

## 7.2: The Method of Moments

### Basic Theory

#### The Method

Suppose that we have a basic random experiment with an observable, real-valued random variable  $X$ . The distribution of  $X$  has  $k$  unknown real-valued parameters, or equivalently, a parameter vector  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_k)$  taking values in a parameter space, a subset of  $\mathbb{R}^k$ . As usual, we repeat the experiment  $n$  times to generate a random sample of size  $n$  from the distribution of  $X$ .

$$\mathbf{X} = (X_1, X_2, \dots, X_n) \quad (7.2.1)$$

Thus,  $\mathbf{X}$  is a sequence of independent random variables, each with the distribution of  $X$ . The *method of moments* is a technique for constructing estimators of the parameters that is based on matching the *sample moments* with the corresponding *distribution moments*. First, let

$$\mu^{(j)}(\boldsymbol{\theta}) = \mathbb{E}(X^j), \quad j \in \mathbb{N}_+ \quad (7.2.2)$$

so that  $\mu^{(j)}(\boldsymbol{\theta})$  is the  $j$ th moment of  $X$  about 0. Note that we are emphasizing the dependence of these moments on the vector of parameters  $\boldsymbol{\theta}$ . Note also that  $\mu^{(1)}(\boldsymbol{\theta})$  is just the mean of  $X$ , which we usually denote simply by  $\mu$ . Next, let

$$M^{(j)}(\mathbf{X}) = \frac{1}{n} \sum_{i=1}^n X_i^j, \quad j \in \mathbb{N}_+ \quad (7.2.3)$$

so that  $M^{(j)}(\mathbf{X})$  is the  $j$ th sample moment about 0. Equivalently,  $M^{(j)}(\mathbf{X})$  is the sample mean for the random sample  $(X_1^j, X_2^j, \dots, X_n^j)$  from the distribution of  $X^j$ . Note that we are emphasizing the dependence of the sample moments on the sample  $\mathbf{X}$ . Note also that  $M^{(1)}(\mathbf{X})$  is just the ordinary sample mean, which we usually just denote by  $M$  (or by  $M_n$  if we wish to emphasize the dependence on the sample size). From our previous work, we know that  $M^{(j)}(\mathbf{X})$  is an unbiased and consistent estimator of  $\mu^{(j)}(\boldsymbol{\theta})$  for each  $j$ . Here's how the method works:

To construct the *method of moments* estimators  $(W_1, W_2, \dots, W_k)$  for the parameters  $(\theta_1, \theta_2, \dots, \theta_k)$  respectively, we consider the equations

$$\mu^{(j)}(W_1, W_2, \dots, W_k) = M^{(j)}(X_1, X_2, \dots, X_n) \quad (7.2.4)$$

consecutively for  $j \in \mathbb{N}_+$  until we are able to solve for  $(W_1, W_2, \dots, W_k)$  in terms of  $(M^{(1)}, M^{(2)}, \dots)$ .

The equations for  $j \in \{1, 2, \dots, k\}$  give  $k$  equations in  $k$  unknowns, so there is hope (but no guarantee) that the equations can be solved for  $(W_1, W_2, \dots, W_k)$  in terms of  $(M^{(1)}, M^{(2)}, \dots, M^{(k)})$ . In fact, sometimes we need equations with  $j > k$ . [Exercise 28](#) below gives a simple example. The method of moments can be extended to parameters associated with bivariate or more general multivariate distributions, by matching sample product moments with the corresponding distribution product moments. The method of moments also sometimes makes sense when the sample variables  $(X_1, X_2, \dots, X_n)$  are not independent, but at least are identically distributed. The [hypergeometric model](#) below is an example of this.

Of course, the method of moments estimators depend on the sample size  $n \in \mathbb{N}_+$ . We have suppressed this so far, to keep the notation simple. But in the applications below, we put the notation back in because we want to discuss asymptotic behavior.

#### Estimates for the Mean and Variance

Estimating the mean and variance of a distribution are the simplest applications of the method of moments. Throughout this subsection, we assume that we have a basic real-valued random variable  $X$  with  $\mu = \mathbb{E}(X) \in \mathbb{R}$  and  $\sigma^2 = \text{var}(X) \in (0, \infty)$ . Occasionally we will also need  $\sigma_4 = \mathbb{E}[(X - \mu)^4]$ , the fourth central moment. We sample from the distribution of  $X$  to produce a sequence  $\mathbf{X} = (X_1, X_2, \dots)$  of independent variables, each with the distribution of  $X$ . For each  $n \in \mathbb{N}_+$ ,  $\mathbf{X}_n = (X_1, X_2, \dots, X_n)$  is a random sample of size  $n$  from the distribution of  $X$ . We start by estimating the mean, which is essentially trivial by this method.

Suppose that the mean  $\mu$  is unknown. The method of moments estimator of  $\mu$  based on  $\mathbf{X}_n$  is the sample mean

$$M_n = \frac{1}{n} \sum_{i=1}^n X_i \quad (7.2.5)$$

1.  $\mathbb{E}(M_n) = \mu$  so  $M_n$  is unbiased for  $n \in \mathbb{N}_+$
2.  $\text{var}(M_n) = \sigma^2/n$  for  $n \in \mathbb{N}_+$  so  $\mathbf{M} = (M_1, M_2, \dots)$  is consistent.

Proof

It does not get any more basic than this. The method of moments works by matching the distribution mean with the sample mean. The fact that  $\mathbb{E}(M_n) = \mu$  and  $\text{var}(M_n) = \sigma^2/n$  for  $n \in \mathbb{N}_+$  are properties that we have seen several times before.

