Statistics 582, Problem Set 4 Solutions

Wellner; 2/1/2006

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

$$R(0, d) = 100\{d_0(.02) + d_1(.13) + d_2(.85)\}$$
$$R(2, d) = 10\{(1 - d_0)(.70) + (1 - d_1)(.27) + (1 - d_2)(.03)\}\,,$$

and, for $\underline{\lambda} = (.2, .8)$, 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 $R(1, d_M) = R(2, d_M) = 444/157 \doteq 2.8280\ldots$. 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:

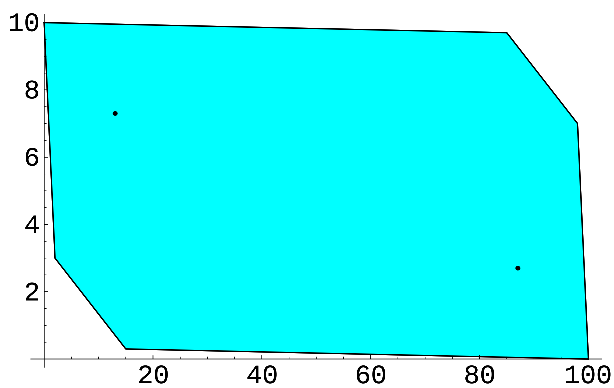

Figure 1: Risk Body.

2. Suppose that $X_n \equiv X \sim \text{Multinomial}_k(n, \underline{\theta})$.
A. Suppose that the prior distribution on θ is given by a Dirichlet distribution, Dirichlet($\underline{\alpha}$):

$$\lambda(\underline{\theta}) = \frac{\Gamma(\alpha_1 + \cdots + \alpha_k)}{\prod_{j=1}^{k} \Gamma(\alpha_j)} \theta_1^{\alpha_1 - 1} \cdots \theta_k^{\alpha_k - 1} 1_{[\underline{\theta}:\ \sum \theta_i = 1]} \,.$$

Verify the computation of the Bayes estimator for squared error loss given in example 4.3.4
B. What is the posterior distribution for θ? Find the mode of the posterior distribution (along the lines of our computations of the MLE of the multinomial)

and compare it with the MLE.
C. Find a minimax estimator d_M of $\underline{\theta}$.

Solution: A. If $\underline{\theta} \sim$ Dirichlet($\underline{\alpha}$) then $\theta_j \sim$ Beta$(\alpha_j, \sum_{j' \neq j} \alpha_{j'})$, and hence from our computations of the mean of a Beta, $E(\theta_j) = \alpha_j / \sum_{i=1}^k \alpha_i$, and as a vector $E(\underline{\theta}) = \underline{\alpha} / \sum_{i=1}^k \alpha_i$ Since the posterior distribution of $\underline{\theta}$ is Dirichlet($\underline{\alpha} + \underline{X}$), the posterior mean is

$$d_\Lambda(\underline{X}) = E(\underline{\theta}|\underline{X}) = (\underline{\alpha} + \underline{X})/(\sum_i \alpha_i + n).$$

B. As noted in A, the posterior density is Dirichlet($\underline{\alpha} + \underline{X}$):

$$\lambda(\underline{\theta}|\underline{X}) = \frac{\Gamma(\alpha_1 + \cdots + \alpha_k + n)}{\prod_{j=1}^k \Gamma(\alpha_j + X_j)} \theta_1^{\alpha_1 + X_1 - 1} \cdots \theta_k^{\alpha_k + X_k - 1} 1_{[\underline{\theta}:\ \sum \theta_j = 1]}.$$

To find the mode of the posterior, we need to find the value of $\underline{\theta}$ which maximizes $\lambda(\underline{\theta}|\underline{X})$ over the set $\sum_j \theta_j = 1$, or equivalently which maximizes

$$\sum_{j=1}^k (\alpha_j + X_j - 1) \log \theta_j + c(\sum_{j=1}^k \theta_j - 1).$$

Thus we need to solve

$$\frac{\alpha_j + X_j - 1}{\theta_j} + c = 0, \qquad j = 1, \ldots, k. \tag{1}$$

and

$$\sum_{j=1}^k \theta_j = 1. \tag{2}$$

The first equation yields

$$\theta_j^{mode} = \frac{\alpha_j + X_j - 1}{-c}, \qquad j = 1, \ldots, k;$$

substitution of this into (2) yields

$$1 = \sum_{j=1}^k \theta_j^{mode} = \frac{1}{-c} \left\{ \sum_{j=1}^k \alpha_j + n - k \right\},$$

and hence $-c = \sum_j \alpha_j + n - k$. Thus the mode of the posterior is given by

$$\underline{\theta}^{mode} = \frac{\underline{\alpha} + \underline{X} - \underline{1}}{\sum \alpha_j + n - k}.$$

When $\underline{\alpha} = \underline{1}$ (the vector of all 1's), then the mode of the posterior equals the MLE $\hat{\theta} = \underline{X}/n$. Note that $\underline{\alpha} = \underline{1}$ yields a uniform prior over θ.

