STAT 4010 – Bayesian Learning

Tutorial 5

Spring 2022

Cheuk Hin (Andy) CHENG (Email | Homepage)

Di SU (Email | Homepage)

1 Admissibility

Definition 1. *(Uniformly dominate & admissible) Given a loss function L, an estimator $\hat{\theta}'$* ***uniformly dominates*** *$\hat{\theta}$ iff*

1. $R(\theta, \hat{\theta}') \leq R(\theta, \hat{\theta})$ *for all θ; and*
2. $R(\theta, \hat{\theta}') < R(\theta, \hat{\theta})$ *for some θ.*

If an estimator $\hat{\theta}$ is uniformly dominated by other estimator, then $\hat{\theta}$ is said to be ***inadmissible****. If there is no such estimator uniformly dominates $\hat{\theta}$, then $\hat{\theta}$ is said to be* ***admissible****.*

Remark 1.1. We use the (Frequentist) risk to compare the estimators, not the Bayesian risk (why?).

Theorem 1.1. *(Uniqueness of Bayes estimator) A Bayes estimator $\hat{\theta}_\pi$ is unique if*

1. $L(\theta, \hat{\theta})$ *is strictly convex function in $\hat{\theta}$;*
2. $R(\pi, \hat{\theta}_\pi) < \infty$*;*
3. *Θ is open and equal to the support of $\pi(\theta)$.*
4. $\Pr(x_1 \leq a_1, \ldots, x_n \leq a_n \mid \theta)$ *is continuous in θ for all $a_1, \ldots, a_n$.*

Theorem 1.2. *(Admissibility of Bayes estimator) Any* ***unique*** *Bayes estimator is* ***admissible****.*

Theorem 1.3. *(Bayes risk under squared loss) Consider the squared loss. If the posterior variance is free of $x_{1:n}$, then the Bayes risk is the posterior variance, i.e. $\mathsf{Var}(\theta \mid x_{1:n})$.*

(Proof of theorem 1.3) Consider the Bayes risk and the Bayes estimator $\hat{\theta}_\pi$. Recall that $\hat{\theta}_\pi = \mathsf{E}(\theta \mid x)$. Note that $\mathsf{Var}(\hat{\theta}_\pi - \theta \mid x) = \mathsf{Var}(\theta \mid x)$. If the posterior variance is free of $x_{1:n}$, we have

$$\begin{aligned} R(\pi, \hat{\theta}_\pi) &= \int_{\mathcal{X}} \int_{\Theta} [\hat{\theta}_\pi - \theta]^2 \mathrm{d}F(\theta \mid x) \mathrm{d}F(x) \\ &= \int_{\mathcal{X}} \mathsf{Var}(\theta \mid x) \mathrm{d}F(x) \\ &= \mathsf{Var}(\theta \mid x_{1:n}) \int_{\mathcal{X}} \mathrm{d}F(x) \\ &= \mathsf{Var}(\theta \mid x_{1:n}). \end{aligned}$$

Example 1.1. Consider the Normal-Normal model,

$$\begin{aligned} x_{1:n} \mid \theta &\overset{\text{IID}}{\sim} \mathrm{N}(\theta, \sigma^2), \\ \theta &\sim \mathrm{N}(\nu_0, \tau_0), \end{aligned}$$

where σ^2 and τ_0^2 are known and finite. Find the Bayes estimator under the squared loss. Is the Bayes estimator admissible?

Solution:

We know that the posterior is $\mathrm{N}(\nu_n, \tau_n)$ and the corresponding Bayes estimator under the squared loss is

$$\hat{\theta}_\pi = \mathsf{E}[\theta \mid x_{1:n}] = \frac{\tau_0^2}{\sigma^2 + \tau_0^2}\bar{x} + \frac{\sigma^2}{\sigma^2 + \tau_0^2}\nu_0.$$

We first show the uniqueness. Then by theorem 1.2, $\hat{\theta}_\pi$ is admissible. The squared loss function is strictly convex. Since the posterior variance

$$\mathsf{Var}(\theta \mid x_{1:n}) = \frac{\sigma^2\tau_0^2}{\tau_0^2 + \sigma^2},$$

is free of x. By theorem 1.3, the Bayes risk is the posterior variance which is finite. Moreover, the Normal-Normal model is regular enough to satisfy the conditions 3 and 4 in theorem 1.2. Therefore, $\hat{\theta}_\pi$ is unique Bayes estimator and thus admissible.

Example 1.2. By previous example, we can see that the Bayes estimator is in the form of $\hat{\theta} = a\bar{X} + b$, where a and b are known constant. Find the range of a and b such that $\hat{\theta}$ is inadmissible.

