

# Math 3215: Lecture 19

Will Perkins

March 27, 2012

## 1 Confidence Intervals

**The Situation:** We're sampling iid copies from a distribution  $X$ , where we know the *form* (say, normal, coin flips, Poisson etc) but don't know the underlying *parameter* (the mean, the variance,  $p$ , etc.).

**The Solution:** Compute an estimator of the unknown parameter (usually an unbiased estimator, often the maximum likelihood estimator).

**What We Report:** We report the estimator of course, but we would also like to give some idea about how confident we are in the estimate. The larger our sample size, the more confident we should be in our estimate of the unknown parameter. So we pick a confidence level, say 99% or 95%, and report a *confidence interval*.

### 1.1 The Slightly Confusing Thing About Confidence Intervals

We have to remember that the true value of the parameter,  $\mu$  or  $\sigma^2$ , is fixed and not random. What's random is the samples we draw to estimate the parameter. Each sample of size  $n$  can give a different value of the estimator, and thus a different confidence interval. So what is a 99% confidence interval? The interval itself is a random variable, depending on our randomly drawn samples, and with probability  $\geq .99$ , the random interval contains the true value of the parameter.

To avoid saying this every time we report a confidence interval, and to avoid saying something wrong, we say that the true parameter is contained in our confidence interval with *confidence* (not probability!) .99.

## 2 Confidence Intervals for Means

### 2.1 Illustrative but less useful case: known variance

Say we are sampling  $X_1, \dots, X_n$  from a normal distribution with unknown mean  $\mu$  but known variance  $\sigma^2 = 4$ . We know that the MLE of  $\mu$  is  $\bar{X} = \frac{X_1 + \dots + X_n}{n}$ . What is a .99% confidence interval for  $\mu$ ?

By the CLT,

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

use this to find the confidence interval.

## 2.2 Useful case: unknown variance

Usually we don't know  $\sigma^2$ . In this case, we can use an estimate for  $\sigma^2$  as we construct the confidence interval for  $\mu$ . We saw last time that an unbiased estimator for  $\sigma^2$  is:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

Here we use a distribution called the Student's  $t$ -distribution. Fact:

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim T_n$$

where  $T_n$  is a  $t$ -distribution with  $n - 1$  degrees of freedom. When  $n$  is large  $T_n \sim N(0, 1)$ .

## 2.3 One-sided confidence intervals

What if we just want to say that with confidence 99%,  $\mu \geq A$ ? Show how we can construct such a one-sided confidence interval.

# 3 Confidence Intervals for the Difference Between two means

Here's a situation that comes up a lot in statistics.

We take random samples from two different populations, and take some measurement of each. Let's say we are measuring the length of rats in New York and the length of squirrels in Georgia. Let  $X_1, \dots, X_n$  be the length of the random rats, and  $Y_1, \dots, Y_m$  be the length of the random squirrels. (notice that  $m$  and  $n$  can be different). We could ask:

- Are the rats longer than the squirrels on average?
- How much longer?
- How confident are we of our answers?

In this case we don't want to estimate just one mean, but instead the difference between the two means,  $\mu_X - \mu_Y$ . The following questions will help you work through this problem:

1. Find an unbiased estimator for  $\mu_X - \mu_Y$ .
2. What is the variance of this estimator?
3. Assuming both  $n$  and  $m$  are large so that we can use a normal approximation, find a 95% confidence interval for  $\mu_X - \mu_Y$