Statistics 582, Problem Set 4 Solutions, Revised

Wellner; 2/7/2018

1. Let $\Theta = \{0, 1\} = \mathbf{A}$ where 0 = a patient has tuberculosis, 1 = a patient does not have tuberculosis. Let X be the number of positive reactions to two different tuberculosis tests, so that $\mathbf{X} = \{0, 1, 2\}$, and suppose that X has the following distributions

x	0	1	2
$p_0(x)$	.02	.13	.85
$p_1(x)$	.70	.27	.03

If the losses are given by $L(1,1) = L(0,0) = 0$, $L(0,1) = 100$, $L(1,0) = 10$, and the prior $\lambda = (\lambda_0, \lambda_1) = (.2, .8)$, find the Bayes rule d_B and the minimax rule d_M. Plot the risk set and label the non-randomized decision rules.

Solution: Let $d = (d_0, d_1, d_2)$ with d_i =prob of action 1 when $x = i$ is observed, $i = 0, 1, 2$. Then the risks are

$$R(0, d) = 100\{d_0(.02) + d_1(.13) + d_2(.85)\}$$
$$R(1, d) = 10\{(1 - d_0)(.70) + (1 - d_1)(.27) + (1 - d_2)(.03)\},$$

and, for $\underset{\sim}{\lambda} = (.2, .8)$, the Bayes risk of d is

$$\begin{aligned} \mathcal{R}(\lambda, d) &= (.2)R(0, d) + (.8)R(1, d) \\ &= 8 + (.01)\{-520d_0 + 44d_1 + 1676d_2\} \end{aligned}$$

which is minimized by $d = (1, 0, 0) \equiv d_B = d_4$ (in the list of nonrandomized rules below).

To find a minimax rule, equate $R(0, d) = R(1, d)$: this yields

$$\{2d_0 + 13d_1 + 85d_2\} = 10 - 7d_0 - 2.7d_1 - .3d_2 .$$

Solving for d_1 yields

$$d_1 = (100 - 90d_0 - 853d_2)/157 ,$$

and plugging this back into $R(0, d)$ yields

$$\begin{aligned} R(0, d) = R(1, d) &= 2d_0 + \frac{130}{157}(10 - 9d_0 - 85.3d_2) + 85d_2 \\ &= \frac{1300}{157} + (2 - \frac{13 \cdot 90}{157})d_0 + (85 - \frac{130 \cdot 85.3}{90})d_2 \end{aligned}$$

which is minimized by $d_0 = 1$, $d_2 = 0$; then $d_1 = 10/157 \doteq .0637\ldots$. Hence the minimax rule is $d_M = (1, 10/157, 0)$, and the corresponding common risk is $R(0, d_M) = R(1, d_M) = 444/157 \doteq 2.8280\ldots$. Note that for the Bayes rule we have $R(0, d_B) = 2$, $R(1, d_B) = 3$.

The nonrandomized rules and their risks are:

x	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8
0	0	0	0	1	1	1	0	1
1	0	0	1	0	1	0	1	1
2	0	1	0	0	0	1	1	1
$R(0,d)$	0	85	13	2	15	87	98	100
$R(1,d)$	10	9.7	7.3	3	0.3	2.7	7	0

Here is a plot of the risk body:

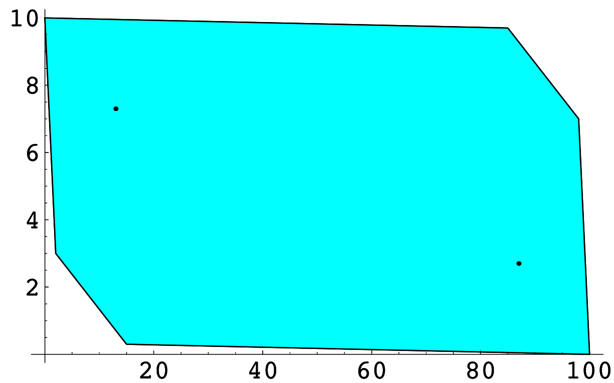

Figure 1: Risk Body.

