Biometrika (2016), **103**, 1, *pp*. 1–9
Printed in Great Britain Advance Access publication on 31 July 2016

Global-local mixtures

BY ANINDYA BHADRA
Department of Statistics, Purdue University,
250 N. University St., West Lafayette, Indiana 47907.
bhadra@purdue.edu.

JYOTISHKA DATTA
Department of Mathematical Sciences, University of Arkansas,
Fayetteville, Arkansas 72701.
jyotishka.datta@gmail.com

NICHOLAS G. POLSON AND BRANDON WILLARD
Booth School of Business, The University of Chicago, 5807 S. Woodlawn Ave., Chicago, Illinois 60637, U.S.A.
ngp@chicagobooth.edu AND brandonwillard@gmail.com

SUMMARY

Global-local mixtures are derived from the Cauchy-Schlömilch and Liouville integral transformation identities. We characterize well-known normal-scale mixture distributions including the Laplace or Lasso, logit and quantile as well as new global-local mixtures. We also apply our methodology to convolutions that commonly arise in Bayesian inference. Finally, we conclude with a conjecture concerning bridge and uniform correlation mixtures.

Some key words: Global-local mixture, Scale mixture, Stable laws, Bayes regularization, Lasso, Quantile, Logistic, Cauchy, Convolutions.

1. INTRODUCTION

Many statistical problems involve regularization penalties derived from global-local mixture distributions (Polson et al., 2011; Hans, 2011; Bhadra et al., 2015). A global-local mixture density, denoted by $p(x_1, \ldots, x_p)$, takes the form

$$p(x_1, \ldots, x_p) = \int_0^\infty \prod_{i=1}^p p(x_i \mid \tau) p(\tau) d\tau,$$

where

$$p(x_i \mid \tau) = \int_0^\infty p(x_i \mid \lambda_i, \tau) p(\lambda_i \mid \tau) d\lambda_i$$

is a local mixture and $p(x_1, \ldots, x_p)$ is a global mixture over $\tau \sim p(\tau)$. There is great interest in analytically calculating $p(x_i \mid \tau)$, and the associated regularization penalty $\phi(x_i, \tau) = -\log p(x_i \mid \tau)$. Convolution mixture of the form $p(x_i \mid \tau) = \int p(x_i - \lambda_i) p(\lambda_i) d\lambda_i$ are also of interest. We show how the Cauchy-Schlömilch and Liouville transformations can be used to derive closed-form global-local mixtures. We start by stating two key integral identities: the Cauchy-Schlömilch transformation

$$\int_0^\infty f\left\{(ax - bx^{-1})^2\right\} dx = \frac{1}{2a}\int_0^\infty f(y^2)dy \ , \tag{1}$$

and the Liouville transformation

$$\int_0^\infty f\left(ax + \frac{b}{x}\right)\frac{dx}{\sqrt{x}} = \frac{1}{\sqrt{a}}\int_0^\infty f\left(2\sqrt{ab} + y\right)\frac{dy}{\sqrt{y}}, \quad \text{for} \quad a, b > 0 \ . \tag{2}$$

See Boros et al. (2006); Baker (2008); Amdeberhan et al. (2010); Jones (2014) for further discussion. Equation (1) follows from the simple transformation $t = b/(ax)$:

$$I = \int_0^\infty f\left\{(ax - b/x)^2\right\} dx = \int_0^\infty f\left\{(at - b/t)^2\right\} \frac{b}{at^2}dt.$$

Adding the two terms in the last equality yields $2I = \int_0^\infty f\left\{(at - b/t)^2\right\}\left\{1 + \frac{b}{at^2}\right\} dt$ and transforming $y = b/t - at$ gives $dy = -a(1 + \frac{b}{at^2})dt$, which yields $I = (2a)^{-1}\int_0^\infty f(y^2)dy$ as required. A useful generalization of the Cauchy-Schlömilch transformation is

$$\int_0^\infty f\left[\{x - s(x)\}^2\right] dx = \int_0^\infty f(y^2)dy \tag{3}$$

where $s(x) = s^{-1}(x)$ is a self-inverse function such as $s(x) = b/x$ or $s(x) = -a^{-1}\log\{1 - \exp(ax)\}$. The Liouville transformation follows in a similar manner. The proof is along the same lines, so, for the sake of brevity, we do not reproduce it here.

