Statistics 210A Lecture 13 Notes

Daniel Raban

October 7, 2021

1 Minimax Estimation

1.1 Bayes risk

If we have a model $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$, then we have a few main ideas for choosing an estimator:

1. Constrain the choice of estimator, e.g. unbiased estimation
2. Minimize average-case risk, i.e. Bayes estimation.

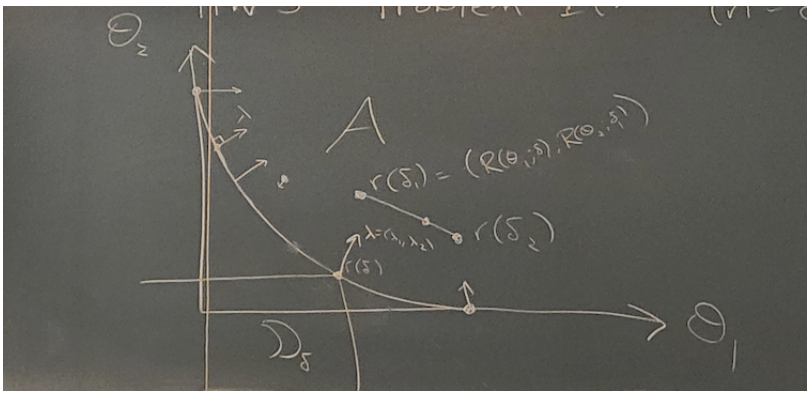
In Bayes estimation, we have a prior Λ with $\Lambda(\Theta) = 1$ (here, Θ is the parameter space). The Bayes estimator (if it exists) minimizes

$$R(\theta; \delta) = \mathbb{E}[L(\theta; \delta(X))].$$

Definition 1.1. The **Bayes risk** for the problem $\Lambda, \mathcal{P}$ is

$$r_\Lambda = \inf_\delta \int R(\theta, \delta)\, d\Lambda(\theta).$$

Example 1.1 (HW 6 Problem 1(c), n=2)**.** In this example, there are only two possible values of θ, θ_1 and θ_2. Then we can plot $r(\delta) = (R(\theta_1; \delta), R(\theta_2; \delta))$.



This is a convex set. The Bayes estimators are the ones on the frontier of this set, the points where the box to the lower left of the point is not in the set. Each vector λ which is normal to this boundary corresponds to a prior.

1.2 Minimax risk, minimax estimators, and least favorable priors

The idea of the minimax risk is to minimize

$$\min_{\delta} \sup_{\theta} R(\theta; \delta).$$

Definition 1.2. The minimal achievable sup-risk is called the **minimax risk**,

$$r^* = \inf_{\delta} \sup_{\theta} R(\theta, \delta),$$

of the problem. An estimator δ^* is **minimax** if it achieves

$$\sup_{\theta} R(\theta, \delta^*) = r^*.$$

There is a game theoretic interpretation: Imagine we pick our δ first, and then nature tries to maximize the risk (i.e. choosing θ adversarially). The interpretation of Bayes estimation is that nature picks θ (via a prior), and then we try to minimize the risk.

For any proper prior Λ, the Bayes risk is

$$\begin{aligned} r_\Lambda &= \inf_{\delta} \int R(\theta; \delta)\, d\Lambda(\theta) \\ &\leq \inf_{\delta} \sup_{\theta} R(\theta; \delta) \\ &= r^*. \end{aligned}$$

Here is the strategy that nature will pick if it can go first.

Definition 1.3. The **least favorable (LF) prior** is the prior distribution Λ^* that gives the best lower bound:

$$r_{\Lambda^*} = \sup_{\Lambda} r_\Lambda.$$

We know that

$$\sup_{\theta} R(\theta; \delta) \geq r^* \geq r_{\Lambda^*} \geq r_\Lambda$$

for any prior Λ. We hope that we can find a prior and an estimator that collapse all these inequalities into equalities.

Theorem 1.1. *If $r_\Lambda = \sup_\theta R(\theta; \delta_\Lambda)$, where δ_Λ is Bayes for Λ, then*

(a) δ_Λ *is minimax.*

(b) If δ_Λ *is the unique Bayes estimator (up to a.s. equality) for* Λ*, then* δ_Λ *is the unique minimax estimator.*

(c) Λ *is the least favorable prior.*

Proof.

