14.30 Introduction to Statistical Methods in Economics
Spring 2009

14.30 Introduction to Statistical Methods in Economics
Lecture Notes 23

Konrad Menzel

May 12, 2009

1 Examples

Example 1 *Assume that babies' weights (in pounds) at birth are distributed according to $X \sim N(7, 1)$. Now suppose that if an obstetrician gave expecting mothers poor advice on diet, this would cause babies to be on average 1 pound lighter (but have same variance). For a sample of 10 live births, we observe $\bar{X}_{10} = 6.2$.*

- *How do we construct a 5% test of the null that the obstetrician is not giving bad advice against the alternative that he is? We have*

$$H_0 : \mu = 7 \text{ against } H_A : \mu = 6$$

We showed that for the normal distribution, it is optimal to base this simple test only on the sample mean, $\bar{X}_{10}$, so that $T(\mathbf{x}) = \bar{x}_{10}$. Under H_0, $\bar{X}_{10} \sim N(7, 0.1)$ and under H_A, $\bar{X}_{10} \sim N(6, 0.1)$. The test rejects H_0 if $\bar{X}_{10} < k$. We therefore have to pick k in a way that makes sure that the test has size 5%, i.e.

$$0.05 = P(\bar{X}_{10} < k|\mu = 7) = \Phi\left(\frac{k-7}{\sqrt{0.1}}\right)$$

where $\Phi(\cdot)$ is the standard normal c.d.f.. Therefore, we can obtain k by inverting this equation

$$k = 7 + \sqrt{0.01}\Phi^{-1}(0.05) \approx 7 - \frac{1.645}{\sqrt{10}} \approx 6.48$$

Therefore, we reject, since $\bar{X}_{10} = 6.2 < 6.48 = k$.

- *What is the power of this test?*

$$P(\bar{X}_{10} < 6.48|\mu = 6) = \Phi\left(\frac{6.48-6}{\sqrt{0.1}}\right) \approx \Phi(1.518) \approx 93.55\%$$

- *Suppose we wanted a test with power of at least 99%, what would be the minimum number n of newborn babies we'd have to observe? The only thing that changes with n is the variance of the sample mean, so from the first part of this example, the critical value is $k_n = 7 - \frac{1.645}{\sqrt{n}}$, whereas the power of a test based on $\bar{X}_n$ and critical value k_n is*

$$1 - \beta = P(\bar{X}_n < k_n|\mu = 6) = \Phi\left(\sqrt{n} - 1.645\right)$$

Setting $1-\beta \geq 0.99$, we get the condition

$$\sqrt{n} - 1.645 \geq \Phi^{-1}(0.99) = 2.326 \Leftrightarrow n \geq 3.971^2 \approx 15.77$$

This type of power calculations is frequently done when planning a statistical experiment or survey - e.g. in order to determine how many patients to include in a drug test in order to be able to detect an effect of a certain size. Often it is very costly to treat or survey a large number of individuals, so we'd like to know beforehand how large the experiment should be so that we will be able to detect any meaningful change with sufficiently high probability.

Example 2 *Suppose we are still in the same setting as in the previous example, but didn't know the variance. Instead, we have an estimate $S^2 = 1.5$. How would you perform a test? As we argued earlier, the statistic*

$$T := \frac{\bar{X}_n - \mu_0}{S/\sqrt{n}} \sim t_{n-1}$$

is student-t distributed with $n-1$ degrees of freedom if the true mean is in fact μ_0. Therefore we reject H_0 if

$$T = \frac{\bar{X}_n - 7}{S/\sqrt{10}} < t_9(5\%)$$

Plugging in the values from the problem, $T = -\frac{0.8}{\sqrt{1.5/10}} \approx -2.066$, which is smaller than $t_9(0.05) = -1.83$.