These identities can be used to construct new global-local mixture distributions. Let $f(x) = 2g\{t(x)\}$ and give $t(x)$ the form $x - s(x)$, where $s : \Re^+ \to \Re^+$ is a self-inverse, onto and monotone decreasing function. Together with the Cauchy-Schlömilch transformation, we have a rather surprising way to represent the resulting $g\{t(x)\}$ as a global-local scale mixture.

Jones (2014) shows that only a few choices of $t(x)$ leads to fully tractable formulae for its inverse $t^{-1} = \Pi$ and the integral $\Pi(y) = \int_{-\infty}^y \pi(\omega)d\omega$. Two special choices are the t-distribution with 2 degrees of freedom and the logistic.

t-distribution $\Pi_T(y) = (1/2)(y + \sqrt{4b + y^2}) \Rightarrow \Pi_T^{-1}(x) = t_T(x) = x - b/x$

Logistic $\Pi_L(y) = a^{-1}\log(1 + e^{ay}) \Rightarrow \Pi_L^{-1}(x) = t_L(x) = a^{-1}\log(e^{ax} - 1)$

Now, the integral identity in (1) shows that if $f(x), x \geq 0$ is a density function, so is $g(x) = 2af(|ax - b/x|), x > 0$, where $f(\cdot)$ and $g(\cdot)$ are called the mother and the daughter density functions, respectively. Chaubey et al. (2010) provide a one-to-one correspondence between f and g.

Apart from simplifying proofs involving global-local mixtures, Cauchy-Schlömilch and Liouville transformations can generate new distributions through suitable scale transformations, $f(x) = 2g\{t(x)\}$, of simple baseline functions, $f(x)$, under suitable conditions. Given a density $f(x)$ we can create a new global-local scale family $f(ax - bx^{-1})$, by effectively reallocating its probability mass. A particularly useful tool for generating univariate and multivariate random variables is Khintchine's theorem for unimodal, univariate distributions. Khintchine's theorem states that any random variable X with a mode at zero can be written as a product $X = ZU$, where $U \sim U(0, 1)$ and Z has the density function $f_Z(z) = -zf_X{}'(z)$. Bryson & Johnson (1982), and successively Jones (2012), discuss how Khintchine's theorem allows us to construct both univariate and multivariate distributions, even with special dependence structure.

Jones (2014) develops an extended Khintchine's theorem that further allows one to generate random variables with unimodal densities of the form $2g(t(x))$.

The rest of the paper is organized as follows: §2 derives scale mixture results for the Lasso, quantile and logistic regression, §3 for convolutions of densities via mixtures and finally §4 concludes with two open problems.

2. GLOBAL-LOCAL SCALE MIXTURES

2·1. *Lasso as a Normal Scale Mixture*

The lasso penalty arises as a Laplace global-local mixture (Andrews & Mallows, 1974). A simple transformation proof follows using Cauchy-Schlömilch with $f(x) = e^{-x}$. Starting with the normal integral identity, $\int_0^\infty f(y^2)dy = \int_0^\infty e^{-y^2}dy = \pi^{1/2}/2$, we obtain

$$\int_0^\infty e^{-(ax)^2-(b/x)^2}dx = \int_0^\infty \exp\left\{-ab\left(\frac{a}{b}x^2 + \frac{b}{a}x^{-2}\right)\right\}dx = \frac{\pi^{1/2}}{2a}e^{-2ab} \ .$$

Substituting $t = (a/b)^{1/2}x$ and $c = ab$ yields the Laplace or Lasso penalty:

$$\int_0^\infty e^{-c(t-t^{-1})^2}dt = \frac{1}{2}(\pi/c)^{1/2} \Rightarrow \int_0^\infty e^{-c(t^2+t^{-2})}dt = \frac{1}{2}(\pi/c)^{1/2}e^{-2c} \ .$$

The Laplace density can be viewed as a transformed normal, via $y = t - t^{-1}$.

