Method of Moments

Francisco Richter and Ernst Wit

Introduction to Data Science (MSc)

1 Introduction to Method of Moments

The Method of Moments represents one of the oldest and most intuitive approaches to parameter estimation in statistical inference, predating maximum likelihood estimation by several decades. Developed by Karl Pearson in the late 19th century, this method provides a natural bridge between descriptive statistics and formal parameter estimation, making it particularly valuable for understanding the foundations of statistical inference and data science methodology.

The theoretical importance of the Method of Moments extends beyond its historical significance. The approach embodies fundamental principles of statistical reasoning, including the use of sample statistics to estimate population parameters and the connection between theoretical moments and empirical observations. Understanding this method provides essential insights into the relationship between data and parameters, forming a foundation for more advanced estimation techniques.

From a practical perspective, the Method of Moments often provides simple, closed-form estimators that are easy to compute and interpret. While these estimators may not always achieve the optimal efficiency of maximum likelihood estimators, they frequently serve as excellent starting values for iterative optimization procedures and provide robust alternatives when likelihood-based methods encounter computational difficulties.

2 Mathematical Foundations of Moment Theory

2.1 Population Moments and Their Properties

Moments provide a systematic way to characterize probability distributions through their location, spread, shape, and other distributional properties. The theoretical framework of moments connects abstract probability distributions to concrete numerical summaries that can be estimated from data.

Definition 2.1 (Population Moments). *Let X be a random variable with distribution function F. The k-th population moment about the origin is:*

$$\mu'_k = E[X^k] = \int_{-\infty}^{\infty} x^k dF(x)$$

provided the integral exists and is finite.

The existence of moments requires careful consideration of the tail behavior of probability distributions. Heavy-tailed distributions may have infinite moments of high order, limiting the applicability of moment-based methods. The Cauchy distribution, for example, has no finite moments, while the Student's t-distribution has finite moments only up to a certain order determined by its degrees of freedom parameter.

Definition 2.2 (Central Moments)**.** *The k-th central moment is defined as:*

$$\mu_k = E[(X - \mu_1')^k] = E[(X - E[X])^k]$$

where $\mu_1' = E[X]$ is the first moment about the origin (the mean).

Central moments provide information about the shape of the distribution relative to its center. The second central moment is the variance, measuring the spread of the distribution. The third central moment relates to skewness, quantifying the asymmetry of the distribution, while the fourth central moment connects to kurtosis, measuring the heaviness of the distribution's tails.

Theorem 2.1 (Moment Generating Function and Moments)**.** *If the moment generating function $M_X(t) = E[e^{tX}]$ exists in a neighborhood of zero, then:*

$$\mu_k' = \left.\frac{d^k M_X(t)}{dt^k}\right|_{t=0}$$

The moment generating function provides a powerful tool for deriving moments analytically and establishes connections between different probability distributions. The uniqueness theorem for moment generating functions ensures that distributions are completely characterized by their moments when the moment generating function exists.

2.2 Sample Moments and Empirical Estimation

Sample moments provide the empirical counterparts to population moments, enabling the estimation of distributional characteristics from observed data. The connection between sample and population moments forms the foundation of the Method of Moments estimation procedure.

Definition 2.3 (Sample Moments)**.** *Given a random sample $X_1, X_2, \ldots, X_n$, the k-th sample moment about the origin is:*

$$m_k' = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

Sample moments are natural estimators of their population counterparts, inheriting many desirable statistical properties through the law of large numbers and central limit theorem. The sample mean $m_1' = \bar{X}$ is an unbiased estimator of the population mean, while higher-order sample moments provide consistent estimators of the corresponding population moments.

Definition 2.4 (Sample Central Moments)**.** *The k-th sample central moment is:*

$$m_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^k$$

The relationship between sample central moments and population central moments requires careful analysis due to the dependence on the sample mean. The sample variance m_2 is a biased estimator of the population variance, leading to the common adjustment factor $\frac{n}{n-1}$ in the unbiased sample variance formula.

