MIT OpenCourseWare
http://ocw.mit.edu

14.30 Introduction to Statistical Methods in Economics
Spring 2009

For information about citing these materials or our Terms of Use, visit: http://ocw.mit.edu/terms.

14.30 Introduction to Statistical Methods in Economics
Lecture Notes 24

Konrad Menzel

May 14, 2009

1 Review

Point Estimation

- Estimator function $\hat{\theta}(X_1, \ldots, X_n)$ of the sample
- bias of estimator is
$$\text{Bias}(\hat{\theta}) = \mathbb{E}_{\theta_0}[\hat{\theta}] - \theta_0$$
- standard error of estimator given by
$$\sigma(\hat{\theta}) = \sqrt{\text{Var}(\hat{\theta})}$$

Important criteria for assessing estimators are

- Unbiasedness
- Efficiency
- Consistency

Methods for constructing Estimators:

1. Method of Moments:
   - mth population moment is $\mathbb{E}_\theta[X_i^m] = \mu_m(\theta)$
   - mth sample moment is $\overline{X^m} = \frac{1}{n}\sum_{i=1}^n X_i^m$
   - compute first k moments, equate $\overline{X^m} \stackrel{!}{=} \mu_m(\hat{\theta})$ for $m = 1, \ldots, k$, and solve for the estimate $\hat{\theta}$.
2. Maximum Likelihood
   - write down likelihood function for the sample $X_1, \ldots, X_n$,
$$L(\theta) = f(X_1, \ldots, X_n|\theta) = \prod_{i=1}^n f(X_i|\theta)$$
   - find value of θ which maximizes $L(\theta)$ or $\log(L(\theta))$.
   - usually find maximum by setting first derivative to zero, but if support of random variable depends on θ may not be differentiable, so you should rather try to see what function looks like, and where the maximum should be.

Confidence Intervals

- find functions of data $A(X_1, \ldots, X_2)$ and $B(X_1, \ldots, X_n)$ such that

$$P_{\theta_0}(A(X_1, \ldots, X_n) \leq \theta_0 \leq B(X_1, \ldots, X_n)) = 1 - \alpha$$

- then $[A(X_1, \ldots, X_n), B(X_1, \ldots, X_n)]$ is a $1 - \alpha$ confidence interval for θ
- for given significance level $1 - \alpha$ many possible valid confidence intervals.

In order to construct confidence intervals, usually proceed as follows:

1. find $a(\theta_0)$ and $b(\theta_0)$ such that

$$P_{\theta_0}(a(\theta_0) \leq T(X_1, \ldots, X_n) \leq b(\theta_0)) = 1 - \alpha$$

   for some statistic $T(X_1, \ldots, X_n)$ (typically will use an estimator $\hat{\theta}$ here).

2. rewrite the event inside the probability in form

$$P(A(X_1, \ldots, X_n) \leq \theta_0 \leq B(X_1, \ldots, X_n)) = 1 - \alpha$$

3. evaluate $A(\cdot)$ and $B(\cdot)$ at the sample values $X_1, \ldots, X_n$ to obtain confidence interval

Some important cases:

- $\hat{\theta}$ unbiased and normally distributed, $\text{Var}(\hat{\theta})$ known:

$$[A(X_1, \ldots, X_n), B(X_1, \ldots, X_n)] = \left[\hat{\theta} + \Phi^{-1}\left(\frac{\alpha}{2}\right)\sqrt{\text{Var}(\hat{\theta})}, \hat{\theta} + \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\sqrt{\text{Var}(\hat{\theta})}\right]$$

- $\hat{\theta}$ unbiased and normally distributed, $\text{Var}(\hat{\theta})$ not known, but have estimator $\hat{S}$:

$$[A(X_1, \ldots, X_n), B(X_1, \ldots, X_n)] = \left[\hat{\theta} + t_{n-1}\left(\frac{\alpha}{2}\right)\sqrt{\text{Var}(\hat{\theta})}, \hat{\theta} + t_{n-1}\left(1 - \frac{\alpha}{2}\right)\sqrt{\text{Var}(\hat{\theta})}\right]$$

- $\hat{\theta}$ not normal, $n > 30$ or so: most estimators we have seen so far turn out to be asymptotically normally distributed, so we'll use that approximation and calculate confidence intervals as in the previous case. Whether or not we know the variance, we use the t-distribution as a way of penalizing the CI for using approximation.
- $\hat{\theta}$ not normal, n small: if (a) we know the p.d.f. of $\hat{\theta}$, can construct CIs from first principles. if (b) we don't know the p.d.f., there's nothing we can do.

