A GENERALIZED MULTIVARIATE LOGISTIC MODEL AND EM ALGORITHM BASED ON THE NORMAL VARIANCE MEAN MIXTURE REPRESENTATION

Jason A. Palmer[1], *Ken Kreutz-Delgado*[2], *and Scott Makeig*[1]

[1]Swartz Center for Computational Neuroscience
Institute for Neural Computation
[2]Dept. of Electrical and Computer Engineering
University of California San Diego

ABSTRACT

We present an EM algorithm for Maximum Likelihood estimation of the location, scale, and skew, and shape parameters of the z distribution, also known as the generalized logistic function (type IV). We use the Barndorff-Nielsen, Kent, and Sørensen representation of the z distribution as a Gaussian location-scale mixture to derive an EM algorithm for estimating the location, scale, skew, and shape parameters. We use a variational bound on the likelihood function to determine a monotonically converging closed form update for the skew (or drift) parameter. The algorithm also extends naturally to multivariate GLSM estimation using the Kolmogorov-Smirnov mixing density in odd dimensions.

Index Terms— Generalized Logistic, Gaussian Location-Scale Mixtures, Multivariate Logistic, Quasiparametric density estimation

1. INTRODUCTION

Fisher's z distribution has the form,

$$p(s) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\exp(s)^{\alpha}}{(1+\exp(s))^{\alpha+\beta}}, \quad \alpha, \beta > 0 \tag{1}$$

with the location, scale family defined by $z(s; \mu, \sigma, \alpha, \beta) = \sigma^{-1} p(\sigma^{-1}(s-\mu))$. This density arises as from the so-called z transformation $\log(X/(1-X))$ when $X \sim B(\alpha, \beta)$ [1]. It also arises from the F distribution as the density of $\log X$ for $X \sim F(\alpha/2, \beta/2)$, and as the distribution of difference of log gamma random variables, $\log X_1 - \log X_2$ with X_1, X_2 independent Gamma distributed [2].

The z distribution has log linear tails, with the slope of the left and right tails being proportional to α and $-\beta$ respectively. The shape varies from Laplacian (double-exponential) to Gaussian with the magnitude of (α, β). It was first described by Fisher [1, 3], and was studied extensively by Prentice [2, 4]. It is also known as the generalized logistic function (Type IV) [5, p.142], and the exponential generalized beta distribution of the second kind (EGB2) [6]. The z distribution is one of the few closed form density models that allow control of skew,[1] and is thus very useful in parametric modeling.

The z distribution was shown to be a Normal Variance Mean Mixture by Barndorff-Nielsen et al. [7]. That is, a z distributed random variable S can be represented in the generative form,

$$S = \xi^{1/2} Z + \xi\, \theta \tag{2}$$

where Z is standard Normal, ξ is a non-negative random variable and independent of Z, and $\theta \in \mathbb{R}$ is a skew or drift parameter. This representation raises the possibility of using an EM algorithm as with (symmetric) GSMs, as first suggested by Dempster, Laird, and Rubin [8]. Prentice [4] considered a standard Newton-Raphson type method for Maximum Likelihood parameter estimation, but he notes that the method suffers from complications in terms of the shape parameters. The EM algorithm has the virtue of being monotonically convergent and only involving first derivatives, as well as extending naturally to higher dimensions.

In §2 we consider the mixing density associated with the z distribution, and the conditions under which the model is extensible to higher dimensions, i.e. for Gaussian $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$, we consider the random vector,

$$\mathbf{s} = \xi^{1/2}\mathbf{z} + \xi\boldsymbol{\theta} + \boldsymbol{\mu} \tag{3}$$

In §3 we derive the complete log likelihood associated with the EM algorithm, along with posterior mixing variable expectations. In §4 we derive the parameter updates, first considering the straightforward location and scale updates μ and σ, and then deriving a variational bound related to the θ objective leading to closed form (constrained) maximum. In §5 we derive the updates arising when this model is embedded in a finite mixture model EM framework.

