Preliminary Statistics

Lecture 4: Estimation

Sophie van Huellen (sv8@soas.ac.uk), September 2013

Outline

- Point Estimation
  - Estimation Process
  - Desirable Properties
    - Small Sample Properties: Linearity, Unbiasedness, Efficiency, BLUE
    - Large Sample Properties: Asymptotic Unbiasedness, Consistency, Asymptotic Efficiency
  - Estimation of Expected Values
- Interval Estimation
  - Confidence Intervals

Estimation Process

Steps	**Example 1** **X = Height of Basketball Players**	**Example 2** **Y_t = Return on a Stock**
1. Model of the process that generates the data.	1. No data generating model (you could say height depends on regional variables, genes, etc.) Assumption: $X \sim N(\mu,\sigma^2)$	1. According to the efficient market theory: $Y_t = \alpha + u_t$ α is the expected returns u_t is the unpredictable random error with: $E(u_t) = 0$; $E(u_t^2) = \sigma^2$; $E(u_t, u_{t-1}) = 0$
2. How can we obtain an estimator of the unknown parameter(s)?	2. Estimatior for μ (expected value of height) Many choices: sample mean $[\bar{x} = \frac{\sum x_i}{N} = \hat{\mu}]$, median, first observation, …	2. Estimator for $\alpha, \hat{\alpha}$: $S = \sum(Y_t - \hat{\alpha})^2 =$ $\sum(Y_t^2 + \hat{\alpha}^2 - 2\,\hat{\alpha}\,Y_t) =$ $\sum Y_t^2 + T\,\hat{\alpha}^2 - 2\,\hat{\alpha}\sum Y_t$ Min S, i.e. $\frac{\delta S}{\delta \hat{\alpha}} = 0$, then $\hat{\alpha} = \frac{\sum Y_t}{T} = \bar{Y}$
3. How good is the estimator?	3. Goodness of $\bar{x} = \hat{\mu}$ - Unbiased? $E(\bar{x}) = \frac{1}{N} E(\sum X_i) = \mu$ - Precise? $Var(\bar{X}) = \frac{\sigma^2}{N}$	3. Goodness of $\hat{\alpha}$? - BLUE?
4. Test hypotheses construct interval estimates.	4. Given our estimate for $\hat{\mu}$ (say, 1.95), what can we infer about μ? -Interval estimate -Test if true mean is 2.00	4. Is the New Economy Hypothesis valid?
5. Check if initial assumptions are true.	5. Is X normally distributed?	5. Check if our assumptions about the residuals are true.

Point Estimation

- An Estimator is a formula that tells us how to calculate the value of an estimate based on the measurements of the sample.
- An Estimate is the computed value when using the formula.
- There could be many possible estimators. How do we choose among them?

Point Estimation (cont.)

An Example

- Assume $Y \sim N(\mu,\sigma^2)$ and that σ^2 is known.
- μ is the true but unknown population mean.
- If we have a random sample, Y_i, $i=1,2,\ldots,n$ from the population, we may use the *sample mean* to estimate the parameter.
- However, we could use other estimators, e.g. median or the first observation.
- How do we know which one is better to use?

Point Estimation (cont.)
An Analogy

- Point estimation is like throwing a dart at the target:
  - Target ->True parameter, μ
  - Player -> Estimator
  - A throw of the dart –> Estimate
- Drawing a single sample and using it to compute an estimate for the true parameter is like throwing a single dart.

Point Estimation (cont.)
An Analogy

- Suppose the player throws a single dart and targets the bull’s eye.
- Can we conclude that the player is an expert dart player? **NO!**
- If 100 shots in succession hit the bull’s eye, we might be more confident regarding the ‘goodness’ of the person as a dart player.

Point Estimation (cont.)

Goodness of an Estimator

- Similarly, we cannot evaluate the ‘goodness’ of an estimator on the basis of a single estimate.
- We must observe the results when the estimation procedure is repeated many times.
- Construct the Probability Distribution of the values of the estimates obtained in repeated sampling (Sampling Distribution).

Point Estimation (cont.)
Sampling Distribution

- Every time we get a different sample, we obtain a different estimate.
- Due to the variability of the sampling process, an estimator is itself a random.
- With many (hypothetical) samples we can form the distribution of the estimator.

Desirable Properties of Estimators

- Small sample properties (finite)
  - Linearity
  - Unbiasedness
  - Efficiency
- Large sample (asymptotic) properties
  - Asymptotic unbiasedness
  - Consistency
  - Asymptotic efficiency

Desirable Properties of Estimators (cont.)

