Instructor Manual
Solved Problems and Sample Exams for

STATISTICS
A CONCISE MATHEMATICAL INTRODUCTION
FOR STUDENTS, SCIENTISTS, AND ENGINEERS

DAVID W. SCOTT
Rice University
Houston, Texas

October 13, 2020

Contents

Chapter 1

Data Analysis and Understanding

1. In a typical introductory statistics course, the construction of a histogram from data boils down to the choice of the number of bins, k. Equivalently, you can focus on the bin width, h, which would be found by dividing the width of the sample range (a, b) by the number of bins: $h = (b - a)/k$. Often you will be taught that the choice of k should be any "convenient" value. In Section 9.1, we learn that the formula $h^* = 1.06\, s_x\, n^{-1/5}$ can be justified for normal data.

   (a) Generate a normal random sample in R by

   ```
   set.seed(246); x = rnorm(10000) .
   ```

   The standard deviation, s_x, is `sd(x) = 1.012`. Examine the default histogram in R by `hist(x,col=2)`. What value of k is used? (Hint: $k = 16$.) Show the corresponding $h = 0.474$. Look at two other histograms choosing bin widths $h/2$ and $2h$, using the optional second argument `breaks` via

   ```
   hist( x, breaks = seq(-4.5,4.5,.948), col=2 )
   ```

   (b) The formula given above results in $h^* = 0.170$. Examine the 3 histograms using $h^*/2$, h^*, and $2h^*$.

   (c) Discuss and evaluate your favorite histogram among these six.

   Hint: Use the on-line R help by invoking `help.start()`.

   Solution:

(a) The default choice for k is called Sturges' Rule, where $k = 1 + \log_2(n)$, which is 14.29 for $n = 10^4$. The sample range is $(-3.88, 3.70)$. Using the interval $(-4.5, 4.5)$ ensures that the sequence used for `break` includes all the data points.

(b) $h^* = 1.06 \times 1.012 \times 10000^{-1/5} = 0.170$.

(c) Sturges' Rule selects few bins, especially as n grows. The resulting histogram omits details in the random sample. For example, when $n = 10^6$, $k = 21$, while $h^* = 0.669$, which roughly corresponds to $k = 120$ if the sample range is $(-4, 4)$. This rule is called Scott's Rule and suggests using 6 times the number of bins.

2. TBA

1.1 Exploring the Distribution of Data

1. TBA

1.1.1 Pearson's Father-Son Height Data

1. TBA

1.1.2 Lord Rayleigh's Data

1. TBA

1.1.3 Discussion

1. TBA

1.2 Exploring Prediction Using Data

1. TBA

1.2.1 Body and Brain Weights of Land Mammals

1. TBA

1.2.2 Space Shuttle Flight 25

1. TBA

1.2.3 Pearson's Father-Son Height Data Revisited

1. TBA

1.2.4 Discussion

1. TBA

1.3 Problems

1. A **frequency histogram** of continuous data is constructed by counting the number of data points that fall into equally-spaced bins of width h. h is called the **bin width**. Typically the bin edges are 0, $\pm h$, $\pm 2h$, $\pm 3h$, and so on. If the bin count in the k^{th} bin is denoted by ν_k, then the frequency histogram is defined as

$$\boxed{\hat{f}(x) = \nu_k \,, \qquad \text{for } x \text{ in the } k^{th} \text{ bin.}} \tag{1.1}$$

   (a) Show that the total area of the frequency histogram is nh, where $n = \sum_k \nu_k$. *Hint: The histogram is made up of rectangular blocks of width h and height v_i.*

   (b) A **probability histogram** is defined to have total area of one. Show that the following definition of a histogram has area equal to one:

$$\boxed{\hat{f}(x) = \frac{\nu_k}{nh} \,, \qquad \text{for } x \text{ in the } k^{th} \text{ bin.}} \tag{1.2}$$

Solution:

   (a) The frequency histogram is a bar chart, where the height of a bar in the k^{th} bin is the bin count, ν_k. We denote the k^{th} bin interval by B_k. The widths of all the bars is the bin width, h. Hence, the total area of the frequency polygon is given by

$$\begin{aligned}\int_{-\infty}^{\infty} \hat{f}(x)\,dx &= \sum_{k=-\infty}^{\infty} \int_{B_k} \hat{f}(x)\,dx \\ &= \sum_{k=-\infty}^{\infty} \nu_k \times h \qquad \text{(the area of the } k^{th} \text{ bar)} \\ &= h \times \sum_{k=-\infty}^{\infty} \nu_k \\ &= h \times n\,.\end{aligned}$$

(b) Thus dividing the frequency histogram by nh makes the total area equal to 1. We call this diagram the probability histogram.

2. One of the most famous epidemiological cases occurred in 1854 when Dr. John Snow successfully tracked down the source of an outbreak of **cholera** in the London suburb of SoHo. He mapped the households of some 500 victims over a 10-day period that lived within a quarter of mile of each other. However, many tens of thousands had died of cholera in England during the prior two decades. Dr. Snow believed contaminated water was a primary cause. Just as in the space shuttle example, there are choices of an appropriate time interval and the geographical extent that can influence your conclusions. Using the descriptions and maps conveniently assembled at

   `http://www.ph.ucla.edu/epi/snow/snowcricketarticle.html` ,

   discuss the evidence and choices that were and could have been made. *Hint: These data have been conveniently collected in CRAN Library* `HistData` *by Michael Friendly. Look at the help file for dataset* `snow` *and its example code.*

   Solution: This is an open-ended problem. You might think about the choice of the boundaries, and the impact of padding the data with 0's. We saw this 0 problem in the context of the Space Shuttle analysis.

3. (a) The Tukey power transformation of a variable x is x^{λ} for any nonzero $\lambda \in \mathbb{R}^1$. To better understand why the $\log(x)$ is used in

place of x^0 when $\lambda = 0$, we consider the linear re-expression of the Tukey transformation given by the formula

$$\frac{x^\lambda - 1}{\lambda}\,. \tag{1.3}$$

Since (1.3) is $0/0$ when $\lambda = 0$, use l'Hôpital's rule to find the limit transformation as $\lambda \to 0$. The scatter diagram using either formula for fixed nonzero λ will be visually identical. Formula (1.3) is referred to as the Box-Cox transformation; see Figure 1.1.

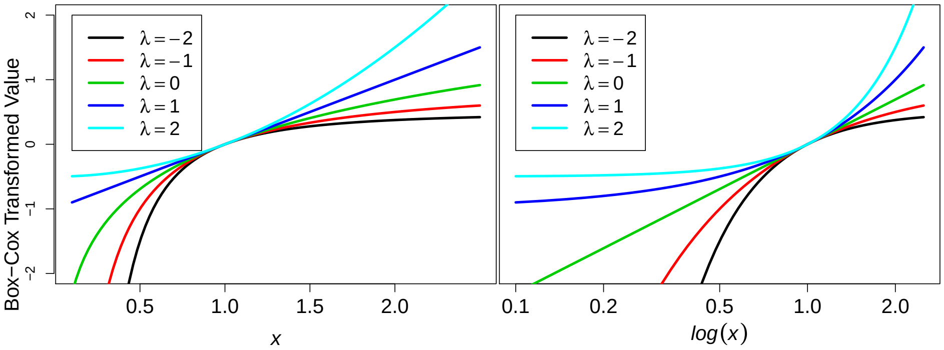


Figure 1.1: Box-Cox Transformation on natural and log scales

(b) Sometimes the transformation $\log(1+x)$ is in place of $\log(x)$ when $x \geq 0$ and x can take on the value 0. In this case, the original and transformed values of 0 are both 0. Try this form on the body-brain data and compare to Figure 1.4.

Solution:

(a) The Box-Cox transformation at $\lambda = 1$ is $0/0$, which is undefined. l'Hôpital's rule is the take the derivate of the numerator and denominator w.r.t. λ and then evaluate at $\lambda = 0$. This gives us

$$\begin{aligned} \frac{\frac{d\,(x^\lambda - 1)}{d\lambda}}{\frac{d\lambda}{\lambda}} &= \frac{\log x \cdot x^\lambda}{1} \\ &= \log x \qquad \text{at } \lambda = 0\,. \end{aligned}$$

(b) The new figure, along with the orignal log-log plot, are shown below. The linear relationship is not as strong. Two other points should be made:

  (i) Assuming that the original x and y values are all non-negative, any positive value of c other than 1 may be (and should be) considered in order to make the new plot as linear as possible. One way is to compute the correlation between $\log(x+c)$ and $\log(y+c)$ and pick the value of c that maximizes the correlation. The correlation is 0.95957 at $c=0$ and 0.95973 at the best choice, $c=-0.00173$. Use $c=0$. Note the smallest value in the data is 0.005 grams.

  (ii) Advocate the use of $\log_{10}(x)$ rather than $\log(x)$ so the values can be understood better when plotted. Does anyone really know what the value 6.3 means on a natural log scale? It is 544.6. This is 2.74 on a log10 scale.

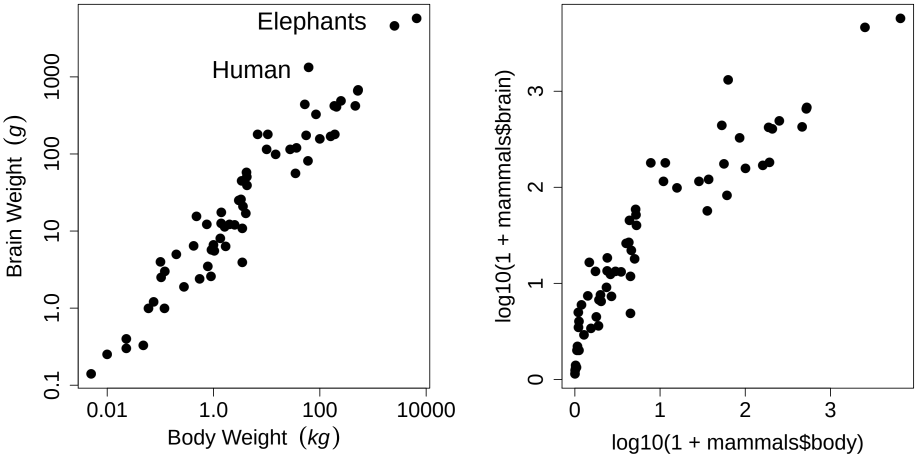


Figure 1.2: Scatter diagrams of the log-transformed body and brain weights of 62 land mammals and the $\log 10(1+\cdot)$ values.

Chapter 2

Classical Probability

2.1 Experiments with Equally Likely Outcomes

1. TBA

2.1.1 Simple Outcomes

1. TBA

2.1.2 Compound Events and Set Operations

1. **Simple Set Operations:** Find the following simple sets:

   (a) If $A \supseteq B$, what is $A \cup B$?
   (b) If $A \supseteq B$, what is $A \cap B$?
   (c) $A \cap A^c$?
   (d) $A \cup A^c$?

   Solution:

   (a) $A \cup B = A$
   (b) $A \cap B = B$
   (c) $A \cap A^c = \Omega$
   (d) $A \cup A^c = \emptyset$

---

2. **Counting Events:** List *all* possible events associated with a single toss of a regular die (6-sided) with faces showing the digits 1, 2, 3, 4, 5, or 6. **Hint:** Record the result of each roll of the die by the digit on the top.

   Solution: Record the result of a single toss of a die. Then $\Omega = \{1,2,3,4,5,6\}$. The list of **all events** includes Ω and $\varnothing$; next, the 6 simple events; and finally, all the compound events:

$$\begin{array}{rl}
1 \text{ null} & \varnothing, \\
6 \text{ w/ } 1 & \{1\},\{2\},\{3\},\{4\},\{5\},\{6\} \\
15 \text{ w/ } 2 & \{1,2\},\{1,3\},\{1,4\},\{1,5\},\{1,6\}, \\
 & \quad \{2,3\},\{2,4\},\{2,5\},\{2,6\}, \\
 & \quad \{3,4\},\{3,5\},\{3,6\},\{4,5\},\{4,6\},\{5,6\}, \\
20 \text{ w/ } 3 & \{1,2,3\},\{1,2,4\},\{1,2,5\},\{1,2,6\},\{1,3,4\}, \\
 & \quad \{1,3,5\},\{1,3,6\},\{1,4,5\},\{1,4,6\},\{1,5,6\}, \\
 & \quad \{2,3,4\},\{2,3,5\},\{2,3,6\},\{2,4,5\},\{2,4,6\}, \\
 & \quad \{2,5,6\},\{3,4,5\},\{3,4,6\},\{3,5,6\},\{4,5,6\}, \\
15 \text{ w/ } 4 & \{1,2,3,4\},\{1,2,3,5\},\{1,2,3,6\},\{1,2,4,5\},\{1,2,4,6\}, \\
 & \quad \{1,2,5,6\},\{1,3,4,5\},\{1,3,4,6\},\{1,3,5,6\},\{1,4,5,6\}, \\
 & \quad \{2,3,4,5\},\{2,3,4,6\},\{2,3,5,6\},\{2,4,5,6\},\{3,4,5,6\}, \\
6 \text{ w/ } 5 & \{1,2,3,4,5\},\{1,2,3,4,6\},\{1,2,3,5,6\}, \\
 & \quad \{1,2,4,5,6\},\{1,3,4,5,6\},\{2,3,4,5,6\} \\
1 \text{ w/ } 6 & \Omega \qquad\qquad (\Omega \text{ same as } \{1,2,3,4,5,6\})
\end{array}$$

3. **Counting Events (Part II):** Show
   (a) $\sum_{k=0}^{n} \binom{n}{k} = 2^n$
   (b) $\sum_{k=0}^{n} (-1)^k \binom{n}{k} = 0$
   (c) What is the application to the previous die rolling problem?

   Solution: By the Binomial Theorem,

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k} \,,$$

(a) $\sum_{k=0}^{n} \binom{n}{k} = (1+1)^n = 2^n$
(b) $\sum_{k=0}^{n} (-1)^k \binom{n}{k} = (-1+1)^n = 0$
(c) Rolling a single die results in one of 6 simple outcomes. The complete list of compound events has all the choices of 6 choose k outcomes, or $2^6 = 64$ events.

4. TBA

2.2 Probability Laws

1. When rolling a pair of die, find

   (a) the probability getting at least 1 six?
   (b) their sum is less than 5?

   Solution: There are 36 equally likely (simple) outcomes. Each has probability 1/36.

   (a) $\Pr(\text{at least 1 six}) = (6+6-1)/36 = 11/36$.
   (b) $\Pr(\text{sum is less than 5}) = \Pr(2,\ 3,\text{ or } 4) = (1+2+3)/36 = 1/6$.

2. TBA

2.2.1 Union and Intersection of Events A and B

1. TBA

2.2.1.1 Case (i):

1. TBA

2.2.1.2 Cases (ii) and (iii):

1. TBA

2.2.1.3 Case (iv):

1. TBA

2.2.2 Conditional Probability

1. TBA

2.2.2.1 Definition of Conditional Probability

1. (a) If $A \cap B = \emptyset$, find $P(A|B)$?
   (b) If $B \subset A$, find $P(A|B)$?

   Solution:

   (a)
   $$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(\emptyset)}{P(B)} = \frac{0}{P(B)} = 0\,.$$

   (b)
   $$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1\,.$$

2. When rolling a pair of dice, define the events
   $$\begin{aligned} A &= \{\text{at least one die shows a 3}\} \\ B &= \{\text{sum is at least 5}\} \end{aligned}$$

   Find

   (a) $P(B|A)$?
   (b) $P(A|B)$?

   Solution: There are 36 equally likely outcomes (x, y).

   (a) There are $6 + 6 - 1 = 11$ pairs that result in event A occurring. Of these, only $(1, 3)$ and $(3, 1)$ do not sum to at least 5. Hence,
   $$P(B|A) = \frac{11-2}{11} = \frac{9/36}{11/36} = \frac{P(B \cap A)}{P(A)},\,.$$

   (b) The events $A \cap B$ and $B \cap A$ are identical, so have the same probability. The probability of B is $30/36$, eliminating the 6 pairs $(1, 1)$, $(1, 2)$, $(1, 3)$, $(2, 1)$, $(2, 2)$, and $(3, 1)$. Hence,
   $$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{9/36}{30/36} = \frac{3}{10},\,.$$

3. TBA

2.2.2.2 Conditional Probability With More Than Two Events

1. **Conditional probability with 3 events:** Suppose a shipment of 25 replacement iPhone screens contains 4 cracked displays. Three are chosen sequentially at random from the shipment box. What is the probability that the first two are OK but the third is broken?

   Solution: Let the event G_i denote a good display is drawn on the *ith* try. Then the desired event is the compound event

   $$G_1 \cap G_2 \cap G_3 . \qquad \text{Its probability is}$$

   $$\begin{aligned} P(G_1 \cap G_2 \cap G_3) &= P(G_1)P(G_2|G_1)P(G_3|G_1 \cap G_2) \\ &= \frac{21}{25} \cdot \frac{20}{24} \cdot \frac{4}{23} \\ &= \frac{14}{115} = 12.17\%. \end{aligned}$$

2. **Alternative Birthday Problem:** How many individuals with random birthdays (of 365 possible) would make the probability of at least one duplicate birthday at least 90%?

   Solution: Following Section 2.2.2.2,

n	Probability at least 1 duplicate birthday
22	47.57%
23	50.73%
40	89.12%
41	90.32%

   so 41 individuals is sufficient.

3. **Scrabble Problem:** In the game of Scrabble, each player draws a tile to see who draws the lowest letter to see who goes first. (If players tie with the lowest letter, they draw again.) Here is a table of all 100 tiles, including two blank ('b') tiles. A blank tile beats the letter 'A.'

b	A	B	C	D	E	F	G	H	I	J	K	L	M
2	9	2	2	4	12	2	3	2	9	1	1	4	2
2	11	13	15	19	31	33	36	38	47	48	49	53	55
	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
	6	8	2	1	6	4	6	4	2	2	1	2	1
	61	69	71	72	78	82	88	92	94	96	97	99	100

(a) In a game with 4 players, what is the probability you would go first if you drew the letter ‘J’? *Hint: This is a conditional probability given that you drew the letter ‘J’.*

(b) For all 27 letters, what is the probability you not be immediately eliminated from going first? Graph these probabilities. *Hint: Unlike part (a) where you cannot be tied, here you can be tied but not eliminated.*

Solution:

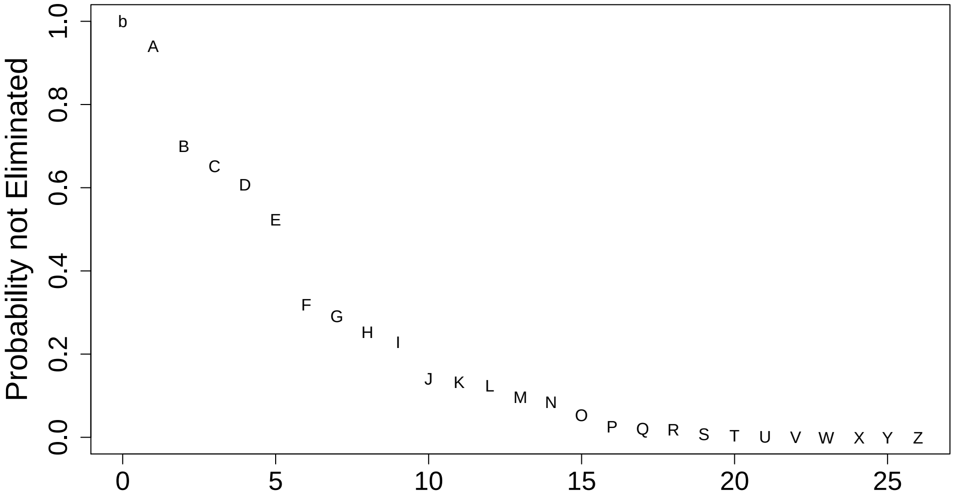
(a) There are 99 tiles remaining, 47 better than the letter ‘J’ which is ranked 48^{th} and 52 worse. There are 99 choose 3 ways your opponents can select their 3 tiles, of which 52 choose 3 are all letters after ‘J’. Hence,

$$\begin{aligned} Pr(\text{go first}|\text{J}) &= \frac{\binom{52}{3}}{\binom{99}{3}} = \frac{\frac{52!}{49!\,3!}}{\frac{99!}{96!\,3!}} \\ &= \frac{52 \cdot 51 \cdot 50}{99 \cdot 98 \cdot 97} = \frac{132{,}600}{941{,}094} \\ &= 14.09\% \,. \end{aligned}$$

(b) For other letters with more than 1 tile, you would include the remaining tiles that could result in a tie. For example, if you drew the letter ‘E’, then there are 19 letters that would beat you but 80 letters that would either lose to you or tie you (the other 11 E’s and 69 letters ‘F’ and above). Hence,

$$\begin{aligned} Pr(\text{not eliminated}|\text{E}) &= \frac{\binom{80}{3}}{\binom{99}{3}} = \frac{80 \cdot 79 \cdot 78}{99 \cdot 98 \cdot 97} = \frac{492{,}960}{941{,}094} \\ &= 52.38\% \,. \end{aligned}$$

The complete plot of the probabilities is given below. Note the probabilities are exactly 0 for the last three letters, since only 1 tile for each.



