

Research Article

Semiparametric Gaussian Variance-Mean Mixtures for Heavy-Tailed and Skewed Data

Kai Cui

Department of Statistical Science, Duke University, Durham, NC 27708, USA

Correspondence should be addressed to Kai Cui, kc52@stat.duke.edu

Received 29 October 2012; Accepted 20 November 2012

Academic Editors: A. Guillou, É. Marchand, and C. Proppe

Copyright © 2012 Kai Cui. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

There is a need for new classes of flexible multivariate distributions that can capture heavy tails and skewness without being so flexible as to fully incur the curse of dimensionality intrinsic to nonparametric density estimation. We focus on the family of Gaussian variance-mean mixtures, which have received limited attention in multivariate settings beyond simple special cases. By using a Bayesian semiparametric approach, we allow the data to infer about the unknown mixing distribution. Properties are considered and an approach to posterior computation is developed relying on Markov chain Monte Carlo. The methods are evaluated through simulation studies and applied to a variety of applications, illustrating their flexible performance in characterizing heavy tails, tail dependence, and skewness.

1. Introduction

There is an increasing awareness of the importance of developing new classes of multivariate distributions that flexibly characterize heavy tails and skewness, while accommodating tail dependence. Such tail dependence arises in many applications and is a natural consequence of dependence in outlying events. Such dependence is well known to occur in financial data, communication networks, weather, and other settings, but is not adequately characterized by common approaches such as Gaussian copula models. Salmon [1] provides a compelling commentary on how reliance on a single measure of correlation in two variables based on a Gaussian copula may have played a substantial role in the financial crisis. We need statistical methods based on new classes of distributions that do not rely on such unrealistic assumptions but that are still tractable to apply even in moderate- to high-dimensional settings.

There is an existing literature relevant to this topic. Wang et al. [2] proposed a class of skew-symmetric distributions having probability density functions (pdfs) of the form

$2f(\mathbf{x})Q(\mathbf{x})$, where f is a continuous symmetric density and $Q : \mathbb{R}^n \rightarrow [0, 1]$ is a skewing function. Choosing f as normal leads to the skew normal class [3, 4], with other special cases corresponding to skew- t [5, 6], skew slash [7] and skew elliptical distributions [8]. These parametric models are useful in providing computationally tractable distributions that have parameters regulating skewness and kurtosis in the data. However, choosing a specific parametric family for f and Q can be challenging in practice, with different choices yielding potentially different results. Although one can potentially conduct model selection or averaging, this adds to the computational burden.

Alternatively, nonparametric approaches have been explored to handle heavy-tailed and skewed observations with more flexibility. For instance, mixtures of normal distributions have been widely used to approximate arbitrary distributions. Venturini et al. [9] uses mixtures of gamma distributions over the shape parameter to model heavy-tailed medical expenditure data. Mixtures of other heavy-tailed distributions have also been proposed. Such nonparametric density estimation approaches face substantial challenges in multivariate cases due to the curse of dimensionality. Fully nonparametric density estimation is almost too flexible in allowing densities that have arbitrary numbers of modes and complex shapes, which are difficult to estimate accurately based on available data in many cases. There has been some attempts to reduce dimensionality in multivariate density estimation using mixtures of factor analyzers [10] and alternative approaches, but nonetheless the curse is only partly thwarted by such efforts.

We focus on Gaussian variance-mean mixtures (GVMMs), introduced by Barndorff-Nielsen [11] as a flexible class of multivariate distributions induced through the following hierarchical model for $\mathbf{y}_i = (y_{i1}, \dots, y_{ip})'$,

$$\mathbf{y}_i = \boldsymbol{\mu} + \boldsymbol{\gamma}V_i + \sqrt{V_i}\boldsymbol{\Sigma}^{1/2}Z_i, \quad Z_i \sim N(0, I_p), V_i \sim G, \quad (1.1)$$

where $\boldsymbol{\mu} \in \mathbb{R}^p$ is a location parameter, $\boldsymbol{\gamma} \in \mathbb{R}^p$ is a drift or skewness parameter, $V_i \sim G$, G is a mixture distribution on $[0, \infty)$, $Z_i \perp V_i$, and $\boldsymbol{\Sigma}$ is a positive definite matrix. Conditions such as $|\boldsymbol{\Sigma}| = 1$ are typically imposed to avoid an unidentifiable scale factor. This model corresponds to a generalization of the multivariate normal distribution, which is obtained in the special case in which $\boldsymbol{\gamma} = 0$ and G is a point mass.

Current literature on Gaussian variance-mean mixtures is mostly focused on univariate models in which the mixing distribution G belongs to a parametric family. Some special cases corresponding to different choices of G include Student's t -test, Laplace, hyperbolic, normal inverse Gaussian, and variance gamma distributions, and such models have been applied broadly (see e.g., [11–16]). However, as with skew-symmetric distributions, limiting to a particular parametric class is clearly restrictive. In addition, some appealing subfamilies of GVMM, such as generalized hyperbolic (GH) distributions, are not analytically tractable. This is partially due to the flatness of the likelihood function [17], which makes it hard to obtain reliable parameter estimates without prior information or some form of penalization, even with large sample sizes [18].

We propose Bayesian semiparametric Gaussian variance-mean mixture models, in which the mixing distribution G is modeled nonparametrically to flexibly accommodate heavy tails and skewness, while letting the data inform about the appropriate distribution choice. Efficient Bayesian computational strategies are developed for reliable inference on parameters.

2. Proposed Modeling Framework

2.1. Semiparametric GVMM

Consider the p -dimensional multivariate continuous heavy-tailed and/or skewed observations $\mathbf{y}_i = (y_{i1}, \dots, y_{ip}) \sim f$, ($i = 1, \dots, N$). Our goal is to obtain a flexible GVMM model for the density f via modeling the mixing distribution G nonparametrically. To achieve this, we use Dirichlet process mixture (DPM) of generalized log-normal (log GN) prior with an unknown order s for G . Mathematically, the model is represented as

$$\begin{aligned} \mathbf{y}_i &= \boldsymbol{\mu} + \gamma V_i + \sqrt{V_i} \boldsymbol{\Sigma}^{1/2} \mathbf{Z}_i, \\ V_i &\sim \text{logGN}(m_i, \psi_i^{-1}, s), \quad \mathbf{Z}_i \sim N(0, I), \\ (m_i, \psi_i) &\sim H, \quad H \sim \text{DP}(\alpha_0 H_0), \end{aligned} \quad (2.1)$$

where α_0 is the DP precision and H_0 is the base measure for m_i and ψ_i , which we choose independent normal and inverse gamma, respectively.

The family of log GN distributions considered here was initially introduced by Vianelli [19], which can be obtained through the exponential transformation of a random variable that follows the generalized normal distribution [20, 21]. The pdf of a $\text{logGN}(m, \psi, s)$ distributed random variable V is given by

$$f(v) = \frac{s}{2v\psi^{1/2}\Gamma(1/s)} \exp\left(-\left|\frac{\log(v) - m}{\psi^{1/2}}\right|^s\right), \quad (2.2)$$

where $v > 0$, $m \in \mathcal{R}$, $\psi > 0$, and $s > 0$. log GN densities are more flexible than log-normal densities in including an additional parameter s controlling tail behavior, with log-normal corresponding to the special case in which $s = 2$ and double-exponential obtained by taking $s = 1$.

