

# STATS300A - Lecture 11

Dominik Rothenhaeusler  
Scribed by Michael Howes

10/25/21

## Contents

<b>1</b>	<b>Announcement</b>	<b>1</b>
<b>2</b>	<b>Admissibility (<math>p = 1</math>)</b>	<b>1</b>
<b>3</b>	<b>Inadmissibility (<math>p \geq 3</math>)</b>	<b>2</b>

## 1 Announcement

Please arrive at least 5 minutes early to the exam on Wednesday.

## 2 Admissibility ( $p = 1$ )

We have seen that unique Bayes estimators are admissible. We wish to boost this result to  $\bar{X}_n$  which is a limit of Bayes estimators.

**Example 1.** Suppose  $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$  where  $\sigma^2$  is unknown and we are using squared error loss. To show that  $\bar{X}_n$  is admissible we will use a limiting Bayes argument. Suppose without loss of generality that  $\sigma^2 = 1$  and that  $\bar{X}_n$  is inadmissible. Note that the risk of  $\bar{X}_n$  is  $\frac{1}{n}$  constantly. Thus if  $\bar{X}_n$  is inadmissible, then there exists an estimator  $\delta$  such that  $R(\theta, \delta) < 1/n$  for some  $\theta$  and  $R(\theta, \delta) \leq 1/n$  for all  $\theta$ .

One can use the dominated convergence theorem to show that  $\theta \mapsto R(\theta, \delta)$  is continuous. Thus there exists an interval  $(\theta_0, \theta_1)$  such that  $\theta_1 - \theta_0 > 0$  and  $R(\theta, \delta) \leq 1/n - \varepsilon$  for all  $\theta \in (\theta_0, \theta_1)$ .

Let  $r'_\tau$  be the average risk of  $\delta$  with respect to the prior  $\theta \sim \mathcal{N}(0, \tau^2)$ . Also let  $r_\tau$  be the average risk of the Bayes estimator with respect to the prior  $\theta \sim \mathcal{N}(0, \tau^2)$ . We know that  $r_\tau$  is the posterior variance of  $\tau$  and thus

$$r_\tau = \frac{1}{n + 1/\tau^2} = \frac{\tau^2}{n\tau^2 + 1}.$$

Thus  $r_\tau$  approaches  $1/n$  as  $\tau \nearrow \infty$ . We also know that  $r'_\tau \leq 1/n$  for all  $\tau$ . We will now look at the

ratio  $\frac{1/n - r'_\tau}{1/n - r_\tau}$ . This is a sort of Taylor's expansion of  $r_\tau$  and  $r'_\tau$  about  $1/n$ . Note that

$$\begin{aligned}
 \frac{1/n - r'_\tau}{1/n - r_\tau} &= \frac{\int_{\mathbb{R}} (1/n - R(\theta, \delta)) \frac{1}{\sqrt{2\pi\tau}} \exp(-1/2\theta^2) d\theta}{\frac{1}{n} - \frac{\tau^2}{n\tau^2+1}} \\
 &= \frac{\int_{\mathbb{R}} (1/n - R(\theta, \delta)) \frac{1}{\sqrt{2\pi\tau}} \exp(-1/2\theta^2) d\theta}{\frac{n\tau^2+1}{n(n\tau^2+1)} - \frac{n\tau^2}{n(n\tau^2+1)}} \\
 &= \frac{\int_{\mathbb{R}} (1/n - R(\theta, \delta)) \frac{1}{\sqrt{2\pi\tau}} \exp(-1/2\theta^2) d\theta}{\frac{1}{n(n\tau^2+1)}} \\
 &= \frac{n(n\tau^2+1)}{\sqrt{2\pi\tau}} \cdot \int_{\mathbb{R}} (1/n - R(\theta, \delta)) \exp(-1/2\theta^2) d\theta \\
 &\geq \frac{n(n\tau^2+1)}{\sqrt{2\pi\tau}} \cdot \int_{\theta_0}^{\theta_1} (1/n - R(\theta, \delta)) \exp(-1/2\theta^2) d\theta \\
 &\geq \frac{n(n\tau^2+1)}{\sqrt{2\pi\tau}} \cdot \varepsilon \int_{\theta_0}^{\theta_1} \exp(-1/2\theta^2) d\theta.
 \end{aligned}$$

