Lecture 4: Method of Moments Estimation

Definition 4.1

(a) **estimation** – the process of estimating the value of a population parameter based on the characteristics of the population.
(b) **estimator** – the statistic used for estimation (e.g., a random sample $X_1, \ldots, X_n$ taken from a population, the estimator of a parameter η is often denoted $\hat{\eta}(X_1, \ldots, X_n)$, $\hat{\eta}_n$, or just $\hat{\eta}$).
(c) **estimate** – the realized value of the estimator (e.g., for $X_1 = x_1, \ldots, X_n = x_n$, $\hat{\eta}(x_1, \ldots, x_n)$).

Definition 4.2 The model for the population distribution is usually assumed to have a probability density function $f(x;\theta)$ with its parameter $\theta \in \Omega$. The goal is to estimate θ or a function of the parameter $\eta = \eta(\theta)$. **Consistency** (a crucial requirement for estimators) states that an estimator $\hat{\eta}_n$ converges to the target parameter $\eta = \eta(\theta)$ as the sample size increases,

$$\hat{\eta}_n \xrightarrow{p} \eta.$$

Example 4.1 If $E(|X_1|) < \infty$, the sample mean $\bar{X}$ is a *consistent* estimator of the population mean μ by the weak law of large number.

Remark 4.1 In addition to consistency, deriving the **asymptotic distribution** of the estimator allows us to construct confidence intervals for the estimator (i.e., do uncertainty quantification for our estimator).

Definition 4.3 When using a random sample $X_1, \ldots, X_n$ to estimate the population mean $\mu = E(X_1)$, it is natural to use the sample mean

$$\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n).$$

For estimating the population r^{th} moment $m_r = E(X_1^r)$, it is natural to use the sample moment

$$\hat{m}_r = \frac{1}{n}(X_1^r + \cdots + X_n^r).$$

In general, when a parameter is a function of the population moments, for example, $\eta = g(m_1, \ldots, m_k)$, we can estimate the parameter by plugging the sample moments into the function,

$$\hat{\eta} = g(\hat{m}_1, \ldots, \hat{m}_k), \quad \hat{m}_r = \frac{1}{n}(X_1^r + \cdots + X_n^r) \quad (r = 1, \ldots, k).$$

This method is called the **method of moments**, and the statistic $\hat{\eta} = g(\hat{m}_1, \ldots, \hat{m}_k)$ is called the **method of moments estimator (MME)**.

Example 4.2 Suppose a random sample $X_1, \dots, X_n$ follows a distribution with variance $\sigma^2 > 0$,

$$\sigma^2 = \mathrm{Var}(X_1) = E(X_1^2) - (E(X_1))^2 = m_2 - m_1^2, \quad \sigma = \sqrt{m_2 - m_1^2}$$

Therefore, the method of moments estimators for population variance is

$$\hat{\sigma}^2_{\text{MME}} = \hat{m}_2 - (\hat{m}_1)^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 - \left(\frac{1}{n}\sum_{i=1}^n X_i\right)^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$$

Example 4.3 For a bivariate random sample $(X_i, Y_i), \dots, (X_n, Y_n)$, $n > 2$, suppose that the population means are μ_1 and μ_2, and the population variances are σ_1^2 and σ_2^2 ($0 < \sigma_1^2 < \infty$ and $0 < \sigma_2^2 < \infty$). We define the correlation coefficient

$$\rho = \frac{\mathrm{Cov}(X_1, Y_1)}{\sqrt{\mathrm{Var}(X_1)\mathrm{Var}(Y_1)}} = \frac{E(X_1Y_1) - E(X_1)E(Y_1)}{\sqrt{E(X_1^2) - [E(X_1)]^2}\sqrt{E(Y_1^2) - [E(Y_1)]^2}},$$

for $-1 < \rho < 1$. If the population moments are replaced with the sample moments, the method of moments estimator for ρ can be expressed as:

$$\hat{\rho}^{\text{MME}} = \frac{\frac{1}{n}\sum_{i=1}^n X_iY_i - \overline{XY}}{\sqrt{\frac{1}{n}\sum_{i=1}^n X_i^2 - (\bar{X})^2}\sqrt{\frac{1}{n}\sum_{i=1}^n Y_i^2 - (\bar{Y})^2}}.$$

The method of moments estimator for ρ is the sample correlation coefficient.

Remark 4.2 It is possible to have multiple method of moments estimators for a single parameter (i.e., non-uniqueness of the MME estimator).

