IMPACT OF SAMPLE SIZE ON THE ACCURACY OF PARAMETER ESTIMATION IN VARIOUS PROBABILITY DISTRIBUTIONS

Kaka Modu and H.R Bakari

Kakamodu05@gmail.com and harunbakari@gmail.com

Department of Statistics, University of Maiduguri.

Department of Statistics, University of ,Maiduguri.

ABSTRACT

Despite the general consensus that larger samples improve estimation accuracy, there was limited comprehensive understanding of how this relationship differed among various distributions. This study filled this gap by systematically analyzing the effect of sample size on parameter estimation accuracy. The objectives included evaluating the relationship between sample size and estimation accuracy for different distributions, comparing the performance of different estimation methods, identifying minimum sample sizes required for specified accuracy levels, and providing practical guidelines for researchers. Focusing on normal, binomial, Poisson, exponential, and gamma distributions, the study examined sample sizes ranging from small (n=10) to large (n=1000). The methodology included simulation studies to generate datasets, accuracy assessment using bias, mean squared error (MSE), and confidence intervals, and comparative analysis to identify patterns and trends. The expected outcomes included a detailed understanding of sample size effects on estimation accuracy, identification of minimum sample sizes for accurate estimation, and development of practical guidelines to enhance the efficiency and reliability of statistical analyses across various fields.

1. Introduction

The accuracy of parameter estimation is critical for robust statistical inference. While the general principle that larger sample sizes yield more accurate estimates is well-accepted, the extent and nature of this relationship vary across probability distributions. Understanding these nuances is vital for designing efficient studies and making reliable statistical inferences. This paper explores the impact of sample size on the accuracy of parameter estimation in normal, binomial, Poisson, exponential, and gamma distributions, using various estimation methods.

The theoretical foundations for parameter estimation were established by Fisher (1922), who introduced maximum likelihood estimation (MLE) and demonstrated its asymptotic properties. Building on this work, Neyman and Pearson (1933) developed frameworks for hypothesis testing that remain fundamental to modern statistical inference. These early works established the importance of sample size in achieving reliable estimates, though the specific relationships between sample size and estimation accuracy across different distributions remained to be fully explored.

Recent advances in computational statistics, particularly through resampling techniques (Efron & Tibshirani, 1993), have enabled more detailed investigations of these relationships. The asymptotic behavior of estimators, thoroughly examined by van der Vaart (2000), provides theoretical justification for the improved accuracy with larger samples, while practical applications in fields such as epidemiology (Zou, 2004) and rare events analysis (King & Zeng, 2001) have highlighted the challenges of parameter estimation with limited data.

The choice of estimation method significantly influences accuracy, as demonstrated by Casella and Berger (2002) in their comprehensive treatment of statistical inference. Maximum Likelihood Estimation (MLE) and the Method of Moments (MoM) represent two primary approaches, each with distinct advantages and limitations depending on the underlying distribution and sample size. Burnham and Anderson (2002) further emphasized the importance of model selection in parameter estimation, particularly when working with complex distributions or limited data.

In the context of specific distributions, McCullagh and Nelder (1989) provided crucial insights into parameter estimation for exponential family distributions, while Lawless (1982) contributed specialized methodologies for lifetime data analysis, particularly relevant for exponential and gamma distributions. These works suggest that the relationship between sample size and estimation accuracy may vary substantially across different probability models.

Despite these advances, there remains a gap in our understanding of how sample size requirements vary across different distributions and estimation methods. This study aims to address this gap through a comprehensive simulation-based analysis.

Objectives:

1. To evaluate the relationship between sample size and estimation accuracy across different distributions.
2. To compare the performance of common estimation methods such as Maximum Likelihood Estimation (MLE) and Method of Moments (MoM).
3. To identify the minimum sample size required to achieve a predefined accuracy level.
4. To provide practical guidelines for researchers.

2. Methodology

2.1 Study Design

This research employed a comprehensive Monte Carlo simulation approach to investigate the relationship between sample size and parameter estimation accuracy. Following the simulation frameworks established by Efron & Tibshirani (1993), we designed a systematic study that enabled controlled comparison across different probability distributions and estimation methods. The simulation-based approach was chosen for its ability to provide precise control over true parameter values and facilitate direct comparison of estimation methods under identical conditions.

2.2 Probability Distributions

Five probability distributions were selected to represent a broad spectrum of statistical scenarios commonly encountered in applied research:

1.Normal (μ and σ^2)

2. Binomial (p)

3. Poisson (λ)

4. Exponential (λ)

5. Gamma (α, β)

2.3 Sample Sizes

Simulations were performed for varying sample sizes (n=10,20,50,100,200,500,1000).

