Math 218 Mathematical Statistics

Prof. D. Joyce, Clark University

13 Feb 2009

Due Monday. Page 230: 11, 13a, 14, 15ab.

Friday. First test on chapters 3, 4, and 5.

Next meeting. Read section 6.3 on hypothesis tests.

Last meeting. Introduced confidence intervals.

Today. We'll look at an example and discuss some of the difficulties of interpreting the meaning of confidence intervals. Apply interval estimates to bent coins. Introduce hypothesis tests.

Interval estimates for Bernoulli distributions. Suppose we have a bent coin with unknown probability p of heads, so the unknown probability of tails is $q = 1 - p$. The mean of this distribution is $\mu = p$, and its variance is $\sigma^2 = pq$.

The sample mean $\overline{X}$, which is the fraction of heads that occur in n trials, has mean $\mu_{\overline{X}} = \mu = p$, variance $\sigma^2_{\overline{X}} = \sigma^2/n = pq/n$, and standard deviation $\sigma_{\overline{X}} = \sigma/\sqrt{n} = \sqrt{pq/n}$.

If n is large, then $\overline{X}$ is approximately normal, so we can apply the results of our last discussion on confidence intervals. We found that

$$P\left(\overline{X} - 1.96\,\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} + 1.96\,\frac{\sigma}{\sqrt{n}}\right) = 0.95.$$

So that a 95% confidence interval for μ is

$$\left[\overline{X} - 1.96\,\frac{\sigma}{\sqrt{n}}, \overline{X} + 1.96\,\frac{\sigma}{\sqrt{n}}\right].$$

We don't know what σ is, but we do know $\sigma^2 = pq = p(1-p)$. Since p is between 0 and 1, therefore $p(1-p)$ is between 0 and $\frac{1}{4}$, and the maximum $\frac{1}{4}$ occurs when $p = \frac{1}{2}$. Therefore, σ^2 is between 0 and $\frac{1}{4}$, so σ is between 0 and $\frac{1}{2}$.

Thus, if we replace $\left[\overline{X} - 1.96\,\frac{\sigma}{\sqrt{n}}, \overline{X} + 1.96\,\frac{\sigma}{\sqrt{n}}\right]$, by $\left[\overline{X} - 1.96\,\frac{1}{2\sqrt{n}}, \overline{X} + 1.96\,\frac{1}{2\sqrt{n}}\right]$, we will have an interval that contains a 95% confidence interval no matter what the value of σ is. Since 1.96 is about 2, therefore

$$P(\overline{X} - 1/\sqrt{n} \le \mu \le \overline{X} + 1/\sqrt{n}) \ge 0.95,$$

so $[\overline{X} - 1/\sqrt{n}, \overline{X} + 1/\sqrt{n}]$ includes the unknown $p = \mu$ at least 95% of the time. Note that the length of this interval is $2/\sqrt{n}$.

Now suppose we have that bent coin with unknown p and we want to estimate p to one digit, with 95% confidence. The phrase "to within one digit" is usually interpreted to mean within 0.05, and that means the length of the interval is 0.1. How many times to we have to flip the coin? We want $2/\sqrt{n} = 0.1$, so that means $n = 400$. Thus, we've justified the rule of thumb that to get one digit of accuracy for the probability of success p, 400 trials are needed. To get two digits of accuracy, 40000 trials are needed, and that's an awful lot of coin flips.

Introduction to hypothesis tests. As stated in our text, it is no an exaggeration to say that, for better or worse, hypothesis testing is the most widely used statistical tool in practice. Unfortunately, it's also one of the most misunderstood and misused of tools.

In this introduction to hypothesis tests, we'll only consider hypothesis tests concerning the population mean μ, but in later chapters we'll look at hypothesis tests that concern other parameters such as σ^2.

