Astrophysical Insights into Black Hole Mass Distribution: A Computational Statistic Inference Approach via Stochastic Modeling under the Likelihood Principle

Ilyan Radjabaly

Boston College, Morrissey School of Arts & Sciences

December 16, 2024

Abstract

Stellar black holes, formed as remnants of massive star collapses, represent one of the most fascinating and enigmatic phenomena in astrophysics. Their mass distributions hold significant implications for understanding stellar evolution, binary interactions, and the influence of environmental factors such as metallicity. However, observational data is inherently incomplete and biased due to detection limitations, necessitating rigorous statistical methods for meaningful analysis. This research explores the parametric modeling of black hole mass distributions through the likelihood principle, focusing on two primary candidate models: the power-law and log-normal distributions. Leveraging maximum likelihood estimation (MLE) alongside advanced tools such as the Cramer-Rao lower bound theorem, the Lehmann-Scheffe theorem, and hypothesis testing frameworks including the Neyman-Pearson lemma and generalized likelihood ratio tests, we aim to investigate the underlying stochastic or hierarchical processes governing black hole mass distrubution. The study includes a rigorous exploration of estimator properties such as efficiency, consistency, sufficiency and completeness to ensure robust parameter inference, as well as an analysis of ordered statistics, convergence, and test power. By linking statistical inference with astrophysical theories, our work offers a comprehensive framework for interpreting black hole mass data, providing robust insights into the processes shaping these extraordinary celestial objects.

Contents

1 Introduction

Throughout the history of statistical inference, many methodologies have been proposed to estimate unknown parameters based on observed data. However, a groundbreaking revolution in statistical theory emerged with the advent of the **Likelihood Principle**. This principle, which asserts that all evidence about a parameter contained in the data is encapsulated within the likelihood function, fundamentally shifted the paradigm of statistical reasoning. Gone were the days of reliance on sampling distributions and ad-hoc methods; instead, a new era of inference, rooted in the likelihood function, began to dominate the landscape of modern statistics.

The inception of the likelihood-based approach can be traced back to the pioneering work of **Ronald Fisher** in the early 20th century. Fisher's introduction of the *Likelihood Function* and the *Maximum Likelihood Estimation* (MLE) method was nothing short of revolutionary. By proposing that inference should be based on the maximization of the likelihood function, Fisher provided statisticians with a powerful and unified tool to estimate unknown parameters in a probabilistic framework. This principle has since been adopted across disciplines, particularly in astrophysics, where it plays a crucial role in extracting meaningful insights from cosmic observations.

In this research, we employ the MLE methodology to analyze black hole mass distributions. The motivation behind this approach lies in its ability to extract the most probable values of astrophysical parameters, given the observed data, under a stochastic modeling framework.

1.1 The Intuition Behind the Maximum Likelihood Estimation (MLE) Method

The Maximum Likelihood Estimation (MLE) method is one of the most fundamental techniques in statistical inference. The core idea of MLE is simple yet profound: given an observed dataset, the best estimate of an unknown parameter θ is the value that maximizes the probability (or density) of the observed data occurring.

Mathematically, if we have a dataset $X_1, X_2, \ldots, X_n$ generated from a probability distribution parameterized by θ, then the likelihood function is given by:

$$L(\theta) = \prod_{i=1}^{n} f(X_i; \theta). \tag{1}$$

MLE finds the value of θ that maximizes this likelihood function, i.e.,

$$\hat{\theta} = \arg\max_{\theta} L(\theta). \tag{2}$$

Taking the logarithm (which simplifies computation), we instead maximize the log-likelihood function:

$$\ell(\theta) = \sum_{i=1}^{n} \ln f(X_i; \theta). \tag{3}$$

The intuition is that this approach selects the parameter value θ that makes the observed data most probable, leading to an estimator that is consistent, asymptotically normal, and efficient under general conditions.

1.2 Theoretical Justification

One of the most profound results in statistical inference is the asymptotic optimality of the Maximum Likelihood Estimator. The following theorem establishes that MLE not only provides an estimate of the true parameter but does so in a way that is asymptotically superior to any other estimation method.

Theorem: Asymptotic Dominance of MLE - Let $X_1, \ldots, X_n$ be a random sample from a distribution with true parameter θ_0. Then for any other $\theta \neq \theta_0$, we have:

$$\lim_{n \to \infty} P_{\theta_0}[L(\theta_0) > L(\theta)] = 1. \tag{4}$$

This result asserts that, asymptotically, the likelihood function is maximized at the true parameter θ_0, meaning that the MLE converges to the actual parameter in probability. This theorem is groundbreaking as it justifies why MLE is not just a reasonable estimator but the best one in an asymptotic sense.

1.3 Proof of the Asymptotic Superiority of MLE

We establish that, asymptotically, the likelihood function is maximized at the true parameter θ_0. This ensures that the Maximum Likelihood Estimator (MLE) $\hat{\theta}$ converges to θ_0 as $n \to \infty$.