4. TBA

2.2.3 Independent Events

1. **Independent events identity:** Suppose $A_1, A_2, \ldots, A_n$ are independent events. Show

$$P\left(\bigcup_{i=1}^{n} A_i\right) = 1 - \prod_{i=1}^{n} P(A_i^c)\,.$$

Solution: Starting with the identity

$$\left(\bigcup_{i=1}^{n} A_i\right)^{\boldsymbol{c}} = \bigcap_{i=1}^{n} A_i^{\boldsymbol{c}}\,, \quad \text{we have}$$

$$\begin{aligned} P\left(\bigcup_{i=1}^{n} A_i\right) &= 1 - P\left(\bigcup_{i=1}^{n} A_i\right)^{\boldsymbol{c}} \\ &= 1 - P\left(\bigcap_{i=1}^{n} A_i^{\boldsymbol{c}}\right) \\ &= 1 - \prod_{i=1}^{n} P(A_i^{\boldsymbol{c}})\,, \end{aligned}$$

since the n events $A_i^{\boldsymbol{c}}$ are also independent.

2. TBA

2.2.4 Bayes Theorem

1. **Equivalence of Independent Events:** Prove that if $\{A, B\}$ are independent events, then so are $\{A, B^c\}$, $\{A^c, B\}$, and $\{A^c, B^c\}$.

   Solution:

   Remark: Take as known that if A is independent of B, then so is B independent of A, i.e., $P(A|B) = P(A) \iff P(B|A) = P(B)$.

   - **(a)** $\{A, B^c\}$

     $$P(B^c|A) = 1 - P(B|A) = 1 - P(B) = P(B^c)$$

   - **(b)** $\{A^c, B\}$ (basically the same)

     $$P(A^c|B) = 1 - P(A|B) = 1 - P(A) = P(A^c)$$

   - **(c)** $\{A^c, B^c\}$ (use complement and Bayes Theorem)

     $$\begin{aligned} P(A^c|B^c) = 1 - P(A|B^c) &= 1 - \frac{P(B^c|A)P(A)}{P(B^c)} \\ &= 1 - \frac{P(B^c)P(A)}{P(B^c)} = 1 - P(A) = P(A^c) \end{aligned}$$

2. TBA

2.2.5 Partitions and Total Probability

1. TBA

2.3 Counting Methods

1. A statistics professor is writing a quiz that covers 5 topics. She plans to reuse questions from previous semesters, one question per topic. If she has 3, 2, 4, 1, and 3 questions for each topic, respectively, how many different exams could she write?

   Solution: This follows the fundamental principle of counting perfectly. Hence,

   $$n(S) = \prod_{i=1}^{5} n(S_i) = 3 \times 2 \times 4 \times 1 \times 3 = 72\,.$$

2. TBA

2.3.1 With Replacement

1. TBA

2.3.2 Without Replacement (Permutations)

1. A research group meets once a month for a beer tasting excuse. One person volunteers to pick the order that the 6 beers will be offered, and to record the scores 1-6. Assuming the order in which the beers is presented matters, how many orderings are there?

   Solution: Since order matters and the $n = 6$ objects are distinguishable, this question is answered by computing the permutation

   $$^{n}P_{r} = \frac{n!}{(n-r)!} = \frac{6!}{0!} = 120\,,$$

   with $n = 6$ and $r = 6$.

2. In the previous problem, suppose 3 of the beers are pale ales, while 3 are its stronger couisin, IPA's. As a result, the pale ales will be offered before the IPA's. How many orderings are there?

   Solution: Beginning with the 3 pale ales, we see once again that order matters and the $n = 3$ objects are distinguishable. Thus there are

   $$^{n}P_{r} = \frac{n!}{(n-r)!} = \frac{3!}{0!} = 6\,,$$

   with $n = 3$ and $r = 3$. There are also 6 orderings for the IPA's. Hence, by the product rule, there are 6 × 6 or 36 orderings that may be offered. Note that "for each" choice of the pale ale orderings, there are 6 orderings of the IPA's. Hence, we have 6 × 6 orderings overall.

3. TBA

2.3.3 Without Replacement Nor Order (Combinations)

1. In a department of statistics with 10 faculty members, the department chair has determined she has sufficient funds for 3 to attend the joint statistical meetings. Including herself, how many groups of 3 must she consider?

   Solution: Since order does not matter but the $n = 10$ objects are distinguishable, this question is answered by computing the combination

   $$^{n}C_{r} = \frac{n!}{r!(n-r)!} = \frac{10!}{3!\,7!} = \frac{10 \times 9 \times 8 \times 7}{6} = 120\,,$$

   with $n = 10$ and $r = 3$. The R function *choose(10,3)* may be used to compute this.

2. In the previous problem, suppose the department chair determines she must attend the meetings in order to represent the department. How many groups must she consider in this scenario?

   Solution: She only has to choose 2 from the 9 remaining faculty. We compute `choose(9,2)=36`.

3. TBA

2.3.4 Examples

1. **Some Poker Hands:** For 5-card poker hands, find
    - (a) ν(3 of a kind);
    - (b) ν(full house);
    - (c) ν(1 pair);
    - (d) ν(Jack high).

Note: $\nu(\cdot)$ *denotes the "number of ways."*

Solution:

- **(a) 3 of a kind:**

$$\underbrace{\binom{13}{1}\cdot\binom{4}{3}}_{\text{3 of a kind}}\times\underbrace{\binom{12}{2}\cdot\binom{4}{1}\cdot\binom{4}{1}}_{\text{remaining 2 cards}}=54{,}912\,.$$

- **(b) full house:**

$$\underbrace{\binom{13}{1}\cdot\binom{4}{3}}_{\text{3 of a kind}}\times\underbrace{\binom{12}{1}\cdot\binom{4}{2}}_{\text{2 of a kind}}=3{,}744\,.$$

- **(c) 1 pair:**

$$\underbrace{\binom{13}{1}\cdot\binom{4}{2}}_{\text{the pair}}\times\underbrace{\binom{12}{3}\cdot\binom{4}{1}\cdot\binom{4}{1}\cdot\binom{4}{1}}_{\text{remaining 3 cards}}=1{,}098{,}240\,.$$

- **(d) Jack high (i.e., no Ace, King, Queen in the hand):**

$$\underbrace{\binom{4}{1}}_{\text{suit of jack}}\times\Bigg[\underbrace{\binom{9}{4}}_{\text{4 other cards}}\times\underbrace{\binom{4}{1}^4}_{\text{suits}}-\underbrace{\binom{9}{4}}_{\text{remove flushes}}$$

$$-\underbrace{4^4}_{\substack{\text{straights}\\ 7\,8\,9\,,10\,J}}+\underbrace{1}_{\substack{\text{add back 1 straight flush of}\\ \text{suit choosen outside brackets}}}\Bigg]=127{,}500\,.$$

Note: "4 other cards" excludes Jacks, Queens, Kings, and Aces.

2. TBA

2.3.5 Extended Combinations (Multinomial)

1. **Counting Committees:** From a Board of Directors with 4 men and 5 women, how many committees can be formed with

   - (a) 2 men and 3 women;
   - (b) 5 members, with at least 3 women?

   Solution: Committees from 4M and 5F:

   - **(a) 2M+3F:**

   $$\binom{4}{2} \times \binom{5}{3} = 60$$

   - **(b) 2M+3F or 1M+4F or 0M+5F (at least 3 women):**

   $$\binom{4}{2} \cdot \binom{5}{3} + \binom{4}{1} \cdot \binom{5}{4} + \binom{4}{0} \cdot \binom{5}{5} = 60 + 20 + 1 = 81\,.$$

2. TBA

2.4 Countable Sets: Implications As $n \to \infty$

1. TBA

2.4.1 Selecting Even or Odd Integers

1. TBA

2.4.2 Selecting Rational Versus Irrational Numbers

1. TBA

2.5 Kolmogorov's Axioms

1. **A Probability Inequality:** Prove $P(A \cup B) \leq P(A) + P(B)$.

   Solution: If we just start with Kolmogorov's Axioms, we may write AB as the disjoint union of 3 sets:

$$\begin{aligned} A \cup B &= AB^* \uplus AB \uplus A^*B\,; \qquad\qquad \text{hence,} \\ P(A \cup B) &= P(AB^*) + P(AB) + P(A^*B) \\ P(A \cup B) &\leq P(AB^*) + P(AB) + P(A^*B) + P(AB) \\ &= P\big(\underbrace{AB^* \uplus AB}_{A}\big) + P\big(\underbrace{A^*B \uplus AB}_{B}\big) \\ &= P(A) + P(B)\,. \end{aligned}$$

2. **A Decomposition of the Union of 3 Sets:** What is $P(A \cup B \cup C)$ in terms of known probabilities of events A, B, C, $A \cap B$, $A \cap C$, $B \cap C$, $A \cap B \cap C$? **Hint:** You probably can guess what the formula is if you draw a general Venn diagram with 3 intersecting circles. But use the formula for $P(U \cup V)$; then let $U = A$ and $V = B \cup C$ to work it all out.

   Solution: Following the hint, let $U = A$ and $V = B \cup C$; then

$$\begin{aligned} (1)\!: \quad P(A \cup B \cup C) &= P(U \cup V) = P(U) + P(V) - P(UV) \\ &= P(A) + \underbrace{P(B \cup C)}_{(2)} - \underbrace{P(A \cap (B \cup C))}_{(3)}\,. \\ (2)\!: \qquad P(B \cup C) &= P(B) + P(C) - P(BC) \\ (3)\!: \; P(A \cap (B \cup C)) &= P(AB \cup AC) \\ &= P(AB) + P(AC) - P(\underbrace{AB \cap AC}_{ABC})\,. \end{aligned}$$

   Combining, we have shown $P(A \cup B \cup C)$ equals

$$\boxed{P(A) + P(B) + P(C) - P(BC) - P(AB) - P(AC) + P(ABC)\,.}$$

3. TBA

2.6 Reliability: Series versus Parallel Networks

1. TBA

2.6.1 Series Network

1. TBA

2.6.2 Parallel Network

1. TBA

2.7 Problems

1. Show that the formula (2.8) for case (iv) also covers cases (i)-(iii).

   Solution: Formula (8.2) is

   $$P(A \cup B) = P(A) + P(B) - P(A \cap B)\,,$$

   (i) If $(A \cap B) = \varnothing$, then $P(A \cap B) = 0$; hence, $P(A \cup B) = P(A) + P(B)$, which is Equation (2.7).

   (ii) If $(A \cap B) = B$, then $P(A \cap B) = P(B)$; hence, $P(A \cup B) = P(A) + P(B) - P(B) = P(A)$.

   (ii) If $(A \cap B) = A$, then $P(A \cap B) = P(A)$; hence, $P(A \cup B) = P(A) + P(B) - P(A) = P(B)$.

2. How does $P(T|C)$ change depending on increasing $P(C)$ versus the PSA test characteristics?

   Solution: Problem 4 gives a better overview.

3. Show $P(A|B) + P(A^c|B) = 1$, as does $P(A|B^c) + P(A^c|B^c) = 1$.

Solution: Using the definition of conditional probability,

$$\begin{aligned}
P(A|B) + P(A^c|B) &= \frac{P(AB)}{P(B)} + \frac{P(A^cB)}{P(B)} \\
&= \frac{P(AB) + P(A^cB)}{P(B)} \\
&= \frac{P(AB \uplus A^cB)}{P(B)} \\
&= \frac{P((A \uplus A^c) \cap B)}{P(B)} \\
&= \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1\,.
\end{aligned}$$

Similarly,

$$\begin{aligned}
P(A|B^c) + P(A^c|B^c) &= \frac{P(AB^c)}{P(B^c)} + \frac{P(A^cB^c)}{P(B^c)} \\
&= \frac{P(AB^c) + P(A^cB^c)}{P(B^c)} \\
&= \frac{P(AB^c \uplus A^cB^c)}{P(B^c)} \\
&= \frac{P((A \uplus A^c) \cap B^c)}{P(B^c)} = \frac{P(B^c)}{P(B^c)} = 1\,.
\end{aligned}$$

4. Examine how $P(C|T)$ improves as a function of $P(C)$, $P(T|C)$, and $P(T^c|C^c)$.

Solution: By Bayes Theorem and the Law of Total Probability, we have

$$\begin{aligned}
P(C|T) &= \frac{P(T|C)P(C)}{P(T)} \\
&= \frac{P(T|C)P(C)}{P(T\cap(C\uplus C^{\mathbf{c}}))} = \frac{P(T|C)P(C)}{P(TC\cap TC^{\mathbf{c}})} \\
&= \frac{P(T|C)P(C)}{P(T|C)P(C)+P(T|C^{\mathbf{c}})P(C^{\mathbf{c}})} \\
&= \frac{P(T|C)P(C)}{P(T|C)P(C)+P(T|C^{\mathbf{c}})[1-P(C)]}
\end{aligned}$$

From the first line, $P(C|T)$ increases directly with $P(C)$ and/or $P(T|C)$. In the first case, you get better results if the group you are analyzing is at “high risk.” Alternatively, if $P(T|C)$ is better, then you have more “true positives.” If $P(T^{\mathbf{c}}|C^{\mathbf{c}})$ is greater, then you have more “true negatives.”

5. Prove these and/or draw Venn Diagrams; Equations (2.4) and (2.5).

   Solution: Graphical “proofs” of these distributive laws:

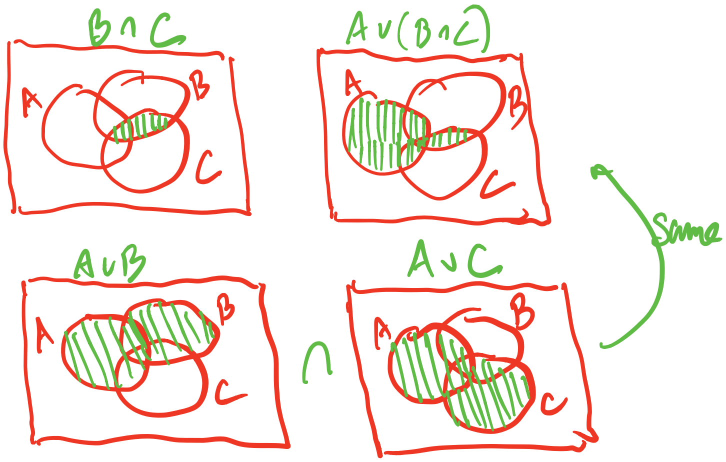


Figure 2.1: $(A\cup B)\cap(A\cup C) = A\cup(B\cap C)$

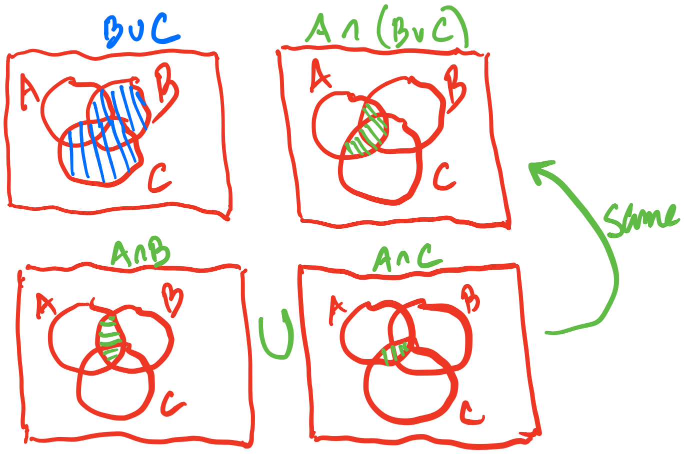


Figure 2.2: $(A \cap B) \cup (A \cap C) = A \cap (B \cup C)$

6. Show that for case (ii), $P(B|A)$ in Equation (2.13) can be expressed in terms of $P(A)$ and $P(B)$ to reach the same conclusion.

   Solution: In case (ii), $n_{A\cap B} = n_B$. Following Equation (2.9),

$$\begin{aligned} P(B|A) &= \frac{n_{B\cap A}}{n_A} \\ &= \frac{n_B}{n_A} = \frac{n_B/n}{n_A/n} \\ &= \frac{P(B)}{P(A)} \\ &> P(B)\,, \qquad \text{since } 0 < P(A) < 1. \end{aligned}$$

7. Show formula (2.22) is correct.

Solution: Just to get started and see the cancelation,

$$\begin{aligned}\binom{n}{n_1}\binom{n-n_1}{n_2}\cdots &= \frac{n!}{n_1!(n-n_1)!}\cdot\frac{(n-n_1)!}{n_2!(n-n_1-n_2)!}\\ &= \frac{n!}{n_1!n_2!}\times\frac{1}{(n-n_1-n_2)!}\cdots\\ &= \frac{n!}{n_1!n_2!\cdots n_k!}\,.\end{aligned}$$

8. Prove that a series network is no more reliable than its weakest component, while a parallel network is more reliable than its strongest component.

Solution: For a series network, suppose the k^{th} component is the least reliable. Then

$$\begin{aligned}P(S) &= \prod_{i=1}^{n} p_i\\ &= p_k \times \prod_{i\neq k}^{n} p_i\\ &< p_k\,,\end{aligned}$$

since the product term is less than 1.

For a parallel network, suppose the k^{th} component is the most reliable. Then

$$\begin{aligned}P(S) &= 1-\prod_{i=1}^{n}(1-p_i)\\ &= 1-(1-p_k)\times\prod_{i\neq k}^{n}(1-p_i)\\ &> 1-(1-p_k)\\ &= p_k\,,\end{aligned}$$

since the product term is less than 1.

Chapter 3

Random Variables and Models Derived From Classical Probability and Postulates

3.1 Random Variables and Probability Distributions: Discrete Uniform Example

1. TBA

3.1.1 Toss of a Single Die

1. TBA

3.1.2 Toss of a Pair of Dice

1. TBA

3.2 Continuous Uniform Example: The Univariate Probability Density Function

1. TBA

3.2.1 Using the PDF to Compute Probabilities

1. TBA

3.2.2 Using the PDF To Compute Relative Odds

1. TBA

3.3 Summary Statistics: Central and Non-Central Moments

1. TBA

3.3.1 Expectation, Average, and Mean

1. Suppose the domain of a continuous random variable, X, is $(0, \infty)$, show
$$EX = \int_0^\infty \Pr(X > x)\, dx.$$
Hint: Use integration by parts.

   Solution: Recall $\int u\, dv = uv - \int v\, du$. Let $u = 1 - F_X(x)$ and $v = x$. Then $du = -f_X(x)$ and
$$\begin{aligned} EX &= \int_0^\infty \Pr(X > x)\, dx \\ &= \int_0^\infty [1 - F_X(x)]\, dx \\ &= [1 - F_X(x)]\, x \Big|_{x=0}^{x=\infty} - \int_0^\infty x\,(-f_X(x))\, dx \\ &= 0 + \int_0^\infty x\, f_X(x)\, dx = \mu\,. \end{aligned}$$

2. TBA

3.3.2 Expectation as a Linear Operator

1. TBA

3.3.3 The Variance of a Random Variable

1. Suppose you are given the two expectations EX and $E[X(X-1)]$ for a random variable X. How can you use these to compute the variance σ_X^2?

   Solution: Recall $EX^2 = \sigma_X^2 + \mu_X^2$. Then

   $$\begin{aligned} E[X(X-1)] &= EX^2 - EX \\ &= \sigma_X^2 + \mu_X^2 - \mu_X \\ &= \sigma_X^2 + \mu_X(\mu_X - 1)\,; \quad \text{thus,} \\ \sigma_X^2 &= E[X(X-1)] - \mu_X(\mu_X - 1)\,. \end{aligned}$$

2. TBA

3.3.4 Standardized Random Variables

1. TBA

3.3.5 Higher Order Moments

1. TBA

3.3.6 Moment Generating Function

1. TBA

3.3.7 Measurement Scales and Units of Measurement

1. TBA

3.3.7.1 The Four Measurement Scales

1. TBA

3.3.7.2 Units of Measurement

1. TBA

3.4 Binomial Experiments

1. TBA

3.5 Success Waiting Time: Geometric PDF

1. TBA

3.6 Waiting Time For r Successes: Negative Binomial

1. TBA

3.7 Poisson Process and Distribution

1. TBA

3.7.1 Moments of the Poisson PMF

1. TBA

3.7.2 Examples

1. An organic honey seller has found that she makes 8 sales on an average Saturday farmer's market over a 4-hour morning. On a slow Saturday, she sells less than 4. What is the probability of a slow Saturday if X is Poisson?

   Solution: An organic honey seller has found that she makes 8 sales on an average Saturday farmer's market over a 4-hour morning. On a slow Saturday, she sells less than 4. What is the probability of a slow Saturday if X is Poisson?

$\lambda = 8$ sales over 4 hours or $\lambda = 2$. Thus $m = \lambda t = 2 \times 4 = 8$ (surprise?). Let $X \sim Pois(m = 8)$. A slow Saturday is $X = 0, 1, 2, 3$ (not 4). Thus the desired probability is

$$\begin{aligned} F(3) = P(X < 4) = P(X \le 3) &= \sum_{x=0}^{3} \frac{e^{-8} 8^x}{x!} \\ &= \texttt{sum(dpois(0:3, 8))} \\ &= 4.24\%\,, \end{aligned}$$

using the R function `dpois` to compute the PMF.

2. TBA

3.8 Waiting Time for Poisson Events: Negative Exponential PDF

1. An organic honey seller has found that she makes 8 sales on an average Saturday farmer's market over a 4-hour morning. She is thinking of going home early because she has had no sales since 8 a.m. and it now 11 a.m. Use the random variable T to measure the time to the first sale (of honey), assuming $\lambda = 8/4$ sales per hour. What is $P(T \ge 3.0\, hrs)$?