Example 3 *Let $X_i \sim \text{Bernoulli}(p)$, $i = 1, 2, 3$. I.e. we are flipping a bent coin three times independently, and $X_i = 1$ if it comes up heads, otherwise $X_i = 0$. We want to test $H_0 : p = \frac{1}{3}$ against $H_A : p = \frac{2}{3}$. Since both hypotheses are simple, can use likelihood ratio test*

$$T = \frac{f_0(X)}{f_A(X)} = \frac{\prod_{i=1}^3 \left(\frac{1}{3}\right)^{X_i} \left(\frac{2}{3}\right)^{1-X_i}}{\prod_{i=1}^3 \left(\frac{2}{3}\right)^{X_i} \left(\frac{1}{3}\right)^{1-X_i}} = \frac{2^{3-\sum_{i=1}^3 X_i}}{2^{\sum_{i=1}^3 X_i}} = 2^{3-2\sum_{i=1}^3 X_i}$$

Therefore, we reject if

$$2^{3-2\sum_{i=1}^3 X_i} \leq k \Leftrightarrow (3 - 2\sum_{i=1}^3 X_i) \log 2 \leq \log k$$

which is equivalent to $\bar{X}_3 \geq \frac{1}{2} - \frac{\log k}{6 \log 2}$. In order to determine k, let's list the possible values of $\bar{X}_3$ and their probabilities under H_0 and H_A, respectively:

$\bar{X}_3$	Prob. under H_0	Prob. under H_A	cumul. prob. under H_0
1	$\frac{1}{27}$	$\frac{8}{27}$	$\frac{1}{27}$
$\frac{2}{3}$	$\frac{6}{27}$	$\frac{12}{27}$	$\frac{7}{27}$
$\frac{1}{3}$	$\frac{12}{27}$	$\frac{6}{27}$	$\frac{19}{27}$
0	$\frac{8}{27}$	$\frac{1}{27}$	1

So if we want the size of the test equal to $\alpha = \frac{1}{27}$, we could reject if and only if $\bar{X}_3 > \frac{2}{3}$, or equivalently we can pick $k = \frac{1}{2}$. The power of this test is equal to

$$1 - \beta = P(\bar{X}_3 = 1 | H_A) = \frac{8}{27} \approx 29.63\%$$

Example 4 *Suppose we have one single observation generated by either*

$$f_0(x) = \begin{cases} 2x & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases} \quad \text{or } f_A(x) = \begin{cases} 2-2x & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

- *Find the testing procedure which minimizes the sum of $\alpha+\beta$ - do we reject if $X = 0.6$? Since we only have one observation X, it's not too complicated to write the critical region directly in terms of X, and there is nothing to be gained by trying to find some clever statistic (though of course Neyman-Pearson would still work here). By looking at a graph of the densities, we can convince ourselves that the test should reject for small values of $X < k$ for some critical level k. The probability of type I and type II error is, respectively,*

  $$\alpha(k) = P(\text{reject}|H_0) = \int_0^k 2x dx = k^2$$

  for $0 \le k \le 1$, and

  $$\beta(k) = P(\text{don't reject}|H_A) = \int_k^1 (2-2x)dx = 2(1-k) - 1 + k^2 = 1 - k(2-k)$$

  Therefore, minimizing the sum of the error probabilities over k,

  $$\min_k\{\alpha(k) + \beta(k)\} = \min_k\{k^2 + 1 - k(2-k)\} = \min_k\{2k^2 + 1 - 2k\}$$

  Setting the first derivative of the minimand to zero,

  $$0 = 4k - 2 \Leftrightarrow k = \frac{1}{2}$$

  Therefore we should reject if $X < \frac{1}{2}$, and $\alpha = \beta = \frac{1}{4}$. Therefore, we would in particular not reject H_0 for $X = 0.6$.

- *Among all tests such that $\alpha \le 0.1$, find the test with the smallest β. What is β? Would you reject if $X = 0.4$? - first we'll solve $\alpha(k) = 0.1$ for k. Using the formula from above, $\bar{k} = \sqrt{0.1}$. Therefore,*

  $$\beta(\bar{k}) = 1 - 2\bar{k} + \bar{k}^2 = 1.1 - 2\sqrt{0.1} \approx 46.75\%$$

  Since $k = \sqrt{0.1} \approx 0.316 < 0.4$, we don't reject H_0 for $X = 0.4$.