Step 1: Reformulating the Likelihood Function and Its Logarithm

We define the likelihood function as:

$$\mathrm{Lik}(\theta) = L(\theta) = \prod_{i=1}^{n} f(X_i; \theta). \tag{5}$$

The goal is to show that:

$$L(\theta_0) > L(\theta), \quad \forall \theta \neq \theta_0, \text{ asymptotically.} \tag{6}$$

Taking the natural logarithm:

$$0 > \ln L(\theta) - \ln L(\theta_0). \tag{7}$$

Expanding using the definition of the likelihood function:

$$\ln L(\theta) = \sum_{i=1}^{n} \ln f(X_i; \theta), \tag{8}$$

which leads to:

$$0 > \sum_{i=1}^{n} \left(\ln f(X_i; \theta) - \ln f(X_i; \theta_0) \right). \tag{9}$$

Rewriting the right-hand side as a sum of log-ratios:

$$0 > \sum_{i=1}^{n} \ln \frac{f(X_i; \theta)}{f(X_i; \theta_0)}. \tag{10}$$

Step 2: Applying the Law of Large Numbers (LLN)

Define the log-likelihood ratio term:

$$W_i = \ln \frac{f(X_i; \theta)}{f(X_i; \theta_0)}. \tag{11}$$

The expression in step 1 now becomes:

$$\sum_{i=1}^{n} W_i < 0. \tag{12}$$

Since $X_1, X_2, \ldots, X_n$ are i.i.d. from $f(X; \theta_0)$, the sequence $\{W_i\}$ is also an i.i.d. sequence with expectation:

$$\mathbb{E}_{\theta_0}[W_i] = \mathbb{E}_{\theta_0}\left[\ln \frac{f(X;\theta)}{f(X;\theta_0)}\right]. \tag{13}$$

By the Law of Large Numbers (LLN), the sample mean converges to this expectation:

$$\frac{1}{n}\sum_{i=1}^{n} W_i \to \mathbb{E}_{\theta_0}[W_i] \quad \text{in probability as } n \to \infty. \tag{14}$$

Step 3: Jensen's Inequality and Convexity of $-\ln x$

To analyze $\mathbb{E}_{\theta_0}[W_i]$, we use Jensen's inequality. The function $g(x) = -\ln x$ is convex because its second derivative is positive:

$$g''(x) = \frac{1}{x^2} > 0 \quad \forall x > 0. \tag{15}$$

Applying Jensen's inequality to the expectation:

$$\mathbb{E}_{\theta_0}\left[-\ln \frac{f(X;\theta)}{f(X;\theta_0)}\right] \geq -\ln \mathbb{E}_{\theta_0}\left[\frac{f(X;\theta)}{f(X;\theta_0)}\right]. \tag{16}$$

We explicitly compute the expectation:

$$\mathbb{E}_{\theta_0}\left[\frac{f(X;\theta)}{f(X;\theta_0)}\right] = \int \frac{f(x;\theta)}{f(x;\theta_0)} f(x;\theta_0)dx. \tag{17}$$

Since the densities $f(x;\theta_0)$ cancel:

$$\int f(x;\theta)dx = 1. \tag{18}$$

Thus, we obtain:

$$\mathbb{E}_{\theta_0}\left[\frac{f(X;\theta)}{f(X;\theta_0)}\right] = 1. \tag{19}$$

Using this result in the Jensen inequality step:

$$-\ln 1 = 0. \tag{20}$$

This implies:

$$\mathbb{E}_{\theta_0}\left[\ln \frac{f(X;\theta)}{f(X;\theta_0)}\right] \leq 0. \tag{21}$$

Since θ_0 is the true parameter and $f(X;\theta_0)$ is the correct density, the expectation is strictly negative for $\theta \neq \theta_0$:

$$\mathbb{E}_{\theta_0}\left[\ln \frac{f(X;\theta)}{f(X;\theta_0)}\right] < 0. \tag{22}$$

Step 4: Convergence to a Negative Sum and Probability Limit

From Step 2, we established that:

$$\frac{1}{n}\sum_{i=1}^{n} \ln \frac{f(X_i;\theta)}{f(X_i;\theta_0)} \to \mathbb{E}_{\theta_0}\left[\ln \frac{f(X;\theta)}{f(X;\theta_0)}\right] < 0 \quad \text{in probability.} \tag{23}$$

Multiplying both sides by n, we obtain:

$$\sum_{i=1}^{n} \ln \frac{f(X_i;\theta)}{f(X_i;\theta_0)} \to -\infty \quad \text{in probability.} \tag{24}$$

Thus, with probability approaching 1,

$$\sum_{i=1}^{n} \ln \frac{f(X_i;\theta)}{f(X_i;\theta_0)} < 0. \tag{25}$$

Equivalently,

$$P_{\theta_0}\left[L(\theta_0) > L(\theta)\right] \to 1 \quad \text{as } n \to \infty. \tag{26}$$

Conclusion: The Asymptotic Superiority of MLE

This result proves that the likelihood function is asymptotically maximized at θ_0. Since the MLE $\hat{\theta}$ is defined as the parameter that maximizes $L(\theta)$, we conclude:

$$\hat{\theta} \to \theta_0 \quad \text{in probability as } n \to \infty. \tag{27}$$

This establishes that MLE is **asymptotically superior** in the sense that it converges to the true parameter θ_0 with probability approaching 1, making it a powerful and justified method for parameter estimation.

This completes the proof, demonstrating that the true parameter θ_0 asymptotically maximizes the likelihood function, justifying the use of Maximum Likelihood Estimation as an optimal inference method.

