# AMS 221 Homework 3

Diana Gerardo

July 3, 2019

1. (Problem 7.5) Find the least favorable priors for the two sampling distributions in Example 7.1. Does it bother you that the prior depends on the sampling?

**Solution:**

By definition, a prior distrbution $\pi^M$ is "least favorable" if

$$\inf_{\delta} r\left(\pi^M, \delta\right) = \sup_{\pi} \inf_{\delta} r\left(\pi, \delta\right)$$

this means that $\pi^M$ produces the largest possible Bayes risk among priors..

In R, we create 200 values for $\pi(\theta_1)$, $\pi(\theta_2)$ in the interval $[0,1]$ and set the constraint to $\pi(\theta_3) = 1 - \pi(\theta_1) - \pi(\theta_2)$. Thus if $\pi(\theta_1) + \pi(\theta_2) > 1$ we discard it and move on to the next combination of $\pi(\theta_1) + \pi(\theta_2)$. In the textbook we are given the following loss function under the two alternative sampling models

| | Under $f_1(x\|\theta)$ | | | Under $f_2(x\|\theta)$ | | |
|---|---|---|---|---|---|---|
| | $\theta_1$ | $\theta_2$ | $\theta_3$ | $\theta_1$ | $\theta_2$ | $\theta_3$ |
| $\delta_1(x)$ | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 1.00 |
| $\delta_2(x)$ | 2.00 | 4.00 | 0.00 | 2.00 | 4.00 | 0.00 |
| $\delta_3(x)$ | 1.60 | 3.60 | 0.25 | 0.80 | 2.80 | 0.75 |
| $\delta_4(x)$ | 0.40 | 0.40 | 0.75 | 1.20 | 1.20 | 0.25 |

Under the first alternative sampling model $f_1(x|\theta)$

In general the Bayes risk for each decision rule and 1 set of priors would be

$$0.00 \cdot \pi_1(\theta_1) + 0.00 \cdot \pi_1(\theta_2) + 1.00 \cdot \pi_1(\theta_3) = r(\pi, \delta_1)$$
$$2.00 \cdot \pi_1(\theta_1) + 4.00 \cdot \pi_1(\theta_2) + 0.00 \cdot \pi_1(\theta_3) = r(\pi, \delta_2)$$
$$1.60 \cdot \pi_1(\theta_1) + 3.60 \cdot \pi_1(\theta_2) + 0.25 \cdot \pi_1(\theta_3) = r(\pi, \delta_3)$$
$$0.40 \cdot \pi_1(\theta_1) + 0.40 \cdot \pi_1(\theta_2) + 0.75 \cdot \pi_1(\theta_3) = r(\pi, \delta_4)$$

Then among the bayes risk for each decision rule we find the minimized risk, $\inf_{\delta} r(\pi_1, \delta)$. So we calculate the Bayes risk for each decision rule and each set of priors $\{\pi_p(\theta_1), \pi_p(\theta_2), \pi_p(\theta_3)\}$ ,for $p = 1, ..., n$. Then we find the set of $\left\{\inf_{\delta} r(\pi_p, \delta)\right\}$. The largest of these minimized risks is selected,

$\sup_{\pi}\left\{\inf_{\delta} r\left(\pi_p,\delta\right)\right\}$, and its corresponding set of priors will be our least favorable priors.

So following this method, $\inf_{\delta} r\left(\pi^M,\delta\right) = \sup_{\pi}\inf_{\delta} r\left(\pi,\delta\right) = 0.6884422$ and our least favorable priors are

$$\pi_{36}^M\left(\theta_1\right) = 0.000000000, \quad \pi_{36}^M\left(\theta_2\right) = 0.1758794, \quad \pi_{36}^M\left(\theta_3\right) = 0.8241206$$
$$\pi_{235}^M\left(\theta_1\right) = 0.005025126, \quad \pi_{235}^M\left(\theta_2\right) = 0.1708543, \quad \pi_{235}^M\left(\theta_3\right) = 0.8241206$$

Under the second alternative sampling model $f_2\left(x|\theta\right)$

Similarly, $\inf_{\delta} r\left(\pi^M,\delta\right) = \sup_{\pi}\inf_{\delta} r\left(\pi,\delta\right) = 0.6884422$ and our least favorable prior are

$$\begin{array}{lll}
\pi_{7732}^M\left(\theta_1\right) = 0.2160804 & \pi_{7732}^M\left(\theta_2\right) = 0.17085427 & \pi_{7732}^M\left(\theta_3\right) = 0.6130653 \\
\pi_{8502}^M\left(\theta_1\right) = 0.2412060 & \pi_{8502}^M\left(\theta_2\right) = 0.14572864 & \pi_{8502}^M\left(\theta_3\right) = 0.6130653 \\
\pi_{9247}^M\left(\theta_1\right) = 0.2663317 & \pi_{9247}^M\left(\theta_2\right) = 0.12060302 & \pi_{9247}^M\left(\theta_3\right) = 0.6130653 \\
\pi_{9967}^M\left(\theta_1\right) = 0.2914573 & \pi_{9967}^M\left(\theta_2\right) = 0.09547739 & \pi_{9967}^M\left(\theta_3\right) = 0.6130653 \\
\pi_{10662}^M\left(\theta_1\right) = 0.3165829 & \pi_{10662}^M\left(\theta_2\right) = 0.07035176 & \pi_{10662}^M\left(\theta_3\right) = 0.6130653 \\
\pi_{11332}^M\left(\theta_1\right) = 0.3417085 & \pi_{11332}^M\left(\theta_2\right) = 0.04522613 & \pi_{11332}^M\left(\theta_3\right) = 0.6130653 \\
\pi_{11977}^M\left(\theta_1\right) = 0.3668342 & \pi_{11977}^M\left(\theta_2\right) = 0.02010050 & \pi_{11977}^M\left(\theta_3\right) = 0.6130653
\end{array}$$

It doesn't bother me that the prior depends on the sampling model. Least favorable priors are different under different sampling models since they are pessimistic in that they favor the worst outcome for that model.

