Decision Theory (Continued)

59. Consider the following decision problem. Let $\mathcal{X} = \Theta = \mathcal{A} = \{0, 1\}$, $L(\theta, a) = |\theta - a|$, $P_1(\{1\}) = P_0(\{0\}) = 1$. Further, there are two class $C_1 = \{\delta_i \mid i \in \mathbb{N}\}$, $C_2 = \{\tilde{\delta}_i \mid i \in \mathbb{N}\}$, where $\delta_i(0) = 1/2^i = 1 - \delta_i(1)$, $\tilde{\delta}_i(0) = 1/3^i = 1 - \tilde{\delta}_i(1)$. Then $R(0, \delta_i) = 1/2^i, R(1, \delta_i) = |1 - (1 - 1/2^i)| = 1/2^i$. Similarly, $R(0, \tilde{\delta}_i) = 1/3^i, R(1, \tilde{\delta}_i) = 1/3^i$. Now let $D = C_1 \cup C_2$. Then $C_1$ and $C_2$ are complete. Because for any behavioral rule $\phi \in D$, its corresponding risk point falls at some point $(\lambda, \lambda)$, where $\lambda > 0$. Which means for any $\phi \in C_1$, there exists large enough $i$ such that $\delta_i$ dominates $\phi$. Same argument holds for $C_2$. However, as we can see, $C_1 \cap C_2 = \emptyset$. Thus intersection of two complete class might even be empty, therefore may not be essentially complete.

60. Recall the definition that the Bayes risk of $\phi$ with respect prior $G$ is $R(G, \phi) = \int_{\Theta} R(\theta, \phi) dG(\theta)$.

(a) Because $L(p, a)$ is convex in $a$, $\forall p \in (0, 1)$ and $(0, 1)$ is a convex set, Lemma 101 tells us that we only need to focus on the non-randomized decision rules when considering Bayes versus any prior in this case. In other words, we only need to look at the optimization problem of

$$\delta^* = \arg \min_{\delta \in D} \int_{\Theta} \int_{\mathcal{X}} L(p, \delta(x)) dP_p(x) dG(p) = \arg \min_{\delta} E\{(p - \delta(X))^2 / \{p(1 - p)\}\},$$

where the above expectation is with respect to the joint distribution of $p$ and $X$. We know that, the solution would be, for each $x$,

$$\delta^*(x) = \arg \min_{\delta} E\{(p - \delta(x))^2 / \{p(1 - p)\}\} | X = x,$$

where the above expectation is with respect to the posterior distribution. We know that the posterior distribution is $Beta(x + 1, n - x + 1)$ in this case. Write out the posterior expectation explicitly, it gives us

$$C(x) \int_{(0,1)} (p - \delta(x))^2 p^{x-1}(1-p)^{n-x} dp,$$

where $C(x)$ is some normalizing constant only depends on $x$. This suggests the minimizer is actually the mean of $Beta(x, n - x)$, which is $x/n$. Thus $X/n$ is Bayes versus uniform prior on $(0, 1)$.

Actually, for any weighted squared error loss function, say $L(\theta, a) = \omega(\theta)(\theta - a)^2$, minimizing $E\{L(\theta, \delta(X)) | X = x\} = \delta(x)^2 E\{\omega(\theta) | X = x\} - 2E\{\omega(\theta) \theta | X = x\} \delta(x) + E\{\omega(\theta) \theta^2 | X = x\}$ gives us the Bayes rule as $\delta^*(X) = E\{\omega(\theta) \theta | X\} / E\{\omega(\theta) | X\}$. 

1
(b) We are going to verify the conditions of Theorem 123 one by one. First, \( \Theta = (0, 1) \) is an open set. Second, \( R(p, \phi) \) is continuous in \( p \) for any \( \phi \in \mathcal{D}^* \). Moreover, the uniform prior has positive mass on any open subset of \((0, 1)\). As \( X/n \) is Bayes versus the uniform prior \( G \) and \( R(G) = R(G, X/n) = \int_{(0,1)} \int_X L(p,X/n) dF_p(x) dp = 1/n < \infty \). By Theorem 123, \( X/n \) is admissible in this decision problem.

(c) Note that \( X/n \) has constant risk function \( R(p,X/n) = 1/n \) and is admissible, thus it is minimax. By theorem 139, the uniform prior is a least favorable prior because \( \Theta = (0,1) \).

