Enhanced Maximum Lq-Likelihood Estimation for the Tail Index of Heavy Tailed Distributions: A New Approach for Small Samples

Silver Jubilee International Pure Mathematics Conference 2025

Nesrine IDIOU*

Abstract

In extreme value theory (EVT), estimating the tail index of heavy-tailed distributions is crucial for understanding rare and extreme events. Traditional estimators such as the Hill and Maximum Likelihood Estimators (MLE) perform well with large samples but struggle with small sample sizes due to increased bias and variance. In this paper, we introduce a novel estimation technique the Maximum Lq-Likelihood Estimator (MLqE), which incorporates a distortion parameter q, making it more robust to extreme observations and more accurate in small-sample scenarios. We demonstrate that the MLqE is consistent and asymptotically normal, outperforming the classical MLE in terms of mean squared error in moderate and small sample sizes. Moreover, we present simulation results that highlight the superior performance of the MLqE, particularly when comparing it to the MLE in tail index estimation. This method not only offers a significant improvement in the accuracy of heavy-tailed distribution parameter estimation but also provides a versatile tool for various real-world applications, including finance, hydrology, and risk management.

Keywords: Heavy-tailed Distributions; Extreme Value Index; Maximum Lq-Likelihood estimation; Excesses over high thresholds.

1 Introduction

In statistical modeling, Maximum Likelihood Estimation (MLE) is a widely used method for estimating the parameters of a statistical model. Given a set of observations, MLE seeks the parameter values that maximize the likelihood function, which represents the probability of observing the given data under the assumed model. However, classical MLE can be sensitive to noise and outliers, which may adversely affect the quality of the parameter estimates.

To address these limitations, the Maximum Likelihood q-Estimator (MLqE) extends the traditional MLE framework by introducing a tuning parameter q. This parameter

*Process Engineering Department, Salah Boubnider University of Constantine 3, Constantine , Algeria. Email: nesrine.idiou@univ-constantine3.dz

modifies the likelihood function to adjust the influence of individual data points, thereby enhancing robustness against atypical observations and improving the estimator's stability in practical scenarios.

This paper focuses on the application of MLE and MLqE to Gaussian Mixture Models (GMMs), which are probabilistic models that assume the observed data are generated from a mixture of several Gaussian distributions. Each Gaussian component is characterized by its own mean vector and covariance matrix, allowing the model to capture complex data distributions and heterogeneity within the dataset.

By comparing the classical MLE approach with the robust MLqE method within the GMM framework, we investigate their respective performances in terms of parameter estimation accuracy, robustness to outliers, and convergence properties. Additionally, we discuss how the Expectation-Maximization (EM) algorithm can be adapted to optimize both estimators effectively.

Our results demonstrate that incorporating the parameter q provides a flexible way to balance sensitivity and robustness, making MLqE a valuable tool for statistical inference in noisy or contaminated data environments.

2 Main Results

Given a set of data $X = \{x_1, x_2, \ldots, x_n\}$, where each x_i is a vector of features, the goal of MLE is to find the parameters

$$\theta = \{\pi_k, \mu_k, \Sigma_k\}$$

that maximize the likelihood function.

In a Gaussian Mixture Model, the likelihood function is given by:

$$L(\theta) = \prod_{i=1}^{n} \sum_{k=1}^{K} \pi_k \mathcal{N}(x_i \mid \mu_k, \Sigma_k)$$

where:

- K is the number of components in the mixture,
- π_k is the mixing weight of the k-th Gaussian component, satisfying

  $$\sum_{k=1}^{K} \pi_k = 1,$$

- μ_k and Σ_k are the mean vector and covariance matrix of the k-th Gaussian component, respectively
- $\mathcal{N}(x_i \mid \mu_k, \Sigma_k)$ is the probability density function of the k-th Gaussian component, defined as:

  $$\mathcal{N}(x_i \mid \mu_k, \Sigma_k) = \frac{1}{(2\pi)^{d/2} |\Sigma_k|^{1/2}} \exp\left(-\frac{1}{2} (x_i - \mu_k)^\top \Sigma_k^{-1} (x_i - \mu_k) \right),$$

  where d is the dimension of the feature vector x_i.

