Math 218 Mathematical Statistics

Prof. D. Joyce, Clark University

9 Feb 2009

Due Today. From page 229, exercises 1, 2ab, 5, 7.

Due Wednesday. From page 658, exercises 1, 3.

First test. Feb. 20.

Last meeting. Finished discussion of point estimators started on the previous meeting and discussed maximum likelihood estimators for discrete distributions.

Next time. Finish maximum likelihood estimators and start confidence intervals. Read section 6.2.

Today. We'll look at maximum likelihood estimators in section 15.1.

Likelihood functions for discrete distributions and maximum likelihood estimators. The setting is that we have a family of distributions parametrized by θ, and we run an experiment to get outcome values $\mathbf{x} = (x_1, \ldots, x_n)$. From this data we want to decide what the value of θ is. The idea of the maximum likelihood estimator is that the best value of θ is the one makes the probability of the outcome the greatest.

In some cases, it's pretty easy to see what the maximum likelihood estimator is. Let's take the Bernoulli case where the unknown parameter of success is p. (So, in this case the parameter θ is p.) Suppose we get a sample of $n = 100$ trials, and 47 of them turn out success. What value of p maximizes the probability

$$P(47 \text{ successes out of 100 trials} \,|\, p)?$$

Intuitively, it's $p = 0.47$, and that's the maximum likelihood estimator. We'll verify that guess is correct after stating some general definitions and principles.

The likelihood of a parameter θ for a given random sample $\mathbf{X} = \mathbf{x}$, that is, $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$, is the probability

$$P(\mathbf{X} = \mathbf{x} \,|\, \theta)$$

but it's denoted

$$L(\theta \,|\, \mathbf{x}).$$

Thus, likelihood is not a probability, but the reverse of a conditional probability. That probability can also be written as the product of values of the probability mass function f as

$$f(x_1 \,|\, \theta) f(x_2 \,|\, \theta) \cdots f(x_n \,|\, \theta).$$

Then the maximum likelihood estimator $\hat{\theta}$ for the unknown parameter θ is just that value of θ that has the highest probability of that outcome, that is, has the greatest likelihood.

So in our Bernoulli example with 47 successes,

$$L(p|\ 47 \text{ successes of } 100) = p^{47}(1-p)^{53}).$$

We want to find the value $\hat{p}$ which maximizes $p^{47}(1-p)^{53}$. We can use calculus to do that. Take the derivative with respect to p, $\frac{d}{dp} p^{47}(1-p)^{53}$, and set that derivative to 0 to find the critical points. There is an easier way. The likelihood function $L(P|\mathbf{x})$ has its maximum at the same place as its natural log $\ln L(P|\mathbf{x})$ does since the log function is an increasing function, and it's easier to take the derivative of the log

$$\begin{aligned}\frac{d}{dp}\ln(p^{47}(1-p)^{53}) &= \frac{d}{dp}(47\ln p + 53\log(1-p)) \\ &= \frac{47}{p} - \frac{53}{1-p}\end{aligned}$$

If we set that to 0 to find the critical points, and simplify the equation, we get $\frac{47}{p} = \frac{53}{1-p}$, so $47(1-p) = 53p$, and so $p = \frac{47}{100}$. Thus, the maximum likelihood estimator is $\hat{p} = 0.47$ just as we expected.

In more complicated cases, we can't see right off what value of θ will maximize the likelihood $L(\theta|\mathbf{x})$, and we'll have to resort to the method above where we take logarithmic derivatives to find $\hat{\theta}$.