INSTRUCTOR'S SOLUTIONS MANUAL

INSTRUCTOR'S SOLUTIONS MANUAL

to accompany

PROBABILITY AND STATISTICS

THIRD EDITION

Morris H. DeGroot

Mark J. Schervish
Carnegie Mellon University

PEARSON
Addison Wesley

Boston San Francisco New York
London Toronto Sydney Tokyo Singapore Madrid
Mexico City Munich Paris Cape Town Hong Kong Montreal

To the memory of Morrie DeGroot MJS

Reproduced by Addison-Wesley from files supplied by the author.

ISBN 0-201-87999-9

1 2 3 4 5 6 7 8 9 10 VG 05 04 03 02

Contents

Preface

This manual contains solutions to all of the exercises in *Probability and Statistics*, 3rd edition, by Morris DeGroot and myself. I have preserved most of Morrie's solutions to the exercises that existed in the 2nd edition. Certainly errors have been introduced, and I will post any errors brought to my attention on my web page `http://www.stat.cmu.edu/ mark/` along with errors in the text itself. Feel free to send me comments.

For instructors who are familiar with the 2nd edition, I hope that you will find the 3rd edition at least as useful. Some new material has been added, and only a small amount has been removed. Assuming that you will be spending the same amount of time using the text as before, something will have to be skipped. I have tried to arrange the material so that instructors can choose what to cover and what not to cover based on the type of course they want. This manual contains commentary on specific sections right before the solutions for those sections. This commentary is intended to explain special features of those sections and help instructors decide which parts they want to require of their students. Special attention is given to more challenging material and how the remainder of the text does or does not depend upon it.

To teach a mathematical statistics course for students with a strong calculus background, one could safely cover all of the material for which one could find time. The Bayesian sections include 4.9, 6.1, 6.2, 6.3, 6.4, 7.6, 8.8, and 10.4. One can choose to skip some or all of this material if one desires, but that would be ignoring one of the unique features of the text. The more challenging material in Sections 6.7, 6.8, 6.9, 8.2, 8.3, and 8.4 is really only suitable for a mathematical statistics course. One should try to make time for some of the material in Sections 11.1 and 11.2 even if it meant cutting back on some of the nonparametrics and two-way ANOVA. To teach a more modern statistics course, one could skip sections 6.7, 6.8, 6.9, 8.2, 8.3, 8.4, 9.8, 10.7, and 10.8. This would leave time to discuss robust estimation (Section 9.6) and simulation (Chapter 11). Section 2.4 on Markov chains is not actually necessary even if one wishes to introduce Markov chain Monte Carlo (Section 11.4), although it is helpful for understanding what this topic is about.

Mark J. Schervish

Chapter 1

1.2 Commentary

It is interesting to have the students determine some of their own subjective probabilities. For example, let X denote the temperature at noon tomorrow outside the building in which the class is being held. Have each student determine a number x_1 such that the student considers the following two possible outcomes to be equally likely: $X \leq x_1$ and $X > x_1$. Also, have each student determine numbers x_2 and x_3 (with $x_2 < x_3$) such that the student considers the following three possible outcomes to be equally likely: $X \leq x_2$, $x_2 < X < x_3$, and $X \geq x_3$. Determinations of more than three outcomes that are considered to be equally likely can also be made. The different values of x_1 determined by different members of the class should be discussed, and also the possibility of getting the class to agree on a common value of x_1.

Similar determinations of equally likely outcomes can be made by the students in the class for quantities such as the following ones which were found in the 1973 World Almanac and Book of Facts: the number of freight cars that were in use by American railways in 1960 (1,690,396), the number of banks in the United States which closed temporarily or permanently in 1931 on account of financial difficulties (2,294), and the total number of telephones which were in service in South America in 1971 (6,137,000).

