Solutions to the Exercises of Section 4.2.

4.2.1. *Reflexivity*: The identity transformation, $\bar{e}(\theta)$, is in $\bar{\mathcal{G}}$ and $\bar{e}(\theta) = \theta$.

Symmetry: If $\theta_1 \equiv \theta_2$, there exists a $\bar{g} \in \bar{\mathcal{G}}$ such that $\bar{g}(\theta_1) = \theta_2$. But then $\bar{g}^{-1} \in \bar{\mathcal{G}}$ and $\bar{g}^{-1}(\theta_2) = \theta_1$.

Transitivity: If $\theta_1 \equiv \theta_2$ and $\theta_2 \equiv \theta_3$, then there exist $\bar{g}_1 \in \bar{\mathcal{G}}$ and $\bar{g}_2 \in \bar{\mathcal{G}}$ such that $\bar{g}_1(\theta_1) = \theta_2$ and $\bar{g}_2(\theta_2) = \theta_3$. But then $\bar{g} = \bar{g}_2\bar{g}_1 \in \bar{\mathcal{G}}$ and $\bar{g}(\theta_1) = \bar{g}_2(\bar{g}_1(\theta_1)) = \bar{g}_2(\theta_2) = \theta_3$.

4.2.2. Example 4.1.1: $g_c(x) = x + c$ and $\tilde{g}_c(a) = a + c$, so d is invariant if $d(g_c(x)) = \tilde{g}_c(d(x))$ or equivalently, $d(x + c) = d(x) + c$ for all real x and all real c. Replacing x by 0, we find $d(c) = d(0) + c$ for all c. Let b denote $d(0)$ and replace c by x. Then $d(x) = x + b$.

Example 4.1.2: $g(x) = n - x$ and $\tilde{g}(a) = 1 - a$, so that d is invariant if $d(n - x) = 1 - d(x)$ for all $x = 0, 1, \ldots, n$.

Example 4.1.3: $g_{b,c}(\mathbf{x}) = b\mathbf{x} + c\mathbf{1}$ and $\tilde{g}_{b,c}(a) = ba + c$, so d is invariant if $d(b\mathbf{x} + c\mathbf{1}) = bd(\mathbf{x}) + c$ for all $\mathbf{x}$, $b \neq 0$ and c.

Example 4.1.4: $g_b(\mathbf{x}) = b\mathbf{x}$ and $\tilde{g}_b(a) = a$, so that d is invariant if $d(b\mathbf{x}) = d(\mathbf{x})$ for all $\mathbf{x}$ and $b > 0$. A behavioral rule δ is invariant if $\delta_{\mathbf{x}}(A) = \delta_{g_b(\mathbf{x})}(\tilde{g}_b(A))$ or equivalently $\delta_{\mathbf{x}}(A) = \delta_{b\mathbf{x}}(A)$ for all $\mathbf{x}$, $b > 0$ and A. Thus a behavioral rule is invariant if the probability it assigns to sets $A \in \mathcal{A}$ depends on $\mathbf{x}$ only through $\mathbf{x}/\|\mathbf{x}\|$.

4.2.3. If $Y_1, \ldots, Y_n$ are the order statistics of a sample of size n from a distribution function F, if $\phi(x)$ is a continuous increasing function of the real line onto itself and if $Z_i = \phi(Y_i)$ for $i = 1, \ldots, n$, then $Z_1, \ldots, Z_n$ are the order statistics of a sample of size n from the distribution function $G(z) = F(\phi^{-1}(z))$. Thus the distributions are invariant under the group $\mathcal{G}$ of transformations of the form $g_\phi(y_1, \ldots, y_n) = (\phi(y_1), \ldots, \phi(y_n))$, with $\bar{g}_\phi(F(x)) = F(\phi^{-1}(x))$. If $L(F, a) = W(F(a))$, then the loss is invariant with $\tilde{g}_\phi(a) = \phi(a)$, because $L(\bar{g}_\phi(F), \tilde{g}_\phi(a)) = W(\bar{g}_\phi(F(\tilde{g}_\phi(a)))) = W(F(\phi^{-1}(\phi(a)))) = W(F(a)) = L(F, a)$. A nonrandomized rule d is invariant if

$$d(g_\phi(\mathbf{y})) = \tilde{g}_\phi(d(\mathbf{y})) \quad \text{or} \quad d(\phi(y_1), \ldots, \phi(y_n)) = \phi(d(y_1, \ldots, y_n)) \quad \text{for all } \phi. \tag{$*$}$$

