Dependency Models based on Generalized Gaussian Scale Mixtures and Normal Variance Mean Mixtures

J. A. Palmer, K. Kreutz-Delgado, and S. Makeig *

September 7, 2011

Abstract

We extend the Gaussian scale mixture model of dependent subspace source densities to include non-radially symmetric densities using Generalized Gaussian random variables linked by a common variance. We also introduce the modeling of skew using the Normal Variance-Mean mixture model. We give closed form expressions for likelihoods and parameter updates in the EM algorithm.

1 Introduction

This paper presents a framework for modeling dependency among multiple random variables based on scale mixtures and location-scale mixtures, drawing on the work of Barndorff-Nielsen and others.

1.1 Non-Gaussian Multivariate Densities

A simple way to construct a non-Gaussian multivariate density is to take an arbitrary scalar function of a quadratic form, subject to integrability constraints. This makes the probability density in the neighborhood of a vector a function of the level ellipsoid on which it lies. We thus have the *elliptically contoured* or *spherically symmetric* densities.

*Jason A. Palmer is Postdoctoral Scholar, Swartz Center for Computational Neuroscience, Institute for Neural Computation, University of California San Diego, La Jolla, CA 92093 (E-mail: japalmer@ucsd.edu). Ken Kreutz-Delgado is Professor, Department of Electrical and Computer Engineering, University of California San Diego, La Jolla, CA 92093 (E-mail: kreutz@ucsd.edu). Scott Makeig is Director and Research Scientist, Swartz Center for Computational Neuroscience, Institute for Neural Computation, University of California San Diego, La Jolla, CA 92093 (E-mail: smakeig@ucsd.edu).

A simple extension of this idea is to make the probability density a function of an arbitrary "gauge-like" [14] function with concentric level sets. For example we might take an arbitrary norm $(\sum_i |x_i|^p)^{1/p}$ instead of the radially symmetric two-norm.

We might generalize the idea of p-norm gauges by taking the sum of general nonlinear functions of the components,

$$\gamma(\mathbf{x}) = \sum_{i=1}^{n} g_i(x_i)$$

The density is then, $p(\mathbf{x}) = f(\sum_i g(x_i))$ for some scalar function f. We might ask under what conditions such a density may represent independent random variables, or for which function f can the density $p(\mathbf{x})$ be factorized into a product of marginal densities. The answer follows from basic theory of functions, and is stated in the following theorem.

Theorem 1.1. *If the density of the random vector* $\mathbf{x}$ *has the form* $p(\mathbf{x}) = f\big(\sum_i g_i(x_i)\big)$, *then the random variables* $x_1, \ldots, x_n$ *are independent if and only if* $f(t) \propto \exp(\lambda\, t)$ *for some* λ, *and thus* $p(\mathbf{x}) \propto \prod_i \exp(\lambda\, g_i(x_i))$.

Proof. The x_i are independent if and only if,

$$p(\mathbf{x}) \;=\; f\big(\textstyle\sum_i g_i(x_i)\big) \;=\; \prod_i \phi_i(x_i)$$

for some scalar functions ϕ_i. In other words,

$$f\big(\textstyle\sum_i h_i(x_i) + C\big) \;=\; \prod_i \phi_i(x_i)$$

where $h_i(x_i) = g_i(x_i) - g_i(0)$, and $C = \sum_i g_i(0)$. Now, for each x_j, setting the others to zero, we have,

$$\phi_j(x_j) \;=\; R_j^{-1} f\big(h_j(x_j) + C\big)$$

where $R_j = \prod_{i \neq j} \phi_i(0)$. Thus,

$$F\big(\textstyle\sum_i h_i(x_i) + C\big) \;=\; \textstyle\sum_i F\big(h_i(x_i) + C\big) - R \tag{1.1}$$

where $F(t) = \log f(t)$, and $R = \sum_i \log R_i$. Evaluating (1.1) at $\mathbf{x} = \mathbf{0}$, we see that $R = (n-1)F(C)$, and thus, putting $y_i = h_i(x_i) + C$, we have,

$$F\big(\textstyle\sum_i y_i - (n-1)C\big) \;=\; \textstyle\sum_i F(y_i) - (n-1)F(C)$$

Thus $F(t) = \log f(t)$ is linear, i.e. $F(t) = \lambda\, t + \mu$, and $f(t) \propto \exp(\lambda\, t)$. □ □

This theorem shows in particular that the only elliptically contoured distributions corresponding to independent random variables are Gaussian. More generally, if $g_i(x_i) = |x_i|^{\rho_i}$, $\rho_i > 0$, then the x_i are independent if and only if $p(\mathbf{x})$ is a product of Generalized Gaussians, which we define presently.

Definition 1.1. *The Generalized Gaussian density has the form,*

$$\mathcal{GG}(x;\alpha) \;=\; \frac{1}{2\Gamma(1+1/\alpha)}\,\exp\big(-|x|^{\alpha}\big) \tag{1.2}$$

The location scale family is denoted $\mathcal{GG}(x;\mu,\sigma,\alpha) \propto \exp(-|\sigma^{-1}(x-\mu)|^{\alpha})$. With this definition, we have $\mathcal{N}(x;\mu,\sigma^2) = \mathcal{GG}\big(x;\mu,\sigma\sqrt{2},2\big)$.

Generalized Gaussians are maximum entropy densities under L^p norm constraint.

Theorem 1.2. *The maximum entropy distribution on* $(-\infty,\infty)$*, subject to,*

$$(E\{|x-\mu|^{\alpha}\})^{1/\alpha} \leq A$$

is $\mathcal{GG}\big(x;\mu,\alpha^{1/\alpha}A,\alpha\big)$.

The maximum entropy Generalized Gaussian distribution tends to a Uniform distribution on $[\mu-A,\mu+A]$ as $\alpha\to\infty$, and becomes proportional to $1/|x-\mu|$ (uniformly on compact subsets not including 0) as $\alpha\to 0$. We use the maximum entropy Generalized Gaussian to define a generalized negentropy for random variables in L^p, $0<p<2$. We use the generalized negentropy and the mutual information as two location-scale invariant measures with which to examine the multivariate dependent densities described in this paper.

1.2 Scale Mixtures

If z is a univariate random variable, $z\sim\mathcal{K}(z)$, and $\sigma>0$ is a constant, then we have,

$$\sigma z \;\sim\; \sigma^{-1}\mathcal{K}\big(\sigma^{-1}z\big)$$

If σ is a nonnegative random variable with distribution function $F(\sigma)$, then $x=\sigma z$ is called a scale mixture, and its density is given by,

$$p(x) = \int p(x|\sigma)\,dF(\sigma) = \int_0^{\infty}\sigma^{-1}\mathcal{K}\big(\sigma^{-1}x\big)\,dF(\sigma)$$

Gaussian scale mixtures have the form,

$$p(x) = \int_0^{\infty}\mathcal{N}(x;0,\xi)\,dF(\xi) \tag{1.3}$$

where $\mathcal{N}(x\,;\mu,\sigma^2)$ denotes the Gaussian density with mean μ and variance σ^2. A Gaussian scale mixture x can be represented as a product $x=\xi^{1/2}z$, where $p(x|\xi) = \mathcal{N}(0,\xi)$.

