Neural Processing Letters (2005) 22:377–389
DOI 10.1007/s11063-005-1385-9

Estimation of Shape Parameter of GGD Function by Negentropy Matching

RAJKISHORE PRASAD⋆, HIROSHI SARUWATARI and KIYOHIRO SHIKANO
Nara Institute of Science and Technology, Takayama-cho, Nara 6300101, Japan.
e-mails: profrkishore@yahoo.com, kishor-p@is.naist.jp

Abstract. In this paper we present a novel method for the estimation of the shape parameter of the Generalized Gaussian Distribution (GGD) function for the leptokurtic and Gaussian signals by matching negentropy of GGD function and that of data approximated by some non-polynomial functions. The negentropy of GGD function is monotonic function of its shape parameter for values corresponding to super-Gaussian and Gaussian distribution family. The simulation results have been compared with those obtained by existing methods such as Mallat's method and Kurtosis matching method. It has been found that the proposed method is effective and useful in the cases where we have a few observation samples and distribution is highly spiky.

Key words. GGD, kurtosis, moment matching, negentropy, shape parameter estimation

1. Introduction

Accurate statistical model for the observed data is a matter of key importance in the many statistical signal processing algorithms. Use of a parametric model for the unknown underlying Probability Density Function (PDF) of the observed data is very common practice. However, accuracy of such a model depends on the accuracy in estimation of defining parameters. One of such parametric models is the Generalized Gaussian Distribution (GGD) function [1] which is widely used to model signals in the different areas. The GGD function is defined in terms of location parameter or mean, scale parameter, and shape parameter. This function is also called as generalized Laplacian distribution and exponential power distribution. The detailed properties of this distribution was first examined in very past in [2] in the development of the Bayesian inferential process and has been used in the different areas of applications. Recently, GGD has been used in the different areas of digital signal processing, e.g., digital image processing [3, 4], speech signal processing [5, 6], digital watermarking [7], blind signal separation [5, 8–10] of speech, images and other arbitrary signals.

The GGD function for different values of shape parameter represents distributions with various shapes such as uniform, normal, Laplacian, and even more highly

⋆Corresponding author.

parsimonious distributions with the exponentially decaying heavy tails. Accordingly, it can be used to model data with symmetric distributions with varying degree of peakedness. In the statistical modeling of the underlying PDF of a data, highly accurate values of these parameters are needed and must be estimated from the data. There have been development of different methods such as Maximum Likelihood (ML) estimation, moment and kurtosis matching methods for the estimation of parameters of GGD function [11–13]. These different methods have been used by different researchers, e.g., in [14], moment matching method has been used to estimate shape parameter for the speech signal. In [10], ML method has been used while in [6, 15] kurtosis matching method has been used. The detailed discussion on relative pros and cones of these methods can be found in [11]. In some applications, such as in [6], very small number of data samples are available or accessible and estimation of GGD parameters from very small number of samples by ML or moment matching methods are less accurate [11]. In this letter, we propose a new method for the estimation of the shape parameter by negentropy matching which has been found suitable for the estimation from a few data samples. Negentropy of the GGD is monotonic function of the shape parameter and inverse mapping of this function can give value of the shape parameter. Also, the proposed method has been applied for the estimation of shape parameter for computer simulated data and some real world signals to compare its performance with that of aforesaid existing methods of kurtosis and moment matching.

2. GGD Function Family

The GGD function is generalized form of generalized gamma distribution family. The PDF for an arbitrary zero mean Random Variable (RV) x in the form of GGD function with mean μ, scale parameter α, and shape parameter β is given by

$$f_{GG}(x;\mu,\alpha,\beta)=\frac{\beta}{2\alpha\Gamma(1/\beta)}\exp(-[|x|/\alpha]^{\beta})=A\exp(-[|x|/\alpha]^{\beta}) \quad \text{for } \mu=0, \quad (1)$$

where $A=\beta/2\alpha\Gamma(1/\beta)$, $\alpha=\sigma\sqrt{\Gamma(1/\beta)/\Gamma(3/\beta)}$; $\Gamma(y)=\int_0^\infty e^{-t}t^{y-1}dt$ = Gamma function; $-\infty<x<\infty$; $\alpha>0$; $\beta>0$; and σ = Standard deviation. The value of β controls rate of exponential decay of the function. Thus the shape of GGD function depends on the value of shape parameter β. For $\beta = 0.5$, $\beta=1$ and $\beta = 2$, GGD function represents, respectively, Gamma Distribution, Laplacian Distribution and Gaussian Distribution and the shape of distribution tends to become uniform as $\beta\to\infty$. For $\beta\to 0$, GGD represents impulsive function with flat tails. The GGD function for zero mean and different values of α and β are shown in Figure 1.

