Decision Theory

For convenience, let us rewrite the definitions of certain notations here.

\[ A = \{a \mid a \text{ is an action}\} \]
\[ D = \{\delta \mid \delta : X \to A \text{ is a non-randomized decision rule}\} \]
\[ D^\ast = \{\phi_x \mid \forall x, \phi_x \text{ is a probability measure on } A\} \]
\[ D^\ast_\ast = \{\psi \mid \psi \text{ is a probability measure on } D\} \]
\[ R(\theta, \delta) = \int_X L(\theta, \delta(x)) dP_\theta(x) \]
\[ R(\theta, \phi) = \int_X \int_A L(\theta, a) d\phi_x(a) dP_\theta(x) \]
\[ R(\theta, \psi) = \int_D R(\theta, \delta) d\psi(\delta) = \int_D \int_X L(\theta, \delta(x)) dP_\theta(x) d\psi(\delta) \]
\[ S^0 = \{y_\delta = (y_1, \ldots, y_k) \mid y_i = R(\theta, \delta), i = 1, \ldots, k\}, \Theta = \{\theta_1, \ldots, \theta_k\} \]
\[ S = \{y_\psi = (y_1, \ldots, y_k) \mid y_i = R(\theta, \psi), i = 1, \ldots, k\} \]

51. Consider \( \Theta = A = \{1, 2, \ldots, k\} \). For any given convex subset \( S \subset [0, \infty)^k \), we are going to show that there exists a collection of behavioral decision rules \( D^\ast \) with \( S \) as its risk set. Suppose \( X \) has a degenerate probability distribution such that \( P_\theta(\{1\}) = 1, \forall \theta \in \Theta \). Consider loss function \( L(\theta, a) = |\theta - a| \), then for any \( y = (y_1, \ldots, y_k) \in S \), define a behavioral decision rule \( \phi^y \) in the manner that \( R(i, \phi^y) = y_i, i = 1, \ldots, k \). This is valid, because \( \phi^y \) would be uniquely determined by the set of linear equations \( \sum_{i=1}^k |\theta - i|\phi^y_1(\{i\}) = y_\theta, \theta = 1, 2, \ldots, k \). In summary, let \( D^* = \{\phi^y \mid y \in S, \sum_{i=1}^k |\theta - i|\phi^y_1(\{i\}) = y_\theta, \theta = 1, 2, \ldots, k\} \), then \( D^* \) would have a risk set as \( S \).

52. Note that \( X = \{0, 1\} \) here.

(a) First of all, the set of non-randomized decision rules is \( D = \{\delta_{i,j} \mid \delta_{i,j}(0) = i, \delta_{i,j}(1) = j, i, j = 1, 2\} \), then \( R(\theta, \delta_{i,j}) = I(\theta \neq i)P_\theta(\{0\}) + I(\theta \neq j)P_\theta(\{1\}) \). \( S^0 = \{(0(1, \delta_{i,j}), R(2, \delta_{i,j})) \mid i, j = 1, 2\} \), which is \( \{(0, 1), (1/4, 1/2), (3/4, 1/2), (1, 0)\} \). The risk set \( S \) would be the convex hull of \( S^0 \). The gray area in the following is then \( S \).
(b) For any elemental \( \phi \in \mathcal{D}^* \), the risk function is \( R(\theta, \phi) = P_0(\{0\})(I\{\theta \neq 1\} \phi_0(\{1\}) + I\{\theta \neq 2\} \phi_0(\{2\})) + P_0(\{1\})(I\{\theta \neq 1\} \phi_1(\{1\}) + I\{\theta \neq 2\} \phi_1(\{2\})) \). And for any \( \psi \in \mathcal{D}_* \), the corresponding risk function is \( R(\theta, \psi) = \sum_{i,j \in [1,2]} \psi(\{\delta_{i,j}\}) I\{\theta \neq \delta_{i,j}(0)\} P_0(\{0\}) + I\{\theta \neq \delta_{i,j}(1)\} P_0(\{1\}) \), which is \( P_0(\{0\}) \sum_{i,j \in [1,2]} \psi(\delta_{i,j}) I\{\theta \neq i\} + P_0(\{1\}) \sum_{i,j \in [1,2]} \psi(\delta_{i,j}) I\{\theta \neq j\} \). This actually motivates the correspondence between \( \phi \) and \( \psi \), which is determined by \( \phi_0(\{i\}) = \psi(\{\delta_{i,1}\}) + \psi(\{\delta_{i,2}\}), \phi_1(\{j\}) = \psi(\{\delta_{1,j}\}) + \psi(\{\delta_{2,j}\}), i,j = 1,2 \). Then for any \( \psi \in \mathcal{D}_* \), we can define \( \phi \) using the above relationship, and they would have the same risk function. On the other hand, for any \( \phi \in \mathcal{D}^* \), we can specify at least one \( \psi \) who satisfies the above relationship and has the same risk function as \( \phi \). The easiest way of thinking about it, in this case, would be thinking about the relationship between joint distribution and marginal distribution of a \( 2 \times 2 \) probability table.

