
Remarks:
1. Each problem is worth 10 points.
2. If you have any questions about the solutions, please contact haiyan@stat.stanford.edu for problems 5, 6 and 12; dsmall@stat.stanford.edu or johanlim@stat.stanford.edu for problems 14, 17, 21, 28 and A.

5
(a)
\[ R(\theta, d_1) = -2(1 - 2\theta(1 - \theta)) + 2\theta(1 - \theta) = -6\theta^2 + 6\theta - 2 \]
\[ R(\theta, d_2) = 6\theta^2 - 6\theta + 1 \]
\[ R(\theta, d_3) = 0 \]

(b) Since \( R(\theta, d_1) \leq -1/2 \) and \( R(\theta, d_2) \leq 1 \), \( d_1 \) is the minimax rule.

(c) Using a uniform prior, we have \( E(R(\theta, d_1)) = -1 \) and \( E(R(\theta, d_2)) = 0 \), so \( d_1 \) is the Bayes rule.

(d)
\[ R(\theta, d_1) = -D(1 - 2\theta(1 - \theta)) + 2\theta(1 - \theta) = -(2D + 2)\theta^2 + (2D + 2)\theta - D \leq -\frac{D}{2} + \frac{1}{2} \]
\[ R(\theta, d_2) = (2D + 2)\theta^2 - (2D + 2)\theta + 1 \leq 1 \]
\[ R(\theta, d_3) = 0 \]

So \( d_1 \) is minimax when \( D > 1 \), and \( d_3 \) is minimax when \( D < 1 \). Using a uniform prior we have \( R(\theta, d_1) = 1/3 - (2D/3) \) and \( R(\theta, d_2) = 2/3 - D/3 \), so \( d_1 \) is the Bayes rule when \( D > 1/2 \) and \( d_3 \) is the Bayes rule when \( D < 1/2 \).

6
(a)
Bayes rule is
\[ d^*(x) = \begin{cases} A & \text{if } \frac{f(x|A)}{f(x|B)} > \frac{\omega_A \pi_A}{\omega_B \pi_B} \\ B & \text{otherwise} \end{cases} \]
where \( \omega_A = l(A, d(x) = B) \); and \( \omega_B = l(B, d(x) = A) \). Here we use 0-1 loss, so \( \frac{\omega_A \pi_A}{\omega_B \pi_B} = 1 \). Since
\[ \frac{f(x|A)}{f(x|B)} = \frac{\exp(-\frac{-2x}{2})}{\exp(-\frac{-2(x-1)}{2})} = \exp(-2(x-1)), \]
we obtain \( d^*(x) = A \) if \( x < 1 \), and \( d^*(x) = B \) if \( x > 1 \).

(b)
We classify \( x = 1.5 \) to class B. \( P(error) = P(A|x = 1.5) = 0.27 \)

(c)
\[ P(error) = 0.5P(x > 1|A) + 0.5P(x < 1|B) = 0.1587 \]

(d)
Now \( \omega_A = 2\omega_B \), so \( \frac{\omega_A \pi_A}{\omega_B \pi_B} = 2 \). Thus \( d^*(x) = A \) if \( \frac{f(x|A)}{f(x|B)} = \exp(-2(x-1)) > 2 \), that is, \( d^*(x) = A \) if \( x < 1.35 \) and \( d^*(x) = B \) if \( x > 1.35 \).
we also classify \( x = 1.5 \) to class B and \( p(\text{error}) = 0.27 \). For the Bayes rule, \( p(\text{error}) = 0.5P(x > 1.35|A) + 0.5P(x < 1.35|B) = 0.17 \)

For \( \pi_A = 2/3 \), we have the same Bayes rule, that is, \( d^*(x) = A \) if \( x < 1.35 \) and \( d^*(x) = B \) if \( x > 1.35 \).