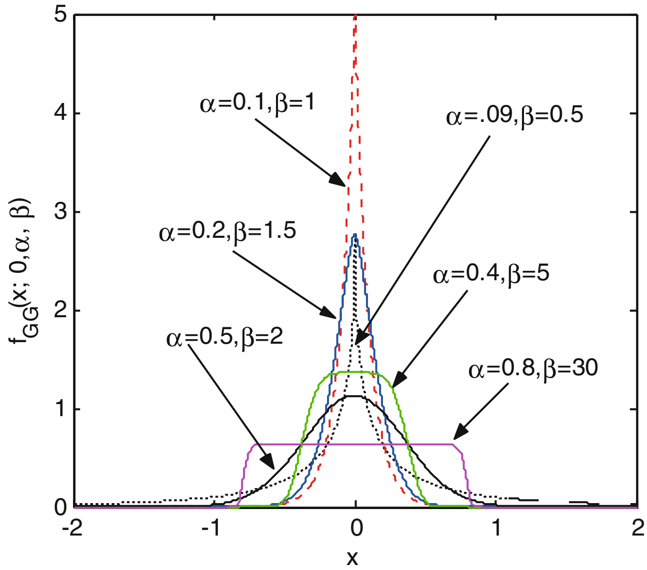

Figure 1. GGD distribution for different values of scale parameter and shape parameter. This family of distribution includes Gaussian, sub-Gaussian and super-Gaussian distributions.

3. Parameter Estimation for GGD

In order to use GGD function as a PDF model for the data $z = \{z_1, z_2, \ldots, z_N\}$ with N samples, the crucial task is the estimation of its defining parameters μ, α, and β from the data. There exist many methods to estimate these parameters. The mean $\bar{z}$ or median $\tilde{z}$ of the data is taken as location parameter μ. The suitability of mean or median as the location parameter depends on the value of β and choice can be judged by the efficiency $\eta(\beta)$ of these estimators defined as follows [16]

$$\eta(\beta) = \frac{\mathrm{var}[\bar{z}]}{\mathrm{var}[\tilde{z}]} \approx \frac{\Gamma(3/\beta)}{\Gamma(1/\beta)\Gamma^2(1+1/\beta)} \tag{2}$$

where var[.] denotes variance of [.]. The plotting of the theoretical relation in Equation (2) is shown in Figure 2 which also contains zoomed-in portion, inside the figure, of the curve for very small values of β. It is evident from Figure 2 that the mean is a good estimator of location parameter for $\beta > 1.41$. For the lower values of β ($\beta < 1.41$), the median is better estimate of the location parameter, and the suitability of both estimators worsen for very small values of β, for as for $\beta \to 0, \eta \to \infty$. The scale parameter is estimated from the knowledge of the shape parameter β and standard deviation σ_z of the data as follows [5]

$$\alpha = \sigma_z \sqrt{\frac{\Gamma(1/\beta)}{\Gamma(3/\beta)}} \quad \text{where } \sigma_z = \frac{1}{N}\sum_{i=1}^{N}(z_i - \mu)^2 \tag{3}$$

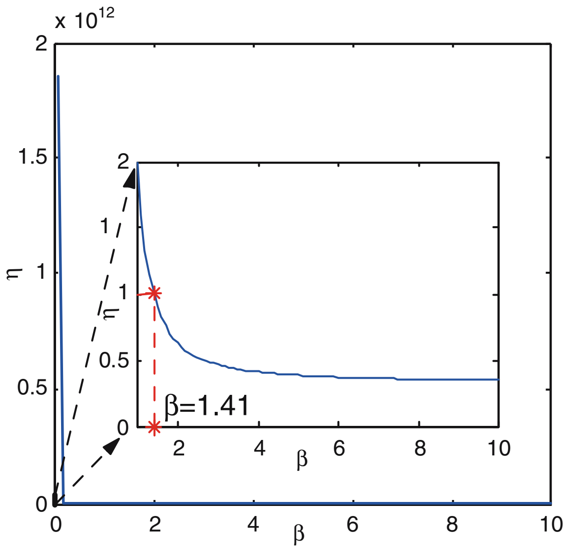

Figure 2. Theoretical value of η for estimators of μ.

As said earlier, there are many other methods such as ML estimation approach, kurtosis and moment matching methods for the estimation of the shape parameter. The comparative study on the performance of these methods can be found in [11] from where it can be imbued that ML approach is suitable for the spiky signals. However, due to computational complexities it is unsuitable for the real time application and data with small number of samples. Apart from ML approach other methods, based on the matching of some monotonic function of the shape parameter of GGD and same of the data, have also been proposed. We take here Mallat's method [13] and kurtosis matching method [14] for study. Mallat's method is a moment matching method in which the Generalized Gaussian Ration (GGR) of the unknown data is matched with the theoretical value of GGR of the GGD function. The GGRs of the data and of GGD function are related as

$$\mathrm{GGR}_z = \frac{\mathrm{E}[|z|]}{\sqrt{\sigma_z^2}} = \Gamma(2/\beta)/\sqrt{\Gamma(3/\beta)\Gamma(1/\beta)} = \gamma(\beta) \tag{4}$$

where γ is some function of the shape parameter of the GGD. The shape parameter is obtained by inverting the above relation as follows

$$\beta = \gamma^{-1}(\mathrm{GGR_z}) \tag{5}$$

Similarly, the generalized kurtosis of the GGD is also monotonic function of its shape parameter. The method of shape parameter estimation by matching of kurtosis of the data and of GGD is also widely used in the signal processing applications [14]. The Generalized kurtosis K_z of the data and shape parameter β of