2. MIXING DENSITY AND MULTIVARIATE DEPENDENT MODELS

Although the mixing density can only be expressed in series form, we don't need the density itself for the EM algorithm. Let $f(\xi; \delta)$ be the mixing density of the symmetric Generalized Logistic function with shape parameter $\alpha = \beta = \delta$. Then for $s = \xi^{1/2} z + \mu$, with $z \sim \mathcal{N}(0, \sigma^2)$,

$$\begin{aligned} p_1(s\,; \delta, \mu, \sigma) &= \int_0^{\infty} \frac{1}{\sigma\sqrt{2\pi\xi}} \exp\left(-\frac{1}{2}\frac{(s-\mu)^2}{\sigma^2\xi}\right) f(\xi; \delta)\, d\xi \\ &= \frac{\Gamma(2\delta)}{\Gamma(\delta)^2} \frac{\sigma^{-1}}{\left(2\cosh\left(\frac{1}{2}\sigma^{-1}(s-\mu)\right)\right)^{2\delta}} \end{aligned} \tag{4}$$

Now, the moment generating function $E\{\exp(\xi t)\} \triangleq \varphi(t)$ of the mixing density is given by,

$$\varphi_1(t; \delta) = \frac{\Gamma\big(\delta + \sqrt{2\,t}\big)\Gamma\big(\delta - \sqrt{2\,t}\big)}{\Gamma(\delta)^2}, \quad t < \tfrac{1}{2}\delta^2$$

We can thus construct a modified form of the mixing density for any $\gamma < \frac{1}{2}\delta^2$,

$$f(\xi; \delta, \gamma) = \frac{\exp(\gamma\xi)}{\varphi(\gamma; \delta)} f(\xi; \delta)$$

This research was partially funded by the Swartz Foundation.

[1]Another being the Generalized Hyperbolic density [7], which involves the Bessel K function.

In particular we shall take $\gamma = \frac{1}{2}\|\boldsymbol{\theta}\|^2_{\Sigma^{-1}}$ in the sequel. The moment generating function of this modified mixing density is then,

$$\varphi(t\,;\delta,\boldsymbol{\theta},\boldsymbol{\Sigma}) = \frac{\Gamma\Big(\delta+\sqrt{\|\boldsymbol{\theta}\|^2_{\Sigma^{-1}}+2\,t}\,\Big)\Gamma\Big(\delta-\sqrt{\|\boldsymbol{\theta}\|^2_{\Sigma^{-1}}+2\,t}\,\Big)}{\Gamma\big(\delta+\|\boldsymbol{\theta}\|_{\Sigma^{-1}}\big)\Gamma\big(\delta-\|\boldsymbol{\theta}\|_{\Sigma^{-1}}\big)} \qquad (5)$$

Using this moment generating function to determine the first moment, we find,

$$E\{\xi\,;\delta,\boldsymbol{\theta},\boldsymbol{\Sigma}\} = \frac{\Psi(\delta+\|\boldsymbol{\theta}\|_{\Sigma^{-1}})-\Psi(\delta-\|\boldsymbol{\theta}\|_{\Sigma^{-1}})}{\|\boldsymbol{\theta}\|_{\Sigma^{-1}}} \qquad (6)$$

As $\|\boldsymbol{\theta}\|_{\Sigma^{-1}} \to 0$, this tends to $2\,\Psi'(\delta)$. Similarly we find that the variance,

$$\mathrm{Var}(\xi\,;\delta,\boldsymbol{\theta},\boldsymbol{\Sigma}) = \frac{\Psi'(\delta-\|\boldsymbol{\theta}\|_{\Sigma^{-1}})+\Psi'(\delta+\|\boldsymbol{\theta}\|_{\Sigma^{-1}})-E\{\xi\,;\delta,\boldsymbol{\theta},\boldsymbol{\Sigma}\}}{\|\boldsymbol{\theta}\|^2_{\Sigma^{-1}}}$$

which tends to $\frac{2}{3}\,\Psi'''(\delta)$ as $\|\boldsymbol{\theta}\|_{\Sigma^{-1}} \to 0$. The mean and variance can be used to normalize the mixing scale and variance.

Now if we consider the density of the random vector defined in (3), we have, $p(\mathbf{s}) = \int \mathcal{N}(\mathbf{s}\,;\boldsymbol{\mu}+\xi\boldsymbol{\theta},\xi\boldsymbol{\Sigma})f(\xi;\delta,\gamma)d\xi$,

$$p(\mathbf{s}) = |\det\boldsymbol{\Sigma}|^{-1/2}\exp\big(\boldsymbol{\theta}^T\boldsymbol{\Sigma}^{-1}(\mathbf{s}-\boldsymbol{\mu})\big)\int_0^\infty (2\pi)^{-n/2}\times \xi^{-n/2}\,\exp\big(-\tfrac{1}{2}\xi^{-1}\|\mathbf{s}-\boldsymbol{\mu}\|^2_{\Sigma^{-1}}-\tfrac{1}{2}\xi\,\|\boldsymbol{\theta}\|^2_{\Sigma^{-1}}\big)f(\xi;\delta,\gamma)\,d\xi$$