A random *vector* can be constructed by multiplying the scalar random variable $\xi^{1/2}$ by a Gaussian random vector,

$$\mathbf{x} = \xi^{1/2}\mathbf{z}$$

where $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. The components of $\mathbf{x}$ are then dependent, while retaining the uncorrelatedness of $\mathbf{z}$. For the density of $\mathbf{x}$ we have,

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}} \int_0^\infty \xi^{-d/2} \exp\left(-\tfrac{1}{2}\,\xi^{-1}\,\|\mathbf{x}\|^2\right) f(\xi)\, d\xi \tag{1.4}$$

We assume throughout the remainder of the paper that the mixing density $f(\xi) = F'(\xi)$ exists.

Similarly, for a Generalized Gaussian scale mixture, with,

$$x_i = \xi^{1/\rho_i} z_i$$

where $z_i \sim \mathcal{GG}(z; 0, 1, \rho_i)$, we have,

$$p(\mathbf{x}) = \frac{1}{\mathcal{Z}_{\boldsymbol{\rho}}} \int_0^\infty \xi^{-d/\bar{\rho}} \exp\left(-\xi^{-1} \textstyle\sum_i |x_i|^{\rho_i}\right) f(\xi)\, d\xi \tag{1.5}$$

where $\bar{\rho}$ is the harmonic mean, $d/\sum_i \rho_i^{-1}$, and

$$\mathcal{Z}_{\boldsymbol{\rho}} \triangleq\ 2^d \prod_{i=1}^{d} \Gamma(1 + 1/\rho_i) \tag{1.6}$$

We shall restrict our attention to Generalized Gaussian scale mixtures as these seem to be the only densities which yield tractable estimation procedures along with the advantages summarized below. Such densities, nevertheless, cover almost all previously proposed multivariate densities, as well as suggest useful new density models.

1.3 The Proposed Framework

We propose three basic models of higher order dependency: positive norm dependency (PND), negative norm dependency (NND), and skew norm dependency (SND). These are described in the following.

1.3.1 Positive Norm Dependence (PND)

If the dependence is of the form,

$$x_i = \xi^{1/\rho_i} z_i\,, \quad i = 1, \ldots, n$$

where $z_i \sim \mathcal{GG}(z; 0, 1, \rho_i)$, then the random variable ξ scales each component x_i of $\mathbf{x}$ similarly (identically for equal ρ_i), increasing or decreasing the magnitude depending on whether ξ is greater than or less than 1. Such a situation arises when for example several independent channels are modulated by a common scaling process, inducing "variance dependency" [9] in the elliptically contoured case, and general positive norm dependent in the Generalized Gaussian case.

1.3.2 Negative Norm Dependence (NND)

It may also happen that there is a "negative" variance dependency between random vectors, such that an increase in amplitude of one is associated with a *decrease* in amplitude of the other, for example in biological inhibitory processes. Such a dependence, say between random variables x_1 and x_2, is modeled by,

$$x_1 = \xi^{1/\rho_1} z_1 \,, \quad x_2 = \xi^{-1/\rho_2} z_2$$

where $z_i \sim \mathcal{GG}(z; 0, 1, \rho_i)$. In this case the same scalar random variable ξ modulates both x and y, but in inverse proportions. If the scale of x_1 is increased ($\xi > 1$), then the scale of x_2 will be deceased ($\xi^{-1} < 1$). In certain cases, the joint density $p(x_1, x_2)$ can be calculated and estimated using a framework similar to that used in PND.

1.3.3 Skew Norm Dependence (SND)

Barndorff-Nielsen [2] proposed the Normal Variance-Mean mixture framework for multivariate density modeling. Here the random vector $\mathbf{x}$ can be represented by,

$$\mathbf{x} = \xi^{1/2}\mathbf{z} + \xi\boldsymbol{\beta}$$

where $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, and $\boldsymbol{\beta} \in \mathbf{R}^n$ is a constant skew, or drift vector. This case leads to tractable estimation procedures only if $\mathbf{z}$ is Gaussian.

Since the scale mixing random variable also controls the amount of skew, or drift, the multivariate density has a conic structure, with the variance increasing as the drift increases, so that the density spreads out in the direction of the skew. This is in contrast to simply adding an independent non-negative scaling of $\boldsymbol{\beta}$, which would simply convolve the density along the direction of the skew. The SND densities cannot be made independent by linear or simple hyperspherical transformation.

1.4 Advantages of the Proposed Framework

The advantages of the GGSM/NVMM model may be be summarized in the following five (mnemonically titled) capacities.

- *Evaluate.* For particular mixing densities, the dependent scale mixture densities can be evaluated in closed form. Marginals and conditional densities can also be evaluated, as well as conditional (posterior) moments. This is important in likelihood based decision theoretic tasks, in which precise discriminatory capabilities may yield significant advantages.

- *Estimate.* The ability to evaluate posterior moments of the scale mixing variable, along with the tractable Gaussian or independent Generalized Gaussian conditional posterior density, allow the model to be estimated in an efficient manner using Expectation-Maximization (EM) based algorithms.

- *Generate.* Since the explicit construction of the multivariate dependency involves on the generation of Generalized Gaussian random vectors, and scalar mixing random variables, the modeled random vectors can be easily generated for purposes of model verification, simulation, and sampling.
- *Calculate.* Since the dependence of the random variables is limited primarily to some type of dependence in norm, hyperspherical coordinates, or generalized hyperspherical coordinates can be used to transform densities into separable functions that may be integrated as a product of univariate integrals. This is important for the calculation of "p-values", or tail probabilities for significance testing. Also, calculation of joint entropy is reduced to the evaluation of a one dimensional integral, which allows the calculation of mutual information and (generalized) negentropy, each of which are properties of density "types" [3], independent of scale or location parameters.
- *Separate.* The generalized hyperspherical coordinates can also be used to transform observations of dependent vectors into independent vectors. This allows in particular the efficient encoding and quantization of observations. The separability of the density in generalized hyperspherical coordinates allows the optimal high rate vector quantization point density [8] to be approximated in a simple manner using scalar quantization.

1.5 Related Work

Hyvärinen [9, 10] has recently proposed such a model for Independent Subspace Analysis of images. A similar approach is developed by Eltoft, Kim et al. [12, 5], which is referred to as Independent Vector Analysis (IVA). In [11] a method is proposed for convolutive blind source separation in reverberative environments using a frequency domain approach with sources having variance (scale) dependency across frequencies.

1.6 Outline of the Paper

The remainder of the paper is organized as follows. In §2 we describe the scale mixing densities that will be used, and develop the properties of univariate Gaussian and Generalized Gaussian scale mixtures, and univariate Normal Variance-Mean mixtures. In §3 we discuss the multivariate Gaussian scale mixtures, and in §4 we present the Generalized Gaussian scale mixture model. In §5 we discuss negative norm dependence models. In §6 we discuss the multivariate Normal Variance-Mean mixture model. In §7 we derive the expressions for posterior moments and EM updates for the proposed models. In §**??** we derive expressions for mutual information and generalized negentropy for the proposed models. These two location-scale invariant measures allow us to plot the curves traced out by the parameterized forms of the proposed densities (e.g. for varying Generalized Gaussian shape parameter) in a two dimensional space, graphically depicting their mutual relationships in terms of dependence (mutual information) and Kullback-Leibler divergence from Gaussian, or Generalized Gaussian (generalized negentropy) in the case of infinite variance densities.