Estimating the variance of the distribution, on the other hand, depends on whether the distribution mean  $\mu$  is known or unknown. First we will consider the more realistic case when the mean is also unknown. Recall that for  $n \in \{2, 3, \dots\}$ , the sample variance based on  $\mathbf{X}_n$  is

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - M_n)^2 \quad (7.2.6)$$

Recall also that  $\mathbb{E}(S_n^2) = \sigma^2$  so  $S_n^2$  is unbiased for  $n \in \{2, 3, \dots\}$ , and that  $\text{var}(S_n^2) = \frac{1}{n} \left( \sigma_4 - \frac{n-3}{n-1} \sigma^4 \right)$  so  $\mathbf{S}^2 = (S_2^2, S_3^2, \dots)$  is consistent.

Suppose that the mean  $\mu$  and the variance  $\sigma^2$  are both unknown. For  $n \in \mathbb{N}_+$ , the method of moments estimator of  $\sigma^2$  based on  $\mathbf{X}_n$  is

$$T_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - M_n)^2 \quad (7.2.7)$$

1.  $\text{bias}(T_n^2) = -\sigma^2/n$  for  $n \in \mathbb{N}_+$  so  $\mathbf{T}^2 = (T_1^2, T_2^2, \dots)$  is asymptotically unbiased.
2.  $\text{mse}(T_n^2) = \frac{1}{n^3} [(n-1)^2 \sigma_4 - (n^2 - 5n + 3) \sigma^4]$  for  $n \in \mathbb{N}_+$  so  $\mathbf{T}^2$  is consistent.

Proof

As before, the method of moments estimator of the distribution mean  $\mu$  is the sample mean  $M_n$ . On the other hand,  $\sigma^2 = \mu^{(2)} - \mu^2$  and hence the method of moments estimator of  $\sigma^2$  is  $T_n^2 = M_n^{(2)} - M_n^2$ , which simplifies to the result above. Note that  $T_n^2 = \frac{n-1}{n} S_n^2$  for  $n \in \{2, 3, \dots\}$ .

1. Note that  $\mathbb{E}(T_n^2) = \frac{n-1}{n} \mathbb{E}(S_n^2) = \frac{n-1}{n} \sigma^2$ , so  $\text{bias}(T_n^2) = \frac{n-1}{n} \sigma^2 - \sigma^2 = -\frac{1}{n} \sigma^2$ .
2. Recall that  $\text{mse}(T_n^2) = \text{var}(T_n^2) + \text{bias}^2(T_n^2)$ . But  $\text{var}(T_n^2) = \left( \frac{n-1}{n} \right)^2 \text{var}(S_n^2)$ . The result follows from substituting  $\text{var}(S_n^2)$  given above and  $\text{bias}(T_n^2)$  in part (a).

Hence  $T_n^2$  is negatively biased and on average underestimates  $\sigma^2$ . Because of this result,  $T_n^2$  is referred to as the *biased sample variance* to distinguish it from the ordinary (unbiased) sample variance  $S_n^2$ .

Next let's consider the usually unrealistic (but mathematically interesting) case where the mean is known, but not the variance.

Suppose that the mean  $\mu$  is known and the variance  $\sigma^2$  unknown. For  $n \in \mathbb{N}_+$ , the method of moments estimator of  $\sigma^2$  based on  $\mathbf{X}_n$  is

$$W_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 \quad (7.2.8)$$

1.  $\mathbb{E}(W_n^2) = \sigma^2$  so  $W_n^2$  is unbiased for  $n \in \mathbb{N}_+$
2.  $\text{var}(W_n^2) = \frac{1}{n} (\sigma_4 - \sigma^4)$  for  $n \in \mathbb{N}_+$  so  $\mathbf{W}^2 = (W_1^2, W_2^2, \dots)$  is consistent.

Proof

These results follow since  $W_n^2$  is the sample mean corresponding to a random sample of size  $n$  from the distribution of  $(X - \mu)^2$ .

We compared the sequence of estimators  $S^2$  with the sequence of estimators  $W^2$  in the introductory section on Estimators. Recall that  $\text{var}(W_n^2) < \text{var}(S_n^2)$  for  $n \in \{2, 3, \dots\}$  but  $\text{var}(S_n^2)/\text{var}(W_n^2) \rightarrow 1$  as  $n \rightarrow \infty$ . There is no simple, general relationship between  $\text{mse}(T_n^2)$  and  $\text{mse}(S_n^2)$  or between  $\text{mse}(T_n^2)$  and  $\text{mse}(W_n^2)$ , but the asymptotic relationship is simple.

$\text{mse}(T_n^2)/\text{mse}(W_n^2) \rightarrow 1$  and  $\text{mse}(T_n^2)/\text{mse}(S_n^2) \rightarrow 1$  as  $n \rightarrow \infty$

Proof

In light of the previous remarks, we just have to prove one of these limits. The first limit is simple, since the coefficients of  $\sigma_4$  and  $\sigma^4$  in  $\text{mse}(T_n^2)$  are asymptotically  $1/n$  as  $n \rightarrow \infty$ .

It also follows that if both  $\mu$  and  $\sigma^2$  are unknown, then the method of moments estimator of the standard deviation  $\sigma$  is  $T = \sqrt{T^2}$ . In the unlikely event that  $\mu$  is known, but  $\sigma^2$  unknown, then the method of moments estimator of  $\sigma$  is  $W = \sqrt{W^2}$ .

