

Assignment 2

Harry Bendekgey
Math 153: Bayesian Statistics

February 8 2018

1. We define the *bias* of an estimator $\hat{\theta}$ as $BIAS = E(\hat{\theta}) - \theta$. Show that the mean (expected) squared error (MSE) of an estimator is the variance plus the squared bias, i.e.

$$E[(\hat{\theta} - \theta)^2] = Var(\hat{\theta}) + BIAS(\hat{\theta})^2$$

We start with the definition of variance and bias:

$$Var(\hat{\theta}) + BIAS(\hat{\theta})^2 = E[(\hat{\theta} - E(\hat{\theta}))^2] + (E(\hat{\theta}) - \theta)^2$$

We then expand:

$$= E[\hat{\theta}^2 - 2\hat{\theta}E(\hat{\theta}) + E(\hat{\theta})^2] + E(\hat{\theta})^2 - 2\theta E(\hat{\theta}) + \theta^2$$

Because θ^2 is a constant, we can replace it with $E(\theta^2)$. Similarly we can pull the 2θ into the third expected value:

$$= E[\hat{\theta}^2 - 2\hat{\theta}E(\hat{\theta}) + E(\hat{\theta})^2] + E(\hat{\theta})^2 + E(-2\theta\hat{\theta}) + E(\theta^2)$$

We rearrange to get:

$$= E[\hat{\theta}^2 - 2\theta\hat{\theta} + \theta^2] + E(-2\hat{\theta}E(\hat{\theta})) + E(E(\hat{\theta})^2) + E(\hat{\theta})^2$$

We note that $E(\hat{\theta})^2$ is just a constant, so can be pulled out of the outer expected value. Similarly, $-2E(\hat{\theta})$ is a constant and can be pulled out. We get:

$$= E[\hat{\theta}^2 - 2\theta\hat{\theta} + \theta^2] + -2E(\hat{\theta})E(\hat{\theta}) + E(\hat{\theta})^2 + E(\hat{\theta})^2$$

The last three terms cancel, so we get:

$$= E[\hat{\theta}^2 - 2\theta\hat{\theta} + \theta^2] = E[(\hat{\theta} - \theta)^2]$$

2. Consider the Bayes' estimator of the form $\hat{p}_b = \frac{X+10}{n+20}$, where $X = \sum X_i$ is our binomial random variable (sum of Bernoulli RVs X_i). Calculate the bias

and variance of the estimator (recall that $Var(aX + b) = a^2 Var(X)$). Recall from Lab 1, we showed that the MLE was unbiased and were given that the variance was $\sigma_{\hat{p}_{MLE}}^2 = \frac{p(1-p)}{n}$. Mathematically justify what you saw in the comparison of these two estimators.

$$Var(\hat{p}_b) = Var\left(\frac{x}{n} \frac{n}{n+20} + \frac{10}{n+20}\right) = \left(\frac{n}{n+20}\right)^2 Var\left(\frac{x}{n}\right) = \left(\frac{n}{n+20}\right)^2 \frac{p(1-p)}{n}$$

This makes sense compared to the variance of the MLE estimator: this estimator varies less at small sample sizes, because we're pushing the estimation closer to $\frac{1}{2}$. As n goes to infinity, the term on the left goes to 1, resulting in the same variance as the MLE estimator.

$$\begin{aligned} BIAS(\hat{p}_b) &= E(\hat{p}_b) - p = E\left(\frac{x}{n} \frac{n}{n+20} + \frac{10}{n+20}\right) - p = \frac{n}{n+20}p + \frac{10}{n+20} - p \\ &= \left(\frac{n}{n+20} - 1\right)p + \frac{10}{n+20} = \frac{-20p}{n+20} + \frac{10}{n+20} = 10 \frac{1-2p}{n+20} \end{aligned}$$

This also makes sense, because as p approaches 0.5, the bias becomes 0. As it deviates from 0.5, however, we get potential for a bias.

3. Consider a more general Bayes' estimator, this time of the form $\hat{p}_b = \frac{X+a}{n+2a}$. Find the value of a that makes the MSE constant (doesn't depend on the value of p).

Using the exact same math as the previous problem, but with a and $2a$ instead of 10 and 20, (and which I have not included for brevity's sake) we see:

$$Var(\hat{p}_b) = \left(\frac{n}{n+2a}\right)^2 \frac{p(1-p)}{n}$$

$$BIAS(\hat{p}_b) = a \frac{1-2p}{n+2a}$$

We know that

$$MSE(\theta) = Var(\hat{\theta}) + BIAS(\hat{\theta})^2 = \left(\frac{n}{n+2a}\right)^2 \frac{p(1-p)}{n} + \left(a \frac{1-2p}{n+2a}\right)^2$$

We want this to be constant with regards to p , so we want the derivative with respect to p to be 0. We set:

$$\frac{\partial}{\partial p} \left[\left(\frac{n}{n+2a}\right)^2 \frac{p(1-p)}{n} + \left(a \frac{1-2p}{n+2a}\right)^2 \right] = 0$$

$$\frac{n^2}{(n+2a)^2} \frac{1-2p}{n} - 4a^2 \frac{1-2p}{(n+2a)^2} = 0$$

The n cancels on the left, and we factor out $1 - 2p$ to get:

$$\frac{(1 - 2p)(n - 4a^2)}{(n + 2a)^2} = 0$$

Thus, this is true if $(n - 4a^2) = 0$, or equivalently $a = \frac{\sqrt{n}}{2}$

4. Show that the value λ that minimizes $E[(X - \lambda)^2]$ is given by $\lambda = E(X)$.

We set the derivative to 0 to find the minimum (which is obviously a minimum because it does not make sense for a maximum to exist)

$$\frac{\partial}{\partial \lambda} E[(X - \lambda)^2] = \frac{\partial}{\partial \lambda} E[X^2 - 2X\lambda + \lambda^2]$$

We know that the derivative of an expected value is the expected value of the derivative (because expected values are just integrals) so we get:

$$= E\left(\frac{\partial}{\partial \lambda}[X^2 - 2X\lambda + \lambda^2]\right) = E(2\lambda - 2X) = 2\lambda - 2E(X)$$

Thus for the derivative to be 0, $\lambda = E(X)$.

5. Consider a row of shops on a street. You have several errands to run, and after visiting each of the several shops that you must go to, you return to your car to drop off your purchased goods before proceeding to the next store. Visualize this problem using the number line and make an argument for where you should park on the street. The shops you visit are not necessarily evenly spaced (as some shops are bigger than others, and there are shops on the street that you will not patronize today).

We can consider the place we park P as well as the location of each shop S_i to be a point on a number line, in which case we're trying to minimize $\sum |P - S_i|$. This is hard to work with because we're used to dealing with squared errors, and absolute values are hard to differentiate. I propose a different approach

Consider any spot on the number line. How many times are you going to have to walk past that point on your journey? Well if you park to the left of it, then the number of times you cross it is equal to the number of shops to its right (times two of course, counting the return journey). If you park to the right of it, then the number of times you cross it is equal to the number of shops to its left. If we want to minimize the total distance walked, we want to minimize the number of times we walk past all points on the number line, meaning that we want to park to the left of all spots that have more shops to their left and to the right of all spots that have more shops to their right. The point with this property is the median shop (or, in the case that there is an even number of shops, anywhere between the two middle-most shops).