STA 623 homework 3

Lingyun Shao

Sep. 24, 2018

Problem Statement

Assume $X \sim N(\theta, 1)$. The loss is the squared error loss $L(\theta, a) = (\theta - a)^2$. Consider a decision rule $\delta_c(x) = cx$, so the risk function here is $R(\theta, \delta_c) = E_\theta^x[(\theta - cx)^2] = c^2 + (1-c)^2\theta^2$. δ_c is inadmissible for $c > 1$. For $c \leq 1$ no decision rule dominates across all θ values.

(1) In the above example, calculate a Bayes estimator for θ under squared error loss and assuming a $N(m, v)$ prior for θ. [Note: a Bayes estimator corresponds to the action that minimizes the expectation of the loss function marginalizing over the posterior.]

Suppose the prior of θ is $\pi(\theta)$. Then we have the posterior of θ given x

$$\begin{aligned}
\pi(\theta|x) &\propto \pi(\theta)p(x|\theta) \\
&\propto \exp\left\{-\frac{(\theta-m)^2}{2v}\right\}\exp\left\{-\frac{(x-\theta)^2}{2}\right\} \\
&= \exp\left\{-\frac{\theta^2 - 2\theta m + m^2 + vx^2 - 2v\theta x + v\theta^2}{2v}\right\} \\
&\propto \exp\left\{-\frac{(1+v)\theta^2 - 2(vx+m)\theta}{2v}\right\} \\
&\propto \exp\left\{-\frac{(\theta - \frac{vx+m}{v+1})^2}{2v/(v+1)}\right\}
\end{aligned}$$

Judging from the kernal of the posterior, we know that $\theta|x \sim N(\frac{vx+m}{v+1}, \frac{v}{v+1})$.

The Bayes estimator $\hat{\theta}_B = \arg\min E^{\theta|x}(\hat{\theta} - \theta)^2$.

$$\begin{aligned}
E^{\theta|x}(\hat{\theta}-\theta)^2 &= \int_\Theta (\hat{\theta}-\theta)^2\pi(\theta|x)d\theta \\
&= \hat{\theta}^2 - 2\hat{\theta}\int_\Theta \theta\pi(\theta|x)d\theta + \int_\Theta \theta^2\pi(\theta|x)d\theta
\end{aligned}$$

Since this is a quadratic function of $\hat{\theta}$ with a positive quadratic coefficient, the minimum can be obtained at $\hat{\theta}_B = \int_\Theta \theta\pi(\theta|x)d\theta = E[\theta|x]$, which is our Bayes estimator. So the Bayes estimator here is the expectation of the posterior distribution $N(\frac{vx+m}{v+1}, \frac{v}{v+1})$. Thus, we have $\hat{\theta}_B = \frac{vx+m}{v+1}$.

(2) Express the Bayes estimator as a decision rule $\delta^{(B)}(x)$. Calculate the frequentist risk function for this rule.

We can express the Bayes estimator as $\delta^{(B)}(x) = \frac{vx+m}{v+1} = \frac{v}{v+1}x + \frac{m}{v+1}$, which takes an input of x and give an output of estimating θ.

