Parameter Estimation of the Extended Generalized Gaussian Family Distributions using Maximum Likelihood Scheme

June-Yule Lee
Department of Marine Engineering, National Kaohsiung Marine University, 482, Chung-Chou 3rd Road, Chi-Jin District, Kaohsiung 805, Taiwan

Abstract: An extended generalized Gaussian distribution which can describe a family of symmetric and asymmetric distributions is considered. Parameter estimation of this function using maximum likelihood scheme is proposed. By measured the tail length and skewness of the observed data, the method integrates a pre-calculated table of initial values for parameters estimation. This allows a fast convergence of the presented model for real-time applications. The simulation results also show that the proposed scheme is an asymptotically unbiased estimator in terms of Cramer-Rao lower bound criterion.

Key words: Generalized Gaussian distribution, maximum likelihood, parameter estimation, Cramer-Rao lower bound

INTRODUCTION

In the field of science and engineering, the family of the Generalized Gaussian Distributions (GGD) are applied to model and detect the Gaussian and non-Gaussian signals (Miller and Thomas, 1972; Kassam, 1988). For example, in natural environments, the performance of communication systems can significantly degrade due to the atmospheric noise caused by the lightning activity in local and distant storm, or the ambient noise in the ocean. In seismic research (Walden, 1985), the non-Gaussian distribution has been used to model the amplitude distribution of the primary reflection coefficients. The aim is to estimate the reflection coefficient sequences as accurately as possible. In image processing, the fundamental work is fitting the image intensity histogram data to gain underlying knowledge of data source. Fan and Lin (2009) developed an expectation maximization algorithm to estimate the parameter of the GGD mixture model that best fit the observed image intensity histogram data. Other useful applications of the GGD also include, video coding (Sharifi and Leon-Garcia, 1995), handwritten character identification (Solihin and Leedham, 1999), speech recognition (Gazor and Zhang, 2003), etc. Therefore, the choice between different families of distribution is important and the estimation of the designed probability density function from observed data is the major task.

In this study, the parameter estimator for the extended generalized Gaussian distribution to include asymmetry is applied. The parameter estimation is determined by the maximum likelihood scheme with iteration method. The key factor of the iteration method depends on the choice of the initial step values. Soderlind (2006) presented and discussed the approaches to step size selection in control theory and signal processing. Tiele-Cuatle et al. (2007) proposed a procedure to control the step-size, where the step size is determined by the eigenvalues of its state matrices. For saving the computing cost and improving the accuracy, we proposed a table look-up method.

THE MODEL

The generalized Gaussian distribution to include asymmetry proposed by Nandi and Mampel (1995) is:

\[
p(x; \alpha, \beta, \alpha_0) = \begin{cases} 
\frac{1}{\beta} \exp \left( -\frac{|x|^\alpha}{\beta} \right), & \text{for } x < 0 \\
\frac{1}{\beta} \exp \left( -\frac{\alpha_0^{\alpha} \Gamma(1/\alpha)}{\beta \Gamma(1/\alpha_0)} |x|^\alpha \right), & \text{for } x \geq 0 
\end{cases}
\]  

(1)

and

\[
\alpha_0 = \frac{\alpha}{2 \beta \Gamma(1/\alpha_0)}
\]  

(2)

where, \( \alpha > 0, \alpha_0 > 0, \beta > 0 \) and \( \Gamma \) is gamma function. By setting:

\[
\beta_0 = \left( \frac{\Gamma(1/\alpha_0)}{\Gamma(1/\alpha)} \right)^{1/2}
\]  

(3)
then the variance of Eq. 1 is chosen as one. By changing the parameters in Eq. 1, one can obtain symmetric \((a_n - a_p)\) and asymmetric \((a_n \neq a_p)\) distributions. For symmetric cases, Fig. 1, \((a_n, a_p)\) are \((1,1)\) and \((2,2)\) represent the Laplace distribution and Gaussian distribution, respectively. When \(a_n\) tends to 0 the distribution tends to impulse and when \(a_n\) tends to infinite the distribution tends to Uniform distribution. For asymmetric cases, (Fig. 2), the right skew distribution can be modeled by reducing the parameter \(a_n\) and as \(a_n\) decreases the tail length in the right side increases. Therefore, Eq 1 provides a wide range of uni-modal distributions with symmetry and asymmetry.