2. Consider testing the simple hypothesis $H_0 : X \sim P_0$ versus the simple alternative $H_1 : X \sim P_1$. Let ϕ be a test of H_0 versus H_1, and let $a \equiv E_1(1-\phi)$, $b \equiv E_0\phi$.
(a) Find a test ϕ which minimizes $a + Db$ where D is a fixed number.
(b) When $D = 1$, relate the minimized total $a + b$ to the risk and to the total variation distance $d_{TV}(P_0, P_1)$ between P_0 and P_1 (or $\int p_0 \wedge p_1 d\mu$ for a dominating measure μ, e.g. $P_0 + P_1$).
(c) Carry the computations of (b) through in the context of problem 1 when the losses are $L(0,0) = L(1,1) = 0$, $L(0,1) = 10 = L(1,0)$, and the prior distribution is $\lambda = (\lambda_0, \lambda_1) = (.3, .7)$.

Solution: (a) Let $p_i \equiv dP_i/d\mu$ where $\mu \equiv P_0 + P_1$, $i = 0, 1$. Now

$$a + Db = E_1(1-\phi) + DE_0\phi = 1 + \int \phi(Dp_0 - p_1)d\mu = 1 - \int \phi(p_1 - Dp_0)d\mu, ,$$

so $a + Db$ is minimized by

$$\phi(x) = \begin{cases} 1 & \text{if } p_1(x) > Dp_0(x) \\ \gamma(x) & \text{if } p_1(x) = Dp_0(x) \\ 0 & \text{if } p_1(x) < Dp_0(x). \end{cases}$$

For any other test ϕ^*,

$$
\begin{aligned}
&\int (\phi - \phi^*)(p_1 - Dp_0) d\mu \\
&= \int_{[p_1 > Dp_0]} (\phi - \phi^*)(p_1 - Dp_0) d\mu + \int_{[p_1 < Dp_0]} (\phi - \phi^*)(p_1 - Dp_0) d\mu \\
&= \int_{[p_1 > Dp_0]} (1 - \phi^*)(p_1 - Dp_0) d\mu + \int_{[p_1 < Dp_0]} (0 - \phi^*)(p_1 - Dp_0) d\mu \\
&\geq 0
\end{aligned}
$$

so that

$$
\int \phi (p_1 - Dp_0) d\mu \geq \int \phi^* (p_1 - Dp_0) d\mu.
$$

This can be reformulated in a Bayesian context by writing

$$
\begin{aligned}
a + Db &= (1+D) \left\{ \frac{1}{1+D} a + \frac{D}{1+D} b \right\} \\
&= (1+D) \{ (1-\lambda) E_1 (1-\phi) + \lambda E_0 \phi \} \\
&= (1+D) \mathcal{R}(\Lambda, \phi),
\end{aligned}
$$

the Bayes risk with respect to the prior distribution Λ given by $\lambda = (\lambda, 1-\lambda)$ with $\lambda \equiv D/(1+D)$. Then minimizing $a + Db$ is equivalent to minimizing the Bayes risk with the prior $1 - \lambda = 1/(1+D)$ on P_1 and $\lambda = D/(1+D)$ on P_0. As we saw in class on 1/30, any rule of the form

$$
\phi(X) = \begin{cases} 1 & \text{if } p_1(X) > \frac{\lambda}{1-\lambda} p_0(X) \\ \gamma(X) & \text{if } p_1(X) = \frac{\lambda}{1-\lambda} p_0(X) \\ 0 & \text{if } p_1(X) < \frac{\lambda}{1-\lambda} p_0(X) \end{cases}
$$

is Bayes wrt λ.

(b) When $D = 1$, the minimized total $a + b$ equals, by using by using our earlier results for total variation distance,

$$
\begin{aligned}
1 + \int_{[p_0 < p_1]} (p_0 - p_1) d\mu &= 1 - d_{TV}(P_0, P_1) \\
&= 1 - \left\{ 1 - \int p_0 \wedge p_1 d\mu \right\} \\
&= \int p_0 \wedge p_1 d\mu;
\end{aligned}
$$

i.e. the test which minimizes the sum of the error probabilities has total error probability equal to $\int p_0 \wedge p_1 d\mu = 1 - d_{TV}(P_0, P_1)$.

