theta^2\right)}$ then it follows that the number of Aa types, given that there are x members of type AA, would follow a Binomial$\left(n-x, \frac{2\theta(1-\theta)}{\left(1-\theta^2\right)}\right)$ distribution.

Exercise 3.5.4

For Example 3.3.3 the conditional probability density function of X given $Y = y$ is

$$f_1(x|y) = \frac{f(x,y)}{f_2(y)} = \frac{x+y}{y+\frac{1}{2}} \quad \text{for } 0 < x < 1 \text{ for each } 0 < y < 1$$

Check:

$$\int_{-\infty}^{\infty} f_1(x|y)\,dx = \int_0^1 \frac{x+y}{y+\frac{1}{2}}dx = \frac{1}{y+\frac{1}{2}}\left(\frac{1}{2}x^2 + xy|_0^1\right) = \frac{1}{y+\frac{1}{2}}\left(\frac{1}{2}+y-0\right) = 1$$

By symmetry the conditional probability density function of Y given $X = x$ is

$$f_2(y|x) = \frac{x+y}{x+\frac{1}{2}} \quad \text{for } 0 < y < 1 \text{ for each } 0 < x < 1$$

and

$$\int_{-\infty}^{\infty} f_2(y|x)\,dy = 1$$

For Exercise 3.3.6 the conditional probability density function of X given $Y = y$ is

$$f_1(x|y) = \frac{f(x,y)}{f_2(y)} = \frac{\frac{2}{(1+x+y)^3}}{\frac{1}{(1+y)^2}} = \frac{2(1+y)^2}{(1+x+y)^3} \quad \text{for } x > 0 \text{ for each } y > 0$$

Check:

$$\begin{aligned}
\int_{-\infty}^{\infty} f_1(x|y)\,dx &= \int_0^{\infty} \frac{2(1+y)^2}{(1+x+y)^3}dx = (1+y)^2 \lim_{a\to\infty}\left[\frac{-1}{(1+x+y)^2}|_0^1\right] \\
&= (1+y)^2 \lim_{a\to\infty}\left[\frac{-1}{(1+a+y)^2} + \frac{1}{(1+y)^2}\right] = (1+y)^2\frac{1}{(1+y)^2} = 1
\end{aligned}$$

By symmetry the conditional probability density function of Y given $X = x$ is

$$f_2(y|x) = \frac{2(1+x)^2}{(1+x+y)^3} \quad \text{for } y > 0 \text{ for each } x > 0$$

and

$$\int_{-\infty}^{\infty} f_2(y|x)\,dy = 1$$

For Exercise 3.3.7 the conditional probability density function of X given $Y = y$ is

$$f_1(x|y) = \frac{f(x,y)}{f_2(y)} = \frac{2e^{-x-y}}{2e^{-y}(1-e^{-y})} = \frac{e^{-x}}{1-e^{-y}} \quad \text{for } 0 < x < y \text{ for each } y > 0$$

Check:

$$\int_{-\infty}^{\infty} f_1(x|y)\,dx = \int_0^y \frac{e^{-x}}{1-e^{-y}}dx = \frac{1}{1-e^{-y}}\left(-e^{-x}|_0^y\right) = \frac{1-e^{-y}}{1-e^{-y}} = 1$$

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

$$f_2(y|x) = \frac{f(x,y)}{f_1(x)} = \frac{2e^{-x-y}}{2e^{-2x}} = e^{x-y} \quad \text{for } y > x \text{ for each } x > 0$$

Check:

$$\int_{-\infty}^{\infty} f_2(y|x)\,dy = \int_x^{\infty} e^{x-y}dy = e^x \lim_{a\to\infty}\left(-e^{-y}|_x^a\right) = e^x\left(0+e^{-x}\right) = 1$$

Exercise 3.6.10

$$\begin{aligned}
E(XY) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)dxdy = 2\int_0^{\infty}\int_0^{y} xye^{-x-y}dxdy \\
&= 2\int_0^{\infty} ye^{-y}\left(\int_0^y xe^{-x}dx\right)dy = 2\int_0^{\infty} ye^{-y}\left[\left(-xe^{-x}-e^{-x}\right)|_0^y\right]dy \\
&= 2\int_0^{\infty} ye^{-y}\left(-ye^{-y}-e^{-y}+1\right)dy = -2\int_0^{\infty}\left(y^2e^{-2y}+ye^{-2y}-ye^{-y}\right)dy \\
&= -2\int_0^{\infty} y^2e^{-2y}dy - \int_0^{\infty}(2y)\,e^{-2y}dy + 2\int_0^{\infty} ye^{-y}dy \\
&= -2\int_0^{\infty} y^2e^{-2y}dy - \int_0^{\infty}(2y)\,e^{-2y}dy + 2\Gamma(2) \quad \text{but } \Gamma(2) = 1! = 1 \\
&= 2 - 2\int_0^{\infty} y^2e^{-2y}dy - \int_0^{\infty}(2y)\,e^{-2y}dy \quad \text{let } u = 2y \text{ or } y = \frac{u}{2} \text{ so } dy = \frac{1}{2}du \\
&= 2 - 2\int_0^{\infty}\left(\frac{u}{2}\right)^2 e^{-u}\frac{1}{2}du - \int_0^{\infty} ue^{-u}\frac{1}{2}du \\
&= 2 - \frac{1}{4}\int_0^{\infty} u^2e^{-u}du - \frac{1}{2}\int_0^{\infty} ue^{-u}du = 2 - \frac{1}{4}\Gamma(3) - \frac{1}{2}\Gamma(2) \quad \text{but } \Gamma(3) = 2! = 2 \\
&= 2 - \frac{1}{2} - \frac{1}{2} = 1
\end{aligned}$$

Since $X \sim$ Exponential$(1/2)$, $E(X) = 1/2$ and $Var(X) = 1/4$.

$$\begin{aligned}
E(Y) &= \int_{-\infty}^{\infty} y f_2(y) dy = 2\int_0^{\infty} y e^{-y}\left(-e^{-y}+1\right) dy \\
&= -\int_0^{\infty} 2y e^{-2y} dy + 2\int_0^{\infty} y e^{-y} dy = -\int_0^{\infty} 2y e^{-2y} dy + 2\Gamma(2) \\
&= 2 - \int_0^{\infty} 2y e^{-2y} dy \quad \text{let } u = 2y \\
&= 2 - \int_0^{\infty} u e^{-u} \frac{1}{2} du = 2 - \frac{1}{2}\int_0^{\infty} u e^{-u} du = 2 - \frac{1}{2}\Gamma(2) = 2 - \frac{1}{2} = \frac{3}{2}
\end{aligned}$$

$$\begin{aligned}
E\left(Y^2\right) &= \int_{-\infty}^{\infty} y^2 f_2(y) dy = 2\int_0^{\infty} y^2 e^{-y}\left(-e^{-y}+1\right) dy \\
&= -2\int_0^{\infty} y^2 e^{-2y} dy + 2\int_0^{\infty} y^2 e^{-y} dy = -2\int_0^{\infty} y^2 e^{-2y} dy + 2\Gamma(3) \\
&= 4 - 2\int_0^{\infty} y^2 e^{-2y} dy \quad \text{let } u = 2y \\
&= 4 - 2\int_0^{\infty} \left(\frac{u}{2}\right)^2 e^{-u} \frac{1}{2} du = 4 - \frac{1}{4}\int_0^{\infty} u^2 e^{-u} du = 4 - \frac{1}{4}\Gamma(3) = 4 - \frac{1}{2} = \frac{7}{2}
\end{aligned}$$

$$Var(Y) = E\left(Y^2\right) - [E(Y)]^2 = \frac{7}{2} - \left(\frac{3}{2}\right)^2 = \frac{14-9}{4} = \frac{5}{4}$$

Therefore

$$Cov(X,Y) = E(XY) - E(X)E(Y) = 1 - \left(\frac{1}{2}\right)\left(\frac{3}{2}\right) = 1 - \frac{3}{4} = \frac{1}{4}$$

and

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = \frac{\frac{1}{4}}{\sqrt{\left(\frac{1}{4}\right)\left(\frac{5}{4}\right)}} = \frac{1}{\sqrt{5}}$$

Exercise 3.7.12

Since $Y \sim \text{Gamma}(\alpha, \frac{1}{\theta})$

$$\begin{aligned} E(Y) &= \alpha\left(\frac{1}{\theta}\right) = \frac{\alpha}{\theta} \\ Var(Y) &= \alpha\left(\frac{1}{\theta}\right)^2 = \frac{\alpha}{\theta^2} \end{aligned}$$

and

$$\begin{aligned} E\left(Y^k\right) &= \int_0^\infty y^k \frac{y^{\alpha-1}\theta^\alpha e^{-\theta y}}{\Gamma(\alpha)} dy \quad \text{let } u = \theta y \\ &= \frac{\theta^\alpha}{\Gamma(\alpha)} \int_0^\infty \left(\frac{u}{\theta}\right)^{k+\alpha-1} e^{-u} \frac{1}{\theta} du = \frac{\theta^\alpha \theta^{-k-\alpha}}{\Gamma(\alpha)} \int_0^\infty u^{k+\alpha-1} e^{-u} du \\ &= \frac{\theta^{-k}\Gamma(\alpha+k)}{\Gamma(\alpha)} \quad \text{for } \alpha + k > 0 \end{aligned}$$

Since $X|Y = y \sim \text{Weibull}(p, y^{-1/p})$

$$E(X|y) = y^{-1/p}\Gamma\left(1 + \frac{1}{p}\right), \quad E(X|Y) = Y^{-1/p}\Gamma\left(1 + \frac{1}{p}\right)$$

and

$$\begin{aligned} Var(X|y) &= \left(y^{-1/p}\right)^2 \left[\Gamma\left(1 + \frac{2}{p}\right) - \Gamma^2\left(1 + \frac{1}{p}\right)\right] \\ Var(X|Y) &= Y^{-2/p} \left[\Gamma\left(1 + \frac{2}{p}\right) - \Gamma^2\left(1 + \frac{1}{p}\right)\right] \end{aligned}$$

Therefore

$$\begin{aligned} E(X) &= E[E(X|Y)] = \Gamma\left(1 + \frac{1}{p}\right) E\left(Y^{-1/p}\right) \\ &= \Gamma\left(1 + \frac{1}{p}\right) \frac{\theta^{1/p}\Gamma\left(\alpha - \frac{1}{p}\right)}{\Gamma(\alpha)} \\ &= \theta^{1/p} \frac{\Gamma\left(1 + \frac{1}{p}\right)\Gamma\left(\alpha - \frac{1}{p}\right)}{\Gamma(\alpha)} \end{aligned}$$

and

$$
\begin{aligned}
Var(X) &= E[Var(X|Y)] + Var[E(X|Y)] \\
&= E\left\{Y^{-2/p}\left[\Gamma\left(1+\frac{2}{p}\right)-\Gamma^2\left(1+\frac{1}{p}\right)\right]\right\} + Var\left[Y^{-1/p}\Gamma\left(1+\frac{1}{p}\right)\right] \\
&= \left[\Gamma\left(1+\frac{2}{p}\right)-\Gamma^2\left(1+\frac{1}{p}\right)\right]E\left(Y^{-2/p}\right) + \Gamma^2\left(1+\frac{1}{p}\right)Var\left(Y^{-1/p}\right) \\
&= \left[\Gamma\left(1+\frac{2}{p}\right)-\Gamma^2\left(1+\frac{1}{p}\right)\right]E\left(Y^{-2/p}\right) \\
&\quad +\Gamma^2\left(1+\frac{1}{p}\right)\left\{E\left[\left(Y^{-1/p}\right)^2\right]-\left[E\left(Y^{-1/p}\right)\right]^2\right\} \\
&= \Gamma\left(1+\frac{2}{p}\right)E\left(Y^{-2/p}\right)-\Gamma^2\left(1+\frac{1}{p}\right)E\left(Y^{-2/p}\right) \\
&\quad +\Gamma^2\left(1+\frac{1}{p}\right)E\left(Y^{-2/p}\right)-\Gamma^2\left(1+\frac{1}{p}\right)\left[E\left(Y^{-1/p}\right)\right]^2 \\
&= \Gamma\left(1+\frac{2}{p}\right)E\left(Y^{-2/p}\right)-\Gamma^2\left(1+\frac{1}{p}\right)\left[E\left(Y^{-1/p}\right)\right]^2 \\
&= \Gamma\left(1+\frac{2}{p}\right)\frac{\theta^{2/p}\Gamma\left(\alpha-\frac{2}{p}\right)}{\Gamma(\alpha)}-\Gamma^2\left(1+\frac{1}{p}\right)\left[\frac{\theta^{1p}\Gamma\left(\alpha-\frac{1}{p}\right)}{\Gamma(\alpha)}\right]^2 \\
&= \theta^{2/p}\left\{\frac{\Gamma\left(1+\frac{2}{p}\right)\Gamma\left(\alpha-\frac{2}{p}\right)}{\Gamma(\alpha)}-\left[\frac{\Gamma\left(1+\frac{1}{p}\right)\Gamma\left(\alpha-\frac{1}{p}\right)}{\Gamma(\alpha)}\right]^2\right\}
\end{aligned}
$$

Exercise 3.7.13

Since $P \sim \text{Beta}(a,b)$

$$E(P)=\frac{a}{a+b},\quad Var(P)=\frac{ab}{(a+b+1)(a+b)^2}$$

and

$$
\begin{aligned}
E\left(P^k\right) &= \int_0^1 p^k\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}p^{a-1}(1-p)^{b-1}\,dp \\
&= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\int_0^1 p^{a+k-1}(1-p)^{b-1}\,dp \\
&= \frac{\Gamma(a+b)\Gamma(a+k)}{\Gamma(a)\Gamma(a+k+b)}\int_0^1\frac{\Gamma(a+k+b)}{\Gamma(a+k)\Gamma(b)}p^{k+a-1}(1-p)^{b-1}\,dp \\
&= \frac{\Gamma(a+b)\Gamma(a+k)}{\Gamma(a)\Gamma(a+k+b)}(1)=\frac{\Gamma(a+b)\Gamma(a+k)}{\Gamma(a)\Gamma(a+k+b)}
\end{aligned}
$$

provided $a+k>0$ and $a+k+b>0$.

Since $Y|P=p \sim$ Geometric(p)

$$E(Y|p)=\frac{1-p}{p},\quad E(Y|P)=\frac{1-P}{P}=\frac{1}{P}-1$$

and

$$Var(Y|p)=\frac{1-p}{p^2},\quad Var(Y|P)=\frac{1-P}{P^2}=\frac{1}{P^2}-\frac{1}{P}$$

Therefore

$$\begin{aligned}
E(Y) &= E[E(Y|P)]=E\left(\frac{1}{P}-1\right)=E\left(P^{-1}\right)-1\\
&= \frac{\Gamma(a+b)\Gamma(a-1)}{\Gamma(a)\Gamma(a-1+b)}-1\\
&= \frac{(a-1+b)\Gamma(a-1+b)\Gamma(a-1)}{(a-1)\Gamma(a-1)\Gamma(a-1+b)}-1\\
&= \frac{a-1+b}{a-1}-1=\frac{b}{a-1}
\end{aligned}$$

provided $a>1$ and $a+b>1$.

Now

$$Var(Y)=E[Var(Y|P)]+Var[E(Y|P)]$$

$$\begin{aligned}
&= E\left(\frac{1}{P^2}-\frac{1}{P}\right)+Var\left(\frac{1}{P}-1\right)\\
&= E\left(\frac{1}{P^2}\right)-E\left(\frac{1}{P}\right)+E\left[\left(\frac{1}{P}-1\right)^2\right]-\left[E\left(\frac{1}{P}-1\right)\right]^2\\
&= E\left(\frac{1}{P^2}\right)-E\left(\frac{1}{P}\right)+E\left(\frac{1}{P^2}\right)-2E\left(\frac{1}{P}\right)+1-\left[E\left(\frac{1}{P}\right)\right]^2+2E\left(\frac{1}{P}\right)-1\\
&= 2E\left(P^{-2}\right)-E\left(P^{-1}\right)-\left[E\left(P^{-1}\right)\right]^2\\
&= 2\frac{\Gamma(a+b)\Gamma(a-2)}{\Gamma(a)\Gamma(a-2+b)}-\frac{\Gamma(a+b)\Gamma(a-1)}{\Gamma(a)\Gamma(a-1+b)}-\left[\frac{\Gamma(a+b)\Gamma(a-1)}{\Gamma(a)\Gamma(a-1+b)}\right]^2\\
&= 2\frac{(a+b-2)(a+b-1)}{(a-2)(a-1)}-\frac{a+b-1}{a-1}-\left(\frac{a+b-1}{a-1}\right)^2\\
&= \left(\frac{a+b-1}{a-1}\right)\left[2\left(\frac{a+b-2}{a-2}\right)-1-\frac{a+b-1}{a-1}\right]\\
&= \frac{ab(a+b-1)}{(a-1)^2(a-2)}
\end{aligned}$$

provided $a>2$ and $a+b>2$.

Exercise 3.9.4

If $T = X_i + X_j$, $i \neq j$, then

$$T \sim \text{ Binomial}(n,\ p_i + p_j)$$

The moment generating function of $T = X_i + X_j$ is

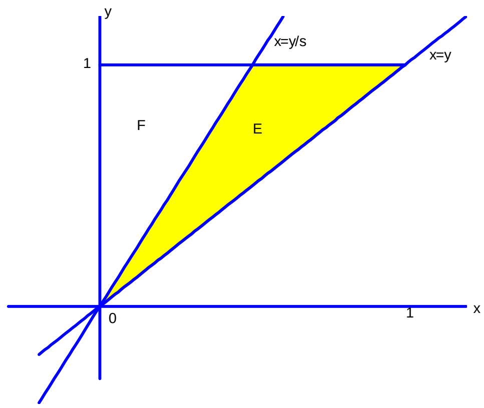
$$\begin{aligned}
M(t) &= E\left(e^{tT}\right) + E\left(e^{t(X_i+X_j)}\right) = E\left(e^{tX_i+tX_j}\right) \\
&= M(0,\dots,0,t,0,\dots,0,t,0,\dots,0) \\
&= \left(p_1 + \cdots + p_i e^t + \cdots + p_j e^t + \cdots + p_{k-1} + p_k\right)^n \quad \text{for } t \in \Re \\
&= \left[(p_i + p_j)\, e^t + (1 - p_i - p_j)\right]^n \quad \text{for } t \in \Re
\end{aligned}$$

which is the moment generating function of a Binomial$(n, p_i + p_j)$ random variable. Therefore by the Uniqueness Theorem for Moment Generating Functions, T has a Binomial$(n, p_i + p_j)$ distribution.

9.3 Chapter 4

Exercise 4.1.2

The support set of (X, Y) is $A = \{(x, y) : 0 < x < y < 1\}$ which is the union of the regions E and F shown in Figure 9.3For $s > 1$



$$\begin{aligned}
G(s) &= P(S \leq s) = P\left(\frac{Y}{X} \leq s\right) = P(Y \leq sX) = P(Y - sX \leq 0) \\
&= \int\limits_{(x,y) \in E}\int 3ydxdy = \int\limits_{y=0}^{1}\int\limits_{x=y/s}^{y} 3ydxdy = \int\limits_{0}^{1} 3y\left(x|_{y/s}^{y}\right)dy \\
&= \int\limits_{0}^{1} 3y\left(y - \frac{y}{s}\right)dy = \left(1 - \frac{1}{s}\right)\int\limits_{0}^{1} 3y^2dy = \left(1 - \frac{1}{s}\right)\left(y^3|_0^1\right) = 1 - \frac{1}{s}
\end{aligned}$$

The cumulative distribution function for S is

$$G(s) = \begin{cases} 0 & s \leq 1 \\ 1 - \frac{1}{s} & s > 1 \end{cases}$$

As a check we note that $\lim\limits_{s \to 1^+}\left(1 - \frac{1}{s}\right) = 0 = G(1)$ and $\lim\limits_{s \to \infty}\left(1 - \frac{1}{s}\right) = 1$ so $G(s)$ is a continuous function for all $s \in \Re$.

For $s > 1$

$$g(s) = \frac{d}{ds}G(s) = \frac{d}{ds}\left(1 - \frac{1}{s}\right) = \frac{1}{s^2}$$

The probability density function of S is

$$g(s) = \frac{1}{s^2} \quad \text{for } s > 1$$

and 0 otherwise.

Exercise 4.2.5

Since $X \sim$ Exponential(1) and $Y \sim$ Exponential(1) independently, the joint probability density function of X and Y is

$$\begin{aligned} f(x,y) &= f_1(x) f_2(y) = e^{-x}e^{-y} \\ &= e^{-x-y} \end{aligned}$$

with support set $R_{XY} = \{(x,y) : x > 0, y > 0\}$ which is shown in Figure 9.9.The transfor-

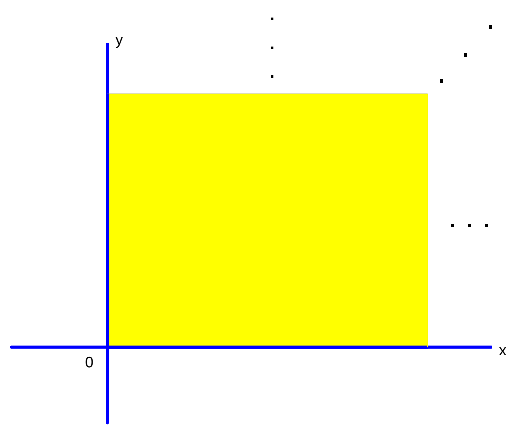


Figure 9.9: Support set R_{XY} for Example 4.2.5

mation

$$S: \; U = X + Y, \quad V = X - Y$$

has inverse transformation

$$X = \frac{U+V}{2}, \quad Y = \frac{U-V}{2}$$

Under S the boundaries of R_{XY} are mapped as

$$\begin{aligned} (k,0) &\rightarrow (k,k) \quad \text{for } k \geq 0 \\ (0,k) &\rightarrow (k,-k) \quad \text{for } k \geq 0 \end{aligned}$$

and the point $(1,2)$ is mapped to and $(3,-1)$. Thus S maps R_{XY} into

$$R_{UV} = \{(u,v) : -u < v < u, \; u > 0\}$$

as shown in Figure 9.10.

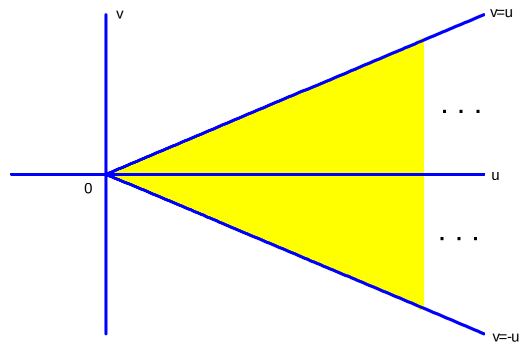


Figure 9.10: Support set R_{UV} for Example 4.2.5

The Jacobian of the inverse transformation is

$$\frac{\partial(x,y)}{\partial(u,v)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{vmatrix} = -\frac{1}{2}$$

The joint probability density function of U and V is given by

$$\begin{aligned} g(u,v) &= f\left(\frac{u+v}{2}, \frac{u-v}{2}\right)\left|-\frac{1}{2}\right| \\ &= \frac{1}{2}e^{-u} \quad \text{for } (u,v) \in R_{UV} \end{aligned}$$

and 0 otherwise.

To find the marginal probability density functions for U we note that the support set R_{UV} is not rectangular and the range of integration for v will depend on u.

$$\begin{aligned} g_1(u) &= \int_{-\infty}^{\infty} g(u,v)\,dv \\ &= \frac{1}{2}e^{-u}\int_{v=-u}^{u} dv \\ &= \frac{1}{2}ue^{-u}(2) \\ &= ue^{-u} \quad \text{for } u > 0 \end{aligned}$$

and 0 otherwise which is the probability density function of a Gamma$(2,1)$ random variable. Therefore $U \sim$ Gamma$(2,1)$.

To find the marginal probability density functions for V we need to consider the two cases $v \geq 0$ and $v < 0$. For $v \geq 0$

$$\begin{aligned}
g_2(v) &= \int_{-\infty}^{\infty} g(u,v)\,du \\
&= \frac{1}{2}\int_{u=v}^{\infty} e^{-u}du \\
&= \frac{-1}{2}\left(\lim_{b\to\infty} e^{-u}|_{v}^{b}\right) \\
&= \frac{-1}{2}\left(\lim_{b\to\infty} e^{-b} - e^{-v}\right) \\
&= \frac{1}{2}e^{-v}
\end{aligned}$$

For $v < 0$

$$\begin{aligned}
g_2(v) &= \int_{-\infty}^{\infty} g(u,v)\,du \\
&= \frac{1}{2}\int_{u=-v}^{\infty} e^{-u}du \\
&= \frac{-1}{2}\left(\lim_{b\to\infty} e^{-u}|_{-v}^{b}\right) \\
&= \frac{-1}{2}\left(\lim_{b\to\infty} e^{-b} - e^{v}\right) \\
&= \frac{1}{2}e^{v}
\end{aligned}$$

Therefore the probability density function of V is

$$g_2(v) = \begin{cases} \frac{1}{2}e^{v} & v < 0 \\ \frac{1}{2}e^{-v} & v \geq 0 \end{cases}$$

which is the probability density function of Double Exponential$(0,1)$ random variable. Therefore $V \sim$ Double Exponential$(0,1)$.

Exercise 4.2.7

Since $X \sim$ Beta(a, b) and $Y \sim$ Beta$(a+b, c)$ independently, the joint probability density function of X and Y is

$$\begin{aligned} f(x,y) &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1} \frac{\Gamma(a+b+c)}{\Gamma(a+b)\Gamma(c)} y^{a+b-1}(1-y)^{c-1} \\ &= \frac{\Gamma(a+b+c)}{\Gamma(a)\Gamma(b)\Gamma(c)} x^{a-1}(1-x)^{b-1} y^{a+b-1}(1-y)^{c-1} \end{aligned}$$

with support set $R_{XY} = \{(x, y) : 0 < x < 1, 0 < y < 1\}$ as shown in Figure 9.11.

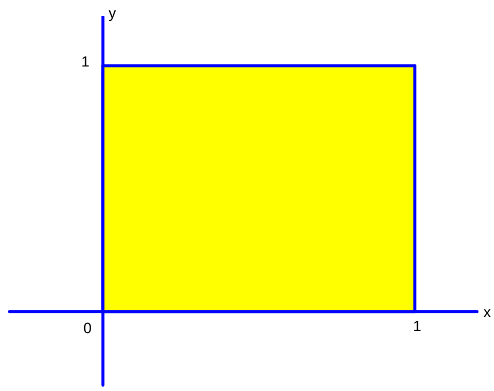


Figure 9.11: Support R_{XY} for Exercise 4.2.5

The transformation

$$S: \; U = XY, \quad V = X$$

has inverse transformation

$$X = V, \quad Y = U/V$$

Under S the boundaries of R_{XY} are mapped as

$$\begin{aligned} (k,0) &\to (0,k) \quad 0 \le k \le 1 \\ (0,k) &\to (0,0) \quad 0 \le k \le 1 \\ (1,k) &\to (k,1) \quad 0 \le k \le 1 \\ (k,1) &\to (k,k) \quad 0 \le k \le 1 \end{aligned}$$

and the point $\left(\frac{1}{2}, \frac{1}{3}\right)$ is mapped to and $\left(\frac{1}{6}, \frac{1}{2}\right)$. Thus S maps R_{XY} into

$$R_{UV} = \{(u, v) : 0 < u < v < 1\}$$

as shown in Figure 9.12

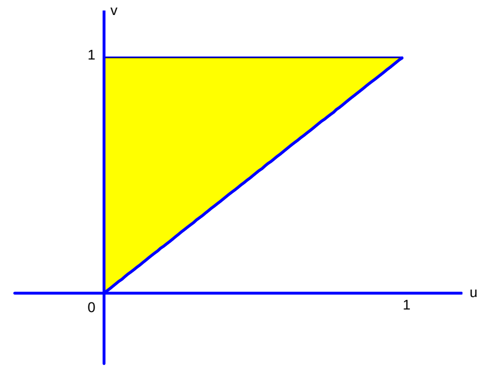


Figure 9.12: Support set R_{UV} for Exercise 4.2.5

The Jacobian of the inverse transformation is

$$\frac{\partial(x,y)}{\partial(u,v)} = \begin{vmatrix} 0 & 1 \\ \frac{1}{v} & \frac{\partial y}{\partial v} \end{vmatrix} = -\frac{1}{v}$$

The joint probability density function of U and V is given by

$$\begin{aligned} g(u,v) &= f\left(v, \frac{u}{v}\right)\left|-\frac{1}{v}\right| \\ &= \frac{\Gamma(a+b+c)}{\Gamma(a)\Gamma(b)\Gamma(c)} v^{a-1}(1-v)^{b-1}\left(\frac{u}{v}\right)^{a+b-1}\left(1-\frac{u}{v}\right)^{c-1}\frac{1}{v} \\ &= \frac{\Gamma(a+b+c)}{\Gamma(a)\Gamma(b)\Gamma(c)} u^{a+b-1}(1-v)^{b-1}v^{-b-1}\left(1-\frac{u}{v}\right)^{c-1} \quad \text{for } (u,v)\in R_{UV} \end{aligned}$$

and 0 otherwise.

To find the marginal probability density functions for U we note that the support set R_{UV} is not rectangular and the range of integration for v will depend on u.

$$\begin{aligned} g(u) &= \int_{-\infty}^{\infty} g(u,v)\,dv = \frac{\Gamma(a+b+c)}{\Gamma(a)\Gamma(b)\Gamma(c)} u^{a+b-1}\int_{v=u}^{1}(1-v)^{b-1}v^{-b-1}\left(1-\frac{u}{v}\right)^{c-1}dv \\ &= \frac{\Gamma(a+b+c)}{\Gamma(a)\Gamma(b)\Gamma(c)} u^{a+b-1}\int_{u}^{1} v^{-b+1}(1-v)^{b-1}\left(1-\frac{u}{v}\right)^{c-1}\frac{1}{v^2}dv \\ &= \frac{\Gamma(a+b+c)}{\Gamma(a)\Gamma(b)\Gamma(c)} u^{a-1}\int_{u}^{1} u^{b-1}\left(\frac{1}{v}-1\right)^{b-1}\left(1-\frac{u}{v}\right)^{c-1}\frac{u}{v^2}dv \\ &= \frac{\Gamma(a+b+c)}{\Gamma(a)\Gamma(b)\Gamma(c)} u^{a-1}\int_{u}^{1}\left(\frac{u}{v}-u\right)^{b-1}\left(1-\frac{u}{v}\right)^{c-1}\frac{u}{v^2}dv \end{aligned}$$

To evaluate this integral we make the substitution

$$t = \frac{\frac{u}{v} - u}{1 - u}$$

Then

$$\begin{aligned}
\frac{u}{v} - u &= (1-u)\,t \\
1 - \frac{u}{v} &= 1 - u - (1-u)\,t = (1-u)(1-t) \\
-\frac{u}{v^2} &= (1-u)\,dt
\end{aligned}$$

When $v = u$ then $t = 1$ and when $v = 1$ then $t = 0$. Therefore

$$\begin{aligned}
g_1(u) &= \frac{\Gamma(a+b+c)}{\Gamma(a)\Gamma(b)\Gamma(c)} u^{a-1} \int_0^1 [(1-u)\,t]^{b-1} [(1-u)(1-t)]^{c-1} (1-u)\,dt \\
&= \frac{\Gamma(a+b+c)}{\Gamma(a)\Gamma(b)\Gamma(c)} u^{a-1} (1-u)^{b+c-1} \int_0^1 t^{b-1} (1-t)^{c-1}\,dt \\
&= \frac{\Gamma(a+b+c)}{\Gamma(a)\Gamma(b)\Gamma(c)} u^{a-1} (1-u)^{b+c-1} \frac{\Gamma(b)\Gamma(c)}{\Gamma(b+c)} \\
&= \frac{\Gamma(a+b+c)}{\Gamma(a)\Gamma(b+c)} u^{a-1} (1-u)^{b+c-1} \quad \text{for } 0 < u < 1
\end{aligned}$$

and 0 otherwise which is the probability density function of a Beta$(a, b+c)$ random variable. Therefore $U \sim$ Beta$(a, b+c)$.

Exercise 4.2.12

(a) Consider the transformation

$$S:\ U = \frac{X/n}{Y/m}, \quad V = Y$$

which has inverse transformation

$$X = \left(\frac{n}{m}\right) UV, \quad Y = V$$

Since $X \sim \chi^2(n)$ independently of $Y \sim \chi^2(m)$ then the joint probability density function of X and Z is

$$\begin{aligned}
f(x,y) &= f_1(x) f_2(y) = \frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-x/2} \frac{1}{2^{m/2}\Gamma(m/2)} y^{m/2-1} e^{-y/2} \\
&= \frac{1}{2^{(n+m)/2}\Gamma(n/2)\Gamma(m/2)} x^{n/2-1} e^{-x/2} y^{m/2-1} e^{-y/2}
\end{aligned}$$

with support set $R_{XY} = \{(x,y) : x > 0, y > 0\}$. The transformation S maps R_{XY} into $R_{UV} = \{(u,v) : u > 0, v > 0\}$.

The Jacobian of the inverse transformation is

$$\frac{\partial (x,y)}{\partial (u,v)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} \left(\frac{n}{m}\right) v & \frac{\partial x}{\partial v} \\ 0 & 1 \end{vmatrix} = \left(\frac{n}{m}\right) v$$

The joint probability density function of U and V is given by

$$\begin{aligned} g(u,v) &= f\left(\left(\frac{n}{m}\right) uv, v\right) \left|\left(\frac{n}{m}\right) v\right| \\ &= \frac{1}{2^{(n+m)/2}\Gamma(n/2)\Gamma(m/2)} \left(\frac{n}{m}\right)^{n/2-1} (uv)^{n/2-1} e^{-\left(\frac{n}{m}\right)uv/2} v^{m/2-1} e^{-v/2} \left(\frac{n}{m}\right) v \\ &= \frac{\left(\frac{n}{m}\right)^{n/2} u^{n/2-1}}{2^{(n+m)/2}\Gamma(n/2)\Gamma(m/2)} v^{(n+m)/2-1} e^{-v\left(\frac{nu}{m}+1\right)/2} \quad \text{for } (u,v) \in R_{UV} \end{aligned}$$

and 0 otherwise.

To determine the distribution of U we still need to find the marginal probability density function for U.

$$\begin{aligned} g_1(u) &= \int_{-\infty}^{\infty} g(u,v)\, dv \\ &= \frac{\left(\frac{n}{m}\right)^{n/2} u^{n/2-1}}{2^{(n+m)/2}\Gamma(n/2)\Gamma(m/2)} \int_0^{\infty} v^{(n+m)/2-1} e^{-v\left(\frac{nu}{m}+1\right)/2} dv \end{aligned}$$

Let $y = \frac{v}{2}\left(1 + \frac{n}{m}u\right)$ so that $v = 2y\left(1 + \frac{n}{m}u\right)^{-1}$ and $dv = 2\left(1 + \frac{n}{m}u\right)^{-1} dy$. Note that when $v = 0$ then $y = 0$, and when $v \to \infty$ then $y \to \infty$. Therefore

$$\begin{aligned} g_1(u) &= \frac{\left(\frac{n}{m}\right)^{n/2} u^{n/2-1}}{2^{(n+m)/2}\Gamma(n/2)\Gamma(m/2)} \int_0^{\infty} \left[2y\left(1+\frac{n}{m}u\right)^{-1}\right]^{(n+m)/2-1} e^{-y} \left[2\left(1+\frac{n}{m}u\right)^{-1}\right] dy \\ &= \frac{\left(\frac{n}{m}\right)^{n/2} u^{n/2-1} 2^{(n+m)/2}}{2^{(n+m)/2}\Gamma(n/2)\Gamma(m/2)} \left[\left(1+\frac{n}{m}u\right)^{-1}\right]^{(n+m)/2} \int_0^{\infty} y^{(n+m)/2-1} e^{-y} dy \\ &= \frac{\left(\frac{n}{m}\right)^{n/2} u^{n/2-1}}{\Gamma(n/2)\Gamma(m/2)} \left(1+\frac{n}{m}u\right)^{-(n+m)/2} \Gamma\left(\frac{n+m}{2}\right) \\ &= \frac{\left(\frac{n}{m}\right)^{n/2} \Gamma\left(\frac{n+m}{2}\right)}{\Gamma(n/2)\Gamma(m/2)} u^{n/2-1} \left(1+\frac{n}{m}u\right)^{-(n+m)/2} \quad \text{for } u > 0 \end{aligned}$$

and 0 otherwise which is the probability density function of a random variable with a $\mathrm{F}(n,m)$ distribution. Therefore $U = \frac{X/n}{Y/m} \sim \mathrm{F}(n,m)$.

(b) To find $E(U)$ we use

$$E(U) = E\left(\frac{X/n}{Y/m}\right) = \frac{m}{n} E(X) E\left(Y^{-1}\right)$$

since $X \sim \chi^2(n)$ independently of $Y \sim \chi^2(m)$. From Example 4.2.10 we know that if $W \sim \chi^2(k)$ then

$$E\left(W^p\right) = \frac{2^p \Gamma\left(k/2 + p\right)}{\Gamma\left(k/2\right)} \quad \text{for } \frac{k}{2} + p > 0$$

Therefore

$$E\left(X\right) = \frac{2\Gamma\left(n/2 + 1\right)}{\Gamma\left(n/2\right)} = n$$

$$E\left(Y^{-1}\right) = \frac{2^{-1}\Gamma\left(m/2 - 1\right)}{\Gamma\left(m/2\right)} = \frac{1}{2\left(m/2 - 1\right)} = \frac{1}{m-2} \quad \text{if } m > 2$$

and

$$E\left(U\right) = \frac{m}{n}\left(n\right) E\left(\frac{1}{m-2}\right) = \frac{m}{m-2} \quad \text{if } m > 2$$

To find $Var\left(U\right)$ we need

$$E\left(U^2\right) = E\left[\frac{\left(X/n\right)^2}{\left(Y/m\right)^2}\right] = \frac{m^2}{n^2} E\left(X^2\right) E\left(Y^{-2}\right)$$

Now

$$E\left(X^2\right) = \frac{2^2\Gamma\left(n/2 + 2\right)}{\Gamma\left(n/2\right)} = 4\left(\frac{n}{2} + 1\right)\left(\frac{n}{2}\right) = n\left(n+2\right)$$

$$E\left(Y^{-2}\right) = \frac{2^{-2}\Gamma\left(m/2 - 2\right)}{\Gamma\left(m/2\right)} = \frac{1}{4\left(\frac{m}{2} - 1\right)\left(\frac{m}{2} - 2\right)} = \frac{1}{\left(m-2\right)\left(m-4\right)} \quad \text{for } m > 4$$

and

$$E\left(U^2\right) = \frac{m^2}{n^2} n\left(n+2\right)\frac{1}{\left(m-2\right)\left(m-4\right)} = \frac{n+2}{n}\frac{m^2}{\left(m-2\right)\left(m-4\right)} \quad \text{for } m > 4$$

Therefore

$$\begin{aligned}
Var\left(U\right) &= E\left(U^2\right) - \left[E\left(U\right)\right]^2 \\
&= \frac{n+2}{n}\frac{m^2}{\left(m-2\right)\left(m-4\right)} - \left(\frac{m}{m-2}\right)^2 \\
&= \frac{m^2}{m-2}\left[\frac{n+2}{n\left(m-4\right)} - \frac{1}{m-2}\right] \\
&= \frac{m^2}{m-2}\left[\frac{\left(n+2\right)\left(m-2\right) - n\left(m-4\right)}{n\left(m-4\right)\left(m-2\right)}\right] \\
&= \frac{m^2}{m-2}\left[\frac{2\left(n+m-2\right)}{n\left(m-4\right)\left(m-2\right)}\right] \\
&= \frac{2m^2\left(n+m-2\right)}{n\left(m-2\right)^2\left(m-4\right)} \quad \text{for } m > 4
\end{aligned}$$

Exercise 4.3.3

$X_1, X_2, \ldots, X_n$ are independent and identically distributed random variables with moment generating function $M(t)$, $E(X_i) = \mu$, and $Var(X_i) = \sigma^2 < \infty$.

The moment generating function of $Z = \sqrt{n}\left(\bar{X} - \mu\right)/\sigma$ is

$$\begin{aligned}
M_Z(t) &= E\left(e^{tZ}\right) = E\left(e^{t\sqrt{n}\left(\bar{X}-\mu\right)/\sigma}\right) \\
&= e^{-\sqrt{n}\mu/\sigma t} E\left[\exp\left(\frac{t\sqrt{n}}{\sigma}\frac{1}{n}\sum_{i=1}^{n} X_i\right)\right] \\
&= e^{-\sqrt{n}\mu/\sigma t} E\left[\exp\left(\frac{t}{\sigma\sqrt{n}}\sum_{i=1}^{n} X_i\right)\right] \\
&= e^{-\sqrt{n}\mu/\sigma t} \prod_{i=1}^{n} E\left[\exp\left(\frac{t}{\sigma\sqrt{n}} X_i\right)\right] \quad \text{since } X_1, X_2, \ldots, X_n \text{ are independent} \\
&= e^{-\sqrt{n}\mu/\sigma t} \prod_{i=1}^{n} M\left(\frac{t}{\sigma\sqrt{n}}\right) \quad \text{since } X_1, X_2, \ldots, X_n \text{ are identically distributed} \\
&= e^{-\sqrt{n}\mu/\sigma t} \left[M\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n
\end{aligned}$$

Exercise 4.3.7

$$\begin{aligned}
\sum_{i=1}^{n} (X_i - \mu)^2 &= \sum_{i=1}^{n} \left(X_i - \bar{X} + \bar{X} - \mu\right)^2 \\
&= \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2 + 2\left(\bar{X} - \mu\right)\sum_{i=1}^{n}\left(X_i - \bar{X}\right) + \sum_{i=1}^{n}\left(\bar{X} - \mu\right)^2 \\
&= \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2 + n\left(\bar{X} - \mu\right)^2
\end{aligned}$$

since

$$\begin{aligned}
\sum_{i=1}^{n} \left(X_i - \bar{X}\right) &= \sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \bar{X} = \sum_{i=1}^{n} X_i - n\bar{X} \\
&= \sum_{i=1}^{n} X_i - n\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\
&= \sum_{i=1}^{n} X_i - \sum_{i=1}^{n} X_i \\
&= 0
\end{aligned}$$

Exercise 4.3.11

Since $X_1, X_2, \ldots, X_n$ are independent $\mathrm{N}(\mu_1, \sigma_1^2)$ random variables then by Theorem 4.3.8

$$U = \frac{(n-1)\,S_1^2}{\sigma_1^2} = \frac{\sum\limits_{i=1}^{n}\left(X_i - \bar{X}\right)^2}{\sigma_1^2} \sim \chi^2(n-1)$$

Since $Y_1, Y_2, \ldots, Y_m$ are independent $\mathrm{N}(\mu_2, \sigma_2^2)$ random variables then by Theorem 4.3.8

$$V = \frac{(m-1)\,S_2^2}{\sigma_2^2} = \frac{\sum\limits_{i=1}^{m}\left(Y_i - \bar{Y}\right)^2}{\sigma_2^2} \sim \chi^2(m-1)$$

U and V are independent random variables since $X_1, X_2, \ldots, X_n$ are independent of $Y_1, Y_2, \ldots, Y_m$.

Now

$$\begin{aligned} \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} &= \frac{\frac{\frac{(n-1)S_1^2}{\sigma_1^2}}{n-1}}{\frac{\frac{(m-1)S_2^2}{\sigma_2^2}}{m-1}} \\ &= \frac{U/\,(n-1)}{V/\,(m-1)} \end{aligned}$$

Therefore by Theorem 4.2.11

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim \mathrm{F}\,(n-1, m-1)$$

9.4 Chapter 5

Exercise 5.4.4

If $X_n \sim$ Binomial(n, p) then

$$M_n(t) = E\left(e^{tX_n}\right) = \left(pe^t + q\right)^n \quad \text{for } t \in \Re \tag{9.7}$$

If $\mu = np$ then

$$p = \frac{\mu}{n} \quad \text{and} \quad q = 1 - \frac{\mu}{n} \tag{9.8}$$

Substituting 9.8 into 9.7 and simplifying gives

$$M_n(t) = \left(\frac{\mu}{n}e^t + 1 - \frac{\mu}{n}\right)^n = \left[1 + \frac{\mu\left(e^t - 1\right)}{n}\right]^n \quad \text{for } t \in \Re$$

Now

$$\lim_{n\to\infty}\left[1 + \frac{\mu\left(e^t - 1\right)}{n}\right]^n = e^{\mu\left(e^t - 1\right)} \quad \text{for } t < \infty$$

by Corollary 5.1.3. Since $M(t) = e^{\mu\left(e^t - 1\right)}$ for $t \in \Re$ is the moment generating function of a Poisson(μ) random variable then by Theorem 5.4.1, $X_n \to_D X \sim$ Poisson(μ).

By Theorem 5.4.2

$$P(X_n = x) = \binom{n}{x} p^x (q)^{n-x} \approx \frac{(np)^x e^{-np}}{x!} \quad \text{for } x = 0, 1, \ldots$$

Exercise 5.4.7

Let $X_i \sim$ Binomial$(1, p)$, $i = 1, 2, \ldots$ independently. Since $X_1, X_2, \ldots$ are independent and identically distributed random variables with $E(X_i) = p$ and $Var(X_i) = p(1-p)$, then

$$\frac{\sqrt{n}\left(\bar{X}_n - p\right)}{\sqrt{p(1-p)}} \to_D Z \sim \mathrm{N}(0,1)$$

by the Central Limit Theorem. Let $S_n = \sum\limits_{i=1}^{n} X_i$. Then

$$\frac{S_n - np}{\sqrt{np(1-p)}} = \frac{n\left(\frac{1}{n}\sum\limits_{i=1}^{n} X_i - p\right)}{\sqrt{n}\sqrt{p(1-p)}} = \frac{\sqrt{n}\left(\bar{X}_n - p\right)}{\sqrt{p(1-p)}} \to_D Z \sim \mathrm{N}(0,1)$$

Now by 4.3.2(1) $S_n \sim$ Binomial(n, p) and therefore Y_n and S_n have the same distribution. It follows that

$$Z_n = \frac{Y_n - np}{\sqrt{np(1-p)}} \to_D Z \sim \mathrm{N}(0,1)$$

Exercise 5.5.3

(a) Let $g(x) = x^2$ which is a continuous function for all $x \in \Re$. Since $X_n \to_p a$ then by 5.5.1(1), $X_n^2 = g(X_n) \to_p g(a) = a^2$ or $X_n^2 \to_p a^2$.

(b) Let $g(x,y) = xy$ which is a continuous function for all $(x,y) \in \Re^2$. Since $X_n \to_p a$ and $Y_n \to_p b$ then by 5.5.1(2), $X_n Y_n = g(X_n, Y_n) \to_p g(a,b) = ab$ or $X_n Y_n \to_p ab$.

(c) Let $g(x,y) = x/y$ which is a continuous function for all $(x,y) \in \Re^2$, $y \neq 0$. Since $X_n \to_p a$ and $Y_n \to_p b \neq 0$ then by 5.5.1(2), $X_n/Y_n = g(X_n, Y_n) \to_p g(a,b) = a/b$, $b \neq 0$ or $X_n/Y_n \to_p a/b$, $b \neq 0$.

(d) Let $g(x,z) = x - 2z$ which is a continuous function for all $(x,z) \in \Re^2$. Since $X_n \to_p a$ and $Z_n \to_D Z \sim \mathrm{N}(0,1)$ then by Slutsky's Theorem, $X_n - 2Z_n = g(X_n, Z_n) \to_D g(a,Z) = a - 2Z$ or $X_n - 2Z_n \to_D a - 2Z$ where $Z \sim \mathrm{N}(0,1)$. Since $a - 2Z \sim \mathrm{N}(a,4)$, therefore $X_n - 2Z_n \to_D a - 2Z \sim \mathrm{N}(a,4)$

(e) Let $g(x,z) = 1/z$ which is a continuous function for all $(x,z) \in \Re^2$, $z \neq 0$. Since $Z_n \to_D Z \sim \mathrm{N}(0,1)$ then by Slutsky's Theorem, $1/Z_n = g(X_n, Z_n) \to_D g(a,z) = 1/Z$ or $1/Z_n \to_D 1/Z$ where $Z \sim \mathrm{N}(0,1)$. Since $h(z) = 1/z$ is a decreasing function for all $z \neq 0$ then by Theorem 2.6.8 the probability density function of $W = 1/Z$ is

$$f(w) = \frac{1}{\sqrt{2\pi}} e^{-1/\left(2w^2\right)} \left(\frac{1}{w^2}\right) \quad \text{for } z \neq 0$$

Exercise 5.5.8

By (5.9)

$$\frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{\sqrt{\mu}} \to_D Z \sim \mathrm{N}(0,1)$$

and by Slutsky's Theorem

$$U_n = \sqrt{n}\left(\bar{X}_n - \mu\right) \to_D \sqrt{\mu} Z \sim \mathrm{N}(0,\mu) \tag{9.9}$$

Let $g(x) = \sqrt{x}$, $a = \mu$, and $b = 1/2$. Then $g'(x) = \frac{1}{2\sqrt{x}}$ and $g'(a) = g'(\mu) = \frac{1}{2\sqrt{\mu}}$. By (9.9) and the Delta Method

$$n^{1/2}\left(\sqrt{\bar{X}_n} - \sqrt{\mu}\right) \to_D \frac{1}{2\sqrt{\mu}}\left(\sqrt{\mu} Z\right) = \frac{1}{2} Z \sim \mathrm{N}\left(0, \frac{1}{4}\right)$$

9.5 Chapter 6

Exercise 6.4.6

The probability density function of a Exponential$(1,\theta)$ random variable is

$$f(x;\theta) = e^{-(x-\theta)} \quad \text{for } x \geq \theta \text{ and } \theta \in \Re$$

and zero otherwise. The support set of the random variable X is $[\theta,\infty)$ which depends on the unknown parameter θ.

The likelihood function is

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{n} f(x_i;\theta) \\ &= \prod_{i=1}^{n} e^{-(x_i-\theta)} \quad \text{if } x_i \geq \theta,\ i=1,2,\ldots,n \text{ and } \theta \in \Re \\ &= \left(\prod_{i=1}^{n} e^{-x_i}\right) e^{n\theta} \quad \text{if } x_i \geq \theta \text{ and } \theta \in \Re \end{aligned}$$

or more simply

$$L(\theta) = \begin{cases} 0 & \text{if } \theta > x_{(1)} \\ e^{n\theta} & \text{if } \theta \leq x_{(1)} \end{cases}$$

where $x_{(1)} = \min(x_1, x_2, \ldots, x_n)$ is the maximum of the sample. (Note: In order to observe the sample $x_1, x_2, \ldots, x_n$ the value of θ must be smaller than all the observed x_i's.) $L(\theta)$ is a increasing function of θ on the interval $(-\infty, x_{(1)]}]$. $L(\theta)$ is maximized at $\theta = x_{(1)}$. The maximum likelihood estimate of θ is $\hat{\theta} = x_{(1)}$ and the maximum likelihood estimator is $\tilde{\theta} = X_{(1)}$.

Note that in this example there is no solution to $\frac{d}{d\theta} l(\theta) = \frac{d}{d\theta}(n\theta) = 0$ and the maximum likelihood estimate of θ is not found by solving $\frac{d}{d\theta} l(\theta) = 0$.

