5.2 The Method of Moments

The objective will be to develop methods that use the observed values of random variables (sample data) in order to gain information about the unknown and unobservable characteristic of the population.

Let \( X_1, \ldots, X_n \) be independent and identically distributed (iid) random variables (in statistical language, a random sample) with a pdf or pf \( f(x, \theta_1, \ldots, \theta_l) \), where \( \theta_1, \ldots, \theta_l \) are the unknown population parameters (characteristics of interest). For example, a normal pdf has parameters \( \mu \) (the mean) and \( \sigma^2 \) (the variance). The actual values of these parameters are not known. The problem in point estimation is to determine statistics \( g_i(X_1, \ldots, X_n), i = 1, \ldots, l \), which can be used to estimate the value of each of the parameters—that is, to assign an appropriate value for the parameters \( \theta = (\theta_1, \ldots, \theta_l) \) based on observed sample data from the population. These statistics are called estimators for the parameters, and the values calculated from these statistics using particular sample data values are called estimates of the parameters. Estimators of \( \theta_i \) are denoted by \( \hat{\theta}_i \), where \( \hat{\theta}_i = g_i(X_1, \ldots, X_n), i = 1, \ldots, l \). Observe that the estimators are random variables. As a result, an estimator has a distribution (which we called sampling distribution in Chapter 4). When we actually run the experiment and observe the data, let the observed values of the random variables be \( X_1, \ldots, X_n \) be \( x_1, \ldots, x_n \); then, \( \hat{\theta}(X_1, \ldots, X_n) \) is an estimator, and its value \( \hat{\theta}(x_1, \ldots, x_n) \) is an estimate. For example, in case of the normal distribution, the parameters of interest are \( \theta_1 = \mu \), and \( \theta_2 = \sigma^2 \), that is, \( \theta = (\mu, \sigma^2) \). If the estimators of \( \mu \) and \( \sigma^2 \) are \( \bar{X} = (1/n) \sum_{i=1}^{n} X_i \) and \( S^2 = (1/(n-1)) \sum_{i=1}^{n} (X_i - \bar{X})^2 \) respectively, then, the corresponding estimates are \( \bar{x} = (1/n) \sum_{i=1}^{n} x_i \) and \( s^2 = (1/(n-1)) \sum_{i=1}^{n} (x_i - \bar{x})^2 \), the mean and variance corresponding to the particular observed sample values. In this book, we use capital letters such as \( \bar{X} \) and \( S^2 \) to represent the estimators, and lowercase letters such as \( \bar{x} \) and \( s^2 \) to represent the estimates.

There are many methods available for estimating the true value(s) of the parameter(s) of interest. Three of the more popular methods of estimation are the method of moments, the method of maximum likelihood, and Bayes’ method. A very popular procedure among econometricians to find a point estimator is the generalized method of moments. In this chapter we study only the method of moments and the method of maximum likelihood for obtaining point estimators and some of their desirable properties. In Chapter 11, we shall discuss Bayes’ method of estimation.

There are many criteria for choosing a desired point estimator. Heuristically, some of them can be explained as follows (detailed coverage is given in Sections 5.2 through 5.5). An estimator, \( \hat{\theta} \), is unbiased if the mean of its sampling distribution is the parameter \( \theta \). The bias of \( \hat{\theta} \) is given by \( B = E(\hat{\theta}) - \theta \). The estimator satisfies the consistency property if the sample estimator has a high probability of being close to the population value \( \theta \) for a large sample size. The concept of efficiency is based on comparing variances of the different unbiased estimators. If there are two unbiased estimators, it is desirable to have the one with the smaller variance. The estimator has the sufficiency property if it fully uses all the sample information. Minimal sufficient statistics are those that are sufficient for the parameter and are functions of every other set of sufficient statistics for those same parameters. A method due to Lehmann and Scheffé can be used to find a minimal sufficient statistic.