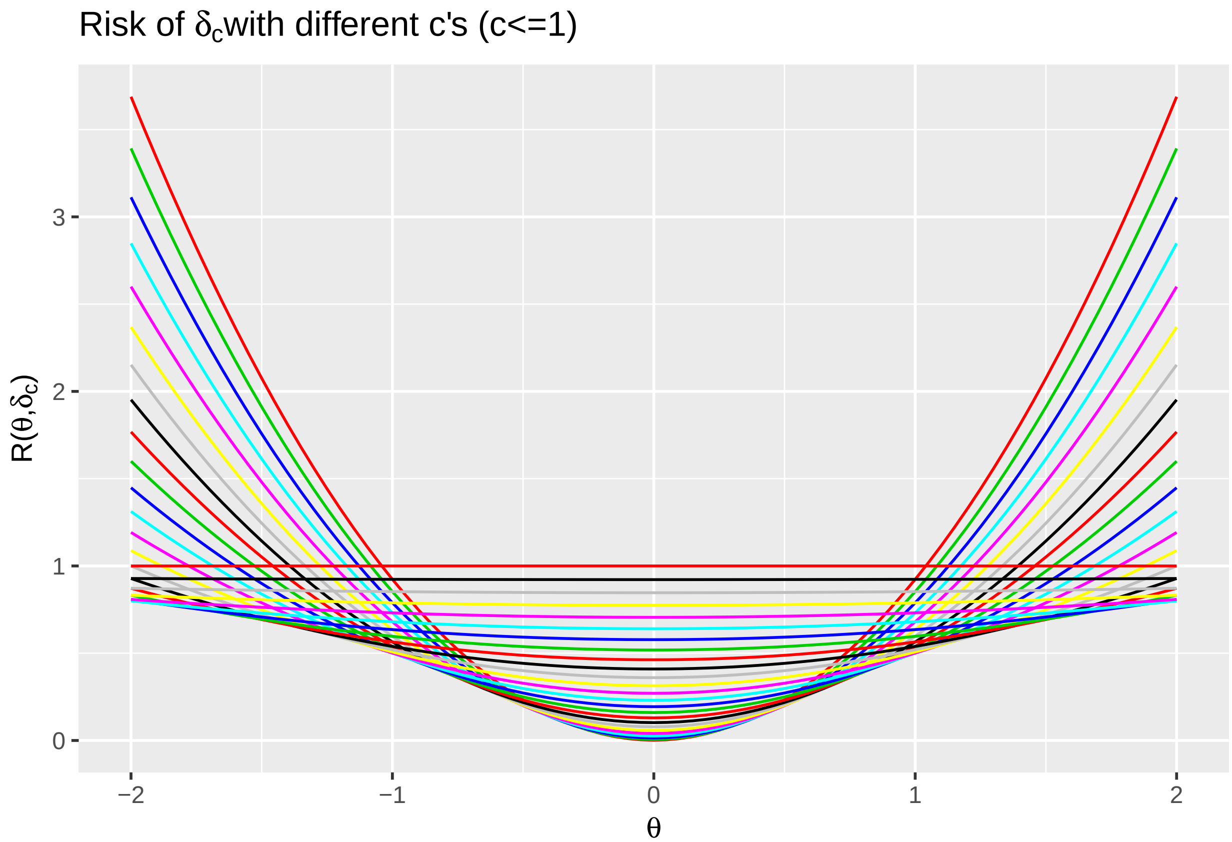
The frequentist risk function for this decision rule is

$$
\begin{aligned}
R(\theta, \delta^{(B)}) &= E_\theta^x[L(\theta, \delta^{(B)})] \\
&= E_\theta^x(\theta - \frac{vx+m}{v+1})^2 \\
&= \theta^2 - \frac{2\theta}{v+1}E_\theta^x[vx+m] + \frac{1}{(v+1)^2}E[(vx+m)^2] \\
&= \theta^2 - \frac{2\theta}{v+1}(v\theta+m) + \frac{1}{(v+1)^2}E[(vx+m)^2] \\
&= \theta^2 - \frac{2\theta}{v+1}(v\theta+m) + \frac{1}{(v+1)^2}E[v^2x^2+2vmx+m^2] \\
&= \theta^2 - \frac{2\theta}{v+1}(v\theta+m) + \frac{1}{(v+1)^2}(v^2(1+\theta^2)+2vm\theta+m^2) \\
&= \theta^2 - \frac{2\theta}{v+1}(v\theta+m) + \frac{1}{(v+1)^2}(v^2+(v\theta+m)^2) \\
&= \theta^2 - 2\theta\frac{v\theta+m}{v+1} + \frac{(v\theta+m)^2}{(v+1)^2} + \frac{v^2}{(v+1)^2} \\
&= (\theta - \frac{v\theta+m}{v+1})^2 + \frac{v^2}{(v+1)^2} \\
&= (\frac{\theta-m}{v+1})^2 + \frac{v^2}{(v+1)^2}
\end{aligned}
$$

(3) Compare $\delta^{(B)}(x)$ to the class of $\delta_c(x)$ decision rules describes above. Is $\delta^{(B)}(x)$ inadmissible? How does m, v [hyperparameters in the prior] impact risk with comparisons to $\delta_c(x)$?