Solution:

We first write the Frequentist risk in terms of function of a and b.

$$\begin{aligned} R(\theta, \hat{\theta}) &= \mathsf{E}[L(\theta, \hat{\theta}) \mid \theta] \\ &= \mathsf{E}[(a\bar{X} + b - \theta)^2 \mid \theta] \\ &= \mathsf{Var}(a\bar{x} + b - \theta \mid \theta) + \{\mathsf{E}[(a\bar{X} + b - \theta) \mid \theta]\}^2 \\ &= a^2\frac{\sigma^2}{n} + (\{a-1\}\theta + b)^2 \\ &=: \rho(a, b). \end{aligned}$$

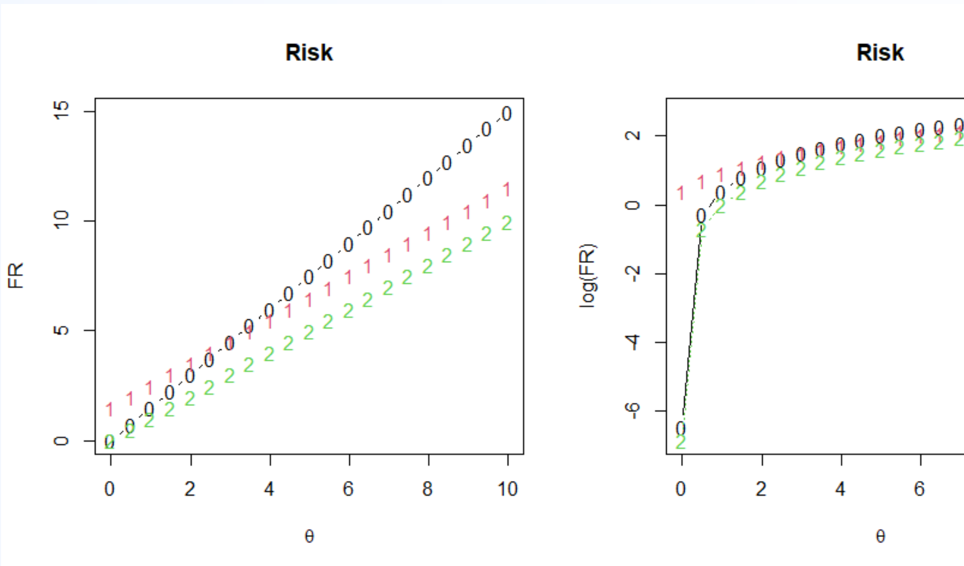
Let $\sigma^2 = \mathsf{Var}(x \mid \theta)$. Consider following cases.

- ($a > 1$) If $a > 1$ and for any value of b, then $\rho(a,b) \geq a^2\sigma^2/n > \sigma^2/n = \rho(1,0)$. Therefore, $a\bar{x} + b$ is uniformly dominated by $\bar{x}$ for any $a > 1$.
- ($a = 1$ and $b \neq 0$) If $a = 1$ and for any value of $b \neq 0$, then $\rho(1,b) = \sigma^2/n + b^2 > \sigma^2/n = \rho(1,0)$. Again, $\bar{x} + b$ is uniformly dominated by $\bar{x}$ for any $b \neq 0$.
- ($0 < a < 1$): Observe that $a\bar{x} + b$ is a convex combination of $\bar{x}$ and b. It is a Bayes estimator with respect to some normal prior on θ. Since we are considering the squared error loss function, which is strictly convex, the Bayes estimator is unique, therefore, it is admissible.
- ($a = 0$): Roughly speaking, in this case, $R(b,b) = 0$, hence it is admissible.
- ($a < 0$) If $a < 0$, then for any value of b

$$\begin{aligned} \rho(a,b) &\geq (\{a-1\}\theta + b)^2 \\ &\geq (a-1)^2(\theta + b/(a-1))^2 \\ &> (\theta + b/(a-1))^2 \\ &= \rho(0, -b/(a-1)). \end{aligned}$$

  Therefore, $a\bar{x} + b$ is uniformly dominated by $-b/(a-1)$ for $a < 0$.