If $n = 12$ and $x_{(1)} = 2$

$$L(\theta) = \begin{cases} 0 & \text{if } \theta > 2 \\ e^{12\theta} & \text{if } \theta \leq 2 \end{cases}$$

The relative likelihood function is

$$R(\theta) = \begin{cases} 0 & \text{if } \theta > 2 \\ e^{12(\theta-2)} & \text{if } \theta \leq 2 \end{cases}$$

is graphed in Figure 9.13 along with lines for determining 10% and 50% likelihood intervals. To determine the value of θ at which the horizontal line $R = p$ intersects the graph of $R(\theta)$ we solve $e^{12(\theta-2)} = p$ to obtain $\theta = 2 + \log p/12$. Since $R(\theta) = 0$ if $\theta > 2$ then a $100p\%$ likelihood interval for θ is of the form $[2 + \log(p)/12, 2]$. For $p = 0.1$ we obtain the 10% likelihood interval $[2 + \log(0.1)/12, 2] = [1.8081, 2]$. For $p = 0.5$ we obtain the 50% likelihood interval $[2 + \log(0.5)/12, 2] = [1.9422, 2]$.

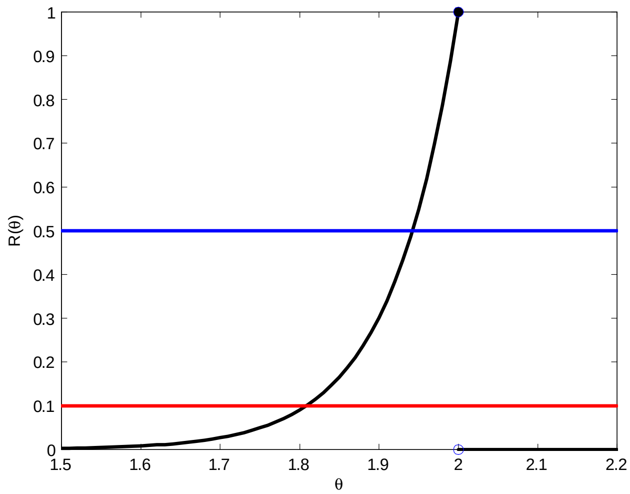


Figure 9.13: Relative likelihood function for Exercise 6.4.6

Exercise 6.7.3

By Chapter 5, Problem 7 we have

$$Q(X_n;\theta) = \frac{\sqrt{n}\left(\frac{X_n}{n} - \theta\right)}{\sqrt{\frac{X_n}{n}\left(1 - \frac{X_n}{n}\right)}} \to_D Z \sim \mathrm{N}(0,1) \tag{9.10}$$

and therefore $Q(X_n;\theta)$ is an asymptotic pivotal quantity.

Let a be the value such that $P\left(Z \leq a\right) = \left(1+p\right)/2$ where $Z \sim \mathrm{N}(0,1)$. Then by (9.10) we have

$$\begin{aligned} p &\approx P\left(-a \leq \frac{\sqrt{n}\left(\frac{X_n}{n} - \theta\right)}{\sqrt{\frac{X_n}{n}\left(1 - \frac{X_n}{n}\right)}} \leq a\right) \\ &= P\left(\frac{X_n}{n} - \frac{a}{\sqrt{n}}\sqrt{\frac{X_n}{n}\left(1 - \frac{X_n}{n}\right)} \leq \theta \leq \frac{X_n}{n} + \frac{a}{\sqrt{n}}\sqrt{\frac{X_n}{n}\left(1 - \frac{X_n}{n}\right)}\right) \end{aligned}$$

and an approximate $100p\%$ equal tail confidence interval for θ is

$$\left[\frac{x_n}{n} - \frac{a}{\sqrt{n}}\sqrt{\frac{x_n}{n}\left(1 - \frac{x_n}{n}\right)}, \frac{x_n}{n} + \frac{a}{\sqrt{n}}\sqrt{\frac{x_n}{n}\left(1 - \frac{x_n}{n}\right)}\right]$$

or

$$\left[\hat{\theta} - a\sqrt{\frac{\hat{\theta}\left(1 - \hat{\theta}\right)}{n}}, \hat{\theta} + a\sqrt{\frac{\hat{\theta}\left(1 - \hat{\theta}\right)}{n}}\right]$$

where $\hat{\theta} = x_n/n$.

Exercise 6.7.12

(a) The likelihood function is

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{n} f(x_i;\theta) = \prod_{i=1}^{n} \frac{e^{-(x-\theta)}}{\left[1+e^{-(x-\theta)}\right]^2} \\ &= \exp\left(\sum_{i=1}^{n} x_i\right) e^{n\theta} \prod_{i=1}^{n} \frac{1}{\left[1+e^{-(x_i-\theta)}\right]^2} \quad \text{for } \theta \in \Re \end{aligned}$$

or more simply

$$L(\theta) = e^{n\theta} \prod_{i=1}^{n} \frac{1}{\left[1+e^{-(x_i-\theta)}\right]^2} \quad \text{for } \theta \in \Re$$

The log likelihood function is

$$l(\theta) = n\theta - 2\sum_{i=1}^{n} \log\left(1+e^{-(x_i-\theta)}\right) \quad \text{for } \theta \in \Re$$

The score function is

$$\begin{aligned} S(\theta) &= \frac{d}{d\theta} l(\theta) = n - 2\sum_{i=1}^{n} \frac{e^{-(x_i-\theta)}}{1+e^{-(x_i-\theta)}} \\ &= n - 2\sum_{i=1}^{n} \frac{1}{1+e^{x_i-\theta}} \quad \text{for } \theta \in \Re \end{aligned}$$

Notice that $S(\theta) = 0$ cannot be solved explicitly. The maximum likelihood estimate can only be determined numerically for a given sample of data $x_1, x_2, \ldots, x_n$. Note that since

$$\frac{d}{d\theta} S(\theta) = -\left[2\sum_{i=1}^{n} \frac{e^{x_i-\theta}}{\left(1+e^{x_i-\theta}\right)^2}\right] \quad \text{for } \theta \in \Re$$

is negative for all values of $\theta > 0$ then we know that $S(\theta)$ is always decreasing so there is only one solution to $S(\theta) = 0$. Therefore the solution to $S(\theta) = 0$ gives the maximum likelihood estimate.

The information function is

$$\begin{aligned} I(\theta) &= -\frac{d^2}{d\theta^2} l(\theta) \\ &= 2\sum_{i=1}^{n} \frac{e^{x_i-\theta}}{\left(1+e^{x_i-\theta}\right)^2} \quad \text{for } \theta \in \Re \end{aligned}$$

(b) If $X \sim \text{Logistic}(\theta, 1)$ then the cumulative distribution function of X is

$$F(x;\theta) = \frac{1}{1+e^{-(x-\theta)}} \quad \text{for } x \in \Re,\ \theta \in \Re$$

Solving

$$u = \frac{1}{1+e^{-(x-\theta)}}$$

gives

$$x = \theta - \log\left(\frac{1}{u} - 1\right)$$

Therefore the inverse cumulative distribution function is

$$F^{-1}(u) = \theta - \log\left(\frac{1}{u} - 1\right) \quad \text{for } 0 < u < 1$$

If u is an observation from the Uniform$(0, 1)$ distribution then $\theta - \log\left(\frac{1}{u} - 1\right)$ is an observation from the Logistic$(\theta, 1)$ distribution by Theorem 2.6.6.

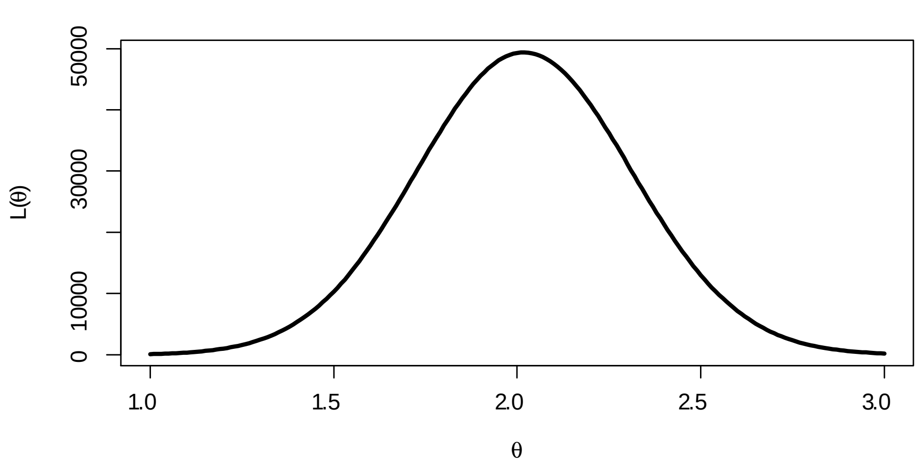
(c) The generated data are

−0.18	0.05	0.32	0.78	1.04	1.11	1.26	1.41	1.50	1.57
1.58	1.60	1.68	1.68	1.71	1.89	1.93	2.02	2.2.5	2.40
2.47	2.59	2.76	2.78	2.87	2.91	4.02	4.52	5.25	5.56

(d) Here is R code for calculating and graphing the likelihood function for these data.

```
# function for calculating Logistic information for data x and theta=th
LOLF<-function(th,x)
{n<-length(x)
L<-exp(n*th)*(prod(1+exp(th-x)))^(-2)
return(L)}
th<-seq(1,3,0.01)
L<-sapply(th,LOLF,x)
plot(th,L,"l",xlab=expression(theta),
   ylab=expression(paste("L(",theta,")")),lwd=3)
```

The graph of the likelihood function is given in Figure 9.5.(e) Here is R code for Newton's



Method. It requires functions for calculating the score and information

```
# function for calculating Logistic score for data x and theta=th
LOSF<-function(th,x)
{n<-length(x)
S<-n-2*sum(1/(1+exp(x-th)))
return(S)}
#
# function for calculating Logistic information for data x and theta=th
LOIF<-function(th,x)
{n<-length(x)
I<-2*sum(exp(x-th)/(1+exp(x-th))^2)
return(I)}
#
# Newton's Method for Logistic Example
NewtonLO<-function(th,x)
{thold<-th
thnew<-th+0.1
while (abs(thold-thnew)>0.00001)
{thold<-thnew
thnew<-thold+LOSF(thold,x)/LOIF(thold,x)
print(thnew)}
return(thnew)}
# use Newton's Method to find the maximum likelihood estimate
# use the mean of the data to begin Newton's Method
# since theta is the mean of the distribution
thetahat<-NewtonLO(mean(x),x)
cat("thetahat = ",thetahat)
```

The maximum estimate found using Newton's Method is

$$\hat{\theta} = 2.018099$$

(f) Here is R code for determining the values of $S(\hat{\theta})$ and $I(\hat{\theta})$.

```
# calculate Score(thetahat) and the observed information
Sthetahat<-LOSF(thetahat,x)
cat("S(thetahat) = ",Sthetahat)
Ithetahat<-LOIF(thetahat,x)
cat("Observed Information = ",Ithetahat)
```

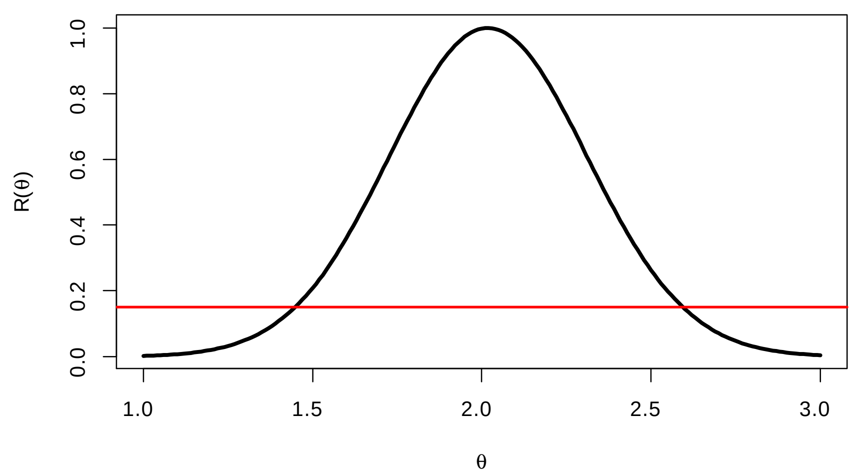
The values of $S(\hat{\theta})$ and $I(\hat{\theta})$ are

$$\begin{aligned} S(\hat{\theta}) &= 3.552714 \times 10^{-15} \\ I(\hat{\theta}) &= 11.65138 \end{aligned}$$

(*g*) Here is R code for plotting the relative likelihood function for θ based on these data.

```
# function for calculating Logistic relative likelihood function
LORLF<-function(th,thetahat,x)
{R<-LOLF(th,x)/LOLF(thetahat,x)
return(R)}
#
# plot the Logistic relative likelihood function
#plus a line to determine the 15% likelihood interval
th<-seq(1,3,0.01)
R<-sapply(th,LORLF,thetahat,x)
plot(th,R,"l",xlab=expression(theta),
   ylab=expression(paste("R(",theta,")")),lwd=3)
abline(a=0.15,b=0,col="red",lwd=2)
```

The graph of the relative likelihood function is given in Figure 9.5.



(*h*) Here is R code for determining the 15% likelihood interval and the approximate 95% confidence interval (6.18)

```
# determine a 15% likelihood interval using uniroot
uniroot(function(th) LORLF(th,thetahat,x)-0.15,lower=1,upper=1.8)$root
uniroot(function(th) LORLF(th,thetahat,x)-0.15,lower=2.2,upper=3)$root
# calculate an approximate 95% confidence intervals for theta
L95<-thetahat-1.96/sqrt(Ithetahat)
U95<-thetahat+1.96/sqrt(Ithetahat)
cat("Approximate 95% confidence interval = ",L95,U95) # display values
```

The 15% likelihood interval is

$$[1.4479, 2.593964]$$

The approximate 95% confidence interval is

$$[1.443893, 2.592304]$$

which are very close due to the symmetric nature of the likelihood function.

9.6 Chapter 7

Exercise 7.1.11

If $x_1, x_2, ..., x_n$ is an observed random sample from the Gamma(α, β) distribution then the likelihood function for α, β is

$$\begin{aligned} L(\alpha,\beta) &= \prod_{i=1}^{n} f(x_i;\alpha,\beta) \\ &= \prod_{i=1}^{n} \frac{x_i^{\alpha-1}}{\Gamma(\alpha)\beta^{\alpha}} e^{-x_i/\beta} \quad \text{for } \alpha>0, \quad \beta>0 \\ &= [\Gamma(\alpha)\beta^{\alpha}]^{-n} \left(\prod_{i=1}^{n} x_i\right)^{\alpha-1} \exp\left(-\frac{t_2}{\beta}\right) \quad \text{for } \alpha>0,\ \beta>0 \end{aligned}$$

where

$$t_2 = \sum_{i=1}^{n} x_i$$

or more simply

$$L(\alpha,\beta) = [\Gamma(\alpha)\beta^{\alpha}]^{-n} \left(\prod_{i=1}^{n} x_i\right)^{\alpha} \exp\left(-\frac{t_2}{\beta}\right) \quad \text{for } \alpha>0,\ \beta>0$$

The log likelihood function is

$$\begin{aligned} l(\alpha,\beta) &= \log L(\alpha,\beta) \\ &= -n\log\Gamma(\alpha) - n\alpha\log\beta + \alpha t_1 - \frac{t_2}{\beta} \quad \text{for } \alpha>0,\ \beta>0 \end{aligned}$$

where

$$t_1 = \sum_{i=1}^{n} \log x_i$$

The score vector is

$$\begin{aligned} S(\alpha,\beta) &= \begin{bmatrix} \frac{\partial l}{\partial\alpha} & \frac{\partial l}{\partial\beta} \end{bmatrix} \\ &= \begin{bmatrix} -n\psi(\alpha) + t_1 - n\log\beta & \frac{t_2}{\beta^2} - \frac{n\alpha}{\beta} \end{bmatrix} \end{aligned}$$

where

$$\psi(z) = \frac{d}{dz}\log\Gamma(z)$$

is the digamma function.

The information matrix is

$$\begin{aligned} I(\alpha,\beta) &= \begin{bmatrix} -\frac{\partial^2 l}{\partial\alpha^2} & -\frac{\partial^2 l}{\partial\alpha\partial\beta} \\ -\frac{\partial^2 l}{\partial\alpha\partial\beta} & -\frac{\partial^2 l}{\partial\beta^2} \end{bmatrix} \\ &= \begin{bmatrix} n\psi'(\alpha) & \frac{n}{\beta} \\ \frac{n}{\beta} & \frac{2t_2}{\beta^3} - \frac{n\alpha}{\beta^2} \end{bmatrix} \\ &= n\begin{bmatrix} \psi'(\alpha) & \frac{1}{\beta} \\ \frac{1}{\beta} & \frac{2\bar{x}}{\beta^3} - \frac{\alpha}{\beta^2} \end{bmatrix} \end{aligned}$$

where

$$\psi'(z) = \frac{d}{dz}\psi(z)$$

is the trigamma function.

The expected information matrix is

$$\begin{aligned} J(\alpha, \beta) &= E[I(\alpha, \beta); X_1, ..., X_n] \\ &= n \begin{bmatrix} \psi'(\alpha) & \frac{1}{\beta} \\ \frac{1}{\beta} & \frac{2E(\bar{X};\alpha,\beta)}{\beta^3} - \frac{\alpha}{\beta^2} \end{bmatrix} \\ &= n \begin{bmatrix} \psi'(\alpha) & \frac{1}{\beta} \\ \frac{1}{\beta} & \frac{2\alpha}{\beta^2} - \frac{\alpha}{\beta^2} \end{bmatrix} \end{aligned}$$

since $E(\bar{X}; \alpha, \beta) = \alpha\beta$.

$S(\alpha, \beta) = (0\ 0)$ must be solved numerically to find the maximum likelihood estimates of α and β.

Exercise 7.1.14

The data are

1.58	2.78	2.81	3.29	3.45	3.64	3.81	4.69	4.89	5.37
5.47	5.52	5.87	6.07	6.11	6.12	6.26	6.42	6.74	7.49
7.93	7.99	8.14	8.31	8.72	9.26	10.10	12.82	15.22	17.82

The maximum likelihood estimates of α and β can be found using Newton's Method

$$\begin{bmatrix} \alpha^{(i+1)} & \beta^{(i+1)} \end{bmatrix} = \begin{bmatrix} \alpha^{(i)} & \beta^{(i)} \end{bmatrix} + S\left(\alpha^{(i)}, \beta^{(i)}\right)\left[I\left(\alpha^{(i)}, \beta^{(i)}\right)\right]^{-1} \quad \text{for } i = 0, 1, \ldots$$

Here is R code for Newton's Method for the Gamma Example.

```
# function for calculating Gamma score for a and b and data x
GASF<-function(a,b,x))
{t1<-sum(log(x))
t2<-sum(x)
n<-length(x)
S<-c(t1-n*(digamma(a)+log(b)),t2/b^2-n*a/b)
return(S)}
#
# function for calculating Gamma information for a and b and data x
GAIF<-function(a,b,x)
{I<-length(x)*cbind(c(trigamma(a),1/b),c(1/b,2*mean(x)/b^3-a/b^2))
return(I)}
#
# Newton's Method for Gamma Example
NewtonGA<-function(a,b,x)
{thold<-c(a,b)
thnew<-thold+0.1
while (sum(abs(thold-thnew))>0.0000001)
{thold<-thnew
thnew<-thold+GASF(thold[1],thold[2],x)%*%solve(GAIF(thold[1],thold[2]))
print(thnew)}
return(thnew)}
thetahat<-NewtonGA(2,2,x)
```

The maximum likelihood estimates are $\hat{\alpha} = 4.118407$ and $\hat{\beta} = 1.657032$.

The score vector evaluated at $(\hat{\alpha}, \hat{\beta})$ is

$$S(\hat{\alpha}, \hat{\beta}) = \begin{bmatrix} 0 & 1.421085 \times 10^{-14} \end{bmatrix}$$

which indicates we have obtained a local extrema.

The observed information matrix is

$$I(\hat{\alpha}, \hat{\beta}) = \begin{bmatrix} 8.239505 & 18.10466 \\ 18.10466 & 44.99752 \end{bmatrix}$$

Note that since $\det[I(\hat{\alpha}, \hat{\beta})] = (8.239505)(44.99752) - (18.10466)^2 > 0$ and $[I(\hat{\alpha}, \hat{\beta})]_{11} = 8.239505 > 0$ then by the second derivative test we have found the maximum likelihood estimates.

Exercise 7.2.3

(a) The following R code generates the required likelihood regions.

```
# function for calculating Gamma relative likelihood for parameters a and b and
data x
GARLF<-function(a,b,that,x)
{t<-prod(x)
t2<-sum(x)
n<-length(x)
ah<-that[1]
bh<-that[2]
L<-((gamma(ah)*bh^ah)/(gamma(a)*b^a))^n
   *t^(a-ah)*exp(t2*(1/bh-1/b))
return(L)}
a<-seq(1,8.5,0.02)
b<-seq(0.2,4.5,0.01)
R<-outer(a,b,FUN = GARLF,thetahat,x)
contour(a,b,R,levels=c(0.01,0.05,0.10,0.50,0.9),xlab="a",ylab="b",lwd=2)
```

The 1%, 5%, 10%, 50%, and 90% likelihood regions for (α, β) are shown in Figure 9.14.

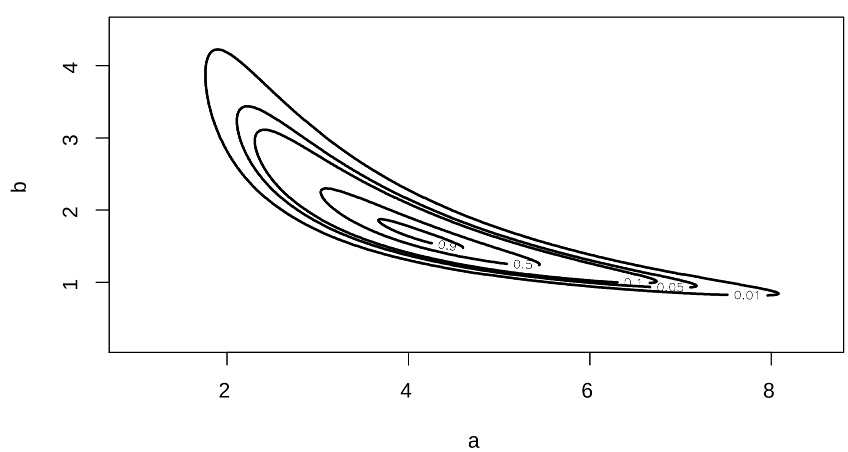


Figure 9.14: Likelihood regions for Gamma for 30 observations

The likelihood contours are not very elliptical in shape. The contours suggest that large values of α together with small values of β or small values of α together with large values of β are plausible given the observed data.

(b) Since $R(3, 2.7) = 0.14$ the point $(3, 2.7)$ lies inside a 10% likelihood region so it is a plausible value of (α, β).

(d) The 1%, 5%, 10%, 50%, and 90% likelihood regions for (α, β) for 100 observations are shown in Figure 9.15. We note that for a larger number of observations the likelihood regions are more elliptical in shape.

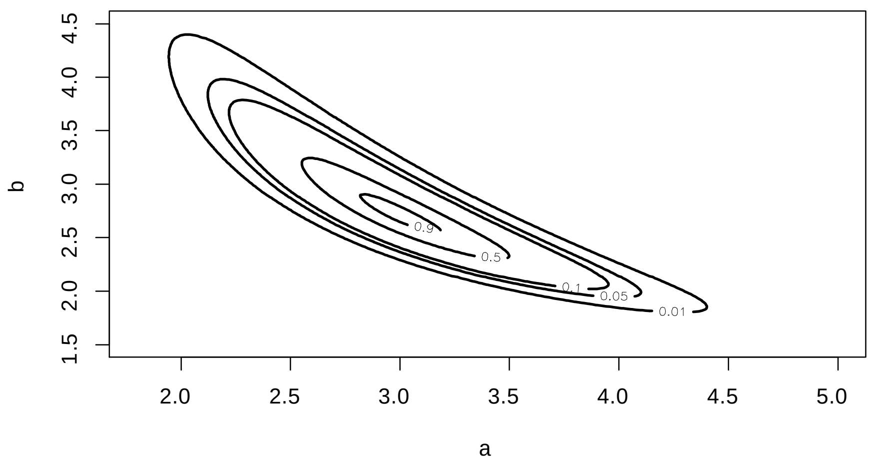


Figure 9.15: Likelihood regions for Gamma for 100 observations

Exercise 7.4.4

The following R code graphs the approximate confidence regions. The function ConfRegion was used in Example 7.4.3.

```
# graph approximate confidence regions
c<-outer(a,b,FUN = ConfRegion,thetahat,Ithetahat)
contour(a,b,c,levels=c(4.61,5.99,9.21),xlab="a",ylab="b",lwd=2)
```

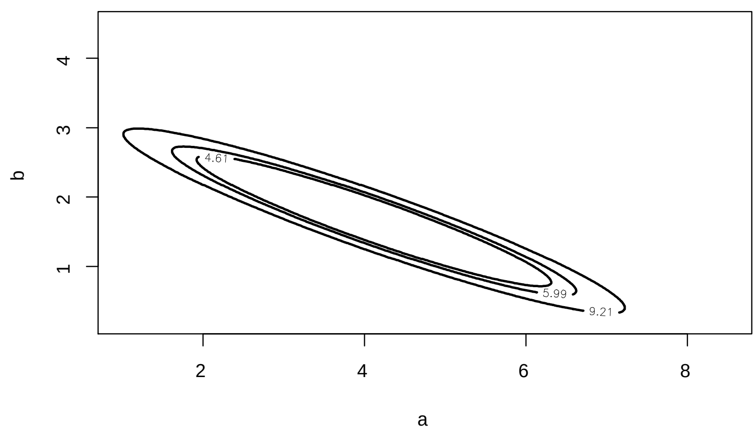


Figure 9.16: Approximate confidence regions for Gamma for 30 observations

These approximate confidence regions which are ellipses are very different than the likelihood regions in Figure 9.14. In particular we note that $(\alpha, \beta) = (3, 2.7)$ lies inside a 10% likelihood region but outside a 99% approximate confidence region.

There are only 30 observations and these differences suggest the Normal approximation is not very good. The likelihood regions are a better summary of the uncertainty in the estimates.

Exercise 7.4.7

Let

$$\left[I(\hat{\alpha},\hat{\beta})\right]^{-1} = \begin{bmatrix} \hat{v}_{11} & \hat{v}_{12} \\ \hat{v}_{12} & \hat{v}_{22} \end{bmatrix}$$

Since

$$\begin{bmatrix} \tilde{\alpha}-\alpha & \tilde{\beta}-\beta \end{bmatrix} [J(\tilde{\alpha},\tilde{\beta})]^{1/2} \to_D Z \sim \text{BVN}\left(\begin{bmatrix} 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right)$$

then for large n, $Var(\tilde{\alpha}) \approx \hat{v}_{11}$, $Var(\tilde{\beta}) \approx \hat{v}_{22}$ and $Cov(\tilde{\alpha},\tilde{\beta}) \approx \hat{v}_{12}$. Therefore an approximate 95% confidence interval for a is given by

$$[\hat{\alpha} - 1.96\sqrt{\hat{v}_{11}},\ \hat{\alpha} + 1.96\sqrt{\hat{v}_{11}}]$$

and an approximate 95% confidence interval for b is given by

$$[\hat{\beta} - 1.96\sqrt{\hat{v}_{22}},\ \hat{\beta} + 1.96\sqrt{\hat{v}_{22}}]$$

For the data in Exercise 7.1.14, $\hat{\alpha} = 4.118407$ and $\hat{\beta} = 1.657032$ and

$$\begin{aligned} &\left[I(\hat{\alpha},\hat{\beta})\right]^{-1} \\ = &\begin{bmatrix} 8.239505 & 18.10466 \\ 18.10466 & 44.99752 \end{bmatrix}^{-1} \\ = &\begin{bmatrix} 1.0469726 & -0.4212472 \\ -0.4212472 & 0.1917114 \end{bmatrix} \end{aligned}$$

An approximate 95% marginal confidence interval for α is

$$[4.118407 + 1.96\sqrt{1.0469726},\ 4.118407 - 1.96\sqrt{1.0469726}] = [2.066341, 6.170473]$$

An approximate 95% confidence interval for β is

$$[1.657032 - 1.96\sqrt{0.1917114},\ 1.657032 + 1.96\sqrt{0.1917114}] = [1.281278, 2.032787]$$

To obtain an approximate 95% marginal confidence interval for $\alpha + \beta$ we note that

$$\begin{aligned} Var(\tilde{\alpha} + \tilde{\beta}) &= Var(\tilde{\alpha}) + Var(\tilde{\beta}) + 2Cov(\tilde{\alpha}, \tilde{\beta}) \\ &\approx \hat{v}_{11} + \hat{v}_{22} + 2\hat{v}_{12} = \hat{v} \end{aligned}$$

so that an approximate 95% confidence interval for $\alpha + \beta$ is given by

$$[\hat{\alpha} + \hat{\beta} - 1.96\sqrt{\hat{v}},\ \hat{\alpha} + \hat{\beta} + 1.96\sqrt{\hat{v}}]$$

For the data in Example 7.1.13

$$\begin{aligned} \hat{\alpha} + \hat{\beta} &= 4.118407 + 1.657032 = 5.775439 \\ \hat{v} &= \hat{v}_{11} + \hat{v}_{22} + 2\hat{v}_{12} = 1.0469726 + 0.1917114 + 2(-0.4212472) = 0.3961895 \end{aligned}$$

and an approximate 95% marginal confidence interval for $\alpha + \beta$ is

$$[5.775439 + 1.96\sqrt{0.3961895},\ 5.775439 - 1.96\sqrt{0.3961895}] = [4.998908, 6.551971]$$

9.7 Chapter 8

Exercise 8.1.7

(a) Let X = number of successes in n trails. Then $X \sim \text{Binomial}(n, \theta)$ and $E(X) = n\theta$. If the null hypothesis is $H_0 : \theta = \theta_0$ and the alternative hypothesis is $H_A : \theta \neq \theta_0$ then a suitable test statistic is $D = |X - n\theta_0|$.

For $n = 100$, $x = 42$, and $\theta_0 = 0.5$ the observed value of D is $d = |x - n\theta_0| = |42 - 50| = 8$. The *p-value* is

$$\begin{aligned}
& P\left(|X - n\theta_0| \geq |x - n\theta_0|\,; H_0 : \theta = \theta_0\right) \\
= \;& P\left(|X - 50| \geq 8\right) \quad \text{where } X \sim \text{Binomial}\,(100, 0.5) \\
= \;& P\left(X \leq 42\right) + P\left(X \geq 58\right) \\
= \;& 0.06660531 + 0.06660531 \\
= \;& 0.1332106
\end{aligned}$$

calculated using R. Since *p-value* > 0.1 there is no evidence based on the data against $H_0 : \theta = 0.5$.

(b) If the null hypothesis is $H_0 : \theta = \theta_0$ and the alternative hypothesis is $H_A : \theta < \theta_0$ then a suitable test statistic is $D = n\theta_0 - X$.

For $n = 100$, $x = 42$, and $\theta_0 = 0.5$ the observed value of D is $d = n\theta_0 - x = 50 - 42 = 8$. The *p-value* is

$$\begin{aligned}
& P\left(n\theta_0 - X \geq n\theta_0 - x; H_0 : \theta = \theta_0\right) \\
= \;& P\left(50 - X \geq 8\right) \quad \text{where } X \sim \text{Binomial}\,(50, 0.5) \\
= \;& P\left(X \leq 42\right) \\
= \;& 0.06660531
\end{aligned}$$

calculated using R. Since $0.05 <$ *p-value* < 0.1 there is weak evidence based on the data against $H_0 : \theta = 0.5$.

Exercise 8.2.5

The model for these data is $(X_1, X_2, \ldots, X_7) \sim \text{Multinomial}(63, \theta_1, \theta_2, \ldots, \theta_7)$ and the hypothesis of interest is $H_0 : \theta_1 = \theta_2 = \cdots = \theta_7 = \frac{1}{7}$. Since the model and parameters are completely specified this is a simple hypothesis. Since $\sum\limits_{j=1}^{7} \theta_j = 1$ there are only $k = 6$ parameters.

The likelihood ratio test statistic, which can be derived in the same way as Example 8.2.4, is

$$\Lambda\left(\mathbf{X}; \boldsymbol{\theta}_0\right) = 2 \sum_{i=1}^{7} X_j \log\left(\frac{X_j}{E_j}\right)$$

where $E_j = 63/7$ is the expected frequency for outcome j.

For these data the observed value of the likelihood ratio test statistic is

$$\begin{aligned}
\lambda(\mathbf{x};\boldsymbol{\theta}_0) &= 2\sum_{i=1}^{7} x_j \log\left(\frac{x_j}{63/7}\right) \\
&= 2\left[22\log\left(\frac{22}{63/7}\right) + 7\log\left(\frac{7}{63/7}\right) + \cdots + 6\log\left(\frac{6}{63/7}\right)\right] \\
&= 23.27396
\end{aligned}$$

The approximate *p-value* is

$$\begin{aligned}
p\text{-}value &\approx P\left(W \geq 23.27396\right) \quad \text{where } W \sim \chi^2(6) \\
&= 0.0007
\end{aligned}$$

calculated using R. Since p-*value* < 0.001 there is strong evidence based on the data against the hypothesis that the deaths are equally likely to occur on any day of the week.

Exercise 8.3.3

(*a*) $\Omega = \{(\theta_1, \theta_2) : \theta_1 > 0,\ \theta_2 > 0\}$ which has dimension $k = 2$ and
$\Omega_0 = \{(\theta_1, \theta_2) : \theta_1 = \theta_2,\ \theta_1 > 0,\ \theta_2 > 0\}$ which has dimension $q = 1$ and the hypothesis $H_0 : \theta_1 = \theta_2$ is composite.

From Example 6.2.5 the likelihood function for an observed random sample $x_1, x_2, \ldots, x_n$ from an Poisson(θ_1) distribution is

$$L_1(\theta_1) = \theta_1^{n\bar{x}} e^{-n\theta_1} \quad \text{for } \theta_1 \geq 0$$

with maximum likelihood estimate $\hat{\theta}_1 = \bar{x}$.

Similarly the likelihood function for an observed random sample $y_1, y_2, \ldots, y_m$ from an Poisson(θ_2) distribution is

$$L_2(\theta_2) = \theta_2^{n\bar{y}} e^{-n\theta_2} \quad \text{for } \theta_2 \geq 0$$

with maximum likelihood estimate $\hat{\theta}_2 = \bar{y}$. Since the samples are independent the likelihood function for (θ_1, θ_2) is

$$L\left(\theta_1, \theta_2\right) = L_1(\theta_1) L_2(\theta_2) \quad \text{for } \theta_1 \geq 0, \theta_2 \geq 0$$

and the log likelihood function

$$l\left(\theta_1, \theta_2\right) = n\bar{x}\log\theta_1 - n\theta_1 + m\bar{y}\log\theta_2 - m\theta_2 \quad \text{for } \theta_1 > 0, \theta_2 > 0$$

The independence of the samples also implies the maximum likelihood estimators are
$\tilde{\theta}_1 = \bar{X}$ and $\tilde{\theta}_2 = \bar{Y}$. Therefore

$$\begin{aligned}
l(\tilde{\theta}_1, \tilde{\theta}_2; \mathbf{X}, \mathbf{Y}) &= n\bar{X}\log\bar{X} - n\bar{X} + m\bar{Y}\log\bar{Y} - m\bar{Y} \\
&= n\bar{X}\log\bar{X} + m\bar{Y}\log\bar{Y} - \left(n\bar{X} + m\bar{Y}\right)
\end{aligned}$$

If $\theta_1 = \theta_2 = \theta$ then the log likelihood function is

$$l(\theta) = (n\bar{x} + m\bar{y})\log\theta - (n+m)\theta \quad \text{for } \theta > 0$$

which is only a function of θ. To determine $\max\limits_{(\theta_1,\theta_2)\in\Omega_0} l(\theta_1,\theta_2;\mathbf{X},\mathbf{Y})$ we note that

$$\frac{d}{d\theta}l(\theta) = \frac{n\bar{x} + m\bar{y}}{\theta} - (n+m)$$

and $\frac{d}{d\theta}l(\theta) = 0$ for $\theta = \frac{n\bar{x}+m\bar{y}}{n+m}$ and therefore

$$\max_{(\theta_1,\theta_2)\in\Omega_0} l(\theta_1,\theta_2;\mathbf{X},\mathbf{Y}) = \left(n\bar{X} + m\bar{Y}\right)\log\left(\frac{n\bar{X}+m\bar{Y}}{n+m}\right) - \left(n\bar{X} + m\bar{Y}\right)$$

The likelihood ratio test statistic is

$$\begin{aligned}
&\Lambda(\mathbf{X},\mathbf{Y};\Omega_0)\\
= \; & 2\left[l(\tilde{\theta}_1,\tilde{\theta}_2;\mathbf{X},\mathbf{Y}) - \max_{(\theta_1,\theta_2)\in\Omega_0} l(\theta_1,\theta_2;\mathbf{X},\mathbf{Y})\right]\\
= \; & 2\left[n\bar{X}\log\bar{X} + m\bar{Y}\log\bar{Y} - \left(n\bar{X}+m\bar{Y}\right) - \left(n\bar{X}+m\bar{Y}\right)\log\left(\frac{n\bar{X}+m\bar{Y}}{n+m}\right) + \left(n\bar{X}+m\bar{Y}\right)\right]\\
= \; & 2\left[n\bar{X}\log\bar{X} + m\bar{Y}\log\bar{Y} - \left(n\bar{X}+m\bar{Y}\right)\log\left(\frac{n\bar{X}+m\bar{Y}}{n+m}\right)\right]
\end{aligned}$$

with corresponding observed value

$$\lambda(\mathbf{x},\mathbf{y};\Omega_0) = 2\left[n\bar{x}\log\bar{x} + m\bar{y}\log\bar{y} - (n\bar{x}+m\bar{y})\log\left(\frac{n\bar{x}+m\bar{y}}{n+m}\right)\right]$$

Since $k - q = 2 - 1 = 1$

$$\begin{aligned}
p\text{-}value &\approx P\left[W \geq \lambda(\mathbf{x},\mathbf{y};\Omega_0)\right] \quad \text{where } W \sim \chi^2(1)\\
&= 2\left[1 - P\left(Z \leq \sqrt{\lambda(\mathbf{x},\mathbf{y};\Omega_0)}\right)\right] \quad \text{where } Z \sim \mathrm{N}(0,1)
\end{aligned}$$

(b) For $n = 10$, $\sum\limits_{i=1}^{10} x_i = 22$, $m = 15$, $\sum\limits_{i=1}^{15} y_i = 40$ the observed value of the likelihood ratio test statistic is $\lambda(\mathbf{x},\mathbf{y};\Omega_0) = 0.5344026$ and

$$\begin{aligned}
p\text{-}value &\approx P\left[W \geq 0.5344026\right] \quad \text{where } W \sim \chi^2(1)\\
&= 2\left[1 - P\left(Z \leq \sqrt{0.5344026}\right)\right] \quad \text{where } Z \sim \mathrm{N}(0,1)\\
&= 0.4647618
\end{aligned}$$

calculated using R. Since *p-value* > 0.1 there is no evidence against $H_0: \theta_1 = \theta_2$ based on the data.

10. Solutions to Selected End of Chapter Problems

10.1 Chapter 2

1(a) Starting with

$$\sum_{x=0}^{\infty} \theta^x = \frac{1}{1-\theta} \quad \text{for } |\theta| < 1$$

it can be shown that

$$\sum_{x=1}^{\infty} x\theta^x = \frac{\theta}{(1-\theta)^2} \quad \text{for } |\theta| < 1 \tag{10.1}$$

$$\sum_{x=1}^{\infty} x^2\theta^{x-1} = \frac{1+\theta}{(1-\theta)^3} \quad \text{for } |\theta| < 1 \tag{10.2}$$

$$\sum_{x=1}^{\infty} x^3\theta^{x-1} = \frac{1+4\theta+\theta^2}{(1-\theta)^4} \quad \text{for } |\theta| < 1 \tag{10.3}$$

(1)

$$\frac{1}{k} = \sum_{x=1}^{\infty} x\theta^x = \frac{\theta}{(1-\theta)^2} \quad \text{using (10.1)} \quad \text{gives} \quad k = \frac{(1-\theta)^2}{\theta}$$

and therefore

$$f(x) = (1-\theta)^2 x\theta^{x-1} \quad \text{for } x = 1, 2, ...; \ 0 < \theta < 1$$

The graph of $f(x)$ in Figure 10.1 is for $\theta = 0.3$.

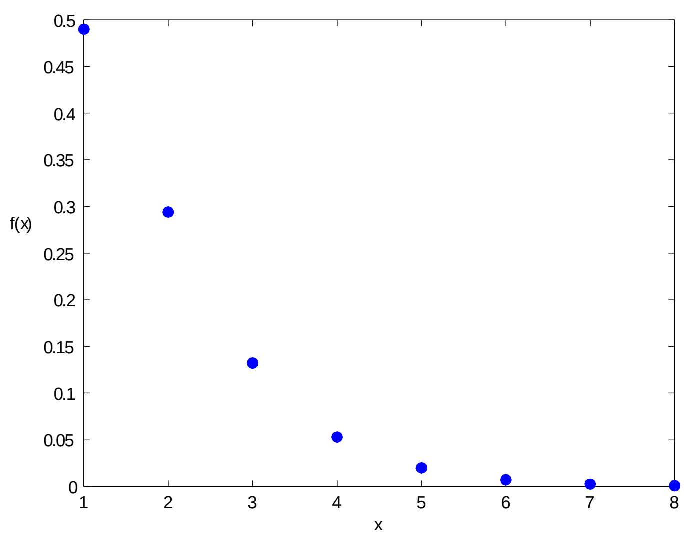


Figure 10.1: Graph of $f(x)$

(2)

$$F(x) = \begin{cases} 0 & \text{if } x < 1 \\ \sum\limits_{t=1}^{x} (1-\theta)^2 t\theta^{t-1} = 1 - (1 + x - x\theta)\theta^x & \text{for } x = 1, 2, ... \end{cases}$$

Note that $F(x)$ is specified by indicating its value at each jump point.

(3)

$$\begin{aligned}
E(X) &= \sum_{x=1}^{\infty} x(1-\theta)^2 x\theta^{x-1} = (1-\theta)^2 \sum_{x=1}^{\infty} x^2\theta^{x-1} \\
&= (1-\theta)^2 \frac{(1+\theta)}{(1-\theta)^3} \qquad \text{using (10.2)} \\
&= \frac{(1+\theta)}{(1-\theta)}
\end{aligned}$$

$$\begin{aligned}
E\left(X^2\right) &= \sum_{x=1}^{\infty} x^2(1-\theta)^2 x^{x-1} = (1-\theta)^2 \sum_{x=1}^{\infty} x^3\theta^{x-1} \\
&= (1-\theta)^2 \frac{(1+4\theta+\theta^2)}{(1-\theta)^4} \qquad \text{using (10.3)} \\
&= \frac{(1+4\theta+\theta^2)}{(1-\theta)^2}
\end{aligned}$$

$$Var(X) = E(X^2) - [E(X)]^2 = \frac{(1+4\theta+\theta^2)}{(1-\theta)^2} - \frac{(1+\theta)^2}{(1-\theta)^2} = \frac{2\theta}{(1-\theta)^2}$$

(4) Using $\theta = 0.3$,

$$\begin{aligned}
P(0.5 < X \le 2) &= P(X=1) + P(X=2) \\
&= 0.49 + (0.49)(2)(0.3) = 0.784
\end{aligned}$$

$$\begin{aligned}
P(X &> 0.5|X \le 2) = \frac{P(X > 0.5, X \le 2)}{P(X \le 2)} \\
&= \frac{P(X=1) + P(X=2)}{P(X \le 2)} = 1
\end{aligned}$$

1.(b) (1)

$$\begin{aligned}
\frac{1}{k} &= \int_{-\infty}^{\infty} \frac{1}{1+(x/\theta)^2} dx = 2\int_{0}^{\infty} \frac{1}{1+(x/\theta)^2} dx \quad \text{because of symmetry} \\
&= 2\theta \int_{0}^{\infty} \frac{1}{1+y^2} dy \quad \text{let } y = \frac{1}{\theta}x, \quad \theta dy = dx \\
&= 2\theta \lim_{b\to\infty} \arctan(b) = 2\theta\left(\frac{\pi}{2}\right) = \pi\theta
\end{aligned}$$

Thus $k = \frac{1}{\pi\theta}$ and

$$f(x) = \frac{1}{\pi\theta\left[1+(x/\theta)^2\right]} \quad \text{for } x \in \Re,\ \theta > 0$$

The graphs for $\theta = 0.5$, 1 and 2 are plotted in Figure 10.2. The graph for each different value of θ is obtained from the graph for $\theta = 1$ by simply relabeling the x and y axes. That is, on the x axis, each point x is relabeled $x\theta$ and on the y axis, each point y is relabeled y/θ. The graph of $f(x)$ below is for $\theta = 1$. Note that the graph of $f(x)$ is symmetric about the y axis.

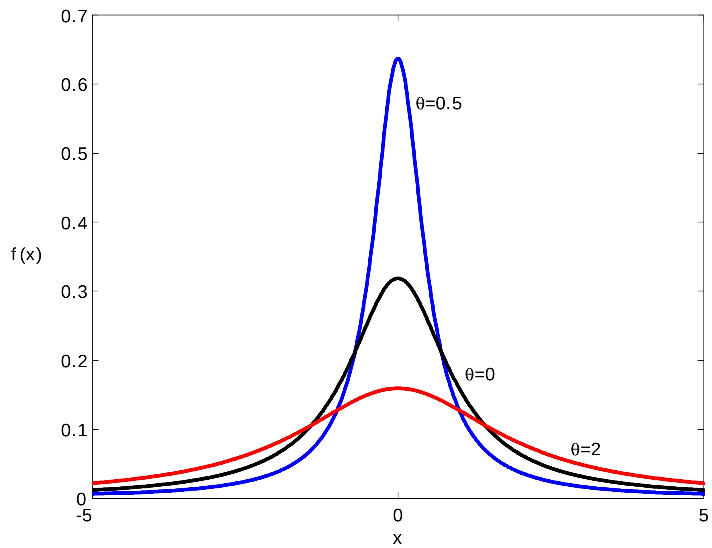


Figure 10.2: Cauchy probability density functions

(2)

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} \frac{1}{\pi\theta\left[1+(t/\theta)^2\right]}dt = \frac{1}{\pi}\left[\lim_{b\to-\infty} \arctan\left(\frac{t}{\theta}\right)|_b^x\right] \\ &= \frac{1}{\pi}\arctan\left(\frac{x}{\theta}\right)+\frac{1}{2} \quad \text{for } x \in \Re \end{aligned}$$

(3) Consider the integral

$$\int_0^{\infty} \frac{x}{\pi\theta\left[1+(x/\theta)^2\right]}dx = \theta\int_0^{\infty}\frac{t}{1+t^2}dt = \theta \lim_{b\to\infty}\frac{1}{2}\ln\left(1+b^2\right) = +\infty$$

Since the integral

$$\int_0^{\infty} \frac{x}{\pi\theta\left[1+(x/\theta)^2\right]}dx$$

does not converge, the integral

$$\int_{-\infty}^{\infty} \frac{x}{\pi\theta\left[1+(x/\theta)^2\right]}dx$$

does not converge absolutely and $E(X)$ does not exist. Since $E(X)$ does not exist $Var(X)$ does not exist.

(4) Using $\theta = 1$,

$$\begin{aligned} P(0.5 &< X \leq 2) = F(2) - F(0.5) \\ &= \frac{1}{\pi}\left[\arctan(2) - \arctan(0.5)\right] \\ &\approx 0.2048 \end{aligned}$$

$$\begin{aligned} P(X &> 0.5 | X \leq 2) = \frac{P(X > 0.5, X \leq 2)}{P(X \leq 2)} \\ &= \frac{P(0.5 < X \leq 2)}{P(X \leq 2)} = \frac{F(2) - F(0.5)}{F(2)} \\ &= \frac{\arctan(2) - \arctan(0.5)}{\arctan(2) + \frac{\pi}{2}} \approx 0.2403 \end{aligned}$$

1.(c) (1)

$$\begin{aligned} \frac{1}{k} &= \int_{-\infty}^{\infty} e^{-|x-\theta|} dx \quad \text{let } y = x - \theta, \text{ then } dy = dx \\ &= \int_{-\infty}^{\infty} e^{-|y|} dy = 2\int_{0}^{\infty} e^{-y} dy \quad \text{by symmetry} \\ &= 2\Gamma(1) = 2(0!) = 2 \end{aligned}$$

Thus $k = \frac{1}{2}$ and

$$f(x) = \frac{1}{2} e^{-|x-\theta|} \quad \text{for } x \in \Re, \ \theta \in \Re$$

The graphs for $\theta = -1$, 0 and 2 are plotted in Figure 10.3. The graph for each different value of θ is obtained from the graph for $\theta = 0$ by simply shifting the graph for $\theta = 0$ to the right θ units if θ is positive and shifting the graph for $\theta = 0$ to the left θ units if θ is negative. Note that the graph of $f(x)$ is symmetric about the line $x = \theta$.

(2)

$$F(x) = \begin{cases} \int\limits_{-\infty}^{x} \frac{1}{2} e^{t-\theta} dt & x \leq \theta \\ \int\limits_{-\infty}^{\theta} \frac{1}{2} e^{t-\theta} dt + \int\limits_{\theta}^{x} \frac{1}{2} e^{-t+\theta} dt & x > \theta \end{cases}$$

$$= \begin{cases} \frac{1}{2} e^{x-\theta} & x \leq \theta \\ \frac{1}{2} + \left(\frac{-1}{2} e^{-t+\theta} |_{\theta}^{x}\right) = 1 - \frac{1}{2} e^{-x+\theta} & x > \theta \end{cases}$$

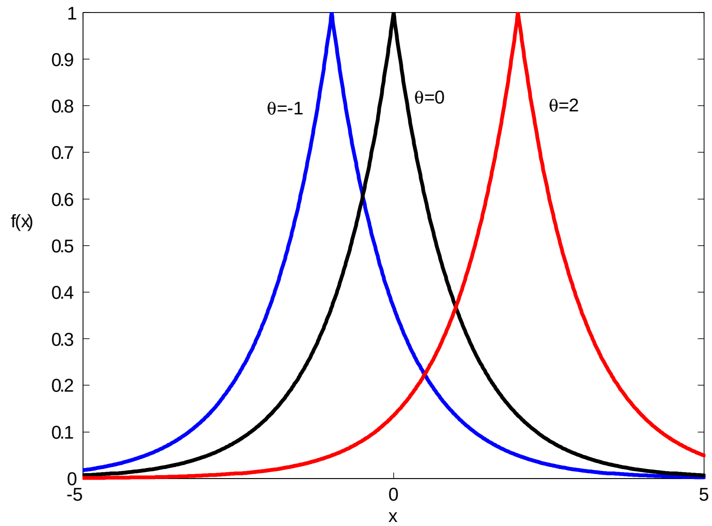


Figure 10.3: Double Exponential probability density functions

(3) Since the improper integral

$$\int_{\theta}^{\infty} (x-\theta)\, e^{-(x-\theta)} dx = \int_{0}^{\infty} y e^{-y} dy = \Gamma(2) = 1! = 1$$

converges, the integral $\frac{1}{2} \int_{-\infty}^{\infty} x e^{-|x-\theta|} dx$ converges absolutely and by the symmetry of $f(x)$ we have $E(X) = \theta$.

$$\begin{aligned}
E\left(X^2\right) &= \frac{1}{2} \int_{-\infty}^{\infty} x^2 e^{-|x-\theta|} dx \qquad \text{let } y = x-\theta, \text{ then } dy = dx \\
&= \frac{1}{2} \int_{-\infty}^{\infty} (y+\theta)^2 e^{-|y|} dy = \frac{1}{2} \int_{-\infty}^{\infty} \left(y^2 + 2y\theta + \theta^2\right) e^{-|y|} dy \\
&= \int_{0}^{\infty} y^2 e^{-y} dy + 0 + \theta^2 \int_{0}^{\infty} e^{-y} dy \ \text{ using the properties of even/odd functions} \\
&= \Gamma(3) + \theta^2 \Gamma(1) = 2! + \theta^2 \\
&= 2 + \theta^2
\end{aligned}$$

Therefore

$$Var(X) = E(X^2) - [E(X)]^2 = 2 + \theta^2 - \theta^2 = 2$$

(4) Using $\theta = 0$,

$$P(0.5 < X \le 2) = F(2) - F(0.5) = \frac{1}{2}(e^{-0.5} - e^{-2}) \approx 0.2356$$

$$\begin{aligned} P(X &> 0.5|X \leq 2) = \frac{P(X > 0.5, X \leq 2)}{P(X \leq 2)} = \frac{P(0.5 < X \leq 2)}{P(X \leq 2)} = \frac{F(2) - F(0.5)}{F(2)} \\ &= \frac{\frac{1}{2}(e^{-0.5} - e^{-2})}{1 - \frac{1}{2}e^{-2}} \approx 0.2527 \end{aligned}$$

1.(e) (1)

$$\begin{aligned} \frac{1}{k} &= \int_0^\infty x^2 e^{-\theta x} dx \quad \text{let } y = \theta x, \;\; \frac{1}{\theta} dy = dx \\ &= \frac{1}{\theta^3} \int_0^\infty y^2 e^{-y} dy = \frac{1}{\theta^3} \Gamma(3) = \frac{2!}{\theta^3} = \frac{2}{\theta^3} \end{aligned}$$

Thus $k = \frac{\theta^3}{2}$ and

$$f(x) = \frac{1}{2}\theta^3 x^2 e^{-\theta x} \quad \text{for } x \geq 0, \; \theta > 0$$

The graphs for $\theta = 0.5$, 1 and 2 are plotted in Figure 10.4. The graph for each different value of θ is obtained from the graph for $\theta = 1$ by simply relabeling the x and y axes. That is, on the x axis, each point x is relabeled x/θ and on the y axis, each point y is relabeled θy.