1.4 Astrophysical Background

Stellar black holes, formed from the gravitational collapse of massive stars, typically possess masses ranging from $5M_\odot$ to $50M_\odot$, where $M_\odot$ denotes the solar mass. These compact objects are often found in binary systems or as isolated remnants of supernovae. Their mass distribution is influenced by various factors, including the initial mass of progenitor stars, metallicity, rotation rates, and the mechanisms of supernova explosions.

Understanding the mass distribution of stellar black holes provides vital clues about stellar evolution and the dynamics of binary star systems. For instance, the prevalence of certain mass ranges can indicate dominant formation channels, while deviations from theoretical predictions may suggest the influence of exotic processes such as pair-instability supernovae or mass transfer in binaries. Consequently, statistical analysis of black hole masses is not only a tool for astrophysics but also a bridge to uncovering deeper physical mechanisms.

1.5 Statistical Motivation

Observations of stellar black hole masses are often derived from gravitational wave detections (e.g., LIGO/Virgo) or X-ray binary systems. However, these datasets are plagued by selection biases and incompleteness, as detection capabilities vary across mass ranges. For example, low-mass black holes may escape detection in gravitational wave observations due to weaker signals, while high-mass black holes may dominate due to hierarchical mergers.

To address these challenges, robust statistical modeling is essential. Parametric models such as the power-law and log-normal distributions offer compelling frameworks to describe black hole mass data, as they align with theoretical predictions from astrophysical processes. For instance, the power-law distribution is consistent with hierarchical mergers, while the log-normal distribution may arise from stochastic multiplicative processes during star formation and collapse. The likelihood principle serves as a unifying framework for parameter estimation and model comparison, enabling a rigorous assessment of these hypotheses.

1.6 Research Objectives

1. Develop and compare parametric models, specifically the power-law and log-normal distributions, for describing black hole mass data.

2. Employ maximum likelihood estimation (MLE) to estimate model parameters, improving these estimates using tools such as the Rao-Blackwell theorem.

3. Analyze the statistical properties of the estimators, including bias, efficiency, consistency, sufficiency and completeness to ensure robustness.

4. Construct confidence intervals using Fisher information and the delta method, providing uncertainty quantification for key parameters.

5. Test hypotheses about the underlying distribution using frameworks such as the likelihood ratio test and Neyman-Pearson lemma.

6. Extend the analysis to include advanced tools from mathematical statistics, such as ordered statistics and convergence properties, to uncover deeper insights.

7. Interpret the results in the context of astrophysical theories, linking statistical findings to physical processes such as supernova mechanisms and binary interactions.

2 Data Preparation and Preliminary Analysis

2.1 Data Sources

The primary data for this study is sourced from two main channels. First, we utilize gravitational wave (GW) detections from the LIGO/Virgo collaboration, specifically focusing on confirmed black hole merger events from observing runs O1 through O3. These observations provide precise measurements of the component masses involved in binary black hole mergers, though they are subject to specific detection biases related to the sensitivity of interferometers to more massive systems.

Second, we incorporate data from X-ray binary (XRB) surveys, particularly focusing on dynamically confirmed stellar-mass black holes in binary systems. These observations offer complementary information about the lower end of the black hole mass spectrum, as XRB detection methods are more sensitive to lower-mass black holes. The combination of these two datasets provides a more complete sampling across the entire mass range of interest.

2.2 Exploratory Data Analysis

Our preliminary analysis begins with a comprehensive histogram analysis to assess the general shape of the black hole mass distribution. We implement adaptive bin widths using the Freedman-Diaconis rule:

$$h = 2\frac{\mathrm{IQR}(x)}{n^{1/3}} \tag{28}$$

where IQR is the interquartile range and n is the sample size. This approach provides an optimal balance between resolution and statistical significance in our histogram representation.

We identify potential outliers using the modified Z-score method, defined as:

$$M_i = \frac{0.6745(x_i - \tilde{x})}{\mathrm{MAD}} \tag{29}$$

where $\tilde{x}$ is the median and MAD is the median absolute deviation. Points with $|M_i| > 3.5$ are flagged for careful examination, particularly focusing on whether they represent genuine massive black holes or potential measurement artifacts.

To account for detection biases, we implement a selection function $S(m)$ that models the probability of detecting a black hole of mass m in our surveys. This function takes the form:

$$S(m) = \begin{cases} \alpha_{\mathrm{GW}}(m/M_\odot)^{2.7} & \text{for GW events} \\ \alpha_{\mathrm{XRB}}(m/M_\odot)^{-0.5} & \text{for XRB observations} \end{cases} \tag{30}$$

where α_{GW} and α_{XRB} are normalization constants determined through detailed sensitivity analysis of each survey.

2.3 Data Transformation

To facilitate our statistical analysis, we apply two key transformations to our dataset. First, we implement a logarithmic transformation:

$$y_i = \ln(x_i/M_\odot) \tag{31}$$

where x_i represents the original mass measurements in solar masses. This transformation is particularly important for testing the log-normal distribution hypothesis and helps stabilize the variance across the mass range.

Subsequently, we standardize the transformed data:

$$z_i = \frac{y_i - \bar{y}}{s_y} \tag{32}$$

where $\bar{y}$ and s_y are the sample mean and standard deviation of the log-transformed data, respectively. This standardization simplifies the comparison of different statistical models and reduces numerical computation issues in our subsequent analyses.

To ensure the robustness of our results, we maintain both the original and transformed datasets throughout our analysis, applying appropriate inverse transformations when necessary to interpret results in physical units. All transformations are carefully documented to maintain transparency and reproducibility in our statistical pipeline.