If ϕ leaves $y_1, \ldots, y_n$ fixed, that is for ϕ in $\Phi_{\mathbf{y}} = \{\phi : \phi(y_i) = y_i, \text{ for } i = 1, \ldots, n\}$, this gives $d(y_1, \ldots, y_n) = \phi(d(y_1, \ldots, y_n))$ which implies that $d(y_1, \ldots, y_n)$ is one of the y_j, but which j may depend on $\mathbf{y}$. Suppose $d(\mathbf{y}') = y_j'$ for some fixed $\mathbf{y}'$ and j. Then, for all ϕ that leave y_j' fixed, Equation ($*$) implies that $d(y_1, \ldots, y_{j-1}, y_j', y_{j+1}, \ldots, y_n) = y_j'$ for all $y_1, \ldots, y_{j-1}, y_{j+1}, \ldots, y_n$. Now, for those ϕ that leave $y_1, \ldots, y_{j-1}, y_{j+1}, \ldots, y_n$ fixed, ($*$) implies that $d(\mathbf{y}) = y_j$ for all $\mathbf{y}$. Thus there are only n different nonrandomized rules.

The risk function of $d_j(\mathbf{Y}) = Y_j$, is $R(F, d_j) = \mathrm{E}\,W(F(Y_j)) = \mathrm{E}\,W(U_j)$, where U_j is the jth order statistic of a sample of size n from a uniform distribution. This is independent of F.

4.2.4. We have $g_{\mathbf{O}}(\mathbf{x}) = \mathbf{O}\mathbf{x}$ for all $\mathbf{x}$ and all orthogonal matrices, $\mathbf{O}$, and $\tilde{g}_{\mathbf{O}}(a) = a$ for all a. A nonrandomized rule, d, is invariant if $d(\mathbf{O}\mathbf{x}) = d(\mathbf{x})$ for all orthogonal $\mathbf{O}$ and $\mathbf{x}$. If $|\mathbf{x}| = |\mathbf{y}|$, then there is an orthogonal $\mathbf{O}$ such that $\mathbf{O}\mathbf{x} = \mathbf{y}$, so that d depends on $\mathbf{x}$ only through its length, $|\mathbf{x}|$. Since $\bar{g}_{\mathbf{O}}(\theta) = \mathbf{O}\theta$, the orbits of Θ are the spheres. The risk function is constant on these orbits and so $R(\theta, d)$ depends on θ only through $|\theta|$.

4.2.5. We have $g_{\mathbf{O}}(\mathbf{X}) = \mathbf{O}\mathbf{X}$, and $\tilde{g}_{\mathbf{O}}(a) = a$. An invariant nonrandomized decision rule, d, satisfies $d(\mathbf{O}\mathbf{X}\mathbf{O}) = d(\mathbf{X})$ for all vectors X and all diagonal matrices $\mathbf{O}$ of determinant ± 1. If we let b_i denote the ith diagonal element of $\mathbf{O}$, then an invariant nonrandomized decision rule satisfies

$$d(x_1, \ldots, x_n) = d(b_1 x_1, \ldots, b_n x_n)$$

for all real $x_1, \ldots, x_n$ and all real $b_1, \ldots, b_n$ such that $\prod_1^n b_i = \pm 1$. If all $x_i \neq 0$, then we may take $b_i = 1/x_i$ for $i = 1, \ldots, n-1$ and $b_n = \mathrm{sgn}(x_n) \prod_1^{n-1} |x_i|$ to find

$$d(x_1, \ldots, x_{n-1}, x_n) = d(1, \ldots, 1, |\textstyle\prod_1^n x_i|)$$

for all non-zero $x_1, \ldots, x_n$, and conversely, any function of $|\prod_1^n x_i|$ is invariant. However, since in the statement of the problem, Σ is allowed to be singular, the X_i may assume the value zero with positive

probability. If some of the x_i are zero, then the same argument shows that d must be constant, but the constant may depend on which of the x_i are zero. The fact that the risk function of an invariant rule is constant on orbits in the parameter space implies that the risk function of an invariant rule depends only on $\det(\mathbf{\Sigma})$, provided $\det(\mathbf{\Sigma}) \neq 0$. When $\det(\mathbf{\Sigma}) = 0$, the risk of an invariant rule may depend on which entries along the diagonal of $\mathbf{\Sigma}$ are zero but not on the non-zero components.