GGD function are related as

$$K_z = \frac{\frac{1}{N}\sum_{i=1}^{N}|z_i|^4}{\left(\frac{1}{N}\sum_{i=1}^{N}|z_i|^2\right)^2} = \frac{\Gamma(5/\beta)\Gamma(1/\beta)}{\Gamma(3/\beta)^2} = \phi(\beta) \tag{6}$$

where ϕ is some function of shape parameter of GGD to represent its kurtosis. The shape parameter is determined by inverting the above relation as follows

$$\beta = \phi^{-1}(K_z) \tag{7}$$

We propose here negentropy based method for the shape parameter estimation. The negentropy J of the GGD can be computed only in terms of shape parameter of the GGD function. The negentropy of the data is a measure of information content in the data and it is always positive and is invariant from scale and linear transformations [17] of the data. The negentropy J is defined in terms of Differential Entropy (DE) of the data. The DE, ΔH of z is given by [17, 18]

$$\Delta H(z) = -\int_{-\infty}^{\infty} p(z)\log p(z)\,\mathrm{d}z, \tag{8}$$

where $p(z)$ represents PDF of the data. Using GGD function for the PDF of the data, Equation (8) an be given by

$$\begin{aligned}\Delta H(z) &= \log(1/A)\int_{-\infty}^{\infty} A\mathrm{e}^{-\frac{|z|^\beta}{\alpha^\beta}}\mathrm{d}z + \int_{-\infty}^{\infty} A\frac{|z|^\beta}{\alpha^\beta}\mathrm{e}^{-\frac{|z|^\beta}{\alpha^\beta}}\mathrm{d}z \\ &= \log(1/A) + 2A\int_{0}^{\infty}\frac{|z|^\beta}{\alpha^\beta}\mathrm{e}^{-\frac{|z|^\beta}{\alpha^\beta}}\mathrm{d}z \end{aligned}\tag{9}$$

Since integral in second term of Equation (9) is for positive values of the variable, using $(z/\alpha)^\beta = m \Rightarrow \mathrm{d}z = \alpha\beta^{-1}m^{(1/\beta)^{-1}}\mathrm{d}m$ in Equation (9) along with use of value of A from Equation (1) and that of α from Equation (3). Equation (9) can be further simplified as

$$\begin{aligned}\Delta H(z) &= \log(1/A) + \frac{2A\alpha}{\beta}\int_0^\infty m^{\frac{1}{\beta}}\mathrm{e}^{-m}\mathrm{d}m = \log(1/A) + \frac{2A\alpha}{\beta}\Gamma\left(\frac{1}{\beta}+1\right) \\ &= f(\alpha,\beta) = \log\left[\frac{2\alpha\Gamma(1/\beta)}{\beta}\right] + \frac{1}{\beta} = \left[\frac{2}{\beta}\sqrt{\frac{\sigma_z^2\Gamma(1/\beta)^3}{\Gamma(3/\beta)}}\right] + \frac{1}{\beta}\end{aligned}\tag{10}$$

Equation (l0) gives DE of data in terms of scale and shape parameters. The negentropy J is computed as the difference of DE $\Delta H(z_{\mathrm{Gauss}})$ of a Gaussian RV z_{Gauss}, with same variance as of that of $z(\sigma^2_{z_{\mathrm{Gauss}}} = \sigma_z^2)$, and DE $\Delta H(z)$ of z modeled by the GGD $f_{GG}(0,\alpha,\beta)$. Accordingly, negentropy J, in light of Equation (10) is given by [19]

$$J = \Delta H(z_{\text{Gauss}}) - \Delta H(z) = f(\alpha_g, \beta_g = 2) - f(\alpha, \beta)$$
$$= \log\left[\frac{\beta}{2}\sqrt{\frac{\Gamma(0.5)^3\Gamma(3/\beta)}{\Gamma(1/\beta)^3\Gamma(1.5)}}\right] + \left(0.5 - \frac{1}{\beta}\right) = \psi(\beta) \quad (11)$$

where α_g and β_g are scale and shape parameters for Gaussian RV z, ψ is some function of shape parameter of the GGD to measure negentropy. The theoretical variation of negentropy of GGD with shape parameter, as in Equation (11), is shown in Figure 3. Since it is monotonically decreasing function of the shape parameter for $0 \leqslant \beta \leqslant 2$ it can be inverted to get β if the negentropy of the data is known. For $\beta > 2$, GGD belongs to sub-Gaussian family and negentropy again begins to rise highly non-linearly with increasing β. It is important to note that the relation between negentropy and shape parameter is monotonic either before or after $\beta = 2$. So it cannot be used for the estimation for all values of shape parameters together, because it will introduce ambiguity. However, it is obvious from curves in Figure 3 that for a very large change in β there occurs very smaller change in negentropy for $\beta > 2$ than for $\beta < 2$. This indicates that the curve for $0 \leqslant \beta \leqslant 2$ will be more sensitive to change in β and will be better for estimation of β by inverse mapping. The shape parameters of super-Gaussian signals, for $0 \leqslant \beta \leqslant 2$ is estimated by inverting the above relation as follows

