

COMPARISON OF SIX NORMAL DISTRIBUTION PARAMETER ESTIMATION METHODS

Pardomuan Robinson Sihombing

BPS-Statistics Indonesia, Jl dr Sutomo No 6-8 Jakarta Pusat

e-mail: robinson@bps.go.id

Abstract

Estimating normal distribution parameters is one of the fundamental problems in statistics that has attracted the attention of researchers for more than two centuries. This study aims to analyse and compare the performance of six normal distribution parameter estimation methods, which include Method of Moments (MoM), Maximum Likelihood Estimation (MLE), Least Squares Estimation (LSE), Bayesian Estimation, Percentile Matching, and Generalised Method of Moments (GMM), through a systematic and comprehensive approach. The methodology of this study combines rigorous theoretical mathematical derivation for each estimation method with empirical evaluation. Through extensive Monte Carlo simulations. Each method was derived mathematically from the basic principle to the final estimator formula, then implemented in a simulation with 500 replications on various sample sizes, i.e. $n = 30, 50, 100$, and 200 , of the normal distribution of $N(5, 4)$. Performance was evaluated in terms of bias, MSE, variance, and MAE to compare six normal distribution parameter estimation methods. The principal findings indicate that MLE and the MoM provide the optimal balance between statistical accuracy and computational efficiency, exhibiting consistent performance across sample sizes. Bayesian Estimation demonstrates remarkable stability and flexibility, proving particularly competitive when prior information is available, whereas the Percentile method serves as a viable alternative when robustness is prioritized, despite showing slightly higher error rates in standard conditions. Practical recommendations suggest prioritizing MLE or MoM for routine applications, while reserving Bayesian approaches for cases requiring comprehensive uncertainty quantification.

Keywords: maximum likelihood, method of moments, Monte Carlo simulation, normal distribution, parameter estimation

1. INTRODUCTION

The normal distribution, first formulated by Abraham de Moivre in 1738 and then comprehensively developed by Carl Friedrich Gauss in the early 19th century, has become the most fundamental and influential probability distribution in modern statistical theory [1]; [2]. This distribution's advantage lies in its elegant mathematical properties and broad applicability in fields ranging from physics to economics, psychology to engineering. Parameters that represent the middle value and that describe the variability of normal distributions play a crucial role in almost all statistical inference procedures, from confidence interval construction to complex hypothesis testing [3]. The history of the development of methods of estimating normal distribution parameters reflects the evolution of statistical thinking over the past nearly three centuries. Each era bringing significant methodological contributions and enriching the arsenal of tools available to practitioners and researchers.

Pearson [4] introduced the Method of Moments as one of the first systematic approaches to parameter estimation, which used the elegant principle of equating the sample moment

with the theoretical moment of the population. This approach, although simple in concept, provided a solid foundation for the later development of estimation theory and is still relevant today due to its ease of implementation and intuitive interpretation. However, the real revolution in estimation theory occurred when Ronald A. Fisher [5] introduced Maximum Likelihood Estimation. This paradigm changed how we look at parameter estimation and laid the foundation for almost all modern statistical inference theories. Maximum Likelihood Estimation, with its optimal asymptotic properties including asymptotic consistency, efficiency, and asymptotic normality, has become the gold standard in parameter estimation for various probability distributions [3]. Knight [6] shows explicitly that for normal distributions, MLE assigns the estimator with the smallest variance among all unbiased estimators, confirming this method's theoretical superiority.

In parallel with the development of MLE, Least Squares Estimation in the context of estimating distribution parameters developed from the pioneering work of Legendre and Gauss in the early 19th century. Later formalised in the probability plot framework by Filliben [7]. This approach uses the elegant linear relationship between empirical and theoretical quantiles, providing parameter estimation and powerful diagnostic tools for validating distribution assumptions [8]. The Bayesian paradigm for the Estimation of normal parameters, developed systematically by Box & Tiao [9], brings a fundamentally different perspective by allowing for the formal incorporation of prior information and providing a complete quantification of uncertainty through posterior distributions. The advantages of Bayesian approaches are becoming increasingly evident in the modern computational era, where techniques such as the Markov Chain Monte Carlo allow for practical implementation of complex models [10].

Meanwhile, Percentile Matching, although seemingly simple, has been shown to have remarkable robustness against outliers and provide satisfactory results in various practical situations, mainly when the data contains extreme observations that can affect estimates from other methods [11]. Generalised Method of Moments, introduced by Hansen [12] in the context of econometrics, provides a sophisticated and flexible framework for parameter estimation when more moment conditions are available than the estimated parameters. GMM extends the methodological toolkit for parameter estimation and enables testing to over-identify constraints that provide diagnostic tools for model specification. This method has become a workhorse in modern econometrics and has wide applications in various fields that require modeling with complex moment conditions.

Although the individual literature for each method is vibrant and extensive, there is a significant gap in comprehensive comparisons that systematically integrate theoretical and empirical aspects. Most existing comparative studies focus on only 2-3 methods or use limited scope simulations, so they do not provide a complete picture of the relative performance of various approaches [13]. Further, many studies do not include detailed

mathematical derivations, which are essential for a fundamental understanding of the characteristics of each method [14]. This gap is becoming increasingly significant, considering that practitioners and researchers often face methodological choices without adequate guidance on the trade-offs and characteristics of the various approaches available. This study is designed to fill this gap by providing a comprehensive comparative framework, which includes empirical evaluation through extensive Monte Carlo simulations and presents a rigorous and complete mathematical derivation for each method. The main objective is to provide evidence-based guidance for practitioners in selecting the most appropriate method based on data characteristics and objective analysis, while also providing theoretical insights that can contribute to future methodological development.

