Review of Statistics

Econometrics

Péter Elek

Statistical Inference

- Population: all units of interest
- Sample: subset drawn from population
- Goal: infer population parameters from data
- Focus here is on independent identically distributed (i.i.d.) samples; parameter: population mean or variance
- Key tools: point estimation, interval estimation, hypothesis testing

Sample Statistics and Point Estimators

- Sample statistic: any function of the sample elements, i.e. $h(X_1, \ldots, X_n)$
- Point estimator: $\hat{\theta} = h(X_1, \ldots, X_n)$ for the estimation of parameter θ
- Sampling distribution: the distribution of the point estimator
- Examples:
    - Sample mean: $\hat{\mu} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ estimates μ (the population mean)
    - (Corrected) sample variance: $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ estimates σ^2 (the population variance)

Properties of the Sample Mean

- Let $X_1, X_2, \ldots, X_n$ be an i.i.d. sample with mean μ and variance σ^2.
- The sample mean $\hat{\mu} = \overline{X}$ is an estimator of μ.
- Properties:
- $E(\overline{X}) = \mu$, hence $\hat{\mu}$ is an unbiased estimator of μ.
- $Var(\overline{X}) = \sigma^2/n$ (why?)
- $sd(\overline{X}) = \sigma/\sqrt{n}$ (hence s.d. is proportional to $1/\sqrt{n}$)

Law of Large Numbers (LLN) and Consistency

- As the sample size gets large ($n \to \infty$), then $\lim_{n\to\infty} sd(\overline{X}) = 0$.
- Thus the sampling distribution of $\overline{X}$ "collapses" (gets more and more concentrated) around the population mean μ.
- The precise statement of this is the law of large numbers.
- We also say that $\overline{X}$ is a consistent estimator of μ.
- Notation: $\text{plim}_{n\to\infty} \overline{X} = E(X) = \mu$ (plim = "probability limit").

Central Limit Theorem (CLT) and Asymptotic Normality

- If $n \to \infty$ then $\overline{X}$ is approximately normally distributed:

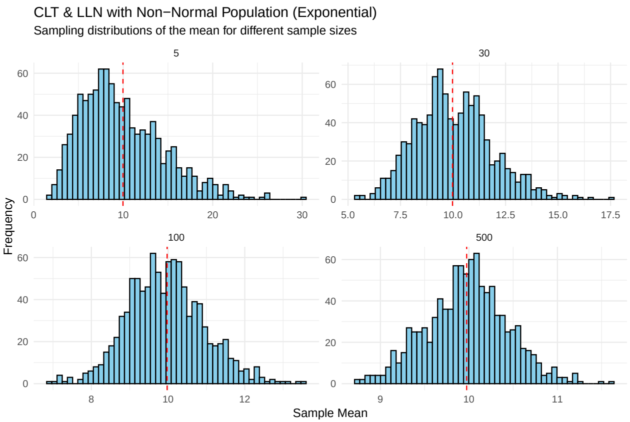
$$\overline{X} \overset{A}{\sim} N\left(\mu, \sigma^2/n\right)$$

- The precise statement of this is the Central Limit Theorem (CLT).
- We also say that $\overline{X}$ is an asymptotically normal estimator of μ.
- (The original sample elements X_i are not necessarily normally distributed.)
- Also, in large samples, σ can be changed to the sample standard deviation:

$$\overline{X} \overset{A}{\sim} N\left(\mu, s^2/n\right)$$

- ($s/\sqrt{n}$ is often called the standard error of $\overline{X}$)

Simulation of LLN and CLT



Large-sample confidence interval (CI) and test for μ

- The large-sample $1-\alpha$ CI for μ is

$$\left[\overline{x} - z_{1-\alpha/2} * \frac{s}{\sqrt{n}}; \overline{x} + z_{1-\alpha/2} * \frac{s}{\sqrt{n}}\right]$$

- where $z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of $N(0,1)$.
- Similarly for hypothesis testing: the usual z-test works in large samples (no need for t-distribution).

Exercise

Given a random sample of size $n = 100$ of adult Hungarian males, the average height is 175.2cm with a sample standard deviation of 7.1cm. Give a 95% confidence interval for the average height of an adult Hungarian male.

Material

- Wooldridge Appendix C.1, C.2 (excl. efficiency), C.3, C.5, C.6 (excl. t-test), C.7