Solutions to the Exercises of Section 4.3.

4.3.1. The parameter space is $\Theta = \{(\theta, j) : 0 \le \theta \le 1, j = 1, \ldots, n\}$, the action space is $\mathcal{A} = [0,1]$, and the loss function is $L((\theta, j), a) = (\theta - a)^2$. Under (θ, j), the observations, $X_1, \ldots, X_n$ are independent, X_j is $\mathcal{B}(1,\theta)$ and X_i is $\mathcal{B}(1,1/2)$ for $i \ne j$.

The problem is invariant under the permutations of the observations, $g_\pi(x_1, \ldots, x_n) = (x_{\pi(1)}, \ldots, x_{\pi(n)})$ for permutations π of $(1, \ldots, n)$, with $\bar{g}_\pi(\theta, j) = (\theta, \pi^{-1}(j))$ and $\tilde{g}_\pi(a) = a$. It is also invariant under the map $g_2(x_1, \ldots, x_n) = (1 - x_1, \ldots, 1 - x_n)$ with $\bar{g}_2(\theta, j) = (1 - \theta, j)$ and $\tilde{g}_2(a) = 1 - a$. We may restrict attention to nonrandomized invariant rules. A rule d is invariant under g_π if $d(g_\pi(x_1, \ldots, x_n)) = \tilde{g}_\pi d(x_1, \ldots, x_n) = d(x_1, \ldots, x_n)$. This means that d depends on the x_i only through the sum $S = X_1 + \cdots + X_n$. We henceforth write invariant rules as $d(s)$. Such a rule is invariant under g_2 if $d(s) = 1 - d(n - s)$. In particular, if n is even then $d(n/2) = 1/2$.

The distribution of S under (θ, j) is independent of j and has mass function,

$$P_\theta(S = s) = \begin{cases} (\frac{1}{2})^{n-1}(1-\theta) & \text{for } s = 0 \\ \binom{n-1}{s-1}(\frac{1}{2})^{n-1}\theta + \binom{n-1}{s}(\frac{1}{2})^{n-1}(1-\theta) & \text{for } s = 1, \ldots, n-1 \\ (\frac{1}{2})^{n-1}\theta & \text{for } s = n. \end{cases}$$

Note that $P_\theta(S = s)$ is linear in θ. The risk function of an invariant rule is $R((\theta, j), d) = \mathrm{E}_\theta(\theta - d(S))^2 = \theta^2 - 2\theta \mathrm{E}_\theta d(S) + \mathrm{E}_\theta d(S)^2$. Now note that $\mathrm{E}_\theta d(S)$ and $\mathrm{E}_\theta d(S)^2$ are linear in θ. This implies that $R((\theta, j), d)$ is quadratic in θ. And since d is invariant under g_2, $R((\theta, j), d)$ is symmetric in θ about 1/2. Therefore the maximum of $R((\theta, j), d)$ over θ occurs at $\theta = 1/2$ or at $\theta = 0$ and $\theta = 1$.

Below, we show that $d_0(s) = s/n$ minimizes $R((0, j), d) = R((1, j), d)$ and then we show that this minimum value is greater than $R((1/2, j), d_0)$. First note that $R((0, j), d) = (1/2)R((0, j), d) + (1/2)R((1, j), d)$:

$$R((0,j),d) = \frac{1}{2}\sum_{s=0}^{n-1}\binom{n-1}{s}(\frac{1}{2})^{n-1}d(s)^2 + \frac{1}{2}\sum_{s=1}^{n}\binom{n-1}{s-1}(\frac{1}{2})^{n-1}(1-d(s))^2$$

We wish to find d to minimize this subject to the restriction that d be invariant, i.e. $d(s) = 1 - d(n - s)$ for $s = 0, \ldots, n$. The overall minimum without regard to the restriction is easily found by setting the derivatives of $R((0, j), d)$ with respect to the $d(s)$ separately to zero. We find $d(s) = s/n$. This satisfies the restriction and so gives the minimum value subject to the restriction.