Remark 1. The usual identity for lasso can also be obtained from Lévy (1940):

$$\int_0^\infty \frac{a}{\sqrt{2\pi}t^{3/2}}e^{-a^2/(2t)}e^{-\lambda t}dt = e^{-a\sqrt{2\lambda}} \tag{4}$$

For $a = 1$, and $\theta = (2\lambda)^{1/2}$, this can be written as

$$E\left(e^{-\theta^2/(2G)}\right) = e^{-\theta} \ , \quad \text{where} \quad G \sim \text{Ga}(1/2, 1/2) \tag{5}$$

Proof. First substitute $t^{-1} = x^2$, which makes the left hand side in (4) equal to

$$\begin{aligned}\int_0^\infty \frac{a}{\sqrt{(2\pi)}t^{3/2}}e^{-a^2/(2t)}e^{-\lambda t}dt &= \sqrt{\frac{2}{\pi}}ae^{-a\sqrt{(2\lambda)}}\int_0^\infty e^{-\left(\frac{a}{\sqrt{2}}x-\lambda x^{-1}\right)^2}dx \\ &= \exp(-a\sqrt{(2\lambda)}) \ .\end{aligned}$$

The last step follows directly from Cauchy-Schlömilch formula. The second relationship (5) follows by fixing $a = 1$, $\theta = (2\lambda)^{1/2}$ and substituting $t = x^{-1}$:

$$\int_0^\infty \frac{a}{\sqrt{(2\pi)}t^{3/2}}e^{-a^2/(2t)}e^{-\lambda t}dt = \int_0^\infty e^{-\frac{\theta^2}{2x}}\sqrt{\left(\frac{1}{2\pi}\right)}x^{-\frac{1}{2}}e^{-\frac{1}{2}x}dx.$$

The left hand side can be identified as $E\left(e^{-\theta^2/(2G)}\right)$ for $G \sim \text{Ga}(1/2, 1/2)$. $\square$

2·2. *Logit and Quantile as Global-local Mixtures*

Logistic modeling can be viewed within the global-local mixture framework via the Pólya-Gamma distribution (Polson et al., 2013). This leads to efficient Markov chain Monte Carlo algorithms for inference. The two key marginal distributions for the hyperbolic generalized inverse

Gaussian (Barndorff-Nielsen et al., 1982) and Pólya-Gamma mixtures are

$$\frac{\alpha^2-\kappa^2}{2\alpha}e^{-\alpha|x-\mu|+\kappa(x-\mu)} = \int_0^\infty \phi(x \mid \mu+\kappa\lambda, \lambda) p_{\mathrm{GIG}}(\lambda \mid 1, 0, \sqrt{(\alpha^2-\kappa^2)}) d\lambda \ , \tag{6}$$

$$\frac{1}{B(\alpha,\kappa)} \frac{e^{\alpha(x-\mu)}}{(1+e^{x-\mu})^{\alpha+\kappa}} = \int_0^\infty \phi(x \mid \mu+\kappa\lambda, \lambda) p_{\mathrm{Polya}}(\lambda \mid \alpha, \kappa) d\lambda \ , \tag{7}$$

where $\phi(\mu+\kappa\lambda, \lambda)$ denotes the normal density function. The functions p_{GIG} and p_{Polya} are the corresponding local mixture densities for the generalized inverse Gaussian and the Pólya-Gamma, respectively. Rather surprisingly, the logit and quantile identities can be derived using Cauchy-Schlömilch transformations. Let $f(x) = e^{-x^2/2}$, $a = \alpha$ and $b = |x-\phi|$ in (1) and we have

$$\sqrt{\frac{2}{\pi}} \int_0^\infty e^{-\frac{1}{2}\left(\alpha y - \frac{|x-\mu|}{y}\right)^2} dy = \frac{1}{\alpha}\frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-\frac{1}{2}y^2} dy = \frac{1}{\alpha} \ .$$

Let $\nu = y^2$ and rearrange the constant terms to get the relation

$$\frac{1}{\alpha}e^{-\alpha|x-\mu|} = \frac{1}{\sqrt{2\pi\nu}} \int_0^\infty \exp\left\{-\left(\frac{(x-\mu)^2}{2\nu} + \frac{\alpha^2}{2}\nu\right)\right\} d\nu \ .$$

Multiplying by $2^{-1}(\alpha^2-\kappa^2)\exp\{\kappa(x-\mu)\}$ and completing the square yields

$$\frac{\alpha^2-\kappa^2}{2\alpha}e^{-\alpha|x-\mu|+\kappa(x-\mu)} = \int_0^\infty \phi(x \mid \mu+\kappa\nu, \nu)\frac{\alpha^2-\kappa^2}{2}e^{-\frac{\alpha^2-\kappa^2}{2}\nu} d\nu.$$