Linearity

- Suppose θ is the true population parameter of the random variable X and $\hat{\theta}$ an estimator for θ
- A linear estimator is a linear function of the realisations of X: $X_1, X_2, \ldots, X_n$
- Linear estimators are generally easier to manipulate mathematically. *E.g.*

$$\hat{\theta} = a_1 X_1 + a_2 X_2 + \cdots + a_n X_n,$$

Desirable Properties of Estimators (cont.)

Unbiasedness

- An estimator is unbiased if its expected value is equal to the true (population) value.
- On average, over lots of hypothetical samples, the estimator gives the true value. It does not systematically under/over estimate the true value.

$$E(\hat{\theta}) = \theta$$

Desirable Properties of Estimators (cont.)

Unbiasedness

- Bias, $B(\hat{\theta}) = E(\hat{\theta}) - \theta$
- If $B(\hat{\theta}) > 0$ the estimator is upward biased (overestimated the true value)
- If $B(\hat{\theta}) < 0$ the estimator is downward biased (underestimates the true value)
- Distinction from:
  - Sampling Error = $\hat{\theta} - \theta$
  - Mean Square Error (MSE) = $E(\hat{\theta} - \theta)^2$

Desirable Properties of Estimators (cont.)
Efficiency

- An efficient estimator is one whose sampling distribution has the smallest variance amongst the unbiased estimators.
- $\hat{\theta}$ is and efficient estimator if
  - $\hat{\theta}$ is unbiased, and
  - $Var(\hat{\theta}) \leq Var(\tilde{\theta})$ where both are unbiased
- Efficiency guarantees that in repeated sampling a high fractions of values of the estimator will be closer to the true value.

Desirable Properties of Estimators (cont.)
BLUE

- If an estimator is:
  - Linear
  - Unbiased
  - Has the minimum variance amongst the set of linear unbiased estimators
- It is the best linear unbiased estimator or

BLUE

Desirable Properties of Estimators (cont.)

Asymptotic Properties

- Sampling Distribution is based on samples whose size approaches infinity
- Asymptotic Distribution
- Some estimators have the same sampling distribution irrespective of the sample size (e.g. Sample mean of a normal population)
- Others have different distributions as the sample size increases. By the CLT, their sampling distribution tends to the normal as $n \rightarrow \infty$

Desirable Properties of Estimators (cont.)

Asymptotic Unbiasedness

- The expected value of the estimator approaches the population parameter as the sample size approaches infinity:

$$\lim_{n\to\infty} E\left(\hat{\theta}\right) = 0$$

- Example, the variance (dividing by n)

$$E(\hat{\sigma}^2) = \left(\frac{n-1}{n}\right)\sigma^2$$

$$\lim_{n\to\infty} E(\hat{\sigma}^2) = \lim_{n\to\infty}\left(\frac{n-1}{n}\right)\sigma^2 = \sigma^2$$

Desirable Properties of Estimators (cont.)

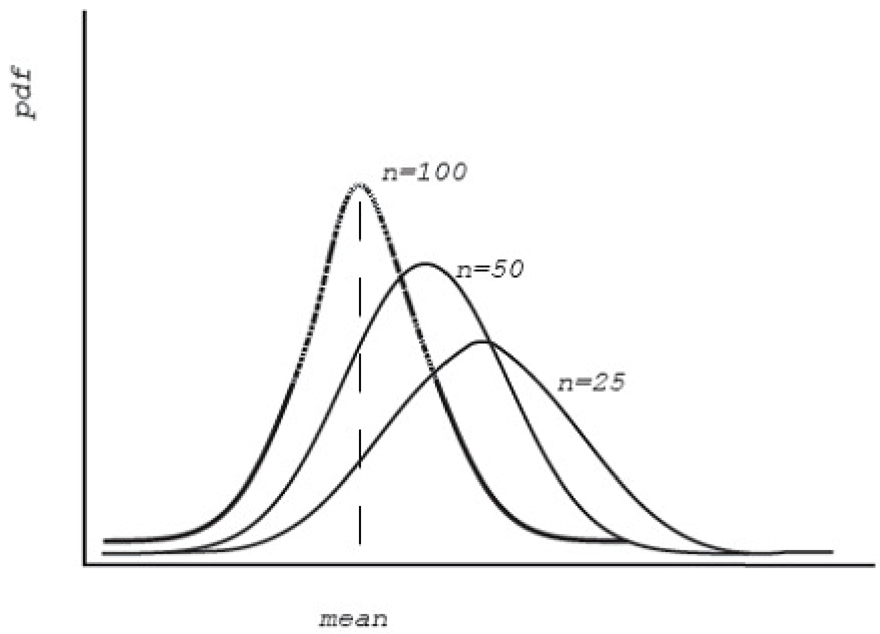
Consistency

- If the sampling distribution collapses on a single value as $n \rightarrow \infty$, then it is said to *converge in probability*.
- The value on which it converges is the probability limit (known as *plim*).
- $\hat{\theta}$ is a consistent estimator of θ if it approaches θ as the sample size gets larger and larger: $p\lim(\hat{\theta}) = \theta$

