

Lecture 17: October 7

Lecturer: Siva Balakrishnan

In the last lecture we discussed Bayes estimators and minimax estimators which are optimal from different standpoints. Minimax estimators have lowest maximum risk, which Bayes estimators have lowest average (with respect to a distribution $\pi(\theta)$) risk.

We then discussed that Bayes estimators are often easy to compute: for instance for the squared-loss the posterior mean is the Bayes estimator. On the other hand, the minimax estimator is often difficult to compute directly.

We will study two ways in which to use Bayes estimators to find minimax estimators. One involves tightly bounding the minimax risk and the other involves identifying what is called a least favorable prior.

It is worth keeping in mind the trade-off: Bayes estimators although easy to compute are somewhat subjective (in that they depend strongly on the prior π). Minimax estimators although more challenging to compute are not subjective, but do have the drawback that they are protecting against the worst-case which might lead to pessimistic conclusions, i.e. the minimax risk might be much higher than the Bayes risk for a “nice” prior.

In 36-708/10-716/..., you will learn about ways to achieve a relaxed goal of computing estimators that achieve the minimax rate, i.e. estimators for which the risk goes to zero at the same rate as the minimax estimator. Formally,

$$\sup_{\theta \in \Theta} R(\theta, \hat{\theta}) \asymp \inf_{\tilde{\theta}} \sup_{\theta \in \Theta} R(\theta, \tilde{\theta}) \quad n \rightarrow \infty \quad (17.1)$$

where $a_n \asymp b_n$ means that both a_n/b_n and b_n/a_n are both bounded as $n \rightarrow \infty$.

17.1 Minimax Estimators through Bayes Estimators

Our goal is to compute a minimax estimator $\hat{\theta}$ that satisfies:

$$\sup_{\theta \in \Theta} R(\theta, \hat{\theta}) \leq \inf_{\tilde{\theta}} \sup_{\theta \in \Theta} R(\theta, \tilde{\theta}).$$

We will let θ_{minimax} denote a minimax estimator.

17.1.1 Bounding the Minimax Risk

One strategy to find the minimax estimator is by finding (upper and lower) bounds on the minimax risk that match. Then the estimator that achieves the upper bound is a minimax

estimator.

Upper bounding the minimax risk is straightforward. Given an estimator $\hat{\theta}_{\text{up}}$ we can compute its maximum risk and use it to upper bound the minimax risk, i.e.

$$\inf_{\tilde{\theta}} \sup_{\theta \in \Theta} R(\theta, \tilde{\theta}) \leq R(\theta, \hat{\theta}_{\text{up}}).$$

The Bayes risk of the Bayes estimator for any prior π lower bounds the minimax risk. Fix a prior π and suppose that $\hat{\theta}_{\text{low}}$ is the Bayes estimator with respect to π , then we have that:

$$B_{\pi}(\hat{\theta}_{\text{low}}) \leq B_{\pi}(\theta_{\text{minimax}}) \leq \sup_{\theta} R(\theta, \theta_{\text{minimax}}) = \inf_{\tilde{\theta}} \sup_{\theta \in \Theta} R(\theta, \tilde{\theta}).$$

Let us see an example of this in action.

Example: We will prove a classical result that if we observe independent draws from a d -dimensional Gaussian, $X_1, \dots, X_n \sim N(\theta, I_d)$, then the average:

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^n X_i,$$

is a minimax estimator of θ with respect to the squared loss. I will do the entire calculation for the d -dimensional case – if you find this confusing try to first work out the case when $d = 1$.

First, let us compute the upper bound. We note that,

$$\hat{\theta} \sim N(\theta, I_d/n),$$

so that its risk:

$$R(\theta, \hat{\theta}) = \mathbb{E}\left[\sum_{i=1}^d (\hat{\theta}_i - \theta_i)^2\right] = \mathbb{E}\left[\sum_{i=1}^d Z_i^2\right],$$

where $Z_i \sim N(0, 1/n)$. This yields that,

$$\inf_{\tilde{\theta}} \sup_{\theta \in \Theta} R(\theta, \tilde{\theta}) \leq R(\theta, \hat{\theta}) = \frac{d}{n}.$$

Let us now try to lower bound the minimax risk using the Bayes risk. Let us take the prior to be zero-mean Gaussian, i.e. we take $\pi = N(0, c^2 I_d)$. You can convince yourself that the likelihood $p(X_1, \dots, X_n | \theta) \propto p(\hat{\theta} | \theta)$ (you can do this directly or appeal to sufficiency). This in turn gives us that the posterior,

$$p(\theta | X_1, \dots, X_n) \propto p(\hat{\theta} | \theta) \pi(\theta) \propto p(\theta | \hat{\theta}),$$

is the same as the posterior in the following setting:

$$\begin{aligned}\theta &\sim N(0, c^2 I_d) \\ \hat{\theta} &\sim N(\theta, I_d/n),\end{aligned}$$

so that in order to compute the posterior mean we note that,

$$\begin{pmatrix} \theta \\ \hat{\theta} \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} c^2 I_d & c^2 I_d \\ c^2 I_d & (c^2 + 1/n) I_d \end{bmatrix} \right]$$