5.2 THE METHOD OF MOMENTS

How do we find a good estimator with desirable properties? One of the oldest methods for finding point estimators is the method of moments. This is a very simple procedure for finding an estimator for one or more population parameters. Let \( \mu_k = E[X^k] \) be the \( k \)th moment about the origin of a
random variable $X$, whenever it exists. Let $m'_k = (1/n) \sum_{i=1}^{n} X_i^k$ be the corresponding $k$th sample moment. Then, the estimator of $\mu'_k$ by the method of moments is $m'_k$. The method of moments is based on matching the sample moments with the corresponding population (distribution) moments and is founded on the assumption that sample moments should provide good estimates of the corresponding population moments. Because the population moments $\mu'_k = h_k(\theta_1, \theta_2, \ldots, \theta_l)$ are often functions of the population parameters, we can equate corresponding population and sample moments and solve for these parameters in terms of the moments.

**METHOD OF MOMENTS**

Choose as estimates those values of the population parameters that are solutions of the equations $\mu'_k = m'_k, k = 1, 2, \ldots, l$. Here $\mu'_k$ is a function of the population parameters.

For example, the first population moment is $\mu'_1 = E(X)$, and the first sample moment is $\bar{X} = \sum_{i=1}^{n} X_i/n$. Hence, the moment estimator of $\mu'_1$ is $\bar{X}$. If $k = 2$, then the second population and sample moments are $\mu'_2 = E(X^2)$ and $m'_2 = (1/n) \sum_{i=1}^{n} X_i^2$, respectively. Basically, we can use the following procedure in finding point estimators of the population parameters using the method of moments.

**THE METHOD OF MOMENTS PROCEDURE**

Suppose there are $l$ parameters to be estimated, say $\theta = (\theta_1, \ldots, \theta_l)$.

1. Find $l$ population moments, $\mu'_k, k = 1, 2, \ldots, l$. $\mu'_k$ will contain one or more parameters $\theta_1, \ldots, \theta_l$.
2. Find the corresponding $l$ sample moments, $m'_k, k = 1, 2, \ldots, l$. The number of sample moments should equal the number of parameters to be estimated.
3. From the system of equations, $\mu'_k = m'_k, k = 1, 2, \ldots, l$, solve for the parameter $\theta = (\theta_1, \ldots, \theta_l)$; this will be a moment estimator of $\hat{\theta}$.

The following examples illustrate the method of moments for population parameter estimation.

**Example 5.2.1**

Let $X_1, \ldots, X_n$ be a random sample from a Bernoulli population with parameter $p$.

(a) Find the moment estimator for $p$.

(b) Tossing a coin 10 times and equating heads to value 1 and tails to value 0, we obtained the following values:

$0 \ 1 \ 1 \ 0 \ 1 \ 0 \ 1 \ 1 \ 1 \ 0$

Obtain a moment estimate for $p$, the probability of success (head).
5.2 The Method of Moments

Solution

(a) For the Bernoulli random variable, \( \mu' = E[X] = p \), so we can use \( m_1' \) to estimate \( p \). Thus,

\[
m_1' = \hat{p} = \frac{1}{n} \sum_{i=1}^{n} X_i.
\]

Let

\[
Y = \sum_{i=1}^{n} X_i.
\]

Then, the method of moments estimator for \( p \) is \( \hat{p} = Y/n \). That is, the ratio of the total number of heads to the total number of tosses will be an estimate of the probability of success.

(b) Note that this experiment results in Bernoulli random variables. Thus, using part (a) with \( Y = 6 \), we get the moment estimate of \( p \) is \( \hat{p} = \frac{6}{10} = 0.6 \).

We would use this value \( \hat{p} = 0.6 \), to answer any probabilistic questions for the given problem. For example, what is the probability of exactly obtaining 8 heads out of 10 tosses of this coin? This can be obtained by using the binomial formula, with \( \hat{p} = 0.6 \), that is,

\[
P(X = 8) = \binom{10}{8} (0.6)^8 (0.4)^{10-8}.
\]

In Example 5.2.1, we used the method of moments to find a single parameter. We demonstrate in Example 5.2.2 how this method is used for estimating more than one parameter.