Although log GN distributions are appealing in providing a simple generalization of the log-normal which is more flexible in the tails, such distributions have been rarely implemented even in simpler settings due to the computational hurdles involved. Fortunately, for Bayesian posterior computation via MCMC we can rely on a data augmentation algorithm based on Fact 1, with the adaption to our setting using DPMs of generalized log-normals shown in Section 3.2.

Fact 1. Let V and W be two random variables such that

- (1) $f(vw) = (1/2\psi^{1/2}w^{1/s}v)I(|\log v - m/\psi^{1/2}|^s < w)$,
- (2) $W \sim \text{Gamma}(1 + 1/s, 1)$

then $V \sim \text{logGN}(m, \psi, s)$.

DPMs of Gaussians provide a highly flexible approximation to arbitrary densities. As a prior for f directly, DPMs of Gaussians have appealing asymptotic properties and lead to minimax optimal adaptive rates of convergence under some conditions on the true data-generating density, with these conditions unfortunately ruling out heavy-tailed densities [22].

As motivated above, our focus is on obtaining a flexible prior for multivariate heavy-tailed and skewed densities, which is not fully nonparametric in the sense of allowing multimodal and other very irregular density shapes but can flexibly approximate a broad class of unimodal densities without inducing restrictive tail constraints. Hence, it is not appropriate to choose a DPM of Gaussians directly for the density of the data f .

Expression (2.1) instead uses a DPM of log GNs for the mixture distribution G within the GVMM framework. By using such DPM prior and setting s unknown, we obtain a large flexibility of the mixing distribution and its tail decay, which can adopt diverse degrees of heavy tailness for the marginal density f of data, while allowing the data to fully infer about the unknown mixing distribution. After considering a broad variety of alternatives, we have found this specification to be excellent at capturing a rich variety of multivariate heavy-tailed and skewed distributions, while also allowing symmetric data. Some basic properties are detailed below.

2.2. Tail Behavior

It is important to understand the relationship between the tail behavior of the mixture distribution G and the induced tail behavior of f , the marginal distribution of y_i . We start by considering the univariate case. Theorem 2.2 of Barndorff-Nielsen and Sorensen [23] describes the relationship in the special case in which $\gamma = 0$ and hence the marginal distribution is symmetric.

Theorem 2.1 (see [13]). *Suppose $f(\mathbf{y})$ is the pdf of a Gaussian variance mixture as described in (1.1) with $\gamma = 0$, and that the tail of the mixing distribution G with pdf g satisfies*

$$g(V) \sim e^{-\psi_+ V} V^{\lambda-1} L(V), \quad \text{as } V \rightarrow +\infty, \quad (2.3)$$

where $\psi_+ \geq 0$ and L is a function of slow variation with $\lim_{V \rightarrow \infty} (L(aV)/L(V)) = 1$, for all $a > 0$. Then

- (1) if $\psi_+ = 0$, $f(\mathbf{y}) \sim |\mathbf{y}|^{2\lambda-1} L(\mathbf{y}^2)$, $|\mathbf{y}| \rightarrow \infty$,
- (2) if $\psi_+ > 0$, $f(\mathbf{y}) \sim |\mathbf{y}|^{\lambda-1} \exp(-(2\psi_+)^{1/2} |\mathbf{y}|) L(|\mathbf{y}|)$, $|\mathbf{y}| \rightarrow \infty$.

Observe that the tail behavior of a Gaussian variance mixture (when $\gamma = 0$) mainly depends on the tail behavior of the mixing distribution. To generalize this to arbitrary Gaussian variance-mean mixtures, we first introduce the following Lemma.

Lemma 2.2. *Suppose $f_G(\mathbf{y})$ is the pdf of a Gaussian variance mixture with mixing distribution G , and let φ denote the moment generating function of G . Then for all $\gamma \in \mathbb{R}$ s.t. $\varphi(\gamma^2/2) < \infty$:*

$$f_{G^*}(\mathbf{y}) = \frac{\exp(\gamma \mathbf{y})}{\varphi(\gamma^2/2)} f_G(\mathbf{y}) \quad (2.4)$$

is the pdf of a Gaussian variance-mean mixture with skewness parameter γ and mixing distribution G^* with pdf g^* satisfying $g^*(V) = (\exp(\gamma^2 V/2)/\varphi(\gamma^2/2))g(V)$, where g is the pdf of G . The converse is also true.

Proofs can be found in the appendix. This lemma provides a link between the tail behavior of a Gaussian variance-mean mixture and that of a Gaussian variance mixture, via the link between tails behaviors of the two mixing distributions. This relationship is used in the following theorem.

Theorem 2.3. *Suppose $f_G(y)$ is the pdf of a Gaussian variance-mean mixture as described in (1.1), with skewness parameter γ and mixing distribution G . If the tail of the mixing density g satisfies*

$$g(V) \sim e^{-\psi_+ V} V^{\lambda-1} L(V) \quad \text{as } V \rightarrow +\infty, \quad (2.5)$$

where $\psi_+ \geq 0$, and L is a function of slow variation with $\lim_{V \rightarrow \infty} (L(aV)/L(V)) = 1$, for all $a > 0$, then:

- (1) $f_G(y) \sim y^{\lambda-1} \exp(-(\sqrt{2\psi_+ + \gamma^2} - \gamma)y)L(y)$, as $y \rightarrow +\infty$,
- (2) $f_G(y) \sim |y|^{\lambda-1} \exp(-(\sqrt{2\psi_+ + \gamma^2} + \gamma)|y|)L(|y|)$, as $y \rightarrow -\infty$.

Theorem 2.3 shows that the tail behavior of a Gaussian variance-mean mixture also depends on that of the mixing distribution. Generally, heavier tails in the mixing distribution induce heavier tails of the Gaussian variance-mean mixture. Thus, by placing a DPM of generalized log-normals on the mixing distribution, we induce a prior on the density f with flexible degrees of tail decay. Another observation is that if the mixing distributions have subexponential tails (such as log-concave distributions), then the Gaussian variance-mean mixture also has sub-exponential tails, which also illustrates the limitation in the flexibility of using particular parametric cases of the Gaussian variance-mean mixtures to fit data.

2.3. Moments

To compute the moments of Gaussian variance-mean mixtures, we can directly apply the law of total cumulance. Let k_1 , k_2 , k_3 , and k_4 denote the expectation, variance, skewness, and kurtosis of the Gaussian variance-mean mixtures described in (1.1), and let k_1^* , k_2^* , k_3^* , and k_4^* denote those of the mixing distribution. Given they all exist, we have simply

$$k_1 = \mu + \gamma k_1^*, \quad k_2 = k_1^* + \gamma^2 k_2^*, \quad k_3 = \gamma k_2^* + \gamma^3 k_3^*. \quad (2.6)$$

More generally, we have $\varphi_f(s) = \exp(\mu s)\varphi(\gamma s + 1/2s^2)$, where φ_f and φ are moment generating functions of GVMM and the corresponding mixing distribution, respectively. So clearly, the existence of moments of mixing distributions indicates the existence of moments of Gaussian variance-mean mixtures, and γ can control expectation, variance, and kurtosis, in addition to being a skewness parameter.

