## 7.6 Method of Moments

The method of moments is a parameter estimation technique that is often easier to implement and compute than the Bayesian or MLE alternatives.

The method itself dates back to the late 19th century, when Karl Pearson published a paper analyzing the distribution of the ratio of forehead width to body length of crabs. Pearson believed observed asymmetrical distribution was resulted from the mixture of two populations, each of which was normal. The proposed mixture model had 5 parameters: 2 means, 2 variances, and a ratio of population sizes. Pearson computed the first 5 sample moments from the data and compared to the theoretical moments, giving a set of 5 equations in 5 unknowns. Pearson then presented this method as an alternative to prevailing Normal approximation method.

**Def:** The $k$th **moment** $\mu_k$ of a random variable $X$ is the value $\mu_k = E[X^k]$. The $k$th **central moment** is the value $E[(X-\mu)^k]$, and the $k$th **standardized moment** is the value $E\left[\left(\frac{X-\mu}{\sigma}\right)^k\right]$.

**Ex 1:** Calculate the 2nd moment of $X \sim \text{Beta}(\alpha, \beta)$.

*Solution.* Recall that $\Gamma(x+1) = x\Gamma(x)$ and

$$1 = \int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}\,dx \qquad \text{so} \qquad \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,dx$$

Then

$$\begin{aligned}
E[X^2] &= \int_0^1 x^2 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}\,dx \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 x^{\alpha+2-1}(1-x)^{\beta-1}\,dx \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+2)\Gamma(\beta)}{\Gamma(\alpha+\beta+2)} \\
&= \frac{\alpha(\alpha+1)\Gamma(\alpha+\beta)\Gamma(\alpha)\Gamma(\beta)}{(\alpha+\beta+1)(\alpha+\beta)\Gamma(\alpha)\Gamma(\beta)\Gamma(\alpha+\beta)} \\
&= \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}
\end{aligned}$$

■

Just as we can estimate the mean and variance (the 2nd central moment) of a distribution by computing corresponding sample moments, we can estimate higher moments using higher sample moments.

**Def:** Let $X_1, \ldots, X_n$ be a random sample. The $k$th **sample moment** is the random variable

$$M_k = \frac{1}{n}\sum_{j=1}^{n} X_j^k$$

The $k$th **central moment** and **standardized moment** are defined analogously.

**Thm:** The $k$th sample moment is a consistent and unbiased estimator for the $k$th moment of a distribution.

*Proof.* By the Law of Large numbers, the sequence $\frac{1}{n}(X_1^k + \cdots + X_n^k)$ converges with probability 1 to the mean of $X_1^k$. But this is exactly the $k$th moment of $X_1$.

Now, by linearity,

$$E[M_k] = \frac{1}{n}\sum E[X_j^k] = \frac{1}{n} n E[X_1^k] = E[X_1^k]$$

which shows that the $k$th sample moment is unbiased. □

**Ex 2:** Caution! While the sample moments are unbiased estimators of the moments of a random variable, this does not mean every estimator built from moments is unbiased!

Recall that the sample variance

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

is a **biased** estimator of the population variance $\sigma^2$, if $X_i \sim N(\mu, \sigma^2)$.

For many families of distributions, the moments (or central moments, or standardized moments) can be expressed as elementary functions of the distribution's parameters. As such, in estimating the moments of a distribution, we also implicitly estimate the values of the parameters (provided we estimate as many moments as we have unknown parameters).

**Def:** To implement the **method of moments** in order to estimate $k$ parameters of a distribution, express the first $k$ moments of the distribution in terms of those parameters, calculate the first $k$ sample moments from observations, set the theoretical moments equal to the sample moment estimates, and solve for the estimates of the parameters.

We'll start with a straightforward example:

**Ex 3:** Suppose $X_1, \ldots, X_n$ form a random sample from Expo($\theta$) with $\theta$ unknown. Find the Method of Moments estimator $\hat{\theta}$ for $\theta$.

*Solution.* To proceed, we need to express the moments of $X \sim \text{Expo}(\theta)$ in terms of the unknown parameter(s). In this case, note that the mean $E[X]$ is the first moment, and $E[X] = \frac{1}{\theta}$. And so solving for $\theta$:

$$\theta = \frac{1}{E[X]}$$

Our estimator for $\theta$ is obtained by replacing the first moment $E[X]$ with the first sample moment $\mathbf{X}$:

$$\hat{\theta} = \frac{1}{\bar{X}}$$

In this case, note that the Method of Moments estimator happens to be the MLE. ∎

**Ex 4:** Suppose $X_1, X_2, \ldots, X_n$ form a random sample from a Beta($\alpha, \beta$) distribution, let $\theta = (\alpha, \beta)$ be the parameter vector. Find the method of moments estimator for $\theta$.

*Solution.* The 1st and 2nd moments $\mu_1$ and $\mu_2$ of the Beta distribution are

$$\mu_1 = \frac{\alpha}{\alpha+\beta} \qquad \mu_2 = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}$$

Now, we need to solve for $\alpha$ and $\beta$ in terms of $\mu_1$ and $\mu_2$.

To start, note that since $\mu_1 = \frac{\alpha}{\alpha+\beta}$, then $\alpha + \beta = \frac{\alpha}{\mu_1}$ and so

$$\mu_2 = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)} = \frac{\alpha(\alpha+1)}{\frac{\alpha}{\mu_1}\left(\frac{\alpha}{\mu_1}+1\right)}$$

Multiplying both sides of the equation by $\frac{1}{\mu_1^2}$:

$$\frac{\mu_2}{\mu_1^2} = \frac{\alpha(\alpha+1)}{\alpha(\alpha+\mu_1)} = \frac{\alpha+1}{\alpha+\mu_1}$$

Then cross-multiplying and solving for $\alpha$ gives

$$\alpha = \frac{\mu_1(\mu_1 - \mu_2)}{\mu_2 - \mu_1^2}$$

To solve for $\beta$, note that $\alpha + \beta = \frac{\alpha}{\mu_1}$ and so substituting the formula for $\alpha$:

$$\beta = \frac{\mu_1 - \mu_2}{\mu_2 - \mu_1^2} - \frac{\mu_1(\mu_1 - \mu_2)}{\mu_2 - \mu_1^2} = \frac{(1 - \mu_1)(\mu_1 - \mu_2)}{\mu_2 - \mu_1^2}$$

Now, our method of moments estimators are obtained by replacing $\mu_1$ and $\mu_2$ with the sample moments $M_1$ and $M_2$:

$$\hat{\alpha} = \frac{M_1(M_1 - M_2)}{M_2 - M_1^2} \qquad \hat{\beta} = \frac{(1 - M_1)(M_1 - M_2)}{M_2 - M_1^2}$$

Finding a formula for the MLE is significantly more complicated and doesn't in general have a closed form expression. ■

**Ex 5:** Let's see how well the Method of Moments estimator performs in a simulation. In R, I'll simulate data for 1000 observations from a particular Beta distribution, then use the formula from the previous part to compute the estimators for $\alpha$ and $\beta$ from the sample, and compare to the known values.