The mixing distribution is exponential with rate parameter $(\alpha^2-\kappa^2)/2$, a special case of the generalized inverse Gaussian distribution introduced by Etienne Halphen circa 1941 (Seshadri, 1997). The density with parameters $(\lambda, \delta, \gamma)$ has the form

$$p_{\mathrm{GIG}}(x \mid \lambda, \delta, \gamma) = \frac{(\gamma/\delta)^\lambda}{2K_\lambda(\delta\gamma)} x^{\lambda-1} \exp\left\{-\frac{1}{2}(\delta^2 x^{-1} + \gamma^2 x)\right\} \ , \quad \text{for} \quad x>0, \lambda>0, \delta>0, p \in \Re \ ,$$

where K_λ is the modified Bessel function of the second kind. The Liouville formula can be used to show that the above is a valid probability density function. When δ or γ is zero, the normalizing constant takes the limit values given by $K_\lambda(u) \sim \Gamma(|\lambda|)2^{|\lambda|-1}u^{|\lambda|}$ for $\lambda > 0$. If $\delta = 0$, the generalized inverse Gaussian is identical to a gamma distribution:

$$p_{\mathrm{GIG}}(x \mid \lambda, \delta=0, \gamma) = \frac{(\alpha)^\lambda}{\Gamma(\lambda)} x^{\lambda-1}\exp\{-\alpha x\} \ , \quad \text{for} \quad x>0, \alpha = \gamma^2/2.$$

We now present a simple proof for the Pólya-Gamma mixture in (7). First, write κ for $a - b/2$:

$$\frac{(e^\psi)^a}{(1+e^\psi)^b} = 2^{-b}e^{\kappa\omega} \int_0^\infty e^{-\omega\psi^2/2} p(\omega) d\omega \ , \tag{8}$$

where $\omega \sim \mathrm{PG}(b, 0)$, the Pólya-Gamma distribution with density is

$$p(\omega \mid b, 0) = \frac{2^{(b-1)}}{\Gamma(b)} \sum_{n=0}^\infty (-1)^n \frac{\Gamma(n+b)}{\Gamma(n+1)} \frac{2n+b}{\sqrt{(2\pi)}\omega^{3/2}} \exp\left(-\frac{(2n+b)^2}{8\omega}\right) .$$

The logit function corresponds to $a = 0, b = 1$ in (8). Cauchy-Schlömilch yields:

$$\frac{1}{1+e^{\psi}} = \frac{1}{2}e^{-\frac{1}{2}\psi}\int_0^{\infty} e^{-\frac{\psi^2}{2}\omega}p(\omega)d\omega, \quad \text{where } p(\omega) = \sum_{n=0}^{\infty}(-1)^n \frac{2n+1}{\sqrt{(2\pi)}\omega^{3/2}}e^{-\frac{(2n+1)^2}{8\omega}} \ . \tag{9}$$

To show (9), write the left-hand side interchanging integral and summation:

$$I = \frac{1}{2}e^{-\psi/2}\sum_{n=0}^{\infty}(-1)^n\frac{2n+1}{\sqrt{(2\pi)}}\int_0^{\infty} e^{-\left(\frac{\psi^2}{2}\omega+\frac{(2n+1)^2}{8\omega}\right)}\frac{1}{\omega^{3/2}}d\omega.$$

Using the change of variable $\omega = t^{-2}$ gives

$$I = \sum_{n=0}^{\infty}(-1)^n e^{-(n+1)\psi}\frac{(2n+1)}{\sqrt{(2\pi)}}\left\{\int_0^{\infty} e^{-\frac{1}{2}\left(\frac{(2n+1)t}{2}-\frac{\psi}{t}\right)^2}dt\right\} \ .$$

Applying Cauchy-Schlömilch to the inner integral yields

$$\int_0^{\infty} e^{-\frac{1}{2}\left(\frac{(2n+1)t}{2}-\frac{\psi}{t}\right)^2}dt = \int_0^{\infty} e^{-\frac{1}{2}y^2}dy\frac{1}{2n+1} = \frac{\sqrt{(2\pi)}}{2n+1},$$

which implies $I = \sum_{n=0}^{\infty}(-1)^n \exp\{-(n+1)\psi\} = \{1+\exp(\psi)\}^{-1}$.