From above example, we can conclude that when a does not lie in $[0,1]$ or when $a = 1$ and $b \neq 0$, the estimator $a\bar{x} + b$ is inadmissible. Let $\hat{\theta}_0 = 1.5\bar{x}$, $\hat{\theta}_1 = \bar{x} + 1.5$ and $\hat{\theta}_2 = \bar{x}$. We can use the following figure to demonstrate that $\hat{\theta}_0$ and $\hat{\theta}_1$ are inadmissible.




```
##setting
nRep = 2^12
n = 100
a = c(1.5,1,1)
b = c(0,1.5,0)
theta.all = seq(0,10,length.out = 21)
sigma = 1

##A function to compute the risk
get_FR = function(x,a,b){
```

```
  FR = a*mean(x)+b
  return(FR)
}

# Simulate the risk of 1.5bar(x), bar(x) + 1.5 and bar(x).
out =array(NA,dim=c(nRep,length(theta.all),3))
dimnames(out) =list(paste0("iRep=",1:nRep),paste0("theta = ",1:length(theta.all)),paste0("q=",0:2))
set.seed(4010)
for(iRep in 1:nRep){
  for(i.t in 1:length(theta.all)){
    theta = theta.all[i.t]
    x = rnorm(n,mean=theta,sd = sigma)
    for(i.q in 1:3){
      out[iRep, i.t, i.q] = get_FR(x,a[i.q],b[i.q])
    }
  }
}
# Results
risk =apply(out, 2:3,mean)
head(risk)
risk_adj = log(risk)
windows(height = 5,width=10)
par(mfrow = c(1,2))
matplot(theta.all, risk, type="b", pch=as.character(q), main="Risk", xlab=expression(theta),ylab = 'FR')
matplot(theta.all, risk_adj, type="b", pch=as.character(q), main="Risk", xlab=expression(theta),ylab = 'log(FR)')
```

2 Minimax Estimator

Definition 2. *(Minimax risk and minimax estimator)*

1. *The maximum risk of an estimator θ is*

$$\bar{R}(\widehat{\theta}) = \sup_{\theta \in \Theta} R(\theta, \widehat{\theta}).$$

2. *The* ***minimax risk*** *for the model parameter $\theta \in \Theta$ is*

$$\bar{R} := \inf_{\widehat{\theta} \in D} \sup_{\theta \in \Theta} R(\theta, \widehat{\theta}).$$

3. *The* ***minimax estimator*** *is an estimator $\widehat{\theta}_{\mathrm{M}}$ that achieves minimax risk, i.e.,*

$$\bar{R}\left(\widehat{\theta}_{\mathrm{M}}\right) = \bar{R}.$$

Remark 2.1. Different principles have different "attention" on the risk curve $R(\theta, \hat{\theta})$.

- Minimax estimator performs well ***at the worst case***.
- Bayes estimator performs well ***on (weighted) average***.

- In STAT4003, we've learned about the uniformly mininal variance estimator (UMVUE) which performs well (w.r.t. the squared loss) ***for all*** θ.

Remark 2.2. Minimax estimator may not be admissible. (Chapter 3 Example 3.19)

Definition 3. *Let* $\widehat{\theta}_\pi, \widehat{\theta}_{\pi'}, \widehat{\theta}_{\pi_m}$ *be Bayes estimators under* π, π', π_m*, respectively, where* $m = 1, 2, \ldots$

1. *The prior* π *is* ***least favorable*** *if for any proper prior* π'
$$R\left(\pi, \widehat{\theta}_\pi\right) \geq R\left(\pi', \widehat{\theta}_{\pi'}\right).$$
2. *The sequence of priors* $\{\pi_m\}$ *is* ***asymptotically least favorable*** *if for any proper prior* π'
$$\lim_{m\to\infty} R\left(\pi_m, \widehat{\theta}_{\pi_m}\right) \geq R\left(\pi', \widehat{\theta}_{\pi'}\right).$$

The following theorems provide us methods of finding minimax estimators.

Theorem 2.1. *If there is* π *such that* $R\left(\theta, \widehat{\theta}_\pi\right)$ ***does not depend on*** θ*, then* $\widehat{\theta}_\pi$ *is minimax for* θ*.*

Theorem 2.2. *If there is* π *such that* $R\left(\pi, \widehat{\theta}_\pi\right) \geq R\left(\theta, \widehat{\theta}_\pi\right)$ ***for all*** θ*, then* $\widehat{\theta}_\pi$ *is minimax for* θ*. And in this case, the prior* π *is least favorable.*

Theorem 2.3. *If there is* $\{\pi_m\}$ *and* $\widehat{\theta}$ *such that* $R\left(\pi_m, \widehat{\theta}_{\pi_m}\right) \to \bar{R}(\widehat{\theta})$ ***as*** $m \to \infty$*, then* $\widehat{\theta}$ *is minimax for* θ*. In this case, the sequence of prior* $\{\pi_m\}$ *is asymptotically least favorable.*

Remark 2.3. The assumption in Theorem 2.2 means that a Bayes estimator has its Bayes risk equal to its maximal (frequentist) risk.

