ON THE USE OF MIXTURES OF GAUSSIANS AND MIXTURES OF GENERALIZED EXPONENTIALS FOR MODELLING AND CLASSIFICATION OF BIOMEDICAL SIGNALS

C.W. Hesse[1], D. Holtackers[2], T. Heskes[2]

[1]F.C. Donders Centre for Cognitive Neuroimaging, Radboud University Nijmegen, The Netherlands

[2]Department of Information and Knowledge Systems, Radboud University Nijmegen, The Netherlands

Abstract

Mixture density models, particularly those based on the Gaussian distribution, are widely used in machine learning tools for data modeling and classification. Gaussian mixture models have also been used in biomedical signal processing applications involving electrophysiological signals such as the electromyogram (EMG) and electro-encephalogram (EEG). In this paper, we consider a generalization of the Gaussian mixture model, which is based on the generalized exponential distribution. We describe a means of fitting such a model to data and explore its utility for bio-signal analysis.

1 Introduction

Parametric probabilistic models form an important aspect of statistical signal processing and analysis. They are generally useful for signal classification and can be used to infer or detect state changes in the underlying system. Parametric models can facilitate these tasks, because the class-conditional posterior probabilities computed using Bayes' rule then have analytic expressions. Mixtures of Gaussian (MOG) distributions [1] are a versatile tool for parametric modeling of arbitrary probability distributions and have been used to classify EMG [2] and EEG [3] signals. Here, we consider an extension of the MOG density model, based on the generalized exponential distribution, and its use in biomedical signal analysis.

This paper is organized as follows. Section 2 reviews the MOG density model and the expectation maximization (EM) algorithm for estimating its parameters. Section 3 introduces the generalized exponential distribution and describes a modification of the EM algorithm for fitting mixtures of generalized exponentials (MGE). Section 4 discusses model selection criteria useful for determining an optimal number of mixture components and for choosing between different mixture models. Section 5 has examples of how MGE density models can be used to analyze electrophysiological signals. Section 6 has a summary.

2 Mixtures of Gaussians

The probability density function (pdf) of a univariate distribution comprising K component densities (here a mixture of Gaussians) has the general form

$$p(x|\theta_K) = \sum_{k=1}^{K} w_k p(x|\mu_k,\sigma_k), \quad (1)$$

where $\theta_K = [w_k,\mu_k,\sigma_k]_{k=1}^K$ and $w_k = p(k)$ is the probability (weighting) that the data sample was generated by the kth mixture component density, with $0 < w_k < 1$ and $w_1 + w_2 + \ldots + w_K = 1$. In the case of Gaussian (Normal) densities, e.g., [4]

$$p(x|\mu_k,\sigma_k) = \frac{1}{\sigma_k\sqrt{2\pi}} \exp\left(-\frac{(x-\mu_k)^2}{2\sigma_k^2}\right), \quad (2)$$

where μ_k and σ_k reflect the mean and standard deviation of the kth Gaussian. An example pdf for a MOG density is shown in Figure 1.

The total log-likelihood of a MOG density model for N samples of data is therefore

$$L(\theta_K) = \sum_{i=1}^{N} \log \sum_{k=1}^{K} w_k p(x_i|\mu_k,\sigma_k). \quad (3)$$

The standard approach for obtaining the maximum likelihood estimate for the MOG parameters is the expectation maximization (EM) algorithm [5]. The EM algorithm alternates between updating the posterior probabilities that each data point was generated by the kth mixture component, the E-step:

$$h_i^k = \frac{w_k p(x_i|\mu_k,\sigma_k)}{\Sigma_{j=1}^K w_j p(x_i|\mu_j,\sigma_j)}, \quad (4)$$

and weighted maximum likelihood updates of the parameters of each mixture component, the M-step:

$$w_k^* = \tfrac{1}{N}\Sigma_{i=1}^N h_i^k \quad (5)$$

$$\mu_k^* = \frac{\Sigma_{i=1}^N h_i^k x_i}{\Sigma_{j=1}^N h_j^k} \quad (6)$$

$$\sigma_k^* = \sqrt{\frac{\Sigma_{i=1}^N h_i^k (x_i-\mu_k^*)^2}{\Sigma_{j=1}^N h_j^k}}. \quad (7)$$