Thus we see that the statement of the solution to the problem is wrong. If the parameter space were restricted to those $\mathbf{\Sigma} > 0$, then the risk function of an invariant rule would depend only on $\det(\mathbf{\Sigma})$. We may try to get this result using invariance under a larger group. In addition to invariance under the transformations $g_{\mathbf{O}}$, the problem is also invariant under the group of permutations of the subscripts of $X_1, \ldots, X_n$. However, an invariant rule d may depend on the number of $x_i = 0$, and the risk function of an invariant rule may still depend on the number of diagonal elements of $\mathbf{\Sigma}$ that are zero.

4.2.6. If $Y = g(X) = -X$, then $P_\theta(Y = 0) = P_\theta(Y = -\theta) = 1/2$, or, equivalently, $P_{-\theta}(Y = 0) = P_{-\theta}(Y = \theta) = 1/2$ for all θ. This shows that $\bar{g}(\theta) = -\theta$. Similarly, since $L(\theta, a) = L(-\theta, -a)$ for all a and θ, we have $\tilde{g}(a) = -a$. So $\bar{\mathcal{G}} = \{\bar{e}, \bar{g}\}$ and $\tilde{\mathcal{G}} = \{\tilde{e}, \tilde{g}\}$.

We may restrict attention to rules that take action $a = x$ with probability 1 when $x \neq 0$. In addition, we may restrict attention to rules that give no weight to action $a = 0$ when $x = 0$. In other words, the following class of rules, indexed by $\pi \in [0, 1]$, forms a complete class of decision rules:

$$\delta_\pi(x) = \begin{cases} 1 & \text{w.p. } 1 & \text{if } x = 1 \\ -1 & \text{w.p. } 1 & \text{if } x = -1 \\ \begin{cases} 1 & \text{w.p. } \pi \\ -1 & \text{w.p. } 1 - \pi \end{cases} & \text{if } x = 0. \end{cases}$$

The risk function of δ_π is

$$R(\theta, \delta_\pi) = \begin{cases} (1/2)(1 - \pi) & \text{if } \theta = 1 \\ (1/2)\pi & \text{if } \theta = -1. \end{cases}$$

We find $\sup_\theta R(\theta, \delta_\pi) = (1/2)\max(\pi, 1 - \pi)$. The minimax rule is δ_π for π that minimizes this supremum, namely $\pi = 1/2$. This gives the behavioral invariant rule stated in the problem. It has value $1/4$.

A nonrandomized rule $d(x)$ is invariant if $d(g(x)) = \tilde{g}(d(x))$, that is, if $d(-x) = -d(x)$ for all x. We have immediately that $d(0) = 0$ for all invariant nonrandomized rules. So we see that there are exactly three such rules depending on the three possible values of $d(1)$, namely, $d_1(x) = x$, $d_2(x) = -x$ and $d_3(x) \equiv 0$. There have constant risks $1/2$, $1/2$ and 1 respectively. The best we can do mixing these (i.e. with randomized invariant rules) is $1/2$.

4.2.7. (a) If $\mathbf{X}$ and $\mathbf{Y}$ are independent $\mathcal{N}(\mathbf{0}, \Sigma)$ and $\mathcal{N}(\mathbf{0}, \Delta\Sigma)$ respectively, then $\mathbf{BX}$ and $\mathbf{BY}$ are independent $\mathcal{N}(\mathbf{0}, \mathbf{B\Sigma B}^T)$ and $\mathcal{N}(\mathbf{0}, \Delta\mathbf{B\Sigma B}^T)$ respectively, so that $\bar{g}_B(\Delta, \Sigma) = (\Delta, \mathbf{B\Sigma B}^T)$. The loss is invariant if we take $\tilde{g}_B(a) = a$.

(b) A (behavioral) rule, δ, is invariant if for all $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{B}$ the distribution $\delta(\mathbf{X}, \mathbf{Y})$ on $\mathcal{A}$ is the same as the distribution $\delta(\mathbf{BX}, \mathbf{BY})$. But but for any two pairs, $(\mathbf{X}, \mathbf{Y})$ and $(\mathbf{X}', \mathbf{Y}')$, of linearly independent vectors, there exists a nonsingular matrix $\mathbf{B}$ such that $\mathbf{X} = \mathbf{BX}'$ and $\mathbf{Y} = \mathbf{BY}'$. Thus, for an invariant δ, the distribution $\delta(\mathbf{X}, \mathbf{Y})$ must be independent of $\mathbf{X}$ and $\mathbf{Y}$.