With $\gamma = \frac{1}{2}\|\boldsymbol{\theta}\|^2_{\boldsymbol{\Sigma}^{-1}}$, the ξ term in the exponent cancels and we are left with,

$$p(\mathbf{s}\,;\delta,\boldsymbol{\mu},\boldsymbol{\theta},\boldsymbol{\Sigma}) = |\det\boldsymbol{\Sigma}|^{-1/2}\frac{\exp\big(\boldsymbol{\theta}^T\boldsymbol{\Sigma}^{-1}(\mathbf{s}-\boldsymbol{\mu})\big)}{\varphi_1\big(\frac{1}{2}\|\boldsymbol{\theta}\|^2_{\Sigma^{-1}};\delta\big)}\times \int_0^\infty (2\pi)^{-n/2}\,\xi^{-n/2}\exp\big(-\tfrac{1}{2}\xi^{-1}\|\mathbf{s}-\boldsymbol{\mu}\|^2_{\Sigma^{-1}}\big)f(\xi;\delta)\,d\xi$$

The integral is seen to have a similar form to that in (4) except for the higher degree of the factor $\xi^{-n/2}$. If we note however the effect of repeated applications of the operator $-\frac{1}{s}\frac{d}{ds}$ under the integral in (4), we see that in the case of n odd, $p_n(\mathbf{s})$ can be written,

$$p_n(\mathbf{s}\,;\delta,\boldsymbol{\mu},\boldsymbol{\theta},\boldsymbol{\Sigma}) = \frac{|\det\boldsymbol{\Sigma}|^{-1/2}\exp\big(\boldsymbol{\theta}^T\boldsymbol{\Sigma}^{-1}(\mathbf{s}-\boldsymbol{\mu})\big)}{B\big(\,\delta-\|\boldsymbol{\theta}\|_{\Sigma^{-1}},\,\delta+\|\boldsymbol{\theta}\|_{\Sigma^{-1}}\big)} \times\left(\frac{-1}{2\pi y}\frac{d}{dy}\right)^{\frac{n-1}{2}}\big(2\cosh\big(\tfrac{1}{2}y\big)\big)^{-2\delta}\Big|_{y=\|\mathbf{s}-\boldsymbol{\mu}\|_{\Sigma^{-1}}} \qquad (7)$$

Define,

$$C_k(y\,;\delta) \triangleq \left(\frac{-1}{y}\frac{d}{dy}\right)^k\big(2\cosh\big(\tfrac{1}{2}y\big)\big)^{-2\delta}$$

Writing out the result in the one, three, and five dimensional cases, we have $C_k(y\,;\delta) = \big(2\cosh(\frac{1}{2}y\big)^{-2\delta}P_k(y\,;\delta)$, where,

$$\begin{aligned} P_0(y) &= 1 \\ P_1(y\,;\delta) &= \frac{\delta\tanh(\frac{1}{2}y)}{y} \\ P_2(y\,;\delta) &= \frac{\delta\tanh(\frac{1}{2}y)}{y}\left(\delta\frac{\cosh(y)-1}{y\sinh(y)}+\frac{1-y/\sinh(y)}{y^2}\right) \end{aligned}$$

For the densities then, we have for $n = 1$,

$$p_1(s\,;\delta,\mu,\theta,\sigma) = \frac{\Gamma(2\,\delta)}{\Gamma\big(\delta-\sigma^{-1}\theta\big)\,\Gamma\big(\delta+\sigma^{-1}\theta\big)}\,\frac{\sigma^{-1}\exp\big(\theta\,\sigma^{-1}(s-\mu)\big)}{\big(2\cosh(\frac{1}{2}\sigma^{-1}(s-\mu)\big)^{2\delta}}$$

and for $n = 2k+1$, $k = 1, 2, \ldots$,

$$p_n(\mathbf{s}\,;\delta,\boldsymbol{\mu},\boldsymbol{\theta},\boldsymbol{\Sigma}) = \frac{\exp\big(\boldsymbol{\theta}^T\boldsymbol{\Sigma}^{-1}(\mathbf{s}-\boldsymbol{\mu})\big)}{B\big(\,\delta-\|\boldsymbol{\theta}\|_{\Sigma^{-1}},\,\delta+\|\boldsymbol{\theta}\|_{\Sigma^{-1}}\big)}\,\frac{C_k\big(\|\mathbf{s}-\boldsymbol{\mu}\|_{\Sigma^{-1}};\delta\big)}{(2\pi)^k(\det\boldsymbol{\Sigma})^{1/2}}$$

where $0 \le \|\boldsymbol{\theta}\|_{\Sigma^{-1}} < \delta$.