3 Statistical Modeling Framework

3.1 Physical Motivation for Model Selection

The choice of our two primary models - the power-law and log-normal distributions - is deeply rooted in the physical processes governing black hole formation. The power-law distribution naturally emerges from scale-invariant processes in astrophysics, particularly in scenarios involving self-similar collapse mechanisms. This behavior is commonly observed in various astrophysical phenomena, from the initial mass function of stars to the mass distribution of molecular clouds.

The log-normal distribution, on the other hand, arises from the multiplicative central limit theorem, which becomes relevant when the final mass of a black hole is determined by the product of multiple independent random factors. These factors include:

- Initial stellar mass
- Mass loss through stellar winds
- Mass transfer in binary systems
- Fallback fraction during supernova

3.2 Power-Law Model Specification

The power-law model is characterized by its probability density function (PDF):

$$f(x; \alpha, x_{\min}) = \frac{\alpha x_{\min}^{\alpha}}{x^{\alpha+1}}, \quad x \geq x_{\min} \tag{33}$$

where:

- α is the power-law index (scaling parameter)
- $x_{\min}$ is the lower mass cutoff
- x represents the black hole mass in solar masses

The cumulative distribution function (CDF) can be derived through integration:

$$F(x; \alpha, x_{\min}) = 1 - \left(\frac{x_{\min}}{x}\right)^{\alpha}, \quad x \geq x_{\min} \tag{34}$$

The expected value and variance of this distribution are:

$$\mathbb{E}[X] = \frac{\alpha x_{\min}}{\alpha - 1}, \quad \alpha > 1 \tag{35}$$

$$\mathrm{Var}[X] = \frac{\alpha x_{\min}^2}{(\alpha - 1)^2(\alpha - 2)}, \quad \alpha > 2 \tag{36}$$

Note that these moments only exist for certain values of α, reflecting the heavy-tailed nature of power-law distributions.

3.3 Log-Normal Model Specification

The log-normal distribution is specified by its PDF:

$$f(x; \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right), \quad x > 0 \tag{37}$$

where:

- μ is the location parameter (mean of $\ln x$)
- σ is the scale parameter (standard deviation of $\ln x$)

The corresponding CDF is:

$$F(x; \mu, \sigma) = \frac{1}{2} + \frac{1}{2}\mathrm{erf}\left(\frac{\ln x - \mu}{\sigma\sqrt{2}}\right) \tag{38}$$

where $\mathrm{erf}(z)$ is the error function defined as:

$$\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt \tag{39}$$

The expected value and variance are:

$$\mathbb{E}[X] = \exp(\mu + \frac{\sigma^2}{2}) \tag{40}$$

$$\mathrm{Var}[X] = [\exp(\sigma^2) - 1] \exp(2\mu + \sigma^2) \tag{41}$$

4 Point Estimation Theory : Finding the MVUE

4.1 Maximum Likelihood Estimation (MLE) for Power-Law Parameters

To estimate the power-law parameter α, we employ a two steps method:

- The **First-Order Condition (FOC)** (also called *stationarity* or the *necessary condition*) requires setting the first derivative of the log-likelihood function to zero to find a critical point.
- The **Second-Order Condition (SOC)** (also called *maximality* or the *sufficient condition*) ensures that this critical point corresponds to a maximum by verifying that the second derivative is negative.

The probability density function (PDF) of a power-law distributed variable is:

$$f(x;\alpha) = \frac{\alpha x_{\min}^{\alpha}}{x^{\alpha+1}}, \quad x \geq x_{\min}, \quad \alpha > 0. \tag{42}$$

Given a sample $X_1, X_2, \ldots, X_n$ of i.i.d. observations, the likelihood function is:

$$\mathrm{Lik}(\alpha) = L(\alpha) = \prod_{i=1}^{n} \frac{\alpha x_{\min}^{\alpha}}{x_i^{\alpha+1}}. \tag{43}$$

Expanding the product:

$$L(\alpha) = \alpha^n x_{\min}^{n\alpha} \prod_{i=1}^{n} x_i^{-(\alpha+1)}. \tag{44}$$

Taking the natural logarithm of the likelihood function:

$$\ln L(\alpha) = n \ln \alpha + n\alpha \ln x_{\min} - (\alpha + 1) \sum_{i=1}^{n} \ln x_i. \tag{45}$$

First-Order Condition (FOC) – Stationarity: We differentiate the log-likelihood function with respect to α:

$$\frac{\partial \ln L(\alpha)}{\partial \alpha} = \frac{n}{\alpha} + n \ln x_{\min} - \sum_{i=1}^{n} \ln x_i. \tag{46}$$

Setting this equation to zero:

$$\frac{n}{\alpha} + n \ln x_{\min} - \sum_{i=1}^{n} \ln x_i = 0. \tag{47}$$

Solving for α:

$$\alpha = \frac{n}{\sum_{i=1}^{n} \ln x_i - n \ln x_{\min}} = \frac{n}{\sum_{i=1}^{n} (\ln x_i - \ln x_{\min})} = \frac{n}{\sum_{i=1}^{n} \ln\left(\frac{x_i}{x_{\min}}\right)} \tag{48}$$

Second-Order Condition (SOC) – Maximality: We check that the second derivative is negative to confirm a maximum:

$$\frac{\partial^2 \ln L(\alpha)}{\partial \alpha^2} = -\frac{n}{\alpha^2}. \tag{49}$$

Since this expression is always negative for $\alpha > 0$, the function is concave at the critical point, confirming that it is indeed a maximum.