![Symmetric model for varying the parameters of \((a_n = a_p)\)](image1)

![Asymmetric model for varying the parameters of \((a_n - a_p)\)](image2)

**THE ESTIMATOR**

Given a series samples \(x_i\), where \(i = 1, 2, 3, ..., N\), the likelihood function of Eq. 1 is:

\[
L = (A_p)^N \prod_{i \in \mathcal{I}} \exp \left( -\frac{x_i}{\beta_p} \right) \times (A_p)^k \prod_{i \in \mathcal{I}} \exp \left( -\frac{\alpha_p \Gamma(1/\alpha_p) x_i}{\beta_p \Gamma(1/\alpha_p)} \right)
\]

(4)

where, \(N_i (N_j)\) is the number of \(x_i\) samples \(\leq 0 (\geq 0)\). Taking the logarithm of Eq. 4, one gets:

\[
\ln L = N \ln A_p - \sum_{i \in \mathcal{I}} \left( -\frac{x_i}{\beta_p} \right) + N \ln A_p - \sum_{i \in \mathcal{I}} \left( \frac{\alpha_p \Gamma(1/\alpha_p) x_i}{\beta_p \Gamma(1/\alpha_p)} \right)
\]

(5)

If \(\partial \ln L / \partial \alpha_p \), \(\partial \ln L / \partial \beta_p\) and \(\partial \ln L / \partial \alpha_p\) are equate to zero, then the maximum likelihood estimates \(\hat{\alpha_p}\), \(\hat{\beta_p}\) and \(\hat{\alpha_p}\) are obtained, i.e.,

\[
\frac{\partial \ln L}{\partial \alpha_p} = N \left( \frac{1}{\alpha_p} - \Psi(1/\alpha_p) \right) + \sum_{i \in \mathcal{I}} \left( -\frac{x_i}{\beta_p} \right) \ln \left( -\frac{x_i}{\beta_p} \right) + N \left( \frac{1}{\alpha_p} - \Psi(1/\alpha_p) \right)
\]

\[
- \sum_{i \in \mathcal{I}} \beta_p \alpha_p \Gamma(1/\alpha_p) \Gamma(1/\alpha_p)
\]

\[
= 0
\]

(6)

\[
\frac{\partial \ln L}{\partial \beta_p} = -N \sum_{i \in \mathcal{I}} \frac{1}{\beta_p} + \sum_{i \in \mathcal{I}} \beta_p \left( -\frac{x_i}{\beta_p} \right) - N \sum_{i \in \mathcal{I}} \beta_p \left( \alpha_p x_i \right) = 0
\]

(7)

\[
\frac{\partial \ln L}{\partial \alpha_p} = -N \sum_{i \in \mathcal{I}} \left( \alpha_p x_i \right) \left( \ln \left( \alpha_p x_i \right) - \frac{1}{\beta_p} \Gamma(1/\alpha_p) \left( \psi(1/\alpha_p) + \frac{1}{\alpha_p} \right) \right)
\]

(8)

Where,

\[
I = \frac{\Gamma(1/\alpha_p)}{\beta_p \alpha_p \Gamma(1/\alpha_p)}
\]

(9)

and \(\Psi(\cdot)\) is digamma function defined as:

\[
\Psi(\cdot) = \frac{\partial \ln \Gamma(\cdot)}{\partial \ln(\cdot)} = \frac{1}{\Gamma(\cdot)} \frac{\partial \Gamma(\cdot)}{\partial \ln(\cdot)}
\]

(10)

To fit the observed data and approximate the solutions \(\alpha_p\), \(\beta_p\) and \(\alpha_p\) one need to solve (Eq. 6-8). In the following two iterated methods are developed.