## Estimating Two Parameters

There are several important special distributions with two parameters; some of these are included in the [computational exercises](#) below. With two parameters, we can derive the method of moments estimators by matching the distribution mean and variance with the sample mean and variance, rather than matching the distribution mean and second moment with the sample mean and second moment. This alternative approach sometimes leads to easier equations. To setup the notation, suppose that a distribution on  $\mathbb{R}$  has parameters  $a$  and  $b$ . We sample from the distribution to produce a sequence of independent variables  $\mathbf{X} = (X_1, X_2, \dots)$ , each with the common distribution. For  $n \in \mathbb{N}_+$ ,  $\mathbf{X}_n = (X_1, X_2, \dots, X_n)$  is a random sample of size  $n$  from the distribution. Let  $M_n$ ,  $M_n^{(2)}$ , and  $T_n^2$  denote the sample mean, second-order sample mean, and biased sample variance corresponding to  $\mathbf{X}_n$ , and let  $\mu(a, b)$ ,  $\mu^{(2)}(a, b)$ , and  $\sigma^2(a, b)$  denote the mean, second-order mean, and variance of the distribution.

If the method of moments estimators  $U_n$  and  $V_n$  of  $a$  and  $b$ , respectively, can be found by solving the first two equations

$$\mu(U_n, V_n) = M_n, \quad \mu^{(2)}(U_n, V_n) = M_n^{(2)} \quad (7.2.9)$$

then  $U_n$  and  $V_n$  can also be found by solving the equations

$$\mu(U_n, V_n) = M_n, \quad \sigma^2(U_n, V_n) = T_n^2 \quad (7.2.10)$$

Proof

Recall that  $\sigma^2(a, b) = \mu^{(2)}(a, b) - \mu^2(a, b)$ . In addition,  $T_n^2 = M_n^{(2)} - M_n^2$ . Hence the equations  $\mu(U_n, V_n) = M_n$ ,  $\sigma^2(U_n, V_n) = T_n^2$  are equivalent to the equations  $\mu(U_n, V_n) = M_n$ ,  $\mu^{(2)}(U_n, V_n) = M_n^{(2)}$ .

Because of this result, the biased sample variance  $T_n^2$  will appear in many of the estimation problems for special distributions that we consider below.

## Special Distributions

### The Normal Distribution

The *normal distribution* with mean  $\mu \in \mathbb{R}$  and variance  $\sigma^2 \in (0, \infty)$  is a continuous distribution on  $\mathbb{R}$  with probability density function  $g$  given by

$$g(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right], \quad x \in \mathbb{R} \quad (7.2.11)$$

This is one of the most important distributions in probability and statistics, primarily because of the central limit theorem. The normal distribution is studied in more detail in the chapter on Special Distributions.

Suppose now that  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  is a random sample of size  $n$  from the normal distribution with mean  $\mu$  and variance  $\sigma^2$ . From our general work [above](#), we know that if  $\mu$  is unknown then the sample mean  $M$  is the method of moments estimator of  $\mu$ , and if in addition,  $\sigma^2$  is unknown then the method of moments estimator of  $\sigma^2$  is  $T^2$ . On the other hand, in the unlikely event that  $\mu$  is known then  $W^2$  is the method of moments estimator of  $\sigma^2$ . Our goal is to see how the comparisons above simplify for the normal distribution.

Mean square errors of  $S_n^2$  and  $T_n^2$ .

1.  $\text{mse}(T^2) = \frac{2n-1}{n^2} \sigma^4$
2.  $\text{mse}(S^2) = \frac{2}{n-1} \sigma^4$
3.  $\text{mse}(T^2) < \text{mse}(S^2)$  for  $n \in \{2, 3, \dots\}$

Proof

Recall that for the normal distribution,  $\sigma_4 = 3\sigma^4$ . Substituting this into the general results gives parts (a) and (b). Part (c) follows from (a) and (b). Of course the asymptotic relative efficiency is still 1, from our previous theorem.

Thus,  $S^2$  and  $T^2$  are multiplies of one another;  $S^2$  is unbiased, but when the sampling distribution is normal,  $T^2$  has smaller mean square error. Surprisingly,  $T^2$  has smaller mean square error even than  $W^2$ .

Mean square errors of  $T^2$  and  $W^2$ .

1.  $\text{mse}(W^2) = \frac{2}{n} \sigma^4$
2.  $\text{mse}(T^2) < \text{mse}(W^2)$  for  $n \in \{2, 3, \dots\}$

Proof

Again, since the sampling distribution is normal,  $\sigma_4 = 3\sigma^4$ . Substituting this into the general formula for  $\text{var}(W_n^2)$  gives part (a).

Run the normal estimation experiment 1000 times for several values of the sample size  $n$  and the parameters  $\mu$  and  $\sigma$ . Compare the empirical bias and mean square error of  $S^2$  and of  $T^2$  to their theoretical values. Which estimator is better in terms of bias? Which estimator is better in terms of mean square error?

Next we consider estimators of the standard deviation  $\sigma$ . As noted in the general discussion above,  $T = \sqrt{T^2}$  is the method of moments estimator when  $\mu$  is unknown, while  $W = \sqrt{W^2}$  is the method of moments estimator in the unlikely event that  $\mu$  is known. Another natural estimator, of course, is  $S = \sqrt{S^2}$ , the usual sample standard deviation. The following sequence, defined in terms of the gamma function turns out to be important in the analysis of all three estimators.