2. RESEARCH METHODOLOGY

This study adopts a Monte Carlo simulation design with a systematic comparative approach, where the methodology consists of two components that complement each other and are seamlessly integrated. The first component involves a rigorous and comprehensive mathematical derivation for each estimation method, presenting the final formula and exploring the underlying assumptions, computational requirements, and theoretical properties of each approach. The second component is an empirical evaluation through extensive Monte Carlo simulations, which is designed to provide insights into finite sample behaviour and relative performance of various methods under controlled and replicable conditions. The six methods that will be used are the Moment Method (MoM), Maximum Likelihood Estimation (MLE), Least Squares Estimation (LSE), Bayesian Estimation, Percentile Matching, and Generalised Method of Moments (GMM)

The Monte Carlo simulation procedure is designed to comprehensively evaluate all six methods across a wide range of sample sizes and conditions. Author used a fixed random seed (`set.seed(1234)`) for reproducibility and data generated. The simulation parameters used include 500 replications per sample size to ensure adequate precision, sample sizes $n = 30, 50, 100, 200$ for capturing small to moderate sample behaviour, actual parameters $\mu = 5$ and $\sigma^2 = 2$ that correspond to the distribution, and implementation in R software version 4.5.1 with appropriate packages for statistical computing and visualisation. For each combination of sample size and replication, random samples were generated from the specified normal distribution, then all six estimation methods were applied to each sample with appropriate error handling for numerical stability. Performance evaluation uses four key metrics that capture different aspects of the quality estimator. Bias is calculated as

$$Bias = E[\hat{\theta}] - \theta = \frac{1}{R} \sum (\hat{\theta}_i - \theta) \quad (1)$$

where $R = 500$ is the number of replications, which measures systematic deviation from the actual parameter value. Mean Squared Error is calculated as

$$MSE = E[(\hat{\theta} - \theta)^2] = \frac{1}{R} \sum (\hat{\theta}_i - \theta)^2 \quad (2)$$

one that combines both bias and variance components and serves as an overall accuracy measure. Variance is calculated as

$$Var = E[(\hat{\theta} - E(\hat{\theta}))^2] = \frac{1}{R} \sum (\hat{\theta}_i - \theta)^2 \quad (3)$$

the variability of an estimator independent of bias. Mean Absolute Error is calculated as

$$MAE = E[|\hat{\theta} - \theta|] = \frac{1}{R} \sum |\hat{\theta}_i - \theta| \quad (4)$$

which provides a robust measure of typical estimation error that is less sensitive to extreme values than MSE.

3. RESULT AND DISCUSSION

3.1 Result

Probability density function, where the main objective is to estimate the parameters based on the random samples observed from the distribution. The selection of normal distributions as the primary focus is based on the fundamental importance of these distributions in statistical theory and widespread applications. Six methods will be used in estimating the parameters of normal distribution functions, namely: $X \sim N(\mu, \sigma^2)$; with

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right) \theta = (\mu, \sigma^2); x_1, x_2, \dots, x_n \quad (5)$$

a. Method of Moments (MoM)

The Method of Moments uses the principle of equating theoretical moments with sample moments. The stages used:

- Step 1: Definition of Theoretical Moment

For normal distributions, the theoretical moments are: $N(\mu, \sigma^2)$

$$\begin{aligned} \mu_k &= E[X^k] \\ \mu_1 &= E[X^1] = \mu \\ \mu_2 &= E[X^2] = \text{Var}(X) + (E[X])^2 = \sigma^2 + \mu^2 \\ \mu_3 &= E[X^3] = \mu^3 + 3\mu\sigma^2 \\ \mu_4 &= E[X^4] = \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4 \end{aligned}$$

- Step 2: Sample Moment

Given observations: $(x_1, x_2, \dots, x_n)$

$$\begin{aligned} m_k &= \frac{1}{n} \sum_{i=1}^n x_i^k \\ m_1 &= \frac{1}{n} \sum_{i=1}^n x_i^1 = \bar{x} \end{aligned} \quad (6)$$

$$\begin{aligned}
m_2 &= \frac{1}{n} \sum_{i=1}^n x_i^2 \\
m_3 &= \frac{1}{n} \sum_{i=1}^n x_i^3 \\
m_4 &= \frac{1}{n} \sum_{i=1}^n x_i^4
\end{aligned} \tag{7}$$

- Step 3: Equation System

Equating a theoretical moment with a sample moment:

$$\mu = m_1 \tag{8}$$

$$\sigma^2 + \mu^2 = m_2 \tag{9}$$

- Step 4: Solution

From eq. (6) and eq. (8), substitute into eq. (9):

$$\begin{aligned}
\hat{\sigma}^2 + (\bar{x})^2 &= m_2 \\
\hat{\sigma}^2 &= m_2 - (\bar{x})^2 \\
\hat{\sigma}^2 &= \frac{1}{n} \sum_{i=1}^n x_i^2 - (\bar{x})^2 \\
\hat{\sigma}^2 &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2
\end{aligned} \tag{10}$$

Thus, the final estimator with the MoM method

$$\hat{\mu}_{\text{MoM}} = \bar{x} \tag{11}$$

$$\hat{\sigma}_{\text{MoM}} = \sqrt{\hat{\sigma}_{\text{MoM}}^2} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \tag{12}$$

b. Maximum Likelihood Estimation (MLE)

MLE looks for parameter values that maximise the likelihood function. The stages used:

- Step 1: Likelihood Function

For independent observation $x_1, x_2, \dots, x_n$ of $N(\mu, \sigma^2)$

$$L(\theta) = f(x_1; \theta) * f(x_2; \theta) * \dots * f(x_n; \theta) = \prod_{i=1}^n f(x_i; \theta)$$

$$L(\mu, \sigma^2) = \prod_{i=1}^n f(x_i; \mu, \sigma^2)$$

$$L(\mu, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right)$$

- Step 2: Likelihood Expansion

$$L(\mu, \sigma^2) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(-\sum_{i=1}^n \frac{(x_i - \mu)^2}{2\sigma^2}\right)$$

$$L(\mu, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right)$$

- Step 3: Log-Likelihood

$$l(\mu, \sigma^2) = \ln L(\mu, \sigma^2)$$

$$l(\mu, \sigma^2) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$$

- Step 4: Partial Derivatives

Towards: μ

$$\frac{\partial l}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu)$$

Towards: σ^2

$$\frac{\partial l}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2$$

- Step 5: First Order Conditions

From: $\frac{\partial l}{\partial \mu} = 0$

$$\sum_{i=1}^n (x_i - \mu) = 0$$

$$\sum_{i=1}^n x_i - n\mu = 0$$

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x} \quad (13)$$

From: $\frac{\partial l}{\partial \sigma^2} = 0$

$$-\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 = 0$$

$$\frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 = \frac{n}{2\sigma^2}$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu})^2 \quad (14)$$