MLE Expression with Random Variables: Finally, rewriting the estimator using the random variables X_i instead of observed values:

$$\hat{\alpha}_{\text{ML}} = \left[\frac{1}{n} \sum_{i=1}^{n} \ln\left(\frac{X_i}{X_{\min}}\right) \right]^{-1}. \tag{50}$$

4.2 Second-Order Method of Moments (MOM) for Log-Normal Parameters

The method of moments (MOM) estimates parameters by equating sample moments to theoretical moments. The second-order MOM refines the estimation by using both the first and second moments of the data, ensuring greater accuracy and robustness. We use the second-order MOM here because the log-normal distribution has two parameters, μ and σ, which require two equations to solve.

The Log-Normal distribution is parameterized by μ and σ^2, where if $X \sim \text{Log-Normal}(\mu, \sigma^2)$, then its logarithm follows a normal distribution:

$$\ln X \sim \mathcal{N}(\mu, \sigma^2). \tag{51}$$

Step 1: Defining the Method of Moments Equations

The first and second theoretical moments of the Log-Normal distribution are:

$$\mathbb{E}[X] = e^{\mu+\frac{\sigma^2}{2}}, \tag{52}$$

$$\mathbb{E}[X^2] = e^{2\mu+2\sigma^2}. \tag{53}$$

Using the variance formula:

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2. \tag{54}$$

To estimate μ and σ^2, we replace the expectations with their empirical sample counterparts. The first and second sample moments are:

$$m_1 = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad m_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2. \tag{55}$$

Thus, the system of equations for MOM estimation becomes:

$$\begin{cases} e^{\mu+\frac{\sigma^2}{2}} = \frac{1}{n}\sum_{i=1}^{n} X_i, \\ e^{2\mu+2\sigma^2} = \frac{1}{n}\sum_{i=1}^{n} X_i^2. \end{cases} \tag{56}$$

Step 2: Expressing σ^2 in Terms of Sample Moments

Rewriting the second equation in terms of $e^{2\mu}$:

$$e^{2\mu} = \left(e^{\mu+\frac{\sigma^2}{2}}\right)^2 = \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2. \tag{57}$$

Substituting this into the second equation:

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 = \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2 e^{2\sigma^2}. \tag{58}$$

Taking the natural logarithm:

$$\ln\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right) = \ln\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2 + 2\sigma^2. \tag{59}$$

Rearranging:

$$\hat{\sigma}^2_{\text{MOM}} = \frac{1}{2}\left[\ln\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right) - 2\ln\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)\right]. \tag{60}$$

Step 3: Solving for μ

From the first equation:

$$e^{\mu+\frac{\hat{\sigma}^2_{\text{MOM}}}{2}} = \frac{1}{n}\sum_{i=1}^{n} X_i. \tag{61}$$

Taking the natural logarithm:

$$\mu + \frac{\hat{\sigma}^2_{\text{MOM}}}{2} = \ln\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right). \tag{62}$$

Solving for μ:

$$\hat{\mu}_{\text{MOM}} = \ln\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) - \frac{\hat{\sigma}^2_{\text{MOM}}}{2}. \tag{63}$$

Final MOM Estimators

Thus, the Method of Moments estimators for the Log-Normal distribution parameters are:

$$\hat{\sigma}^2_{\text{MOM}} = \frac{1}{2}\left[\ln\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right) - 2\ln\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)\right], \tag{64}$$

$$\hat{\mu}_{\text{MOM}} = \ln\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) - \frac{\hat{\sigma}^2_{\text{MOM}}}{2}. \tag{65}$$

These estimators provide a practical approach to estimating the parameters of a Log-Normal distribution using only the first two sample moments.

4.3 Exponential Family of Distributions

A probability distribution belongs to the *exponential family* if its probability density function (pdf) or probability mass function (pmf) can be written in the form:

$$f(x \mid \theta) = h(x)\exp\left(\eta(\theta)T(x) - A(\theta)\right), \tag{66}$$

where:

- $T(x)$ is the *sufficient statistic*,

- $\eta(\theta)$ is the *natural parameter*,
- $A(\theta)$ is the *log-partition function*, ensuring normalization,
- $h(x)$ is a function independent of θ.

Distributions in the exponential family have useful properties, such as:

- The existence of sufficient statistics,
- Simplified maximum likelihood estimation (MLE),
- Convenient mathematical properties for Bayesian inference.

4.3.1 Log-Normal Distribution in the Exponential Family

The log-normal distribution is defined as:

$$f(x \mid \mu, \sigma^2) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right), \quad x > 0. \tag{67}$$

Rewriting this in exponential family form:

$$f(x \mid \mu, \sigma^2) = \exp\left(-\frac{(\ln x)^2}{2\sigma^2} + \frac{\mu \ln x}{\sigma^2} - \left(\frac{\mu^2}{2\sigma^2} + \ln(x\sqrt{2\pi\sigma^2})\right)\right). \tag{68}$$

Comparing with the general form, we identify:

- $T(x) = (\ln x, (\ln x)^2)$,
- $\eta(\mu, \sigma^2) = \left(\frac{\mu}{\sigma^2}, -\frac{1}{2\sigma^2}\right)$,
- $A(\mu, \sigma^2) = \frac{\mu^2}{2\sigma^2} + \ln(x\sqrt{2\pi\sigma^2})$.