   Solution:

   $T \sim \lambda \exp(-\lambda\, t)$ with $\lambda = 2$. Can verify that $F(t) = 1 - \exp(-\lambda\, t)$; hence,

$$\begin{aligned} P(T \ge 3) = 1 - F(3) &= \int_3^\infty f(t)\, dt \\ &= 1 - \Big[1 - \exp\left(-2 \cdot 3\right) \Big] \\ &= e^{-6} = 0.248\%\,. \end{aligned}$$

   Note that $T > 3$ is equivalent to zero Poisson events from 8:00 a.m. until

11:00 a.,m., which has a mean of $m = \lambda t = 2 \cdot 3 = 6$. Now

$$\begin{aligned} p_X(x) &= \frac{e^{-m} m^x}{x!} \qquad \text{therefore} \\ p_X(0) &= \frac{e^{-6}\,(6)^0}{0!} \\ &= e^{-6} \quad \text{as before.} \end{aligned}$$

2. Suppose the lifetime of a light bulb follows a negative exponential distribution with a mean of 1000 hours. Find the probability the light bulb
   (a) lasts 1000 hours?
   (b) lasts 2000 hours?
   (c) burns out before 500 hours?

   Solution: The CDF of T is $F_T(t) = 1 - \exp(t/1000)$. Hence,

   (a) $\Pr(X \geq 1000) = \exp(-1000/1000) = 36.79\%$
   (b) $\Pr(X \geq 2000) = \exp(-2000/1000) = 13.53\%$
   (c) $\Pr(X \leq 500) = 1 - \exp(-500/1000) = 39.34\%$

3. TBA

3.9 Normal or Gaussian Distribution

1. Suppose your laptop battery will run for X hours before recharging, where $X \sim N(7, 1/2^2)$. Find
   (a) The probability it will run more than 9 hours?
   (b) At least 8 hours?
   (c) Between 6 and 7 hours?

   Solution:

   (a) $\Pr(X \geq 9) =$ `1 - pnorm(9,7,1/2)` $= 0.32\%\%$

   (b) $\Pr(X \leq 8) =$ `pnorm(8,7,1/2)` $= 97.72\%$

   (c) $\Pr(6 \leq X \leq 7) =$ `pnorm(7,7,1/2) - pnorm(6,7,1/2)` $= 47.72\%$

2. TBA

3.9.1 Standard Normal Distribution

1. What is the probability a Gaussian random variable is *further than* 1, 2, 3, 4, 5 or 6 standard deviations of its mean? *(Hint: Use the R function* `pnorm(q,mu,sd)`*.)*

   Solution:

   ```
   > ans = 2 * pnorm( seq(-1,-6,-1) )   # Trick: double left side
   > signif(ans,3)
   > 3.17e-01 4.55e-02 2.70e-03 6.33e-05 5.73e-07 1.97e-09
   ```

   Note:

   $$\begin{aligned} &P(-1 < Z < 1) = 0.6827 \\ &P(-2 < Z < 2) = 0.9545 \quad \text{worth memorizing} \\ &P(-3 < Z < 3) = 0.9973 \\ &P(-4 < Z < 4) = 0.9999 \quad \text{etc.} \end{aligned}$$

2. TBA

3.9.2 Sums of Independent Normal Random Variables

1. TBA

3.9.3 Normal Approximation to the Poisson Distribution

1. Plot the PMF of a Poisson PMF with $m = 36$, and overlay the approximating Gaussian PDF given by Eqn (3.44). Explain why these curve are close (without having to rescale the vertical scales) simply knowing that the former sums to 1 and latter integrates to 1. (Note the R function `dnorm(x,mu,sd)` computes the Gaussian/Normal PDF.)

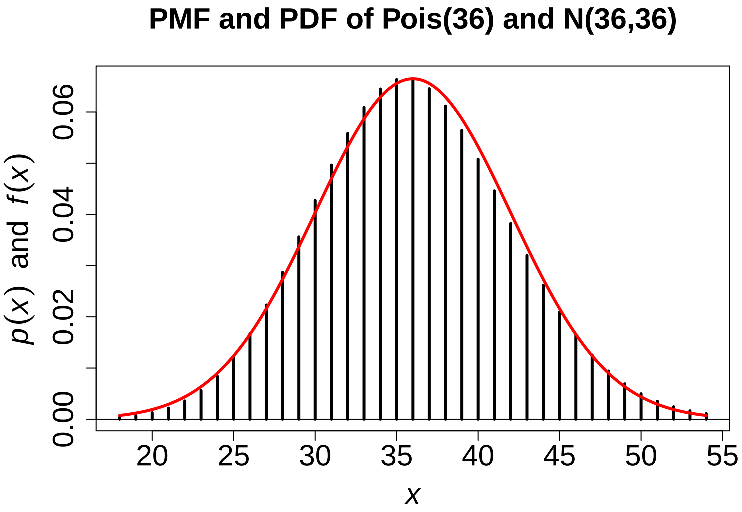
   Solution:

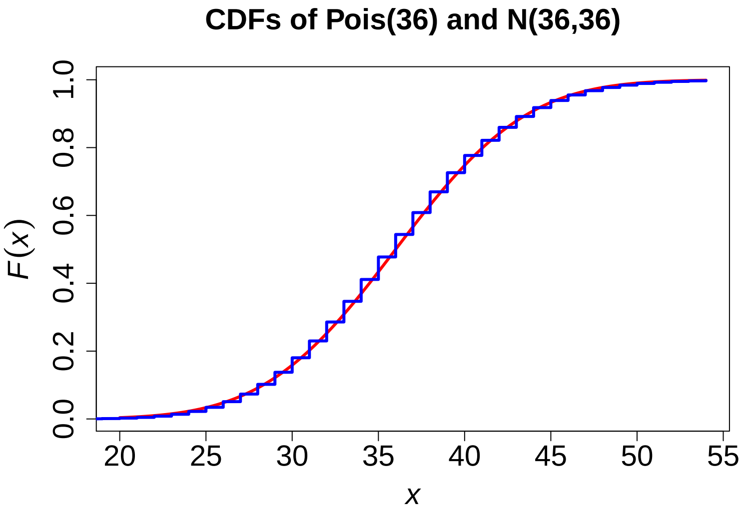
   For a $Pois(36)$, $\mu = \sigma^2 = 36$. So we compare to the $N(36, 36)$ PDF.

   Why are the curves close? The discrete PMF gives probabilities for $x = 0, 1, 2, ...$. The normal PDF is defined continuously there. But its curve has area over the interval $(x - 1/2, x + 1/2)$ approximately given by

$$\int_{s=x-1/2}^{s=x+1/2} \phi(s|\mu, \sigma)\, ds \approx \phi(x|\mu, \sigma) \cdot 1$$

   by the rectangle rule. Hence, the normal curve should be close to the discrete probabilities! Can do this both for the PMF/PDF and CDFs (not asked for).





```
> pmf = dpois( 18:55, 36 ); plot( 18:55, pmf, type="h" )
> xx = seq(18,55,.01); lines( xx, dnorm(xx,36,6), col=2 )
```

Note: $N(36, 6^2)$ versus `dnorm(xx,36,6)`

2. TBA

3.10 Problems

1. Throw 3 dice and label the results X_1, X_2, and X_3. What is the PMF of the total number of pips, $X = X_1 + X_2 + X_3$?

   Solution: There are $6 \times 6 \times 6$ or 218 equally likely outcomes such as the triple $(1, 1, 1)$. There is one way for the total number of pips to be 3, so the probability is $1/218$. Similar for a total of 18 (all 6's). The 3 triples $(2, 1, 1)$, $(1, 2, 1)$, and $(1, 1, 2)$ give a total of 4. Counting the others "by hand," we have:

3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	3	6	10	15	21	25	27	27	25	21	15	10	6	3	1

Table 3.1: Number of equally likely ways to see a total $\in [3, 18]$

2. Find the skewness and kurtosis coefficients in terms of μ and the non-central moments μ'_2, μ'_3, and μ'_4.

   Solution: Using the fact that the expectation of a sum is the sum of the expectations:

$$\begin{aligned}
\text{skew} &= E[(X-\mu)^3] = E[X^3 - 3X^2\mu + 3X\mu^2 - \mu^3] \\
&= E[X^3] - 3\mu E[X^2] + 3\mu^2 E[X] - E[\mu^3] \\
&= \mu'_3 - 3\mu\mu'_2 + 3\mu^2\mu - \mu^3 \\
&= \mu'_3 - 3\mu\mu'_2 + 2\mu^3\,;
\end{aligned}$$

$$\begin{aligned}
\text{kurt} &= E[(X-\mu)^4] = E[X^4 - 4X^3\mu + 6X^2\mu^2 - 4X\mu^3 + \mu^4] \\
&= \mu'_4 - 4\mu'_3\mu + 6\mu'_2\mu^2 - 4\mu\mu^3 + \mu^4 \\
&= \mu'_4 - 4\mu'_3\mu + 6\mu'_2\mu^2 - 3\mu^4\,.
\end{aligned}$$

3. If $X \sim NegBinom(r, p)$, find its MGF, and its mean and variance.

   Solution: The moment generating function and moments given in Appendix B may be confirmed by executing the Mathematica commands

```
fx = Binomial[r + x - 1, x] p^r (1 - p)^x

mgf = Sum[Exp[t x] fx, {x, 0, Infinity},
Assumptions -> {Element[r, Integers], p > 0, p < 1}]

m1 = D[mgf,t] /.  t -> 0
m2 = D[mgf,t,2] /.  t -> 0
var = Factor[ m2 - m1^2 ]
```

4. Use Mathematica to find the Taylor Series approximation to the difference of the exact MGF in Equation (3.52) and the MGF of a standard Normal. (We used a Taylor Series in the exponent when deriving Equation (3.53).

   Solution: The 2 moment generating functions are (in Mathematica):

```
mgf1 = Exp[ m ( Exp[t/Sqrt[m]] - 1 ) - Sqrt[m] t ]
mgf2 = Exp[ t^2 /2 ]
Series[ mgf1-mgf2, {t,0,4} ]
```

   which equals

$$\frac{t^3}{6\sqrt{m}} + \frac{t^4}{24m} + \cdots .$$

   Again, the two moment generating functions are equal as $m \to \infty$.

5. Use the MGF technique to prove the result in Equation (3.50). You should use the results from Section 5.4.1.

   Solution: Starting with the fact that the MGF of a $N(\mu_i, \sigma_i^2)$ r.v. is

$$MGF_{X_i}(t) = \exp(\mu_i t + \tfrac{1}{2}\sigma_i^2 t^2)\,,$$

the MGF of $Y = \sum a_i X_i$ may be derived and simplified to obtain

$$\begin{aligned}
MGF_Y(t) &= E[e^{tY}] \\
&= E[e^{t(a_1X_1 + a_2X_2 + \cdots a_nX_n)}] \\
&= E[e^{a_1tX_1 + a_2tX_2 + \cdots a_ntX_n)}] \\
&= E\left[\prod_{i=1}^{n} e^{a_itX_i}\right] \\
&= \prod_{i=1}^{n} E[e^{(a_it)X_i}] \\
&= \prod_{i=1}^{n} \exp\left[\mu_i(a_it) + \tfrac{1}{2}\sigma_i^2(a_it)^2\right] \\
&= \exp\left[t\sum a_i\mu_i + \tfrac{1}{2}t^2\sum a_i^2\sigma_i^2\right] ,
\end{aligned}$$

which confirms equation (3.50). Note that we used the fact that the expectation of the product of independent r.v.'s is the product of the expectations, something discussed in detail in Section 5.4.1.

Chapter 4

Bivariate Random Variables, Transformations, and Simulations

4.1 Bivariate Continuous Random Variables

1. Let $f(x, y)$ be a uniform PDF over the region bounded by straight lines connecting the 5 points $(0, 0)$, $(4, 0)$, $(4, 1)$, $(2, 1)$, back to $(0, 0)$. *Remark: These problems cover all of this Section.*

   (a) What is the PDF $f(x, y)$?

   (b) Find the marginal pdf of X and its mean and variance?

   (c) Same for Y?

   (d) What is $m(x) = E[Y|X = x]$? Sketch it over the support of (X, Y).
   (Hint: This will be defined in a piecewise fashion.)

   (e) What is conditional variance of $m(x)$? Compare it to the unconditional variance of Y? Sketch $m(x)$ and add the 2 curves that are the conditional mean plus/minus one conditional standard deviation.

   (f) What is the covariance of X and Y? The correlation?

   Solution:

   (a) The bivariate density is equally likely (uniform) over the trapezoid shown in the figure. Its area is 3, so the density is $f(x, y) = 1/3$

there, and zero, elsewhere.

(b) The marginal PDF of X is found in two intervals, $(0, 2)$ and $(2, 4)$ (why)? See figures below.

$$f_X(x) = \int_y f_{X,Y}(x,y)\,dy = \begin{cases} \int_{y=0}^{x/2} \frac{1}{3}\,dy = \frac{x}{6} & 0 \le x \le 2 \\ \int_{y=0}^{1} \frac{1}{3}\,dy = \frac{1}{3} & 2 \le x \le 4 \end{cases}$$

$$\begin{aligned} E[X] &= \int_{x=0}^{4} x\, f_X(x)\,dx \\ &= \int_{x=0}^{2} x \cdot \frac{x}{6}\,dx + \int_{x=2}^{4} x \cdot \frac{1}{3}\,dx \\ &= \frac{x^3}{18}\Big|_{x=0}^{x=2} + \frac{x^2}{6}\Big|_{x=2}^{x=4} \\ &= \left[\frac{2^3}{18} - 0\right] + \left[\frac{4^2}{6} - \frac{2^2}{6}\right] \\ &= \frac{4}{9} + 2 = 2.444\,. \end{aligned}$$

$$\begin{aligned} E[X^2] &= \int_{x=0}^{4} x^2\, f_X(x)\,dx \\ &= \int_{x=0}^{2} x^2 \cdot \frac{x}{6}\,dx + \int_{x=2}^{4} x^2 \cdot \frac{1}{3}\,dx \\ &= \frac{x^4}{24}\Big|_{x=0}^{x=2} + \frac{x^3}{9}\Big|_{x=2}^{x=4} \\ &= \left[\frac{2^4}{24} - 0\right] + \left[\frac{4^3}{9} - \frac{2^3}{9}\right] \\ &= \frac{2}{3} + \frac{64-8}{9} = 6.889\,. \quad \text{hence;} \end{aligned}$$

$$Var[X] = E[X^2] - \mu^2 = 6.889 - 2.444^2 = 0.914\,.$$

(c) The marginal of Y is easier as we'll see. Note the lower limit is given by the line $y = x/2$ or $x = 2y$; therefore,

$$\begin{aligned} f_Y(y) &= \int_x f_{X,Y}(x,y)\,dx \\ &= \int_{x=2y}^{4} \frac{1}{3}\,dx = \frac{4-2y}{3}\,. \end{aligned}$$

$$\begin{aligned} E[Y] &= \int_y y\, f_Y(y)\,dy \\ &= \int_{y=0}^{y=1} y\,\frac{4-2y}{3}\,dy \\ &= \frac{4\cdot y^2}{3\cdot 2}\bigg|_{y=0}^{y=1} - \frac{2\cdot y^3}{3\cdot 3}\bigg|_{y=0}^{y=1} \\ &= \frac{2}{3} - \frac{2}{9} = \frac{4}{9} = 0.444 \end{aligned}$$

$$\begin{aligned} E[Y^2] &= \int_y y^2\, f_Y(y)\,dy \\ &= \int_{y=0}^{y=1} y^2\,\frac{4-2y}{3}\,dy \\ &= \frac{4\cdot y^3}{3\cdot 3}\bigg|_{y=0}^{y=1} - \frac{2\cdot y^4}{3\cdot 4}\bigg|_{y=0}^{y=1} \\ &= \frac{4}{9} - \frac{1}{6} = \frac{8-3}{18} = 0.278 \end{aligned}$$

$$Var[Y] = E[Y^2] - \mu^2 = 0.080\,.$$

(d)

$$f_{Y|X=x}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \begin{cases} \frac{2}{x}\, I_{[0,x/2]}(y) & 0 \le x \le 2 \\ 1\cdot I_{[0,1]}(y) & 2 \le x \le 4\,. \end{cases} \quad ; \quad \text{hence,}$$

$$E[Y|X=x] = \begin{cases} \int_{y=0}^{y=x/2} y\,\frac{2}{x}\,dy = \frac{x}{4} & 0 \le x \le 2 \\ \int_{y=0}^{y=1} y\,1\,dy = \frac{1}{2} & 2 \le x \le 4 \end{cases}$$

(e)

$$E[Y^2|X=x] = \begin{cases} \int_{y=0}^{y=x/2} y^2\, \frac{2}{x}\, dy = \frac{x^2}{12} & 0 \le x \le 2 \\ \int_{y=0}^{y=1} y^2\, 1\, dy = \frac{1}{3} & 2 \le x \le 4 \end{cases} \qquad \text{hence,}$$

$$\begin{aligned} Var[Y|X=x] &= E[Y^2|X=x] - \Big(E[Y|X=x]\Big)^2 \\ &= \begin{cases} \frac{x^2}{12} - \left(\frac{x}{4}\right)^2 = 0.0208\, x^2 & 0 \le x \le 2 \\ \frac{1}{3} - \left(\frac{1}{2}\right)^2 = 0.0833 & 2 \le x \le 4\,. \end{cases} \end{aligned}$$

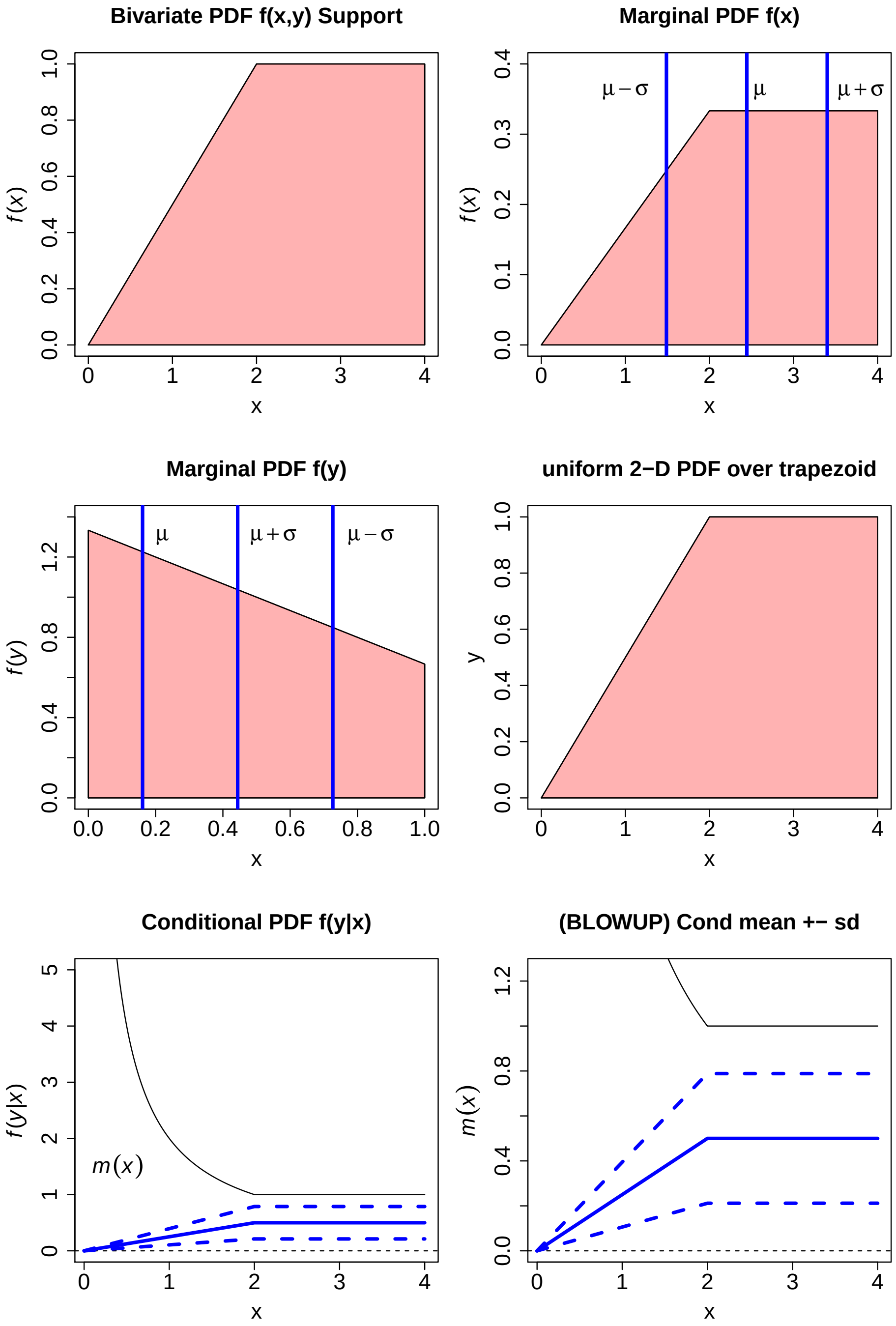
(f) Choose the order of integration carefully:

$$\begin{aligned} E[XY] &= \int_y \int_x xy\, f_{X,Y}(x,y)\, dx\, dy \\ &= \int_{y=0}^{y=1} \int_{x=2y}^{x=4} xy\, \frac{1}{3}\, dx\, dy \\ &= \int_{y=0}^{y=1} y \left[\int_{x=2y}^{x=4} x\, \frac{1}{3}\, dx \right] dy \\ &= \int_{y=0}^{y=1} y \left[\frac{x^2}{6} \Big|_{x=2y}^{x=4} \right] dy \\ &= \int_{y=0}^{y=1} y \left[\frac{4^2 - (2y)^2}{6} \right] dy \\ &= \int_{y=0}^{y=1} y \left[\frac{8}{3} - \frac{2y^2}{3} \right] dy \\ &= \int_{y=0}^{y=1} \left[\frac{8}{3} y - \frac{2}{3} y^3 \right] dy \\ &= \frac{8}{3} \cdot \frac{1}{2} - \frac{2}{3} \cdot \frac{1}{4} = \frac{4}{3} - \frac{1}{6} \\ &= \frac{7}{6}\,. \end{aligned}$$

$$Cov(X,Y) = E[XY] - E[X] \cdot E[Y] = 0.0802\,.$$

$$Cor(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = 0.297\,.$$

Check all these with Mathematica! Which is easier?