2. (Problem 7.6) In order to choose an action $a$ based on the loss $L\left(\theta,a\right)$, you elicit the opinion of two experts about the probabilities of the various outcomes of $\theta$ (you can assume if you wish that $\theta$ is discrete random variable). The two experts give you distributions $\pi_1$ and $\pi_2$. Suppose that the action $a^*$is Bayes for both distributions $\pi_1$ and $\pi_2$. Is it true that $a^*$must be Bayes for all weighted averages of $\pi_1$and $\pi_2$; that is, for all the distributions of the form $\alpha\pi_1 + (1-\alpha)\pi_2$, with $0 < \alpha < 1$?

   **Solution:**

   Suppose that the action $a^*$is Bayes for both distributions $\pi_1$ and $\pi_2$, thus we have

$$a^* = \arg\min\ r\left(\pi_1,a\right) = \arg\min\ r\left(\pi_2,a\right)$$
$$r\left(\pi_1,a^*\right) = \inf_{a} r\left(\pi_1,a\right)$$
$$r\left(\pi_2,a^*\right) = \inf_{a} r\left(\pi_2,a\right)$$

For $0 < \alpha < 1$

$$
\begin{aligned}
\inf_a \left[ r\left( \alpha\pi_1 + (1-\alpha)\pi_2,\, a \right) \right] &= \inf_a \left[ \int_\Theta R\left(\theta, a\right) \cdot \left( \alpha\pi_1 + (1-\alpha)\pi_2 \right) d\theta \right] \\
&= \inf_a \left[ \int_\Theta R\left(\theta, a\right) \alpha\pi_1 d\theta + \int_\Theta R\left(\theta, a\right) (1-\alpha)\pi_2 d\theta \right] \\
&\geq \alpha \cdot \inf_a \left[ \int_\Theta R\left(\theta, a\right) \pi_1 d\theta \right] + (1-\alpha) \inf_a \left[ \int_\Theta R\left(\theta, a\right) \pi_2 d\theta \right] \\
&= \alpha \cdot \int_\Theta R\left(\theta, a^*\right) \pi_1 d\theta + (1-\alpha) \int_\Theta R\left(\theta, a^*\right) \pi_2 d\theta \\
&= \int_\Theta R\left(\theta, a^*\right) \cdot \left( \alpha\pi_1 + (1-\alpha)\pi_2 \right) d\theta \\
&= r\left( \alpha\pi_1 + (1-\alpha)\pi_2,\, a^* \right)
\end{aligned}
$$

Hence $a^*$ is Bayes for $\alpha\pi_1 + (1-\alpha)\pi_2$ .

3. (Problem 7.10) Consider a point estimation problem in which you observe $x_1, ..., x_n$ as iid random variables from the Poisson Distribution

$$
f\left(x|\theta\right) = \frac{1}{x!}\theta^x e^{-\theta}
$$

Assume a squared error of estimation loss $L\left(\theta, a\right) = \left(a - \theta\right)^2$, and assume a prior distribution on $\theta$ given by the gamma density

$$
\pi\left(\theta\right) = \frac{1}{\Gamma\left(\alpha\right)\beta^\alpha}\theta^{\alpha-1} \exp\left\{ -\frac{\theta}{\beta} \right\}
$$

(a) Show that the Bayes decision rule with respect tp the prior above is of the form

$$
\delta^*\left(x_1, ..., x_n\right) = a + b\bar{x}
$$

where $a > 0$, $b \in (0, 1)$, and $\bar{x} = \frac{1}{n}\sum_i x_i$. You may use the fact that the distribution of $\sum_i x_i$ is Poisson with parameter $n\theta$ without proof.

**Solution:**

For a quadratic loss function the optimal Bayes is the posterior mean

$$
\begin{aligned}
f\left(x|\theta\right) &= \prod \frac{1}{x_i!}\theta^{x_i} e^{-\theta} \\
&\propto \theta^{\sum x_i} \exp\left\{ -n\theta \right\}
\end{aligned}
$$

$$
\pi\left(\theta\right) = \frac{1}{\Gamma\left(\alpha\right)}\theta^{\alpha-1} \exp\left\{ -\frac{\theta}{\beta} \right\}
$$

$$
\begin{aligned}
\implies \pi\left(\theta|x\right) &\propto f\left(x|\theta\right)\pi\left(\theta\right) \\
&\propto \theta^{\left(\alpha + \sum x_i\right) - 1} \exp\left\{ -\theta(n + \beta^{-1}) \right\} \\
&\sim Ga\left( \alpha + \sum x_i,\, n + \beta^{-1} \right)
\end{aligned}
$$

Thus the posterior mean is

$$E\left(\theta|x\right)=\frac{\alpha+\sum x_i}{n+\beta^{-1}}=\frac{\alpha+\bar{x}}{n+\beta^{-1}}=\frac{\alpha}{n+\beta^{-1}}+\frac{n}{n+\beta^{-1}}\bar{x}=a+b\bar{x}$$

Then Bayes rules is $\delta(x)=a+b\bar{x}$, where $a=\frac{\alpha}{n+\beta^{-1}}$ and $b=\frac{n}{n+\beta^{-1}}$.

