Electron. Commun. Probab. **0** (2016), no. 0, 1–9.
DOI: 10.1214/YY-TN
ISSN: 1083-589X

ELECTRONIC COMMUNICATIONS in PROBABILITY

A Short Note on Global-Local Mixtures

Anindya Bhadra[*] Jyotishka Datta[†] Nicholas G. Polson[‡]
Brandon Willard[§]

Abstract

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.

Keywords: Bayes regularization; Cauchy; Convolution; Global-local mixture; Lasso; Logistic; Quantile; Stable law..
AMS MSC 2010: NA.
Submitted to ECP on January 2, 2013, final version accepted on December 13, 2014.
Supersedes arXiv:1604.07487.

Abstract

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.

Keywords: Bayes regularization; Cauchy; Convolution; Global-local mixture; Lasso; Logistic; Quantile; Stable law.

1 Introduction

Many statistical problems involve regularization penalties derived from global-local mixture distributions [20, 10, 5]. A global-local mixture density, denoted by $p(x_1, \dots, x_p)$, takes the form

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

[*]Department of Statistics, Purdue University. E-mail: bhadra@stat.purdue.edu
[†]Department of Mathematical Sciences, University of Arkansas.
E-mail: jd033@uark.edu
[‡]The University of Chicago Booth School of Business.
E-mail: ngp@chicagobooth.edu
[§]The University of Chicago Booth School of Business.
E-mail: brandonwillard@gmail.com

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 mixtures 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:

$$\int_0^\infty f\left\{(ax - bx^{-1})^2\right\} dx = \frac{1}{2a} \int_0^\infty f(y^2) dy, \quad a, b > 0 \; , \tag{1.1}$$

and the Liouville transformation:

$$\int_0^\infty f\left(ax + \frac{b}{x}\right) x^{-1/2} dx = a^{-1/2} \int_0^\infty f\left\{2(ab)^{1/2} + y\right\} y^{-1/2} dy, \quad a, b > 0 \; . \tag{1.2}$$

See [6], [2] and [11] for further discussion. Identity (1.1) follows from the simple transformation $t = b/(ax)$ as

$$I = \int_0^\infty f\left\{(ax - b/x)^2\right\} dx = \int_0^\infty f\left\{(at - b/t)^2\right\} \frac{b}{at^2} dt, \quad a, b > 0.$$

Adding the two terms in the last equality yields

$$2I = \int_0^\infty f\left\{(at - b/t)^2\right\} \left\{1 + b/(at^2)\right\} dt,$$

and transforming $y = b/t - at$ gives $dy = -a\{1 + b/(at^2)\} dt$, yielding $I = (2a)^{-1} \int_0^\infty f(y^2) dy$, as required. A useful generalization of the Cauchy-Schlömilch transformation is as follows:

$$\int_0^\infty f\left[\{x - s(x)\}^2\right] dx = \int_0^\infty f(y^2) dy, \tag{1.3}$$

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 proof for the Liouville transformation identity follows in a similar manner, and is omitted for the sake of brevity.

We can use these results to generate new probability distributions with different choices of simple baseline function $g(\cdot)$ and derive new scale mixture representations that are useful in Bayesian global-local modeling. The Cauchy-Schlömilch and Liouville transformations can generate new distributions via scale transformations that can take the form $f(x) = 2g\{t(x)\}$ for certain $f(x)$ under suitable conditions. The simplest example is creating a new global-local scale family, $f(ax - b/x)$ by effectively reallocating the probability mass of a given density $f(x)$.

More generally, let $f(x) = 2g\{t(x)\}$ and let $t(x)$ be of the form $x - s(x)$, where $s : \Re^+ \to \Re^+$ is a self-inverse, onto and monotone decreasing function. [11] 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, as shown below:

$$\Pi_T(y) = (1/2)\{y + (4b + y^2)^{1/2}\}, \quad \Pi_T^{-1}(x) = t_T(x) = x - b/x, \quad b > 0,$$
$$\Pi_L(y) = a^{-1}\log(1 + e^{ay}), \quad \Pi_L^{-1}(x) = t_L(x) = a^{-1}\log(e^{ax} - 1), \quad a > 0.$$

Now, the integral identity in (1.3) shows that if $f(x)$, $x \geq 0$ is a density function, so is $g(x) = f\{|x - s(x)|\}$, $x \geq 0$. The functions $f(\cdot)$ and $g(\cdot)$ are called mother and daughter density functions, respectively.