3. Bayesian Computation

3.1. Priors

In the semiparametric GVMM framework, the $|\Sigma| = 1$ constraint is typically imposed to guarantee model identifiability. To improve computational efficiency in our proposed

Bayesian computational algorithm, we use a parameter-expansion approach in which priors are placed on parameters in an unidentifiable working model without the $|\Sigma| = 1$ constraint. We then include a postprocessing step to transform the parameters back to an identifiable inferential model, which includes the $|\Sigma| = 1$ constraint. A related strategy was used in Gaussian factor models by Ghosh and Dunson [24]. We use fairly diffuse priors for unidentifiable parameters, as an aid to mixing and because it is difficult to elicit priors for these parameters. However, we avoid completely noninformative priors because the flatness of the likelihood function in some GVMM models [13] can lead to unreliable inferences in the absence of some prior information or penalization. To address this problem, we propose an empirical Bayes approach to incorporate skewness information from the data in estimating hyperparameters for the skewness parameter γ .

In particular, we transform the original data to have a positive sample skewness and a unit sample variance. The data are first normalized and the sample skewness is calculated. If it is negative, we multiply the normalized data by a negative one. In conducting inferences, we transform back to the scale and sign of the original data. As GVMMs are closed under linear transformations, this will induce a GVMM. Because the transformed data are more likely to be right skewed or symmetric, we can more easily elicit a default weakly informative prior for the skewness parameter γ , which we choose to be normal with positive mean μ_γ , with a gamma hyperprior placed on μ_γ to improve prior robustness. For the order parameter s , a truncated inverse-gamma prior is used. Diffuse priors are placed on the remaining unknowns. To summarize

$$\begin{aligned} \gamma &\sim N(\mu_\gamma, \phi_\gamma), & \mu_\gamma &\sim \text{Gamma}(\alpha_\gamma, \beta_\gamma), & \phi_\gamma &\sim \text{Gamma}(a_\phi, b_\phi), & s &\sim \text{Gamma}(\alpha_s, \beta_s), \\ \mu &\sim N(0, \phi_\mu), & \Sigma &\sim \text{Inv-Wishart}(m_0, \Psi_0). \end{aligned} \quad (3.1)$$

These priors were used in an unconstrained working model for the transformed data and induce priors on the parameters in the identifiable inferential model having the $|\Sigma| = 1$ constraint.

As for the DPM of log GN prior for the mixing distribution G , for ease in computation, we use a similar stick-breaking representation of the DPM as proposed by Sethuraman [25] but with generalized log-normal instead of normal components, and truncated at M components following Ishwaran and James [26]. Furthermore, as pointed out before, since it is hard to directly update parameters of the generalized log-normal distributions, we further utilize Fact 1 to introduce augmented data $\mathbf{w} = (w_1, \dots, w_N)$ to improve the computational efficiency. To summarize, the data-augmented stick-breaking representation of the DPM of log GN prior is shown as follows:

$$\begin{aligned} f(v_i | \mathbf{m}, \boldsymbol{\psi}, \mathbf{K}, s, \mathbf{w}) &= \frac{1}{2\psi_{K_i}^{1/2} \omega_i^{1/s} v_i} I\left(\left|\frac{\log v_i - m_{K_i}}{\psi_{K_i}^{1/2}}\right|^s < w_i\right), \\ w_i | s &\sim \text{Gamma}\left(1 + \frac{1}{s}, 1\right). \end{aligned}$$

Note that these are equivalent to $V_i | \mathbf{m}, \boldsymbol{\psi}, \mathbf{K}, s \sim \text{logGN}(V_i; m_{K_i}, \psi_{K_i}^{-1}, s)$,

$$\begin{aligned} K_i | \mathbf{p} &\sim \sum_{k=1}^M p_k \delta_k(\cdot), \quad i = 1, \dots, N, \\ (m_k, \psi_k) &\sim H_0(m, \psi), \quad k = 1, \dots, M, \\ p_1 = \nu_1, \quad p_k &= \nu_k \prod_{l=1}^{k-1} (1 - \nu_l), \quad k = 2, \dots, M-1, \\ \nu_k &\sim \text{Beta}(1, \alpha_0), \quad k = 1, \dots, M-1, \quad \nu_M = 1, \end{aligned} \quad (3.2)$$

where $\mathbf{K} = (K_1, \dots, K_N)$, $\mathbf{m} = (m_1, \dots, m_M)$, $\boldsymbol{\psi} = (\psi_1, \dots, \psi_M)$, $\mathbf{p} = (p_1, \dots, p_M)$ and for H_0 , we use independent normal and inverse gamma for m and ψ , respectively. Here, K_i is augmented data representing the mixture component index for observation i .

3.2. Full Conditionals and Posterior Analysis

Given the model and priors as specified by (2.1), (3.1), and (3.2), we use a data-augmented Gibbs sampler to update the unknown quantities, including parameters in the GVMM framework (μ , γ , and Σ), parameters in the DPM of log GN (\mathbf{m} , $\boldsymbol{\psi}$, and s), augmented data (\mathbf{K} and \mathbf{w}), hyperparameters, and the mixing variables (V_i). The Gibbs sampler is computationally efficient and mixes rapidly as most of the full conditional distributions have closed forms, except those for μ_γ , V_i , and s , which are all univariate and updated using Metropolis-Hastings steps within the Gibbs sampler. Key steps in each Gibbs sampler iteration are listed as follows.

- (I) Samples μ , γ , and Σ : given normal priors for (μ, γ) and inverse-Wishart prior for Σ and the model that $y_i | V_i \sim N(\mu + \gamma V_i, V_i \Sigma)$, (μ, γ) is sampled from conditionally normal distribution, and Σ from a conditionally inverse Wishart.
- (II) Samples \mathbf{m} and $\boldsymbol{\psi}$: given the following priors are used for m_k and ψ_k for the k th log-normal component, $k = 1, \dots, M$:

$$m_k \sim N(\mu_m, \phi_m), \quad \psi_k \sim \text{Inv-Gamma}(\alpha_\psi, \beta_\psi). \quad (3.3)$$

Sample m_k , $k = 1, \dots, M$, from conditionally truncated normal distribution:

$$m_k | \dots \sim N(\mu_m, \phi_m) I\left(\max_{l:K_l=k} \left\{ \log V_l - \psi_k^{1/2} \omega_l^{1/s} \right\} < m_k < \min_{l:K_l=k} \left\{ \log V_l + \psi_k^{1/2} \omega_l^{1/s} \right\}\right) \quad (3.4)$$

and sample ψ_k from conditionally truncated inverse-gamma distribution:

$$\psi_k | \dots \sim \text{Inv-Gamma}\left(\alpha_\psi + \frac{N_k}{2}, \beta_\psi\right) I\left(\psi^{1/2} > \max_{l:K_l=k} \left\{ \frac{|\log V_l - m_k|}{\omega_l^{1/s}} \right\}\right), \quad (3.5)$$

where N_k is the total number of V_i in the k th mixture component.