Example 5 *Suppose we observe an i.i.d. sample $X_1, \ldots, X_n$, where $X_i \sim U[0, \theta]$, and we want to test*

$$H_0 : \theta = \theta_0 \text{ against } H_A : \theta \neq \theta_0, \theta > 0$$

There are two options: we can either construct a $1-\alpha$ confidence interval for θ and reject if it doesn't cover θ_0. Alternatively, we could construct a GLRT test statistic

$$T = \frac{L(\theta_0)}{\max_{\theta \in \mathbb{R}_+} L(\theta)}$$

The likelihood function is given by

$$L(\theta) = \prod_{i=1}^n f_X(X_i|\theta) = \begin{cases} \left(\frac{1}{\theta}\right)^n & \text{for } 0 \le X_i \le \theta, i = 1, \ldots, n \\ 0 & \text{otherwise} \end{cases}$$

The denominator of T is given by the likelihood evaluated at the maximizer, which is the maximum likelihood estimator, $\hat{\theta}_{MLE} = X_{(n)} = \max\{X_1, \ldots, X_n\}$, so that

$$\max_{\theta \in \mathbb{R}_+} L(\theta) = L(\hat{\theta}_{MLE}) = \left(\frac{1}{X_{(n)}}\right)^n$$

Therefore,

$$T = \frac{L(\theta_0)}{\max_{\theta \in \mathbb{R}_+} L(\theta)} = \left(\frac{X_{(n)}}{\theta_0}\right)^n$$

In order to find the critical value k of the statistic which makes the size of the test equal to the desired level, we'd have to figure out the distribution under the null $\theta = \theta_0$ - could look this up in the section on order statistics.

As an aside, even though we said earlier that for large n, the GLRT statistic is χ^2-distributed under the null, this turns out not to be true for this particular example because the density has a discontinuity at the true parameter value.

2 Other Special Tests

2.1 Two-Sample Tests

Suppose we have two i.i.d. samples $X_1, \ldots, X_{n_1}$ and $Z_1, \ldots, Z_{n_2}$, potentially of different sizes n_1 and n_2, and may be from two different distributions.

$$\begin{aligned} X_i &\sim N(\mu_X, \sigma_X^2) \\ Z_i &\sim N(\mu_Z, \sigma_Z^2) \end{aligned}$$

Two types of hypothesis tests we might want to do are

1. $H_0 : \mu_X = \mu_Z$ against $H_A : \mu_X \neq \mu_Z$, or
2. $H_0' : \sigma_X^2 = \sigma_Z^2$ against $H_A' : \sigma_X^2 \neq \sigma_Z^2$

How should we test these hypotheses?

1. Here, we will only consider the case in which σ_X^2 and σ_Z^2 are known (see the book for a discussion of the other case). Under $H_0 : \mu_X = \mu_Z$,

$$T = \frac{\bar{X} - \bar{Z}}{\sqrt{\frac{\sigma_X^2}{n_1} + \frac{\sigma_Z^2}{n_2}}} \sim N(0, 1)$$

   Intuitively, T should be large (in absolute value) if the null is not true. Therefore, a size α test of H_0 against H_A rejects H_0 if

$$|T| > -\Phi^{-1}\left(\frac{\alpha}{2}\right)$$

2. For the test on the variances, need to recall distributional results:

$$\frac{(n_1 - 1)s_X^2}{\sigma_X^2} \sim \chi^2_{n_1 - 1}, \text{ and } \frac{(n_2 - 1)s_Z^2}{\sigma_Z^2} \sim \chi^2_{n_2 - 1}$$

independently of another. Also recall that a ratio of independent chi-squares divided by their degrees of freedom is distributed F,

$$S' = \frac{\frac{(n_1-1)s_X^2}{\sigma_X^2}/(n_1-1)}{\frac{(n_2-1)s_Z^2}{\sigma_Z^2}/(n_2-1)} \sim F_{n_1-1,n_2-1}$$

We clearly don't know σ_X^2 and σ_Z^2, but under $H_0 : \sigma_X^2 = \sigma_Z^2$, this expression simplifies to

$$\tilde{S} = \frac{s_X^2}{s_Z^2}$$

Therefore, a size α test rejects if $\tilde{S} > F_{n_1-1,n_2-1}(1-\alpha/2)$ or $\tilde{S} < F_{n_1-1,n_2-1}(\alpha/2)$.

