Statistics 582, Problem Set 3 Solutions

Wellner; 1/28/2015

1. Consider the zero-inflated Poisson distribution p_θ as described in Example 3 of the handout on M- and Z- theorems. Suppose that $X_1, \ldots, X_n$ i.i.d. p_θ are observed.
(a) Set up alternative estimating equations for $\theta = (\gamma, \lambda)$ where $\gamma \in [0,1]$ and $\lambda > 0$ based on $g_1(x) = x$ and $g_2(x) = x^2$. Express your alternative estimator $\hat\theta_n = (\hat\gamma_n, \hat\lambda_n)$ of θ explicitly in terms of the first and second moments, $\overline{X}_n$ and $\overline{X^2}_n$, of the data, and show that your estimators are consistent when the model holds.
(b) Use Huber's Z-theorem to show that $\sqrt{n}(\hat\theta_n - \theta_0) \to_d N_2(0, \Sigma)$ and give the form of Σ.
(c) What happens if the X_i's are i.i.d. $p \notin \mathcal{P} = \{p_\theta : \theta \in \Theta\}$? Describe the parameter $\theta(P)$ to which $\hat\theta_n$ converges in probability and use Huber's theorem to establish a limit theorem for $\sqrt{n}(\hat\theta_n - \theta(P))$ in this case. What other methods do you have available in this case?

Solution: (a) For $X \sim p_\theta$ we compute

$$\begin{aligned} E_\theta X &= (1-\gamma)\lambda, \\ E_\theta X^2 &= (1-\gamma)(\lambda + \lambda^2). \end{aligned}$$

Thus the method of moments estimators $\hat\theta_n$ of θ based on $g_1(x) = x$ and $g_2(x) = x^2$ are given by

$$\Psi_n(\hat\theta_n) = \mathbb{P}_n \psi(X; \hat\theta_n) = 0$$

where

$$\begin{aligned} \psi_{\theta,1}(x) &= \psi_1(x;\theta) \equiv x - t_1(\theta) = x - (1-\gamma)\lambda, \\ \psi_{\theta,2}(x) &= \psi_2(x;\theta) \equiv x^2 - t_2(\theta) = x^2 - (1-\gamma)(\lambda+\lambda^2). \end{aligned}$$

The population versions of the estimating equations, $\Psi(\theta) = P\psi(X;\theta) = 0$ can be rewritten as

$$E_P X = (1-\gamma)\lambda, \quad \text{and} \quad E_P X^2 = (1-\gamma)(\lambda+\lambda^2),$$

and hence we find that

$$\frac{E_P X^2}{E_P X} = 1 + \lambda, \quad \text{or} \quad \lambda = \lambda(P) = \frac{E_P X^2}{E_P X} - 1.$$

This leads to

$$(1-\gamma)\lambda = E_P(X) \quad \text{or} \quad \gamma = 1 - \frac{E_P(X)}{E_P(X^2) - E_P(X)} = \frac{Var_P(X) - E_P(X)}{E_P(X^2) - E_P(X)}.$$

The corresponding estimators become

$$\begin{aligned} \hat\gamma_n &= 1 - \frac{\overline{X}_n^2}{\overline{X^2}_n - \overline{X}_n} = \frac{S_n^2 - \overline{X}_n}{\overline{X^2}_n - \overline{X}_n}, \quad \text{and} \\ \hat\lambda_n &= \frac{\overline{X^2}_n}{\overline{X}_n} - 1. \end{aligned}$$

When the model holds, these are consistent:

$$\hat{\lambda}_n \to_p \frac{(1-\gamma)(\lambda+\lambda^2)}{(1-\gamma)\lambda} - 1 = \lambda, \quad \text{and}$$
$$\hat{\gamma}_n \to_p \frac{(1-\gamma)[1+\gamma\lambda] - (1-\gamma)\lambda}{(1-\gamma)[\lambda+\lambda^2] - (1-\gamma)\lambda} = \frac{\gamma(1-\gamma)\lambda^2}{(1-\gamma)\lambda^2} = \gamma.$$

