Solutions Manual for

Statistical Inference, Second Edition

George Casella
University of Florida

Roger L. Berger
North Carolina State University

Damaris Santana
University of Florida

"When I hear you give your reasons," I remarked, "the thing always appears to me to be so ridiculously simple that I could easily do it myself, though at each successive instance of your reasoning I am baffled until you explain your process."

Dr. Watson to Sherlock Holmes
A Scandal in Bohemia

0.1 Description

This solutions manual contains solutions for *all odd numbered problems* plus a large number of solutions for even numbered problems. Of the 624 exercises in *Statistical Inference, Second Edition*, this manual gives solutions for 484 (78%) of them. There is an obtuse pattern as to which solutions were included in this manual. We assembled all of the solutions that we had from the first edition, and filled in so that all odd-numbered problems were done. In the passage from the first to the second edition, problems were shuffled with no attention paid to numbering (hence no attention paid to minimize the new effort), but rather we tried to put the problems in logical order.

A major change from the first edition is the use of the computer, both symbolically through Mathematica^{tm} and numerically using *R*. Some solutions are given as code in either of these languages. Mathematica^{tm} can be purchased from Wolfram Research, and *R* is a free download from `http://www.r-project.org/`.

Here is a detailed listing of the solutions included.

Chapter	Number of Exercises	Number of Solutions	Missing
1	55	51	26, 30, 36, 42
2	40	37	34, 38, 40
3	50	42	4, 6, 10, 20, 30, 32, 34, 36
4	65	52	8, 14, 22, 28, 36, 40 48, 50, 52, 56, 58, 60, 62
5	69	46	2, 4, 12, 14, 26, 28 all even problems from 36 − 68
6	43	35	8, 16, 26, 28, 34, 36, 38, 42
7	66	52	4, 14, 16, 28, 30, 32, 34, 36, 42, 54, 58, 60, 62, 64
8	58	51	36, 40, 46, 48, 52, 56, 58
9	58	41	2, 8, 10, 20, 22, 24, 26, 28, 30 32, 38, 40, 42, 44, 50, 54, 56
10	48	26	all even problems except 4 and 32
11	41	35	4, 20, 22, 24, 26, 40
12	31	16	all even problems

0.2 Acknowledgement

Many people contributed to the assembly of this solutions manual. We again thank all of those who contributed solutions to the first edition – many problems have carried over into the second edition. Moreover, throughout the years a number of people have been in constant touch with us, contributing to both the presentations and solutions. We apologize in advance for those we forget to mention, and we especially thank Jay Beder, Yong Sung Joo, Michael Perlman, Rob Strawderman, and Tom Wehrly. Thank you all for your help.

And, as we said the first time around, although we have benefited greatly from the assistance and comments of others in the assembly of this manual, we are responsible for its ultimate correctness. To this end, we have tried our best but, as a wise man once said, "You pays your money and you takes your chances."

George Casella
Roger L. Berger
Damaris Santana

December, 2001

Chapter 1

Probability Theory

"If any little problem comes your way, I shall be happy, if I can, to give you a hint or two as to its solution."

Sherlock Holmes
The Adventure of the Three Students

1.1 a. Each sample point describes the result of the toss (H or T) for each of the four tosses. So, for example THTT denotes T on 1st, H on 2nd, T on 3rd and T on 4th. There are $2^4 = 16$ such sample points.

b. The number of damaged leaves is a nonnegative integer. So we might use $S = \{0, 1, 2, \ldots\}$.

c. We might observe fractions of an hour. So we might use $S = \{t : t \geq 0\}$, that is, the half infinite interval $[0, \infty)$.

d. Suppose we weigh the rats in ounces. The weight must be greater than zero so we might use $S = (0, \infty)$. If we know no 10-day-old rat weighs more than 100 oz., we could use $S = (0, 100]$.

e. If n is the number of items in the shipment, then $S = \{0/n, 1/n, \ldots, 1\}$.