As  $\tau \rightarrow \infty$ ,  $\frac{n(n\tau^2+1)}{\sqrt{2\pi\tau}} \rightarrow \infty$  and by the dominated convergence theorem  $\int_{\theta_0}^{\theta_1} \exp(-1/2\theta^2) d\theta \rightarrow \int_{\theta_0}^{\theta_1} 1 d\theta = \theta_1 - \theta_0 > 0$ . Thus we have

$$\lim_{\tau \rightarrow \infty} \frac{1/n - r'_\tau}{1/n - r_\tau} = \infty.$$

In particular there exists  $\tau > 0$  such that  $\frac{1/n - r'_\tau}{1/n - r_\tau} > 1$ . This implies that  $r'_\tau < r_\tau$  which is a contradiction.

### 3 Inadmissibility ( $p \geq 3$ )

We will now look at an example of simultaneous estimation and look at the James-Stein estimator. The take away will be that minimax estimators and UMRUES need not be admissible.

Suppose  $X \in \mathbb{R}^p$  and  $X \sim \mathcal{N}(\theta, I_p)$  for some  $\theta \in \mathbb{R}^p$ . Our goal is to estimate  $\theta$  under the loss  $L(\theta, d) = \sum_{j=1}^p (\theta_j - d_j)^2 = \|\theta - d\|_2^2$ . The estimator  $\delta(X) = X$  is

- A minimax estimator for  $\theta$ .
- The UMRUES for  $\theta$ .
- The MLE for  $\theta$ , that is  $X = \arg \max_{\theta} p(x; \theta)$ .

From many perspectives  $X$  seems like the best estimator but  $X$  is inadmissible for  $p \geq 3$ . Recall the empirical Bayes estimator for  $\theta$

$$\left(1 - \frac{p}{\sum_i X_i^2}\right) X.$$

This came up in the setting  $\theta_i \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \tau^2)$  and  $X_i | \theta_i \stackrel{\text{ind}}{\sim} \mathcal{N}(\theta_i, 1)$ . We will see that a similar estimator will outperform  $\delta(X) = X$  uniformly in  $\theta$  in a frequentist setting.

For intuition one may ask what is the problem with  $\delta(X) = X$ ? The problem is that  $\|X\|_2^2$  is normally much larger  $\|\theta\|_2^2$  since  $\mathbb{E}[\|X\|_2^2] = \sum_{i=1}^p (\theta_i^2 + 1) = \|\theta\|_2^2 + p \gg \|\theta\|_2^2$ .

**Theorem 1.** [TPE 5.5.1] Define the estimator  $\delta^0$  by

$$\delta_j^0(X) = \left(1 - \frac{p-2}{\|X\|_2^2}\right) X_j.$$

The estimator  $\delta^0$  had strictly smaller risk than  $\delta(X) = X$  for all  $\theta$ . Thus  $\delta(X) = X$  is inadmissible.

We call  $\delta^0$  a *James-Stein estimator*.

*Proof.* We know that  $R(\theta, \delta) = p$  when  $\delta(X) = X$ . Now note that

$$\begin{aligned} R(\theta, \delta^0) &= \mathbb{E}_\theta \left[ \sum_j \left( \theta_j - \left(1 - \frac{p-2}{\|X\|_2^2}\right) X_j \right)^2 \right] \\ &= \mathbb{E}_\theta \left[ \sum_j \left( \theta_j - X_j + \frac{p-2}{\|X\|_2^2} X_j \right)^2 \right] \\ &= \sum_j \mathbb{E}_\theta [(\theta_j - X_j)^2] - 2 \sum_j \mathbb{E}_\theta \left[ (X_j - \theta_j) \left( \frac{p-2}{\|X\|_2^2} X_j \right) \right] + \sum_j \mathbb{E}_\theta \left[ \frac{(p-2)^2 X_j^2}{\|X\|_2^4} \right] \\ &= p - 2 \sum_j \mathbb{E}_\theta \left[ (X_j - \theta_j) \left( \frac{p-2}{\|X\|_2^2} X_j \right) \right] + \sum_j \mathbb{E}_\theta \left[ \frac{(p-2)^2 X_j^2}{\|X\|_2^4} \right]. \end{aligned}$$