$$\beta = \psi^{-1}(J) \quad (12)$$

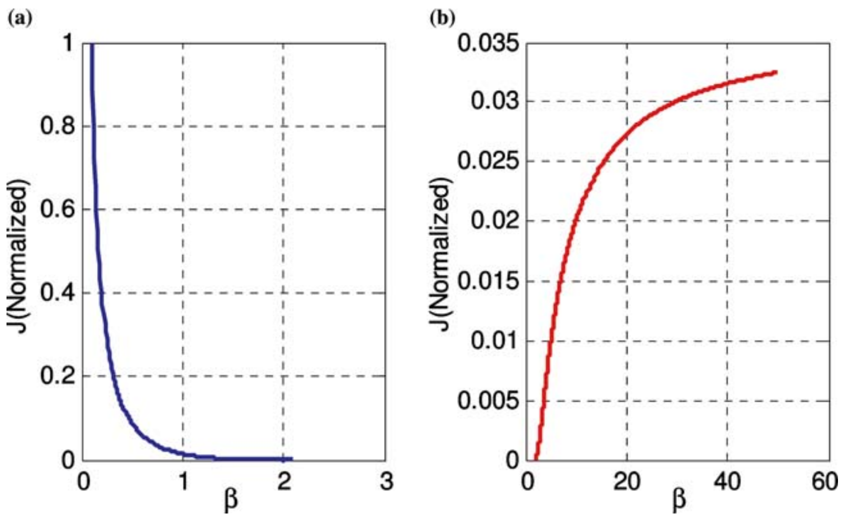

Figure 3. Shape parameter versus negentropy of the GGD. It is zero for Gaussian distribution and positive for the spiky distribution (a) for $0.1 \leqslant \beta \leqslant 2.1$ (b) for $2 \leqslant \beta \leqslant 60$.

The negentropy of the standardized data z can be approximated in terms non-polynomial density functions given by Hyvarinen *et al.* [17]

$$J(z) = k_1 \left[E\left\{ z \exp\left(-\frac{z^2}{2} \right) \right\} \right]^2 + k_2 \left[E\left\{ \exp\left(-\frac{z^2}{2} \right) \right\} - \sqrt{\frac{1}{2}} \right]^2 \tag{13}$$

where $k_1 = 7.412, k_2 = 33.67, E\{.\}$ denotes expectation of $\{.\}$. Cumulants based approximation of negentropy can also be used but it is worst for super-Gaussian distribution because it gives too much weight to the tail of the distribution [17]. In the Equations (5), (7) and (13) it is impossible or may be difficult to find exact inverse functions γ^{-1}, ϕ^{-1} and ψ^{-1}. However, with the help of look-up table it is possible to approximate the inverse relation and same will be used here. The used look-up table can be simple $(2 \times R)$ array in which one row (or column) contains R values of shape parameters and others row (or column) contains corresponding values of negentropy estimated from Equation (11). Since the entries in table include shape parameter values up to 2, it can be made smaller or larger subject to required tolerance and need of real time application. However, it can not be always expected and guaranteed that the entries for β–J pair will always match for the unknown test data. Under such a case the entries in look-up table can be interpolated to get an approximate value of the shape parameter at the cost of extra but easier computations.

4. Experiments and Results

For the experiment random variables of different lengths, ranging from order of 10 to 10^5, were generated with GGD parameters mean $= 0$, standard deviation $= 1$ and shape parameters ranging from 0.02 to 2. The look-up table for negentropy, kurtosis and moment matching were prepared for different values of β ranging from 0.02 to 2 (of size 2×100) for negentropy matching method and 0.02 to 5 for the moment and kurtosis matching. Two sets of data, namely (1) small number of samples (20–80), and other of large number of samples (10^2 to 10^5) were prepared. The estimated and true shape parameters for the first data set are shown in Figure 4 for β ranging from 0.02 to 0.3 (up to 2 are not shown only to keep clarity in figure). The related, error in estimation, for the value of β ranging from 0.02 to 2, is plotted in Figure 5 from where estimation performance for β ranging from 0.3 to 2 (not shown in Figure 5) can also be imbued. The similar estimation results for large sample size are shown in Figures 6 and 7. It is evident from these figures that the proposed method provides lesser error, in comparison to moment or kurtosis matching methods, in estimation from small number of samples in case when β is less than 1. The reason behind such behavior of estimator can be explained using Figure 8 which shows slopes of negentropy, kurtosis and of GGR (inverted) curves for related range of β value. The slope of negentropy curve is more conducive than that of kurtosis or GGR curve to notice and reflect small mutual changes in abscissa and ordinate. Thus even for

