

## STAT 4010 Bayesian Learning

## TUTORIAL 5

Spring 2023

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  $\theta$ ; and
2.  $R(\theta, \hat{\theta}') < R(\theta, \hat{\theta})$  for some  $\theta$ .

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.  $\Theta$  is open and equal to the support of  $\pi(\theta)$ .
4.  $\Pr(x_1 \leq a_1, \dots, x_n \leq a_n \mid \theta)$  is continuous in  $\theta$  for all  $a_1, \dots, a_n$ .

**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.  $\text{Var}(\theta \mid x_{1:n})$ .

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

we have

$$\begin{aligned}
 R(\pi, \hat{\theta}_\pi) &= \int_x \int_{\Theta} [\hat{\theta}_\pi - \theta]^2 dF(\theta | x) dF(x) \\
 &= \int_x \text{Var}(\theta | x) dF(x) \\
 &= \text{Var}(\theta | x_{1:n}) \int_x dF(x) \\
 &= \text{Var}(\theta | x_{1:n}).
 \end{aligned}$$

**Example 1.1.** Consider the Normal-Normal model,

$$\begin{aligned}
 x_{1:n} | \theta &\stackrel{\text{iid}}{\sim} N(\theta, \sigma^2), \\
 \theta &\sim N(\nu_0, \tau_0^2),
 \end{aligned}$$

where  $\sigma^2$  and  $\tau_0^2$  are known and finite. Find the Bayes estimator under the squared loss. Is the Bayes estimator admissible?

**SOLUTION:**

We know that the posterior is  $N(\nu_n, \tau_n^2)$  and the corresponding Bayes estimator under the squared loss is

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

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

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

is free of  $x_{1:n}$ . 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  $b$ .

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

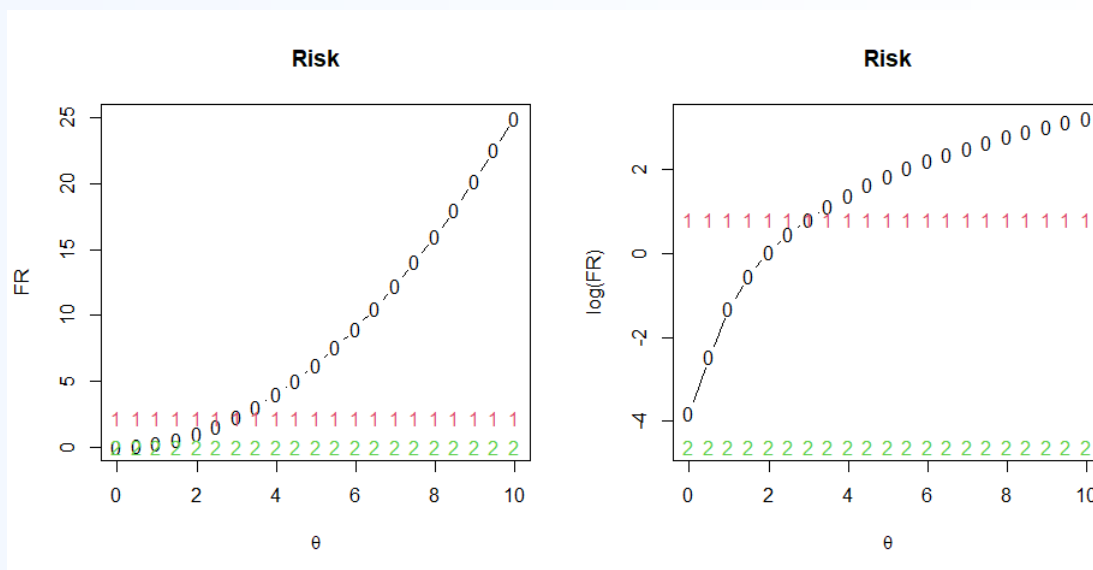
Let  $\sigma^2 = \text{Var}(x | \theta)$ . Consider following cases.