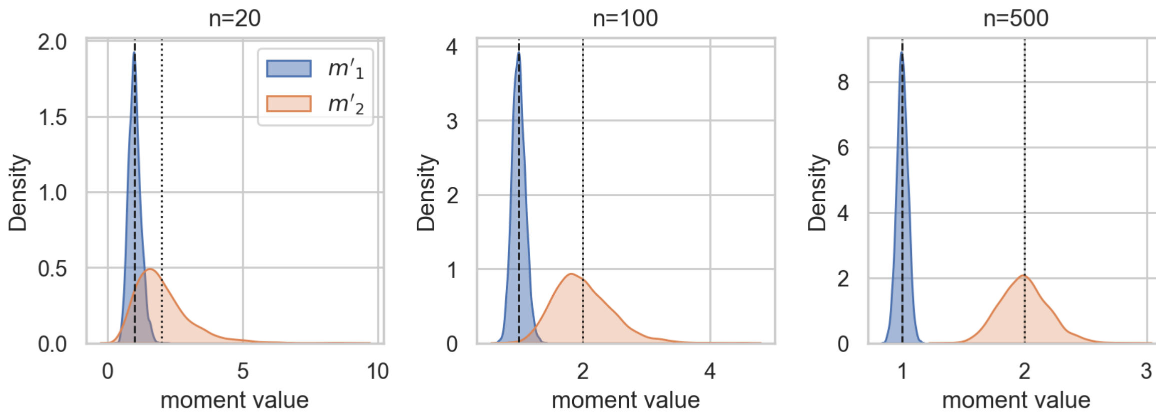

Figure 1: Empirical distributions of the first two raw moments across repeated samples (exponential population). Convergence toward population values increases with n.

3 Method of Moments Estimation Procedure

3.1 Basic Method of Moments

The Method of Moments estimation procedure provides a systematic approach to parameter estimation by equating sample moments to their theoretical counterparts expressed in terms of unknown parameters.

Definition 3.1 (Method of Moments Estimator)**.** *Let $X_1, \ldots, X_n$ be a random sample from a distribution with parameter vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)^T$. The Method of Moments estimators $\hat{\boldsymbol{\theta}}_{MM}$ are obtained by solving the system of equations:*

$$m'_j = \mu'_j(\boldsymbol{\theta}), \quad j = 1, 2, \ldots, k$$

where m'_j are sample moments and $\mu'_j(\boldsymbol{\theta})$ are population moments expressed as functions of the parameters.

The choice of which moments to use in the estimation procedure affects both the computational complexity and the statistical properties of the resulting estimators. Typically, the first k moments are used for a k-parameter distribution, though alternative moment combinations may be preferred in specific situations.

Theorem 3.1 (Existence and Uniqueness of Method of Moments Estimators)**.** *If the system of moment equations has a unique solution in the parameter space for all possible values of the sample moments, then the Method of Moments estimator exists and is unique.*

The existence and uniqueness of Method of Moments estimators depend on the invertibility of the mapping from parameters to moments. Some distributions may have multiple parameter values that produce the same moments, leading to identification problems that require careful analysis.

3.2 Classical Examples and Applications

Example 3.1 (Normal Distribution)**.** *For $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$, the population moments are:*

$$\mu'_1 = \mu \tag{1}$$
$$\mu'_2 = \sigma^2 + \mu^2 \tag{2}$$

Setting sample moments equal to population moments:

$$\bar{X} = \mu \tag{3}$$

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 = \sigma^2 + \mu^2 \tag{4}$$

Solving yields the Method of Moments estimators:

$$\hat{\mu}_{MM} = \bar{X} \tag{5}$$

$$\hat{\sigma}^2_{MM} = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 \tag{6}$$

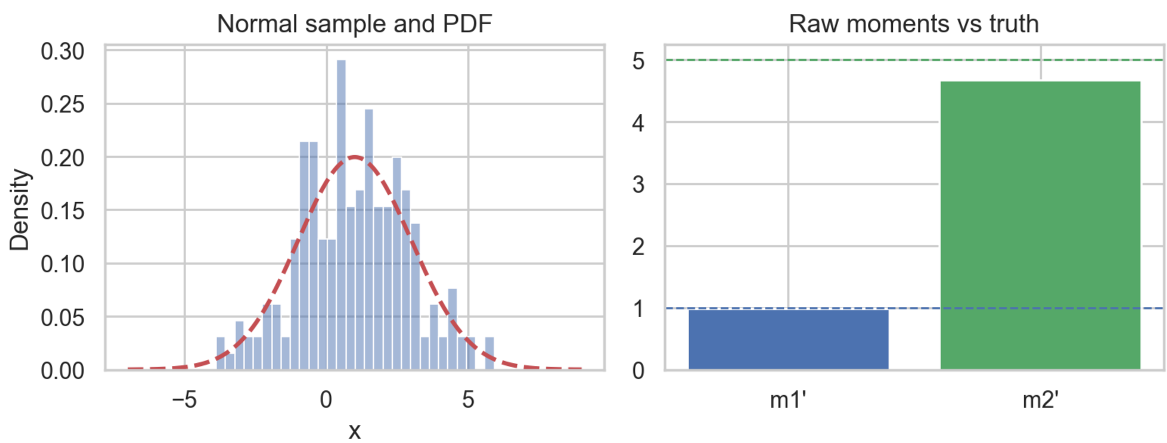

Figure 2: Normal sample with overlaid PDF and comparison of empirical raw moments to their theoretical values.