1.2 For each of these equalities, you must show containment in both directions.

a. $x \in A \backslash B \Leftrightarrow x \in A$ and $x \notin B \Leftrightarrow x \in A$ and $x \notin A \cap B \Leftrightarrow x \in A \backslash (A \cap B)$. Also, $x \in A$ and $x \notin B \Leftrightarrow x \in A$ and $x \in B^c \Leftrightarrow x \in A \cap B^c$.

b. Suppose $x \in B$. Then either $x \in A$ or $x \in A^c$. If $x \in A$, then $x \in B \cap A$, and, hence $x \in (B \cap A) \cup (B \cap A^c)$. Thus $B \subset (B \cap A) \cup (B \cap A^c)$. Now suppose $x \in (B \cap A) \cup (B \cap A^c)$. Then either $x \in (B \cap A)$ or $x \in (B \cap A^c)$. If $x \in (B \cap A)$, then $x \in B$. If $x \in (B \cap A^c)$, then $x \in B$. Thus $(B \cap A) \cup (B \cap A^c) \subset B$. Since the containment goes both ways, we have $B = (B \cap A) \cup (B \cap A^c)$. (Note, a more straightforward argument for this part simply uses the Distributive Law to state that $(B \cap A) \cup (B \cap A^c) = B \cap (A \cup A^c) = B \cap S = B$.)

c. Similar to part a).

d. From part b).
$A \cup B = A \cup [(B \cap A) \cup (B \cap A^c)] = A \cup (B \cap A) \cup A \cup (B \cap A^c) = A \cup [A \cup (B \cap A^c)] = A \cup (B \cap A^c)$.

1.3 a. $x \in A \cup B \;\Leftrightarrow\; x \in A$ or $x \in B \;\Leftrightarrow\; x \in B \cup A$
$x \in A \cap B \;\Leftrightarrow\; x \in A$ and $x \in B \;\Leftrightarrow\; x \in B \cap A$.

b. $x \in A \cup (B \cup C) \Leftrightarrow\; x \in A$ or $x \in B \cup C \Leftrightarrow\; x \in A \cup B$ or $x \in C \Leftrightarrow\; x \in (A \cup B) \cup C$.
(It can similarly be shown that $A \cup (B \cup C) = (A \cup C) \cup B$.)
$x \in A \cap (B \cap C) \Leftrightarrow\; x \in A$ and $x \in B$ and $x \in C \Leftrightarrow\; x \in (A \cap B) \cap C$.

c. $x \in (A \cup B)^c \Leftrightarrow\; x \notin A$ or $x \notin B \Leftrightarrow\; x \in A^c$ and $x \in B^c \Leftrightarrow\; x \in A^c \cap B^c$
$x \in (A \cap B)^c \Leftrightarrow\; x \notin A \cap B \Leftrightarrow\; x \notin A$ and $x \notin B \Leftrightarrow\; x \in A^c$ or $x \in B^c \Leftrightarrow\; x \in A^c \cup B^c$.

1.4 a. "A or B or both" is $A \cup B$. From Theorem 1.2.9b we have $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Then,

$$
\begin{aligned}
f_U(u) &= \frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma)} u^{\alpha-1} \int_u^1 v^{\beta-1}(1-v)^{\gamma-1}\left(\frac{v-u}{v}\right)^{\beta-1} dv \\
&= \frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma)} u^{\alpha-1}(1-u)^{\beta+\gamma-1} \int_0^1 y^{\beta-1}(1-y)^{\gamma-1} dy \quad \left(y = \frac{v-u}{1-u}, dy = \frac{dv}{1-u}\right) \\
&= \frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma)} u^{\alpha-1}(1-u)^{\beta+\gamma-1} \frac{\Gamma(\beta)\Gamma(\gamma)}{\Gamma(\beta+\gamma)} \\
&= \frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma(\alpha)\Gamma(\beta+\gamma)} u^{\alpha-1}(1-u)^{\beta+\gamma-1}, \qquad 0 < u < 1.
\end{aligned}
$$

Thus, $U \sim \text{gamma}(\alpha, \beta+\gamma)$.

b. Let $x = \sqrt{uv}$, $y = \sqrt{\frac{u}{v}}$ then

$$
J = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial x}{\partial v} \end{vmatrix} = \begin{vmatrix} \frac{1}{2}v^{1/2}u^{-1/2} & \frac{1}{2}u^{1/2}v^{-1/2} \\ \frac{1}{2}v^{-1/2}u^{-1/2} & -\frac{1}{2}u^{1/2}v^{-3/2} \end{vmatrix} = \frac{1}{2v}.
$$

$$
f_{U,V}(u,v) = \frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma)} (\sqrt{uv}^{\alpha-1}(1-\sqrt{uv})^{\beta-1} \left(\sqrt{\frac{u}{v}}\right)^{\alpha+\beta-1} \left(1-\sqrt{\frac{u}{v}}\right)^{\gamma-1} \frac{1}{2v}.
$$