(Thomas, 1997, Chapter 5)

Desirable Properties of Estimators (cont.)
Consistency

- As the sample size increases biases and variance reduces (i.e. estimator gets asymptotically unbiased and its asymptotic variance gets 0).
- The estimator converges to its true value and it can be estimated exactly.



Desirable Properties of Estimators (cont.)
Asymptotic Efficiency

- An estimator is asymptotically efficient if
  - It is consistent
  - It has the smallest asymptotic variance amongst the set of all consistent estimators
- Given two consistent estimators, the estimator with the smaller asymptotic variance will converge on the population parameter for smaller sample sizes

Estimating the Expected Value of a Random Variable

- We want to estimate the true mean, μ of a random variable Y, from a sample Y_i, $i=1,2,\ldots,n$
- A possible estimator is the sample mean,

$$\hat{\mu} = \bar{Y} = \frac{\sum Y_i}{n}$$

- It can be shown that the sample mean is BLUE
  - Linear $\hat{\mu} = \frac{1}{n}Y_1 + \frac{1}{n}Y_2 + \cdots + \frac{1}{n}Y_n$
  - Unbiased $E(\hat{\mu}) = \mu$
  - Minimum Variance $\min Var(\hat{\mu})$

Properties of the Sample Mean

- Linear function of the sample observations:

$$\hat{\mu} = \frac{\sum_{i=1}^{n} Y_i}{n} = \frac{1}{n}Y_1 + \frac{1}{n}Y_2 + \cdots + \frac{1}{n}Y_n$$

- Unbiased: $E(\hat{\mu}) = \mu$

$$E(\hat{\mu}) = E\left(\frac{\sum Y_i}{n}\right) = \frac{1}{n}E\left(\sum Y_i\right) = \frac{1}{n}E(Y_1 + Y_2 + \cdots + Y_n)$$

$$= \frac{1}{n}(\mu + \mu + \cdots + \mu) = \frac{n\mu}{n} = \mu$$

Properties of the Sample Mean (cont.)

- The Sample Mean has the Smallest Variance:

$$
\begin{aligned}
Var(\hat{\mu}) = Var\left(\frac{\sum Y_i}{n}\right) &= \frac{1}{n^2} Var\left(\sum Y_i\right) \\
&= \frac{1}{n^2} Var(Y_1 + Y_2 + \cdots + Y_n) \\
&= \frac{1}{n^2}\left(Var(Y_1) + Var(Y_2) + \cdots + Var(Y_n)\right) \\
&= \frac{1}{n^2}(\sigma^2 + \sigma^2 + \cdots + \sigma^2) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}
\end{aligned}
$$

Means vs Median

- Both the mean and median are linear estimators of the true mean.
- Both the mean and median are unbiased.
- Mean has a smaller variance than the median

Distribution of the Sample Mean

- **Theorem 1**
  - If Y_i, $i=1,2,\ldots,n$ is a random sample from a random variable Y with mean μ and variance σ^2 the estimator for μ, (Eq.) is the random variable with mean μ and variance σ^2/n:

$$Y \sim (\mu, \sigma^2) \Rightarrow \hat{\mu} \sim \left(\mu, \frac{\sigma^2}{n}\right)$$

- **Theorem 2**
  - If Y is normally distributed, then $\hat{\mu}$ will also be normally distributed:

$$Y \sim N(\mu, \sigma^2) \Rightarrow \hat{\mu} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Distribution of the Sample Mean (cont.)