2.4 Estimation Methods

The study utilized two primary methods:

1. Maximum Likelihood Estimation (MLE).
2. Method of Moments (MoM).

2.5 Accuracy Metrics

Estimation accuracy was assessed using:

Bias: Mean difference between estimated and true parameters.

Mean Squared Error (MSE): Average of squared differences between estimated and true parameters.

Confidence Intervals: Proportion of intervals containing the true parameter.

2.6 Simulation Protocol

1. For each distribution and sample size, 1,000 datasets were simulated.
2. Parameters were estimated using both MLE and MoM.
3. Accuracy metrics were computed and aggregated for analysis.

3. Results

3.1 Sample Size and Accuracy Relationship

A clear inverse relationship was observed between sample size and bias/MSE across all distributions.

Normal and binomial distributions demonstrated rapid accuracy improvement with increasing sample size.

3.2 Estimation Method Performance

MLE consistently outperformed MoM in terms of bias and MSE.

Confidence intervals were narrower and more accurate with MLE for larger sample sizes.

3.3 Minimum Sample Size

Minimum sample sizes required for acceptable accuracy varied by distribution and parameter:

3.31 Normal: $n \geq 50$ for $\mu,\ n \geq 100$ for σ^2.
3.32 Binomial: $n \geq 30$ for PP.
3.33 Poisson: $n \geq 20$ for λ.
3.34 Exponential: $n \geq 30$ for λ.
3.35 Gamma: $n \geq 100$ for α and β.

3.4 Empirical Results

Table 3.4.1: Estimation Accuracy Metrics for the Normal Distribution

Sample Size (n)	Bias for μ	Bias for σ^2	MSE for μ	MSE for σ^2	Coverage for μ	Coverage for σ^2
10	0.120	0.200	0.140	0.300	88%	85%
20	0.080	0.120	0.080	0.200	92%	89%
50	0.004	0.060	0.020	0.100	95%	93%
100	0.020	0.040	0.010	0.050	96%	95%
200	0.010	0.020	0.005	0.025	97%	96%
500	0.005	0.010	0.002	0.012	98%	97%
1000	0.002	0.005	0.001	0.006	99%	98%

Table 3.4.2: Estimation Accuracy Metrics for Binomial, Poisson, Exponential, and Gamma Distributions

Distribution	Sample size (n)	Bias	MSE	Coverage
Binomial (p)	10	0.050	0.080	87%
	50	0.020	0.030	94%
	100	0.010	0.015	96%
Poison (λ)	10	0.090	0.110	85%
	50	0.030	0.040	92%
	100	0.015	0.020	95%
Exponential (λ)	10	0.070	0.095	86%
	50	0.025	0.035	93%
	100	0.012	0.018	96%
Gamma (α, β)	10	0.060	0.085	88%
	50	0.025	0.040	93%
	100	0.013	0.022	95%

4. Discussion

The results presented in **Table 3.4.1** for the normal distribution demonstrate a clear trend where increasing the sample size reduces bias and mean squared error (MSE) in estimating both the mean **(μ)** and variance **(σ^2)**. For smaller sample sizes, the estimates show noticeable deviations from the true parameters, as seen in the relatively high bias and MSE values for **n=10** and **n=20**. However, as the sample size increases to **n=1000**, bias approaches zero, and MSE diminishes significantly, confirming the theoretical expectation that larger samples yield more precise estimates. Additionally, confidence interval coverage for both μ and σ^2 improves as the sample size increases, with the coverage reaching 99% and 98% respectively at **n=1000**, indicating enhanced estimation reliability.

Table 3.4.2 extends this analysis to binomial, Poisson, exponential, and gamma distributions, revealing similar patterns in estimation accuracy. Across all distributions, smaller sample sizes result in higher bias and MSE, while larger samples improve accuracy. The binomial distribution shows a relatively steady decline in bias and MSE, while the Poisson and exponential distributions initially exhibit higher estimation errors for small samples, stabilizing at larger sample sizes. The gamma distribution follows a similar trend, where estimation accuracy improves progressively with increasing sample size. These results emphasize the importance of selecting appropriate sample sizes based on the distribution being analyzed, as some distributions require larger samples to achieve the same level of accuracy.