The set $\{0 < x < 1, 0 < y < 1\}$ is mapped onto the set $\{0 < u < v < \frac{1}{u}, 0 < u < 1\}$. Then,

$$
\begin{aligned}
&f_U(u) \\
&= \int_u^{1/u} f_{U,V}(u,v) dv \\
&= \underbrace{\frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma)} u^{\alpha-1}(1-u)^{\beta+\gamma-1}}_{\text{Call it A}} \int_u^{1/u} \left(\frac{1-\sqrt{uv}}{1-u}\right)^{\beta-1} \left(\frac{1-\sqrt{u/v}}{1-u}\right)^{\gamma-1} \frac{(\sqrt{u/v})^{\beta}}{2v(1-u)} dv.
\end{aligned}
$$

To simplify, let $z = \frac{\sqrt{u/v}-u}{1-u}$. Then $v = u \Rightarrow z = 1$, $v = 1/u \Rightarrow z = 0$ and $dz = -\frac{\sqrt{u/v}}{2(1-u)v} dv$. Thus,

$$
\begin{aligned}
f_U(u) &= A \int z^{\beta-1}(1-z)^{\gamma-1} dz \qquad (\text{ kernel of beta}(\beta,\gamma)) \\
&= \frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma)} u^{\alpha-1}(1-u)^{\beta+\gamma-1} \frac{\Gamma(\beta)\Gamma(\gamma)}{\Gamma(\beta+\gamma)} \\
&= \frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma(\alpha)\Gamma(\beta+\gamma)} u^{\alpha-1}(1-u)^{\beta+\gamma-1}, \qquad 0 < u < 1.
\end{aligned}
$$

That is, $U \sim \text{beta}(\alpha, \beta+\gamma)$, as in a).

4.24 Let $z_1 = x + y$, $z_2 = \frac{x}{x+y}$, then $x = z_1 z_2$, $y = z_1(1-z_2)$ and

$$
|J| = \begin{vmatrix} \frac{\partial x}{\partial z_1} & \frac{\partial x}{\partial z_2} \\ \frac{\partial y}{\partial z_1} & \frac{\partial y}{\partial z_2} \end{vmatrix} = \begin{vmatrix} z_2 & z_1 \\ 1-z_2 & -z_1 \end{vmatrix} = z_1.
$$

The set $\{x > 0, y > 0\}$ is mapped onto the set $\{z_1 > 0, 0 < z_2 < 1\}$.

$$
\begin{aligned}
f_{Z_1,Z_2}(z_1,z_2) &= \frac{1}{\Gamma(r)}(z_1 z_2)^{r-1} e^{-z_1 z_2} \cdot \frac{1}{\Gamma(s)}(z_1 - z_1 z_2)^{s-1} e^{-z_1 + z_1 z_2} z_1 \\
&= \frac{1}{\Gamma(r+s)} z_1^{r+s-1} e^{-z_1} \cdot \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} z_2^{r-1}(1-z_2)^{s-1}, \qquad 0 < z_1, 0 < z_2 < 1.
\end{aligned}
$$

The likelihood function is

$$L(\theta|\mathbf{x}) = \prod_{i=1}^{n} \frac{1}{\theta} I_{[0,\theta]}(x_i) = \frac{1}{\theta^n} I_{[0,\theta]}(x_{(n)}) I_{[0,\infty)}(x_{(1)}),$$

where $x_{(1)}$ and $x_{(n)}$ are the smallest and largest order statistics. For $\theta \ge x_{(n)}$, $L = 1/\theta^n$, a decreasing function. So for $\theta \ge x_{(n)}$, L is maximized at $\hat\theta = x_{(n)}$. $L = 0$ for $\theta < x_{(n)}$. So the overall maximum, the MLE, is $\hat\theta = X_{(n)}$. The pdf of $\hat\theta = X_{(n)}$ is nx^{n-1}/θ^n, $0 \le x \le \theta$. This can be used to calculate

$$\mathrm{E}\,\hat\theta = \frac{n}{n+1}\theta, \quad \mathrm{E}\,\hat\theta^2 = \frac{n}{n+2}\theta^2 \quad \text{and} \quad \mathrm{Var}\,\hat\theta = \frac{n\theta^2}{(n+2)(n+1)^2}.$$