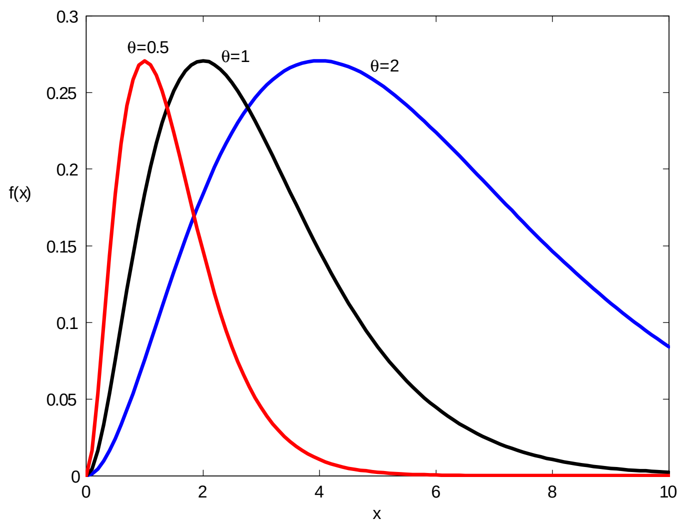


Figure 10.4: Gamma probabilty density functions

(2)

$$F(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ \frac{1}{2}\int_0^x \theta^3 t^2 e^{-\theta t} dt & \text{if } x > 0 \end{cases}$$

Using integration by parts twice we have

$$\begin{aligned}
\frac{1}{2}\int_0^x \theta^3 t^2 e^{-\theta t} dt &= \frac{1}{2}\left[-(\theta t)^2 e^{-\theta t}|_0^x + 2\int_0^x \theta^2 t e^{-\theta t} dt\right] \\
&= \frac{1}{2}\left[-(\theta x)^2 e^{-\theta x} + 2\left\{-(\theta t) e^{-\theta t}|_0^x + \int_0^x e^{-\theta t} dt\right\}\right] \\
&= \frac{1}{2}\left[-(\theta x)^2 e^{-\theta x} - 2(\theta x) e^{-\theta x} - 2\left\{e^{-\theta t}|_0^x\right\}\right] \\
&= \frac{1}{2}\left[-(\theta x)^2 e^{-\theta x} - 2(\theta x) e^{-\theta x} - 2e^{-\theta x} + 2\right] \\
&= 1 - \frac{1}{2}e^{-\theta x}\left(\theta^2 x^2 + 2\theta x + 2\right) \quad \text{for } x > 0
\end{aligned}$$

Therefore

$$F(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ 1 - \frac{1}{2}e^{-\theta x}\left(\theta^2 x^2 + 2\theta x + 2\right) & \text{if } x > 0 \end{cases}$$

(3)

$$\begin{aligned}
E(X) &= \frac{1}{2}\int_0^\infty \theta^3 x^3 e^{-\theta x} dx \quad \text{let } y = \theta x, \ \ \frac{1}{\theta}dy = dx \\
&= \frac{1}{2\theta}\int_0^\infty y^3 e^{-y} dy = \frac{1}{2\theta}\Gamma(4) = \frac{3!}{2\theta} = \frac{3}{\theta}
\end{aligned}$$

$$\begin{aligned}
E\left(X^2\right) &= \frac{1}{2}\int_0^\infty \theta^3 x^4 e^{-\theta x} dx \quad \text{let } y = \theta x, \ \ \frac{1}{\theta}dy = dx \\
&= \frac{1}{2\theta^2}\int_0^\infty y^4 e^{-y} dy = \frac{1}{2\theta^2}\Gamma(5) = \frac{4!}{2\theta^2} = \frac{12}{\theta^2}
\end{aligned}$$

and

$$Var(X) = E\left(X^2\right) - [E(X)]^2 = \frac{12}{\theta^2} - \left(\frac{3}{\theta}\right)^2 = \frac{3}{\theta^2}$$

(4) Using $\theta = 1$,

$$\begin{aligned}
P(0.5 &< X \leq 2) = F(2) - F(0.5) \\
&= \left[1 - \frac{1}{2}e^{-x}\left(x^2 + 2x + 2\right)\right]|_{0.5}^{2} \\
&= \left[1 - \frac{1}{2}e^{-2}(4 + 4 + 2)\right] - \left[1 - \frac{1}{2}e^{-0.5}\left(\frac{1}{4} + 1 + 2\right)\right] \\
&= \frac{1}{2}\left[e^{-0.5}\left(\frac{13}{4}\right) - e^{-2}(10)\right] \\
&\approx 0.3089
\end{aligned}$$

$$\begin{aligned}
P(X &> 0.5|X \leq 2) = \frac{P(X > 0.5, X \leq 2)}{P(X \leq 2)} \\
&= \frac{P(0.5 < X \leq 2)}{P(X \leq 2)} = \frac{F(2) - F(0.5)}{F(2)} \\
&= \frac{\frac{1}{2}\left[e^{-0.5}\left(\frac{13}{4}\right) - e^{-2}(10)\right]}{1 - \frac{1}{2}e^{-2}(10)} \\
&\approx 0.9555
\end{aligned}$$

1.(g) (1) Since

$$f(-x) = \frac{e^{x/\theta}}{\theta\left(1 + e^{x/\theta}\right)^2} = \frac{e^{x/\theta}}{\theta e^{2x/\theta}\left(1 + e^{-x/\theta}\right)^2} = \frac{e^{-x/\theta}}{\theta\left(1 + e^{-x/\theta}\right)^2} = f(x)$$

therefore f is an even function which is symmetric about the y axis.

$$\begin{aligned}
\frac{1}{k} &= \int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{\infty} \frac{e^{-x/\theta}}{\left(1 + e^{-x/\theta}\right)^2}dx \\
&= 2\int_{0}^{\infty} \frac{e^{-x/\theta}}{\left(1 + e^{-x/\theta}\right)^2}dx \qquad \text{by symmetry} \\
&= 2\theta\left[\lim_{b\to\infty} \frac{1}{1 + e^{-x/\theta}}|_0^b\right] = 2\theta\left[\lim_{b\to\infty} \frac{1}{1 + e^{-b/\theta}} - \frac{1}{2}\right] = 2\theta\left(\frac{1}{2}\right) = \theta
\end{aligned}$$

Therefore $k = 1/\theta$.

(2)

$$\begin{aligned}
F(x) &= \int_{-\infty}^{x} \frac{e^{-t/\theta}}{\left(1 + e^{-t/\theta}\right)^2}dt = \lim_{a\to-\infty} \frac{1}{1 + e^{-t/\theta}}|_a^x \\
&= \frac{1}{1 + e^{-x/\theta}} - \lim_{a\to-\infty} \frac{1}{1 + e^{-a/\theta}} \\
&= \frac{1}{1 + e^{-x/\theta}} \qquad \text{for } x \in \Re
\end{aligned}$$

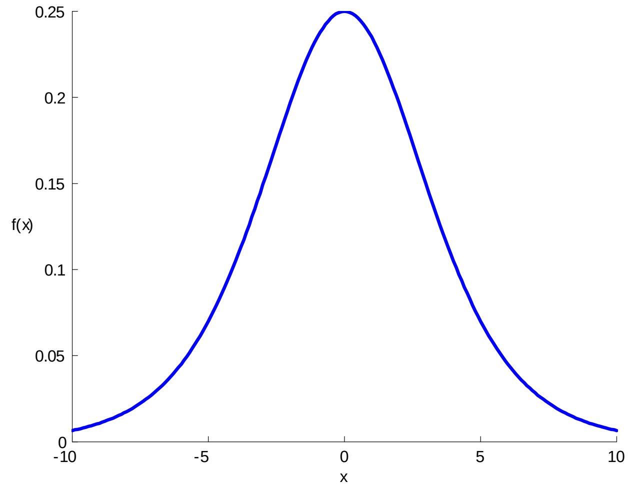


Figure 10.5: Graph of $f(x)$ for $\theta = 2$

(3) Since f is a symmetric function about $x = 0$ then if $E(|X|)$ exists then $E(X) = 0$. Now

$$E(|X|) = 2\int_0^\infty \frac{xe^{-x/\theta}}{\left(1+e^{-x/\theta}\right)^2}dx$$

Since

$$\int_0^\infty \frac{x}{\theta}e^{-x/\theta}dx = \theta\int_0^\infty ye^{-y}dy = \theta\Gamma(2)$$

converges and

$$\frac{x}{\theta}e^{-x/\theta} \geq \frac{xe^{-x/\theta}}{\left(1+e^{-x/\theta}\right)^2} \quad \text{for } x \geq 0$$

therefore by the Comparison Test for Integrals the improper integral

$$E(|X|) = 2\int_0^\infty \frac{xe^{-x/\theta}}{\left(1+e^{-x/\theta}\right)^2}dx$$

converges and thus $E(X) = 0$.

By symmetry

$$\begin{aligned} E\left(X^2\right) &= \int_{-\infty}^{\infty} \frac{x^2 e^{-x/\theta}}{\left(1+e^{-x/\theta}\right)^2} dx \\ &= 2\int_{-\infty}^{0} \frac{x^2 e^{-x/\theta}}{\left(1+e^{-x/\theta}\right)^2} dx \qquad \text{let } y = x/\theta \\ &= 2\theta^2 \int_{-\infty}^{0} \frac{y^2 e^{-y}}{(1+e^{-y})^2} dy \end{aligned}$$

Using integration by parts with

$$u = y^2, \quad du = 2y dy, \quad dv = \frac{e^{-y}}{(1+e^{-y})^2}, \quad v = \frac{1}{(1+e^{-y})}$$

we have

$$\begin{aligned} \int_{-\infty}^{0} \frac{y^2 e^{-y}}{(1+e^{-y})^2} dy &= \lim_{a\to-\infty} \frac{y^2}{(1+e^{-y})}\Big|_a^0 - \int_{-\infty}^{0} \frac{2y}{1+e^{-y}} dy \\ &= \lim_{a\to-\infty} \frac{-a^2}{(1+e^{-a})} - 2\int_{-\infty}^{0} \frac{y}{1+e^{-y}} dy \qquad \lim_{a\to-\infty} e^{-a} = \infty \\ &= \lim_{a\to-\infty} \frac{-2a}{-e^{-a}} - 2\int_{-\infty}^{0} \frac{y}{1+e^{-y}} dy \qquad \text{by L'Hospital's Rule} \\ &= \lim_{a\to-\infty} \frac{2}{e^{-a}} - 2\int_{-\infty}^{0} \frac{y}{1+e^{-y}} dy \qquad \text{by L'Hospital's Rule} \\ &= 0 - 2\int_{-\infty}^{0} \frac{y}{1+e^{-y}} dy \qquad \text{multiply by } \frac{e^y}{e^y} \\ &= -2\int_{-\infty}^{0} \frac{ye^y}{1+e^y} dy \end{aligned}$$

Let

$$u = e^y, \quad du = e^y dy, \quad \log u = y$$

to obtain

$$\int_{-\infty}^{0} \frac{ye^y}{1+e^y} dy = \int_{0}^{1} \frac{\log u}{1+u} du = -\frac{\pi^2}{12}$$

This definite integral can be found at *https://en.wikipedia.org/wiki/List_ of_ definite_ integrals*.

$$\int_0^1 \frac{\log u}{1+u} du = -\frac{\pi^2}{12}$$

Therefore

$$\begin{aligned} E\left(X^2\right) &= 2\theta^2 \left[-2\left(-\frac{\pi^2}{12}\right)\right] \\ &= \frac{\theta^2\pi^2}{3} \end{aligned}$$

and

$$\begin{aligned} Var\left(X\right) &= E\left(X^2\right) - \left[E\left(X\right)\right]^2 = \frac{\theta^2\pi^2}{3} - 0^2 \\ &= \frac{\theta^2\pi^2}{3} \end{aligned}$$

(4)

$$\begin{aligned} P(0.5 &< X \leq 2) = F(2) - F(0.5) \\ &= \frac{1}{1+e^{-2/2}} - \frac{1}{1+e^{-0.5/2}} \quad \text{using } \theta = 2 \\ &= \frac{1}{1+e^{-1}} - \frac{1}{1+e^{-0.25}} \\ &\approx 0.1689 \end{aligned}$$

$$\begin{aligned} P(X &> 0.5|X \leq 2) = \frac{P(X > 0.5, X \leq 2)}{P(X \leq 2)} \\ &= \frac{P(0.5 < X \leq 2)}{P(X \leq 2)} \\ &= \frac{F(2) - F(0.5)}{F(2)} \\ &= \left(\frac{1}{1+e^{-1}} - \frac{1}{1+e^{-0.25}}\right)\left(1+e^{-1}\right) \\ &\approx 0.2310 \end{aligned}$$

2.(*a*) Since

$$f(x;\theta) = (1-\theta)^2 \, x\theta^{x-1}$$

therefore

$$f_0(x) = f(x;\theta = 0) = 0$$

and

$$f_1(x) = f(x;\theta = 1) = 0$$

Since

$$f(x;\theta) \neq f_0(x-\theta) \quad \text{and} \quad f(x;\theta) \neq \frac{1}{\theta} f_1\left(\frac{x}{\theta}\right)$$

therefore θ is neither a location nor scale parameter.

2. (*b*) Since

$$f(x;\theta) = \frac{1}{\pi\theta\left[1 + (x/\theta)^2\right]} \quad \text{for } x \in \Re,\ \theta > 0$$

therefore

$$f_1(x) = f(x;\theta = 1) = \frac{1}{\pi\left(1 + x^2\right)} \quad \text{for } x \in \Re$$

Since

$$f(x;\theta) = \frac{1}{\theta} f_1\left(\frac{x}{\theta}\right) \quad \text{for all } x \in \Re,\ \theta > 0$$

therefore θ is a scale parameter for this distribution.

2. (*c*) Since

$$f(x;\theta) = \frac{1}{2} e^{-|x-\theta|} \quad \text{for } x \in \Re,\ \theta \in \Re$$

therefore

$$f_0(x) = f(x;\theta = 0) = \frac{1}{2} e^{-|x|} \quad \text{for } x \in \Re$$

Since

$$f(x;\theta) = f_0(x-\theta) \quad \text{for } x \in \Re,\ \theta \in \Re$$

therefore θ is a location parameter for this distribution.

2. (*e*) Since

$$f(x;\theta) = \frac{1}{2}\theta^3 x^2 e^{-\theta x} \quad \text{for } x \geq 0,\ \theta > 0$$

therefore

$$f_1(x) = f(x;\theta = 1) = \frac{1}{2} x^2 e^{-x} \quad \text{for } x \geq 0$$

Since

$$f(x;\theta) = \theta f_1(\theta x) \quad \text{for } x \geq 0,\ \theta > 0$$

therefore $1/\theta$ is a scale parameter for this distribution.

4. (a) Note that $f(x)$ can be written as

$$f(x) = \begin{cases} ke^{c^2/2}e^{c(x-\theta)} & x < \theta - c \\ ke^{-(x-\theta)^2/2} & \theta - c \leq x \leq \theta + c \\ ke^{c^2/2}e^{-c(x-\theta)} & x > \theta + c \end{cases}$$

Therefore

$$\begin{aligned}
\frac{1}{k} &= e^{c^2/2}\int_{-\infty}^{\theta-c} e^{c(x-\theta)}dx + \int_{\theta-c}^{\theta+c} e^{-(x-\theta)^2/2}dx + e^{c^2/2}\int_{\theta+c}^{\infty} e^{-c(x-\theta)}dx \\
&= 2e^{c^2/2}\int_{c}^{\infty} e^{-cu}du + \sqrt{2\pi}\int_{-c}^{c} \frac{1}{\sqrt{2\pi}}e^{-z^2/2}dz \\
&= 2e^{c^2/2}\left[\lim_{b\to\infty}\left(-\frac{1}{c}e^{-cu}|_c^b\right)\right] + \sqrt{2\pi}P(|Z| \leq c) \quad \text{where } Z \sim \mathrm{N}(0,1) \\
&= \frac{2}{c}e^{c^2/2}e^{-c^2} + \sqrt{2\pi}\left[2\Phi(c) - 1\right] \quad \text{where } \Phi \text{ is the } \mathrm{N}(0,1) \text{ c.d.f.} \\
&= \frac{2}{c}e^{-c^2/2} + \sqrt{2\pi}\left[2\Phi(c) - 1\right]
\end{aligned}$$

as required.

4. (b) If $x < \theta - c$ then

$$\begin{aligned}
F(x) &= ke^{c^2/2}\int_{-\infty}^{x} e^{c(u-\theta)}du, \qquad \text{let } y = u - \theta \\
&= ke^{c^2/2}\int_{-\infty}^{x-\theta} e^{cy}dy \\
&= ke^{c^2/2}\left[\lim_{a\to-\infty}\left(\frac{1}{c}e^{cy}|_a^{x-\theta}\right)\right] \\
&= \frac{k}{c}e^{c^2/2+c(x-\theta)}
\end{aligned}$$

and $F(\theta - c) = \frac{k}{c}e^{-c^2/2}$.

If $\theta - c \le x \le \theta + c$ then

$$\begin{aligned}
F(x) &= \frac{k}{c}e^{-c^2/2} + k\sqrt{2\pi}\int_{\theta-c}^{x}\frac{1}{\sqrt{2\pi}}e^{-(u-\theta)^2/2}du \quad \text{let } z = u-\theta \\
&= \frac{k}{c}e^{-c^2/2} + k\sqrt{2\pi}\int_{-c}^{x-\theta}\frac{1}{\sqrt{2\pi}}e^{-z^2/2}dz \\
&= \frac{k}{c}e^{-c^2/2} + k\sqrt{2\pi}\left[\Phi(x-\theta) - \Phi(-c)\right] \\
&= \frac{k}{c}e^{-c^2/2} + k\sqrt{2\pi}\left[\Phi(x-\theta) + \Phi(c) - 1\right].
\end{aligned}$$

If $x > \theta + c$ then

$$\begin{aligned}
F(x) &= 1 - ke^{c^2/2}\int_{x}^{\infty}e^{-c(u-\theta)}du \quad \text{let } y = u-\theta \\
&= 1 - ke^{c^2/2}\int_{x-\theta}^{\infty}e^{-cy}dy \\
&= 1 - ke^{c^2/2}\left[\lim_{b\to\infty}\left(-\frac{1}{c}e^{-cy}\Big|_{x-\theta}^{b}\right)\right] \\
&= 1 - \frac{k}{c}e^{c^2/2-c(x-\theta)}
\end{aligned}$$

Therefore

$$F(x) = \begin{cases} \frac{k}{c}e^{c^2/2+c(x-\theta)} & x < \theta - c \\ \frac{k}{c}e^{-c^2/2} + k\sqrt{2\pi}\left[\Phi(x-\theta) + \Phi(c) - 1\right] & \theta - c \le x \le \theta + c \\ 1 - \frac{k}{c}e^{c^2/2-c(x-\theta)} & x > \theta + c \end{cases}$$

Since θ is a location parameter (see part (c))

$$\begin{aligned}
E\left(X^k\right) &= \int_{-\infty}^{\infty}x^k f(x)\,dx = \int_{-\infty}^{\infty}x^k f_0(x-\theta)\,dx \quad \text{let } y = u-\theta \qquad (10.4) \\
&= \int_{-\infty}^{\infty}(y+\theta)^k f_0(y)\,dy
\end{aligned}$$

In particular

$$\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} (y+\theta) f_0(y)\, dy = \int_{-\infty}^{\infty} y f_0(y)\, dy + \theta \int_{-\infty}^{\infty} f_0(y)\, dy \\
&= \int_{-\infty}^{\infty} y f_0(y)\, dy + \theta(1) \quad \text{since } f_0(y) \text{ is a p.d.f.} \\
&= \int_{-\infty}^{\infty} y f_0(y)\, dy + \theta
\end{aligned}$$

Now

$$\begin{aligned}
\frac{1}{k}\int_{-\infty}^{\infty} y f_0(y)\, dy &= e^{c^2/2}\int_{-\infty}^{-c} ye^{cy}dy + \int_{-c}^{c} ye^{-y^2/2}dy + e^{c^2/2}\int_{c}^{\infty} ye^{-cy}dy \\
\text{let } y &= -u \text{ in the first integral} \\
&= -e^{c^2/2}\int_{c}^{\infty} ue^{-cu}du + \int_{-c}^{c} ye^{-y^2/2}dy + e^{c^2/2}\int_{c}^{\infty} ye^{-cy}dy
\end{aligned}$$

By integration by parts

$$\begin{aligned}
\int_{c}^{\infty} ye^{-cy}dy &= \lim_{b\to\infty}\left[-\frac{1}{c}ye^{-cy}|_c^b + \frac{1}{c}\int_{c}^{b} e^{-cy}dy\right] \\
&= \lim_{b\to\infty}\left[-\frac{1}{c}ye^{-cy}|_c^b - \frac{1}{c^2}e^{-cy}|_c^b\right] \\
&= \left(1+\frac{1}{c^2}\right)e^{-c^2} \qquad (10.5)
\end{aligned}$$

Also since $g(y) = ye^{-y^2/2}$ is a bounded odd function and $[-c, c]$ is a symmetric interval about 0

$$\int_{-c}^{+c} ye^{-y^2/2}dy = 0$$

Therefore

$$\frac{1}{k}\int_{-\infty}^{\infty} y f_0(y)\, dy = -\left(1+\frac{1}{c^2}\right)e^{-c^2} + 0 + \left(1+\frac{1}{c^2}\right)e^{-c^2} = 0$$

and

$$E(X) = \int_{-\infty}^{\infty} y f_0(y)\, dy + \theta = 0 + \theta = \theta$$

To determine $Var\,(X)$ we note that

$$\begin{aligned}
Var\,(X) &= E\left(X^2\right) - [E\,(X)]^2 \\
&= \int_{-\infty}^{\infty} (y+\theta)^2 f_0\,(y)\,dy - \theta^2 \qquad \text{using } (10.4) \\
&= \int_{-\infty}^{\infty} y^2 f_0\,(y)\,dy + 2\theta \int_{-\infty}^{\infty} y f_0\,(y)\,dy + \theta^2 \int_{-\infty}^{\infty} f_0\,(y)\,dy - \theta^2 \\
&= \int_{-\infty}^{\infty} y^2 f_0\,(y)\,dy + 2\theta\,(0) + \theta^2\,(1) - \theta^2 \\
&= \int_{-\infty}^{\infty} y^2 f_0\,(y)\,dy
\end{aligned}$$

Now

$$\begin{aligned}
\frac{1}{k}\int_{-\infty}^{\infty} y^2 f_0\,(y)\,dy &= e^{c^2/2}\int_{-\infty}^{-c} y^2 e^{cy}dy + \int_{-c}^{c} y^2 e^{-y^2/2}dy + e^{c^2/2}\int_{c}^{\infty} y^2 e^{-cy}dy \\
(\text{let } y &= -u \text{ in the first integral}) \\
&= e^{c^2/2}\int_{c}^{\infty} u^2 e^{-cu}du + \int_{-c}^{c} y^2 e^{-y^2/2}dy + e^{c^2/2}\int_{c}^{\infty} y^2 e^{-cy}dy \\
&= 2e^{c^2/2}\int_{c}^{\infty} y^2 e^{-cy}dy + \int_{-c}^{c} y^2 e^{-y^2/2}dy
\end{aligned}$$

By integration by parts and using (10.5) we have

$$\begin{aligned}
\int_{c}^{\infty} y^2 e^{-cy}dy &= \lim_{b\to\infty}\left[-\frac{1}{c}y^2 e^{-cy}|_c^b + \frac{2}{c}\int_{c}^{b} y e^{-cy}dy\right] \\
&= ce^{-c^2} + \frac{2}{c}\left[\left(1+\frac{1}{c^2}\right)e^{-c^2}\right] \\
&= \left(c + \frac{2}{c} + \frac{2}{c^3}\right)e^{-c^2}
\end{aligned}$$

Also

$$\begin{aligned}
\int_{-c}^{+c} y^2 e^{-y^2/2} dy &= 2\int_{0}^{c} y^2 e^{-y^2/2} dy \\
&= 2\left[-ye^{-y^2/2}|_0^c + \sqrt{2\pi}\int_0^c \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy \right] \quad \text{using integration by parts} \\
&= 2\left\{ -ce^{-c^2/2} + \sqrt{2\pi}\left[\Phi(c) - 0.5\right] \right\} \\
&= \sqrt{2\pi}\left[2\Phi(c) - 1\right] - 2ce^{-c^2/2}
\end{aligned}$$

Therefore

$$\begin{aligned}
Var(X) &= \int_{-\infty}^{\infty} y^2 f_0(y)\, dy \\
&= k\left\{ 2\left(c + \frac{2}{c} + \frac{2}{c^3}\right) e^{-c^2} + \sqrt{2\pi}\left[2\Phi(c) - 1\right] - 2ce^{-c^2/2} \right\} \\
&= \left\{ \frac{1}{\frac{2}{c}e^{-c^2/2} + \sqrt{2\pi}\left[2\Phi(c) - 1\right]} \right\} \times \\
&\quad \left\{ 2\left(c + \frac{2}{c} + \frac{2}{c^3}\right) e^{-c^2} + \sqrt{2\pi}\left[2\Phi(c) - 1\right] - 2ce^{-c^2/2} \right\}
\end{aligned}$$

4. (c) Let

$$\begin{aligned}
f_0(x) &= f(x; \theta = 0) \\
&= \begin{cases} ke^{-x^2/2} & \text{if } |x| \le c \\ ke^{-c|x| + c^2/2} & \text{if } |x| > c \end{cases}
\end{aligned}$$

Since

$$\begin{aligned}
f_0(x - \theta) &= \begin{cases} ke^{-(x-\theta)^2/2} & \text{if } |x - \theta| \le c \\ ke^{-c|x-\theta| + c^2/2} & \text{if } |x - \theta| > c \end{cases} \\
&= f(x)
\end{aligned}$$

therefore θ is a location parameter for this distribution.

4. (d) On the graph in Figure 10.6 we have graphed $f(x)$ for $c = 1$, $\theta = 0$ (red), $f(x)$ for $c = 2$, $\theta = 0$ (blue) and the N$(0, 1)$ probability density function (black).

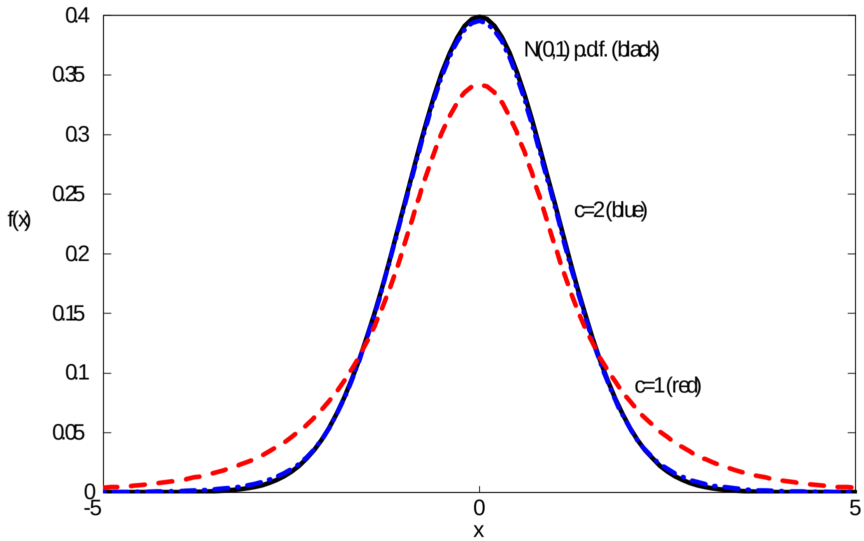


Figure 10.6: Graphs of $f(x)$ for $c = 1$, $\theta = 0$ (red), $c = 2$, $\theta = 0$ and the N$(0, 1)$ p.d.f. (black).

We note that there is very little difference between the graphs of the N$(0, 1)$ probability density function and the graph for $f(x)$ for $c = 2$, $\theta = 0$ however as c becomes smaller $(c = 1)$ the "tails" of the probability density function become much "fatter" relative to the N$(0, 1)$ probability density function.

5. (a) Since $X \sim$ Geometric(p)

$$\begin{aligned} P(X \geq k) &= \sum_{x=k}^{\infty} p(1-p)^x = \frac{p(1-p)^k}{1-(1-p)} \quad \text{by the Geometric Series} \\ &= (1-p)^k \quad \text{for } k = 0, 1, \ldots \end{aligned} \tag{10.6}$$

Therefore

$$\begin{aligned} P(X \geq k+j | X \geq k) &= \frac{P(X \geq k+j, X \geq k)}{P(X \geq k)} = \frac{P(X \geq k+j)}{P(X \geq k)} \\ &= \frac{(1-p)^{k+j}}{(1-p)^k} \quad \text{by (10.6)} \\ &= (1-p)^j \\ &= P(X \geq j) \quad \text{for } j = 0, 1, \ldots \end{aligned} \tag{10.7}$$

Suppose we have a large number of items which are to be tested to determine if they are defective or not. Suppose a proportion p of these items are defective. Items are tested one after another until the first defective item is found. If we let the random variable X be the number of good items found before observing the first defective item then $X \sim$ Geometric(p) and (10.6) holds. Now $P(X \geq j)$ is the probability we find at least j good items before observing the first defective item and $P(X \geq k+j|X \geq k)$ is the probability we find at least j more good items before observing the first defective item given that we have already observed at least k good items before observing the first defective item. Since these probabilities are the same for all nonnegative integers by (10.7), this implies that, no matter how many good items we have already observed before observing the first defective item, the probability of finding at least j more good items before observing the first defective item is the same as when we first began testing. It is like we have "forgotten" that we have already observed at least k good items before observing the first defective item. In other words, conditioning on the event that we have already observed at least k good items before observing the first defective item does not affect the probability of observing at least j more good items before observing the first defective item.

5. (b) If $Y \sim$ Exponential(θ) then

$$P(Y \geq a) = \int_a^{\infty} \frac{1}{\theta} e^{-y/\theta} dy = e^{-a/\theta} \quad \text{for } a > 0 \tag{10.8}$$

Therefore

$$\begin{aligned} P(Y \geq a+b | Y \geq a) &= \frac{P(Y \geq a+b, Y \geq a)}{P(Y \geq a)} = \frac{P(Y \geq a+b)}{P(Y \geq a)} \\ &= \frac{e^{-(a+b)/\theta}}{e^{-a/\theta}} \quad \text{by (10.8)} \\ &= e^{-b/\theta} = P(Y \geq b) \quad \text{for all } a, b > 0 \end{aligned}$$

as required.

6. Since $f_1(x)$, $f_2(x), \ldots, f_k(x)$ are probability density functions with support sets A_1, $A_2, \ldots, A_k$ then we know that $f_i(x) > 0$ for all $x \in A_i$, $i = 1, 2, \ldots, k$. Also since $0 < p_1, p_2, \ldots, p_k \leq 1$ with $\sum\limits_{i=1}^{k} p_i = 1$, we have that $g(x) = \sum\limits_{i=1}^{k} p_i f_i(x) > 0$ for all $x \in A = \bigcup\limits_{i=1}^{k} A_i$ and $A =$ support set of X. Also

$$\int\limits_{-\infty}^{\infty} g(x)dx = \sum_{i=1}^{k} p_i \int\limits_{-\infty}^{\infty} f_i(x)\,dx = \sum_{i=1}^{k} p_i(1) = \sum_{i=1}^{k} p_i = 1$$

Therefore $g(x)$ is a probability density function.

Now the mean of X is given by

$$\begin{aligned} E(X) &= \int\limits_{-\infty}^{\infty} xg(x)dx = \sum_{i=1}^{k} p_i \int\limits_{-\infty}^{\infty} xf_i(x)\,dx \\ &= \sum_{i=1}^{k} p_i \mu_i \end{aligned}$$

As well

$$\begin{aligned} E\left(X^2\right) &= \int\limits_{-\infty}^{\infty} x^2 g(x)dx = \sum_{i=1}^{k} p_i \int\limits_{-\infty}^{\infty} x^2 f_i(x)\,dx \\ &= \sum_{i=1}^{k} p_i\left(\sigma_i^2 + \mu_i^2\right) \end{aligned}$$

since

$$\begin{aligned} \int\limits_{-\infty}^{\infty} x^2 f_i(x)\,dx &= \int\limits_{-\infty}^{\infty} (x-\mu_i)^2 f_i(x)\,dx + 2\mu_i \int\limits_{-\infty}^{\infty} xf_i(x)\,dx - \mu_i^2 \int\limits_{-\infty}^{\infty} f_i(x)\,dx \\ &= \sigma_i^2 + 2\mu_i^2 - \mu_i^2 \\ &= \sigma_i^2 + \mu_i^2 \end{aligned}$$

Thus the variance of X is

$$Var(X) = \sum_{i=1}^{k} p_i\left(\sigma_i^2 + \mu_i^2\right) - \left(\sum_{i=1}^{k} p_i \mu_i\right)^2$$

7.(a) Since $X \sim \text{Gamma}(\alpha, \beta)$ the probability density function of X is

$$f(x) = \frac{x^{\alpha-1}e^{-x/\beta}}{\Gamma(\alpha)\beta^{\alpha}} \quad \text{for } x > 0$$

and 0 otherwise. Let $A = \{x : f(x) > 0\} = \{x : x > 0\}$. Now $y = e^x = h(x)$ is a one-to-one function on A and h maps the set A to the set $B = \{y : y > 1\}$. Also

$$x = h^{-1}(y) = \log y \quad \text{and} \quad \frac{d}{dy}h^{-1}(y) = \frac{1}{y}$$

The probability density function of Y is

$$\begin{aligned} g(y) &= f\left(h^{-1}(y)\right)\left|\frac{d}{dy}h^{-1}(y)\right| \\ &= \frac{(\log y)^{\alpha-1}e^{-\log y/\beta}}{\Gamma(\alpha)\beta^{\alpha}}\left(\frac{1}{y}\right) \\ &= \frac{(\log y)^{\alpha-1}y^{-1/\beta-1}}{\Gamma(\alpha)\beta^{\alpha}} \quad \text{for } y \in B \end{aligned}$$

and 0 otherwise.

7.(b) Since $X \sim \text{Gamma}(\alpha, \beta)$ the probability density function of X is

$$f(x) = \frac{x^{\alpha-1}e^{-x/\beta}}{\Gamma(\alpha)\beta^{\alpha}} \quad \text{for } x > 0$$

and 0 otherwise. Let $A = \{x : f(x) > 0\} = \{x : x > 0\}$. Now $y = 1/x = h(x)$ is a one-to-one function on A and h maps the set A to the set $B = \{y : y > 0\}$. Also

$$x = h^{-1}(y) = \frac{1}{y} \quad \text{and} \quad \frac{d}{dy}h^{-1}(y) = \frac{-1}{y^2}$$

The probability density function of Y is

$$\begin{aligned} g(y) &= f\left(h^{-1}(y)\right)\left|\frac{d}{dy}h^{-1}(y)\right| \\ &= \frac{y^{-\alpha-1}e^{-1/(\beta y)}}{\Gamma(\alpha)\beta^{\alpha}} \quad \text{for } y \in B \end{aligned}$$

and 0 otherwise. This is the probability density function of an Inverse Gamma(α, β) random variable. Therefore $Y = X^{-1} \sim$ Inverse Gamma(α, β).

7.(c) Since $X \sim \text{Gamma}(k, \beta)$ the probability density function of X is

$$f(x) = \frac{x^{k-1}e^{-x/\beta}}{\Gamma(k)\beta^{k}} \quad \text{for } x > 0$$

and 0 otherwise. Let $A = \{x : f(x) > 0\} = \{x : x > 0\}$. Now $y = 2x/\beta = h(x)$ is a one-to-one function on A and h maps the set A to the set $B = \{y : y > 0\}$. Also

$$x = h^{-1}(y) = \frac{\beta y}{2} \quad \text{and} \quad \frac{d}{dy}h^{-1}(y) = \frac{\beta}{2}$$

The probability density function of Y is

$$\begin{aligned} g(y) &= f\left(h^{-1}(y)\right)\left|\frac{d}{dy}h^{-1}(y)\right| \\ &= \frac{y^{k-1}e^{-y/2}}{\Gamma(k)\,2^k} \quad \text{for } y \in B \end{aligned}$$

and 0 otherwise for $k = 1, 2, \ldots$ which is the probability density function of a $\chi^2(2k)$ random variable. Therefore $Y = 2X/\beta \sim \chi^2(2k)$.

7.(*d*) Since $X \sim \mathrm{N}(\mu, \sigma^2)$ the probability density function of X is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{\frac{-1}{2\sigma^2}(x-\mu)^2} \quad \text{for } x \in \Re$$

Let $A = \{x : f(x) > 0\} = \Re$. Now $y = e^x = h(x)$ is a one-to-one function on A and h maps the set A to the set $B = \{y : y > 0\}$. Also

$$x = h^{-1}(y) = \log y \quad \text{and} \quad \frac{d}{dy}h^{-1}(y) = \frac{1}{y}$$

The probability density function of Y is

$$\begin{aligned} g(y) &= f\left(h^{-1}(y)\right)\left|\frac{d}{dy}h^{-1}(y)\right| \\ &= \frac{1}{y\sqrt{2\pi}\sigma}e^{\frac{-1}{2\sigma^2}(\log y-\mu)^2} \quad y \in B \end{aligned}$$

Note this distribution is called the Lognormal distribution.

7.(*e*) Since $X \sim \mathrm{N}(\mu, \sigma^2)$ the probability density function of X is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{\frac{-1}{2\sigma^2}(x-\mu)^2} \quad \text{for } x \in \Re$$

Let $A = \{x : f(x) > 0\} = \Re$. Now $y = x^{-1} = h(x)$ is a one-to-one function on A and h maps the set A to the set $B = \{y : y \neq 0,\ y \in \Re\}$. Also

$$x = h^{-1}(y) = \frac{1}{y} \quad \text{and} \quad \frac{d}{dy}h^{-1}(y) = \frac{-1}{y^2}$$

The probability density function of Y is

$$\begin{aligned} g(y) &= f\left(h^{-1}(y)\right)\left|\frac{d}{dy}h^{-1}(y)\right| \\ &= \frac{1}{\sqrt{2\pi}\sigma y^2}e^{\frac{-1}{2\sigma^2}\left[\left(\frac{1}{y}\right)-\mu\right]^2} \quad \text{for } y \in B \end{aligned}$$

7.(f) Since $X \sim \text{Uniform}\left(\frac{-\pi}{2}, \frac{\pi}{2}\right)$ the probability density function of X is

$$f(x) = \frac{1}{\pi} \quad \text{for } \frac{-\pi}{2} < x < \frac{\pi}{2}$$

and 0 otherwise. Let $A = \{x : f(x) > 0\} = \left\{x : \frac{-\pi}{2} < x < \frac{\pi}{2}\right\}$. Now $y = \tan(x) = h(x)$ is a one-to-one function on A and h maps A to the set $B = \{y : -\infty < y < \infty\}$. Also $x = h^{-1}(y) = \arctan(y)$ and $\frac{d}{dy}h^{-1}(y) = \frac{1}{1+y^2}$. The probability density function of Y is

$$\begin{aligned} g(y) &= f\left(h^{-1}(y)\right) \left|\frac{d}{dy}h^{-1}(y)\right| \\ &= \frac{1}{\pi}\frac{1}{1+y^2} \quad \text{for } y \in \Re \end{aligned}$$

and 0 otherwise. This is the probability density function of a Cauchy$(1, 0)$ random variable. Therefore $Y = \tan(X) \sim \text{Cauchy}(1, 0)$.

7.(g) Since $X \sim \text{Pareto}(\alpha, \beta)$ the probability density function of X is

$$f(x) = \frac{\beta\alpha^\beta}{x^{\beta+1}} \quad \text{for } x \geq \alpha,\ \alpha, \beta > 0$$

and 0 otherwise. Let $A = \{x : f(x) > 0\} = \{x : x \geq \alpha > 0\}$. Now $y = \beta \log(x/\alpha) = h(x)$ is a one-to-one function on A and h maps A to the set $B = \{y : 0 \leq y < \infty\}$. Also $x = h^{-1}(y) = \alpha e^{y/\beta}$ and $\frac{d}{dy}h^{-1}(y) = \frac{\alpha}{\beta}e^{y/\beta}$. The probability density function of Y is

$$\begin{aligned} g(y) &= f\left(h^{-1}(y)\right) \left|\frac{d}{dy}h^{-1}(y)\right| \\ &= \frac{\alpha}{\beta}e^{y/\beta}\frac{\beta\alpha^\beta}{\left(\alpha e^{y/\beta}\right)^{\beta+1}} = e^{-y} \quad \text{for } y \in B \end{aligned}$$

and 0 otherwise. This is the probability density function of a Exponential(1) random variable. Therefore $Y = \beta \log(X/\alpha) \sim \text{Exponential}(1)$.

7.(h) If $X \sim \text{Weibull}(2, \theta)$ the probability density function of X is

$$f(x) = \frac{2xe^{-(x/\theta)^2}}{\theta^2} \quad \text{for } x > 0,\ \theta > 0$$

and 0 otherwise. Let $A = \{x : f(x) > 0\} = \{x : x > 0\}$. Now $y = x^2 = h(x)$ is a one-to-one function on A and h maps the set A to the set $B = \{y : y > 0\}$. Also

$$x = h^{-1}(y) = y^{1/2} \text{ and } \frac{d}{dy}h^{-1}(y) = \frac{1}{2y^{1/2}}$$

The probability density function of Y is

$$\begin{aligned} g(y) &= f\left(h^{-1}(y)\right) \left|\frac{d}{dy}h^{-1}(y)\right| \\ &= \frac{e^{-y/\theta^2}}{\theta^2} \quad \text{for } y \in B \end{aligned}$$

and 0 otherwise for $k = 1, 2, \ldots$ which is the probability density function of a Exponential$\left(\theta^2\right)$ random variable. Therefore $Y = X^2 \sim$ Exponential(θ^2)

7.(i) Since $X \sim$ Double Exponential$(0,1)$ the probability density function of X is

$$f(x) = \frac{1}{2}e^{-|x|} \quad \text{for } x \in \Re$$

The cumulative distribution function of $Y = X^2$ is

$$\begin{aligned} G(y) &= P(Y \le y) = P\left(X^2 \le y\right) = P\left(-\sqrt{y} \le X \le \sqrt{y}\right) \\ &= \int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{2} e^{-|x|} dx \\ &= \int_0^{\sqrt{y}} e^{-x} dx \qquad \text{by symmetry for } y > 0 \end{aligned}$$

By the First Fundamental Theorem of Calculus and the chain rule the probability density function of Y is

$$\begin{aligned} g(y) &= \frac{d}{dy} G(y) = e^{-\sqrt{y}} \frac{d}{dy} \sqrt{y} \\ &= \frac{1}{2\sqrt{y}} e^{-\sqrt{y}} \quad \text{for } y > 0 \end{aligned}$$

and 0 otherwise.

7.(j) Since $X \sim$ t(k) the probability density function of X is

$$f(x) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)} \frac{1}{\sqrt{k\pi}} \left(1 + \frac{x^2}{k}\right)^{-\left(\frac{k+1}{2}\right)} \qquad \text{for } x \in \Re$$

which is an even function. The cumulative distribution function of $Y = X^2$ is

$$\begin{aligned} G(y) &= P(Y \le y) = P\left(X^2 \le y\right) = P\left(-\sqrt{y} \le X \le \sqrt{y}\right) \\ &= \int_{-\sqrt{y}}^{\sqrt{y}} \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)} \frac{1}{\sqrt{k\pi}} \left(1 + \frac{x^2}{k}\right)^{-\left(\frac{k+1}{2}\right)} dx \\ &= 2\int_0^{\sqrt{y}} \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)} \frac{1}{\sqrt{k\pi}} \left(1 + \frac{x^2}{k}\right)^{-\left(\frac{k+1}{2}\right)} dx \qquad \text{by symmetry for } y > 0 \end{aligned}$$

By the First Fundamental Theorem of Calculus and the chain rule the probability density function of Y is

$$
\begin{aligned}
g(y) &= \frac{d}{dy} G(y) \\
&= 2\frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)} \frac{1}{\sqrt{k\pi}} \left(1+\frac{y}{k}\right)^{-\left(\frac{k+1}{2}\right)} \frac{d}{dy}\sqrt{y} \quad \text{for } y>0 \\
&= 2\frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\sqrt{\pi}} \frac{1}{\sqrt{k}} \left(1+\frac{y}{k}\right)^{-\left(\frac{k+1}{2}\right)} \frac{1}{2\sqrt{y}} \quad \text{for } y>0 \\
&= \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{1}{2}\right)} \left(\frac{1}{k}\right)^{1/2} y^{\frac{1}{2}-1} \left(1+\frac{1}{k}y\right)^{-\left(\frac{k+1}{2}\right)} \quad \text{for } y>0 \quad \text{since } \Gamma\left(\frac{1}{2}\right)=\sqrt{\pi}
\end{aligned}
$$

and 0 otherwise. This is the probability density function of a $\mathrm{F}(1,k)$ random variable. Therefore $Y = X^2 \sim \mathrm{F}(1,k)$.

8.(a) Let

$$c_n = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \frac{1}{\sqrt{n\pi}}$$

If $T \sim t(n)$ then T has probability density function

$$f(t) = c_n \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2} \quad \text{for } t \in \Re, \quad n = 1, 2, \ldots$$

Since $f(-t) = f(t)$, f is an even function whose graph is symmetric about the y axis. Therefore if $E(|T|)$ exists then due to symmetry $E(T) = 0$. To determine when $E(|T|)$ exists, again due to symmetry, we only need to determine for what values of n the integral

$$\int_0^\infty t \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2} dt$$

converges.

There are two cases to consider: $n = 1$ and $n > 1$.

For $n = 1$ we have

$$\int_0^\infty t\left(1 + t^2\right)^{-1} dt = \lim_{b \to \infty} \frac{1}{2} \ln\left(1 + t^2\right)|_0^b = \lim_{b \to \infty} \frac{1}{2} \ln\left(1 + b^2\right) = \infty$$

and therefore $E(T)$ does not exist.

For $n > 1$

$$\begin{aligned}
\int_0^\infty t \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2} dt &= \lim_{b \to \infty} \frac{-n}{n-1} \left(1 + \frac{t^2}{n}\right)^{-(n-1)/2} |_0^b \\
&= \frac{n}{n-1} \left[1 - \lim_{b \to \infty} \left(1 + \frac{b^2}{n}\right)^{-(n-1)/2}\right] \\
&= \frac{n}{n-1}
\end{aligned}$$

and the integral converges. Therefore $E(T) = 0$ for $n > 1$.

8.(b) To determine whether $Var(T) = E\left(T^2\right)$ (since $E(T) = 0$) exists we need to determine for what values of n the integral

$$\int_0^\infty t^2 \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2} dt$$

converges.

Now

$$\int_0^\infty t^2\left(1+\frac{t^2}{n}\right)^{-(n+1)/2} dt$$
$$= \int_0^{\sqrt{n}} t^2\left(1+\frac{t^2}{n}\right)^{-(n+1)/2} dt + \int_{\sqrt{n}}^\infty t^2\left(1+\frac{t^2}{n}\right)^{-(n+1)/2} dt$$

The first integral is finite since it is the integral of a finite function over the finite interval $[0, \sqrt{n}]$. We will show that the second integral

$$\int_{\sqrt{n}}^\infty t^2\left(1+\frac{t^2}{n}\right)^{-(n+1)/2} dt$$

diverges for $n = 1, 2$.

Now

$$\int_{\sqrt{n}}^\infty t^2\left(1+\frac{t^2}{n}\right)^{-(n+1)/2} dt \quad \text{let } y = t/\sqrt{n}$$

$$= n^{3/2}\int_1^\infty y^2\left(1+y^2\right)^{-(n+1)/2} dy \tag{10.9}$$

For $n = 1$

$$\frac{y^2}{(1+y^2)} \geq \frac{y^2}{(y^2+y^2)} = \frac{1}{2} \quad \text{for } y \geq 1$$

and since

$$\int_1^\infty \frac{1}{2} dy$$

diverges, therefore by the Comparison Test for Improper Integrals, (10.9) diverges for $n = 1$. (Note: For $n = 1$ we could also argue that $Var(T)$ does not exist since $E(T)$ does not exist for $n = 1$.)

For $n = 2$,

$$\frac{y^2}{(1+y^2)^{3/2}} \geq \frac{y^2}{(y^2+y^2)^{3/2}} = \frac{1}{2^{3/2}y} \quad \text{for } y \geq 1$$

and since

$$\frac{1}{2^{3/2}}\int_1^\infty \frac{1}{y} dy$$

diverges, therefore by the Comparison Test for Improper Integrals, (10.9) diverges for $n = 2$.

Now for $n > 2$,

$$\begin{aligned} E\left(T^2\right) &= \int_{-\infty}^{\infty} c_n t^2 \left(1+\frac{t^2}{n}\right)^{-(n+1)/2} dt \\ &= 2c_n \int_0^{\infty} t^2 \left(1+\frac{t^2}{n}\right)^{-(n+1)/2} dt \end{aligned}$$

since the integrand is an even function. Integrate by parts using

$$\begin{aligned} u &= t, \quad dv = t\left(1+\frac{t^2}{n}\right)^{-(n+1)/2} dt \\ du &= dt, \quad v = \frac{-n}{n-1}\left(1+\frac{t^2}{n}\right)^{-(n-1)/2} \end{aligned}$$

Then

$$\begin{aligned} E\left(T^2\right) &= 2c_n \left[\lim_{b\to\infty} t\left(\frac{-n}{n-1}\right)\left(1+\frac{t^2}{n}\right)^{-(n+1)/2} |_0^b\right] \\ &\quad +2c_n\left(\frac{n}{n-1}\right)\int_0^{\infty}\left(1+\frac{t^2}{n}\right)^{-(n-1)/2} dt \\ &= -2c_n\left(\frac{n}{n-1}\right)\lim_{b\to\infty}\frac{b}{\left(1+\frac{b^2}{n}\right)^{(n+1)/2}} + c_n\left(\frac{n}{n-1}\right)\int_{-\infty}^{\infty}\left(1+\frac{t^2}{n}\right)^{-(n-1)/2} dt \end{aligned}$$

where we use symmetry on the second integral.