(c) For the two distributions given in problem 2, the given prior $\lambda = (\lambda_0, \lambda_1) = (.3, .7)$, and the given losses, the Bayes risk for an arbitrary rule $d = \phi$ is given by

$$\begin{aligned}\mathcal{R}(\lambda, \phi) &= 10\left\{\lambda_1 E_1(1-\phi) + \lambda_0 E_0 \phi\right\} \\ &= 10\left\{\lambda_1 + \int (\lambda_0 p_0 - \lambda_1 p_1)\phi d\mu\right\}.\end{aligned}$$

This is minimized by any rule ϕ_λ of the form

$$\phi_\lambda(x) = \begin{cases} 1, & \text{if } \lambda_1 p_1(x) > \lambda_0 p_0(x), \\ \gamma(x), & \text{if } \lambda_1 p_1(x) = \lambda_0 p_0(x), \\ 0, & \text{if } \lambda_1 p_1(x) < \lambda_0 p_0(x). \end{cases}$$

Thus we compute

x	0	1	2
$.3p_0(x)$	.006	.039	.255
$.7p_1(x)$	.490	.189	.021
$.3p_0(x) - .7p_1(x)$	-.484	-.15	+ .234
$\phi_\lambda(x)$	1	1	0

This yields the Bayes Risk

$$\mathcal{R}(\lambda, \phi_\lambda) = 10\{.7 - .484 - .150\} = .66.$$

For the two distributions in problem 2,

$$\eta(P_0, P_1) = \int p_0 \wedge p_1 d\mu = .02 + .13 + .03 = .18,$$

$$d_{TV}(P_0, P_1) = 1 - .18 = .82 = 2^{-1} \int |p_0 - p_d| d\mu.$$

The rule which minimizes $a + b = 2((1/2)a + (1/2)b)$ is the Bayes rule with respect to the prior $\lambda = (1/2, 1/2)$, and it is given by $\phi(X) = 1\{X \in \{0, 1\}\}$. The total risk is twice the Bayes risk for the prior $(.5, .5)$ and it equals $\rho(P_0, P_1) = .18$. Thus the Bayes risk equals $10 \times (1/2)(.18) = .9$ for the prior $\lambda = (1/2, 1/2)$ with $D = 1$.

3. Let $\mathcal{X} = \{0, 1\}$, $\mathcal{A} = \Theta = \{1, 2\}$, and assume that the losses are given by $L(1, 1) = L(2, 2) = 0$, $L(1, 2) = a$, $L(2, 1) = b$. Suppose that the statistician can observe either X or Y where

$$\begin{aligned} p_1(1) &= P_1(X = 1) = 2/3, \qquad & p_2(1) &= P_2(X = 1) = 1/2, \\ p_1^*(1) &= P_1(Y = 1) = 3/4, \qquad & p_2^*(1) &= P_2(Y = 1) = 1/2. \end{aligned}$$

Let $\underline{\lambda} = (\lambda, 1 - \lambda)$, $\lambda \in [0, 1]$ be the prior distribution over Θ.
(a) Find the Bayes risk when X is observed, and similarly for Y.
(b) In the case $a = b$, $\lambda = 1/2$, would the statistician prefer to observe X or Y?
(c) For general $a \neq b$, $\lambda \in (0, 1)$ would the statistician prefer to observe X or Y?

Solution: Let d_i =probability of action 1 given that i is observed.
(i) The Bayes risks for observing X or Y are:

$$\mathcal{R}_X(\lambda, d) = \lambda a\{(1-d_0)\frac{1}{3} + (1-d_1)\frac{2}{3}\} + (1-\lambda)b\{d_0\frac{1}{2} + d_1\frac{1}{2}\},$$

and

$$\mathcal{R}_Y(\lambda, d) = \lambda a\{(1-d_0)\frac{1}{4} + (1-d_1)\frac{3}{4}\} + (1-\lambda)b\{d_0\frac{1}{2} + d_1\frac{1}{2}\}.$$

(ii) When $a = b$ and $\lambda = 1/2$,

$$\mathcal{R}_X(\lambda, d) = a\frac{1}{2}\left\{1 + \frac{1}{6}d_0 - \frac{1}{6}d_1\right\},$$
$$\mathcal{R}_Y(\lambda, d) = a\frac{1}{2}\left\{1 + \frac{1}{4}d_0 - \frac{1}{4}d_1\right\},$$

and these are both minimized by choosing $d = (0,1) \equiv d_\lambda$. Then $\mathcal{R}_X(\lambda, d_\lambda) = (5/12)a > (3/8)a = \mathcal{R}_Y(\lambda, d_\lambda)$, so we would prefer to observe Y.
(iii) The risks are

$$R_X(1,d) = a\left\{(1-d_0)\tfrac{1}{3} + (1-d_1)\tfrac{2}{3}\right\}, \qquad R_X(2,d) = b\left\{d_0\tfrac{1}{2} + d_1\tfrac{1}{2}\right\}$$
$$R_Y(1,d) = a\left\{(1-d_0)\tfrac{1}{4} + (1-d_1)\tfrac{3}{4}\right\}, \qquad R_Y(2,d) = b\left\{d_0\tfrac{1}{2} + d_1\tfrac{1}{2}\right\}.$$