$\tilde\theta$ is an unbiased estimator of θ; $\hat\theta$ is a biased estimator. If n is large, the bias is not large because $n/(n+1)$ is close to one. But if n is small, the bias is quite large. On the other hand, $\mathrm{Var}\,\hat\theta < \mathrm{Var}\,\tilde\theta$ for all θ. So, if n is large, $\hat\theta$ is probably preferable to $\tilde\theta$.

7.10 a. $f(\mathbf{x}|\theta) = \prod_i \frac{\alpha}{\beta^\alpha} x_i^{\alpha-1} I_{[0,\beta]}(x_i) = \left(\frac{\alpha}{\beta^\alpha}\right)^n \left(\prod_i x_i\right)^{\alpha-1} I_{(-\infty,\beta]}(x_{(n)}) I_{[0,\infty)}(x_{(1)}) = L(\alpha,\beta|\mathbf{x})$. By the Factorization Theorem, $(\prod_i X_i, X_{(n)})$ are sufficient.

b. For any fixed α, $L(\alpha,\beta|\mathbf{x}) = 0$ if $\beta < x_{(n)}$, and $L(\alpha,\beta|\mathbf{x})$ a decreasing function of β if $\beta \ge x_{(n)}$. Thus, $X_{(n)}$ is the MLE of β. For the MLE of α calculate

$$\frac{\partial}{\partial\alpha}\log L = \frac{\partial}{\partial\alpha}\left[n\log\alpha - n\alpha\log\beta + (\alpha-1)\log\prod_i x_i\right] = \frac{n}{\alpha} - n\log\beta + \log\prod_i x_i.$$

Set the derivative equal to zero and use $\hat\beta = X_{(n)}$ to obtain

$$\hat\alpha = \frac{n}{n\log X_{(n)} - \log\prod_i X_i} = \left[\frac{1}{n}\sum_i (\log X_{(n)} - \log X_i)\right]^{-1}.$$

The second derivative is $-n/\alpha^2 < 0$, so this is the MLE.

c. $X_{(n)} = 25.0$, $\log\prod_i X_i = \sum_i \log X_i = 43.95 \Rightarrow \hat\beta = 25.0$, $\hat\alpha = 12.59$.

7.11 a.

$$\begin{aligned} f(\mathbf{x}|\theta) &= \prod_i \theta x_i^{\theta-1} = \theta^n \left(\prod_i x_i\right)^{\theta-1} = L(\theta|\mathbf{x}) \\ \frac{d}{d\theta}\log L &= \frac{d}{d\theta}\left[n\log\theta + (\theta-1)\log\prod_i x_i\right] = \frac{n}{\theta} + \sum_i \log x_i. \end{aligned}$$

Set the derivative equal to zero and solve for θ to obtain $\hat\theta = (-\frac{1}{n}\sum_i \log x_i)^{-1}$. The second derivative is $-n/\theta^2 < 0$, so this is the MLE. To calculate the variance of $\hat\theta$, note that $Y_i = -\log X_i \sim$ exponential$(1/\theta)$, so $-\sum_i \log X_i \sim$ gamma$(n, 1/\theta)$. Thus $\hat\theta = n/T$, where $T \sim$ gamma$(n, 1/\theta)$. We can either calculate the first and second moments directly, or use the fact that $\hat\theta$ is inverted gamma (page 51). We have

$$\begin{aligned} \mathrm{E}\frac{1}{T} &= \frac{\theta^n}{\Gamma(n)}\int_0^\infty \frac{1}{t} t^{n-1} e^{-\theta t}\, dt = \frac{\theta^n}{\Gamma(n)}\frac{\Gamma(n-1)}{\theta^{n-1}} = \frac{\theta}{n-1}. \\ \mathrm{E}\frac{1}{T^2} &= \frac{\theta^n}{\Gamma(n)}\int_0^\infty \frac{1}{t^2} t^{n-1} e^{-\theta t}\, dt = \frac{\theta^n}{\Gamma(n)}\frac{\Gamma(n-2)}{\theta^{n-2}} = \frac{\theta^2}{(n-1)(n-2)}, \end{aligned}$$

