MICHIGAN STATE UNIVERSITY

AFRE 802
Statistical Methods for Agricultural, Food, & Resource Economists

More Properties of Point Estimators & Methods of Estimation
(WMS Ch. 9.1, 9.3, 9.6-9.7, 9.9)

November 7, 2017

Mary Doidge & Nicole Mason
Michigan State University
Fall 2017

MICHIGAN STATE UNIVERSITY

GAME PLAN

- Collect take-home graded exercises
- Reminder: Ch. 8 HW due Thursday
- Review
- Finish Ch. 8 – selecting the sample size

- Ch. 9
  - Another desirable property of estimators: Consistency
  - Intuition on some common methods of estimation
    - Method of moments
    - Maximum likelihood
    - Least squares

Review: Confidence intervals (CIs)

- **Definition**: A rule used to construct a random interval so that a certain percentage of all data sets yields an interval that contains the population value (target parameter)

$$\boxed{[\hat{\theta}_L, \hat{\theta}_U] \text{ such that } P(\hat{\theta}_L \le \theta \ \le \hat{\theta}_U) = 1 - \alpha}$$

- The percentage we choose for the confidence interval (1-α) is called the **confidence level** (**confidence coefficient** in WMS)
- **Large-sample CIs**: if N is large, then we can invoke the CLT

$$\boxed{\hat{\theta} \pm z_{\alpha/2}\sigma_{\hat{\theta}}}$$

Also, when N is large, can replace σ by S without much loss of accuracy (and still use Z-statistic)

- **Small-sample CI for μ**

$$\boxed{\bar{Y} \pm t_{\alpha/2}\left(\frac{S}{\sqrt{N}}\right)}$$

**For data drawn from approximately normal/bell-shaped distributions – refer to our discussion of t-stats and note that this was something we assumed there as well

Review: Confidence intervals (CIs) – cont'd

$$\boxed{[\hat{\theta}_L, \hat{\theta}_U] \text{ such that } P(\hat{\theta}_L \le \theta \ \le \hat{\theta}_U) = 1 - \alpha}$$

- *What is the interpretation of a 95% confidence interval?*

  From Wooldridge (2009, p. 138):

  - "If random samples were obtained over and over again, with $\hat{\theta}_L$ and $\hat{\theta}_U$ computed each time, then the unknown population value θ would lie in the interval for 95% of the samples."
  - "Unfortunately, for a single sample that we use to construct the CI, we do not know whether θ is actually contained in the interval. We hope we have obtained one of the 95% of all samples where the interval estimate contains θ, but we have no guarantee".

REVIEW

MICHIGAN STATE UNIVERSITY

Selecting the sample size (simple random sampling case)

- **How large of a sample** you need depends on **how accurate** you want your estimates to be (and your **budget constraint**)
- E.g., say you want the sample mean to be within *"a"* units of the population mean with 95% probability, then you could solve for *N* such that:

$$P(|\bar{Y}-\mu| \leq a) = P(-a \leq \bar{Y}-\mu \leq a) = 0.95$$

$$= P(\underbrace{\frac{-a}{\sigma/\sqrt{N}}}_{=-Z_{0.025}} \leq Z \leq \underbrace{\frac{a}{\sigma/\sqrt{N}}}_{=Z_{0.025}}) = 0.95$$

Solve for N, then plug in σ if you have it (to get exact sample size), or S if you don't (to get approximate sample size). **Round up** to the nearest N.

MICHIGAN STATE UNIVERSITY

For the sample mean,

$$N = \left(\frac{z_{\alpha/2}\sigma}{a}\right)^2$$

This formula comes from solving the formula on the previous slide for N

where a is the acceptable distance from μ

(Rounded up to the nearest whole integer)

- Intuitively, how would you expect the sample size to change with increases in the standard deviation? Confidence level? Acceptable distance from the true population parameter?
- Is the formula above consistent this these expectations?