(III) Sample \mathbf{w} . w_i , $i = 1, \dots, N$, is sample from conditionally truncated exponential distribution:

$$w_i | \dots \sim \text{Exp}(1) I \left(w_i > \left(\frac{|\log V_i - m_{K_i}|}{\psi_{K_i}^{1/2}} \right)^s \right). \quad (3.6)$$

(IV) Sample s : given the Gamma(α_s, β_s) for s , the full conditional distribution is sampled using Metropolis-Hastings algorithm, from the following full conditional truncated kernel:

$$f(s | \dots) \propto \frac{s^{\alpha_s + N - 1} \exp(-\beta_s s)}{\Gamma^N(1/s)} I \left(\max_{i \in S^-} (0, a_i) < s < \min_{i \in S^+} (a_i) \right), \quad (3.7)$$

where $a_i = (\log w_i) / \log(|\log V_i - m_{K_i}| / \psi_{K_i}^{1/2})$, $i = 1, \dots, N$.

$S^- = \{i : \log(|\log V_i - m_{K_i}| / \psi_{K_i}^{1/2}) < 0\}$, and $S^+ = \{i : \log(|\log V_i - m_{K_i}| / \psi_{K_i}^{1/2}) > 0\}$.

(V) Sampling \mathbf{K} : K_i , $i = 1, \dots, N$ is sampled from the conditionally multinomial (MN) distribution, with

$$\Pr(K_i = k) | \dots \propto \frac{p_k}{\psi_k^{1/2}} \exp \left\{ - \left(\frac{|\log V_i - m_k|}{\psi_k^{1/2}} \right)^s \right\}. \quad (3.8)$$

(VI) ν_k , $k = 1, \dots, M - 1$ are updated from conditionally beta distribution as shown in Ishwaraman and James [26]. Also update univariate V_i , $i = 1, \dots, N$ using a Metropolis-Hastings step within the Gibbs sampler.

4. Simulation Study and Real Data Analysis

4.1. Univariate Semiparametric GVMM

To test the semiparametric framework, a simulated dataset from univariate GVMM is first modeled. Specifically, observations y_i ($i = 1, \dots, N = 1000$) are generated from model (2.1), with the true values of the parameters as shown in

$$\mu = 2, \quad \gamma = 2, \quad \Sigma = 1, \quad V_i \sim \text{Gamma}(3, 1), \quad \forall i = 1, \dots, N. \quad (4.1)$$

The histogram of simulated data is shown in Figure S2(B) in Supplementary Material available online at doi:10.5402/2012/345784, which shows significant heavy tailness and skewness, with sample kurtosis and skewness being 4.60 and 1.01, respectively.

For Bayesian inference, we preprocess the original simulated data and place priors as described previously. We run the MCMC for 10000 iterations with the first 5000 as burn-in. Several aspects of the posterior distributions are analyzed to evaluate the model fitting. First of all, posterior samples of V_i and parameters allow us to reconstruct and visualize the unknown mixing distribution G . As shown in Figure S1, a comparison between the true

Table 1: Posterior quantile estimation to show the model fitting. Posterior quantile C.I.s are obtained by simulating 200 reconstructed datasets based (each consisting of 5000 data points) on posterior samples of unknown quantities, each dataset giving one set of quantile point estimates. Observed quantiles are obtained from the 1000 observed simulated data.

Quantiles	Observed quantile	Posterior mean	Posterior 95% C.I.
2.5%	-1.474	-1.492	[-1.535, -1.450]
5%	-1.328	-1.322	[-1.363, -1.284]
25%	-0.738	-0.711	[-0.742, -0.676]
50%	-0.175	-0.162	[-0.190, -0.124]
75%	0.555	0.534	[0.492, 0.584]
95%	1.783	1.871	[1.776, 1.985]
97.5%	2.394	2.414	[2.283, 2.566]

Table 2: Quantile estimates are obtained from models fitted to the observed dataset using maximum likelihood estimation. To obtain the maximum likelihood estimators in skewed-Gaussian and t -distributions, the `snR` package is used [27].

Quantiles	Observed	Gaussian	Skewed Gaussian	Skewed t
2.5%	-1.474	-1.960	-1.451	-1.446
5%	-1.328	-1.645	-1.306	-1.300
25%	-0.738	-0.674	-0.751	-0.749
50%	-0.175	0.000	-0.173	-0.185
75%	0.555	0.674	0.590	0.566
95%	1.783	1.645	1.887	1.895
97.5%	2.394	1.960	2.339	2.379

mixing distribution (Gamma(3,1), panel (B)) and 100 reconstructed mixing distributions (panel (A)) shows significant similarity, indicating that the model can effectively capture the underlying structure. This is further illustrated by the posterior distributions of μ and γ , which have 95% C.I.s being [0.8676, 3.2390] and [1.2532, 2.9660], and posterior means being 2.0804 and 2.1523, respectively, which also very well cover the true values. It is worth mentioning that later in the real data analysis where the true values are unknown, the posterior distribution of γ provides us with a useful tool to assess skewness, with significant positive values of γ indicating skewness.

Furthermore, we can reconstruct the dataset based on the posterior samples of model parameters and the mixing distribution G and directly visualize whether the posterior predictive distribution resembles the observed one. We reconstructed 200 such datasets, each containing 5000 data points. While the reconstructed datasets resemble the observed one (shown in Figure S2), we specifically looked at the posterior quantile estimates of the fitted semiparametric GVMM, which is done by getting the 95% posterior C.I.s for a series of quantiles based on the 200 reconstructed datasets. This is shown in Table 1 and compared to the quantile estimates obtained by fitting other models, such as normal, skewed normal, and skewed t -distributions (Table 2). Compared to the fact that simple Gaussian model fails to capture neither heavy tailness and skewness, our GVMM model fits the data as well as skewed-Gaussian and skewed t -distributions, while maintaining unique advantages of relatively easy forms, convenient sampling, and interpretability.

Table 3: Posterior quantile estimation to show the model fitting. Posterior quantile C.I.s are obtained by simulating 200 reconstructed datasets based (each consisting of 5470 data points) on posterior samples of unknown quantities, each dataset giving one set of quantile point estimates. Real data quantiles are obtained from the 5470 observed S&P 500 returns.