So, the final estimator with the MLE method is:

$$\hat{\mu}_{MLE} = \bar{x} \quad (15)$$

$$\hat{\sigma}_{MLE} = \sqrt{\hat{\sigma}_{MLE}^2} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (16)$$

c. Least Squares Estimation (LSE)

LSE uses a linear relationship between theoretical and empirical quantiles through probability plots. The stages used:

- Step 1: Sorted Statistics and Plotting Positions

From the sample, create the sequenced statistics: $(x_1, x_2, \dots, x_n)$ ($x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$) and calculate plotting positions: for $(p_i = \frac{i-0.5}{n})$ ($i = 1, 2, \dots, n$)

- Step 2: Theoretical Relationships

For the fourth quantile is

$$Q(p) = \mu + \sigma \cdot \Phi^{-1}(p) \quad (17)$$

Where $\Phi^{-1}(p)$ is the inverse of the standard normal distribution CDF.

- Step 3: Linear Regression Model

The relationship between ordered statistics and theoretical quantiles:

$$x_{(i)} \approx \mu + \sigma \cdot \Phi^{-1}(p_i) + \varepsilon_i \quad (18)$$

These are linear models: $Y = \beta_0 + \beta_1 X + \varepsilon$

Where:

$$Y = x_{(i)} (\text{ordered statistics})$$

$$X = \Phi^{-1}(p_i) (\text{theoretical quantiles})$$

$$\beta_0 = \mu$$

$$\beta_1 = \sigma$$

- Step 4: Least Squares Method

Minimisation:

$$SSE = \sum_{i=1}^n [x_{(i)} - (\beta_0 + \beta_1 \cdot \Phi^{-1}(p_i))]^2 \quad (19)$$

Normal equation:

$$\frac{\partial SSE}{\partial \beta_0} = -2 \sum_{i=1}^n [x_{(i)} - \beta_0 - \beta_1 \cdot \Phi^{-1}(p_i)] = 0$$

$$\frac{\partial SSE}{\partial \beta_1} = -2 \sum_{i=1}^n \Phi^{-1}(p_i) [x_{(i)} - \beta_0 - \beta_1 \cdot \Phi^{-1}(p_i)] = 0$$

- Step 5: Normal Equation Solution

From eq. (18):

$$\sum_{i=1}^n x_{(i)} = n\beta_0 + \beta_1 \sum_{i=1}^n \Phi^{-1}(p_i)$$

Because for symmetric plotting positions: $\sum_{i=1}^n \Phi^{-1}(p_i) \approx 0$

$$\hat{\beta}_0 = \frac{1}{n} \sum_{i=1}^n x_{(i)} = \bar{x} \quad (20)$$

From the second equation:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n \Phi^{-1}(p_i) x_{(i)}}{\sum_{i=1}^n [\Phi^{-1}(p_i)]^2} \quad (21)$$

Thus, the final estimator with the LSE method is:

$$\hat{\mu}_{LSE} = \hat{\beta}_0 = \frac{1}{n} \sum_{i=1}^n x_{(i)} = \bar{x} \quad (22)$$

$$\hat{\sigma}_{LSE} = \sqrt{\hat{\sigma}_{LSE}^2} = \hat{\beta}_1 = \frac{\sum_{i=1}^n \Phi^{-1}(p_i) x_{(i)}}{\sum_{i=1}^n [\Phi^{-1}(p_i)]^2} \quad (23)$$

d. Bayesian Estimation

Bayesian Estimation combines prior information with the likelihood of data using Bayes' theorem, where the prior is informative. The stages used:

- Step 1: Prior Specification

Prior conjugate assumptions:

$\mu \sim N(\mu_0, \tau_0^2)$ Prior to the mean

$\sigma^2 \sim \text{InvGamma}(\alpha_0, \beta_0)$ Prior inverse gamma for variance

- Step 2: Likelihood

$$L(x_1, \dots, x_n | \mu, \sigma^2) \propto (\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right) \quad (24)$$

- Step 3: Bayes' Theorem

$$p(\mu, \sigma^2 | x_1, \dots, x_n) \propto p(x_1, \dots, x_n | \mu, \sigma^2) \times p(\mu) \times p(\sigma^2) \quad (25)$$

- Step 4: Posterior to (μ) given (σ^2)

$$p(\mu | \sigma^2, \text{data}) \propto \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right) \times \exp\left(-\frac{1}{2\tau_0^2} (\mu - \mu_0)^2\right) \quad (26)$$

Exponential expansion:

$$\begin{aligned} & -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 - \frac{1}{2\tau_0^2} (\mu - \mu_0)^2 \\ &= -\frac{1}{2} \left[\frac{n}{\sigma^2} (\mu - \bar{x})^2 + \frac{1}{\tau_0^2} (\mu - \mu_0)^2 \right] \\ &= -\frac{1}{2} \left[\left(\frac{n}{\sigma^2} + \frac{1}{\tau_0^2} \right) \mu^2 - 2\mu \left(\frac{n\bar{x}}{\sigma^2} + \frac{\mu_0}{\tau_0^2} \right) + c \right] \end{aligned}$$