The frequentist risk function of δ_c is $R(\theta, \delta_c) = (1-c)^2\theta^2 + c^2$. The frequantist risk function of $\delta^{(B)}$ is $R(\theta, \delta^{(B)}) = (\frac{\theta-m}{v+1})^2 + \frac{v^2}{(v+1)^2}$. Notice that if we let $m = 0, v = \frac{c}{1-c}$, then we have $R(\theta, \delta^{(B)}) = (\frac{\theta-0}{1/1-c})^2 + \frac{c^2/(1-c)^2}{1/(1-c)^2} = (1-c)^2\theta + c^2 = R(\theta, \delta_c)$. Given appropriate hyperparameters, these two decision rules can be equivalent.

```
n = 25
c = (0:n)/n
x = seq(-2, 2, by = 0.05)
r = c^2 + (1-c)^2 %*% t(x^2)
r = data.frame(cbind(t(r), x) )
library(ggplot2)
p = ggplot()
for(i in 2:(ncol(r)-1)){
  p = p + geom_line(data = r, aes_string(x = x, y = r[[i]]), color = i)
}
p + labs(x = expression(theta),
         y = expression(paste('R(',theta, ',', delta[c], ')', sep = '')),
         title = expression(paste('Risk of ', delta[c], 'with different c\'s (c<=1)', sep = '')))
```



We drew the risk functions of δ_c w.r.t different c's that are no greater than 1. From the figure, we know that none of these curves is uniformly above the others for $\theta \in \mathbb{R}$. Since we have already reached the conclusion that when $c > 1$, δ_c is inadmissible, then according to this figure we know that no decision rule dominates across all θ values. If we only consider δ_c and $\delta^{(B)}$, we can always convert the $\delta_{(B)}$ to δ_c by letting $m = 0, v = c/(1-c)$. Therefore, $\delta^{(B)}(x)$ is not inadmissible.

Even if $m \neq 0$, the curve of risk function is just horizontally shifted and $c = \frac{v}{1+v} \leq 1$. So $\delta_{(B)}$ is still admissible when we only considers δ_c's and $\delta^{(B)}$'s. The hyperparameter m affects the value of optimal θ to obtain the minimal risk. The hyperparameter v controls the curvature of the risk function.

(4) Calculate the Bayesian expected loss of $\delta_c(x)$ in the above setting. Does $\delta_c(x)$ have lower expected loss than $\delta^{(B)}(x)$ for any choice of c?

By definition, the Bayesian expected loss of δ_c is

$$
\begin{aligned}
&E^{\theta|x}[(cx-\theta)^2] \\
=&E^{\theta|x}[c^2x^2 - 2cx\theta + \theta^2] \\
=&c^2x^2 - 2cxE^{\theta|x}\theta + [(E^{\theta|x}\theta)^2 + Var^{\theta|x}(\theta)] \\
=&c^2x^2 - 2cx\frac{vx+m}{v+1} + (\frac{vx+m}{v+1})^2 + \frac{v}{v+1} \\
=&(cx - \frac{vx+m}{v+1})^2 + \frac{v}{v+1}
\end{aligned}
$$

By definition, the Bayesian expected loss of $\delta^{(B)} is$

$$
\begin{aligned}
&E^{\theta|x}[(\frac{vx+m}{v+1}-\theta)^2]\\
=&(\frac{vx+m}{v+1})^2-2\frac{vx+m}{v+1}E^{\theta|x}\theta+[(E^{\theta|x}\theta)^2+Var^{\theta|x}(\theta)]\\
=&(\frac{vx+m}{v+1})^2-2(\frac{vx+m}{v+1})^2+(\frac{vx+m}{v+1})^2+Var^{\theta|x}(\theta)\\
=&Var^{\theta|x}(\theta)\\
=&\frac{v}{v+1}
\end{aligned}
$$

By comparing these two Bayesian expected loss, we can easily find that no matter what c we choose, the Bayesian expected loss of δ_c can never be lower than that of $\delta^{(B)}$.