---

2. TBA

4.1.1 Joint CDF and PDF Functions

1. TBA

4.1.2 Marginal PDF

1. TBA

4.1.3 Conditional Probability Density Function

1. TBA

4.1.4 Independence of Two Random Variables

1. TBA

4.1.5 Expectation, Correlation, and Regression

1. TBA

4.1.5.1 Covariance and Correlation

1. Suppose X is a discrete random variable that takes on the values $-2, -1, 1, 2$ with equal probabilities $p_x = 1/4$. If $Y = X^2$, compute the correlation between X and Y. *Hint: Compute the means and variances of X and Y, and finally their covariance.*

Solution: Computing the various moments of X and $Y = X^2$:

$$\begin{aligned}
\mu_X &= EX = \sum_x x\,p_x = 0 \quad \text{(by symmetry)}\\
\sigma_X^2 &= E(X-\mu_X)^2 = EX^2 = [(-2)^2 + (-1)^2 + 1^2 + 2^2]\cdot\frac{1}{4} = \frac{5}{2}\\
\mu_Y &= EY = EX^2 = \frac{5}{2}\\
\sigma_Y^2 &= E[(Y-\mu_Y)^2] = EY^2 - \mu_Y^2 = EX^4 - \left(\frac{5}{2}\right)^2\\
&= [(-2)^4 + (-1)^4 + 1^4 + 2^4]\cdot\frac{1}{4} - \frac{25}{4} = \frac{34}{4} - \frac{25}{4} = \frac{9}{4}\\
\gamma_{XY} &= E[(X-\mu_X)(Y-\mu_Y)] = EXY - \mu_X\mu_Y\\
&= [(-2)(-2)^2 + (-1)(-1)^2 + 1(1)^2 + 2(2)^2]\cdot\frac{1}{4} - 0\cdot\frac{5}{2} = 0\,.
\end{aligned}$$

Therefore, the correlation is also 0. Note that $EXY = EX^3$, which is 0 by symmetry.

2. Repeat the previous problem but with $Y = X^3$.

Solution: Computing the various moments of X and $Y = X^3$:

$$\begin{aligned}
\mu_X &= EX = \sum_x x\,p_x = 0 \quad \text{(by symmetry)}\\
\sigma_X^2 &= E(X-\mu_X)^2 = EX^2 = [(-2)^2 + (-1)^2 + 1^2 + 2^2]\cdot\frac{1}{4} = \frac{5}{2}\\
\mu_Y &= EY = EX^3 = [(-2)^3 + (-1)^3 + 1^3 + 2^3]\cdot\frac{1}{4} = 0\\
\sigma_Y^2 &= E[(Y-\mu_Y)^2] = EY^2 - \mu_Y^2 = EX^6\\
&= [(-2)^6 + (-1)^6 + 1^6 + 2^6]\cdot\frac{1}{4} = \frac{130}{4} = \frac{65}{2}\\
\gamma_{XY} &= E[(X-\mu_X)(Y-\mu_Y)] = EXY - \mu_X\mu_Y = EX^4\\
&= [(-2)^4 + (-1)^4 + (1)^4 + (2)^4]\cdot\frac{1}{4} = \frac{34}{4} = \frac{17}{2}\\
\rho_{XY} &= \frac{\gamma}{\sigma_X\sigma_Y} = \frac{\frac{17}{2}}{\sqrt{\frac{5}{2}\cdot\frac{65}{2}}} = \frac{17}{5\sqrt{13}}\,,
\end{aligned}$$

or the correlation is 0.943. Note this is almost +1 (plot (x,y)?).

---

3. TBA

4.1.5.2 Regression Function

1. TBA

4.1.6 Independence of n Random Variables

1. TBA

4.1.7 Bivariate Normal PDF

1. TBA

4.1.8 Correlation, Independence, and Confounding Variables

1. TBA

4.2 Change of Variables

1. TBA

4.2.1 Examples: Two Uniform Transformations

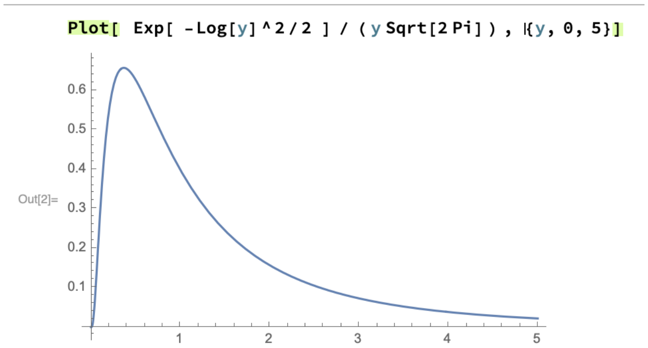
1. TBA

4.2.2 One-Dimensional Transformations

1. Let X be a standard Gaussian r.v. and let $Y = exp(X)$ be the transformed r.v. The PDF of Y is called the **log-normal**. Using the change-of-variables technique, find the formula for $f_Y(y)$? Plot the PDF for $0 \leq y \leq 5$.

Solution: The transformations are

$$
\begin{aligned}
y &= y(x) = \exp(x) && \text{for } x \in \Re^1 \text{ and } y > 0 \\
x &= x(y) = \log(y) && \text{is the inverse transformation and} \\
x'(y) &= \frac{1}{y} > 0 && \text{is the Jacobian; now} \\
f_X(x) &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) && \text{is the PDF of } X\text{; hence} \\
f_Y(y) &= f_X\big[x(y)\big] \cdot \big|J\big| \\
&= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x(y)^2\right) \cdot \left|\frac{1}{y}\right| \\
&= \frac{1}{y\sqrt{2\pi}} \exp\left(-\frac{1}{2}\log(y)^2\right)
\end{aligned}
$$



```
Plot[ Exp[ -Log[y]^2/2 ] / ( y Sqrt[2 Pi] ) , {y, 0, 5}]
```

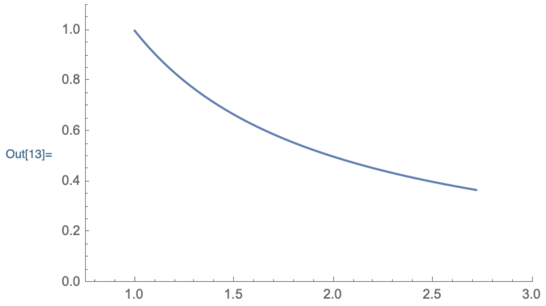
2. If X is distributed $Unif(0,1)$, what is the CDF of $Y = e^X$? Find the PDF and plot it.

Solution: Since the domain of X is $(0,1)$, the domain of $Y = e^X$ is

$(1, e)$. Now $f_X(x) = 1$ and $F_X(x) = x$ for $0 < x < 1$. Hence,

$$\begin{aligned} F_Y(y) &= \Pr(Y \le y) \\ &= \Pr(e^X \le y) \\ &= \Pr(X \le \log y) \\ &= \log y\,. \qquad \text{Thus} \\ f_Y(y) &= \frac{1}{y} \quad \text{for } 1 < y < e\,. \end{aligned}$$

It is easy to check that $F_Y(1) = 0$ and $F_Y(e) = 1$. The PDF is

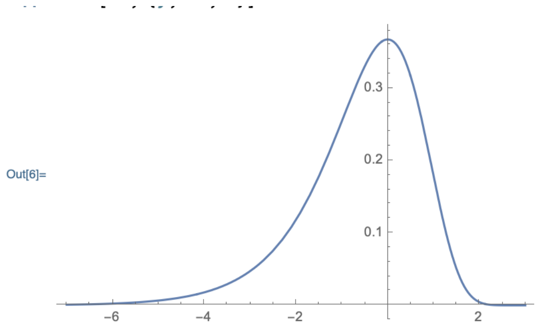


3. If X is distributed as negative exponential with paramter $\lambda = 1$, what is the CDF of $Y = \log X$? Find the PDF and plot it.

   Solution: Since the domain of X is $(0, \infty)$, the domain of $Y = \log X$ is the entire real line. Now $f_X(x) = e^{-x}$ and $F_X(x) = 1 - e^{-x}$ for $x > 0$. Hence,

$$\begin{aligned} F_Y(y) &= \Pr(Y \le y) \\ &= \Pr(\log X \le y) \\ &= \Pr(X \le e^y) \\ &= 1 - e^{-e^y}\,. \qquad \text{Thus} \\ f_Y(y) &= (-e^{-e^y}) \cdot (-e^y) = e^{y - e^y}\,, \end{aligned}$$

   which is a form of the extreme value distribution. It is easy to check that $F_Y(-\infty) = 0$ and $F_Y(\infty) = 1$. The PDF is



4. TBA

4.2.2.1 Example 1: Exponential PDF

1. TBA

4.2.2.2 Example 2: Cauchy PDF

1. TBA

4.2.2.3 Example 3: Chi-Squared PDF With 1 Degree of Freedom

1. TBA

4.2.3 Two-Dimensional Transformations

1. TBA

4.3 Simulations

1. TBA

4.3.1 Generating Uniform Pseudo-Random Numbers

1. TBA

4.3.1.1 Reproducibility

1. TBA

4.3.1.2 RANDU

1. TBA

4.3.2 Probability Integral Transformation

1. We want to generate samples from the $Beta(1,3)$ PDF

$$f(x) = 3(1-x)^2\,,$$

where the support is $0 < x < 1$, using the Probability Integral Transformation.

   (a) Show the CDF is $F(x) = 3x - 3x^2 + x^3$

   (b) Find the closed form solution for x in $u = F(x)$, given a random number $u \sim U(0,1)$. (Note: Obtaining $x = F^{-1}(u)$.)

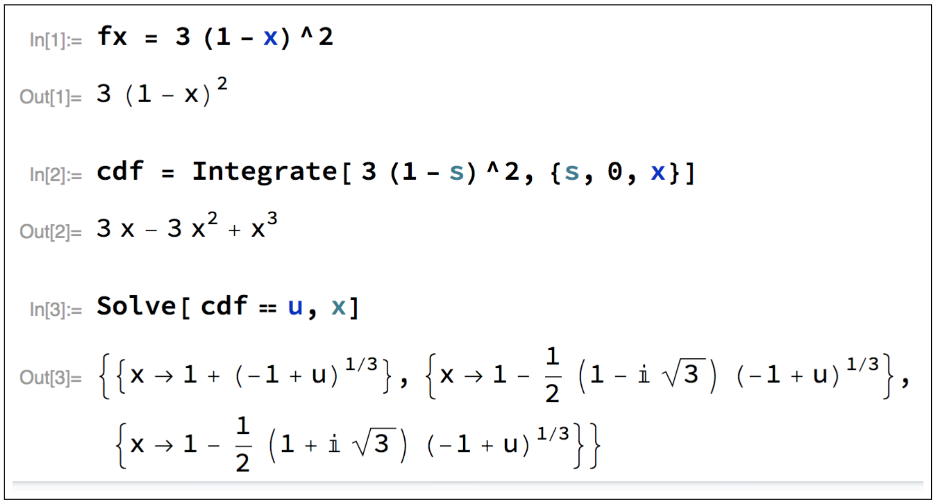
   (c) Generate a sample of size $n = 10^4$ and plot a histogram and superimpose the PDF. How does it look to you? **Hints:**

```
> hist( 1-(1-runif(1e4))**(1/3), prob=50, T, col=2 )
> xs = seq(0,1,.01); lines( xs, 3*(1-xs)**2, col=4, lwd=3
)
```

Solution: We use Mathematica for parts (a) and (b). Note that for part (a) we use a dummy argument, s, in the integral. We let Mathematica find the inverse roots. Only one is "real", the other two are imaginary. Mathematica gives the solution as

$$x = 1 + (-1+u)^{1/3} \qquad \text{which we re-write as}$$

$$\boxed{x = 1 - (1-u)^{1/3}}$$



```
In[1]:= fx = 3 (1 - x) ^2

Out[1]= 3 (1 - x)^2

In[2]:= cdf = Integrate[ 3 (1 - s) ^2, {s, 0, x}]

Out[2]= 3 x - 3 x^2 + x^3

In[3]:= Solve[ cdf == u, x]

Out[3]= {{x → 1 + (-1 + u)^(1/3)}, {x → 1 - 1/2 (1 - ⅈ √3) (-1 + u)^(1/3)},
         {x → 1 - 1/2 (1 + ⅈ √3) (-1 + u)^(1/3)}}
```

Figure 4.1: Mathematica Code

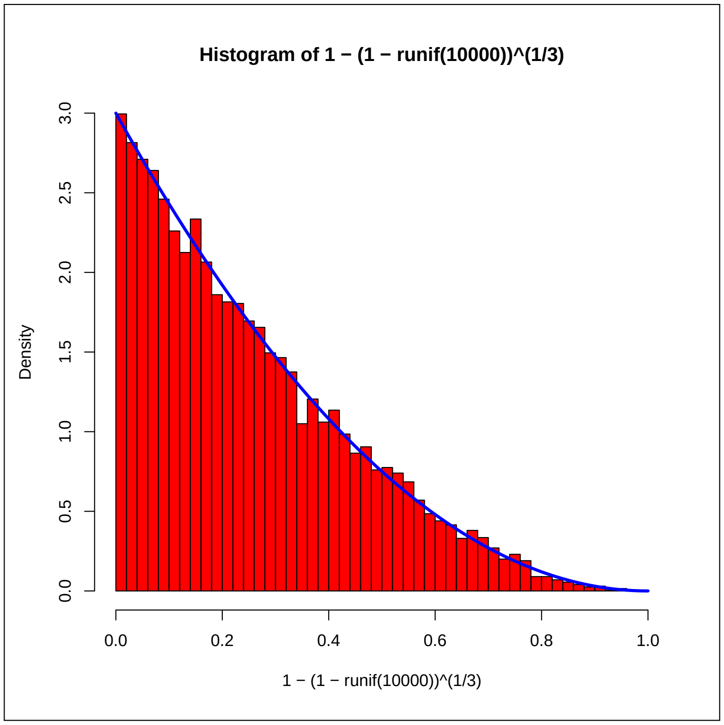


Figure 4.2: Histogram and Beta PDF

2. TBA

4.3.3 Event-Driven Simulation

1. TBA

4.4 Problems

1. An alternative and informal derivation of Equation (4.5) may be illuminating. As we did in Figure (3.5), we write

$$\begin{aligned} P(X \approx x) &= P\Big(X \in \big(x - \tfrac{\Delta}{2}, x + \tfrac{\Delta}{2}\big)\Big) \\ &= P\Big(X \in \big(x - \tfrac{\Delta}{2}, x + \tfrac{\Delta}{2}\big), Y \in \big(-\infty, \infty\big)\Big) \\ &= \int_{y=\infty}^{\infty} \left[\int_{s=x-\Delta/2}^{x+\Delta/2} f_{X,Y}(s,y)\, ds \right] dy \\ &\approx \int_{y=\infty}^{\infty} \big[f_{X,Y}(x,y) \cdot \Delta \big]\, dy\,. \end{aligned}$$

Since $P(X \approx x) \approx \Delta \cdot f_X(x)$, verify this sequence of equations and show that Equation (4.5) follows. Draw the relevant bivariate event.

Solution: Continuing,

$$\begin{aligned} P(X \approx x) &\approx \Delta \cdot f_X(x) \\ &= P\Big(X \in \big(x - \tfrac{\Delta}{2}, x + \tfrac{\Delta}{2}\big)\Big) \\ &= P\Big(X \in \big(x - \tfrac{\Delta}{2}, x + \tfrac{\Delta}{2}\big), Y \in \big(-\infty, \infty\big)\Big) \\ &= \int_{y=\infty}^{\infty} \left[\int_{s=x-\Delta/2}^{x+\Delta/2} f_{X,Y}(s,y)\, ds \right] dy \\ &\approx \int_{y=\infty}^{\infty} \big[f_{X,Y}(x,y) \cdot \Delta \big]\, dy \\ &\approx \Delta \cdot \int_{y=\infty}^{\infty} \big[f_{X,Y}(x,y) \big]\, dy\,. \end{aligned}$$

Equating the first and last lines gives Equation (4.5). Note we "integrate out" the second variable, y.

---

2. Show that the sequence of pseudo-random variables $\{(x_n, x_{n+1}, x_{n+2}),\ n = 1, 2, \dots\}$ defined in Equation (4.32) satisfy the equation
$$x_{n+2} = 6\,x_{n+1} - 9\,x_n \qquad \bmod 2^{31}\,;$$
hence, the points (y_n, y_{n+1}, y_{n+2}), where $y_n = x_n/2^{31}$, fall on 15 planes in $\mathbb{R}^3$.

Solution: We first observe that $65{,}539 = 2^{16}+3$; hence, (all $\bmod\ 2^{31}$)
$$\begin{aligned}
x_{n+1} &= 65{,}539x_n\\
x_{n+2} &= 65{,}539x_{n+1} = 65{,}539^2x_n = (2^{16}+3)^2x_n\\
x_{n+2} &= (2^{32}+6\cdot 2^{16}+9)x_n\\
x_{n+2} &= (0+6\cdot 2^{16}+9)x_n \qquad \text{since } 2^{32} \bmod 2^{31} = 0\\
x_{n+2} &= (6\cdot(2^{16}+3-3)+9)x_n\\
x_{n+2} &= (6\cdot(2^{16}+3)-18+9)x_n\\
x_{n+2} &= 6\cdot(2^{16}+3)x_n - 9\cdot x_n\\
&\boxed{x_{n+2} = 6\cdot x_{n+1} - 9\cdot x_n\,,}
\end{aligned}$$

which is the equation for a plane in $\Re^3$. In fact, with the integer coefficients $(1, 6, 9)$, there are only 15 parallel planes.

3. Use the bivariate change-of-variables approach to find the PDF of a bivariate normal, where the r.v.'s are **not** in standard form. Conclude the correlation coefficient is unchanged. *Hint:* You know $f_{X,Y}(x,y)$ where X and Y are in standard form. Define two new r.v.'s, $U = \mu_x + \sigma_x X$ and $V = \mu_y + \sigma_y Y$. Note that $U \sim N(\mu_X, \sigma_X^2)$ and $V \sim N(\mu_Y, \sigma_Y^2)$. Find $g_{U,V}(u,v)$, which is the general form of the bivariate Normal PDF with all 5 parameters.

Solution: The one-parameter bivariate normal pdf of (X,Y) is
$$f_{X,Y}(x,y|\rho) = \frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left[-\frac{x^2-2\rho xy+y^2}{2(1-\rho^2)}\right].$$

Given the new random variables (U,V),
$$\begin{aligned}
U &= \mu_x + \sigma_x X & \qquad V &= \mu_y + \sigma_y Y\\
X &= \frac{U-\mu_x}{\sigma_x} & \qquad Y &= \frac{V-\mu_y}{\sigma_y},
\end{aligned}$$

the Jacobian is

$$J = \begin{pmatrix} 1/\sigma_x & 0 \\ 0 & 1/\sigma_y \end{pmatrix} = \frac{1}{\sigma_x \sigma_y} > 0\,.$$

Hence, replacing x with $x(u,v)$ and y with $y(u,v)$, we have

$$g_{U,V}(u,v) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left[-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}\right],$$

and the correlation coefficient ρ is maintained.

Chapter 5

Approximations and Asymptotics

5.1 Why Do We Like Random Samples?

5.1.1 When $u(\boldsymbol{X})$ Takes a Product Form

5.1.2 When $u(\boldsymbol{X})$ Takes a Summation Form

5.2 Useful Inequalities

5.2.1 Markov's Inequality

5.2.2 Chebyshev's Inequality

1. Using Chebyschev's inequality to find n: Supposedly Asian Elephants are 200 pounds at birth. We think $\sigma = 5$ pounds. How large a sample n should we collect so that the probability $\bar{X}$ is within one (1) pound of μ is at least 99%?

Solution: Use Chebyshev's Inequality (5.11). Now we desire

$$\begin{aligned}
Prob(|\bar{X}_n - \mu| < 1) &\geq 0.99 && \text{equivalent to} \\
Prob(|\bar{X}_n - \mu| > 1) &\leq 0.01 && \text{compare to (5.11)} \\
\boxed{Prob\left(|\bar{X}_n - \mu| > \frac{k \cdot 5}{\sqrt{n}}\right) \leq \frac{1}{k^2}} & && \text{so equating things} \\
\frac{k \cdot 5}{\sqrt{n}} &= 1 && \text{or} \\
k^2 &= \frac{n}{25} && \text{next equate to desire error 0.01} \\
0.01 &= \frac{1}{k^2} = \frac{25}{n} && \text{with the result} \\
& \boxed{n = \frac{25}{0.01} = 2{,}500\,.}
\end{aligned}$$

That's a lot of elephants!

