Mathematical Statistics Review

Tyler J. Brough

February 8, 2016

Introduction

This is a review of fundamental mathematical statistics that will be essential for learning econometrics. The coverage is based on Wooldridge (Appendix C).

What is Statistics?

Statistical inference is the process of learning something about a population given a sample from that population. Using the tools of statistics we will seek to *infer* something about the population, given only a sample.

A **population** is a well defined group of subjects, such as individuals, firms, cities, etc.

By learning, we mainly mean two things:

- Estimation
- Hypothesis Testing

Example 1

An example of a population is all working adults in the US. Labor economists are interested in learning about the return to education, measured by the average increase in earnings given another year of education. It is impractical or impossible to gather data on the entire population, but she can obtain data on a subset of the population. Using the data collected a labor economist may report that her best estimate of the return to another year of education is 7.5%. This is an example of a **point estimate**. Or she may report a range, such as "the return to education is between 5.6% and 9.4%." This is an example of an **interval estimate**.

Example 2

An urban economist might want to know whether neighborhood crime watch programs are associated with lower crime rates. After comparing crime rates of neighborhoods with and without such programs in a sample from the population, he can draw one of two conclusions: neighborhood watch programs do affect crime, or they do not. This is an example of **hypothesis testing**.

Population, Models, & Parameters

The first step in statistical inference is to identify the population of interest. Once a population has been identified, a model for the population relationship of interest may be specified. Models involve probability distributions or features of probability distributions, and these depend on unknown parameters. **Parameters** are constants that determine the directions and strengths of relationships among variables.

In the labor economics example, the parameter of interest is the return to education in the population.

Sampling

Let Y be a random variable representing a population with PDF $f(y;\theta)$, which depends on a single parameter θ. The PDF is assumed to be known, except for θ. Different values of θ imply different population distributions, and therefore we are interested in θ. If we can obtain samples from the population we can learn something about θ.

Random sampling

- If $Y_1, Y_2, \dots Y_n$ are independent random variables with a common PDF $f(y;\theta)$ then $\{Y_1, Y_2, \dots, Y_n\}$ is said to be a *random sample* from $f(y,\theta)$ (a random sample represented by $f(y;\theta)$).
- When $\{Y_1, Y_2, \dots, Y_n\}$ is a random sample from $f(y,\theta)$, we also say that the Y_i are **independent and identically distributed** (or iid) random variables from $f(y;\theta)$.

If family income is obtained for $n = 100$ families in the US, the incomes we observe will differ for each sample of 100 that we choose. Once a sample is obtained we have a set of number $\{y_1, y_2, \dots, y_3\}$, which constitute the data that we work with.

Random Sample from a Bernoulli Distribution

Random samples from Bernoulli distributions are often used to illustrate statistical concepts. If $Y_1, Y_2, \ldots, Y_n$ are iid Bernoulli(θ), such that $P(Y_i = 1) = \theta$ and $P(Y_i = 0) = 1 - \theta$ then $\{Y_1, Y_2, \ldots, Y_n\}$ constitute a random sample from a Bernoull(θ) distribution.

The Airline Example

Consider the airline example: Each Y_i denotes whether or not passenger i shows up. θ is the probability that a randomly drawn individual from the population shows up.

Random Samples from the Normal Distribution

For many applications, random samples can be assumed to be drawn from a normal distribution. If $\{Y_1, Y_2, \ldots, Y_n\}$ is a random sample from the Normal(μ, σ^2) population, the population is characterized by two parameters, the mean mu and the variance σ^2.

Finite Sample Properties

Finite sample properties are properties that hold for a sample of any size, no matter how small or large (sometimes called "small sample properties" to distinguish from "asymptotic properties").

Estimation and Estimators

Given a random sample drawn from a population distribution that depends on an unknown parameter θ. An **estimator** of θ is a rule that assigns each possible outcome of the sample a value of θ. The rule is specified before any sampling is carried out (regardless of the data collected).