Plotting these for $d_1 = (1,1)$, $d_2 = (1,0)$, $d_3 = (0,1)$, and $d_4 = (0,0)$ yields the following plot of the risk bodies:

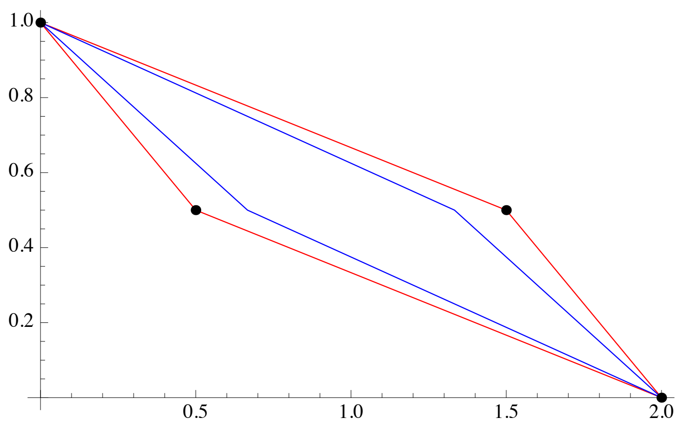

Figure 2: Comparison of the risk bodies $\mathcal{R}_Y$ (red), $\mathcal{R}_X$ (blue), for $a = 2$, $b = 1$

This plot makes it clear that we will always prefer to observe Y. To confirm this, let $r = a/b$, and write

$$\mathcal{R}_X = a\lambda + b\left\{(1-\lambda)\frac{1}{2} - r\lambda\frac{1}{3}\right\}d_0 + b\left\{(1-\lambda)\frac{1}{2} - r\lambda\frac{2}{3}\right\}d_1,$$
$$\mathcal{R}_Y = a\lambda + b\left\{(1-\lambda)\frac{1}{2} - r\lambda\frac{1}{4}\right\}d_0 + b\left\{(1-\lambda)\frac{1}{2} - r\lambda\frac{3}{4}\right\}d_1.$$

First consider $\mathcal{R}_X(\lambda, d)$:
For $0 \le \lambda \le (1 + 4r/3)^{-1}$, both coefficients are > 0, so $d_\lambda = (0, 0)$ and $\mathcal{R}_X(\lambda, d_\lambda) = a\lambda$.
For $(1 + 4r/3)^{-1} \le \lambda(1 + 2r/3)^{-1}$, $d_\lambda = (0, 1)$ and $\mathcal{R}_X(\lambda, d_\lambda) = a\lambda/3 + b(1 - \lambda)/2$.
For $(1 + 2r/3)^{-1} \le \lambda \le 1$, $d_\lambda = (1, 1)$ and $\mathcal{R}_X(\lambda, d_\lambda) = b(1 - \lambda)$.

Now consider $\mathcal{R}_Y(\lambda, d)$:
For $0 \le \lambda \le (1 + 2r/3)^{-1}$, both coefficients are > 0, so $d_\lambda = (0, 0)$ and $\mathcal{R}_Y(\lambda, d_\lambda) = a\lambda$.
For $(1 + 3r/2)^{-1} \le \lambda(1 + r/2)^{-1}$, $d_\lambda = (0, 1)$ and $\mathcal{R}_Y(\lambda, d_\lambda) = a\lambda/4 + b(1 - \lambda)/2$.
For $(1 + r/2)^{-1} \le \lambda \le 1$, $d_\lambda = (1, 1)$ and $\mathcal{R}_Y(\lambda, d_\lambda) = b(1 - \lambda)$.

These Bayes risks are plotted in Figure **??**.Thus the Bayes risk is *always* small for Y.

