Lecture 24: November 7

Lecturer: Donlapark Pornnopparath

24.1 Bayesian estimator as a point estimator

We go back to the previous framework and assume that θ is a fixed parameter. From the previous lecture, the Bayes estimator for a Bernoulli parameter p with prior Beta(a, b) is given by

$$\delta_B = \left[\frac{n}{n+a+b}\right]\frac{X}{n} + \left[\frac{a+b}{n+a+b}\right]\frac{a}{a+b}$$

which is a weighted average between the UMVU statistic and the prior mean. Thus δ_B is biased unless $a = b = 0$. Since we obtained a similar result for normal conjugates as well, we might we wondering if the Bayes estimator is biased in general. The following proposition gives a condition on which this is the case.

Proposition 24.1. Assume that a parameter θ has a prior $\pi(\theta)$ which gives rise to a Bayesian estimator $\delta(X)$ under the squared error loss. Then it is biased in a sense that the set of θ satisfying $\mathbb{E}[\delta(X)|\theta] \neq g(\theta)$ has positive probability, unless

$$\mathbb{E}[(\delta(X) - g(\theta))^2] = 0,$$

where the expected value is taken with respect to the distributions of X and θ, or equivalently, $\delta(X) = g(\theta)$ with probability 1.

Proof: Recall that the bayes estimator under the squared error loss is given by

$$\delta(X) = \mathbb{E}[g(\theta)|X].$$

Thus, by smoothing,

$$\mathbb{E}[\delta(X)g(\theta)] = \mathbb{E}[\delta(X)\mathbb{E}[g(\theta)|X]] = \mathbb{E}[(\delta(X))^2].$$

By conditioning of θ instead, we have

$$\mathbb{E}[\delta(X)g(\theta)] = \mathbb{E}[g(\theta)\mathbb{E}[\delta(X)|\theta]] = \mathbb{E}[(g(\theta))^2].$$

Therefore,

$$\mathbb{E}[(\delta(X) - g(\theta))^2] = \mathbb{E}[(\delta(X))^2] + \mathbb{E}[(g(\theta))^2] - 2\mathbb{E}[\delta(X)g(\theta)] = 0.$$

■

We can use this proposition to check if an estimator can be a Bayes estimator.

Example 24.2. Let $X \sim N(\mu, \sigma^2)$. Then $\overline{X}$ is a UMVU estimator of μ and we cannot have $\overline{X} = \mu$ with probability one. Therefore, μ is not a Bayes estimator. ◇

We can use a Bayes estimator as a point estimator. We have already seen the posterior mean estimator $\delta_{\text{PM}}(X) = \mathbb{E}[\theta|X]$ and the MAP estimator $\delta_{\text{MAP}}(X) = \arg\max_\theta \pi(\theta|X)$. Another loss function that we could consider is the absolute error loss:

$$L(\theta, \delta) = |\theta - \delta|.$$

Then the Bayes estimator δ, which satisfies $\delta(x) = \min_d \mathbb{E}[|\theta - d| X = x]$, is the median of the posterior distribution.

24.2 Minimax estimator

In this section, we give another way to define an estimator from the loss function. Instead of computing from the average risk, we could look at the maximum risk:

$$\overline{R}(\delta) = \sup_{\theta \in \Theta} R(\theta, \delta)$$

and try to minimize this quantity. An estimator that minimizes the maximum risk is called a *minimax estimator* of θ.

$$\delta_m = \inf_{\delta} \overline{R}(\delta).$$

Generally, it is quite hard to find a minimax estimator. The main goal of this section is to provide a condition on which a Bayes estimator is also a minimax estimator. To prove this, we need the following proposition.

Proposition 24.3. Let δ be the Bayes estimator for some prior π such that

$$R(\theta, \delta) \leq r(\pi, \delta) \quad \text{for all } \theta,$$

then δ is a minimax estimator.

Proof: Suppose that δ is not a minimax, then there exists another estimator δ_0 such that $\overline{R}(\delta_0) < \overline{R}(\delta)$. Since δ minimized the Bayes risk,

$$\begin{aligned} r(\pi, \delta) \leq r(\pi, \delta_0) &= \int R(\theta, \delta_0)\pi(\theta)\ d\theta \\ &\leq \overline{R}(\delta_0) \int \pi(\theta)\ d\theta \\ &< \overline{R}(\delta) \\ &\leq r(\pi, \delta), \end{aligned}$$

which is a contradiction. Therefore, δ is a minimax estimator. ■

Theorem 24.4. Let δ be a Bayes estimator for some prior π. If its risk $R(\theta, \delta) = c$ is constant for all θ, then δ is a minimax estimator.