The mother and daughter density functions, $f(\cdot)$ and $g(\cdot)$ are linked via a dual relationship with respect to symmetry and reciprocal symmetry for densities supported on the whole real line or its positive half $\Re^+$, respectively. The density function $f(\cdot)$ on $\Re^+$ is defined to have reciprocal symmetry (or, R-symmetry) if $f(\theta y) = f(\theta/y)$ for all $y > 0$ and some $\theta > 0$. It turns out that if $f(x)$ is the pdf of a symmetric real-valued random variable X, the daughter pdf $g(x) = f(x - 1/x), x > 0$ is an R-symmetric density, and vice-versa, there exists a symmetric density $f(x) = g(x + \sqrt{1 + x^2})$ for every R-symmetric density $g(x)$. Furthermore, $f(\cdot)$ is unimodal if and only if $g(\cdot)$ is unimodal. [8] provide a few examples of generating R-symmetric densities g starting from well-known symmetric densities f. The most well-known example of this duality is perhaps the normal density as f that gives rise to the root reciprocal inverse Gaussian, abbreviated as RRIG, distribution, with density given by:

$$g(x) = \sqrt{\frac{2\lambda}{\pi}} \exp\left\{ -\frac{\lambda}{2} \left(x - \frac{1}{x} \right)^2 \right\}, x > 0.$$

Once again, the Cauchy-Schlömilch transformation $y = x - x^{-1}$ guarantees that this is a valid probability density function.

A particularly useful tool for generating univariate and multivariate random variables is Khintchine's theorem, which states that any random variable X with a unimodal, univariate distribution and a mode at zero can be written as a product $X = ZU$, where $U \sim \mathcal{U}(0, 1)$ and Z has the density function $f_Z(z) = -z f_X'(z), z \in \Re$. [7], and subsequently [12], discuss how Khintchine's theorem allows us to construct both univariate and multivariate densities, even with special dependence structure. [11] develops an extended Khintchine's theorem that further allows us 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 [1]. 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:

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

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

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

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

Proposition 2.1. *The usual identity for the Lasso also follows from [13] as*

$$\int_0^\infty \frac{a}{(2\pi)^{1/2} t^{3/2}} e^{-a^2/(2t)} e^{-\lambda t} dt = e^{-a(2\lambda)^{1/2}} . \tag{2.1}$$

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

$$E\left[\exp\{-\theta^2/(2G)\}\right] = \exp(-\theta), \quad \textit{where } G \sim \mathcal{G}(1/2, 1/2). \tag{2.2}$$

Proof. First substitute $t^{-1} = x^2$, which makes the left hand side in (2.1) equal to

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

The last step follows from Cauchy-Schlömilch formula. The second relationship (2.2) follows by fixing $a = 1$, $\theta = (2\lambda)^{1/2}$ and substituting $t = x^{-1}$.

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

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

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 [18]. As [18] show, this mixture representation leads to efficient Markov chain Monte Carlo algorithms for inference.

Proposition 2.2. *The two key marginal distributions for the hyperbolic generalized inverse Gaussian [3] and Pólya-Gamma mixtures are*

$$\frac{\alpha^2 - \kappa^2}{2\alpha} e^{-\alpha|x-\mu| + \kappa(x-\mu)} = \int_0^\infty \phi(x \mid \mu + \kappa\lambda, \lambda) p_{\mathrm{GIG}} \left\{\lambda \mid 1, 0, (\alpha^2 - \kappa^2)^{1/2}\right\} d\lambda, \ \alpha \geq \kappa \geq 0, \tag{2.3}$$

$$\frac{1}{B(\alpha, \kappa)} \frac{e^{\alpha(x-\mu)}}{(1 + e^{x-\mu})^{\alpha+\kappa}} = \int_0^\infty \phi(x \mid \mu + \kappa\lambda, \lambda) p_{\mathrm{Polya}}(\lambda \mid \alpha, \kappa) d\lambda \ , \tag{2.4}$$

where $\phi(\mu + \kappa\lambda, \lambda)$ denotes the normal density function with mean $(\mu + \kappa\lambda)$ and variance λ. The functions p_{GIG} and p_{Polya} are the corresponding local mixture densities for the generalized inverse Gaussian and the Pólya-Gamma, respectively. The logit and quantile identities can be derived using Cauchy-Schlömilch identity.