Thus, the log-normal distribution belongs to the exponential family.

4.3.2 Power-Law Distribution in the Exponential Family

The power-law model we consider is:

$$f(x \mid \alpha) = \frac{\alpha x_{\min}^{\alpha}}{x^{\alpha+1}}, \quad x \geq x_{\min}. \tag{69}$$

Taking the logarithm:

$$\ln f(x \mid \alpha) = \ln \alpha + \alpha \ln x_{\min} - (\alpha + 1) \ln x. \tag{70}$$

Rewriting:

$$f(x \mid \alpha) = \exp\left(-(\alpha + 1) \ln x + \alpha \ln x_{\min} + \ln \alpha\right). \tag{71}$$

Comparing with the exponential family form, we identify:

- $T(x) = \ln x$,
- $\eta(\alpha) = -(\alpha + 1)$,
- $A(\alpha) = -\alpha \ln x_{\min} - \ln \alpha$.

Thus, the power-law distribution also belongs to the exponential family.

4.4 Sufficiency and Sufficient Statistics

In statistics, the concept of sufficiency is crucial as it provides a way to summarize data without losing relevant information about the parameter of interest. The goal of finding a sufficient statistic is to reduce the dimensionality of the data while preserving all the information needed to estimate the parameter efficiently.

When working with large datasets, using sufficient statistics can greatly simplify inference procedures, allowing for more efficient estimation and hypothesis testing. A sufficient statistic captures all the relevant information contained in the sample regarding the parameter, meaning that once the sufficient statistic is known, the sample provides no additional information about the parameter.

Definition: A statistic $T(X)$ is sufficient for a parameter θ if the conditional distribution of the sample given $T(X)$ does not depend on θ. In practical terms, sufficiency plays a key role in statistical inference, particularly in constructing estimators, simplifying likelihood functions, and developing Bayesian methods.

4.4.1 Sufficient Statistic for the Log-Normal Model

One of the most useful results for determining sufficiency is the *factorization theorem*.

(Factorization) Theorem: A statistic $T(X)$ is sufficient for a parameter θ if and only if the likelihood function can be factorized as:

$$L(\theta \mid x_1, \ldots, x_n) = g(T(x) \mid \theta) h(x_1, \ldots, x_n), \tag{72}$$

where $g(T(x) \mid \theta)$ is a function that depends only on the sufficient statistic and the parameter, while $h(x_1, \ldots, x_n)$ is a function that depends only on the sample but not on the parameter.

Applying the factorization theorem to the log-normal model, the likelihood function is given by:

$$L(\mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{x_i\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\ln x_i - \mu)^2}{2\sigma^2}\right). \tag{73}$$

Taking the logarithm:

$$\ln L(\mu, \sigma^2) = -\sum_{i=1}^{n} \ln x_i - \frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(\ln x_i - \mu)^2. \tag{74}$$

We rewrite the sum inside the exponential as:

$$\sum_{i=1}^{n}(\ln x_i - \mu)^2 = \sum_{i=1}^{n}(\ln x_i)^2 - 2\mu\sum_{i=1}^{n}\ln x_i + n\mu^2. \tag{75}$$

Factoring the likelihood function, we identify:

$$g(T(x) \mid \mu, \sigma^2) = \exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(\ln x_i)^2 - 2\mu\sum_{i=1}^{n}\ln x_i + n\mu^2\right)\right), \tag{76}$$

which depends on $T(X) = (\sum_{i=1}^{n}\ln x_i, \sum_{i=1}^{n}(\ln x_i)^2)$. Thus, this statistic is sufficient.

4.4.2 Sufficient Statistic for the Power-Law Model

An alternative method for proving sufficiency is direct computation of the conditional probability. Unlike the factorization theorem, which allows us to identify a sufficient statistic by factorizing the likelihood function, the conditional probability approach requires making an educated hypothesis about the form of the sufficient statistic and then verifying whether it satisfies the definition.

Given the power-law model with:

$$f(x \mid \alpha) = \frac{\alpha x_{\min}^{\alpha}}{x^{\alpha+1}}, \quad x \geq x_{\min}, \tag{77}$$

we hypothesize that $T(X) = \sum_{i=1}^{n} \ln X_i$ is sufficient based on the structure of the MLE estimator.

To check sufficiency, we compute the conditional probability:

$$P(X_1 = x_1, \ldots, X_n = x_n \mid T(X)) = \frac{P(X_1 = x_1, \ldots, X_n = x_n)}{P(T(X))}. \tag{78}$$

The joint probability is:

$$P(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} \frac{\alpha x_{\min}^{\alpha}}{x_i^{\alpha+1}}, \tag{79}$$

which simplifies to:

$$P(X_1 = x_1, \ldots, X_n = x_n) = \alpha^n x_{\min}^{n\alpha} \prod_{i=1}^{n} x_i^{-\alpha-1}. \tag{80}$$

The marginal probability $P(T(X))$ requires integrating over all possible values of x_i while keeping $T(X)$ fixed.

$$P(T(X)) = \int_{x_{\min}}^{\infty} \cdots \int_{x_{\min}}^{\infty} P(X_1 = x_1, \ldots, X_n = x_n) \, dx_1 \ldots dx_n, \tag{81}$$

while ensuring that the constraint $T(X) = \sum_{i=1}^{n} \ln x_i$ is satisfied.