1.4 Solutions to Exercises

1. Assume that $x \in B^c$. We need to show that $x \in A^c$. We shall show this indirectly. If $x \in A$, then $x \in B$ because $A \subset B$. This contradicts $x \in B^c$. Hence $x \in A$ is false and $x \in A^c$.

2. First, show that $A(B \cup C) \subset (AB) \cup (AC)$. Let $x \in A(B \cup C)$. Then $x \in A$ and $x \in B \cup C$. That is, $x \in A$ and either $x \in B$ or $x \in C$ (or both). So either ($x \in A$ and $x \in B$) or ($x \in A$ and $x \in C$) or both. That is, either $x \in AB$ or $x \in AC$. This is what it means to say that $x \in (AB) \cup (AC)$. Thus $A(B \cup C) \subset (AB) \cup (AC)$. Basically, running these steps backwards shows that $(AB) \cup (AC) \subset A(B \cup C)$.

3. To prove the first result, let $x \in (A \cup B)^c$. This means that x is not in $A \cup B$. In other words, x is neither in A nor in B. Hence $x \in A^c$ and $x \in B^c$. So $x \in A^c \cap B^c$. This proves that $(A \cup B)^c \subset A^c \cap B^c$. Next, suppose that $x \in A^c \cap B^c$. Then $x \in A^c$ and $x \in B^c$. So x is neither in A nor in B, so it can't be in $A \cup B$. Hence $x \in (A \cup B)^c$. This shows that $A^c \cap B^c \subset (A \cup B)^c$. The second result follows from the first by applying the first result to A^c and B^c and then taking complements of both sides.

4. To see that AB and AB^c are disjoint, let $x \in AB$. Then $x \in B$, hence $x \notin B^c$ and so $x \notin AB^c$. So no element of AB is in AB^c, hence the two events are disjoint. To prove that $A = (AB) \cup (AB^c)$, we shall show that each side is a subset of the other side. First, let $x \in A$. Either $x \in B$ or $x \in B^c$. If $x \in B$, then $x \in AB$. If $x \in B^c$, then $x \in AB^c$. Either way, $x \in (AB) \cup (AB^c)$. So every element of A is an element of $(AB) \cup (AB^c)$ and we conclude that $A \subset (AB) \cup (AB^c)$. Finally, let $x \in (AB) \cup (AB^c)$.

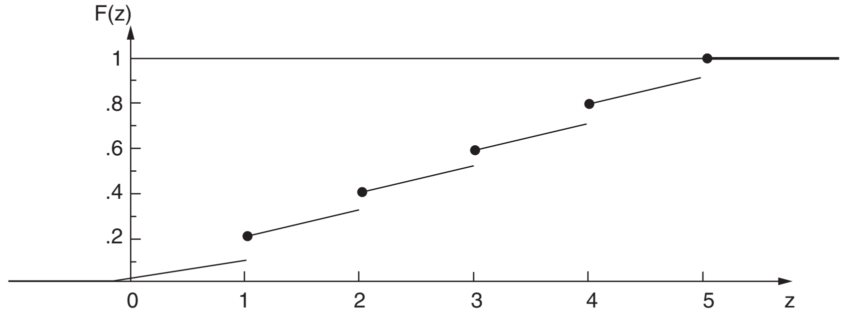

Figure S.3.33: Graph of d.f. for Exercise 1 of Section 3.10.

2. Let $x_1, \dots, x_k$ be the finitely many values for which $f_1(x) > 0$. Since X and Y are independent, the conditional distribution of $Z = X + Y$ given $X = x$ is the same as the distribution of $x + Y$, which has the p.d.f. $f_2(z - x)$, and the d.f. $F_2(z - x)$. By the law of total probability the d.f. of Z is $\sum_{i=1}^{k} F_2(z - x_i) f_1(x_i)$. Notice that this is a weighted average of continuous functions of z, $F_2(z - x_i)$ for $i = 1, \dots, k$, hence it is a continuous function. The p.d.f. of Z can easily be found by differentiating the d.f. to obtain $\sum_{i=1}^{k} f_2(z - x_i) f_1(x_i)$.

