Chapter 7. Statistical Estimation

7.3: Method of Moments Estimation

Slides (Google Drive) Alex Tsun Video (YouTube)

7.3.1 Sample Moments

Maximum likelihood estimation (MLE) as you saw had a nice intuition but mathematically is a bit tedious to solve. We'll learn a different technique for estimating parameters called the Method of Moments (MoM). The early definitions and strategy may be confusing at first, but we provide several examples which hopefully makes things clearer!

Recall the definition of a moment from 5.6:

Definition 7.3.1: Moments (Review)

Let X be a random variable and $c \in \mathbb{R}$ a scalar. Then: The kth moment of X is:

$$\mathbb{E}\left[X^k\right]$$

and the kth moment of X (about c) is:

$$\mathbb{E}\left[(X - c)^k\right]$$

Usually, we are interested in the first moment of X: $\mu = \mathbb{E}[X]$, and the second moment of X about μ: $\text{Var}(X) = \mathbb{E}\left[(X - \mu)^2\right]$.

Now since we are in the statistics portion of the class, we will define a sample moment.

Definition 7.3.2: Sample Moments

Let X be a random variable, and $c \in \mathbb{R}$ a scalar. Let $x_1, \ldots, x_n$ be iid realizations (samples) from X. The kth **sample moment of** X is

$$\frac{1}{n}\sum_{i=1}^{n} x_i^k$$

The kth **sample moment of** X **(about** c**)** is

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - c)^k$$

For example, the first sample moment is just the **sample mean**, and the second sample moment about the sample mean is the **sample variance**.

7.3.2   Method of Moments (MoM)

Recall that the first four moments tell us a lot about the distribution (see 5.6). The first moment is the expectation or mean, and the second moment tells us the variance.

Suppose we only need to estimate one parameter θ (you might have to estimate two for example $\theta = (\mu, \sigma^2)$ for the $\mathcal{N}(\mu, \sigma^2)$ distribution). The idea behind **Method of Moments (MoM)** estimation is that: to find a good estimator, we should have the true and sample moments match as best we can. That is, I should choose the parameter θ such that the first true moment $\mathbb{E}[X]$ is equal to the first sample moment $\bar{x}$. Examples always make things clearer!

> Example(s)
>
> Let's say $x_1, x_2, \ldots, x_n$ are iid samples from $X \sim \text{Unif}(0, \theta)$ (continuous). (These values might look like $x_1 = 3.21, x_2 = 5.11, x_3 = 4.33$, etc.) What is the MoM estimator of θ?

Solution **We then set the first true moment to the first sample moment** as follows (recall that $\mathbb{E}[\text{Unif}(a, b)] = \frac{a+b}{2}$):

$$\mathbb{E}[X] = \frac{\theta}{2} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Solving for θ we get:

$$\hat{\theta}_{MoM} = \frac{2}{n} \sum_{i=1}^{n} x_i$$

That's all there is to it! Much simpler than MLE right?

This estimator makes sense intuitively once you think about it for a bit: if we take the sample mean of a bunch of $\text{Unif}(0, \theta)$ rvs, we expect to get close to the true mean: $\frac{1}{n} \sum_{i=1}^{n} x_i \to \theta/2$ (by the Law of Large Numbers). Hence, a good estimator for θ would just be twice the sample mean!

Notice that in this case, the MoM estimator **disagrees with the MLE** we derived in 7.2!

$$\frac{2}{n} \sum_{i=1}^{n} x_i = \hat{\theta}_{MoM} \neq \hat{\theta}_{MLE} = x_{\max}$$

$\square$

What if you had two parameters instead of just one? Well, then you would set the first true moment equal to the first sample moment (as we just did), but also the second true moment equal to the second sample moment! We'll see an example of this below. But basically, if we have k parameters to estimate, we need k equations to solve for these k unknowns!

Definition 7.3.3: Method of Moments Estimation

Let $\mathbf{x} = (x_1, \ldots, x_n)$ be iid realizations (samples) from probability mass function $p_X(t; \theta)$ (if X is discrete), or from density $f_X(t; \theta)$ (if X continuous), where θ is a parameter (or vector of parameters).