(c) All admissible risk vectors would be those points, which is the lower boundary of \( \mathcal{S} \), i.e., \( \lambda(\mathcal{S}) = \{ y \mid Q_y \cap \mathcal{S} = \{ y \} \} \), as indicated by the blue lines in the above figure. Notice further that \( A(\mathcal{D}^*) \), the admissible rules in \( \mathcal{D}^* \), is complete here by just looking at the above risk set plot. Thus it is minimum complete.

(d) Note that for each \( \phi \in \mathcal{D}^* \), \( R(\mathbf{g}, \phi) = p R(1, \phi) + (1-p) R(2, \phi) \). Now consider the line \((1-p)y + px = b\). When \( p = 1 \), the risk vector corresponding to the Bayes with respect to \( \mathbf{g} \) would be \((0,1)\). For \( 0 \leq p \leq 1 \), the line becomes \( y = -p/(1-p) \cdot x + b/(1-p) \).

For simplicity, define \( \mathcal{L} = \{(x,y) \in \mathbb{R}^2 \mid (1-p)y + px = b, b \in \mathbb{R} \} \), then the set of risk vectors that are Bayes versus \( \mathbf{g} \) is \( \{(x,y) \in \mathcal{S} \cap \mathcal{L} \mid Q_{(x,y)} \cap (\mathcal{S} \cap \mathcal{L}) = \{(x,y)\} \} \), where \( Q_{(x,y)} \) is the lower quadrant of \((x,y)\). Then we see that two particular priors, corresponding to \(-p/(1-p) = -2\) and \(-p/(1-p) = -2/3\) have more than one decision rules. In all, when \( 0 \leq p < 2/5 \), then answer is \((1,0)\); when \( p = 2/5 \), the answer is \((\lambda/4 + (1-\lambda), \lambda/2) \mid 0 \leq \lambda \leq 1\); when \( 2/5 < p < 2/3 \), then answer is \((1/4, 1/2)\); when \( p = 2/3 \), the answer is \((\lambda/4, \lambda/2 + (1-\lambda)) \mid 0 \leq \lambda \leq 1\); when \( 2/3 < p \leq 1 \), the answer is \((0, 1)\).

(e) The posterior distribution with respect to \( \mathbf{g} = (1/2, 1/2) \) is \( \pi_{\theta \mid X=x}(\{1\}) = P_1(\{x\})/P_1(\{x\}) + P_2(\{x\}) \), and the posterior expected loss is \( E\{L(\theta, \phi_x) | X = x\} = E\{\phi_x(\{1\}) I\{\theta = 1\} + \phi_x(\{2\}) I\{\theta = 1\} | X = x\} = \{\phi_x(\{2\}) P_1(\{x\}) + \phi_x(\{1\}) P_2(\{x\})\}/P_1(\{x\}) + P_2(\{x\}) \}. Thus the Bayes with respect to \( \mathbf{g} = (1/2, 1/2) \) is \( \phi^*_x(\{1\}) = I\{x \neq 1\} \). Based on the formula de-
rived in part (b), we know that \((R(1, \phi^*), R(2, \phi^*)) = (1/4, 1/2)\), which is actually the correct answer due to our analysis in part (d).

53. The joint probability measure of \(X = (X_1, X_2)\) is \(P_\theta \times P_\theta = P_\theta^2\). Then the risk function corresponding to \(\phi\) defined in the context of this problem is \(R(\theta, \phi) = L(\theta, 1)P_\theta(\{0\}) + P_\theta(\{1\})/2\). Therefore, its risk vector is \((1/8, 3/4)\). Now let's consider \(\phi'_X(\{1\}) = I\{x_1 + x_2 \leq 1\}\), then \(R(1, \phi') = \int \phi'_x(\{2\})dP_\theta(x) = P_1(\{1\})^2 = 1/16\). \(R(2, \phi') = \int \phi'_x(\{1\})dP_\theta(x) = 1 - P_2(\{1\})^2 = 3/4\), which means \(\phi'\) is better than \(\phi\) in terms of risk functions. Therefore, \(\phi\) is inadmissible.