(b) Compute and graph the risk functions of $\delta^*(x_1,...,x_n)$ and that of the MLE $\delta(x_1,...,x_n)=\bar{x}$

**Solution:**

Risk function of the MLE $\delta(x)=\bar{x}$

$$\begin{aligned}
R\left(\theta,\bar{x}\right)&=E_X\left[L\left(\theta,\bar{x}\right)\right]\\
&=E_X\left[\left(\bar{x}-\theta\right)^2\right]\\
&=E_X\left[\bar{x}^2-2\theta\bar{x}+\theta^2\right]\\
&=E_X\left(\bar{x}\right)-2\theta E_X\left(\bar{x}\right)+\theta^2
\end{aligned}$$

We know $\sum x_i\sim Pois\left(n\theta\right)$, so

$$E\left(\bar{x}\right)=\frac{1}{n}\cdot n\theta=\theta$$

$$Var(\bar{x})=\frac{1}{n^2}\cdot n\theta=\frac{\theta}{n}$$

$$\begin{aligned}
\implies E\left(\bar{x}^2\right)&=Var\left(\bar{x}\right)+E\left(\bar{x}\right)^2\\
&=\frac{\theta}{n}+\theta^2
\end{aligned}$$

Then

$$\begin{aligned}
R\left(\theta,\bar{x}\right)&=E_X\left(\bar{x}\right)-2\theta E_X\left(\bar{x}\right)+\theta^2\\
&=\theta^2+\frac{\theta}{n}-2\theta^2+\theta^2\\
&=\frac{\theta}{n}
\end{aligned}$$

Risk Function of $\delta^*(x) = a + b\bar{x}$

$$\begin{aligned}
R(\theta, \delta^*) &= E_X\left[(\delta^* - \theta)^2\right] \\
&= E_X\left[\theta^2 - 2\theta(a + b\bar{x}) + (a + b\bar{x})^2\right] \\
&= E_X\left[\theta^2 - 2\theta a - 2\theta b\bar{x} + a^2 + 2ab\bar{x} + b^2\bar{x}^2\right] \\
&= \theta^2 - 2\theta a - 2\theta bE(\bar{x}) + a^2 + 2abE(\bar{x}) + b^2E\left(\bar{x}^2\right) \\
&= \theta^2 - 2\theta a - 2\theta^2 b + a^2 + 2ab\theta + b^2\left(\frac{\theta}{n} + \theta^2\right) \\
&= \theta^2\left(1 - 2b + b^2\right) + \theta\left(-2a + 2ab + \frac{b^2}{n}\right) + a^2 \\
&= \theta^2(1 - b)^2 + \theta\left(\frac{b^2}{n} - 2a + 2ab\right) + a^2
\end{aligned}$$

The following Figure are the graphs of both risk functions

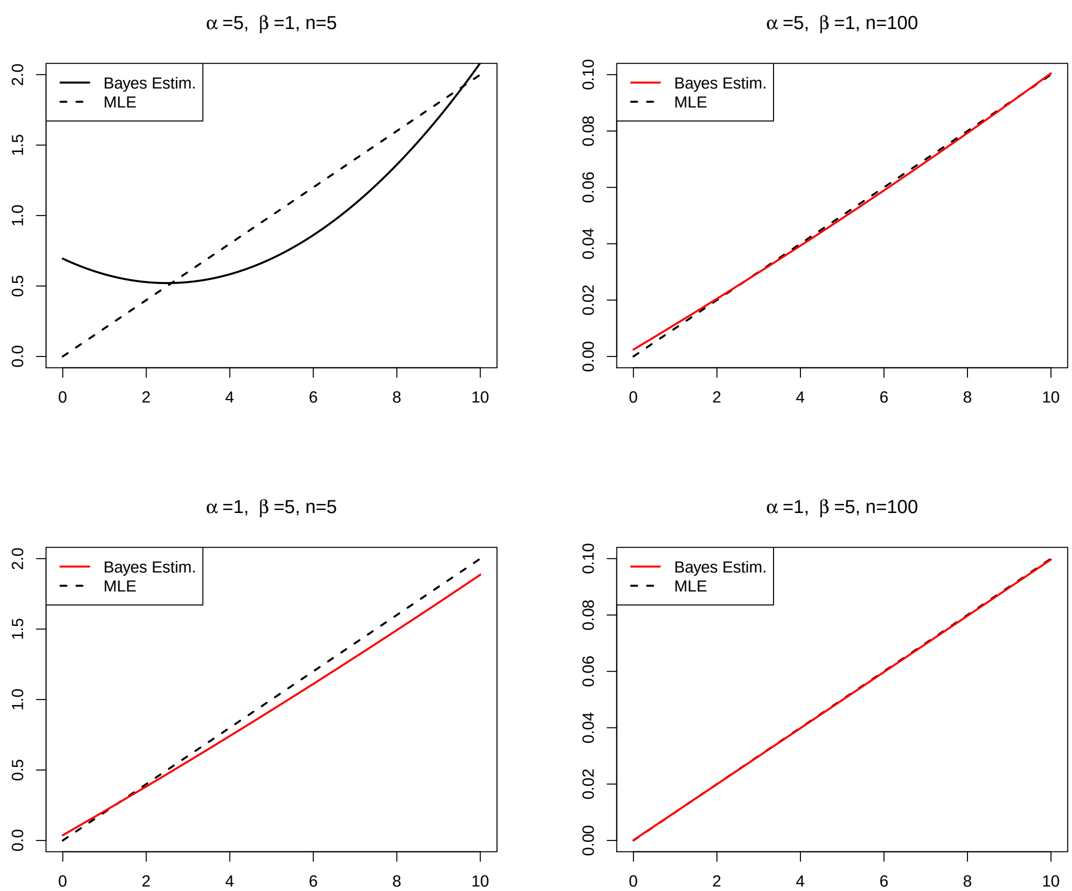


Figure 1:

(c) Compute the Bayes risk of $\delta^*(x_1, ..., x_n)$ and show that it is (a) decreasing in $n$ and (b) it goes to 0 as $n$ gets large.