2 Mixing Densities and Univariate Scale Mixtures

If $p(x)$ is a Gaussian scale mixture density, then its characteristic function $\hat{p}(\omega)$ is given by,

$$\hat{p}(\omega) \;=\; \int_{-\infty}^{\infty} \exp(i\omega x) \int_0^{\infty} \mathcal{N}(x; 0, \xi)\, f(\xi)\, d\xi\, dx \;=\; \int_0^{\infty} \exp\big(-\tfrac{1}{2}\,\xi\,\omega^2\big)\, f(\xi)\, d\xi \quad (2.1)$$

So we see that the characteristic function of a Gaussian scale mixture is also (proportional to) a Gaussian scale mixture, with a transformed mixing density [6],

$$\hat{p}(\omega) = \int_0^{\infty} \xi^{-1/2} \exp(-\tfrac{1}{2}\xi^{-1}\omega^2)\, \xi^{-3/2} f(\xi^{-1})\, d\xi \quad (2.2)$$

Defining $\varphi(t) \triangleq E\{\exp(\xi t)\}$ to be the moment generating function of the mixing density, we have, from (2.1),

$$\hat{p}(\omega) = \varphi\big(-\tfrac{1}{2}\,\omega^2\big) \quad \Rightarrow \quad \varphi(t) = \hat{p}\big(i\sqrt{2t}\big) \quad (2.3)$$

Thus for Gaussian scale mixtures, the moment generating function of the mixing density, when it exists, is related in a simple way to the characteristic function of the scale mixture [2]. The mixing density itself can be found formally using the inverse transform of the moment generating function. This was pointed out in the Andrews and Mallows (1974) paper. The mixing density can also be found (formally) using the Mellin transform [4].

2.1 Mixing Densities

In this section we give examples of the densities of scale mixing random variables. We give the expressions for the moments of these densities, and generalize the densities to include as a multiplicative factor a half-integral power function when this does not already exist as part of the density parametrization. The moments of the scale mixture are linearly related to the moments of the scale mixing density

2.1.1 Stable Densities

Using the Mellin transform, the moments of the positive α-stable distribution are found to be,

$$\int_0^{\infty} \xi^p\, S_\alpha^+(\xi)\, d\xi = \frac{\Gamma(1 - p/\alpha)}{\Gamma(1-p)}, \quad 0 < \alpha < 1, \;\; p < \alpha$$

where S_α^+ denotes the positive α-stable density of order α. Thus for the moments of the mixing density of the Generalized Gaussian density with shape parameter α, we have,

$$E\{\xi^p\} = \frac{\sqrt{\pi}\,\Gamma\big(\frac{2p+1}{\alpha}\big)}{2^p\,\Gamma\big(\frac{1}{\alpha}\big)\Gamma\big(\frac{2p+1}{2}\big)}, \quad 0 < \alpha < 2, \;\; p > -(\alpha+1)/2 \tag{2.4}$$

Using (2.3), and the series representation of the α-stable density [13], we have for the moment generating function of the mixing density of the Generalized Gaussian with shape parameter α,

$$E\{\exp(\xi t)\} = \frac{1}{\pi\alpha}\sum_{k=0}^{\infty}\frac{\Gamma\big((2k+1)/\alpha\big)}{\Gamma(2k+1)}\,(2t)^k\,, \quad 1 < \alpha \le 2 \tag{2.5}$$

The mixing density for the Generalized Gaussian density with shape parameter α is related to the positive α-stable density of order $\alpha/2$ [6],

$$f(\xi) = \frac{\sqrt{2\pi}}{4\Gamma(1+1/\alpha)}\,\xi^{-3/2}\,S_{\alpha/2}^{+}\big(\tfrac{1}{2}\xi^{-1}\big), \quad \xi > 0 \tag{2.6}$$

2.1.2 Generalized Inverse Gaussian

The mixing density of the Generalized Hyperbolic density is the Generalized Inverse Gaussian density, which has the form,

$$\mathcal{N}^\dagger\big(\xi\,;\lambda,\,\delta^2,\kappa^2\big) = \frac{(\kappa/\delta)^\lambda}{2\,K_\lambda(\delta\kappa)}\,\xi^{\lambda-1}\exp\big(-\tfrac{1}{2}\big(\delta^2\xi^{-1}+\kappa^2\xi\big)\big), \quad \xi > 0 \tag{2.7}$$

where K_λ is the Bessel K function, or modified Bessel function of the second kind. The moments of the Generalized Inverse Gaussian are easily found by direct integration, using the fact that (2.7) integrates to one,

$$E\{\xi^a\} = \left(\frac{\delta}{\kappa}\right)^a\frac{K_{\lambda+a}(\delta\kappa)}{K_\lambda(\delta\kappa)} \tag{2.8}$$

Similarly, by direct integration, we have for the moment generating function,

$$E\{\exp(\xi t)\} = \frac{\kappa^\lambda}{(\kappa^2-t)^{\lambda/2}}\frac{K_\lambda\big(\delta\sqrt{\kappa^2-t}\big)}{K_\lambda(\delta\kappa)}\,, \quad t < \kappa^2 \tag{2.9}$$

2.1.3 Beta and Pareto

2.1.4 Kolmogorov-Smirnov

The scale mixing density is related to the Kolmogorov-Smirnov distance statistic [1, 2, 6]. The Generalized Kolmogorov density [2] is given by,

$$f(\xi) = \frac{1}{\Gamma(\nu)^2}\sum_{k=0}^{\infty}(-1)^k\frac{\Gamma(k+2\nu)}{\Gamma(k+1)}\,(k+\nu)\exp\big(-\tfrac{1}{2}(k+\nu)^2\xi\big), \quad \xi > 0 \tag{2.10}$$

The moments are given by,

$$E\{\xi^a\} = \frac{2^{a+1}\Gamma(a+1)}{\Gamma(\nu)^2} \sum_{k=0}^{\infty} (-1)^k \frac{\Gamma(k+2\nu)}{\Gamma(k+1)} \frac{1}{(k+\nu)^{2a+1}}$$

Half integral moments, $a = m/2$, for $m \geq -1$, can be written,

$$E\big\{\xi^{m/2}\big\} \;=\; \frac{2^{m/2+1}\Gamma(m/2+1)}{B(\nu,\nu)\,\nu^{m+1}}\; {}_{m+2}F_{m+1}\big([2\nu,\nu,\ldots,\nu]\,;[\nu{+}1,\ldots,\nu{+}1]\,;-1\big) \quad (2.11)$$

where ${}_pF_q$ is the generalized hypergeometric function, and $B(x,y)$ is the Beta function.

Using (2.3) and the characteristic function of the Generalized Logistic density [2], we have for the moment generating function of the Generalized Kolmogorov,

$$E\{\exp(\xi t)\} = \frac{\Gamma\big(\nu+\sqrt{2\,t}\big)\Gamma\big(\nu-\sqrt{2\,t}\big)}{\Gamma(\nu)^2}\,, \quad t < \tfrac{1}{2}\nu^2 \quad (2.12)$$

2.2 Gaussian scale mixtures

In this section we give some examples of Gaussian scale mixtures.

2.2.1 Generalized Gaussian

The Generalized Gaussian density has the form,

$$\mathcal{GG}(x;\alpha) \;=\; \frac{1}{2\Gamma(1+1/\alpha)}\,\exp\big(-|x|^\alpha\big) \quad (2.13)$$

The Generalized Gaussian is a Gaussian scale mixture for $0 < \alpha < 2$.

2.2.2 Generalized Logistic

The Generalized Logistic, also referred to as (the symmetric) Fisher's z distribution [2], has the form,

$$\mathcal{GL}(x;\nu) \;=\; \frac{1}{B(\nu,\nu)}\,\frac{e^{-\nu x}}{(1+e^{-x})^{2\nu}} \;=\; \frac{1}{4^\nu B(\nu,\nu)}\,\frac{1}{\cosh^{2\nu}\big(\frac{1}{2}x\big)} \quad (2.14)$$

The Generalized Logistic is a Gaussian scale mixture for all $\nu > 0$.