**Iterated method 1**: Given a reasonable initial values \(\alpha_0\), \(\beta_0\) and \(\alpha_0\) (Eq. 6-8) can be solved iteratively for \(\alpha_n\), \(\beta_n\) and \(\alpha_p\). First step, compute Eq. 6 with \((\alpha_0, \Delta \alpha_0)\) by fixed \(\beta_0\) and
α_{p0}, then α_{n} is selected by the value of (α_{p0}±Δα_{p}) or
(α_{n0}±Δα_{n}) so that the likelihood function of Eq. 6
approaches zero. Second step, compute Eq. 7 with
(β_{p0}±Δβ_{p}) by fixed α_{p0} and α_{n0}, then β_{p0} can be updated by the
value of (β_{p0}+Δβ_{p}) or (β_{p0}−Δβ_{p}) so that Eq. 7
approaches zero. Third step, compute Eq. 8 with (α_{p0}±Δα_{p})
by fixed α_{p0} and β_{p0}, then α_{p0} can be updated by the value of
(α_{p0}+Δα_{p}) or (α_{p0}−Δα_{p}) so that Eq. 8 approaches zero.
Thus, one iteration is completed for updating α_{p0}, β_{p0}, and
α_{n0}. This process continues until Eq. 6-8 simultaneously
approach zero with some error tolerances. The accuracy of
the estimation of this method depends on the choice of the
step values Δα_{p0}, Δβ_{p0} and Δα_{n0}. Therefore, small step
value leads to accuracy of the estimation but requires
more computing cost.

**Iterated method 2:** In this section the gradient method for
iteration is developed. In the j-th iteration the Taylor
series of Eq. 11 is applied to update the parameters α_{p0}, β_{p0}
and α_{n0}.

\[ \hat{\theta}^{(j)} = \hat{\theta}^{(0)} - V(\hat{\theta}^{(0)})^{-1} \text{grad } \ln L(\hat{\theta}^{(0)}) \]  

where, \( \hat{\theta} = (\alpha_{p0}, \beta_{p0}, \alpha_{n0})^{T} \), \text{grad } \ln L = (\partial \ln L / \partial \alpha_{p0}, \partial \ln L / \partial \beta_{p0}, \partial \ln L / \partial \alpha_{n0})^{T} \)

and:

\[ V(\hat{\theta}) = \begin{pmatrix}
\frac{\partial^{2} \ln L}{\partial \alpha_{p0}^{2}} & \frac{\partial^{2} \ln L}{\partial \alpha_{p0} \partial \beta_{p0}} & \frac{\partial^{2} \ln L}{\partial \alpha_{p0} \partial \alpha_{n0}} \\
\frac{\partial^{2} \ln L}{\partial \beta_{p0} \partial \alpha_{p0}} & \frac{\partial^{2} \ln L}{\partial \beta_{p0}^{2}} & \frac{\partial^{2} \ln L}{\partial \beta_{p0} \partial \alpha_{n0}} \\
\frac{\partial^{2} \ln L}{\partial \alpha_{n0} \partial \alpha_{p0}} & \frac{\partial^{2} \ln L}{\partial \alpha_{n0} \partial \beta_{p0}} & \frac{\partial^{2} \ln L}{\partial \alpha_{n0}^{2}}
\end{pmatrix} \]  

is a Fisher information matrix. The inverse of the Fisher
information matrix plays a key role for the iteration of
Eq. 11. Given N samples of the observed data, the Fisher
matrix can be obtained from Eq. 12. For each iteration,
the expected value of the information matrix is calculated with
updating parameters α_{p0}, β_{p0} and α_{n0}. The estimation of
the parameter is obtained as the value \( V(\hat{\theta}^{(0)})^{-1} \)
converge to zero with some error tolerances.

**Cramer-Rao lower bound:** It is important to understand
the unbiasedness of an estimator. The most common way
to decide whether or not an estimator's distribution is
unbiased is by looking at its expected value. The Cramer-Rao Lower Bound (CRLB) for the standard
deviation of every unbiased estimator is defined as
(Hogg and Tanis, 1989):

\[ \text{std}(\hat{\theta}_{p0}) \geq \left[ \text{NE} \left( \frac{\partial \ln L}{\partial \alpha_{p0}} \right)^{2} \right]^{-1/2} \]  

\[ \text{std}(\hat{\alpha}_{n0}) \geq \left[ \text{NE} \left( \frac{\partial \ln L}{\partial \alpha_{n0}} \right)^{2} \right]^{-1/2} \]  

where, N is the samples of the every realizations and the
\( \text{E}[*] \) is the expected value. Note that Eq. 13-15 are
equivalence to the inverse of the square root of the
diagonal elements in Eq. 12.