Remark 2. When $\alpha = \kappa$, we have the limiting result

$$(\alpha^2-\kappa^2)^{-1}p_{\text{GIG}}\{1, 0, \sqrt{\alpha^2-\kappa^2}\} \equiv 1 \ ,$$

or equivalently in terms of densities, with a marginal improper uniform prior, $p(\lambda) = 1$,

$$\int_0^{\infty}\phi(b \mid -a\lambda, c\lambda)d\lambda = a^{-1}\exp\{-2\max(ab/c, 0)\} \ . \tag{10}$$

This pseudo-likelihood represents support vector machines as a global-local mixture.

Polson et al. (2011) derive this as a direct consequence of the Lasso identity

$$\int_0^{\infty}\frac{p}{\sqrt{2\pi\lambda}}\exp\left\{-\frac{1}{2}\left(p^2\lambda+q^2\lambda^{-1}\right)\right\}d\lambda = \exp(-|pq|) \ .$$

We apply Liouville formula and obtain

$$\int_0^{\infty} f\left(ax+\frac{b}{x}\right)\frac{dx}{\sqrt{x}} = \frac{1}{\sqrt{a}}\int_0^{\infty} f\left(2\sqrt{ab}+y\right)\frac{dy}{\sqrt{y}} \quad \text{for} \quad a, b > 0 \ .$$

Setting $f(x) = e^{-x}$, $a = p^2/2$, and $b = q^2/2$ we get

$$\begin{aligned}\int_0^{\infty}\lambda^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(p^2\lambda+q^2\lambda^{-1})\right\}d\lambda &= \frac{2^{\frac{1}{2}}}{p}\int_0^{\infty} e^{-|pq|+y}y^{-\frac{1}{2}}dy \\ &= \frac{2^{\frac{1}{2}}e^{-|pq|}}{p}\int_0^{\infty} e^{-y}y^{-\frac{1}{2}}dy = \frac{(2\pi)^{\frac{1}{2}}e^{-|pq|}}{p} \ .\end{aligned}$$

Hans (2011) shows that an elastic-net regression can be recast as a global-local mixture with a mixing density belonging to the orthant-normal family of distributions. The orthant-normal prior on a single regression coefficient, β, given hyper-parameters λ_1 and λ_2, has a density function with the following form:

$$p(\beta \mid \lambda_1, \lambda_2) = \begin{cases} \phi(\beta \mid \frac{\lambda_1}{2\lambda_2}, \frac{\sigma^2}{\lambda_2})/2\Phi(\frac{-\lambda_1}{2\sigma\sqrt{\lambda_2}}) & \text{if } \beta < 0 \\ \phi(\beta \mid \frac{-\lambda_1}{2\lambda_2}, \frac{\sigma^2}{\lambda_2})/2\Phi(\frac{-\lambda_1}{2\sigma\sqrt{\lambda_2}}) & \text{if } \beta \geq 0 \end{cases}. \tag{11}$$

3. CONVOLUTION MIXTURES

Another interesting area of application is convolution mixtures and marginal densities for location-scale mixture problems. We show that the Cauchy convolution (Pillai & Meng, 2015) and Inverse-Gamma convolution can be derived similarly (Gelman, 2006; Polson & Scott, 2012). In a recent article, Bhadra et al. (2016) shows that the regularly varying tails of half-Cauchy priors are also key to learning many-to-one functions of normal vector mean, where the flat prior gives poorly calibrated inference.

LEMMA 1. *Let* $X_i \sim \text{Cauchy}(0,1)$, *then for* $i = 1, 2$ *we have* $Z = w_1 X_1 + w_2 X_2 \sim C(0, w_1 + w_2)$.

LEMMA 2. *Let* $X_i \sim \text{IG}(\alpha t_i, \alpha t_i^2)$, *then for* $i = 1, 2$ *we have* $Z \doteq X_1 + X_2 \sim \text{IG}(\alpha(t_1 + t_2), \alpha(t_1^2 + t_2^2)$, *where* $\text{IG}(\alpha t, \alpha t^2)$ *denotes the inverse-Gaussian density given by*

$$f(t) = \frac{t\sqrt{\alpha}e^t}{\sqrt{(2\pi)}x^{3/2}} \exp\left(-\frac{\alpha t^2}{2x} - \frac{x}{2\alpha}\right) 1_{(0,\infty)}(x)$$

Both of these results will follow from a straight-forward application of the Cauchy-Schlömilch transformation. We give the proof for the Cauchy convolution identity below.