Example 5.2.2

Let \( X_1, \ldots, X_n \) be a random sample from a gamma probability distribution with parameters \( \alpha \) and \( \beta \). Find moment estimators for the unknown parameters \( \alpha \) and \( \beta \).

Solution

For the gamma distribution (see Section 3.2.5),

\[
E[X] = \alpha \beta \quad \text{and} \quad E\left[X^2\right] = \alpha \beta^2 + \alpha^2 \beta^2.
\]

Because there are two parameters, we need to find the first two moment estimators. Equating sample moments to distribution (theoretical) moments, we have

\[
\frac{1}{n} \sum_{i=1}^{n} X_i = \overline{X} = \alpha \beta, \quad \text{and} \quad \frac{1}{n} \sum_{i=1}^{n} X_i^2 = \alpha \beta^2 + \alpha^2 \beta^2.
\]

Solving for \( \alpha \) and \( \beta \) we obtain the estimates as \( \alpha = (\overline{X}/\beta) \) and \( \beta = \left[\frac{\{(1/n) \sum_{i=1}^{n} X_i^2 - \overline{X}^2\}}{\overline{X}}\right]^{1/2} \).
Therefore, the method of moments estimators for $\alpha$ and $\beta$ are

$$\hat{\alpha} = \frac{\bar{X}}{\hat{\beta}}$$

and

$$\hat{\beta} = \frac{1}{n} \sum_{i=1}^{n} X_i^2 - \bar{X}^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n\bar{X}},$$

which implies that

$$\hat{\alpha} = \frac{\bar{X}}{\hat{\beta}} = \frac{\bar{X}^2}{\frac{1}{n} \sum_{i=1}^{n} X_i^2 - \bar{X}^2} = \frac{\bar{X}^2}{\sum_{i=1}^{n} (X_i - \bar{X})^2}.$$

Thus, we can use these values in the gamma pdf to answer questions concerning the probabilistic behavior of the r.v. $X$.

---

**Example 5.2.3**

Let the distribution of $X$ be $N(\mu, \sigma^2)$.

(a) For a given sample of size $n$, use the method of moments to estimate $\mu$ and $\sigma^2$.

(b) The following data (rounded to the third decimal digit) were generated using Minitab from a normal distribution with mean 2 and a standard deviation of 1.5.

<table>
<thead>
<tr>
<th>3.163</th>
<th>1.883</th>
<th>3.252</th>
<th>3.716</th>
<th>-0.049</th>
<th>-0.653</th>
<th>0.057</th>
<th>2.987</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.098</td>
<td>1.670</td>
<td>1.396</td>
<td>2.332</td>
<td>1.838</td>
<td>3.024</td>
<td>2.706</td>
<td>0.231</td>
</tr>
<tr>
<td>3.830</td>
<td>3.349</td>
<td>-0.230</td>
<td>1.496</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Obtain the method of moments estimates of the true mean and the true variance.

**Solution**

(a) For the normal distribution, $E(X) = \mu$, and because $\text{Var}(X) = EX^2 - \mu^2$, we have the second moment as $E(X^2) = \sigma^2 + \mu^2$.

Equating sample moments to distribution moments we have

$$\frac{1}{n} \sum_{i=1}^{n} X_i = \mu' = \mu$$

and

$$\mu'_2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2 = \sigma^2 + \mu^2.$$
Solving for $\mu$ and $\sigma^2$, we obtain the moment estimators as

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

and

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2 - \overline{X}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2.$$ 

(b) Because we know that the estimator of the mean is $\hat{\mu} = \overline{X}$ and the estimator of the variance is $\hat{\sigma}^2 = (1/n) \sum_{i=1}^{n} X_i^2 - \overline{X}^2$, from the data the estimates are $\hat{\mu} = 2.005$, and $\hat{\sigma}^2 = 6.12 - (2.005)^2 = 2.1$. Notice that the true mean is 2 and the true variance is 2.25, which we used to simulate the data.

