Lecture 39: The method of moments

The method of moments is the oldest method of deriving point estimators.
It almost always produces some asymptotically unbiased estimators, although they may not be the best estimators.

Consider a parametric problem where $X_1, ..., X_n$ are i.i.d. random variables from P_θ, $\theta \in \Theta \subset \mathcal{R}^k$, and $E|X_1|^k < \infty$.
Let $\mu_j = EX_1^j$ be the jth moment of P and let

$$\hat{\mu}_j = \frac{1}{n}\sum_{i=1}^{n} X_i^j$$

be the jth *sample moment*, which is an unbiased estimator of μ_j, $j = 1, ..., k$.
Typically,

$$\mu_j = h_j(\theta), \qquad j = 1, ..., k, \tag{1}$$

for some functions h_j on $\mathcal{R}^k$.
By substituting μ_j's on the left-hand side of (1) by the sample moments $\hat{\mu}_j$, we obtain a *moment estimator* $\hat{\theta}$, i.e., $\hat{\theta}$ satisfies

$$\hat{\mu}_j = h_j(\hat{\theta}), \qquad j = 1, ..., k,$$

which is a sample analogue of (1).
This method of deriving estimators is called the *method of moments*.
An important statistical principle, the *substitution principle*, is applied in this method.

Let $\hat{\mu} = (\hat{\mu}_1, ..., \hat{\mu}_k)$ and $h = (h_1, ..., h_k)$.
Then $\hat{\mu} = h(\hat{\theta})$.
If the inverse function h^{-1} exists, then the unique moment estimator of θ is $\hat{\theta} = h^{-1}(\hat{\mu})$.
When h^{-1} does not exist (i.e., h is not one-to-one), any solution of $\hat{\mu} = h(\hat{\theta})$ is a moment estimator of θ;
if possible, we always choose a solution $\hat{\theta}$ in the parameter space Θ.
In some cases, however, a moment estimator does not exist (see Exercise 111).

Assume that $\hat{\theta} = g(\hat{\mu})$ for a function g.
If h^{-1} exists, then $g = h^{-1}$.
If g is continuous at $\mu = (\mu_1, ..., \mu_k)$, then $\hat{\theta}$ is strongly consistent for θ, since $\hat{\mu}_j \to_{a.s.} \mu_j$ by the SLLN.
If g is differentiable at μ and $E|X_1|^{2k} < \infty$, then $\hat{\theta}$ is asymptotically normal, by the CLT and Theorem 1.12, and

$$\text{amse}_{\hat{\theta}}(\theta) = n^{-1}[\nabla g(\mu)]^\tau V_\mu \nabla g(\mu),$$

where V_μ is a $k \times k$ matrix whose (i, j)th element is $\mu_{i+j} - \mu_i\mu_j$.
Furthermore, the n^{-1} order asymptotic bias of $\hat{\theta}$ is

$$(2n)^{-1}\text{tr}\left(\nabla^2 g(\mu) V_\mu\right).$$

Example 3.24. Let $X_1, ..., X_n$ be i.i.d. from a population P_θ indexed by the parameter $\theta = (\mu, \sigma^2)$, where $\mu = EX_1 \in \mathcal{R}$ and $\sigma^2 = \text{Var}(X_1) \in (0, \infty)$.
This includes cases such as the family of normal distributions, double exponential distributions, or logistic distributions (Table 1.2, page 20).
Since $EX_1 = \mu$ and $EX_1^2 = \text{Var}(X_1) + (EX_1)^2 = \sigma^2 + \mu^2$, setting $\hat{\mu}_1 = \mu$ and $\hat{\mu}_2 = \sigma^2 + \mu^2$ we obtain the moment estimator

$$\hat{\theta} = \left(\bar{X},\ \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2 \right) = \left(\bar{X},\ \frac{n-1}{n} S^2 \right).$$

Note that $\bar{X}$ is unbiased, but $\frac{n-1}{n} S^2$ is not.
If X_i is normal, then $\hat{\theta}$ is sufficient and is nearly the same as an optimal estimator such as the UMVUE.
On the other hand, if X_i is from a double exponential or logistic distribution, then $\hat{\theta}$ is not sufficient and can often be improved.

