Vincent's CS70 Discussion 12B Notes

April 21, 2016

1 A Quick Review of Confidence Intervals

1.1 The main idea

We once again return to our favorite example (flipping coins). Suppose we are given a coin of unknown bias. The coin has some true probability p of landing on heads for any given flip, and we want to estimate what p may be. From now on, we will let $\hat{p}$ denote our estimate of p.

How can we go about finding a good guess for $\hat{p}$? Recall that the Law of Large Numbers gives us that if X is the number of heads that we get in n flips of our coin (so $X \sim Binom(n, p)$) , then

$$\lim_{n\to\infty} \frac{1}{n}X = p.$$

It follows that as long as we take n to be sufficiently large, letting $\hat{p} = \frac{1}{n}X$ be our estimation should work out pretty well for us.

1.2 "Sufficiently large"?

It's necessary to introduce a few more variables to formalize what we mean by a "good" estimation. Suppose we want our estimate to be within some maximum error $\varepsilon \in (0, 1)$ of the true value, and we want to bound the probability of being incorrect above by $\delta \in (0, 1)$. More formally, we want

$$\mathbb{P}(|\hat{p} - p| \geq \varepsilon) \leq \delta \iff \mathbb{P}(|\hat{p} - p| < \varepsilon) \geq 1 - \delta.$$

It turns out that the left hand side of the equivalence above is more useful to us, so let's proceed to examine it. By Chebyshev's inequality and the fact that $x(1 - x)$ is maximal at $x = 1/2$, we have that

$$\mathbb{P}(|\hat{p} - p| \geq \varepsilon) \leq \frac{Var(\hat{p})}{\varepsilon^2} = \frac{Var(\frac{1}{n}X)}{\varepsilon^2} = \frac{Var(X)}{n^2\varepsilon^2} = \frac{np(1-p)}{n^2\varepsilon^2} = \frac{p(1-p)}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2}.$$

By the transitivity of $\leq$, it follows that we obtain our desired conditions if

$$\frac{1}{4n\varepsilon^2} \leq \delta \implies n \geq \frac{1}{4\delta\varepsilon^2}.$$

2 Covariance

2.1 The definition and where it comes from

Recall that we proved some time ago that if X and Y are two independent random variables, then

$$Var(X+Y) = Var(X) + Var(Y).$$

In the more general case, a straightforward computation shows that

$$\begin{aligned} Var(X+Y) = E[(X+Y-E[X+Y])^2] &= E[((X-E[X]) + (Y-E[Y]))^2] \\ &= E[(X-E[X])^2] + E[(Y-E[Y])^2] + 2E[(X-E[X])(Y-E[Y])] \\ &= Var(X) + Var(Y) + 2E[(X-E[X])(Y-E[Y])] \end{aligned}$$

This brings us to the following definition.

Definition 1. *The covariance of two random variables X and Y is defined to be*

$$Cov(X,Y) = E[(X-E[X])(Y-E[Y])]$$

We can now write $Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$ for any random variables X and Y, and an immediate result is that if X and Y are independent random variables, then $Cov(X,Y) = 0$.

A good exercise is to consider the converse of this statement: If X, Y are two random variables with $Cov(X,Y) = 0$, then is it always true that X and Y are independent?

2.2 Computing covariances

Similarly to how we have that it is often easier to compute the variance of a random variable X by the equation $Var(X) = E[X^2] - (E[X])^2$ instead of directly by the definition, we also have a computationally convenient form for covariance.

Theorem 1. *For two random variables X and Y, $Cov(X,Y) = E[XY] - E[X]E[Y]$*

Proof. By the definition of covariance and linearity of expectation, we have

$$Cov(X,Y) = E[(X-E[X])(Y-E[Y])] = E[XY - XE[Y] - E[X]Y + E[X]E[Y]]$$

$$= E[XY] - 2E[X]E[Y] + E[X]E[Y] = E[XY] - E[X]E[Y]$$

□

As a quick exercise, see if you can use the result above to find an alternate proof of the fact that if X and Y are independent, then $Cov(X,Y) = 0$ (Hint: citing a previous result involving $E[XY]$ makes this proof a one-liner).