Let $\{Y_1, Y_2, \ldots, Y_n\}$ be a random sample from a population with mean μ. A natural estimator of μ is the average of the random sample:

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$$

$\bar{Y}$ is called the **sample average**; unlike earlier when we defined the average as a descriptive statistics, $\bar{Y}$ is now viewed as an estimator. Given any outcome of the random variables $Y_1, Y_2, \ldots, Y_n$, we use the same rule to estimate μ: we average them. For actual outcomes $\{y_1, y_2, \ldots, y_n\}$, the estimate is just the average in the sample.

$$\bar{y} = \frac{(y_1 + y_2 + \cdots + y_n)}{n}$$

More generally an estimator W of a parameter θ can be expressed as:

$$W = h(Y_1, Y_2, \ldots, Y_n)$$

for some known function h of the random variables $Y_1, Y_2, \ldots, Y_n$. W is a random variable because it depends on the random sample: as we obtain different random samples from the population, the value of W can change.

When a particular set of numbers $\{y_1, y_2, \ldots, y_n\}$ is plugged into h, we obtain an *estimate* of θ, denoted $w = h(y_1, y_2, \ldots, y_n)$.

So we have that:

- W is a point estimator
- w is a point estimate

to evaluate estimation procedures we study various properties of the PDF of W. The distribution of an estimator is called its **sampling distribution**. In mathematical statistics, we study the sampling distributions of estimators.

Unbiasedness: an estimator, W of θ, is an unbiased estimator if

$$E(W) = \theta$$

for all possible values of θ.

Remarks:

- If an estimator is unbiased, then its PDF has an expected value equal to the parameter it is estimating. However, in any given sample $E(W)$ may not equal θ.
- Rather, if we could indefinitely draw samples from the population, getting an estimate each time, and then average these estimates over all random samples we would obtain θ.
- This is just a thought experiment, because in reality we have only one sample to work with. But this “what if” property is desirable for estimators.

If W is a **biased estimator** of θ, its bias is defined as

$$Bias(W) = E(W) - \theta$$

Example: $\bar{Y}$ is an unbiased estimator of the population mean, μ

$$\begin{aligned} E(\bar{Y}) &= E\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) \\ &= \frac{1}{n}E\left(\sum_{i=1}^{n} Y_i\right) = \frac{1}{n}\left(\sum_{i=1}^{n} E(Y_i)\right) \\ &= \frac{1}{n}\sum_{i=1}^{n} \mu = \frac{1}{n}n\mu = \mu \end{aligned}$$

Example: s^2 is an unbiased estimator of σ^2.

Let $\{Y_1, Y_1, \ldots, Y_n\}$ denote a random sample from the population with

- $E(Y) = \mu$
- $Var(Y)\sigma^2$

then

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

This is usually called the **sample variance**.

Note: the division by $n-1$ accounts for the fact that μ is estimated by $\bar{Y}$ and not known. If μ were known $\frac{1}{n}\sum\limits_{i=1}^{n}(Y_i - \mu)^2$, would be an unbiased estimator.

Unbiasedness has some weaknesses:

- Some very reasonable, even very good, estimates are not unbiased.
- Some unbiased estimates are quite poor.

For example:

$$W = Y_1 (\text{i.e. discard all other observations})$$

It is an unbiased estimator $E(Y_1) = \mu$.

Example: If $n = 100$, we have one hundred observation of the random variable Y, but we discard all but the first to estimate $E(Y)$.

The weaknesses of unbiasedness show that we need additional criteria to evaluate estimators. Unbiasedness ensures that the sampling distribution of an estimator has a mean value equal to the parameter it is estimating.

We also want to know how spread out it is. The variance of an estimator is called its **sampling variance** because it is the variance associated with the sampling distribution.

Example:

$$\begin{aligned} Var(\bar{Y}) &= Var\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) \\ &= \frac{1}{n^2} Var\left(\sum_{i=1}^{n} Y_i\right) = \frac{1}{n^2}\left(\sum_{i=1}^{n} Var(Y_i)\right) \\ &= \frac{1}{n^2}\left(\sum_{i=1}^{n} \sigma^2\right) = \frac{1}{n^2} n\sigma^2 = \frac{1}{n}\sigma^2 \end{aligned}$$

References

Wooldridge, Jeffrey (2009) *Introductory Econometrics: A Modern Approach 4th Edition.*.