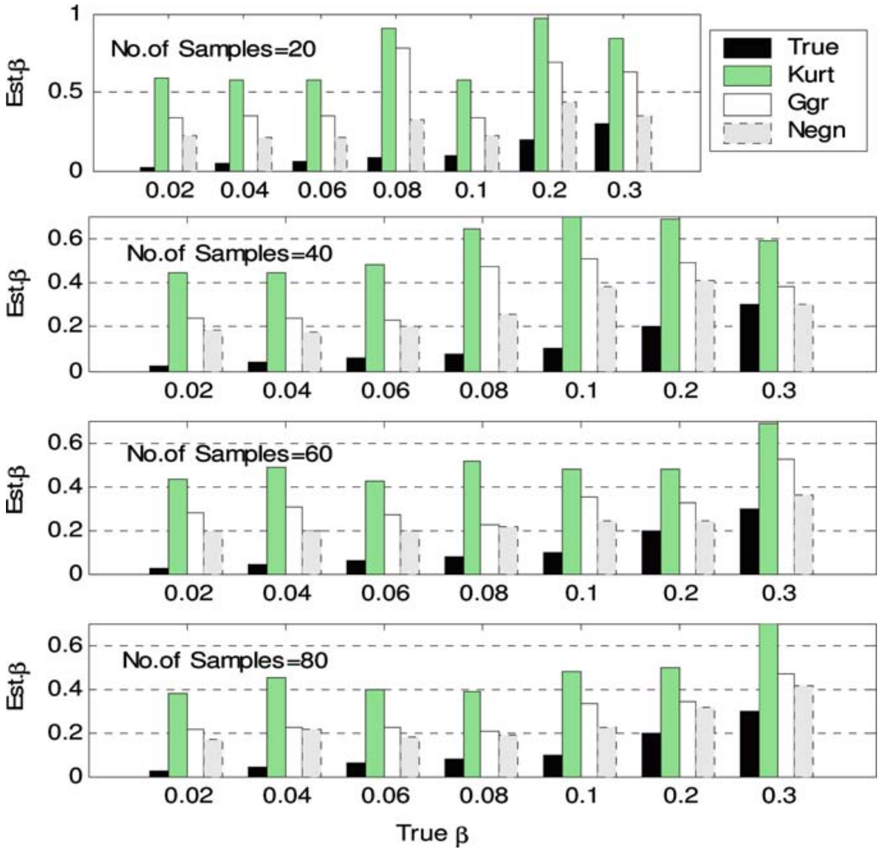

Figure 4. True and estimated β for different sample sizes. The legend indication is same for all subplots. (Kurt-kurtosis matching, Ggr-Mallat's method, Negn = Negentropy matching.)

a large change in kurtosis or GGR, β remains same, but that change is noticed by the negentropy curve. However, this property of kurtosis and GGR curves, arising due to non-linearity, can make it robust to outlier. It is important to mention here that if the $J - \beta$ relation for the sub-Gaussian region will be used to obtain shape parameter (of course for sub-Gaussian signal) similar non-linear relation, as shown in Figure 3(b), between abscissa and ordinate will result in inaccurate estimation because J–β relation is saturating very early. Also, the J–β mapping relation for both the leptokurtic and platykurtic signals cannot be kept in one look-up-table because it will give two values of shape parameters for a given value of negentropy and estimation will be ambiguous. It can be imbued from Figure 9 that such ambiguity will exist for $0.75 < \beta < 2$ before and after mesokurtic value ($\beta = 2$) because for $\beta \gg 2$ negentropy curve begins to saturate to 0.035. In Figure 10 the results of estimation of β for a highly spiky noise, machine gun noise taken from the NOISEX-92 database, are presented. Sub-plots in that figure present fittings of GGD function,

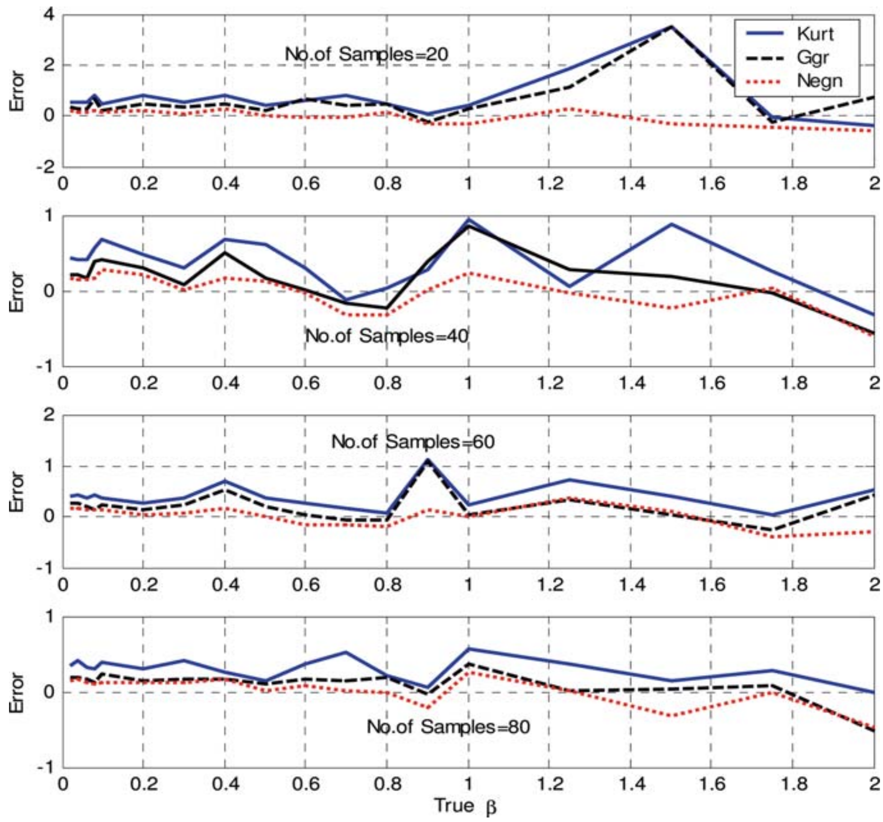

Figure 5. Error in estimation of shape parameter for small value of shape parameters with small no. of data samples. The proposed method gives less error than kurtosis and moment matching.