Proof. Exploiting the symmetry and the Lagrange identity $(a^2+b^2)(c^2+d^2) = (ac + bd)^2 + (ad - bc)^2$, leads us to the convolution density

$$\begin{aligned} f_Z(z) &= 2\int_0^\infty \frac{1}{\pi w_1\{(1 + x^2/w_1^2)\}} \frac{1}{\pi w_2\{1 + (z-x)^2/w_2^2\}} dx, \\ &= \frac{2}{\pi^2 w_1 w_2} \int_0^\infty \frac{1}{\{1 + w_1^{-1}w_2^{-1}x(z-x)\}^2 + \{w_2^{-1}z - (w_1^{-1} + w_2^{-1})x\}^2} dx. \end{aligned}$$

Transforming $x \mapsto x + w_2^{-1}z(w_1^{-1} + w_2^{-1})^{-1}$, using $a = 1 + z^2(w_1+w_2)^{-2}$, $b = (w_1 w_2)^{-1}$, $c = z(w_2 - w_1)\{(w_1+w_2)w_1w_2\}^{-1}$ and $d = z(w_2 - w_1)\{(w_1+w_2)w_1w_2\}^{-1}$, we obtain

$$\begin{aligned} f_Z(z) &= \frac{2}{\pi^2 w_1 w_2} \int_0^\infty \left[\left\{1 + \frac{z^2}{(w_1+w_2)^2} - \frac{x^2}{w_1 w_2} + xz\frac{w_2 - w_1}{(w_1+w_2)w_1w_2}\right\}^2 + x^2\left(\frac{w_1+w_2}{w_1w_2}\right)^2\right]^{-1} dx \\ &= \frac{2}{\pi^2 w_1 w_2} \int_0^\infty \frac{dx}{(a - bx^2 + cx)^2 + x^2 d^2} = \frac{2}{\pi^2 w_1 w_2} \int_0^\infty \frac{dx/x^2}{(a/x - bx + c)^2 + d^2}. \end{aligned}$$

Transforming $y = x^{-1}$, and applying Cauchy-Schlömilch transformation we arrive at

$$f_Z(z) = \frac{2}{\pi w_1 w_2} \int_0^\infty \frac{dy}{2a(y^2 + d^2)} = \frac{1}{\pi w_1 w_2} \frac{1}{ad} = \frac{1}{\pi(w_1+w_2)} \frac{1}{1 + z^2/(w_1+w_2)^2}.$$

A simple induction argument proves that the sum of any number of independent Cauchy random variates is also another Cauchy. □

One can also use the characteristic function of $X \sim \text{C}(\mu, \sigma)$, $\psi_X(t) = \exp(it\mu - |t|\sigma^2)$, and the relation $\psi_{X+Y}(t) = \psi_X(t)\psi_Y(t)$ to derive the result in just one step. For $X = \sum_{i=1}^p \omega_i C_i$

and $C_i \sim \mathrm{C}(0,1)$, when $\sum_{i=1}^{p} \omega_i = 1$ we have

$$\phi_X(t) = \exp\left(-\sum_{i=1}^{p} \omega_i |t|\right) = \exp(-|t|) = \phi_C(t) \ ,$$

where $C \sim \mathrm{C}(0,1)$.

The most general result in this category is due to Pillai & Meng (2015), who they showed the following: Let $(X_1, \cdots, X_m)$ and $(Y_1, \cdots, Y_m)$ be independent and identically distributed. $\mathcal{N}(0,\Sigma)$ for an arbitrary positive definite matrix Σ, then $Z = \sum_{j=1}^{m} w_j X_j (Y_j)^{-1} \sim \mathrm{C}(0,1)$, as long as $(w_1, \cdots, w_m)$ is independent of (X, Y), $w_j \geq 0$ $(j = 1, \cdots, m)$ and $\sum_{j=1}^{m} w_j = 1$.

4. Discussion

The Cauchy-Schlömilch and Liouville transformations not only guarantee an 'astonishingly simple' normalizing constant for $f(\cdot)$, it also establishes the wide class of unimodal densities as global-local scale mixtures. Global-local scale mixtures that are conditionally Gaussian hold a special place in statistical modeling and can be rapidly fit using an expectation-maximization algorithm as pointed out by Polson & Scott (2013). Palmer et al. (2011) provides a similar tool for modeling multivariate dependence by writing general non-Gaussian multivariate densities as multivariate Gaussian scale mixtures. Our future goal is to extend the Cauchy-Schlömilch transformation to express the wide multivariate Gaussian scale mixture models as global-local mixtures that also facilitate easy computation.