**RESULTS AND DISCUSSION**

In the following simulation, the digamma function of
Eq. 10 is computed using the algorithm in Char et al.
(1991) and for further details of the algorithms one can
refer to Abramowitz and Stegun (1972). The step value for
method 1 is 0.005 and the error tolerance for method 2 is
0.05. All simulation results are based on 50 Monte Carlo
runs.

To obtain a good initial value, the tail length and
skewness of the observed data are measured and
indicated by Q_{t} (tailness) and Q_{s} (skewness) (Lee, 2003):

\[ Q_{t} = \frac{\bar{u}(0.05) - L(0.05)}{\bar{L}(0.05) - L(0.05)} \]  

and

\[ Q_{s} = \frac{\bar{u}(0.05) - m(0.25)}{m(0.25) - L(0.05)} \]  

where, \( \bar{u}(\alpha_{0}) \) and \( L(\alpha_{0}) \) are the means of right \( \alpha_{0} \) and left
\( \alpha_{0} \) intervals and \( m(\alpha) \) is \( \alpha \)-trimmed mean. The \( Q_{t} \) for
Gaussian distribution is 2.58 and for Laplace distribution
is 3.30, and \( Q_{s} \) is 1 for all symmetric cases. Distributions with
\( Q_{s} < 1 \) have relatively longer right tail (right
skewness), while those with \( 0 < Q_{s} < 1 \) have relatively longer
left tail (left skewness). By choosing the parameter values
(with \( \alpha_{s} = \alpha_{s} \) for symmetry and \( \alpha_{s} \neq \alpha_{s} \) for asymmetry),
the Q_{t}-Q_{s} plot (Fig. 3) for various tail length and skewness
can be pre-calculated which corresponds to the
parameters under the Eq. 1. For example, the values for
Gaussian model the (Q_{t}, Q_{s}) is (2.58, 1) and the
corresponding (\alpha_{p0}, \beta_{p0}, \alpha_{n0}) is (2, 1.14, 2). The results shown
in Fig. 3 are various Q_{t}-Q_{s} plots which cover the tailness
up to 5.5 and right skewness up to 10. Using the table
look-up of Q_{t}-Q_{s} plots and the corresponding (\alpha_{p0}, \beta_{p0}, \alpha_{n0}), a
reasonable initial value of estimated parameter can be
interpolated.

The samples applied in Table 1 are Gaussian
distributions, the results shown both iterated methods are
reach the exact values of the parameters with low bias. As
the sample size increases the standard deviation of the
estimation decreases. It appears that the estimation is asymptotically unbiased and satisfies the Cramer-Rao lower bound, where the standard deviations approach but greater than the CRLB (Fig. 4). The results are shown in Table 2 and Fig. 5 are for the samples of Laplace distributions. They are also in agreement with the theoretical values of the parameters.

The parameter estimation of extended GGD model include asymmetric distribution is achieved in this study. This differs from the previous study of GGD mode, where more data can be fitted in the proposed model. Furthermore, the pre-calculated table of the Q₁-Q₂ plots save the computing cost and improve the accuracy of the estimation. The key factor of the iteration method depends on the choice of the initial and step values, in this study a table look-up method is provided for fast and accuracy estimation.