This example demonstrates that Method of Moments estimators often coincide with intuitive estimators. The estimator for the mean is simply the sample mean, while the estimator for the variance is the sample variance (with divisor n rather than $n-1$).

Example 3.2 (Gamma Distribution)**.** *For the Gamma distribution with shape parameter α and rate parameter β, the population moments are:*

$$\mu_1' = \frac{\alpha}{\beta} \tag{7}$$

$$\mu_2' = \frac{\alpha(\alpha+1)}{\beta^2} \tag{8}$$

The Method of Moments estimators are:

$$\hat{\alpha}_{MM} = \frac{(\bar{X})^2}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2} \tag{9}$$

$$\hat{\beta}_{MM} = \frac{\bar{X}}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2} \tag{10}$$

The Gamma distribution example illustrates how Method of Moments can provide explicit estimators for distributions where maximum likelihood estimation requires numerical optimization.

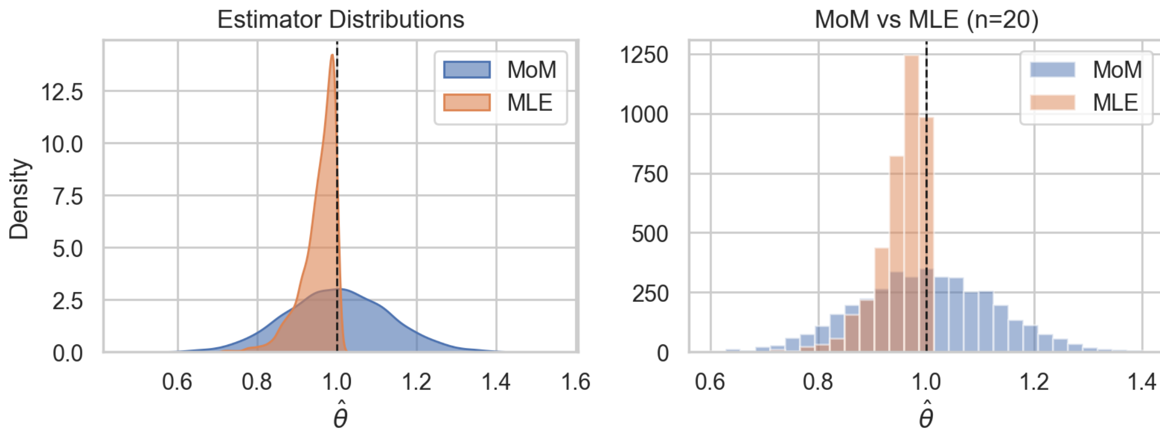

Figure 3: Uniform$(0, \theta)$ estimation: Method of Moments $(2\bar{X})$ versus MLE $(\max X_i)$ distributions across repeated samples $(n = 20)$.

4 Asymptotic Properties of Method of Moments Estimators

4.1 Consistency Theory

The consistency of Method of Moments estimators follows from the consistency of sample moments and the continuity of the parameter-to-moment mapping. This theoretical foundation ensures that Method of Moments estimators improve with larger sample sizes.

Theorem 4.1 (Consistency of Method of Moments Estimators). *Let $\hat{\boldsymbol{\theta}}_{MM}$ be the Method of Moments estimator obtained by solving $m'_j = \mu'_j(\boldsymbol{\theta})$ for $j = 1, \ldots, k$. If:*

1. *The population moments $\mu'_j(\boldsymbol{\theta})$ exist and are finite*
2. *The mapping $\boldsymbol{\theta} \mapsto (\mu'_1(\boldsymbol{\theta}), \ldots, \mu'_k(\boldsymbol{\theta}))$ is one-to-one*
3. *The inverse mapping is continuous at the true parameter value*

Then $\hat{\boldsymbol{\theta}}_{MM} \xrightarrow{P} \boldsymbol{\theta}_0$ as $n \to \infty$.

Proof. By the strong law of large numbers, $m'_j \xrightarrow{a.s.} \mu'_j(\boldsymbol{\theta}_0)$ for each j. The continuous mapping theorem then implies that $\hat{\boldsymbol{\theta}}_{MM} \xrightarrow{P} \boldsymbol{\theta}_0$. □

The consistency result provides theoretical justification for the use of Method of Moments estimators in large-sample applications. The conditions required for consistency are generally mild and are satisfied by most common parametric families used in practice.