2.2.3 Generalized Hyperbolic

The Generalized Hyperbolic density [2] has the form,

$$\mathcal{GH}(x\,;\delta,\kappa,\lambda) = \frac{1}{\sqrt{2\pi}}\,\frac{\kappa^{1/2}}{\delta^\lambda K_\lambda(\delta\kappa)}\,\frac{K_{\lambda-1/2}\big(\kappa\sqrt{\delta^2+x^2}\,\big)}{\big(\delta^2+x^2\big)^{1/4-\lambda/2}} \quad (2.15)$$

Limiting cases when $\delta \to 0$ or $\kappa \to 0$:

1. ***McKay's Bessel K*** If the mixing density is Gamma distributed, then the scale mixture is given by,

$$p(x;\nu) = \frac{\nu^{1/2}}{\pi^{1/2}\Gamma(\frac{1}{2}\nu+\frac{1}{2})}\left(\frac{\nu^{1/2}|x|}{2}\right)^{\nu/2} K_{\nu/2}\big(\nu^{1/2}|x|\big)$$

2. ***Student's t***. If the scale mixing density is that of a random variable whose inverse is Gamma distributed, then the scale mixture is the Student's t density,

$$p(x;\nu) \;=\; \frac{1}{B(\frac{1}{2},\nu)}\,\frac{1}{(1+x^2)^{1/2+\nu}}$$

2.3 Generalized Gaussian scale mixtures

For a GSM, $p(x)$, evaluated at $|x|^{\rho/2}$, we have,

$$p\big(|x|^{\rho/2}\big) \;=\; \frac{1}{\sqrt{2\pi}}\int_0^\infty \xi^{-1/2}\,\exp\big(-\tfrac{1}{2}\xi^{-1}|x|^{\rho}\big)\;dF(\xi) \tag{2.16}$$

Integrating over x, we get,

$$\int_{-\infty}^{\infty} p\big(|x|^{\rho/2}\big)\,dx \;=\; \frac{\mathcal{Z}_\rho}{\sqrt{2\pi}}\int_0^\infty \xi^{\,1/\rho-1/2}\,dF(\xi) \;=\; \frac{\mathcal{Z}_\rho M_{2/\rho}}{\sqrt{2\pi}} \tag{2.17}$$

where,

$$\mathcal{Z}_\rho \triangleq 2^{1+1/\rho}\,\Gamma(1+1/\rho), \quad M_a \triangleq \int_0^\infty \xi^{\,a/2-1/2}\,dF(\xi) \tag{2.18}$$

We can thus construct the Generalized Gaussian scale mixture (GGSM),

$$x = (2\,\xi)^{1/\rho} z$$

where $z \sim \mathcal{GG}(z;\rho)$, and we have,

$$p(x;\rho) = \frac{\sqrt{2\pi}}{\mathcal{Z}_\rho M_{2/\rho}}\,p\big(|x|^{\rho/2}\big), \quad f(\xi;\rho) = M_{2/\rho}^{-1}\,\xi^{1/\rho-1/2} f(\xi) \tag{2.19}$$

2.3.1 Hypergeneralized Hyperbolic density

In the case of the Generalized Inverse Gaussian mixing density, the integration required to form the Generalized Gaussian scale mixture can be evaluated in terms of the Bessel K function. If the mixing density is $\mathcal{N}^\dagger$, then the posterior density of ξ given x is also $\mathcal{N}^\dagger$,

$$f(\xi|x) = \mathcal{N}^\dagger\big(\xi\,;\sqrt{\delta^2+|x|^{\rho}}\,,\kappa,\lambda-d/\rho\big) \tag{2.20}$$

We then get the "hypergeneralized hyperbolic distribution",

$$\mathcal{HH}(x\,;\delta,\kappa,\lambda,\rho) = \frac{1}{\mathcal{Z}_\rho}\frac{\kappa^{1/\rho}}{\delta^\lambda K_\lambda(\delta\kappa)}\frac{K_{\lambda-1/\rho}\big(\kappa\sqrt{\delta^2+|x|^\rho}\,\big)}{\big(\delta^2+|x|^\rho\big)^{(1/\rho-\lambda)/2}} \tag{2.21}$$

where $\mathcal{Z}_\rho$ is defined in (2.18). Using (2.8) with (2.20), we get,

$$E\{\xi^{-1}|x\} = \frac{\kappa}{\sqrt{\delta^2+|x|^\rho}}\frac{K_{\lambda-1/\rho-1}\big(\kappa\sqrt{\delta^2+|x|^\rho}\,\big)}{K_{\lambda-1/\rho}\big(\kappa\sqrt{\delta^2+|x|^\rho}\,\big)} \tag{2.22}$$

We have the following limiting cases of the Hypergeneralized Hyperbolic density.

1. ***Generalized Cauchy***. Inverse Gamma mixing of Generalized Gaussian random variables yields the Generalized Cauchy density,

   $$\mathcal{GC}(x;\alpha,\nu) \;=\; \frac{\alpha}{2\,B(1/\alpha,\nu)}\,\frac{1}{(1+|x|^\alpha)^{1/\alpha+\nu}}$$

   where $B(x,y) = \Gamma(x)\Gamma(y)/\Gamma(x+y)$ is the Beta function. The Generalized Cauchy is a Gaussian scale mixture for $\nu > 0$ and $0 < \alpha \le 2$. The scale mixing density is the scale convolution of the inverse Gamma density with a positive α-stable density of order $\alpha/2$.

2. **McKay's Bessel *K*** If the mixing density is Gamma distributed, then the scale mixture is,

   $$p(x;\nu) = \frac{\nu^{1/2}}{\pi^{1/2}\Gamma((\nu+1)/2)}\left(\frac{\nu^{1/2}|x|}{2}\right)^{\nu/2} K_{\nu/2}\big(\nu^{1/2}|x|\big)$$

3 Multivariate Gaussian Scale Mixtures

In this section we show how general dependent multivariate densities can be derived using univariate Gaussian scale mixtures. Throughout this section, $\|\mathbf{x}\|$ will denote the 2-norm in the assoociated d-dimensional Euclidean space.

3.1 Multivariate analogues of univariate Gaussian Scale Mixtures

A Gaussian scale mixture x can be represented as a product $\xi^{1/2}z$, where $z \sim \mathcal{N}(0,\xi)$. We can construct a random *vector* by multiplying the same scalar random variable $\xi^{1/2}$ by a Gaussian random vector,

$$\mathbf{x} = \xi^{1/2}\mathbf{z}$$

where $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. The components $\mathbf{x}$ then become dependent, while retaining the uncorrelatedness of $\mathbf{z}$. For the density of $\mathbf{x}$ we have,

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}} \int_0^\infty \xi^{-d/2} \exp\left(-\tfrac{1}{2}\,\xi^{-1}\,\|\mathbf{x}\|^2\right) f(\xi)\, d\xi \tag{3.1}$$

If ξ is a Generalized Inverse Gaussian, then the density of $\mathbf{x}$ can be written in terms of the modified Bessel function of the second kind.