(b) If the models holds, we know $X \sim P_{\theta_0} \in \mathcal{P}$. Then with $P_0 \equiv P_{\theta_0}$ and $E_0 h(X) \equiv E_{\theta_0} h(X)$,

$$E_0 X^4 = (1-\gamma) E_{Poiss(\lambda)} X^4 = (1-\gamma)(\lambda + 7\lambda^2 + 6\lambda^3 + \lambda^4) < \infty,$$

so $X^2 \in L_2(P_0)$ and the covariance matrix $\Sigma_0 \equiv Cov_0(X, X^2)$ is well defined. But the multivariate CLT it follows that

$$\sqrt{n}(\Psi_n - \Psi)(\theta_0) = \sqrt{n}\begin{pmatrix} \overline{X}_n - E_0(X) \\ \overline{X^2}_n - E_0 X^2 \end{pmatrix} \to_d N_2(0, \Sigma_0).$$

so A2 holds. As noted in class on 21 January, condition A3 holds trivially for method of moment estimators. To verify A4 we write

$$\Psi(\theta) = P\psi(X;\theta) = P\begin{pmatrix} X - t_1(\theta) \\ X^2 - t_2(\theta) \end{pmatrix},$$

and hence

$$\dot{\Psi}(\theta_0) \quad = \quad -\dot{t}(\theta) = \begin{pmatrix} \lambda & -(1-\gamma) \\ \lambda + \lambda^2 & -(1-\gamma)(1+2\lambda) \end{pmatrix}$$

where $\det(-\dot{\Psi}(\theta_0)) = (1-\gamma)\lambda^2 > 0$ if $0 \le \gamma < 1$. Thus we conclude from Huber's theorem that

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \to_d -\dot{\Psi}_0^{-1} Z \sim N_2(0, B\Sigma_0 B^T)$$

where $B \equiv -\dot{\Psi}_0^{-1}$.

(c) As can be seen from (a), when $P \notin \mathcal{P}$, if we assume that $E_P(X^2) < \infty$, then

$$\hat{\gamma}_n \to_p (Var_P(X) - E_P(X))/(E_P(X^2) - E_P(X)), \quad \text{and}$$
$$\hat{\lambda}_n \to_p \frac{E_P(X^2)}{E_P(X)} - 1 \equiv \lambda(P).$$

Note that this $\lambda(P) \ge 0$ for any distribution on $\{0, 1, 2, \ldots\}$ with equality holding if P is Bernoulli(p) for any $p \in (0, 1]$. Supposing that $\theta_0(P)$ is well-defined, Huber's theorem holds if $E_P(X^4) < \infty$. Then $\sqrt{n}(\Psi_n - \Psi)(\theta_0) \to_d N_2(0, \Sigma_P)$ where $\Sigma_P = Cov_P(X, X^2)$ and $\dot{\Psi}(\theta_0)$ remains as in (b) but with $\theta_0 = \theta(P)$. Thus Huber's theorem yields $\sqrt{n}(\hat{\theta}_n - \theta(P)) \to_d N_2(0, B_P \Sigma_P B_P^T)$.

2. 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 $\underline{\lambda} = (.2, .8)$, 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 $R(0,d_M) = R(1,d_M) = 444/157 \doteq 2.8280\ldots$. 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:

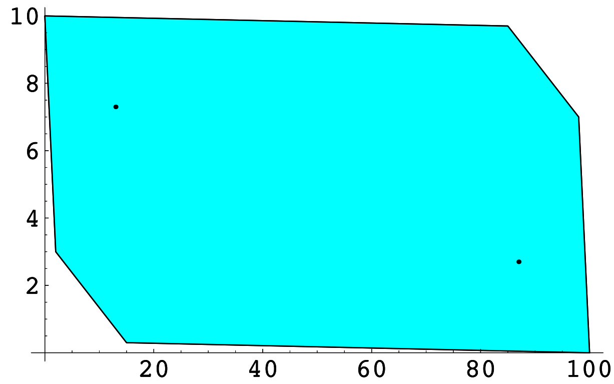

Figure 1: Risk Body.

3. 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 2 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

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

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

(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 $D = 1$.