3. COMPLETE LOG LIKELIHOOD AND POSTERIOR MOMENTS

Define the sample mean $\mathbf{m} \triangleq N^{-1}\sum_{t=1}^N \mathbf{s}_t$. The complete log likelihood $\log\prod_t p(\mathbf{s}_t|\xi_t)p(\xi_t)$ of N i.i.d. samples, scaled by N^{-1}, is,

$$-\tfrac{1}{2}\log\det\boldsymbol{\Sigma} + \boldsymbol{\theta}^T\boldsymbol{\Sigma}^{-1}(\mathbf{m}-\boldsymbol{\mu}) - \log\varphi_1\big(\tfrac{1}{2}\|\boldsymbol{\theta}\|^2_{\Sigma^{-1}};\delta\big) - N^{-1}\sum_{t=1}^N \tfrac{1}{2}\,\xi_t^{-1}\|\mathbf{s}_t-\boldsymbol{\mu}\|^2_{\Sigma^{-1}} \qquad (8)$$

Note that we only need to compute the posterior expectation of ξ_t^{-1}. But using the differentiation trick again, we have,

$$\begin{aligned} E\{\xi_t^{-1}|\mathbf{s}_t\} &= \frac{\int \xi_t^{-1}p(\mathbf{s}_t|\xi_t)p(\xi_t)d\xi_t}{p(\mathbf{s}_t)} \\ &= \frac{V^{k+1}p_1(y\,;\delta)}{V^k p_1(y\,;\delta)}\bigg|_{y=\|\mathbf{s}_t-\boldsymbol{\mu}\|_{\Sigma^{-1}}} = \frac{P_{k+1}\big(\|\mathbf{s}_t-\boldsymbol{\mu}\|_{\Sigma^{-1}};\delta\big)}{P_k\big(\|\mathbf{s}_t-\boldsymbol{\mu}\|_{\Sigma^{-1}};\delta\big)} \end{aligned} \qquad (9)$$

In the one dimensional case, we have,

$$E\{\xi_t^{-1}|s_t\,;\delta,\mu,\sigma\} = \frac{\delta\tanh\big(\frac{1}{2}y\big)}{y}\bigg|_{y=\sigma^{-1}(s_t-\mu)}$$

and in $\mathbb{R}^3$, we have,

$$E\{\xi_t^{-1}|\mathbf{s}_t\,;\delta,\boldsymbol{\mu},\boldsymbol{\Sigma}\} = \delta\,\frac{\cosh(y)-1}{y\sinh(y)}+\frac{1-y/\sinh(y)}{y^2}\bigg|_{y=\|\mathbf{s}_t-\boldsymbol{\mu}\|_{\Sigma^{-1}}}$$

In this formulation, the expected value of ξ_t^{-1} given $\mathbf{s}_t$ does not depend on $\boldsymbol{\theta}$ in any dimension n. Let $\nu_t \triangleq E\{\xi_t^{-1}|\mathbf{s}_t\}$. Then the complete log likelihood (8), scaled by N^{-1}, can be written,

$$-\tfrac{1}{2}\log|\boldsymbol{\Sigma}| + \boldsymbol{\theta}^T\boldsymbol{\Sigma}^{-1}(\mathbf{m}-\boldsymbol{\mu}) - \log\varphi_1\big(\tfrac{1}{2}\|\boldsymbol{\theta}\|^2_{\Sigma^{-1}};\delta\big) - N^{-1}\sum_{t=1}^N \tfrac{1}{2}\,\nu_t\,\|\mathbf{s}_t-\boldsymbol{\mu}\|^2_{\Sigma^{-1}} \qquad (10)$$

4. PARAMETER UPDATES

We first consider the updates for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ assuming $\boldsymbol{\theta}$ is given. We then derive an update for $\boldsymbol{\theta}$ given $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. Finally we derive a variational update the shape parameter δ which monotonically increases the log likelihood as part of the EM algorithm.

4.1. Location update

Setting the gradient of (10) with respect to $\boldsymbol{\mu}$ equal to zero, we find that the optimal $\boldsymbol{\mu}$ satisfies,

$$\frac{1}{N}\sum_{t=1}^{N}\nu_t(\mathbf{s}_t-\boldsymbol{\mu})=\boldsymbol{\theta}$$

Defining $\hat{\nu}\triangleq N^{-1}\sum_{t=1}^{N}\nu_t$, and $\mathbf{m}_\nu\triangleq(N\hat{\nu})^{-1}\sum_{t=1}^{N}\nu_t\mathbf{s}_t$, we have,

$$\boldsymbol{\mu}=\mathbf{m}_\nu-\hat{\nu}^{-1}\boldsymbol{\theta} \tag{11}$$

4.2. Drift update

We would like to maximize the posterior likelihood (10) over $\boldsymbol{\theta}$ by solving,

$$\max_{\boldsymbol{\theta}}\,(\mathbf{m}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}-\log\big(\Gamma(\delta+\|\boldsymbol{\theta}\|_{\boldsymbol{\Sigma}^{-1}})\Gamma(\delta-\|\boldsymbol{\theta}\|_{\boldsymbol{\Sigma}^{-1}})\big)$$