Completing the square:

$$\text{Posterior precision: } \tau_n^{-2} = \frac{n}{\sigma^2} + \frac{1}{\tau_0^2}$$

$$\text{Posterior red: } \mu_n = \frac{n\bar{x}/\sigma^2 + \mu_0/\tau_0^2}{n/\sigma^2 + 1/\tau_0^2}$$

$$\text{Therefore: } (\mu | \sigma^2, \text{data}) \sim N(\mu_n, \tau_n^2)$$

- Step 5: Posterior to σ^2

$$p(\sigma^2 | \text{data}) \propto (\sigma^2)^{-n/2} \exp\left(-\frac{SS}{2\sigma^2}\right) \times (\sigma^2)^{-\alpha_0-1} \exp\left(-\frac{\beta_0}{\sigma^2}\right) \quad (27)$$

Where

$$SS = \sum_{i=1}^n (x_i - \bar{x})^2 + \frac{n \cdot \tau_0^{-2}}{n/\sigma^2 + \tau_0^{-2}} (\bar{x} - \mu_0)^2 \quad (28)$$

It provides: $(\sigma^2 | \text{data}) \sim \text{InvGamma}(\alpha_n, \beta_n)$

$$\alpha_n = \alpha_0 + \frac{n}{2} \quad (29)$$

$$\beta_n = \beta_0 + \frac{SS}{2} + \frac{n \cdot \tau_0^{-2} (\bar{x} - \mu_0)^2}{2(n/\sigma^2 + \tau_0^{-2})} \quad (30)$$

Thus, the final estimator with Bayes method is

$$\hat{\mu}_{\text{Bayes}} = E[\mu|\text{data}] = \mu_n \quad (31)$$

$$\hat{\sigma}_{\text{Bayes}}^2 = E[\sigma^2|\text{data}] = \frac{\beta_n}{\alpha_n - 1} \text{ untuk } \alpha_n > 1 \quad (32)$$

e. Percentile Matching

This method equates the percentile of the sample with the theoretical percentile. The stages used:

- Step 1: Percentile Selection

Select two percentiles and generally $p_1 = 0.25, p_2 = 0.75$. Calculate the percentile of the sample: x_{p_1} and x_{p_2}

- Step 2: Theoretical Quantile Functions

For the p quotient is: $X \sim N(\mu, \sigma^2)$

$$Q_X(p) = \mu + \sigma \cdot Q_Z(p) \quad (33)$$

Where $Q_Z(p) = \Phi^{-1}(p)$ is the p-quotient of $N(0,1)$

- Step 3: Equation System

Equalising sample and theoretical quantities:

$$x_{p_1} = \mu + \sigma \cdot \Phi^{-1}(p_1) \quad (34)$$

$$x_{p_2} = \mu + \sigma \cdot \Phi^{-1}(p_2) \quad (35)$$

- Step 4: Solution to σ

Subtract eq. (33) from eq. (34):

$$x_{p_2} - x_{p_1} = \sigma(\Phi^{-1}(p_2) - \Phi^{-1}(p_1)) \quad (36)$$

Therefore:

$$\hat{\sigma} = \frac{x_{p_2} - x_{p_1}}{\Phi^{-1}(p_2) - \Phi^{-1}(p_1)} \quad (37)$$

- Step 5: Solution to μ

Substitute eq. (73) into eq. (34):

$$\begin{aligned} x_{p_1} &= \mu + \hat{\sigma} \cdot \Phi^{-1}(p_1) \\ \hat{\mu} &= x_{p_1} - \hat{\sigma} \cdot \Phi^{-1}(p_1) \end{aligned} \quad (38)$$

Thus, the final estimator with the Percentile Matching method is

$$\hat{\sigma}_{PM} = \frac{x_{p_2} - x_{p_1}}{\Phi^{-1}(p_2) - \Phi^{-1}(p_1)} \quad (39)$$

$$\hat{\mu}_{PM} = x_{p_1} - \hat{\sigma}_{PM} \cdot \Phi^{-1}(p_1) \quad (40)$$

For and $p_1 = 0.25, p_2 = 0.75$

$$\Phi^{-1}(0.25) = -0.6745, \Phi^{-1}(0.75) = 0.6745$$

$$\hat{\sigma}_{PM} = \frac{x_{0.75} - x_{0.25}}{1.3490} \text{ and } \hat{\mu}_{PM} = x_{0.25} + 0.6745 \cdot \hat{\sigma}_{PM}$$

f. Generalised Method of Moments (GMM)

GMM uses more moment conditions than the estimated parameters. The stages used:

- Step 1: Moment Conditions

For the normal distribution, the theoretical moments: $N(\mu, \sigma^2)$

$$E[X] = \mu \text{ and } E[(X - \mu)^2] = \sigma^2$$

$$E[(X - \mu)^3] = 0 \text{ (skewness} = 0) \text{ and } E[(X - \mu)^4] = 3\sigma^4 \text{ (kurtosis} = 3)$$

- Step 2: Formulation of Moment Conditions

The conditions of the moment are:

$$g_1(\theta) = E[X - \mu] = 0$$

$$g_2(\theta) = E[(X - \mu)^2 - \sigma^2] = 0$$

$$g_3(\theta) = E[(X - \mu)^3] = 0$$

$$g_4(\theta) = E[(X - \mu)^4 - 3\sigma^4] = 0$$

where $N(\mu, \sigma^2)$

- Step 3: Sample Moment Conditions

Replace expectations with sample averages:

$$\hat{g}_1(\theta) = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)$$

$$\hat{g}_2(\theta) = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 - \sigma^2$$

$$\hat{g}_3(\theta) = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^3$$

$$\hat{g}_4(\theta) = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^4 - 3\sigma^4$$