4. 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

$$p_1(1) = P_1(X = 1) = 2/3, \qquad p_2(1) = P_2(X = 1) = 1/2,$$
$$p_1^*(1) = P_1(Y = 1) = 3/4, \qquad p_2^*(1) = P_2(Y = 1) = 1/2.$$

Let $\underset{\sim}{\lambda} = (\lambda, 1 - \lambda)$, $\lambda \in [0, 1]$ 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

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

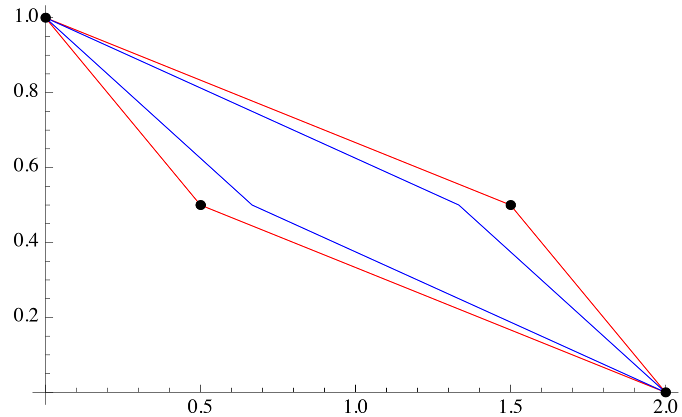

Figure 2: Comparison of the risk bodies $\mathcal{R}_Y$ (red), $\mathcal{R}_X$ (blue), for $a = 2$, $b = 1$

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:

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 3.Thus the Bayes risk is *always* small for Y.

5. **Optional bonus problem 1:** 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.

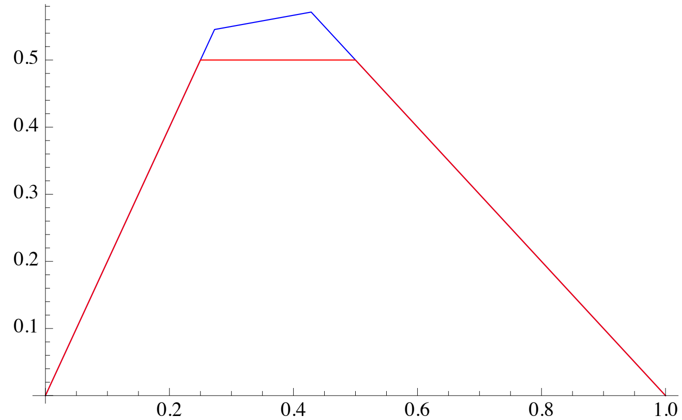

Figure 3: Comparison of the Bayes risks $\mathcal{R}_Y(\Lambda, d_\Lambda)$ (red), $\mathcal{R}_X(\Lambda, d_\Lambda)$ (blue), for $a = 2$, $b = 1$

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

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

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 $\underset{\sim}{\lambda}$ we have

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

hence by Theorem 6.2, $\underset{\sim}{\lambda} = (1/3, 2/3)$ 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 $R(1, d) = 2^{-1}(1 - \alpha)$.

6. **Optional bonus problem 2:** Let X be a random variable with distribution function F having finite first moment: $E|X| < \infty$.
(a) 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 $F(m) = P(X \le m) \ge 1/2$ and $1 - F(m-) = P(X \ge m) \ge 1/2$; see Lehmann and Casella, TPE, page 62, problems 1.7 and 1.8.]
(b) For $0 < \tau < 1$, let $\rho_\tau(x) = x(\tau - 1_{(-\infty,0)}(x))$. Consider minimizing

$$M_\tau(\theta) = E\rho_\tau(X - \theta)$$

with respect to θ. Show that the solution $\theta_0 = \theta_0(F)$ is given by the $\tau-$th quantile of F: $\theta_0(F) = F^{-1}(\tau)$.

Solution: (a) Suppose that m is a median of F. From Lehmann and Casella problem 1.7, it follows that $m_0 \le m \le m_1$ 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 $m_0 \equiv \inf\{x : F(x) \ge 1/2\} \in \mathcal{M}$. If $\mathcal{M} = \{m_0\}$, then $[m_0, m_0] = \{m_0\}$ is closed. If If $a, b \in \mathcal{M}$ with $a < b$, then if $c \in (a, b)$ we have $P(X \le c) \ge P(X \ge a) \ge 1/2$ (since $a \in \mathcal{M}$), and $P(X \ge c) \ge P(X \ge b) \ge 1/2$ (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 $m \in (m_0, m_1)$