MICHIGAN STATE UNIVERSITY

Selecting the sample size – free software

- Optimal Design (out of U of M)
- Works for simple random sampling and more complex sampling schemes
- https://sites.google.com/site/optimaldesignsoftware/home
- See http://blogs.worldbank.org/impactevaluations/power-calculations-what-software-should-i-use for more on pros/cons of Optimal Design and other software options

6

Summary of Chapter 8

- Major objective of statistics is to **make inferences about population parameters based on sample data**
- Often inferences take the form of estimates – either **point estimates or interval estimates**
- We prefer **unbiased** estimators with **small variance**
- **MSE** gives us a **combined measure** of the **bias** and **variance** of an estimator
- **Confidence intervals** for many parameters can be derived from the **normal distribution b/c of the CLT**
- If **sample size is small** and we don't know the population variance, then can use the ***t* distribution** when deriving **confidence intervals for μ**

MICHIGAN STATE UNIVERSITY

More Properties of Point Estimators & Methods of Estimation

Unbiasedness vs. consistency

- Previously: **unbiasedness**
- *What is unbiasedness and how do we measure the bias of an estimator?*

$$\boxed{\text{If } E(\hat{\theta}) = \theta, \text{ then } \hat{\theta} \text{ is an "}\textbf{unbiased estimator}\text{" of } \theta.}$$

$$\boxed{\text{Bias: } B(\hat{\theta}) = E(\hat{\theta}) - \theta}$$

- **Unbiasedness** is a **finite sample property**, but some very useful estimators in econometrics are always biased (e.g., instrumental variables, two-stage least squares)
- **While unbiasedness is desirable, "consistency" is non-negotiable**

Consistency

Nobel-prize winner, Clive Granger on consistency: "if you can't get it right as N goes to infinity, you shouldn't be in this business" (Wooldridge, 2003: 163).

- Gist: an estimator is **consistent** if it **converges (in probability) to the population parameter as *N*→∞**
- Consistency is an asymptotic property (i.e., concerns the behavior of a statistic as *N*→∞)
- Formally,

$\hat{\theta}$ is a consistent estimator of θ if:

$P(|\hat{\theta}-\theta| \; > \varepsilon) \rightarrow 0$

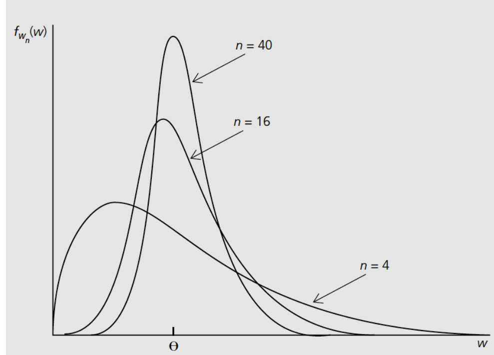
as $N \rightarrow \infty$ for any $\varepsilon > 0$

$\Leftrightarrow$

The "probability limit" (plim) of $\hat{\theta}$ is θ: $\text{plim}\,(\hat{\theta}) = \theta$.

$\Leftrightarrow \hat{\theta} \xrightarrow{a} \theta$

The sampling distributions of a consistent estimator for three sample sizes.



Source: Wooldridge (2003)

What does this remind you of?

The Law of Large Numbers

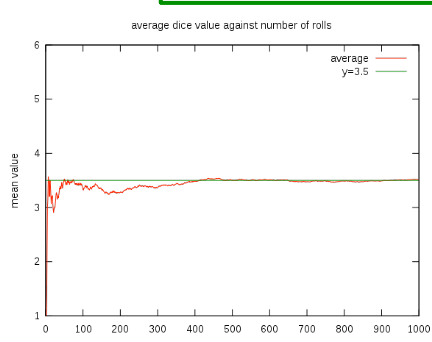
- **As *N*→∞, the sample mean converges (in probability) to the population mean**

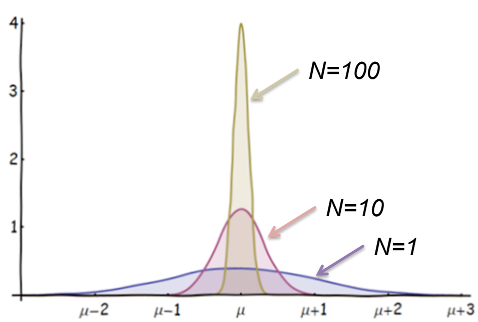
$$P(|\bar{Y}_N - \mu| \; > \varepsilon) \rightarrow 0 \text{ as } N \rightarrow \infty \text{ for any } \varepsilon > 0$$

$$\Leftrightarrow$$

$$\text{plim}(\bar{Y}_N) = \mu$$

Thus $\bar{Y}_N$ is a consistent estimator of μ.