Now

$$\lim_{b\to\infty}\frac{b}{\left(1+\frac{b^2}{n}\right)^{(n+1)/2}} = \lim_{b\to\infty}\frac{1}{\left(\frac{n+1}{n}\right)b\left(1+\frac{b^2}{n}\right)^{(n-1)/2}} = 0$$

by L'Hopital's Rule. Also

$$c_n\left(\frac{n}{n-1}\right)\int_{-\infty}^{\infty}\left(1+\frac{t^2}{n}\right)^{-(n-1)/2} dt \qquad \text{let } \frac{y}{\sqrt{n-2}} = \frac{t}{\sqrt{n}}$$

$$\begin{aligned} &= \frac{c_n}{c_{n-2}}\left(\frac{n}{n-1}\right)\left(\frac{n}{n-2}\right)^{1/2}\int_{-\infty}^{\infty} c_{n-2}\left(1+\frac{y^2}{n-2}\right)^{-(n-2+1)/2} dy \\ &= \frac{c_n}{c_{n-2}}\left(\frac{n}{n-1}\right)\left(\frac{n}{n-2}\right)^{1/2} \end{aligned}$$

where the integral equals one since the integrand is the p.d.f. of a $t\,(n-2)$ random variable.

Finally

$$
\begin{aligned}
& \frac{c_n}{c_{n-2}}\left(\frac{n}{n-1}\right)\left(\frac{n}{n-2}\right)^{1/2} \\
= & \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{n\pi}} \frac{\Gamma\left(\frac{n-2}{2}\right)\sqrt{(n-2)\,\pi}}{\Gamma\left(\frac{n-2+1}{2}\right)}\left(\frac{n}{n-1}\right)\left(\frac{n}{n-2}\right)^{1/2} \\
= & \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n+1}{2}-1\right)} \frac{\Gamma\left(\frac{n}{2}-1\right)}{\Gamma\left(\frac{n}{2}\right)}\left(\frac{n-2}{n}\right)^{1/2}\left(\frac{n}{n-1}\right)\left(\frac{n}{n-2}\right)^{1/2} \\
= & \frac{\left(\frac{n+1}{2}-1\right)\Gamma\left(\frac{n+1}{2}-1\right)}{\Gamma\left(\frac{n+1}{2}-1\right)} \frac{\Gamma\left(\frac{n}{2}-1\right)}{\left(\frac{n}{2}-1\right)\Gamma\left(\frac{n}{2}-1\right)}\left(\frac{n}{n-1}\right) \\
= & \frac{(n-1)}{2} \frac{1}{(n-2)\,/2}\left(\frac{n}{n-1}\right) \\
= & \frac{n}{n-2}
\end{aligned}
$$

Therefore for $n > 2$

$$
Var(T) = E\left(T^2\right) = \frac{n}{n-2}
$$

9. (a) To find $E\left(X^k\right)$ we first note that since

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1} \text{ for } 0 < x < 1$$

and 0 otherwise then

$$\int_0^1 \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}\,dx = 1$$

or

$$\int_0^1 x^{a-1}(1-x)^{b-1}\,dx = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \tag{10.10}$$

for $a > 0$, $b > 0$. Therefore

$$\begin{aligned}
E\left(X^k\right) &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 x^k x^{a-1}(1-x)^{b-1}\,dx \\
&= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 x^{a+k-1}(1-x)^{b-1}\,dx \\
&= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a+k)\Gamma(b)}{\Gamma(a+b+k)} \quad \text{by (10.10)} \\
&= \frac{\Gamma(a+k)}{\Gamma(a)} \frac{\Gamma(a+b)}{\Gamma(a+b+k)} \quad \text{for } k = 1, 2, \ldots
\end{aligned}$$

For $k = 1$ we have

$$\begin{aligned}
E\left(X^k\right) &= E(X) = \frac{\Gamma(a+1)}{\Gamma(a)} \frac{\Gamma(a+b)}{\Gamma(a+b+1)} \\
&= \frac{a\Gamma(a)}{\Gamma(a)} \frac{\Gamma(a+b)}{(a+b)\Gamma(a+b)} \\
&= \frac{a}{a+b}
\end{aligned}$$

For $k = 2$ we have

$$\begin{aligned}
E\left(X^2\right) &= \frac{\Gamma(a+2)}{\Gamma(a)} \frac{\Gamma(a+b)}{\Gamma(a+b+2)} \\
&= \frac{(a+1)(a)\Gamma(a)}{\Gamma(a)} \frac{\Gamma(a+b)}{(a+b+1)(a+b)\Gamma(a+b)} \\
&= \frac{a(a+1)}{(a+b)(a+b+1)}
\end{aligned}$$

Therefore

$$\begin{aligned}
Var(X) &= E\left(X^2\right) - [E(X)]^2 \\
&= \frac{a(a+1)}{(a+b)(a+b+1)} - \left(\frac{a}{a+b}\right)^2 \\
&= \frac{a(a+1)(a+b) - a^2(a+b+1)}{(a+b)^2(a+b+1)} \\
&= \frac{a\left[a^2+ab+a+b-\left(a^2+ab+a\right)\right]}{(a+b)^2(a+b+1)} \\
&= \frac{ab}{(a+b)^2(a+b+1)}
\end{aligned}$$

9. (b)

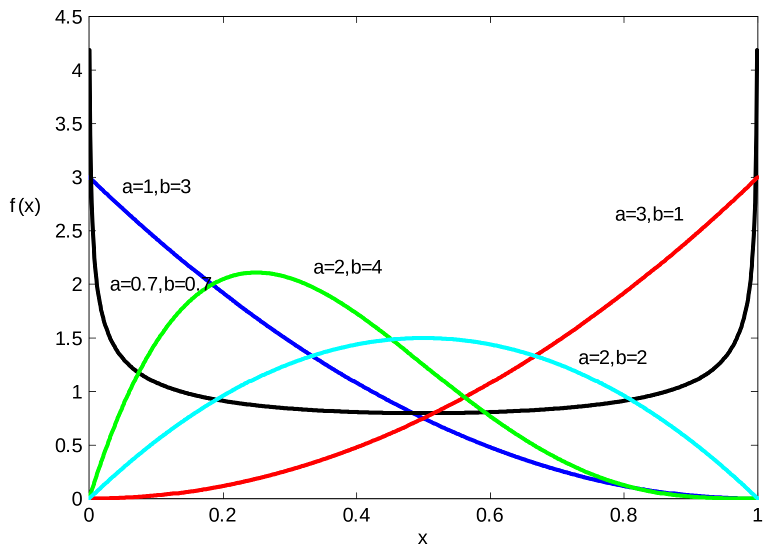


Figure 10.7: Graphs of Beta probability density functions

9. (c) If $a = b = 1$ then

$$f(x) = 1 \quad \text{for } 0 < x < 1$$

and 0 otherwise. This is the Uniform$(0, 1)$ probability density function.

10. We will prove this result assuming X is a continuous random variable. The proof for X a discrete random variable follows in a similar manner with integrals replaced by sums.

Suppose X has probability density function $f(x)$ and $E\left(|X|^k\right)$ exists for some integer $k > 1$. Then the improper integral

$$\int_{-\infty}^{\infty} |x|^k f(x)\,dx$$

converges. Let $A = \{x : |x| \geq 1\}$. Then

$$\int_{-\infty}^{\infty} |x|^k f(x)\,dx = \int_{A} |x|^k f(x)\,dx + \int_{\bar{A}} |x|^k f(x)\,dx$$

Since

$$0 \leq |x|^k f(x) \leq f(x) \quad \text{for } x \in \bar{A}$$

we have

$$0 \leq \int_{\bar{A}} |x|^k f(x)\,dx \leq \int_{\bar{A}} f(x)\,dx = P\left(X \in \bar{A}\right) \leq 1 \tag{10.11}$$

Convergence of $\int_{-\infty}^{\infty} |x|^k f(x)\,dx$ and (10.11) imply the convergence of $\int_{A} |x|^k f(x)\,dx$.

Now

$$\int_{-\infty}^{\infty} |x|^j f(x)\,dx = \int_{A} |x|^j f(x)\,dx + \int_{\bar{A}} |x|^j f(x)\,dx \quad \text{for } j = 1, 2, ..., k-1 \tag{10.12}$$

and

$$0 \leq \int_{\bar{A}} |x|^j f(x)\,dx \leq 1$$

by the same argument as in (10.11). Since $\int_{A} |x|^k f(x)\,dx$ converges and

$$|x|^k f(x) \geq |x|^j f(x) \quad \text{for } x \in A,\ j = 1, 2, ..., k-1$$

then by the Comparison Theorem for Improper Integrals $\int_{A} |x|^j f(x)\,dx$ converges. Since both integrals on the right side of (10.12) exist, therefore

$$E\left(|X|^j\right) = \int_{-\infty}^{\infty} |x|^j f(x)\,dx \ \text{ exists for } j = 1, 2, ..., k-1$$

11. If $X \sim \text{Binomial}(n, \theta)$ then

$$E(X) = n\theta$$

and

$$\sqrt{Var(X)} = \sqrt{n\theta(1-\theta)}$$

Let $W = X/n$. Then

$$E(W) = \theta$$

and

$$Var(W) = \frac{\theta(1-\theta)}{n}$$

From the result in Section 2.9 we wish to find a transformation $Y = g(W) = g(X/n)$ such that $Var(Y) \approx$ constant then we need g such that

$$\frac{dg}{d\theta} = \frac{k}{\sqrt{\theta(1-\theta)}}$$

where k is chosen for convenience. We need to solve the separable differential equation

$$\int dg = k \int \frac{1}{\sqrt{\theta(1-\theta)}} d\theta \tag{10.13}$$

Since

$$\begin{aligned}
\frac{d}{dx} \arcsin\left(\sqrt{x}\right) &= \frac{1}{\sqrt{1-\left(\sqrt{x}\right)^2}} \frac{d}{dx}\left(\sqrt{x}\right) \\
&= \frac{1}{\sqrt{1-x}}\left(\frac{1}{2\sqrt{x}}\right) \\
&= \frac{1}{2\sqrt{x(1-x)}}
\end{aligned}$$

then the solution to (10.13) is

$$g(\theta) = k \cdot 2 \arcsin(\sqrt{\theta}) + C$$

Letting $k = 1/2$ and $C = 0$ we have $g(\theta) = \arcsin(\sqrt{\theta})$.

Therefore if $X \sim \text{Binomial}(n, \theta)$ and $Y = \arcsin(\sqrt{X/n})$ then $Var(Y) = Var\left[\arcsin(\sqrt{X/n})\right] \approx$ constant.

13.(b)

$$\begin{aligned}
M(t) &= E(e^{tx}) = \sum_{x=0}^{\infty} e^{tx}\frac{\mu^x e^{-\mu}}{x!} \\
&= e^{-\mu}\sum_{x=0}^{\infty}\frac{\left(\mu e^t\right)^x}{x!} \\
&= e^{-\mu}e^{\mu e^t} \quad \text{by the Exponential Series} \\
&= e^{\mu\left(e^t-1\right)} \quad \text{for } t \in \Re
\end{aligned}$$

$$\begin{aligned}
M'(t) &= e^{\mu\left(e^t-1\right)}\mu e^t \\
E(X) &= M'(0) = \mu
\end{aligned}$$

$$\begin{aligned}
M''(t) &= e^{\mu\left(e^t-1\right)}(\mu e^t)^2 + e^{\mu\left(e^t-1\right)}\mu e^t \\
E(X^2) &= M''(0) = \mu^2 + \mu \\
Var(X) &= E(X^2) - [E(X)]^2 = \mu^2 + \mu - \mu^2 = \mu
\end{aligned}$$

13.(c)

$$\begin{aligned}
M(t) &= E(e^{tx}) = \int_{\theta}^{\infty} e^{tx}\frac{1}{\beta}e^{-(x-\theta)/\beta}dx \\
&= \frac{e^{\theta/\beta}}{\beta}\int_{\theta}^{\infty} e^{-x\left(\frac{1}{\beta}-t\right)}dx \quad \text{which converges for } \left(\frac{1}{\beta}-t\right) > 0 \ \text{ or } \ t < \frac{1}{\beta}
\end{aligned}$$

Let

$$y = \left(\frac{1}{\beta}-t\right)x, \quad dy = \left(\frac{1}{\beta}-t\right)dx$$

to obtain

$$\begin{aligned}
M(t) &= \frac{e^{\theta/\beta}}{\beta}\int_{\theta}^{\infty} e^{-x\left(\frac{1}{\beta}-t\right)}dx = \frac{e^{\theta/\beta}}{\beta}\int_{\theta}^{\infty} e^{-x\left(\frac{1}{\beta}-t\right)}dx \\
&= \frac{e^{\theta/\beta}}{\beta\left(\frac{1}{\beta}-t\right)}\int_{\theta\left(\frac{1}{\beta}-t\right)}^{\infty} e^{-y}dy = \frac{e^{\theta/\beta}}{(1-\beta t)}\lim_{b\to\infty}\left(-e^{-y}\Big|_{\theta\left(\frac{1}{\beta}-t\right)}^{b}\right) \\
&= \frac{e^{\theta/\beta}}{(1-\beta t)}\lim_{b\to\infty}\left[e^{-\theta\left(\frac{1}{\beta}-t\right)} - e^{-b}\right] = \frac{e^{\theta/\beta}}{(1-\beta t)}\left[e^{-\theta\left(\frac{1}{\beta}-t\right)}\right] \\
&= \frac{e^{\theta t}}{(1-\beta t)} \quad \text{for } \ t < \frac{1}{\beta}
\end{aligned}$$

$$\begin{aligned}M'(t) &= \frac{\theta e^{\theta t}}{(1-\beta t)} + \frac{\beta e^{\theta t}}{(1-\beta t)^2} \\ &= \frac{e^{\theta t}}{(1-\beta t)^2}\left[\theta(1-\beta t)+\beta\right] \\ E(X) &= M'(0) = \theta + \beta\end{aligned}$$

$$\begin{aligned}M''(t) &= \frac{e^{\theta t}}{(1-\beta t)^2}(-\theta\beta) + \left[\frac{\theta e^{\theta t}}{(1-\beta t)^2} + \frac{2\beta e^{\theta t}}{(1-\beta t)^3}\right]\left[\theta(1-\beta t)+\beta\right] \\ E(X^2) &= M''(0) = -\theta\beta + (\theta+2\beta)(\theta+\beta) \\ &= -\theta\beta + \theta^2 + 3\beta\theta + 2\beta^2 = \theta^2 + 2\beta\theta + 2\beta^2 \\ &= (\theta+\beta)^2 + \beta^2 \\ Var(X) &= E(X^2) - [E(X)]^2 \\ &= (\theta+\beta)^2 + \beta^2 - (\theta+\beta)^2 \\ &= \beta^2\end{aligned}$$

13.(d)

$$\begin{aligned}M(t) &= E(e^{tx}) = \frac{1}{2}\int_{-\infty}^{\infty} e^{tx}e^{-|x-\theta|}dx \\ &= \frac{1}{2}\left[\int_{-\infty}^{\theta} e^{tx}e^{x-\theta}dx + \int_{\theta}^{\infty} e^{tx}e^{-(x-\theta)}dx\right] \\ &= \frac{1}{2}\left[e^{-\theta}\int_{-\infty}^{\theta} e^{x(t+1)}dx + e^{\theta}\int_{\theta}^{\infty} e^{-x(1-t)}dx\right] \\ &= \frac{1}{2}\left[e^{-\theta}\left(\frac{e^{\theta(t+1)}}{t+1}\right) + e^{\theta}\left(\frac{e^{-\theta(1-t)}}{1-t}\right)\right] \quad \text{for } t+1>0 \text{ and } 1-t>0 \\ &= \frac{1}{2}\left[\frac{e^{\theta t}}{t+1} + \frac{e^{\theta t}}{1-t}\right] \quad \text{for } t \in (-1,1) \\ &= \frac{e^{\theta t}}{1-t^2} \quad \text{for } t \in (-1,1)\end{aligned}$$

$$\begin{aligned}M'(t) &= \frac{\theta e^{\theta t}}{1-t^2} + \frac{e^{\theta t}(2t)}{(1-t^2)^2} \\ &= \frac{e^{\theta t}}{(1-t^2)^2}\left[\left(1-t^2\right)\theta + 2t\right] \\ E(X) &= M'(0) = \theta\end{aligned}$$

$$\begin{aligned}
M''(t) &= \frac{e^{\theta t}}{(1-t^2)^2}\left[-2t\theta+2\right]+\left[\frac{\theta e^{\theta t}}{(1-t^2)^2}+\frac{e^{\theta t}(4t)}{(1-t^2)^4}\right]\left[(1-t^2)\theta+2t\right] \\
E(X^2) &= M''(0)=2+\theta^2 \\
Var(X) &= E(X^2)-[E(X)]^2 \\
&= 2+\theta^2-\theta^2 \\
&= 2
\end{aligned}$$

13.(e)

$$M\left(t\right)=E\left(e^{tX}\right)=\int\limits_0^1 2xe^{tx}xdx$$

Since

$$\int xe^{tx}dx=\frac{1}{t}\left(x-\frac{1}{t}\right)e^{tx}+C$$

$$\begin{aligned}
M\left(t\right) &= \frac{2}{t}\left(x-\frac{1}{t}\right)e^{tx}|_0^1 \\
&= \frac{2}{t}\left(1-\frac{1}{t}\right)e^{t}-\frac{2}{t}\left(-\frac{1}{t}\right) \\
&= \frac{2\left[\left(t-1\right)e^t+1\right]}{t^2}, \qquad \text{if } t\neq 0
\end{aligned}$$

For $t=0$, $M\left(0\right)=E\left(1\right)=1$. Therefore

$$M\left(t\right)=\begin{cases}1 & \text{if } t=0\\ \frac{2\left[(t-1)e^t+1\right]}{t^2} & \text{if } t\neq 0\end{cases}$$

Note that

$$\begin{aligned}
\lim_{t\to 0}M\left(t\right) &= \lim_{t\to 0}\frac{2\left[\left(t-1\right)e^t+1\right]}{t^2} \quad \text{indeterminate of the form } \left[\frac{0}{0}\right] \\
&= \lim_{t\to 0}\frac{2\left[e^t+\left(t-1\right)e^t\right]}{2t} \quad \text{by Hospital's Rule} \\
&= \lim_{t\to 0}\frac{e^t+\left(t-1\right)e^t}{t} \quad \text{indeterminate of the form } \left[\frac{0}{0}\right] \\
&= \lim_{t\to 0}\left[e^t+e^t+\left(t-1\right)e^t\right] \quad \text{by Hospital's Rule} \\
&= 1+1-1=1 \\
&= M\left(0\right)
\end{aligned}$$

Therefore $M\left(t\right)$ exists and is continuous for all $t\in\Re$.

Using the Exponential series we have for $t \neq 0$

$$\begin{aligned}
\frac{2\left[(t-1)e^t+1\right]}{t^2} &= \frac{2}{t^2}\left[(t-1)\sum_{i=0}^{\infty}\frac{t^i}{i!}+1\right] \\
&= \frac{2}{t^2}\left[\sum_{i=0}^{\infty}\frac{t^{i+1}}{i!}-\sum_{i=0}^{\infty}\frac{t^i}{i!}+1\right] \\
&= \frac{2}{t^2}\left[t+\sum_{i=1}^{\infty}\frac{t^{i+1}}{i!}-\left(1+t+\sum_{i=2}^{\infty}\frac{t^i}{i!}\right)+1\right] \\
&= \frac{2}{t^2}\left[\sum_{i=1}^{\infty}\frac{t^{i+1}}{i!}-\sum_{i=2}^{\infty}\frac{t^i}{i!}\right] \\
&= \frac{2}{t^2}\left[\sum_{i=1}^{\infty}\frac{t^{i+1}}{i!}-\sum_{i=1}^{\infty}\frac{t^{i+1}}{(i+1)!}\right] \\
&= \frac{2}{t^2}\sum_{i=1}^{\infty}\left[\frac{1}{i!}-\frac{1}{(i+1)!}\right]t^{i+1}
\end{aligned}$$

and since

$$\frac{2}{t^2}\sum_{i=1}^{\infty}\left[\frac{1}{i!}-\frac{1}{(i+1)!}\right]t^{i+1}|_{t=0}=1$$

therefore $M(t)$ has a Maclaurin series representation for all $t \in \Re$ given by

$$\begin{aligned}
&\frac{2}{t^2}\sum_{i=1}^{\infty}\left[\frac{1}{i!}-\frac{1}{(i+1)!}\right]t^{i+1} \\
&= 2\sum_{i=1}^{\infty}\left[\frac{1}{i!}-\frac{1}{(i+1)!}\right]t^{i-1} \\
&= \sum_{i=0}^{\infty}2\left[\frac{1}{(i+1)!}-\frac{1}{(i+2)!}\right]t^{i}
\end{aligned}$$

Since $E\left(X^k\right)=k!\times$ the coefficient of t^k in the Maclaurin series for $M(t)$ we have

$$E(X)=(1!)(2)\left[\frac{1}{(1+1)!}-\frac{1}{(1+2)!}\right]=2\left(\frac{1}{2}-\frac{1}{6}\right)=\frac{2}{3}$$

and

$$E\left(X^2\right)=(2!)(2)\left[\frac{1}{(2+1)!}-\frac{1}{(2+2)!}\right]=4\left(\frac{1}{6}-\frac{1}{24}\right)=\frac{1}{2}$$

Therefore

$$Var(X)=E\left(X^2\right)-\left[E(X)\right]^2=\frac{1}{2}-\left(\frac{2}{3}\right)^2=\frac{1}{18}$$

Alternatively we could find $E(X) = M'(0)$ using the limit definition of the derivative

$$\begin{aligned} M'(0) &= \lim_{t \to 0} \frac{M(t) - M(0)}{t} \\ &= \lim_{t \to 0} \frac{\frac{2\left[(t-1)e^t + 1\right]}{t^2} - 1}{t} \\ &= \lim_{t \to 0} \frac{2\left[(t-1)e^t + 1\right] - t^2}{t^3} \\ &= \frac{2}{3} \quad \text{using L'Hospital's Rule} \end{aligned}$$

Similarly $E\left(X^2\right) = M''(0)$ could be found using

$$M''(0) = \lim_{t \to 0} \frac{M'(t) - M'(0)}{t}$$

where

$$M'(t) = \frac{d}{dt}\left(\frac{2\left[(t-1)e^t + 1\right]}{t^2}\right)$$

for $t \neq 0$.

13.(f)

$$\begin{aligned} M(t) &= E\left(e^{tX}\right) = \int_0^1 e^{tx} x dx + \int_1^2 e^{tx}(2-x) dx \\ &= \int_0^1 x e^{tx} dx + 2\int_1^2 e^{tx} dx - \int_1^2 x e^{tx} dx. \end{aligned}$$

Since

$$\int x e^{tx} dx = \frac{1}{t}\left(x - \frac{1}{t}\right) e^{tx} + C$$

$$\begin{aligned} M(t) &= \frac{1}{t}\left(x - \frac{1}{t}\right) e^{tx}\Big|_0^1 + \frac{2}{t}\left(e^{tx}\Big|_1^2\right) - \frac{1}{t}\left(x - \frac{1}{t}\right) e^{tx}\Big|_1^2 \\ &= \frac{1}{t}\left(1 - \frac{1}{t}\right) e^t - \frac{1}{t}\left(-\frac{1}{t}\right) + \frac{2}{t}\left(e^{2t} - e^t\right) - \frac{1}{t}\left[\left(2 - \frac{1}{t}\right) e^{2t} - \left(1 - \frac{1}{t}\right) e^t\right] \\ &= e^{2t}\left[\frac{2}{t} + \frac{1}{t}\left(\frac{1}{t} - 2\right)\right] + e^t\left[\frac{1}{t}\left(1 - \frac{1}{t}\right) - \frac{2}{t} + \frac{1}{t}\left(1 - \frac{1}{t}\right)\right] + \frac{1}{t^2} \\ &= \frac{e^{2t} - 2e^t + 1}{t^2} \quad \text{for } t \neq 0 \end{aligned}$$

For $t = 0$, $M(0) = E(1) = 1$. Therefore

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

Note that

$$\begin{aligned}
\lim_{t\to 0} M(t) &= \lim_{t\to 0} \frac{e^{2t}-2e^t+1}{t^2}, \text{ indeterminate of the form } \left[\frac{0}{0}\right], \text{ use Hospital's Rule} \\
&= \lim_{t\to 0} \frac{2e^{2t}-2e^t}{2t}, \text{ indeterminate of the form } \left[\frac{0}{0}\right], \text{ use Hospital's Rule} \\
&= \lim_{t\to 0} \frac{2e^{2t}-e^t}{1} = 2-1 = 1
\end{aligned}$$

and therefore $M(t)$ exists and is continuous for $t \in \Re$.

Using the Exponential series we have for $t \neq 0$

$$\begin{aligned}
\frac{e^{2t}-2e^t+1}{t^2} &= \frac{1}{t^2}\{1+2t+\frac{(2t)^2}{2!}+\frac{(2t)^3}{3!}+\frac{(2t)^4}{4!}+\cdots \\
&\quad -2\left[1+t+\frac{t^2}{2!}+\frac{t^3}{3!}+\frac{t^4}{4!}+\cdots\right]+1\} \\
&= 1+t+\frac{7}{12}t^2+\cdots \qquad (10.14)
\end{aligned}$$

and since

$$\left(1+t+\frac{7}{12}t^2+\cdots\right)|_{t=0} = 1$$

(10.14) is the Maclaurin series representation for $M(t)$ for $t \in \Re$.

Since $E\left(X^k\right) = k!\times$ the coefficient of t^k in the Maclaurin series for $M(t)$ we have

$$E(X) = 1! \times 1 = 1 \quad \text{and} \quad E\left(X^2\right) = 2! \times \frac{7}{12} = \frac{7}{6}$$

Therefore

$$Var(X) = E\left(X^2\right) - [E(X)]^2 = \frac{7}{6} - 1 = \frac{1}{6}$$

Alternatively we could find $E(X) = M'(0)$ using the limit definition of the derivative

$$\begin{aligned}
M'(0) &= \lim_{t\to 0} \frac{M(t)-M(0)}{t} = \lim_{t\to 0} \frac{\frac{e^{2t}-2e^t+1}{t^2}-1}{t} \\
&= \lim_{t\to 0} \frac{e^{2t}-2e^t+1-t^2}{t^3} = \lim_{t\to 0} \frac{\left(t^2+t^3+\frac{7}{12}t^4+\cdots\right)-t^2}{t^3} \\
&= \lim_{t\to 0} \left(1+\frac{7}{12}t+\cdots\right) = 1
\end{aligned}$$

Similarly $E\left(X^2\right) = M''(0)$ could be found using

$$M''(0) = \lim_{t\to 0} \frac{M'(t)-M'(0)}{t}$$

where

$$M''(t) = \frac{d}{dt}\left(\frac{e^{2t}-2e^t+1}{t^2}\right) = \frac{2\left[t\left(e^{2t}-e^t\right)-e^{2t}+2e^t-1\right]}{t^3} \quad \text{for } t \neq 0$$

14.(a)

$$
\begin{aligned}
K(t) &= \log M(t) \\
K'(t) &= \frac{M'(t)}{M(t)} \\
K'(0) &= \frac{M'(0)}{M(0)} \\
&= \frac{E(X)}{1} \\
&= E(X) \quad \text{since } M(0)=1
\end{aligned}
$$

$$
\begin{aligned}
K''(t) &= \frac{M(t)M''(t) - [M'(t)]^2}{[M(t)]^2} \\
K''(0) &= \frac{M(0)M''(0) - [M'(0)]^2}{[M(0)]^2} \\
&= \frac{E(X^2) - [E(X)]^2}{1} \\
&= Var(X)
\end{aligned}
$$

14.(b) If $X \sim$ Negative Binomial(k, p) then

$$
M(t) = \left[\frac{p}{1-qe^t}\right]^k \quad \text{for } t < -\log q, \quad q = 1-p
$$

Therefore

$$
\begin{aligned}
K(t) &= \log M(t) = k \log\left(\frac{p}{1-qe^t}\right) \\
&= k\log p - k\log\left(1-qe^t\right) \quad \text{for } t < -\log q
\end{aligned}
$$

$$
\begin{aligned}
K'(t) &= -k\frac{\left(-qe^t\right)}{1-qe^t} = \frac{kqe^t}{1-qe^t} \\
E(X) &= K'(0) = \frac{kq}{1-q} = \frac{kq}{p}
\end{aligned}
$$

$$
\begin{aligned}
K''(t) &= kq\left[\frac{\left(1-qe^t\right)e^t - e^t\left(-qe^t\right)}{\left(1-qe^t\right)^2}\right] \\
Var(X) &= K''(0) = kq\left[\frac{1-q+q}{(1-q)^2}\right] = \frac{kq}{p^2}
\end{aligned}
$$

15.(b)

$$\begin{aligned}
M(t) &= \frac{1+t}{1-t} \\
&= (1+t)\sum_{k=0}^{\infty} t^k \quad \text{for } |t|<1 \text{ by the Geometric series} \\
&= \sum_{k=0}^{\infty} t^k + \sum_{k=0}^{\infty} t^{k+1} \\
&= \left(1+t+t^2+...\right)\left(t+t^2+t^3+...\right) \\
&= 1+\sum_{k=1}^{\infty} 2t^k \quad \text{for } |t|<1 \qquad (10.15)
\end{aligned}$$

Since

$$M(t)=\sum_{k=0}^{\infty}\frac{M^{(k)}(0)}{k!}t^k=\sum_{k=0}^{\infty}\frac{E(X^k)}{k!}t^k \qquad (10.16)$$

then by matching coefficients in the two series (10.15) and (10.16) we have

$$\frac{E(X^k)}{k!}=2 \quad \text{for } k=1,2,...$$

or

$$E(X^k)=2k! \quad \text{for } k=1,2,...$$

15.(c)

$$\begin{aligned}
M(t) &= \frac{e^t}{1-t^2} \\
&= \left(\sum_{i=0}^{\infty}\frac{t^i}{i!}\right)\left(\sum_{k=0}^{\infty} t^{2k}\right) \quad \text{for } |t|<1 \text{ by the Geometric series} \\
&= \left(1+\frac{t}{1!}+\frac{t^2}{2!}+\frac{t^3}{3!}+...\right)\left(1+t^2+t^4+...\right) \quad \text{for } |t|<1 \\
&= 1+\left(\frac{1}{1!}\right)t+\left(1+\frac{1}{2!}\right)t^2+\left(\frac{1}{1!}+\frac{1}{3!}\right)t^3 \\
&\quad +\left(1+\frac{1}{2!}+\frac{1}{4!}\right)t^4+\left(\frac{1}{1!}+\frac{1}{3!}+\frac{1}{5!}\right)t^5+... \quad \text{for } |t|<1
\end{aligned}$$

Since

$$M(t)=\sum_{k=0}^{\infty}\frac{M^{(k)}(0)}{k!}t^k=\sum_{k=0}^{\infty}\frac{E(X^k)}{k!}t^k$$

then by matching coefficients in the two series we have

$$\begin{aligned}
E\left(X^{2k}\right) &= (2k)!\sum_{i=0}^{k}\frac{1}{(2i)!} \quad \text{for } k=1,2,... \\
E\left(X^{2k+1}\right) &= (2k+1)!\sum_{i=0}^{k}\frac{1}{(2i+1)!} \quad \text{for } k=1,2,...
\end{aligned}$$

16. (a)

$$\begin{aligned}
M_Y(t) &= E\left(e^{tY}\right) = E\left(e^{t|Z|}\right) = \int_{-\infty}^{\infty} e^{t|z|}\frac{1}{\sqrt{2\pi}}e^{-z^2/2}dz \\
&= \frac{2}{\sqrt{2\pi}}\int_0^{\infty} e^{tz}e^{-z^2/2}dz \quad \text{since } e^{t|z|}e^{-z^2/2} \text{ is an even function} \\
&= \frac{2}{\sqrt{2\pi}}\int_0^{\infty} e^{-\left(z^2-2zt\right)/2}dz = \frac{2e^{t^2/2}}{\sqrt{2\pi}}\int_0^{\infty} e^{-\left(z^2-2zt-t^2\right)/2}dz \\
&= 2e^{t^2/2}\int_0^{\infty}\frac{1}{\sqrt{2\pi}}e^{-(z-t)^2/2}dz \quad \text{let } y=-(z-t)\,, \quad dy=-dz \\
&= 2e^{t^2/2}\int_{-\infty}^{t}\frac{1}{\sqrt{2\pi}}e^{-y^2/2}dy \\
&= 2e^{t^2/2}\Phi(t) \quad \text{for } t\in\Re
\end{aligned}$$

where Φ is the $\mathrm{N}(0,1)$ cumulative distribution function.

16. (b) To find $E(Y) = E(|Z|)$ we first note that

$$\Phi(0) = \frac{1}{2}$$

$$\frac{d}{dt}\Phi(t) = \phi(t) = \frac{1}{\sqrt{2\pi}}e^{-t^2/2}$$

and

$$\phi(0) = \frac{1}{\sqrt{2\pi}}$$

Then

$$\begin{aligned}
\frac{d}{dt}M_Y(t) &= \frac{d}{dt}\left[2e^{t^2/2}\Phi(t)\right] \\
&= 2te^{t^2/2}\Phi(t) + 2e^{t^2/2}\phi(t)
\end{aligned}$$

Therefore

$$\begin{aligned}
E(Y) &= E(|Z|) \\
&= \frac{d}{dt}M_Y(t)\,|_{t=0} \\
&= \left[2te^{t^2/2}\Phi(t) + 2e^{t^2/2}\phi(t)\right]|_{t=0} \\
&= 0 + \frac{2}{\sqrt{2\pi}} = \sqrt{\frac{2}{\pi}}
\end{aligned}$$

To find $Var(Y) = Var(|Z|)$ we note that

$$\begin{aligned}\frac{d}{dt}\phi(t) &= \phi'(t)\\ &= \frac{d}{dt}\left(\frac{1}{\sqrt{2\pi}}e^{-t^2/2}\right)\\ &= \frac{-te^{-t^2/2}}{\sqrt{2\pi}}\end{aligned}$$

and

$$\phi'(0) = 0$$

Therefore

$$\begin{aligned}\frac{d^2}{dt^2}M_Y(t) &= \frac{d}{dt}\left[\frac{d}{dt}M_Y(t)\right]\\ &= \frac{d}{dt}\left[2te^{t^2/2}\Phi(t) + 2e^{t^2/2}\phi(t)\right]\\ &= 2\frac{d}{dt}\left(e^{t^2/2}\right)[t\Phi(t) + \phi(t)]\\ &= 2\left\{\left(e^{t^2/2}\right)\left[t\phi(t) + \Phi(t) + \phi'(t)\right] + \left(te^{t^2/2}\right)[t\Phi(t) + \phi(t)]\right\}\end{aligned}$$

and

$$\begin{aligned}E\left(Y^2\right) &= \frac{d^2}{dt^2}M_Y(t)\,|_{t=0}\\ &= 2\left\{(1)\left[0 + \Phi(0) + \phi'(0)\right] + (0)\left[(0)\Phi(0) + \phi(0)\right]\right\}\\ &= 2\Phi(0)\\ &= 2\left(\frac{1}{2}\right) = 1\end{aligned}$$

Therefore

$$\begin{aligned}Var(Y) &= Var(|Z|)\\ &= E\left(Y^2\right) - [E(Y)]^2\\ &= 1 - \left(\sqrt{\frac{2}{\pi}}\right)^2\\ &= 1 - \frac{2}{\pi}\\ &= \frac{\pi - 2}{\pi}\end{aligned}$$

18. Since

$$\begin{aligned} M_X(t) &= E(e^{tx}) \\ &= \sum_{j=0}^{\infty} e^{jt} p_j \quad \text{for } |t| < h, \quad h > 0 \end{aligned}$$

then

$$\begin{aligned} M_X(\log s) &= \sum_{j=0}^{\infty} e^{j \log s} p_j \\ &= \sum_{j=0}^{\infty} s^j p_j \quad \text{for } |\log s| < h, \quad h > 0 \end{aligned}$$

which is a power series in s. Similarly

$$M_Y(\log s) = \sum_{j=0}^{\infty} s^j q_j \quad \text{for } |\log s| < h, \quad h > 0$$

which is also a power series in s.

We are given that $M_X(t) = M_Y(t)$ for $|t| < h$, $h > 0$. Therefore

$$M_X(\log s) = M_Y(\log s) \quad \text{for } |\log s| < h, \quad h > 0$$

and

$$\sum_{j=0}^{\infty} s^j p_j = \sum_{j=0}^{\infty} s^j q_j \quad \text{for } |\log s| < h, \quad h > 0$$

Since two power series are equal if and only if their coefficients are all equal we have $p_j = q_j$, $j = 0, 1, ...$ and therefore X and Y have the same distribution.

19. (a) Since X has moment generating function

$$M(t) = \frac{e^t}{1-t^2} \quad \text{for } |t| < 1$$

the moment generating function of $Y = (X-1)/2$ is

$$\begin{aligned} M_Y(t) &= E\left(e^{tY}\right) = E\left[e^{t(X-1)/2}\right] \\ &= e^{-t/2} E\left(e^{\left(\frac{t}{2}\right)X}\right) \\ &= e^{-t/2} M\left(\frac{t}{2}\right) \\ &= e^{-t/2} \frac{e^{t/2}}{1-\left(\frac{t}{2}\right)^2}, = \quad \text{for } \left|\frac{t}{2}\right| < 1 \\ &= \frac{1}{1-\frac{1}{4}t^2} \quad \text{for } |t| < 2 \end{aligned}$$

19. (b)

$$\begin{aligned} M'(t) &= (-1)\left(1-\frac{1}{4}t^2\right)^{-2}\left(\frac{-1}{2}t\right) = \frac{1}{2}t\left(1-\frac{1}{4}t^2\right)^{-2} \\ E(X) &= M'(0) = 0 \end{aligned}$$

$$\begin{aligned} M''(t) &= \frac{1}{2}\left[\left(1-\frac{1}{4}t^2\right)^{-2} + t(-2)\left(1-\frac{1}{4}t^2\right)^{-3}\left(\frac{-1}{2}t\right)\right] \\ E(X^2) &= M''(0) = \frac{1}{2} \\ Var(X) &= E(X^2) - [E(X)]^2 = \frac{1}{2} - 0 = \frac{1}{2} \end{aligned}$$

19. (c) Since the moment generating function of a Double Exponential(μ, β) random variable is

$$M(t) = \frac{e^{\mu t}}{1-\beta^2 t^2} \quad \text{for } |t| < \frac{1}{\beta}$$

and the moment generating function of Y is

$$M_Y(t) = \frac{1}{1-\frac{1}{4}t^2} \quad \text{for } |t| < 2$$

therefore by the Uniqueness Theorem for Moment Generating Functions Y has a Double Exponential$\left(0, \frac{1}{2}\right)$ distribution.

10.2 Chapter 3

1.(a)

$$\begin{aligned}
\frac{1}{k} &= \sum_{y=0}^{\infty}\sum_{x=0}^{\infty} q^2 p^{x+y} = q^2 \sum_{y=0}^{\infty} p^y \left(\sum_{x=0}^{\infty} p^x\right) \\
&= q^2 \sum_{y=0}^{\infty} p^y \left(\frac{1}{1-p}\right) \quad \text{by the Geometric Series since } 0<p<1 \\
&= q \sum_{y=0}^{\infty} p^y \quad \text{since } q = 1-p \\
&= q\left(\frac{1}{1-p}\right) \quad \text{by the Geometric Series} \\
&= 1
\end{aligned}$$

Therefore $k = 1$.

1.(b) The marginal probability function of X is

$$\begin{aligned}
f_1(x) &= P(X=x) = \sum_y f(x,y) = \sum_{y=0}^{\infty} q^2 p^{x+y} = q^2 p^x \left(\sum_{y=0}^{\infty} p^y\right) \\
&= q^2 p^x \left(\frac{1}{1-p}\right) \quad \text{by the Geometric Series} \\
&= qp^x \quad \text{for } x = 0,1,\ldots
\end{aligned}$$

By symmetry marginal probability function of Y is

$$f_2(y) = qp^y \quad \text{for } y = 0,1,\ldots$$

The support set of (X,Y) is $A = \{(x,y) : x = 0,1,\ldots; y = 0,1,\ldots\}$. Since

$$f(x,y) = f_1(x)\, f_2(y) \quad \text{for } (x,y) \in A$$

therefore X and Y are independent random variables.

1.(c)

$$P(X=x|X+Y=t) = \frac{P(X=x, X+Y=t)}{P(X+Y=t)} = \frac{P(X=x, Y=t-x)}{P(X+Y=t)}$$

Now

$$\begin{aligned}
P(X+Y=t) &= \sum_{(x,y):}\sum_{x+y=t} q^2 p^{x+y} = q^2 \sum_{x=0}^{t} p^{x+(t-x)} = q^2 p^t \sum_{x=0}^{t} 1 \\
&= q^2 p^t (t+1) \quad \text{for } t = 0,1,\ldots
\end{aligned}$$

Therefore

$$P(X=x|X+Y=t) = \frac{q^2 p^{x+(t-x)}}{q^2 p^t (t+1)} = \frac{1}{t+1} \quad \text{for } x = 0,1,\ldots,t$$

2.(a)

$$f(x,y) = \frac{e^{-2}}{x!\,(y-x)!} \quad \text{for } x = 0,1,...,y; \quad y = 0,1,...$$

OR

$$f(x,y) = \frac{e^{-2}}{x!\,(y-x)!} \quad \text{for } y = x, x+1,...; \quad x = 0,1,...$$

$$\begin{aligned}
f_1(x) &= \sum_y f(x,y) = \sum_{y=x}^{\infty} \frac{e^{-2}}{x!\,(y-x)!} \\
&= \frac{e^{-2}}{x!} \sum_{y=x}^{\infty} \frac{1}{(y-x)!} \quad \text{let } k = y - x \\
&= \frac{e^{-2}}{x!} \sum_{k=0}^{\infty} \frac{1}{k!} = \frac{e^{-2}e^{1}}{x!} \\
&= \frac{e^{-1}}{x!} \quad x = 0,1,... \text{ by the Exponential Series}
\end{aligned}$$

Note that $X \sim$ Poisson(1).

$$\begin{aligned}
f_2(y) &= \sum_x f(x,y) = \sum_{x=0}^{y} \frac{e^{-2}}{x!\,(y-x)!} \\
&= \frac{e^{-2}}{y!} \sum_{x=0}^{y} \frac{y!}{x!(y-x)!} \\
&= \frac{e^{-2}}{y!} \sum_{x=0}^{y} \binom{y}{x} 1^x \\
&= \frac{e^{-2}}{y!} (1+1)^y \quad \text{by the Binomial Series} \\
&= \frac{2^y e^{-2}}{y!} \quad \text{for } y = 0,1,...
\end{aligned}$$

Note that $Y \sim$ Poisson(2).

2.(b) Since (for example)

$$f(1,2) = e^{-2} \neq f_1(1) f_2(2) = e^{-1} \frac{2^2 e^{-2}}{2!} = 2e^{-3}$$

therefore X and Y are not independent random variables.

OR

The support set of X is $A_1 = \{x : x = 0,1,...\}$, the support set of Y is $A_2 = \{y : y = 0,1,...\}$ and the support set of (X,Y) is $A = \{(x,y) : x = 0,1,...,y; \; y = 0,1,...\}$. Since $A \neq A_1 \times A_2$ therefore by the Factorization Theorem for Independence X and Y are not independent random variables.

3.(a) The support set of (X,Y)

$$\begin{aligned} A &= \left\{(x,y): 0<y<1-x^2,\ \ -1<x<1\right\} \\ &= \left\{(x,y): -\sqrt{1-y}<x<\sqrt{1-y},\ \ 0<y<1\right\} \end{aligned}$$

is pictured in Figure 10.8.

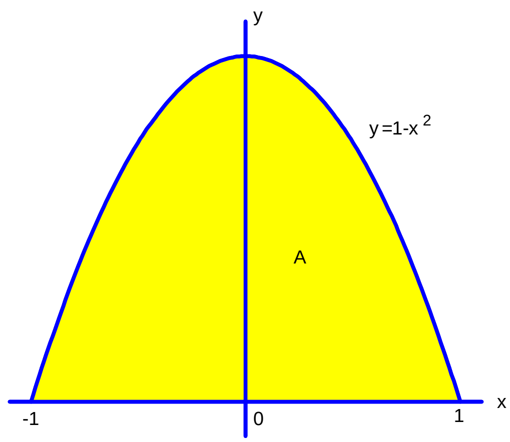


Figure 10.8: Support set for Problem 3(a)

$$\begin{aligned} 1 &= \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} f(x,y)dxdy = k\int\limits_{(x,y)\,\in A}\int \left(x^2+y\right)dydx = k\int\limits_{x=-1}^{1}\int\limits_{y=0}^{1-x^2}\left(x^2+y\right)dydx \\ &= k\int\limits_{-1}^{1}\left[x^2y+\frac{1}{2}y^2|_0^{1-x^2}\right]dx = k\int\limits_{-1}^{1}\left[x^2\left(1-x^2\right)+\frac{1}{2}\left(1-x^2\right)^2\right]dx \\ &= k\int\limits_{0}^{1}\left[2x^2\left(1-x^2\right)+\left(1-x^2\right)^2\right]dx \qquad \text{by symmetry} \\ &= k\int\limits_{0}^{1}\left(1-x^4\right)dx = k\left[x-\frac{1}{5}x^5|_0^1\right] = \frac{4}{5}k \ \ \text{and thus } k=\frac{5}{4} \end{aligned}$$

Therefore

$$f(x,y) = \frac{5}{4}\left(x^2+y\right) \quad \text{for } (x,y)\in A$$

3.(*b*) The marginal probability density function of X is

$$\begin{aligned} f_1(x) &= \int_{-\infty}^{\infty} f(x,y)dy \\ &= \frac{5}{4}\int_{0}^{1-x^2} \left(x^2+y\right) dy \\ &= \frac{5}{8}\left(1-x^4\right) \quad \text{for } -1<x<1 \end{aligned}$$

and 0 otherwise. The support set of X is $A_1 = \{x : -1 < x < 1\}$.

The marginal probability density function of Y is

$$\begin{aligned} f_2(y) &= \int_{-\infty}^{\infty} f(x,y)dx \\ &= \frac{5}{4}\int_{-\sqrt{1-y}}^{\sqrt{1-y}} \left(x^2+y\right) dx \\ &= \frac{5}{2}\int_{0}^{\sqrt{1-y}} \left(x^2+y\right) dx \qquad \text{because of symmetry} \\ &= \frac{5}{2}\left[\frac{1}{3}x^3 + yx|_0^{\sqrt{1-y}}\right] \\ &= \frac{5}{2}\left[\frac{1}{3}(1-y)^{3/2} + y(1-y)^{1/2}\right] \\ &= \frac{5}{6}(1-y)^{1/2}\left[(1-y)+3y\right] \\ &= \frac{5}{6}(1-y)^{1/2}(1+2y) \quad \text{for } 0<y<1 \end{aligned}$$

and 0 otherwise. The support set of Y is $A_2 = \{y : 0 < y < 1\}$.

3.(*c*) The support set A of (X, Y) is not rectangular. To show that X and Y are not independent random variables we only need to find $x \in A_1$, and $y \in A_2$ such that $(x, y) \notin A$.

Let $x = \frac{3}{4}$ and $y = \frac{1}{2}$. Since

$$f\left(\frac{3}{4},\frac{1}{2}\right) = 0 \neq f_1\left(\frac{3}{4}\right) f_2\left(\frac{1}{2}\right) > 0$$

therefore X and Y are not independent random variables.

3.(d) The region of integration is pictured in Figure 10.9.

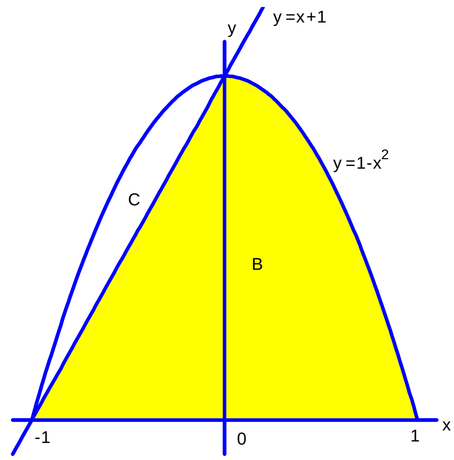


Figure 10.9: Region of integration for Problem 3$\,(d)$

$$\begin{aligned}
P\left(Y \leq X+1\right) &= \int\limits_{(x,y)\,\in B}\int \frac{5}{4}\left(x^2+y\right)dydx = 1-\int\limits_{(x,y)\,\in C}\int \frac{5}{4}\left(x^2+y\right)dydx \\
&= 1-\int\limits_{x=-1}^{0}\int\limits_{y=x+1}^{1-x^2}\frac{5}{4}\left(x^2+y\right)dydx \\
&= 1-\frac{5}{4}\int\limits_{x=-1}^{0}\left[x^2y+\frac{1}{2}y^2|_{x+1}^{1-x^2}\right]dx \\
&= 1-\frac{5}{8}\int\limits_{-1}^{0}\left\{[2x^2(1-x^2)+\left(1-x^2\right)^2]-[2x^2\left(x+1\right)+\left(x+1\right)^2]\right\}dx \\
&= 1-\frac{5}{8}\int\limits_{-1}^{0}\left[-x^4-2x^3-3x^2-2x\right]dx = 1+\frac{5}{8}\left[\frac{1}{5}x^5+\frac{1}{2}x^4+x^3+x^2|_{-1}^{0}\right] \\
&= 1-\frac{5}{8}\left[\frac{1}{5}\left(-1\right)+\frac{1}{2}+\left(-1\right)+1\right] = 1-\frac{5}{8}\left(\frac{-2+5}{10}\right) = 1-\frac{3}{16} \\
&= \frac{13}{16}
\end{aligned}$$

4.(a) The support set of (X, Y)

$$\begin{aligned} A &= \left\{(x, y) : x^2 < y < 1, \quad -1 < x < 1\right\} \\ &= \{(x, y) : -\sqrt{y} < x < \sqrt{y}, \quad 0 < y < 1\} \end{aligned}$$

is pictured in Figure 10.10.

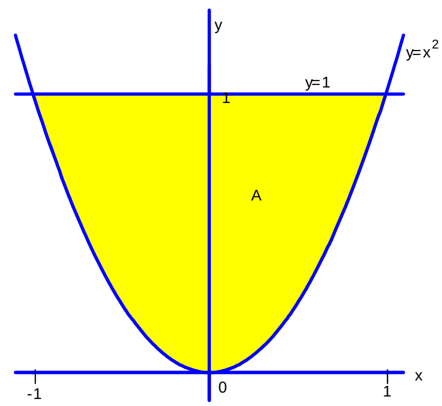


Figure 10.10: Support set for Problem 4 (a)

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)dxdy = k \int\int_{(x,y)\,\in A} x^2 ydydx \\ &= k \int_{x=-1}^{1}\int_{y=x^2}^{1} x^2 ydydx = k \int_{x=-1}^{1} x^2 \left[\frac{1}{2} y^2|_{x^2}^{1}\right] dx \\ &= \frac{k}{2}\int_{x=-1}^{1} x^2\left(1 - x^4\right) dx = \frac{k}{2}\left[\frac{1}{3}x^3 - \frac{1}{7}x^7|_{-1}^{1}\right] \\ &= \frac{k}{2}\left[\frac{1}{3} - \frac{1}{7} - \frac{1}{3}(-1) + \frac{1}{7}(-1)\right] = k\left(\frac{1}{3} - \frac{1}{7}\right) \\ &= \frac{4k}{21} \end{aligned}$$

Therefore $k = 21/4$ and

$$f(x, y) = \frac{21}{4}x^2 y \qquad \text{for } (x, y) \in A$$

and 0 otherwise.