(a) For any other δ,

$$\begin{aligned}
\sup_\theta R(\theta;\delta) &\geq \int R(\theta;\delta)\, d\Lambda(\theta) \\
&\geq \int R(\theta;\delta_\Lambda)\, d\Lambda(\theta) \qquad (*) \\
&= r_\Lambda \\
&= \sup_\theta R(\theta;\delta_\Lambda).
\end{aligned}$$

(b) Replace $\geq$ with $>$ in the step $(*)$.

(c) If $\widetilde{\Lambda}$ is any other prior, then

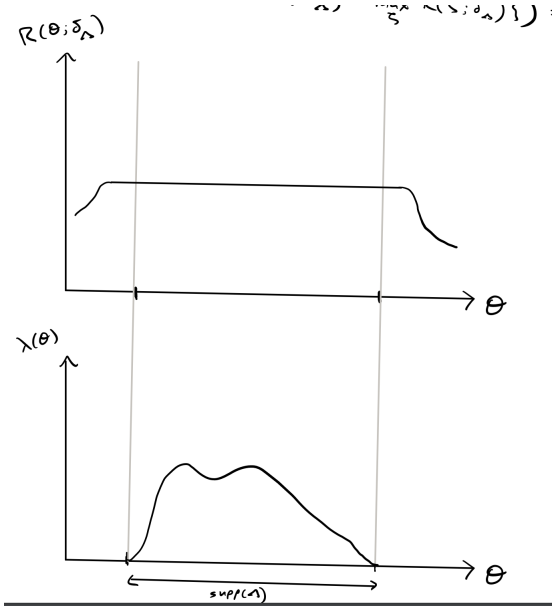
$$\begin{aligned}
r_{\widetilde{\Lambda}} &= \inf_\delta \int R(\theta;\delta)\, d\widetilde{\Lambda} \\
&\leq \int R(\theta;\delta_\Lambda)\, d\widetilde{\Lambda} \\
&\leq \sup_\theta R(\theta;\delta_\Lambda) \\
&= r_\Lambda.
\end{aligned}$$

□

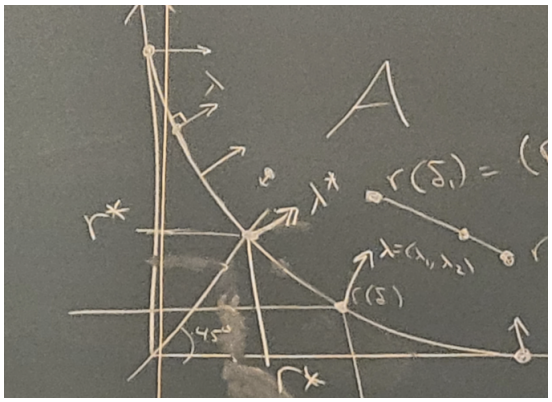
Here are sufficient conditions for a minimax estimator:

1. δ is a Bayes estimator whose risk function is constant.

2. δ_Λ is a Bayes estimator with $1 = \Lambda(\{\theta : R(\theta; \delta_\Lambda) = \max_\zeta R(\zeta; \delta_\Lambda)\})$.



In our picture of Bayes estimation, a 45 degree line denotes the points corresponding to estimators with constant risk. The least favorable prior is the corresponding normal vector at the point where this line reaches the boundary of possible risks.



Example 1.2 (Binomial)**.** Suppose $X \sim \text{Binom}(n, \theta)$ with $\theta \in [0, 1]$. We want to estimate θ using the MSE for our risk. Try $\theta \sim \text{Beta}(\alpha, \beta)$, so the Bayes estimator will be

$$\delta_{\alpha,\beta}(X) = \frac{\alpha + X}{\alpha + \beta + n}.$$

Then the Bayes risk is

$$\begin{aligned}\mathrm{MSE}(\theta;\delta_{\alpha,\beta}) &= \mathbb{E}_\theta\left[\left(\frac{\alpha+X}{\alpha+\beta+n}-\Theta\right)^2\right]\\ &\propto_\theta [(\alpha+\beta)^2-n]\theta^2+[n-2\alpha(\alpha+\beta)]\theta+\alpha^2.\end{aligned}$$

To get a minimax estimator, we want to pick α and β to make this constant in θ. So we set $(\alpha+\beta)^2=n$ and $2\alpha(\alpha+\beta)=n$ and get $\alpha=\beta=\sqrt{n}/2$. So $\mathrm{Beta}(\sqrt{n}/2,\sqrt{n}/2)$ is the least favorable prior.