Consider the sequence

$$a_n = \sqrt{\frac{2}{n} \frac{\Gamma[(n+1)/2]}{\Gamma(n/2)}}, \quad n \in \mathbb{N}_+ \quad (7.2.12)$$

Then  $0 < a_n < 1$  for  $n \in \mathbb{N}_+$  and  $a_n \uparrow 1$  as  $n \uparrow \infty$ .

First, assume that  $\mu$  is known so that  $W_n$  is the method of moments estimator of  $\sigma$ .

For  $n \in \mathbb{N}_+$ ,

1.  $\mathbb{E}(W) = a_n \sigma$
2.  $\text{bias}(W) = (a_n - 1) \sigma$
3.  $\text{var}(W) = (1 - a_n^2) \sigma^2$
4.  $\text{mse}(W) = 2(1 - a_n) \sigma^2$

Proof

Recall that  $U^2 = nW^2/\sigma^2$  has the chi-square distribution with  $n$  degrees of freedom, and hence  $U$  has the chi distribution with  $n$  degrees of freedom. Solving gives

$$W = \frac{\sigma}{\sqrt{n}} U \quad (7.2.13)$$

From the formulas for the mean and variance of the chi distribution we have

$$\begin{aligned}\mathbb{E}(W) &= \frac{\sigma}{\sqrt{n}} \mathbb{E}(U) = \frac{\sigma}{\sqrt{n}} \sqrt{2} \frac{\Gamma[(n+1)/2]}{\Gamma(n/2)} = \sigma a_n \\ \text{var}(W) &= \frac{\sigma^2}{n} \text{var}(U) = \frac{\sigma^2}{n} \{n - [\mathbb{E}(U)]^2\} = \sigma^2 (1 - a_n^2)\end{aligned}$$

Thus  $W$  is negatively biased as an estimator of  $\sigma$  but asymptotically unbiased and consistent. Of course we know that in general (regardless of the underlying distribution),  $W^2$  is an unbiased estimator of  $\sigma^2$  and so  $W$  is negatively biased as an estimator of  $\sigma$ . In the normal case, since  $a_n$  involves no unknown parameters, the statistic  $W/a_n$  is an unbiased estimator of  $\sigma$ . Next we consider the usual sample standard deviation  $S$ .

For  $n \in \{2, 3, \dots\}$ ,

1.  $\mathbb{E}(S) = a_{n-1} \sigma$
2.  $\text{bias}(S) = (a_{n-1} - 1) \sigma$
3.  $\text{var}(S) = (1 - a_{n-1}^2) \sigma^2$
4.  $\text{mse}(S) = 2(1 - a_{n-1}) \sigma^2$

Proof

Recall that  $V^2 = (n-1)S^2/\sigma^2$  has the chi-square distribution with  $n-1$  degrees of freedom, and hence  $V$  has the chi distribution with  $n-1$  degrees of freedom. The proof now proceeds just as in the previous theorem, but with  $n-1$  replacing  $n$ .

As with  $W$ , the statistic  $S$  is negatively biased as an estimator of  $\sigma$  but asymptotically unbiased, and also consistent. Since  $a_{n-1}$  involves no unknown parameters, the statistic  $S/a_{n-1}$  is an unbiased estimator of  $\sigma$ . Note also that, in terms of bias and mean square error,  $S$  with sample size  $n$  behaves like  $W$  with sample size  $n-1$ . Finally we consider  $T$ , the method of moments estimator of  $\sigma$  when  $\mu$  is unknown.

For  $n \in \{2, 3, \dots\}$ ,

1.  $\mathbb{E}(T) = \sqrt{\frac{n-1}{n}} a_{n-1} \sigma$
2.  $\text{bias}(T) = \left( \sqrt{\frac{n-1}{n}} a_{n-1} - 1 \right) \sigma$
3.  $\text{var}(T) = \frac{n-1}{n} (1 - a_{n-1}^2) \sigma^2$
4.  $\text{mse}(T) = \left( 2 - \frac{1}{n} - 2 \sqrt{\frac{n-1}{n}} a_{n-1} \right) \sigma^2$

Proof

The results follow easily from the previous theorem since  $T_n = \sqrt{\frac{n-1}{n}} S_n$ .

## The Bernoulli Distribution

Recall that an *indicator variable* is a random variable  $X$  that takes only the values 0 and 1. The distribution of  $X$  is known as the *Bernoulli distribution*, named for Jacob Bernoulli, and has probability density function  $g$  given by

$$g(x) = p^x (1-p)^{1-x}, \quad x \in \{0, 1\} \quad (7.2.14)$$

where  $p \in (0, 1)$  is the *success parameter*. The mean of the distribution is  $p$  and the variance is  $p(1-p)$ .

Suppose now that  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  is a random sample of size  $n$  from the Bernoulli distribution with unknown success parameter  $p$ . Since the mean of the distribution is  $p$ , it follows from our general work [above](#) that the method of moments estimator of  $p$  is  $M$ , the sample mean. In this case, the sample  $\mathbf{X}$  is a sequence of Bernoulli trials, and  $M$  has a scaled version of the binomial distribution with parameters  $n$  and  $p$ :

$$\mathbb{P}\left(M = \frac{k}{n}\right) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k \in \{0, 1, \dots, n\} \quad (7.2.15)$$

Note that since  $X^k = X$  for every  $k \in \mathbb{N}_+$ , it follows that  $\mu^{(k)} = p$  and  $M^{(k)} = M$  for every  $k \in \mathbb{N}_+$ . So any of the method of moments equations would lead to the sample mean  $M$  as the estimator of  $p$ . Although very simple, this is an important application,

since Bernoulli trials are found embedded in all sorts of estimation problems, such as empirical probability density functions and empirical distribution functions.