Consider now the estimation of σ^2 when we know that $\mu = 0$.
Obviously we cannot use the equation $\hat{\mu}_1 = \mu$ to solve the problem.
Using $\hat{\mu}_2 = \mu_2 = \sigma^2$, we obtain the moment estimator $\hat{\sigma}^2 = \hat{\mu}_2 = n^{-1} \sum_{i=1}^{n} X_i^2$.
This is still a good estimator when X_i is normal, but is not a function of sufficient statistic when X_i is from a double exponential distribution.
For the double exponential case one can argue that we should first make a transformation $Y_i = |X_i|$ and then obtain the moment estimator based on the transformed data.
The moment estimator of σ^2 based on the transformed data is $\bar{Y}^2 = (n^{-1} \sum_{i=1}^{n} |X_i|)^2$, which is sufficient for σ^2.
Note that this estimator can also be obtained based on absolute moment equations.

Example 3.25. Let $X_1, ..., X_n$ be i.i.d. from the uniform distribution on (θ_1, θ_2), $-\infty < \theta_1 < \theta_2 < \infty$.
Note that

$$EX_1 = (\theta_1 + \theta_2)/2$$

and

$$EX_1^2 = (\theta_1^2 + \theta_2^2 + \theta_1\theta_2)/3.$$

Setting $\hat{\mu}_1 = EX_1$ and $\hat{\mu}_2 = EX_1^2$ and substituting θ_1 in the second equation by $2\hat{\mu}_1 - \theta_2$ (the first equation), we obtain that

$$(2\hat{\mu}_1 - \theta_2)^2 + \theta_2^2 + (2\hat{\mu}_1 - \theta_2)\theta_2 = 3\hat{\mu}_2,$$

which is the same as

$$(\theta_2 - \hat{\mu}_1)^2 = 3(\hat{\mu}_2 - \hat{\mu}_1^2).$$

Since $\theta_2 > EX_1$, we obtain that

$$\hat{\theta}_2 = \hat{\mu}_1 + \sqrt{3(\hat{\mu}_2 - \hat{\mu}_1^2)} = \bar{X} + \sqrt{\tfrac{3(n-1)}{n} S^2}$$

and

$$\hat{\theta}_1 = \hat{\mu}_1 - \sqrt{3(\hat{\mu}_2 - \hat{\mu}_1^2)} = \bar{X} - \sqrt{\tfrac{3(n-1)}{n} S^2}.$$

These estimators are not functions of the sufficient and complete statistic $(X_{(1)}, X_{(n)})$.

Example 3.26. Let $X_1, ..., X_n$ be i.i.d. from the binomial distribution $Bi(p, k)$ with unknown parameters $k \in \{1, 2, ...\}$ and $p \in (0, 1)$.
Since

$$EX_1 = kp$$

and

$$EX_1^2 = kp(1-p) + k^2p^2,$$

we obtain the moment estimators

$$\hat{p} = (\hat{\mu}_1 + \hat{\mu}_1^2 - \hat{\mu}_2)/\hat{\mu}_1 = 1 - \tfrac{n-1}{n} S^2/\bar{X}$$

and

$$\hat{k} = \hat{\mu}_1^2/(\hat{\mu}_1 + \hat{\mu}_1^2 - \hat{\mu}_2) = \bar{X}/(1 - \tfrac{n-1}{n} S^2/\bar{X}).$$

The estimator $\hat{p}$ is in the range of $(0, 1)$.
But $\hat{k}$ may not be an integer.
It can be improved by an estimator that is $\hat{k}$ rounded to the nearest positive integer.