More generally, let x be distributed according to the Gaussian scale mixture density $p(x)$. Then,

$$p(\sqrt{x}) = \frac{1}{(2\pi)^{1/2}} \int_0^\infty \xi^{-1/2} \exp\left(-\tfrac{1}{2}\,\xi^{-1} x\right) f(\xi)\, d\xi \tag{3.2}$$

Taking the nth derivative of both sides of (3.2), we find,

$$\frac{d^n}{dx^n} p(\sqrt{x}) = \frac{(-2)^{-n}}{(2\pi)^{1/2}} \int_0^\infty \xi^{-n-1/2} \exp\left(-\tfrac{1}{2}\,\xi^{-1} x\right) f(\xi)\, d\xi \tag{3.3}$$

Derivatives of the univariate density $p(x)$ are used to construct densities in higher dimensions $d \geq 2$. We consider the cases of even and odd dimension d separately, then formulate a combined equation for the relationship.

3.1.1 Odd d

If d is odd, then with $n = (d-1)/2$ in (3.3), we have,

$$\pi^{-(d-1)/2}(-D)^{(d-1)/2} p(\sqrt{x}) = \frac{1}{(2\pi)^{d/2}} \int_0^\infty \xi^{-d/2} \exp\left(-\tfrac{1}{2}\,\xi^{-1} x\right) f(\xi)\, d\xi$$

and we can write the density of $p(\mathbf{x})$

$$d \textbf{ odd}: \qquad p(\mathbf{x}) = \pi^{-(d-1)/2}(-D)^{(d-1)/2} p(\sqrt{t})\big|_{t=\|\mathbf{x}\|^2} \tag{3.4}$$

If we define the linear operator V by,

$$V g(x) \triangleq -2\, D g(\sqrt{t}\,)|_{t=x^2} = -x^{-1} D g(x) \tag{3.5}$$

then we have,

$$d \textbf{ odd}: \qquad p(\mathbf{x}) = (2\pi)^{-(d-1)/2}\, V^{(d-1)/2} p(t)\big|_{t=\|\mathbf{x}\|} \tag{3.6}$$

In particular, for $d = 3$, we have,

$$d = 3: \qquad p(\mathbf{x}) = -\frac{1}{2\pi} \frac{p'\big(\|\mathbf{x}\|\big)}{\|\mathbf{x}\|} \tag{3.7}$$

3.1.2 Even d

For even d, the multivariate density arising from applying the original scale mixing variable to a Gaussian vector of dimension d can be written formally in terms of the Weyl fractional derivative [4]. However as the fractional derivative is not generally obtainable in closed form, we consider a modification of the original univariate scale density $f(\xi)$,

$$f(\xi;0) = \frac{\xi^{-1/2} f(\xi)}{\int_0^\infty \xi^{-1/2} f(\xi)\, d\xi} \tag{3.8}$$

With this modified scale density, the density of x evaluated at $\sqrt{x}$ becomes,

$$p(\sqrt{x}) = \frac{M_0}{(2\pi)^{1/2}} \int_0^\infty \exp\left(-\tfrac{1}{2}\,\xi^{-1}x\right) f(\xi;0)\, d\xi \tag{3.9}$$

where $M_0 = \int_0^\infty \xi^{-1/2} f(\xi)\, d\xi$ as in (2.18).

Proceeding as we did for odd d, taking the nth derivative of both sides of (3.9), with $n = d/2$, we get,

$$\begin{aligned} d \textbf{ even}: \quad p(\mathbf{x}) &= M_0^{-1}\, 2^{1/2} \pi^{-(d-1)/2} (-D)^{d/2} p(\sqrt{t})\big|_{t=\|\mathbf{x}\|^2} \\ &= M_0^{-1} (2\pi)^{-(d-1)/2}\, V^{d/2} p(t)\big|_{t=\|\mathbf{x}\|} \end{aligned} \tag{3.10}$$

where the operator V was defined in (3.5). The formula for even d is thus equivalent to that for the odd $d+1$ given by (3.6) except for the constant factor $M_0^{-1}\sqrt{2\pi}$. In particular, we have

$$d = 2: \qquad p(\mathbf{x}) = -\frac{1}{M_0\sqrt{2\pi}} \frac{p'\big(\|\mathbf{x}\|\big)}{\|\mathbf{x}\|} \tag{3.11}$$

3.1.3 General modified mixing density

We can generalize the modification (3.8) by including an arbitrary half integral moment,

$$f(\xi;m) \triangleq \frac{\xi^{(m-1)/2} f(\xi)}{\int_0^\infty \xi^{(m-1)/2} f(\xi)\, d\xi} = M_m^{-1}\, \xi^{(m-1)/2} f(\xi) \tag{3.12}$$

where M_m is defined as in (2.18). In terms of this modified scale mixing density, the density of x evaluated at $\sqrt{x}$ becomes,

$$p(\sqrt{x}) = \frac{M_m}{(2\pi)^{1/2}} \int_0^\infty \xi^{-m/2} \exp\left(-\tfrac{1}{2}\,\xi^{-1}x\right) f(\xi;m)\, d\xi \tag{3.13}$$

Taking the nth derivative of both sides of (3.13), with $n = (d-m)/2$, we get,

$$\frac{d^n}{dx^n} p(\sqrt{x}) = \frac{M_m}{(2\pi)^{1/2}}\, (-2)^{-(d-m)/2} \int_0^\infty \xi^{-d/2} \exp\left(-\tfrac{1}{2}\,\xi^{-1}x\right) f(\xi;m)\, d\xi$$

Thus, given a univariate Gaussian scale mixture, we can construct a density in $\mathbf{R}^d$ for all $d-m$ even such that the scalar moment M_m exists,

$$\begin{aligned} d-m\ \textbf{even}: \quad p(\mathbf{x}) &= M_m^{-1}\,(2\pi)^{-(d-1)/2}\,(-2)^{(d-m)/2} D^{(d-m)/2} p(\sqrt{t})\big|_{t=\|\mathbf{x}\|^2} \\ &= M_m^{-1}\,(2\pi)^{-(d-1)/2}\, V^{(d-m)/2} p(t)\big|_{t=\|\mathbf{x}\|} \end{aligned} \tag{3.14}$$

Letting $n = (d-m)/2$, we have the following form for densities in $\mathbf{R}^d$,

$$p(\mathbf{x}; n) = M_{d-2n}^{-1}\,(2\pi)^{-(d-1)/2}\,(-2)^n D^n p(\sqrt{t})\big|_{t=\|\mathbf{x}\|^2} \tag{3.15}$$

The mixing density $f(\xi; m)$ of (3.14) is given by (3.12), where $f(\xi) = f(\xi; 1)$ is the mixing density of the univariate Gaussian scale mixture. The mixing density in (3.15) is $f(\xi; d-2n)$.