For a general behavioral decision rule \(\tilde{\phi}\), note that \(R(1, \tilde{\phi}) = \frac{9}{16} \tilde{\phi}_{(0,0)}(\{2\}) + \frac{3}{12} \tilde{\phi}_{(0,1)}(\{2\}) + \frac{3}{10} \tilde{\phi}_{(1,0)}(\{2\}) + \frac{1}{10} \tilde{\phi}_{(1,1)}(\{2\}); R(2, \tilde{\phi}) = \{\tilde{\phi}_{(0,0)}(\{1\}) + \tilde{\phi}_{(0,1)}(\{1\}) + \tilde{\phi}_{(1,0)}(\{1\}) + \tilde{\phi}_{(1,1)}(\{1\})\}/4.\) One can actually solve an equation using \(\tilde{\phi}'_x(\{2\}) = aI\{x_1 + x_2 = 0\} + bI\{x_1 + x_2 = 1\} + cI\{x_1 + x_2 = 2\}.\) This suggests \(\tilde{\phi}'_x(\{2\}) = \frac{1}{8}I\{x_1 + x_2 = 0\} + \frac{7}{8}I\{x_1 + x_2 = 2\}.\) One can verify that \(\tilde{\phi}\) here is risk equivalent to \(\phi\).

54. \(R(p, \delta_1) = E(p - X/n)^2 = Var(X/n) = p(1 - p)/n, R(p, \delta_2) = E\left\{ p - \frac{X/n + 1/2}{2} \right\}^2 = E(p - X/n)^2/4 + (p - 1/2)^2/4 = p(1 - p)/(4n) + (p - 1/2)^2/4, \) and \(R(p, \psi) = R(p, \delta_1)\psi(\{\delta_2\}) + R(p, \delta_2)\psi(\{\delta_2\}) = 5p(1 - p)/(8n) + (p - 1/2)^2/8.\) To make a behavioral rule equivalent to psi, notice that \(R(p, \phi) = \int_X \int_{\{a\}} (p - a)^2 d\phi_x(a)dP_p(x),\) the most natural choice would be \(\phi_x(\{x/n\}) = \phi_x(\{(x/n + 1/2)/2\}) = 1/2.\) In this way, \(R(p, \phi) = \int_X (p - x/n)dP_p(x)/2 + \int_X (p - x/n + 1/2)/2)dP_p(x)/2 = R(p, \psi).\)

55. Suppose that \(\phi\) is inadmissible when the parameter space is \(\Theta,\) that is, there exists \(\phi'\) such that \(R(\theta, \phi') \leq R(\theta, \phi), \forall \theta = (\theta_1, \theta_2) \in \Theta = \Theta_1 \times \Theta_2\) and the inequality holds strictly for at least one \(\theta^* = (\theta_1^*, \theta_2^*) \in \Theta.\) Then clearly \(\phi'\) is better than \(\phi\) when the parameter space is \(\Theta_1 \times \{\theta_2^*\},\) which is a contradiction. Therefore, \(\phi\) is admissible when the parameter space is \(\Theta.\)

56. For any two behavioral decision rules \(\phi, \phi', \int_X \int_A L(\theta, a)d\phi_x(a)dP_\theta(x) \leq \int_X \int_A L(\theta, a)d\phi'_x(a)dP_\theta(x)\) is equivalent to \(\int_X \int_A w(\theta)L(\theta, a)d\phi_x(a)dP_\theta(x) \leq \int_X \int_A w(\theta)L(\theta, a)d\phi'_x(a)dP_\theta(x),\) for any \(\theta.\) This means all inequalities would remain the same when multiplying the loss function by \(w(\theta) > 0.\) And this suffices to support the fact that admissibility would not change when changing loss function in this way.

57. \(T_0(X) = I\{\bar{X} > 1/2\} + I\{\bar{X} = 1/2\}/2,\) then \(R(\theta, T_0) = E\{T_0(X) - \theta\}^2 = \theta^2 + (1 - 2\theta)P_\theta(\bar{X} > 1/2) + (1/4 - \theta)P_\theta(\bar{X} = 1/2),\) where \(\sum X_i \sim Binomial(n, \theta).\) Now \(R(\theta, T_1) = R(\theta, \bar{X})/2 + R(\theta, T_0)/2 = Var(\bar{X})/2 + R(\theta, T_0)/2 = \theta(1 - \theta)/(2n) + R(\theta, T_0).\) And \(R(\theta, T_2) = E\theta(\theta - \bar{X})^2\bar{X} + (\theta - 1/2)^2(1 - \bar{X}) = (\theta - 1/2)^2(1 - \theta) + (1 - \theta)\theta^2/n + E\theta(\bar{X} - \theta)^3}.\)

58. \(T_0\) is a non-randomized decision rule. By Theorem 2.5 from Shao, we can consider \(\delta_1(X) = (\bar{X} + T_0)/2\) for \(T_1\) and \(\delta_2(X) = \bar{X}^2 + (1 - \bar{X})/2\) for \(T_2.\)