Quantiles	Real data quantile	Posterior mean	Posterior 95% C.I.
2.5%	-1.965	-1.965	[-2.097, -1.867]
5%	-1.511	-1.513	[-1.589, -1.440]
25%	-0.493	-0.495	[-0.534, -0.469]
50%	-0.0315	-0.0175	[-0.0421, 0.0087]
75%	0.467	0.476	[0.440, 0.512]
95%	1.537	1.560	[1.482, 1.649]
97.5%	2.056	2.064	[1.943, 2.225]

4.2. Modeling the S&P 500 Returns

It is well known that stock returns do not always confirm well with a Gaussian distribution. Modeling both heavy tailness and asymmetry of returns is becoming important in economics and finance. Here, we look at daily returns of Standard & Poor's 500 Composite (S&P 500) index from 01/02/1990 to 09/13/2011. Totally 5470 observations are shown in Figure S3(A), with sample skewness being 0.189 (after preprocessing), which suggests that the return distribution may be slightly right skewed.

Similar univariate semiparametric GVMM and prior setup are applied to the dataset to access the capability of the model in capturing the return distribution. To evaluate the model fitting, we also reconstructed 200 datasets based on the posterior samples of unknown quantities (each consisting of 5470 observations), and a quick comparison (Figure S3) between the observed and reconstructed datasets shows a significant similarity, indicating that our model captures the return distribution well. We also look at posterior quantile estimates based on the 200 simulated datasets (Table 3), which further illustrate the goodness of fit of the model.

Furthermore, we specifically look at the posterior distribution of γ . Generally speaking, when the sample skewness is relatively small (as in this case), it is difficult to claim whether the true distribution is skewed or symmetric, because samples from symmetric heavy-tailed distributions can also exhibit significant sample skewness due to the presence of extreme values with a finite sample size. However, one feature of the Bayesian semiparametric GVMM framework is that the skewness can be simultaneously inferred by looking at the sign of the γ parameter, which can test the actual existence of skewness directly against an "artificial sample skewness" due to heavy tailness. To illustrate this, the histogram of 5000 posterior samples of γ is shown in Figure 1, which gives a 95% posterior C.I. [0.0033, 0.0838] and thus claims that the distribution is slightly right skewed at the 95% confidence level. As a striking comparison, we generated 5000 samples from t -distribution ($df = 3$, shown in the appendix). The sample skewness is 0.228, which appears to suggest a slight right skewness. However, when we fit the data using our Bayesian semiparametric GVMM and look at the posterior distribution of γ (shown in Figure 2, as compared to Figure 1), the posterior 95% C.I. is [-0.0371, 0.0757], with posterior mean at 0.0213, which argue against the existence of skewness and suggest that the sample skewness is due to the fact that t -distribution is heavy tailed.

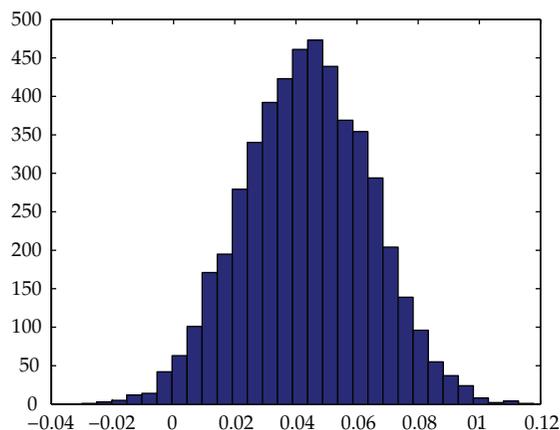


Figure 1: Posterior distribution of γ is presented. S&P 500 daily returns are modeled via Bayesian semiparametric GVMM, and γ confirms the existence of skewness with the sample skewness being 0.189.

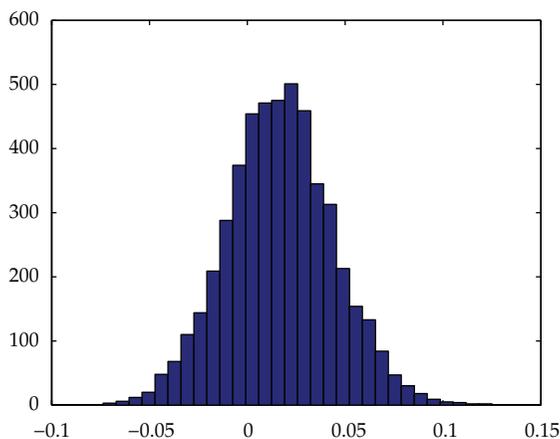


Figure 2: Posterior distribution of γ is presented. A set of random samples from t -distribution with $df = 3$ are modeled via Bayesian semiparametric GVMM, and γ confirms no skewness although the sample skewness is 0.228.

4.3. Modeling Multivariate Monthly Precipitation

There has also been a growing interest for flexible families of non-Gaussian distributions allowing skewness and heavy tails in environmental science and climatology, as more heavy-tailed and skewed data are observed practically. Specifically, it is well known that monthly rainfall is strongly skewed to the right with high positive values of skewness coefficients (e.g., [28]). Various distributions have been suggested to model the precipitation data, among which there are the exponential, gamma (e.g., [29, page 98]), log-normal (e.g., [30]), and log-skewed-normal/ t -distributions [31]. However, most of the studies are focusing on univariate modeling, and there has been a little physical justification to why a specific distribution is used. Although skewed-Gaussian/ t -type distributions have been extended to the multivariate setting (see e.g., [27, 31]), they do not handle substantial skewness well

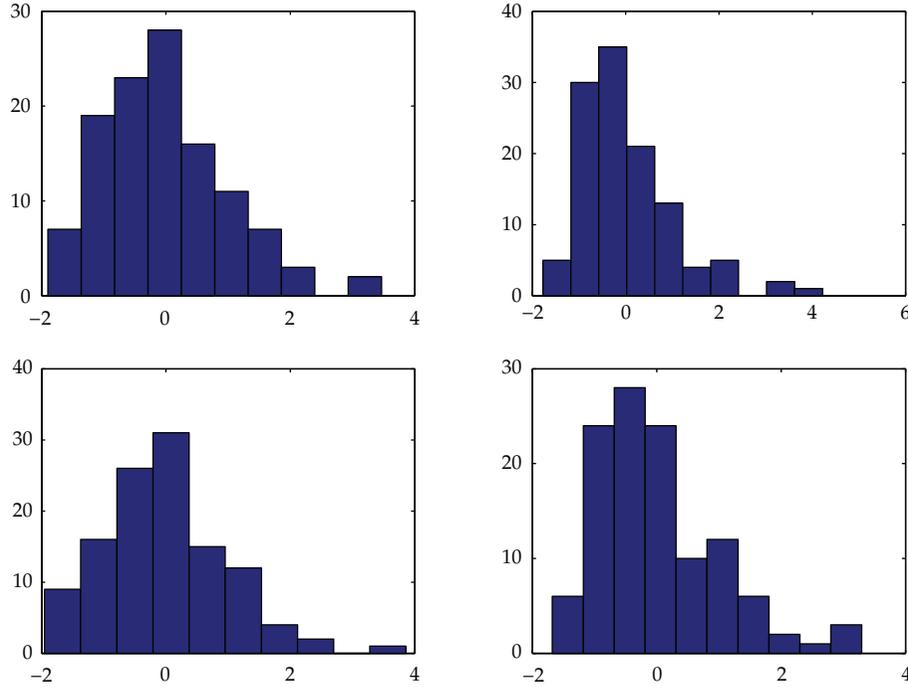


Figure 3: Monthly log precipitation data for July from 1895 to 2010 (116 observations) obtained from four stations in North Carolina show heavy right skewness.