According to [7, §0.4331], we have, for arbitrary $F(x)$ smooth on $(0, x)$, and $n \geq 1$,

$$\frac{d^n}{dx^n} F(\sqrt{x}) = \frac{F^{(n)}(\sqrt{x})}{(2\sqrt{x})^n} - \frac{n(n-1)}{1!}\frac{F^{(n-1)}(\sqrt{x})}{(2\sqrt{x})^{n+1}} + \frac{(n+1)n(n-1)(n-2)}{2!}\frac{F^{(n-2)}(\sqrt{x})}{(2\sqrt{x})^{n+2}} - \cdots$$

which may be written,

$$\begin{aligned} \frac{d^n}{dx^n} F(\sqrt{x}) &= \sum_{j=0}^{n-1} \frac{(-1)^j}{j!}\frac{\Gamma(n+j)}{\Gamma(n-j)}\frac{F^{(n-j)}(\sqrt{x})}{(2\sqrt{x})^{n+j}} \\ &= \sum_{k=1}^{n} \frac{(-1)^{n-k}}{(n-k)!}\frac{\Gamma(2n-k)}{\Gamma(k)}\frac{F^{(k)}(\sqrt{x})}{(2\sqrt{x})^{2n-k}} \end{aligned} \tag{3.16}$$

Thus with the definition (3.5), we have,

$$V^n F(x) = (-2)^n \frac{d^n}{dt^n} F(\sqrt{t})\Big|_{t=x^2} = \sum_{k=1}^{n} (-1)^k\, C_{k,n}\, \frac{F^{(k)}(x)}{x^{2n-k}} \tag{3.17}$$

where the (integer valued) coefficients $C_{k,n}$ are given by,

$$C_{k,n} \triangleq \frac{2^{-(n-k)}\Gamma(2n-k)}{\Gamma(n-k+1)\Gamma(k)} \tag{3.18}$$

Using (3.14), we can write the density $p(\mathbf{x}; n)$, for $n = 1, 2, \ldots$,

$$p(\mathbf{x}; n) = M_{d-2n}^{-1}\,(2\pi)^{-(d-1)/2} \sum_{k=1}^{n} (-1)^k\, C_{k,n} \frac{p^{(k)}\big(\|\mathbf{x}\|\big)}{\|\mathbf{x}\|^{2n-k}} \tag{3.19}$$

where $p^{(k)}(x)$ denotes the kth derivative of the univariate Gaussian scale mixture $p(x)$, which has mixing density $f(\xi; 1)$.

3.2 Examples

Let $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ in $\mathbf{R}^3$.

3.2.1 Dependent Generalized Gaussian in $\mathbf{R}^3$

If the scale mixing random variable ξ is inverse α-stable of order $\alpha/2$, then the density of $\mathbf{x} = \xi^{1/2}\mathbf{z}$ will be a dependent multivariate generalization of the Generalized Gaussian density. In $\mathbf{R}^3$, using formula (3.7), we have,

$$p(\mathbf{x}) = \frac{1}{4\pi}\frac{\alpha}{\Gamma(1+1/\alpha)}\,\|\mathbf{x}\|^{\alpha-2}\exp\big(-\|\mathbf{x}\|^{\alpha}\big)\,,\quad 0<\alpha<2 \tag{3.20}$$

3.2.2 Dependent Logistic in $\mathbf{R}^3$

Suppose we wish to formulate a dependent Logistic type density on $\mathbf{R}^3$. The scale mixing density in the Gaussian scale mixture representation for the Logistic density is related to the Kolmogorov-Smirnov distance statistic [1, 2, 6], which is only expressible in series form. However, we may determine the multivariate density produced from the product $\mathbf{x} = \xi^{1/2}\mathbf{z}$. Using formula (3.7), we get,

$$p(\mathbf{x}) = \frac{1}{8\pi}\,\frac{\sinh\big(\frac{1}{2}\|\mathbf{x}\|\big)}{\|\mathbf{x}\|\cosh^3\big(\frac{1}{2}\|\mathbf{x}\|\big)} \tag{3.21}$$

For the Generalized Logistic, we have

$$p(\mathbf{x}) = \frac{1}{2\pi}\,\frac{\nu}{4^{\nu}B(\nu,\nu)}\,\frac{\sinh\big(\frac{1}{2}\|\mathbf{x}\|\big)}{\|\mathbf{x}\|\cosh^{2\nu+1}\big(\frac{1}{2}\|\mathbf{x}\|\big)} \tag{3.22}$$

3.2.3 Generalized Hyperbolic density

The isotropic generalized hyperbolic distribution [2] in dimension d,

$$\mathcal{GH}(\mathbf{x}\,;\delta,\kappa,\lambda) = \frac{1}{(2\pi)^{d/2}}\frac{\kappa^{d/2}}{\delta^{\lambda}K_{\lambda}(\delta\kappa)}\frac{K_{\lambda-d/2}\Big(\kappa\sqrt{\delta^2+\|\mathbf{x}\|^2}\Big)}{\big(\delta^2+\|\mathbf{x}\|^2\big)^{d/4-\lambda/2}} \tag{3.23}$$

is derived as a Gaussian scale mixture with $\mathcal{N}^{\dagger}$ mixing density [2, 5].

4 Multivariate Generalized Gaussian scale mixtures

A possible limitation of the Gaussian scale mixture dependent subspace model is the implied radial symmetry of vectors in the subspace, which leads to non-identifiability of features within the subspace; only the subspace itself can be identified. However, a similar approach using multivariate Generalized Gaussian scale mixtures can be developed, in which the multivariate density becomes a function of the p-norm of the subspace vector rather than the radially symmetric 2-norm. We are thereby able to maintain the directionality and identifiability of the dependent within-subspace features, while preserving their (non-affine) dependence.

4.1 Hypergeneralized Hyperbolic density

For a Generalized Gaussian scale mixture, we have,

$$p(\mathbf{x}) = \frac{1}{\mathcal{Z}_{\boldsymbol{\rho}}} \int_0^\infty \xi^{-d/\bar{\rho}} \exp\left(-\tfrac{1}{2}\,\xi^{-1} \textstyle\sum_i |x_i|^{\rho_i}\right) f(\xi)\, d\xi \tag{4.1}$$

where $\bar{\rho}$ is the harmonic mean, $d/\sum_i \rho_i^{-1}$, and

$$\mathcal{Z}_{\boldsymbol{\rho}} \triangleq\ 2^{d+d/\bar{\rho}} \prod_{i=1}^{d} \Gamma(1+1/\rho_i) \tag{4.2}$$

If the mixing density $f(\xi)$ is $\mathcal{N}^\dagger$, then the posterior density of ξ given $\mathbf{x}$ is also $\mathcal{N}^\dagger$,

$$f(\xi|\mathbf{x}) = \mathcal{N}^\dagger\Big(\xi\,;\sqrt{\delta^2 + \|\mathbf{x}\|_{\boldsymbol{\rho}}^{\bar{\rho}}}\,, \kappa, \lambda - d/\bar{\rho}\Big) \tag{4.3}$$

where we define the pseudo-norm,

$$\|\mathbf{x}\|_{\boldsymbol{\rho}} \triangleq \Big(\textstyle\sum_{i=1}^{d} |x_i|^{\rho_i}\Big)^{1/\bar{\rho}}$$

which is only a true norm if $1 \leq \rho_i = \rho_j$ for all i, j. For $\mathbf{x}$ we then get the anisotropic *hypergeneralized hyperbolic distribution*,

$$\mathcal{HH}(\mathbf{x}\,;\delta,\kappa,\lambda,\boldsymbol{\rho}) = \frac{1}{\mathcal{Z}_{\boldsymbol{\rho}}}\,\frac{\kappa^{d/\bar{\rho}}}{\delta^\lambda K_\lambda(\delta\kappa)}\,\frac{K_{\lambda-d/\bar{\rho}}\Big(\kappa\sqrt{\delta^2 + \|\mathbf{x}\|_{\boldsymbol{\rho}}^{\bar{\rho}}}\,\Big)}{\Big(\delta^2 + \|\mathbf{x}\|_{\boldsymbol{\rho}}^{\bar{\rho}}\Big)^{(d/\bar{\rho}-\lambda)/2}} \tag{4.4}$$

Using (2.8) with (4.3), we get,

$$E\big(\xi^{-1}|\mathbf{x}\big) = \frac{\kappa}{\sqrt{\delta^2 + \|\mathbf{x}\|_{\boldsymbol{\rho}}^{\bar{\rho}}}}\,\frac{K_{\lambda-d/\bar{\rho}-1}\Big(\kappa\sqrt{\delta^2 + \|\mathbf{x}\|_{\boldsymbol{\rho}}^{\bar{\rho}}}\,\Big)}{K_{\lambda-d/\bar{\rho}}\Big(\kappa\sqrt{\delta^2 + \|\mathbf{x}\|_{\boldsymbol{\rho}}^{\bar{\rho}}}\,\Big)} \tag{4.5}$$

We have the following limiting cases of the Hypergeneralized Hyperbolic density.

