

Statistics 582, Problem Set 4 Solutions

Wellner; 2/7/2007

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
$p_0(x)$.04	.16	.80
$p_1(x)$.75	.20	.05

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 $d_i = \text{prob of action 1 when } x = i \text{ is observed}$, $i = 0, 1, 2$. Then the risks are

$$R(0, d) = 100\{d_0(.04) + d_1(.16) + d_2(.80)\}$$

$$R(1, d) = 10\{(1 - d_0)(.75) + (1 - d_1)(.20) + (1 - d_2)(.05)\},$$

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

$$\begin{aligned} \mathcal{R}(\lambda, d) &= (.2)R(0, d) + (.8)R(1, d) \\ &= 8 + \{-520d_0 + 1.6d_1 + 15.6d_2\} \end{aligned}$$

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

$$\{4d_0 + 16d_1 + 80d_2\} = 10 - 7.5d_0 - 2.0d_1 - .5d_2.$$

Solving for d_0 yields

$$d_0 = (10 - 18d_1 - 80.5d_2)/11.5,$$

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

$$\begin{aligned} R(0, d) = R(1, d) &= 4 \frac{10 - 18d_1 - 80.5d_2}{11.5} + 16d_1 + 80d_2 \\ &= \frac{40}{11.5} + \left(16 - \frac{4(18)}{11.5}\right)d_1 + \left(80 - \frac{4 \cdot 80.5}{11.5}\right)d_2 \end{aligned}$$

which is minimized by $d_1 = 0$, $d_2 = 0$; then $d_0 = 10/11.5 \approx 0.869565$. Hence the minimax rule is $d_M = (1, 20/23, 0)$, and the corresponding common risk is $R(1, d_M) = R(2, d_M) = 80/23 \doteq 3.47826\dots$. 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	80	16	4	20	84	96	100
$R(1, d)$	10	9.5	8.0	2.5	0.5	2.0	7.5	0

Here is a plot of the risk body:

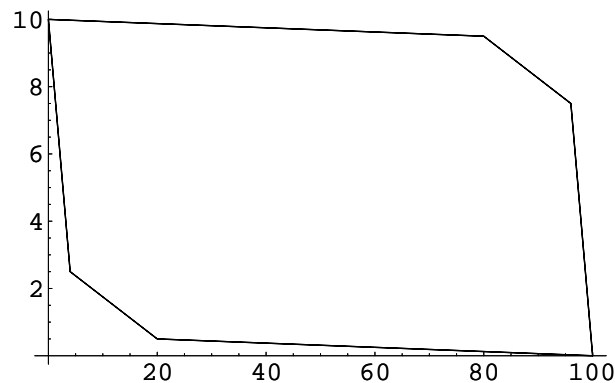


Figure 1: Risk Body.

- Let X be a random variable with finite first moment: $E|X| < \infty$. Show that $f(b) \equiv E|X - b|$ is minimized by $b =$ any median of the distribution F of X . [A median m of F is any value satisfying $F(m) = P(X \leq m) \geq 1/2$ and $1 - F(m-) = P(X \geq m) \geq 1/2$; see Lehmann and Casella, TPE, page 62, problems 1.7 and 1.8.]

Solution: Suppose that m is a median of F . From Lehmann and Casella problem 1.7, it follows that $m_0 \leq m \leq m_1$ so that the set of medians is a closed interval. This is easily proved as follows: suppose that \mathcal{M} is the set of medians of F . Note that \mathcal{M} is always non-empty since $m_0 \equiv \inf\{x : F(x) \geq 1/2\} \in \mathcal{M}$. If $\mathcal{M} = \{m_0\}$, then $[m_0, m_0] = \{m_0\}$ is closed. If $a, b \in \mathcal{M}$ with $a < b$, then if $c \in (a, b)$ we have $P(X \leq c) \geq P(X \geq a) \geq 1/2$ (since $a \in \mathcal{M}$), and