This is not such a great estimator, however, since it put a lot of weight around $1/2$. So the pessimistic perspective of minimax estimation can lead us astray for some values of θ.

1.3 Least favorable sequences of priors

Example 1.3. Suppose $X\sim N(\theta,1)$, and we are estimating θ with the MSE risk. To find the least favorable prior, we would want a flat prior, but this does not give a probability distribution. So we can take, say, $\Lambda_n=N(0,n)$ as a sequence of priors.

Definition 1.4. As sequence $\Lambda_1,\Lambda_2,\dots$ of priors is **least favorable** if $r_{\Lambda_n}\to\sup_\Lambda r_\Lambda$.

Theorem 1.2. *Suppose $\Lambda_1,\Lambda_2,\dots$ is any sequence of priors, and suppose δ satisfies*

$$\sup_\theta R(\theta;\delta)=\lim_n r_{\Lambda_n}.$$

Then

(a) δ is minimax.

(b) $\Lambda_1,\Lambda_2,\dots$ is least favorable.

Proof.

(a) Suppose $\widetilde{\delta}$ is another estimator. Then for all n,

$$\begin{aligned}\sup_\theta R(\theta;\widetilde{\delta}) &\geq \int R(\theta;\widetilde{\delta})\,d\Lambda_n\\ &\geq r_{\Lambda_n}.\end{aligned}$$

Then

$$\sup_\theta R(\theta;\widetilde{\delta})\geq\lim_n r_{\Lambda_n}=\sup_\theta R(\theta;\delta).$$

(b) If Λ is a prior, then

$$\begin{aligned} r_\Lambda &\le \int R(\theta;\delta)\, d\Lambda \\ &\le \sup_\Theta R(\theta;\delta) \\ &= \lim_n r_{\Lambda_n}. \end{aligned}$$

So we get

$$\lim_n r_{\Lambda_n} = \sup_\Lambda r_\Lambda. \qquad \square$$

Remark 1.1. If we find the Bayes risk, then we get a lower bound on the minimax risk, and if we provide an estimator, we can get an upper bound on the minimax risk. If these are close, this gives a good estimate of the hardness of a problem.

This is not a very useful measure if your parameter space has some bad corner which you never encounter in practice.

1.4 Bayes estimation example: the Gaussian sequence model

Here is an example of Bayes estimation we did not have time to cover before:

Example 1.4 (Gaussian sequence model)**.** Suppose $X \sim N_d(\theta, I_d)$ for $\theta \in \mathbb{R}^d$. Then the Jeffreys prior on θ is flat. The objective Bayes estimator for Θ is X because the posterior distribution is

$$\lambda(\theta \mid X) \propto_\theta p_\theta(X) \propto_\theta e^{-\|X-\theta\|^2/2} \propto_\theta N_d(X, I_d).$$

What about $\rho^2 = \|\Theta\|^2$? Since $\Theta_i \sim N(X_i, 1)$, $\mathbb{E}[\Theta_i^2 \mid X_i] = 1 + X_i^2$, so

$$\widehat{\rho^2} = \mathbb{E}[\|\Theta\|^2 \mid X] = d + \|X\|^2.$$

The UMVU estimator is $\widehat{\rho^2}_{\text{UMVU}} = \|X\|^2 - d$ because

$$\mathbb{E}_\theta[\|X\|^2] = d + \|\theta\|^2.$$

Finally, we have the MLE

$$\widehat{\rho^2}_{\text{MLE}} = \|X\|^2.$$

Which one of these estimators is the best? The UMVU estimator is inadmissible because it is negative, but we may not want to rule it out. These all have the same variance, d, and the UMVU estimator has no bias. This serves as a cautionary tale about constructing

objective priors. Suppose we took the prior $\Theta \sim N(0, n)$, so $\rho^2 \sim n\chi^2_d$. Then picking an "objective prior" may not produce a good result. In this case, $\lambda(\rho^2) \propto_{\rho^2} (\rho^2)^{(d-1)/2}$.