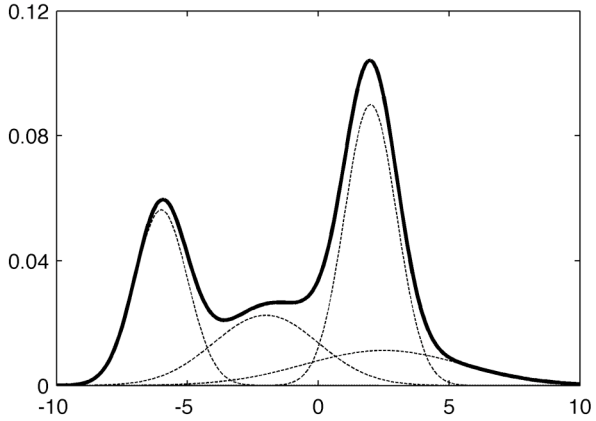

Figure 1. The probability density function of a mixture of several Gaussian distributions.

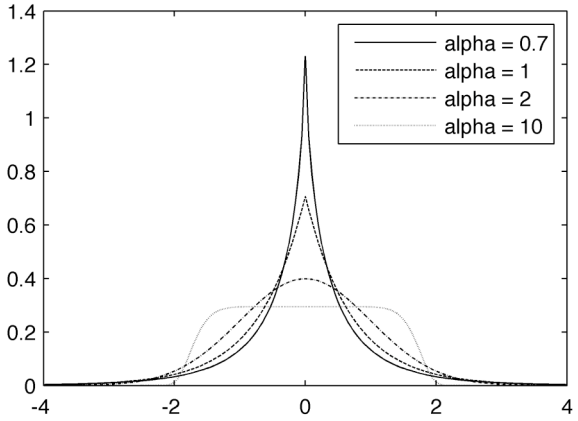

Figure 2. The probability density function of the exponential power (or generalized exponential) distribution for different values of α.

The EM algorithm (especially in cases of high-dimensional multivariate Gaussian mixtures) has a tendency to converge to spurious solutions when there are singularities in the log-likelihood function due to small sample sizes, outliers, repeated data points or rank deficiencies leading to "variance collapse"; these problems along with possible preventative measures have been described by several authors [6]-[8]. However, these issues are beyond the scope of this paper, which is focused on applications with large samples of univariate data.

3 Mixtures of Generalized Exponentials

A generalized formulation of the pdf of the univariate Gaussian distribution (2) where the power in the exponential term can take strictly positive values other than 2 was proposed by [9] as the exponential power distribution, whose pdf is symmetric about the mean and has a general expression, e.g. [10],

$$p(x|\mu,\sigma,\alpha) = \frac{c_1(\alpha)}{\sigma}\exp\left(-c_2(\alpha)\left|\frac{x-\mu}{\sigma}\right|^{\alpha}\right), \qquad (8)$$

where μ is the mean and $\sigma > 0$ is a scale parameter corresponding to the standard deviation. The shape parameter $\alpha > 0$ reflects kurtosis and controls the degree of deviation from normality. Thus, as shown in Figure 2, for $\alpha = 2$ the distribution is Gaussian, for $\alpha = 1$ it is Laplacian, as $\alpha \to \infty$ it becomes uniform, and as $\alpha \to 0$ it approaches a delta function at μ. The scaling terms $c_1(\alpha)$ and $c_2(\alpha)$ are

$$c_1(\alpha) = \frac{\alpha\Gamma(3/\alpha)^{1/2}}{2\Gamma(1/\alpha)^{3/2}} \qquad (9)$$

$$c_2(\alpha) = \left[\frac{\Gamma(3/\alpha)}{\Gamma(1/\alpha)}\right]^{\alpha/2}. \qquad (10)$$

The exponential power distribution has been variously referred to as the generalized Gaussian, generalized Laplacian and generalized exponential distribution (the preferred term here) [9]-[12].