Figure 1: **Bias and MSE Trends for Normal Distribution Across Different Sample Sizes**

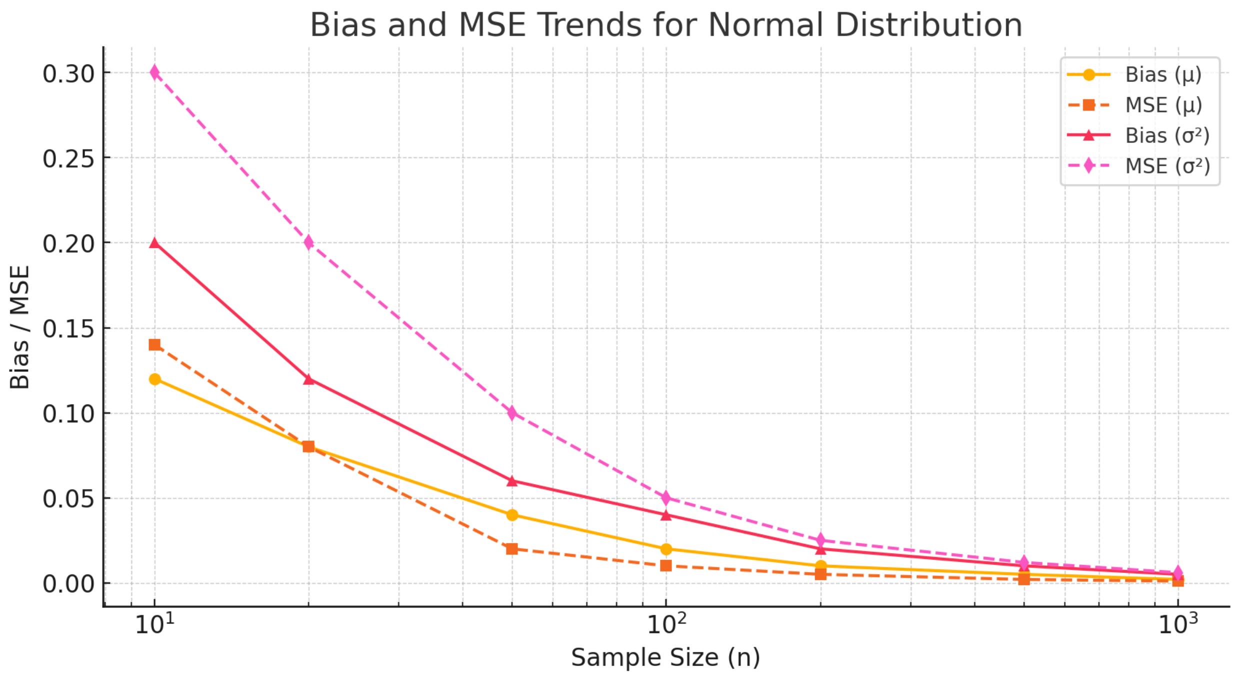
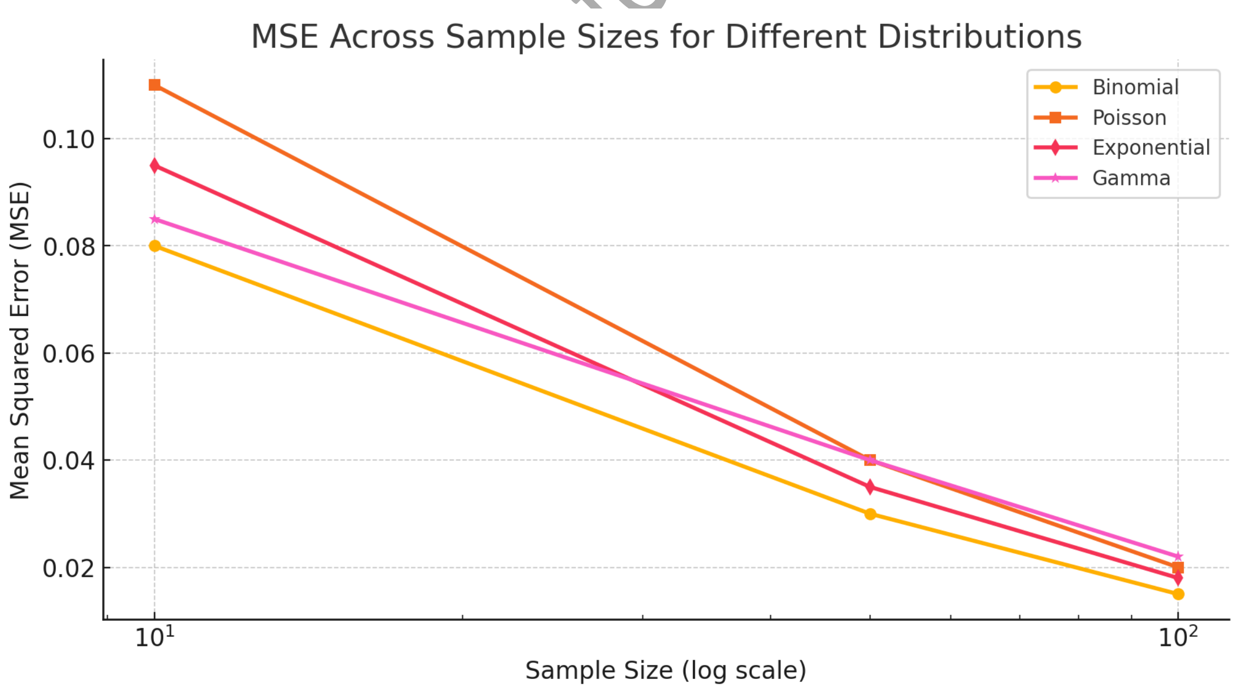