Proof. Let $f(x) = e^{-x^2/2}$, $a = \alpha$ and $b = |x - \phi|$ in (1.1). Then,

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

Let $\nu = y^2$. Rearranging the constant terms yields

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

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

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

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 [21]. The density with parameters $(\lambda, \delta, \gamma)$ has the form

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

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 limiting values given by $K_\lambda(u) \asymp \Gamma(|\lambda|)2^{|\lambda|-1}u^{|\lambda|}$ for $\lambda > 0$. If $\delta = 0$, the generalized inverse Gaussian is identical to a gamma distribution:

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

We now present a simple proof for the Pólya-Gamma mixture in (2.4). First, write κ for $a - b/2$:

$$\frac{(e^\psi)^a}{(1+e^\psi)^b} = 2^{-b} e^{\kappa\omega} \int_0^\infty e^{-\omega\psi^2/2} p(\omega) d\omega \ , \tag{2.5}$$

where $\omega \sim \mathrm{PG}(b, 0)$, a Pólya-Gamma random variable with density

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

The logit function corresponds to $a = 0, b = 1$ in (2.5). The Cauchy-Schlömilch identity yields

$$\frac{1}{1+e^\psi} = \frac{1}{2} e^{-\psi/2} \int_0^\infty e^{-(\psi^2\omega)/2} p(\omega) d\omega, \text{ where } p(\omega) = \sum_{n=0}^\infty (-1)^n \frac{2n+1}{(2\pi\omega^3)^{1/2}} e^{-(2n+1)^2/(8\omega)} \ . \tag{2.6}$$

To show (2.6), write the right-hand side interchanging the integral and summation:

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

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

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

Applying the Cauchy-Schlömilch identity to the inner integral yields

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

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

Remark 2.3. When $\alpha = \kappa$, we have the limiting result $(\alpha^2 - \kappa^2)^{-1} p_{\mathrm{GIG}}\{1, 0, (\alpha^2 - \kappa^2)^{1/2}\} = 1$, or equivalently in terms of densities, with a marginal improper uniform prior, $p(\lambda) = 1$,

$$\int_0^\infty \phi(b \mid -a\lambda, c\lambda) d\lambda = a^{-1} \exp\left\{-2\max(ab/c, 0)\right\} \ . \tag{2.7}$$

This pseudo-likelihood represents support vector machines as a global-local mixture. The identity for quantile regression, which is a limiting case of the above identities by applying Fatou-Lebesgue theorem, is the following:

$$c^{-1} \exp\{2c^{-1}\rho_q(b)\} = \int_0^\infty \phi(b \mid \lambda - 2\tau\lambda, c\lambda) \mathrm{e}^{-2\tau(1-\tau)\lambda} d\lambda, \quad c, \tau > 0,$$

where $\rho_q(b) = |b|/2 + (q - 1/2)b$ is the check-loss function [17].

[20] derive this as a direct consequence of the Lasso identity

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

Applying the Liouville identity yields

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

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

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

[10] shows that the 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:

$$p(\beta \mid \lambda_1, \lambda_2) = \begin{cases} \phi(\beta \mid \frac{\lambda_1}{2\lambda_2}, \frac{\sigma^2}{\lambda_2})/2\Phi\left(-\frac{\lambda_1}{2\sigma\lambda_2^{1/2}}\right), & \beta < 0, \\ \phi(\beta \mid \frac{-\lambda_1}{2\lambda_2}, \frac{\sigma^2}{\lambda_2})/2\Phi\left(-\frac{\lambda_1}{2\sigma\lambda_2^{1/2}}\right), & \beta \geq 0. \end{cases} \tag{2.8}$$

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 [15] and inverse-gamma convolution can be derived similarly [16]. [4] show that the regularly varying tails of half-Cauchy priors work well for low-dimensional functions of normal vector mean, where flat priors give poorly calibrated inference.

Lemma 3.1. *Let $X_i \sim \mathcal{C}(0,1)$ $(i = 1, 2)$ be Cauchy distributed random variates, then $Z = w_1 X_1 + w_2 X_2 \sim \mathcal{C}(0, w_1 + w_2)$. where $w_1, w_2 > 0$.*

Lemma 3.2. *Let $X_i \sim \mathcal{IG}(\alpha t_i, \alpha t_i^2)$ $(i = 1, 2)$, then $Z = X_1 + X_2 \sim \mathcal{IG}\{\alpha(t_1 + t_2), \alpha(t_1^2 + t_2^2)\}$, where $\alpha, t_1, t_2 \geq 0$, and $\mathcal{IG}(\alpha t, \alpha t^2)$ is an inverse-Gaussian random variable with density*

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

Both of these results follow from straightforward applications of the Cauchy-Schlömilch transformation. We give a proof for the Cauchy convolution identity below.

Proof. Exploiting symmetry and the Lagrange identity $(a^2 + b^2)(c^2 + d^2) = (ac + bd)^2 + (ad - bc)^2$, leads 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 to $x+w_2^{-1}z(w_1^{-1}+w_2^{-1})^{-1}$ and letting $a=1+z^2(w_1+w_2)^{-2}$, $b=(w_1w_2)^{-1}$, $c=z(w_2-w_1)\{(w_1+w_2)w_1w_2\}^{-1}$, $d=z(w_2-w_1)\{(w_1+w_2)w_1w_2\}^{-1}$ gives

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

If we let $y=x^{-1}$ and apply the 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 $X\sim\mathcal{C}(\mu,\sigma)$, $\psi_X(t)=\exp(it\mu-|t|\sigma^2)$, 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 $C_i\sim\mathcal{C}(0,1)$, 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 $C\sim\mathcal{C}(0,1)$.