2.2 Nonparametric Inference

So far, we have mostly considered problems where the data generating process is of form $f(x|\theta)$ (family of distributions) and known up to a finite dimensional parameter θ. Testing in that setting is called *parametric inference.*

As exceptions, we noted in estimation that sample means, variances and other moments had favorable properties for estimation of means, variances and higher-order moments of *any* distribution.

Since the entire distribution of a random variable can be characterized by its c.d.f., it may seem like a good idea to estimate the c.d.f. from the data without imposing any restrictions (except that it should be a valid c.d.f. of course, i.e. monotone and continuous from the right).

The *sample distribution function* $F_n(x)$ is given by

$$F_n(x) = \frac{j}{n} \text{ for } X_{(j)} \leq x < X_{(j+1)}$$

where $X_{(j)}$ is the jth order statistic (remember that this is the jth smallest value in the sample), and $X_{(0)} \equiv -\infty$ and $X_{(n+1)} \equiv \infty$.

Example 6 *For a sample $\{-1, 3, 1, 1, .5, 2, 0\}$, the ordered sample is $\{-1, 0, 0.5, 1, 1, 2, 3\}$, and we can graph the sample distribution function $F_n(x)$:*

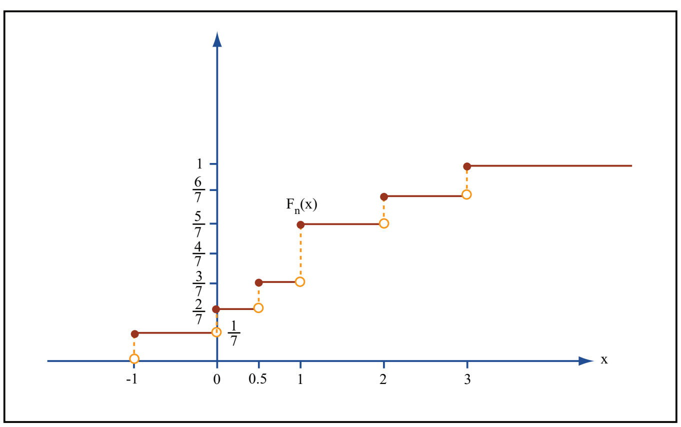

Image by MIT OpenCourseWare.

We are interested in the inference problem in which we have a random sample $X_1, \ldots, X_n$ from an unknown family of distributions and want to test whether it has been generated by a particular distribution with c.d.f. $F(x)$ (e.g. $F(x) = \Phi(x)$ for the standard normal distribution). Since there are no specific parameters for which we can do any of the tests outlined in the previous discussion, the idea for a test is to check whether $F_n(x)$ does not deviate "too much" from $F(x)$.

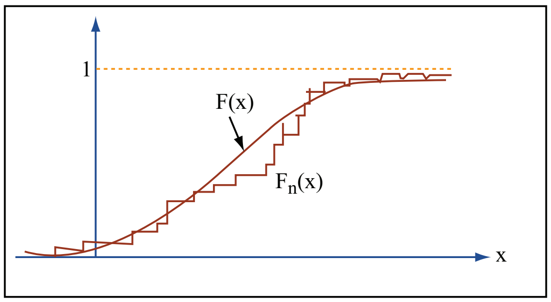

Image by MIT OpenCourseWare.

2.3 The Kolmogorov-Smirnov Test

In order to test whether an observed sample was generated by a distribution $F(x)$, we reject for large values of the Kolmogorov-Smirnov statistic which is defined as

$$D_n = \sup_x |F_n(x) - F(x)|$$

where $\sup_x F(x)$ denotes the supremum, i.e. the smallest upper bound on $\{F(x) : x \in \mathbb{R}\}$ - for continuous functions on compact sets, this is the same as the maximum, but since the Kolmogorov-Smirnov statistic involves the sample distribution function which has jumps of size $\frac{1}{n}$, and the supremum is taken over the entire real line, it may in fact not be attained at any particular value of x.