To show that $R((0, j), d_0) = \mathrm{E}_0(S/n)^2$ is greater than $R((1/2, j), d_0) = \mathrm{E}_{1/2}(S/n - 1/2)^2$, we evaluate both. When $\theta = 1/2$, S has a binomial distribution sample size n and probability of success 1/2, so $R((1/2, j), d_0) = \mathrm{Var}_{1/2}(S/n) = 1/(4n)$. When $\theta = 0$, S has a binomial distirbution sample size $n - 1$ and success probability 1/2, so

$$R((0,j),d_0) = \mathrm{Var}_0(S/n) + (\mathrm{E}_0(S/n))^2 = \frac{n-1}{4n^2} + \frac{(n-1)^2}{4n^2} = \frac{n-1}{4n}$$

This is greater than $1/4n$ for all $n > 1$.

4.3.2. From Lemma 2.11.1, the least favorable distribution must give all its weight to points p for which $R(p, 1/4) = v = 1/4$. As in Figure 4.1, this occurs only at the three points 0, 1/2 and 1. From Theorem 3(c), we may restrict attention to invariant prior distributions, those that distribute weight symmetrically about 1/2. Thus a least favorable distribution τ must be of the form $\tau(0) = \tau(1) = z$ and $\tau(1/2) = 1 - 2z$ for some $0 \le z \le 1/2$. The Bayes risk of such a prior is

$$r(\tau, x) = z[R(0, x) + R(1, x)] + (1 - 2z)R(1/2, x) = z2x + (1 - 2z)|1 - 2x|.$$

Since this is increasing in x for $x > 1/2$ we may assume $x \le 1/2$, and write $r(\tau, x) = z2x + (1-2z)(1-2x) = (1/2) - z + x(1 - 4z)$. If $z = 1/4$, this is constant in x with value 1/4. Since 1/4 is the minimax value, τ is least favorable.

4.3.3. From Theorem 3, we may search among the invariant priors for a least favorable τ_0. If a prior is invariant under $\bar{g}_1$, then for all θ it must assign equal weight to $(\theta, 1)$ and $(\theta, 2)$. If it is invariant under

$\bar{g}_2$, then for $i = 1, 2$ and all θ it must assign equal weight to (θ, i) and $(1-\theta, i)$. However, the risk of a nonrandomized invariant rule, z, was found to be $R((\theta, i), z) = 2z\theta^2 - 2z\theta + z^2/2 + 1/8$ and to be maximized at $\theta = 0$ and $\theta = 1$. Therefore the invariant prior, τ_0, that gives mass $1/4$ to each of $(0,1)$, $(0,2)$, $(1,1)$, and $(1,2)$, is least favorable: Its Bayes risk is the average of $R((\theta, i), z)$ over these four points and so is $r(\tau_0, z) = z^2/2 + 1/8$, whose minimum over z is $1/8$, the minimax value.

4.3.4. From Exercise 4.2.7(b), we know that a behavioral invariant rule chooses an action at random independent of the observations. For any such distribution, δ, we may find, for a given $\epsilon > 0$, a number Δ such that δ assigns $1-\epsilon$ of its mass to the interval $(0, \Delta/2)$. Then, $R((\Delta, \Sigma), \delta) \geq 1 - \epsilon$. This shows that $\sup_\theta R(\theta, \delta) = 1$. Yet, if $d(\mathbf{X}, \mathbf{Y}) = (Y_1/X_1)^2$ (note the correction of the text), then

$$R(\theta, d) = 1 - \mathrm{P}_{\Delta,\Sigma}\{|\Delta - (Y_1/X_1)^2| \leq \Delta/2\} = 1 - \mathrm{P}_{1,\Sigma}\{|1 - (Y_1/X_1)^2| \leq 1/2\}.$$

This is independent of Σ, and so is a constant less than one.

4.3.5. (a) If X is $\mathcal{B}(n, \theta)$ (the binomial distribution), then $n - X$ is $\mathcal{B}(n, 1-\theta)$, so $\bar{g}\theta = 1 - \theta$. Moreover, if $\tilde{g}a = 1 - a$, then $L(\bar{\theta}, \tilde{a}) = (1-(1-\theta))(1-a) + (1-\theta)(1-(1-a)) = L(\theta, a)$. So the problem is invariant.

