5

Method of Moments

As you have no doubt realized, if θ is a parameter of interest, then it is not easy to "guess" unbiased estimators, let alone determine the minimum variance unbiased estimator of θ. We will now learn the oldest method for deriving point estimators, namely the *method of moments*, introduced in 1894 by Karl Pearson.

Suppose that $Y_1, \ldots, Y_n$ is a random sample from a population having common density $f(y|\theta)$ depending on a parameter θ. The *kth sample moment* is given by

$$\hat{\mu}_k := \frac{1}{n} \sum_{i=1}^{n} Y_i^k$$

and the *kth population moment* is given by

$$\mu_k := \mathbb{E}(Y_1^k) = \int_0^\infty y^k f(y|\theta)\, \mathrm{d}y.$$

Notice that $\hat{\mu}_k$ is a random variable, whereas μ_k is not random but rather depends on the parameter θ. In fact, $\hat{\mu}_k$ is an unbiased estimator of μ_k since

$$\mathbb{E}(\hat{\mu}_k) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i^k\right) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}(Y_i^k) = \frac{1}{n}\sum_{i=1}^{n} \mu_k = \frac{n\mu_k}{n} = \mu_k.$$

The method of moments estimator of θ is the value of θ solving

$$\mu_1 = \hat{\mu}_1.$$

Call the solution $\hat{\theta}_{\mathrm{MOM}}$, the *method of moments estimator of θ*.

Remark. It might be the case that $\mu_1 = \hat{\mu}_1$ has no solutions, or more than one solution. In the first situation, there is no method of moments estimator. In the second situation, we choose the most reasonable solution given the problem at hand. Examples 5.3 and 5.4 below illustrate these situations.

Example 5.1. Suppose that $Y_1, \dots, Y_n$ is a random sample from a $\mathcal{N}(\theta, \sigma^2)$ population where σ^2 is known and $\theta \in \mathbb{R}$ is a parameter. Observe that

$$\mu_1 := \mathbb{E}(Y_1) = \theta$$

and

$$\hat{\mu}_1 := \frac{1}{n}\sum_{i=1}^{n} Y_i = \overline{Y}.$$

Notice that $\hat{\mu}_1$ is a random variable, whereas μ_1 is not random but rather depends on the parameter θ. Setting $\mu_1 = \hat{\mu}_1$ implies that $\theta = \overline{Y}$ and so

$$\hat{\theta}_{\mathrm{MOM}} = \overline{Y}.$$

Example 5.2. Suppose that $Y_1, \dots, Y_n$ is a random sample from an $\mathrm{Exp}(1/\theta)$ population so that their common density is

$$f(y|\theta) = \theta e^{-\theta y}, \quad y > 0,$$

where $\theta > 0$ is a parameter. Observe that

$$\mu_1 := \mathbb{E}(Y_1) = \theta \int_0^\infty y e^{-\theta y}\, \mathrm{d}y = \frac{1}{\theta}$$

and

$$\hat{\mu}_1 := \frac{1}{n}\sum_{i=1}^{n} Y_i = \overline{Y}.$$

Setting $\mu_1 = \hat{\mu}_1$ implies that $1/\theta = \overline{Y}$ and so solving for θ yields $\theta = 1/\overline{Y}$. Therefore,

$$\hat{\theta}_{\mathrm{MOM}} = \frac{1}{\overline{Y}} = n\left(\sum_{i=1}^{n} Y_i\right)^{-1}.$$

As the following example shows, it might not be possible to solve $\mu_1 = \hat{\mu}_1$ for θ.

Example 5.3. Suppose that $Y_1, \dots, Y_n$ is a random sample from a $\mathrm{Uniform}[-\theta, \theta]$ population so that their common density is

$$f(y|\theta) = \frac{1}{2\theta}, \quad -\theta \le y \le \theta,$$

where $\theta > 0$ is a parameter. Observe that

$$\mu_1 := \mathbb{E}(Y_1) = \frac{1}{2\theta}\int_{-\theta}^{\theta} y\, \mathrm{d}y = 0$$

and

$$\hat{\mu}_1 := \frac{1}{n}\sum_{i=1}^{n} Y_i = \overline{Y}.$$

The equation $\mu_1 = \hat{\mu}_1$, namely $0 = \overline{Y}$, has no solutions for θ. Thus, a method of moments estimator of θ does not exist in this example.

We now provide an example where $\mu_1 = \hat{\mu}_1$ has more than one solution for θ.