[18]. We are motivated to consider applying the Bayesian semiparametric GVMM framework to model multivariate precipitation data. One appealing feature is that the extension to multivariate case is both straightforward and interpretable.

US national and regional precipitation data are publicly available from the United States Historical Climatology Network (USHCN). For the purpose of exposition, we used monthly precipitation data measured in inches from four local stations (Albemarle, Chapel Hill, Edenton, and Elizabeth City) in the state of North Carolina, for the period from 1895 through 2010 (116 data per station for each month). Figure 3 presents the histogram of log monthly precipitation data for July every year. Data from all four stations exhibit right skewness, with sample skewness coefficients being 0.675, 1.440, 0.632, and 0.971, respectively.

We fit the semiparametric multivariate GVMM (2.1) to the data (dimension $p = 4$).

We run the Markov Chain for 10000 iterations, which shows good mixing and convergence, and discarded the first 5000 as burn-in. To illustrate the model fitting, we reconstruct a precipitation dataset with 5000 observations based on posterior samples of all unknown quantities and compare the reconstructed posterior predictive distribution with the observed. Specifically, we test both whether the marginal univariate distributions of each dimension and the covariance structure are captured correctly. As shown in Figure 4, the curves represent the corresponding univariate kernel smoothing density estimates of the posterior predictive distributions based on 5000 samples from the fitted model, which show good fitting to the observed dataset shown with the histograms. The goodness of fit is further illustrated by the PP-plot between the observed and fitted distributions (Figure 5). Covariance structure is also modeled well (a comparison between Figures 6 and 7).

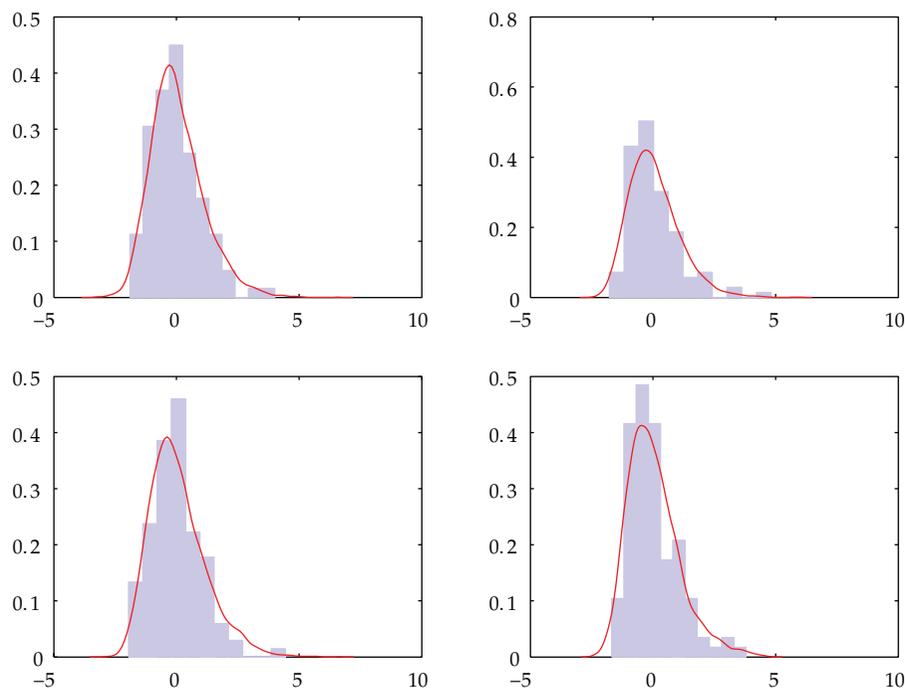


Figure 4: Monthly log-precipitation data for July from 1895 to 2010 (116 observations) obtained from four stations in North Carolina (shown in histogram) are fitted using Bayesian semiparametric GVMM. Red line shows the kernel density of fitted distributions for the stations, estimated from 5000 posterior predictive samples of the fitted GVMM.

All of these suggest that our framework effectively captures the underlying structure of the precipitation data.

As a comparison, we also fitted the log-precipitation data with multivariate skewed t -distributions [32], which have also been used to model skewed and heavy-tailed data, with two tails behaving as polynomials. The fitted model is obtained via maximum likelihood estimation using the `snR` package provided by Azzalini [27], and both the marginal distributions and covariance structure of the fitted multivariate skewed t -model are compared to those of the observed dataset. We can find that although the multivariate skewed t -distribution captures most of the univariate marginal distributions with a slightly lighter right tails and heavier left tails (Figure S4), it failed to find the correct covariance structure (Figure S5 compared to Figures 6 and 7). This could result from getting stuck in a local maximum and obtaining a suboptimal solution due to the relatively complicated likelihood function of multivariate skewed t -distributions. In any case, our Bayesian semiparametric multivariate GVMM seems to provide a computationally and structurally simpler yet more effective family of distributions for multivariate skewed and heavy-tailed data.

5. Discussion

This paper proposes the use of Bayesian semiparametric Gaussian variance-mean mixtures as a flexible, interpretable, and computationally tractable model for heavy-tailed and skewed