- Step 4: GMM Objective Function

Minimisation of square shapes:

$$Q(\theta) = \hat{g}(\theta)'W^{-1}\hat{g}(\theta)$$

For identity weighting matrix: $W: W = I$

$$Q(\theta) = \hat{g}_1^2 + \hat{g}_2^2 + \hat{g}_3^2 + \hat{g}_4^2$$

- Step 5: First Order Conditions

$$\frac{\partial Q}{\partial \mu} = 2\hat{g}_1 \left(\frac{\partial \hat{g}_1}{\partial \mu} \right) + 2\hat{g}_2 \left(\frac{\partial \hat{g}_2}{\partial \mu} \right) + 2\hat{g}_3 \left(\frac{\partial \hat{g}_3}{\partial \mu} \right) + 2\hat{g}_4 \left(\frac{\partial \hat{g}_4}{\partial \mu} \right) = 0$$

$$\frac{\partial Q}{\partial \sigma^2} = 2\hat{g}_2 \left(\frac{\partial \hat{g}_2}{\partial \sigma^2} \right) + 2\hat{g}_4 \left(\frac{\partial \hat{g}_4}{\partial \sigma^2} \right) = 0$$

- Step 6: Partial Derivatives

$$\frac{\partial \hat{g}_1}{\partial \mu} = -1$$

$$\frac{\partial \hat{g}_2}{\partial \mu} = -\frac{2}{n} \sum_{i=1}^n (x_i - \mu)$$

$$\frac{\partial \hat{g}_3}{\partial \mu} = -\frac{3}{n} \sum_{i=1}^n (x_i - \mu)^2$$

$$\frac{\partial \hat{g}_4}{\partial \mu} = -\frac{4}{n} \sum_{i=1}^n (x_i - \mu)^3$$

$$\frac{\partial \hat{g}_1}{\partial \sigma^2} = 0$$

$$\frac{\partial \hat{g}_2}{\partial \sigma^2} = -1$$

$$\frac{\partial \hat{g}_3}{\partial \sigma^2} = 0$$

$$\frac{\partial \hat{g}_4}{\partial \sigma^2} = -12\sigma^2$$

The resulting system numerical solution is nonlinear and requires numerical methods:

$$\hat{\theta}_{\text{GMM}} = \arg \min Q(\theta) \quad (41)$$

From the above analytical proof, the following table can be extracted:

Table 1. Characteristic Estimator

Characteristic	Mom	MLE	LSE	Bayesian	Percentile	GMM
Consistency	✓	✓	✓	✓	✓	✓
Asymptotic Efficiency	X	✓	X	✓*	X	✓
Unbiased (μ)	✓	✓	✓	**	**	✓
Unbiased (σ^2)	X	✓	~	**	**	✓
Asymptotic Normality	✓	✓	✓	✓	✓	✓
Computational Complexity	Low	Keep	Keep	Tall	Low	Tall

*With the right prior; **Depends on specifications; Asymptotic

Furthermore, a comprehensive analysis of 500 Monte Carlo simulation replications yielded valuable insights into the comparative performance of six estimation methods across different sample sizes and evaluation metrics. The overall performance summary shows patterns consistent with theoretical expectations while revealing some surprising findings about practical behaviour from various approaches. Bayesian Estimation showed remarkable consistency with the mean MSE for the parameter μ of 0.068 and for the parameter σ of 0.035, reflecting excellent stability of the posterior-based Inference even with non-informative priors used in this simulation. As predicted by the theory, the Method of Moments and Maximum Likelihood Estimation showed virtually identical performance with almost the same mean MSE of 0.068 μ for and 0.033 for σ , confirming the equivalence of the two approaches to the normal distribution that belongs to the exponential family. Least Squares Estimation shows competitive performance with a slight advantage for parameter estimation, achieving the lowest mean MSE of 0.033 for these σ parameters. This result reflects the effectiveness of the probability plot approach in capturing distributional characteristics, particularly for scale parameters that are critically important in many applications.

Table 2. Evaluation Performance Estimator

n	Method	Parameters	bias	MSE	variance	MAE
30	MoM	μ	-0.021	0.127	0.127	0.284
30	MoM	σ	-0.039	0.067	0.066	0.207
30	MLE	μ	-0.021	0.127	0.127	0.284
30	MLE	σ	-0.039	0.067	0.066	0.207
30	LSE	μ	-0.021	0.127	0.127	0.284
30	LSE	σ	-0.029	0.067	0.067	0.207
30	Bayesian	μ	-0.022	0.127	0.127	0.284
30	Bayesian	σ	0.031	0.072	0.071	0.215
30	Percentile	μ	-0.007	0.155	0.155	0.311
30	Percentile	σ	-0.088	0.162	0.154	0.324
30	GMM	μ	-0.021	0.130	0.130	0.287

n	Method	Parameters	bias	MSE	variance	MAE
30	GMM	σ	-0.031	0.078	0.078	0.217
50	Mom	μ	0.009	0.081	0.081	0.229
50	Mom	σ	-0.006	0.037	0.037	0.151
50	MLE	μ	0.009	0.081	0.081	0.229
50	MLE	σ	-0.006	0.037	0.037	0.151
50	LSE	μ	0.009	0.081	0.081	0.229
50	LSE	σ	-0.004	0.037	0.037	0.151
50	Bayesian	μ	0.009	0.081	0.081	0.229
50	Bayesian	σ	0.036	0.039	0.038	0.157
50	Percentile	μ	0.009	0.102	0.102	0.253
50	Percentile	σ	-0.014	0.100	0.100	0.255
50	GMM	μ	0.009	0.081	0.081	0.229
50	GMM	σ	-0.003	0.038	0.038	0.153
100	MoM	μ	-0.007	0.043	0.043	0.164
100	MoM	σ	-0.019	0.019	0.019	0.109
100	MLE	μ	-0.007	0.043	0.043	0.164
100	MLE	σ	-0.019	0.019	0.019	0.109
100	LSE	μ	-0.007	0.043	0.043	0.164
100	LSE	σ	-0.018	0.019	0.019	0.109
100	Bayesian	μ	-0.007	0.043	0.043	0.164
100	Bayesian	σ	0.002	0.019	0.019	0.108
100	Percentile	μ	-0.008	0.050	0.050	0.179
100	Percentile	σ	-0.026	0.055	0.055	0.186
100	GMM	μ	-0.007	0.043	0.043	0.164
100	GMM	σ	-0.019	0.019	0.019	0.109
200	MoM	μ	-0.009	0.021	0.021	0.115
200	MoM	σ	-0.003	0.010	0.010	0.078
200	MLE	μ	-0.009	0.021	0.021	0.115
200	MLE	σ	-0.003	0.010	0.010	0.078
200	LSE	μ	-0.009	0.021	0.021	0.115
200	LSE	σ	-0.004	0.010	0.010	0.079
200	Bayesian	μ	-0.009	0.021	0.021	0.115
200	Bayesian	σ	0.007	0.010	0.010	0.079
200	Percentile	μ	-0.007	0.026	0.026	0.127
200	Percentile	σ	-0.017	0.024	0.024	0.125
200	GMM	μ	-0.009	0.021	0.021	0.115
200	GMM	σ	-0.003	0.010	0.010	0.078
average	Bayesian	μ	-0.007	0.068	0.068	0.198
average	Bayesian	σ	0.019	0.035	0.035	0.140
average	GMM	μ	-0.007	0.069	0.069	0.199
average	GMM	σ	-0.014	0.036	0.036	0.139
average	LSE	μ	-0.007	0.068	0.068	0.198
average	LSE	σ	-0.014	0.033	0.033	0.137
average	MLE	μ	-0.007	0.068	0.068	0.198
average	MLE	σ	-0.017	0.033	0.033	0.137

