

Definition of Estimator and Choosing among Estimators: Lecture XIX

Charles B. Moss

October 12, 2026

Reading Question

- Suppose that we know that $m_1 = 0.73678$ and $m_2 = 1.07054$. What is the method of moments estimator of the parameters of the normal distribution? What are the method of moments estimator for the parameters of the gamma distribution? [Hint: The first, second, and third moments of the gamma distribution are mean $k\theta$, variance $k\theta^2$, skewness $2/\sqrt{k}$, and excess kurtosis $6/k$]

Moment	1	2	3	4
Non-Central	0.73678	1.07054	2.36991	6.86821
Central	0.73678	0.53481	2.09636	8.93008

- 1 What is an Estimator?
 - Properties of the Sample Mean
- 2 Estimators in General
 - Nonparametric Estimation
 - Properties of Estimators
 - Measures of Closeness

What is an Estimator?

- In the next several lectures we will be discussing statistical estimators and estimation. The book divides this discussion into the estimation of a single number such as a mean or standard deviation or the estimation of a range such as a confidence interval.
- At the most basic level, the definition of an estimator involves the distinction between a sample and a population.
 - In general we assume that we have a random variable (X) with some distribution function.
 - Next, we assume that we want to estimate something about that population, for example we may be interested in estimating the mean of the population or probability that the outcome will lie between two numbers.
 - For example, in a farm-planning model we may be interested in estimating the expected return for a particular crop.
 - In a regression context, we may be interested in estimating the average effect of price or income on the quantity of goods consumed.

- Common point estimators are the sample moments
 - Sample Mean

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad (1)$$

- Sample Variance

$$S_X^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n-1} \sum_{i=1}^n X_i^2 - (\bar{X})^2 \quad (2)$$

- The k^{th} sample moment around zero

$$M^k = \frac{1}{n} \sum_{i=1}^n X_i^k \quad (3)$$

- Continued

- The k^{th} moment around the mean

$$\tilde{M}^k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k \quad (4)$$

- Sample Covariance

$$\text{Cov}(X, Y) = S_{xy} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) \quad (5)$$

- Sample Correlation

$$\rho_{xy} = \frac{S_{xy}}{S_{xx}S_{yy}} \quad (6)$$

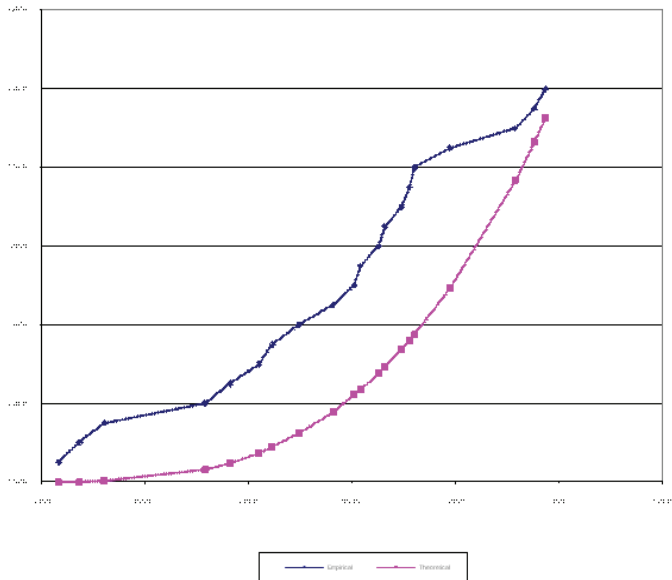
- Focusing on the sample versus population dichotomy for a moment, the sample image of X , denoted X^* and the empirical distribution function for can be depicted as a discrete distribution function with probability $1/n$.
 - Using the sample distribution from the first exam as an example. The sample distribution is presented in Table 8

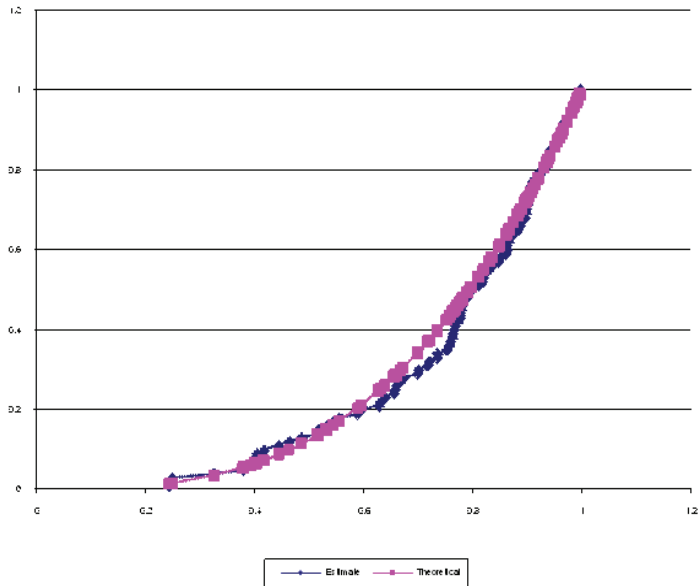
Empirical and Sample Distribution

Obs	X	Probability Density Function	Cumulative Density Function	
			Empirical	Theoretical
1	0.0317	0.0500	0.0500	0.0000
2	0.0720	0.0500	0.1000	0.0004
3	0.1208	0.0500	0.1500	0.0018
4	0.3148	0.0500	0.2000	0.0312
5	0.3644	0.0500	0.2500	0.0484
6	0.4200	0.0500	0.3000	0.0741
7	0.4458	0.0500	0.3500	0.0886
8	0.4971	0.0500	0.4000	0.1229
9	0.5633	0.0500	0.4500	0.1787
10	0.6044	0.0500	0.5000	0.2208
11	0.6163	0.0500	0.5500	0.2341
12	0.6522	0.0500	0.6000	0.2774
13	0.6637	0.0500	0.6500	0.2924
14	0.6965	0.0500	0.7000	0.3378
15	0.7113	0.0500	0.7500	0.3598
16	0.7210	0.0500	0.8000	0.3748
17	0.7892	0.0500	0.8500	0.4915
18	0.9154	0.0500	0.9000	0.7670
19	0.9528	0.0500	0.9500	0.8650
20	0.9747	0.0500	1.0000	0.9260

- Continued

- The empirical distribution function is depicted in Figure 1.
- Figure 2 depicts the same relationship extending the sample to 100.