Remark 2.4. Theorem 2.1 is a special case of Theorem 2.2

Remark 2.5. We can find a minimax estimator by finding a Bayes estimator with its Bayes risk equal to its maximal risk, or with its Bayes risk being a constant. It requires us to design the prior and the support of the prior wisely.

Example 2.1. (Proof of Theorem 2.2) Consider a prior π. Assume $R\left(\pi, \widehat{\theta}_\pi\right) \geq R\left(\theta, \widehat{\theta}_\pi\right)$ for all θ, want to show that $\widehat{\theta}_\pi$ is minimax and that π is least favorable.
Solution:

Let $\widehat{\theta}$ be another estimator of θ. We have

$$\begin{aligned}
\sup_{\theta\in\Theta} R(\theta,\widehat{\theta}_\pi) &\le R(\pi,\widehat{\theta}_\pi) && \text{Assumption}\\
&= \int_\Theta R(\theta,\widehat{\theta}_\pi)\pi(\theta)\mathrm{d}\theta && \text{Definition of Bayes risk}\\
&\le \int_\Theta R\left(\theta,\widehat{\theta}\right)\pi(\theta)\mathrm{d}\theta && \text{Definition of Bayes Estimator}\\
&\le \sup_{\theta\in\Theta} R\left(\theta,\widehat{\theta}\right) && \text{Average } \le \text{ Maximum}
\end{aligned}$$

Since $\widehat{\theta}$ is arbitrary, we have proved that $\widehat{\theta}_\pi$ is minimax for θ. Next, let η be another prior on θ. Then

$$\begin{aligned}
R(\pi,\widehat{\theta}_\pi) &\ge \sup_{\theta\in\Theta} R(\theta,\widehat{\theta}_\pi) && \text{Assumption}\\
&\ge \int_\Theta R(\theta,\widehat{\theta}_\pi)\eta(\theta)\mathrm{d}\theta && \text{Maximum } \ge \text{ Average}\\
&\ge \int_\Theta R\left(\theta,\widehat{\theta}_\eta\right)\eta(\theta)\mathrm{d}\theta && \text{Definition of Bayes Estimator}\\
&= R\left(\eta,\widehat{\theta}_\eta\right) && \text{Definition of Bayes risk}
\end{aligned}$$

Since η is arbitrary, π is least favorable. This completes the proof. □

Example 2.2. (Minimax estimator of Binomial model) Suppose $[x_1,\ldots,x_n \mid \theta] \sim \text{Bern}(\theta)$ for some $\theta\in(0,1)$ and $L(\theta,\hat{\theta}) = (\theta-\hat{\theta})^2$. Find a minimax estimator of θ.

Solution:

- Proposal 1: *The Bayes estimator under a conjugate prior. Does it have constant risk?* Denote $\bar{x}_n$. By Example 3.24 in the lecture note, under the prior $\theta \sim \text{Beta}(\sqrt{n/4},\sqrt{n/4})$, the Bayes estimator is given by

$$\hat{\theta}_M = \frac{n\bar{x}_n + \sqrt{n}/2}{n+\sqrt{n}},$$

  and it has a constant risk $R(\theta,\hat{\theta}_M) = 1/[4(\sqrt{n}+1)^2]$. Hence $\hat{\theta}_M$ is a minimax estimator.

- Proposal 2: *The sample average $\tilde{\theta} = \bar{x}_n$. Can we find a suitable prior for it?* We want a prior $\pi(\theta)$ so that $\tilde{\theta}$ is the corresponding Bayes estimator and $R(\pi,\tilde{\theta}) \ge R(\theta,\tilde{\theta})$. Notice that the risk of $\tilde{\theta}$ is

$$R\left(\theta,\frac{\sum_{i=1}^n x_i}{n}\right) = \frac{\theta(1-\theta)}{n}.$$

  The risk has a unique maximum at $\theta^* = \frac{1}{2}$. Hence the maximum risk of $\tilde{\theta}$ is:

$$\sup_{\theta\in\Omega} R\left(\theta,\tilde{\theta}\right) = R\left(\frac{1}{2},\frac{\sum_{i=1}^n x_i}{n}\right) = \frac{1}{4n}.$$