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

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

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

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 $m_0 \equiv \inf\{x : F(x) \ge 1/2\}$ 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

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

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

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

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

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

and taking expectations yields

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

Thus $E|X-b|$ is minimized by any median of the distribution F of X.
(b) First rewrite $M_\tau(\theta)$ as

$$\begin{aligned}
M_\tau(\theta) &= E\rho_\tau(X-\theta) = (\tau-1)\int_{(-\infty,\theta]}(x-\theta)dF(x) + \tau\int_{(\theta,\infty)}(x-\theta)dF(x) \\
&= (\tau-1)\int_{(-\infty,\theta]}(x-\theta)f(x)dx + \tau\int_{(\theta,\infty)}(x-\theta)f(x)dx
\end{aligned}$$

if F is absolutely continuous with density f. Thus in this case

$$\begin{aligned}
M'_\tau(\theta) &= -(\tau-1)\int_{(-\infty,\theta]}dF(x) - \tau\int_{(\theta,\infty)}dF(x) \\
&= (1-\tau)F(\theta) - \tau(1-F(\theta)) = F(\theta)-\tau \\
&= 0
\end{aligned}$$

if θ is a $\tau-$quantile of F, $\theta=\theta_0(F)=F^{-1}(\tau)$. If F is an arbitrary distribution, then an argument as in (a) shows that any $\tau-$th quantile of F minimizes $M_\tau(\theta)$.

7. **Optional bonus problem 3:** 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 $j = 1, 2$.
(a) Find the decision risk set $\mathcal{R}$.
(b) Find the decision rules that are Bayes with respect to the prior distribution $\lambda = (1, 0)$.
(c) 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.
(d) Relate this example to our theorem about admissibility of Bayes rules.

Solution: (a) Let $d = (d_1, d_2, d_3, d_4)$ where $d_j \equiv d(j, 0)$ and $d_1 + d_2 + d_3 + d_4 = 1$. Since $P_{\theta_j}(X = 0) = 1$, it is easily seen that

$$\begin{aligned} R(\theta_1, d) &= 1 \cdot d_1 + 1 \cdot d_2 + 2 \cdot d_3 + 2 \cdot d_4 = d_1 + d_2 + 2d_3 + 2d_4, \\ R(\theta_2, d) &= 1 \cdot d_1 + 0 \cdot d_2 + 1 \cdot d_3 + 0 \cdot d_4 = d_1 + d_3. \end{aligned}$$

Using the fact that $\sum_{j=1}^{4} d_j = 1$ we find that

$$\begin{aligned} R(1, d) \equiv R(\theta_1, d) &= 1 + d_3 + d_4, \\ R(2, d) \equiv R(\theta_2, d) &= 1 - d_2 - d_4. \end{aligned}$$

The nonrandomized rules δ_i, $i = 1, \ldots, 4$ and their risks are:

a_j	δ_1	δ_2	δ_3	δ_4
1	1	0	0	0
2	0	1	0	0
3	0	0	1	0
4	0	0	0	1
$R(1, d)$	1	1	2	2
$R(2, d)$	1	0	1	0

Thus the risk set $\mathcal{R}$ is a box all sides of length 1 with its lower edge along the x (or $R(\theta_1, d)$) axis between $(1, 0)$ and $(2, 0)$.
(b) for the prior $\lambda = (1, 0)$, all the point in the risk body on the left edge of the box (i.e. all the points of the form $(1, y)$ with $0 \le y \le 1$) are the risks corresponding to Bayes rules of the form $\gamma\delta_1 + (1 - \gamma)\delta_2$ with $\gamma \in [0, 1]$.
(c) One of the points on the line in (b) is the point $(1, 1)$. This the risk point corresponding to the decision rule $d_0 = \delta_1$. It is Bayes with respect to $\lambda = (1, 0)$, but it is clearly not admissible, since the nonrandomized rule δ_2 with risk coordinates $(1, 0)$ dominates it. d_0 is also minimax, but not admissible, while δ_2 is minimax and admissible.
(d) Our theorem says that any Bayes rule with $\lambda_i > 0$ for all i yields an admissible rule, but this is clearly violated by the prior $\lambda = (1, 0)$.