4.(b) The marginal probability density function of X is

$$\begin{aligned}
f_1(x) &= \frac{21}{4}\int\limits_{x^2}^{1} x^2 y\,dy \\
&= \frac{21x^2}{8}\left[y^2\big|_{x^2}^{1}\right] \\
&= \frac{21x^2}{8}\left(1-x^4\right) \quad \text{for } -1<x<1
\end{aligned}$$

and 0 otherwise. The support set of X is $A_1 = \{x : -1 < x < 1\}$.

The marginal probability density function of Y

$$\begin{aligned}
f_2(y) &= \frac{21}{4}\int\limits_{x=-\sqrt{y}}^{\sqrt{y}} x^2 y\,dx \\
&= \frac{21}{2}\int_0^{\sqrt{y}} x^2 y\,dx \quad \text{because of symmetry} \\
&= \frac{7}{2}y\left[x^3\big|_0^{\sqrt{y}}\right] \\
&= \frac{7}{2}y\left(y^{3/2}\right) \\
&= \frac{7}{2}y^{5/2} \quad \text{for } 0<y<1
\end{aligned}$$

and 0 otherwise. The support set of Y is is $A_2 = \{y : 0 < y < 1\}$.

The support set A of (X, Y) is not rectangular. To show that X and Y are not independent random variables we only need to find $x \in A_1$, and $y \in A_2$ such that $(x, y) \notin A$.

Let $x = \frac{1}{2}$ and $y = \frac{1}{10}$. Since

$$f\left(\frac{1}{2}, \frac{1}{10}\right) = 0 \neq f_1\left(\frac{1}{2}\right) f_2\left(\frac{1}{10}\right) > 0$$

therefore X and Y are not independent random variables.

4.(c) The region of integration is pictured in Figure 10.11.

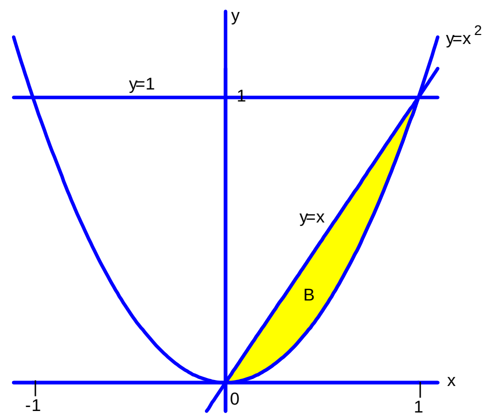


Figure 10.11: Region of intergration for Problem 4 (c)

$$\begin{aligned}
P(X \geq Y) &= \int_{(x,y)\in B}\int \frac{21}{4}x^2 y\,dy\,dx \\
&= \int_{x=0}^{1}\int_{y=x^2}^{x} \frac{21}{4}x^2 y\,dy\,dx \\
&= \frac{21}{8}\int_0^1 x^2\left[y^2|_{x^2}^{x}\right]dx \\
&= \frac{21}{8}\int_0^1 x^2\left(x^2 - x^4\right)dx \\
&= \frac{21}{8}\int_0^1 \left(x^4 - x^6\right)dx \\
&= \frac{21}{8}\left[\frac{1}{5}x^5 - \frac{1}{7}x^7|_0^1\right] \\
&= \frac{21}{8}\left(\frac{7-5}{35}\right) \\
&= \frac{3}{20}
\end{aligned}$$

4.(d) The conditional probability density function of X given $Y = y$ is

$$\begin{aligned} f_1(x|y) &= \frac{f(x,y)}{f_2(y)} \\ &= \frac{\frac{21}{4}x^2y}{\frac{7}{2}y^{5/2}} \\ &= \frac{3}{2}x^2y^{-3/2} \quad \text{for } -\sqrt{y} < x < \sqrt{y}, \quad 0 < y < 1 \end{aligned}$$

and 0 otherwise. Check:

$$\begin{aligned} \int_{-\infty}^{\infty} f_1(x|y)\,dx &= \frac{1}{2}y^{-3/2}\int_{-\sqrt{y}}^{\sqrt{y}} 3x^2dx \\ &= y^{-3/2}\left[x^3|_0^{\sqrt{y}}\right] \\ &= y^{-3/2}y^{3/2} \\ &= 1 \end{aligned}$$

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

$$\begin{aligned} f_2(y|x) &= \frac{f(x,y)}{f_1(x)} \\ &= \frac{\frac{21}{4}x^2y}{\frac{21x^2}{8}(1-x^4)} \\ &= \frac{2y}{(1-x^4)} \quad \text{for } x^2 < y < 1, \quad -1 < x < 1 \end{aligned}$$

and 0 otherwise. Check:

$$\begin{aligned} \int_{-\infty}^{\infty} f_2(y|x)\,dy &= \frac{1}{(1-x^4)}\int_{x^2}^{1} 2ydy \\ &= \frac{1}{(1-x^4)}\left[y^2|_{x^2}^{1}\right] \\ &= \frac{1-x^4}{1-x^4} \\ &= 1 \end{aligned}$$

6.(d) (i) The support set of (X, Y)

$$A = \{(x, y) : 0 < x < y < 1\}$$

is pictured in Figure 10.12.

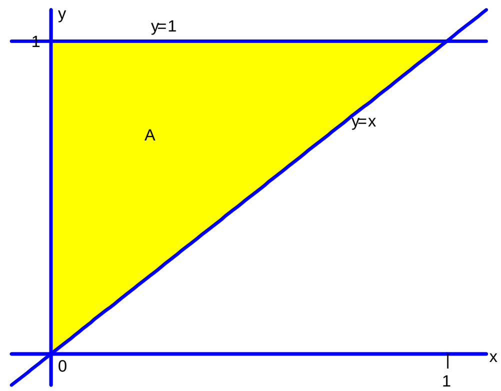


Figure 10.12: Graph of support set for Problem 6(d) (i)

$$\begin{aligned}
1 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\, dxdy = k \int\int_{(x,y)\,\in A} (x + y)\, dydx \\
&= \int_{y=0}^{1}\int_{x=0}^{y} k(x + y)\, dxdy = k \int_{y=0}^{1} \left[\left(\frac{1}{2}x^2 + xy \right) |_{x=0}^{y} \right] dy \\
&= k \int_{0}^{1} \left(\frac{1}{2}y^2 + y^2 \right) dx \\
&= k \int_{0}^{1} \frac{3}{2} y^2 dy = \frac{k}{2} \left[y^3|_0^1 \right] \\
&= \frac{k}{2}
\end{aligned}$$

Therefore $k = 2$ and

$$f(x, y) = 2(x + y) \quad \text{for } (x, y) \in A$$

and 0 otherwise.

6.(d) (ii) The marginal probability density function of X is

$$\begin{aligned}
f_1(x) &= \int_{-\infty}^{\infty} f(x,y)\,dy = \int_{y=x}^{1} 2(x+y)\,dy \\
&= \left(2xy+y^2\right)\Big|_{y=x}^{1} \\
&= (2x+1)-\left(2x^2+x^2\right) \\
&= 1+2x-3x^2 \quad \text{for } 0<x<1
\end{aligned}$$

and 0 otherwise. The support set of X is $A_1=\{x:0<x<1\}$.

The marginal probability density function of Y is

$$\begin{aligned}
f_2(y) &= \int_{-\infty}^{\infty} f(x,y)\,dx = \int_{x=0}^{y} 2(x+y)\,dx \\
&= \left(x^2+2yx\right)\Big|_{x=0}^{y} \\
&= y^2+2y^2-0 \\
&= 3y^2 \quad \text{for } 0<y<1
\end{aligned}$$

and 0 otherwise. The support set of Y is $A_2=\{y:0<y<1\}$.

6.(d) (iii) The conditional probability density function of X given $Y=y$ is

$$\begin{aligned}
f_1(x|y) &= \frac{f(x,y)}{f_2(y)} \\
&= \frac{2(x+y)}{3y^2} \quad \text{for } 0<x<y<1
\end{aligned}$$

Check:

$$\int_{-\infty}^{\infty} f_1(x|y)\,dx = \int_{x=0}^{y} \frac{2(x+y)}{3y^2}dx = \frac{1}{3y^2}\int_{x=0}^{y} 2(x+y)\,dx = \frac{3y^2}{3y^2} = 1$$

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

$$\begin{aligned}
f_2(y|x) &= \frac{f(x,y)}{f_1(x)} \\
&= \frac{2(x+y)}{1+2x-3x^2} \quad \text{for } 0<x<y<1
\end{aligned}$$

and 0 otherwise. Check:

$$\int_{-\infty}^{\infty} f_2(y|x)\,dy = \int_{y=x}^{1} \frac{2(x+y)}{1+2x-3x^2}dy = \frac{1}{1+2x-3x^2}\int_{y=x}^{1} 2(x+y)\,dx = \frac{1+2x-3x^2}{1+2x-3x^2} = 1$$

6.(d) (iv)

$$\begin{aligned}
E\left(X|y\right) &= \int_{-\infty}^{\infty} x f_1\left(x|y\right) dx \\
&= \int_{x=0}^{y} x \left[\frac{2\left(x+y\right)}{3y^2}\right] dx \\
&= \frac{1}{3y^2} \int_{x=0}^{y} 2\left(x^2+yx\right) dx \\
&= \frac{1}{3y^2}\left[\frac{2}{3}x^3 + yx^2|_{x=0}^{y}\right] \\
&= \frac{1}{3y^2}\left[\frac{2}{3}y^3 + y^3\right] \\
&= \frac{1}{3y^2}\left[\frac{5}{3}y^3\right] \\
&= \frac{5}{9}y \quad \text{for } 0<y<1
\end{aligned}$$

$$\begin{aligned}
E\left(Y|x\right) &= \int_{-\infty}^{\infty} y f_2\left(y|x\right) dy \\
&= \int_{y=x}^{1} y \left[\frac{2\left(x+y\right)}{1+2x-3x^2}\right] dy \\
&= \frac{1}{1+2x-3x^2}\left[\int_{y=x}^{1} 2\left(xy+y^2\right) dy\right] \\
&= \frac{1}{1+2x-3x^2}\left[xy^2 + \frac{2}{3}y^3|_{y=x}^{1}\right] \\
&= \frac{1}{1+2x-3x^2}\left[\left(x+\frac{2}{3}\right) - \left(x^3+\frac{2}{3}x^3\right)\right] \\
&= \frac{\frac{2}{3}+x-\frac{5}{3}x^3}{1+2x-3x^2} \\
&= \frac{2+3x-5x^3}{3\left(1+2x-3x^2\right)} \quad \text{for } 0<x<1
\end{aligned}$$

6.(f) (i) The support set of (X,Y)

$$\begin{aligned} A &= \{(x,y) : 0 < y < 1-x,\ 0 < x < 1\} \\ &= \{(x,y) : 0 < x < 1-y,\ 0 < y < 1\} \end{aligned}$$

is pictured in Figure 10.13.

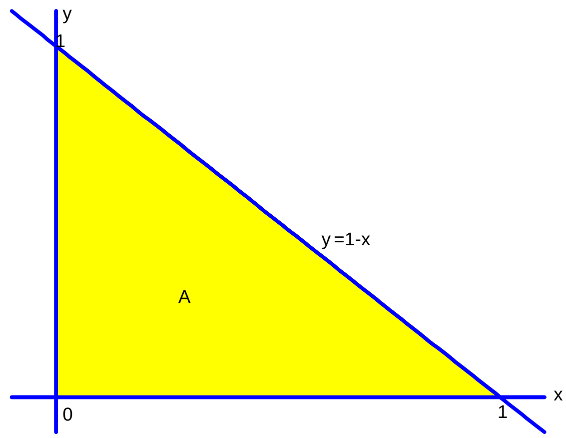


Figure 10.13: Support set for Problem 6 (f) (i)

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dxdx = k \int\int_{(x,y)\,\in A} x^2 y dy dx \\ &= k \int_{x=0}^{1}\int_{y=0}^{1-x} x^2 y dy dx = k \int_0^1 x^2 \left[\frac{1}{2} y^2|_0^{1-x}\right] dx = \frac{k}{2}\int_0^1 x^2 (1-x)^2\, dx \\ &= \frac{k}{2}\int_0^1 \left(x^2 - 2x^3 + x^4\right) dx = \frac{k}{2}\left[\frac{1}{3}x^3 - \frac{1}{2}x^4 + \frac{1}{5}x^5|_0^1\right] \\ &= \frac{k}{2}\left[\frac{1}{3} - \frac{1}{2} + \frac{1}{5}\right] = \frac{k}{2}\left[\frac{10-15+6}{30}\right] \\ &= \frac{k}{60} \end{aligned}$$

Therefore $k = 60$ and

$$f(x,y) = 60x^2 y \quad \text{for } (x,y) \in A$$

and 0 otherwise.

6.(f) (ii) The marginal probability density function of X is

$$\begin{aligned} f_1(x) &= \int_{-\infty}^{\infty} f(x,y)\,dy \\ &= 60x^2 \int_0^{1-x} y dy \\ &= 30x^2 \left[y^2|_0^{1-x} \right] \\ &= 30x^2 (1-x)^2 \quad \text{for } 0<x<1 \end{aligned}$$

and 0 otherwise. The support of X is $A_1 = \{x : 0 < x < 1\}$. Note that $X \sim \text{Beta}(3,3)$.

The marginal probability density function. of Y is

$$\begin{aligned} f_2(y) &= \int_{-\infty}^{\infty} f(x,y)\,dx \\ &= 60y \int_0^{1-y} x^2 dx \\ &= 20y \left[x^3|_0^{1-y} \right] \\ &= 20y (1-y)^3 \quad \text{for } 0<y<1 \end{aligned}$$

and 0 otherwise. The support of Y is $A_2 = \{y : 0 < y < 1\}$. Note that $Y \sim \text{Beta}(2,4)$.

6.(f) (iii) The conditional probability density function of X given $Y = y$ is

$$\begin{aligned} f_1(x|y) &= \frac{f(x,y)}{f_2(y)} \\ &= \frac{60x^2 y}{20y(1-y)^3} \\ &= \frac{3x^2}{(1-y)^3} \quad \text{for } 0<x<1-y,\ 0<y<1 \end{aligned}$$

Check:

$$\int_{-\infty}^{\infty} f_1(x|y)\,dx = \frac{1}{(1-y)^3} \int_0^{1-y} 3x^2 dx = \frac{1}{(1-y)^3} \left[x^3|_0^{1-y} \right] = \frac{(1-y)^3}{(1-y)^3} = 1$$

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

$$\begin{aligned}
f_2(y|x) &= \frac{f(x,y)}{f_1(x)} \\
&= \frac{60x^2 y}{30x^2(1-x)^2} \\
&= \frac{2y}{(1-x)^2} \quad \text{for } 0 < y < 1-x,\ 0 < x < 1
\end{aligned}$$

and 0 otherwise. Check:

$$\int_{-\infty}^{\infty} f_2(y|x)\,dy = \frac{1}{(1-x)^2}\int_0^{1-x} 2y dy = \frac{1}{(1-x)^2}\left[y^2|_0^{1-x}\right] = \frac{(1-x)^2}{(1-x)^2} = 1$$

6.(f) (iv)

$$\begin{aligned}
E(X|y) &= \int_{-\infty}^{\infty} x f_1(x|y)\,dx \\
&= 3(1-y)^{-3}\int_0^{1-y} x\left(x^2\right)dx \\
&= 3(1-y)^{-3}\left[\frac{1}{4}x^4|_0^{1-y}\right] \\
&= \frac{3}{4}(1-y)^{-3}(1-y)^4 \\
&= \frac{3}{4}(1-y) \quad \text{for } 0 < y < 1
\end{aligned}$$

$$\begin{aligned}
E(Y|x) &= \int_{-\infty}^{\infty} y f_2(y|x)\,dy \\
&= 2(1-x)^{-2}\int_0^{1-x} y(y)\,dy \\
&= 2(1-x)^{-2}\left[\frac{1}{3}y^3|_0^{1-x}\right] \\
&= \frac{2}{3}(1-x)^{-2}(1-x)^3 \\
&= \frac{2}{3}(1-x) \quad \text{for } 0 < x < 1
\end{aligned}$$

6.(g) (i) The support set of (X, Y)

$$A = \{(x, y) : 0 < y < x\}$$

is pictured in Figure 10.14.

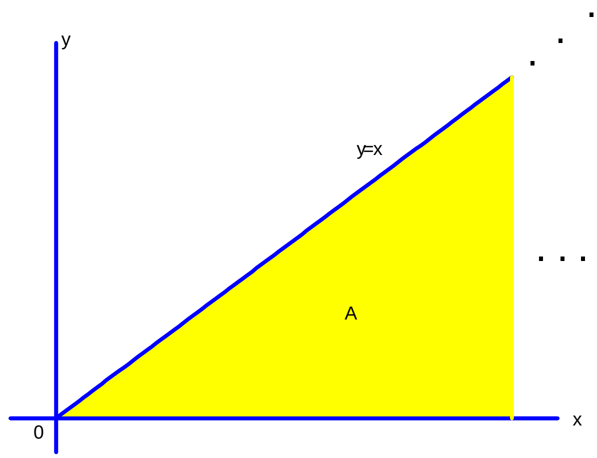


Figure 10.14: Support set for Problem 6 (g) (i)

$$\begin{aligned}
1 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\, dxdx = k \int\int_{(x,y)\,\in A} e^{-x-2y} dydx \\
&= k \int_{y=0}^{\infty}\int_{x=y}^{\infty} e^{-x-2y} dxdy = k \int_{0}^{\infty} e^{-2y} \lim_{b\to\infty} \left[-e^{-x}|_{y}^{b} \right] dy \\
&= k \int_{0}^{\infty} e^{-2y} \left[e^{-y} - \lim_{b\to\infty} e^{-b} \right] dy = k \int_{0}^{\infty} e^{-3y} dy \\
&= \frac{k}{3} \int_{0}^{\infty} 3e^{-3y} dy
\end{aligned}$$

But $3e^{-3y}$ is the probability density function of a Exponential$\left(\frac{1}{3}\right)$ random variable and therefore the integral is equal to 1. Therefore $1 = k/3$ or $k = 3$.

Therefore

$$f(x, y) = 3e^{-x-2y} \quad \text{for} \;\; (x, y) \in A$$

and 0 otherwise.

6.(g) (ii) The marginal probability density function of X is

$$\begin{aligned} f_1(x) &= \int_{-\infty}^{\infty} f(x,y)\,dy = \int_{0}^{x} 3e^{-x-2y}dy \\ &= 3e^{-x}\frac{-1}{2}e^{-2y}|_0^x \\ &= \frac{3}{2}e^{-x}\left(1-e^{-2x}\right) \quad \text{for } x>0 \end{aligned}$$

and 0 otherwise. The support of X is $A_1 = \{x : x > 0\}$.

The marginal probability density function of Y is

$$\begin{aligned} f_2(y) &= \int_{-\infty}^{\infty} f(x,y)\,dx \\ &= \int_{y}^{\infty} 3e^{-x-2y}dx \\ &= 3e^{-2y}\left[\lim_{b\to\infty} -e^{-x}|_y^b\right] \\ &= 3e^{-2y}\left[e^{-y} - \lim_{b\to\infty} e^{-b}\right] \\ &= 3e^{-3y} \quad \text{for } y>0 \end{aligned}$$

and 0 otherwise. The support of Y is $A_2 = \{y : y > 0\}$. Note that $Y \sim$ Exponential$\left(\frac{1}{3}\right)$.

6.(g) (iii) The conditional probability density function of X given $Y = y$ is

$$\begin{aligned} f_1(x|y) &= \frac{f(x,y)}{f_2(y)} \\ &= \frac{3e^{-x-2y}}{3e^{-3y}} \\ &= e^{-(x-y)} \quad \text{for } x>y>0 \end{aligned}$$

Note that $X|Y = y \sim$ Two Parameter Exponential$(y, 1)$.

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

$$\begin{aligned} f_2(y|x) &= \frac{f(x,y)}{f_1(x)} \\ &= \frac{3e^{-x-2y}}{\frac{3}{2}e^{-x}\left(1-e^{-2x}\right)} \\ &= \frac{2e^{-2y}}{1-e^{-2x}} \quad \text{for } 0<y<x \end{aligned}$$

and 0 otherwise. Check:

$$\int_{-\infty}^{\infty} f_2(y|x)\,dy = \int_0^x \frac{2e^{-2y}}{1-e^{-2x}}dy = \frac{1}{1-e^{-2x}}\left[-e^{-2y}|_0^x\right] = \frac{1-e^{-2x}}{1-e^{-2x}} = 1$$

6.(g) (iv)

$$\begin{aligned}
E(X|y) &= \int_{-\infty}^{\infty} x f_1(x|y)\,dx \\
&= \int_y^{\infty} x e^{-(x-y)}dx \\
&= e^y \lim_{b\to\infty}\left[-(x+1)\,e^{-x}|_y^b\right] \\
&= e^y\left[(y+1)\,e^{-y} - \lim_{b\to\infty}\frac{b+1}{e^b}\right] \\
&= y+1 \quad \text{for } y>0
\end{aligned}$$

$$\begin{aligned}
E(Y|x) &= \int_{-\infty}^{\infty} y f_2(y|x)\,dy \\
&= \int_0^x \frac{2ye^{-2y}}{1-e^{-2x}}dy \\
&= \frac{1}{1-e^{-2x}}\left[-\left(y+\frac{1}{2}\right)e^{-2y}|_0^x\right] \\
&= \frac{1}{1-e^{-2x}}\left[\frac{1}{2}-\left(x+\frac{1}{2}\right)e^{-2x}\right] \\
&= \frac{1-(2x+1)\,e^{-2x}}{2\,(1-e^{-2x})} \qquad \text{for } x>0
\end{aligned}$$

7.(a) Since $X \sim$ Uniform$(0,1)$

$$f_1(x) = 1 \quad \text{for } 0 < x < 1$$

The joint probability density function of X and Y is

$$\begin{aligned} f(x,y) &= f_2(y|x)\, f_1(x) \\ &= \frac{1}{1-x}(1) \\ &= \frac{1}{1-x} \quad \text{for } 0 < x < y < 1 \end{aligned}$$

and 0 otherwise.

7.(b) The marginal probability density function of Y is

$$\begin{aligned} f_2(y) &= \int\limits_{-\infty}^{\infty} f(x,y)\, dx \\ &= \int\limits_{x=0}^{y} \frac{1}{1-x} dx \\ &= -\log(1-x)\,|_0^y \\ &= -\log(1-y) \quad \text{for } 0 < y < 1 \end{aligned}$$

and 0 otherwise.

7.(c) The conditional probability density function of X given $Y = y$ is

$$\begin{aligned} f_1(x|y) &= \frac{f(x,y)}{f_2(y)} \\ &= \frac{\frac{1}{1-x}}{-\log(1-y)} \\ &= \frac{1}{(x-1)\log(1-y)} \quad \text{for } 0 < x < y < 1 \end{aligned}$$

and 0 otherwise.

9. Since $Y|\theta \sim$ Binomial(n, θ) then

$$E(Y|\theta) = n\theta \quad \text{and} \quad Var(Y|\theta) = n\theta(1-\theta)$$

Since $\theta \sim$ Beta(a, b) then

$$E(\theta) = \frac{a}{a+b}, \quad Var(\theta) = \frac{ab}{(a+b+1)(a+b)^2}$$

and

$$\begin{aligned}
E\left(\theta^2\right) &= Var(\theta) + [E(\theta)]^2 \\
&= \frac{ab}{(a+b+1)(a+b)^2} + \left(\frac{a}{a+b}\right)^2 \\
&= \frac{ab + a^2(a+b+1)}{(a+b+1)(a+b)^2} \\
&= \frac{a[b + a(a+b) + a]}{(a+b+1)(a+b)^2} = \frac{a(a+b)(a+1)}{(a+b+1)(a+b)^2} \\
&= \frac{a(a+1)}{(a+b+1)(a+b)}
\end{aligned}$$

Therefore

$$E(Y) = E[E(Y|\theta)] = E(n\theta) = nE(\theta) = n\left(\frac{a}{a+b}\right)$$

and

$$\begin{aligned}
Var(Y) &= E[var(Y|\theta)] + Var[E(Y|\theta)] \\
&= E[n\theta(1-\theta)] + Var(n\theta) \\
&= n\left[E(\theta) - E\left(\theta^2\right)\right] + n^2 Var(\theta) \\
&= n\left[\frac{a}{a+b} - \frac{a(a+1)}{(a+b+1)(a+b)}\right] + \frac{n^2 ab}{(a+b+1)(a+b)^2} \\
&= na\left[\frac{(a+b+1) - (a+1)}{(a+b+1)(a+b)}\right] + \frac{n^2 ab}{(a+b+1)(a+b)^2} \\
&= na\left[\frac{b(a+b)}{(a+b+1)(a+b)^2}\right] + \frac{n^2 ab}{(a+b+1)(a+b)^2} \\
&= \frac{nab(a+b+n)}{(a+b+1)(a+b)^2}
\end{aligned}$$

10. (a) Since $Y|\mu \sim$ Poisson(μ), $E(Y|\mu) = \mu$ and since $\mu \sim$ Gamma(α, β), $E(\mu) = \alpha\beta$.

$$E(Y) = E[E(Y|\mu)] = E(\mu) = \alpha\beta$$

Since $Y|\mu \sim$ Poisson(μ), $Var(Y|\mu) = \mu$ and since $\mu \sim$ Gamma(α, β), $Var(\mu) = \alpha\beta^2$.

$$\begin{aligned} Var(Y) &= E[Var(Y|\mu)] + Var[E(Y|\mu)] \\ &= E(\mu) + Var(\mu) \\ &= \alpha\beta + \alpha\beta^2 \end{aligned}$$

10. (b) Since $Y|\mu \sim$ Poisson(μ) and $\mu \sim$ Gamma(α, β) we have

$$f_2(y|\mu) = \frac{\mu^y e^{-\mu}}{y!} \quad \text{for } y = 0, 1, \ldots; \quad \mu > 0$$

and

$$f_1(\mu) = \frac{\mu^{\alpha-1} e^{-\mu/\beta}}{\beta^\alpha \Gamma(\alpha)} \quad \text{for } \mu > 0$$

and by the Product Rule

$$f(\mu, y) = f_2(y|\mu) f_1(\mu) = \frac{\mu^y e^{-\mu}}{y!} \frac{\mu^{\alpha-1} e^{-\mu/\beta}}{\beta^\alpha \Gamma(\alpha)} \quad \text{for } y = 0, 1, \ldots; \quad \mu > 0$$

The marginal probability function of Y is

$$\begin{aligned} f_2(y) &= \int_{-\infty}^{\infty} f(\mu, y)\, d\mu \\ &= \frac{1}{y!\beta^\alpha \Gamma(\alpha)} \int_0^\infty \mu^{y+\alpha-1} e^{-\mu(1+1/\beta)} d\mu \qquad \text{let } x = \mu\left(1 + \frac{1}{\beta}\right) = \mu\left(\frac{\beta+1}{\beta}\right) \\ &= \frac{1}{y!\beta^\alpha \Gamma(\alpha)} \left(\frac{\beta}{\beta+1}\right)^{y+\alpha} \int_0^\infty x^{y+\alpha-1} e^{-x} dx \\ &= \frac{\beta^y}{(1+\beta)^{y+\alpha}} \frac{\Gamma(y+\alpha)}{y!\Gamma(\alpha)} \\ &= \frac{\beta^y}{(1+\beta)^{y+\alpha}} \frac{(y+\alpha-1)(y+\alpha-2)\cdots(\alpha)\Gamma(\alpha)}{y!\Gamma(\alpha)} \\ &= \binom{y+\alpha-1}{y} \left(\frac{1}{1+\beta}\right)^\alpha \left(1 - \frac{1}{1+\beta}\right)^y \quad \text{for } y = 0, 1, \ldots \end{aligned}$$

If α is a nonnegative integer then we recognize this as the probability function of a Negative Binomial$\left(\alpha, \frac{1}{1+\beta}\right)$ random variable.

11. (a) First note that

$$\frac{\partial^{j+k}}{\partial t_1^j \partial t_2^k} e^{t_1 x + t_2 y} = x^j y^k e^{t_1 x + t_2 y}$$

Therefore

$$\frac{\partial^{j+k}}{\partial t_1^j \partial t_2^k} M(t_1, t_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\partial^{j+k}}{\partial t_1^j \partial t_2^k} e^{t_1 x + t_2 y} f(x, y)\, dx dx$$

(assuming the operations of integration and differentiation can be interchanged)

$$\begin{aligned} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^j y^k e^{t_1 x + t_2 y} f(x, y)\, dx dx \\ &= E\left(X^j Y^k e^{t_1 X + t_2 Y}\right) \end{aligned}$$

and

$$\frac{\partial^{j+k}}{\partial t_1^j \partial t_2^k} M(t_1, t_2)\,|_{(t_1, t_2) = (0,0)} = E\left(X^j Y^k\right)$$

as required. Note that this proof is for the case of (X, Y) continuous random variables. The proof for (X, Y) discrete random variables follows in a similar manner with integrals replaced by summations.

11. (b) Suppose that X and Y are independent random variables then

$$M(t_1, t_2) = E\left(e^{t_1 X + t_2 Y}\right) = E\left(e^{t_1 X} e^{t_2 Y}\right) = E\left(e^{t_1 X}\right) E\left(e^{t_2 Y}\right) = M_X(t_1)\, M_Y(t_2)$$

Suppose that $M(t_1, t_2) = M_X(t_1)\, M_Y(t_2)$ for all $|t_1| < h$, $|t_2| < h$ for some $h > 0$. Then

$$\begin{aligned} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{t_1 x + t_2 y} f(x, y)\, dx dx &= \int_{-\infty}^{\infty} e^{t_1 x} f_1(x)\, dx \int_{-\infty}^{\infty} e^{t_2 Y} f_2(y)\, dy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{t_1 x + t_2 y} f_1(x)\, f_2(y)\, dx dx \end{aligned}$$

But by the Uniqueness Theorem for Moment Generating Functions this can only hold if $f(x, y) = f_1(x)\, f_2(y)$ for all (x, y) and therefore X and Y are independent random variables.

Thus we have shown that X and Y are independent random variables if and only if $M_X(t_1)\, M_Y(t_2) = M(t_1, t_2)$.

Note that this proof is for the case of (X, Y) continuous random variables. The proof for (X, Y) discrete random variables follows in a similar manner with integrals replaced by summations.

11. (c) If $(X_1, X_2, X_3) \sim$ Multinomial(n, p_1, p_2, p_3) then

$$M(t_1, t_2) = E\left(e^{t_1 X_1} + e^{t_2 X_2}\right) = \left(p_1 e^{t_1} + p_2 e^{t_2} + p_3\right)^n \quad \text{for } t_1 \in \Re,\ t_2 \in \Re$$

By 11 (a)

$$\begin{aligned}
E(X_1 X_2) &= \frac{\partial^2}{\partial t_1 \partial t_2} M(t_1, t_2) |_{(t_1,t_2)=(0,0)} \\
&= \frac{\partial^2}{\partial t_1 \partial t_2} \left(p_1 e^{t_1} + p_2 e^{t_2} + p_3\right)^n |_{(t_1,t_2)=(0,0)} \\
&= n(n-1) p_1 e^{t_1} p_2 e^{t_2} \left(p_1 e^{t_1} + p_2 e^{t_2} + p_3\right)^{n-2} |_{(t_1,t_2)=(0,0)} \\
&= n(n-1) p_1 p_2
\end{aligned}$$

Also

$$\begin{aligned}
M_{X_1}(t) &= M(t, 0) = \left(p_1 e^t + p_2 + p_3\right)^n \\
&= \left(p_1 e^t + 1 - p_1\right)^n \quad \text{for } t \in \Re
\end{aligned}$$

and

$$\begin{aligned}
E(X_1) &= \frac{d}{dt} M_{X_1}(t) |_{t=0} \\
&= np_1 \left(p_1 e^t + 1 - p_1\right)^{n-1} |_{t=0} \\
&= np_1
\end{aligned}$$

Similarly

$$E(X_2) = np_2$$

Therefore

$$\begin{aligned}
Cov(X_1, X_2) &= E(X_1 X_2) - E(X_1) E(X_2) \\
&= n(n-1) p_1 p_2 - np_1 np_2 \\
&= -np_1 p_2
\end{aligned}$$

13.(*a*) Note that since Σ is a symmetric matrix $\Sigma^T = \Sigma$ so that $\Sigma^{-1}\Sigma = I$ (2×2 identity matrix) and $\Sigma^T\Sigma^{-1} = I$. Also $\boldsymbol{\mu}\mathbf{t}^T = \mathbf{t}\boldsymbol{\mu}^T$ since $\boldsymbol{\mu}\mathbf{t}^T$ is a scalar and $\mathbf{x}\mathbf{t}^T = \mathbf{t}\mathbf{x}^T$ since $(\mathbf{x}-\boldsymbol{\mu})\,\mathbf{t}^T$ is a scalar.

$$
\begin{aligned}
& [\mathbf{x}-(\boldsymbol{\mu}+\mathbf{t}\Sigma)]\,\Sigma^{-1}\,[\mathbf{x}-(\boldsymbol{\mu}+\mathbf{t}\Sigma)]^T - 2\boldsymbol{\mu}\mathbf{t}^T - \mathbf{t}\Sigma\mathbf{t}^T \\
= \; & [(\mathbf{x}-\boldsymbol{\mu})-\mathbf{t}\Sigma]\,\Sigma^{-1}\,[(\mathbf{x}-\boldsymbol{\mu})-\mathbf{t}\Sigma]^T - 2\boldsymbol{\mu}\mathbf{t}^T - \mathbf{t}\Sigma\mathbf{t}^T \\
= \; & [(\mathbf{x}-\boldsymbol{\mu})-\mathbf{t}\Sigma]\,\Sigma^{-1}\left[(\mathbf{x}-\boldsymbol{\mu})^T-\Sigma\mathbf{t}^T\right] - 2\boldsymbol{\mu}\mathbf{t}^T - \mathbf{t}\Sigma\mathbf{t}^T \\
= \; & (\mathbf{x}-\boldsymbol{\mu})\,\Sigma^{-1}\,(\mathbf{x}-\boldsymbol{\mu})^T - (\mathbf{x}-\boldsymbol{\mu})\,\Sigma^{-1}\Sigma\mathbf{t}^T - \mathbf{t}\Sigma\Sigma^{-1}\,(\mathbf{x}-\boldsymbol{\mu})^T + \mathbf{t}\Sigma\Sigma^{-1}\Sigma\mathbf{t}^T - 2\boldsymbol{\mu}\mathbf{t}^T - \mathbf{t}\Sigma\mathbf{t}^T \\
= \; & (\mathbf{x}-\boldsymbol{\mu})\,\Sigma^{-1}\,(\mathbf{x}-\boldsymbol{\mu})^T - (\mathbf{x}-\boldsymbol{\mu})\,\mathbf{t}^T - \mathbf{t}\,(\mathbf{x}-\boldsymbol{\mu})^T + \mathbf{t}\Sigma\mathbf{t}^T - 2\boldsymbol{\mu}\mathbf{t}^T - \mathbf{t}\Sigma\mathbf{t}^T \\
= \; & (\mathbf{x}-\boldsymbol{\mu})\,\Sigma^{-1}\,(\mathbf{x}-\boldsymbol{\mu})^T - \mathbf{x}\mathbf{t}^T + \boldsymbol{\mu}\mathbf{t}^T - \mathbf{t}\mathbf{x}^T + \mathbf{t}\boldsymbol{\mu}^T - 2\boldsymbol{\mu}\mathbf{t}^T \\
= \; & (\mathbf{x}-\boldsymbol{\mu})\,\Sigma^{-1}\,(\mathbf{x}-\boldsymbol{\mu})^T - \mathbf{x}\mathbf{t}^T + \boldsymbol{\mu}\mathbf{t}^T - \mathbf{x}\mathbf{t}^T + \boldsymbol{\mu}\mathbf{t}^T - 2\boldsymbol{\mu}\mathbf{t}^T \\
= \; & (\mathbf{x}-\boldsymbol{\mu})\,\Sigma^{-1}\,(\mathbf{x}-\boldsymbol{\mu})^T - 2\mathbf{x}\mathbf{t}^T \\
= \; & (\mathbf{x}-\boldsymbol{\mu})\,\Sigma^{-1}\,(\mathbf{x}-\boldsymbol{\mu})^T - 2\mathbf{x}\mathbf{t}^T
\end{aligned}
$$

as required.

Now

$$
\begin{aligned}
& M(t_1,t_2) = E\left(e^{t_1X_1+t_2X_2}\right) = E\left[\exp\left(\mathbf{X}\mathbf{t}^T\right)\right] \\
= \; & \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{2\pi|\Sigma|^{1/2}}\exp(\mathbf{x}\mathbf{t}^T)\exp\left[-\frac{1}{2}\,(\mathbf{x}-\boldsymbol{\mu})\,\Sigma^{-1}\,(\mathbf{x}-\boldsymbol{\mu})^T\right]dx_1dx_2 \\
= \; & \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{2\pi|\Sigma|^{1/2}}\exp\left\{-\frac{1}{2}\left[(\mathbf{x}-\boldsymbol{\mu})\,\Sigma^{-1}\,(\mathbf{x}-\boldsymbol{\mu})^T - 2\mathbf{x}\mathbf{t}^T\right]\right\}dx_1dx_2 \\
= \; & \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{2\pi|\Sigma|^{1/2}}\exp\left\{-\frac{1}{2}\left([\mathbf{x}-(\boldsymbol{\mu}+\mathbf{t}\Sigma)]\,\Sigma^{-1}\,[\mathbf{x}-(\boldsymbol{\mu}+\mathbf{t}\Sigma)]^T - 2\boldsymbol{\mu}\mathbf{t}^T - \mathbf{t}\Sigma\mathbf{t}^T\right)\right\}dx_1dx_2 \\
= \; & \exp\left(\boldsymbol{\mu}\mathbf{t}^T + \frac{1}{2}\mathbf{t}\Sigma\mathbf{t}^T\right)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{2\pi|\Sigma|^{1/2}}\exp\left\{-\frac{1}{2}\,[\mathbf{x}-(\boldsymbol{\mu}+\mathbf{t}\Sigma)]\,\Sigma^{-1}\,[\mathbf{x}-(\boldsymbol{\mu}+\mathbf{t}\Sigma)]^T\right\}dx_1dx_2 \\
= \; & \exp\left(\boldsymbol{\mu}\mathbf{t}^T + \frac{1}{2}\mathbf{t}\Sigma\mathbf{t}^T\right) \quad \text{for all } \mathbf{t} = (t_1,t_2) \in \Re^2
\end{aligned}
$$

since

$$
\frac{1}{2\pi|\Sigma|^{1/2}}\exp\left\{-\frac{1}{2}\,[\mathbf{x}-(\boldsymbol{\mu}+\mathbf{t}\Sigma)]\,\Sigma^{-1}\,[\mathbf{x}-(\boldsymbol{\mu}+\mathbf{t}\Sigma)]^T\right\}
$$

is a BVN$(\boldsymbol{\mu}+\mathbf{t}\Sigma, \Sigma)$ probability density function and therefore the integral is equal to one.

13.(b) Since

$$M_{X_1}(t) = M(t,0) = \exp\left(\mu_1 t + \frac{1}{2}t^2\sigma_1^2\right) \quad \text{for } t \in \Re$$

which is the moment generating function of a $\mathrm{N}(\mu_1, \sigma_1^2)$ random variable, then by the Uniqueness Theorem for Moment Generating Functions $X_1 \sim \mathrm{N}(\mu_1, \sigma_1^2)$. By a similar argument $X_2 \sim \mathrm{N}(\mu_2, \sigma_2^2)$.

13.(c) Since

$$\begin{aligned}
& \frac{\partial^2}{\partial t_1 \partial t_2} M(t_1, t_2) = \frac{\partial^2}{\partial t_1 \partial t_2} \exp\left(\boldsymbol{\mu}^T \mathbf{t} + \frac{1}{2}\mathbf{t}^T \Sigma \mathbf{t}\right) \\
= \quad & \frac{\partial^2}{\partial t_1 \partial t_2} \exp\left(\mu_1 t_1 + \mu_2 t_2 + \frac{1}{2}t_1^2\sigma_1^2 + t_1 t_2 \rho\sigma_1\sigma_2 + \frac{1}{2}t_2^2\sigma_2^2\right) \\
= \quad & \frac{\partial}{\partial t_2}\left[\left(\mu_1 + t_1\sigma_1^2 + t_2\rho\sigma_1\sigma_2\right) M(t_1, t_2)\right] \\
= \quad & \rho\sigma_1\sigma_2 M(t_1, t_2) + \left(\mu_1 + t_1\sigma_1^2 + t_2\rho\sigma_1\sigma_2\right)\left(\mu_2 + t_2\sigma_2^2 + t_1\rho\sigma_1\sigma_2\right) M(t_1, t_2)
\end{aligned}$$

therefore

$$\begin{aligned}
E(XY) &= \frac{\partial^2}{\partial t_1 \partial t_2} M(t_1, t_2)\big|_{(t_1,t_2)=(0,0)} \\
&= \rho\sigma_1\sigma_2 + \mu_1\mu_2
\end{aligned}$$

From (b) we know $E(X_1) = \mu_1$ and $E(X_2) = \mu_2$. Therefore

$$Cov(X,Y) = E(XY) - E(X)E(Y) = \rho\sigma_1\sigma_2 + \mu_1\mu_2 - \mu_1\mu_2 = \rho\sigma_1\sigma_2$$

13.(d) By Theorem 3.8.6, X_1 and X_2 are independent random variables if and only if

$$M(t_1, t_2) = M_{X_1}(t_1) M_{X_2}(t_2)$$

then X_1 and X_2 are independent random variables if and only if

$$\begin{aligned}
& \exp\left(\mu_1 t_1 + \mu_2 t_2 + \frac{1}{2}t_1^2\sigma_1^2 + t_1 t_2 \rho\sigma_1\sigma_2 + \frac{1}{2}t_2^2\sigma_2^2\right) \\
= \quad & \exp\left(\mu_1 t_1 + \frac{1}{2}t_1^2\sigma_1^2\right) \exp\left(\mu_2 t_2 + \frac{1}{2}t_2^2\sigma_2^2\right)
\end{aligned}$$

for all $(t_1, t_2) \in \Re^2$ or

$$\mu_1 t_1 + \mu_2 t_2 + \frac{1}{2}t_1^2\sigma_1^2 + t_1 t_2 \rho\sigma_1\sigma_2 + \frac{1}{2}t_2^2\sigma_2^2 = \mu_1 t_1 + \frac{1}{2}t_1^2\sigma_1^2 + \mu_2 t_2 + \frac{1}{2}t_2^2\sigma_2^2$$

or

$$t_1 t_2 \rho\sigma_1\sigma_2 = 0$$

for all $(t_1, t_2) \in \Re^2$ which is true if and only if $\rho = 0$. Therefore X_1 and X_2 are independent random variables if and only if $\rho = 0$.

13.(e) Since

$$E\left[\exp(\mathbf{X}\mathbf{t}^T)\right] = \exp\left(\boldsymbol{\mu}\mathbf{t}^T + \frac{1}{2}\mathbf{t}\Sigma\mathbf{t}^T\right) \quad \text{for } t \in \Re^2$$

therefore

$$\begin{aligned}
& E\{\exp[(\mathbf{X}A + \mathbf{b})\mathbf{t}^T]\} \\
= \; & E\{\exp[\mathbf{X}A\mathbf{t}^T + \mathbf{b}\mathbf{t}^T]\} \\
= \; & \exp(\mathbf{b}\mathbf{t}^T)E\left\{\exp\left[\mathbf{X}\left(\mathbf{t}A^T\right)^T\right]\right\} \\
= \; & \exp(\mathbf{b}\mathbf{t}^T)\exp\left(\boldsymbol{\mu}\left(\mathbf{t}A^T\right)^T + \frac{1}{2}\left(\mathbf{t}A^T\right)\Sigma\left(\mathbf{t}A^T\right)^T\right) \\
= \; & \exp\left(\mathbf{b}\mathbf{t}^T + \boldsymbol{\mu}A\mathbf{t}^T + \frac{1}{2}\mathbf{t}\left(A^T\Sigma A\right)\mathbf{t}^T\right) \\
= \; & \exp\left[\left(\boldsymbol{\mu}A + \mathbf{b}\right)\mathbf{t}^T + \frac{1}{2}\mathbf{t}(A^T\Sigma A)\mathbf{t}\right] \quad \text{for } t \in \Re^2
\end{aligned}$$

which is the moment generating function of a $\text{BVN}\left(\boldsymbol{\mu}A + \mathbf{b}, A^T\Sigma A\right)$ random variable, then by the Uniqueness Theorem for Moment Generating Functions, $\mathbf{X}A+\mathbf{b} \sim \text{BVN}\left(\boldsymbol{\mu}A + \mathbf{b}, A^T\Sigma A\right)$.

13.(f) First note that

$$\Sigma^{-1} = \frac{1}{\sigma_1^2\sigma_2^2\left(1-\rho^2\right)}\begin{bmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix} = \frac{1}{\left(1-\rho^2\right)}\begin{bmatrix} \frac{1}{\sigma_1^2} & -\frac{\rho}{\sigma_1\sigma_2} \\ -\frac{\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{bmatrix}$$

$$|\Sigma|^{1/2} = \left|\begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}\right|^{1/2} = \left[\sigma_1^2\sigma_2^2 - \left(\rho\sigma_1\sigma_2\right)^2\right]^{1/2} = \sigma_1\sigma_2\sqrt{1-\rho^2}$$

$$\left(\mathbf{x}-\boldsymbol{\mu}\right)\Sigma^{-1}\left(\mathbf{x}-\boldsymbol{\mu}\right)^T = \frac{1}{\left(1-\rho^2\right)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]$$

and

$$\begin{aligned}
& \left(\mathbf{x}-\boldsymbol{\mu}\right)\Sigma^{-1}\left(\mathbf{x}-\boldsymbol{\mu}\right)^T - \left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 \\
= \; & \frac{1}{\left(1-\rho^2\right)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right] - \left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 \\
= \; & \frac{1}{\left(1-\rho^2\right)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2 - \left(1-\rho^2\right)\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2\right] \\
= \; & \frac{1}{\sigma_2^2\left(1-\rho^2\right)}\left[\left(x_2-\mu_2\right)^2 - 2\rho\frac{\sigma_2}{\sigma_1}\left(x_1-\mu_1\right)\left(x_2-\mu_2\right) + \frac{\rho^2\sigma_2^2}{\sigma_1^2}\left(x_1-\mu_1\right)^2\right] \\
= \; & \frac{1}{\sigma_2^2\left(1-\rho^2\right)}\left[\left(x_2-\mu_2\right) - \frac{\rho\sigma_2}{\sigma_1}\left(x_1-\mu_1\right)\right]^2 \\
= \; & \frac{1}{\sigma_2^2\left(1-\rho^2\right)}\left\{x_2 - \left[\mu_2 + \frac{\rho\sigma_2}{\sigma_1}\left(x_1-\mu_1\right)\right]\right\}^2
\end{aligned}$$

The conditional probability density function of X_2 given $X_1 = x_1$ is

$$\begin{aligned}
& \frac{f(x_1, x_2)}{f_1(x_1)} \\
= & \frac{\frac{1}{2\pi|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})^T\right]}{\frac{1}{\sqrt{2\pi}\sigma_1} \exp\left[-\frac{1}{2}\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2\right]} \\
= & \frac{\sigma_1}{\sqrt{2\pi}\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{\frac{-1}{2}\left[(\mathbf{x}-\boldsymbol{\mu})\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})^T - \left(\frac{x_1-\mu_1}{\sigma_1}\right)^2\right]\right\} \\
= & \frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2}\frac{1}{\sigma_2^2(1-\rho^2)}\left\{x_2 - \left[\mu_2 + \frac{\rho\sigma_2}{\sigma_1}(x_1-\mu_1)\right]\right\}^2\right] \quad \text{for } x \in \Re^2
\end{aligned}$$

which is the probability density function of a $\mathrm{N}\left(\mu_2 + \frac{\rho\sigma_2}{\sigma_1}(x_1-\mu_1),\ \sigma_2^2(1-\rho^2)\right)$ random variable.

14.(a)

$$\begin{aligned}
& M(t_1, t_2) = E\left(e^{t_1X+t_2Y}\right) \\
= & \int_{y=0}^{\infty}\int_{x=0}^{y} e^{t_1x+t_2y}2e^{-x-y}dxdy = 2\int_0^{\infty} e^{-(1-t_2)y}\left[\int_0^y e^{-(1-t_1)x}dx\right]dy \\
= & 2\int_0^{\infty} e^{-(1-t_2)y}\left[\frac{-e^{-(1-t_1)x}}{1-t_1}\Big|_0^y\right]dy \qquad \text{if } t_1 < 1 \\
= & \frac{2}{(1-t_1)}\int_0^{\infty} e^{-(1-t_2)y}\left[1-e^{-(1-t_1)y}\right]dy \qquad \text{which converges if } t_1 < 1,\ t_2 < 1 \\
= & \frac{2}{(1-t_1)}\int_0^{\infty}\left[e^{-(1-t_2)y} - e^{-(2-t_1-t_2)y}\right]dy \qquad \text{which converges if } t_1+t_2 < 2,\ t_2 < 1 \\
= & \frac{2}{(1-t_1)}\lim_{b\to\infty}\left[\frac{-e^{-(1-t_2)y}}{(1-t_2)} + \frac{e^{-(2-t_1-t_2)y}}{(2-t_1-t_2)}\Big|_0^b\right] \\
= & \frac{2}{(1-t_1)}\lim_{b\to\infty}\left[\frac{-e^{-(1-t_2)b}}{(1-t_2)} + \frac{e^{-(2-t_1-t_2)b}}{(2-t_1-t_2)} + \frac{1}{(1-t_2)} - \frac{1}{(2-t_1-t_2)}\right] \\
= & \frac{2}{(1-t_1)}\left[\frac{1}{(1-t_2)} - \frac{1}{(2-t_1-t_2)}\right] \\
= & \frac{2}{(1-t_1)}\left[\frac{2-t_1-t_2-(1-t_2)}{(1-t_2)(2-t_1-t_2)}\right] \\
= & \frac{2}{(1-t_1)}\left[\frac{1-t_1}{(1-t_2)(2-t_1-t_2)}\right] \\
= & \frac{2}{(1-t_2)(2-t_1-t_2)} \qquad \text{for } t_1+t_2 < 2,\ t_2 < 1
\end{aligned}$$

14.(*b*)

$$M_X(t) = M(t,0) = \frac{2}{2-t} = \frac{1}{1-\frac{1}{2}t} \quad \text{for } t<1$$

which is the moment generating function of an Exponential$\left(\frac{1}{2}\right)$ random variable. Therefore by the Uniqueness Theorem for Moment Generating Functions $X \sim$ Exponential$\left(\frac{1}{2}\right)$.

$$M_Y(t) = M(0,t) = \frac{2}{(1-t)(2-t)} = \frac{2}{2-3t+t^2} \quad \text{for } t<1$$

which is not a moment generating function we recognize. We find the probability density function of Y using

$$\begin{aligned} f_2(y) &= \int_0^y 2e^{-x-y}dx = 2e^{-y}\left[-e^{-x}|_0^y\right] \\ &= 2e^{-y}\left(1-e^{-y}\right) \quad \text{for } y>0 \end{aligned}$$

and 0 otherwise.

14.(*c*) $E(X) = \frac{1}{2}$ since $X \sim$ Exponential$\left(\frac{1}{2}\right)$.

