2. The Method of Moments

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 parameters, or equivalently, a parameter vector $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_k)$ taking values in a parameter space $\Theta \subseteq \mathbb{R}^k$. As usual, we repeat the experiment n times to generate a random sample of size n from the distribution of X.

$$\boldsymbol{X} = (X_1, X_2, ..., X_n)$$

Thus, $\boldsymbol{X}$ 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

$$\mu_i(\theta) = \mathbb{E}_{\boldsymbol{\theta}}\left(X^i\right), \quad i \in \{1, 2, ..., k\}$$

so that $\mu_i(\boldsymbol{\theta})$ is the i^{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 μ. Next, let

$$M_i(\boldsymbol{X}) = \frac{1}{n} \sum_{j=1}^{n} X_j{}^i, \quad i \in \{1, 2, ..., k\}$$

so that $M_i(\boldsymbol{X})$ is the i^{th} sample moment about 0. Equivalently, $M_i(\boldsymbol{X})$ is the sample mean for the random sample $\left(X_1{}^i, X_2{}^i, ..., X_n{}^i\right)$ from the distribution of X^i. Note that we are emphasizing the dependence of the sample moments on the sample $\boldsymbol{X}$. Note also that $M_1(\boldsymbol{X})$ is just the ordinary sample mean, which we usually just denote by $M(\boldsymbol{X})$.

From our previous work, we know that $M_i(\boldsymbol{X})$ is an unbiased and consistent estimator of $\mu_i(\boldsymbol{\theta})$ for each i. Thus, to construct estimators $(W_1, W_2, ..., W_k)$ for our parameters $(\theta_1, \theta_2, ..., \theta_k)$ respectively, we attempt to solve the set of simultaneous equations

$$\mu_i(W_1, W_2, ..., W_k) = M_i(X_1, X_2, ..., X_n), \quad i \in \{1, 2, ..., k\}$$

for $(W_1, W_2, ..., W_k)$ in terms of $(X_1, X_2, ..., X_n)$. Note that we have k equations in k unknowns, so there is hope that the equations can be solved.

Estimates for the Mean and Variance

1. Suppose that $\boldsymbol{X} = (X_1, X_2, ..., X_n)$ is a random sample of size n from a distribution with unknown mean μ and variance σ^2. Show that the method of moments estimators for μ and σ^2 are, respectively

$$M = \frac{1}{n} \sum_{i=1}^n X_i, \quad T^2 = \frac{1}{n} \sum_{i=1}^n (X_i - M)^2$$

Of course, M is just the ordinary sample mean, but $T^2 = \frac{n-1}{n} S^2$ where S^2 is the usual sample variance. In the remainder of this subsection, we will compare the estimators S^2 and T^2. Recall that S^2 is unbiased and consistent, with $\mathrm{var}(S^2) = \frac{1}{n}\left(d_4 - \frac{n-3}{n-1} \sigma^4\right)$

2. Show that $\mathrm{bias}(T^2) = -\frac{\sigma^2}{n}$. Thus, T^2 is negatively biased, and so on average underestimates σ^2.

3. Show that T^2 is asymptotically unbiased.

4. Let $d_4 = \mathbb{E}\left((X - \mu)^4\right)$ denote the 4th central moment. Show that

$$\mathrm{MSE}(T^2) = \frac{(n-1)^2}{n^3}\left(d_4 - \frac{n-3}{n-1}\sigma^4\right) + \frac{\sigma^4}{n^2}$$

5. Show that the asymptotic relative efficiency of T^2 to S^2 is 1.

6. Suppose that the sampling distribution is normal, so that $d_4 = 3\sigma^4$. Show that in this case

a. $\mathrm{MSE}(T^2) = \frac{2n-1}{n^2}\sigma^4$
b. $\mathrm{MSE}(S^2) = \frac{2}{n-1}\sigma^4$
c. $\mathrm{MSE}(T^2) < \mathrm{MSE}(S^2)$ for $n \in \{2, 3, ...\}$

Thus, S^2 and T^2 are multiplies of one another; S^2 is unbiased, but at least when the sampling distribution is normal, T^2 has smaller mean square error. Next, recall that under the (artificial) assumption that μ is known, a natural estimator of σ^2 is

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

Moreover, W^2 is unbiased and consistent, with $\mathrm{var}(W^2) = \frac{d_4 - \sigma^4}{n}$. Amazingly, when the sampling distribution is normal, T^2 has smaller mean square error even than W^2.

7. Suppose again that the sampling distribution is normal. Show that

a. $\mathrm{MSE}(W^2) = \frac{2\sigma^4}{n}$
b. $\mathrm{MSE}(T^2) < \mathrm{MSE}(W^2)$ for $n \in \{2, 3, ...\}$

8. Run the normal estimation experiment 1000 times, updating every 10 runs, for several values of the sample size n and the parameters μ and σ. 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?

There are several important one-parameter families of distributions for which the parameter is the mean, including the Bernoulli distribution with parameter p and the Poisson distribution with parameter a. For these families, the method of moments estimator of the parameter is the sample mean M. Similarly, the parameters of the normal distribution are the mean μ and the variance σ^2, so the method of moments estimators are M and T^2, respectively.

Additional Exercises

9. Suppose that $\boldsymbol{X} = (X_1, X_2, ..., X_n)$ is a random sample from the gamma distribution with shape parameter k and scale parameter b. Show that the method of moments estimators of k and b are respectively

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

10. Run the gamma estimation experiment 1000 times, updating every 10 runs for several different values of the sample size n, the shape parameter k and scale parameter b. Note the empirical bias and mean square error of the estimators U and V.

11. Suppose that $\boldsymbol{X} = (X_1, X_2, ..., X_n)$ is a random sample of size n from the beta distribution with left parameter a and right parameter 1. Show that the method of moments estimator of a is

$$U = \frac{M}{1 - M}$$

.

12. Run the beta estimation experiment 1000 times, updating every 10 runs, for several different values of the sample size n and the parameter a. Note the empirical bias and mean square error of the estimator U.

13. Suppose that $\boldsymbol{X} = (X_1, X_2, ..., X_n)$ is a random sample of size n from the Pareto distribution with shape parameter $a > 1$. Show that the method of moments estimator of a is

$$U = \frac{M}{M - 1}$$

.

14. Run the Pareto estimation experiment 1000 times, updating every 10 runs, for several different values of the sample size n and the parameter a. Note the empirical bias and mean square error of the estimator U.