n	Method	Parameters	bias	MSE	variance	MAE
average	MoM	μ	-0.007	0.068	0.068	0.198
average	MoM	σ	-0.017	0.033	0.033	0.137
average	Percentile	μ	-0.003	0.083	0.083	0.218
average	Percentile	σ	-0.036	0.085	0.083	0.223

Although theoretically sophisticated, the Generalised Method of Moments showed no substantial improvement over simpler methods, with a mean MSE of 0.069 for μ and 0.036 for σ . These results suggest that for a straightforward normal distribution, the additional moment conditions utilised. GMM do not provide a significant advantage over standard approaches, potentially due to redundancy in information content for this particular distribution. Percentile Matching shows inferior performance compared to other methods, with the highest mean MSE values being 0.083 for μ and 0.085 for σ . This poor performance is particularly pronounced for scale parameter estimation, where the method only utilises two sample percentiles and consequently loses substantial information available in the full sample. However, it is important to note that Percentile Matching has inherent robustness properties not captured by MSE metrics in an ideal simulation environment without outliers or distributional violations.

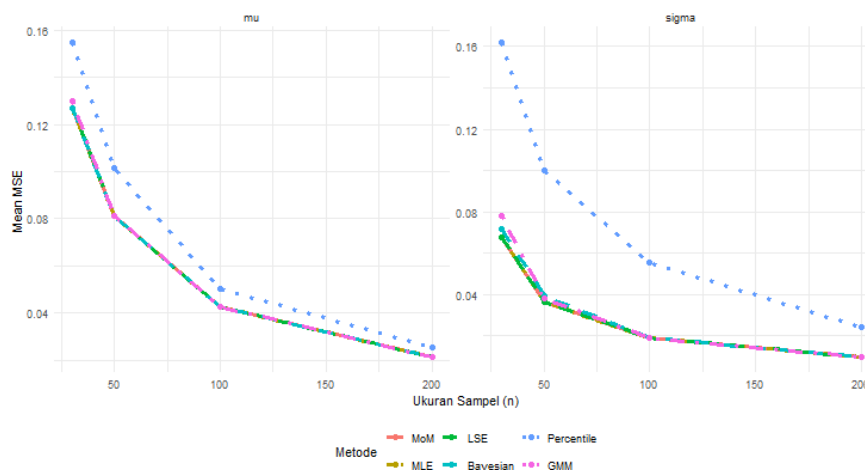


Figure 1. MSE Versus n Trend

Small sample performance with $n = 30$ shows significant differential behaviour across methods and highlights the importance of finite sample considerations in practical applications. The Method of Moments and Maximum Likelihood Estimation maintain their equivalence even in small samples. The value of them with identical bias values of -0.021 for μ and -0.039 σ for, and identical MSE values of 0.127 μ for and 0.067 for σ . Bayesian Estimation shows slightly different behaviour with a bias of -0.022 for μ but a positive bias of 0.031 for σ , reflecting the influence of prior specifications on posterior estimates. The MSE values for the Bayesian approach were 0.127 for μ and 0.072 for σ , showing competitive performance close to MLE/MoM. Least Squares Estimation in small samples showed performance remarkably similar to MLE/MoM for μ parameters, but slightly better for σ Estimation with a bias of -0.029 and MSE 0.067. This result suggests that

probability plot approaches can be efficient for scale parameter estimation even in limited data situations.

The Generalised Method of Moments shows comparable performance with slight degradation, reflecting additional variability introduced by more complex estimation procedures in finite samples. Percentile matching performance in small samples is particularly concerning for Estimation, with a substantial negative bias of -0.088 and a high MSE of 0.162, which is substantially larger than other methods. The bias for Estimation is relatively modest -0.007, but MSE 0.155 is still the highest among all approaches. The variance values for Percentile Matching are 0.155 and 0.154, respectively, indicating high variability contributing significantly to poor MSE performance. Medium sample performance with $n = 50$ and $n = 100$ consistently improved across all methods, with convergence patterns aligned with asymptotic theory predictions. For $n = 50$, the Method of Moments and Maximum Likelihood show identical performance with bias values improving to 0.009 for μ and -0.006 for σ , accompanied by substantial MSE reductions to 0.081 for μ and 0.037 for σ . Bayesian Estimation shows similar improvement with comparable MSE values, while maintaining slightly positive bias for parameters that reflect prior influence.