### The Geometric Distribution

The *geometric distribution* on  $\mathbb{N}_+$  with success parameter  $p \in (0, 1)$  has probability density function  $g$  given by

$$g(x) = p(1-p)^{x-1}, \quad x \in \mathbb{N}_+ \quad (7.2.16)$$

The geometric distribution on  $\mathbb{N}_+$  governs the number of trials needed to get the first success in a sequence of Bernoulli trials with success parameter  $p$ . The mean of the distribution is  $\mu = 1/p$ .

Suppose that  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  is a random sample of size  $n$  from the geometric distribution on  $\mathbb{N}_+$  with unknown success parameter  $p$ . The method of moments estimator of  $p$  is

$$U = \frac{1}{M} \quad (7.2.17)$$

Proof

The method of moments equation for  $U$  is  $1/U = M$ .

The geometric distribution on  $\mathbb{N}$  with success parameter  $p \in (0, 1)$  has probability density function

$$g(x) = p(1-p)^x, \quad x \in \mathbb{N} \quad (7.2.18)$$

This version of the geometric distribution governs the number of failures before the first success in a sequence of Bernoulli trials. The mean of the distribution is  $\mu = (1-p)/p$ .

Suppose that  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  is a random sample of size  $n$  from the geometric distribution on  $\mathbb{N}$  with unknown parameter  $p$ . The method of moments estimator of  $p$  is

$$U = \frac{1}{M+1} \quad (7.2.19)$$

Proof

The method of moments equation for  $U$  is  $(1-U)/U = M$ .

### The Negative Binomial Distribution

More generally, the *negative binomial distribution* on  $\mathbb{N}$  with shape parameter  $k \in (0, \infty)$  and success parameter  $p \in (0, 1)$  has probability density function

$$g(x) = \binom{x+k-1}{k-1} p^k (1-p)^x, \quad x \in \mathbb{N} \quad (7.2.20)$$

If  $k$  is a positive integer, then this distribution governs the number of failures before the  $k$ th success in a sequence of Bernoulli trials with success parameter  $p$ . However, the distribution makes sense for general  $k \in (0, \infty)$ . The negative binomial distribution is studied in more detail in the chapter on Bernoulli Trials. The mean of the distribution is  $k(1-p)/p$  and the variance is  $k(1-p)/p^2$ . Suppose now that  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  is a random sample of size  $n$  from the negative binomial distribution on  $\mathbb{N}$  with shape parameter  $k$  and success parameter  $p$

If  $k$  and  $p$  are unknown, then the corresponding method of moments estimators  $U$  and  $V$  are

$$U = \frac{M^2}{T^2 - M}, \quad V = \frac{M}{T^2} \quad (7.2.21)$$

Proof

Matching the distribution mean and variance to the sample mean and variance gives the equations

$$U \frac{1-V}{V} = M, \quad U \frac{1-V}{V^2} = T^2 \quad (7.2.22)$$

As usual, the results are nicer when one of the parameters is known.

Suppose that  $k$  is known but  $p$  is unknown. The method of moments estimator  $V_k$  of  $p$  is

$$V_k = \frac{k}{M + k} \quad (7.2.23)$$

Proof

Matching the distribution mean to the sample mean gives the equation

$$k \frac{1 - V_k}{V_k} = M \quad (7.2.24)$$

Suppose that  $k$  is unknown but  $p$  is known. The method of moments estimator of  $k$  is

$$U_p = \frac{p}{1 - p} M \quad (7.2.25)$$

1.  $\mathbb{E}(U_p) = k$  so  $U_p$  is unbiased.
2.  $\text{var}(U_p) = \frac{k}{n(1-p)}$  so  $U_p$  is consistent.

Proof

Matching the distribution mean to the sample mean gives the equation  $U_p \frac{1-p}{p} = M$ .

1.  $E(U_p) = \frac{p}{1-p} E(M)$  and  $E(M) = \frac{1-p}{p} k$
2.  $\text{var}(U_p) = \left(\frac{p}{1-p}\right)^2 \text{var}(M)$  and  $\text{var}(M) = \frac{1}{n} \text{var}(X) = \frac{1-p}{np^2}$

## The Poisson Distribution

The *Poisson distribution* with parameter  $r \in (0, \infty)$  is a discrete distribution on  $\mathbb{N}$  with probability density function  $g$  given by

$$g(x) = e^{-r} \frac{r^x}{x!}, \quad x \in \mathbb{N} \quad (7.2.26)$$

The mean and variance are both  $r$ . The distribution is named for Simeon Poisson and is widely used to model the number of “random points” in a region of time or space. The parameter  $r$  is proportional to the size of the region, with the proportionality constant playing the role of the average *rate* at which the points are distributed in time or space. The Poisson distribution is studied in more detail in the chapter on the Poisson Process.

Suppose now that  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  is a random sample of size  $n$  from the Poisson distribution with parameter  $r$ . Since  $r$  is the mean, it follows from our general work [above](#) that the method of moments estimator of  $r$  is the sample mean  $M$ .

