A Review of Estimating the Shape Parameter of Generalized Gaussian Distribution

Shaoquan YU, Anyi ZHANG, Hongwei LI*

School of Mathematics and Physics, China University of Geosciences, Wuhan 430074, China

Abstract

In this paper, some existing methods for estimating the shape parameter $\alpha$ of generalized Gaussian distribution (GGD) and their main features are summarized and compared with each other theoretically and practically. Some problems for the existing methods are put forward and some suggestions are also given for the further study on this problem.

Keywords: Generalized Gaussian Distribution; Maximum Likelihood Estimation; Moment-based Estimation; Entropy Matching; Globally Convergent Method

1 Introduction

The key issue of many signal processing problems is to build a statistical model for the observed data. In generally, some parametric model of probability density function (PDF) for the observed data is built, and then parameters in the model are estimated. The most prominent and widely used family of parametric distributions is that of generalized Gaussian distribution (GGD) which is also called as an exponential power distribution. Box[1] first discussed its characters as well as the Bayes inference. At present, GGD has been widely adopted in signal processing field. For example, Gharavi and Tabatabai[2] adopted it to approximate the marginal distribution in image processing; Joshi and Fischer[3] fitted the AC coefficients of DCT using GGD with zero mean, and drew a conclusion that the shape parameters in AC coefficients mostly lied in 1.0~2.0. In speech signal processing, Prasad[4] showed that the distributions of the speech signals were light tailed GGD, which is based on to establish the VAD algorithm of GGD speech signal detection by Chang[5]. In addition, GGD has also been adopted to data modeling in the fields such as digital watermarking[6], blind signal separation(BSS)[7, 8], synthetic aperture radar (SAR)[9], ultrasonic cardio-gram images[10], face recognition[11], power system load demand[12], image subband signals[13], and independent component analysis (ICA) [14-16], etc.

The PDF of GGD is given by [1-5, 7, 8, 17]

*Project supported by the National Natural Science Foundation of China (No. 61071188, 61102103, and 11126274).

*Corresponding author.

Email address: hwli@cug.edu.cn (Hongwei LI).
\[ f(x; \mu, \alpha, \beta) = \left[ \frac{\alpha}{2\beta \Gamma(1/\alpha)} \right] \exp \left\{ - \left[ \frac{|x - \mu|}{\beta} \right]^\alpha \right\} \] (1)

where \( \Gamma(\bullet) \) defines the gamma function given by \( \Gamma(z) = \int_0^{+\infty} e^{-t} t^{z-1} dt \) \((z > 0)\), \( \mu, \alpha > 0, \beta \) denote the mean, shape and scale parameters respectively, and \( \beta \) is given by \( \beta = \sigma \sqrt{\Gamma(1/\alpha)/\Gamma(3/\alpha)} \) (\( \sigma \) is the standard deviation of the GGD).

Fig.1 illustrates the PDFs of GGD for various \( \alpha \) with \( \mu = 0 \) and \( \sigma = 1 \). The shape parameter \( \alpha \) denotes the rate of decay: the smaller \( \alpha \), the more peaked for the PDF, and the larger \( \alpha \), the flatter for the PDF, so it is also called as the decay rate. Two special cases are the Laplacian distribution for \( \alpha = 1 \) and the Gaussian distribution for \( \alpha = 2 \). For the limiting cases of \( \alpha \), the GGD approaches to an impulse function and the uniform distribution as \( \alpha \to 0 \) and \( \alpha \to \infty \) respectively.

It is important to estimate the parameters in the model so that we can model the signal efficiently with GGD and put it into practice. There are three parameters needed to be estimate in Eq. (1) which includes the most important parameter \( \alpha \) for the mean and variance can be obtained by the sample average and sample variance respectively, and the scale parameter \( \beta \) is determined by the shape one. Many studies focus on obtaining more accurate estimation of \( \alpha \), and the methods proposed for this purpose can be summarized as four categories as follows:


2. Moment-based method (MM). The original MM about parameters estimation of GGD model was discussed in detail in [17]. Further researches are developed to obtain \( \alpha \) according to the structuring of its monotone function based on the absolute moments of \( \alpha \), such as Mallat’s generalized Gaussian rate method (MRM)[19-22], Kurtosis generalized Gaussian rate method (KRM), etc. To obtain more accurate estimator of \( \alpha \), Krupiński[23] further adopted various generalized Gaussian rate functions on different intervals.


(4) Global convergence method (GCM). This method was first proposed by Song[26]. It is essentially a moment method, but it is different from the ordinary moment method that the orders in GCM are variable about \( \alpha \) while it is a constant as one in the ordinary method. The solution can be obtained by iteration which converges more easily by GCM than by MLE.