Progression to $n = 100$ demonstrates continued improvement with further MSE reductions across all methods. MLE and MoM achieve MSE values of 0.043 for μ and 0.019 for σ , representing substantial improvement in smaller sample performance. Bias values become increasingly negligible, with -0.007 for μ and -0.019 for σ , confirming asymptotic unbiasedness properties. Bayesian Estimation achieves very similar performance with MSE 0.043 for μ and 0.019 for σ , showing convergence toward frequentist results as sample size increases and prior influence diminishes. Large sample performance with $n = 200$ demonstrates convergence properties predicted by asymptotic theory, with all methods achieving excellent performance and minimal differences between approaches. The Method of Moments and Maximum Likelihood maintain their equivalence with MSE values dropping to 0.021 for μ and 0.010 for σ , accompanied by minimal bias values of -0.009 for μ and -0.003 for σ . Variance components dominated MSE in large samples, with values of 0.021 for μ and 0.010 for σ the approach theoretical asymptotic values.

Bayesian Estimation in large samples shows virtually identical performance to frequentist approaches, with an MSE of 0.021 for μ and 0.010 for σ , confirming the asymptotic equivalence expected from theory. The bias remains small at -0.009, while the bias becomes slightly positive at 0.007, but overall performance is excellent. Least Squares Estimation maintains competitive performance with MSE values that are very close to other methods, demonstrating the robustness of the probability plot approach across sample sizes. Even Percentile Matching shows substantial improvement in large samples, remaining inferior to other approaches. MSE values improve dramatically to 0.026 for μ and 0.024 for σ , substantially reducing small sample performance. Bias values also improve significantly to -0.007 for μ and -0.017 for σ , showing that asymptotic properties

eventually manifest even for this approach. Trend analysis reveals consistent patterns across all methods, with MSE decreasing approximately proportionally to $1/n$ as predicted by statistical theory for consistent estimators. The convergence rate appears fastest for MLE and MoM, followed closely by Bayesian and LSE approaches. GMM shows slightly slower convergence, particularly for σ estimation, possibly due to additional complexity in the optimisation procedure. Percentile Matching shows improvement but at a slower rate, reflecting fundamental information loss from using only selected quantiles rather than complete sample information.

3.2 Discussion

The results of this comprehensive simulation provide valuable insights into the practical performance of different parameter estimation approaches for the normal distribution, while also confirming many theoretical predictions and revealing some nuanced behaviours that are important for practitioners. The observed equivalence between Maximum Likelihood Estimation and the Method of Moments for normal distribution estimation represents one of the most elegant examples of convergence between different statistical philosophies, reflecting fundamental properties of exponential family distributions where sufficient statistics have a special role in both approaches [3]. The theoretical explanation for this equivalence lies in the fact that the normal distribution belongs to an exponential family, where MLE automatically reduces to the method of moments when sufficient statistics are used as the basis for Estimation, creating an inherent connection between approaches that might otherwise seem quite different in their underlying philosophies. Performance superiority demonstrated by MLE and MoM across virtually all simulation conditions reflects deep theoretical foundations established over more than a century of statistical development. MLE achieves the Cramer-Rao lower bound for unbiased Estimation in large samples, making it optimal among all possible unbiased estimators regarding variance minimisation [15]. Simultaneously, consistency properties ensure estimates converge to actual parameters as sample size increases, providing a reliable foundation for inference procedures.

Method of Moments, while not necessarily achieving the same optimality properties, provides computational simplicity and intuitive interpretation that makes it attractive for many practical applications, particularly when computational resources are limited or quick exploratory analysis is needed. Bayesian estimation performance observed in simulation represents interesting case studies in modern statistical practice, where posterior-based Inference is increasingly gaining acceptance across many disciplines. Stability across different sample sizes reflects the robustness of the Bayesian framework to incorporate uncertainty systematically, even when prior information is minimal or non-informative [10]. Slight differences were observed compared with frequentist approaches, particularly in small samples, illustrating the influence of prior specifications on posterior estimates, highlighting the importance of careful prior elicitation in practical applications. However, the convergence observed in large samples confirms asymptotic equivalence between Bayesian and frequentist approaches to the normal distribution,

reassuring that different philosophical frameworks ultimately lead to similar conclusions when data is abundant.

Least Squares Estimation performance, particularly its competitive showing for scale parameter estimation, reflects the sophisticated nature of the probability plot approach that simultaneously provides parameter estimates and diagnostic information about distributional assumptions. Superior performance for σ estimation likely reflects that the probability plot approach directly utilises the relationship between theoretical and empirical quantiles, particularly sensitive to distributional characteristics that affect scale parameters [16]. This dual functionality makes LSE particularly valuable in exploratory data analysis contexts where parameter estimation and distributional validation are important considerations. Visual diagnostics provided by probability plots offer substantial additional value beyond simple point estimates, enabling practitioners to assess goodness of fit and identify potential departures from the normality assumption. The poor performance of Percentile Matching, particularly for scale parameter estimation, reflects the fundamental limitation of approaches that utilise only selected portions of available data. While the robustness properties of quantile-based methods are well-established for situations involving outliers or heavy-tailed distributions, a simulation environment with ideal normal data penalises information loss inherent in percentile-based approaches [11].

However, practical applications often involve data that deviates from ideal conditions, where robustness can become more important than pure efficiency. The trade-off between efficiency and robustness represents a fundamental consideration in statistical practice, and the choice between methods should reflect characteristics of specific application domains and data quality considerations. While theoretically sophisticated, the generalised Method of Moments performance illustrates that additional complexity does not automatically translate to improved performance for straightforward problems. Normal distribution estimation represents a case where fundamental moment conditions are sufficient to identify parameters uniquely, and additional moment restrictions primarily test distributional assumptions rather than improve estimation efficiency [12]. This finding has broader implications for methodological choices in statistical practice, suggesting that parsimony and simplicity often provide better practical solutions than sophisticated approaches that may introduce additional sources of variability without corresponding benefits.