4.2 Asymptotic Normality and Efficiency

The asymptotic distribution of Method of Moments estimators can be derived using the delta method, providing the foundation for confidence interval construction and hypothesis testing.

Theorem 4.2 (Asymptotic Normality of Method of Moments Estimators). *Under regularity conditions, the Method of Moments estimator satisfies:*

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{MM} - \boldsymbol{\theta}_0) \xrightarrow{D} N(\mathbf{0}, \mathbf{G}^{-1}\boldsymbol{\Sigma}\mathbf{G}^{-T})$$

where $\mathbf{G}$ *is the Jacobian matrix of the moment functions and* $\boldsymbol{\Sigma}$ *is the covariance matrix of the sample moments.*

The asymptotic covariance matrix $\mathbf{G}^{-1}\boldsymbol{\Sigma}\mathbf{G}^{-T}$ generally differs from the inverse Fisher information matrix, indicating that Method of Moments estimators are typically less efficient than maximum likelihood estimators. However, the efficiency loss may be small in many practical situations, and the computational advantages of Method of Moments may outweigh the efficiency considerations.

Definition 4.1 (Asymptotic Relative Efficiency)**.** *The asymptotic relative efficiency of the Method of Moments estimator relative to the maximum likelihood estimator is:*

$$ARE = \frac{tr(\mathbf{I}^{-1}(\boldsymbol{\theta}_0))}{tr(\mathbf{G}^{-1}\boldsymbol{\Sigma}\mathbf{G}^{-T})}$$

where $\mathbf{I}(\boldsymbol{\theta}_0)$ *is the Fisher information matrix.*

The asymptotic relative efficiency provides a measure of the information loss incurred by using Method of Moments instead of maximum likelihood estimation. For many common distributions, this efficiency loss is modest, making Method of Moments an attractive alternative when computational simplicity is important.

5 Generalized Method of Moments

5.1 Overidentified Systems and Optimal Weighting

When more moment conditions are available than parameters to be estimated, the system becomes overidentified, requiring a criterion for choosing among the potentially conflicting moment conditions. The Generalized Method of Moments (GMM) provides a systematic approach to this problem.

Definition 5.1 (Generalized Method of Moments)**.** *Let* $\mathbf{g}(\mathbf{X}_i, \boldsymbol{\theta})$ *be a vector of moment functions with* $\dim(\mathbf{g}) \geq \dim(\boldsymbol{\theta})$*. The GMM estimator minimizes:*

$$Q_n(\boldsymbol{\theta}) = \left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}(\mathbf{X}_i, \boldsymbol{\theta})\right)^T \mathbf{W}_n \left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}(\mathbf{X}_i, \boldsymbol{\theta})\right)$$

where $\mathbf{W}_n$ *is a positive definite weighting matrix.*

The choice of weighting matrix $\mathbf{W}_n$ affects the efficiency of the GMM estimator. The optimal weighting matrix that minimizes the asymptotic variance is the inverse of the covariance matrix of the moment conditions.

Theorem 5.1 (Optimal GMM Weighting)**.** *The asymptotically efficient GMM estimator uses the weighting matrix:*

$$\mathbf{W}_n = \boldsymbol{\Omega}^{-1}$$

where $\boldsymbol{\Omega} = E[\mathbf{g}(\mathbf{X}_i, \boldsymbol{\theta}_0)\mathbf{g}(\mathbf{X}_i, \boldsymbol{\theta}_0)^T]$ *is the covariance matrix of the moment conditions.*

In practice, the optimal weighting matrix must be estimated, leading to a two-step or iterative estimation procedure. The first step uses an arbitrary weighting matrix (often the identity) to obtain preliminary estimates, which are then used to estimate the optimal weighting matrix for the second step.

5.2 Specification Testing and Model Validation

The overidentifying restrictions in GMM provide a natural framework for testing the validity of the moment conditions and, by extension, the underlying economic or statistical model.

Theorem 5.2 (Hansen's J-Test). *Under the null hypothesis that the moment conditions are correctly specified, the test statistic:*

$$J = n \cdot Q_n(\hat{\boldsymbol{\theta}}_{GMM}) \xrightarrow{D} \chi^2_r$$

where r is the number of overidentifying restrictions.

The J-test provides a formal test of model specification that can detect violations of the moment conditions. Rejection of the null hypothesis suggests that the model is misspecified, either through incorrect functional form assumptions or omitted variables.