3. Since $F(x)$ is continuous and differentiable everywhere except at the points $x = 0$, 1, and 2,

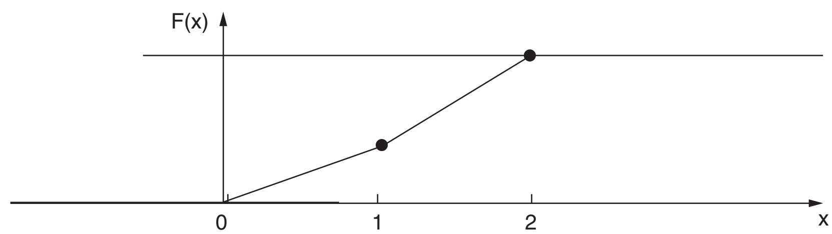

Figure S.3.34: Graph of d.f. for Exercise 3 of Section 3.10.

$$f(x) = \frac{dF(x)}{dx} \begin{cases} \frac{2}{5} & \text{for } 0 < x < 1, \\ \frac{3}{5} & \text{for } 1 < x < 2, \\ 0 & \text{otherwise.} \end{cases}$$

4. Since $f(x)$ is symmetric with respect to $x = 0$, $F(0) = Pr(X \le 0) = 0.5$. Hence,

$$\int_0^{x_0} f(x)\,dx = \frac{1}{2} \int_0^{x_0} \exp(-x)\,dx = .4.$$

It follows that $\exp(-x_0) = .2$ and $x_0 = \log 5$.

5. X_1 and X_2 have a uniform distribution over the square, which has area 1. The area of the quarter circle in Figure S.3.35, which is the required probability, is $\pi/4$.

6. (a) $$\Pr(X \text{ divisible by } n) = f(n) + f(2n) + f(3n) + \cdots = \sum_{x=1}^{\infty} \frac{1}{c(p)} \frac{1}{(nx)^p} = \frac{1}{n^p}.$$

6.4 Commentary

We introduce the fundamental concepts of Bayesian decision theory. The use of a loss function arises again in Bayesian hypothesis testing in Section 8.8. This section ends with foundational discussion of the limitations of Bayes estimators. This material is included for those instructors who want their students to have both a working and a critical understanding of the topic.

6.4 Solutions to Exercises

1. The posterior distribution of θ would be a beta distribution with parameters 2 and 1. The mean of the posterior distribution is $2/3$, which would be the Bayes estimate under squared error loss. The median of the posterior distribution would be the Bayes estimate under absolute error loss. To find the median, write the d.f. as

$$F(\theta) = \int_0^\theta 2t dt = \theta^2,$$

for $0 < \theta < 1$. The quantile function is then $F^{-1}(p) = p^{1/2}$, so the median is $(1/2)^{1/2} = 0.7071$.

2. The posterior distribution of θ is a beta distribution with parameters $5 + 1 = 6$ and $10 + 19 = 29$. The mean of this distribution is $6/(6 + 29) = 6/35$. Therefore, the Bayes estimate of θ is $6/35$.

3. If y denotes the number of defective items in the sample, then the posterior distribution of θ will be a beta distribution with parameters $5 + y$ and $10 + 20 - y = 30 - y$. The variance V of this beta distribution is

$$V = \frac{(5 + y)(30 - y)}{(35)^2(36)}.$$

Since the Bayes estimate of θ is the mean μ of the posterior distribution, the mean squared error of this estimate is $E[(\theta - \mu)^2 \mid \boldsymbol{x}]$, which is the variance V of the posterior distribution.

(a) V will attain its maximum at a value of y for which $(5+y)(30-y)$ is a maximum. By differentiating with respect to y and setting the derivative equal to 0, we find that the maximum is attained when $y = 12.5$. Since the number of defective items y must be an integer, the maximum of V will be attained for $y = 12$ or $y = 13$. When these values are substituted into $(5 + y)(30 - y)$, it is found that they both yield the same value.