## The Gamma Distribution

The *gamma distribution* with shape parameter  $k \in (0, \infty)$  and scale parameter  $b \in (0, \infty)$  is a continuous distribution on  $(0, \infty)$  with probability density function  $g$  given by

$$g(x) = \frac{1}{\Gamma(k)b^k} x^{k-1} e^{-x/b}, \quad x \in (0, \infty) \quad (7.2.27)$$

The gamma probability density function has a variety of shapes, and so this distribution is used to model various types of positive random variables. The gamma distribution is studied in more detail in the chapter on Special Distributions. The mean is  $\mu = kb$  and the variance is  $\sigma^2 = kb^2$ .

Suppose now that  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  is a random sample from the gamma distribution with shape parameter  $k$  and scale parameter  $b$ .

Suppose that  $k$  and  $b$  are both unknown, and let  $U$  and  $V$  be the corresponding method of moments estimators. Then

$$U = \frac{M^2}{T^2}, \quad V = \frac{T^2}{M} \quad (7.2.28)$$

### Proof

Matching the distribution mean and variance with the sample mean and variance leads to the equations  $UV = M$ ,  $UV^2 = T^2$ . Solving gives the results.

The method of moments estimators of  $k$  and  $b$  given in the previous exercise are complicated, nonlinear functions of the sample mean  $M$  and the sample variance  $T^2$ . Thus, computing the bias and mean square errors of these estimators are difficult problems that we will not attempt. However, we can judge the quality of the estimators empirically, through simulations.

When one of the parameters is known, the method of moments estimator of the other parameter is much simpler.

Suppose that  $k$  is unknown, but  $b$  is known. The method of moments estimator of  $k$  is

$$U_b = \frac{M}{b} \quad (7.2.29)$$

1.  $\mathbb{E}(U_b) = k$  so  $U_b$  is unbiased.
2.  $\text{var}(U_b) = k/n$  so  $U_b$  is consistent.

### Proof

If  $b$  is known, then the method of moments equation for  $U_b$  is  $bU_b = M$ . Solving gives (a). Next,  $\mathbb{E}(U_b) = \mathbb{E}(M)/b = kb/b = k$ , so  $U_b$  is unbiased. Finally  $\text{var}(U_b) = \text{var}(M)/b^2 = kb^2/(nb^2) = k/n$ .

Suppose that  $b$  is unknown, but  $k$  is known. The method of moments estimator of  $b$  is

$$V_k = \frac{M}{k} \quad (7.2.30)$$

1.  $\mathbb{E}(V_k) = b$  so  $V_k$  is unbiased.
2.  $\text{var}(V_k) = b^2/kn$  so that  $V_k$  is consistent.

### Proof

If  $k$  is known, then the method of moments equation for  $V_k$  is  $kV_k = M$ . Solving gives (a). Next,  $\mathbb{E}(V_k) = \mathbb{E}(M)/k = kb/k = b$ , so  $V_k$  is unbiased. Finally  $\text{var}(V_k) = \text{var}(M)/k^2 = kb^2/(nk^2) = b^2/kn$ .

Run the gamma estimation experiment 1000 times for several different values of the sample size  $n$  and the parameters  $k$  and  $b$ . Note the empirical bias and mean square error of the estimators  $U$ ,  $V$ ,  $U_b$ , and  $V_k$ . One would think that the estimators when one of the parameters is known should work better than the corresponding estimators when both parameters are unknown; but investigate this question empirically.

## The Beta Distribution

The *beta distribution* with left parameter  $a \in (0, \infty)$  and right parameter  $b \in (0, \infty)$  is a continuous distribution on  $(0, 1)$  with probability density function  $g$  given by

$$g(x) = \frac{1}{B(a, b)} x^{a-1} (1-x)^{b-1}, \quad 0 < x < 1 \quad (7.2.31)$$

The beta probability density function has a variety of shapes, and so this distribution is widely used to model various types of random variables that take values in bounded intervals. The beta distribution is studied in more detail in the chapter on Special Distributions. The first two moments are  $\mu = \frac{a}{a+b}$  and  $\mu^{(2)} = \frac{a(a+1)}{(a+b)(a+b+1)}$ .

Suppose now that  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  is a random sample of size  $n$  from the beta distribution with left parameter  $a$  and right parameter  $b$ .

Suppose that  $a$  and  $b$  are both unknown, and let  $U$  and  $V$  be the corresponding method of moments estimators. Then

$$U = \frac{M(M - M^{(2)})}{M^{(2)} - M^2}, \quad V = \frac{(1 - M)(M - M^{(2)})}{M^{(2)} - M^2} \quad (7.2.32)$$

Proof

The method of moments equations for  $U$  and  $V$  are

$$\frac{U}{U+V} = M, \quad \frac{U(U+1)}{(U+V)(U+V+1)} = M^{(2)} \quad (7.2.33)$$

Solving gives the result.

The method of moments estimators of  $a$  and  $b$  given in the previous exercise are complicated nonlinear functions of the sample moments  $M$  and  $M^{(2)}$ . Thus, we will not attempt to determine the bias and mean square errors analytically, but you will have an opportunity to explore them empirically through a simulation.

Suppose that  $a$  is unknown, but  $b$  is known. Let  $U_b$  be the method of moments estimator of  $a$ . Then

$$U_b = b \frac{M}{1-M} \quad (7.2.34)$$

Proof

If  $b$  is known then the method of moments equation for  $U_b$  as an estimator of  $a$  is  $U_b/(U_b+b) = M$ . Solving for  $U_b$  gives the result.

Suppose that  $b$  is unknown, but  $a$  is known. Let  $V_a$  be the method of moments estimator of  $b$ . Then

$$V_a = a \frac{1-M}{M} \quad (7.2.35)$$

Proof

If  $a$  is known then the method of moments equation for  $V_a$  as an estimator of  $b$  is  $a/(a+V_a) = M$ . Solving for  $V_a$  gives the result.