Using the power-law joint probability density:

$$P(X_1 = x_1, \ldots, X_n = x_n) = \alpha^n x_{\min}^{n\alpha} \prod_{i=1}^{n} x_i^{-\alpha-1}, \tag{82}$$

we perform a change of variables:

$$U = \sum_{i=1}^{n} \ln X_i = T(X). \tag{83}$$

The Jacobian for this transformation can be derived from order statistics methods, but a direct approach is to recognize that the new density must be integrated over all partitions of $T(X)$, which introduces a factor:

$$P(T(X)) = \alpha^n x_{\min}^{n\alpha} \int_{x_{\min}}^{\infty} \cdots \int_{x_{\min}}^{\infty} e^{-\alpha T(X)} \prod_{i=1}^{n} x_i^{-1} dx_1 \dots dx_n. \tag{84}$$

Using the constraint $T(X) = \sum_{i=1}^{n} \ln x_i$, the integral is rewritten as an integration over a simplex:

$$P(T(X)) \propto \alpha^n x_{\min}^{n\alpha} e^{-\alpha T(X)}. \tag{85}$$

The symbol $\propto$ means that the left-hand side is proportional to the right-hand side, which can be rewritten as $P(T(X)) = C \cdot \alpha^n x_{\min}^{n\alpha} e^{-\alpha T(X)}$ for some constant C.

Thus, we obtain the marginal probability. Dividing the numerator by the denominator:

$$P(X_1 = x_1, \dots, X_n = x_n \mid T(X)) \propto e^{-\alpha T(X)} \prod_{i=1}^{n} x_i^{-1}. \tag{86}$$

Since this does not depend on α, we conclude that $T(X) = \sum_{i=1}^{n} \ln X_i$ is sufficient for α.

4.5 Improved Estimators Using Rao-Blackwellization Theorem

One main reason for identifying sufficient statistics in the previous section was to apply the Rao-Blackwellization theorem, which provides a method for improving estimators. The Rao-Blackwell theorem is a fundamental result in statistical estimation theory, formulated by Calyampudi Radhakrishna Rao and David Blackwell. It states that given an estimator and a sufficient statistic, we can construct a new estimator that is at least as good, and often strictly better, in terms of mean squared error or variance.

This improvement arises because conditioning on a sufficient statistic eliminates unnecessary variability, leading to a more efficient estimator. In the context of our models, we seek to improve the estimators obtained from MLE or MOM using sufficient statistics derived earlier.

Theorem: Rao-Blackwell - Given $\hat{\theta}$ an unbiased estimator of θ, and $T(X)$ a sufficient statistic for θ,

$$\tilde{\theta} = \mathbb{E}[\hat{\theta} \mid T(X)] \tag{87}$$

is an improved (lower MSE) estimator of θ. We call this operation the *Rao-Blackwellization*, and $\tilde{\theta}$ is called the *Rao-Blackwellized* estimator.

4.5.1 Rao-Blackwellization for the Log-Normal Model

We apply the Rao-Blackwell theorem to the Method of Moments (MOM) estimators of μ and σ^2. We recall the MOM estimators:

$$\hat{\sigma}^2_{\text{MOM}} = \frac{1}{2}\left[\ln\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right) - 2\ln\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)\right], \tag{88}$$

$$\hat{\mu}_{\text{MOM}} = \ln\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) - \frac{\hat{\sigma}^2_{\text{MOM}}}{2}. \tag{89}$$

We use the sufficient statistic:

$$T(X) = \left(\sum_{i=1}^{n} \ln X_i, \sum_{i=1}^{n} (\ln X_i)^2\right). \tag{90}$$

Symmetry Method for Rao-Blackwellization The key idea behind the symmetry method is to first express the conditional expectation of the estimator in terms of $T(X)$, then use properties of expectation to simplify. We begin by computing:

$$A = E[\hat{\mu}_{\text{MOM}} \mid T(X)]. \tag{91}$$

Multiplying by n and using the i.i.d. assumption:

$$nA = E\left[\sum_{i=1}^{n}\left(\ln X_i - \frac{\hat{\sigma}^2_{\text{MOM}}}{2}\right) \mid T(X)\right]. \tag{92}$$

Expanding the expectation inside the summation:

$$nA = E\left[\sum_{i=1}^{n} \ln X_i \mid T(X)\right] - \frac{1}{2}E\left[\sum_{i=1}^{n} \hat{\sigma}^2_{\text{MOM}} \mid T(X)\right]. \tag{93}$$

Since $T(X)$ includes $\sum_{i=1}^{n} \ln X_i$, we can simplify:

$$E\left[\sum_{i=1}^{n} \ln X_i \mid T(X)\right] = \sum_{i=1}^{n} \ln X_i. \tag{94}$$

Thus,

$$nA = \sum_{i=1}^{n} \ln X_i - \frac{1}{2}E\left[\sum_{i=1}^{n} \hat{\sigma}^2_{\text{MOM}} \mid T(X)\right]. \tag{95}$$

For $\hat{\sigma}^2_{\text{MOM}}$, applying expectation:

$$E[\hat{\sigma}^2_{\text{MOM}} \mid T(X)] = \frac{1}{2}\left[\ln\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right) - 2\ln\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)\right]. \tag{96}$$

