Math 218 Mathematical Statistics

Prof. D. Joyce, Clark University

18 Feb 2008

Due Wednesday. Page 231: 17–20, 22.

Test Friday. Thru chapter 5.

Next week. Begin chapter 7. Basically, it goes into detail about the statistical tests that were introduced in chapter 6. Here's a brief outline of the chapter. Throughout the chapter we have a single population with an unknown parameter. A the beginning of the chapter, we'll look at how to make inferences about the mean μ.

In 7.1, we'll look at when (1) we have a normal distribution with a known variance σ^2, or (2) any distribution with a large sample. With a large sample, say $n \geq 30$, the central limit theorem says that $\overline{X}$ is approximately normal, and we can use the sample variance S^2 to approximate σ^2 if we don't happen to know σ^2.

In 7.2, we'll look at the case when n is small, but we have a normal distribution with unknown σ^2. As we saw in chapter 6, the T-distribution is used in that situation. (Incidentally, for small samples from distributions other than normal distributions, the T-distribution is replaced by other specialized distributions, but we won't discuss any of those.)

In 7.3, we'll switch from making inferences about μ to inferences about σ^2, but only in the case that the population distribution is a normal distribution. The statistical tests will all involve the χ^2 distribution.

Last meeting. Introduced hypothesis tests. A hypothesis test has two hypotheses concerning the value of an unknown parameter. 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.

Today. More on hypothesis tests.

Type I and Type II errors, α-risks and β-risks. Errors in hypothesis tests have these two types.

A type I error occurs when the null hypothesis holds, but we reject it. Hypothesis tests are designed to control for type I errors. For instance, if the test is designed at the 95% confidence level, then when the null hypothesis actually holds, then 95% of the time we won't reject it, but $\alpha = 5\%$ of the time we will, so make this type I error 5% of the time. The probability of a type I error is denoted α, and it's sometimes called *level of significance* or the *α-risk*. To reduce the α-risk, just design the test to a higher confidence level.

A type II error occurs when the null hypothesis does not hold, but we don't reject it. Typically, we can't tell how often type II errors occur, because the frequency depends on the unknown parameter, and in the worst case they can occur 95% of the time when the null hypotheses does not hold.

Let's take an example to see this more clearly. Suppose the population distribution is a Bernoulli distribution with unknown parameter p. Last time we saw how to construct a test for a fair coin. The null hypothesis H_0 was that $p = \frac{1}{2}$, while the alternative hypothesis H_1 was that $p \neq \frac{1}{2}$. The hypothesis test said

> 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 $1/\sqrt{n}$.

To pin down this example, let's take $n = 10000$. Then we will reject H_0 if $|\overline{X} - \frac{1}{2}| > 0.01$.

Now, if $p = \frac{1}{2}$, then $P(|\overline{X} - \frac{1}{2}| > 0.01)$ is 0.05, leading to a type I error.

But if $p \neq \frac{1}{2}$, then the probability of a type II error,

$$\beta = P\left(|\overline{X} - \tfrac{1}{2}| \leq 0.01\right)$$

depends on what the value of p is. That is, the β-risk is actually a function that depends on the parameter. For instance, if p is very close to $\frac{1}{2}$, then this probability of a type II error will be very close to 0.95. But if p is near 0 or 1, this probability will be nearly 0.

Let's compute the probability of a type II error if $p = 0.49$. The distribution of $\overline{X}$ is almost normal with a mean of $p = 0.49$ and a variance of

$$\sigma^2_{\overline{X}} = \frac{pq}{n} = \tfrac{1}{10000} \cdot 0.49 \cdot 0.51 \approx \tfrac{1}{40000},$$

so a standard deviation of $\sigma_{\overline{X}} \approx \frac{1}{200} = 0.005$. Thus, standardizing the condition, we have

$$\begin{aligned} & P(0.49 \leq \overline{X} \leq 0.51) \\ = \; & P\left(\frac{0.49 - 0.49}{0.005} \leq \frac{\overline{X} - 0.49}{0.005} \leq \frac{0.51 - 0.49}{0.005}\right) \\ = \; & P(0 \leq Z \leq 4) \approx 0.5. \end{aligned}$$

In other words, a coin whose probability is heads is 0.49 will pass this fairness test half the time giving 50% type II errors.

This function β that depends on the unknown parameter θ, or rather the function that gives the probability that H_0 will not be rejected, is called the *operating characteristic function* of the test, and 1 minus it, that is the function that gives the probability that H_0 will be rejected, is called the *power function* $\pi(\theta)$ of the test. From either one, you can read off the α-risk and the β-risks for various values of θ.

The observed level of significance, called the P-value. Sometimes a hypothesis test just barely ends up rejecting or accepting H_0, and sometimes it clearly rejects or accepts H_0. The *observed level of significance*, or *P-value*, is a way of recording the information of near and far hits and misses. The value P is the smallest level of α for which H_0 is accepted; any lower and H_0 would be rejected.

Suppose we do a 95% confidence level test (so the level of significance is $\alpha = 0.05$). If H_0 is just barely accepted, then the observed level of significance is $P = 0.05$ or slightly larger. But if H_0 is just barely rejected, then P is just slightly smaller than 0.05. If the test accepts H_0 without question, then the P-value is higher, perhaps much higher than 0.05, while if the test clearly rejects H_0, then the P-value is smaller than 0.05, perhaps nearly 0.

P-values are fairly easy to compute. The fair-coin test is a special case of a two-sided hypothesis test on μ. We'll look at this in more detail in section 7.1. The test statistic for such tests is $z = \dfrac{\overline{x} - \mu_0}{\sigma/\sqrt{n}}$ when σ is known, or, when n is large and σ not known, it's $z = \dfrac{\overline{x} - \mu_0}{s/\sqrt{n}}$, where s is the sample variance. This z has a standard normal distribution. The null hypothesis is $H_0 : \mu = \mu_0$, and the alternative hypothesis is $H_1 : \mu \neq \mu_0$. The P-value is the probability $P(Z \leq z \,|\, H_0)$, which is $2(1 - \Phi(|z|))$, which can be looked up in the standard normal table.