The function $h(t)=\log(\Gamma(\delta+t)\Gamma(\delta-t))$ is increasing and unbounded above on the interval $[\,0,\delta)$. While the convexity of function ensures a unique solution, we can only formulate the solution in terms of inverse Digamma or Psi functions, which are not readily computable. We can however bound the cost function variationally in terms of a similar function that is tractable, to derive a monotonic coordinate ascent algorithm. Specifically, we have,

$$\log(\Gamma(\delta+t)\Gamma(\delta-t))=\inf_v\; v\log\frac{1}{\delta^2-t^2}-h(v)$$

for a certain function h which it is not necessary to compute, and the optimal v is given by,

$$v=\frac{\Psi(\delta+t)-\Psi(\delta-t)}{2\,t(\delta^2-t^2)^{-1}}=\tfrac{1}{2}\hat{\xi}(\delta^2-t^2) \tag{12}$$

where $t=\|\boldsymbol{\theta}\|_{\Sigma^{-1}}$. We thus have the surrogate problem,

$$\max_{\boldsymbol{\theta}}\,(\mathbf{m}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}+v\log\big(\delta^2-\|\boldsymbol{\theta}\|^2_{\boldsymbol{\Sigma}^{-1}}\big)$$

This solution may be derived in closed form

$$\boldsymbol{\theta}=\left(\frac{\sqrt{v^2+\delta^2\|\mathbf{m}-\boldsymbol{\mu}\|^2_{\Sigma^{-1}}}-v}{\|\mathbf{m}-\boldsymbol{\mu}\|_{\Sigma^{-1}}}\right)\frac{\mathbf{m}-\boldsymbol{\mu}}{\|\mathbf{m}-\boldsymbol{\mu}\|_{\Sigma^{-1}}} \tag{13}$$

4.3. Structure matrix update

We use the same variational representation used to derive the drift update. Let us define $\boldsymbol{\Sigma}_\nu\triangleq(N\hat{\nu})^{-1}\sum_{t=1}^{N}\nu_t(\mathbf{s}_t-\boldsymbol{\mu})(\mathbf{s}_t-\boldsymbol{\mu})^T$. At a stationary point then, $\boldsymbol{\Sigma}$ satisfies,

$$\boldsymbol{\Sigma}=\hat{\nu}\,\boldsymbol{\Sigma}_\nu-(\mathbf{m}-\boldsymbol{\mu})\boldsymbol{\theta}^T-\boldsymbol{\theta}(\mathbf{m}-\boldsymbol{\mu})^T+\frac{2v}{\delta^2-\|\boldsymbol{\theta}\|^2_{\Sigma^{-1}}}\boldsymbol{\theta}\boldsymbol{\theta}^T \tag{14}$$

This equation can be solved for $t\triangleq\|\boldsymbol{\theta}\|^2_{\Sigma^{-1}}$ by using the Woodbury identity three times to invert (14).

$$\boldsymbol{\Sigma}^{-1}=\mathbf{A}^{-1}-\frac{2v\mathbf{A}^{-1}\boldsymbol{\theta}\boldsymbol{\theta}^T\mathbf{A}^{-1}}{\delta^2-t+2v\,\boldsymbol{\theta}^T\mathbf{A}^{-1}\boldsymbol{\theta}} \tag{15}$$

This leads to the update,

$$t=\frac{\delta^2+(2v+1)t_0-\sqrt{\big(\delta^2+(2v+1)t_0\big)^2-4\delta^2t_0}}{2} \tag{16}$$

where we define,

$$a\triangleq\boldsymbol{\theta}^T(\hat{\nu}\boldsymbol{\Sigma}_\nu)^{-1}\boldsymbol{\theta},\quad b\triangleq\boldsymbol{\theta}^T(\hat{\nu}\boldsymbol{\Sigma}_\nu)^{-1}\mathbf{d},\quad d\triangleq\mathbf{d}^T(\hat{\nu}\boldsymbol{\Sigma}_\nu)^{-1}\mathbf{d}$$

and

$$t_0\triangleq\boldsymbol{\theta}^T\mathbf{A}^{-1}\boldsymbol{\theta}=\frac{a}{(1-b)^2-ad} \tag{17}$$