In general, using the population pdf we evaluate the lower order moments, finding expressions for the moments in terms of the corresponding parameters. Once we have population (theoretical) moments, we equate them to the corresponding sample moments to obtain the moment estimators.

**Example 5.2.4**
Let $X_1, \ldots, X_n$ be a random sample from a uniform distribution on the interval $[a, b]$. Obtain method of moment estimators for $a$ and $b$.

**Solution**
Here, $a$ and $b$ are treated as parameters. That is, we only know that the sample comes from a uniform distribution on some interval, but we do not know from which interval. Our interest is to estimate this interval.

The pdf of a uniform distribution is

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{otherwise} \end{cases}$$

Hence, the first two population moments are

$$\mu_1 = E(X) = \int_{a}^{b} \frac{x}{b-a} dx = \frac{a+b}{2} \quad \text{and} \quad \mu_2 = E(X^2) = \int_{a}^{b} \frac{x^2}{b-a} dx = \frac{a^2 + ab + b^2}{3}.$$ 

The corresponding sample moments are

$$\hat{\mu}_1 = \overline{X} \quad \text{and} \quad \hat{\mu}_2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2.$$ 

Equating the first two sample moments to the corresponding population moments, we have

$$\hat{\mu}_1 = \frac{a+b}{2} \quad \text{and} \quad \hat{\mu}_2 = \frac{a^2 + ab + b^2}{3}$$
which, solving for $a$ and $b$, results in the moment estimators of $a$ and $b$,

\[ \hat{a} = \hat{\mu}_1 - \sqrt{3} (\hat{\mu}_2 - \hat{\mu}_1^2) \quad \text{and} \quad \hat{b} = \hat{\mu}_1 + \sqrt{3} (\hat{\mu}_2 - \hat{\mu}_1^2). \]

In Example 5.2.4, if $a = -b$, that is, $X_1, \ldots, X_n$ is a random sample from a uniform distribution on the interval $(-b, b)$, the problem reduces to a one-parameter estimation problem. However, in this case $E(X_i) = 0$, so the first moment cannot be used to estimate $b$. It becomes necessary to use the second moment. For the derivation, see Exercise 5.2.3.

It is important to observe that the method of moments estimators need not be unique. The following is an example of the nonuniqueness of moment estimators.

**Example 5.2.5**

Let $X_1, \ldots, X_n$ be a random sample from a Poisson distribution with parameter $\lambda > 0$. Show that both $(1/n) \sum_{i=1}^n X_i$ and $(1/n) \sum_{i=1}^n X_i^2 - (1/n \sum_{i=1}^n X_i)^2$ are moment estimators of $\lambda$.

**Solution**

We know that $E(X) = \lambda$, from which we have a moment estimator of $\lambda$ as $(1/n) \sum_{i=1}^n X_i$. Also, because we have $\text{Var}(X) = \lambda$, equating the second moments, we can see that

\[ \lambda = E(X^2) - (EX)^2, \]

so that

\[ \hat{\lambda} = \frac{1}{n} \sum_{i=1}^n X_i^2 - \left( \frac{1}{n} \sum_{i=1}^n X_i \right)^2. \]

Thus,

\[ \hat{\lambda} = \frac{1}{n} \sum_{i=1}^n X_i \]

and

\[ \hat{\lambda} = \frac{1}{n} \sum_{i=1}^n X_i^2 - \left( \frac{1}{n} \sum_{i=1}^n X_i \right)^2. \]

Both are moment estimators of $\lambda$. Thus, the moment estimators may not be unique. We generally choose $\bar{X}$ as an estimator of $\lambda$, for its simplicity.
It is important to note that, in general, we have as many moment conditions as the parameters. In Example 5.2.5, we have more moment conditions than parameters, because both the mean and variance of Poisson random variables are the same. Given a sample, this results in two different estimates of a single parameter. One of the questions could be, can these two estimators be combined in some optimal way? This is done by the so-called generalized method of moments (GMM). We will not deal with this topic.