(d) Let \( \omega(p) = 1/(p(1-p)) \), \( p \in (0,1) \), then \( \omega(p) > 0 \) and \( \omega(p)(p - a)^2 = L(p,a) \). By problem 56, because \( X/n \) is admissible under loss function \( L(p,a) \), it is also admissible under squared loss function.

61. (a) A formal non-randomized Bayes rule versus \( G \) is a \( \delta(x) \) which minimizes the posterior expectation of the loss function. In this case, the posterior distribution is \( \Pi_{\theta \mid X=x} \{\{0\}\} = G(\{0\})f_0(x)/\{G(\{0\})f_0(x) + G(\{1\})f_1(x)\} \). Therefore the posterior expectation of the loss function is

\[
\frac{G(\{0\})f_0(x)L(0, \delta(x))}{G(\{0\})f_0(x) + G(\{1\})f_1(x)} + \frac{G(\{1\})f_1(x)L(1, \delta(x))}{G(\{0\})f_0(x) + G(\{1\})f_1(x)}
\]

Then the formal Bayes rule versus prior \( G \) would be

\[
I\{G(\{0\})f_0(x)L(0,1) + G(\{1\})f_1(x)L(1,1) \leq G(\{0\})f_0(x)L(0,0) + G(\{1\})f_1(x)L(1,0)\}.
\]

(b) When \( L(\theta, a) = I\{\theta \neq a\} \), we have \( \delta(x) = I\{G(\{0\})f_0(x) \leq G(\{1\})f_1(x)\} \). In the simple versus simple hypothesis testing context, say \( H_0 : \theta = 0 vs H_1 : \theta = 1 \), we would tend to reject the null hypothesis when \( f_1/f_0 \) is large.

62. (a) \( R(p, \delta) = |p - 1/4|(1 - p)p - 1/4| = (-p/2 + 1/4)I\{0 \leq p \leq 1/4\} + \{-2(p - 1/2)^2 + 1/4\}I\{1/4 < p \leq 3/4\} + \{-1 - p)/2 + 1/4\}I\{3/4 < p \leq 1\} \leq 1/4\).

(b) The logic here is like this. First, we assume that there exists some prior distribution \( G \) placing all its mass on \( \{0,1/2,1\} \), say \( G(\{0\}) = p_1, G(\{1/2\}) = p_2, G(\{1\}) = 1 - p_1 - p_2 \) such that \( \delta \) defined in this problem is Bayes versus \( G \). We know that a non-randomized rule which minimizes the posterior mean of the loss function would serve as a candidate. In our case here, under the \( L_1 \) loss function, the solution would be the posterior median. Now we try to get the posterior median under prior \( G \), and they should be functions of \( p_1, p_2 \), then consider the correspondence with \( \delta(x) \), we are left to solve this equation and if we are lucky to get solutions, say \( p_1^*, p_2^* \), then the problem is solved. First of all, using our set-up here, the posterior distribution is

\[
(P_{\theta \mid X=0}(\{0\}), P_{\theta \mid X=0}(\{1/2\}), P_{\theta \mid X=0}(\{1\})) = \left(\frac{p_1}{p_1 + p_2/2}, \frac{p_2/2}{p_1 + p_2/2}, 0\right),
\]

\[
(P_{\theta \mid X=1}(\{0\}), P_{\theta \mid X=1}(\{1/2\}), P_{\theta \mid X=1}(\{1\})) = \left(0, \frac{p_2/2}{1 - p_1 - p_2/2}, \frac{1 - p_1 - p_2}{1 - p_1 - p_2/2}\right).
\]

Note that for any random variable \( Z \), whose probability distribution is \( P_Z \), the median of \( Z \) is defined as \( z_{median} \) such that \( P_Z(\{z < z_{median}\}) \geq 1/2 \) and \( P_Z(\{z \geq z_{median}\}) \leq 1/2 \). When \( x = 0 \), we want the median to be 1/4, then \( p_1/(p_1 + p_2/2) = 1/2 \); when \( x = 1 \), we want the median to be 3/4, which means \( p_2/2/(1 - p_1 - p_2/2) = 1/2 \). Solving those two equations, we get \( p_1^* = 1/4, p_2^* = 1/2 \). Thus \( \delta \) is Bayes versus prior \( G \) where \( (G(\{0\}), G(\{1/2\}), G(\{1\})) = (1/4, 1/2, 1/4) \).
(c) Using the prior we find in $G$, as $\delta$ is Bayes versus $G$, we know that $R(G) = R(G, \delta) = 1/4$. In part (a), we proved that $R(\theta, \delta) \leq 1/4 = R(G), \forall \theta$. Therefore, by Corollary 137, $\delta$ is minimax. By Theorem 139, $G$ is least favorable.