**Solution:**

Note that $E(\theta) = \alpha\beta$, $Var(\theta) = \alpha\beta^2$, and $E(\theta^2) = \alpha\beta^2 + \alpha^2\beta^2$. Then the Bayes risk of $\delta^*$is

$$\begin{aligned}
r(\pi, \delta^*) &= \int R(\theta, \delta^*)\, \pi(\theta)\, d\theta \\
&= E_\theta [R(\theta, \delta^*)] \\
&= E[\theta^2](1-b)^2 + E[\theta]\left(\frac{b^2}{n} - 2a + 2ab\right) + a^2 \\
&= (\alpha\beta^2 + \alpha^2\beta^2)\left(1 - \frac{n}{n+\beta^{-1}}\right)^2 + (\alpha\beta)\left[\frac{\left(\frac{n}{n+\beta^{-1}}\right)^2}{n} - \frac{2\alpha}{n+\beta^{-1}} + \frac{2\alpha n}{(n+\beta^{-1})^2}\right] + \left(\frac{\alpha}{n+\beta^{-1}}\right)^2 \\
&= (\alpha\beta^2 + \alpha^2\beta^2)\left(\frac{n+\beta^{-1}-n}{n+\beta^{-1}}\right)^2 + \frac{[\alpha\beta n - 2\alpha^2\beta(n+\beta^{-1}) + 2\alpha^2\beta n + \alpha^2]}{(n+\beta^{-1})^2} \\
&= \frac{[\alpha + \alpha^2 + \alpha\beta n - 2\alpha^2\beta n - 2\alpha^2 + 2\alpha^2\beta n + \alpha^2]}{(n+\beta^{-1})^2} \\
&= \frac{[\alpha + \alpha\beta n]}{(n+\beta^{-1})^2} \\
&= \frac{\alpha\beta(n+\beta^{-1})}{(n+\beta^{-1})^2} \\
&= \frac{\alpha\beta}{(n+\beta^{-1})^2}
\end{aligned}$$

Hence the Bayes risk of $(x_1, ..., x_n)$ is decreasing in n and it goes to 0 as n gets large.

(d) Suppose an investigator wants to collect a sample that is large enough that the Bayess risk after the experiment is half of the Bayes risk before the experiment. Find that sample size.

**Solution:**

The Bayes risk before the experiment under squared error loss is the prior variance,that is,

$$r_{\text{before}}(\pi, \delta^*) = \alpha\beta^2$$

We want to find $n$ such that

$$\frac{\alpha\beta}{(n+\beta^{-1})} < \frac{\alpha\beta^2}{2}$$

Now we solve for $n$ and get $n > \beta^{-1}$

4. (Problem 7.11) Take $x$ to be binomial with unknown $\theta$ and $n = 2$, and consider testing $\theta_0 = 1/3$, versus $\theta_1 = 1/2$. Draw the points corresponding to every possible decision rule in the risk space with coordinates $R(\theta_0, \delta)$ and $R(\theta_1, \delta)$. Identify minimax randomized and nonrandomized rules and the expected utility rule for prior $\pi(\theta_0) = 1/4$ . Identify the set of rules that are not dominated by any other rule (those are called admissible in Ch.8).

**Solution:**

Our hypothesis test is

$$H_0 : \theta_0 = 1/3 \qquad H : \theta_1 = 1/2$$

Our loss function is

| | $\theta=\theta_0$ | $\theta=\theta_1$ |
|---|---|---|
| $a_0$ | $0$ | $L_0$ |
| $a_1$ | $L_1$ | $0$ |

Where $L_1$ and $L_2$ are positive numbers. Then, in general (on pg. 134), our risk function is

$$R(\theta,\delta)=\begin{cases} L_1 P(\delta(x)=a_1|\theta) & \text{if } \theta=\theta_0=1/3 \\ L_0 P(\delta(x)=a_0|\theta) & \text{if } \theta=\theta_1=1/2 \end{cases}$$

Take $x$ to be binomial with unknown $\theta$ and $n=2$. Then for each $x=0,1,2$ there are two possible actions $a_0$ and $a_1$. Thus, in total, we have 8 possible decision rules $\delta_i(x_1,x_2,x_3)$: $(\delta_1,...,\delta_8)=(a_0,a_0,a_0)$, $(a_0,a_0,a_1)$, $(a_0,a_1,a_0)$, $(a_1,a_0,a_0)$, $(a_1,a_1,a_0)$, $(a_1,a_0,a_1)$, $(a_0,a_1,a_1)$, $(a_1,a_1,a_1)$.

For $\theta=\theta_0=1/3$:

$$\begin{aligned}
R(\theta_0,(a_0,a_0,a_0)) &= L(\theta_0,a_0) f(x=0|\theta_0)+L(\theta_0,a_0) f(x=1|\theta_0)+L(\theta_0,a_0) f(x=2|\theta_0)=0 \\
R(\theta_0,(a_0,a_0,a_1)) &= 0+0+L(\theta_0,a_1) f(x=2|\theta_0)=L_1\cdot\binom{2}{2}\theta_0^2(1-\theta_0)^{2-2}=\frac{L_1}{9} \\
R(\theta_0,(a_0,a_1,a_0)) &= 0+L(\theta_0,a_1) f(x=1|\theta_0)+0=\frac{4L_1}{9} \\
R(\theta_0,(a_1,a_0,a_0)) &= \frac{4L_1}{9}+0+0=\frac{4L_1}{9} \\
R(\theta_0,(a_1,a_1,a_0)) &= \frac{4L_1}{9}+\frac{4L_1}{9}+0=\frac{8L_1}{9} \\
R(\theta_0,(a_1,a_0,a_1)) &= \frac{4L_1}{9}+0+\frac{L_1}{9}=\frac{5L_1}{9} \\
R(\theta_0,(a_0,a_1,a_1)) &= 0+\frac{4L_1}{9}+\frac{L_1}{9}=\frac{5L_1}{9} \\
R(\theta_0,(a_1,a_1,a_1)) &= \frac{4L_1}{9}+\frac{4L_1}{9}+\frac{L_1}{9}=L_1
\end{aligned}$$