As we have seen, the method of moments finds estimators of unknown parameters by equating the corresponding sample and population moments. This method often provides estimators when other methods fail to do so or when estimators are harder to obtain, as in the case of a gamma distribution. Compared to other methods, method of moments estimators are easy to compute and have some desirable properties that we will discuss in ensuing sections. The drawback is that they are usually not the “best estimators” (to be defined later) available and sometimes may even be meaningless.

**EXERCISES 5.2**

**5.2.1.** Let $X_1, \ldots, X_n$ be a random sample of size $n$ from the geometric distribution for which $p$ is the probability of success.

(a) Use the method of moments to find a point estimator for $p$.

(b) Use the following data (simulated from geometric distribution) to find the moment estimator for $p$:

\[
2 \quad 5 \quad 7 \quad 43 \quad 18 \quad 19 \quad 16 \quad 11 \quad 22 \\
4 \quad 34 \quad 19 \quad 21 \quad 23 \quad 6 \quad 21 \quad 7 \quad 12
\]

How will you use this information? [The pdf of a geometric distribution is $f(x) = p(1 - p)^{x-1}$, for $x = 1, 2, \ldots$. Also $\mu = 1/p$.]

**5.2.2.** Let $X_1, \ldots, X_n$ be a random sample of size $n$ from the exponential distribution whose pdf (by taking $\theta = 1/\beta$ in Definition 2.3.7) is

\[
f(x, \theta) = \begin{cases} 
\theta e^{-\theta x}, & x \geq 0 \\
0, & x < 0.
\end{cases}
\]

(a) Use the method of moments to find a point estimator for $\theta$.

(b) The following data represent the time intervals between the emissions of beta particles.

\[
0.9 \quad 0.1 \quad 0.1 \quad 0.8 \quad 0.9 \quad 0.1 \quad 0.7 \quad 1.0 \quad 0.2 \\
0.1 \quad 0.1 \quad 0.1 \quad 2.3 \quad 0.8 \quad 0.3 \quad 0.2 \quad 0.1 \quad 1.0 \quad 0.9 \\
0.1 \quad 0.5 \quad 0.4 \quad 0.6 \quad 0.2 \quad 0.4 \quad 0.2 \quad 0.1 \quad 0.8 \quad 0.2 \\
0.5 \quad 3.0 \quad 1.0 \quad 0.5 \quad 0.2 \quad 2.0 \quad 1.7 \quad 0.1 \quad 0.3 \quad 0.1 \\
0.4 \quad 0.5 \quad 0.8 \quad 0.1 \quad 0.1 \quad 1.7 \quad 0.1 \quad 0.2 \quad 0.3 \quad 0.1
\]
Assuming the data follow an exponential distribution, obtain a moment estimate for the parameter \( \theta \). Interpret.

5.2.3. Let \( X_1, \ldots, X_n \) be a random sample from a uniform distribution on the interval \((\theta - 1, \theta + 1)\).
   (a) Find a moment estimator for \( \theta \).
   (b) Use the following data to obtain a moment estimate for \( \theta \):

   \[
   11.72 \quad 12.81 \quad 12.09 \quad 13.47 \quad 12.37
   \]

5.2.4. The probability density of a one-parameter Weibull distribution is given by

   \[
   f(x) = \begin{cases} 
   2\alpha x e^{-\alpha x^2}, & x > 0 \\
   0, & \text{otherwise.}
   \end{cases}
   \]

   (a) Using a random sample of size \( n \), obtain a moment estimator for \( \alpha \).
   (b) Assuming that the following data are from a one-parameter Weibull population,

   \[
   1.87 \quad 1.60 \quad 2.36 \quad 1.12 \quad 0.15 \\
   1.83 \quad 0.64 \quad 1.53 \quad 0.73 \quad 2.26
   \]

   obtain a moment estimate of \( \alpha \).

5.2.5. Let \( X_1, \ldots, X_n \) be a random sample from the truncated exponential distribution with pdf

   \[
   f(x) = \begin{cases} 
   e^{-(x-\theta)}, & x \geq \theta \\
   0, & \text{otherwise.}
   \end{cases}
   \]

   Find the method of moments estimate of \( \theta \).