The sample mean and standard deviation can be taken as method of moments estimates (MME) for the location and scale parameters μ and σ. However, the relationship between the sample kurtosis, as measured by $\kappa = E(x-\mu)^4/\sigma^4 - 3$, and the shape parameter α is more complex

$$\kappa = \frac{\Gamma(5/\alpha)\Gamma(1/\alpha)}{[\Gamma(3/\alpha)]^2} - 3. \qquad (11)$$

Nevertheless, one could use numerical optimization approaches to estimate to α with respect to kurtosis, e.g., a golden section search or a look-up table mapping the relation in (11), which is independent of μ and σ. Maximum likelihood estimates (MLE) for μ, σ and α require the solution of a system of three coupled nonlinear equations, e.g. [11].

By adding parameter subscript indices and substituting equation (8) for the Gaussian pdf term in equation (1) we obtain a general expression for the pdf of a mixture of generalized exponentials (MGE)

$$p(x|\theta_K) = \sum_{k=1}^{K} w_k\, p(x|\mu_k,\sigma_k,\alpha_k), \qquad (12)$$

where the parameters are $\theta_K = [w_k,\mu_k,\sigma_k,\alpha_k]_{k=1}^{K}$ and the mixture component weights w_k have the same constraints as before. Figure 3 shows the pdf of an MGE with 4 components. The MGE parameters can be estimated using a suitably modified (approximate) version of the EM algorithm, where the E-step is

$$h_i^k = \frac{w_k\, p(x_i|\mu_k,\sigma_k,\alpha_k)}{\Sigma_{j=1}^{K} w_j\, p(x_i|\mu_j,\sigma_j,\alpha_j)}. \qquad (13)$$

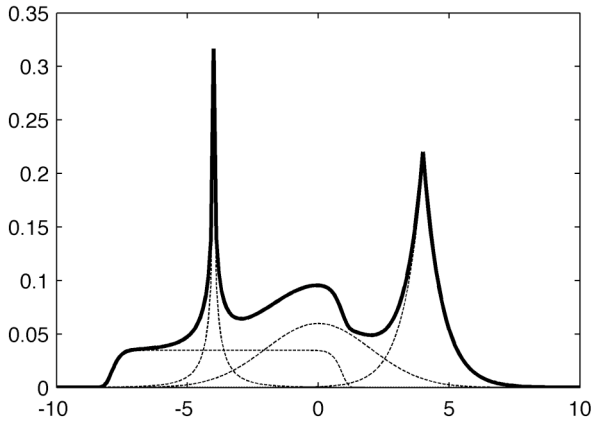

Figure 3. The probability density function of a mixture of several generalized exponential distributions (each with different kurtosis).

The M-step update for w_k follows equation (5); precise M-step updates for μ_k, σ_k and α_k require numerical optimization of the weighted log-likelihood function with respect to each of the parameters

$$(\mu_k^*,\sigma_k^*,\alpha_k^*) = \arg\max_{(\mu_k \sigma_k \alpha_k)} \frac{\Sigma_{i=1}^N h_i^k \log p(x_i|\mu_k,\sigma_k,\alpha_k)}{\Sigma_{j=1}^N h_j^k} \quad (14)$$

A simple alternative is to use approximate M-step updates based on (weighted) method of moments estimates of μ_k and σ_k as in equations (6) and (7), and the kurtosis of the kth mixture component

$$\kappa_k^* = \frac{\Sigma_{i=1}^N h_i^k (x_i - \mu_k^*)^4}{(\sigma_k^*)^4 \Sigma_{j=1}^N h_j^k} - 3, \quad (15)$$

so that α_k^* can be determined with respect to κ_k^* using the relationship in (11). Note that numerical optimization using kurtosis is computationally far more efficient than numerical optimization of (14), which may be an important factor for some practical applications. For the examples reported in section 5 we used numerical optimization of (14).