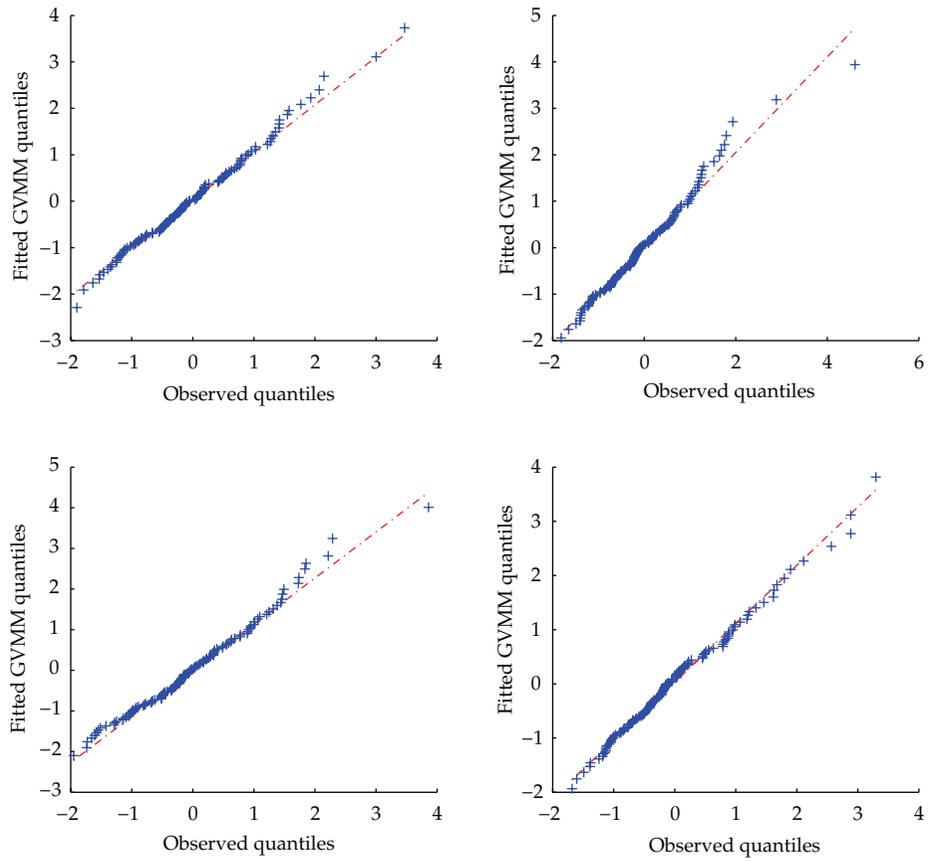


Figure 5: PP-plots for the Bayesian semiparametric GVMM model fitted to log-precipitation for July in all four stations.

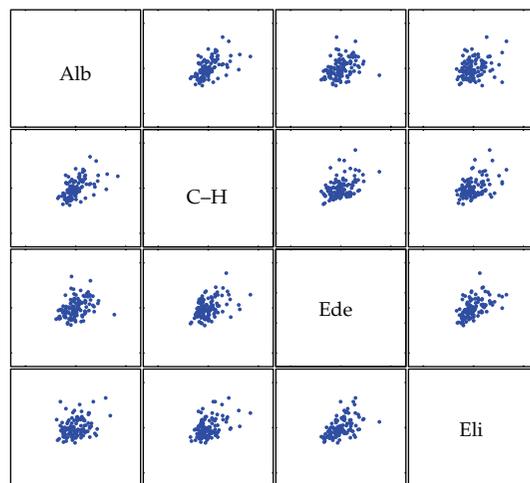


Figure 6: Sample covariance structure of monthly log-precipitation data from four local stations in the state of North Carolina. Alb: Albemarle, C-H: Chapel Hill, Ede: Edenton, and Eli: Elizabeth City.

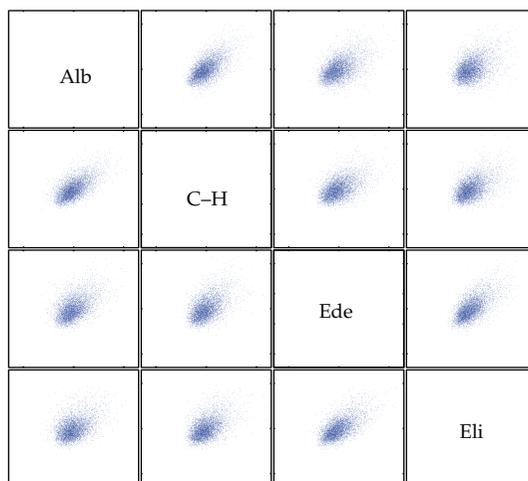


Figure 7: Covariance structure of monthly log-precipitation when fitted with Bayesian semiparametric GVMM. Alb: Albemarle, C-H: Chapel Hill, Ede: Edenton, and Eli: Elizabeth City are four stations in the state of North Carolina.

observations. The model assumes the mixing distribution G in the general Gaussian variance-mean mixtures to be unknown, so the data inform about the appropriate distribution choice, the degree of skewness, heaviness of tails, and the shape of the distributions and thus provide a more flexible framework for heavy-tailed and skewed data analysis. Although we test the model with univariate and multivariate simulated and real data, the broader question of scaling the framework to higher dimensions, possibly combined with factor models for sparse modeling, is interesting and challenging. Under the same scenario, the assumption that a single mixing variable is controlling the tails and skewness of all dimensions seems restrictive, although this may be an acceptable assumption with low dimensions. Some hierarchical nonparametric models that allow multiple mixing variable/distributions may help in this case, but how to define an identifiable parametrization will definitely be an issue whenever multiple mixing variables are assumed, which is also worth more thoughts and theory to support. On the other hand, we consider time-varying semiparametric GVMM a natural extension to the GVMM framework, taking advantage of the interpretability of model parameters. However, the more general class of spatial-temporal models, where additional structures and sources of information are included, is still challenging and yet unexplored.

Appendix

Proofs

Proof of Fact 1. Let V and W be two random variables such that

$$(1) f(v | w) = (1/2\psi^{1/2}w^{1/s}v)I(|\log v - m/\psi^{1/2}|^s < w),$$

$$(2) W \sim \text{Gamma}(1 + 1/s, 1)$$

then

$$\begin{aligned}
f(v) &= \int f(v | w) f(w) dw \\
&\propto \int \frac{1}{w^{1/s} v} I\left(\left|\frac{\log v - m}{\psi^{1/2}}\right|^s < w\right) \cdot w^{1/s} \exp(-w) dw \\
&\propto \int \frac{1}{v} I\left(\exp(-w) < \exp\left(-\left|\frac{\log v - m}{\psi^{1/2}}\right|^s\right)\right) \exp(-w) dw \\
&\propto \frac{1}{v} \exp\left(-\left|\frac{\log v - m}{\psi^{1/2}}\right|^s\right)
\end{aligned} \tag{A.1}$$

which is the kernel of a log GN(m, ψ , and s) as defined in (2.2). \square

Proof of Lemma 2.2. Given p -dimensional observation y from a Gaussian variance-mean mixtures described in (1.1), let f_G^* denote the pdf of the Gaussian variance-mean mixture with mixing distribution G^* , and let f_G denote the pdf of the Gaussian variance mixture ($\gamma = 0$) with mixing distribution G . Without the loss of generality, we considered the case with $\mu = 0$ (otherwise, a linear transformation $\mu + y$ will complete the proof)

$$\begin{aligned}
f(y) &\propto \int_0^\infty V^{-p/2} \exp\left(-\frac{1}{2} y' \frac{1}{V} \Sigma^{-1} y g(V) dV\right) \\
&\propto \int_0^\infty V^{-p/2} \exp\left(-\frac{1}{2} y' \frac{1}{V} \Sigma^{-1} y\right) \exp\left(-\frac{1}{2} \gamma' \Sigma^{-1} \gamma V\right) \exp\left(\frac{1}{2} \gamma' \Sigma^{-1} \gamma V\right) g(V) dV.
\end{aligned} \tag{A.2}$$