$P(X \geq c) \geq P(X \geq b) \geq 1/2$ (since $b \in \mathcal{M}$). Thus $c \in \mathcal{M}$ and hence $(a, b) \subset \mathcal{M}$. Let $(m_0, m_1) = \cup_{a, b \in \mathcal{M}} (a, b)$ be the union of all the open intervals contained in \mathcal{M} . Then if $m \in (m_0, m_1)$

$$\begin{aligned} 1/2 \leq P(X \leq m) &= E1\{X \leq m\} \rightarrow E1\{X < m_1\} \leq E1\{X \leq m_1\} = P(X \leq m_1), \quad \text{and} \\ 1/2 \leq P(X \geq m) &= E1\{X \geq m\} \rightarrow E1\{X \geq m_1\} = P(X \geq m_1) \end{aligned}$$

as $m \nearrow m_1$ by the dominated convergence theorem. Thus $m_1 \in \mathcal{M}$. Similarly,

$$\begin{aligned} 1/2 \leq P(X \leq m) &= E1\{X \leq m\} \rightarrow E1\{X \leq m_0\} \leq P(X \leq m_0), \quad \text{and} \\ 1/2 \leq P(X \geq m) &= E1\{X \geq m\} \rightarrow E1\{X > m_0\} \leq P(X \geq m_0) \end{aligned}$$

as $m \searrow m_0$ by the dominated convergence theorem. Thus $m_0 \in \mathcal{M}$ and we conclude that $[m_0, m_1] \subset \mathcal{M}$. On the other hand $\mathcal{M} \subset [m_0, m_1]$ with $m_0 \equiv \inf\{x : F(x) \geq 1/2\}$ and $m_1 \equiv \inf\{x : F(x) > 1/2\}$.

Suppose that $c > m_1$. Then by examining the graphs of $|x - c|$ and $|x - m|$ we see that

$$\begin{aligned} |x - c| - |x - m| &= (m - c)1_{[x \geq c]} + (c - m)1_{[x \leq m]} + \{(c - x) - (x - m)\}1_{[m < x < c]} \\ &= (c - m) \{1_{[x \leq m]} - 1_{[x \leq c]}\} + (c + m - 2x)1_{[m < x < c]} \\ &= (c - m) \{1_{[x \leq m]} - 1_{[x \leq c]}\} + 2(c - x)1_{[m < x < c]} - (c - m)1_{[m < x < c]} \\ &= (c - m) \{1_{[x \leq m]} - 1_{[x > m]}\} + 2(c - x)1_{[m < x < c]}. \end{aligned}$$

Replacing x by X and taking expectations across the identity with respect to X yields

$$\begin{aligned} E|X - c| - E|X - m| &= (c - m)\{P(X \leq m) - P(X > m)\} + 2E\{(c - X)1_{[m < X < c]}\} \\ &> 0 + 0 = 0 \end{aligned}$$

since m is a median of F implies that $P(X \leq m) - P(X > m) \geq 0$ and $c > m_1 \geq m_0$ implies that $E\{(c - X)1_{[m < X < c]}\} = E\{(c - X)1_{[m_1 < X < c]}\} > 0$. Similarly, if $c < m_0$,

$$|x - c| - |x - m| = (m - c)(1_{[x \geq m]} - 1_{[x < m]}) + 2(x - c)1_{[c < x < m]},$$

and taking expectations yields

$$\begin{aligned} E|X - c| - E|X - m| &= (m - c)\{P(X \geq m) - P(X < m)\} + 2E\{(X - c)1_{[c < X < m]}\} \\ &> 0. \end{aligned}$$

Thus $E|X - b|$ is minimized by any median of the distribution F of X .