(c) If $\mathcal{A} = (0, \infty)$ and $L(\Delta, a) = W(a/\Delta)$, then the problem is invariant under the group of transformations $g_{B,c}(\mathbf{X}, \mathbf{Y}) = (\mathbf{BX}, c\mathbf{BY})$, with $\bar{g}_{B,c}(\Delta, \Sigma) = (c\Delta, \mathbf{B\Sigma B}^T)$ and $\tilde{g}_{B,c}(a) = ca$. As in part (b), for an invariant rule δ, the distribution $\delta(\mathbf{X}, \mathbf{Y})$ is independent of $\mathbf{X}$ and $\mathbf{Y}$. Here we have in addition that the distribution must be independent of scale change. This is true on;y of the distribution degenerate at zero. But zero is not in $\mathcal{A}$ so there are no invariant rules.

4.2.8. (a) Since the pair (X, Y) are independent $\mathcal{P}(\theta_1), \mathcal{P}(\theta_2)$, then $g(X, Y) = (Y, X)$ are independent $\mathcal{P}(\theta_2), \mathcal{P}(\theta_1)$, so $\bar{g}(\theta_1, \theta_2) = (\theta_2, \theta_1)$. With $\tilde{g}a = 1 - a$, we have $L(\bar{g}(\theta_1, \theta_2), \tilde{g}a) = L((\theta_2, \theta_1), 1 - a) = L((\theta_1, \theta_2), a)$.

(b) A nonrandomized rule satisfies $d(x, y) = 1 - d(y, x)$. But for $x = y$, which occurs with positive probability, we have $d(x, x) = 1 - d(x, x)$ which implies $d(x, x) = 1/2$ which is not in the action space, $\mathcal{A}$. Thus, no nonrandomized rules exist.

(c) Let p_{xy} denote the probability of taking action 0 if $X = x, Y = y$ is observed. This rule is invariant if and only if $p_{xy} = 1 - p_{yx}$. So all invariant rules can be specified by specifying p_{xy} for $x < y$, and then letting $p_{xy} = 1 - p_{yx}$ for $x > y$ and letting $p_{xx} = 1/2$.

(d) The risk function of an invariant rule δ satisfies $R((\theta, \theta), \delta) = 0$, and $R((\theta_1, \theta_2), \delta) = R((\theta_2, \theta_1), \delta)$. So assume $\theta_1 < \theta_2$. Then,

$$R((\theta_1, \theta_2), \delta) = \sum_{x,y} p_{xy} \frac{e^{-\theta_1} \theta_1^x}{x!} \frac{e^{-\theta_1} \theta_1^x}{x!}$$

$$= e^{-\theta_1 - \theta_2} \left[\sum_x p_{xx} \frac{\theta_1^x}{x!} \frac{\theta_2^x}{x!} + \sum_{x<y} p_{xy} \frac{\theta_1^x}{x!} \frac{\theta_2^y}{y!} + \sum_{x>y}(1 - p_{yx}) \frac{\theta_1^x}{x!} \frac{\theta_2^y}{y!} \right]$$

$$= e^{-\theta_1 - \theta_2} \left[\frac{1}{2} \sum_x \frac{\theta_1^x}{x!} \frac{\theta_2^x}{x!} + \sum_{x<y} p_{xy} \frac{\theta_1^x}{x!} \frac{\theta_2^y}{y!} + \sum_{x>y} \frac{\theta_1^x}{x!} \frac{\theta_2^y}{y!} - \sum_{y>x} p_{xy} \frac{\theta_1^y}{y!} \frac{\theta_2^x}{x!} \right]$$

$$= (\text{something independent of } p_{xy}) + e^{-\theta_1 - \theta_2} \sum_{x<y} p_{xy} \frac{1}{x!y!} (\theta_1^x \theta_2^y - \theta_1^y \theta_2^x).$$

When $\theta_1 < \theta_2$ and $x < y$, we have $\theta_1^x \theta_2^y > \theta_1^y \theta_2^x$. So $R((\theta_1, \theta_2), \delta)$ is minimized by choosing $p_{xy} = 0$ for $x < y$, and hence $p_{xy} = 1$ for $x > y$ and $p_{xx} = 1/2$.