12
We first show that \( X_{(1)} \) follows exponential \((\theta/n)\), because

\[
P(X_{(1)} > a) = \prod_{i=1}^{n} P(X_i > a) = \exp\left(-\frac{n}{\theta}a\right)
\]

Thus we obtain

\[
R(\theta, d_1) = E(\bar{X} - \theta)^2 = \text{Var}(\bar{X}) = \frac{\theta^2}{n}
\]

\[
R(\theta, d_1) = E(nX_{(1)} - \theta)^2 = n^2 E(X_{(1)} - \theta/n)^2 = n^2 \text{var}(X_{(1)}) = n^2\frac{\theta^2}{n^2} = \theta^2
\]

\[
R(\theta, d_3) = E\left(\frac{n}{n+1}\bar{X} - \theta\right)^2 = E\left(\frac{n}{n+1}\bar{X} - \frac{n}{n+1}\theta - \frac{\theta}{n+1}\right)^2 = \text{var}\left(\frac{n}{n+1}\bar{X}\right) + E\left(\frac{\theta^2}{(n+1)^2}\right)
\]

\[
= \left(\frac{n}{n+1}\right)^2\frac{\theta^2}{n} + \frac{\theta^2}{(n+1)^2} = \frac{\theta^2}{n+1}
\]

Thus \( d_3 \) dominates \( d_1 \) and \( d_2 \).

14
(a)
The posterior distribution of \( p \) is

\[
h(p|x) = \frac{f(x)p^f(p)}{f(x)} = \frac{(1-p)^{x-1}p}{\int_0^1 (1-p)^{x-1}pdp}
\]

\[
= \frac{(1-p)^{x-1}p}{\frac{\Gamma(2)\Gamma(x)}{\Gamma(x+2)}} = x(x+1)p(1-p)^{x-1}
\]

The third line uses the beta integral on page 594. We conclude that the posterior distribution of \( p \) is beta\((2, x)\).

(b)
The Bayes estimate of \( p \) under squared error loss is the posterior mean. The Bayes estimate of \( p \) is thus \( \frac{2}{x+2} \) since the posterior distribution of \( p \) is beta\((2, x)\). The posterior mean can be derived directly -

\[
\int_0^1 px(x+1)p(1-p)^{x-1}dp = x(x+1)\int_0^1 p^2(1-p)^{x-1}dp
\]

\[
= x(x+1)\frac{\Gamma(3)\Gamma(x)}{\Gamma(3+x)}
\]

\[
= \frac{2}{x+2}
\]
(c) The log likelihood of \( p \) is
\[
   l(p|x) = (x - 1) \log(1 - p) + \log p
\]
Thus,
\[
   l'(p|x) = \frac{1 - x}{1 - p} + \frac{1}{p}
\]
Setting the above equation to 0 and solving for \( p \) gives \( p = \frac{1}{x} \). We have \( l''(p|x) = \frac{1-x}{(1-p)^2} - \frac{1}{p^2} < 0 \). Thus, the MLE is \( \hat{p}_{MLE} = \frac{1}{x} \).

17
The posterior probabilities of \( H \) and \( K \) are
\[
   P(H|X) = \frac{f(x|H)P(H)}{f(x|H)P(H) + f(x|K)P(K)}
\]
\[
   P(K|X) = \frac{f(x|K)P(K)}{f(x|H)P(H) + f(x|K)P(K)}
\]
Thus, the posterior odds of \( H \) to \( K \) are
\[
   \frac{P(H|X)}{P(K|X)} = \frac{\frac{f(x|H)P(H)}{f(x|H)P(H) + f(x|K)P(K)}}{\frac{f(x|K)P(K)}{f(x|H)P(H) + f(x|K)P(K)}}
   = \frac{P(H)}{P(K)} \frac{f(x|H)}{f(x|K)}
\]
Thus, the posterior odds of \( H \) to \( K \) are equal to the prior odds multiplied by the likelihood ratio.