C. As shown in class, if $\underline{X} \sim \text{Mult}_k(n; \underline{\theta})$ and $\underline{\theta} \sim \text{Dirichlet}(\underline{\alpha})$, then the Bayes estimator of $\underline{\theta}$ for squared error loss is $d_\Lambda(\underline{X}) = (\underline{\alpha} + \underline{X})/(\sum \alpha_i + n)$ For $\alpha_1 = \cdots = \alpha_k = \alpha$, this yields the Bayes estimator

$$d_\Lambda(\underline{X}) \;=\; \frac{\alpha\underline{1} + \underline{X}}{k\alpha + n} = \frac{k\alpha}{k\alpha + n}\frac{\underline{1}}{k} + \frac{n}{k\alpha + n}\frac{\underline{X}}{n}.$$

Note that $d_{\Lambda,i}(\underline{X}) = (\alpha + X_i)/(k\alpha + n)$ has

$$\begin{aligned}
Var_{\underline{\theta}}(d_{\Lambda,i}(X) &= \frac{n\theta_i(1-\theta_i)}{(k\alpha+n)^2},\\
E_{\underline{\theta}}(d_{\Lambda,i}(X)) &= \frac{\alpha + n\theta_i}{k\alpha + n},\\
\text{bias}_{\underline{\theta}}(d_{\Lambda,i}(X)) &= \frac{\alpha - k\alpha\theta_i}{k\alpha + n}.
\end{aligned}$$

Thus the risk is

$$\begin{aligned}
R(\underline{\theta}, \underline{d}_\Lambda) &= E_{\underline{\theta}}|\underline{\theta} - \underline{d}_\Lambda(\underline{X})|^2\\
&= \sum_{i=1}^{k}\left\{Var_{\underline{\theta}}(d_{\Lambda,i}(\underline{X})) + \text{bias}^2_{\underline{\theta}}(d_{\Lambda,i})\right\}\\
&= \frac{1}{(k\alpha+n)^2}\sum_{i=1}^{k}\left\{n\theta_i(1-\theta_i) + (\alpha - k\alpha\theta_i)^2\right\}\\
&= \frac{1}{(k\alpha+n)^2}\left\{n - k\alpha^2 + (\alpha^2k^2 - n)\sum_{i=1}^{k}\theta_i^2\right\} \qquad \text{since} \quad \sum\theta_i = 1\\
&= \frac{(1-1/k)}{(1+\sqrt{n})^2} \qquad \text{if} \quad \alpha = \frac{\sqrt{n}}{k}.
\end{aligned}$$

which is constant in $\underline{\theta}$. Hence by corollary 5.6.3

$$\begin{aligned}
d_\Lambda(\underline{X}) &= \frac{\sqrt{n}}{\sqrt{n}+n}\frac{\underline{1}}{k} + \frac{n}{\sqrt{n}+n}\frac{\underline{X}}{n}\\
&= (1-\lambda_n)\frac{\underline{1}}{k} + \lambda_n\underline{\hat{p}}_n
\end{aligned}$$

is minimax for estimation of $\underline{\theta}$.

3. Find the limit distribution of the minimax estimator d_M in problem 2 (i.e. $\sqrt{n}(d_M(\underline{X}_n) - \underline{\theta}) \to_d \;\; ??$).

Solution: Note that $\sqrt{n}(1-\lambda_n) = \lambda_n \to 1$. Hence

$$\begin{aligned}
\sqrt{n}(d_M(\underline{X}_n) - \underline{\theta}) &= \sqrt{n}\{\lambda_n \underline{\hat{p}}_n + (1-\lambda_n)\frac{1}{k} - (\lambda_n + 1 - \lambda_n)\underline{\theta}\} \\
&= \lambda_n \sqrt{n}(\underline{\hat{\theta}}_n - \underline{\theta}) + \sqrt{n}(1-\lambda_n)(\frac{1}{k} - \underline{\theta}) \\
&\to_d N_k(0, \Sigma) + \frac{1}{k} - \underline{\theta} \\
&= N_k(\frac{1}{k} - \underline{\theta}, \Sigma)
\end{aligned}$$

where $\Sigma = \text{diag}(\underline{\theta}) - \underline{\theta}\underline{\theta}^T$.

4. 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 $L(1,1) = L(2,2) = 0$, $L(1,2) = 10 = L(2,1)$, and the prior distribution is $\lambda = (\lambda_0, \lambda_1) = (.5, .5)$.

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

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

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 $\int p_0 \wedge p_1 d\mu = 1 - d_{TV}(P_0, P_1)$