with estimated parameters from noise data of different sample size, e.g., 20, 100, and 1000 samples, in the normalized histograms of data. The shape parameters were estimated using kurtosis, moment and negentropy matching methods. In each case it can be seen how the negentropy based estimation provides better fit. In one of the sub-plots in Figure 10 the Chi-Square scores [20] between the GGD with estimated parameters and data are also shown. It is evident from the figure that even for the smaller number of samples negentropy based method gives better estimation than that of the moment or kurtosis based methods.

5. Conclusions

We proposed a novel method for the shape parameter estimation of the GGD by negentropy matching. This method provides better estimation of β, in comparison to moment or kurtosis matching, when the data samples are smaller and shape parameter is also less than 1. For $\beta > 2$ negentropy matching method can not be used as the negentropy again begins to increase after $\beta = 2$, terminating

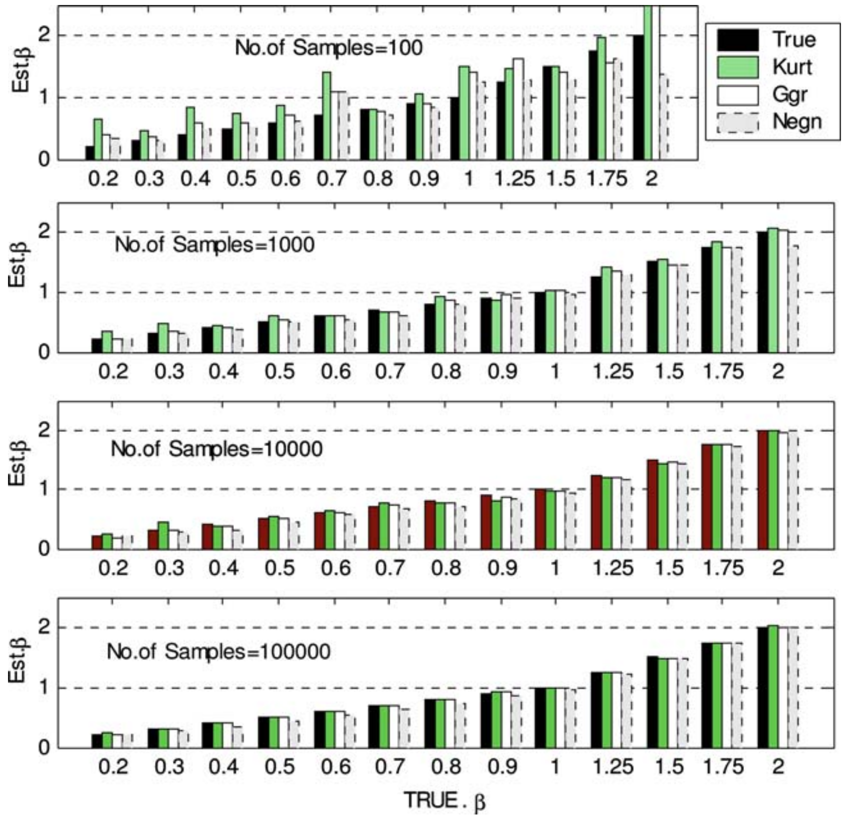

Figure 6. True and estimated β for different but large sample sizes (legend is same for all plots).

the monotonicity of the negentropy curve. Thus the proposed method can give ambiguous result if entire values of shape parameter of GGD are put in the same look-up table. This is happening due to break of monotonic relation between negentropy and shape parameter for $\beta = 2$. However, most of the natural signals, e.g., speech, image are leptokurtic, so proposed method is still useful. Anonymous reviewers suggested an interesting idea to mitigate such an ambiguity by use of normalized kurtosis of the data as indicator of the super-Gaussianity or sub-Gaussianity of data to select suitable part of $J - \beta$ relation in the look-up table. However, relative performance of $J - \beta$ relation for sub-Gaussian and super-Gaussian region for the estimation β is still unexplored, but due to early saturation in the $J - \beta$ relation for sub-Gaussian side it will not give better result. We have approximated negentropy in terms of non-polynomial function. It can also be approximated in terms of cumulants. Estimation using cumulants based approximation of negentropy is left unexplored, however, cumulants based approximation of negentropy is worst for the super-Gaussian data.