(b) Since $(5+y)(30-y)$ is a quadratic function of y and the coefficient of y^2 is negative, its minimum value over the interval $0 \le y \le 20$ will be attained at one of the endpoints of the interval. It is found that the value for $y = 0$ is smaller than the value for $y = 20$.

4. Suppose that the parameters of the prior beta distribution of θ are α and β. Then $\mu_0 = \alpha/(\alpha+\beta)$. As shown in Example 6.4.1, the mean of the posterior distribution of θ is

$$\frac{\alpha + \sum_{i=1}^n X_i}{\alpha + \beta + n} = \frac{\alpha + \beta}{\alpha + \beta + n}\mu_0 + \frac{n}{\alpha + \beta + n}\overline{X}_n.$$

Hence, $\gamma_n = n/(\alpha + \beta + n)$ and $\gamma_n \to 1$ *as* $n \to \infty$.

5. It was shown in Exercise 5 of Sec. 6.3 that the posterior distribution of θ is a gamma distribution with parameters $\alpha = 16$ and $\beta = 6$. The Bayes estimate of θ is the mean of this distribution and is equal to $16/6 = 8/3$.

The statistic Q is then 12.96. The two p-values for 10 and 11 degrees of freedom are 0.2258 and 0.2959.

7. There is no single correct answer to this problem. The M.L.E.'s $\hat{\mu} = \overline{X}_n$ and $\hat{\sigma}^2 = S_n^2/n$ should be calculated from the given observations. These observations should then be grouped into intervals and the observed number in each interval compared with the expected number in that interval if each of the 50 observations had a normal distribution with mean $\overline{X}_n$ and variance S_n^2/n. If the number of intervals is k, then when H_0 is true, the approximate distribution of the statistic Q will lie between the χ^2 distribution with $k-3$ degrees of freedom and the χ^2 distribution with $k-1$ degrees of freedom.

8. There is no single correct answer to this problem. The M.L.E. $\hat{\beta} = 1/\overline{X}_n$ of the parameter of the exponential distribution should be calculated from the given observations. These observations should then be grouped into intervals and the observed number in each interval compared with the expected number in that interval if each of the 50 observations had an exponential distribution with parameter $1/\overline{X}_n$. If the number of intervals is k, then when H_0 is true, the approximate distribution of the statistic Q will lie between a χ^2 distribution with $k-2$ degrees of freedom and a χ^2 distribution with $k-1$ degrees of freedom.

9.3 Solutions to Exercises.

1. Table S.9.1 contains the expected counts for this example. The value of the χ^2 statistic Q calculated

Table S.9.1: Expected cell counts for Exercise 1 of Section 9.3.

	Good grades	Athletic ability	Popularity
Boys	117.3	42.7	67.0
Girls	129.7	47.3	74.0

from these data is $Q = 21.5$. This should be compared to a χ^2 distribution with two degrees of freedom. The tail area can be calculated using statistical software as 2.2×10^{-5}.

2. $Q = \sum_{i=1}^{R}\sum_{j=1}^{C} \frac{(N_{ij} - \hat{E}_{ij})^2}{\hat{E}_{ij}} = \sum_{i=1}^{R}\sum_{j=1}^{C} \left(\frac{N_{ij}^2}{\hat{E}_{ij}} - 2N_{ij} + \hat{E}_{ij}\right) = \left(\sum_{i=1}^{R}\sum_{j=1}^{C} \frac{N_{ij}^2}{\hat{E}_{ij}}\right) - 2n + n$
$= \left(\sum_{i=1}^{R}\sum_{j=1}^{C} \frac{N_{ij}^2}{\hat{E}_{ij}}\right) - n.$