3. Suppose that X_1, \dots, X_n are i.i.d. Exponential(θ) (so the X 's have density $p_\theta(x) = \theta e^{-\theta x} 1_{(0, \infty)}(x)$. with respect to Lebesgue measure on R , and that $\theta \sim \Gamma(\alpha, \beta)$:

$$\lambda(\theta) = \beta \frac{(\beta\theta)^{\alpha-1}}{\Gamma(\alpha)} \exp(-\beta\theta) 1_{[0, \infty)}(\theta).$$

(a) Find the Bayes rule $d_B(\underline{X})$ for estimation of θ with squared error loss $L(\theta, a) = |\theta - a|^2$. Find the Bayes rule $d_{Bw}(\underline{X})$ for estimation of θ with weighted squared error loss $L(\theta, a) = (\theta - a)^2/\theta$. Is the maximum likelihood estimator among either of these families of Bayes estimators?

(b) Are the Bayes estimators d_B and d_{Bw} consistent? What are the limit distributions of d_B and d_{Bw} ? Compare them with the maximum likelihood estimator.

(c) Suppose that instead of the Gamma prior distribution, θ has the Pareto(θ_0, α) distribution with density λ given by

$$\lambda(\theta) = \left(\frac{\alpha}{\theta_0}\right) \left(\frac{\theta_0}{\theta}\right)^{\alpha+1} 1_{(\theta_0, \infty)}(\theta);$$

here $E(\theta) = \frac{\alpha}{\alpha-1}\theta_0$ where $\alpha > 1$ and $\theta_0 > 0$ are known. What can you say about the Bayes estimator for squared error loss with this prior? For what values of θ_0 is the Bayes rule consistent?

Solution: (a) The posterior distribution is Gamma($\alpha + n, \beta + \sum X_i$). Thus the Bayes rule for $L(\theta, a) = (\theta - a)^2$ is

$$d_B(\underline{X}) = \frac{\alpha + n}{\beta + \sum X_i}.$$

For $L(\theta, a) = (\theta - a)^2/\theta$, the Bayes rule is

$$d_{Bw}(\underline{X}) = \frac{E(\theta K(\theta) | \underline{X})}{E(K(\theta) | \underline{X})} = \frac{1}{E(1/\theta | \underline{X})} = \frac{\alpha + n - 1}{\beta + \sum X_i}$$

since, for $\theta \sim \text{Gamma}(\alpha, \beta)$ we have

$$E(1/\theta) = \frac{\beta}{\alpha - 1}$$

if $\alpha > 1$. Thus the MLE $1/\bar{X}_n$ is *not* among either of these families of estimators.

(b) Both d_B and d_{Bw} are consistent and asymptotically equivalent to the MLE $1/\bar{X}_n$:

$$\begin{aligned} \sqrt{n} \{d_B(\underline{X}) - 1/\bar{X}_n\} &= \sqrt{n} \left\{ \frac{1 + n^{-1}\alpha}{\bar{X}_n + n^{-1}\beta} - \frac{1}{\bar{X}_n} \right\} \\ &= n^{-1/2} \frac{\alpha \bar{X}_n - \beta}{\bar{X}_n(\bar{X}_n + n^{-1}\beta)} = O(n^{-1/2})O_p(1) = o_p(1), \end{aligned}$$

and similarly for d_{Bw} . Thus, for $d = d_B$ or $d = d_{Bw}$ we have, since $I(\theta) = \theta^{-2}$,

$$\sqrt{n}(d(\underline{X}) - \theta) = \sqrt{n}\left(\frac{1}{\bar{X}_n} - \theta\right) + o_p(1) \rightarrow_d N(0, 1/I(\theta)) = N(0, \theta^2).$$