1. ***Generalized Cauchy***. Inverse Gamma mixing of Generalized Gaussian random variables yields the Generalized Cauchy density,

$$\mathcal{GC}(x;\alpha,\nu) \;=\; \frac{\alpha}{2B(1/\alpha,\nu)}\,\frac{1}{(1+|x|^{\alpha})^{1/\alpha+\nu}}$$

   where $B(x,y) = \Gamma(x)\Gamma(y)/\Gamma(x+y)$ is the Beta function. The Generalized Cauchy is a Gaussian scale mixture for $\nu > 0$ and $0 < \alpha \le 2$. The scale mixing density is the scale convolution of the inverse Gamma density with a positive α-stable density of order $\alpha/2$.

2. **McKay's Bessel *K*** If the mixing density is Gamma distributed, then the scale mixture

$$p(x;\nu) = \frac{\nu^{1/2}}{\pi^{1/2}\Gamma((\nu+1)/2)}\left(\frac{\nu^{1/2}|x|}{2}\right)^{\nu/2} K_{\nu/2}\big(\nu^{1/2}|x|\big)$$

4.2 Hypergeneralized Gaussian scale mixtures

Given a radially symmetric multivariate Gaussian scale mixture $p(\mathbf{x})$, we can formulate a non-radially symmetric "Hypergeneralized Gaussian scale mixture" $\tilde{p}(\mathbf{x})$ in terms of the radially symmetric density $p(\mathbf{x})$, and a moment of the scale mixing density $p(\xi)$. If we define $\mathbf{x}^{\boldsymbol{\rho}}$ to be the vector with components $|x_i|^{\rho_i}$, then we have,

$$p\big(\mathbf{x}^{\boldsymbol{\rho}/2}\big) \;=\; \frac{1}{(2\pi)^{d/2}}\int_0^{\infty} \xi^{-d/2}\,\exp\big(-\tfrac{1}{2}\xi^{-1}\textstyle\sum_i |x_i|^{\rho_i}\big)\,f(\xi)\,d\xi \tag{4.6}$$

Integrating over $\mathbf{x}$, we get,

$$\int p\big(\mathbf{x}^{\boldsymbol{\rho}/2}\big)\,d\mathbf{x} \;=\; \frac{\mathcal{Z}_{\boldsymbol{\rho}}}{(2\pi)^{d/2}}\int_0^{\infty} \xi^{\,d/\bar{\rho}-d/2} f(\xi)\,d\xi \tag{4.7}$$

where $\mathcal{Z}_{\boldsymbol{\rho}}$ is given by (4.2), and $\bar{p}$ is the harmonic mean of the components of $\boldsymbol{\rho}$. We can thus construct a general dependent anisotropic density,

$$p(\mathbf{x};\boldsymbol{\rho}) = \frac{(2\pi)^{d/2}}{\mathcal{Z}_{\boldsymbol{\rho}} M_{\bar{\rho}/2}}\,p\big(\mathbf{x}^{\boldsymbol{\rho}/2}\big) \tag{4.8}$$

5 Negative Norm Dependence

6 Skew Norm Dependence

6.1 Construction of multivariate skew densities from Gaussian scale mixtures

Given a Gaussian scale mixture $\mathbf{x} = \xi^{1/2}\mathbf{z}$,

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}}\int_0^{\infty} \xi^{-d/2}\exp\big(-\tfrac{1}{2}\xi^{-1}\mathbf{x}^T\Sigma^{-1}\mathbf{x}\big)p(\xi)\,d\xi$$

we have, trivially, for arbitrary β,

$$\frac{p(\mathbf{x})\exp\left(\beta^T\Sigma^{-1}\mathbf{x}\right)}{\varphi\left(\frac{1}{2}\beta^T\Sigma^{-1}\beta\right)} = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}}\times$$
$$\int_0^\infty \xi^{-d/2}\exp\left(-\tfrac{1}{2}\,\xi^{-1}\mathbf{x}^T\Sigma^{-1}\mathbf{x}+\beta^T\Sigma^{-1}\mathbf{x}-\tfrac{1}{2}\,\xi\,\beta^T\Sigma^{-1}\beta\right)\frac{p(\xi)\exp\left(\frac{1}{2}\,\xi\,\beta^T\Sigma^{-1}\beta\right)}{\varphi\left(\frac{1}{2}\beta^T\Sigma^{-1}\beta\right)}\,d\xi \tag{6.1}$$

where $\varphi(t) = E\{\exp(\xi\,t)\}$ is the moment generating function of ξ. Now, (6.1) can be written,

$$p(\mathbf{x};\beta,\Sigma) = \int_0^\infty \mathcal{N}\big(\mathbf{x}\,;\xi\beta,\xi\Sigma\big)p(\xi;\beta,\Sigma)\,d\xi \tag{6.2}$$

where,

$$p(\mathbf{x};\boldsymbol{\beta},\Sigma) = \frac{p(\mathbf{x})\,\exp\big(\beta^T\Sigma^{-1}\mathbf{x}\big)}{\varphi\big(\frac{1}{2}\|\beta\|^2_{\Sigma^{-1}}\big)}\,,\quad p(\xi;\boldsymbol{\beta},\Sigma) = \frac{p(\xi)\,\exp\big(\frac{1}{2}\xi\,\|\beta\|^2_{\Sigma^{-1}}\big)}{\varphi\big(\frac{1}{2}\|\beta\|^2_{\Sigma^{-1}}\big)}$$

We have thus constructed a skewed density $p(\mathbf{x};\beta)$ in terms of the isotropic density $p(\mathbf{x}) = p(\mathbf{x};\mathbf{0})$ and the moment generating function φ of the scale mixing density $p(\xi)$. The skewed density is a location-scale mixture [2] of the Gaussian $\mathbf{z}\sim\mathcal{N}\big(\mathbf{0},\Sigma\big)$,

$$\mathbf{x} = \xi^{1/2}\mathbf{z} + \xi\boldsymbol{\beta}$$

7 Posterior moments and the EM algorithm

7.1 Posterior moments and EM algorithms

To use scale mixtures in the EM context, it is necessary to calculate posterior moments of the scaling random variable. Using (3.14), we have,

$$\begin{aligned} E\{\xi^{-1}|\mathbf{x};n\} &= \frac{\int_0^\infty \xi^{-1}p(\mathbf{x},\xi;n)\,d\xi}{p(\mathbf{x};n)} \\ &= \frac{M^{-1}_{d/2-n}(2\pi)^{-(d-1)/2}VV^np(t)|_{t=\|\mathbf{x}\|}}{p(\mathbf{x};n)} \\ &= \left.\frac{V^{n+1}p(t)}{V^np(t)}\right|_{t=\|\mathbf{x}\|} \end{aligned}$$

where V is the operator defined by (3.5). In general, whenever the kth posterior moment exists, we have,

$$E\{\xi^k|\mathbf{x};n\} = \left.\frac{V^{n-k}p(t)}{V^np(t)}\right|_{t=\|\mathbf{x}\|} \tag{7.1}$$