Proof: Since $r(\theta, \delta) = \int R(\theta, \delta)\pi(\theta)\ d\theta = c = R(\theta, \delta)$ for all $theta$, this follows easily from Proposition 24.3. ■

Example 24.5. Let $X_1, X_2, \ldots, X_n$ be iid Bernoulli(p), $\hat{p}$ be the MLE estimator of p and $\hat{p}_B$ be the Bayes estimator of p for prior Beta(a, b). Recall that

$$\begin{aligned} \hat{p} &= \frac{\sum_i X_i}{n} = \frac{Y}{n} \\ \hat{p}_B &= \frac{Y + a}{a + b + n}. \end{aligned}$$

We can choose a and b so that $\hat{p}_B$ become a minimax estimator under the loss squared error. Thus we compute the risk

$$\begin{aligned} R(p,\hat{p}_B) = \mathbb{E}_p[(\hat{p}_b - p)]^2 &= \text{Var}_p(\hat{p}_B - p) + (\mathbb{E}_p[\hat{p}_B - p])^2 \\ &= \text{Var}_p(\hat{p}_B) + (\mathbb{E}_p[\hat{p}_B - p])^2 \\ &= \text{Var}_p\left(\frac{Y+a}{a+b+n}\right) + \left(\mathbb{E}_p\left[\frac{Y+a}{a+b+n}\right] - p\right)^2 \\ &= \frac{np(1-p)}{(a+b+n)^2} + \left(\frac{np+a}{a+b+n} - p\right)^2 \end{aligned}$$

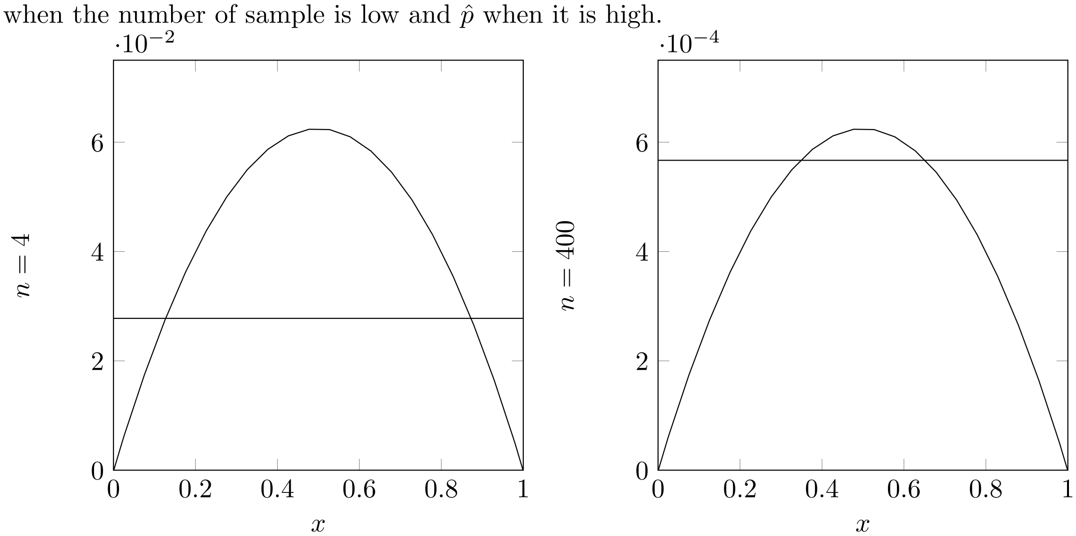
It turns out that if we let $a = b = \sqrt{n/4}$ then the risk becomes a constant.

$$R(p,\hat{p}_B) = \frac{1}{4(\sqrt{n}+1)^2},$$

and so this is a minimax estimator of p. Let us compare this with the risk of $\hat{p}$, which is

$$\mathbb{E}_p[(\hat{p} - p)^2] = \text{Var}_p\left(\frac{Y}{n}\right) = \frac{p(1-p)}{n}.$$

The graph the risk functions below can help us decide which estimator is better. We might want to pick $\hat{p}_B$ when the number of sample is low and $\hat{p}$ when it is high.



◇