4 Model Selection Criteria

Given the tools for estimating the parameters of MOG or MGE models comprising a specified number of mixture components K, a remaining problem is to estimate the "correct" model order $\hat{K} \approx K$ from the data sample. Since the log-likelihood for these models can readily be computed, it makes sense to use a standard likelihood-based model selection criterion such as Akaike's information criterion (AIC)

$$AIC(\theta_K) = -2L(\theta_K) + 2P, \quad (16)$$

or the Bayes information criterion (BIC)

$$BIC(\theta_K) = -2L(\theta_K) + P \log N, \quad (17)$$

which both aim to optimally balance model accuracy and parsimony by penalizing for excessive numbers of parameters P; BIC also accounts for sample size. For MOG models $P = 2K + K - 1$ and for MGE models $P = 3K + K - 1$. Thus, an optimal choice for $\hat{K} \approx K$ is the one which minimizes (16) or (17). Such model selection criteria also facilitate comparisons between MOG and MGE models fitted to the same data.

5 Biomedical Signal Processing Applications

We investigated the relative utility of MOG and MGE densities for modeling and analysis of single channel electrophysiological signals such as EMG, EOG and EEG. We were particularly interested to see whether the MGE distribution would yield a more optimal description (i.e. more accurate and/or parsimonious) of the data than the MOG model, when comparing the number of mixture components determined for each model using BIC.

We used data from the experiment described in [13] and selected two trials from the condition involving right hand movement for analysis. We fitted MEG and MGE models to the signals from the right extensor EMG channel, the vertical EOG, and the EEG electrode C3. In each case we found that the optimal MGE model has a lower BIC than the optimal MOG model, as shown in Figures 4, 5 and 6.

We also wanted to illustrate how mixture density models can be used for signal classification (e.g., event detection) when different mixture components of the model can be identified with different states of the measured system, such as muscle contraction (movement) in EMG, eye-blinks in the EOG, and changes in brain state during a behavioral task (movement) reflected in the EEG.

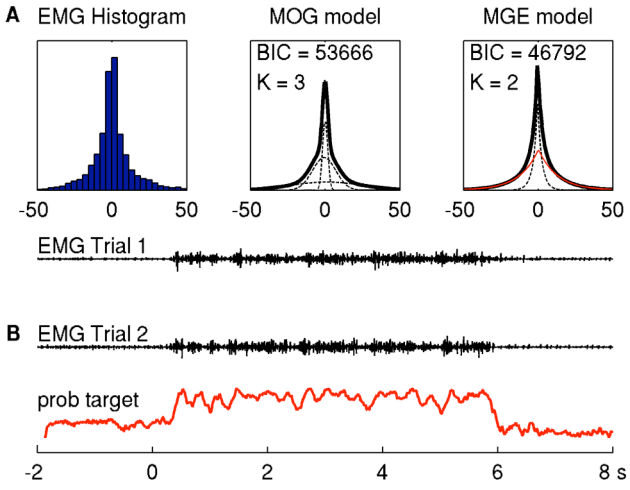

Figure 4. Reference EMG signal with MOG and MGE models (A). The "activation" probability of one MGE (red) mixture component on EMG from a second trial tracks muscle contraction (B).

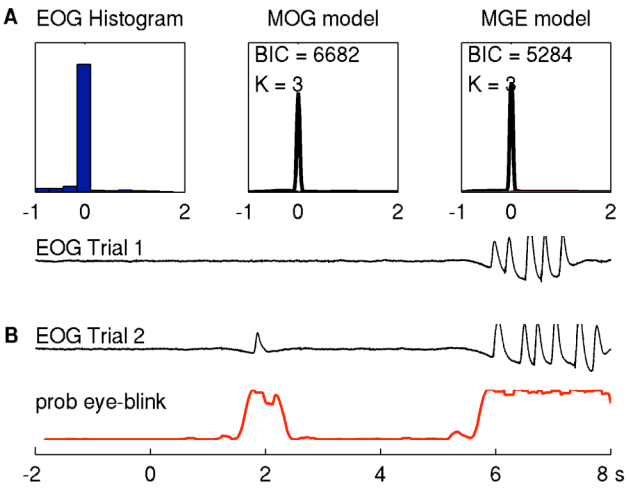

Figure 5. Reference EOG signal with MOG and MGE models (A). The "activation" probability of one MGE (red) mixture component on EOG from a second trial tracks eye-blinks (B).

Using the MGE models from the first trial, we (manually) identified the dominant component during the movement period for EMG and EEG, and for eye-blinks in the EOG. We then analyzed the signals from the second trial by averaging the component posterior probabilities computed using equation (13) over a short moving window. This gives a relatively smooth estimate of the probability that the signal reflects components of interest.