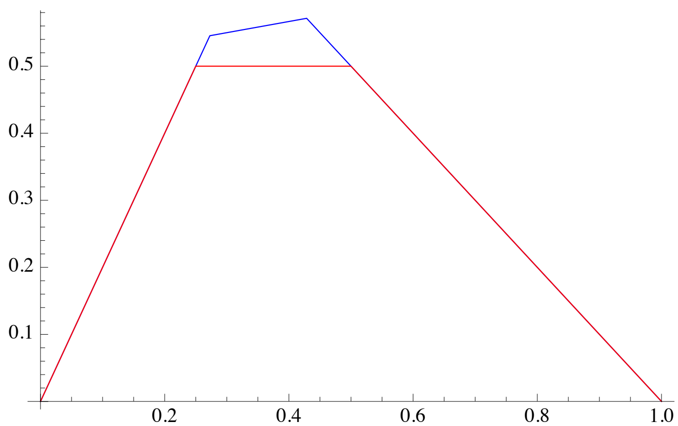

Figure 3: Comparison of the Bayes risks $\mathcal{R}_Y(\Lambda, d_\Lambda)$ (red), $\mathcal{R}_X(\Lambda, d_\Lambda)$ (blue), for $a = 2$, $b = 1$

4. Let $\Theta = \mathcal{A} = \{0, 1\}$, assume the (hypothesis testing) loss function $L(0, 0) = L(1, 1) = 0$, $L(0, 1) = L(1, 0) = 1$. Suppose that we observe the random variable X with the discrete distribution $P_\theta(X = x) = 2^{-(x+\theta)} 1_{\mathbb{Z}^+ \cap [1-\theta, \infty)}(x)$. (a) Describe the set of all non-randomized decision rules.
(b) Plot the risk set $\mathcal{R}$ in the plane. Which non-randomized rules are admissible? Why?
(c) Can you find a non-randomized minimax rule?
(d) What decision rules result from a Neyman-Pearson approach?
Hint: every number $z \in [0, 1]$ has a diadic representation of the form $\sum_{x=1}^\infty d(x) 2^{-x}$ where $d(x) = 0$ or 1.

Solution: Let $d(x)$ = probability of choosing $\theta = 1$ given that $X = x$ is observed. Thus an arbitrary rule $d = (d(0), d(1), d(2), \ldots)$, and the nonrandomized rules have 0 or 1 in each coordinate. Hence the nonrandomized rules consist of all infinite sequences of 0's and 1's. For any rule d

$$R(0, d) = \sum_{x=1}^{\infty} d(x) 2^{-x},$$

$$
\begin{aligned}
R(1,d) &= \sum_{x=0}^{\infty}(1-d(x))2^{-(x+1)} = \frac{1}{2}\sum_{x=0}^{\infty}(1-d(x))2^{-x} \\
&= 1 - \frac{1}{2}d(0) - \frac{1}{2}R(0,d).
\end{aligned}
$$

Note that as d ranges over all non-randomized rules, $R(0,d) = \sum_1^\infty d(x)2^{-x}$ ranges over all real numbers between 0 and 1; every $z \in [0,1]$ has a diadic expansion. Hence the non-randomized rules have risk points on the lines obtained by taking $d(0) = 0$ or $d(0) = 1$; Those with $d(0) = 0$ are inadmissible since the rule with $d(0) = 1$ is better (i.e. has strictly smaller $R(1,d)$); note that $X = 0$ is never observed when $\theta = 0$. When $d(0) = 1$, $R(1,d) = 2^{-1}(1 - R(0,d))$, and hence the minimax risk is $R(0,d) = R(1,d) = 1/3$. This risk is attained by the non-randomized rule $d_M(2x) = 1$, $d_M(2x+1) = 0$, $x \geq 0$. In fact, there are infinitely many randomized minimax rules; e.g. $(1, 2/3, 0, \ldots)$, $(1, 0, 1, 2/3, 0, \ldots)$, $(1, 0, 1, 0, 1, 2/3, 0, \ldots)$, and so forth.

To find a least-favorable prior, for rules d with $d(0) = 1$, write the Bayes risk as

$$
\begin{aligned}
\mathcal{R}(\lambda, d) &= \lambda R(0,d) + (1-\lambda)R(1,d) = \lambda R(0,d) + (1-\lambda)\frac{1}{2}(1 - R(0,d)) \\
&= \frac{1}{2}(1-\lambda) + \frac{1}{2}(3\lambda - 1)R(0,d).
\end{aligned}
$$

For $0 \leq \lambda < 1/3$, $3\lambda - 1 < 0$, so the Bayes risk is minimized by rules d with $R(0,d) = 1$, e.g. $d = (1, 1, \ldots)$, and the Bayes risk is λ, $0 \leq \lambda \leq 1/3$.
For $1/3 < \lambda \leq 1$, $3\lambda - 1 > 0$, so the Bayes risk is minimized by rules d with $R(0,d) = 0$; e.g. $d = (1, 0, 0, \ldots)$, with Bayes risk $\mathcal{R}(\lambda, d_\lambda) = (1-\lambda)/2$.
When $\lambda = 1/3$, $3\lambda - 1 = 0$, and all rules with $d(0) = 1$ are Bayes wrt λ, with Bayes risk $\mathcal{R}(\lambda, d_\lambda) = 1/3$. Hence the least favorable prior is $\lambda = (1/3, 2/3)$.