- **Central Limited Theorem**
  - If the random variable Y is not normally distributed, $\hat{\mu}$ has a distribution which approaches the normal as the sample size approaches infinity:

$$Y \sim (\mu, \sigma^2) \Rightarrow \hat{\mu} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \qquad n \rightarrow \infty$$

Estimating the Variance of a Random Variable

- There are two common estimators for the variance of Y:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(Y_i - \hat{\mu})^2}{n}, s^2 = \frac{\sum_{i=1}^{n}(Y_i - \hat{\mu})^2}{n-1}$$

- $\hat{\sigma}^2$ is a biased estimator for σ^2 (*underestimates* σ^2):

$$E(\sigma^2) = \frac{n-1}{n}\sigma^2 < \sigma^2$$

- s^2 is an unbiased estimator for σ^2:

$$E(s^2) = \sigma^2$$

Standard Deviation vs Standard Error

- Standard Deviation:

$$s = \sqrt{s^2}$$

- Estimate for the Standard Error of the sample mean:

$$SE(\hat{\mu}) = \sqrt{s^2/n} = s/\sqrt{n}$$

Interval Estimation

- Interval Estimator is a formula that tells us how to calculate two endpoints that form the interval which is intended to enclose the value of the true parameter.
- Interval Estimators are commonly known as confidence intervals confidence intervals.
- The probability that a confidence interval will enclose the true parameter value is called the confidence coefficient.

Calculation of Confidence Intervals

Steps

- Find the distribution of the estimator, $\hat{\theta}$.
- Standardise the estimator, $Z = \frac{\hat{\theta} - E(\hat{\theta})}{SE(\hat{\theta})}$.
- Find the distribution of the standardised estimator.
- Find the 2 values in the tails of the distribution of z, such that they give, say, 95% of the distribution.
- Rearrange back from Z to θ.

Calculation of Confidence Intervals
for the Mean

- Find the distribution of the estimator, $\hat{\mu}$.
- Standardise the estimator, $Z = \frac{\widehat{\mu} - E(\widehat{\mu})}{SE(\widehat{\mu})}$.
- Find the distribution of the standardised estimator:
  - Large sample ($n>30$) Z follows the SN(0,1)
  - Small sample and unknown variance, Z follows the t-distribution.

Calculation of Confidence Intervals
for the Mean

- Find the 2 values in the tails of the distribution of z, such that:

$$P\left(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}\right) = (1 - \alpha)$$
$$P\left(-z_{\alpha/2} \leq \frac{\hat{\mu}-\mu}{SE(\hat{\mu})} \leq z_{\alpha/2}\right) = (1 - \alpha)$$

- Rearrange (rescale back from Z to Y):

$$P\left(\hat{\mu} - z_{\alpha/2}SE(\hat{\mu}) \leq \theta \leq \hat{\mu} + z_{\alpha/2}SE(\hat{\mu})\right) = (1 - \alpha)$$

95% Confidence Interval
for the Mean

- If we assume we have a large enough sample, Z will be standard normally distributed, hence the critical values would be:

$$\pm z_{\alpha/2} = 1.96.$$

- Hence, the 95% Confidence Interval for the mean, would be an interval around the estimate for the mean:

$$P\left(\hat{\mu} - 1.96\, SE(\hat{\mu}) \leq \mu \leq \hat{\mu} + 1.96\, SE(\hat{\mu})\right) = (1 - \alpha)$$

- If the sample size is small, Z is distributed as t, and the critical value would be a bit bigger than 1.96, depending on the degrees of freedom.

Confidence Interval

Z	$\boldsymbol{\mu = \hat{\mu} + SE(\hat{\mu})Z}$	Prob.	What it means
$P(-1 < Z < 1)$	$P(\hat{\mu} - SE(\hat{\mu}) < \mu < \hat{\mu} + SE(\hat{\mu}))$	$= 0.6826$	Prob. of being 1 s.d. from sample mean is 68.26%. **OR** The 68% confidence interval for the true mean.
$P(-1.96 < Z < 1.96)$	$P(\hat{\mu} - 1.96\ SE(\hat{\mu}) < \mu < \hat{\mu} + 1.96\ SE(\hat{\mu}))$	$= 0.95$	95% of the normal distribution lies within 1.96 s.d. from mean. **OR** The 95% confidence interval for the mean.
$P(-2.57 < Z < 2.57)$	$P(\hat{\mu} - 2.57\ SE(\hat{\mu}) < \mu < \hat{\mu} + 2.57\ SE(\hat{\mu}))$	$= 0.99$	Prob. of being within 2.57 s.d. from mean is 99%. **OR** The 99% confidence interval for the mean.

Confidence Interval

An Example

- A machine is set up such that the average content of juice per bottle equals μ. A sample of 100 bottles yields an average content of 48cl with a standard deviation $s = 5$cl. Calculate a 95% confidence interval for the average content.