Hypothesis Testing

- hypotheses $H_0 : \theta \in \Theta_0$ against $H_A : \theta \in \Theta_A$
- test statistic $T(X)$ some function of data which takes different distributions under the null and the alternative
- critical region C: regions of realizations of $T(X)$ for which we reject the null hypothesis

- testing procedure: reject H_0 if $T(X) \in C$

The choice of C determines

$$\begin{aligned} \alpha = & \quad P(\text{type I error}) & = P(\text{reject}|H_0) \\ \beta = & \quad P(\text{type II error}) & = P(\text{don't reject}|H_A) \end{aligned}$$

alpha is called the *size*, and $1-\beta$ the *power* of the test.

- for two tests with same size α, prefer test with greater power $1-\beta$
- if $\beta = \beta(\theta)$ would prefer test with lower $\beta(\theta)$ for all values of θ, uniformly most powerful test
- H_0 and H_A both simple: by Neyman-Pearson Lemma, most powerful test is of form "reject if $\frac{f(x|\theta_0)}{f(X|\theta_A)} < k$"
- test of the form "reject if $g(T(X)) < g(k)$ for some monotone function identical to test "reject if $T(X) < k$"

Construction of tests depends on form of hypotheses H_0 and H_A:

1. both H_0 and H_A simple: likelihood ratio test
$$T(\mathbf{x}) = \frac{f_0(\mathbf{x})}{f_A(\mathbf{x})}$$
and reject if $T(\mathbf{X}) < k$ for some appropriately chosen value k (most powerful by Neyman-Pearson Lemma)
2. $H_0 : \theta = \theta_0$ simple, $H_A : \theta \in \Theta_A$ composite and 2-sided: construct a $1-\alpha$ confidence interval $[A(X), B(X)]$ and reject if $\theta_0 \notin [A(\mathbf{X}), B(\mathbf{X})]$
3. $H_0 : \theta = \theta_0$ simple, $H_A : \theta > \theta_0$ composite and one-sided: construct a $1-2\alpha$ confidence interval $[A(X), B(X)]$ for θ and reject if $\theta_0 < A(X)$.
4. general case: Generalized Likelihood Ratio Test statistic
$$T(\mathbf{x}) = \frac{\max_{\theta \in \Theta_0} L(\theta)}{\max_{\theta \in \Theta_A \cup \Theta_0} L(\theta)} = \frac{\max_{\theta \in \Theta_0} f(\mathbf{x}|\theta)}{\max_{\theta \in \Theta_A \cup \Theta_0} f(\mathbf{x}|\theta)}$$
and reject if $T(\mathbf{X}) < k$ for some appropriately chosen constant k

Two-Sample Tests

- have two independent samples $X_1, \ldots, X_{n1}$ and $Z_1, \ldots, Z_{n2}$ where $X_i \sim N(\mu_X, \sigma_X^2)$ and $Z_i \sim N(\mu_Z, \sigma_Z^2)$ i.i.d.
- want to test either (i) $H_0 : \mu_X = \mu_Z$ vs. $H_A : \mu_X \neq \mu_Z$ *or* (ii) $H_0 : \sigma_X^2 = \sigma_Z^2$ vs. $H_A : \sigma_X^2 \neq \sigma_Z^2$
- in case (i) form statistic
$$T = \frac{\bar{X}_{n1} - \bar{Z}_{n2}}{\sqrt{\frac{\sigma_X^2}{n_1} + \frac{\sigma_Z^2}{n_2}}}$$
which is $N(0,1)$ under the null hypothesis.