We end our paper with conjectures that two other remarkable identities arise as corollaries of such transformation identities. The first one is a recent result by Zhang et al. (2014) that proves the uniform correlation mixture of the bivariate Gaussian density with unit variance is a function of the maximum norm:

$$\int_{-1}^{1} \frac{1}{2} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{x_1^2 + x_2^2 - 2\rho x_1 x_2}{2(1-\rho^2)}\right\} d\rho = \frac{1}{2}(1 - \Phi(||x||_\infty)), \tag{12}$$

where $\Phi(\cdot)$ is the standard normal distribution function and $||x||_\infty = \max\{x_1, x_2\}$. The bivariate density on the right side of (12) was introduced before in Bryson & Johnson (1982) as uniform mixtures of a χ_3 random variate, but the representation as a uniform correlation mixture is a surprising new find. We make a couple of remarks connected to the Erdelyi's integral identity, which is key to the proof of the uniform correlation mixture (12):

$$\int_{\frac{1}{2}}^{\infty} \frac{\mathrm{e}^{-x^2 z}}{4\pi z\sqrt{2z-1}} \mathrm{d}z = \frac{1}{2}(1 - \Phi(x)) \ , \quad \text{where} \quad x \geq 0 \tag{13}$$

Erdelyi's identity in (13) follows from the Laplace transformation $(1+u)^{-1} = \int_0^\infty \mathrm{e}^{-v(1+u)} dv$.

Proof. Let $u = 2z - 1$ on the left hand side of (13), denoted by I, to obtain:

$$I = \int_0^\infty \frac{\mathrm{e}^{-x^2/(2(1+u))}}{4\pi\sqrt{u}(1+u)} \mathrm{d}u.$$

Using the Laplace transformation $(1+u)^{-1} = \int_0^\infty \mathrm{e}^{-v(1+u)} dv$,

$$
\begin{aligned}
I &= \int_0^\infty \frac{\mathrm{e}^{-x^2/(2(1+u))}}{4\pi\sqrt{u}} \int_0^\infty \mathrm{e}^{-v(1+u)} \mathrm{d}v\mathrm{d}u = \int_{v=0}^\infty \int_{u=0}^\infty \frac{\mathrm{e}^{-(\frac{x^2}{2}+v)(1+u)}}{4\pi\sqrt{u}} \mathrm{d}v\mathrm{d}u \\
&= \int_{v=0}^\infty \frac{1}{4\pi} \mathrm{e}^{-(\frac{x^2}{2}+v)} \int_{u=0}^\infty u^{-1/2} \mathrm{e}^{-(\frac{x^2}{2}+v)u} \mathrm{d}u\mathrm{d}v = \int_{v=0}^\infty \frac{\mathrm{e}^{-\frac{1}{2}(x^2+2v)}}{2(2\pi)^{1/2}} \frac{1}{(x^2+2v)^{1/2}} \mathrm{d}v
\end{aligned}
$$

Transforming $z^2 = x^2 + 2v$, we get

$$
I = \frac{1}{2} \int_{z=|x|}^\infty \frac{1}{(2\pi)^{1/2}} e^{-\frac{1}{2}z^2} \mathrm{d}z = \frac{1}{2}(1 - \Phi(|x|)) \ .
$$

Secondly, Erdelyi's identity in (13) follows from (9.2) of Amdeberhan et al. (2010):

$$
\int_0^\infty \frac{\mathrm{e}^{-\mu^2(x^2+\beta^2)}}{x^2+\beta^2} \mathrm{d}x = \frac{\pi}{2\beta}\{1 - \mathrm{erf}(\mu\beta)\}.
$$

If we let $\beta = 2^{-1/2}$ and $x^2 + 1/2 = z$, the above identity reduces to (13) in μ.

The second candidate is the symmetric stable distribution, defined by its characteristic function $\phi(t) = \exp\{-|t|^\alpha\}$ for $\alpha \in (0,2]$. It admits a normal scale mixture representation with mixing density

$$
f(v) = \frac{1}{2} s_{\alpha/2}\left(\frac{v}{2}\right), v > 0
$$

$$
\exp(-|x|^\alpha) = \int_0^\infty \exp(-xs) g(s) ds, \ g(s) = \sum_{j=1}^\infty (-1)^j \frac{s^{-j\alpha-1}}{j!\Gamma(-\alpha j)} \ ,
$$

when $s_{\alpha/2}$ is the density of the positive stable distribution with index $\alpha/2$. An important application of this representation is found in Bayesian bridge regression (Polson et al., 2014). Regularization, in this case, is an outcome of a normal scale mixture with respect to an α-stable random variable. We conjecture that these two results follow as upshots of the Cauchy-Schlömilch formula (1). Other fruitful areas could be unearthed by applications of the Liouville formula to recognize and generate global-local mixtures, and other applications such as calculating higher-order 'closed-form' moments $E(X^n)$ for random variables X that admit a global-local representation.