Recall Stein's identity if  $X \sim \mathcal{N}(\mu, \sigma^2)$  we have

$$\mathbb{E}[g(X)(X - \mu)] = \sigma^2 \mathbb{E}[g'(X)].$$

By conditioning this gives

$$\mathbb{E}_\theta \left[ (X_j - \theta_j) \left( \frac{p-2}{\|X\|_2^2} X_j \right) \right] = \mathbb{E}_\theta \left[ (p-2) \frac{\partial}{\partial X_j} \left( \frac{X_j}{\|X\|_2^2} \right) \right].$$

If we make this substitution we have

$$\begin{aligned} R(\theta, \delta^0) &= p - 2 \sum_j \mathbb{E}_\theta \left[ (p-2) \frac{\partial}{\partial X_j} \left( \frac{X_j}{\|X\|_2^2} \right) \right] + \sum_j \mathbb{E}_\theta \left[ \frac{(p-2)^2 X_j^2}{\|X\|_2^4} \right] \\ &= p - 2 \sum_j \mathbb{E}_\theta \left[ (p-2) \frac{\|X\|_2^2 - 2X_j^2}{\|X\|_2^4} \right] + \sum_j \mathbb{E}_\theta \left[ \frac{(p-2)^2 X_j^2}{\|X\|_2^4} \right] \\ &= p - 2(p-2) \mathbb{E}_\theta \left[ \sum_j \frac{\|X\|_2^2 - 2X_j^2}{\|X\|_2^4} \right] + (p-2)^2 \mathbb{E}_\theta \left[ \sum_j \frac{X_j^2}{\|X\|_2^4} \right] \\ &= p - 2(p-2) \mathbb{E}_\theta \left[ \frac{p\|X\|_2^2 - 2\|X\|_2^2}{\|X\|_2^4} \right] + (p-2)^2 \mathbb{E}_\theta \left[ \frac{\|X\|_2^2}{\|X\|_2^4} \right] \\ &= p - 2(p-2)^2 \mathbb{E}_\theta \left[ \frac{1}{\|X\|_2^2} \right] + (p-2)^2 \mathbb{E}_\theta \left[ \frac{1}{\|X\|_2^2} \right] \\ &= p - (p-2)^2 \mathbb{E}_\theta \left[ \frac{1}{\|X\|_2^2} \right] \\ &< p. \end{aligned}$$

Thus  $\delta^0$  uniformly outperforms  $\delta$ . □

**Remark 1.** (a) Why only  $p > 2$ ? The random variable  $\frac{1}{\|X\|_2^2}$  is not integrable when  $p = 2$ . Thus the regularity conditions of Stein's identity are not met.

(b) The above theorem is surprising! We can improve the risk by sharing information across different dimensions. The components  $X_{\setminus j}$  are used to estimate  $\theta_j$  even though  $X_j$  are independent and  $\theta_j$  do not have any restrictions.

(c) Define  $\delta_j^v = \left(1 - \frac{p-2}{\|X_v\|_2^2}\right) (X_j - v_j)$  for  $v \in \mathbb{R}^p$ . Then the above proof shows that  $\delta^v$  also outperforms  $\delta$ .

(d) If  $\sigma \neq 1$ , then we can use  $\hat{\theta}_{JS} = \left(1 - \frac{\sigma^2(p-2)}{\|X\|_2^2}\right) X$ .

(e)  $\hat{\theta}_{JS}$  is inadmissible, we can improve by using

$$\delta'_j(X) = \left(1 - \frac{p-2}{\|X\|_2^2}\right)_+ X_j.$$

(f) A warning: For some  $j$ ,  $\mathbb{E}[(\delta_j^0(X) - \theta_j)^2]$  may be larger than  $\mathbb{E}[(X_j - \theta_j)^2]$ . An example is "Baseball batting averages - James-Stein estimator"