63. (a) Consider $G$ such that $G(\{5\}) = G(\{6\}) = 1/2$, then the Bayes versus $G$ is $\delta(x) = I\{f(5 - x) \leq f(6 - x)\} = I\{x \geq 5.5\}$, where $f$ is the density of standard normal distribution. $R(\theta, \delta) = P_\theta(X < 5.5)I\{\theta < 5\} + P_\theta(X \geq 5.5)I\{\theta \leq 5\} = F(5.5 - \theta)I\{\theta < 5\} + (1 - F(5.5 - \theta))I\{\theta \leq 5\}$, where $F$ is the CDF of standard normal distribution. $R(G, \delta) = (1 - F(0.5))/2 + F(-0.5)/2 = F(-0.5)$. For any $\theta \in \Theta$, as $F(5.5 - \theta)$ is decreasing in $\theta$, $1 - F(5.5 - \theta) = F(\theta - 5.5)$ is increasing in $\theta$, $\sup_{\theta \in \Theta} R(\theta, \delta) \leq R(G, \delta)$. By Corollary 137 and Theorem 139, $G$ is a least favorable prior and its corresponding Bayes rule is $\delta(x) = I\{x \geq 5.5\}$, which is also minimax.

(b) Motivated by part (a), consider a sequence of priors $\{G_i\}_{i \geq 1}$ such that $G_i(\{5\}) = G_i(\{5+2/i\}) = 1/2$, then the Bayes rule for each $G_i$ is $\delta_i = I\{x \geq 5+1/i\}$ and its corresponding Bayes risk is $R(G_i, \delta_i) = (F(-1/i) + F(0))/2$, where $F$ is the CDF of standard normal distribution. Then $\lim_{i \to \infty} R(G_i, \delta_i) = F(0) = 1/2 < \infty$. Now consider $\delta^i(x) = I\{x \geq 5\}$, then $R(\theta, \delta^i) = I\{\theta < 5\}F(\theta - 5) + I\{\theta \leq 5\}F(\theta - 5) \leq F(0) = 1/2$. By theorem 135, $\delta^i(x) = I\{x \geq 5\}$ is minimax.

64. (a) $R(\lambda, \delta) = E\{(\lambda - X)^2\}/\lambda = 1$, thus $\delta$ is an equalizer rule.

(b) Consider the optimization problem

$$\arg \min_a \int_{\Lambda} L(\lambda, a)f_\lambda(x)dm(\lambda) = \arg \min_a \{a^2\Gamma(x) - 2a\Gamma(x + 1) + \Gamma(x + 2)\} = x,$$

thus $\delta(X) = X$ is generalized Bayes versus Lebesgue measure on $\Lambda$.

(c) Let $\omega(\lambda) = \lambda^{-1}$, by our analysis in 60 (a), we know that the Bayes estimator with respect to prior $\Gamma(\alpha, \beta)$ is

$$E\{\omega(\lambda)\lambda|X\} / E\{\omega(\lambda)|X\} = \left\{\int_{\Lambda} e^{-\lambda X}e^{-\lambda/\beta}\lambda^{\alpha-1}d\lambda\right\}^{-1} \cdot \left\{\int_{\Lambda} e^{-\lambda X}e^{-\lambda/\beta}\lambda^{\alpha-1}d\lambda\right\}.$$

And the result is $\beta(X + \alpha - 1)/(1 + \beta)$.

(d) Consider a sequence of priors $G_i = \Gamma(1, i)$, then $\delta_i(X) = iX/(i + 1)$ is Bayes versus $G_i$ and $R(G_i, \delta_i) = E(X, \lambda)\{\lambda^{-1}(\lambda - iX/(i + 1))^2\} = E_i\{\lambda/(i + 1)^2 + i^2/(i + 1)^2\} = (i^2 + i)/(i + 1)^2 \to 1, i \to \infty$. Note that $\delta(X) = X$ has $R(\theta, X) = 1, \forall \theta$. By theorem 135, $\delta(X) = X$ is minimax.