STA 250/MTH 342 – Intro to Mathematical Statistics

Lecture 7

More general criteria for selecting "good" estimators

Recall from last time the goal of constructing an estimator $\delta(\mathbf{X})$ so that $\delta(\mathbf{X})$ will *likely to be close to* θ.

- Again, we need a notion of distance between the estimate and the parameter.
- Can again use a loss function just like before such as
  1. The absolute error loss: $L(\theta,a) = |\theta - a|$.
  2. The squared error loss: $L(\theta,a) = (\theta - a)^2$.
  3. The step error loss: $L(\theta,a) = \mathbf{1}(|\theta - a| > \Delta)$.

  So we can again try to choose an estimator that minimizes the *expected loss*.

But now the expectation is taken over the distribution of the estimator given θ:

$$
\begin{aligned}
E(L(\theta,\delta(\mathbf{X}))|\theta) = E_\theta(L(\theta,\delta)) &= \int_{-\infty}^{\infty} L(\theta,u) f_\delta(u|\theta)\,du \\
&= \int_{-\infty}^{\infty} L(\theta,\delta(x)) f(x|\theta)\,dx
\end{aligned}
$$

where $f_\delta(u|\theta)$ denotes the p.d.f of the estimator $\delta(\mathbf{X})$.

This expectation is also called the *risk function* of estimator δ:

$$
R_\delta(\theta) := E_\theta(L(\theta,\delta))
$$

Examples of risk functions

- For absolute error loss

$$\begin{aligned} R_\delta(\theta) = ME_\delta(\theta) &:= E\left[|\delta(X) - \theta||\theta\right] \\ &= E_\theta\left[|\delta - \theta|\right] \\ &= \int_{-\infty}^{\infty} |\delta(x) - \theta| f(x|\theta)\,dx. \end{aligned}$$

  This is called the *mean error (risk)*.

- For squared error loss

$$\begin{aligned} R_\delta(\theta) = MSE_\delta(\theta) &:= E\left[(\delta(X) - \theta)^2|\theta\right] \\ &= E_\theta\left[(\delta - \theta)^2\right] \\ &= \int_{-\infty}^{\infty} (\delta(x) - \theta)^2 f(x|\theta)\,dx. \end{aligned}$$

  This is called the *mean squared error (risk)*.

Note that these expectations are computed using only information available *before the experiment*, not the observed value of the data.

- These are the *average* distances between the estimator $\delta(\mathbf{X})$ and the parameter θ *if the experiment is repeated many times* under fixed parameter value θ. (Recall the "frequentist" viewpoint.)
- Given an observation $\mathbf{X} = \mathbf{x}$, the corresponding estimate is $\delta(\mathbf{x})$. Note that the estimator $\delta(\mathbf{X})$ is chosen using information available *before the experiment*. The estimate given data $\mathbf{X} = \mathbf{x}$ is simply the plug-in value of that estimator.
- Contrast this with the Bayesian approach to estimation, where we use the posterior given the data to find the estimate, and then find the corresponding estimator.
- Under the sampling perspective, we know nothing about how far our realized estimate $\delta(\mathbf{x})$ is from the underlying θ.
- This is a price to pay when treating parameters as fixed quantities.

While the mean error (ME) seems to be the most natural criterion for judging "average closeness", the mean squared error (MSE) is the most popular one to use due to ease of computation.

$$\begin{aligned}
MSE_\delta(\theta) &= E\left((\delta(X)-\theta)^2|\theta\right) = E_\theta(\delta-\theta)^2 \\
&= E_\theta[(\delta - E_\theta(\delta)) + (E_\theta(\delta)-\theta)]^2 \\
&= E_\theta(\delta - E_\theta(\delta))^2 + 2(E_\theta(\delta)-\theta)E_\theta(\delta - E_\theta(\delta)) + (E_\theta(\delta)-\theta)^2 \\
&= E_\theta(\delta - E_\theta(\delta))^2 + (E_\theta(\delta)-\theta)^2 \\
&= \mathrm{Var}_\theta(\delta) + B_\delta(\theta)^2.
\end{aligned}$$

This is sometimes referred to as the *Bias-Variance trade-off*. We want estimators that strike a balance between small bias and small variability.

- It may be worth it to take a biased estimator if its variance is much smaller than an alternative unbiased one. (Draw a figure.)

Example: Political poll

We have data $X \sim \text{Binomial}(n, \theta)$. Let us now consider three estimators for θ.

$$\delta_1(X) = \frac{X}{n} \quad \text{(The sample mean. It's unbiased.)}$$

$$\delta_2(X) = \frac{1}{2} \quad \text{(A "stubborn" estimator. It has zero variance.)}$$

$$\delta_3(X) = \frac{X+12}{n+24} \quad \text{(What esimator is this?)}$$

$$\delta_4(X) = \frac{X+1}{n+2} \quad \text{(What estimator is this?)}$$

Recall that the Bayes estimators δ_3 and δ_4 are weighted averages of δ_1 and δ_2. What are the weights?

Next, let us evaluate and compare these three esimators in their mean squared error.