4.4. Shape parameter update

Note that $\log p(\mathbf{s}\,;\delta)$ has the general form,

$$G(\delta)=F(\delta)-2\,\delta\big\langle\log\big(2\cosh(\tfrac{1}{2}\|\mathbf{s}_t-\boldsymbol{\mu}\|_{\Sigma^{-1}})\big)\big\rangle_N+\text{const.}$$

where $F_k(\delta)=\log P_k-\log\varphi_1\big(\tfrac{1}{2}\|\boldsymbol{\theta}\|^2_{\boldsymbol{\Sigma}^{-1}};\delta\big)$, i.e.,

$$F_k(\delta)\;=\;\log\frac{\Gamma(2\,\delta)\,\exp\big\langle\log P_k(y_t\,;\delta)\big\rangle_N}{\Gamma\big(\delta+\|\boldsymbol{\theta}\|_{\Sigma^{-1}}\big)\Gamma\big(\delta-\|\boldsymbol{\theta}\|_{\Sigma^{-1}}\big)},\quad\delta>\|\boldsymbol{\theta}\|_{\Sigma^{-1}}$$

We note that F is increasing and concave for $\delta>\|\boldsymbol{\theta}\|$, but F is also convex with respect to the logarithm $\log(\delta^2-\|\boldsymbol{\theta}\|^2)$. Thus,

$$F_k(\delta)=\sup_u\;u\log\big(\delta^2-\|\boldsymbol{\theta}\|^2_{\Sigma^{-1}}\big)-H(u)$$

for a certain relative conjugate function $H(u)$, where the unique optimal u is given by,

$$u=(2\,\delta)^{-1}\big(\delta^2-\|\boldsymbol{\theta}\|^2_{\Sigma^{-1}}\big)F_k'(\delta) \tag{18}$$

where we have $F_0'(\delta)=2\Psi(2\delta)-\Psi(\delta+\|\boldsymbol{\theta}\|)-\Psi(\delta-\|\boldsymbol{\theta}\|)$, $F_1'(\delta)=F_0'(\delta)+\delta^{-1}$. If we define,

$$\hat{\eta}=2\log 2+N^{-1}\sum_{t=1}^{N}2\log\cosh\big(\tfrac{1}{2}\|\mathbf{s}_t-\boldsymbol{\mu}\|_{\Sigma^{-1}}\big) \tag{19}$$

then we have the surrogate problem for δ,

$$\max_{\delta>\|\boldsymbol{\theta}\|_{\boldsymbol{\Sigma}^{-1}}}\;u\log\big(\delta^2-\|\boldsymbol{\theta}\|^2_{\Sigma^{-1}}\big)-\hat{\eta}\,\delta$$

which has the optimum,

$$\delta=\frac{u+\sqrt{u^2+\hat{\eta}^2\|\boldsymbol{\theta}\|^2_{\Sigma^{-1}}}}{\hat{\eta}} \tag{20}$$

5. FINITE MIXTURE MODEL

We can readily formulate a finite mixture model EM algorithm to estimate the model,

$$p\big(\mathbf{s}\,;\{\alpha_j,\delta_j,\boldsymbol{\mu}_j,\boldsymbol{\theta}_j,\boldsymbol{\Sigma}_j\}_{j=1}^{M}\big)=\sum_{j=1}^{M}\alpha_j\,p\big(\mathbf{s}\,;\delta_j,\boldsymbol{\mu}_j,\boldsymbol{\theta}_j,\boldsymbol{\Sigma}_j\big)$$

Define the hidden model index for time t, $j_t\in\{1,\ldots,M\}$ and the hidden model indicator random variables,

$$e_{jt}=\begin{cases}1, & j_t=j\\ 0, & j_t\neq j\end{cases}$$

so that the mixture random variable representation has the form,

$$\mathbf{s}_t=\sum_{j=1}^{M}e_{jt}\big(\boldsymbol{\mu}_j+\xi_{jt}\boldsymbol{\theta}_j+\xi_{jt}^{1/2}\mathbf{z}_{jt}\big) \tag{21}$$

where $\mathbf{z}_{jt} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_j)$, $\xi_{jt} \sim f\big(\xi; \delta_j, \frac{1}{2}\|\boldsymbol{\theta}_j\|^2\big)$, independent, for $t = 1, \ldots, N$, $0 \leq \|\boldsymbol{\theta}_j\| < \delta_j$, $j = 1, \ldots, M$, $\sum_{j=1}^{M} \alpha_j(\boldsymbol{\mu}_j + \hat{\xi}_j \boldsymbol{\theta}_j) = \mathbf{0}$, and $\hat{\xi}_j$ defined by (6).