3 Estimation problems

3.1 A guessing game

Suppose we want to guess the value of some random variable Y. For example, suppose we are randomly choosing people from a population and we want to guess the chosen person's income.

Without any other information, a good first guess is to say that the random variable will take on value $E[Y]$ (in this case the average income within our population). To justify why this is the case, we show that setting $a = E[Y]$ minimizes $E[(Y-a)^2]$. We give a proof different from the one in lecture note 26. Consider

$$E[(Y-a)^2] = E[Y^2] - 2aE[Y] + a^2.$$

Treating $E[(Y-a)^2]$ as a function of a, we differentiate with respect to a and set our function equal to 0 to see that $2a - 2E[Y] = 0 \implies a = E[Y]$ is either a maximum or minimum point of $E[(Y-a)^2]$.

As the second derivative of $E[(Y-a)^2]$ is

$$\frac{d^2}{da^2}\left(E[Y^2] - 2aE[Y] + a^2\right) = 2 > 0$$

it follows that our critical point is indeed a minimum.

3.2 The LLSE of Y given X

Now, what happens if we have a bit more information? To be more precise, we want to guess the value of Y given the value of some other random variable X. Continuing with our example from above, suppose we are still trying to guess a person's income Y, but we are also given the value X of the person's car. In many cases, we would be wise to guess a person driving a new Porsche makes more money than a person driving a 15 year old Camry.

One thing that we can try, given X, is to estimate that Y is equal to $\hat{Y} = a + bX$ for some constants $a, b \in \mathbb{R}$. Of course, we could just as well guess that Y is a quadratic, cubic, or some other arbitrary function of X, but let's start assuming that it is linear. Our problem has now become to find the values of a, b that minimize $E[(Y-a-bX)^2]$. This brings us to the following definition.

Definition 2. *The Linear Least Squares Estimate of Y given X, written $LLSE[Y|X]$, is the function $a+bX$ that minimizes $E[(Y-a-bX)^2]$.*

A bit more calculus allows us to find an explicit formula for the $LLSE$ of Y given X, and the proof is quite similar to the proof that $E[(Y-a)^2]$ is minimized at $a = E[Y]$ given above.

(NOTE: the proof below is included for completeness but can be treated as "bonus material" as this course does not assume knowledge of multivariable calc. Of course, while fully understanding the proof is optional, we expect you to know the result itself.)

Theorem 2. *Given two random variables X and Y, we have that*

$$LLSE[Y|X] = E[Y] + \frac{Cov(X,Y)}{Var(X)}(X - E[X])$$

Proof. We want to choose a, b to minimize $E[(Y - a - bX)^2]$. Viewing our expression as a function of a and b and expanding it, we consider

$$E[(Y - a - bX)^2] = E[Y^2 + a^2 + b^2X^2 - 2aY - 2bXY + 2abX]$$
$$= E[Y^2] + a^2 + b^2E[X^2] - 2aE[Y] - 2bE[XY] + 2abE[X]$$

To minimize this function, we take partial derivatives with respect to a and b and set them equal to zero to obtain

$$\frac{\partial}{\partial a}(E[(Y - a - bX)^2]) = 2a - 2E[Y] + 2bE[X] = 0$$
$$\frac{\partial}{\partial b}(E[(Y - a - bX)^2]) = 2bE[X^2] - 2E[XY] + 2aE[X] = 0$$

Equivalently,

$$a + bE[X] = E[Y], \text{ and}$$
$$aE[X] + bE[X^2] = E[XY].$$

Solving the system of equations above and recalling that $Var(X) = E[X^2] - E[X]^2$ and $Cov(X,Y) = E[XY] - E[X]E[Y]$ gives us that

$$a = E[Y] - \frac{Cov(X,Y)}{Var(X)}E[X], \quad b = \frac{Cov(X,Y)}{Var(X)},$$

and it is also easily shown using the second partial derivative test that this critical point is indeed a minimum. Thus,

$$LLSE[Y|X] = a + bX = E[Y] - \frac{Cov(X,Y)}{Var(X)}E[X] + \frac{Cov(X,Y)}{Var(X)}X$$
$$= E[Y] + \frac{Cov(X,Y)}{Var(X)}(X - E[X])$$

□