Similarly, for $\theta=\theta_0=1/2$

$$\begin{aligned}
R(\theta_1,(a_0,a_0,a_0)) &= \frac{L_0}{4}+\frac{2L_0}{4}+\frac{L_0}{4}=L_0 \\
R(\theta_1,(a_0,a_0,a_1)) &= \frac{L_0}{4}+\frac{2L_0}{4}+0=\frac{3L_0}{4} \\
R(\theta_1,(a_0,a_1,a_0)) &= \frac{L_0}{4}+0+\frac{L_0}{4}=\frac{2L_0}{4} \\
R(\theta_1,(a_1,a_0,a_0)) &= 0+\frac{2L_0}{4}+\frac{L_0}{4}=\frac{3L_0}{4} \\
R(\theta_1,(a_1,a_1,a_0)) &= 0+0+\frac{L_0}{4}=\frac{L_0}{4} \\
R(\theta_1,(a_1,a_0,a_1)) &= 0+\frac{2L_0}{4}+0=\frac{2L_0}{4} \\
R(\theta_1,(a_0,a_1,a_1)) &= \frac{L_0}{4}+0+0=\frac{L_0}{4} \\
R(\theta_1,(a_1,a_1,a_1)) &= 0
\end{aligned}$$

Then we can calculate the bayes risk. Given prior $\pi(\theta_0) = 1/4$ implies $\pi(\theta_1) = 3/4$, the 8 Bayes risks are

$$
\begin{aligned}
r(\pi, \delta_1) &= R(\theta_0, \delta_1)\pi(\theta_0) + R(\theta_1, \delta_1)\pi(\theta_1) = 0 + \frac{3L_0}{4} = \frac{3L_0}{4} \\
r(\pi, \delta_2) &= R(\theta_0, \delta_2)\pi(\theta_0) + R(\theta_1, \delta_2)\pi(\theta_1) = \frac{L_1}{9}\left(\frac{1}{4}\right) + \frac{3L_0}{4}\left(\frac{3}{4}\right) = \frac{L_1}{36} + \frac{9L_0}{16} \\
r(\pi, \delta_3) &= \frac{4L_1}{9}\left(\frac{1}{4}\right) + \frac{2L_0}{4}\left(\frac{3}{4}\right) = \frac{4L_1}{36} + \frac{6L_0}{16} \\
r(\pi, \delta_4) &= \frac{4L_1}{9}\left(\frac{1}{4}\right) + \frac{3L_0}{4}\left(\frac{3}{4}\right) = \frac{4L_1}{36} + \frac{9L_0}{16} \\
r(\pi, \delta_5) &= \frac{8L_1}{9}\left(\frac{1}{4}\right) + \frac{L_0}{4}\left(\frac{3}{4}\right) = \frac{8L_1}{36} + \frac{3L_0}{16} \\
r(\pi, \delta_6) &= \frac{5L_1}{9}\left(\frac{1}{4}\right) + \frac{2L_0}{4}\left(\frac{3}{4}\right) = \frac{5L_1}{36} + \frac{6L_0}{16} \\
r(\pi, \delta_7) &= \frac{5L_1}{9}\left(\frac{1}{4}\right) + \frac{L_0}{4}\left(\frac{3}{4}\right) = \frac{5L_1}{36} + \frac{3L_0}{16} \\
r(\pi, \delta_8) &= \frac{L_1}{4}
\end{aligned}
$$

For simplicity, suppose $L_0 = L_1 = 1$. Then Bayes risks are

$$
\begin{aligned}
r(\pi, \delta_1) &= \frac{3}{4} = 0.75 \\
r(\pi, \delta_2) &= \frac{1}{36} + \frac{9}{16} = 0.5903 \\
r(\pi, \delta_3) &= \frac{4}{36} + \frac{6}{16} = 0.4861 \\
r(\pi, \delta_4) &= \frac{4}{36} + \frac{9}{16} = 0.6736 \\
r(\pi, \delta_5) &= \frac{8}{36} + \frac{3}{16} = 0.4097 \\
r(\pi, \delta_6) &= \frac{5}{36} + \frac{6}{16} = 0.5139 \\
r(\pi, \delta_7) &= \frac{5}{36} + \frac{3}{16} = 0.3264 \\
r(\pi, \delta_8) &= \frac{1}{4} = 0.25
\end{aligned}
$$