Here is another way: Note that for any minimax rule d_M, d_M is Bayes wrt $\lambda = (1/3, 2/3)$, and for this $\underline{\lambda}$ we have

$$
\mathcal{R}(\lambda, d_M) = \frac{1}{3} = \sup_{\theta=0,1} R(\theta, d_M);
$$

hence by Theorem 6.2, $\underline{\lambda} = (1/3, 2/3)$ is least favorable.

The Neyman - Pearson approach would be to find d so that $R(0,d) = \alpha < 1$, and so that $R(1,d) =$ probability of type 2 error is as small as possible. Thus we are lead to a rule of the form $d = (1, d_1, d_2, \ldots)$ where $\alpha = \sum_1^\infty d_x 2^{-x} = R(0,d)$, and then $R(1,d) = 2^{-1}(1-\alpha)$.

5. Consider Example 1.3 as given on page 4 of the Chapter 5 notes.
(a) Consider calculation of the Bayes rule for the prior $\lambda_1 = \lambda$, $\lambda_2 = 1 - \lambda$ with $\lambda \in [0,1]$ via computation of the posterior risk(s) as in Theorem 5.1: in particular, compute the posterior risks

$$
E\{L(\boldsymbol{\theta}, d(\cdot|X))|X = x\}
$$

for $x \in \{R, B, G\}$.
(b) Specialize your calculation in (a) to the particular prior $\lambda = 1/2$ and thereby show that the non-randomized rule $d_2 = (0, 0, 1)$ is the Bayes rule.
(c) Determine a prior which yields all the risk points on the line between $(1, 3.6) = (R(1, d_7), R(2, d_7))$ and $(0, 6) = (R(1, d_8), R(2, d_8))$ as risks of Bayes rules.

Solution: (a) We first compute the posterior risks conditional on $X = R, B, G$ for a general rule d and prior λ:

$$\begin{aligned}
E\{L(\boldsymbol{\theta}, d(\cdot|X))|X = R\} &= 10\left\{(1-d_R)\frac{\lambda}{\lambda + 4(1-\lambda)} + 6d_R\frac{4(1-\lambda)}{\lambda+4(1-\lambda)}\right\} \\
&= \frac{10\lambda}{\lambda+4(1-\lambda)} + \frac{24(1-\lambda)-10\lambda}{\lambda+4(1-\lambda)}d_R \\
&= \frac{10\lambda}{\lambda+4(1-\lambda)} + \frac{24-34\lambda}{\lambda+4(1-\lambda)}d_R.
\end{aligned}$$

Similarly,

$$\begin{aligned}
E\{L(\boldsymbol{\theta}, d(\cdot|X))|X = B\} &= \frac{20\lambda}{2\lambda+4(1-\lambda)} + \frac{24-44\lambda}{2\lambda+4(1-\lambda)}d_B, \\
E\{L(\boldsymbol{\theta}, d(\cdot|X))|X = G\} &= \frac{70\lambda}{7\lambda+2(1-\lambda)} + \frac{12-82\lambda}{7\lambda+2(1-\lambda)}d_G.
\end{aligned}$$

(b) When $\lambda = (1/2, 1/2)$, these become

$$E\{L(\boldsymbol{\theta}, d(\cdot|X))|X = x\} = \begin{cases} \frac{10}{7} + 2d_R, & x = R, \\ \frac{10}{5} + \frac{2}{5}d_B, & x = B, \\ \frac{70}{9} - \frac{58}{9}d_G, & x = G. \end{cases}$$

It follow that the Bayes rule for the prior $\lambda = (1/2, 1/2)$ is $d = (0, 0, 1)$.
(c) When $\lambda = 24/34 = 12/17$ we see that the posterior risk for $x = R$ is the same for all values of d_R; thus all the rules of the form $d = (d_R, 1, 1)$ are Bayes for $\underline{\lambda} = (12/17, 5/17)$. Taking $d_R = 0$ gives the nonrandomized rule $d = (0, 1, 1) = \delta_7$; taking $d_R = 1$ gives the nonrandomized rule $d = (1, 1, 1) = \delta_8$.