Run the beta estimation experiment 1000 times for several different values of the sample size  $n$  and the parameters  $a$  and  $b$ . Note the empirical bias and mean square error of the estimators  $U$ ,  $V$ ,  $U_b$ , and  $V_a$ . One would think that the estimators when one of the parameters is known should work better than the corresponding estimators when both parameters are unknown; but investigate this question empirically.

The following problem gives a distribution with just one parameter but the second moment equation from the method of moments is needed to derive an estimator.

Suppose that  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  is a random sample from the *symmetric* beta distribution, in which the left and right parameters are equal to an unknown value  $c \in (0, \infty)$ . The method of moments estimator of  $c$  is

$$U = \frac{2M^{(2)}}{1-4M^{(2)}} \quad (7.2.36)$$

Proof

Note that the mean  $\mu$  of the symmetric distribution is  $\frac{1}{2}$ , independently of  $c$ , and so the first equation in the method of moments is useless. However, matching the second distribution moment to the second sample moment leads to the equation

$$\frac{U+1}{2(2U+1)} = M^{(2)} \quad (7.2.37)$$

Solving gives the result.

## The Pareto Distribution

The *Pareto distribution* with shape parameter  $a \in (0, \infty)$  and scale parameter  $b \in (0, \infty)$  is a continuous distribution on  $(b, \infty)$  with probability density function  $g$  given by

$$g(x) = \frac{ab^a}{x^{a+1}}, \quad b \leq x < \infty \quad (7.2.38)$$

The Pareto distribution is named for Vilfredo Pareto and is a highly skewed and *heavy-tailed* distribution. It is often used to model income and certain other types of positive random variables. The Pareto distribution is studied in more detail in the chapter on Special Distributions. If  $a > 2$ , the first two moments of the Pareto distribution are  $\mu = \frac{ab}{a-1}$  and  $\mu^{(2)} = \frac{ab^2}{a-2}$ .

Suppose now that  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  is a random sample of size  $n$  from the Pareto distribution with shape parameter  $a > 2$  and scale parameter  $b > 0$ .

Suppose that  $a$  and  $b$  are both unknown, and let  $U$  and  $V$  be the corresponding method of moments estimators. Then

$$U = 1 + \sqrt{\frac{M^{(2)}}{M^{(2)} - M^2}} \quad (7.2.39)$$

$$V = \frac{M^{(2)}}{M} \left( 1 - \sqrt{\frac{M^{(2)} - M^2}{M^{(2)}}} \right) \quad (7.2.40)$$

Proof

The method of moments equations for  $U$  and  $V$  are

$$\frac{UV}{U-1} = M \quad (7.2.41)$$

$$\frac{UV^2}{U-2} = M^{(2)} \quad (7.2.42)$$

Solving for  $U$  and  $V$  gives the results.

As with our previous examples, the method of moments estimators are complicated nonlinear functions of  $M$  and  $M^{(2)}$ , so computing the bias and mean square error of the estimator is difficult. Instead, we can investigate the bias and mean square error empirically, through a simulation.

Run the Pareto estimation experiment 1000 times for several different values of the sample size  $n$  and the parameters  $a$  and  $b$ . Note the empirical bias and mean square error of the estimators  $U$  and  $V$ .

When one of the parameters is known, the method of moments estimator for the other parameter is simpler.

Suppose that  $a$  is unknown, but  $b$  is known. Let  $U_b$  be the method of moments estimator of  $a$ . Then

$$U_b = \frac{M}{M-b} \quad (7.2.43)$$

Proof

If  $b$  is known then the method of moment equation for  $U_b$  as an estimator of  $a$  is  $bU_b/(U_b - 1) = M$ . Solving for  $U_b$  gives the result.

Suppose that  $b$  is unknown, but  $a$  is known. Let  $V_a$  be the method of moments estimator of  $b$ . Then

$$V_a = \frac{a-1}{a} M \quad (7.2.44)$$

1.  $\mathbb{E}(V_a) = b$  so  $V_a$  is unbiased.
2.  $\text{var}(V_a) = \frac{b^2}{na(a-2)}$  so  $V_a$  is consistent.

Proof

If  $a$  is known then the method of moments equation for  $V_a$  as an estimator of  $b$  is  $aV_a/(a-1) = M$ . Solving for  $V_a$  gives

(a). Next,  $\mathbb{E}(V_a) = \frac{a-1}{a} \mathbb{E}(M) = \frac{a-1}{a} \frac{ab}{a-1} = b$  so  $V_a$  is unbiased. Finally,

$$\text{var}(V_a) = \left(\frac{a-1}{a}\right)^2 \text{var}(M) = \frac{(a-1)^2}{a^2} \frac{ab^2}{n(a-1)^2(a-2)} = \frac{b^2}{na(a-2)}.$$

### The Uniform Distribution

The (continuous) *uniform distribution* with location parameter  $a \in \mathbb{R}$  and scale parameter  $h \in (0, \infty)$  has probability density function  $g$  given by

$$g(x) = \frac{1}{h}, \quad x \in [a, a+h] \quad (7.2.45)$$

The distribution models a point chosen “at random” from the interval  $[a, a+h]$ . The mean of the distribution is  $\mu = a + \frac{1}{2}h$  and the variance is  $\sigma^2 = \frac{1}{12}h^2$ . The uniform distribution is studied in more detail in the chapter on Special Distributions. Suppose now that  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  is a random sample of size  $n$  from the uniform distribution.

