Statistics 582, Problem Set 6 Solutions

Wellner; 2/17/2010

1. (a) Suppose that $X \sim F$ on $\mathbb{R}$. Let $\rho_p(x) \equiv x(p - 1_{(-\infty,0]}(x))$ for $p \in (0,1)$ and $x \in \mathbb{R}$. Show that $h_p(t) \equiv E_F \rho_p(X - t)$ is minimized by $F^{-1}(p)$. (Note that $\rho_{1/2}(x) = |x|/2$, so $h_{1/2}(t)$ is minimized by any median of F as we already know.)
(b) Suppose that $(X|\boldsymbol{\theta} = \theta) \sim P_\theta$ on $\mathbb{R}$ and that $\boldsymbol{\theta} \sim \Lambda$. Fix $p \in (0,1)$, and let $L(\theta, a) \equiv \rho_p(\theta - a)$. Show that the Bayes rule with respect to the prior Λ for this loss function is the $p-$th quantile of the posterior distribution $\Lambda(\theta|X)$.

   Solution: (a) First note that ρ_p has derivative $p - 1$ for $x < 0$ and derivative p for $x > 0$, while the derivative is not defined at $x = 0$. If we assume that F is continuous, then we conclude that

$$\begin{aligned} h_p'(t) &= E_F\{(p-1)1_{[X\le t]} + p1_{[X>t]}\} = E_F\{(p-1)1_{[X\le t]} + p(1 - 1_{[X\le t]})\} \\ &= E_F\{p - 1_{[X\le t]}\} = p - F(t). \end{aligned}$$

   Thus $h_p'(t) = 0$ for $t = F^{-1}(p)$ if F is strictly increasing, and more generally for any $p-$th quantile of F. A complete proof of this can be carried out in general along the lines of the proof of the bonus problem on problem set 5 for the case $p = 1/2$. This view of the $p-$th quantile as the solution of an optimization problem is the starting point for the theory of "quantile regression" which was initiated by Bassett and Koenker (1986); see R. Koenker's (2005) book, *Quantile Regression*.
(b) By Theorem 5.5, it suffices to minimize the posterior risk

$$E\{L(\boldsymbol{\theta}, d(X))|X\} = \int \rho_p(\theta - d(X))d\Lambda(\theta|X)$$

   almost surely (for fixed X). But by (a), this posterior risk is minimized by the $p-$th quantile of the posterior distribution of $\Lambda(\theta|X)$.

2. 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 $d_i \equiv$ probability of action 1 given that i is observed.
(i) The Bayes risks for observing X or Y are:

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

and

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

(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 $\mathcal{R}_X(\lambda, d_\lambda) = \frac{5}{12} a > \frac{3}{8} a = \mathcal{R}_Y(\lambda, d_\lambda)$, so we would prefer to observe Y.
(iii) The risks are:

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

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

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

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

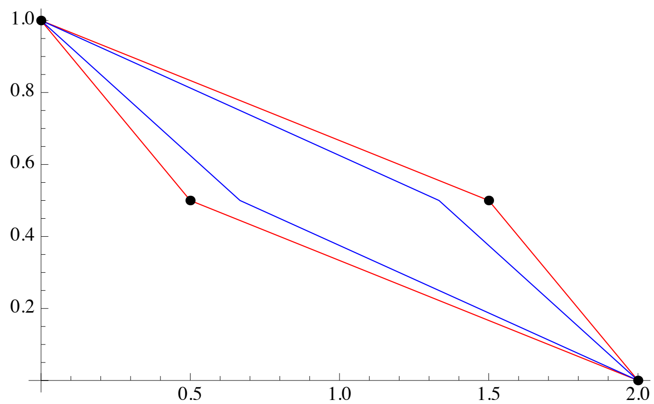

Figure 1: Comparison of the risk bodies $\mathcal{R}_Y$ (red), $\mathcal{R}_X$ (blue), for $a = 2$, $b = 1$

First consider $\mathcal{R}_X(\lambda, d)$:
For $0 \leq \lambda \leq (1 + 4r/3)^{-1}$, both coefficients are > 0, so $d_\lambda = (0, 0)$, and $\mathcal{R}_X(\lambda, d_\lambda) = a\lambda$.
For $(1 + 4r/3)^{-1} \leq \lambda \leq (1 + 2r/3)^{-1}$, $d_\lambda = (0, 1)$ and $\mathcal{R}_X(\lambda, d_\lambda) = a\lambda/3 + b(1 - \lambda)/2$.
For $(1 + 2r/3)^{-1} \leq \lambda \leq 1$, $d_\lambda = (1, 1)$ and $\mathcal{R}_X(\lambda, d_\lambda) = b(1 - \lambda)$.
Now consider $\mathcal{R}_Y(\lambda, d)$:
For $0 \leq \lambda \leq (1 + 3r/2)^{-1}$, both coefficients are > 0, so $d_\lambda = (0, 0)$, and $\mathcal{R}_Y(\lambda, d_\lambda) = a\lambda$.
For $(1 + 3r/2)^{-1} \leq \lambda \leq (1 + r/2)^{-1}$, $d_\lambda = (0, 1)$, and $\mathcal{R}_Y(\lambda, d_\lambda) = a\lambda/4 + (b(1 - \lambda)/2$.
For $(1 + r/2)^{-1} \leq \lambda \leq 1$, $d_\lambda = (1, 1)$ and $\mathcal{R}_Y(\lambda, d_\lambda) = b(1 - \lambda)$.
Graphing this as a function of λ yields:

Thus the Bayes risk is *always* smaller for Y.