Define the posterior expectations $\hat{e}_{jt} = E\{e_{jt}|\,\mathbf{s}_t\}$. Then we have the standard mixture model updates,

$$\hat{e}_{jt} = \frac{\alpha_j^{\ell} p(\,\mathbf{s}_t\,; \delta_j, \boldsymbol{\mu}_j, \boldsymbol{\theta}_j, \mathbf{\Sigma}_j)}{\sum_{j'=1}^{M} \alpha_{j'}^{\ell} p(\,\mathbf{s}_t\,; \delta_{j'}, \boldsymbol{\mu}_{j'}, \boldsymbol{\theta}_{j'}, \mathbf{\Sigma}_{j'})} \tag{22}$$

and $\alpha_j^{\ell+1} = \frac{1}{N}\sum_{t=1}^{N} \hat{e}_{jt}$, where we use ℓ to indicate the iteration number. Now defining $\nu_{jt} \triangleq E\{\xi_{jt}^{-1}|\mathbf{s}_t, j_t = j\}$ and $\mathbf{m}_j \triangleq (N\alpha_j^{\ell+1})^{-1}\sum_{t=1}^{N} \hat{e}_{jt}\mathbf{s}_t$, the complete log likelihood is,

$$\sum_{j=1}^{M} \alpha_j^{\ell+1}\Big(\boldsymbol{\theta}_j^T \mathbf{\Sigma}_j^{-1}(\mathbf{m}_j - \boldsymbol{\mu}_j) - \log \varphi\big(\tfrac{1}{2}\|\boldsymbol{\theta}_j\|^2_{\Sigma_j^{-1}}; \delta_j\big)$$
$$- \tfrac{1}{2}\log|\mathbf{\Sigma}_j| - (N\alpha_j^{\ell+1})^{-1}\sum_{t=1}^{N} \tfrac{1}{2}\,\hat{e}_{jt}\nu_{jt}\|\mathbf{s}_t - \boldsymbol{\mu}_j\|^2_{\Sigma_j^{-1}}\Big)$$

Then for the location parameter updates, we have,

$$\boldsymbol{\mu}_j = \mathbf{m}_{j\nu} - \hat{\nu}_j^{-1}\boldsymbol{\theta}_j \tag{23}$$

For the drift parameters updates, define $\mathbf{d}_j \triangleq \mathbf{m}_j - \boldsymbol{\mu}_j$. With the variational parameter estimates $v_j = \frac{1}{2}\hat{\xi}_j\big(\delta_j^2 - \|\boldsymbol{\theta}_j\|^2_{\Sigma_j^{-1}}\big)$, we have,

$$\boldsymbol{\theta}_j = \left(\frac{\sqrt{v_j^2 + \delta_j^2\|\mathbf{d}_j\|^2_{\Sigma_j^{-1}}} - v_j}{\|\mathbf{d}_j\|^2_{\Sigma_j^{-1}}}\right)\mathbf{d}_j \tag{24}$$

For the structure matrix update, let $\mathbf{y}_{jt} \triangleq \mathbf{s}_t - \boldsymbol{\mu}_j$, and define,

$$\mathbf{\Sigma}_{j\nu} \triangleq (N\alpha_j^{\ell+1}\hat{\nu}_j)^{-1}\sum_{t=1}^{N} \hat{\mathbf{e}}_{jt}\nu_{jt}\mathbf{y}_{jt}\mathbf{y}_{jt}^T \tag{25}$$

then $\mathbf{\Sigma}_j$ is updated as in (16). And finally, for the shape parameters, we have,

$$\delta_j = \frac{u_j + \sqrt{u_j^2 + \hat{\eta}_j^2\|\boldsymbol{\theta}_j\|^2}}{\hat{\eta}_j} \tag{26}$$

where $u_j = (2\delta_j)^{-1}(\delta_j^2 - \|\boldsymbol{\theta}_j\|^2)F'(\delta_j)$ and $\hat{\eta}_j = 2\log 2 + 2N^{-1}\sum_{t=1}^{N}\log\cosh\big(\frac{1}{2}\|\mathbf{y}_{jt}\|_{\Sigma_j^{-1}}\big)$

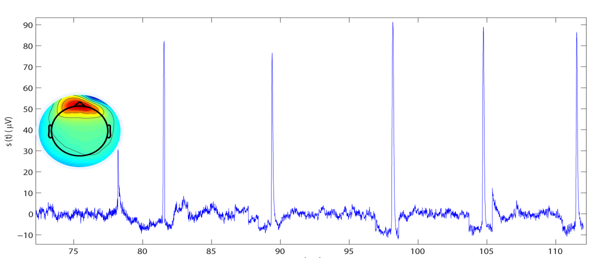

Fig. 1. Topographic map and plot of eyeblink electric potential. Spikes correspond to blinks.