$\delta_1(X) = \frac{X}{n}$ (The sample mean.)

$$\begin{aligned} MSE_{\delta_1}(\theta) &= E_\theta(\delta_1(X) - \theta)^2 = E_\theta\left(\frac{X}{n} - \theta\right)^2 \\ &= \mathrm{Var}_\theta\left(\frac{X}{n}\right) = \frac{\mathrm{Var}_\theta(X)}{n^2} = \frac{n\theta(1-\theta)}{n^2} = \frac{\theta(1-\theta)}{n}. \end{aligned}$$

$\delta_2(X) = \frac{1}{2}$ (A "stubborn" estimator.)

$$MSE_{\delta_2}(\theta) = E_\theta\left(\frac{1}{2} - \theta\right)^2 = \left(\frac{1}{2} - \theta\right)^2 = \frac{1}{4} - \theta(1-\theta).$$

- Note that this is $B_{\delta_2}(\theta)^2$ as this estimator has variance 0.

$\delta_3(X) = \frac{X+12}{n+24}$: Bayes estimator under Beta(12,12) prior.

$$
\begin{aligned}
MSE_{\delta_3}(\theta) &= E_\theta\Big(\delta_3(X) - \theta\Big)^2 = \mathrm{Var}_\theta(\delta_3) + (B_{\delta_3}(\theta))^2 \\
&= \mathrm{Var}_\theta\Big(\frac{X+12}{n+24}\Big) + \Big(\frac{12-24\theta}{n+24}\Big)^2 \\
&= \frac{\mathrm{Var}_\theta(X)}{(n+24)^2} + \frac{144(1-2\theta)^2}{(n+24)^2} \\
&= \frac{n\theta(1-\theta) + 144(1-2\theta)^2}{(n+24)^2}
\end{aligned}
$$

Note that

$$\delta_3 = \frac{n}{n+24}\delta_1 + \frac{24}{n+24}\delta_2.$$

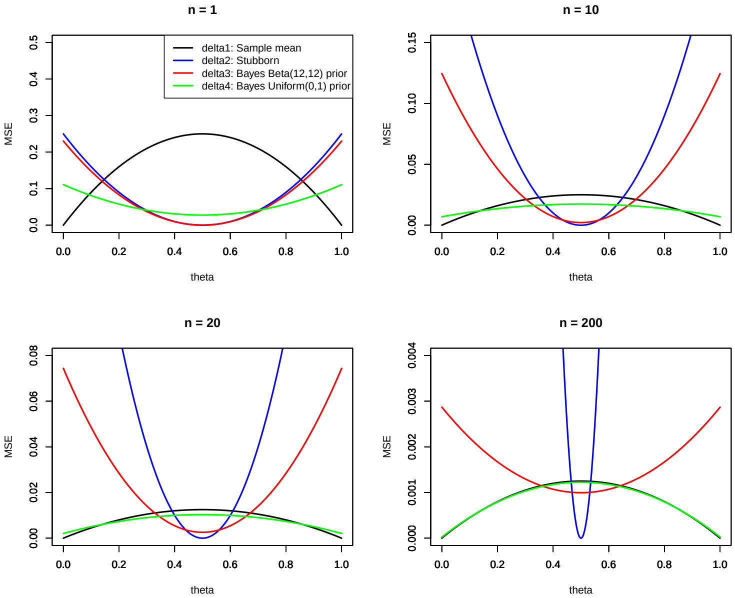
$\delta_4(X) = \frac{X+1}{n+2}$: Bayes estimator under Uniform(0,1) prior.

$$MSE_{\delta_4}(\theta) = \frac{n\theta(1-\theta)+(1-2\theta)^2}{(n+2)^2} \quad \text{(Exercise.)}$$

Note that

$$\delta_4 = \frac{n}{n+2}\delta_1 + \frac{2}{n+2}\delta_2.$$

Plots of MSE(θ) for the four estimators



- There is no champion over all possible values of θ.
- The range of θ values over which the stubborn estimate does better than the sample mean shrinks as n increases. The more data you have, the more costly it is to ignore them.
- The Bayes estimators are a compromise between sample mean and the stubborn estimator. Depending on the strength of the prior belief, the Bayes estimators will be closer to the sample mean or the stubborn estimator.
- For the binomial experiment, if no prior information is incorporated, then it is "hardest" to estimate θ is when θ is about 1/2.
- Bayes estimators tend to behave better for θ values that is likely according to prior knowledge.

For unbiased estimators,

$$MSE_\delta(\theta) = \text{Var}(\delta(X)|\theta) = \text{Var}_\theta(\delta(X)).$$

Thus the "champion" among all unbiased estimators is the one with the smallest variance for *all* θ. Such a champion may not exist, but when it does, it is called the *"Minimum Variance Unbiased Estimator" (MVUE)*.

- We define the *standard error* of an estimator $\delta(X)$ to be

$$\sigma_\delta = \sqrt{\text{Var}_\theta(\delta(X))}.$$

- Many of the "good" estimators (in the sampling theory sense) we will encounter are slightly biased. (But they tend to have small bias and variance, and hence MSE, over all possible values of θ.)

Next . . .

- We will learn a general procedure to construct such "good" estimators under the sampling view.