- ( $a > 1$ ) 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  $\theta$ . 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.



```
1 ##setting
2 nRep = 2^12
3 n = 100
4 a = c(1.5, 1, 1)
5 b = c(0, 1.5, 0)
6 theta.all = seq(0, 10, length.out = 21)
7 q = 0:2
8 sigma = 1
9
10 ##A function to compute the risk
```

```

11 get_FR = function(a,b,theta,sigma,n){
12   FR = a^2*sigma^2/n+((a-1)*theta+b)^2
13   return(FR)
14 }
15
16 # Simulate the risk of 1.5bar(x), bar(x) + 1.5 and bar(x) .
17 out =array(NA,dim=c(nRep,length(theta.all),3))
18 dimnames(out) =list(paste0("iRep=",1:nRep),paste0("theta = ",1:length(theta.
19   all)),paste0("q=",0:2))
20 set.seed(4010)
21 for(iRep in 1:nRep){
22   for(i.t in 1:length(theta.all)){
23     theta = theta.all[i.t]
24     x = rnorm(n,mean=theta,sd = sigma)
25     for(i.q in 1:3){
26       out[iRep, i.t, i.q] = get_FR(a[i.q],b[i.q],theta,sigma,n)
27     }
28   }
29 }
30 # Results
31 risk =apply(out, 2:3,mean)
32 head(risk)
33 risk_adj= log(risk)
34 windows(height = 5,width=10)
35 par(mfrow = c(1,2))
36 matplot(theta.all, risk, type="b", pch=as.character(q), main="Risk", xlab=
37   expression(theta),ylab = 'FR')
38 matplot(theta.all, risk_adj, type="b", pch=as.character(q), main="Risk", xlab=
39   expression(theta),ylab = 'log(FR)')

```

## 2 Minimax Estimator

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

1. The maximum risk of an estimator  $\hat{\theta}$  is

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

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

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

3. The **minimax estimator** is an estimator  $\hat{\theta}_M$  that achieves minimax risk, i.e.,

$$\bar{R}(\hat{\theta}_M) = \bar{R}.$$

**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 minimal variance estimator (UMVUE) which performs well (w.r.t. the squared loss) for all  $\theta$ .

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

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

1. The prior  $\pi$  is least favorable if for any proper prior  $\pi'$

$$R(\pi, \hat{\theta}_\pi) \geq R(\pi', \hat{\theta}_{\pi'}).$$

2. The sequence of priors  $\{\pi_m\}$  is asymptotically least favorable if for any proper prior  $\pi'$

$$\lim_{m \rightarrow \infty} R(\pi_m, \hat{\theta}_{\pi_m}) \geq R(\pi', \hat{\theta}_{\pi'}).$$

The following theorems provide us methods of finding minimax estimators.

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

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

**Theorem 2.3.** If there is  $\{\pi_m\}$  and  $\hat{\theta}$  such that  $R(\pi_m, \hat{\theta}_{\pi_m}) \rightarrow \bar{R}(\hat{\theta})$  as  $m \rightarrow \infty$ , then  $\hat{\theta}$  is minimax for  $\theta$ . 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  $\pi$ . Assume  $R(\pi, \hat{\theta}_\pi) \geq R(\theta, \hat{\theta}_\pi)$  for all  $\theta$ , want to show that  $\hat{\theta}_\pi$  is minimax and that  $\pi$  is least favorable.

**SOLUTION:**

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

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

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

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

Since  $\eta$  is arbitrary,  $\pi$  is least favorable. This completes the proof.  $\square$

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

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  $R(\pi, \tilde{\theta}) \geq R(\theta, \tilde{\theta})$ . 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:

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

If we want to use Theorem 2.2, we need the prior  $\pi$  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 estimator.

 **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  $E(g(\theta)^2) < \infty$ . Then under the squared loss, if  $\hat{g}(\theta)$  is Bayes, the Bayes risk must be zero, i.e.,

$$E_{\theta, X} [\{\hat{g}(\theta) - g(\theta)\}^2] = 0.$$

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

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

Thus,

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

On the other hand, due to unbiasedness,

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

Observe that

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

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

**Example 2.3.** (Minimax estimator of Normal model with unknown mean  $\theta$ ) Let  $[x_1, x_2, \dots, x_n | \theta] \stackrel{iid}{\sim} N(\theta, \sigma^2)$  with  $\sigma^2$  known. Suppose  $L(\theta, \hat{\theta}) = (\theta - \hat{\theta})^2$ . Find a minimax estimator of  $\theta$ .

**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  $\text{Bias}(\bar{x}_n) = 0$ . The risk of  $\bar{x}_n$  is

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

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  $R(\theta, \hat{\theta}_M) \rightarrow R(\theta, \bar{x}_n)$ . Thus  $\bar{x}_n$  is indeed a minimax estimator under the squared loss.

 **Takeaway:** Minimax estimator is not necessarily a Bayes estimator.