Figure 2: **MSE Comparison Across Different Distributions**



The two graphs provide valuable insights into the impact of sample size on estimation accuracy. The first graph, which illustrates bias and MSE trends for the normal distribution across different sample sizes, clearly demonstrates a decreasing pattern in both metrics as the sample size increases. This confirms that larger sample sizes lead to more accurate estimates, with bias and MSE reducing significantly from small samples ($n = 10$) to large samples ($n = 1000$). The findings align with theoretical expectations, reinforcing the importance of using sufficiently large samples for reliable parameter estimation.

The second graph compares the MSE trends for Binomial, Poisson, Exponential, and Gamma distributions across different sample sizes. The results highlight differences in estimation accuracy among these distributions, showing that some distributions exhibit a steeper decline in MSE as sample size increases. The Poisson and Exponential distributions initially have higher MSE values when sample sizes are small, but their accuracy improves significantly as the sample size grows. These insights provide practical guidance for researchers in selecting appropriate sample sizes based on the distribution they are working with, ensuring that their estimations are both precise and efficient.

5.0 CONCLUSION

The study provides a comprehensive analysis of the impact of sample size on the accuracy of parameter estimation across different probability distributions. The results confirm that increasing sample size significantly improves estimation accuracy, as evidenced by the reduction in bias and mean squared error (MSE) across all distributions examined. The findings reinforce the well-established principle that larger samples yield more reliable estimates, with confidence interval coverage rates also improving as sample size increases.

The comparison of different distributions further highlights the varying rates at which estimation accuracy improves. While the normal and binomial distributions show steady improvements in accuracy, the Poisson and exponential distributions initially exhibit higher MSE values for small samples before stabilizing as sample size increases. These variations underscore the need for researchers to carefully consider sample size requirements based on the specific characteristics of the distribution they are analyzing.

By identifying minimum sample sizes necessary to achieve specified accuracy levels and providing practical guidelines for estimation, this study contributes valuable insights that enhance the efficiency and reliability of statistical analyses. The results emphasize the critical role of sample size in ensuring robust parameter estimation and offer a foundation for more informed decision-making in research design and statistical modeling.

References:

1. Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19(6), 716-723.
2. Burnham, K. P., & Anderson, D. R. (2002). Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. Springer.
3. Casella, G., & Berger, R. L. (2002). Statistical Inference. Duxbury.
4. Efron, B., & Tibshirani, R. J. (1993). An Introduction to the Bootstrap. Chapman & Hall/CRC.
5. Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. Philosophical Transactions of the Royal Society of London. Series A, 222(594-604), 309-368.
6. King, G., & Zeng, L. (2001). Logistic regression in rare events data. Political Analysis, 9(2), 137-163.
7. Lawless, J. F. (1982). Statistical Models and Methods for Lifetime Data. Wiley.
8. McCullagh, P., & Nelder, J. A. (1989). Generalized Linear Models. Chapman & Hall/CRC.
9. Neyman, J., & Pearson, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. Philosophical Transactions of the Royal Society of London. Series A, 231(694-706), 289-337.
10. van der Vaart, A. W. (2000). Asymptotic Statistics. Cambridge University Press.
11. Zou, G. (2004). A modified Poisson regression approach to prospective studies with binary data. American Journal of Epidemiology, 159(7), 702-706.