(b) A nonrandomized decision rule d is invariant if $d(g(x)) = \tilde{g}d(x)$, that is, if $d(n-x) = 1 - d(x)$ for all $x = 0, 1, \ldots, n$. So if $d(x)$ is specified for $x < n/2$, then it is determined for $x \geq n/2$ by $d(n-x) = 1 - d(x)$, which implies, if n is even, that $d(n/2) = 1/2$).

(c) We may compute the risk function for an invariant rule d, using $d(n-x) = d(x)$ to reduce the dependence of the risk to $d(x)$ for $x < n/2$ as follows.

$$\begin{aligned}
R(\theta, d) &= \sum_{x=0}^{n} \binom{n}{x} \theta^x (1-\theta)^{n-x} [(1-\theta)d(x) + \theta(1-d(x))] \\
&= \sum_{x<n/2} \Big\{ \binom{n}{x} \theta^x (1-\theta)^{n-x} [(1-\theta)d(x) + \theta(1-d(x))] \\
&\qquad + \binom{n}{n-x} \theta^{n-x} (1-\theta)^x [(1-\theta)(1-d(x)) + \theta d(x)] \Big\} + \text{(possible term with } x = n/2) \\
&= \sum_{x<n/2} \binom{n}{x} (1-2\theta) d(x) [\theta^x (1-\theta)^{n-x} - \theta^{n-x}(1-\theta)^x] + \text{(terms not involving d(x))}.
\end{aligned}$$

The coefficients of $d(x)$ are $(1-2\theta)[1-(\theta/(1-\theta)^{n-2x}] \geq 0$ for all $\theta \in [0,1]$ when $x < n/2$. Therefore, the risk is minimized by choosing $d(x) = 0$ for $x < n/2$, and hence $d(x) = 1$ for $x > n/2$ (and, if n is even, $d(n/2) = 1/2$. This is the best invariant rule.

By Exercise 2.11.15, the rule $d(x) \equiv 1/2$ is minimax. Since this rule is invariant, the best invariant rule is also minimax. But the best invariant rule has much smaller risk function.

4.3.6. Let τ_0 be invariant ($\bar{g}\tau_0 = \tau_0$ for all $g \in \mathcal{G}$) and let δ be Bayes with respect to τ_0 ($r(\tau_0, \delta) \leq r(\tau_0, \delta')$ for all δ'). Define

$$\delta_0 = \frac{1}{N} \sum_{g \in \mathcal{G}} \delta^g$$

where N is the number of elements in $\mathcal{G}$. Then δ_0 is invariant (in fact, it is the same as δ^I in the proof of Theorem 1). We will show that δ_0 is also Bayes with respect to τ_0.

$$\begin{aligned}
r(\tau_0, \delta_0) &= \frac{1}{N} \sum_{g \in \mathcal{G}} r(\tau_0, \delta^g) \\
&= \frac{1}{N} \sum_{g \in \mathcal{G}} r(\bar{g}\tau_0, \delta) \qquad \text{as in Theorem 3(a)} \\
&= \frac{1}{N} \sum_{g \in \mathcal{G}} r(\tau_0, \delta) \qquad \text{since } \delta_0 \text{ invariant} \\
&= r(\tau_0, \delta).
\end{aligned}$$

Thus δ_0 has the same Bayes risk as δ as so is also Bayes with respect to τ_0.

4.3.7. Let τ be least favorable ($\inf_\delta r(\tau, \delta) \geq \inf_\delta r(\tau', \delta)$ for all τ'). Define $\tau_0 = (1/N) \sum_g \bar{g}\tau$ as in the proof of Theorem 3(a). Then

$$\begin{aligned}
\inf_\delta r(\tau_0, \delta) &= \inf_\delta \frac{1}{N} \sum_{g \in \mathcal{G}} r(\bar{g}\tau, \delta) \\
&= \inf_\delta \frac{1}{N} \sum_{g \in \mathcal{G}} r(\tau, \delta^g) \\
&\geq \frac{1}{N} \sum_{g \in \mathcal{G}} \inf_\delta r(\tau, \delta^g) \\
&= \frac{1}{N} \sum_{g \in \mathcal{G}} \inf_\delta r(\tau, \delta) \\
&= \inf_\delta r(\tau, \delta).
\end{aligned}$$

Thus, τ_0 is also least favorable.