We can now compute the posterior (using standard conditional Gaussian formulae), and obtain its mean:

$$\mathbb{E}[\theta|\hat{\theta}] = \frac{c^2}{c^2 + 1/n} \hat{\theta}.$$

Now, the Bayes risk of this estimator provides us a lower bound on the minimax risk. To compute the Bayes risk we note that,

$$R(\theta, \hat{\theta}) = \mathbb{E}_{X_1, \dots, X_n} \left\| \frac{c^2}{c^2 + 1/n} \hat{\theta} - \theta \right\|_2^2.$$

Above we noted that $\hat{\theta} = \theta + Z$, where $Z \sim N(0, I_d/n)$, so we have

$$R(\theta, \hat{\theta}) = \mathbb{E}_Z \left\| \frac{c^2}{c^2 + 1/n} Z - \frac{\theta}{n(c^2 + 1/n)} \right\|_2^2.$$

Let us denote $\beta := c^2 + 1/n$. Then we obtain that,

$$R(\theta, \hat{\theta}) = \frac{\|\theta\|_2^2}{n^2 \beta^2} + \frac{c^4}{\beta^2} \mathbb{E} \|Z\|_2^2 = \frac{\|\theta\|_2^2}{n^2 \beta^2} + \frac{c^4 d}{\beta^2 n}.$$

The Bayes risk further averages this over $\theta \sim N(0, c^2 I_d)$ to obtain that,

$$B_\pi \left(\frac{c^2}{c^2 + 1/n} \hat{\theta} \right) = \frac{c^2 d}{n^2 \beta^2} + \frac{c^4 d}{\beta^2 n} = \frac{c^2 d}{n \beta} = \frac{d}{n(1 + 1/(nc^2))}.$$

Since c was arbitrary we can take the limit as $c \rightarrow \infty$ to obtain that the minimax risk is upper and lower bounded by d/n and conclude that the average $\hat{\theta}$ is minimax.

17.1.2 Least Favorable Prior

The other way to obtain Bayes estimators is by constructing what are called least favorable priors.

Theorem 17.1 Let $\hat{\theta}$ be the Bayes estimator for some prior π . If

$$R(\theta, \hat{\theta}) \leq B_{\pi}(\hat{\theta}) \quad \text{for all } \theta \quad (17.2)$$

then $\hat{\theta}$ is minimax and π is called a **least favorable prior**.

Proof:

Suppose that $\hat{\theta}$ is not minimax. Then there is another estimator $\hat{\theta}_0$ such that $\sup_{\theta} R(\theta, \hat{\theta}_0) < \sup_{\theta} R(\theta, \hat{\theta})$. Since the average of a function is always less than or equal to its maximum, we have that $B_{\pi}(\hat{\theta}_0) \leq \sup_{\theta} R(\theta, \hat{\theta}_0)$. Hence,

$$B_{\pi}(\hat{\theta}_0) \leq \sup_{\theta} R(\theta, \hat{\theta}_0) < \sup_{\theta} R(\theta, \hat{\theta}) \leq B_{\pi}(\hat{\theta}) \quad (17.3)$$

which is a contradiction.

A previous student's alternative Proof: Since we have that,

$$R(\theta, \hat{\theta}) \leq B_{\pi}(\hat{\theta}) \quad \text{for all } \theta$$

we can see that,

$$\sup_{\theta} R(\theta, \hat{\theta}) \leq B_{\pi}(\hat{\theta}),$$

where the LHS is an upper bound on the minimax risk, and the RHS is a lower bound on the minimax risk, so it must be the case that $\sup_{\theta} R(\theta, \hat{\theta})$ is equal to the minimax risk, and hence that $\hat{\theta}$ is a minimax estimator.

Theorem 17.2 Suppose that $\hat{\theta}$ is the Bayes estimator with respect to some prior π . If the risk is constant then $\hat{\theta}$ is minimax.

Proof:

The Bayes risk is $B_{\pi}(\hat{\theta}) = \int R(\theta, \hat{\theta})\pi(\theta)d\theta = c$ and hence $R(\theta, \hat{\theta}) \leq B_{\pi}(\hat{\theta})$ for all θ . Now apply the previous theorem.

Example 17.3 Consider the Bernoulli model with squared error loss. We showed previously that the estimator

$$\hat{p} = \frac{\sum_{i=1}^n X_i + \sqrt{n/4}}{n + \sqrt{n}}$$

has a constant risk function. This estimator is the posterior mean, and hence the Bayes estimator, for the prior Beta(α, β) with $\alpha = \beta = \sqrt{n/4}$. Hence, by the previous theorem, this estimator is minimax.