The parameters $\theta = \{\pi_k, \mu_k, \Sigma_k\}$ are estimated by maximizing the likelihood function using the Expectation-Maximization (EM) algorithm. The EM algorithm alternates between two steps:

E-step: In this step, we compute the posterior probability (or responsibility) of each data point x_i belonging to each Gaussian component k:

$$\gamma_{ik} = P(z_i = k \mid x_i, \theta) = \frac{\pi_k \mathcal{N}(x_i \mid \mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_i \mid \mu_j, \Sigma_j)}.$$

M-step: In this step, we update the parameters θ based on the responsibilities computed in the E-step:

- Update the mixing weights:
$$\pi_k = \frac{1}{n} \sum_{i=1}^{n} \gamma_{ik}.$$
- Update the means:
$$\mu_k = \frac{\sum_{i=1}^{n} \gamma_{ik} x_i}{\sum_{i=1}^{n} \gamma_{ik}}.$$
- Update the covariance matrices:
$$\Sigma_k = \frac{\sum_{i=1}^{n} \gamma_{ik} (x_i - \mu_k)(x_i - \mu_k)^\top}{\sum_{i=1}^{n} \gamma_{ik}}.$$

These steps are repeated iteratively until convergence, i.e., when the change in the log-likelihood function is below a certain threshold.

The Maximum Likelihood q-Estimator (MLqE) introduces a parameter q that modifies the likelihood function to enhance robustness against noise and outliers. This approach generalizes the classical Maximum Likelihood Estimation (MLE) by incorporating an additional degree of freedom via q, which adjusts the influence of individual data points. The q-modified likelihood function is defined as:

$$L_q(\theta) = \prod_{i=1}^{n} \left(\sum_{k=1}^{K} \pi_k \mathcal{N}(x_i \mid \mu_k, \sigma_k) \right)^q,$$

where:

- x_i are the observed data points,
- $\mathcal{N}(x_i \mid \mu_k, \sigma_k)$ represents the Gaussian density function with mean μ_k and standard deviation σ_k,
- π_k denotes the mixing coefficients,
- q is a tuning parameter that alters the sensitivity of the likelihood function.

The introduction of q influences the estimator's behavior as follows:

- When $q = 1$, MLqE reduces to the standard MLE formulation.
- When $q > 1$, the method downweights highly probable data points, making the estimation more sensitive to less frequent observations. This property is useful for handling outliers and improving robustness.

- When $q < 1$, it emphasizes dominant data points, which can be beneficial in scenarios where rare observations should have less impact.

To estimate the parameters θ, the optimization problem consists of maximizing the log-transformed q-likelihood function:

$$\log L_q(\theta) = q \sum_{i=1}^{n} \log \left(\sum_{k=1}^{K} \pi_k \mathcal{N}(x_i \mid \mu_k, \sigma_k) \right).$$

Optimization techniques such as gradient-based methods or the Expectation-Maximization (EM) algorithm can be employed. Depending on q, modifications to the standard EM algorithm may be required to ensure convergence and numerical stability.

3 Conclusion

In this work, we have explored the parameter estimation methods for Gaussian mixture models, focusing on the classical Maximum Likelihood Estimation (MLE) and its robust extension, the Maximum Likelihood q-Estimator (MLqE). While MLE provides an effective framework for estimating model parameters under ideal conditions, MLqE introduces a tuning parameter q that enhances robustness against noise and outliers, making it well-suited for real-world data that often deviate from ideal assumptions.

The Expectation-Maximization (EM) algorithm plays a central role in efficiently optimizing these estimators. Modifications to the EM algorithm are sometimes necessary when employing MLqE to ensure stable convergence. Overall, the choice between MLE and MLqE depends on the specific characteristics of the data and the desired balance between sensitivity and robustness.

Future work may include extending these approaches to more complex models or exploring alternative robust estimation frameworks to further improve performance in challenging data environments.

References

[1] Altun, Y. & Smola, A. (2006). Unifying divergence minimization and statistical inference via convex duality. In Learning Theory. Lecture Notes in Computer Science 4005 139–153. Springer, Berlin. DOI: 10.1007/1177642013

[2] Balkema, A. A., & De Haan, L. (1974). Residual lifetime at great age. The Annals of Probability, 792-804. DOI 10.1214/aop/1176996548

[3] Ferrari, D. & Yang, Y. (2010). Maximum Lq-likelihood estimation. The Annals of Statistics, 38, 753–783. DOI: 10.1214/09-AOS687

[4] Haan, L de. & Ferreira, A. (2006). Extreme Value Theory: An Introduction. Springer. DOI: 10.1007/0-387-34471-3

[5] Kouider, M. R. & Idiou, N. & Benatia, F. (2023). Adaptive Estimators of the General Pareto Distribution Parameters under Random Censorship and Application. Journal of Science and Arts, No 2, pp. 395-412. Link

[6] Reiss, R. D., & Thomas, M. (1997). Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology, and Other Fields. Basel: Birkhüser. DOI: 10.1007/978-3-7643-7399-3

[7] Smith, R. L. (1987). Estimating tails of probability distributions. The Annals of Statistics, 1174-1207. DOI: 10.1214/aos/1176350499

[8] Hyndman, R. J., & Fan, Y. (1996). Sample Quantiles in Statistical Packages. The American Statistician, 50(4), 361–365. DOI: 10.2307/2684934