Alternatively

$$\begin{aligned} E(X) &= \frac{d}{dt}M_X(t)|_{t=0} = \frac{d}{dt}\left(\frac{2}{2-t}\right)|_{t=0} \\ &= \frac{2(-1)(-1)}{(2-t)^2}|_{t=0} = \frac{2}{4} = \frac{1}{2} \end{aligned}$$

Similarly

$$\begin{aligned} E(Y) &= \frac{d}{dt}M_Y(t)|_{t=0} = \frac{d}{dt}\left(\frac{2}{2-3t+t^2}\right)|_{t=0} \\ &= \frac{2(-1)(-3+2t)}{(2-3t+t^2)^2}|_{t=0} = \frac{6}{4} = \frac{3}{2} \end{aligned}$$

Since

$$\begin{aligned} \frac{\partial^2}{\partial t_1 \partial t_2}M(t_1,t_2) &= 2\frac{\partial^2}{\partial t_1 \partial t_2}\left[\frac{1}{(1-t_2)(2-t_1-t_2)}\right] \\ &= 2\frac{\partial}{\partial t_2}\left[\frac{1}{(1-t_2)(2-t_1-t_2)^2}\right] \\ &= 2\left[\frac{1}{(1-t_2)^2}\frac{1}{(2-t_1-t_2)^2} + \frac{1}{(1-t_2)}\frac{2}{(2-t_1-t_2)^3}\right] \end{aligned}$$

we obtain

$$\begin{aligned} E(XY) &= \frac{\partial^2}{\partial t_1 \partial t_2}M(t_1,t_2)|_{(t_1,t_2)=(0,0)} \\ &= 2\left[\frac{1}{(1-t_2)^2}\frac{1}{(2-t_1-t_2)^2} + \frac{1}{(1-t_2)}\frac{2}{(2-t_1-t_2)^3}\right]|_{(t_1,t_2)=(0,0)} \\ &= 2\left(\frac{1}{4}+\frac{2}{8}\right) = 1 \end{aligned}$$

Therefore

$$\begin{aligned} Cov(X,Y) &= E(XY) - E(X)E(Y) \\ &= 1 - \left(\frac{1}{2}\right)\left(\frac{3}{2}\right) \\ &= \frac{1}{4} \end{aligned}$$

10.3 Chapter 4

1. We are given that X and Y are independent random variables and we want to show that $U = h(X)$ and $V = g(Y)$ are independent random variables. We will assume that X and Y are continuous random variables. The proof for discrete random variables is obtained by replacing integrals by sums.

Suppose X has probability density function $f_1(x)$ and support set A_1, and Y has probability density function $f_2(y)$ and support set A_2. Then the joint probability density function of X and Y is

$$f(x,y) = f_1(x) f_2(y) \quad \text{ for } (x,y) \in A_1 \times A_2$$

Now for any $(u,v) \in \Re^2$

$$\begin{aligned} P(U \le u, V \le v) &= P(h(X) \le u, g(Y) \le v) \\ &= \iint\limits_{B} f_1(x) f_2(y)\, dxdy \end{aligned}$$

where

$$B = \{(x,y) : h(x) \le u,\ g(y) \le v\}$$

Let $B_1 = \{x : h(x) \le u\}$ and $B_2 = \{y : g(y) \le v\}$. Then $B = B_1 \times B_2$.

Since

$$\begin{aligned} P(U \le u, V \le v) &= \iint\limits_{B_1 \times B_2} f_1(x) f_2(y)\, dxdy \\ &= \int\limits_{B_1} f_1(x)\, dx \int\limits_{B_2} f_2(y)\, dy \\ &= P(h(X) \le u) P(g(Y) \le v) \\ &= P(U \le u) P(V \le v) \quad \text{ for all } (u,v) \in \Re^2 \end{aligned}$$

therefore U and V are independent random variables.

2.(a) The transformation

$$S: \; U = X + Y, \;\; V = X$$

has inverse transformation

$$X = V, \quad Y = U - V$$

The support set of (X, Y), pictured in Figure 10.15, is

$$R_{XY} = \{(x, y) : 0 < x + y < 1, \; 0 < x < 1, \; 0 < y < 1\}$$

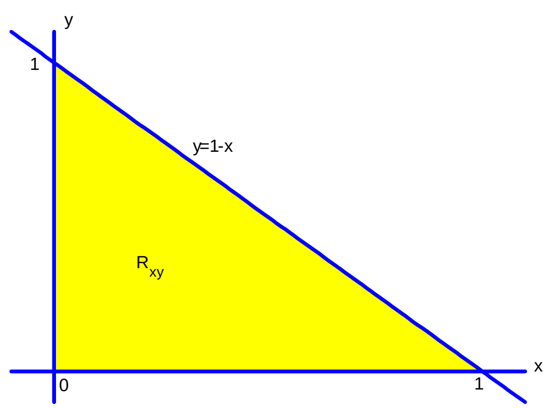


Figure 10.15: R_{XY} for Problem 2 (a)

Under S

$$\begin{aligned} (k, 0) &\rightarrow (k, k) \quad 0 < k < 1 \\ (k, 1-k) &\rightarrow (1, k) \quad 0 < k < 1 \\ (0, k) &\rightarrow (k, 0) \quad 0 < k < 1 \end{aligned}$$

and thus S maps R_{XY} into

$$R_{UV} = \{(u, v) : 0 < v < u < 1\}$$

which is pictured in Figure 10.16.

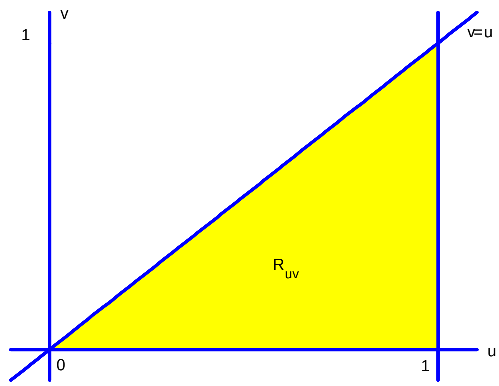


Figure 10.16: R_{UV} for Problem 2 (a)

The Jacobian of the inverse transformation is

$$\frac{\partial(x,y)}{\partial(u,v)} = \begin{vmatrix} 0 & 1 \\ 1 & -1 \end{vmatrix} = -1$$

The joint probability density function of U and V is given by

$$\begin{aligned} g(u,v) &= f(v, u-v)\,|-1| \\ &= 24v(u-v) \quad \text{for } (u,v) \in R_{UV} \end{aligned}$$

and 0 otherwise.

2. (b) The marginal probability density function of U is given by

$$\begin{aligned} g_1(u) &= \int_{-\infty}^{\infty} g(u,v)\,dv \\ &= \int_0^u 24v(u-v)\,dv \\ &= 12uv^2 - 8v^3|_0^u \\ &= 4u^3 \quad \text{for } 0 < u < 1 \end{aligned}$$

and 0 otherwise.

6.(a)

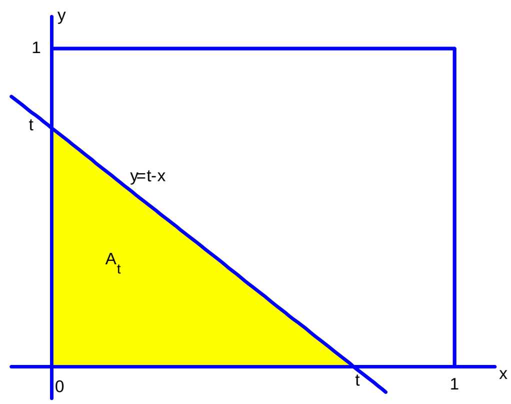


Figure 10.17: Region of integration for $0 \leq t \leq 1$

For $0 \leq t \leq 1$

$$\begin{aligned}
G(t) &= P(T \leq t) = P(X+Y \leq t) \\
&= \int\int_{(x,y) \in A_t} 4xydydx = \int_{x=0}^{t}\int_{y=0}^{t-x} 4xydydx \\
&= \int_0^t 2x\left[y^2|_0^{t-x}\right]dx \\
&= \int_0^t 2x(t-x)^2\,dx \\
&= \int_0^t 2x\left(t^2-2tx+x^2\right)dx \\
&= x^2t^2 - \frac{4}{3}tx^3 + \frac{1}{2}x^4|_0^t \\
&= t^4 - \frac{4}{3}t^4 + \frac{1}{2}t^4 \\
&= \frac{1}{6}t^4
\end{aligned}$$

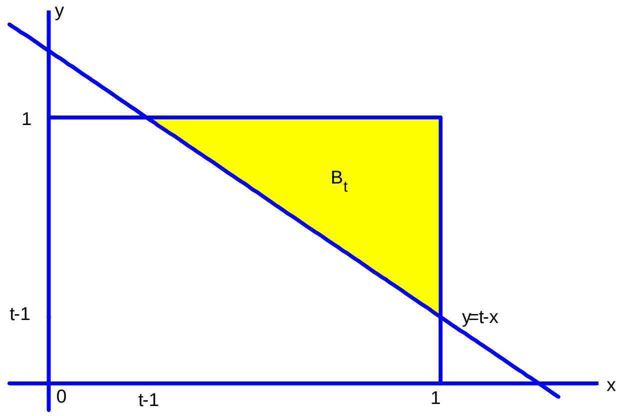


Figure 10.18: Region of integration for $1 \leq t \leq 2$

For $1 \leq t \leq 2$ we use

$$G(t) = P(T \leq t) = P(X + Y \leq t) = 1 - P(X + Y > t)$$

where

$$\begin{aligned}
P(X+Y>t) &= \int\limits_{(x,y)\in\ B_t}\int 4xydydx = \int\limits_{x=t-1}^{1}\int\limits_{y=t-x}^{1} 4xydydx = \int\limits_{t-1}^{1} 2x\left[y^2|_{t-x}^{1}\right]dx \\
&= \int\limits_{t-1}^{1} 2x\left[1-(t-x)^2\right]dx = \int\limits_{t-1}^{1} 2x\left(1-t^2+2tx-x^2\right)dx \\
&= x^2\left(1-t^2\right)+\frac{4}{3}tx^3-\frac{1}{2}x^4|_{t-1}^{1} \\
&= \left(1-t^2\right)+\frac{4}{3}t-\frac{1}{2}-\left[(t-1)^2\left(1-t^2\right)+\frac{4}{3}t(t-1)^3-\frac{1}{2}(t-1)^4\right] \\
&= 1-t^2+\frac{4}{3}t-\frac{1}{2}+\frac{1}{6}(t-1)^3(t+3)
\end{aligned}$$

so

$$G(t) = t^2 - \frac{4}{3}t + \frac{1}{2} - \frac{1}{6}(t-1)^3(t+3) \quad \text{for } 1 \leq t \leq 2$$

The probability density function of $T = X + T$ is

$$\begin{aligned}
g(t) &= \frac{dG(t)}{dt} = \begin{cases} \frac{2}{3}t^3 & \text{if } 0 \leq t \leq 1 \\ 2t - \frac{4}{3} - \left[\frac{1}{2}(t-1)^2(t+3) + \frac{1}{6}(t-1)^3\right] & \text{if } 1 < t \leq 2 \end{cases} \\
&= \begin{cases} \frac{2}{3}t^3 & \text{if } 0 \leq t \leq 1 \\ \frac{2}{3}\left(-t^3+6t-4\right) & \text{if } 1 < t \leq 2 \end{cases}
\end{aligned}$$

and 0 otherwise.

6. (c) The transformation

$$S: \; U = X^2, \quad V = XY$$

has inverse transformation

$$X = \sqrt{U}, \quad Y = V/\sqrt{U}$$

The support set of (X, Y), pictured in Figure 10.19, is

$$R_{XY} = \{(x, y) : 0 < x < 1, \; 0 < y < 1\}$$

Under S

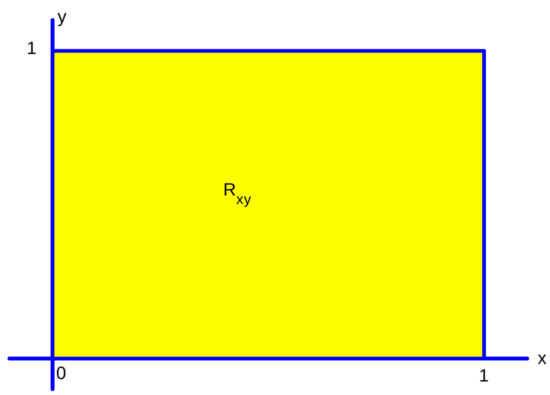


Figure 10.19: Support set of (X, Y) for Problem 6 (c)

$$\begin{aligned}
(k, 0) &\rightarrow \left(k^2, 0\right) \quad 0 < k < 1 \\
(0, k) &\rightarrow (0, 0) \quad 0 < k < 1 \\
(1, k) &\rightarrow (1, k) \quad 0 < k < 1 \\
(k, 1) &\rightarrow \left(k^2, k\right) \quad 0 < k < 1
\end{aligned}$$

and thus S maps R_{XY} into the region

$$\begin{aligned}
R_{UV} &= \left\{(u, v) : 0 < v < \sqrt{u}, \; 0 < u < 1\right\} \\
&= \left\{(u, v) : v^2 < u < 1, \; 0 < v < 1\right\}
\end{aligned}$$

which is pictured in Figure 10.20.

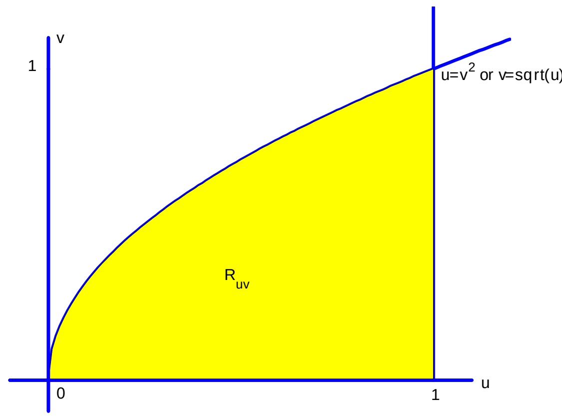


Figure 10.20: Support set of (U, V) for Problem 6 (c)

The Jacobian of the inverse transformation is

$$\frac{\partial (x, y)}{\partial (u, v)} = \begin{vmatrix} \frac{1}{2\sqrt{u}} & 0 \\ \frac{\partial y}{\partial u} & \frac{1}{\sqrt{u}} \end{vmatrix} = \frac{1}{2u}$$

The joint probability density function of U and V is given by

$$\begin{aligned} g(u, v) &= f\left(\sqrt{u}, v/\sqrt{u}\right) \left|\frac{1}{2u}\right| \\ &= 4\sqrt{u}\left(v/\sqrt{u}\right)\left(\frac{1}{2u}\right) \\ &= \frac{2v}{u} \quad \text{for } (u, v) \in R_{UV} \end{aligned}$$

and 0 otherwise.

6.(d) The marginal probability density function of U is

$$\begin{aligned} g_1(u) &= \int_{-\infty}^{\infty} g(u, v)\, dv \\ &= \int_0^{\sqrt{u}} \frac{2v}{u} dv = \frac{1}{u}\left[v^2|_0^{\sqrt{u}}\right] \\ &= 1 \quad \text{for } 0 < u < 1 \end{aligned}$$

and 0 otherwise. Note that $U \sim \text{Uniform}(0, 1)$.

The marginal probability density function of V is

$$\begin{aligned}
g_2(v) &= \int_{-\infty}^{\infty} g(u,v)\,du \\
&= \int_{v^2}^{1} \frac{2v}{u} du \\
&= 2v\left[\log u|_{v^2}^{1}\right] \\
&= -2v\log\left(v^2\right) \\
&= -4v\log(v) \quad \text{for } 0 < v < 1
\end{aligned}$$

and 0 otherwise.

6.(e) The support set of X is $A_1 = \{x : 0 < x < 1\}$ and the support set of Y is $A_2 = \{y : 0 < y < 1\}$. Since

$$f(x,y) = 4xy = 2x(2y)$$

for all $(x,y) \in R_{XY} = A_1 \times A_2$, therefore by the Factorization Theorem for Independence X and Y are independent random variables. Also

$$f_1(x) = 2x \quad \text{for } x \in A_1$$

and

$$f_2(y) = 2y \quad \text{for } y \in A_2$$

so X and Y have the same distribution. Therefore

$$\begin{aligned}
E\left(V^3\right) &= E\left[(XY)^3\right] \\
&= E\left(X^3\right)E\left(Y^3\right) \\
&= \left[E\left(X^3\right)\right]^2 \\
&= \left[\int_0^1 x^3(2x)\,dx\right]^2 \\
&= \left[\frac{2}{5}x^5|_0^1\right]^2 \\
&= \left(\frac{2}{5}\right)^2 \\
&= \frac{4}{25}
\end{aligned}$$

7.(a) The transformation

$$S: \; U = X/Y, \quad V = XY$$

has inverse transformation

$$X = \sqrt{UV}, \quad Y = \sqrt{V/U}$$

The support set of (X, Y), picture in Figure 10.21, is

$$R_{XY} = \{(x, y) : 0 < x < 1, \; 0 < y < 1\}$$

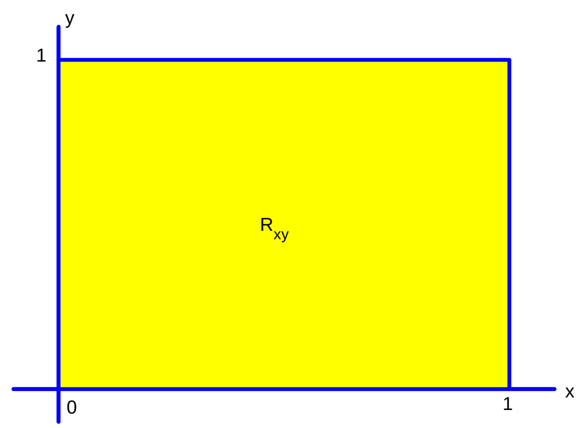


Figure 10.21: Support set of (X, Y) for Problem 7$\,(a)$

Under S

$$\begin{array}{lcll}
(k, 0) & \rightarrow & (\infty, 0) & 0 < k < 1 \\
(0, k) & \rightarrow & (0, 0) & 0 < k < 1 \\
(1, k) & \rightarrow & \left(\dfrac{1}{k}, k\right) & 0 < k < 1 \\
(k, 1) & \rightarrow & (k, k) & 0 < k < 1
\end{array}$$

and thus S maps R_{XY} into

$$R_{UV} = \left\{(u, v) : v < u < \frac{1}{v}, \; 0 < v < 1\right\}$$

which is picture in Figure 10.22.

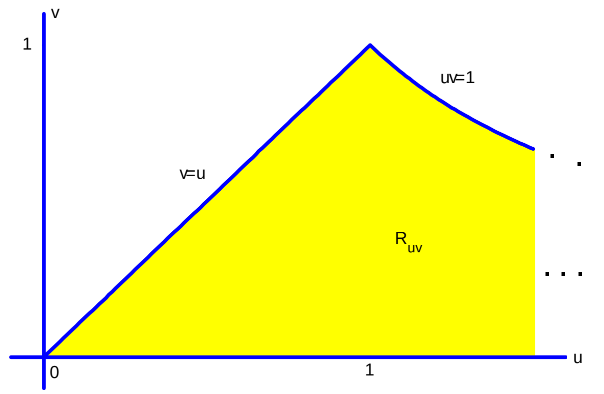


Figure 10.22: Support set of (U, V) for Problem 7 (a)

The Jacobian of the inverse transformation is

$$\frac{\partial(x, y)}{\partial(u, v)} = \begin{vmatrix} \frac{\sqrt{v}}{2\sqrt{u}} & \frac{\sqrt{u}}{2\sqrt{v}} \\ \frac{-\sqrt{v}}{2u^{3/2}} & \frac{1}{2\sqrt{u}\sqrt{v}} \end{vmatrix} = \frac{1}{4u} + \frac{1}{4u} = \frac{1}{2u}$$

The joint probability density function of U and V is given by

$$\begin{aligned} g(u, v) &= f\left(\sqrt{uv}, \sqrt{v/u}\right)\left|\frac{1}{2u}\right| \\ &= 4\sqrt{uv}\sqrt{v/u}\left(\frac{1}{2u}\right) \\ &= \frac{2v}{u} \quad \text{for } (u, v) \in R_{UV} \end{aligned}$$

and 0 otherwise.

7.(b) The support set of (U, V) is $R_{UV} = \left\{(u, v) : v < u < \frac{1}{v},\ 0 < v < 1\right\}$ which is not rectangular. The support set of U is $A_1 = \{u : 0 < u < \infty\}$ and the support set of V is $A_2 = \{v : 0 < v < 1\}$.

Consider the point $\left(\frac{1}{2}, \frac{3}{4}\right) \notin R_{UV}$ so $g\left(\frac{1}{2}, \frac{3}{4}\right) = 0$. Since $\frac{1}{2} \in A_1$ then $g_1\left(\frac{1}{2}\right) > 0$. Since $\frac{3}{4} \in A_2$ then $g_2\left(\frac{3}{4}\right) > 0$. Therefore $g_1\left(\frac{1}{2}\right) g_2\left(\frac{3}{4}\right) > 0$ so

$$g\left(\frac{1}{2}, \frac{3}{4}\right) = 0 \neq g_1\left(\frac{1}{2}\right) g_2\left(\frac{3}{4}\right)$$

and U and V are not independent random variables.

7.(c) The marginal probability density function of V is given by

$$\begin{aligned} g_2(v) &= \int_{-\infty}^{\infty} g(u,v)\,du \\ &= 2v \int_{v}^{1/v} \frac{1}{u} du \\ &= 2v\left[\ln v\right]\Big|_{v}^{1/v} \\ &= -4v\ln v \quad \text{for } 0<v<1 \end{aligned}$$

and 0 otherwise.

The marginal probability density function of U is given by

$$g_1(u) = \int_{-\infty}^{\infty} g(u,v)\,dv$$

$$= \begin{cases} \frac{1}{u}\int_0^u 2v dv & \text{if } \ 0<u<1 \\ \frac{1}{u}\int_0^{1/u} 2v dv & \text{if } \ u \geq 1 \end{cases}$$

$$= \begin{cases} u & \text{if } \ 0<u<1 \\ \frac{1}{u^3} & \text{if } \ u \geq 1 \end{cases}$$

and 0 otherwise.

8. Since $X \sim$ Uniform$(0, \theta)$ and $Y \sim$ Uniform$(0, \theta)$ independently the joint probability density function of X and Y is

$$f(x, y) = \frac{1}{\theta^2} \quad \text{for } (x, y) \in R_{XY}$$

where

$$R_{XY} = \{(x, y) : 0 < x < \theta,\ 0 < y < \theta\}$$

which is pictured in Figure 10.23.The transformation

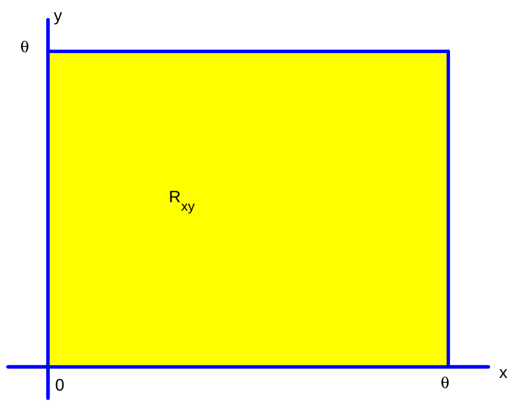


Figure 10.23: Support set of (X, Y) for Problem 8

$$S: \ U = X - Y, \quad V = X + Y$$

has inverse transformation

$$X = \frac{1}{2}(U + V), \quad Y = \frac{1}{2}(V - U)$$

Under S

$$\begin{aligned}
(k, 0) &\rightarrow (k, k) \quad 0 < k < \theta \\
(0, k) &\rightarrow (-k, k) \quad 0 < k < \theta \\
(\theta, k) &\rightarrow (\theta - k, \theta + k) \quad 0 < k < \theta \\
(k, \theta) &\rightarrow (k - \theta, k + \theta) \quad 0 < k < \theta
\end{aligned}$$

S maps R_{XY} into

$$R_{UV} = \{(u, v) : -u < v < 2\theta + u, \ -\theta < u \leq 0 \ \text{ or } \ u < v < 2\theta - u,\ 0 < u < \theta\}$$

which is pictured in Figure 10.24.

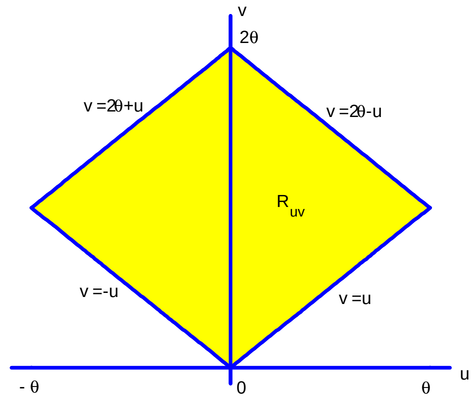


Figure 10.24: Support set of (U, V) for Problem 8

The Jacobian of the inverse transformation is

$$\frac{\partial(x,y)}{\partial(u,v)} = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{vmatrix} = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

The joint probability density function of U and V is given by

$$\begin{aligned} g(u,v) &= f\left(\frac{1}{2}(u+v), \frac{1}{2}(v-u)\right)\left|\frac{1}{2}\right| \\ &= \frac{1}{2\theta^2} \quad \text{for } (u,v) \in R_{UV} \end{aligned}$$

and 0 otherwise.

The marginal probability density function of U is

$$
\begin{aligned}
g_1(u) &= \int_{-\infty}^{\infty} g(u,v)\,dv \\
&= \begin{cases} \frac{1}{2\theta^2} \int\limits_{-u}^{2\theta+u} dv = \frac{1}{2\theta^2}(2\theta+u+u) & \text{for } -\theta < u \le 0 \\ \frac{1}{2\theta^2} \int\limits_{u}^{2\theta-u} dv = \frac{1}{2\theta^2}(2\theta-u-u) & \text{for } 0 < u < \theta \end{cases} \\
&= \begin{cases} \frac{u+\theta}{\theta^2} & \text{for } -\theta < u \le 0 \\ \frac{\theta-u}{\theta^2} & \text{for } 0 < u < \theta \end{cases} \\
&= \frac{\theta-|u|}{\theta^2} \quad \text{for } -\theta < u < \theta
\end{aligned}
$$

and 0 otherwise.

9. (a) The joint probability density function of (Z_1, Z_2) is

$$\begin{aligned} f(z_1, z_2) &= \frac{1}{\sqrt{2\pi}} e^{-z_1^2/2} \frac{1}{\sqrt{2\pi}} e^{-z_2^2/2} \\ &= \frac{1}{2\pi} e^{-\left(z_1^2+z_2^2\right)/2} \quad \text{for } (z_1, z_2) \in \Re^2 \end{aligned}$$

The support set of (Z_1, Z_2) is $\Re^2$.

The transformation

$$X_1 = \mu_1 + \sigma_1 Z_1, \quad X_2 = \mu_2 + \sigma_2 \left[\rho Z_1 + \left(1-\rho^2\right)^{1/2} Z_2\right]$$

has inverse transformation

$$Z_1 = \frac{X_1 - \mu_1}{\sigma_1}, \quad Z_2 = \frac{1}{\sqrt{1-\rho^2}} \left[\frac{X_2 - \mu_2}{\sigma_2} - \rho\left(\frac{X_1 - \mu_1}{\sigma_1}\right)\right]$$

The Jacobian of the inverse transformation is

$$\frac{\partial(z_1, z_2)}{\partial(x_1, x_2)} = \begin{vmatrix} \frac{1}{\sigma_1} & 0 \\ \frac{\partial z_2}{\partial x_2} & \frac{1}{\sigma_2(1-\rho^2)^{1/2}} \end{vmatrix} = \frac{1}{\sigma_1 \sigma_2 \sqrt{1-\rho^2}} = \frac{1}{|\Sigma|^{1/2}}$$

Note that

$$\begin{aligned} & z_1^2 + z_2^2 \\ = & \left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 + \frac{1}{(1-\rho^2)} \left[\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \rho^2\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2\right] \\ = & \frac{1}{(1-\rho^2)} \left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right] \\ = & (\mathbf{x} - \boldsymbol{\mu}) \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})^T \end{aligned}$$

where

$$\mathbf{x} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \text{ and } \boldsymbol{\mu} = \begin{bmatrix} \mu_1 & \mu_2 \end{bmatrix}$$

Therefore the joint probability density function of $\mathbf{X} = \begin{bmatrix} X_1 & X_2 \end{bmatrix}$ is

$$g(\mathbf{x}) = \frac{1}{2\pi |\Sigma|^{1/2}} \exp\left[\frac{-1}{2} (\mathbf{x} - \boldsymbol{\mu}) \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})^T\right] \quad \text{for } \mathbf{x} \in \Re^2$$

and thus $X \sim \text{BVN}(\mu, \Sigma)$.

9. (b) Since $Z_1 \sim \text{N}(0,1)$ and $Z_1 \sim \text{N}(0,1)$ independently we know $Z_1^2 \sim \chi^2(1)$ and $Z_2^2 \sim \chi^2(1)$ and $Z_1^2 + Z_2^2 \sim \chi^2(2)$. From (a) $Z_1^2 + Z_2^2 = (\mathbf{X} - \boldsymbol{\mu}) \Sigma^{-1} (\mathbf{X} - \boldsymbol{\mu})^T$. Therefore $(\mathbf{X} - \boldsymbol{\mu}) \Sigma^{-1} (\mathbf{X} - \boldsymbol{\mu})^T \sim \chi^2(2)$.

10. (a) The joint moment generating function of U and V is

$$\begin{aligned}
M(s,t) &= E\left(e^{sU+tV}\right) \\
&= E\left[e^{s(X+Y)+t(X-Y)}\right] \\
&= E\left[e^{(s+t)X}e^{(s-t)Y}\right] \\
&= E\left[e^{(s+t)X}\right]E\left[e^{(s-t)Y}\right] \quad \text{since } X \text{ and } Y \text{ are independent random variables} \\
&= M_X(s+t)\,M_Y(s-t) \\
&= \left(e^{\mu(s+t)+\sigma^2(s+t)^2/2}\right)\left(e^{\mu(s-t)+\sigma^2(s-t)^2/2}\right) \quad \text{since } X \sim \mathrm{N}\left(\mu,\sigma^2\right) \text{ and } Y \sim \mathrm{N}\left(\mu,\sigma^2\right) \\
&= e^{2\mu s+\sigma^2(2s^2+2t^2)/2} \\
&= \left(e^{(2\mu)s+(2\sigma^2)s^2/2}\right)\left(e^{(2\sigma^2)t^2/2}\right) \quad \text{for } s \in \Re,\ t \in \Re
\end{aligned}$$

10. (b) The moment generating function of U is

$$\begin{aligned}
M_U(s) &= M(s,0) \\
&= e^{(2\mu)s+(2\sigma^2)s^2/2} \quad \text{for } s \in \Re
\end{aligned}$$

which is the moment generating function of a $\mathrm{N}\left(2\mu, 2\sigma^2\right)$ random variable. The moment generating function of V is

$$\begin{aligned}
M_V(t) &= M(0,t) \\
&= e^{(2\sigma^2)t^2/2} \quad \text{for } t \in \Re
\end{aligned}$$

which is the moment generating function of a $\mathrm{N}\left(0, 2\sigma^2\right)$ random variable.

Since

$$M(s,t) = M_U(s)\,M_V(t) \quad \text{for all } s \in \Re,\ t \in \Re$$

therefore by Theorem 3.8.6, U and V are independent random variables.

Also by the Uniqueness Theorem for Moment Generating Functions $U \sim \mathrm{N}\left(2\mu, 2\sigma^2\right)$ and $V \sim \mathrm{N}\left(0, 2\sigma^2\right)$.

12. The transformation defined by

$$Y_1 = \frac{X_1}{X_1 + X_2}, \quad Y_2 = \frac{X_1 + X_2}{X_1 + X_2 + X_3}, \quad Y_3 = X_1 + X_2 + X_3$$

has inverse transformation

$$X_1 = Y_1 Y_2 Y_3, \quad X_2 = Y_2 Y_3 (1 - Y_1), \quad X_3 = Y_3 (1 - Y_2)$$

Let

$$R_X = \{(x_1, x_2, x_3) : 0 < x_1 < \infty, \quad 0 < x_2 < \infty, \quad 0 < x_3 < \infty\}$$

and

$$R_Y = \{(y_1, y_2, y_3) : 0 < y_1 < 1, \quad 0 < y_2 < 1, \quad 0 < y_3 < \infty\}$$

The transformation from $(X_1, X_2, X_3) \to (Y_1, Y_2, Y_3)$ maps R_X into R_Y.

The Jacobian of the transformation from $(X_1, X_2, X_3) \to (Y_1, Y_2, Y_3)$ is

$$\begin{aligned}
\frac{\partial (x_1, x_2, x_3)}{\partial (y_1, y_2, y_3)} &= \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \frac{\partial x_1}{\partial y_3} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \frac{\partial x_2}{\partial y_3} \\ \frac{\partial x_3}{\partial y_1} & \frac{\partial x_3}{\partial y_2} & \frac{\partial x_3}{\partial y_3} \end{vmatrix} \\
&= \begin{vmatrix} y_2 y_3 & y_1 y_3 & y_1 y_2 \\ -y_2 y_3 & y_3 (1 - y_1) & y_2 (1 - y_1) \\ 0 & -y_3 & 1 - y_2 \end{vmatrix} \\
&= y_3 \left[(1 - y_1) y_2^2 y_3 + y_1 y_2^2 y_3 \right] + (1 - y_2) \left[y_2 y_3^2 (1 - y_1) + y_1 y_2 y_3^2 \right] \\
&= y_3 \left(y_2^2 y_3 - y_1 y_2^2 y_3 + y_1 y_2^2 y_3 \right) + (1 - y_2) \left(y_2 y_3^2 - y_1 y_2 y_3^2 + y_1 y_2 y_3^2 \right) \\
&= y_2^2 y_3^2 + y_2 y_3^2 - y_2^2 y_3^2 \\
&= y_2 y_3^2
\end{aligned}$$

Since X_1, X_2, X_3 are independent Exponential(1) random variables the joint probability density function of (X_1, X_2, X_3) is

$$f(x_1, x_2, x_3) = e^{-x_1 - x_2 - x_3} \quad \text{for } (x_1, x_2, x_3) \in R_X$$

The joint probability density function of (Y_1, Y_2, Y_3) is

$$\begin{aligned}
g(y_1, y_2, y_3) &= e^{-y_3} \left| y_2 y_3^2 \right| \\
&= y_2 y_3^2 e^{-y_3} \quad \text{for } (y_1, y_2, y_3) \in R_Y
\end{aligned}$$

and 0 otherwise.

Let

$$\begin{aligned} A_1 &= \{y_1 : 0 < y_1 < 1\} \\ A_2 &= \{y_2 : 0 < y_2 < 1\} \\ A_3 &= \{y_3 : 0 < y_3 < \infty\} \end{aligned}$$

$$\begin{aligned} g_1(y_1) &= 1 \quad \text{for } y_1 \in A_1 \\ g_2(y_2) &= 2y_2 \quad \text{for } y_2 \in A_2 \end{aligned}$$

and

$$g_3(y_3) = \frac{1}{2} y_3^2 \exp(-y_3) \quad \text{for } y_3 \in A_3$$

Since $g(y_1, y_2, y_3) = g_1(y_1) g_2(y_2) g_3(y_3)$ for all $(y_1, y_2, y_3) \in A_1 \times A_2 \times A_3$ therefore by the Factorization Theorem for Independence, (Y_1, Y_2, Y_3) are independent random variables.

Since

$$\int\limits_{A_i} g_i(y_i)\, dy_i = 1 \quad \text{for } i = 1, 2, 3$$

therefore the marginal probability density function of Y_i is equal to $g_i(y_i)$, $i = 1, 2, 3$.

Note that $Y_1 \sim \text{Uniform}(0, 1) = \text{Beta}(1, 1)$, $Y_2 \sim \text{Beta}(2, 1)$, and $Y_3 \sim \text{Gamma}(3, 1)$ independently.

13. Let

$$R_Y = \{(y_1, y_2, y_3) : 0 < y_1 < \infty, \quad 0 < y_2 < 2\pi, \quad 0 < y_3 < \pi\}$$

Consider the transformation from $(X_1, X_2, X_3) \to (Y_1, Y_2, Y_3)$ defined by

$$X_1 = Y_1 \cos Y_2 \sin Y_{3,} \quad X_2 = Y_1 \sin Y_2 \sin Y_3, \quad X_3 = Y_1 \cos Y_3$$

The Jacobian of the transformation from $(X_1, X_2, X_3) \to (Y_1, Y_2, Y_3)$ is

$$\begin{aligned}
\frac{\partial (x_1, x_2, x_3)}{\partial (y_1, y_2, y_3)} &= \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \frac{\partial x_1}{\partial y_3} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \frac{\partial x_2}{\partial y_3} \\ \frac{\partial x_3}{\partial y_1} & \frac{\partial x_3}{\partial y_2} & \frac{\partial x_3}{\partial y_3} \end{vmatrix} \\
&= \begin{vmatrix} \cos y_2 \sin y_3 & -y_1 \sin y_2 \sin y_3 & y_1 \cos y_2 \cos y_3 \\ \sin y_2 \sin y_3 & y_1 \cos y_2 \sin y_3 & y_1 \sin y_2 \cos y_3 \\ \cos y_3 & 0 & -\sin y_3 \end{vmatrix} \\
&= y_1^2 \sin y_3 \begin{vmatrix} \cos y_2 \sin y_3 & -\sin y_2 & \cos y_2 \cos y_3 \\ \sin y_2 \sin y_3 & \cos y_2 & \sin y_2 \cos y_3 \\ \cos y_3 & 0 & -\sin y_3 \end{vmatrix} \\
&= y_1^2 \sin y_3 [\cos y_3 \left(-\sin^2 y_2 \cos y_3 - \cos^2 y_2 \cos y_3\right) \\
&\quad - \sin y_3 \left(\cos^2 y_2 \sin y_3 + \sin^2 y_2 \sin y_3\right)] \\
&= y_1^2 \sin y_3 \left(-\cos^2 y_3 - \sin^2 y_3\right) \\
&= -y_1^2 \sin y_3
\end{aligned}$$

Since the entries in $\frac{\partial(x_1,x_2,x_3)}{\partial(y_1,y_2,y_3)}$ are all continuous functions for $(y_1, y_2, y_3) \in R_Y$ and $\frac{\partial(x_1,x_2,x_3)}{\partial(y_1,y_2,y_3)} \neq 0$ for $(y_1, y_2, y_3) \in R_Y$ therefore by the Inverse Mapping Theorem the transformation has an inverse in the neighbourhood of each point in R_Y.

Since X_1, X_2, X_3 are independent $\mathrm{N}(0, 1)$ random variables the joint probability density function of (X_1, X_2, X_3) is

$$f(x_1, x_2, x_3) = (2\pi)^{-3/2} \exp\left[-\frac{1}{2}\left(x_1^2 + x_2^2 + x_2^2\right)\right] \quad \text{for } (x_1, x_2, x_3) \in \Re^3$$

Now

$$\begin{aligned}
x_1^2 + x_2^2 + x_2^2 &= (y_1 \cos y_2 \sin y_{3,})^2 + (y_1 \sin y_2 \sin y_3)^2 + (y_1 \cos y_3)^2 \\
&= y_1^2 \left(\cos^2 y_2 \sin^2 y_{3,} + \sin^2 y_2 \sin^2 y_3 + \cos^2 y_3\right) \\
&= y_1^2 \left[\sin^2 y_{3,} \left(\cos^2 y_2 + \sin^2 y_2\right) + \cos^2 y_3\right] \\
&= y_1^2 \left(\sin^2 y_{3,} + \cos^2 y_3\right) \\
&= y_1^2
\end{aligned}$$

The joint probability density function of (Y_1, Y_2, Y_3) is

$$\begin{aligned}
g(y_1, y_2, y_3) &= (2\pi)^{-3/2} \exp\left(-\frac{1}{2} y_1^2\right) \left|-y_1^2 \sin y_3\right| \\
&= (2\pi)^{-3/2} \exp\left(-\frac{1}{2} y_1^2\right) y_1^2 \sin y_3 \quad \text{for } (y_1, y_2, y_3) \in R_Y
\end{aligned}$$

and 0 otherwise.

Let

$$\begin{aligned} A_1 &= \{y_1 : y_1 > 0\} \\ A_2 &= \{y_2 : 0 < y_2 < 2\pi\} \\ A_3 &= \{y_3 : 0 < y_3 < \pi\} \end{aligned}$$

$$\begin{aligned} g_1(y_1) &= \frac{2}{\sqrt{2\pi}} y_1^2 \exp\left(-\frac{1}{2} y_1^2\right) \quad \text{for } y_1 \in A_1 \\ g_2(y_2) &= \frac{1}{2\pi} \quad \text{for } y_2 \in A_2 \end{aligned}$$

and

$$g_3(y_3) = \frac{1}{2} \sin y_3 \quad \text{for } y_3 \in A_3$$

Since $g(y_1, y_2, y_3) = g_1(y_1) g_2(y_2) g_3(y_3)$ for all $(y_1, y_2, y_3) \in A_1 \times A_2 \times A_3$ therefore by the Factorization Theorem for Independence, (Y_1, Y_2, Y_3) are independent random variables.

Since

$$\int_{A_i} g_i(y_i)\, dy_i = 1 \quad \text{for } i = 1, 2, 3$$

therefore the marginal probability density function of Y_i is equal to $g_i(y_i)$, $i = 1, 2, 3$.

15. Since $X \sim \chi^2(n)$, the moment generating function of X is

$$M_X(t) = \frac{1}{(1-2t)^{n/2}} \quad \text{for } t < \frac{1}{2}$$

Since $U = X + Y \sim \chi^2(m)$, the moment generating function of U is

$$M_U(t) = \frac{1}{(1-2t)^{m/2}} \quad \text{for } t < \frac{1}{2}$$

The moment generating function of U can also be obtained as

$$\begin{aligned} M_U(t) &= E\left(e^{tU}\right) \\ &= E\left(e^{t(X+Y)}\right) \\ &= E\left(e^{tX}\right) E\left(e^{tY}\right) \quad \text{since } X \text{ and } Y \text{ are independent random variables} \\ &= M_X(t)\, M_Y(t) \end{aligned}$$

By rearranging $M_U(t) = M_X(t)\, M_Y(t)$ we obtain

$$\begin{aligned} M_Y(t) &= \frac{M_U(t)}{M_X(t)} \\ &= \frac{(1-2t)^{-m/2}}{(1-2t)^{-n/2}} \\ &= \frac{1}{(1-2t)^{(m-n)/2}} \quad \text{for } t < \frac{1}{2} \end{aligned}$$

which is the moment generating function of a $\chi^2(m-n)$ random variable. Therefore by the Uniqueness Theorem for Moment Generating Functions $Y \sim \chi^2(m-n)$.

16.(a) Note that

$$\sum_{i=1}^{n}\left(X_i-\bar{X}\right)=\left(\sum_{i=1}^{n}X_i\right)-n\bar{X}=0 \quad\text{and}\quad \sum_{i=1}^{n}\left(s_i-\bar{s}\right)=0$$

Therefore

$$\begin{aligned}\sum_{i=1}^{n}t_iX_i &= \sum_{i=1}^{n}\left(s_i-\bar{s}+\frac{s}{n}\right)\left(X_i-\bar{X}+\bar{X}\right) \\ &= \sum_{i=1}^{n}\left[s_i\left(X_i-\bar{X}\right)+\left(\frac{s}{n}-\bar{s}\right)\left(X_i-\bar{X}\right)+\left(s_i-\bar{s}\right)\bar{X}+\frac{s}{n}\bar{X}\right] \\ &= \sum_{i=1}^{n}s_iU_i+\left(\frac{s}{n}-\bar{s}\right)\sum_{i=1}^{n}\left(X_i-\bar{X}\right)+\bar{X}\sum_{i=1}^{n}\left(s_i-\bar{s}\right)+s\bar{X} \\ &= \sum_{i=1}^{n}s_iU_i+s\bar{X}\end{aligned} \tag{10.17}$$

Also since $X_1, X_2, ..., X_n$ are independent $\mathrm{N}\left(\mu,\sigma^2\right)$ random variables

$$\begin{aligned}E\left[\exp\left(\sum_{i=1}^{n}t_iX_i\right)\right] &= \prod_{i=1}^{n}E\left[\exp\left(t_iX_i\right)\right]=\prod_{i=1}^{n}\exp\left(\mu t_i+\frac{1}{2}\sigma^2t_i^2\right) \\ &= \exp\left(\mu\sum_{i=1}^{n}t_i+\frac{1}{2}\sigma^2\sum_{i=1}^{n}t_i^2\right)\end{aligned} \tag{10.18}$$

Therefore by (10.17) and (10.18)

$$\begin{aligned}E\left[\exp\left(\sum_{i=1}^{n}s_iU_i+s\bar{X}\right)\right] &= E\left[\exp\left(\sum_{i=1}^{n}t_iX_i\right)\right] \\ &= \exp\left(\mu\sum_{i=1}^{n}t_i+\frac{1}{2}\sigma^2\sum_{i=1}^{n}t_i^2\right)\end{aligned} \tag{10.19}$$

16.(b)

$$\sum_{i=1}^{n}t_i=\sum_{i=1}^{n}\left(s_i-\bar{s}+\frac{s}{n}\right)=\sum_{i=1}^{n}\left(s_i-\bar{s}\right)+n\frac{s}{n}=0+s=s \tag{10.20}$$

$$\begin{aligned}\sum_{i=1}^{n}t_i^2 &= \sum_{i=1}^{n}\left(s_i-\bar{s}+\frac{s}{n}\right)^2 \\ &= \sum_{i=1}^{n}\left(s_i-\bar{s}\right)^2+2\frac{s}{n}\sum_{i=1}^{n}\left(s_i-\bar{s}\right)+\sum_{i=1}^{n}\left(\frac{s}{n}\right)^2 \\ &= \sum_{i=1}^{n}\left(s_i-\bar{s}\right)^2+0+\frac{s^2}{n} \\ &= \sum_{i=1}^{n}\left(s_i-\bar{s}\right)^2+\frac{s^2}{n}\end{aligned} \tag{10.21}$$

16.(c)

$$\begin{aligned}
M\left(s_1,...,s_n,s\right) &= E\left[\exp\left(\sum_{i=1}^{n} s_iU_i + s\bar{X}\right)\right] = E\left[\exp\left(\sum_{i=1}^{n} t_iX_i\right)\right] \\
&= \exp\left(\mu\sum_{i=1}^{n} t_i + \frac{\sigma^2}{2}\sum_{i=1}^{n} t_i^2\right) \quad \text{by (10.19)} \\
&= \exp\left\{\mu s + \frac{1}{2}\sigma^2\left[\sum_{i=1}^{n}\left(s_i-\bar{s}\right)^2 + \frac{s^2}{n}\right]\right\} \quad \text{by (10.20) and (10.21)} \\
&= \exp\left[\mu s + \frac{1}{2}\sigma^2\left(\frac{s^2}{n}\right)\right]\exp\left[\frac{1}{2}\sigma^2\sum_{i=1}^{n}\left(s_i-\bar{s}\right)^2\right]
\end{aligned}$$

16.(d) Since

$$M_{\bar{X}}\left(s\right) = M\left(0,...,0,s\right) = \exp\left[\mu s + \frac{1}{2}\sigma^2\left(\frac{s^2}{n}\right)\right]$$

and

$$M_U\left(s_1,...,s_n\right) = M\left(s_1,...,s_n,0\right) = \exp\left[\frac{1}{2}\sigma^2\sum_{i=1}^{n}\left(s_i-\bar{s}\right)^2\right]$$

we have

$$M\left(s_1,...,s_n,s\right) = M_{\bar{X}}\left(s\right)M_U\left(s_1,...,s_n\right)$$

By the Independence Theorem for Moment Generating Functions, $\bar{X}$ and $U = (U_1, U_2, \ldots, U_n)$ are independent random variables. Therefore by Chapter 4, Problem 1, $\bar{X}$ and $\sum_{i=1}^{n} U_i^2 = \sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$ are independent random variables.

10.4 Chapter 5

1. (a) Since $Y_i \sim$ Exponential$(\theta, 1)$, $i = 1, 2, \ldots$ independently then

$$P(Y_i > x) = \int_x^\infty e^{-(y-\theta)} dy = e^{-(x-\theta)} \quad \text{for } x > \theta,\ i = 1, 2, \ldots \tag{10.22}$$

and for $x > \theta$

$$\begin{aligned} F_n(x) &= P(X_n \le x) = P(\min(Y_1, Y_2, \ldots, Y_n) \le x) = 1 - P(Y_1 > x, Y_2 > x, \ldots, Y_n > x) \\ &= 1 - \prod_{i=1}^n P(Y_i > x) \quad \text{since } Y_1, Y_2, \ldots, Y_n \text{ are independent random variables} \\ &= 1 - e^{-n(x-\theta)} \quad \text{using (10.22)} \end{aligned} \tag{10.23}$$

Since

$$\lim_{n\to\infty} F_n(x) = \begin{cases} 1 & \text{if } x > \theta \\ 0 & \text{if } x < \theta \end{cases}$$

therefore

$$X_n \to_p \theta \tag{10.24}$$

(b) By (10.24) and the Limit Theorems

$$U_n = \frac{X_n}{\theta} \to_p 1$$

(c) Now

$$\begin{aligned} P(V_n \le v) &= P[n(X_n - \theta) < v] = P\left(X_n \le \frac{v}{n} + \theta\right) \\ &= 1 - e^{-n(v/n+\theta-\theta)} \quad \text{using (10.23)} \\ &= 1 - e^{-v} \quad \text{for } v \ge 0 \end{aligned}$$

which is the cumulative distribution function of an Exponential(1) random variable. Therefore $V_n \sim$ Exponential(1) for $n = 1, 2, ...$ which implies

$$V_n \to_D V \sim \text{Exponential}(1)$$

(d) Since

$$\begin{aligned} P(W_n \le w) &= P\left(n^2(X_n - \theta) < w\right) = P\left(X_n \le \frac{v}{n^2} + \theta\right) \\ &= 1 - e^{-n\left(w/n^2+\theta-\theta\right)} \\ &= 1 - e^{-w/n} \quad \text{for } w \ge 0 \end{aligned}$$

therefore

$$\lim_{n\to\infty} P(W_n \le w) = 0 \quad \text{for all } w \in \Re$$

which is not a cumulative distribution function. Therefore W_n has no limiting distribution.