Example 5.4. Suppose that $Y_1, \ldots, Y_n$ is a random sample from an $\text{Exp}(\theta^2)$ population so that their common density is

$$f(y|\theta) = \frac{1}{\theta^2} e^{-y/\theta^2}, \quad y > 0,$$

where $\theta \neq 0$ is a parameter. (We will be a bit more specific about θ momentarily.) Observe that

$$\mu_1 := \mathbb{E}(Y_1) = \frac{1}{\theta^2} \int_0^\infty y e^{-y/\theta^2} \, \mathrm{d}y = \theta^2$$

and

$$\hat{\mu}_1 := \frac{1}{n} \sum_{i=1}^n Y_i = \overline{Y}.$$

Setting $\mu_1 = \hat{\mu}_1$ implies that $\theta^2 = \overline{Y}$. Solving this equation for θ implies that there are two solutions, namely

$$\theta \in \left\{ -\sqrt{\overline{Y}}, \sqrt{\overline{Y}} \right\}.$$

(The fact that the support of $f(y|\theta)$ is $y > 0$ implies that $\overline{Y} > 0$ so that there is no issue taking the square root of $\overline{Y}$.) If nothing is known about θ (other than the fact that $\theta \neq 0$), then there is no unique solution to $\theta^2 = \overline{Y}$ implying that there are two method of moments estimators of θ, namely

$$\hat{\theta}_{\text{MOM},1} = -\sqrt{\overline{Y}} \quad \text{and} \quad \hat{\theta}_{\text{MOM},2} = \sqrt{\overline{Y}}.$$

However, if it is also known that $\theta > 0$, then there *is* a unique solution to $\theta^2 = \overline{Y}$ subject to the constraint $\theta > 0$, namely $\theta = \sqrt{\overline{Y}}$, in which case we have

$$\hat{\theta}_{\text{MOM}} = \sqrt{\overline{Y}}.$$

Multiple parameters

One of the advantages of the method of moments is that it often works in situations where there are more than one parameter. In fact, if $Y_1, \ldots, Y_n$ is a random sample from a population having common density depending on k parameters, say $f(y|\theta_1, \ldots, \theta_k)$, then the method of moments estimators of $\theta_1, \ldots, \theta_k$ are the values of $\theta_1, \ldots, \theta_k$ solving the system of k equations

$$\begin{aligned} \mu_1 &= \hat{\mu}_1 \\ &\vdots \\ \mu_k &= \hat{\mu}_k. \end{aligned}$$

Call the solution $\hat{\theta}_{1,\text{MOM}}, \ldots, \hat{\theta}_{k,\text{MOM}}$, the *method of moments estimator of* $\theta_1, \ldots, \theta_k$.

Example 5.5. Suppose that $Y_1, \dots, Y_n$ is a random sample from a $\mathcal{N}(\theta, \sigma^2)$ population where both θ and σ^2 are parameters. Determine the method of moment estimators $\hat{\theta}_{\mathrm{MOM}}$ and $\hat{\sigma}^2_{\mathrm{MOM}}$.

Solution. If $Y \sim \mathcal{N}(\theta, \sigma^2)$, then $\mathbb{E}(Y) = \theta$ and $\mathbb{E}(Y^2) = \mathrm{Var}(Y) + [\mathbb{E}(Y)]^2 = \sigma^2 + \theta^2$, and so it follows that $\mu_1 := \mathbb{E}(Y) = \theta$ and $\mu_2 := \mathbb{E}(Y^2) = \sigma^2 + \theta^2$. Moreover,

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

Setting $\mu_1 = \hat{\mu}_1$ and $\mu_2 = \hat{\mu}_2$ implies that

$$\theta = \frac{1}{n}\sum_{i=1}^{n} Y_i \quad \text{and} \quad \sigma^2 + \theta^2 = \frac{1}{n}\sum_{i=1}^{n} Y_i^2$$

and so solving this system yields the method of moments estimators

$$\hat{\theta}_{\mathrm{MOM}} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

and

$$\hat{\sigma}^2_{\mathrm{MOM}} = \frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \frac{1}{n^2}\left(\sum_{i=1}^{n} Y_i\right)^2.$$

We can simplify the expression for $\hat{\sigma}^2_{\mathrm{MOM}}$ with a little bit of algebra, namely

$$\begin{aligned}\sum_{i=1}^{n}(Y_i - \overline{Y})^2 = \sum_{i=1}^{n}(Y_i^2 - 2\overline{Y}Y_i + \overline{Y}^2) &= \sum_{i=1}^{n} Y_i^2 - 2\overline{Y}\sum_{i=1}^{n} Y_i + \overline{Y}^2\sum_{i=1}^{n} 1 \\ &= \sum_{i=1}^{n} Y_i^2 - n\overline{Y}^2\end{aligned}$$

so that

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \frac{1}{n^2}\left(\sum_{i=1}^{n} Y_i\right)^2 &= \frac{1}{n}\left(\sum_{i=1}^{n} Y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} Y_i\right)^2\right) \\ &= \frac{1}{n}\left(\sum_{i=1}^{n} Y_i^2 - n\overline{Y}^2\right) \\ &= \frac{1}{n}\sum_{i=1}^{n}(Y_i - \overline{Y})^2 \\ &= \frac{n-1}{n}S^2.\end{aligned}$$

That is,

$$\hat{\sigma}^2_{\mathrm{MOM}} = \frac{n-1}{n}S^2.$$