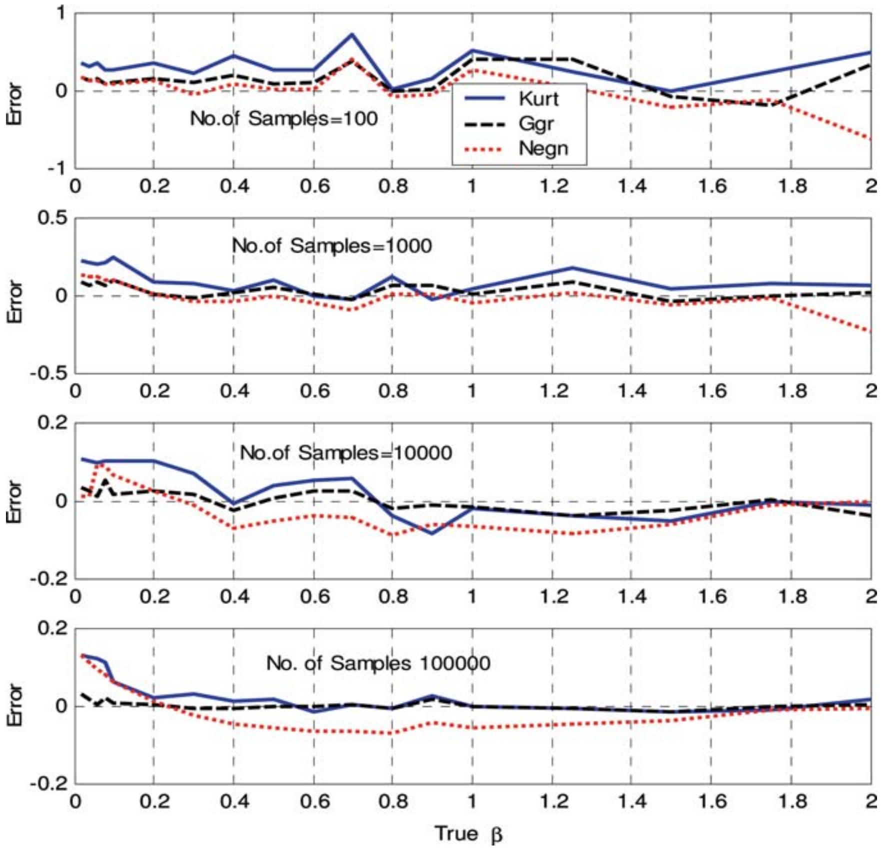

Figure 7. Error in estimation of shape parameter for small value of shape parameters with small no. of data samples. The proposed method gives less error than kurtosis and moment matching.

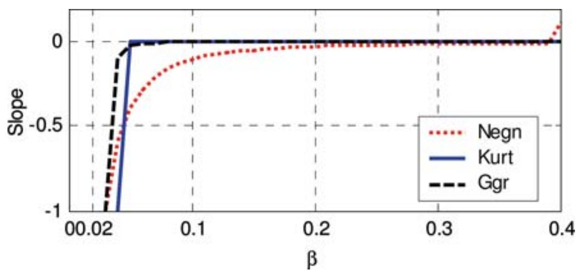

Figure 8. Slope of the negentropy, kurtosis and GGR of the GGD for small value (0.02–0.4) of β.

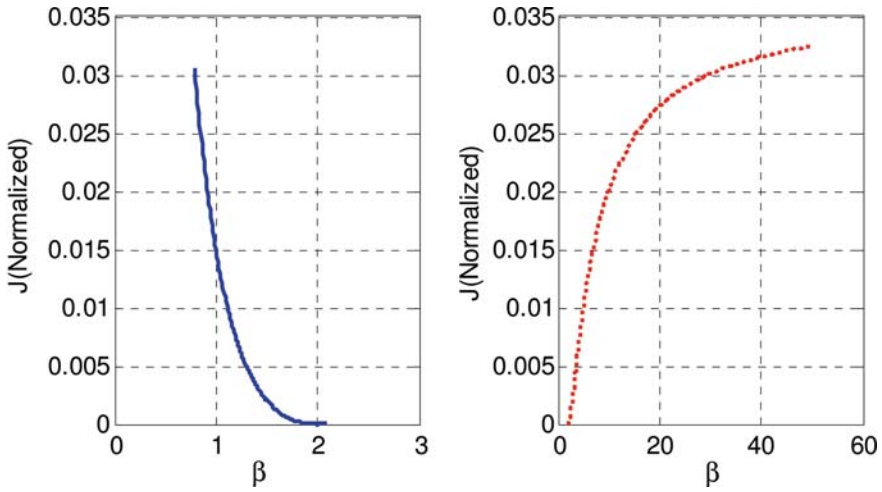

Figure 9. Showing ambiguity in inverse mapping of $J - \beta$ relation in estimation of shape parameter. For the same value of negentropy two values of shape parameters are possible which correspond to super-Gaussian and sub-Gaussian distribution. However for $\beta \gg 2$, $J - \beta$ relation saturates to very low value of negentropy.