Suppose that  $a$  and  $h$  are both unknown, and let  $U$  and  $V$  denote the corresponding method of moments estimators. Then

$$U = 2M - \sqrt{3}T, \quad V = 2\sqrt{3}T \quad (7.2.46)$$

Proof

Matching the distribution mean and variance to the sample mean and variance leads to the equations  $U + \frac{1}{2}V = M$  and  $\frac{1}{12}V^2 = T^2$ . Solving gives the result.

As usual, we get nicer results when one of the parameters is known.

Suppose that  $a$  is known and  $h$  is unknown, and let  $V_a$  denote the method of moments estimator of  $h$ . Then

$$V_a = 2(M - a) \quad (7.2.47)$$

1.  $\mathbb{E}(V_a) = h$  so  $V$  is unbiased.
2.  $\text{var}(V_a) = \frac{h^2}{3n}$  so  $V_a$  is consistent.

Proof

Matching the distribution mean to the sample mean leads to the equation  $a + \frac{1}{2}V_a = M$ . Solving gives the result.

1.  $\mathbb{E}(V_a) = 2[\mathbb{E}(M) - a] = 2(a + h/2 - a) = h$
2.  $\text{var}(V_a) = 4\text{var}(M) = \frac{h^2}{3n}$

Suppose that  $h$  is known and  $a$  is unknown, and let  $U_h$  denote the method of moments estimator of  $a$ . Then

$$U_h = M - \frac{1}{2}h \quad (7.2.48)$$

1.  $\mathbb{E}(U_h) = a$  so  $U_h$  is unbiased.
2.  $\text{var}(U_h) = \frac{h^2}{12n}$  so  $U_h$  is consistent.

Proof

Matching the distribution mean to the sample mean leads to the equation  $U_h + \frac{1}{2}h = M$ . Solving gives the result.

1.  $\mathbb{E}(U_h) = \mathbb{E}(M) - \frac{1}{2}h = a + \frac{1}{2}h - \frac{1}{2}h = a$
2.  $\text{var}(U_h) = \text{var}(M) = \frac{h^2}{12n}$

### The Hypergeometric Model

Our basic assumption in the method of moments is that the sequence of observed random variables  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  is a random sample from a distribution. However, the method makes sense, at least in some cases, when the variables are identically distributed but dependent. In the *hypergeometric model*, we have a population of  $N$  objects with  $r$  of the objects *type 1* and the remaining  $N - r$  objects *type 0*. The parameter  $N$ , the *population size*, is a positive integer. The parameter  $r$ , the *type 1 size*, is a

nonnegative integer with  $r \leq N$ . These are the basic parameters, and typically one or both is unknown. Here are some typical examples:

1. The objects are devices, classified as *good* or *defective*.
2. The objects are persons, classified as *female* or *male*.
3. The objects are voters, classified as *for* or *against* a particular candidate.
4. The objects are wildlife or a particular type, either *tagged* or *untagged*.

We sample  $n$  objects from the population at random, without replacement. Let  $X_i$  be the type of the  $i$ th object selected, so that our sequence of observed variables is  $\mathbf{X} = (X_1, X_2, \dots, X_n)$ . The variables are identically distributed indicator variables, with  $P(X_i = 1) = r/N$  for each  $i \in \{1, 2, \dots, n\}$ , but are dependent since the sampling is without replacement. The number of type 1 objects in the sample is  $Y = \sum_{i=1}^n X_i$ . This statistic has the *hypergeometric distribution* with parameter  $N$ ,  $r$ , and  $n$ , and has probability density function given by

$$P(Y = y) = \frac{\binom{r}{y} \binom{N-r}{n-y}}{\binom{N}{n}} = \binom{n}{y} \frac{r^{(y)} (N-r)^{(n-y)}}{N^{(n)}}, \quad y \in \{\max\{0, N-n+r\}, \dots, \min\{n, r\}\} \quad (7.2.49)$$

The hypergeometric model is studied in more detail in the chapter on Finite Sampling Models.

As above, let  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  be the observed variables in the hypergeometric model with parameters  $N$  and  $r$ . Then

1. The method of moments estimator of  $p = r/N$  is  $M = Y/n$ , the sample mean.
2. The method of moments estimator of  $r$  with  $N$  known is  $U = NM = NY/n$ .
3. The method of moments estimator of  $N$  with  $r$  known is  $V = r/M = rn/Y$  if  $Y > 0$ .

Proof

These results all follow simply from the fact that  $\mathbb{E}(X) = \mathbb{P}(X = 1) = r/N$ .

In the voter example (3) above, typically  $N$  and  $r$  are both unknown, but we would only be interested in estimating the ratio  $p = r/N$ . In the reliability example (1), we might typically know  $N$  and would be interested in estimating  $r$ . In the wildlife example (4), we would typically know  $r$  and would be interested in estimating  $N$ . This example is known as the *capture-recapture* model.

Clearly there is a close relationship between the hypergeometric model and the [Bernoulli trials model](#) above. In fact, if the sampling is *with* replacement, the Bernoulli trials model would apply rather than the hypergeometric model. In addition, if the population size  $N$  is large compared to the sample size  $n$ , the hypergeometric model is well approximated by the Bernoulli trials model.

This page titled [7.2: The Method of Moments](#) is shared under a [CC BY 2.0](#) license and was authored, remixed, and/or curated by [Kyle Siegrist \(Random Services\)](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.