- in case (ii) form statistic

$$T = \frac{(n_1 - 1)\hat{s}_X^2}{(n_2 - 1)\hat{s}_Z^2}$$

which is $F(n_1 - 1, n_2 - 1)$ distributed under the null hypothesis, and we reject if either $T < F^{-1}\left(\frac{\alpha}{2}\right)$ or if $T > F^{-1}\left(1 - \frac{\alpha}{2}\right)$.

Sample Questions (Spring 2000 Exam)

1. Method of Moments The random variables $X_1, \dots, X_n$ are independent draws from a continuous uniform distribution with support $[0, \theta]$. You know from you time in 14.30 that you can use either method of moments or maximum likelihood to derive an estimator for θ from the sample of X_i. But you want a small challenge, so you define new random variables $Y_1, \dots, Y_n$ such that

$$Y_i = \begin{cases} 0 & \text{if } X_i \leq k \\ 1 & \text{if } X_i > k \end{cases}$$

where k is a constant determined by and known to you. you can estimate θ only using $Y_1, \dots, Y_n$.

(a) Assume $k \in (0, \theta)$. Use thee method of moments to derive an estimator for θ as a function of the Y_i. Also explain why you need k to be in the interval $(0, \theta)$.

(b) Now assume $k \in (0, \infty)$ and that k may be greater than or less than the unknown parameter θ. What can you say about the relationship between k and θ if you observe a random sample with $\bar{Y}_n = 0$?

(c) Derive the maximum likelihood estimator for θ (remember that you can still only use $Y_1, \dots, Y_n$ for the estimation).

Answers:

(a) Since θ is one-dimensional, only have to use first moment of Y_i. The population expectation is given by

$$\mathbb{E}_\theta[Y_i] = P_\theta(X_i \geq k) = \max\left\{1 - \frac{k}{\theta}, 0\right\}$$

If $k < \theta$, the method of moments estimator is obtained by solving

$$\bar{Y}_n = \mathbb{E}_\theta[Y_i] = 1 - \frac{k}{\hat{\theta}} \Leftrightarrow \hat{\theta} = \frac{k}{1 - \bar{Y}_n}$$

(b) If $k > \theta$, $\mathbb{E}_\theta[Y_i] = P(X_i \geq k) = \max\left\{1 - \frac{k}{\theta}, 0\right\} = 0$ does not depend on θ anymore. If we don't know whether k is greater or less than θ, we can use the same reasoning as in the construction of the method of moments estimator to bound the parameter θ setting

$$\bar{Y}_n = \max\left\{1 - \frac{k}{\hat{\theta}}, 0\right\} \geq 1 - \frac{k}{\hat{\theta}} \Leftrightarrow \hat{\theta} \leq \frac{k}{1 - \bar{Y}_n}$$

If $\bar{Y}_n = 0$ for a large sample, k is likely to be greater than θ.

(c) In order to derive the likelihood function, note that

$$P_\theta(Y_i = 1) = P_\theta(X_i \geq k) = 1 - \frac{k}{\theta}$$

Therefore,

$$L(\theta) = \prod_{i=1}^{n} \left(1 - \frac{k}{\theta}\right)^{Y_i} \left(\frac{k}{\theta}\right)^{1-Y_i} = \left(1 - \frac{k}{\theta}\right)^{\sum_{i=1}^{n} Y_i} \left(\frac{k}{\theta}\right)^{\sum_{i=1}^{n}(1-Y_i)}$$

Taking logs,

$$\mathcal{L}(\theta) = \log L(\theta) = \left(\sum_{i=1}^{n} Y_i\right) \log\left(1 - \frac{k}{\theta}\right) + \left(n - \sum_{i=1}^{n} Y_i\right) \log\left(\frac{k}{\theta}\right)$$

Setting the derivative with respect to θ to zero,

$$0 = \frac{\left(\sum_{i=1}^{n} Y_i\right)}{1 - \frac{k}{\theta}} \frac{k}{\theta^2} - \frac{\left(n - \sum_{i=1}^{n} Y_i\right)}{\frac{k}{\theta}} \frac{k}{\theta^2} \Leftrightarrow \left(\sum_{i=1}^{n} Y_i\right) k = \left(n - \sum_{i=1}^{n} Y_i\right)(\theta - k)$$

Solving for θ,

$$\hat{\theta}_{ML} = \frac{nk}{n - \sum_{i=1}^{n} Y_i} = \frac{k}{1 - \bar{Y}_n}$$

Notice that this estimator works even if $k > \theta$.