Thus $r(\pi, \delta^*) = \inf_{\delta} r(\pi, \delta) = r(\pi, \delta_8) = 0.25$ which implies that the Bayesian decision rule is $\delta^*(x) = \delta_8(x)$ which takes action $a_1$ if $x = 0, 1, 2$. The dashed line in Figure blah is the equation

$$
\begin{aligned}
k &= \frac{1}{4}R(\theta_0, \delta_8) + \frac{3}{4}R(\theta_1, \delta_8) \\
\implies R(\theta_1, \delta_8) &= -\frac{1}{3}R(\theta_0, \delta_8) + \frac{4}{3}k
\end{aligned}
$$

where $k = 0.25$, the minimum in correspondence with $\delta_8$.
Next we want to find the non randomized minimax rule. By definition, a decision rule say $y_1$ is preferred to a rule $y_2$ if

$$
\sup_{\theta} R(\theta, y_1) < \sup_{\theta} R(\theta, y_1)
$$

Thus we have

$$
\begin{array}{lll}
R(\theta_0,\delta_1)\approx 0.00; & R(\theta_1,\delta_1)\approx 1.00 & \sup_\theta R(\theta,\delta_1)=1.00\\
R(\theta_0,\delta_2)\approx 0.11; & R(\theta_1,\delta_2)\approx 0.75 & \sup_\theta R(\theta,\delta_2)=0.75\\
R(\theta_0,\delta_3)\approx 0.44; & R(\theta_1,\delta_3)\approx 0.50 & \sup_\theta R(\theta,\delta_3)=0.50\\
R(\theta_0,\delta_4)\approx 0.44; & R(\theta_1,\delta_4)\approx 0.75 & \sup_\theta R(\theta,\delta_4)=0.75\\
R(\theta_0,\delta_5)\approx 0.89; & R(\theta_1,\delta_5)\approx 0.25 & \sup_\theta R(\theta,\delta_5)=0.89\\
R(\theta_0,\delta_6)\approx 0.56; & R(\theta_1,\delta_6)\approx 0.50 & \sup_\theta R(\theta,\delta_6)=0.56\\
R(\theta_0,\delta_7)\approx 0.56; & R(\theta_1,\delta_7)\approx 0.25 & \sup_\theta R(\theta,\delta_7)=0.56\\
R(\theta_0,\delta_8)\approx 1.00; & R(\theta_1,\delta_8)\approx 0.00 & \sup_\theta R(\theta,\delta_8)=1.00
\end{array}
$$

Then $\inf_\delta \sup_\theta R(\theta,\delta) = 0.50$ which implies that the non randomized minimax rule is $\delta^M(x) = \delta_3(x)$ which takes action $a_0$if $x = 0, 2$ and $a_1$ if $x = 1$.

Next we want to find a randomized minimax rule. To find a minimax randomized action we move the wedge of the indifference set until it contacts the point of intersection of the line segment joining points $\delta_2$,$\delta_7$ and the 45 degree line. The equation of the line that passing through points $\delta_2$ and $\delta_7$ is $R(\theta_1,\delta) = -9/8 * R(\theta_0,\delta) + 7/8$. Thus the point of intersection that we speak of is (0.411764705, 0.411764705). Thus the randomized minimax rule is $(1-p)\,\delta_2 + p\delta_7$ such that the corresponding risk is $(1-p)\cdot(1/9) + p\cdot(5/9) = 0.411764705$. Solving for $p$ we get that $p \approx 0.676471$. Hence the randomized minimax rule is $(0.323529)\delta_2 + (0.676471)\delta_7$.

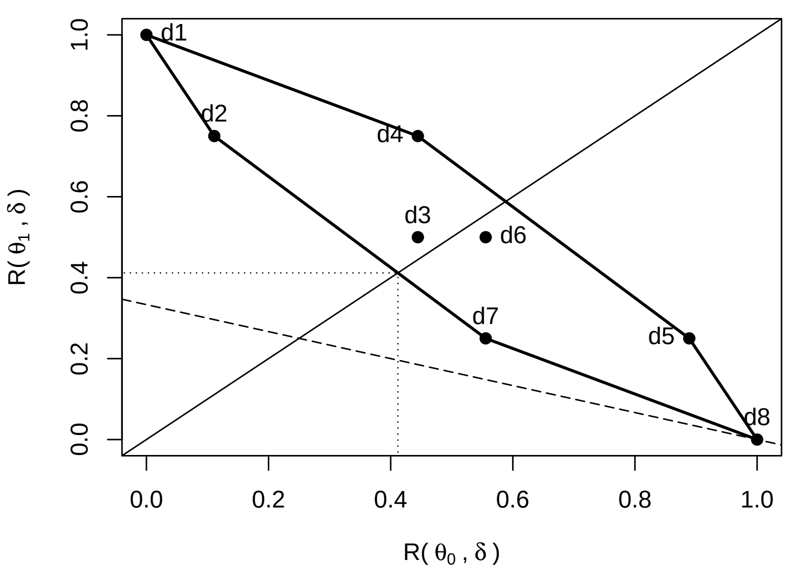


Figure 2:

Lastly, we will find the admissible set. By definition, a decision rule $y$ is R-better than $y'$ if

$$R(\theta, y) \leq R(\theta, y'), \ \forall\theta$$

and $R(\theta, y) < R(\theta, y')$ for some $\theta$. (i.e. $y$ dominates $y'$). A decision rule $y$ is "admissible" if no other rule dominates it. Thus, notice in our case that

$$\begin{array}{lll} R(\theta,\delta_2) \leq R(\theta,\delta_4)\ \forall\theta & \text{and} & R(\theta_0,\delta_2) < R(\theta_0,\delta_4) \\ R(\theta,\delta_3) \leq R(\theta,\delta_4)\ \forall\theta & \text{and} & R(\theta_1,\delta_3) < R(\theta_1,\delta_4) \\ R(\theta,\delta_3) \leq R(\theta,\delta_6)\ \forall\theta & \text{and} & R(\theta_0,\delta_3) < R(\theta_0,\delta_6) \\ R(\theta,\delta_7) \leq R(\theta,\delta_6)\ \forall\theta & \text{and} & R(\theta_1,\delta_7) < R(\theta_1,\delta_6) \\ R(\theta,\delta_7) \leq R(\theta,\delta_5)\ \forall\theta & \text{and} & R(\theta_0,\delta_7) < R(\theta_0,\delta_5) \end{array}$$

This implies that $\delta_4$and $\delta_6$ are inadmissible. The rest of the decision rules have no R-better rules. Thus the admissible set of decision rules is $\{\delta_1, \delta_2, \delta_3, \delta_7, \delta_8\}$.