Table 1: Gaussian distribution with 50 realizations

<table>
<thead>
<tr>
<th>Samples</th>
<th>Parameters</th>
<th>$w_0$</th>
<th>$\hat{\mu}_1$</th>
<th>$\hat{\mu}_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exact</td>
<td>2.000</td>
<td>1.414</td>
<td>2.000</td>
<td></td>
</tr>
<tr>
<td>128</td>
<td>Method 1</td>
<td>2.121±0.032</td>
<td>1.407±0.092</td>
<td>1.995±0.343</td>
</tr>
<tr>
<td></td>
<td>Method 2</td>
<td>2.154±0.564</td>
<td>1.406±0.122</td>
<td>2.026±0.423</td>
</tr>
<tr>
<td>256</td>
<td>Method 1</td>
<td>2.013±0.196</td>
<td>1.409±0.068</td>
<td>2.034±0.326</td>
</tr>
<tr>
<td></td>
<td>Method 2</td>
<td>2.100±0.234</td>
<td>1.404±0.095</td>
<td>2.041±0.383</td>
</tr>
<tr>
<td>512</td>
<td>Method 1</td>
<td>2.017±0.215</td>
<td>1.414±0.564</td>
<td>2.034±0.199</td>
</tr>
<tr>
<td></td>
<td>Method 2</td>
<td>2.018±0.261</td>
<td>1.411±0.793</td>
<td>2.034±0.243</td>
</tr>
<tr>
<td>1024</td>
<td>Method 1</td>
<td>1.995±0.156</td>
<td>1.405±0.049</td>
<td>1.988±0.150</td>
</tr>
<tr>
<td></td>
<td>Method 2</td>
<td>1.991±0.172</td>
<td>1.402±0.054</td>
<td>1.981±0.170</td>
</tr>
<tr>
<td>2048</td>
<td>Method 1</td>
<td>2.026±0.105</td>
<td>1.420±0.027</td>
<td>2.013±0.093</td>
</tr>
<tr>
<td></td>
<td>Method 2</td>
<td>2.026±0.125</td>
<td>1.419±0.037</td>
<td>2.015±0.110</td>
</tr>
<tr>
<td>4096</td>
<td>Method 1</td>
<td>2.008±0.071</td>
<td>1.414±0.019</td>
<td>1.997±0.064</td>
</tr>
<tr>
<td></td>
<td>Method 2</td>
<td>2.005±0.086</td>
<td>1.143±0.027</td>
<td>1.999±0.077</td>
</tr>
</tbody>
</table>

Fig. 3: The $Q_1$-$Q_2$ plots

Fig. 4: Standard deviation of the estimated parameters of Gaussian samples
Fig. 5: Standard deviation of the estimated parameters of Laplace samples

Table 2: Laplace distribution with 50 realizations

<table>
<thead>
<tr>
<th>Samples</th>
<th>Parameters</th>
<th>$\alpha_1$</th>
<th>$\beta_1$</th>
<th>$\sigma_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exact values</td>
<td>1.000</td>
<td>0.797</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>128 Method 1</td>
<td>1.055±0.125</td>
<td>0.766±0.100</td>
<td>1.087±0.170</td>
<td></td>
</tr>
<tr>
<td>Method 2</td>
<td>1.129±0.224</td>
<td>0.827±0.182</td>
<td>1.154±0.249</td>
<td></td>
</tr>
<tr>
<td>256 Method 1</td>
<td>1.008±0.072</td>
<td>0.714±0.066</td>
<td>1.015±0.079</td>
<td></td>
</tr>
<tr>
<td>Method 2</td>
<td>1.018±0.128</td>
<td>0.719±0.128</td>
<td>1.020±0.112</td>
<td></td>
</tr>
<tr>
<td>512 Method 1</td>
<td>1.007±0.051</td>
<td>0.708±0.060</td>
<td>0.999±0.074</td>
<td></td>
</tr>
<tr>
<td>Method 2</td>
<td>1.011±0.094</td>
<td>0.710±0.108</td>
<td>1.002±0.107</td>
<td></td>
</tr>
<tr>
<td>1024 Method 1</td>
<td>1.008±0.052</td>
<td>0.718±0.054</td>
<td>1.018±0.059</td>
<td></td>
</tr>
<tr>
<td>Method 2</td>
<td>1.006±0.051</td>
<td>0.717±0.052</td>
<td>1.015±0.052</td>
<td></td>
</tr>
<tr>
<td>2048 Method 1</td>
<td>1.000±0.024</td>
<td>0.708±0.024</td>
<td>1.000±0.030</td>
<td></td>
</tr>
<tr>
<td>Method 2</td>
<td>0.999±0.044</td>
<td>0.707±0.045</td>
<td>1.002±0.040</td>
<td></td>
</tr>
<tr>
<td>4096 Method 1</td>
<td>1.001±0.018</td>
<td>0.708±0.019</td>
<td>1.000±0.023</td>
<td></td>
</tr>
<tr>
<td>Method 2</td>
<td>1.001±0.035</td>
<td>0.707±0.038</td>
<td>0.999±0.035</td>
<td></td>
</tr>
</tbody>
</table>

CONCLUSION

The maximum likelihood estimator is applied to fit three-parameter probability density function. The results showed that the proposed estimator is asymptotically unbiased for generalized Gaussian family distributions. They are useful for modeling and detecting of the observed data in real-time applications.

ACKNOWLEDGMENT

The author would like to acknowledge the valued comments made by Prof. A.K. Nandi.

REFERENCES