3. Consider Example 5.5.4 on pages 16 and 17 of the Chapter 5 notes.
(a) Show that the variance of $\hat{\psi}$ is given by

$$Var(\hat{\psi}_n) = \frac{1}{n}\left\{\frac{1}{B}\sum_{j=1}^{B}\frac{\theta_j}{\xi_j} - \psi(\theta)^2\right\}.$$

[Hint: use the formula $Var(Y) = EVar(Y|X) + Var[E(Y|X)]$ twice.]
(b) Use the result of (a) to show that

$$Var(\hat{\psi}_n) \leq \frac{1}{n\delta}$$

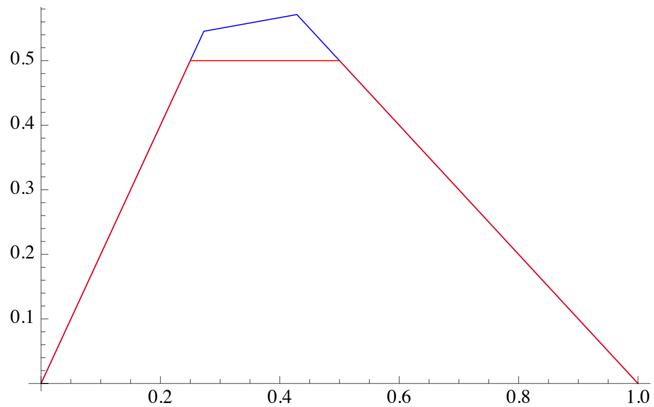

Figure 2: Comparison of the Bayes risks $\mathcal{R}_Y(\Lambda, d_\Lambda)$ (red), $\mathcal{R}_X(\Lambda, d_\Lambda)$ (blue), for $a = 2$, $b = 1$

under the assumption that $\xi_j \geq \delta > 0$ for all $1 \leq j \leq B$.

Solution: (a) Since the (X_i, R_i, Y_i)'s are i.i.d.,

$$\begin{aligned}
Var(\hat{\psi}_n) &= n^{-1} Var\left(\frac{R_1 Y_1}{\xi_{X_1}}\right) \\
&= n^{-1}\left\{ EVar\left(\frac{R_1 Y_1}{\xi_{X_1}} | R_1, X_1\right) + Var\left(E\left(\frac{R_1 Y_1}{\xi_{X_1}} | R_1, X_1\right)\right)\right\} \\
&= n^{-1}\left\{ E\left(\frac{R_1^2}{\xi_{X_1}^2}\theta_{X_1}(1-\theta_{X_1})\right) + Var\left(\frac{R_1}{\xi_{X_1}}\theta_{X_1}\right)\right\} \\
&= n^{-1}\left\{ EE\left(\frac{R_1^2}{\xi_{X_1}^2}\theta_{X_1}(1-\theta_{X_1}) | X_1\right)\right. \\
&\qquad \left. + EVar\left(\frac{R_1}{\xi_{X_1}}\theta_{X_1} | X_1\right) + Var\left(E\left(\frac{R_1}{\xi_{X_1}}\theta_{X_1} | X_1\right)\right)\right\} \\
&= n^{-1}\left\{ E\left(\frac{\theta_{X_1}(1-\theta_{X_1})}{\xi_{X_1}}\right)\right. \\
&\qquad \left. + E\left(\frac{\theta_{X_1}^2}{\xi_{X_1}^2}\xi_{X_1}(1-\xi_{X_1})\right) + Var(\theta_{X_1})\right\} \\
&= n^{-1}\left\{ \frac{1}{B}\sum_{j=1}^{B}\frac{\theta_j(1-\theta_j)}{\xi_j} + \frac{1}{B}\sum_{j=1}^{B}\theta_j^2\frac{1-\xi_j}{\xi_j} + \frac{1}{B}\sum_{j=1}^{B}(\theta_j - \overline{\theta})^2\right\} \\
&= n^{-1}\left\{ \frac{1}{B}\sum_{j=1}^{B}\frac{\theta_j}{\xi_j} - \psi(\theta)^2\right\}.
\end{aligned}$$

(b) Since $\xi_j \geq \delta$ and $\theta_j \leq 1$ for all j, it follows that

$$Var(\hat{\psi}_n) \leq n^{-1} \frac{1}{B} \sum_{j=1}^{B} \frac{1}{\delta} = \frac{1}{n\delta}.$$