2. TBA

5.2.3 Jensen's Inequality

5.2.4 Cauchy-Schwarz Inequality

5.3 Sequences of Random Variables

5.3.1 Weak Law of Large Numbers

5.3.2 Consistency of the Sample Variance

5.3.3 Relationships Among the Modes of Convergence

5.3.3.1 Proof of Result (5.21)

5.3.3.2 Proof of Result (5.22)

5.4 Central Limit Theorem

5.4.1 Moment Generating Function for Sums

1. The MGF technique and the CLT:

   (a) Show by the MGF technique that the random variable $Y = \sum_{i=1}^n X_i$ will be exactly $\chi^2(n)$ if the X_i are a random sample from the $\chi^2(1)$ PDF.

   (b) May we conclude the PDF $\chi^2(n)$ is approximately Gaussian?

   Solution:

   (a) The MGF of a general $\chi^2(p)$ PDF is

   $$M_{\chi^2(p)}(t) = \left(\frac{1}{1-2t}\right)^{p/2}$$

   So the MGF of a general $\chi^2(1)$ PDF is

   $$M_{\chi^2(1)}(t) = \left(\frac{1}{1-2t}\right)^{1/2}$$

   For a sum $Y = \sum_{i=1}^n X_i$,

   $$\begin{aligned} M_Y(t) &= [M_X(t)]^n \\ &= \left[\left(\frac{1}{1-2t}\right)^{1/2}\right]^n \\ &= \left(\frac{1}{1-2t}\right)^{n/2} \\ &\sim \chi^2(n)\,, \qquad \text{as was to be shown} \end{aligned}$$

   (b) Yes, the CLT, an i.i.d. sum of random variables converges to a Gaussian r.v. $N(n, 2n)$.

2. **Analyzing the Average of a Normal Random Sample**

   - Use the moment generating function technique to find the exact PDF of the random variable $\bar{X}$ for a $N(\mu, \sigma^2)$ random sample.

- First, use Mathematica to compute the MGF of a $N(\mu, \sigma^2)$ r.v.:

$$\begin{aligned} M_X(t) &= E[e^{tX}] \\ &= \exp\left(\mu\, t + \frac{1}{2}\sigma^2\, t^2\right) \,. \end{aligned}$$

- Next use the MGF approach to find the PDF of $\bar{X}$. *Hint:*

$$\begin{aligned} M_{\bar{X}}(t) &= E[e^{t\bar{X}}] \\ &= E[e^{(t/n)(X_1+X_2+\cdots+X_n)}] \\ &= E\prod_{\ell=1}^{n} e^{(t/n)X_\ell} = \prod_{\ell=1}^{n} E[e^{(t/n)X_\ell}] = \prod_{\ell=1}^{n} M_{X_\ell}(t/n) \\ &= [M_X(t/n)]^n \,. \end{aligned}$$

Solution:

- Continuing, we now use the MGF approach to find the PDF of $\bar{X}$:

$$\begin{aligned} M_{\bar{X}}(t) &= E[e^{t\bar{X}}] = E[e^{t/n)(X_1+X_2+\cdots+X_n)}] \\ &= E\prod_{\ell=1}^{n} e^{(t/n)X_\ell} = \prod_{\ell=1}^{n} E[e^{(t/n)X_\ell}] = \prod_{\ell=1}^{n} M_{X_\ell}(t/n) \\ &= [M_X(t/n)]^n \\ &= \left[\exp\left(\mu(t/n) + \tfrac{1}{2}\sigma^2(t/n)^2\right)\right]^n \\ &= \exp\left(\mu t + \tfrac{1}{2}(\sigma^2/n)t^2\right) \\ &\sim N\left(\mu, \frac{\sigma^2}{n}\right) \,. \end{aligned}$$

- Thus the sample mean of a Normal random sample is exactly a Normal random variable for which

$$E\bar{X} = EX = \mu \quad \text{and} \quad \text{var}\bar{X} = \frac{\text{var}X}{n} \,.$$

- Of course, these relationships between $\mu_{\bar{X}}$ and $\sigma^2_{\bar{X}}$ with μ_X and σ^2_X are true in general.

3. **Analyzing the Average of a Cauchy Random Sample**

- We wish to attempt to use the moment generating function technique to see if we can find the PDF of the random variable $\bar{X}$ for a Cauchy random sample. This cannot be done, as the Cauchy PDF, where

$$f_X(x) = \frac{1}{\pi(1+x^2)} \qquad x \in \mathbb{R}$$

does not have a well-defined MGF. Instead, we need to use the Fourier (rather than LaPlace) transform, which is called the Characteristic Function (CF) in statistics.

- First, use Mathematica to compute the CF of a Cauchy r.v.:

$$\begin{aligned} CF_X(t) &= E[e^{itX}] \qquad \text{where } i = \sqrt{-1} \\ &= \int_{-\infty}^{\infty} e^{itx} \frac{1}{\pi(1+x^2)}\, dx \\ &= e^{-|t|} \,. \end{aligned}$$

- Next use the CF approach to find the PDF of $\bar{X}$. *Hint:*

$$\begin{aligned} CF_{\bar{X}}(t) &= E[e^{it\bar{X}}] = E[e^{i(t/n)(X_1+X_2+\cdots+X_n)}] \\ &= E\prod_{\ell=1}^{n} e^{i(t/n)X_\ell} = \prod_{\ell=1}^{n} E[e^{i(t/n)X_\ell}] = \prod_{\ell=1}^{n} CF_{X_\ell}(t/n) \\ &= [CF_X(t/n)]^n \,. \end{aligned}$$

- **Bonus:** Use the CF approach to show that if $U \sim Unif(-\pi/2, \pi.2)$, then $X = \tan(U)$ is Cauchy.

Solution:

- Use Mathematica to compute the MGF

  `fx` $= 1/(1+x^2)/Pi$

  `Integrate[ Exp[I t x] fx, {x,-Infinity,Infinity} ]`

  and adding the constraint that the constant t is a real number

  `Assumptions -> Element[t,Reals]`

  we obtain,

$$\begin{aligned} CF_X(t) &= E[e^{itX}] \qquad \text{obtaining} \\ &= e^{-|t|} \qquad \text{for } t \in \mathbb{R} \,. \end{aligned}$$

- Now use the CF approach to find the PDF of $\bar{X}$:

$$\begin{aligned} CF_{\bar{X}}(t) &= E[e^{it\bar{X}}] = E[e^{i(t/n)(X_1+X_2+\cdots+X_n)}] \\ &= E\prod_{\ell=1}^{n} e^{i(t/n)X_\ell} = \prod_{\ell=1}^{n} E[e^{i(t/n)X_\ell}] = \prod_{\ell=1}^{n} CF_{X_\ell}(t/n) \\ &= [CF_X(t/n)]^n \\ &= \left[e^{-\left|(t/n)\right|}\right]^n \\ &= e^{-|t|}\,, \end{aligned}$$

which is the CF of a single Cauchy random variable.

- Thus unlike most random variables with finite variable for which

$$\mathrm{var}\bar{X} = \frac{\mathrm{var}X}{n}\,,$$

the sample mean of a Cauchy random sample shows no central tendency. Of course, no moments of a Cauchy random variable exist (i.e., are finite). That the sample mean is exactly another standard Cauchy r.v. is a nice and memorable finding.

- **Bonus Problem:** We find the CF of X by

$$\begin{aligned} CF_X(t) &= E[e^{itX}] = E[e^{it\tan(U)}] \\ &= \int_{-\pi/2}^{\pi/2} e^{it\tan(u)}\,\frac{1}{\pi}\,du \\ &= e^{-|t|}\,, \end{aligned}$$

which is the CF of the Cauchy PDF.

5.4.2 Standardizing the Sum S_n

1. An old way of generating (approximately) $N(0,1)$ samples was to use the formula

$$Y = \sum_{i=1}^{12} U_i - 6\,.$$

(a) Why should this work if $U_i \sim U(0,1)$?
Hints: Compute the moments of Y *and invoke the CLT.*

(b) Try it 1000 times and plot a histogram of the pseudo-random sample. **Hint:**

```
> y=NULL; for(i in 1:1000) {y=c(y,sum(runif(12))-6)}; hist(y,40,T,col=2)
```

Solution:

(a) Now the mean and variance of U_i are $1/2$ and $1/12$. From Section 5.4.2

$$E\,S_n = n \cdot \mu = 12 \cdot \frac{1}{2} = 6$$

$$Var\,S_n = n \cdot \sigma^2 = 12 \cdot \frac{1}{12} = 1$$

So the r.v. Y has mean 6-6=0 and variance = 1. By the CLT, Y is approximately $N(0,1)$.

(b) Looks pretty normal???

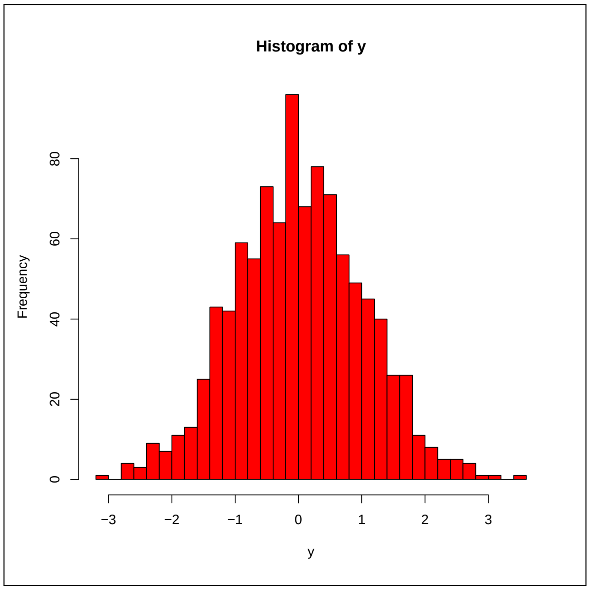


Figure 5.1: Histogram of $\sum_{i=1}^{12} U_i - 6$

---

2. The Central Limit Theorem says that for any random sample $\{X_i\}$,

$$Y_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\frac{X_i - \mu_X}{\sigma_X} \right) \xrightarrow{\mathscr{D}} N(0,1)\,.$$

If, in fact, the sample is itself Normal, i.e. $X_i \sim N(\mu_X, \sigma_X^2)$, use the MGF technique to show that Y_n is exactly $N(0,1)$ for all sample sizes n.

Solution: We give a long and short version of this fact. The short version takes three easy steps.

In general, if X_i is a random sample, then so are $X_i - \mu_X$, $(X_i - \mu_X)/\sigma_X$, and $(X_i - \mu_X)/(\sqrt{n}\sigma_X)$.

Recall that the MGF of the $X \sim N(\mu_X, \sigma_X^2)$ random variable is

$$MGF_X(t) = \exp\left(\mu_X t + \frac{1}{2}\sigma_X^2 t^2 \right)\,.$$

Hence, any linear function of X, say $V = a + bX$ is also Normal. This follows (as in class) since

$$\begin{aligned} M_V(t) = E[e^{tV}] = E[e^{t(a+bX)}] = E[e^{ta}e^{tbX}] &= e^{ta} \cdot E[e^{(tb)X}] \\ &= e^{ta} \cdot \exp\left(\mu_x(tb) + \tfrac{1}{2}\sigma_X(tb)^2 \right) \\ &= \exp\left((a + \mu_x b)t + \tfrac{1}{2}b^2\sigma_X^2 t^2 \right) \\ &\sim N\left(a + b\mu_X, b^2\sigma_X^2 \right)\,. \end{aligned}$$

We'll start with the shortcut solution. First we observe that

$$\frac{X_i - \mu_X}{\sigma_X} \sim N(0,1) \quad \text{call it, } Z_i\text{; so can write } Y_n \text{ as}$$

$$Y_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i\,; \quad \text{now the MGF technique shows}$$

$$\sum_{i=1}^{n} Z_i \sim N(0,n)\,, \quad \text{since } (\exp(t^2/2))^n = \exp(n \cdot t^2/2)\text{. Finally,}$$

$$\frac{1}{\sqrt{n}} \cdot N(0,n) \sim N\left(0, \frac{n}{\sqrt{n}^{\,2}}\right) = N(0,1)\,.$$

The long version computing the MGF of Y_n directly is given now. We compute the MGF of Y_n as (carefully)

$$\begin{aligned}
MGF_{Y_n}(t) &= E\left[e^{tY_n}\right] \\
&= E\left[\exp\left\{\frac{t}{\sqrt{n}} \sum_{i=1}^{n} \left(\frac{X_i - \mu_X)}{\sigma_X}\right)\right\}\right] \\
&= E\left[\exp\left\{\frac{t}{\sigma_X\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu_X)\right\}\right] \\
&= E\left[\exp\left\{\frac{t}{\sigma_X\sqrt{n}} \left(\sum_{i=1}^{n} X_i - n\mu_X\right)\right\}\right] \\
&= E\left[\exp\left\{\frac{t}{\sigma_X\sqrt{n}} \sum_{i=1}^{n} X_i - \frac{t}{\sigma_X\sqrt{n}} n\mu_X\right\}\right] \\
&= E\left[\exp\left\{\frac{t}{\sigma_X\sqrt{n}} \sum_{i=1}^{n} X_i\right\}\right] \cdot \exp\left\{\frac{-t\sqrt{n}\,\mu_X}{\sigma_X}\right\} \\
&= \exp\left\{\frac{-t\sqrt{n}\,\mu_X}{\sigma_X}\right\} \times \prod_{i=1}^{n} E\left[\exp\left\{\frac{t}{\sigma_X\sqrt{n}} X_i\right\}\right] \\
&= \exp\left\{\frac{-t\sqrt{n}\,\mu_X}{\sigma_X}\right\} \times \left(E\left[\exp\left\{\frac{t}{\sigma_X\sqrt{n}} X_i\right\}\right]\right)^n \qquad (5.1)
\end{aligned}$$

since the expectation of a constant is that constant, and the expectation of the product is the product of the expectation (of independent r.v.'s).

Now one of the expectations in Equation (5.1) is the MGF of a Normal random $N(\mu_X, \sigma_X^2)$ variable, but with t replaced by $t/(\sigma_X \sqrt{n}\,)$;

$$
\begin{aligned}
E\left[\exp\left\{\frac{t}{\sigma_X\sqrt{n}}\,X_i\right\}\right] &= \exp\left\{\left(\frac{t}{\sigma_X\sqrt{n}}\right)\,\mu_X + \frac{1}{2}\left(\frac{t}{\sigma_X\sqrt{n}}\right)^2\,\sigma_X^2\right\} \\
&= \exp\left\{\frac{t\,\mu_X}{\sigma_X\sqrt{n}} + \frac{1}{2}\frac{t^2\sigma_X^2}{\sigma_X^2\,n}\right\} \\
&= \exp\left\{\frac{t\,\mu_X}{\sigma_X\sqrt{n}}\right\} \times \exp\left\{\frac{1}{2}\frac{t^2}{n}\right\} \\
\text{and the } n^{th} \text{ power is} \qquad &= \exp\left\{\frac{t\,\mu_X\cdot n}{\sigma_X\sqrt{n}}\right\} \times \exp\left\{\frac{1}{2}\,\frac{t^2\cdot n}{n}\right\} \\
&= \exp\left\{\frac{t\sqrt{n}\,\mu_X}{\sigma_X}\right\} \times \exp\left\{\frac{1}{2}\,t^2\right\} \qquad (5.2)
\end{aligned}
$$

Substituting Equation (5.2) into Equation (5.1) gives

$$
\begin{aligned}
MGF_{Y_n)}(t) &= \exp\left\{\frac{-t\sqrt{n}\,\mu_X}{\sigma_X}\right\} \times \left(E\left[\exp\left\{\frac{t}{\sigma_X\sqrt{n}}X_i\right\}\right]\right)^n \\
&= \exp\left\{\frac{-t\sqrt{n}\,\mu_X}{\sigma_X}\right\} \times \exp\left\{\frac{t\sqrt{n}\,\mu_X}{\sigma_X}\right\} \times \exp\left\{\frac{1}{2}\,t^2\right\} \\
&= \exp\left\{\frac{1}{2}\,t^2\right\}\,,
\end{aligned}
$$

which is the MGF of a standard Normal r.v., as was to be shown.

3. **Where we derive a general formula for the first two moments of a linear combination of independent, but not identically distributed, random variables.** Given n independent r.v.'s X_i with means μ_i and variances σ_i^2, consider the random variable

$$
Y = \sum_{i=1}^{n} \left(a_i\,X_i + b_i\right),
$$

where a_i and b_i are known constants.

   (a) Find the mean of Y.

(b) Find the variance of Y.

Hint: Define $Y_i = X_i - \mu_i$ *and show* $Y - \mu_Y = \sum_i a_i\, Y_i$. *Then easy to compute the variance of* Y, *since the* Y_i'*s are independent with* $EY_i = 0$ *and* $EY_i^2 = \sigma_i^2$.

Solution:

(a) By the E[Sum] equals Sum of E's,

$$\begin{aligned} E[Y] &= \sum_{i=1}^{n} E\left(a_i\, X_i + b_i\right) \\ &= \sum_{i=1}^{n} \left(a_i\, EX_i + b_i\right) \\ &= \sum_{i=1}^{n} \left(a_i\, \mu_i + b_i\right) \\ &\boxed{\mu_Y = \sum_{i=1}^{n} a_i\, \mu_i + \sum_{i=1}^{n} b_i\,.} \end{aligned}$$

(b) Define

$Y_i = X_i - \mu_i\,;$ then $EY_i = 0$; $Var\, Y_i = \sigma_i^2$; and $E[Y_iY_j] = 0$ if $i \neq j$.

Next, we can re-write $Y - \mu_Y$ in a more convenient form, starting with

$$\begin{aligned} Y &= \sum_{i=1}^{n} \left(a_i\, X_i + b_i\right) \\ Y - \mu_Y &= \sum_{i=1}^{n} \left(a_i\, X_i + b_i\right) - \left[\sum_{i=1}^{n} a_i\, \mu_i + \sum_{i=1}^{n} b_i\right] \\ &= \sum_{i=1}^{n} a_i \left(X_i - \mu_i\right) + \sum_{i=1}^{n} b_i - \sum_{i=1}^{n} b_i \\ &= \sum_{i=1}^{n} a_i Y_i\,. \end{aligned}$$

Since $EY_i = 0$, $Var\,Y_i = \sigma_i^2$,

$$\begin{aligned} Var\,Y &= E(Y-\mu_Y)^2 = E\left[\sum_{i=1}^n a_i Y_i\right]^2 \\ &= E\left[\sum_{i=1}^n a_i^2 Y_i^2 + \sum_{i\neq j} a_i a_j Y_i Y_j\right] \\ &= \sum_{i=1}^n a_i^2 E\,Y_i^2 + \sum_{i\neq j} a_i a_j\, E\left[Y_i Y_j\right] \\ &= \sum_{i=1}^n a_i^2 \sigma_i^2 + \sum_{i\neq j} a_i a_j EY_i \cdot E[Y_j] \end{aligned}$$

$$\boxed{Var\,Y = \sum_{i=1}^n a_i^2 \sigma_i^2} \qquad \text{since } EY_i = 0.$$

Thus the variance of the weighted sum is the sum of the variances with squared weights.

4. **This problem is a generalization of the previous problem. Here, we derive the formula for the first two moments of a linear combination of dependent, and non-identically distributed, random variables.** The n r.v.'s X_i have means μ_i and variances σ_i^2, and the correlation coefficient between X_i and X_j is ρ_{ij}. Note that $cov(X_i, X_j) = \sigma_i \sigma_j \rho_{ij}$. Consider the random variable

$$Y = \sum_{i=1}^n \left(a_i\, X_i + b_i\right)\,, \quad \text{where } a_i \text{ and } b_i \text{ are known constants.}$$

   (a) Find the mean of Y.

   (b) Find the variance of Y.

   (c) What is the variance of Y when $n = 2$?