Using the sufficiency property, this simplifies to:

$$E[\hat{\sigma}^2_{\text{MOM}} \mid T(X)] = \sigma^2. \tag{97}$$

Thus,

$$A = \frac{1}{n}\sum_{i=1}^{n} \ln X_i - \frac{\sigma^2}{2}. \tag{98}$$

This gives the Rao-Blackwellized estimator:

$$\hat{\mu}_{\text{RB}} = \frac{1}{n}\sum_{i=1}^{n} \ln X_i - \frac{\sigma^2}{2}. \tag{99}$$

Similarly, for $\hat{\sigma}^2_{\text{MOM}}$:

$$E[\hat{\sigma}^2_{\text{MOM}} \mid T(X)] = \sigma^2, \tag{100}$$

so the Rao-Blackwellized estimator is simply:

$$\hat{\sigma}^2_{\text{RB}} = \sigma^2. \tag{101}$$

Unexpected Convergence and Theoretical Foundations The Rao-Blackwellization process for the σ^2 estimator yields a surprising result:

$$\hat{\sigma}^2_{\text{RB}} = \sigma^2 \tag{102}$$

This outcome emerges from the intricate properties of the sufficient statistic $T(X)$. When conditioning on this statistic, we extract all parameter-relevant information from the sample, leading to a profound statistical phenomenon...

Definition: Mechanism of Convergence - The convergence occurs through the following critical steps:

- The conditional expectation $E[\hat{\sigma}^2_{\text{MOM}} \mid T(X)]$ evaluates exactly to σ^2
- Conditioning on the sufficient statistic eliminates the estimator's variability

- The Rao-Blackwellization process "collapses" the estimator to the true parameter value

Limitations of the Constant Estimator The Rao-Blackwellized σ^2 estimator suffers from critical drawbacks:

1. Loss of sample-dependent variability
2. Inability to generate meaningful confidence intervals
3. Elimination of the estimator's adaptive characteristics

The original Method of Moments estimator:

$$\hat{\sigma}^2_{\text{MOM}} = \frac{1}{2}\left[\ln\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right) - 2\ln\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)\right] \tag{103}$$

retains crucial statistical properties that the Rao-Blackwellized version loses.

Comparative Performance with μ Estimator In stark contrast, the Rao-Blackwellized μ estimator:

- Maintains meaningful variability
- Provides robust statistical inference
- Demonstrates the nuanced potential of Rao-Blackwellization

Conclusion: This analysis reveals the delicate nature of statistical estimation. While Rao-Blackwellization proves highly effective for the μ estimator, it demonstrates significant limitations when applied to the σ^2 estimator. We will thus retain the original Method of Moments estimator for σ^2.

The process underscores the critical importance of context-specific statistical methodologies and the need for nuanced approaches in parameter estimation. The fundamental lesson emerges: statistical techniques are not universal prescriptions but sophisticated tools requiring careful, context-aware application.

4.5.2 Rao-Blackwellization for the Power-Law Model

For the power-law model, we recall that the MLE for α is given by:

$$\hat{\alpha}_{MLE} = 1 + n\left(\sum_{i=1}^{n} \ln\frac{X_i}{x_{\min}}\right)^{-1}. \tag{104}$$

Since we previously found that $T(X) = \sum_{i=1}^{n} \ln X_i$ is a sufficient statistic, we consider applying the Rao-Blackwell theorem by computing:

$$\mathbb{E}[\hat{\alpha}_{MLE} \mid T(X)]. \tag{105}$$

However, we notice that $\hat{\alpha}_{MLE}$ is already a function of $T(X)$. That is, $\hat{\alpha}_{MLE}$ can be rewritten entirely in terms of $T(X)$:

$$\hat{\alpha}_{MLE} = f(T(X)), \tag{106}$$

for some function f. In such cases, the conditional expectation simplifies to:

$$\mathbb{E}[f(T(X)) \mid T(X)] = f(T(X)), \tag{107}$$

which implies that the Rao-Blackwellized estimator is identical to the MLE. This result demonstrates another important property of MLE: in many cases, it already depends on the sufficient statistic, making it an optimal estimator that cannot be further improved via Rao-Blackwellization. This reinforces the superiority of MLE in estimation theory.

4.6 Fisher's Theorem for Approximate Confidence Interval for Power-Law parameter

4.7 Delta Method for Approximate Confidence Interval for Log-Normal parameters

.

5 Analysis of the estimates

5.1 Precision analysis : biasedness

5.2 Efficiency analysis : MSE

5.3 Consistency analysis : convergence

5.4 Sufficiency and completeness

5.5 Cramer-Rao Lower Bound

5.6 Lehmann-Scheffe theorem and sufficiency theorem

.

6 Advanced Applications of Mathematical Statistics

6.1 Ordered Statistics

6.2 Jensen's Inequality

6.3 Convergence Analysis

.

7 Hypothesis Testing : Finding the UMPT

7.1 Set theory and framework generalization : parameter space and universe

7.2 Likelihood Ratio Test (LRT)

7.3 Neyman-Pearson Lemma

7.4 Generalized Likelihood Ratio Test (GLRT)

7.5 Wilks Theorem for Chi-Squared approximation

.

8 Model Comparisons and Implications

8.1 Goodness-of-Fit Evaluation

8.2 Power Analysis

.

9 Results and Discussion

9.1 Parameter Estimates

9.2 Model Insights

9.3 Broader Implications