Consistency (cont'd)

THEOREM

An unbiased estimator $\hat{\theta}$ is a consistent estimator of θ if:

$$\lim_{N \to \infty} V(\hat{\theta}) = 0$$

See pp. 450-451 in WMS for proof.

EXAMPLE **9.2**

Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample from a distribution with mean μ and variance $\sigma^2 < \infty$. Show that $\overline{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$ is a consistent estimator of μ. (*Note:* We use the notation $\overline{Y}_n$ to explicitly indicate that $\overline{Y}$ is calculated by using a sample of size n.)

You'll learn more about consistency in AFRE 835 and PhD-level econometrics – e.g., how to determine if a biased estimator is consistent

Example of an estimator that is unbiased but not consistent (pretty rare)

$$\tilde{\mu} = Y_1 \quad \text{where } E(Y_i) = \mu \quad \text{and} \quad V(Y_i) = \sigma^2 \quad \text{for all } i$$

Why is this this estimator an unbiased but inconsistent estimator of μ?

Some estimators are consistent but biased – e.g., instrumental variables and 2SLS; sample correlation estimator (we'll see this later today)

MICHIGAN STATE UNIVERSITY

Rules for probability limits (plims)

- Key point: unlike the expected value, *plims* pass through both **linear AND nonlinear functions**

1. If plim $\hat{\theta} = \theta$ and $g(\hat{\theta})$ is a continuous function of $\hat{\theta}$, then plim $g(\hat{\theta}) = g(\theta)$

2. plim $c = c$ for any constant c

3. If plim $\hat{\theta}_1 = \theta_1$ and plim $\hat{\theta}_2 = \theta_2$, then
   a. plim $\hat{\theta}_1 + \hat{\theta}_2 = \theta_1 + \theta_2$
   b. plim $\hat{\theta}_1 \hat{\theta}_2 = \theta_1 \theta_2$
   c. plim $\dfrac{\hat{\theta}_1}{\hat{\theta}_2} = \dfrac{\theta_1}{\theta_2}$

MICHIGAN STATE UNIVERSITY

3 general approaches to estimation

1. Method of moments
2. Maximum likelihood
3. Least squares

We'll go over the intuition for #1 and #2, then focus on #3 in Chapter 11

Estimation method #1: Method of moments

- **Gist: replace the population "moment" with its sample analogue.**
- The "moments" of Y are $E(Y)=\mu$, $E(Y^2)$, …, $E(Y^k)$
- EX) Suppose that the parameter of interest, θ, is a function of the population mean: $\theta=g(\mu)$

Since the sample mean, $\bar{Y}$, is an unbiased and consistent estimator of μ, it is natural to replace μ with $\bar{Y}$ to obtain estimator $\hat{\theta}=\widehat{g(\mu)}=g(\bar{Y})$.

How could we apply the method of moments to obtain an estimator of $\sigma_{XY}=Cov(X,Y)=E\left[(X-\mu_x)(Y-\mu_y)\right]$?

$$\hat{\sigma}_{XY}^{mom}=\frac{1}{N}\sum_{i=1}^{N}(X_i-\bar{X})(Y_i-\bar{Y})$$

Consistent but biased for similar reason as S'^2. Divide by N-1 instead of N to get unbiased estimator (sample covariance).

Method of moments (cont'd) - examples

- **Sample covariance** (unbiased & consistent):

$$S_{XY}=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})$$

- **Sample correlation**: recall the population correlation coefficient $\rho_{XY}=\sigma_{XY}/(\sigma_X\sigma_Y)$

$$R_{XY}=\frac{S_{XY}}{S_X S_Y}=\frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\left(\sum_{i=1}^{n}(X_i-\bar{X})^2\right)^{1/2}\left(\sum_{i=1}^{n}(Y_i-\bar{Y})^2\right)^{1/2}}$$

R_{XY} is biased but consistent. No unbiased estimator exists for the correlation coefficient!