2. We first note that

$$\begin{aligned}
P(Y_n \le y) &= P(\max(X_1, X_2, \ldots, X_n) \le y) \\
&= P(X_1 \le y, X_2 \le y \ldots, X_n \le y) \\
&= \prod_{i=1}^{n} P(X_i \le y) \quad \text{since } X_1, X_2, \ldots, X_n \text{ are independent random variables} \\
&= \prod_{i=1}^{n} F(y) \\
&= [F(y)]^n \quad \text{for } y \in \Re \qquad (10.25)
\end{aligned}$$

Since F is a cumulative distribution function, F takes on values between 0 and 1. Therefore the function $n[1 - F(\cdot)]$ takes on values between 0 and n. $G_n(z) = P(Z_n \le z)$, the cumulative distribution function of Z_n, equals 0 for $z \le 0$ and equals 1 for $z \ge n$. For $0 < z < n$

$$\begin{aligned}
G_n(z) &= P(Z_n \le z) \\
&= P(n[1 - F(Y_n)] \le z) \\
&= P\left(F(Y_n) \ge 1 - \frac{z}{n}\right)
\end{aligned}$$

Let A be the support set of X_i. $F(x)$ is an increasing function for $x \in A$ and therefore has an inverse, F^{-1}, which is defined on the interval $(0, 1)$. Therefore for $0 < z < n$

$$\begin{aligned}
G_n(z) &= P\left(F(Y_n) \ge 1 - \frac{z}{n}\right) \\
&= P\left(Y_n \ge F^{-1}\left(1 - \frac{z}{n}\right)\right) \\
&= 1 - P\left(Y_n < F^{-1}\left(1 - \frac{z}{n}\right)\right) \\
&= 1 - \left[F\left(F^{-1}\left(1 - \frac{z}{n}\right)\right)\right]^n \quad \text{by (10.25)} \\
&= 1 - \left(1 - \frac{z}{n}\right)^n
\end{aligned}$$

Since

$$\lim_{n\to\infty} \left[1 - \left(1 - \frac{z}{n}\right)^n\right] = 1 - e^{-z} \quad \text{for } z > 0$$

therefore

$$\lim_{n\to\infty} G_n(z) = \begin{cases} 0 & \text{if } z < 0 \\ 1 - e^{-z} & \text{if } z > 0 \end{cases}$$

which is the cumulative distribution function of a Exponential(1) random variable. Therefore by the definition of convergence in distribution

$$Z_n \to_D Z \sim \text{Exponential}(1)$$

3. The moment generating function of a Poisson(μ) random variable is

$$M(t) = \exp\left[\mu\left(e^t - 1\right)\right] \quad \text{for } t \in \Re$$

The moment generating function of

$$Y_n = \sqrt{n}\left(\bar{X} - \mu\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i - \sqrt{n}\mu$$

is

$$\begin{aligned}
M_n(t) &= E\left(e^{tY_n}\right) \\
&= E\left\{\exp\left[-\sqrt{n}\mu t + \left(\frac{t}{\sqrt{n}}\sum_{i=1}^{n} X_i\right)\right]\right\} \\
&= e^{-\sqrt{n}\mu t} E\left[\exp\left(\frac{t}{\sqrt{n}}\sum_{i=1}^{n} X_i\right)\right] \\
&= e^{-\sqrt{n}\mu t} \prod_{i=1}^{n} E\left[\exp\left(\frac{t}{\sqrt{n}} X_i\right)\right] \\
&\quad \text{since } X_1, X_2, \ldots, X_n \text{ are independent random variables} \\
&= e^{-\sqrt{n}\mu t}\left[M\left(\frac{t}{\sqrt{n}}\right)\right]^n \\
&= e^{-\sqrt{n}\mu t}\exp\left[n\mu\left(e^{t/\sqrt{n}} - 1\right)\right] \\
&= \exp\left[-\sqrt{n}\mu t + n\mu\left(e^{t/\sqrt{n}} - 1\right)\right] \quad \text{for } t \in \Re
\end{aligned}$$

and

$$\log M_n(t) = -\sqrt{n}\mu t + n\mu\left(e^{t/\sqrt{n}} - 1\right) \quad \text{for } t \in \Re$$

By Taylor's Theorem

$$e^x = 1 + x + \frac{x^2}{2} + \frac{e^c}{3!}x^3$$

for some c between 0 and x. Therefore

$$\begin{aligned}
e^{t/\sqrt{n}} &= 1 + \frac{t}{\sqrt{n}} + \frac{1}{2}\left(\frac{t}{\sqrt{n}}\right)^2 + \frac{1}{3!}\left(\frac{t}{\sqrt{n}}\right)^3 e^{c_n} \\
&= 1 + \frac{t}{\sqrt{n}} + \frac{1}{2}\left(\frac{t^2}{n}\right) + \frac{1}{3!}\left(\frac{t^3}{n^{3/2}}\right) e^{c_n}
\end{aligned}$$

for some c_n between 0 and $t/\sqrt{n}$.

Therefore

$$\begin{aligned}
\log M_n(t) &= -\sqrt{n}\mu t + n\mu\left(e^{t/\sqrt{n}} - 1\right) \\
&= -\sqrt{n}\mu t + n\mu\left[\frac{t}{\sqrt{n}} + \frac{1}{2}\left(\frac{t^2}{n}\right) + \frac{1}{3!}\left(\frac{t^3}{n^{3/2}}\right)e^{c_n}\right] \\
&= \frac{1}{2}\mu t^2 + \left(\frac{\mu t^3}{3!}\right)\left(\frac{1}{\sqrt{n}}\right)e^{c_n} \quad \text{for } t \in \Re
\end{aligned}$$

Since

$$\lim_{n\to\infty} c_n = 0$$

it follows that

$$\lim_{n\to\infty} e^{c_n} = e^0 = 1$$

Therefore

$$\begin{aligned}
\lim_{n\to\infty} \log M_n(t) &= \lim_{n\to\infty}\left[\frac{1}{2}\mu t^2 + \left(\frac{\mu t^3}{3!}\right)\left(\frac{1}{\sqrt{n}}\right)e^{c_n}\right] \\
&= \frac{1}{2}\mu t^2 + \left(\frac{\mu t^3}{3!}\right)(0)(1) \\
&= \frac{1}{2}\mu t^2 \quad \text{for } t \in \Re
\end{aligned}$$

and

$$\lim_{n\to\infty} M_n(t) = e^{\frac{1}{2}\mu t^2} \quad \text{for } t \in \Re$$

which is the moment generating function of a $\mathrm{N}(0, \mu)$ random variable.

Therefore by the Limit Theorem for Moment Generating Functions

$$Y_n \to_D Y \sim \mathrm{N}(0, \mu)$$

4. The moment generating function of an Exponential(θ) random variable is

$$M(t) = \frac{1}{1-\theta t} \quad \text{for } t < \frac{1}{\theta}$$

Since $X_1, X_2, \ldots, X_n$ are independent random variables, the moment generating function of

$$\begin{aligned} Z_n &= \frac{1}{\sqrt{n}} \left(\sum_{i=1}^{n} X_i - n\theta \right) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i - \sqrt{n}\theta \end{aligned}$$

is

$$\begin{aligned} M_n(t) &= E\left(e^{tZ_n}\right) \\ &= E\left\{ \exp\left[t \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i - \sqrt{n}\theta \right) \right] \right\} \\ &= E\left[e^{-\sqrt{n}\theta t} \exp\left(\frac{t}{\sqrt{n}} \sum_{i=1}^{n} X_i \right) \right] \\ &= e^{-\sqrt{n}\theta t} \prod_{i=1}^{n} E\left[\exp\left(\frac{t}{\sqrt{n}} X_i \right) \right] \\ &= e^{-\sqrt{n}\theta t} \left[M\left(\frac{t}{\sqrt{n}} \right) \right]^n \\ &= e^{-\sqrt{n}\theta t} \left(\frac{1}{1 - \frac{\theta t}{\sqrt{n}}} \right)^n \\ &= \left[e^{\theta t/\sqrt{n}} \left(1 - \frac{\theta t}{\sqrt{n}} \right) \right]^{-n} \quad \text{for } t < \frac{\sqrt{n}}{\theta} \end{aligned}$$

By Taylor's Theorem

$$e^x = 1 + x + \frac{x^2}{2} + \frac{e^c}{3!} x^3$$

for some c between 0 and x. Therefore

$$\begin{aligned} e^{\theta t/\sqrt{n}} &= 1 + \frac{\theta t}{\sqrt{n}} + \frac{1}{2!} \left(\frac{\theta t}{\sqrt{n}} \right)^2 + \frac{1}{3!} \left(\frac{\theta t}{\sqrt{n}} \right)^3 e^{c_n} \\ &= 1 + \frac{\theta t}{\sqrt{n}} + \frac{1}{2} \left(\frac{\theta^2 t^2}{n} \right) + \frac{1}{3!} \left(\frac{\theta^3 t^3}{n^{3/2}} \right) e^{c_n} \end{aligned}$$

for some c_n between 0 and $\theta t/\sqrt{n}$.

Therefore

$$\begin{aligned}
M_n(t) &= \left\{\left[1+\frac{\theta t}{\sqrt{n}}+\frac{1}{2}\left(\frac{\theta^2 t^2}{n}\right)+\frac{1}{3!}\left(\frac{\theta^3 t^3}{n^{3/2}}\right)e^{c_n}\right]\left(1-\frac{\theta t}{\sqrt{n}}\right)\right\}^{-n} \\
&= [1+\frac{\theta t}{\sqrt{n}}+\frac{1}{2}\left(\frac{\theta^2 t^2}{n}\right)+\frac{1}{3!}\left(\frac{\theta^3 t^3}{n^{3/2}}\right)e^{c_n} \\
&\quad -\frac{\theta t}{\sqrt{n}}-\left(\frac{\theta t}{\sqrt{n}}\right)\left(\frac{\theta t}{\sqrt{n}}\right)-\frac{1}{2}\left(\frac{\theta^2 t^2}{n}\right)\left(\frac{\theta t}{\sqrt{n}}\right)-\frac{1}{3!}\left(\frac{\theta^3 t^3}{n^{3/2}}\right)e^{c_n}\left(\frac{\theta t}{\sqrt{n}}\right)]^{-n} \\
&= \left[1-\frac{1}{2}\left(\frac{\theta^2 t^2}{n}\right)-\frac{1}{2}\left(\frac{\theta^3 t^3}{n^{3/2}}\right)+\frac{1}{3!}\left(\frac{\theta^3 t^3}{n^{3/2}}-\frac{\theta^4 t^4}{n^2}\right)e^{c_n}\right]^{-n} \\
&= \left[1-\frac{1}{2}\left(\frac{\theta^2 t^2}{n}\right)+\frac{\psi(n)}{n}\right]^{-n}
\end{aligned}$$

where

$$\psi(n) = -\frac{1}{2}\left(\frac{\theta^3 t^3}{n^{1/2}}\right)+\frac{1}{3!}\left(\frac{\theta^3 t^3}{\sqrt{n}}-\frac{\theta^4 t^4}{n}\right)e^{c_n}$$

Since

$$\lim_{n\to\infty} c_n = 0$$

it follows that

$$\lim_{n\to\infty} e^{c_n} = e^0 = 1$$

Also

$$\lim_{n\to\infty}\frac{\theta^3 t^3}{\sqrt{n}} = 0, \quad \lim_{n\to\infty}\frac{\theta^4 t^4}{n} = 0$$

and therefore

$$\lim_{n\to\infty}\psi(n) = 0$$

Thus by Theorem 5.1.2

$$\begin{aligned}
\lim_{n\to\infty} M_n(t) &= \lim_{n\to\infty}\left[1-\frac{1}{2}\left(\frac{\theta^2 t^2}{n}\right)+\frac{\psi(n)}{n}\right]^{-n} \\
&= e^{\theta^2 t^2/2} \quad \text{for } t \in \Re
\end{aligned}$$

which is the moment generating function of a $\mathrm{N}(0,\theta^2)$ random variable.

Therefore by the Limit Theorem for Moment Generating Functions

$$Z_n \to_D Z \sim \mathrm{N}(0,\theta^2)$$

6. We can rewrite S_n^2 as

$$\begin{aligned}
S_n^2 &= \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\bar{X}_n\right)^2 \\
&= \frac{1}{n-1}\sum_{i=1}^{n}\left[(X_i-\mu)-\left(\bar{X}_n-\mu\right)\right]^2 \\
&= \frac{1}{n-1}\left[\sum_{i=1}^{n}(X_i-\mu)^2-2\left(\bar{X}_n-\mu\right)\sum_{i=1}^{n}(X_i-\mu)+n\left(\bar{X}_n-\mu\right)^2\right] \\
&= \frac{1}{n-1}\left[\sum_{i=1}^{n}(X_i-\mu)^2-2\left(\bar{X}_n-\mu\right)\left(\sum_{i=1}^{n}X_i-n\mu\right)+n\left(\bar{X}_n-\mu\right)^2\right] \qquad (10.26) \\
&= \frac{1}{n-1}\left[\sum_{i=1}^{n}(X_i-\mu)^2-2\left(\bar{X}_n-\mu\right)n\left(\bar{X}_n-\mu\right)+n\left(\bar{X}_n-\mu\right)^2\right] \qquad (10.27) \\
&= \frac{1}{n-1}\left[\sum_{i=1}^{n}(X_i-\mu)^2-n\left(\bar{X}_n-\mu\right)^2\right] \\
&= \sigma^2\left(\frac{n}{n-1}\right)\left[\frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2-\left(\frac{\bar{X}_n-\mu}{\sigma}\right)^2\right] \qquad (10.28)
\end{aligned}$$

Since $X_1, X_2, \ldots$ are independent and identically distributed random variables with $E(X_i)=\mu$ and $Var(X_i)=\sigma^2<\infty$, then

$$\bar{X}_n \to_p \mu \qquad (10.29)$$

by the Weak Law of Large Numbers.

By (10.29) and the Limit Theorems

$$\left(\frac{\bar{X}_n-\mu}{\sigma}\right)^2 \to_p 0 \qquad (10.30)$$

Let $W_i=\left(\frac{X_i-\mu}{\sigma}\right)^2$, $i=1,2,\ldots$ with

$$E(W_i)=E\left[\left(\frac{X_i-\mu}{\sigma}\right)^2\right]=1$$

and $Var(W_i)<\infty$ since $E\left(X_i^4\right)<\infty$.

Since $W_1, W_2, \ldots$ are independent and identically distributed random variables with $E(W_i)=1$ and $Var(W_i)<\infty$, then

$$\bar{W}_n=\frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 \to_p 1 \qquad (10.31)$$

by the Weak Law of Large Numbers.

By (10.28), (10.30), (10.31) and the Limit Theorems

$$S_n^2 \to_p \sigma^2(1)(1+0)=\sigma^2$$

and therefore

$$\frac{S_n}{\sigma} \to_p 1 \tag{10.32}$$

by the Limit Theorems.

Since $X_1, X_2, \ldots$ are independent and identically distributed random variables with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2 < \infty$, then

$$\frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{\sigma} \to_D Z \sim \mathrm{N}(0,1) \tag{10.33}$$

by the Central Limit Theorem.

By (10.32), (10.33) and Slutsky's Theorem

$$T_n = \frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{S_n} = \frac{\frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{\sigma}}{\frac{S_n}{\sigma}} \to_D \frac{Z}{1} = Z \sim \mathrm{N}(0,1)$$

7. (a) Let $Y_1, Y_2, \ldots$ be independent Binomial$(1, \theta)$ random variables with

$$E(Y_i) = \theta$$

and

$$Var(Y_i) = \theta(1-\theta)$$

for $i = 1, 2, \ldots$
By 4.3.2(1)

$$\sum_{i=1}^{n} Y_i \sim \text{Binomial}(n, \theta)$$

Since $Y_1, Y_2, \ldots$ are independent and identically distributed random variables with $E(Y_i) = \theta$ and $Var(Y_i) = \theta(1-\theta) < \infty$, then by the Weak Law of Large Numbers

$$\frac{1}{n}\sum_{i=1}^{n} Y_i = \bar{Y}_n \to_p \theta$$

Since X_n and $\sum_{i=1}^{n} Y_i$ have the same distribution

$$T_n = \frac{X_n}{n} \to_p \theta \tag{10.34}$$

(b) By (10.34) and the Limit Theorems

$$U_n = \frac{X_n}{n}\left(1 - \frac{X_n}{n}\right) \to_p \theta(1-\theta) \tag{10.35}$$

(c) Since $Y_1, Y_2, \ldots$ are independent and identically distributed random variables with $E(Y_i) = \theta$ and $Var(Y_i) = \theta(1-\theta) < \infty$, then by the Central Limit Theorem

$$\frac{\sqrt{n}\left(\bar{Y}_n - \theta\right)}{\sqrt{\theta(1-\theta)}} \to_D Z \sim \text{N}(0,1)$$

Since X_n and $\sum_{i=1}^{n} Y_i$ have the same distribution

$$S_n = \frac{\sqrt{n}\left(\frac{X_n}{n} - \theta\right)}{\sqrt{\theta(1-\theta)}} \to_D Z \sim \text{N}(0,1) \tag{10.36}$$

By (10.36) and Slutsky's Theorem

$$W_n = \sqrt{n}\left(\frac{X_n}{n} - \theta\right) = S_n\sqrt{\theta(1-\theta)} \to_D W = \sqrt{\theta(1-\theta)}Z$$

Since $Z \sim \text{N}(0,1)$, $W = \sqrt{\theta(1-\theta)}Z \sim \text{N}(0, \theta(1-\theta))$ and therefore

$$W_n \to_D W \sim \text{N}(0, \theta(1-\theta)) \tag{10.37}$$

(d) By (10.35), (10.37) and Slutsky's Theorem

$$Z_n = \frac{W_n}{\sqrt{U_n}} \to_D \frac{W}{\sqrt{\theta(1-\theta)}} = \frac{Z\sqrt{\theta(1-\theta)}}{\sqrt{\theta(1-\theta)}} = Z \sim \mathrm{N}(0,1)$$

(e) To determine the limiting distribution of

$$V_n = \sqrt{n}\left[\arcsin\left(\sqrt{\frac{X_n}{n}}\right) - \arcsin\left(\sqrt{\theta}\right)\right]$$

let $g(x) = \arcsin(\sqrt{x})$ and $a = \theta$. Then

$$\begin{aligned} g'(x) &= \frac{1}{\sqrt{1-(\sqrt{x})^2}}\left(\frac{1}{2\sqrt{x}}\right) \\ &= \frac{1}{2\sqrt{x(1-x)}} \end{aligned}$$

and

$$\begin{aligned} g'(a) &= g'(\theta) \\ &= \frac{1}{2\sqrt{\theta(1-\theta)}} \end{aligned}$$

By (10.37) and the Delta Method

$$V_n \to_D \frac{1}{2\sqrt{\theta(1-\theta)}} Z\sqrt{\theta(1-\theta)} = \frac{Z}{2} \sim \mathrm{N}\left(0, \frac{1}{4}\right)$$

(f) The limiting variance of W_n is equal to $\theta(1-\theta)$ which depends on θ. The limiting variance of Z_n is 1 which does not depend on θ. The limiting variance of V_n is $1/4$ which does not depend on θ. The transformation $g(x) = \arcsin(\sqrt{x})$ is a variance-stabilizing transformation for the Binomial distribution.

8. $X_1, X_2, \ldots$ are independent Geometric(θ) random variables with

$$E(X_i) = \frac{1-\theta}{\theta}$$

and

$$Var(X_i) = \frac{1-\theta}{\theta^2}$$

$i = 1, 2, \ldots$

(a) Since $X_1, X_2, \ldots$ are independent and identically distributed random variables with $E(X_i) = \frac{1-\theta}{\theta}$ and $Var(X_i) = \frac{1-\theta}{\theta^2} < \infty$, then by the Weak Law of Large Numbers

$$\bar{X}_n = \frac{Y_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i \to_p \frac{1-\theta}{\theta} \tag{10.38}$$

(b) Since $X_1, X_2, \ldots$ are independent and identically distributed random variables with $E(X_i) = \frac{1-\theta}{\theta}$ and $Var(X_i) = \frac{1-\theta}{\theta^2} < \infty$, then by the Central Limit Theorem

$$\frac{\sqrt{n}\left(\bar{X}_n - \frac{1-\theta}{\theta}\right)}{\sqrt{\frac{1-\theta}{\theta^2}}} \to_D Z \sim \mathrm{N}(0,1) \tag{10.39}$$

By (10.39) and Slutsky's Theorem

$$W_n = \sqrt{n}\left(\bar{X}_n - \frac{1-\theta}{\theta}\right) = \sqrt{\frac{1-\theta}{\theta^2}}\frac{\sqrt{n}\left(\bar{X}_n - \frac{1-\theta}{\theta}\right)}{\sqrt{\frac{1-\theta}{\theta^2}}} \to_D W = \sqrt{\frac{1-\theta}{\theta^2}}Z$$

Since $Z \sim \mathrm{N}(0,1)$, $W = \sqrt{\frac{1-\theta}{\theta^2}}Z \sim \mathrm{N}\left(0, \frac{1-\theta}{\theta^2}\right)$ and therefore

$$W_n = \sqrt{n}\left(\bar{X}_n - \frac{1-\theta}{\theta}\right) \to_D W \sim \mathrm{N}\left(0, \frac{1-\theta}{\theta^2}\right) \tag{10.40}$$

(c) By (10.38) and the Limit Theorems

$$V_n = \frac{1}{1+\bar{X}_n} \to_p \frac{1}{1+\frac{1-\theta}{\theta}} = \frac{\theta}{\theta + 1 - \theta} = \theta \tag{10.41}$$

(d) To find the distribution of

$$Z_n = \frac{\sqrt{n}\left(V_n - \theta\right)}{\sqrt{V_n^2\left(1 - V_n\right)}}$$

we first note that

$$\begin{aligned}
W_n &= \sqrt{n}\left[\bar{X}_n - \left(\frac{1-\theta}{\theta}\right)\right] \\
&= \sqrt{n}\left[\bar{X}_n - \left(\frac{1}{\theta} - 1\right)\right] \\
&= \sqrt{n}\left(\bar{X}_n + 1 - \frac{1}{\theta}\right) \\
&= \sqrt{n}\left(\frac{1}{V_n} - \frac{1}{\theta}\right)
\end{aligned}$$

Therefore by (10.40)

$$\sqrt{n}\left(\frac{1}{V_n}-\frac{1}{\theta}\right)\rightarrow_D W\sim \mathrm{N}\left(0,\frac{1-\theta}{\theta^2}\right) \tag{10.42}$$

Next we determine the limiting distribution of

$$\sqrt{n}\,(V_n-\theta)$$

which is the numerator of Z_n. Let $g(x)=x^{-1}$ and $a=\theta^{-1}$. Then

$$g'(x)=-x^{-2}$$

and

$$g'(a)=g'\left(\theta^{-1}\right)=-\theta^2$$

By (10.42) and the Delta Theorem

$$\sqrt{n}\,(V_n-\theta)\rightarrow_D \left(-\theta^2\right)\sqrt{\frac{1-\theta}{\theta}}Z=-\theta\sqrt{1-\theta}Z\sim \mathrm{N}\left(0,\theta^2(1-\theta)\right) \tag{10.43}$$

By (10.43) and Slutsky's Theorem

$$\frac{1}{\sqrt{\theta^2(1-\theta)}}\sqrt{n}\,(V_n-\theta)\rightarrow_D \frac{1}{\sqrt{\theta^2(1-\theta)}}\left(-\theta\sqrt{1-\theta}Z\right)=-Z\sim \mathrm{N}(0,1)$$

or

$$\frac{\sqrt{n}\,(V_n-\theta)}{\sqrt{\theta^2(1-\theta)}}\rightarrow_D Z\sim \mathrm{N}(0,1) \tag{10.44}$$

since if $-Z\sim \mathrm{N}(0,1)$ then $Z\sim \mathrm{N}(0,1)$ by symmetry of the $\mathrm{N}(0,1)$ distribution.
Since

$$Z_n=\frac{\sqrt{n}\,(V_n-\theta)}{\sqrt{V_n^2(1-V_n)}}=\frac{\frac{\sqrt{n}(V_n-\theta)}{\sqrt{\theta^2(1-\theta)}}}{\sqrt{\frac{V_n^2(1-V_n)}{\theta^2(1-\theta)}}} \tag{10.45}$$

then by (10.41) and the Limit Theorems

$$\sqrt{\frac{V_n^2(1-V_n)}{\theta^2(1-\theta)}}\rightarrow_p \sqrt{\frac{\theta^2(1-\theta)}{\theta^2(1-\theta)}}=1 \tag{10.46}$$

By (10.44), (10.45), (10.46), and Slutsky's Theorem

$$Z_n\rightarrow_D \frac{Z}{1}=Z\sim \mathrm{N}(0,1)$$

9. $X_1, X_2, \ldots, X_n$ are independent Gamma$(2, \theta)$ random variables with

$$E(X_i) = 2\theta \text{ and } Var(X_i) = 2\theta^2 \quad \text{for } i = 1, 2, \ldots, n$$

(a) Since $X_1, X_2, \ldots$ are independent and identically distributed random variables with $E(X_i) = 2\theta$ and $Var(X_i) = 2\theta^2 < \infty$ then by the Weak Law of Large Numbers

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \rightarrow_p 2\theta \tag{10.47}$$

By (10.47) and the Limit Theorems

$$\frac{\bar{X}_n}{\sqrt{2}} \rightarrow_p \frac{2\theta}{\sqrt{2}} = \sqrt{2}\theta$$

and

$$\frac{\sqrt{2}\theta}{\bar{X}_n/\sqrt{2}} \rightarrow_p 1 \tag{10.48}$$

(b) Since $X_1, X_2, \ldots$ are independent and identically distributed random variables with $E(X_i) = 2\theta$ and $Var(X_i) = 2\theta^2 < \infty$ then by the Central Limit Theorem

$$W_n = \frac{\sqrt{n}\left(\bar{X}_n - 2\theta\right)}{\sqrt{2\theta^2}} = \frac{\sqrt{n}\left(\bar{X}_n - 2\theta\right)}{\sqrt{2}\theta} \rightarrow_D Z \sim \mathrm{N}(0,1) \tag{10.49}$$

By (10.49) and Slutsky's Theorem

$$V_n = \sqrt{n}\left(\bar{X}_n - 2\theta\right) \rightarrow_D \sqrt{2}\theta Z \sim \mathrm{N}\left(0, 2\theta^2\right) \tag{10.50}$$

(c) By (10.48), (10.49) and Slutsky's Theorem

$$Z_n = \frac{\sqrt{n}\left(\bar{X}_n - 2\theta\right)}{\bar{X}_n/\sqrt{2}} = \left[\frac{\sqrt{n}\left(\bar{X}_n - 2\theta\right)}{\sqrt{2}\theta}\right]\left[\frac{\sqrt{2}\theta}{\bar{X}_n/\sqrt{2}}\right] \rightarrow_D Z(1) = Z \sim \mathrm{N}(0,1)$$

(d) To determine the limiting distribution of

$$U_n = \sqrt{n}\left[\log\left(\bar{X}_n\right) - \log(2\theta)\right]$$

let $g(x) = \log x$ and $a = 2\theta$. Then $g'(x) = 1/x$ and $g'(a) = g'(2\theta) = 1/(2\theta)$. By (10.50) and the Delta Method

$$U_n \rightarrow_D \frac{1}{2\theta}\sqrt{2}\theta Z = \frac{Z}{\sqrt{2}} \sim \mathrm{N}\left(0, \frac{1}{2}\right)$$

(e) The limiting variance of Z_n is 1 which does not depend on θ. The limiting variance of U_n is 1/2 which does not depend on θ. The transformation $g(x) = \log x$ is a variance-stabalizing transformation for the Gamma distribution.

10.5 Chapter 6

3. (a) If $X_i \sim$ Geometric(θ) then

$$f(x;\theta) = \theta(1-\theta)^x \quad \text{for } x = 0,1,\ldots;\ 0<\theta<1$$

The likelihood function is

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{n} f(x_i;\theta) \\ &= \prod_{i=1}^{n} \theta(1-\theta)^{x_i} \\ &= \theta^n (1-\theta)^t \quad \text{for } 0<\theta<1 \end{aligned}$$

where

$$t = \sum_{i=1}^{n} x_i$$

The log likelihood function is

$$l(\theta) = \log L(\theta) = n\log\theta + t\log(1-\theta) \quad \text{for } 0<\theta<1$$

The score function is

$$\begin{aligned} S(\theta) &= l'(\theta) = \frac{n}{\theta} - \frac{t}{1-\theta} \\ &= \frac{n-(n+t)\theta}{\theta(1-\theta)} \quad \text{for } 0<\theta<1 \end{aligned}$$

$$S(\theta) = 0 \ \text{ for } \ \theta = \frac{n}{n+t}$$

Since $S(\theta) > 0$ for $0 < \theta < n/(n+t)$ and $S(\theta) < 0$ for $n/(n+t) < \theta < 1$, therefore by the first derivative test the maximum likelihood estimate of θ is

$$\hat{\theta} = \frac{n}{n+t}$$

and the maximum likelihood estimator is

$$\hat{\theta} = \frac{n}{n+T} \quad \text{where } \ T = \sum_{i=1}^{n} X_i$$

The information function is

$$\begin{aligned} I(\theta) &= -S'(\theta) = -l''(\theta) \\ &= \frac{n}{\theta^2} + \frac{t}{(1-\theta)^2} \quad \text{for } 0<\theta<1 \end{aligned}$$

Since $I(\theta) > 0$ for all $0<\theta<1$, the graph of $l(\theta)$ is concave down and this also confirms that $\hat{\theta} = n/(n+t)$ is the maximum likelihood estimate.

3. (b) The observed information is

$$\begin{aligned} I(\hat{\theta}) &= \frac{n}{\hat{\theta}^2} + \frac{t}{(1-\hat{\theta})^2} = \frac{n}{\left(\frac{n}{n+t}\right)^2} + \frac{t}{(\frac{t}{n+t})^2} \\ &= \frac{(n+t)^2}{n} + \frac{(n+t)^2}{t} = \frac{(n+t)^3}{nt} \\ &= \frac{n}{\hat{\theta}^2(1-\hat{\theta})} \end{aligned}$$

The expected information is

$$\begin{aligned} E\left[\frac{n}{\theta^2} + \frac{T}{(1-\theta)^2}\right] &= \frac{n}{\theta^2} + \frac{E(T)}{(1-\theta)^2} \\ &= \frac{n}{\theta^2} + \frac{n(1-\theta)/\theta}{(1-\theta)^2} \\ &= n\left[\frac{(1-\theta)+\theta}{\theta^2(1-\theta)}\right] \\ &= \frac{n}{\theta^2(1-\theta)} \quad \text{for } 0 < \theta < 1 \end{aligned}$$

3. (c) Since

$$\tau = E(X_i) = \frac{(1-\theta)}{\theta}$$

then by the invariance property of maximum likelihood estimators the maximum likelihood estimator of $\tau = E(X_i)$ is

$$\hat{\tau} = \frac{1-\hat{\theta}}{\hat{\theta}} = \frac{T}{n} = \bar{X}$$

3. (d) If $n = 20$ and $t = \sum\limits_{i=1}^{20} x_i = 40$ then the maximum likelihood estimate of θ is

$$\hat{\theta} = \frac{20}{20+40} = \frac{1}{3}$$

The relative likelihood function of θ is given by

$$R(\theta) = \frac{L(\theta)}{L(\hat{\theta})} = \frac{\theta^{20}(1-\theta)^{40}}{\left(\frac{1}{3}\right)^{20}\left(\frac{2}{3}\right)^{40}} \quad \text{for } 0 \le \theta \le 1$$

A graph of $R(\theta)$ is given in Figure 10.25A 15% likelihood interval is found by solving $R(\theta) = 0.15$. The 15% likelihood interval is $[0.2234,\ 0.4570]$.

$R(0.5) = 0.03344$ implies that $\theta = 0.5$ is outside a 10% likelihood interval and we would conclude that $\theta = 0.5$ is not a very plausible value of θ given the data.

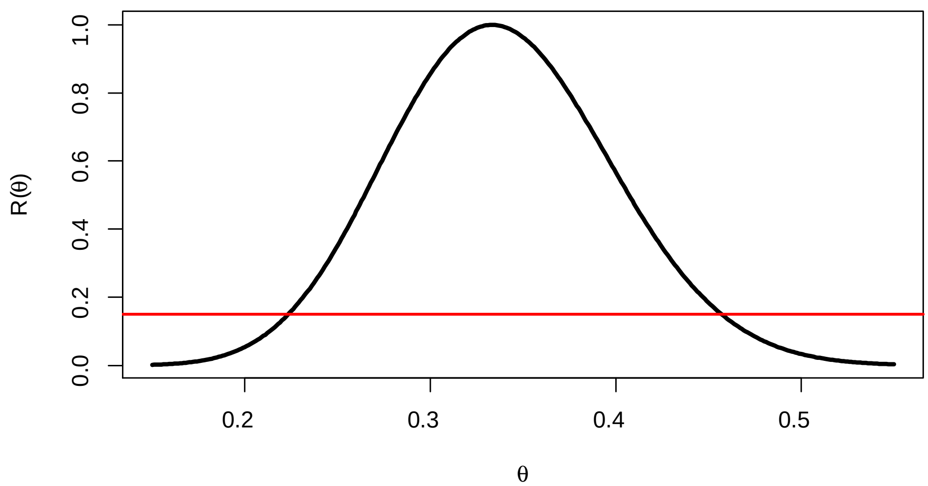


Figure 10.25: Relative likelihood function for Problem 3

4. Since $(X_1, X_2, X_3) \sim \text{Multinomial}\left(n, \theta^2, 2\theta(1-\theta), (1-\theta)^2\right)$ the likelihood function is

$$\begin{aligned} L(\theta_1, \theta_2) &= \frac{n!}{x_1! x_2! (n - x_1 - x_2)!} \left(\theta^2\right)^{x_1} \left[2\theta(1-\theta)\right]^{x_2} \left[(1-\theta)^2\right]^{n-x_1-x_2} \\ &= \frac{n!}{x_1! x_2! (n - x_1 - x_2)!} 2^{x_2} \theta^{2x_1+x_2} (1-\theta)^{2n-2x_1-x_2} \quad \text{for } 0 < \theta < 1 \end{aligned}$$

The log likelihood function is

$$\begin{aligned} l(\theta) &= \log\left(\frac{n!}{x_1! x_2! (n - x_1 - x_2)!}\right) + x_2 \log 2 \\ &\quad + (2x_1 + x_2) \log\theta + (2n - 2x_1 - x_2) \log(1-\theta) \quad \text{for } 0 < \theta < 1 \end{aligned}$$

The score function is

$$\begin{aligned} S(\theta) &= \frac{2x_1 + x_2}{\theta} - \frac{2n - (2x_1 + x_2)}{1-\theta} \\ &= \frac{(2x_1 + x_2)(1-\theta) - [2n - (2x_1 + x_2)]\theta}{\theta(1-\theta)} \\ &= \frac{(2x_1 + x_2) - (2x_1 + x_2)\theta - 2n + (2x_1 + x_2)\theta}{\theta(1-\theta)} \\ &= \frac{(2x_1 + x_2) - 2n}{\theta(1-\theta)} \quad \text{for } 0 < \theta < 1 \end{aligned}$$

$$S(\theta) = 0 \quad \text{if} \quad \theta = \frac{2x_1 + x_2}{2n}$$

Since

$$S(\theta) > 0 \text{ for } 0 < \theta < \frac{2x_1 + x_2}{2n}$$

and

$$S(\theta) < 0 \text{ for } 1 > \theta > \frac{2x_1 + x_2}{2n}$$

therefore by the first derivative test, $l(\theta)$ has an absolute maximum at $\theta = (2x_1 + x_2) / (2n)$. Thus the maximum likelihood estimate of θ is

$$\hat{\theta} = \frac{2x_1 + x_2}{2n}$$

and the maximum likelihood estimator of θ is

$$\hat{\theta} = \frac{2X_1 + X_2}{2n}$$

The information function is

$$I(\theta) = \frac{2x_1 + x_2}{\theta^2} + \frac{2n - (2x_1 + x_2)}{(1-\theta)^2} \quad \text{for } 0 < \theta < 1$$

and the observed information is

$$\begin{aligned} I(\hat{\theta}) &= I\left(\frac{2x_1 + x_2}{2n}\right) = \frac{(2n)^2 (2x_1 + x_2)}{(2x_1 + x_2)^2} + \frac{(2n)^2 [2n - (2x_1 + x_2)]}{[2n - (2x_1 + x_2)]^2} \\ &= \frac{(2n)^2}{(2x_1 + x_2)} + \frac{(2n)^2}{[2n - (2x_1 + x_2)]} = \frac{2n}{\left(\frac{2x_1+x_2}{2n}\right)\left(1 - \frac{2x_1+x_2}{2n}\right)} \\ &= \frac{2n}{\hat{\theta}\left(1 - \hat{\theta}\right)} \end{aligned}$$

Since

$$X_1 \sim \text{Binomial}\left(n, \theta^2\right) \quad \text{and} \quad X_2 \sim \text{Binomial}\left(n, 2\theta(1-\theta)\right)$$

$$E(2X_1 + X_2) = 2n\theta^2 + n[2\theta(1-\theta)] = 2n\theta$$

The expected information is

$$\begin{aligned} J(\theta) &= E\left[\frac{2X_1 + X_2}{\theta^2} + \frac{2n - (2X_1 + X_2)}{(1-\theta)^2}\right] \\ &= \frac{2n\theta}{\theta^2} + \frac{2n(1-\theta)}{(1-\theta)^2} = 2n\left(\frac{1}{\theta} + \frac{1}{1-\theta}\right) \\ &= \frac{2n}{\theta(1-\theta)} \quad \text{for } 0 < \theta < 1 \end{aligned}$$

6. (a) Given

$$\begin{aligned} P(k \text{ children in family; } \theta) &= \theta^k \quad \text{for } k = 1, 2, \ldots \\ P(0 \text{ children in family; } \theta) &= \frac{1 - 2\theta}{1 - \theta} \quad \text{for } 0 < \theta < \frac{1}{2} \end{aligned}$$

and the observed data

No. of children	0	1	2	3	4	Total
Frequency observed	17	22	7	3	1	50

the appropriate likelihood function for θ is based on the Multinomial model:

$$\begin{aligned} L(\theta) &= \frac{50!}{17!22!7!3!1!}\left(\frac{1-2\theta}{1-\theta}\right)^{17}\theta^{22}\left(\theta^2\right)^7\left(\theta^3\right)^3\left(\theta^4\right)^1\left(\theta^5+\theta^6+\cdots\right)^0 \\ &= \frac{50!}{17!22!7!3!1!}\left(\frac{1-2\theta}{1-\theta}\right)^{17}\theta^{49} \quad \text{for } 0<\theta<\frac{1}{2} \end{aligned}$$

or more simply

$$L(\theta) = \left(\frac{1-2\theta}{1-\theta}\right)^{17}\theta^{49} \quad \text{for } 0<\theta<\frac{1}{2}$$

The log likelihood function is

$$l(\theta) = 17\log(1-2\theta) - 17\log(1-\theta) + 49\log\theta \quad \text{for } 0<\theta<\frac{1}{2}$$

The score function is

$$\begin{aligned} S(\theta) &= \frac{-34}{1-2\theta} + \frac{17}{1-\theta} + \frac{49}{\theta} \\ &= \frac{-34\left(\theta-\theta^2\right)+17\left(\theta-2\theta^2\right)+49\left(1-3\theta+2\theta^2\right)}{\theta(1-\theta)(1-2\theta)} \\ &= \frac{98\theta^2-164\theta+49}{\theta(1-\theta)(1-2\theta)} \quad \text{for } 0<\theta<\frac{1}{2} \end{aligned}$$

The information function is

$$I(\theta) = \frac{68}{(1-2\theta)^2} - \frac{17}{(1-\theta)^2} + \frac{49}{\theta^2} \quad \text{for } 0<\theta<\frac{1}{2}$$

6. (b) $S(\theta)=0$ if

$$98\theta^2-164\theta+49=0 \quad \text{or} \quad \theta^2-\frac{82}{49}\theta+\frac{1}{2}=0$$

Therefore $S(\theta)=0$ if

$$\begin{aligned} \theta &= \frac{\frac{82}{49}\pm\sqrt{\left(\frac{82}{49}\right)^2-4\left(\frac{1}{2}\right)}}{2} = \frac{41}{49}\pm\frac{1}{2}\sqrt{\left(\frac{82}{49}\right)^2-2} = \frac{41}{49}\pm\frac{1}{2}\sqrt{\frac{(82)^2-2(49)^2}{(49)^2}} \\ &= \frac{41}{49}\pm\frac{1}{98}\sqrt{6724-4802} = \frac{41}{49}\pm\frac{1}{98}\sqrt{1922} \end{aligned}$$

Since $0<\theta<\frac{1}{2}$ and $\theta=\frac{41}{49}+\frac{1}{98}\sqrt{1922}>1$, we choose

$$\theta = \frac{41}{49}-\frac{1}{98}\sqrt{1922}$$

Since

$$S(\theta) > 0 \text{ for } 0 < \theta < \theta = \frac{41}{49} - \frac{1}{98}\sqrt{1922} \text{ and } S(\theta) < 0 \text{ for } \theta = \frac{41}{49} - \frac{1}{98}\sqrt{1922} < \theta < \frac{1}{2}$$

therefore the maximum likelihood estimate of θ is

$$\hat{\theta} = \theta = \frac{41}{49} - \frac{1}{98}\sqrt{1922} \approx 0.389381424147286 \approx 0.3894$$

The observed information for the given data is

$$I(\hat{\theta}) = \frac{68}{(1-2\hat{\theta})^2} - \frac{17}{(1-\hat{\theta})^2} + \frac{49}{\hat{\theta}^2} \approx 1666.88$$

6. (c) A graph of the relative likelihood function is given in Figure 10.26.A 15% likelihood

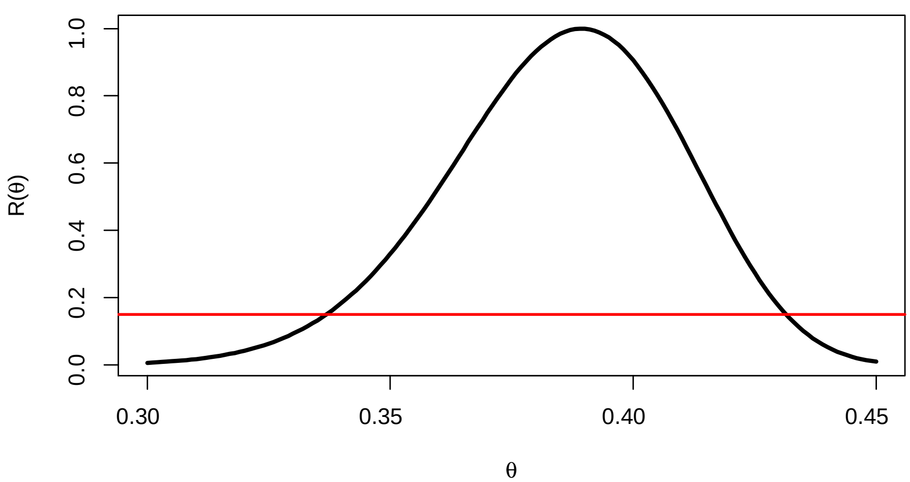


Figure 10.26: Relative likelihood function for Problem 6

interval for θ is $[0.34, 0.43]$.

6. (d) Since $R(0.45) \approx 0.0097$, $\theta = 0.45$ is outside a 1% likelihood interval and therefore $\theta = 0.45$ is not a plausible value of θ for these data.

6. (e) The expected frequencies are calculated using

$$e_0 = 50\left(\frac{1-2\hat{\theta}}{1-\hat{\theta}}\right) \quad \text{and} \quad e_k = 50\hat{\theta}^k \quad \text{for } k = 1, 2, \ldots$$

The observed and expected frequencies are

No. of children	0	1	2	3	4	Total
Observed Frequency: f_k	17	22	7	3	1	50
Expected Frequency: e_k	18.12	19.47	7.58	2.95	1.15	50

We see that the agreement between observed and expected frequencies is very good and the model gives a reasonable fit to the data.

7. Show $\hat{\theta} = x_{(1)}$ is the maximum likelihood estimate. By Example 2.5.3(a) θ is a location parameter for this distribution. By Theorem 6.6.4 $Q(\mathbf{X};\theta) = \tilde{\theta} - \theta = X_{(1)} - \theta$ is a pivotal quantity.

$$\begin{aligned}
P(Q(\mathbf{X};\theta) \le q) &= P\left(\tilde{\theta} - \theta \le q\right) \\
&= P\left(X_{(1)} - \theta \le q\right) \\
&= 1 - P\left(X_{(1)} \ge q + \theta\right) \\
&= 1 - \prod_{i=1}^{n} e^{-(q+\theta-\theta)} \quad \text{since } P(X_i > x) = e^{-(x-\theta)} \quad \text{for } x > \theta \\
&= 1 - e^{-nq} \quad \text{for } q \ge 0
\end{aligned}$$

Since

$$\begin{aligned}
& P\left(\tilde{\theta} + \frac{1}{n}\log(1-p) \le \theta \le \tilde{\theta}\right) \\
= \; & P\left(0 \le \tilde{\theta} - \theta \le -\frac{1}{n}\log(1-p)\right) \\
= \; & P\left(0 \le Q(\mathbf{X};\theta) \le -\frac{1}{n}\log(1-p)\right) \\
= \; & P\left(Q(\mathbf{X};\theta) \le -\frac{1}{n}\log(1-p)\right) - P(Q(\mathbf{X};\theta) \le 0) \\
= \; & 1 - e^{-\log(1-p)} - 0 \\
= \; & 1 - (1-p) = p
\end{aligned}$$

$\left[\hat{\theta} + \frac{1}{n}\log(1-p), \hat{\theta}\right]$ is a $100p\%$ confidence interval for θ.

Since

$$\begin{aligned}
& P\left[\tilde{\theta} + \frac{1}{n}\log\left(\frac{1-p}{2}\right) \le \theta \le \tilde{\theta} + \frac{1}{n}\log\left(\frac{1+p}{2}\right)\right] \\
= \; & P\left[-\frac{1}{n}\log\left(\frac{1+p}{2}\right) \le \tilde{\theta} - \theta \le -\frac{1}{n}\log\left(\frac{1-p}{2}\right)\right] \\
= \; & P\left[-\frac{1}{n}\log\left(\frac{1+p}{2}\right) \le Q(\mathbf{X};\theta) \le -\frac{1}{n}\log\left(\frac{1-p}{2}\right)\right] \\
= \; & 1 - e^{\log\left(\frac{1-p}{2}\right)} - \left(1 - e^{\log\left(\frac{1+p}{2}\right)}\right) \\
= \; & -\frac{1}{2} + \frac{p}{2} + \frac{1}{2} + \frac{p}{2} = p
\end{aligned}$$

$\left[\hat{\theta}+\frac{1}{n}\log\left(\frac{1-p}{2}\right),\hat{\theta}+\frac{1}{n}\log\left(\frac{1+p}{2}\right)\right]$ is a $100p\%$ confidence interval for θ.

The interval $\left[\hat{\theta}+\frac{1}{n}\log(1-p),\hat{\theta}\right]$ is a better choice since it contains $\hat{\theta}$ while the interval $\left[\hat{\theta}+\frac{1}{n}\log\left(\frac{1-p}{2}\right),\hat{\theta}+\frac{1}{n}\log\left(\frac{1+p}{2}\right)\right]$ does not.

8.(*a*) If $x_1, x_2, \ldots, x_n$ is an observed random sample from the Gamma$\left(\frac{1}{2},\frac{1}{\theta}\right)$ distribution then the likelihood function is

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{n} f(x_i;\theta) \\ &= \prod_{i=1}^{n} \frac{\theta^{1/2}x_i^{-1/2}e^{-\theta x_i}}{\Gamma\left(\frac{1}{2}\right)} \\ &= \left(\prod_{i=1}^{n} x_i\right)^{-1/2}\left[\Gamma\left(\frac{1}{2}\right)\right]^{n}\theta^{n/2}e^{-\theta t} \quad \text{for } \theta>0 \end{aligned}$$

where

$$t=\sum_{i=1}^{n} x_i$$

or more simply

$$L(\theta)=\theta^{n/2}e^{-\theta t} \quad \text{for } \theta>0$$

The log likelihood function is

$$\begin{aligned} l(\theta) &= \log L(\theta) \\ &= \frac{n}{2}\log\theta-\theta t \quad \text{for } \theta>0 \end{aligned}$$

and the score function is

$$S(\theta)=\frac{d}{d\theta}l(\theta)=\frac{n}{2\theta}-t=\frac{n-2\theta t}{2\theta} \quad \text{for } \theta>0$$

$$S(\theta)=0 \ \text{ for } \ \theta=\frac{n}{2t}$$

Since

$$S(\theta)>0 \ \text{ for } \ 0<\theta<\frac{n}{2t}$$

and

$$S(\theta)<0 \ \text{ for } \ \theta>\frac{n}{2t}$$

therefore by the first derivative test $l(\theta)$ has a absolute maximum at $\theta=\frac{n}{2t}$. Thus

$$\hat{\theta}=\frac{n}{2t}=\frac{1}{2\bar{x}}$$

is the maximum likelihood estimate of θ and

$$\tilde{\theta}=\frac{1}{2\bar{X}}$$

is the maximum likelihood estimator of θ.

8.(b) If $X_i \sim \text{Gamma}\left(\frac{1}{2}, \frac{1}{\theta}\right)$, $i = 1, 2, \ldots, n$ independently then

$$E(X_i) = \frac{1}{2\theta} \quad \text{and} \quad Var(X_i) = \frac{1}{2\theta^2}$$

and by the Weak Law of Large Numbers

$$\bar{X} \to_p \frac{1}{2\theta}$$

and by the Limit Theorems

$$\tilde{\theta} = \frac{1}{2\bar{X}} \to_p \frac{1}{2\left(\frac{1}{2\theta}\right)} = \theta$$

as required.

8.(c) By the Invariance Property of Maximum Likelihood Estimates the maximum likelihood estimate of

$$\tau = Var(X_i) = \frac{1}{2\theta^2}$$

is

$$\hat{\tau} = \frac{1}{2\hat{\theta}^2} = \frac{1}{2\left(\frac{n}{2t}\right)^2} = \frac{4t^2}{2n^2} = 2\left(\frac{t}{n}\right)^2 = 2\bar{x}^2$$

8.(d) The moment generating function of X_i is

$$M(t) = \frac{1}{\left(1 - \frac{t}{\theta}\right)^{1/2}} \quad \text{for } t < \theta$$

The moment generating function of $Q = 2\theta \sum\limits_{i=1}^{n} X_i$ is

$$\begin{aligned}
M_Q(t) &= E\left(e^{tQ}\right) = E\left[\exp\left(2t\theta \sum_{i=1}^{n} X_i\right)\right] \\
&= \prod_{i=1}^{n} E\left[\exp\left(2tX_i\right)\right] = \prod_{i=1}^{n} M(2t\theta) \\
&= \left(\frac{1}{1 - \frac{2t\theta}{\theta}}\right)^{n/2} \quad \text{for } 2t\theta < \theta \\
&= \frac{1}{(1-2t)^{n/2}} \quad \text{for } t < \frac{1}{2}
\end{aligned}$$

which is the moment generating function of a $\chi^2(n)$ random variable. Therefore by the Uniqueness Theorem for Moment Generating Functions, $Q \sim \chi^2(n)$.

To construct a 95% equal tail confidence interval for θ we find a and b such that $P(Q \le a) = 0.025 = P(Q > b)$ so that

$$P(a < Q < b) = P(a < 2T\theta < b) = 0.95$$

or

$$P\left(\frac{a}{2T} < \theta < \frac{b}{2T}\right) = 0.95$$

so that $\left(\frac{a}{2t}, \frac{b}{2t}\right)$ is a 95% equal tail confidence interval for θ.

For $n = 20$ we have

$$P(Q \le 9.59) = 0.025 = P(Q > 34.17)$$

For $t = \sum\limits_{i=1}^{20} x_i = 6$ a 95% equal tail confidence interval for θ is

$$\left[\frac{9.59}{2(6)}, \frac{34.17}{2(6)}\right] = [0.80,\ 2.85]$$

Since $\theta = 0.7$ is not in the 95% confidence interval it is not a plausible value of θ in light of the data.