Let $\delta = (1/2) \gamma' \Sigma^{-1} \gamma$, with $\delta = (1/2) \gamma^2 \geq 0$ in the univariate case, and define

$$g^*(V) = \frac{g(V) \exp(\delta V)}{\varphi(\delta)}, \tag{A.3}$$

where g and g^* are densities for G and G^* , respectively, and φ is the moment generating function for g , with $\varphi(\delta) < \infty$. Plug g^* in, and we can see that

$$f_{g^*}(y) = \frac{\exp(\gamma y)}{\varphi(\gamma^2/2)} f_g(y). \tag{A.4}$$

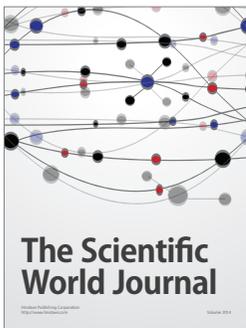
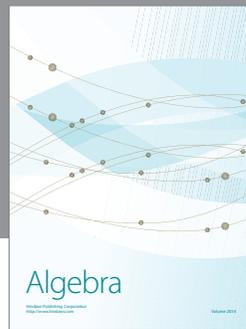
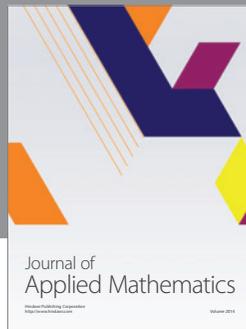
\square

References

- [1] F. Salmon, "The formula that killed wall street," *Significance*, vol. 9, no. 1, pp. 16–20, 2012.
- [2] J. Wang, J. Boyer, and M. G. Genton, "A skew-symmetric representation of multivariate distributions," *Statistica Sinica*, vol. 14, pp. 1259–1270, 2004.
- [3] A. Azzalini, "A class of distributions which includes the normal ones," *Scandinavian Journal of Statistics*, vol. 12, no. 2, pp. 171–178, 1985.

- [4] A. Azzalini and A. Dalla Valle, "The multivariate skew-normal distribution," *Biometrika*, vol. 83, no. 4, pp. 715–726, 1996.
- [5] S. K. Sahu, D. K. Dey, and M. D. Branco, "A new class of multivariate skew distributions with applications to Bayesian regression models," *The Canadian Journal of Statistics*, vol. 31, no. 2, pp. 129–150, 2003.
- [6] A. K. Gupta, "Multivariate skew t -distribution," *Statistics*, vol. 37, no. 4, pp. 359–363, 2003.
- [7] J. Wang and M. G. Genton, "The multivariate skew-slash distribution," *Journal of Statistical Planning and Inference*, vol. 136, no. 1, pp. 209–220, 2006.
- [8] M. G. Genton and N. M. R. Loperfido, "Generalized skew-elliptical distributions and their quadratic forms," *Annals of the Institute of Statistical Mathematics*, vol. 57, no. 2, pp. 389–401, 2005.
- [9] S. Venturini, F. Dominici, and G. Parmigiani, "Gamma shape mixtures for heavy-tailed distributions," *The Annals of Applied Statistics*, vol. 2, no. 2, pp. 756–776, 2008.
- [10] M. Chen, J. Silva, J. Paisley, C. Wang, D. Dunson, and L. Carin, "Compressive sensing on manifolds using a nonparametric mixture of factor analyzers: algorithm and performance bounds," *IEEE Transactions on Signal Processing*, vol. 58, no. 12, pp. 6140–6155, 2010.
- [11] O. Barndorff-Nielsen, "Exponentially decreasing distributions for the logarithm of particle size," *Proceedings of the Royal Society of London A*, vol. 353, no. 1674, pp. 401–419, 1977.
- [12] O. E. Barndorff-Nielsen, J. L. Jensen, and M. Sorensen, "Parametric modelling of turbulence," *Philosophical Transactions*, vol. 332, no. 1627, pp. 439–455, 1990.
- [13] O. Barndorff-Nielsen, "The hyperbolic distribution in statistical physics," *Scandinavian Journal of Statistics*, vol. 9, no. 1, pp. 43–46, 1982.
- [14] O. E. Barndorff-Nielsen, "Normal inverse Gaussian distributions and stochastic volatility modelling," *Scandinavian Journal of Statistics*, vol. 24, no. 1, pp. 1–13, 1997.
- [15] O. Arslan and A. İ. Genç, "The skew generalized t (SGT) distribution as the scale mixture of a skew exponential power distribution and its applications in robust estimation," *Statistics*, vol. 43, no. 5, pp. 481–498, 2009.
- [16] O. Arslan, "An alternative multivariate skew Laplace distribution: properties and estimation," *Statistical Papers*, vol. 51, no. 4, pp. 865–887, 2010.
- [17] E. Eberlein and K. Prause, "The generalized hyperbolic model: financial derivatives and risk measures," in *Mathematical Finance C Bachelier Congress 2000*, pp. 245–267, Springer, Berlin, Germany, 1998.
- [18] K. Aas and I. H. Haff, "The generalized hyperbolic skew students t -distribution," *Journal of Financial Econometrics*, vol. 4, no. 2, pp. 275–309, 2006.
- [19] S. Vianelli, "The family of normal and lognormal distributions of order r ," *Metron*, vol. 41, no. 1-2, pp. 3–10, 1983.
- [20] M. Varanasi and B. Aazhang, "Parametric generalized gaussian density estimation," *Journal of the Acoustical Society of America*, vol. 86, no. 4, pp. 1404–1415, 1989.
- [21] S. Nadarajah, "A generalized normal distribution," *Journal of Applied Statistics*, vol. 32, no. 7, pp. 685–694, 2005.
- [22] W. Shen and S. Ghosal, "Adaptive bayesian multivariate density estimation with dirichlet mixtures," submitted.
- [23] O. E. Barndorff-Nielsen, J. Kent, and M. Sørensen, "Normal variance-mean mixtures and z distributions," *International Statistical Review*, vol. 50, no. 2, pp. 145–159, 1982.
- [24] J. Ghosh and D. B. Dunson, "Default prior distributions and efficient posterior computation in Bayesian factor analysis," *Journal of Computational and Graphical Statistics*, vol. 18, no. 2, pp. 306–320, 2009.
- [25] J. Sethuraman, "A constructive definition of Dirichlet priors," *Statistica Sinica*, vol. 4, no. 2, pp. 639–650, 1994.
- [26] H. Ishwaran and L. F. James, "Gibbs sampling methods for stick-breaking priors," *Journal of the American Statistical Association*, vol. 96, no. 453, pp. 161–173, 2001.
- [27] A. Azzalini, *R Package sn: The Skew-Normal and Skew-t Distributions (Version 0.4–17)*, Università di Padova, Padova, Italia, 2011.
- [28] H. Ishwaran and L. F. James, "Gibbs sampling methods for stick breaking priors," *Journal of the American Statistical Association*, vol. 96, no. 453, pp. 161–173, 2001.

- [29] D. S. Wilks, *Statistical Methods in the Atmospheric Sciences*, vol. 91 of *International Geophysics*, Academic Press, 2nd edition, 2005.
- [30] E. L. Crow and K. Shimizu, *Lognormal Distributions: Theory and Applications*, Marcel Dekker, 1988.
- [31] Y. V. Marchenko and M. G. Genton, "Multivariate log-skew-elliptical distributions with applications to precipitation data," *Environmetrics*, vol. 21, no. 3-4, pp. 318–340, 2010.
- [32] B. E. Hansen, "Autoregressive conditional density estimation," *International Economic Review*, vol. 35, no. 3, pp. 705–730, 1994.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