The rest of the paper is organized as follows. The methods listed above are reviewed in Section 2. In Section 3, some numerical experiments are given to compare the above methods theoretically and practically. Finally, we conclude this paper in Section 4.

2 Estimation Methods for the Shape Parameter in GGD

Given the observed data \( \{x_1, x_2, \ldots, x_n\} \) of a signal \( x \) which is GGD with \( \mu = 0 \) and \( \sigma = 1 \), we review the four methods for the shape parameter estimation of the GGD model as follows.

2.1 Maximum likelihood estimation (MLE)

The maximum likelihood estimator of \( \alpha \) is given by the logarithm likelihood function[3, 17, 18] as follows:

\[
L = \log \prod_{i=1}^{n} f(x_i; \alpha, \beta) = n \log \alpha - n \log[2\beta \Gamma(1/\alpha)] - \sum_{i=1}^{n} \left(\frac{|x_i|}{\beta}\right)^{\alpha} \tag{2}
\]

The MLE solution of the shape parameter can be obtained by solving the following equation:

\[
1 + \frac{\psi(1/\hat{\alpha})}{\hat{\alpha}} - \frac{\sum_{i=1}^{n} |x_i|^\hat{\alpha} \log |x_i|}{\sum_{i=1}^{n} |x_i|^\hat{\alpha}} + \frac{\log(\frac{n}{n} \sum_{i=1}^{n} |x_i|^{\hat{\alpha}})}{\hat{\alpha}} = 0
\]

The equation can be solving by the iteration method of Newton-Raphson with the formula:

\[
\alpha_{k+1} = \alpha_k - \frac{g(\alpha_k)}{g'(\alpha_k)} \tag{3}
\]

where

\[
g(\alpha) = 1 + \frac{\psi(1/\alpha)}{\alpha} - \frac{\sum_{i=1}^{n} |x_i|^\alpha \log |x_i|}{\sum_{i=1}^{n} |x_i|^\alpha} + \frac{\log(\frac{n}{n} \sum_{i=1}^{n} |x_i|^\alpha)}{\alpha} \tag{4}
\]

\[
g'(\alpha) = -\frac{\psi(1/\alpha)}{\alpha^2} - \frac{\psi'(1/\alpha)}{\alpha^3} + \frac{1}{\alpha^2} - \frac{\sum_{i=1}^{n} |x_i|^\alpha (\log |x_i|)^2}{\sum_{i=1}^{n} |x_i|^\alpha} + \frac{\sum_{i=1}^{n} |x_i|^\alpha \log |x_i|}{\sum_{i=1}^{n} |x_i|^\alpha} - \frac{\log(\frac{n}{n} \sum_{i=1}^{n} |x_i|^\alpha)}{\alpha^2} \tag{5}
\]
where $\psi(z) = \Gamma'(z)/\Gamma(z) = -\gamma + \int_0^1 (1 - x^{z-1})(1-x)^{-1}dx$ denotes the digamma function, and $
abla = 0.577$ is Euler's constant. $\psi'(z)$ denotes the trigamma function. The moment estimation was suggested as the initial estimator [32-34]. It showed in [17] that there exists unique root for the equation.

### 2.2 Moment-based method (MM)

The $k$-order absolute central moments of $x$ for GGD and the random variable $T$ are defined respectively as follows[20]:

$$m_k = E(|x|^k) = \beta^{k} \frac{\Gamma(k + 1/\alpha)}{\Gamma(1/\alpha)} \quad (k = 1, 2 \ldots)$$

$$T = \left[\frac{m_k}{m_p}\frac{[m_m]^n}{[m_p]^q[m_r]^s}\right]$$

where $k, m, p$ and $r$ are positive integers, $l, n, q$ and $s$ are nonnegative integers. As $kl + mn = pq + rs$, there is only one parameter $\alpha$ remained in $T$, so $T$ can be calculated as:

$$T = \left[\frac{m_k}{m_p}\frac{[m_m]^n}{[m_p]^q[m_r]^s}\right] = \frac{\Gamma(k + 1/\alpha)}{\Gamma(1/\alpha)\Gamma(3/\alpha)} \Gamma^{q+s-l-n}(\frac{1}{\alpha}) \triangleq F(\alpha)$$