Hint: Let $Y_i = X_i - \mu_i$*; hence,* $Y - \mu_Y = \sum_i a_i\, Y_i$*. Then it is easier to compute the* $var(Y)$*, since the* Y_i'*s satisfy* $EY_i = 0$*,* $EY_i^2 = \sigma_i^2$*, and* $cov(Y_i, Y_j) = E[Y_i Y_j] = \sigma_i \sigma_j \rho_{ij}$ *as well.*

Solution:

(a) The fact that the random variables are not independent does not change the expectation of the sum. (The expectation of the sum is the sum of the expectations does not require a random sample.) Hence, the answer is the same as on the previous problem:

$$EY = \sum_{i=1}^{n} \left(a_i\mu_i + b_i\right) .$$

(b) Let $Y_i = X_i - \mu_i$; then

$$\begin{aligned} EY_i &= E[X_i - \mu_i] = E[X_i] - \mu_i = \mu_i - \mu_i = 0 \\ Var\,Y_i &= E\big[Y_i - \mu_{Y_i}\big]^2 = EY_i^2 = E\big[X_i - \mu_i\big]^2 = \sigma_i^2 \\ Cov(Y_i, Y_j) &= E\big[(Y_i - \mu_{Y_i})(Y_j - \mu_{Y_j})\big] = E\big[Y_iY_j\big] \\ &= E\big[(X_i - \mu_i)(X_j - \mu_j)\big] \\ &= Cov(X_i, X_j) . \end{aligned}$$

Continuing,

$$\begin{aligned} Cor(X_i, X_j) = \rho_{ij} &= E\left[\frac{X_i - \mu_i}{\sigma_i} \cdot \frac{X_j - \mu_j}{\sigma_j}\right] \\ &= E\left[\frac{Y_i}{\sigma_i} \cdot \frac{Y_j}{\sigma_j}\right] = \frac{1}{\sigma_i\sigma_j} E[Y_iY_j] ; \quad \text{hence,} \\ \boxed{E[Y_iY_j] = \sigma_i\sigma_j\rho_{ij}} & \quad \text{per the hint.} \end{aligned}$$

Now proceeding as in the previous problem:

$$
\begin{aligned}
Y - \mu_Y &= \sum_{i=1}^{n} a_i Y_i \\
Var\, Y = E\big[Y - \mu_y\big]^2 &= E\left[\sum_{i=1}^{n} a_i Y_i\right]^2 \\
&= E\left[\sum_{i=1}^{n} a_i^2 Y_i^2 + \sum_{i\neq j} a_i a_j Y_i Y_j\right] \\
&= \left[\sum_{i=1}^{n} a_i^2\, E\big(Y_i^2\big) + \sum_{i\neq j} a_i a_j E\big(Y_i Y_j\big)\right] \\
&= \left[\sum_{i=1}^{n} a_i^2\, \sigma_i^2 + \sum_{i\neq j} a_i a_j \sigma_i \sigma_j \rho_{ij}\right] .
\end{aligned}
$$

(c) For $n = 2$, the $Var\, Y$ written out is

$$\boxed{Var\, Y = a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + 2\, a_1 a_2\, \sigma_1\sigma_2\, \rho_{12}\, .}$$

5. Consider a random sample of n Bernoulli events, that is, $X_i \sim B(1, p)$.

(a) Show $X_i^2 = X_i$. Then show

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\big(X_i - \bar{X}\big)^2 = \frac{n}{n-1}\bar{X}\big(1 - \bar{X}\big)\, .$$

(b) By computing the first two non-central moments of $\bar{X}$, show that

$$E\, S^2 = \sigma_X^2 = p(1-p)\, , \qquad \text{i.e., } S^2 \text{ is unbiased for } \sigma_X^2.$$

Hint: First show that $E\bar{X} = \mu_{\bar{X}} = \mu_X = p$ *and* $E\bar{X}^2 = Var\,\bar{X} + (E\bar{X})^2$.

Solution:

(a) Since $X_i = 0$ or 1, $X_i^2 = X_i$ as well.

$$\begin{aligned} S^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \\ &= \frac{1}{n-1}\left[\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right] \\ &= \frac{1}{n-1}\left[\sum_{i=1}^{n} X_i - n\bar{X}^2\right] \\ &= \frac{1}{n-1}\left[n\bar{X} - n\bar{X}^2\right] \\ &= \frac{n}{n-1}\bar{X}\left(1-\bar{X}\right). \end{aligned}$$

(b) By linearity (or the expectation of the sum is the sum of the expectations),

$$\begin{aligned} ES^2 &= \frac{n}{n-1}\left[E\bar{X} - E\bar{X}^2\right] \\ &= \frac{n}{n-1}\left\{\mu_{\bar{X}} - \left[Var\,\bar{X} + \mu_{\bar{X}}^2\right]\right\} \\ &= \frac{n}{n-1}\left[p - \frac{p(1-p)}{n} - p^2\right] \\ &= \frac{n}{n-1}\left[p(1-p) - \frac{p(1-p)}{n}\right] \\ &= \frac{n}{n-1}\left[\frac{n-1}{n}\cdot p(1-p)\right] \\ &\boxed{ES^2 = p(1-p) = \sigma_X^2\,.} \qquad \text{i.e., unbiased} \end{aligned}$$

6. TBA

5.4.3 Proof of Central Limit Theorem

5.5 Delta Method and Variance-Stabilizing Transformations

5.6 Problems

1. Prove Equation (5.8) assuming $f(\boldsymbol{x}) = \prod_{i=1}^{n} f_{X_i}(x_i)$, i.e. $\{X_1, X_2, \ldots, X_n\}$ is an i.i.d. sample. Note: This is a stronger assumption than required.

   Solution: This is ...

2. Show Chebyshev's inequality can also be expressed as

$$P\left(|X - \mu_X| > c\right) \leq \frac{\sigma_X^2}{c^2}\,.$$

   Solution: This is ...

3. Show the delta method given in Equation (5.37) for a Poisson r.v. with $g(x) = \sqrt{x}$ gives the same answer as in Equation (5.36).

   Solution: This is ...

Chapter 6

Parameter Estimation

6.1 Desirable Properties of an Estimator

1. TBA

6.2 Moments of the Sample Mean and Variance

1. TBA

6.2.1 Theoretical Mean and Variance of the Sample Mean

1. TBA

6.2.2 Theoretical Mean of the Sample Variance

1. TBA

6.2.3 Theoretical Variance of the Sample Variance

1. TBA

6.3 Method of Moments (MoM)

1. TBA

6.4 Sufficient Statistics and Data Compression

1. For a random sample of size n from the Geometric PMF, $Geom(p)$:

   (a) Find the sufficient statistic, $T(\boldsymbol{X})$, for the parameter $\theta = p$.

   (b) Find a Method of Moments estimator for p.

   Solution:

   (a) Now $f(x) = pq^{x-1}$; hence,

$$\begin{aligned} L(p|\boldsymbol{X}) &= \prod_{i=1}^{n} pq^{X_i - 1} \\ &= p^n q^{\sum_{i=1}^{n} X_i - n} \\ &= p^n q^{-n} q^{\sum_{i=1}^{n} X_i}\,; \qquad \text{hence,} \end{aligned}$$

$$\boxed{T(\boldsymbol{X}) = \sum_{i=1}^{n} X_i\,, \qquad \text{is the sufficient statistic for } p.}$$

   (b) Now $EX = 1/p$; hence,

$$\begin{aligned} E\bar{X} &= \mu_{\bar{X}} = \mu_X = \frac{1}{p}\,. \qquad \text{Thus the MoM equation is} \\ \bar{X} &= \frac{1}{p} \qquad \text{or} \end{aligned}$$

$$\boxed{\hat{p}_{MoM} = \frac{1}{\bar{X}}\,.}$$

2. Consider the $Gamma(\alpha, \beta)$ PDF, which may also be found in the Appendix. What are the two sufficient statistics, $T_1(\boldsymbol{X})$ and $T_2(\boldsymbol{X})$, for the parameters α and β?

Solution: The Gamma PDF is given by

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

$$L(\alpha,\beta|\boldsymbol{x}) = \prod_{i=1}^{n} f(x_i)$$

$$= \Gamma(\alpha)^{-n}\beta^{-n\alpha}\left(\prod x_i\right)^{\alpha-1} e^{-\sum x_i/\beta}\,. \quad \text{Thus,}$$

$$\boxed{T_1(\boldsymbol{X}) = \prod X_i \quad \text{and}}$$

$$\boxed{T_2(\boldsymbol{X}) = \sum X_i \quad \text{are the sufficient statistics}}$$

Note that any invertible transformations of the sufficient statistics also work, for example,

$$\boxed{\tilde{T}_1(\boldsymbol{X}) = \sum \log X_i \quad \text{and}}$$

$$\boxed{\tilde{T}_2(\boldsymbol{X}) = \bar{X}\,.}$$

3. TBA

6.5 Bayesian Parameter Estimation

1. TBA

6.6 Maximum Likelihood Parameter Estimation

1. For a random sample of size n from the Geometric PMF, $Geom(p)$:
   (a) Find the MLE $\hat{p}_{MLE}$.
   (b) Show that it truly is a maximum.

Solution:

(a) From an earlier, $f(x) = pq^{x-1}$, $x = 1, 2, \ldots$, and

$$L(p|\boldsymbol{X}) = \prod_{i=1}^{n} pq^{X_i - 1} = p^n q^{\sum_{i=1}^{n} X_i - n} .$$

Switching to the log-likelihood:

$$\begin{aligned}
\ell(p|\boldsymbol{X}) &= n \log p + \left(\sum X_i - n \right) \log q \\
&= n \log p + (n\bar{X} - n) \log q \\
&= n \log p + n(\bar{X} - 1) \log(1 - p) \,;\text{hence,} \\
\frac{d\,\ell(p|\boldsymbol{X})}{dp} &= \frac{n}{p} - \frac{n(\bar{X} - 1)}{1 - p} \,, \qquad \text{which vanishes when} \\
0 &= n(1 - p) - p \cdot n(\bar{X} - 1) \\
&= n - np - np\bar{X} + np \\
&= n - np\bar{X} \,, \qquad \text{which gives}
\end{aligned}$$

$$\boxed{\hat{p} = \frac{1}{\bar{X}}} \qquad \text{(same as the MoM estimator)}$$

(b) Checking the sign of the second derivative at $p = \hat{p}$,

$$\begin{aligned}
\frac{d^2\,\ell(p|\boldsymbol{X})}{dp^2} &= \frac{-n}{p^2} - \frac{n(\bar{X} - 1)}{(1 - p)^2} = -n\bar{X}^2 - \frac{n(\bar{X} - 1)}{\left(1 - \frac{1}{\bar{X}}\right)^2} \\
&= -n\bar{X}^2 - \frac{n(\bar{X} - 1)}{\left(\frac{\bar{X} - 1}{\bar{X}}\right)^2} = -n\bar{X}^2 - \frac{\bar{X}^2 \cdot n(\bar{X} - 1)}{(\bar{X} - 1)^2} \\
&= -n\bar{X}^2 - \frac{n\bar{X}^2}{\bar{X} - 1} = \frac{-n\bar{X}^3 + n\bar{X}^2 - n\bar{X}^2}{\bar{X} - 1} \\
&= \frac{-n\bar{X}^3}{\bar{X} - 1} \,; \qquad \text{therefore,}
\end{aligned}$$

$$\boxed{\left. \frac{d^2\,\ell(p|\boldsymbol{X})}{dp^2} \right|_{p = \hat{p}} < 0 \qquad \text{assuming } \bar{X} > 1 \;\; \text{(it is, right?),}}$$

which demonstrates $\hat{p}$ is a maximizer.

2. Consider a random sample of size n from the $Unif(\theta, 1)$ PDF, where $\theta < 1$.

(a) What is the likelihood function $L(\theta \mid \boldsymbol{X})$?

(b) What is the MLE for θ?

Solution:

(a) Now $X \sim Unif(\theta, 1)$; hence,

$$f(x) = \begin{cases} \frac{1}{1-\theta} & \text{for } \theta \le x \le 1 \\ 0 & \text{elsewhere} \end{cases}$$

$$= \frac{1}{1-\theta}\, I\big(x \in [\theta, 1]\big) \qquad \text{thus the likelihood is}$$

$$L(\theta \mid \boldsymbol{X}) = \prod_{i=1}^{n} (1-\theta)^{-1}\, I\big(X_i \in [\theta, 1]\big)$$

$$= (1-\theta)^{-n}\, I\big(X_{(1)} \in [\theta, 1]\big)\,,$$

where $X_{(1)}$ is the 1^{st} order statistic (i.e., smallest value of the data).

(b) The likelihood is maximized by making the interval $(1-\theta)$ as narrow as possible, subject to the constraint that $\theta \le X_{(1)}$. Hence,

$$\boxed{\hat{\theta}_{MLE} = X_{(1)}\,.}$$

3. TBA

6.6.1 Relationship to Bayesian Parameter Estimation

1. TBA

6.6.2 Poisson MLE Example

1. TBA

6.6.3 Normal MLE Example

1. TBA

6.6.4 Uniform MLE Example

1. TBA

6.7 Information Inequalities and the Cramér-Rao Lower Bound

1. TBA

6.7.1 Score Function

1. TBA

6.7.2 Asymptotics of MLE

1. TBA

6.7.3 Minimum Variance of Unbiased Estimators

1. TBA

6.7.4 Examples

1. TBA

6.8 Problems

1. Given n independent r.v.'s X_i with means μ_i and variances σ_i^2, consider the random variable

$$Y = \sum_{i=1}^{n} \big(a_i\, X_i + b_i\big)\,,$$

where a_i and b_i are known constants. Show

$$EY = \sum_{i=1}^{n} (a_i \, \mu_i + b_i)$$
$$\operatorname{var} Y = \sum_{i=1}^{n} a_i^2 \sigma_i^2 \, .$$

Hint: Define $Y_i = X_i - \mu_i$ and show $Y - \mu_Y = \sum a_i Y_i$.

Solution: Following the hint, let $Y_i = X_i - \mu_i$, where $EX_i = \mu_i$. Clearly, the variance of X_i and Y_i are the same, σ_i^2. Then

$$\begin{aligned}
EY = \mu_Y &= \sum_i E[a_i X_i + b_i] \\
&= \sum_i \big(E[a_i X_i] + E[b_i] \big) \\
&= \sum_i \big(a_i E[X_i] + b_i \big) \\
&\boxed{EY = \sum_i \big(a_i \mu_i + b_i \big) \, . \quad \text{Next,}} \\
Y - \mu_y &= \sum_i \big(a_i X_i + b_i \big) - \sum_i \big(a_i \mu_i + b_i \big) \\
&= \sum_i a_i (X_i - \mu_i) \\
&= \sum_i a_i Y_i \, ; \quad \text{hence,} \\
\operatorname{var} Y &= \sum_I a_i^2 \mathrm{Var} Y_i \\
&\boxed{\operatorname{var} Y = \sum_I a_i^2 \sigma_i^2 \, .}
\end{aligned}$$

2. For a Normal random sample, show that $\operatorname{var} S^2 = 2\sigma_X^4/(n-1)$ using the fact that $(n-1)S^2/\sigma_X^2 \sim \chi^2(n-1)$; see Equation (6.9) and following.

Solution: Recall that the variance of a χ^2_p r.v. is $2p$. Hence,

$$\begin{aligned}
\operatorname{var}\left[\frac{(n-1)S^2}{\sigma_X^2}\right] &= 2(n-1)\\
\left[\frac{(n-1)}{\sigma_X^2}\right]^2 \operatorname{var} S^2 &= 2(n-1)\\
\operatorname{var} S^2 &= 2(n-1)\times\frac{\sigma_X^4}{(n-1)^2}\\
&\boxed{\operatorname{var} S^2 = \frac{2\sigma_X^4}{n-1}\,.}
\end{aligned}$$

3. In the MoM example 3 in Section (6.3), show that $F(x) = 1-\theta/x$, so that you can generate random samples as $x = \theta/(1-u)$ by the probability integral transformation method, where u is a pseudo-independent $\mathit{Unif}(0,1)$ sample. Choose $\theta = 3$. For a sample of size 10^3, plot a histogram of your sample. Does the default **R** histogram show much structure? Does the alternative MoM estimator of θ seem to work? Verify that $\hat\theta$ is a MoM estimator.

   Solution: Given the pdf:

$$\begin{aligned}
f(x) &= \theta x^{-2} && x>\theta\\
F(x) &= \int_{s=\theta}^{x}\frac{\theta}{s^2}ds = -\frac{\theta}{s}\bigg|_{\theta}^{x} = 1-\frac{\theta}{x} && x>\theta
\end{aligned}$$

   Thus $u = F(x)$ if and only if

$$\begin{aligned}
u &= 1-\frac{\theta}{x}\\
x &= \frac{\theta}{1-u} \quad \text{the PIT}\,.
\end{aligned}$$

   We can check the inverse MoM estimator by `set.seed(234); u=runif(1e3); x=3/(1-u); 0.5/mean(1/x)=2.944`, which is close to the true value $\theta=3$. The histogram of x_i shows lots of large values (highly skewed).

---

4. Show that if the sampling density $f(x|\boldsymbol{\theta})$ has m sufficient statistics, then the MLE $\hat{\boldsymbol{\theta}}$ must be a function of the data through the sufficient statistics alone.

   Solution: Consider the case where there is just one parameter, θ. The MLE maximizes the likelihood, i.e.

   $$\begin{aligned}\hat{\theta} &= \arg\max_{\theta} f(\boldsymbol{x}|\theta) \\ &= \arg\max_{\theta} \big[g(T(\boldsymbol{x})|\theta) \cdot h(\boldsymbol{x})\big] \\ &= \arg\max_{\theta} g(T(\boldsymbol{x})|\theta)\,,\end{aligned}$$

   since the function $h(\cdot)$ does not involve θ. Hence, $\hat{\theta}$ can only depend on the data through the sufficient statistic, $T(\boldsymbol{x})$.

Chapter 7

Hypothesis Testing

7.1 Setting up a Hypothesis Test

1. TBA

7.1.1 Example of a Critical Region

1. TBA

7.1.2 Accuracy and Errors in Hypothesis Testing

1. TBA

7.2 Best Critical Region for Simple Hypotheses

1. TBA

7.2.1 Simple Example Continued

1. TBA

7.2.2 Gaussian Shift Model with Common Variance

1. TBA

7.3 Best Critical Region for a Composite Alternative Hypothesis

1. Suppose we have a random sample of size n from a Poisson $Pois(m)$ PMF. We wish to test

$$\begin{aligned} H_0 : &\quad m = m_0 && \text{versus} \\ H_1 : &\quad m = m_1 > m_0 \,. \end{aligned}$$

   (a) Following Section 7.2, find the form of the best critical region.

   (b) The number of murders in Houston for the 4 years from 2014 to 2017 was 242, 297, 303, and 269. What is the best critical region for testing

$$\begin{aligned} H_0 : &\quad m = 250 && \text{versus} \\ H_1 : &\quad m = 275 \end{aligned}$$

   at the 5% significance level?

   (c) What is the power of your test?

   (d) What is your decision for the data given?

Solution:

(a) To find the form of the best critical region:

$$\begin{aligned} p_X(x|m) = \frac{m^x}{x!}\,e^{-m} \quad &\Longrightarrow \quad \log\, p_X(x|m) = x\log m - m - \log\, x! \\ \ell(m|\boldsymbol{x}) = \sum p_X(x_i|m) &= n\bar{x}\log m - nm - \sum \log x_i! \\ \ell(m_0|\boldsymbol{x}) - \ell(m_1|\boldsymbol{x}) &= n\bar{x}\log(m_0/m_1) - n(m_0 - m_1) - 0 < k \quad \text{when} \\ &\boxed{C : \boldsymbol{x} \,|\, \bar{x} > k' \,, \qquad \text{since } \log(m_0/m_1) < 0.} \end{aligned}$$

(b) Since we can do things exactly with the Poisson PMF, we choose to write the critical region as $Y = \sum X_i > k$. Under the null hypothesis, $X_i \sim Pois(250)$ so that $Y \sim Pois(4 \times 250) = Pois(1000)$.

The 95^{th} percentile is $\boxed{\texttt{> qpoiss(0.95,1000) = 1052}}$. We note

$$\sum_{i\geq 1052} p_X(i|1000) = 5.26\% \quad \text{and}$$
$$\sum_{i\geq 1053} p_X(i|1000) = 4.93\%\,.$$

So we see there is not exactly a 5% level test, so we use 1053 as the threshold and the level is 4.93%:

$$\boxed{C = \{\boldsymbol{x} \,|\, \sum x_i \geq 1053\}.}$$

(c) What is the probability of making a type II error? The power function computes lots of probabilities, one of which can give you the type II error:

$$\beta(\theta) = Pr(\boldsymbol{X} \in C \,|\, \theta) = Pr(\sum X_i \geq 1053 \,|\, m) \qquad \text{Note: } \theta = m \text{ here,}$$

$$Pr(\text{Type II error}) = 1 - \beta(m_1) = Pr(\sum X_i \leq 1052 \,|\, m = 4 \times 275)$$

$$\boxed{Pr(\text{Type II error}) = 7.527\% \qquad \texttt{> sum(dpois(0:1052, 4*275) = 0.07527186}}$$

(d) For the Houston data, $\sum x_i = 1{,}111 \geq 1{,}053$. This is in the critical region, so we reject the null hypothesis at the 5% level.

2. Consider a hypothesis test of the mean of normal data with known variance of 1:

$$H_0: \quad \mu = 0 \qquad\qquad \text{versus}$$
$$H_1: \quad \mu \neq 0.$$

(a) If $n = 16$, what is the best critical region?

(b) Perform the hypothesis test of these simulated samples. Generate 16 normal samples in R via

$$set.seed(123); \quad x = rnorm(16, 1, 1)\,.$$

What is your decision?

Solution:

(a) $X \sim N(0,1)$ under the null hypothesis. Note the alternative is two-sided; therefore, we must use the MLE $\bar{x}$ in place of a known value for μ_1. To invoke the Neyman-Pearson Lemma, we compute the log-likelihood

$$\begin{aligned}
\ell(\mu \,|\, \boldsymbol{x}, \sigma = 1) &= -\frac{n}{2}\log 2\pi - \frac{1}{2}\sum (x_i - \mu)^2 \\
\ell(\mu_0 = 0) - \ell(\mu_1 = \bar{x}) &= 0 - \frac{1}{2}\sum (x_i - 0)^2 + \frac{1}{2}\sum (x_i - \bar{x})^2 \\
&= \frac{1}{2}\sum \Big[- x_i^2 + x_i^2 - 2x_i\bar{x} + \bar{x}^2 \Big] \\
&= \frac{1}{2}\Big[0 - 2n\bar{x}\bar{x} + n\bar{x}^2 \Big] = -\frac{n}{2}\bar{x}^2 < k \qquad \text{or } \bar{x}^2 > k', \text{ or} \\
& \boxed{C = \{\boldsymbol{x} : |\bar{x}| > k''\} \qquad \text{critical region}}
\end{aligned}$$

Now, under the null hypothesis with $n = 16$,

$$\bar{X} \sim N\left(0, \frac{1}{16}\right) = N\left(0, \frac{1}{4^2}\right); \qquad \text{we find } k'' \text{ by}$$

$k'' = \Phi^{-1}(0.975 \,|\, 0, 0.25) = 0.490$ `> qnorm( 0.975, 0, 0.25 ) = 0.489991.`

Note that the critical region is two-sided, even though the MLE is just on one side of the null hypothesis. (Magic?)