REFERENCES

AMDEBERHAN, T., GLASSER, M. L., JONES, M. C., MOLL, V. H., POSEY, R. & VARELA, D. (2010). The Cauchy-Schlomilch transformation. *arXiv:1004.2445 [math]* .

ANDREWS, D. & MALLOWS, C. (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society Series B Methodological* **36**, 99–102.

BAKER, R. (2008). Probabilistic applications of the Schlömilch transformation. *Communications in Statistics—Theory and Methods* **37**, 2162–2176.

BARNDORFF-NIELSEN, O., KENT, J. & SØRENSEN, M. (1982). Normal variance-mean mixtures and z distributions. *International Statistical Review/Revue Internationale de Statistique* **50**, 145–159.

BHADRA, A., DATTA, J., POLSON, N. G. & WILLARD, B. (2015). The Horseshoe+ Estimator of Ultra-Sparse Signals. *arXiv preprint arXiv:1502.00560* .

BHADRA, A., DATTA, J., POLSON, N. G. & WILLARD, B. T. (2016). Default Bayesian Analysis with Global-Local Shrinkage Priors. *Biometrika* .

Boros, G., Moll, V. H. & Foncannon, J. (2006). Irresistible integrals: symbolics, analysis and experiments in the evaluation of integrals. *The Mathematical Intelligencer* **28**, 65–68.

Bryson, M. C. & Johnson, M. E. (1982). Constructing and simulating multivariate distributions using Khintchine's theorem. *Journal of Statistical Computation and Simulation* **16**, 129–137.

Chaubey, Y. P., Mudholkar, G. S. & Jones, M. (2010). Reciprocal symmetry, unimodality and Khintchine's theorem. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 466. The Royal Society.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian analysis* **1**, 515–534.

Hans, C. (2011). Comment on Article by Polson and Scott. *Bayesian Analysis* **6**, 37–41.

Jones, M. (2012). On Khintchine's theorem and its place in random variate generation. *The American Statistician* .

Jones, M. C. (2014). Generating distributions by transformation of scale. *Statist. Sinica* **24**, 749–772.

Lévy, P. (1940). Sur certains processus stochastiques homogènes. *Compositio mathematica* **7**, 283–339.

Palmer, J. A., Kreutz-Delgado, K. & Makeig, S. (2011). AMICA: An adaptive mixture of independent component analyzers with shared components. Tech. rep., San Diego, CA: Technical report, Swartz Center for Computational Neuroscience.

Pillai, N. S. & Meng, X.-L. (2015). An unexpected encounter with Cauchy and L\'evy. *arXiv preprint arXiv:1505.01957* .

Polson, N. G. & Scott, J. G. (2012). On the Half-Cauchy Prior for a Global Scale Parameter. *Bayesian Analysis* **7**, 887–902.

Polson, N. G. & Scott, J. G. (2013). Data augmentation for non-Gaussian regression models using variance-mean mixtures. *Biometrika* **100**, 459–471.

Polson, N. G., Scott, J. G. & Windle, J. (2013). Bayesian inference for logistic models using Pólya–Gamma latent variables. *Journal of the American Statistical Association* **108**, 1339–1349.

Polson, N. G., Scott, J. G. & Windle, J. (2014). The Bayesian bridge. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* **76**, 713–733.

Polson, N. G., Scott, S. L. & others (2011). Data augmentation for support vector machines. *Bayesian Analysis* **6**, 1–23.

Seshadri, V. (1997). Halphen's laws. *Encyclopedia of statistical sciences* .

Zhang, K., Brown, L. D., George, E. & Zhao, L. (2014). Uniform Correlation Mixture of Bivariate Normal Distributions and Hypercubically Contoured Densities That Are Marginally Normal. *The American Statistician* **68**, 183–187.

[*Received April* 2016. *Revised September* 2016]