3. By Exercise 2,

$$Q = \sum_{i=1}^{R} \frac{N_{i1}^2}{\hat{E}_{i1}} + \sum_{i=1}^{R} \frac{N_{i2}^2}{\hat{E}_{i2}} - n.$$

But

$$\sum_{i=1}^{R} \frac{N_{i2}^2}{\hat{E}_{i2}} = \sum_{i=1}^{R} \frac{(N_{i+} - N_{i1})^2}{\hat{E}_{i2}} = \sum_{i=1}^{R} \frac{N_{i+}^2}{\hat{E}_{i2}} - 2\sum_{i=1}^{R} \frac{N_{i+}N_{i1}}{\hat{E}_{i2}} + \sum_{i=1}^{R} \frac{N_{i1}^2}{\hat{E}_{i2}}.$$

2. The plots for this exercise are formed the same way as that in Exercise 1 except we replace the normal pseudo-random values by the appropriate gamma pseudo-random values and we replace Φ^{-1} by the quantile function of the appropriate gamma distribution. Two of the plots are in Figure S.11.8. The plots are pretty straight except in the extreme upper tail, where things are expected to be highly variable.

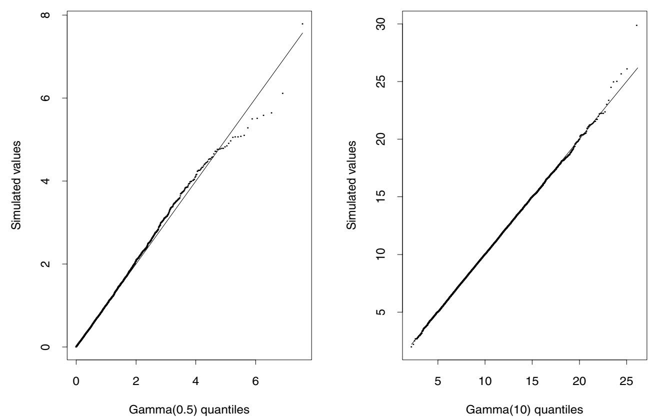

Figure S.11.8: Gamma quantile plots for Exercise 2 in Section 11.6. The left plot has parameters 0.5 and 1 and the right plot has parameters 10 and 1. Straight lines have been added for reference.

3. Once again, the plots are drawn in a fashion similar to Exercise 1. This time, we notice that the plot with one degree of freedom has some really serious non-linearity. This is the Cauchy distribution which has very long tails. The extreme observations from a Cauchy sample are very variable. Two of the plots are in Figure S.11.9.

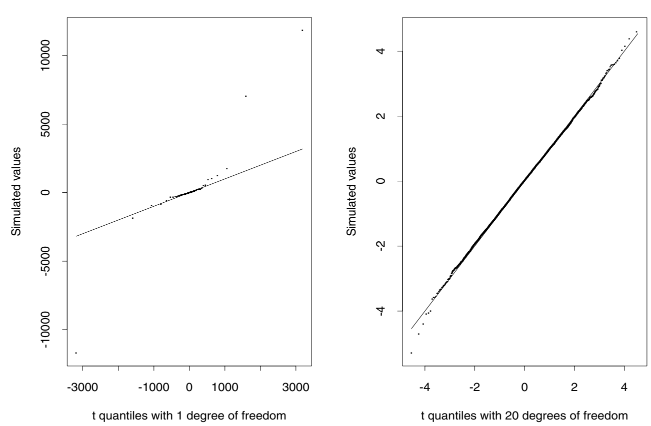

Figure S.11.9: Two t quantile plots for Exercise 3 in Section 11.6. The left plot has 1 degree of freedom, and the right plot has 20 degrees of freedom. Straight lines have been added for reference.

4. (a) I simulated 1000 pairs three times and got the following average values: 1.478, 1.462, 1.608. It looks like 1000 is not enough to be very confident of getting the average within 0.01.