- Properties of the sample mean
 - Using Theorem 4.1.6, we know that

$$E[\bar{X}] = E[X] \quad (7)$$

which means that the population mean is close to a “center” of the distribution of the sample mean.

- Suppose $V(X) = \sigma^2$ is finite. Then using Theorem 4.3.3, we know that

$$V(\bar{X}) = \frac{\sigma^2}{n} \quad (8)$$

which shows that the degree of dispersion of the distribution of the sample mean around the population mean is inversely related to the sample size n .

- Continued

- Using Theorem 6.2.1 (Khinchine's law of large numbers), we know that

$$\text{plim}_{n \rightarrow \infty} \bar{X} = E[X] \quad (9)$$

If $V(X)$ is finite, the same result follows from Equations 7 and 8 above because of Theorem 6.1.1 (Chebyshev).

Estimators in General

- In general, an estimator is a function of the sample, not based on population parameters.
 - First, the estimator is a known function of random variables

$$\hat{\theta} = \phi(X_1, X_2, X_3, \dots X_n) \quad (10)$$

The value of an estimator is then a random variable. As any other random variable, it is possible to define the distribution of the estimator based on distribution of the random variables in the sample. These distributions will be used in the next section to define confidence intervals.

- Continued

- Any function of the sample is referred to as a statistic.
- Most of the time in econometrics, we focus on the moments as sample statistics. Specifically, we may be interested in the sample means, or may use the sample covariances with the sample variances to define least squares estimators.
- Other statistics may be important. For example, we may be interested in the probability of a given die role (for example the probability of a three). If we define a new set of variables, Y , such that $Y = 1$ if $X = 3$ and $Y = 0$ otherwise, the probability of a three becomes

$$\hat{p}_3 = \frac{1}{n} \sum_{i=1}^n Y_i \quad (11)$$

- Amemiya demonstrates that this probability could also be derived from the moments of the distribution.
 - Assume that you have a sample of 50 die roles. Compute the sample distribution for each moment $k = 0, 1, 2, 3, 4, 5$

$$m_k = \frac{1}{50} \sum_{i=1}^n X_i^k \quad (12)$$

where X_i is the value of the die role.

- Continued

- The method of moments estimate of each probability p_i is defined by the solution of the five equation system

$$m_0 = 1.00, \mu_0 = \sum_{j=1}^6 p_j, m_3 = \frac{1}{50} \sum_{i=1}^{50} r_i^3 = 61.32, \mu_3 = \sum_{j=1}^6 j^3 p_j$$

$$m_1 = \frac{1}{50} \sum_{i=1}^{50} r_i = 3.12, \mu_1 = \sum_{j=1}^6 j p_j, m_4 = \frac{1}{50} \sum_{i=1}^{50} r_i^4 = 316.32, \mu_4 = \sum_{j=1}^6 j^4 p_j \quad (13)$$

$$m_2 = \frac{1}{50} \sum_{i=1}^{50} r_i^2 = 12.84, \mu_2 = \sum_{j=1}^6 j^2 p_j, m_5 = \frac{1}{50} \sum_{i=1}^{50} r_i^5 = 1705.32, \mu_5 = \sum_{j=1}^6 j^5 p_j$$

Nonparametric Estimation

- **Distribution-Specific Method:** In these procedures, the distribution is assumed to belong to a certain class of distributions such as the negative exponential or normal distribution. These procedures can tailor the estimator to the estimation of specific distribution parameters.
- **Distribution-Free Method:** In these procedures, we do not specify a priori a distribution and typically restrict our estimation to the estimation of the first few moments of the distribution.

Properties of Estimators

- In this section, Amemiya compares three estimators of the probability of a Bernoulli distribution. The Bernoulli variable is simply $X = 1$ with probability p and $X = 0$ with probability $1 - p$. Given a sample of 2, the estimators are defined as

$$\begin{aligned} T &= \frac{(X_1 + X_2)}{2} \\ S &= X_1 \\ W &= \frac{1}{2} \end{aligned} \tag{14}$$

- The question (loosely phrased) is then which is the best estimator of p ? In answering this question, however, two kinds of ambiguities occur
 - For a particular value of the parameter, say $p = 3/4$, it is not clear which of the three estimators is preferred.
 - T dominates W for $p = 0$, but W dominates T for $p = 1/2$.

Measures of Closeness

- $P(|X - \theta| \leq |Y - \theta|) = 1$
- $E[g(X - \theta)] \leq E[g(Y - \theta)]$ for every continuous function $g(\cdot)$ which is nonincreasing for $x < 0$ and nondecreasing for $x > 0$.
- $E[g(|X - \theta|)] \leq E[g(|Y - \theta|)]$ for every continuous function $g(\cdot)$ which is nondecreasing.
- $P(|X - \theta| > \epsilon) \leq P(|Y - \theta| > \epsilon)$ for every ϵ .
- $E(X - \theta)^2 \leq E(Y - \theta)^2$.
- $P(|X - \theta| < |Y - \theta|) \geq P(|X - \theta| > |Y - \theta|)$.

Reading

- Moss 7.3.2 - 7.4.