For a hypothesis test, we assume the population distribution comes from a known family of distributions, but an unknown mean μ. We have under consideration two hypotheses concerning the value of μ. One hypothesis, H_0, is called the *null* hypothesis, the other, H_1, is called the *alternative hypothesis*. A test is designed to determine whether to reject or not reject H_0 at some prespecified confidence level. (In practice, these tests are designed to show that the null hypothesis is false.) After the test is performed, there are two possible results, either the data strongly contradict H_0, in which case we reject H_0 and accept H_1, or the data are consistent with H_0, in which case we don't reject H_0. In the second case, not rejecting H_0 does not mean we accept H_0 or reject H_1 as the data may not be strong enough for those conclusions.

There are different forms for these hypotheses. Here are four of them. In each, μ_0 is some specified constant.

- Single population. The mean is μ_0.

  H_0: $\mu = \mu_0$; H_1: $\mu \neq \mu_0$.

  This form will require a two-sided test.

- Single population. The mean is at most μ_0.

  H_0: $\mu \leq \mu_0$; H_1: $\mu > \mu_0$.

  This form will require a one-sided test.

  Of course, there's an analogous one-sided test to see if the mean is at least μ_0.

- Two populations. The means of the two populations are the same.

  H_0: $\mu_1 = \mu_2$; H_1: $\mu_1 \neq \mu_2$.

  This form will require a two-sided test.

- Two populations. The mean of the first population is less than or equal to the mean of the second population.

  H_0: $\mu_1 \leq \mu_2$; H_1: $\mu_1 > \mu_2$.

  This form will require a one-sided test.

We'll look at a couple of examples in the text.

The form of the hypothesis test for the mean is usually to evaluate the sample mean $\overline{X}$ and determine whether $\overline{X}$ falls in a *rejection region* or its complement, an *acceptance region*. The boundaries between these two regions are called *critical constants*. In a one-sided test, there is only one critical constant and each of the regions is a half-infinite interval. In a two-sided test, there are two constants, the acceptance region is the interval between them, and the rejection region is the union of two half-infinite intervals.

A test for a fair coin. Let's design a test for a fair coin. We want a test that will reject or not reject the null hypothesis H_0 that the coin is fair, and let's choose the confidence level to be 95%. The alternative hypothesis H_1 is that the coin is not fair.

So, H_0 is that p, which is μ, equals $\frac{1}{2}$, while H_1 is that it doesn't. Now, we know that for large n,

$$P(\overline{X} - 1/\sqrt{n} \leq p \leq \overline{X} + 1/\sqrt{n}) \geq 0.95.$$

We developed this probability above when we were looking at confidence intervals. We found a 95% confidence interval for p was the interval

$$[\overline{X} - 1/\sqrt{n}, \overline{X} + 1/\sqrt{n}].$$

Hypothesis tests are directly related to confidence intervals, and we can turn this confidence interval into this hypothesis test:

> Reject the null hypothesis H_0 that $p = \frac{1}{2}$ in favor of the alternative hypothesis H_1 if $\frac{1}{2} \notin [\overline{X} - 1/\sqrt{n}, \overline{X} + 1/\sqrt{n}]$.

In other words, we conclude, at the 95% confidence level, that the coin is unfair if $\overline{X}$ is further from $\frac{1}{2}$ than $1/\sqrt{n}$.

How big does n have to be to make this conclusion? You get to decide, but n shouldn't be too small, or it won't be possible to conclude the coin is unfair. For instance, if you take $n = 4$, the test says never says that the coin is unfair, since $\frac{1}{2}$ is never further from $\frac{1}{2}$ than $1/\sqrt{4}$. But suppose you let $n = 100$. Then you could say the coin is unfair if $\overline{X}$ lies outside the interval $[0.4, 0.6]$. That's a pretty big interval, but with $n = 10000$, you could say that the coin is unfair if $\overline{X}$ lies outside the interval $[0.49, 0.51]$. Even so, at confidence level 95%, you'd be wrong 5% of the time.