21
Let \( Y \) be distributed according to the posterior distribution of \( \theta|X = x \). Suppose that \( m \) is a median of the posterior distribution if \( m_0 \leq m \leq m_1 \) and suppose \( c > m_1 \). We will show that \( E(|Y - m|) < E(|Y - (m + c)|) \). We have
\[
   E(|Y - m|) - E(|Y - (m + c)|) = \int_{-\infty}^{m} -cf(y)dy + \int_{m}^{m+c} [(y - m) - (c + m - y)]f(y)dy + \int_{m+c}^{\infty} cf(y)dy
\]
\[
   = \int_{-\infty}^{m} -cf(y)dy + \int_{m}^{m+c} (y - m) - (c + m - y) - cf(y)dy
   + \int_{m}^{m+c} cf(y)dy + \int_{m+c}^{\infty} cf(y)dy
\]
\[
   = \int_{-\infty}^{m} -cf(y)dy + \int_{m}^{\infty} cf(y)dy + \int_{m}^{m+c} (2y - 2m - 2c)f(y)dy
   = \int_{m}^{m+c} (2y - 2m - 2c)f(y)dy
   > 0
\]
In the second and third lines, we add and subtract the term \( \int_{m}^{m+c} cf(y)dy \). In the fifth line, we use the fact that \( m \) is a median and hence \( \int_{-\infty}^{m} -cf(y)dy + \int_{m}^{\infty} cf(y)dy = 0 \). We can use a similar
argument to show that if $c < m_0$, we have $E(|Y - m|) - E(|Y - (m + c)|) < 0$. If $m_0 < m_2 < m_1$, we have

$$E(|Y - m|) - E(|Y - m_2|) = \int_{-\infty}^{m} -(m - m_2)f(y)dy + \int_{m}^{m_2} [(y - m) - (m_2 - y)]f(y)dy$$

$$+ \int_{m_2}^{\infty} (m_2 - m) f(y)dy$$

$$= \int_{-\infty}^{m} -(m - m_2)f(y)dy + \int_{m_2}^{\infty} (m_2 - m) f(y)dy$$

$$= 0$$

Hence the posterior loss is minimized by any median of the posterior distribution.

28

(a) Suppose that $\theta$ has a Gamma$(a, b)$ distribution. The posterior distribution of $\theta$ is proportional to

$$f(\theta|X_1, \ldots, X_n) \propto f(X_1, \ldots, X_n|\theta)f(\theta)$$

$$\propto e^{-n\theta} \sum_{i=1}^{n} X_i^a \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta}$$

$$\propto \theta^{a-1} \sum_{i=1}^{n} X_i e^{-(n+b)\theta}$$

The last expression is proportional to the Gamma$(a + \sum_{i=1}^{n} X_i, n + b)$ density. Hence the posterior distribution of $\theta$ is Gamma$(a + \sum_{i=1}^{n} X_i, n + b)$.

(b) The Bayes estimate of $\theta$ under squared error loss is the posterior mean (Theorem A, page 584). Hence from part (a), the Bayes estimate of $\theta$ is the mean of a Gamma$(a + \sum_{i=1}^{n} X_i, n + b)$ random variable, i.e.,

$$\hat{\theta}_{Bayes} = \frac{a + \sum_{i=1}^{n} X_i}{b + n}$$

We have

$$\frac{a + \sum_{i=1}^{n} X_i}{b + n} = \frac{a}{b} \frac{ab}{a(b + n)} + \frac{\sum_{i=1}^{n} X_i}{n} \frac{an}{a(b + n)}$$

Thus the Bayes estimate is a weighted average of the prior mean and $\bar{X}$.

(c) The risk of $\bar{X}$ under squared error loss is

$$E(\bar{X} - \theta)^2 = \frac{Var(X)}{n} = \frac{\theta}{n}$$

The risk of the Bayes estimate under squared error loss is

$$E\left(\frac{a + \sum_{i=1}^{n} X_i}{b + n} - \theta\right)^2 = [E\left(\frac{a + \sum_{i=1}^{n} X_i}{b + n} - \theta\right)]^2 + Var\left(\frac{a + \sum_{i=1}^{n} X_i}{b + n}\right)$$

$$= \left(\frac{a - b\theta}{b + n}\right)^2 + \frac{n\theta}{(b + n)^2}$$

For values of $\theta$ near $\frac{a}{b}$, the Bayes estimate has smaller risk than $\bar{X}$. For values of $\theta$ far from $\frac{a}{b}$, $\bar{X}$ has smaller risk than the Bayes estimate.
**Extra Problem A**