5.2.6. Let \( X_1, \ldots, X_n \) be a random sample from a distribution with pdf

   \[
   f(x, \alpha) = \frac{1 + \alpha x}{2}, \quad -1 \leq x \leq 1, \text{ and } -1 \leq \alpha \leq 1.
   \]

   Find the moment estimators for \( \alpha \).

5.2.7. Let \( X_1, \ldots, X_n \) be a random sample from a population with pdf

   \[
   f(x) = \begin{cases} 
   \frac{2\alpha^2}{x^3}, & x \geq \alpha \\
   0, & \text{otherwise.}
   \end{cases}
   \]

   Find a method of moments estimator for \( \alpha \).

5.2.8. Let \( X_1, \ldots, X_n \) be a random sample from a negative binomial distribution with pmf

   \[
   p(x, r, p) = \binom{x + r - 1}{r - 1} p^x (1 - p)^r, 0 \leq p \leq 1, x = 0, 1, 2, \ldots
   \]
Find method of moments estimators for $r$ and $p$. [Here $E[X] = r(1 - p)/p$ and $E[X^2] = r(1 - p)(r - rp + 1)/p^2$.]

5.2.9. Let $X_1, \ldots, X_n$ be a random sample from a distribution with pdf

$$f(x) = \begin{cases} (\theta + 1)x^\theta, & 0 \leq x \leq 1; \theta > -1 \\ 0, & \text{otherwise.} \end{cases}$$

Use the method of moments to obtain an estimator of $\theta$.

5.2.10. Let $X_1, \ldots, X_n$ be a random sample from a distribution with pdf

$$f(x) = \begin{cases} \frac{2\beta-2x}{\beta^2}, & 0 < x < \beta \\ 0, & \text{otherwise.} \end{cases}$$

Use the method of moments to obtain an estimator of $\beta$.

5.2.11. Let $X_1, \ldots, X_n$ be a random sample with common mean $\mu$ and variance $\sigma^2$. Obtain a method of moments estimator for $\sigma$.

5.2.12. Let $X_1, \ldots, X_n$ be a random sample from the beta distribution with parameters $\alpha$ and $\beta$. Find the method of moments estimator for $\alpha$ and $\beta$.

5.2.13. Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with unknown mean $\mu$ and variance $\sigma^2$. Show that the method of moments estimators for $\mu$ and $\sigma^2$ are, respectively, the sample mean $\bar{X}$ and $S^2 = \left(\frac{1}{n}\right) \sum_{i=1}^{n} (X - \bar{X})^2$. Note that $S^2 = \left(\frac{n-1}{n}\right) S^2$ where $S^2$ is the sample variance.

5.3 THE METHOD OF MAXIMUM LIKELIHOOD

It is highly desirable to have a method that is generally applicable to the construction of statistical estimators that have “good” properties. In this section we present an important method for finding estimators of parameters proposed by geneticist/statistician Sir Ronald A. Fisher around 1922 called the method of maximum likelihood. Even though the method of moments is intuitive and easy to apply, it usually does not yield “good” estimators. The method of maximum likelihood is intuitively appealing, because we attempt to find the values of the true parameters that would have most likely produced the data that we in fact observed. For most cases of practical interest, the performance of maximum likelihood estimators is optimal for large enough data. This is one of the most versatile methods for fitting parametric statistical models to data. First, we define the concept of a likelihood function.

**Definition 5.3.1** Let $f(x_1, \ldots, x_n; \theta), \theta \in \Theta \subseteq \mathbb{R}^k$, be the joint probability (or density) function of $n$ random variables $X_1, \ldots, X_n$ with sample values $x_1, \ldots, x_n$. The **likelihood function** of the sample is given by

$$L(\theta; x_1, \ldots, x_n) = f(x_1, \ldots, x_n; \theta), \; [= L(\theta), \text{in a briefer notation}].$$

We emphasize that $L$ is a function of $\theta$ for fixed sample values.