Therefore, the UMVUE is

$$\mathrm{E}\left(T\left|\sum_{i=1}^{n+1} X_i = y\right.\right) = \begin{cases} 0 & \text{if } y=0 \\ \frac{\binom{n}{y}p^y(1-p)^{n-y}(1-p)}{\binom{n+1}{y}p^y(1-p)^{n-y+1}} = \frac{\binom{n}{y}}{\binom{n+1}{y}} = \frac{1}{(n+1)(n+1-y)} & \text{if } y=1 \text{ or } 2 \\ \frac{\left(\binom{n}{y}+\binom{n}{y-1}\right)p^y(1-p)^{n-y+1}}{\binom{n+1}{y}p^y(1-p)^{n-y+1}} = \frac{\binom{n}{y}+\binom{n}{y-1}}{\binom{n+1}{y}} = 1 & \text{if } y>2. \end{cases}$$

7.59 We know $T=(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$. Then

$$\mathrm{E}\,T^{p/2} = \frac{1}{\Gamma\left(\frac{n-1}{2}\right)2^{\frac{n-1}{2}}}\int_0^\infty t^{\frac{p+n-1}{2}-1}e^{-\frac{t}{2}}\,dt = \frac{2^{\frac{p}{2}}\Gamma\left(\frac{p+n-1}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} = C_{p,n}.$$

Thus

$$\mathrm{E}\left(\frac{(n-1)S^2}{\sigma^2}\right)^{p/2} = C_{p,n},$$

so $(n-1)^{p/2}S^p\Big/C_{p,n}$ is an unbiased estimator of σ^p. From Theorem 6.2.25, $(\bar{X}, S^2)$ is a complete, sufficient statistic. The unbiased estimator $(n-1)^{p/2}S^p\Big/C_{p,n}$ is a function of $(\bar{X}, S^2)$. Hence, it is the best unbiased estimator.

7.61 The pdf for $Y \sim \chi^2_\nu$ is

$$f(y) = \frac{1}{\Gamma(\nu/2)2^{\nu/2}}y^{\nu/2-1}e^{-y/2}.$$

Thus the pdf for $S^2 = \sigma^2 Y/\nu$ is

$$g(s^2) = \frac{\nu}{\sigma^2}\frac{1}{\Gamma(\nu/2)2^{\nu/2}}\left(\frac{s^2\nu}{\sigma^2}\right)^{\nu/2-1}e^{-s^2\nu/(2\sigma^2)}.$$

Thus, the log-likelihood has the form (gathering together constants that do not depend on s^2 or σ^2)

$$\log L(\sigma^2|s^2) = \log\left(\frac{1}{\sigma^2}\right) + K\log\left(\frac{s^2}{\sigma^2}\right) - K'\frac{s^2}{\sigma^2} + K'',$$

where $K>0$ and $K'>0$.

The loss function in Example 7.3.27 is

$$L(\sigma^2, a) = \frac{a}{\sigma^2} - \log\left(\frac{a}{\sigma^2}\right) - 1,$$

so the loss of an estimator is the negative of its likelihood.

7.63 Let $a = \tau^2/(\tau^2+1)$, so the Bayes estimator is $\delta^\pi(x) = ax$. Then $R(\mu, \delta^\pi) = (a-1)^2\mu^2 + a^2$. As τ^2 increases, $R(\mu, \delta^\pi)$ becomes flatter.

7.65 a. Figure omitted.

b. The posterior expected loss is $\mathrm{E}\left(L(\theta,a)|x\right) = e^{ca}\mathrm{E}\,e^{-c\theta} - c\mathrm{E}(a-\theta) - 1$, where the expectation is with respect to $\pi(\theta|x)$. Then

$$\frac{d}{da}\mathrm{E}\left(L(\theta,a)|x\right) = ce^{ca}\mathrm{E}\,e^{-c\theta} - c \stackrel{\text{set}}{=} 0,$$

and $a = -\frac{1}{c}\log\mathrm{E}\,e^{-c\theta}$ is the solution. The second derivative is positive, so this is the minimum.