6 Summary

We introduced a mixture density model based on the generalized exponential distribution as an analysis tool for biomedical signal processing. We described the EM algorithm for maximum likelihood parameter estimation and showed that the MGE model can provide a more optimal description of data than a MOG model and is useful for signal classification.

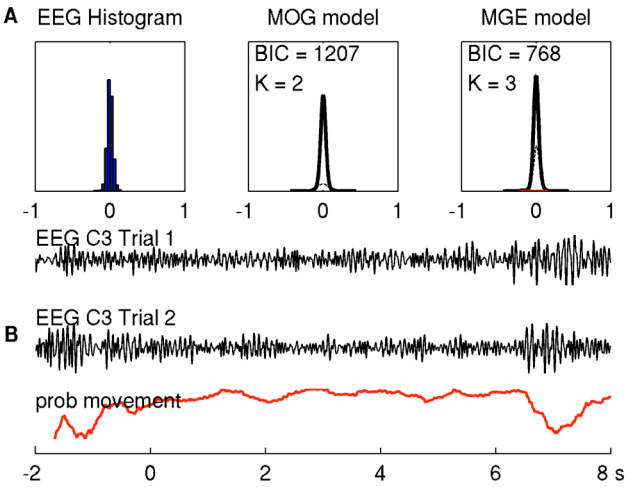

Figure 6. MOG and MGE models for reference EEG (A). The "activation" probability relates to movement related activity (B).

References

[1] Duda, R.O. and Hart, P.E., *Pattern classification and scene analysis*, Wiley, 1973.

[2] Huang, Y., Engelhart, K.B., Hudins, B., and Chan, A.D.C., A Gaussian mixture model based classification scheme for myoelectric control of powered upper limb prostheses, *IEEE Trans. Biomed. Eng.*, 52(11), 1801-1811, 2005.

[3] Tsuji, T., Fukuda, O., Ichinobe, H., and Kaneko, M., A log-linearized Gaussian mixture network and its application to EEG pattern classification. *IEEE Transactions on Systems, Man and Cybernetics, Part C*, 29(1), 60-72, 1999.

[4] Roweis, S. and Ghahramani, Z., A unifying review of linear Gaussian models, *Neural Computation*, 11, 305-345, 1999.

[5] Dempster, A.P., Laird, N.M. and Rubin, D.B., Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society B*, 39, 1-38, 1977.

[6] Redner, R.A. and Walker, H.F., Mixture densities, maximum likelihood and the EM algorithm, *SIAM Review*, 26, 195-239, 1984.

[7] Ormoneit, D. and Tresp, V., Averaging, maximum penalized likelihood and Bayesian estimation for improving Gaussian mixture probability density estimates, *IEEE Trans. on Neural Networks*, 9(4), 639-650, 1998.

[8] Archambeau, C., Lee, J.A. and Verleysen, M., On convergence problems of the EM algorithm for finite Gaussian mixtures, *Proc. Euro. Symp. on Artificial Neural Networks (ESANN 2003)*, Bruges, Belgium, 23-25 April 2003, pp. 99-106.

[9] Box, G. and Tiao, G., *Bayesian inference in statistical analysis*, John Wiley and Sons, 1973.

[10] Chiappa, S. and Barber, D., EEG classification using generative independent component analysis, *Neurocomputing*, 69, 769-777, 2006.

[11] Everson, R. and Roberts, S., Independent component analysis: a flexible non-linearity and decorrelating manifold approach, *Neural Computation*, 11(8), 1957-1983, 1999.

[12] Lee, T.-W. and Lewicki, M.S., The generalized Gaussian mixture model using ICA, *in Proc. Int. Conf. on BSS and ICA (ICA 2000)*, Helsinki, Finland, June 2000, pp. 239-244.

[13] Hesse, C.W., Brouwer, C.L. and Jensen, O., Identification of task-related brain activity in the multichannel EEG through joint diagonalization of single-trial cross-covariance matrices. *Proc. 1st Ann. Symp. IEEE EMBS Benelux Chapter*, Brussels, Belgium, 7-8 December 2006.