We then define the **method of moments (MoM)** estimator $\hat{\theta}_{MoM}$ of $\theta = (\theta_1, \ldots, \theta_k)$ to be a solution (if it exists) to the k simultaneous equations where, for $j = 1, \ldots, k$, we set the jth (true) moment equal to the jth sample moment:

$$\mathbb{E}[X] = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$\ldots$$

$$\mathbb{E}[X^k] = \frac{1}{n} \sum_{i=1}^{n} x_i^k$$

Example(s)

Let's say $x_1, x_2, \ldots, x_n$ are iid samples from $X \sim \text{Exp}(\theta)$. (These values might look like $x_1 = 3.21, x_2 = 5.11, x_3 = 4.33$, etc.) What is the MoM estimator of θ?

Solution We have $k = 1$ (since only one parameter). **We then set the first true moment to the first sample moment** as follows (recall that $\mathbb{E}[\text{Exp}(\lambda)] = \frac{1}{\lambda}$):

$$\mathbb{E}[X] = \frac{1}{\theta} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Solving for θ (just taking inverse), we get:

$$\hat{\theta}_{MoM} = \frac{1}{\frac{1}{n} \sum_{i=1}^{n} x_i}$$

Notice that in this case, the MoM estimator **agrees with the MLE** (Maximum Likelihood Estimator), hooray!

$$\hat{\theta}_{MoM} = \hat{\theta}_{MLE} = \frac{1}{\frac{1}{n} \sum_{i=1}^{n} x_i}$$

Isn't this way better/easier than MLE?   □

Example(s)

Let's say $x_1, x_2, \ldots, x_n$ are iid samples from $X \sim \text{Poi}(\theta)$. (These values might look like $x_1 = 13, x_2 = 5, x_3 = 4$, etc.) What is the MoM estimator of θ?

Solution We have $k = 1$ (since only one parameter). **We then set the first true moment to the first**

sample moment as follows (recall that $\mathbb{E}\left[\text{Poi}(\lambda)\right] = \lambda$):

$$\mathbb{E}\left[X\right] = \theta = \frac{1}{n}\sum_{i=1}^{n} x_i$$

"Solving" for θ, we get:

$$\hat{\theta}_{MoM} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

In this case, again, the MoM estimator **agrees with the MLE**! Again, much easier than MLE :).   □

Now, we'll do an example where there is more than one parameter.

Example(s)

Let's say $x_1, x_2, \ldots, x_n$ are iid samples from $X \sim \mathcal{N}(\theta_1, \theta_2)$. (These values might look like $x_1 = -2.321, x_2 = 1.112, x_3 = -5.221$, etc.) What is the MoM estimator of the vector $\theta = (\theta_1, \theta_2)$ (θ_1 is the mean, and θ_2 is the variance)?

Solution We have $k = 2$ (since now we have two parameters $\theta_1 = \mu$ and $\theta_2 = \sigma^2$). Notice $\text{Var}\left(X\right) = \mathbb{E}\left[X^2\right] - \mathbb{E}\left[X\right]^2$, so rearranging we get $\mathbb{E}\left[X^2\right] = \text{Var}\left(X\right) + \mathbb{E}\left[X\right]^2$. Let's solve for θ_1 first:

Again, we set the first true moment to the first sample moment:

$$\mathbb{E}\left[X\right] = \theta_1 = \frac{1}{n}\sum_{i=1}^{n} x_i$$

"Solving" for θ_1, we get:

$$\hat{\theta}_1 = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Now let's use our result for $\hat{\theta}_1$ to solve for $\hat{\theta}_2$ (recall that $\mathbb{E}\left[X^2\right] = \text{Var}\left(X\right) + \mathbb{E}\left[X\right]^2 = \theta_2 + \theta_1^2$)

$$\mathbb{E}\left[X^2\right] = \theta_2 + \theta_1^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2$$

Solving for θ_2, and plugging in our result for $\hat{\theta}_1$, we get:

$$\hat{\theta}_2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2$$

If you were to use maximum likelihood to estimate the mean and variance of a Normal distribution, you would get the same result!   □