Example 4.4 Suppose $X_1, \dots, X_n \overset{iid}{\sim} \text{Poisson}(\lambda)$ with $\lambda > 0$. The population mean and population variance are both λ in this case. Therefore, $\lambda = E(X_1) = m_1$ and $\lambda = \mathrm{Var}(X_1) = m_2 - m_1^2$. Depending on which function of the moments is used to express λ, two method of moments estimators can be obtained as follows.

$$\hat{\lambda}_1^{\text{MME}} = \hat{m}_1 = \bar{X}, \quad \hat{\lambda}_2^{\text{MME}} = \hat{m}_2 - (\hat{m}_1)^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2.$$

Remark 4.3 From the lecture 3, we can prove that the sample counterpart $\hat{m}_k$ of the k-th moment m_k for each k will be consistent (by WLLN) and also calculate the asymptotic distribution (by CLT), if it exists. We can use the theorems in the lecture 3 to show the **consistency and asymptotic distribution of the MME** $\hat{\eta}^{\text{MME}} = g(\hat{m}_1, \dots, \hat{m}_k)$ for a continuous function g.

Theorem 4.1 *Suppose a sample $X_1, \ldots, X_n$ drawn from a population, let $m_r = E(X^r)$ (for $r = 1, \ldots, k$) denote the moments of the population. If the estimator η can be expressed as a function of these moments, $\eta = g(m_1, \ldots, m_k)$, the method of moments estimator, which is given by $\hat{\eta}_n^{MME} = g(\hat{m}_1, \ldots, \hat{m}_k)$, converges to the true parameter value η as the sample size n increases. In order words, $\hat{\eta}_n^{MME} \xrightarrow{P} \eta$*

Proof. Using the WLLN and the continuous mapping theorem, it can be shown that the method of moments estimator converges in probability to the true parameter η. □

Theorem 4.2 ***(Multivariate Central Limit Theorem)*** *Let $\{\boldsymbol{X}_n\}_{n=1}^{\infty}$ be a sequence of iid random vectors with common mean $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$ (positive definite). Assume the common mgf, $M(\boldsymbol{t})$ exists in an open neighborhood of $\mathbf{0}$. Then*

$$\sqrt{n}(\bar{\boldsymbol{X}}_n - \boldsymbol{\mu}) \xrightarrow{D} \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma}).$$

Theorem 4.3 *Suppose a sample $X_1, \ldots, X_n$ and the method of moments estimator for the parameter $\eta = g(m_1, \ldots, m_k)$ is given by $\hat{\eta}_n^{MME} = g(\hat{m}_1, \ldots, \hat{m}_k)$. If $E(X^{2k}) < \infty$ and the function $g(\boldsymbol{m}) = g(m_1, \ldots, m_k)$ is continuously differentiable (g is differentiable, and g' is continuous), then*

$$\sqrt{n}(\hat{\eta}_n^{MME} - \eta) \xrightarrow{D} Z, \quad Z \sim \mathcal{N}(0, \sigma^2), \quad \sigma^2 = g'(\boldsymbol{m}) \cdot \boldsymbol{\Sigma} \cdot [g'(\boldsymbol{m})]^T,$$

where $\boldsymbol{\Sigma}_{r,s} = \mathbb{E}(X_i^{r+s}) - \mathbb{E}(X_i^r)\mathbb{E}(X_i^s)$, $r, s \leq k$, is the covariance matrix, and $g'(\boldsymbol{m})$ is the row vector of partials.

Proof. Let $\boldsymbol{Y}_i = (X_i, X_i^2, \ldots, X_i^k)^T$. We know that $\mathbb{E}(\boldsymbol{Y}_i) = \boldsymbol{m} = (m_1, m_2, \ldots, m_k)$ and $\mathrm{Cov}(\boldsymbol{Y}_i) = \boldsymbol{\Sigma}$, where $\boldsymbol{\Sigma}_{r,s} = \mathbb{E}(X_i^{r+s}) - \mathbb{E}(X_i^r)\mathbb{E}(X_i^s) = m_{r+s} - m_r m_s$, for all $i \in \{1, \ldots, n\}$. By the multivariate CLT, we know

$$\sqrt{n}(\bar{\boldsymbol{Y}}_n - \boldsymbol{m}) \xrightarrow{D} \mathcal{N}_k(\mathbf{0}, \boldsymbol{\Sigma}).$$

By assumption, we know $g : \mathbb{R}^k \to \mathbb{R}$ is continuously differentiable. Thus, if we assume the vector of partial derivatives $g'(\boldsymbol{m}) \neq \mathbf{0} \in \mathbb{R}^{1\times k}$, then by the multivariate Delta Method,

$$\sqrt{n}(g(\bar{\boldsymbol{Y}}_n) - g(\boldsymbol{m})) = \sqrt{n}(\hat{\eta}_n^{\mathrm{MME}} - \eta) \xrightarrow{D} \mathcal{N}(0,\, g'(\boldsymbol{m}) \cdot \boldsymbol{\Sigma} \cdot [g'(\boldsymbol{m})]^T).$$

□

[Refer to *Question 4.b* in the Homework 4 as good example]