Example 3.27. Suppose that $X_1, ..., X_n$ are i.i.d. from the Pareto distribution $Pa(a, \theta)$ with unknown $a > 0$ and $\theta > 2$ (Table 1.2, page 20).
Note that

$$EX_1 = \theta a/(\theta - 1)$$

and

$$EX_1^2 = \theta a^2/(\theta - 2).$$

From the moment equation,

$$\tfrac{(\theta-1)^2}{\theta(\theta-2)} = \hat{\mu}_2/\hat{\mu}_1^2.$$

Note that $\frac{(\theta-1)^2}{\theta(\theta-2)} - 1 = \frac{1}{\theta(\theta-2)}$.
Hence

$$\theta(\theta - 2) = \hat{\mu}_1^2/(\hat{\mu}_2 - \hat{\mu}_1^2).$$

Since $\theta > 2$, there is a unique solution in the parameter space:

$$\hat{\theta} = 1 + \sqrt{\hat{\mu}_2/(\hat{\mu}_2 - \hat{\mu}_1^2)} = 1 + \sqrt{1 + \tfrac{n}{n-1} \bar{X}^2/S^2}$$

and

$$\begin{aligned} \hat{a} &= \frac{\hat{\mu}_1(\hat{\theta} - 1)}{\hat{\theta}} \\ &= \bar{X} \sqrt{1 + \tfrac{n}{n-1} \bar{X}^2/S^2} \Big/ \left(1 + \sqrt{1 + \tfrac{n}{n-1} \bar{X}^2/S^2}\right). \end{aligned}$$

Exercise 108. Let $X_1, ..., X_n$ be a random sample from the following discrete distribution:

$$P(X_1 = 1) = \frac{2(1-\theta)}{2-\theta}, \quad P(X_1 = 2) = \frac{\theta}{2-\theta},$$

where $\theta \in (0,1)$ is unknown.
Note that

$$EX_1 = \frac{2(1-\theta)}{2-\theta} + \frac{2\theta}{2-\theta} = \frac{2}{2-\theta}.$$

Hence, a moment estimator of θ is $\hat{\theta} = 2(1 - \bar{X}^{-1})$, where $\bar{X}$ is the sample mean.
Note that

$$\text{Var}(X_1) = \frac{2(1-\theta)}{2-\theta} + \frac{4\theta}{2-\theta} - \frac{4}{(2-\theta)^2} = \frac{4\theta - 2\theta^2 - 4}{(2-\theta)^2},$$

$$\theta = 2(1 - \mu^{-1}) = g(\mu),$$

$$g'(\mu) = 2/\mu^2 = 2/[2/(2-\theta)]^2 = (2-\theta)^2/2.$$

By the central limit theorem and δ-method,

$$\sqrt{n}(\hat{\theta} - \theta) \to_d N\left(0, \frac{(2-\theta)^2(2\theta - \theta^2 - 2)}{2}\right).$$

The method of moments can also be applied to nonparametric problems.
Consider, for example, the estimation of the central moments

$$c_j = E(X_1 - \mu_1)^j, \qquad j = 2, ..., k.$$

Since

$$c_j = \sum_{t=0}^{j} \binom{j}{t} (-\mu_1)^t \mu_{j-t},$$

the moment estimator of c_j is

$$\hat{c}_j = \sum_{t=0}^{j} \binom{j}{t} (-\bar{X})^t \hat{\mu}_{j-t},$$

where $\hat{\mu}_0 = 1$.
It can be shown (exercise) that

$$\hat{c}_j = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^j, \qquad j = 2, ..., k, \tag{2}$$

which are sample central moments.
From the SLLN, $\hat{c}_j$'s are strongly consistent.
If $E|X_1|^{2k} < \infty$, then

$$\sqrt{n}\,(\hat{c}_2 - c_2, ..., \hat{c}_k - c_k) \to_d N_{k-1}(0, D) \tag{3}$$

where the (i,j)th element of the $(k-1) \times (k-1)$ matrix D is

$$c_{i+j+2} - c_{i+1}c_{j+1} - (i+1)c_i c_{j+2} - (j+1)c_{i+2}c_j + (i+1)(j+1)c_i c_j c_2.$$