2. Hypothesis Testing Suppose that $X_1, \ldots, X_n$ form a random sample from a normal distribution with an unknown mean μ and a known variance σ^2 equal to 1.

(a) State the critical region that yields the most powerful test of

$$\begin{aligned} H_0 : &\quad \mu = 0 \\ H_A : &\quad \mu = 1 \end{aligned}$$

at the 5% significance level. Calculate the power of this test.

(b) State the critical region that yields the most powerful test of

$$\begin{aligned} H_0 : &\quad \mu = 1 \\ H_A : &\quad \mu = 0 \end{aligned}$$

at the 5% significance level.

(c) For what values of n and $\bar{X}_n$ will you accept the hypothesis that $\mu = 0$ in part **(a)** and simultaneously accept the hypothesis that $\mu = 1$ in part **(b)**?

(d) State the critical region that yields the uniformly most powerful test of

$$\begin{aligned} H_0 : &\quad \mu = 0 \\ H_A : &\quad \mu > 1 \end{aligned}$$

at the 5% significance level. State the formula for, and then graph the power function $1 - \beta(\mu)$ of this test.

(e) How are the critical regions of the tests in parts **(a)** and **(d)** related? How are the probabilities of committing a type II error related?

Answers:

(a) According to the Neyman-Pearson lemma, the most powerful test is based on the likelihood ratio

$$r(x) = \frac{\prod_{i=1}^n f(x_i|\mu = 0)}{\prod_{i=1}^n f(x_i|\mu = 1)} = \frac{\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\sum_{i=1}^n x_i^2\right\}}{\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\sum_{i=1}^n (x_i - 1)^2\right\}} = \exp\left\{\frac{n}{2} - \sum_{i=1}^n x_i\right\}$$

The most powerful test rejects if the likelihood ratio is less than the critical value, or equivalently, if $\bar{X}_n > k$ for some appropriately chosen value of k (on the exam it's sufficient to point out that we already derived this in class).
Since under the null hypothesis, $\bar{X}_n \sim N(0, 1/n)$,

$$P\left(\bar{X}_n > k \middle| \mu = 0\right) = 1 - \Phi(\sqrt{n}k)$$

so choosing $k = \frac{\Phi^{-1}(1-\alpha)}{\sqrt{n}}$ gives a test of size 5%. The power of this test is given by

$$P\left(\bar{X}_n > k \middle| \mu = 1\right) = 1 - \Phi(\sqrt{n}(k-1))$$

(b) By similar reasoning as in part **(a)**, the most powerful test rejects if $\bar{X}_n < k'$, where $k' = 1 + \frac{\Phi^{-1}(\alpha)}{\sqrt{n}}$.

(c) We'll accept in both tests if

$$k' \leq \bar{X}_n \leq k \Leftrightarrow \sqrt{n} + \Phi^{-1}(\alpha) \leq \sqrt{n}\bar{X}_n \leq \Phi^{-1}(1-\alpha)$$

For sufficiently large n, there will be no values of $\bar{X}_n$ for which none of the tests rejects.

(d) This test is the same as in part **(a)**, because for any value of $\mu > 1$, the likelihood ratio is a strictly decreasing function in the sample mean $\bar{X}_n$, and the critical value k for a size α test is determined only by the distribution under the null hypothesis, which is the same as in part **(a)**.

(e) The critical regions are the same, but the probability of a type II error in part (d) is smaller, since all alternatives are further away from the null than $\mu = 1$.