5. (Problem 7.12) State a probabilistic model for the situation decribed by Pratt, and specify a prior distribution and a loss function for the point estimation of the mean voltage. Writing code if necessary, compute the risk fucntion R of the Bayes rule and that of your favorite frequentist rule in two scenarios: when the high range voltmeter is available and when it is not. Does examining the risk function help you select a decision rule once the data are observed.

**Solution:**

Based on Pratt's description when the voltmeter is available, the data follows a normal distribution. To create a general model we will assume a conjugate prior distribution instead of a non-informative prior distribution. (pp.67, Gelman 2004) Our model is

$$\begin{aligned} X_i &\sim N\left(\mu, \sigma^2\right) \\ \mu|\sigma^2 &\sim N\left(\mu_0, \frac{\sigma^2}{k_0}\right) \\ \sigma^2 &\sim Inv\chi^2\left(v_0, \sigma_0^2\right) \end{aligned}$$

and our posterior distribution is

$$\begin{aligned} \mu|x &\sim t_{v_n}\left(\mu_n, \sigma_n^2/k_n\right) \\ \text{where } \mu_n &= \frac{k_0}{k_0+n}\mu_0 + \frac{n}{k_0+n}\bar{x} \\ k_n &= k_0 + n \\ v_n &= v_0 + n \\ v_n\sigma_n^2 &= v_0\sigma_0^2 + (n-1)\,s^2 + \frac{k_0 n}{k_0+n}\left(\bar{x}-\mu_0\right)^2 \\ s^2 &= \frac{1}{n-1}\sum(x_i-\bar{x})^2 \end{aligned}$$

Suppose we have a quadratic loss function: $L(\mu, \delta) = (\mu - \delta)^2$. Then the optimal Bayes is the posterior mean, $\delta^B = \mu_n = \frac{\mu_0 k_0}{k_0+n} + \frac{n}{k_0+n}\bar{x} = a + b\bar{x}$. The frequentist decision rule is the mle of $x = (x_1, ..., x_n)$

which is $\delta_{mle} = \bar{x}$. Now we can calculate the risk function for the frequentist rule and bayes rule

$$\begin{aligned} R\left(\mu, \delta_{mle}\right) &= E_x\left[\left(\mu - \delta_{mle}\right)^2\right] \\ &= \frac{\mu}{n} \end{aligned}$$

$$\begin{aligned} R\left(\mu, \delta^B\right) &= E_x\left[\left(\mu - \delta^B\right)^2\right] \\ &= \mu^2\left(1-b\right)^2 + \mu\left(\frac{b^2}{n} - 2a + 2ab\right) + a^2 \end{aligned}$$

Now we will focus on the scenario when the high-range voltmeter is not available. The voltmeter used reads only as far as 100. That means all values above 100 are censored (which there are none) and the rest of the data is uncensored as if the high-range voltmeter is available. (pp.226, Gelman 2004) The complete-data likelihood is

$$p\left(x|\mu\right) = \prod_{i=1}^{n} N\left(x_i|\mu, 1\right),$$

and the likelihood of the inclusion vector, given the complete data, has i.i.d form:

$$\begin{aligned} p\left(I|x,\phi\right) &= \prod_{i=1}^{n} p\left(I_i|x_i,\phi\right) \\ &= \prod_{i=1}^{n} \begin{cases} 1 & \text{if } \left(I_i = 1 \text{ and } x_i \leq \phi\right) \text{ or } \left(I_i = 0 \text{ and } x_i > \phi\right) \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

For valid Bayesian inference we must condition on all observed data, which means we need the joint likelihood of $x_{\text{obs}}$ and $I$, which we obtain mathematically by integrating out $x_{\text{miss}}$ from the complete-data likelihood:

$$\begin{aligned} p\left(x_{\text{obs}}, I|\mu,\phi\right) &= \int p\left(x, I|\mu,\phi\right) dx_{\text{miss}} \\ &= \int p\left(x|\mu,\phi\right) \cdot p\left(I|x,\mu,\phi\right) dx_{\text{miss}} \\ &= \prod_{i:I_i=1} N\left(x_i|\mu,\sigma^2\right) \prod_{i:I_i=0} \int_{\phi}^{\infty} N\left(x_i|\mu,\sigma^2\right) p\left(I_i|x_i,\phi\right) dx_i \\ &= \prod_{i:I_i=1} N\left(x_i|\mu,\sigma^2\right) \prod_{i:I_i=0} \Phi\left(\frac{\mu-\phi}{\sigma}\right) \\ &= \left(\prod_{i=1}^{n} N\left(x_{\text{obs}}|\mu,\sigma^2\right)\right)\left[\Phi\left(\frac{\mu-\phi}{\sigma}\right)\right]^0 \\ &= \left(\prod_{i=1}^{n} N\left(x_{\text{obs}}|\mu,\sigma^2\right)\right) \\ &= N(\bar{x}, \sigma^2/n) \end{aligned}$$

Note that none of the observations were censored hence the term $\left[\Phi\left(\frac{\mu-\phi}{\sigma}\right)\right]^0$. Thus it is an uncersored data set as if the high-range voltmeter is available. By the conditionality principle, our bayes rule would still be the posterior mean $\delta^B = \mu_n$ and our frequentist rule would still be $\delta_{mle} = \bar{x}$ for $x_{\text{obs}}$. However for $x_{\text{miss}}$we will use $\phi = 100$ as our decision rule since no other information is given. Then our risk function in general is

$$\begin{aligned} R(\mu, \delta) &= \int_{-\infty}^{100} (\mu - \delta)^2 N(\bar{x}, \sigma^2/n)dx + \int_{100}^{\infty} (\mu - \phi)^2 N(\bar{x}, \sigma^2/n)dx \\ &= \int_{-\infty}^{100} (\mu - \delta)^2 N(\bar{x}, \sigma^2/n)dx + (\mu - 100)^2 \Phi\left(\frac{\mu - 100}{\sigma}\right)^n \end{aligned}$$