  If we want to use Theorem 2.2, we need the prior π to satisfy $\pi(\theta) = \mathbb{1}(\theta = 1/2)$. However, in this case, the Bayes estimator is $\hat{\theta}_\pi = 1/2$, and it is not equal to $\tilde{\theta}$. Therefore,

we still do not know whether the sample average $\tilde{\theta}$ is a minimax estimator. (Notice that we cannot say $\tilde{\theta}$ is not a minimax estimator at this stage, because Theorem 2.2 is a sufficient but not necessary condition for minimax estimator.)
On the other hand, observe that the maximum risk of Proposal 1 is $R(\theta, \hat{\theta}_M) = 1/[4(\sqrt{n}+1)^2]$, and $R(\theta, \hat{\theta}_M) < R(\theta, \tilde{\theta}) = 1/[4(\sqrt{n})^2]$, hence the sample average $\tilde{\theta}$ is not a minimax estimaotr.

Takeaway: The commonly used sample average is no longer favorable in the minimax sense.

Lemma 2.4. *Suppose $\hat{g}(\theta)$ is an unbiased estimator of $g(\theta)$ with finite Bayesian risk and $\mathsf{E}(g(\theta)^2) < \infty$. Then under the squared loss, if $\hat{g}(\theta)$ is Bayes, the Bayes risk must be zero, i.e.,*

$$\mathsf{E}_{\theta,X}\left[\{\hat{g}(\theta) - g(\theta)\}^2\right] = 0.$$

Proof. Let $\hat{\theta}$ be an unbiased estimator under the squared loss function. Then by Theorem 3.2 in the lecture,

$$\hat{\theta} = E(g(\theta) \mid X).$$

Thus,

$$\begin{aligned} E(\hat{g}(\theta)g(\theta)) &= E(E(\hat{g}(\theta)g(\theta) \mid X)) \\ &= E(\hat{g}(\theta)E(g(\theta) \mid X)) \\ &= E\left(\hat{g}(\theta)^2\right). \end{aligned}$$

On the other hand, due to unbiasedness,

$$\begin{aligned} E(\hat{g}(\theta)g(\theta)) &= E(E(\hat{g}(\theta)g(\theta) \mid \theta)) \\ &= E(g(\theta)E(\hat{g}(\theta) \mid \theta)) \\ &= E\left(g^2(\theta)\right) \end{aligned}$$

Observe that

$$\begin{aligned} E\left(\{\hat{g}(\theta) - g(\theta)\}^2\right) &= E\left(\hat{g}(\theta)\right) - 2E(\hat{g}(\theta)g(\theta)) + E\left(g^2(\theta)\right) \\ &= E\left(\hat{g}(\theta)\right) - E(\hat{g}(\theta)g(\theta)) + E\left(g^2(\theta)\right) - E(\hat{g}(\theta)g(\theta)) \\ &= E\left(\hat{g}(\theta)\right) - E\left(\hat{g}(\theta)\right) + E\left(g^2(\theta)\right) - E\left(g^2(\theta)\right) \\ &= 0 \end{aligned}$$

Thus we have $E\left(\{\hat{g}(\theta) - g(\theta)\}^2\right) = 0$.

Example 2.3. (Minimax estimator of Normal model with unknown mean θ) Let $[x_1, x_2, \ldots, x_n \mid \theta] \overset{iid}{\sim} \mathrm{N}(\theta, \sigma^2)$ with σ^2 known. Suppose $L(\theta, \hat{\theta}) = (\theta, \hat{\theta})^2$. Find a minimax estimator of θ.

Solution:

As in the previous example, consider the sample average $\bar{x}_n = \sum_{i=1}^n x_i/n$.

- *Can we find a suitable prior for it?*

Notice that $\mathsf{Bias}(\bar{x}_n) = 0$. The risk of $\bar{x}_n$ is

$$R(\theta, \bar{x}_n) = \mathsf{E}_\theta \left[(\bar{x}_n - \theta)^2 \right] = \mathsf{MSE}(\bar{x}_n) = \mathsf{Var}(\bar{x}_n) = \frac{\sigma^2}{n},$$

which is a constant. This suggests that $\bar{X}$ "can" be a minimax estimator. However, by Lemma 2.4, $\bar{x}_n$ is not Bayes for any prior. Theorem 2.1 and 2.2 are not applicable in this case.

- *Can we find a sequence of priors whose Bayes risk will converge to its risk?*
  Theorem 2.3 can still be helpful. By the calculations in Example 3.22, if we consider the sequence of priors $\pi_m = N(0, m^2)$, then $R(\theta, \hat{\theta}_M) \to R(\theta, \bar{x}_n)$. Thus $\bar{x}_n$ is indeed a minimax estimator under the squared loss.

Takeaway: Minimax estimator is not necessarily a Bayes estimator.