When $k = k_1, l = 1, p = k_2, q = k_1/k_2, n = s = 0$, we have[23] :

$$\frac{m_{k_1}}{(m_{k_2})_{k_1/k_2}} = \frac{\Gamma((k_1 + 1)/\alpha)}{\Gamma((k_2 + 1)/\alpha)_{k_1/k_2}} \frac{\Gamma(1/\alpha)^{1-(k_1/k_2)}}{\Gamma(3/\alpha)}$$

so different monotonic function of parameter $\alpha$ can be constructed with various $k_1, k_2$ as follows:

#### 2.2.1 Mallat’s generalized Gaussian ratio method (MRM)

This method was first put forward by Mallat[19], and was discussed in deep by Sharifi[21]. Mallat generalized Gaussian ratio (Mallat’s ratio) is given by

$$M(\alpha) = \frac{m_1^2}{m_2} = \frac{\Gamma^2(2/\alpha)}{\Gamma(1/\alpha)\Gamma(3/\alpha)}$$

or the equivalent form which is usually employed in practical application as:

$$M(\alpha) = \frac{m_1}{\sqrt{m_2}} = \sqrt{\frac{\Gamma^2(2/\alpha)}{\Gamma(1/\alpha)\Gamma(3/\alpha)}}$$

It can be seen that Eq. (11) corresponds to the situation that $k_1 = 1, k_2 = 2$ in Eq. (9). We can get the estimator of $\alpha$ by solving the inverse function as below:

$$\hat{\alpha} \triangleq M^{-1}\left(\frac{\hat{m}_1}{\sqrt{\hat{m}_2}}\right)$$
Practically, the estimator $\hat{\alpha}$ is usually obtained through curve-fitting of the inverse function $M^{-1}(\bullet)$ [20, 23] for the display expression of $M^{-1}$ is difficult to work out. The matching method was used to find the estimator $\hat{\alpha}$ in [25]. Firstly, a look-up table of $M(\alpha)$ was established, then the generalized Gaussian ratio was calculated. Finally, the estimator $\hat{\alpha}$ was obtained by finding the corresponding value $M(\alpha)$ of $\alpha$ in the table which was most close to $M(\hat{\alpha})$. An estimator $\hat{\alpha}$ with $E(|x|)/E(x^2)$ was suggested in [31].

2.2.2 Kurtosis generalized Gaussian ratio method (KRM)

It can be seen from Eq. (1) that the shape parameter $\alpha < 2, \alpha = 2, \alpha > 2$ correspond to the three situations of the distribution of sup-Gaussian, Gaussian, sub-Gaussian respectively, so the values of $\alpha$ can be expressed by the kurtosis of PDF in Eq. (1). Similar with the definition of the Mallat’s ratio in subsection 2.2.1, the kurtosis ratio can be obtained by:

$$K(\alpha) = \frac{m_2}{\sqrt{m_4}} = \sqrt{\frac{\Gamma^2(3/\alpha)}{\Gamma(1/\alpha)\Gamma(5/\alpha)}}$$ (13)

and

$$\hat{\alpha} \Delta = K^{-1}(\frac{\hat{m}_2}{\sqrt{\hat{m}_4}})$$ (14)

To obtain the more precise estimator of $\alpha$, Krupiński[23] made an improvement by adopting different ratios on diverse intervals. The simulation results showed that the improved estimator had smaller mean square error (MSE) and lower computational complexity than the one of MRM. But the estimators were still not satisfactory in the case of short data due to the inherent shortage of MM which is based on the empirical distribution approaching the real distribution when the sample is large.

2.3 Comentropy-based method (CM)

2.3.1 Generalized entropy matching method

The differential entropy of a signal $x$ is defined by

$$H(x) = -\int_{-\infty}^{\infty} p(x) \log p(x) dx$$ (15)

where $p(x)$ is the probability density function of $x$. When $x$ is a GGD signal, its differential entropy is given by [18, 24, 25]

$$H(x) = -\log A \int_{-\infty}^{\infty} A e^{-(|x|/\beta)^\alpha} dx + \int_{-\infty}^{\infty} A\left(\frac{|x|}{\beta}\right)^\alpha e^{-(|x|/\beta)^\alpha} dx = \log\left(\frac{2}{\alpha}\sqrt{\frac{\Gamma^3(1/\alpha)}{\Gamma(3/\alpha)}}\right) + \frac{1}{\alpha}$$ (16)

where
\[ A = \frac{\alpha}{2\beta \Gamma(1/\alpha)} \]