Method of moments

- **Pros**:
  - Easy & intuitive to use
  - Consistent
- **Cons**:
  - Often biased
  - Typically not very efficient

Estimation method #2: Maximum likelihood

Source: Wooldridge (2003: 714-715)

Let $Y_1, Y_2, ..., Y_N$ be a random sample from a population with PDF $f(y;\theta)$.

Because this is a random sample, we can write the joint (multivariate) PDF as

$$f(y_1;\theta)f(y_2;\theta)\cdots f(y_N;\theta)$$

Define the **likelihood function** (L) as:

$$L(\theta;Y_1,Y_2,...,Y_N) = f(Y_1;\theta)f(Y_2;\theta)\cdots f(Y_N;\theta)$$

- The maximum likelihood estimator **(MLE) of θ is the value of θ that maximizes L**. In practice, often easier to maximize log L.
- Pros of MLEs:
  - Usually **consistent** and **sometimes unbiased**
  - **Often the most (asymptotically) efficient** estimator
  - If adjusted to be unbiased, **often the minimum-variance unbiased estimator**
- No major cons for MLEs
- Common MLEs: probit, logit, Tobit, Poisson

Maximum likelihood - example

Poisson Distribution

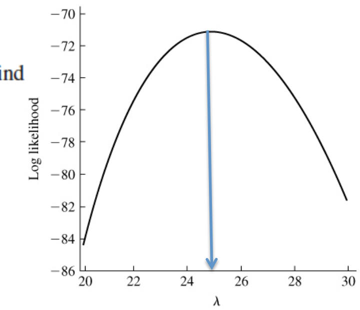
If X follows a Poisson distribution with parameter λ, then

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

If $X_1, \ldots, X_n$ are i.i.d. and Poisson, their joint frequency function is the product of the marginal frequency functions. The log likelihood is thus

$$l(\lambda) = \sum_{i=1}^{n} (X_i \log \lambda - \lambda - \log X_i!)$$
$$= \log \lambda \sum_{i=1}^{n} X_i - n\lambda - \sum_{i=1}^{n} \log X_i!$$

MLE of λ is 24.9 in this example



Setting the first derivative of the log likelihood equal to zero, we find

$$l'(\lambda) = \frac{1}{\lambda} \sum_{i=1}^{n} X_i - n = 0$$

The mle is then

$$\hat{\lambda} = \overline{X}$$

MICHIGAN STATE UNIVERSITY

Estimation method #3: Least squares

- Quick intro today; 3 classes on this after later in course; covered extensively in AFRE 835
- Least squares estimators: estimate the population parameter by the value that makes the sum of squared deviations between the observed values and the estimated values as small as possible
- EX) To find least squares estimator for μ, choose estimator that minimizes: $\sum_{i=1}^{N} \left(Y_i - \hat{\mu}\right)^2 \Rightarrow \hat{\mu} = \overline{Y}$
- EX) Linear regression via least squares (a few classes from now)

MICHIGAN STATE UNIVERSITY

Summary

- 3 important properties of estimators:
  1. Unbiasedness
  2. Efficiency
  3. Consistency
  - Also discussed mean square error (combined measure of bias and variance)

- 3 common methods of estimation:
  - Method of moments
  - Maximum likelihood
  - Least squares

MICHIGAN STATE UNIVERSITY

In-class exercise: method of moments & showing consistency

Suppose you have a random sample of N observations, $Y_1, Y_2, ..., Y_N$, from a population in which Y_i is uniformly distributed over $[0, \theta]$. θ is unknown. Recall from our discussion of the uniform distribution that $E(Y_i) = \mu = \frac{\theta}{2}$. Find the method of moments estimator for θ, noting that $\theta = 2\mu$. Then show that this estimator is consistent.

MICHIGAN STATE UNIVERSITY

Homework:

- WMS Ch. 8 HW due Thursday
- WMS Ch. 9
  - No problems but please review today's lecture notes & read for next class

Next class:

- Intro to hypothesis testing
- Elements of a statistical test
- Common large-sample tests

Reading for next class:

- WMS Ch. 10 (sections 10.1-10.3)