Sample size effects observed throughout the simulation confirm the fundamental importance of asymptotic theory in understanding estimator behaviour while highlighting the practical relevance of finite sample properties. Dramatic improvements in performance as sample size increases reflect consistent estimator properties, but substantial differences observed in small samples emphasise the importance of choosing appropriate methods for specific sample size contexts. Practitioners working with limited data should pay particular attention to finite sample properties and consider robust

alternatives, while those with large datasets can focus on asymptotic efficiency considerations. The observed convergence patterns provide practical guidance for sample size planning in research contexts. Results suggest that for normal distribution parameter estimation, samples of size 100 or larger generally provide adequate precision for most practical purposes, while samples smaller than 50 may require careful consideration of method choice and uncertainty quantification. These findings align with the general rules of thumb commonly used in statistical practice, providing empirical support for conventional wisdom about minimum sample size requirements.

Implications for practical applications extend beyond simple method selection to broader considerations about statistical methodology in applied research. Results suggest that for straightforward normal distribution problems, computational complexity is rarely justified by performance improvements, supporting preference for interpretable and straightforward approaches. However, complexity can be warranted when additional functionality, such as diagnostic capabilities or robustness properties, provides value beyond pure estimation accuracy. The methodological limitations of the current simulation study include restrictions to ideal normal distribution conditions, the use of a single parameter configuration, and the focus on standard performance metrics. Real-world applications often involve data that deviates from normality assumptions, parameters that vary across broader ranges, and loss functions that differ from standard quadratic forms. Future research should address these limitations through expanded simulation studies incorporating distributional violations, robustness analysis, and alternative evaluation criteria. While not explicitly analysed in the current study, computational considerations represent important practical factors for method selection. MLE and MoM require minimal computational resources and can be implemented easily in any statistical software environment.

Bayesian approaches require more sophisticated computational techniques, particularly for complex models, but modern software makes implementation straightforward for standard problems. GMM requires numerical optimisation that can be computationally intensive and may have convergence issues in challenging applications. These considerations should influence method selection in resource-constrained environments or high-throughput applications. Extensions to more complex distributional settings, such as multivariate normal distributions or mixture models, would provide additional insights into the relative merits of different approaches. Many practical applications involve situations where normal distribution assumption is an approximation rather than the exact truth, and robustness to distributional misspecification becomes a critical consideration. Understanding performance in these more realistic settings would enhance the practical value of comparative analysis and provide better guidance for practitioners working with imperfect data

4. CONCLUSION

The principal findings demonstrate that Maximum Likelihood Estimation (MLE) and the Method of Moments provide the optimal balance between statistical accuracy and computational efficiency, making them the superior default choices for routine normal distribution parameter estimation. Bayesian Estimation exhibits remarkable stability and is highly recommended when prior information is available or when comprehensive uncertainty quantification is critical for decision-making, despite its slightly higher computational demands. Based on this comprehensive analysis, practitioners are advised to prioritize MLE or Method of Moments for general applications, utilize Least Squares when distributional validation is required, and select Percentile Matching when robustness is paramount.

However, this study is primarily limited to the normal distribution under controlled simulation conditions, which may not fully capture the complexities of real-world data exhibiting heavy tails or significant outliers. Future research should extend this comparative framework to non-normal distributions (such as Gamma or Weibull families) and investigate the robustness of these estimators under conditions of severe model misspecification.

DAFTAR PUSTAKA

- [1] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions, Volume 1*, 2nd ed. John Wiley & Sons, 1994.
- [2] R. E. Walpole, *Probability & Statistics for Engineers & Scientists*. USA: Pearson, 2012.
- [3] G. Casella and R. L. Berger, *Statistical Inference*, 2nd ed. Australia: DUXBURY Thomson Learning, 2002.
- [4] K. Pearson, "Contributions to the mathematical theory of evolution," *Philos. Trans. R. Soc. A*, vol. 185, pp. 71–110, 1894.
- [5] B. Efron, "R. A. Fisher in the 21st century: Invited paper presented at the 1996 R. A. Fisher lecture," *Stat. Sci.*, vol. 13, no. 2, pp. 95 – 122, 1998, doi: 10.1214/ss/1028905930.
- [6] K. Knight, *MATHEMATICAL STATISTICS*. Washington, D.C: CHAPMAN & HALL/CRC, 2000.
- [7] J. J. Filliben, "The probability plot correlation coefficient test for normality," *Technometrics*, vol. 17, no. 1, pp. 111–117, 1975, doi: 10.1080/00401706.1975.10489279.
- [8] R. B. D'Agostino and M. A. Stephens, *Goodness-of-Fit Techniques*. Texas: Marcel Dekker Inc., 1986.
- [9] G. E. P. Box and G. C. Tiao, *Bayesian Inference In Statistical Analysis*. New York: JOHN WILEY AND SONS, INC, 1992.
- [10] A. Gelman, J. Carlin, H. Stern, D. Dunson, A. Vehtari, and D. Rubin, *Bayesian Data Analysis Third edition (electronic edition)*, 3rd ed. Chapman and Hall/CRC, 2013. [Online]. Available: <https://sites.stat.columbia.edu/gelman/book/BDA3.pdf>
- [11] S. Kang and J. Song, "Parameter and quantile estimation for the generalized Pareto distribution in peaks over threshold framework," *J. Korean Stat. Soc.*, vol. 46, no. 4, pp. 487–501, Dec. 2017, doi: <https://doi.org/10.2307/1269343>.

- [12] L. P. Hansen, "Large Sample Properties of Generalized Method of Moments Estimators Author(s): Lars Peter Hansen Source:," *Econometrica*, vol. 50, no. 4, pp. 1029–1054, 1982.
- [13] R. V. Hogg, E. A. Tanis, and D. L. Zimmerman, *Probability and Statistical Inference*, 10th ed. New York: Pearson, 1978. doi: 10.2307/2346239.
- [14] M. DeGroot and M. Schervish, *Probability and Statistics Morris*, 4th ed. Edinburgh: Pearson Education Limited Edinburgh, 2014.
- [15] E. L. Lehmann and G. & Casella, *Theory of Point Estimation*, 2nd ed. Springer-Verlag, 1998.
- [16] R. B. D'Agostino, A. J. Belanger, and R. B. D. Jr, "A suggestion for using powerful and informative tests of normality," *Am. Stat.* 44 ., vol. 44, pp. 316–321, 1990.