First we will compute the risk function of $\delta^B = \mu_n = \frac{\mu_0 k_0}{k_0+n} + \frac{n}{k_0+n}\bar{x} = a + b\bar{x}$:

$$\begin{aligned} R\left(\mu, \delta^B\right) &= \int_{-\infty}^{100} \left(\mu - \delta^B\right)^2 N(\bar{x}, \sigma^2/n)dx + (\mu - 100)^2 \Phi\left(\frac{\mu - 100}{\sigma}\right)^n \\ &= \int_{-\infty}^{100} (\mu - (a + b\bar{x}))^2 N(\bar{x}, \sigma^2/n)dx + (\mu - 100)^2 \Phi\left(\frac{\mu - 100}{\sigma}\right)^n \end{aligned}$$

Lets focus on the first term, for organization's sake, since the second term is a constant.

$$\begin{aligned} \int_{-\infty}^{100} (\mu - (a + b\bar{x}))^2 N(\bar{x}, \sigma^2/n)dx &= \int_{-\infty}^{100} (\mu - (a + b\bar{x}))^2 N(\bar{x}, \sigma^2/n)dx \\ &= \int_{-\infty}^{100} (\mu - (a + b\bar{x}))^2 N(\bar{x}, \sigma^2/n)dx \\ &= \int_{-\infty}^{100} \left(\mu^2 - 2\mu a - 2\mu b\bar{x} + a^2 + 2ab\bar{x} + b^2\bar{x}^2\right) N(\bar{x}, \sigma^2/n)dx \\ &= \left(\mu^2 - 2\mu a + a^2\right) \Phi\left(\frac{100 - \mu}{\sigma}\right)^n \\ &\quad + b^2 \int_{-\infty}^{100} \bar{x}^2 N(\bar{x}, \sigma^2/n)dx + 2(ab - \mu b) \int_{-\infty}^{100} \bar{x} N(\bar{x}, \sigma^2/n)dx \end{aligned}$$

Let $\mu^*$ and $\sigma_*^2$ be the mean and variance for each truncated normal x, $x_i \sim TN\left(\mu, \sigma^2\right)$ for $-\infty < x_i < 100$. Then

$$\begin{aligned} \int_{-\infty}^{100} (\mu - (a + b\bar{x}))^2 N(\bar{x}, \sigma^2/n)dx &= \left(\mu^2 - 2\mu a + a^2\right) \Phi\left(\frac{100 - \mu}{\sigma}\right)^n + b^2 E\left[\bar{x}^2\right] + 2ab\mu^* - 2\mu b\mu^* \\ &= \left(\mu^2 - 2\mu a + a^2\right) \Phi(z)^n + b^2\left(\frac{\sigma_*^2}{n} + \mu^*\right) + 2ab\mu^* - 2\mu b\mu^* \\ &= (\mu - a)^2 \Phi(z)^n + b^2\left(\frac{\sigma_*^2}{n} + \mu^*\right) + 2ab\mu^* - 2\mu b\mu^* \end{aligned}$$

Then the risk function for the bayes rule is

$$R\left(\mu, \delta^B\right) = (\mu - a)^2 \Phi(z)^n + b^2\left(\frac{\sigma_*^2}{n} + \mu^*\right) + 2ab\mu^* - 2\mu b\mu^* + (\mu - 100)^2 \Phi\left(\frac{\mu - 100}{\sigma}\right)^n$$

where

$$z = \frac{100 - \mu}{\sigma}$$
$$\mu^* = \mu - \sigma \frac{\psi(z)}{\Phi(z)}$$
$$\sigma_*^2 = \sigma^2 \left[ 1 - z \frac{\psi(z)}{\Phi(z)} - \left( \frac{\psi(z)}{\Phi(z)} \right)^2 \right]$$

Note that $\psi(\cdot)$ is the probability density function of the standard normal distribution and $\Phi(\cdot)$ is its cumulative distribution function.

Next we compute the risk function for $\delta_{mle} = \bar{x}$

$$\begin{aligned}
R(\mu, \delta_{mle}) &= \int_{-\infty}^{100} (\mu - \bar{x})^2 N(\bar{x}, \sigma^2/n) dx + (\mu - 100)^2 \Phi \left( \frac{\mu - 100}{\sigma} \right)^n \\
&= \int_{-\infty}^{100} \left( \mu^2 - 2\mu\bar{x} + \bar{x}^2 \right) N(\bar{x}, \sigma^2/n) dx + (\mu - 100)^2 \Phi \left( \frac{\mu - 100}{\sigma} \right)^n \\
&= \mu^2 \Phi(z) - 2\mu \int_{-\infty}^{100} \bar{x} N(\bar{x}, \sigma^2/n) dx + \int_{-\infty}^{100} \bar{x}^2 N(\bar{x}, \sigma^2/n) dx + (\mu - 100)^2 \Phi \left( \frac{\mu - 100}{\sigma} \right)^n \\
&= \mu^2 \Phi(z) - 2\mu\mu^* + \left( \frac{\sigma_*^2}{n} + \mu^* \right) + (\mu - 100)^2 \Phi \left( \frac{\mu - 100}{\sigma} \right)^n
\end{aligned}$$

Would I prefer the bayes rule or frequentist rule? I would prefer Bayes rule since it incorporates prior information and doesn't change if the high voltmeter is available or not.