6. EXPERIMENTS

Monotonic convergence of the algorithm without the need to set or modify step sizes has been verified. Figure 1 shows a case study application to the eyeblink muscle signal component typically seen in EEG recordings. In Figure 2 we plot the Fisher's z mixture fits, showing the log histogram, the model, and the mixture components constituting the each model. In Figure 3 we show the determination of model order using the Generalized Likelihood Ratio Test approach, where twice the change in log likelihood is distributed $\chi^2(k)$ where k is the difference in the number of degrees of freedom.

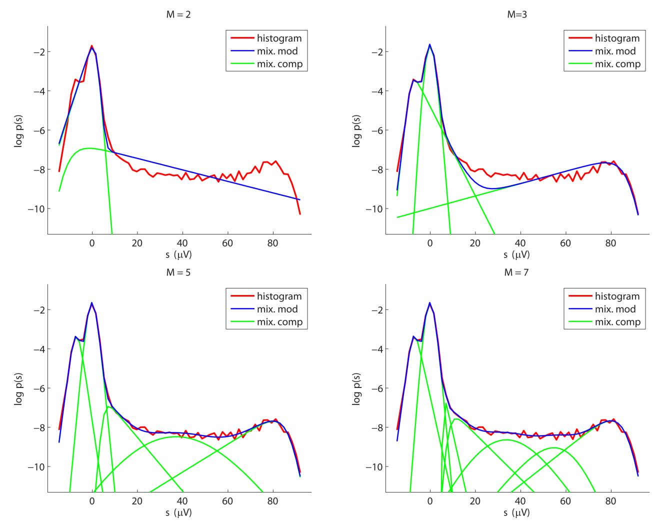

Fig. 2. Fisher's z mixture model fits of eyeblink source distribution with $M = 2, 3, 5, 7$.

7. CONCLUSION

We have derived a multivariate density model generalizing the non-symmetric Generalized Logistic, or Fisher's z distribution, based on the GLSM model of Barndorff-Nielsen et al. [7]. We derived an EM algorithm to update the location, scale, and drift parameters, using novel variational representations, and we used the explicit likelihood formula to derive an EM algorithm to fit a finite mixture model.

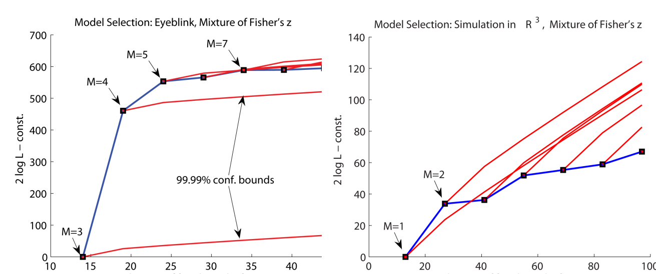

Fig. 3. Plots of twice log likelihood versus model degrees of freedom. Red lines plot boundary of region with models having significant likelihood increase, using the Generalized Likelihood Ratio Test.

REFERENCES

[1] R. A. Fisher, "On the 'probable error' of a coefficient of correlation deduced from a small sample," *Metron*, vol. 1, pp. 3–32, 1921, reprinted in *Collected Papers of R.A. Fisher*, The University of Adelaide, 1971.

[2] R. L. Prentice, "Discrimination among some parametric models," *Biometrika*, vol. 62, no. 3, pp. 607–614, 1975.

[3] R. A. Fisher, "The mathematical distributions used in the common tests of significance," *Econometrica*, vol. 3, pp. 353–363, 1935, reprinted in *Contributions to Mathematical Statistics* by R.A. Fisher, New York: Wiley, 1950.

[4] R. L. Prentice, "A generalization of the probit and logit methods for dose response curves," *Biometrics*, vol. 32, no. 4, pp. 761–768, 1976.

[5] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous univariate distributions, Volume 2*, John Wiley & Sons, Inc.: New York, 1995.

[6] J. B. McDonald, "Parametric models for partially adaptive estimation with skewed and leptokurtic residuals," *Economics Letters*, vol. 37, pp. 273–278, 1991.

[7] O. Barndorff-Nielsen, J. Kent, and M. Sørensen, "Normal variance-mean mixtures and z distributions," *International Statistical Review*, vol. 50, pp. 145–159, 1982.

[8] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Iteratively reweighted least squares for linear regression when errors are Normal/Independent distributed," in *Multivariate Analysis V*, P. R. Krishnaiah, Ed. 1980, pp. 35–57, North Holland Publishing Company.