8.(e) The information function is

$$I(\theta) = -\frac{d}{d\theta} S(\theta) = \frac{n}{2\theta^2} \quad \text{for } \theta > 0$$

The expected information is

$$J(\theta) = E[I(\theta; X_1, \ldots, X_n)] = E\left(\frac{n}{2\theta^2}\right) = \frac{n}{2\theta^2} \quad \text{for } \theta > 0$$

and

$$\left[J(\tilde{\theta})\right]^{1/2}\left(\tilde{\theta} - \theta\right) = \frac{\sqrt{n}}{\sqrt{2}\tilde{\theta}}\left(\tilde{\theta} - \theta\right) = \frac{\sqrt{n}}{\sqrt{2}\tilde{\theta}}\left(\frac{1}{2\bar{X}} - \theta\right)$$

By the Central Limit Theorem

$$\frac{\sqrt{n}\left(\bar{X} - \frac{1}{2\theta}\right)}{\frac{1}{\sqrt{2}\theta}} = \sqrt{2n}\theta\left(\bar{X} - \frac{1}{2\theta}\right) \to_D Z \sim \mathrm{N}(0,1) \tag{10.51}$$

Let

$$g(x) = \frac{1}{2x} \quad \text{and} \quad a = \frac{1}{2\theta}$$

then

$$g'(x) = \frac{-1}{2x^2}$$

$$g(a) = g\left(\frac{1}{2\theta}\right) = \frac{1}{2\left(\frac{1}{2\theta}\right)} = \theta$$

and

$$g'(a) = g'\left(\frac{1}{2\theta}\right) = \frac{-1}{2\left(\frac{1}{2\theta}\right)^2} = -2\theta^2$$

By the Delta Theorem and (10.51)

$$\sqrt{2n}\theta\left(\frac{1}{2\bar{X}} - \theta\right) \rightarrow_D -2\theta^2 Z \sim \mathrm{N}\left(0, 4\theta^4\right)$$

or

$$\sqrt{2n}\theta\left(\tilde{\theta} - \theta\right) \rightarrow_D -2\theta^2 Z \sim \mathrm{N}\left(0, 4\theta^4\right)$$

By Slutsky's Theorem

$$\frac{\sqrt{2n}\theta}{2\theta^2}\left(\tilde{\theta} - \theta\right) \rightarrow_D -Z \sim \mathrm{N}(0,1)$$

But if $-Z \sim \mathrm{N}(0,1)$ then $Z \sim \mathrm{N}(0,1)$ and thus

$$\frac{\sqrt{n}}{\sqrt{2}\theta}\left(\tilde{\theta} - \theta\right) \rightarrow_D Z \sim \mathrm{N}(0,1) \tag{10.52}$$

Since $\tilde{\theta} \rightarrow_p \theta$ then by the Limit Theorems

$$\frac{\tilde{\theta}}{\theta} \rightarrow_p 1 \tag{10.53}$$

Thus by (10.52), (10.53) and Slutsky's Theorem

$$\left[J(\tilde{\theta})\right]^{1/2}\left(\tilde{\theta} - \theta\right) = \frac{\sqrt{n}\left(\tilde{\theta} - \theta\right)}{\sqrt{2}\tilde{\theta}}$$

$$= \frac{\frac{\sqrt{n}}{\sqrt{2}\theta}\left(\tilde{\theta} - \theta\right)}{\frac{\tilde{\theta}}{\theta}} \rightarrow_D \frac{Z}{1} = Z \sim \mathrm{N}(0,1)$$

An approximate 95% confidence interval is given by

$$\left[\hat{\theta} - \frac{1.96}{\sqrt{J\left(\hat{\theta}\right)}},\quad \hat{\theta} + \frac{1.96}{\sqrt{J\left(\hat{\theta}\right)}}\right]$$

For $n = 20$ and $t = \sum_{i=1}^{20} x_i = 6$,

$$\hat{\theta} = \frac{1}{2\left(\frac{6}{20}\right)} = \frac{5}{3} \quad \text{and} \quad J\left(\hat{\theta}\right) = \frac{n}{2\hat{\theta}^2} = \frac{20}{2\left(\frac{5}{3}\right)^2} = 3.6$$

and an approximate 95% confidence interval is given by

$$\left[\frac{5}{3} - \frac{1.96}{\sqrt{3.6}},\quad \frac{5}{3} + \frac{1.96}{\sqrt{3.6}}\right] = [0.63,\ 2.70]$$

For $n = 20$ and $t = \sum_{i=1}^{20} x_i = 6$ the relative likelihood function of θ is given by

$$R(\theta) = \frac{L(\theta)}{L(\hat{\theta})} = \frac{\theta^{10} e^{-6\theta}}{\left(\frac{5}{3}\right)^{10} e^{-10}} = \left(\frac{3\theta}{5}\right)^{10} e^{10-6\theta} \quad \text{for } \theta > 0$$

A graph of $R(\theta)$ is given in Figure 10.27.A 15% likelihood interval is found by solving

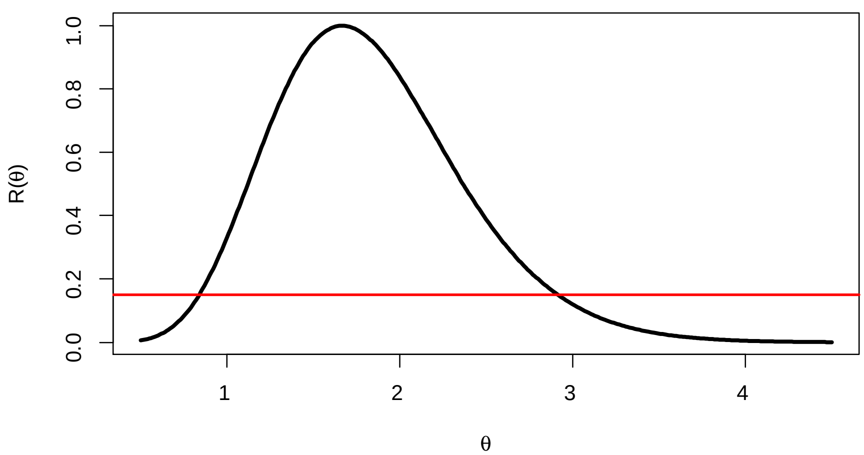


Figure 10.27: Relative likelihood function for Problem 8

$R(\theta) = 0.15$. The 15% likelihood interval is $[0.84,\ 2.91]$.

The exact 95% equal tail confidence interval $[0.80,\ 2.85]$, the approximate 95% confidence interval $[0.63,\ 2.70]$, and the 15% likelihood interval $[0.84,\ 2.91]$ are all approximately of the same width. The exact confidence interval and likelihood interval are skewed to the right while the approximate confidence interval is symmetric about the maximum likelihood estimate $\hat{\theta} = 5/3$. Approximate confidence intervals are symmetric about the maximum likelihood estimate because they are based on a Normal approximation. Since $n = 20$ the approximation cannot be completely trusted. Therefore for these data the exact confidence interval and the likelihood interval are both better interval estimates for θ.

$R(0.7) = 0.056$ implies that $\theta = 0.7$ is outside a 10% likelihood interval so based on the likelihood function we would conclude that $\theta = 0.7$ is not a very plausible value of θ given the data. Previously we noted that $\theta = 0.7$ is also not contained in the exact 95% confidence interval. Note however that $\theta = 0.7$ is contained in the approximate 95% confidence interval and so based on the approximate confidence interval we would conclude that $\theta = 0.7$ is a reasonable value of θ given the data. Again the reason for the disagreement is because $n = 20$ is not large enough for the approximation to be a good one.

10.6 Chapter 7

1. Since X_i has cumulative distribution function

$$F(x;\theta_1,\theta_2) = 1 - \left(\frac{\theta_1}{x}\right)^{\theta_2} \quad \text{for } x \ge \theta_1,\ \theta_1 > 0,\ \theta_2 > 0$$

the probability density function of X_i is

$$\begin{aligned} f(x;\theta_1,\theta_2) &= \frac{d}{dx} F(x;\theta_1,\theta_2) \\ &= \frac{\theta_2}{x}\left(\frac{\theta_1}{x}\right)^{\theta_2} \quad \text{for } x \ge \theta_1,\ \theta_1 > 0,\ \theta_2 > 0 \end{aligned}$$

The likelihood function is

$$\begin{aligned} L(\theta_1,\theta_2) &= \prod_{i=1}^{n} f(x_i;\theta_1,\theta_2) \\ &= \prod_{i=1}^{n} \frac{\theta_2}{x_i}\left(\frac{\theta_1}{x_i}\right)^{\theta_2} \quad \text{if } 0 < \theta_1 \le x_i; \quad i = 1,2,\ldots,n \text{ and } \theta_2 > 0 \\ &= \theta_2^n \theta_1^{n\theta_2}\left(\prod_{i=1}^{n} x_i\right)^{-\theta_2 - 1} \quad \text{if } 0 < \theta_1 \le x_{(1)} \text{ and } \theta_2 > 0 \end{aligned}$$

For each value of θ_2 the likelihood function is maximized over θ_1 by taking θ_1 to be as large as possible subject to $0 < \theta_1 \le x_{(1)}$. Therefore for fixed θ_2 the likelihood is maximized for $\theta_1 = x_{(1)}$. Since this is true for all values of θ_2 the value of (θ_1, θ_2) which maximizes $L(\theta_1,\theta_2)$ will necessarily have $\theta_1 = x_{(1)}$.

To find the value of θ_2 which maximizes $L\left(x_{(1)},\theta_2\right)$ consider the function

$$L_2(\theta_2) = L\left(x_{(1)},\theta_2\right) = \theta_2^n x_{(1)}^{n\theta_2}\left(\prod_{i=1}^{n} x_i\right)^{-\theta_2 - 1} \quad \text{for } \theta_2 > 0$$

and its logarithm

$$l_2(\theta_2) = \log L_2(\theta_2) = n\log\theta_2 + n\theta_2 \log x_{(1)} - (\theta_2 + 1)\sum_{i=1}^{n} \log x_i$$

Now

$$\begin{aligned} \frac{d}{d\theta_2} l_2(\theta_2) &= l_2'(\theta_2) = \frac{n}{\theta_2} + n\log x_{(1)} - \sum_{i=1}^{n} \log x_i \\ &= \frac{n}{\theta_2} - \sum_{i=1}^{n} \log\left(\frac{x_i}{x_{(1)}}\right) \\ &= \frac{n}{\theta_2} - t \\ &= \frac{n - \theta_2 t}{\theta_2} \end{aligned}$$

where

$$t=\sum_{i=1}^{n}\log\left(\frac{x_i}{x_{(1)}}\right)$$

Now $l_2'(\theta_2)=0$ for $\theta_2=n/t$. Since $l_2'(\theta_2)>0$ for $0<\theta_2<n/t$ and $l_2'(\theta_2)<0$ for $\theta_2>n/t$ therefore by the first derivative test $l_2(\theta_2)$ is maximized for $\theta_2=n/t=\hat{\theta}_2$. Therefore $L_2(\theta_2)=L\left(x_{(1)},\theta_2\right)$ is also maximized for $\theta_2=\hat{\theta}_2$. Therefore the maximum likelihood estimates are

$$\hat{\theta}_1=x_{(1)}\quad\text{and}\quad\hat{\theta}_2=\frac{n}{\sum\limits_{i=1}^{n}\log\left(\frac{x_i}{x_{(1)}}\right)}$$

and the maximum likelihood estimators are

$$\tilde{\theta}_1=X_{(1)}\quad\text{and}\quad\tilde{\theta}_2=\frac{n}{\sum\limits_{i=1}^{n}\log\left(\frac{X_i}{X_{(1)}}\right)}$$

4. (a) If the events S and H are independent events then $P(S\cap H)=P(S)P(H)=\alpha\beta$, $P(S\cap\bar{H})=P(S)P(\bar{H})=\alpha(1-\beta)$, etc.

The likelihood function is

$$L(\alpha,\beta)=\frac{n!}{x_{11}!x_{12}!x_{21}!x_{22}!}(\alpha\beta)^{x_{11}}\left[\alpha(1-\beta)\right]^{x_{12}}\left[(1-\alpha)\beta\right]^{x_{21}}\left[(1-\alpha)(1-\beta)\right]^{x_{22}}$$

or more simply (ignoring constants with respect to α and β)

$$L(\alpha,\beta)=\alpha^{x_{11}+x_{12}}(1-\alpha)^{x_{21}+x_{22}}\beta^{x_{11}+x_{21}}(1-\beta)^{x_{12}+x_{22}}\quad\text{for } 0\le\alpha\le1,\ 0\le\beta\le1$$

The log likelihood is

$$\begin{aligned} l(\alpha,\beta) &= (x_{11}+x_{12})\log\alpha+(x_{21}+x_{22})\log(1-\alpha)+(x_{11}+x_{21})\log\beta+(x_{12}+x_{22})\log(1-\beta)\\ \text{for } 0 &< \alpha<1,\ 0<\beta<1\end{aligned}$$

Since

$$\begin{aligned}\frac{\partial l}{\partial\alpha} &= \frac{x_{11}+x_{12}}{\alpha}-\frac{x_{21}+x_{22}}{1-\alpha}=\frac{x_{11}+x_{12}}{\alpha}-\frac{n-(x_{11}+x_{12})}{1-\alpha}=\frac{x_{11}+x_{12}-n\alpha}{\alpha(1-\alpha)}\\ \frac{\partial l}{\partial\beta} &= \frac{x_{11}+x_{21}}{\beta}-\frac{x_{12}+x_{22}}{1-\beta}=\frac{x_{11}+x_{21}}{\beta}-\frac{n-(x_{11}+x_{21})}{1-\beta}=\frac{x_{11}+x_{21}-n\beta}{\beta(1-\beta)}\end{aligned}$$

the score vector is

$$\begin{aligned} S(\alpha,\beta) &= \begin{bmatrix}\frac{x_{11}+x_{12}-n\alpha}{\alpha(1-\alpha)} & \frac{x_{11}+x_{21}-n\beta}{\beta(1-\beta)}\end{bmatrix}\\ \text{for } 0 &< \alpha<1,\ 0<\beta<1\end{aligned}$$

Solving $S(\alpha,\beta)=(0,0)$ gives the maximum likelihood estimates

$$\hat{\alpha}=\frac{x_{11}+x_{12}}{n}\quad\text{and}\quad\hat{\beta}=\frac{x_{11}+x_{21}}{n}$$

The information matrix is

$$I(\alpha,\beta) = \begin{bmatrix} -\frac{\partial^2 l}{\partial\alpha^2} & -\frac{\partial^2 l}{\partial\beta\partial\alpha} \\ -\frac{\partial^2 l}{\partial\beta\partial\alpha} & -\frac{\partial^2 l}{\partial\beta^2} \end{bmatrix} = \begin{bmatrix} \frac{x_{11}+x_{12}}{\alpha^2} + \frac{n-(x_{11}+x_{12})}{(1-\alpha)^2} & 0 \\ 0 & \frac{x_{11}+x_{21}}{\beta^2} + \frac{n-(x_{11}+x_{21})}{(1-\beta)^2} \end{bmatrix}$$

4.(b) Since $X_{11}+X_{12}=$ the number of times the event S is observed and $P(S)=\alpha$, then the distribution of $X_{11}+X_{12}$ is Binomial(n,α). Therefore $E(X_{11}+X_{12}) = n\alpha$ and

$$E\left(\frac{X_{11}+X_{12}}{\alpha^2} + \frac{n-(X_{11}+X_{12})}{(1-\alpha)^2}\right) = \frac{n}{\alpha(1-\alpha)}$$

Since $X_{11}+X_{21}=$ the number of times the event H is observed and $P(H)=\beta$, then the distribution of $X_{11}+X_{21}$ is Binomial(n,β). Therefore $E(X_{11}+X_{21}) = n\beta$ and

$$E\left(\frac{X_{11}+X_{21}}{\beta^2} + \frac{n-(X_{11}+X_{21})}{(1-\beta)^2}\right) = \frac{n}{\beta(1-\beta)}$$

Therefore the expected information matrix is

$$J(\alpha,\beta) = \begin{bmatrix} \frac{n}{\alpha(1-\alpha)} & 0 \\ 0 & \frac{n}{\beta(1-\beta)} \end{bmatrix}$$

The inverse matrix is

$$[J(\alpha,\beta)]^{-1} = \begin{bmatrix} \frac{\alpha(1-\alpha)}{n} & 0 \\ 0 & \frac{\beta(1-\beta)}{n} \end{bmatrix}$$

Also $Var(\tilde{\alpha}) = \frac{\alpha(1-\alpha)}{n}$ and $Var(\tilde{\beta}) = \frac{\beta(1-\beta)}{n}$ so the diagonal entries of $[J(\alpha,\beta)]^{-1}$ give us the variances of the maximum likelihood estimators.

7.(a) If $Y_i \sim$ Binomial$(1,p_i)$ where $p_i = \left(1+e^{-\alpha-\beta x_i}\right)^{-1}$ and $x_1,x_2,\ldots,x_n$ are known constants, the likelihood function for (α,β) is

$$L(\alpha,\beta) = \prod_{i=1}^{n} \binom{1}{y_i} p_i^{y_i}(1-p_i)^{1-y_i}$$

or more simply

$$L(\alpha,\beta) = \prod_{i=1}^{n} p_i^{y_i}(1-p_i)^{1-y_i}$$

The log likelihood function is

$$l(\alpha,\beta) = \log L(\alpha,\beta) = \sum_{i=1}^{n} \left[y_i \log(p_i) + (1-y_i)\log(1-p_i)\right]$$

Note that

$$
\begin{aligned}
\frac{\partial p_i}{\partial \alpha} &= \frac{\partial}{\partial \alpha}\left(1+e^{-\alpha-\beta x_i}\right)^{-1} = \frac{e^{-\alpha-\beta x_i}}{\left(1+e^{-\alpha-\beta x_i}\right)^2} \\
&= \frac{1}{\left(1+e^{-\alpha-\beta x_i}\right)}\frac{e^{-\alpha-\beta x_i}}{\left(1+e^{-\alpha-\beta x_i}\right)} = p_i\left(1-p_i\right)
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{\partial p_i}{\partial \beta} &= \frac{\partial}{\partial \beta}\left(1+e^{-\alpha-\beta x_i}\right)^{-1} = \frac{x_i e^{-\alpha-\beta x_i}}{\left(1+e^{-\alpha-\beta x_i}\right)^2} \\
&= x_i\frac{1}{\left(1+e^{-\alpha-\beta x_i}\right)}\frac{e^{-\alpha-\beta x_i}}{\left(1+e^{-\alpha-\beta x_i}\right)} = x_i p_i\left(1-p_i\right)
\end{aligned}
$$

Therefore

$$
\begin{aligned}
\frac{\partial l}{\partial \alpha} &= \frac{\partial l}{\partial p_i}\frac{\partial p_i}{\partial \alpha} = \sum_{i=1}^{n}\left[\frac{y_i}{p_i} - \frac{\left(1-y_i\right)}{\left(1-p_i\right)}\right]\frac{\partial p_i}{\partial \alpha} \\
&= \sum_{i=1}^{n}\left[\frac{y_i\left(1-p_i\right)-\left(1-y_i\right)p_i}{p_i\left(1-p_i\right)}\right]p_i\left(1-p_i\right) \\
&= \sum_{i=1}^{n}\left[y_i\left(1-p_i\right)-\left(1-y_i\right)p_i\right] \\
&= \sum_{i=1}^{n}\left(y_i-p_i\right)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial l}{\partial \beta} &= \frac{\partial l}{\partial p_i}\frac{\partial p_i}{\partial \beta} = \sum_{i=1}^{n}\left[\frac{y_i}{p_i} - \frac{\left(1-y_i\right)}{\left(1-p_i\right)}\right]\frac{\partial p_i}{\partial \beta} \\
&= \sum_{i=1}^{n}\left[\frac{y_i\left(1-p_i\right)-\left(1-y_i\right)p_i}{p_i\left(1-p_i\right)}\right]x_i p_i\left(1-p_i\right) \\
&= \sum_{i=1}^{n}x_i\left[y_i\left(1-p_i\right)-\left(1-y_i\right)p_i\right] \\
&= \sum_{i=1}^{n}x_i\left(y_i-p_i\right)
\end{aligned}
$$

The score vector is

$$
S\left(\alpha,\beta\right) = \begin{bmatrix} \frac{\partial l}{\partial \alpha} \\ \frac{\partial l}{\partial \beta} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n}\left(y_i-p_i\right) \\ \sum_{i=1}^{n}x_i\left(y_i-p_i\right) \end{bmatrix}
$$

To obtain the expected information we first note that

$$
\begin{aligned}
\frac{\partial^2 l}{\partial \alpha^2} &= \frac{\partial}{\partial \alpha}\left(\frac{\partial l}{\partial \alpha}\right) = \frac{\partial}{\partial \alpha}\left[\sum_{i=1}^{n}\left(y_i-p_i\right)\right] = -\sum_{i=1}^{n}\frac{\partial p_i}{\partial \alpha} = -\sum_{i=1}^{n}p_i\left(1-p_i\right) \\
\frac{\partial^2 l}{\partial \beta \partial \alpha} &= \frac{\partial}{\partial \beta}\left(\frac{\partial l}{\partial \alpha}\right) = \frac{\partial}{\partial \beta}\left[\sum_{i=1}^{n}\left(y_i-p_i\right)\right] = -\sum_{i=1}^{n}\frac{\partial p_i}{\partial \beta} = -\sum_{i=1}^{n}x_i p_i\left(1-p_i\right)
\end{aligned}
$$

and

$$\frac{\partial^2 l}{\partial \beta^2} = \frac{\partial}{\partial \beta}\left(\frac{\partial l}{\partial \beta}\right) = \frac{\partial}{\partial \beta}\left[\sum_{i=1}^{n} x_i\left(y_i - p_i\right)\right] = -\sum_{i=1}^{n} x_i \cdot \frac{\partial p_i}{\partial \beta} = -\sum_{i=1}^{n} x_i^2 p_i\left(1 - p_i\right)$$

The information matrix is

$$\begin{aligned} I\left(\alpha, \beta\right) &= \begin{bmatrix} -\frac{\partial^2 l}{\partial \alpha^2} & -\frac{\partial^2 l}{\partial \beta \partial \alpha} \\ -\frac{\partial^2 l}{\partial \beta \partial \alpha} & -\frac{\partial^2 l}{\partial \beta^2} \end{bmatrix} \\ &= \begin{bmatrix} \sum\limits_{i=1}^{n} p_i\left(1 - p_i\right) & \sum\limits_{i=1}^{n} x_i p_i\left(1 - p_i\right) \\ \sum\limits_{i=1}^{n} x_i p_i\left(1 - p_i\right) & \sum\limits_{i=1}^{n} x_i^2 p_i\left(1 - p_i\right) \end{bmatrix} \end{aligned}$$

which is a constant function of the random variables $Y_1, Y_2, ..., Y_n$ and therefore the expected information is $J\left(\alpha, \beta\right) = I\left(\alpha, \beta\right)$

5.(b) The maximum likelihood estimates of α and β are found by solving the equations

$$\begin{aligned} S\left(\alpha, \beta\right) &= \begin{bmatrix} \frac{\partial l}{\partial \alpha} & \frac{\partial l}{\partial \beta} \end{bmatrix} \\ &= \begin{bmatrix} \sum\limits_{i=1}^{n} \left(y_i - p_i\right) & \sum\limits_{i=1}^{n} x_i\left(y_i - p_i\right) \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 \end{bmatrix} \end{aligned}$$

which must be done numerically.

Newton's method is given by

$$\begin{bmatrix} \alpha^{(i+1)} & \beta^{(i+1)} \end{bmatrix} = \begin{bmatrix} \alpha^{(i)} & \beta^{(i)} \end{bmatrix} + S\left(\alpha^{(i)}, \beta^{(i)}\right)\left[I\left(\alpha^{(i)}, \beta^{(i)}\right)\right]^{-1} \qquad i = 0, 1, \ldots \text{ convergence}$$

where $\left(\alpha^{(0)}, \beta^{(0)}\right)$ is an initial estimate of (α, β).

10.7 Chapter 8

1.(a) The hypothesis $H_0 : \theta = \theta_0$ is a simple hypothesis since the model is completely specified.

From Example 6.3.6 the likelihood function is

$$L(\theta) = \theta^n \left(\prod_{i=1}^n x_i\right)^\theta \quad \text{for } \theta > 0$$

The log likelihood function is

$$l(\theta) = n \log \theta + \theta \sum_{i=1}^n \log x_i \quad \text{for } \theta > 0$$

and the maximum likelihood estimate is

$$\hat{\theta} = \frac{-n}{\sum\limits_{i=1}^n \log x_i}$$

The relative likelihood function is

$$R(\theta) = \frac{L(\theta)}{L(\hat{\theta})} = \left(\frac{\theta}{\tilde{\theta}}\right)^n \left(\prod_{i=1}^n x_i\right)^{\theta - \hat{\theta}} \quad \text{for } \theta \geq 0$$

The likelihood ratio test statistic for $H_0 : \theta = \theta_0$ is

$$\begin{aligned}
\Lambda(\theta_0; \mathbf{X}) &= -2 \log R(\theta_0; \mathbf{X}) \\
&= -2 \log \left[\left(\frac{\theta_0}{\tilde{\theta}}\right)^n \left(\prod_{i=1}^n X_i\right)^{\theta_0 - \tilde{\theta}}\right] \\
&= 2\left[-n \log\left(\frac{\theta_0}{\tilde{\theta}}\right) - \left(\theta_0 - \tilde{\theta}\right) \sum_{i=1}^n \log X_i\right] \\
&= 2\left[-n \log\left(\frac{\theta_0}{\tilde{\theta}}\right) + n\left(\theta_0 - \tilde{\theta}\right)\left(\frac{-1}{n} \sum_{i=1}^n \log X_i\right)\right] \\
&= 2n\left[-\log\left(\frac{\theta_0}{\tilde{\theta}}\right) + \left(\theta_0 - \tilde{\theta}\right)\frac{1}{\tilde{\theta}}\right] \quad \text{since } \frac{-\sum\limits_{i=1}^n \log X_i}{n} = \frac{1}{\tilde{\theta}} \\
&= 2n\left[\left(\frac{\theta_0}{\tilde{\theta}} - 1\right) - \log\left(\frac{\theta_0}{\tilde{\theta}}\right)\right]
\end{aligned}$$

The observed value of the likelihood ratio test statistic is

$$\begin{aligned}
\lambda(\theta_0; \mathbf{x}) &= -2 \log R(\theta_0; \mathbf{X}) \\
&= 2n\left[\left(\frac{\theta_0}{\hat{\theta}} - 1\right) - \log\left(\frac{\theta_0}{\hat{\theta}}\right)\right]
\end{aligned}$$

The parameter space is $\Omega = \{\theta : \theta > 0\}$ which has dimension 1 and thus $k = 1$. The approximate *p-value* is

$$\begin{aligned} p\text{-}value &\approx P\left(W \geq \lambda\left(\theta_0; \mathbf{x}\right)\right) \quad \text{where } W \sim \chi^2\left(1\right) \\ &= 2\left[1 - P\left(Z \leq \sqrt{\lambda\left(\theta_0; \mathbf{x}\right)}\right)\right] \quad \text{where } Z \sim \mathrm{N}\left(0,1\right) \end{aligned}$$

(b) If $n = 20$ and $\sum\limits_{i=1}^{20} \log x_i = -25$ and $H_0 : \theta = 1$ then $\hat{\theta} = 20/25 = 0.8$ the observed value of the likelihood ratio test statistic is

$$\begin{aligned} \lambda\left(\theta_0; \mathbf{x}\right) &= -2\log R\left(\theta_0; \mathbf{X}\right) \\ &= 2\left(20\right)\left[\left(\frac{1}{0.8} - 1\right) - \log\left(\frac{1}{0.8}\right)\right] \\ &= 40\left[0.25 - \log\left(1.25\right)\right] \\ &= 1.074258 \end{aligned}$$

and

$$\begin{aligned} p - value &\approx P\left(W \geq 1.074258\right) \quad \text{where } W \sim \chi^2\left(1\right) \\ &= 2\left[1 - P\left(Z \leq \sqrt{1.074258}\right)\right] \quad \text{where } Z \sim \mathrm{N}\left(0,1\right) \\ &= 0.2999857 \end{aligned}$$

calculated using R. Since $p - value > 0.1$ there is no evidence against $H_0 : \theta = 1$ based on the data.

4. Since $\Omega = \{(\theta_1, \theta_2) : \theta_1 > 0,\ \theta_2 > 0\}$ which has dimension $k = 2$ and
$\Omega_0 = \{(\theta_1, \theta_2) : \theta_1 = \theta_2,\ \theta_1 > 0,\ \theta_2 > 0\}$ which has dimension $q = 1$ and the hypothesis is composite.

From Example 6.5.2 the likelihood function for an observed random sample $x_1, x_2, \ldots, x_n$ from an Weibull$(2, \theta_1)$ distribution is

$$L_1(\theta_1) = \theta_1^{-2n} \exp\left(-\frac{1}{\theta_1^2} \sum_{i=1}^{n} x_i^2\right) \quad \text{for } \theta_1 > 0$$

with maximum likelihood estimate $\hat{\theta}_1 = \left(\frac{1}{n} \sum\limits_{i=1}^{n} x_i^2\right)^{1/2}$.

Similarly the likelihood function for an observed random sample $y_1, y_2, \ldots, y_m$ from a Weibull$(2, \theta_2)$ distribution is

$$L_2(\theta_2) = \theta_2^{-2m} \exp\left(-\frac{1}{\theta_2^2} \sum_{i=1}^{m} y_i^2\right) \quad \text{for } \theta_2 > 0$$

with maximum likelihood estimate $\hat{\theta}_2 = \left(\frac{1}{m} \sum\limits_{i=1}^{m} y_i^2\right)^{1/2}$.

Since the samples are independent the likelihood function for (θ_1, θ_2) is

$$L(\theta_1, \theta_2) = L_1(\theta_1)L_2(\theta_2) \quad \text{for } \theta_1 > 0, \theta_2 > 0$$

and the log likelihood function

$$l(\theta_1, \theta_2) = -2n \log \theta_1 - \frac{1}{\theta_1^2} \sum_{i=1}^{n} x_i^2 - 2m \log \theta_2 - \frac{1}{\theta_2^2} \sum_{i=1}^{m} y_i^2 \quad \text{for } \theta_1 > 0, \theta_2 > 0$$

The independence of the samples implies the maximum likelihood estimators are

$$\tilde{\theta}_1 = \left(\frac{1}{n} \sum_{i=1}^{n} X_i^2 \right)^{1/2} \qquad \tilde{\theta}_2 = \left(\frac{1}{m} \sum_{i=1}^{m} Y_i^2 \right)^{1/2}$$

Therefore

$$l(\tilde{\theta}_1, \tilde{\theta}_2; \mathbf{X}, \mathbf{Y}) = -n \log \left(\frac{1}{n} \sum_{i=1}^{n} X_i^2 \right) - m \log \left(\frac{1}{m} \sum_{i=1}^{m} Y_i^2 \right) - (n+m)$$

If $\theta_1 = \theta_2 = \theta$ then the log likelihood function is

$$l(\theta) = -2(n+m) \log \theta - \frac{1}{\theta^2} \left(\sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{m} y_i^2 \right) \quad \text{for } \theta > 0$$

which is only a function of θ. To determine $\max\limits_{(\theta_1, \theta_2) \in \Omega_0} l(\theta_1, \theta_2; \mathbf{X}, \mathbf{Y})$ we note that

$$\frac{d}{d\theta} l(\theta) = \frac{-2(n+m)}{\theta} + \frac{2}{\theta^3} \left(\sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{m} y_i^2 \right)$$

and $\frac{d}{d\theta} l(\theta) = 0$ for

$$\theta = \left[\frac{1}{n+m} \left(\sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{m} y_i^2 \right) \right]^{1/2}$$

and therefore

$$\max_{(\theta_1, \theta_2) \in \Omega_0} l(\theta_1, \theta_2; \mathbf{X}, \mathbf{Y}) = -(n+m) \log \left[\frac{1}{n+m} \left(\sum_{i=1}^{n} X_i^2 + \sum_{i=1}^{m} Y_i^2 \right) \right] - (n+m)$$

The likelihood ratio test statistic is

$$\begin{aligned}
& \Lambda(\mathbf{X}, \mathbf{Y}; \Omega_0) \\
= \; & 2 \left[l(\tilde{\theta}_1, \tilde{\theta}_2; \mathbf{X}, \mathbf{Y}) - \max_{(\theta_1, \theta_2) \in \Omega_0} l(\theta_1, \theta_2; \mathbf{X}, \mathbf{Y}) \right] \\
= \; & 2\left[-n \log \left(\frac{1}{n} \sum_{i=1}^{n} X_i^2 \right) - m \log \left(\frac{1}{m} \sum_{i=1}^{m} Y_i^2 \right) - (n+m) \right. \\
& \left. + (n+m) \log \left[\frac{1}{n+m} \left(\sum_{i=1}^{n} X_i^2 + \sum_{i=1}^{m} Y_i^2 \right) \right] + (n+m) \right] \\
= \; & 2\left[(n+m) \log \left[\frac{1}{n+m} \left(\sum_{i=1}^{n} X_i^2 + \sum_{i=1}^{m} Y_i^2 \right) \right] \right. \\
& \left. - n \log \left(\frac{1}{n} \sum_{i=1}^{n} X_i^2 \right) - m \log \left(\frac{1}{m} \sum_{i=1}^{m} Y_i^2 \right) \right]
\end{aligned}$$

with corresponding observed value

$$\begin{aligned}\lambda(\mathbf{x},\mathbf{y};\Omega_0) &= 2[\,(n+m)\log\left[\frac{1}{n+m}\left(\sum_{i=1}^{n}x_i^2+\sum_{i=1}^{m}y_i^2\right)\right]\\ &\quad -n\log\left(\frac{1}{n}\sum_{i=1}^{n}x_i^2\right)-m\log\left(\frac{1}{m}\sum_{i=1}^{m}y_i^2\right)]\end{aligned}$$

Since $k-q=2-1=1$

$$\begin{aligned}p\text{-}value &\approx P\left[W\geq\lambda(\mathbf{x},\mathbf{y};\Omega_0)\right] \quad \text{where } W\sim\chi^2(1)\\ &= 2\left[1-P\left(Z\leq\sqrt{\lambda(\mathbf{x},\mathbf{y};\Omega_0)}\right)\right] \quad \text{where } Z\sim \mathrm{N}\,(0,1)\end{aligned}$$

11. Summary of Named Distributions

Summary of Discrete Distributions

Notation and Parameters	**Probability Function** $f(x)$	**Mean** $E(X)$	**Variance** $Var(X)$	**Moment Generating Function** $M(t)$
Discrete Uniform(a,b) $b \geq a$ a,b integers	$\frac{1}{b-a+1}$ $x = a, a+1, \ldots, b$	$\frac{a+b}{2}$	$\frac{(b-a+1)^2-1}{12}$	$\frac{1}{b-a+1}\sum_{x=a}^{b} e^{tx}$ $t \in \Re$
Hypergeometric(N,r,n) $N = 1, 2, \ldots$ $n = 0, 1, \ldots, N$ $r = 0, 1, \ldots, N$	$\frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$ $x = \max(0, n-N+r),$ $\ldots, \min(r,n)$	$\frac{nr}{N}$	$\frac{nr}{N}(1-\frac{r}{N})\frac{N-n}{N-1}$	Not tractable
Binomial(n,p) $0 \leq p \leq 1,\ q = 1-p$ $n = 1, 2, \ldots$	$\binom{n}{x}p^x q^{n-x}$ $x = 0, 1, \ldots, n$	np	npq	$(pe^t + q)^n$ $t \in \Re$
Bernoulli(p) $0 \leq p \leq 1,\ q = 1-p$	$p^x q^{1-x}$ $x = 0, 1$	p	pq	$pe^t + q$ $t \in \Re$
Negative Binomial(k,p) $0 < p \leq 1,\ q = 1-p$ $k = 1, 2, \ldots$	$\binom{x+k-1}{x}p^k q^x$ $= \binom{-k}{x}p^k(-q)^x$ $x = 0, 1, \ldots$	$\frac{kq}{p}$	$\frac{kq}{p^2}$	$\left(\frac{p}{1-qe^t}\right)^k$ $t < -\ln q$
Geometric(p) $0 < p \leq 1,\ q = 1-p$	pq^x $x = 0, 1, \ldots$	$\frac{q}{p}$	$\frac{q}{p^2}$	$\frac{p}{1-qe^t}$ $t < -\ln q$
Poisson(μ) $\mu \geq 0$	$\frac{\mu^x e^{-\mu}}{x!}$ $x = 0, 1, \ldots$	μ	μ	$e^{\mu(e^t-1)}$ $t \in \Re$
Multinomial$(n; p_1, p_2, \ldots, p_k)$ $0 \leq p_i \leq 1$ $i = 1, 2, \ldots, k$ and $\sum_{i=1}^{k} p_i = 1$	$f(x_1, x_2, \ldots, x_k) =$ $\frac{n!}{x_1! x_2! \cdots x_k!} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}$ $x_i = 0, 1, \ldots, n$ $i = 1, 2, \ldots, k$ and $\sum_{i=1}^{k} x_i = n$	$E(X_i) = np_i$ $i = 1, 2, \ldots, k$	$Var(X_i)$ $= np_i(1-p_i)$ $i = 1, 2, \ldots, k$	$M(t_1, t_2, \ldots, t_{k-1}) =$ $(p_1 e^{t_1} + p_2 e^{t_2} + \cdots$ $+ p_{k-1} e^{t_{k-1}} + p_k)^n$ $t_i \in \Re$ $i = 1, 2, \ldots, k-1$

Summary of Continuous Distributions

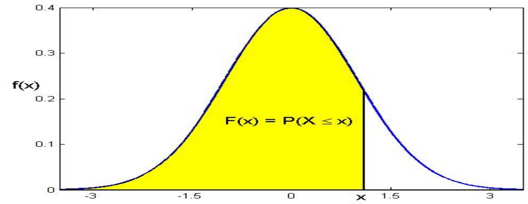
Notation and Parameters	**Probability Density Function** $f(x)$	**Mean** $E(X)$	**Variance** $Var(X)$	**Moment Generating Function** $M(t)$
Uniform(a,b) $b > a$	$\frac{1}{b-a}$ $a < x < b$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{bt}-e^{at}}{(b-a)t} \quad t \neq 0$ $1 \quad t = 0$
Beta(a,b) $a > 0,\ b > 0$	$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}$ $0 < x < 1$ $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx$	$\frac{a}{a+b}$	$\frac{ab}{(a+b+1)(a+b)^2}$	$1 + \sum_{k=1}^{\infty}\left(\prod_{i=0}^{k-1}\frac{a+i}{a+b+i}\right)\frac{t^k}{k!}$ $t \in \Re$
N(μ,σ^2) $\mu \in \Re,\ \sigma^2 > 0$	$\frac{e^{-(x-\mu)^2/(2\sigma^2)}}{\sqrt{2\pi}\,\sigma}$ $x \in \Re$	μ	σ^2	$e^{\mu t+\sigma^2 t^2/2}$ $t \in \Re$
Lognormal(μ,σ^2) $\mu \in \Re,\ \sigma^2 > 0$	$\frac{e^{-(\log x-\mu)^2/(2\sigma^2)}}{\sqrt{2\pi}\,\sigma x}$ $x > 0$	$e^{\mu+\sigma^2/2}$	$e^{2(\mu+\sigma^2)}$ $-e^{2\mu+\sigma^2}$	DNE
Exponential(θ) $\theta > 0$	$\frac{1}{\theta}e^{-x/\theta}$ $x \geq 0$	θ	θ^2	$\frac{1}{1-\theta t}$ $t < \frac{1}{\theta}$
Two Parameter Exponential(α,β) $\alpha \in \Re,\ \beta > 0$	$\frac{1}{\beta}e^{-(x-\alpha)/\beta}$ $x \geq \alpha$	$\alpha + \beta$	β^2	$\frac{e^{\alpha t}}{(1-\beta t)}$ $t < \frac{1}{\beta}$
Double Exponential(μ,β) $\mu \in \Re,\ \beta > 0$	$\frac{1}{2\beta}e^{-\lvert x-\mu\rvert/\beta}$ $x \in \Re$	μ	$2\beta^2$	$\frac{e^{\mu t}}{(1-\beta^2 t^2)}$ $\lvert t\rvert < \frac{1}{\beta}$
Extreme Value(μ,β) $\mu \in \Re,\ \beta > 0$	$\frac{1}{\beta}e^{[(x-\mu)/\beta - e^{(x-\mu)/\beta}]}$ $x \in \Re$	$\mu - \gamma\beta$ $\gamma \approx 0.5772$ Euler's constant	$\frac{\pi^2\beta^2}{6}$	$e^{\mu t}\Gamma(1+\beta t)$ $t > -1/\beta$
Gamma(α,β) $\alpha > 0,\ \beta > 0$	$\frac{x^{\alpha-1}e^{-x/\beta}}{\beta^\alpha\Gamma(\alpha)}$ $x > 0$	$\alpha\beta$	$\alpha\beta^2$	$(1-\beta t)^{-\alpha}$ $t < \frac{1}{\beta}$
Inverse Gamma(α,β) $\alpha > 0,\ \beta > 0$	$\frac{x^{-\alpha-1}e^{-1/(\beta x)}}{\beta^\alpha\Gamma(\alpha)}$ $x > 0$	$\frac{1}{\beta(\alpha-1)}$ $\alpha > 1$	$\frac{1}{\beta^2(\alpha-1)^2(\alpha-2)}$ $\alpha > 2$	DNE

Summary of Continuous Distributions Continued

Notation and Parameters	**Probability Density Function** $f(x)$	**Mean** $E(X)$	**Variance** $Var(X)$	**Moment Generating Function** $M(t)$
$\chi^2(k)$ $k = 1,2,\ldots$	$\frac{x^{(k/2)-1}e^{-x/2}}{2^{k/2}\Gamma(k/2)}$ $x > 0$	k	$2k$	$(1-2t)^{-k/2}$ $t < \frac{1}{2}$
Weibull(α,β) $\alpha > 0,\ \beta > 0$	$\frac{\alpha}{\beta^\alpha}x^{\alpha-1}e^{-(x/\beta)^\alpha}$ $x > 0$	$\beta\Gamma(1+\frac{1}{\alpha})$	$\beta^2[\Gamma(1+\frac{2}{\alpha})$ $-\Gamma^2(1+\frac{1}{\alpha})]$	Not tractable
Pareto(α,β) $\alpha > 0,\ \beta > 0$	$\frac{\beta\alpha^\beta}{x^{\beta+1}}$ $x > \alpha$	$\frac{\alpha\beta}{\beta-1}$ $\beta > 1$	$\frac{\alpha^2\beta}{(\beta-1)^2(\beta-2)}$ $\beta > 2$	DNE
Logistic(μ,β) $\mu \in \Re,\ \beta > 0$	$\frac{e^{-(x-\mu)/\beta}}{\beta\left[1+e^{-(x-\mu)/\beta}\right]^2}$ $x \in \Re$	μ	$\frac{\beta^2\pi^2}{3}$	$e^{\mu t}\Gamma(1-\beta t)\Gamma(1+\beta t)$
Cauchy(μ,β) $\mu \in \Re,\ \beta > 0$	$\frac{1}{\beta\pi\left\{1+[(x-\mu)/\beta]^2\right\}}$ $x \in \Re$	DNE	DNE	DNE
t(k) $k = 1,2,\ldots$	$\frac{\Gamma\left(\frac{k+1}{2}\right)\left(1+\frac{x^2}{k}\right)^{-(k+1)/2}}{\sqrt{k\pi}\,\Gamma\left(\frac{k}{2}\right)}$ $x \in \Re$	0 $k = 2,3,\ldots$	$\frac{k}{k-2}$ $k = 3,4,\ldots$	DNE
F(k_1,k_2) $k_1 = 1,2,\ldots$ $k_2 = 1,2,\ldots$	$\frac{\left(\frac{k_1}{k_2}\right)^{\frac{k_1}{2}}\Gamma\left(\frac{k_1+k_2}{2}\right)}{\Gamma\left(\frac{k_1}{2}\right)\Gamma\left(\frac{k_2}{2}\right)} \times$ $x^{\frac{k_1}{2}-1}\left(1+\frac{k_1}{k_2}x\right)^{-\frac{k_1+k_2}{2}}$ $x > 0$	$\frac{k_2}{k_2-2}$ $k_2 > 2$	$\frac{2k_2^2(k_1+k_2-2)}{k_1(k_2-2)^2(k_2-4)}$ $k_2 > 4$	DNE
$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim$ BVN(μ,Σ) $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$ $\mu_1 \in \Re,\ \mu_2 \in \Re$ $\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$ $\sigma_1 > 0,\ \sigma_2 > 0$ $-1 < \rho < 1$	$f(x_1,x_2) =$ $\frac{1}{2\pi\lvert\Sigma\rvert^{1/2}}e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}$ $x_1 \in \Re,\ x_2 \in \Re$	μ	Σ	$M(t_1,t_2)$ $= e^{\mu^T t+\frac{1}{2}t^T\Sigma t}$ $t_1 \in \Re,\ t_2 \in \Re$

12. Distribution Tables

N(0,1) Cumulative Distribution Function



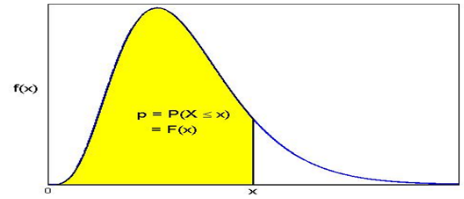
This table gives values of $F(x) = P(X \le x)$ for $X \sim N(0,1)$ and $x \ge 0$

x	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.99900
3.1	0.99903	0.99906	0.99910	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976
3.5	0.99977	0.99978	0.99978	0.99979	0.99980	0.99981	0.99981	0.99982	0.99983	0.99983

N(0,1) Quantiles: This table gives values of $F^{-1}(p)$ for $p \ge 0.5$

p	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.075**	**0.08**	**0.09**	**0.095**
0.5	0.0000	0.0251	0.0502	0.0753	0.1004	0.1257	0.1510	0.1764	0.1891	0.2019	0.2275	0.2404
0.6	0.2533	0.2793	0.3055	0.3319	0.3585	0.3853	0.4125	0.4399	0.4538	0.4677	0.4959	0.5101
0.7	0.5244	0.5534	0.5828	0.6128	0.6433	0.6745	0.7063	0.7388	0.7554	0.7722	0.8064	0.8239
0.8	0.8416	0.8779	0.9154	0.9542	0.9945	1.0364	1.0803	1.1264	1.1503	1.1750	1.2265	1.2536
0.9	1.2816	1.3408	1.4051	1.4758	1.5548	1.6449	1.7507	1.8808	1.9600	2.0537	2.3263	2.5758

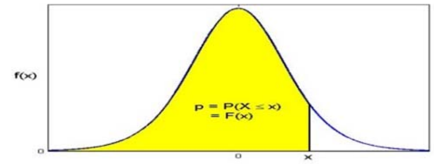
Chi-Squared Quantiles



This table gives values of x for p = P(X ≤ x) = F(x)

df\p	0.005	0.01	0.025	0.05	0.1	0.9	0.95	0.975	0.99	0.995
1	0.000	0.000	0.001	0.004	0.016	2.706	3.842	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.992	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.146	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.647	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.054	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.391	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.430	104.210
80	51.172	53.540	57.153	60.391	64.278	96.578	101.880	106.630	112.330	116.320
90	59.196	61.754	65.647	69.126	73.291	107.570	113.150	118.140	124.120	128.300
100	67.328	70.065	74.222	77.929	82.358	118.500	124.340	129.560	135.810	140.170

Student t Quantiles



This table gives values of x for $p = P(X \le x) = F(x)$, for $p \ge 0.6$

df \ p	0.6	0.7	0.8	0.9	0.95	0.975	0.99	0.995	0.999	0.9995
1	0.3249	0.7265	1.3764	3.0777	6.3138	12.7062	31.8205	63.6567	318.3088	636.6192
2	0.2887	0.6172	1.0607	1.8856	2.9200	4.3027	6.9646	9.9248	22.3271	31.5991
3	0.2767	0.5844	0.9785	1.6377	2.3534	3.1824	4.5407	5.8409	10.2145	12.9240
4	0.2707	0.5686	0.9410	1.5332	2.1318	2.7764	3.7469	4.6041	7.1732	8.6103
5	0.2672	0.5594	0.9195	1.4759	2.0150	2.5706	3.3649	4.0321	5.8934	6.8688
6	0.2648	0.5534	0.9057	1.4398	1.9432	2.4469	3.1427	3.7074	5.2076	5.9588
7	0.2632	0.5491	0.8960	1.4149	1.8946	2.3646	2.9980	3.4995	4.7853	5.4079
8	0.2619	0.5459	0.8889	1.3968	1.8595	2.3060	2.8965	3.3554	4.5008	5.0413
9	0.2610	0.5435	0.8834	1.3830	1.8331	2.2622	2.8214	3.2498	4.2968	4.7809
10	0.2602	0.5415	0.8791	1.3722	1.8125	2.2281	2.7638	3.1693	4.1437	4.5869
11	0.2596	0.5399	0.8755	1.3634	1.7959	2.2010	2.7181	3.1058	4.0247	4.4370
12	0.2590	0.5386	0.8726	1.3562	1.7823	2.1788	2.6810	3.0545	3.9296	4.3178
13	0.2586	0.5375	0.8702	1.3502	1.7709	2.1604	2.6503	3.0123	3.8520	4.2208
14	0.2582	0.5366	0.8681	1.3450	1.7613	2.1448	2.6245	2.9768	3.7874	4.1405
15	0.2579	0.5357	0.8662	1.3406	1.7531	2.1314	2.6025	2.9467	3.7328	4.0728
16	0.2576	0.5350	0.8647	1.3368	1.7459	2.1199	2.5835	2.9208	3.6862	4.0150
17	0.2573	0.5344	0.8633	1.3334	1.7396	2.1098	2.5669	2.8982	3.6458	3.9651
18	0.2571	0.5338	0.8620	1.3304	1.7341	2.1009	2.5524	2.8784	3.6105	3.9216
19	0.2569	0.5333	0.8610	1.3277	1.7291	2.0930	2.5395	2.8609	3.5794	3.8834
20	0.2567	0.5329	0.8600	1.3253	1.7247	2.0860	2.5280	2.8453	3.5518	3.8495
21	0.2566	0.5325	0.8591	1.3232	1.7207	2.0796	2.5176	2.8314	3.5272	3.8193
22	0.2564	0.5321	0.8583	1.3212	1.7171	2.0739	2.5083	2.8188	3.5050	3.7921
23	0.2563	0.5317	0.8575	1.3195	1.7139	2.0687	2.4999	2.8073	3.4850	3.7676
24	0.2562	0.5314	0.8569	1.3178	1.7109	2.0639	2.4922	2.7969	3.4668	3.7454
25	0.2561	0.5312	0.8562	1.3163	1.7081	2.0595	2.4851	2.7874	3.4502	3.7251
26	0.2560	0.5309	0.8557	1.3150	1.7056	2.0555	2.4786	2.7787	3.4350	3.7066
27	0.2559	0.5306	0.8551	1.3137	1.7033	2.0518	2.4727	2.7707	3.4210	3.6896
28	0.2558	0.5304	0.8546	1.3125	1.7011	2.0484	2.4671	2.7633	3.4082	3.6739
29	0.2557	0.5302	0.8542	1.3114	1.6991	2.0452	2.4620	2.7564	3.3962	3.6594
30	0.2556	0.5300	0.8538	1.3104	1.6973	2.0423	2.4573	2.7500	3.3852	3.6460
40	0.2550	0.5286	0.8507	1.3031	1.6839	2.0211	2.4233	2.7045	3.3069	3.5510
50	0.2547	0.5278	0.8489	1.2987	1.6759	2.0086	2.4033	2.6778	3.2614	3.4960
60	0.2545	0.5272	0.8477	1.2958	1.6706	2.0003	2.3901	2.6603	3.2317	3.4602
70	0.2543	0.5268	0.8468	1.2938	1.6669	1.9944	2.3808	2.6479	3.2108	3.4350
80	0.2542	0.5265	0.8461	1.2922	1.6641	1.9901	2.3739	2.6387	3.1953	3.4163
90	0.2541	0.5263	0.8456	1.2910	1.6620	1.9867	2.3685	2.6316	3.1833	3.4019
100	0.2540	0.5261	0.8452	1.2901	1.6602	1.9840	2.3642	2.6259	3.1737	3.3905
>100	0.2535	0.5247	0.8423	1.2832	1.6479	1.9647	2.3338	2.5857	3.1066	3.3101