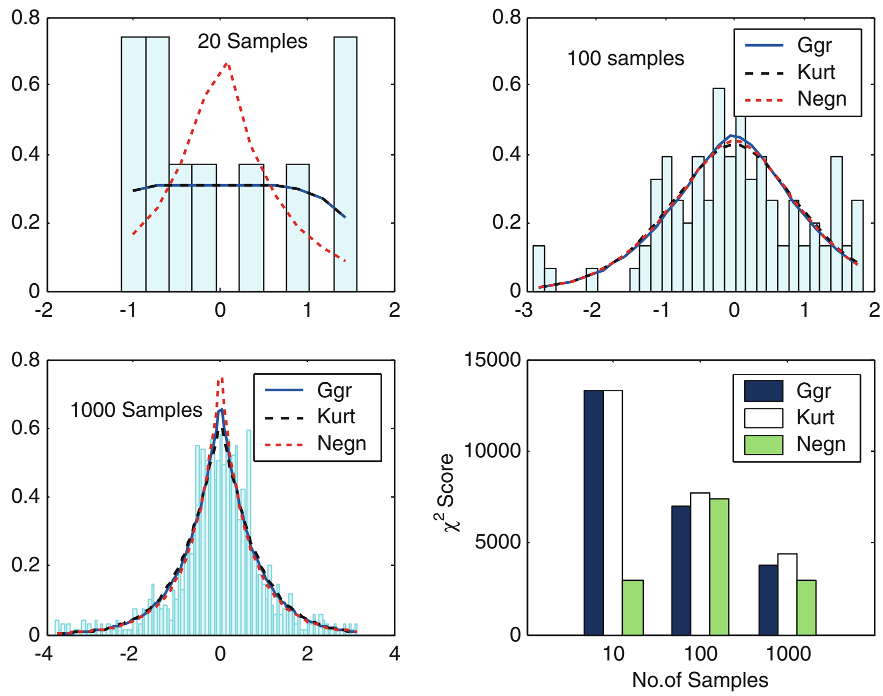

Figure 10. Estimation of shape parameter of machine gun noise for different sample size. Chi-Square scores between GGD with estimated parameters and data are also shown.

References

1. Miller, J. H. and Thomson, J. B.: Detector for discrete-time signal in non-Gaussian noise, *IEEE Transactions on IT*, IT-18 **2** (1972), 241–250.
2. Box, G. E. P. and TIAO, G. C.: *Bayesian Inference in Statistical Analysis*. Addision-Wesley, Reading, MA, 1972.
3. Gharavi, H. and Tabatabai, A.: Sub-band coding of monochrome and color images, *IEEE Transactions on Circuits System*, **35** (1988), 207–214.
4. Keith, A. B. and Thomas, R. F.: On the modeling of DCT and subband image data for compression, *IEEE Transactions IP* **4**(2) (1995), 186–193.
5. Prasad, R. K., Saruwatari, H., and Shikano, K.: Probability distribution of time-series of speech spectral components, *IEICE Transactions Fundamental of Electronics Communications Computer Science* **E87-A**(3) (2004).
6. Prasad, R. K., Saruwatari, H., and Shikano, K.: MAP estimation of Speech Spectral Component under GGD a Priori, In: *Proceedings of the SAPA 2004, Korea*.
7. Sviatoslav, V. and Thierry, P.: Stochastic digital image watermarking, In: *IEEE/ISSPA ISSPA 2001, 6th International Symposium on Signal Processing and its Applications*, Kuala-Lumpur, Malaysia, 2001.
8. Heinz, M., Marcel, J., and George, S. M.: A simple threshold non-linearity for blind signal separation of sub-Gaussian signals, In: *IEEE International Symposium on Circuits and Systems* (2000), 489–492.
9. Choi, S., Chichocki, A., and Amari, S.: Flexible independent component analysis, In: *Proceedings of IEEE Workshop on Neural Network for Signal Processing (NNSP"98)*, 83–92, UK.
10. Everson, R. and Roberts, S.: Independent component analysis: a flexible nonlinearity and decorrelating manifold approach. *Neural Computation* **11**(8) (1999), 1957–1983.
11. Varanasi, M. K. and Aazhang, B.: Parametric generalized Gaussian density estimation, *Journal of Acoustations Society American* **86**(4) (1989), 1404–1415.
12. Mallat, S. G.: A theory of multiresolution signal decomposition: the wavelet representation, *IEEE Transactions Pattern Analysis and Machine Intelligence* **11**(7) (1989), 674–693.
13. Simoncell, E. and Adelson, E.: Noise removal via Bayesian wavelet coring, In: *Proceedings of IEEE International Conference Image Processing* **1**: 379–382, 1996.
14. Gazor, S.: Speech probability distribution, *IEEE, SPL* (2003).
15. Chang, S. G., Yu, B., and Vetterli, M.: Adaptive wavelet thresholding for image denoising and compression, *IEEE Transactions on Image Processing*, **9**(9) (2000), 1532–1546.
16. D'Agostino, R. B and Lee, A. F. S.: Robustness of location estimators under changes of population kurtosis, *Journal of American Statistical Association* **72** (1977), 393–396.
17. Hyvarinen, A., Karuhnen, J., and Oja, E.: *Independent Component Analysis*. Wiley, USA, 2001.
18. Farvardan, N., and James, W. M.: Optimum quantizer performance for a class of non-Gaussian memoryless sources, *IEEE Transactions Information Theory* **IT-30**(3) (1984).
19. Prasad, R. K., Saruwatari, H., and Shikano, K.: Noise Estimation using negentropy based voice activity detector, In: *Proceedings of 47th IEEE Midwest Symposium on Circuits and Systems*, Hiroshima, Japan 149–152.
20. Populis, A., Pillai, S. U.: Probability, random variables and stochastic processes, 4th edn, Mc-Graw Hill, 2002.