When $n-k \leq -1$, (7.1) involves repeated application of the linear operator V^{-1}, the inverse of the operator V, which is given by,

$$V^{-1} f(x) \;=\; -Ixf(x) \;=\; -\int_{-\infty}^{x} t f(t)\, dt$$

7.2 Skew posterior updates

We now assume arbitrary location vector μ, along with drift vector β, and structure matrix Σ. The posterior expectation of ξ^{-1} is the same as in the non-skew case, since,

$$E\{\xi^{-1}|\mathbf{x};\boldsymbol{\mu},\boldsymbol{\beta},\Sigma\} \;=\; \frac{\int_0^\infty \xi^{-1} p(\mathbf{x},\xi;\boldsymbol{\mu},\boldsymbol{\beta},\Sigma)\, d\xi}{p(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\beta},\Sigma)} \;=\; \frac{\int_0^\infty \xi^{-1} p(\mathbf{x},\xi;\boldsymbol{\mu},\Sigma)\, d\xi}{p(\mathbf{x};\boldsymbol{\mu},\Sigma)} \;=\; E\{\xi^{-1}|\mathbf{x};\boldsymbol{\mu},\Sigma\}$$

Thus, we have,

$$E\{\xi^{-1}|\mathbf{x};\boldsymbol{\mu},\boldsymbol{\beta},\Sigma\} = \left.\frac{V^{(d-m+1)/2} p(t)}{V^{(d-m-1)/2} p(t)}\right|_{t=\|\mathbf{x}-\boldsymbol{\mu}\|_{\Sigma^{-1}}} \tag{7.2}$$

as in §7.1.

7.3 Closed form parameter updates

Given N observations $\{\mathbf{x}_k\}_{k=1}^N$ and fixed current estimate $\boldsymbol{\beta}^l$, the location parameter $\boldsymbol{\mu}$ that maximizes the complete log likelihood is found to be,

$$\boldsymbol{\mu}^{l+1} = \frac{\frac{1}{N}\sum_k \gamma_k^l \mathbf{x}_k - \boldsymbol{\beta}^l}{\frac{1}{N}\sum_k \gamma_k^l} \tag{7.3}$$

where $\gamma_k^l \triangleq E\big\{\xi_k^{-1} \,|\, \mathbf{x}_k; \boldsymbol{\mu}^l\big\}$ does not depend on β^l.

Then the estimation equation to be solved for $\boldsymbol{\beta}^{l+1}$, which does not involve the posterior estimates of the ξ_k^{-1} given $\boldsymbol{\mu}^{l+1}$, is,

$$\frac{\varphi'\big(\frac{1}{2}\|\boldsymbol{\beta}\|^2_{\Sigma^{-1}}\big)}{\varphi\big(\frac{1}{2}\|\boldsymbol{\beta}\|^2_{\Sigma^{-1}}\big)}\,\boldsymbol{\beta} \;=\; \mathbf{c} - \boldsymbol{\mu}^{l+1}$$

where $\mathbf{c} \triangleq \frac{1}{N}\sum_k \mathbf{x}_k$. Thus the direction of $\boldsymbol{\beta}^{l+1}$ is the same as that of $\mathbf{c} - \boldsymbol{\mu}^{l+1}$. Given $\boldsymbol{\mu}^{l+1}$, the optimal $\boldsymbol{\beta}^{l+1}$ may be found by first determining $\zeta^{l+1} \triangleq \frac{1}{2}\|\boldsymbol{\beta}^{l+1}\|^2_{\Sigma^{-1}}$ by solving,

$$h(\zeta) \;\triangleq\; \left(\frac{\varphi'(\zeta)}{\varphi(\zeta)}\right)^2 \zeta \;=\; \tfrac{1}{2}\left\|\mathbf{c} - \boldsymbol{\mu}^{l+1}\right\|^2_{\Sigma^{-1}} \tag{7.4}$$

for ζ^{l+1}. Then $\boldsymbol{\beta}^{l+1}$ is given as,

$$\boldsymbol{\beta}^{l+1} = \sqrt{2\zeta^{l+1}}\,\frac{\mathbf{c} - \boldsymbol{\mu}^{l+1}}{\|\mathbf{c} - \boldsymbol{\mu}^{l+1}\|_{\Sigma^{-1}}} \tag{7.5}$$

Repeated iteration constitutes a coordinate ascent EM algorithm for μ and β.

8 Conclusion

We have shown how to derive general multivariate Gaussian scale mixtures in terms of scalar Gaussian scale mixtures, and how to optimize them using an EM algorithm. We generalized the spherically (or ellipsoidally) symmetric Gaussian scale mixture by introducing a generalization of Barndorff-Nielsen's generalized hyperbolic density using Generalized Gaussian scale mixtures, yielding a multivariate dependent anisotropic model. We also introduced the modeling of skew in ICA sources, deriving a general form of skewed multivariate Gaussian scale mixture, and an EM algorithm to update the location, drift, and structure parameters.

References

[1] D. F. Andrews and C. L. Mallows. Scale mixtures of normal distributions. *J. Roy. Statist. Soc. Ser. B*, 36:99–102, 1974.

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

[3] T. Cover and J. Thomas. *Elements of Information Theory.* John Wiley and Sons, Inc., 1991.

[4] L. Debnath. *Integral transforms and their applications.* CRC Press, 1995.

[5] T. Eltoft, T. Kim, and T.-W. Lee. Multivariate scale mixture of Gaussians modeling. In J. Rosca et al., editor, *Proceedings of the 6th International Conference on Independent Component Analysis*, Lecture Notes in Computer Science, pages 799–806. Springer-Verlag, 2006.

[6] T. Gneiting. Normal scale mixtures and dual probability densities. *J. Statist. Comput. Simul.*, 59:375–384, 1997.

[7] I. S. Gradshteyn and I. M. Ryzhik. *Table of integrals, series, and products.* Academic Press, Inc.: Orlando, 1980. Corrected and enlarged edition prepared by Alan Jeffrey.

[8] R. M. Gray. *Source Coding Theory.* Kluwer Academic Publishers: Boston, 1990.

[9] A. Hyvärinen and P. O. Hoyer. Emergence of phase- and shift-invariant features by decomposition of natural images into independent feature subspaces. *Neural Computation*, 12:1705–1720, 2000.

[10] A. Hyvärinen, P. O. Hoyer, and M. Inki. Topographic independent component analysis. *Neural Computation*, 13(7):1527–1558, 2001.

[11] T. Kim, H. Attias, S.-Y. Lee, and T.-W. Lee. Blind source separation exploiting higher-order frequency dependencies. *IEEE Transactions on Speech and Audio Processing*, 15(1), 2007.

[12] T. Kim, T. Eltoft, and T.-W. Lee. Independent vector analysis: An extension of ICA to multivariate components. In J. Rosca et al., editor, *Proceedings of the 6th International Conference on Independent Component Analysis*, Lecture Notes in Computer Science, pages 165–172. Springer-Verlag, 2006.

[13] C. L. Nikias and M. Shao. *Signal Processing with Alpha-Stable Distributions and Applications*. John Wiley & Sons, Inc., New York, 1995.

[14] R. T. Rockafellar. *Convex Analysis*. Princeton, 1970.