Because the prior distribution of \( p \) is

\[
f(p) = \begin{cases} 
\frac{1}{2} & \text{w.p. } Q \\
\text{Uniform}(0, 1) & \text{w.p. } 1-Q
\end{cases}
\]

and the data \( \{X_i\}_{i=1}^n \) are i.i.d. Bernoulli \( p \), the posterior distribution is evaluated as

\[
f(p|\text{Obs.}) = f(\text{Obs.}|p) \cdot f(p) / f(\text{Obs.})
\]

\[
= \left( \frac{n}{x} \right) \left( \frac{1}{2} \right)^n \cdot Q / g(x) \quad \text{when } p = \frac{1}{2}
\]

\[
= \left( \frac{n}{x} \right) p^x (1-p)^{(n-x)} \cdot (1-Q) / g(x) \quad \text{when } p \neq \frac{1}{2},
\]

where \( x = \sum_{i=1}^n X_i \) and \( g(x) \) denotes the marginal distribution of \( x \). Hence, the posterior probabilities for each hypothesis are respectively,

\[
P(H_0|\text{Obs.}) = \frac{1}{g(x)} \left( \frac{n}{x} \right) \left( \frac{1}{2} \right)^n \cdot Q
\]

\[
P(H_1|\text{Obs.}) = \frac{1}{g(x)} \int_0^1 \left( \frac{n}{x} \right) p^x (1-p)^{(n-x)} dp \cdot (1-Q)
\]

\[
= \frac{1}{g(x)} \left( \frac{n}{x} \right) \Gamma(x+1) \Gamma(n-x+1) / \Gamma(n+2) \cdot (1-Q) = \frac{1}{g(x)} (1-Q)/(n+1).
\]

Using zero-one loss, the Bayes test will reject the null hypothesis when

\[
P(H_1|\text{Obs.}) > P(H_0|\text{Obs.}).
\]

or equivalently

\[
(1-Q)/(n+1) > \left( \frac{n}{x} \right) \left( \frac{1}{2} \right)^n \cdot Q.
\]

Since, according to Stirlings’ formula, \( 1/(n) \) can be approximated by,

\[
\sqrt{2\pi n} \left( \frac{x}{n} \right)^{x+1/2} \left( 1 - \frac{x}{n} \right)^{-x+1/2},
\]

the rejection region for the Bayes test is approximately

\[
\left( \frac{x}{n} \right)^{x+1/2} \left( 1 - \frac{x}{n} \right)^{-x+1/2} \cdot \sqrt{\frac{2\pi}{n}} \cdot \frac{1-Q}{Q} > \left( \frac{1}{2} \right)^n
\]

or equivalently

\[
\left( \frac{x}{n} \right)^x \left( 1 - \frac{x}{n} \right)^{-x} / \left( \frac{1}{2} \right)^n > \sqrt{\frac{n}{x/n(1-x/n)}} \cdot \sqrt{\frac{1}{2\pi} \cdot \frac{Q}{1-Q}}.
\]

The likelihood ratio is

\[
\Lambda = \left( \frac{1}{2} \right)^n / \left( \frac{x}{n} \right)^x \left( 1 - \frac{x}{n} \right)^{-x}
\]
Under the null hypothesis, we have that $-2\log \Lambda$ has approximately a $\chi^2(1)$ distribution (page 310). Thus, the approximate rejection region for the likelihood ratio test is

$$\left(\frac{x}{n}\right)^x \left(1 - \frac{x}{n}\right)^{n-x} \left(\frac{1}{2}\right)^n > \chi^2_{1-a}/2.$$  

Noting that the rejection region for the Bayes test is $O(\sqrt{n})$ whereas the rejection region for the likelihood ratio test with a fixed significance level is $O(1)$, we may conclude that the Bayes test is rather conservative compared to a fixed significance level likelihood ratio test.