(b) `set.seed(123); mean(rnorm(16,1,1)) = 1.255`, which is greater than 0.490, so in the critical region, i.e., reject the null hypothesis.

3. TBA

7.3.1 Negative Exponential Composite Hypothesis Test

1. TBA

7.3.1.1 Example

1. TBA

7.3.1.2 Alternative Critical Regions

1. TBA

7.3.1.3 Mount St. Helens Example

1. TBA

7.3.2 Gaussian Shift Model with Common But Unknown Variance: The t-test

1. Consider a hypothesis test of the mean of normal data with unknown variance:

$$\begin{aligned} H_0 : &\quad \mu = 0 \qquad\qquad \text{versus} \\ H_1 : &\quad \mu \neq 0. \end{aligned}$$

   (a) If $n = 9$, what is the best critical region?

   (b) Perform the hypothesis test of these simulated samples. Generate 9 normal samples in R via

$$set.seed(456); \quad x = rnorm(9, 1, 2)\,.$$

   What is your decision?

Solution:

   (a) This is the subject of Section 7.3.4. The critical region is (see Equation 7.11)

$$\left|T_{n-1}\right| = \left|\frac{\bar{X} - \mu_0}{\sqrt{S^2/n}}\right| \geq t_{9-1}(0.975) = 2.306\,,$$

   where S^2 is the unbiased estimator of the variance.

   (b) For the dataset, $\bar{x} = 0.9112$ and $\sqrt{s^2} = 1.859$. So we compute

$$\boxed{t_{9-1} = \frac{0.9112 - 0}{1.859/3} = 1.470\,,}$$

   which is *not* in the critical region. Thus, we $\boxed{\textit{fail to reject the null hypothesis.}}$

2. TBA

7.3.3 The Random Variable T_{n-1}

1. TBA

7.3.3.1 Where We Show $\bar{X}$ and S^2 are Independent

1. TBA

7.3.3.2 Where We Show That S^2 Scaled is $\chi^2(n-1)$

1. TBA

7.3.3.3 Where We Finally Derive the T PDF

1. TBA

7.3.4 The One-Sample t-test

1. TBA

7.3.5 Example

1. TBA

7.3.6 Other t-tests

1. TBA

7.3.6.1 Paired t-test

1. TBA

7.3.6.2 Two-Sample t-test

1. TBA

7.3.6.3 Example Two-Sample t-test: Lord Rayleigh's Data

1. TBA

7.4 Reporting Results: p-Values and Power

1. TBA

7.4.1 Example When the Null Hypothesis Is Rejected

1. TBA

7.4.2 When the Null Hypothesis Is Not Rejected

1. TBA

7.4.3 The Power Function

1. TBA

7.5 Multiple Testing and the Bonferroni Correction

1. TBA

7.6 Problems

1. Show that the sum of n i.i.d. negative exponential r.v.'s with parameter λ has a $Gamma(n, \lambda)$ exactly. Hint: Use the MGF technique of Section 5.4.1.

   Solution: From Appendix B, the MGF of the negative exponential pdf $f(x) = e^{-x/\lambda}/\lambda$ is $M_X(t) = 1/(1 - \lambda t)$. Therefore, the MGF of

$Y = \sum_i X_i$ is

$$\begin{aligned} M_Y(t) &= E^{tY} = E^{\sum_i X_i} \\ &= E\left[\prod_i e^{tX_i}\right] \\ &= [M_X(t)]^n \\ &= \left(\frac{1}{1-\lambda t}\right)^n , \end{aligned}$$

which is the MGF of a $Gamma(n, \lambda)$.

2. Show that among the contiguous intervals (a, b) containing 95% of the probability for our example in Section 7.3.1.1, a necessary condition to minimize the width of the interval, $b - a$, is that the sampling density must be equal at a and b. This result holds generally if the sampling density is unimodal and monotone on either side of the mode.

   Solution: Finding the optimal interval (a, b) in this setting is a constrained optimization problem to minimize the interval width $b - a$ subject to the constraint that the interval contains the desired probability. Hence, the Lagrangian is

   $$\begin{aligned} L(a, b, \lambda) &= (b-a) + \lambda\left(0.95 - \int_a^b f(x)\,dx\right) ; \\ \frac{\partial L}{\partial a} &= -1 + \lambda(-(-f(a)) = -1 + \lambda f(a) = 0 \\ \frac{\partial L}{\partial b} &= 1 + \lambda(-f(b)) = 1 - \lambda f(b) = 0 \end{aligned}$$

   are the necessary conditions. It follows that whatever the value of λ is that $f(a) = f(b)$ must be satisfied. Usually, a nonlinear search is required to find the value off the density $f(a)$ so that the interval contains 95% of the probability. Because the density was assumed to be unimodal and smooth, this interval will be unique in general.

3. Suppose the distribution of the r.v. used to find the various intervals (a, b) in Section 7.3.1.2 was not only unimodal and monotone on either side of the mode, but was also symmetric. Show that the equal-tail-area and narrowest-width intervals are identical.

Solution: Without loss of generality, assume that the mode (and mean) of the symmetric pdf is 0. Hence, $f(-a) = f(a)$, so that the optimal interval by problem 2 is $(-a, a)$. This interval is also symmetric around the mean and has equal tail area as well.k

4. Using the MGF technique, show that

$$Y = \sum_{i=1}^{n} Z_i^2 \sim \chi^2(n)\,,$$

where Z_i is a random sample from the $N(0,1)$ PDF.

Solution: We know that $Z^2 \sim \chi^2(1)$. Hence,

$$M_Y(t) = \left[M_{Z^2}(t)\right]^n = \left[\left(\frac{1}{1-2t}\right)^{1/2}\right]^n = \left(\frac{1}{1-2t}\right)^{n/2}\,,$$

as claimed.

5. Using the MGF technique, show that if $U = S + T$, where (a) $U \sim \chi^2(n)$; (b) $T \sim \chi^2(1)$; and (c) S and T are independent, then $S \sim \chi^2(n-1)$.

Solution: Since everything is independent, the expectation of the product is the product of the expectation. Hence,

$$\begin{aligned} Ee^{tU} &= E[e^{t(S+T)}] = Ee^{tS} \cdot Ee^{tT} \quad \text{or} \\ M_U(t) &= M_S(t) \cdot M_T(t)\,. \quad \text{Substituting, we have} \\ \left(\frac{1}{1-2t}\right)^n &= M_S(t) \cdot \left(\frac{1}{1-2t}\right)^1\,, \quad \text{hence, we solve for} \\ M_S(t) &= \left(\frac{1}{1-2t}\right)^{n-1}\,, \end{aligned}$$

which is the MGF of a $\chi^2(n-1)$ random variable.

6. Verify the critical regions and sample sizes for the example at the end of Section 7.4.3.

Solution: Since $\sigma = 5$ is assumed known, then following Sections 7.2.2 and 7.4.1, the "acceptance interval" of the best test at the 95% confidence level is of the form

$$\mu_0 \mp 1.96 \frac{\sigma}{\sqrt{n}} \qquad \text{or} \qquad C^c = 10 \mp \frac{1.96 \times 5}{\sqrt{n}} = 10 \mp \frac{9.80}{\sqrt{n}} \,.$$

The probability $\bar{X}$ will fall in this interval has probability 95% for any sample size n under the null hypothesis. The power of the test at some other value of $\mu \neq 10$ would be the probability $\bar{X}$ does not land in this interval (i.e., correctly rejects H_0). Thus, if we wish the power at $\mu = 12$ to be 90%, then we want its complement to be 10% or

$$\begin{aligned} 1 - 0.90 &= Pr(\bar{X} \in C^c | \mu = 12) \\ &= Pr(10 - \frac{9.80}{\sqrt{n}} \leq \bar{X} \leq 10 + \frac{9.80}{\sqrt{n}} | \mu_{\bar{x}} = 12, \sigma_{\bar{x}} = 5/\sqrt{n}) \,. \end{aligned}$$

Finding n requires a simple nonlinear search, for which we find in **R**

$$\begin{aligned} &= \texttt{pnorm}\left(10 + \frac{9.8}{\sqrt{n}}, 12, \frac{5}{\sqrt{n}}\right) - \texttt{pnorm}\left(10 - \frac{9.8}{\sqrt{n}}, 12, \frac{5}{\sqrt{n}}\right) \\ &= 0.1001 - 10^{-7} \qquad \text{at } n = 65.67 \end{aligned}$$

for which $C^c = (8.791.11.209)$. Rounding n to 66, we find the power is slightly better than 90%, namely, 90.14%.

A similar calulation shows that $C^c = (8.89, 11.11)$ and $n = 133.96$ gives a significance level of $\alpha = 1\%$ and power $\beta(12) = 98\%$.

7. The 1969 military draft lottery numbers are shown in Figure 7.9. Individuals with numbers over 195 were not drafted. There appears to be a nonrandom downward trend. For example, only 5 birthdays in December had a lottery number above average. Run a two-sample T-test in **R** with the command `t.test(x[1:183],x[184:366])` and show the p-value is 5×10^{-5}. What do we conclude? Perform a simulation using data from `x = sample(366)` to verify the T-distribution is the appropriate test statistic.

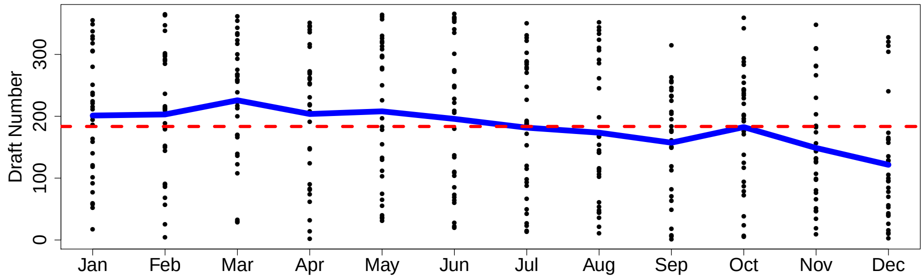


Figure 7.1: Draft lottery numbers 1–366 by month. The monthly average is the blue line. The overall average of 183.5 is the red dotted line.

Solution: The `t.test(x[1:183],x[184:366],var.equal=T)` computes the value 4.502 with 364 degrees of freedom, and a p-value of 10^{-5}. One might also fit a straight line to these data and test if the slope is 0 or not.

Chapter 8

Confidence Intervals and Other Hypothesis Tests

8.1 Confidence Intervals

1. TBA

8.1.1 Confidence Interval for μ: Normal Data, σ^2 Known

1. TBA

8.1.2 Confidence Interval for μ: σ^2 Unknown

1. Suppose we have a random sample of size 50 from a $N(\mu, \sigma^2)$ PDF. We wish to test

$$\begin{aligned} H_0 : &\quad \mu = 10 \qquad\qquad \text{versus} \\ H_1 : &\quad \mu \neq 10\,. \end{aligned}$$

   The sample moments are $\bar{x} = 13.4508$ and $s^2 = 65.8016$.

   (a) Find the critical region C and test the null hypothesis at the 5% level. What is your decision?

   (b) What is the p-value for your decision?

   (c) What is a 95% confidence interval for μ?

Solution:

(a) This starts out the same as the first problem for Section 7.3.2. The test statistic is

$$\boxed{t_{50-1} = \frac{13.4508 - 10}{\sqrt{65.8016/50}} = 3.008\,.}$$

The 97.5^{th} percentile is $t_{50-1}(0.975) = 2.0096$. So

$$3.008 \notin (-2.0096, 2.0096) \qquad \text{so in critical region, or write as}$$
$$\boxed{3.008 \in (-2.0096, 2.0096)^c \qquad \text{so in critical region;}}$$

therefore, the *decision is to reject the null hypothesis* at the 5% level. Note the critical region is $C = (-2.0096, 2.0096)^c$, the complement of the interval above.

(b) The p-value is the significance level, α, that makes the critical region match the observed test statistic value; that is, what choice for α results in the critical region (interval)

$$C = (-3.008, 3.008)^c\,.$$

With this critical region, the type I error would be the sum of the two tail areas:

$$\begin{aligned} p\text{-value} &= Pr(T_{49} < -3.008) + Pr(T_{49} > 3.008) \\ &= 2 \times Pr(T_{49} < -3.008) = 2 \times 0.00297155 \\ &\boxed{p\text{-value} = 0.414\%\,.} \end{aligned}$$

Note the 1% test would use the interval $(-2.68, 2.68)^c$, using `> qt(.995,49)`.

(c) From Equation (8.4), a 95% confidence interval for μ is

$$\begin{aligned} \mu &\in \bar{X} \mp t_{0.975,49}\,\frac{S}{\sqrt{n}} \\ &\in 13.4508 \mp 2.0096\,\frac{\sqrt{65.8016}}{\sqrt{49}} \\ &\boxed{\mu \in (11.122, 15.780)\,.} \end{aligned}$$

---

2. Use the same data as in the previous Problem.

   (a) Test the null hypothesis that $\sigma^2 = 64$ versus a two-sided alternative. First find the critical region and then give your decision.

   (b) Find a 95% confidence interval for σ^2?

   (c) If you are worried about performing 2 statistical tests on the same data and that the overall type I error might not be 5%, how might you modify your approach?

Solution:

(a) The test is derived in Section 8.2.2. You *fail to reject the null hypothesis* at the 5% level if

$$a < \frac{(n-1)\,S^2}{\sigma_0^2} < b \qquad a = \chi^2_{50-1}(.025) = 31.55 \;\&\; b = \chi^2_{50-1}(.975) = 70.22$$

$$31.55 < \frac{49 \times 65.8016}{64} < 70.22$$

$$\boxed{31.55 < 50.38 < 70.22 \qquad \text{which is true, so fail to reject } H_0}$$

(b) The confidence interval starts the same, but leaves the quantity σ^2 as an "unknown." Thus a 95% confidence interval may be derived as

$$
\begin{aligned}
0.95 &= Pr\left(31.55 < \frac{(n-1)\,S^2}{\sigma^2} < 70.22\right) \\
&= Pr\left(\frac{1}{31.55} > \frac{\sigma^2}{(n-1)\,S^2} > \frac{1}{70.22}\right) \\
&= Pr\left(\frac{(n-1)\,S^2}{31.55} > \sigma^2 > \frac{(n-1)\,S^2}{70.22}\right) \\
&= Pr\left(\frac{(n-1)\,S^2}{70.22} < \sigma^2 < \frac{(n-1)\,S^2}{31.55}\right) \qquad \text{Note: } (n-1)S^2 = 3224.278
\end{aligned}
$$

$$\boxed{0.95 = Pr\Big(45.917 < \sigma^2 < 102.196\Big) \qquad \text{C.I. for the variance.}}$$

$$\boxed{0.95 = Pr\Big(6.776 < \sigma < 10.110\Big) \qquad \text{C.I. for the standard deviation.}}$$

(c) The simplest modification to have an overall 5% level pair of tests is to perform each of the 2 tests at the 2.5% level. A little conservative, but ... (See the JAMA instructions again.)

3. TBA

8.1.3 Confidence Intervals and p-Values

1. TBA

8.2 Hypotheses About the Variance and the F-Distribution

1. TBA

8.2.1 The F-Distribution

1. TBA

8.2.2 Hypotheses About the Value of σ^2

1. TBA

8.2.3 Confidence Interval for σ^2

1. TBA

8.2.4 Two-Sided Alternative for Testing $\sigma^2 = \sigma_0^2$

1. TBA

8.3 Pearson's Chi-Squared Tests

1. TBA

8.3.1 The Multinomial PMF

1. TBA

8.3.2 Goodness-of-Fit (GoF) Tests

1. TBA

8.3.3 Two-Category Binomial Case

1. TBA

8.3.4 m-Category Multinomial Case

1. TBA

8.3.5 Goodness-of-Fit Test for a Parametric Model

1. One hundred points were simulated using the R function `rexp(100,1)` for the negative exponential PDF with $\lambda = 1$. We are interested in testing a goodness-of-fit of the random numbers to the true model. A histogram of these data gave bin counts over the 9 bins

   $$\{(0,0.5),\ (0.5,1),\ \ldots,\ (4,4.5)\} \qquad \text{of} \qquad (42,\ 16,\ 15,\ 11,\ 3,\ 5,\ 3,\ 1,\ 4)\,.$$

   Also, $\boxed{\bar{x} = 1.0865}$.

   (a) Perform a Pearson goodness-of-fit test at the 5% level. *Hint: Be careful that the expected counts all are greater than 5 and that they add up to* $n = 100$.

   (b) What is the p-value? How do your interpret this number?

   (c) If our null hypothesis was that the data came from a negative exponential PDF, but we did not know that the mean was 1, how would you perform the goodness-of-fit test? (Do it.)

Solution:

(a) So that the bin counts are at least 5, we must combine some of the smaller bins together. Working from the left, we end up with

$$\begin{aligned} \text{use bins} &= (0.0.5),\ (0.5,1),\ (1,1.5),\ (1.5,2),\ (2,3),\ (3,4.5(\infty)) \\ \text{new counts} &= (42,\ 16,\ 15,\ 11,\ 3+5,\ 3+1+4) \end{aligned}$$

$$\boxed{\boldsymbol{o} = (42,\ 16,\ 15,\ 11,\ 8,\ 8) \qquad \text{for the 6 new bins}}$$

To get the bin probabilities, we compute

$$\begin{aligned} \texttt{pexp(c(0,.5,1,1.5,2,3,Inf), 1)} &= \texttt{0.000 0.393 0.632 0.777 0.865 0.95} \\ \boldsymbol{p} = \texttt{diff(previous vector)} &= \texttt{0.393 0.239 0.145 0.088 0.086 0.05} \end{aligned}$$

$$\boxed{\boldsymbol{e} = \ 39.3\ 23.9\ 14.5\ 8.8\ 8.6\ 5.0 \qquad \text{that is, } \boldsymbol{e} = 100 \times \boldsymbol{p}}$$

The Pearson χ^2-test measures the difference between $\boldsymbol{o}$ and $\boldsymbol{e}$ as

$$\begin{aligned} \mathcal{X}_6 &= \sum_{k=1}^{6} \frac{(o_k - e_k)^2}{e_k} \sim \chi^2_{6-1} \\ &= 5.221\,. \end{aligned}$$

The critical value at the 5% level is $\texttt{chisq}(.95, 6-1) = \texttt{11.071}$. Therefore, we *fail to reject the null hypothesis* since $5.221 < 11.081$.

(b) To find the p-value, we seek the significance level of the critical region that has the test statistic on the boundary, namely,

$$\begin{aligned} C &= (5.221, \infty)\,. \qquad \text{from which the tail area is} \\ Pr(\chi^2_5 > 5.221 \,|\, H_0 \ true) &= 1 - \texttt{pchisq}(5.221, 5) \\ &\boxed{p\text{-value} = 38.95\%} \end{aligned}$$

Thus, there is no strong evidence to reject the exponential model.

(c) If the null hypothesis does not specify the parameter, we use the MLE and reduce the degrees of freedom by 1. The MLE is $\bar{x} = 1.0865$. The observed counts do not change, but we change the expected counts using λ value of 1.0865 or 1/1.0865 rather than 1 before. The **R** function `pexp` uses a different definition of the PDF than our text. This leads to

$$\begin{aligned} \texttt{pexp(c(0,.5,1,1.5,2,3,Inf), 1/1.0865)} &= \texttt{0.000 0.393 0.632 0.777 0.865 0.950} \\ \boldsymbol{p} = \texttt{diff(previous vector)} &= \texttt{0.369 0.233 0.147 0.093 0.095 0.063} \\ &\boxed{\boldsymbol{e} = \ 36.9\ 23.3\ 14.7\ 9.3\ 9.5\ 6.3 \qquad \text{that is, } \boldsymbol{e} = 100 \times \boldsymbol{p}} \end{aligned}$$

Now

$$\begin{aligned} \mathcal{X}_6 = \sum_{k=1}^{6} \frac{(o_k - e_k)^2}{e_k} &\sim \chi^2_{6-1-1} \\ &= 4.01\,. \\ &< 9.488 = \chi^2_4(0.95)\,. \end{aligned}$$

so again fail to reject the null hypothesis.

2. TBA

8.3.6 Tests for Independence in Tables

1. **Pearson Test for Independence in Tables.** One hundred middle school students studying Spanish were randomly assigned to 2 classrooms. One classroom used Rosetta Stone and the other used Babbel

software in the language lab and on their personal computers. At the end of one semester, the final grades turned out to be:

Grade	A	B	C	D	E	Totals
Classroom I	8	13	16	10	3	50
Classroom II	4	9	14	16	7	50

Table 8.1: Hypothetical Data Comparing Foreign Language Software

(a) Test the null hypothesis that the software options are equally effective.