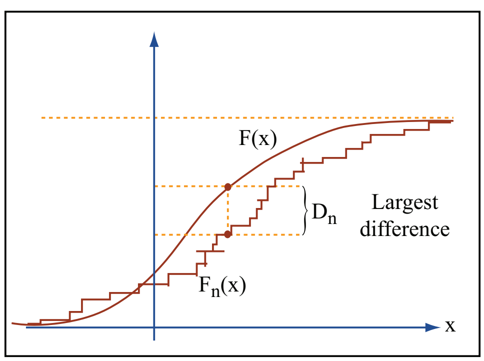

Image by MIT OpenCourseWare.

The critical values of the statistic can be obtained from its asymptotic (i.e. for large n) distribution function

$$G(D_n) = \lim_{n\to\infty} P(D_n < x) = 1 - 2\sum_{i=1}^{\infty}(-1)^{n-1}e^{-2n^2x^2}$$

The argument leading to this expression is not at all obvious and very technical since it involves distributions over functions rather than real numbers (random functions are usually called *stochastic processes*).

Since the calculation of the c.d.f. requires that we compute an infinite series, using the formula is not straightforward. However, most statistics texts tabulate the most common critical values.

Example 7 *Suppose we toss four coins repeatedly, say 160 times and want to test at the significance level $\alpha = 0.2$ whether the sample was generated by a $B(4, 0.5)$ distribution. Let's say we observed the following sample frequencies:*
Then the Kolmogorov-Smirnov statistic equals

number of heads	0	1	2	3	4
sample frequency	10	33	61	43	13
cumulative *sample* frequency $F_n(\cdot)$	10	43	104	147	160
cumulative frequency *under* H_0 $F(\cdot)$	10	50	110	150	160
differences	0	7	6	3	0

$$D_n = \frac{1}{160} \max\{0.7, 6, 3, 0\} = \frac{7}{160} \approx 0.044$$

Using the asymptotic formula from the book, $C_{0.20} = \frac{1.07}{\sqrt{160}} \approx 0.85$*. Since* $0.44 < 0.85$*, we don't reject the null at the 20% level.*

2.4 2-Sample Kolmogorov-Smirnov Test

Suppose we have two independent random samples, $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ from unknown families of distributions, and we want to test whether both samples were generated by the same distribution. The idea is that we should test whether $F_n(x)$ and $G_n(x)$ are not "too far" apart.
We can construct a test statistic

$$D = \sup_x |F_n(x) - G_n(x)|$$

and reject the null for large values of D. A good asymptotic approximation for the critical value for a size α test is

$$\text{reject if } D > \sqrt{-\frac{1}{2}\left(\frac{1}{m} + \frac{1}{n}\right) \log \frac{\alpha}{2}}$$

2.5 Pearson's χ^2 Test

Suppose each X_i from a sample of n i.i.d. observations is classified into one of k categories, $A_1, \ldots, A_k$. Let $p_1, \ldots, p_k$ be the probabilities of each category, and $f_1, \ldots, f_k$ be the observed frequencies. Suppose we want to test the joint hypothesis

$$H_0 : p_1 = \pi_1, p_2 = \pi_2, \ldots, p_k = \pi_k$$

against the alternative that at least two or more of these equalities don't hold (note that since the probabilities have to add up to one, it can't be that exactly one equality is violated). We can use the statistic

$$T = \sum_{i=1}^{k} \frac{(f_i - n\pi_i)^2}{n\pi_i}$$

and reject for large values of T. In order to determine the appropriate critical value, we'd have to know how T is distributed. Unfortunately this distribution depends on the underlying model. However, under H_0 the distribution is asymptotically independent of model, and for large samples n $T \sim \chi^2_{k-1}$ approximately. As a rule of thumb, the chi-squared approximation works well if $n \geq 4k$.