Suppose \( H(X) \) is the entropy value of the optimum entropy constrained by uniform threshold quantizer (UTQ), then \cite{27}

\[ H(X) = H(x) - \log \Delta \]

where \( \Delta \) is the step size of UTQ, then we can get

\[ H(X) - \frac{1}{k} \log \frac{m_k}{\Delta} = \log \left[ \frac{2\Gamma(1/\alpha)^{(k+1)/k}}{\alpha \Gamma((k+1)/\alpha)^{1/k}} \right] + \frac{1}{\alpha} \triangleq f_k(\alpha) \]  \hspace{1cm} (17)

where \( f_k(\alpha) \) is called as generalized entropy matching estimator, and it can reach the theoretical maximum. According to Eq. (17),

\[ \hat{\alpha} \Delta = f_k^{-1}(H(\hat{X}) - \frac{1}{k} \log \hat{m}_k) \]  \hspace{1cm} (18)

which is the same with the entropy matching method\cite{24} when \( k = 2 \) in Eq. (18). It should be point out that the larger \( k \), larger sample is required for the estimation.

### 2.3.2 Negentropy matching (NM) method \cite{25}

Suppose that \( x_{\text{Gauss}} \) is a Gaussian variable with unit variance just as \( x \). NM is define as follows:

\[ D = H(x_{\text{Gauss}}) - H(x) = \log \left( \frac{\alpha}{2} \sqrt{\frac{3}{\Gamma(3/2)\Gamma(3/\alpha)}} \right) + \frac{1}{2} - \frac{1}{\alpha} \triangleq f(\alpha) \]  \hspace{1cm} (19)

The traditional NM method is based on the high order moments and the estimator \( \hat{D} \) based on the observed data of \( x \) is given by:

\[ \hat{D} = \frac{1}{12} \left[ E(x^3) \right]^2 + \frac{1}{48} \left[ kurt(x) \right]^2 \]  \hspace{1cm} (20)

Since the kurtosis term \( kurt(x) \) of \( x \) in Eq. (20) is very sensitive to the sample, a more stable estimation of \( \hat{D} \) was proposed in \cite{28} as follows:

\[ \hat{D} = k_1 \left[ E(x e^{-x^2/\tau}) \right]^2 + k_2 \left[ E(e^{-x^2/\tau}) - \sqrt{1/2} \right]^2 \]  \hspace{1cm} (21)

where \( k_1 = 7.412, k_2 = 33.67 \).

### 2.4 Global convergence method (GCM)

Suppose \( x \) is generalized Gaussian distributed, so

\[ Z(\alpha) \triangleq \frac{E|x|^{2\alpha}}{(E|x|^\alpha)^2} - (1 + \alpha) = 0 \]  \hspace{1cm} (22)
where \( Z(\alpha) \) is a convex function and the Eq. (22) has unique root in \((0, \infty)\) [26]. It pointed out in [26] that one consistent estimator of \( Z(\alpha) \) according to the law of large numbers and the continuous mapping theorem can be given by

\[
Z_n(\alpha) \triangleq \frac{\frac{1}{n} \sum_{i=1}^{n} |x_i|^{2\alpha}}{\left(\frac{1}{n} \sum_{i=1}^{n} |x_i|^\alpha\right)^2} - (1 + \alpha)
\]

(23)

so we can take the root of \( Z_n(\alpha) = 0 \) as the estimator of \( \alpha \). The Newton-Raphson iteration formula for solving the root of \( Z_n(\alpha) = 0 \) is given as follow[26]:

\[
\alpha_{k+1} = \alpha_k - \frac{Z_n(\alpha_k)}{Z'_n(\alpha_k)}
\]

(24)

where

\[
Z'_n(\alpha_k) = \frac{\left(\frac{2}{n} \sum_{i=1}^{n} |x_i|^{2\alpha} \log |x_i|\right)\left(\frac{1}{n} \sum_{i=1}^{n} |x_i|^\alpha\right)^2 - \left(\frac{1}{n} \sum_{i=1}^{n} |x_i|^\alpha \log |x_i|\right)\left(\frac{1}{n} \sum_{i=1}^{n} |x_i|^{2\alpha}\right)\left(\frac{2}{n} \sum_{i=1}^{n} |x_i|^\alpha\right)}{\left(\frac{1}{n} \sum_{i=1}^{n} |x_i|^\alpha\right)^4} - 1
\]

(25)

It was proved in [26] that the estimator \( \hat{\alpha} \) converges to the truth value \( \alpha \) in probability when the initial iteration value is greater than \( \alpha \).

3 Comparisons of the Estimation Methods

3.1 Theoretical comparison

(1) Comparison of MM and MLE. Moreover, the choice of initial value was critical for the iteration of MLE for the negative value of \( \alpha_k \) would occur if the initial value was unsuitable, especially, it would be impossible for the gamma function to carry out the iteration and keep proceeding.