(b) What is the p-value?

Solution:

(a) We use Equation (8.22)

$$e_{ij} = n\hat{p}_{ij} = n\,\widehat{P(A_i)}\widehat{P(B_j)} = n \times \frac{o_{i\bullet}}{n} \times \frac{o_{\bullet j}}{n} = \frac{o_{i\bullet} \times o_{\bullet j}}{n}$$

to find

Grade	A	B	C	D	E	Totals
Classroom I	6.0	11.0	15.0	13.0	5.0	50.0
Classroom II	6.0	11.0	15.0	13.0	5.0	50.0

Table 8.2: Table of Expectations $\{e_{ij}\}$.

Hence,

$$\mathcal{X}_{rs} = \sum_{i=1}^{r}\sum_{j=1}^{s}\frac{(o_{ij}-e_{ij})^2}{e_{ij}} = 5.179 \sim \chi^2_{(2-1)(5-1)}$$
$$\leq 9.49 = \chi^2_4(.95)\,;$$

so we fail to reject.

(b) The p-value is

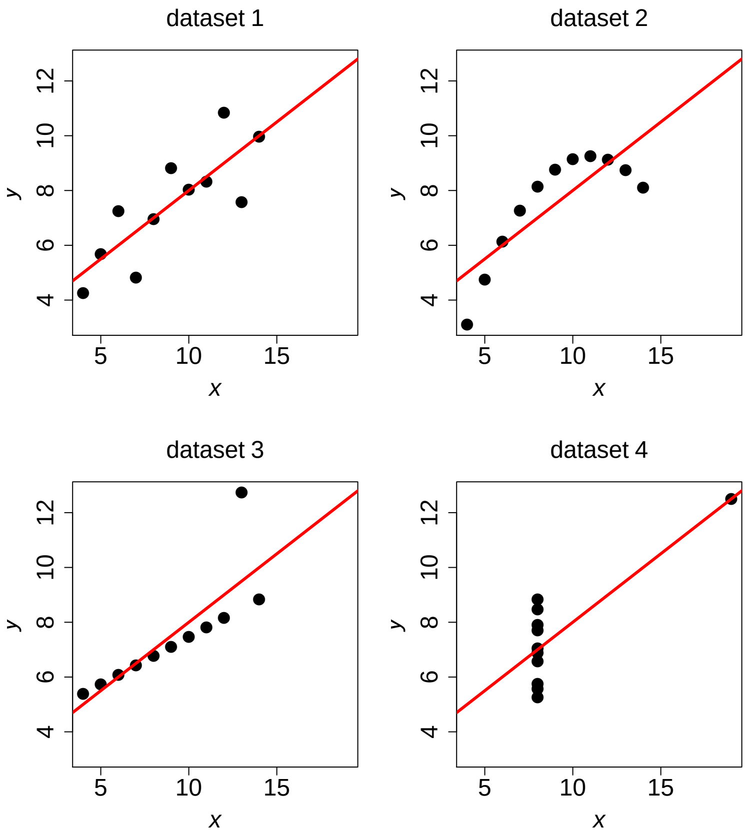
$$\boxed{p\text{-value} = Pr(\chi^2_4 > 5.179) = 26.95\%\,.}$$

2. TBA

8.4 Correlation Coefficient Tests and C.I.'s

1. The correlation coefficient measures the degree to which the random variable Y is related to X in a linear fashion. If the relationship is in fact nonlinear, the intuition may be faculty. In, 1973, Professor Francis Anscombe devised a clever set of 4 datasets (each with $n = 11$) that illustrate this fact. These data are available in R as the object `anscombe`. Compute the correlation coefficient by the R expression `cor(x,y)` for each set of data and plot. Add the least squares regression line to each figure by using the R expression `abline( lsfit(x,y) )` . What do you conclude?

Solution: Performing `> cor(anscombe$x1,anscombe$y1)` returns the value 0.816, the same value as with the other 3 datasets. Adding the least squares regression line is accomplished by `> abline( lsfit(x,y) )` . The equation of the line is $y = 3 + x/2$ in each case.



2. Consider the full 5-parameter bivariate normal PDF. Given a random sample $\{(x_i, y_i),\ i = 1, \ldots, n\}$, assume you know the MLE's for μ_x, μ_y, σ_x, and σ_y are

$$\hat{\mu}_x = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \hat{\sigma}_x = s_x = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$\hat{\mu}_y = \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \qquad \hat{\sigma}_y = s_y = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2}\,.$$

Verify the formula for the MLE of the correlation coefficient ρ in Equation (8.23)

$$
\begin{aligned}
\hat{\rho} = r &= \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \qquad \text{or equivalently} \\
r &= \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{s_x\, s_y}
\end{aligned}
$$

Hints: Now the pdf is

$$
\begin{aligned}
f(x,y) &= \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\ \exp\left[\frac{-Q(x,y)}{2(1-\rho^2)}\right] \quad \text{where} \\
Q(x,y) &= \left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 .
\end{aligned}
$$

Thus the average (i.e. divided by n) log-likelihood for ρ is given by

$$
\begin{aligned}
\ell(\rho|\boldsymbol{x},\boldsymbol{y}) &= -\log 2\pi - \log\sigma_x - \log\sigma_y - \frac{1}{2}\log(1-\rho^2) - \frac{\bar{Q}(\boldsymbol{x},\boldsymbol{y})}{2(1-\rho^2)} \\
&\qquad \text{where} \\
\bar{Q}(\boldsymbol{x},\boldsymbol{y}) &= \frac{1}{n}\sum_{i=1}^{n} Q(x_i,y_i)\,.
\end{aligned}
$$

Plugging in the 4 MLE's we know, we find

$$
\begin{aligned}
Q(x,y) &= \frac{(x-\bar{x})^2}{s_x^2} - 2\rho\frac{(x-\bar{x})(y-\bar{y})}{s_x\, s_y} + \frac{(y-\bar{y})^2}{s_y^2}\,; \quad \text{hence,} \\
\bar{Q}(\boldsymbol{x},\boldsymbol{y}) &= \frac{1}{n}\sum_{i=1}^{n} Q(x_i,y_i) \\
&= \frac{\frac{1}{n}\sum_i (x_i-\bar{x})^2}{s_x^2} - 2\rho\frac{\frac{1}{n}\sum_i (x_i-\bar{x})(y_i-\bar{y})}{s_x\, s_y} + \frac{\frac{1}{n}\sum_i (y_i-\bar{y})^2}{s_y^2} \\
&= \frac{s_x^2}{s_x^2} - 2\rho\frac{r s_x s_y}{s_x s_y} + \frac{s_y^2}{s_y^2} \\
&= 2(1-\rho\, r)\,.
\end{aligned}
$$

Thus the relevant terms in the average log-likelihood to be maximized are

$$\tilde{\ell}(\rho|\boldsymbol{x},\boldsymbol{y}) = -\frac{1}{2}\log(1-\rho^2) - \frac{2(1-\rho\, r)}{2(1-\rho^2)}$$

Use Mathematica to show

$$\frac{\partial\tilde{\ell}(\rho|\boldsymbol{x},\boldsymbol{y})}{\partial\rho} = \frac{r-\rho+r\rho^2-\rho^3}{(1-\rho^2)^2}$$

which equals 0 when $\hat{\rho}=r$ (also $\pm i$).

Solution: The Mathematica command

```
Solve[ r - rho + r rho**2 - rho**3 == 0, rho ]
```

returns the 3 solutions

```
{{ rho -> -i }, { rho -> i }, { rho -> r }}
```

Deriving all 5 MLE's can be accomplished in 2 steps. First, find $\hat{\mu}_x$ and $\hat{\mu}_y$ using Mathematica. Then re-express the likelihood using $\bar{x}$ and $\bar{y}$ and solve simultaneous for the remaining 3 parameters. When writing down $\bar{Q}(\boldsymbol{x},\boldsymbol{y})$, use the identities (adding and subtracting the sample means)

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n (x_i-\mu_x)^2 &= \frac{1}{n}\sum_{i=1}^n [(x_i-\bar{x})-(\mu_x-\bar{x})]^2\\ &= \frac{1}{n}\sum_{i=1}^n \left[(x_i-\bar{x})^2 - 2(x_i-\bar{x})(\mu_x-\bar{x}) + (\mu_x-\bar{x})^2\right]\\ &= \frac{1}{n}\sum_{i=1}^n (x_i-\bar{x})^2 + (\mu_x-\bar{x})^2\\ &= s_x^2 + (\mu_x-\bar{x})^2\,,\end{aligned}$$

and similarly,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n (x_i-\mu_x)(y_i-\mu_y) &= \frac{1}{n}\sum_{i=1}^n (x_i-\bar{x})(y_i-\bar{y}) + (\mu_x-\bar{x})(\mu_y-\bar{y})\\ &= r s_x s_y + (\mu_x-\bar{x})(\mu_y-\bar{y})\,.\end{aligned}$$

Be sure to optimize w.r.t. σ and not the variance σ^2. Looking at the first portion, notice that the parameters μ_x and μ_y occur only in $\bar{Q}(\boldsymbol{x}, \boldsymbol{y})$.

$$\bar{Q}(\mu_x, \mu_y) = \frac{s_x^2 + (\mu_x - \bar{x})^2}{\sigma_x^2} - \frac{2\rho[r s_x s_y + (\mu_x - \bar{x})(\mu_y - \bar{y})]}{\sigma_x \sigma_y} + \frac{s_y^2 + (\mu_y - \bar{y})^2}{\sigma_y^2}$$

Next, we find that $\hat{\mu}_x = \bar{x}$ and $\hat{\mu}_y = \bar{y}$ by solving the pair of equations

$$\begin{aligned} \frac{\partial \bar{Q}(\mu_x, \mu_y)}{\partial \mu_x} &= \frac{2(\mu_x - \bar{x})}{\sigma_x^2} - \frac{2\rho(\mu_y - \bar{y})}{\sigma_x \sigma_y} \\ \frac{\partial \bar{Q}(\mu_x, \mu_y)}{\partial \mu_y} &= -\frac{2\rho(\mu_x - \bar{x})}{\sigma_x \sigma_y} + \frac{2(\mu_y - \bar{y})}{\sigma_y^2} \end{aligned}$$

for μ_x and μ_y in Mathematica. Similarly, re-expressing $\bar{Q}(\sigma_x, \sigma_y, \rho | \boldsymbol{x}, \boldsymbol{y})$ leads to the remaining 3 MLE's.

3. TBA

8.4.1 How to Test if the Correlation $\rho = 0$

1. TBA

8.4.2 Confidence Intervals and Tests for a General Correlation Coefficient

1. **Correlation Coefficient Test and Confidence Interval.** This case study examined the relationship between an isometric strength test, X, and job performance, Y, for 50 warehouse workers. You may assume these data follow a bivariate normal PDF. The following statistics were

calculated from these hypothetical data:

$$\sum_{i=1}^{50} x_i = 250.0109$$
$$\sum_{i=1}^{50} x_i^2 = 1335.2427$$
$$\sum_{i=1}^{50} y_i = 494.6882$$
$$\sum_{i=1}^{50} y_i^2 = 4934.5076$$
$$\sum_{i=1}^{50} x_i y_i = 2515.2075\,.$$

(a) What is the sample correlation coefficient?

(b) Test the null hypothesis that $H_0 : \rho = 0$ versus a two-side alternative. First find the critical region and then give your decision.

(c) Find a 95% confidence interval for ρ?

Solution:

(a) The sample correlation coefficient expanded out is

$$\begin{aligned}\hat{R} &= \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2\sum(y_i-\bar{y})^2}} \\ &= \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sqrt{\left(\sum x_i^2 - n\bar{x}^2\right)\left(\sum y_i^2 - n\bar{y}^2\right)}}\end{aligned}$$

$$\boxed{\hat{R} = 0.7123\,.}$$

(b) Using the equation after (8.25), we **fail to reject** H_0 if

$$z_{\alpha/2} \le \frac{\frac{1}{2}\log\frac{1+R}{1-R} - 0}{\frac{1}{\sqrt{n-3}}} \le z_{1-\alpha/2} \qquad \text{or}$$

$$\boxed{-1.96 < 6.114 < 1.96\ ???}$$

This is false, so we reject H_0.

(c) The confidence interval is given by Equation (8.28):

$$\frac{e^{\frac{-2z_{1-\alpha/2}}{\sqrt{n-3}}}-\frac{1-R}{1+R}}{e^{\frac{-2z_{1-\alpha/2}}{\sqrt{n-3}}}+\frac{1-R}{1+R}} \le \rho \le \frac{e^{\frac{-2z_{\alpha/2}}{\sqrt{n-3}}}-\frac{1-R}{1+R}}{e^{\frac{-2z_{\alpha/2}}{\sqrt{n-3}}}+\frac{1-R}{1+R}} \qquad \text{or}$$

$$\boxed{0.541 \le \rho \le 0.827\,.}$$

2. TBA

8.5 Linear Regression

1. TBA

8.5.1 Least Squares Regression

1. TBA

8.5.2 Distribution of the Least-Squares Parameters

1. TBA

8.5.3 A Confidence Interval for the Slope

1. TBA

8.5.4 A Two-Side Hypothesis Test for the Slope

1. TBA

8.5.5 Predictions at a New Value

1. TBA

8.5.6 Population Interval at a New Value

1. **Linear Regression Confidence Intervals.** A chemical engineering graduate student ran 11 experiments as follows. Given 500 grams of input chemicals by weight, a process was run at 11 temperatures $(50, 60, 70, \ldots, 140, 150)$. The output material was separated into product (y_i) and waste. The weight of the product was also measured in grams; see the Figure on the next page, on which both the data and the least-squares line are displayed.

   You may assume the ϵ_i are normally distributed. The following statistics were calculated from these hypothetical data:

$$\sum_{i=1}^{11} x_i = 1100$$

$$\sum_{i=1}^{11} x_i^2 = 121{,}000$$

$$\sum_{i=1}^{11} y_i = 1{,}844.812$$

$$\sum_{i=1}^{11} y_i^2 = 369{,}372.483$$

$$\sum_{i=1}^{11} x_i y_i = 205{,}743.937\,.$$

   (a) Consider the pointwise confidence interval for $Y|x$ at the 95% level. Sketch the (upper and lower) curves onto the graph provided over the range of the data. (Compute these CI's at enough points to sketch the continuous lines. Never mind that the overall type I error may be wrong.)

   (b) Next, consider the pointwise confidence interval for the predictions, also at the 95% level for each x. Sketch the (upper and lower) curves onto the graph provided over the range of the data. (Again, ignore the fact that the overall type I error will be wrong.)

Solution:

(a) The least-squares estimates of the intercept and slope are $\boxed{\hat{a} = \bar{y} = 167.71}$ and slope (Eqn (8.34))

$$\hat{b} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2} = \boxed{1.9330 = \hat{b}}$$

The notation in Section 8.5.6 is easy to mis-read. The corrected Equation (8.45) gives us the T_{n-2} random variable we need to form the confidence interval at a "new point" $x = x_{n+1}$

$$\boxed{T_{n-2} = \frac{Y_{n+1} - \widehat{Y}_{n+1}}{\sqrt{\hat{\sigma_\epsilon^2}\left(1 + \frac{1}{n} + \frac{(x_{n+1}-\bar{x})^2}{\sum(x_i-\bar{x})^2}\right)}}\,.}$$

Let $t = T_{11-2}(0.975) = 2.262$. Then rearranging, we have the CI is

$$Y_{n+1} \in \left[\hat{a} + \hat{b}(x_{n+1} - \bar{x})\right] \mp t \times \sqrt{\hat{\sigma_\epsilon^2}\left(1 + \frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{\sum(x_i - \bar{x})^2}\right)}$$

Now $\boxed{\bar{x} = 100}$ and $\boxed{\sum(x_i - \bar{x})^2 = \sum x_i^2 - n\bar{x}^2 = 11{,}000}$, but $\hat{\sigma_\epsilon^2}$ requires some work.

$$\begin{aligned}\hat{\sigma_\epsilon^2} &= \frac{1}{n-2}\sum_{i=1}^{n}\left[y_i - \hat{a} - \hat{b}(x_i - \bar{x})\right]^2 \\ &= \frac{1}{n-2}\sum_{i=1}^{n}\left[y_i^2 + \hat{a}^2 + \hat{b}^2(x_i - \bar{x})^2 - 2\hat{a}y_i - 2\hat{b}y_i(x_i - \bar{x}) + 2\hat{a}\hat{b}(x_i - \bar{x})\right] \\ &= \frac{1}{n-2}\left[\sum y_i^2 + n\hat{a}^2 + \hat{b}^2\sum(x_i - \bar{x})^2 - 2n\hat{a}\bar{y} - 2\hat{b}\sum y_i(x_i - \bar{x}) + 0\right] \\ &\boxed{= 2097.5961 = 45.80^2\,.}\end{aligned}$$

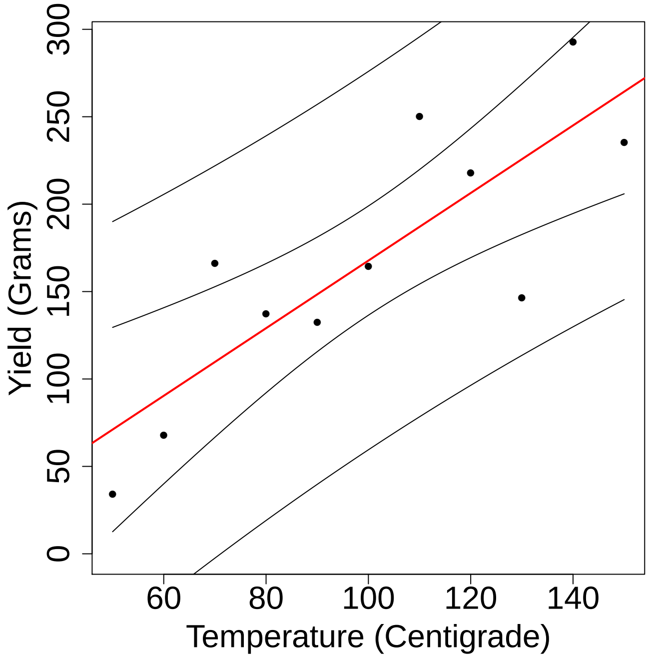
Therefore, the prediction interval is

$$Y_{n+1} \in \left[167.71 + 1.933(x_{n+1} - 100)\right] \mp 2.262 \times \sqrt{2097.60\left(1 + \frac{1}{11} + \frac{(x_{n+1} - 100)^2}{11000}\right)},$$

which is plotted in the figure.

(b) For part (b), the variance term drops the 1, giving

$$m(x) \in \Big[167.71 + 1.933(x-100)\Big] \mp 2.262 \times \sqrt{2097.60\left(\frac{1}{11} + \frac{(x-100)^2}{11000}\right)},$$



2. TBA

8.6 Analysis of Variance

1. TBA

8.7 Problems

1. Find confidence intervals for paired and two-sample t-tests.

Solution: Looking back at Sections 7.3.6.1 and 7.3.6.2, we find the two T statistics that we wish to use (as pivots) to form the confidence intervals. In the first case of paired data (x_i, y_i), we form the n differences $d_i = y_i - x_i$, and use

$$T_{n-1} = \frac{\bar{D} - \mu_D}{S_D/\sqrt{n}}$$

to obtain the CI

$$\mu_D \in \bar{D} \mp t_{n-1}(1-\alpha/2)\frac{S_D}{\sqrt{n}} \, .$$

In the second case, we follow Equations (7.20) and (7.21) and obtain

$$(\mu_Y - \mu_X) \in (\bar{Y} - \bar{X}) \mp t_{n-2}(1-\alpha/2)\sqrt{S_P^2\left(\frac{1}{n_X} + \frac{1}{n_Y}\right)} \, .$$

2. Show that the inequality in Equation (8.6) may be written in the form

$$Prob\left(\frac{\sqrt{n}\,\bar{X}}{\sqrt{\frac{(n-1)\,S^2}{n-1}}} < t_{0.975,n-1}\right) = 20\% \, ,$$

where $\sqrt{n}\bar{X} \sim N(\sqrt{n}\mu_1, 1)$. The quantity follows the non-central T distribution, where the numerator is $N(\mu, 1)$ rather than $N(0, 1)$. For the parameters in the example,

```
> pt( qt(.975,15), 15, 2.997865 ) = 0.20000;
```

where the noncentrality parameter $\mu = 2.997865$ was found by interpolation. Hence, $4\mu_1 = 2.997865$ or $\mu_1 = 0.7495$.

Solution: Follow the steps given to confirm the 80% calculation.

Chapter 9

Topics in Statistics

9.1 MSE and Histogram Bin Width Selection

9.1.1 MSE Criterion for Biased Estimators

9.1.2 Case Study: Optimal Histogram Bin Widths

9.1.3 Examples with Gaussian Data

9.1.4 Normal Reference Rules for the Histogram Bin Width

9.1.4.1 Scott's Rule

9.1.4.2 Friedman-Diaconis Rule

9.1.4.3 Sturges' Rule

9.1.4.4 Comparison of the Three Rules

9.2 Optimal Stopping Time Problem

9.3 Compound Random Variables

9.3.1 Computing Expectations with Conditioning

9.3.2 Compound Random Variables

9.4 Simulation and the Bootstrap

9.5 Multiple Linear Regression

9.6 Experimental Design

9.7 Logistic Regression, Poisson Regression, and the Generalized Linear Model

9.8 Robustness

9.9 Conclusions

Subject Index