(c) When the prior is $\text{Pareto}(\theta_0, \alpha)$, the posterior density is of the form

$$\begin{aligned} \lambda(\theta|\underline{X}) &= \frac{\theta^n \exp(-\theta \sum X_i) (\alpha \theta_0^{-1}) (\theta_0/\theta)^{\alpha+1} 1_{(\theta_0, \infty)}(\theta)}{\int_{\theta_0}^{\infty} s^n \exp(-s \sum X_i) (\alpha \theta_0^{-1}) (\theta_0/s)^{\alpha+1} ds} \\ &= \frac{\theta^{n-\alpha-1} \exp(-\theta \sum X_i) 1_{(\theta_0, \infty)}(\theta)}{\int_{\theta_0}^{\infty} s^{n-\alpha-1} \exp(-s \sum X_i) ds}, \end{aligned}$$

which is concentrated on (θ_0, ∞) . Thus the Bayes rule $d_B(\underline{X}) = E(\theta|\underline{X})$ takes values in (θ_0, ∞) a.s.. Similar to the argument in class in the Bernoulli(θ) example, $Z_n = d_B(\underline{X}) = E(\theta|X_1, \dots, X_n)$ is a martingale and hence $Z_n = d_B(\underline{X}) \rightarrow E(\theta|X_1, X_2, \dots)$. But $\hat{\theta} = \bar{X}_n^{-1} \rightarrow_{a.s.} \theta$ for each fixed $\theta \in (0, \infty)$, and hence

$$P_\Lambda(\hat{\theta}_n \rightarrow \theta) = \int P_\theta(\hat{\theta}_n \rightarrow \theta) d\Lambda(\theta) = 1.$$

Hence $\hat{\theta}_n \rightarrow \theta$ a.s. P_Λ , and this implies that θ is $\mathcal{F}_\infty \equiv \sigma(X_1, X_2, \dots)$ measurable. Therefore $E(\theta|X_1, X_2, \dots) = \theta$ a.s. and $d_B(\underline{X}) \rightarrow \theta$ a.s. P_Λ . This in turn implies that $d_B(\underline{X}) \rightarrow_{a.s.} \theta$ for Λ -a.e. θ . This suggests that d_B might be inconsistent for $\theta \in (0, \theta_0)$, and this is in fact the case since $d_B(\underline{X}) < \theta_0$. When the true $\theta < \theta_0$, it is possible to show that $d_B(\underline{X}) \rightarrow_{a.s.} \theta_0 > \theta$ and that the posterior distributions converge to point mass at θ_0 .

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 $L(0, 0) = L(1, 1) = 0$, $L(0, 1) = 10 = L(1, 0)$, 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

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

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,

$$1 + \int_{[p_0 < p_1]} (p_0 - p_1) d\mu = 1 - d_{TV}(P_0, P_1)$$

$$\begin{aligned}
&= 1 - \left\{ 1 - \int p_0 \wedge p_1 d\mu \right\} \\
&= \int p_0 \wedge p_1 d\mu;
\end{aligned}$$

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 in problem 1,

$$\begin{aligned}
\rho(P_0, P_1) &= \int p_0 \wedge p_1 d\mu = .4 + .16 + .05 = .25, \\
d_{TV}(P_0, P_1) &= 1 - .25 = .75 = 2^{-1} \int |p_0 - p_1| d\mu.
\end{aligned}$$

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 $\rho(P_0, P_1) = .25$. Thus the Bayes risk equals $10 \times .125 = 1.25$ in this case.

5. **Optional bonus problem:** Suppose that $\Theta = \{\theta_1, \theta_2\}$, $\mathbf{A} = \{a_1, a_2, a_3, a_4\}$, and that the loss function $L(\theta, a)$ is given by the following table:

θ/a	a_1	a_2	a_3	a_4
θ_1	1	1	2	2
θ_2	0	1	0	1

Further suppose that $P_{\theta_j}(X = 0) = 1$ for $j = 1, 2$.

- Find the decision risk set \mathcal{R} .
- Find the decision rules that are Bayes with respect to the prior distribution $\lambda = (1, 0)$.
- Show that the rule d_0 for which $R(\theta_1, d_0) = 1$ and $R(\theta_2, d_0) = 1$ is Bayes with respect to $\lambda = (1, 0)$ and also minimax, but that it is not admissible.
- Relate this example to our theorem about admissibility of Bayes rules.