(2) Comparison of GEM and MLE. The GEM which is based on \( f_2(\alpha), f_3(\alpha)(\Delta = 1) \) is compared with MLE in [18] that when \( \alpha \) was small and the sample was less, MLE is the best. Comparing \( f_2(\alpha) \) with \( f_3(\alpha) \), \( f_3(\alpha) \) was superior to \( f_2(\alpha) \) for the estimations of lighter-tail distribution situations (\( \alpha \geq 0.7 \)) and it is especially obvious when \( \alpha = 2 \). The expected estimations were difficult to obtain by \( f_2(\alpha) \) even if the sample size was large (\( n = 1000 \)) and \( \alpha \geq 0.7 \).

(3) Comparison of MM and NM. In comparison with MM, NM has weaker requirement than MM on the sample size. Since many signals are usually peaky and the sample size is finite in practice, it will attach significance importance in the application of NM.

(4) Comparison of GCM and other methods. It was proved that in [26] \( \hat{\alpha} \) will converge in probability to the truth value \( \alpha \) for any initial estimate greater than \( \alpha \) and different samples. Comparisons of MLE and GCME for various data size are shown in table 1 and table 2. Additionally, the calculation of GCM is a little more complex than the matching methods, but it occupies less memory space, therefore GCM is suitable for real-time data processing.
Table 1: Comparisons of MLE and GCM for various sample size when $\alpha = 0.4$ and the initial estimate is 1

<table>
<thead>
<tr>
<th>n</th>
<th>150</th>
<th>200</th>
<th>300</th>
<th>350</th>
<th>450</th>
<th>500</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCM</td>
<td>0.4099</td>
<td>0.4183</td>
<td>0.4140</td>
<td>0.4152</td>
<td>0.4023</td>
<td>0.4123</td>
</tr>
<tr>
<td>MLE</td>
<td>-1.947</td>
<td>0.1455</td>
<td>0.1751</td>
<td>0.1368</td>
<td>0.1908</td>
<td>0.3671</td>
</tr>
</tbody>
</table>

Table 2: Comparisons of MLE and GCM when $n=500$, $\sigma = 1$ and the initial estimate is 3

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>0.4</th>
<th>0.6</th>
<th>0.8</th>
<th>1.0</th>
<th>1.2</th>
<th>1.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCM</td>
<td>0.40388</td>
<td>0.58834</td>
<td>0.79971</td>
<td>0.99022</td>
<td>1.2759</td>
<td>1.4979</td>
</tr>
<tr>
<td>MLE</td>
<td>NaN</td>
<td>39.159</td>
<td>-47.983</td>
<td>-0.6715</td>
<td>-0.06645</td>
<td>0.5845</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>1.6</th>
<th>1.8</th>
<th>2.0</th>
<th>2.2</th>
<th>2.4</th>
<th>2.6</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCM</td>
<td>1.5967</td>
<td>1.9541</td>
<td>2.0909</td>
<td>2.2535</td>
<td>2.5205</td>
<td>2.7084</td>
</tr>
<tr>
<td>MLE</td>
<td>1.5787</td>
<td>1.864</td>
<td>2.0115</td>
<td>2.1752</td>
<td>2.4029</td>
<td>2.6701</td>
</tr>
</tbody>
</table>

3.2 Practical comparison

In practical modeling of GGD data, we can first roughly determine the interval containing the parameter $\alpha$ in it through some method such as calculating the sample kurtosis, or approximately estimating $\alpha$ using a suitable method based on experience.

(1) For Gaussian signals ($\alpha = 2$), MM is superior to MLE, but GEM ($k = 3$) is much more precise.

(2) For sup-Gaussian signals ($\alpha < 2$), MLE is superior to MM, GEM and other methods, but it is very sensitive to the samples and the initial estimator. In comparison, GCM is more robust for various initial estimator. NM can still accurately estimate the parameters of peaky signals ($\alpha < 1$) under the condition of small sample.

(3) For sub-Gaussian signals ($\alpha > 2$), we can employ GEM, MLE or GCM to estimate the parameter $\alpha$, and MLE is closer to the truth value of $\alpha$ than GCM as is shown in table 1.

4 Conclusions

Four well-known estimation techniques for the shape parameter of GGD, the moment-based method, the maximum likelihood method, the comementropy-based method, and the global convergence method are reviewed, and the summary of each method is described, at the same time their relative metrics are compared in detail. We make some proposals to choose suitable method under different conditions.

References