The most general result in this category is due to [15], who they showed the following: Let $(X_1,\ldots,X_m)$ and $(Y_1,\ldots,Y_m)$ be independent and identically distributed $\mathcal{N}(0,\Sigma)$ for an arbitrary positive definite matrix Σ, then $Z=\sum_{j=1}^m w_jX_j/Y_j\sim\mathcal{C}(0,1)$, as long as $(w_1,\ldots,w_m)$ is independent of (X,Y), $w_j\geq 0$ $(j=1,\ldots,m)$ and $\sum_{j=1}^m w_j=1$.

4 Discussion

The Cauchy-Schlömilch and Liouville transformations not only guarantee simple normalizing constants for $f(\cdot)$, they also establish 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 [17]. [14] provide 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 [22] that proves a uniform correlation mixture of a bivariate Gaussian density with unit variance is a function of the maximum norm:

$$\int_{-1}^1 \frac{1}{4\pi(1-\rho^2)^{1/2}}\exp\left\{-\frac{x_1^2+x_2^2-2\rho x_1x_2}{2(1-\rho^2)}\right\}d\rho=\frac{1}{2}\left\{1-\Phi(||x||_\infty)\right\}, \tag{4.1}$$

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 (4.1) was introduced by [7] as uniform mixtures of a chi random variate with 3 degrees of freedom, but the representation as a uniform correlation mixture is a new find. We make a few remarks connected to the Erdelyi's integral identity, which is key to the proof of the uniform correlation mixture of (4.1).

Lemma 4.1. *Erdelyi's identity, defined by*

$$\int_{1/2}^\infty \frac{e^{-x^2z}}{4\pi z(2z-1)^{1/2}}dz=\frac{1}{2}\left\{1-\Phi(x)\right\},\quad x\geq 0, \tag{4.2}$$

follows from the Laplace transformation $(1+u)^{-1}=\int_0^\infty\exp\{-v(1+u)\}dv$.

Proof. Apply the transform $u = 2z - 1$ to the left hand side of (4.2), denoted by I, to obtain

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

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

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

and letting $z^2 = x^2 + 2v$ we get

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

□

The second candidate is the symmetric stable distribution, defined by its characteristic function $\phi(t) = \exp(-|t|^\alpha), 0 < \alpha \leq 2$. It admits a normal scale mixture representation with mixing density as $f(v) = 2^{-1} s_{\alpha/2}(v/2), v > 0$, where $s_{\alpha/2}$ is the positive stable density with index $\alpha/2$ [9]. The exponential power density arising as a dual of the symmetric stable density also has a normal scale mixture representation with important application in Bayesian bridge regression [19].

$$e^{-|x|^\alpha} = \int_0^\infty e^{-x\eta} g(\eta) d\eta, \quad g(\eta) = \sum_{j=1}^\infty (-1)^j \frac{\eta^{-j\alpha-1}}{j!\Gamma(-\alpha j)} \ ,$$

[19] derive this as a limiting result of the scale-mixture of beta representation for k-montone densities and utilizing the complete monotonicity of exponential power density. 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 from the Cauchy-Schlömilch formula (1.1). Other potential applications include using Liouville formula to recognize and generate global-local mixtures, and to calculate higher-order closed-form moments $E(X^n)$ for random variables X that admit a global-local representation.

References

[1] D.F. Andrews and C.L. Mallows, *Scale mixtures of normal distributions*, Journal of the Royal Statistical Society. Series B: Statistical Methodology **36** (1974), no. 1, 99–102.

[2] Rose Baker, *Probabilistic applications of the Schlömilch transformation*, Communications in Statistics – Theory and Methods **37** (2008), no. 14, 2162–2176.

[3] O Barndorff-Nielsen, John Kent, and Michael Sørensen, *Normal variance-mean mixtures and z distributions*, International Statistical Review **50** (1982), 145–159.

[4] Anindya Bhadra, Jyotishka Datta, Nicholas G Polson, and Brandon Willard, *Default bayesian analysis with global-local shrinkage priors*, Biometrika **103** (2016), no. 4, 955–969.

[5] Anindya Bhadra, Jyotishka Datta, Nicholas G Polson, Brandon Willard, et al., *The horseshoe+ estimator of ultra-sparse signals*, Bayesian Analysis **12** (2017), no. 4, 1105–1131.

[6] George Boros, Victor H Moll, and JJ Foncannon, *Irresistible integrals: symbolics, analysis and experiments in the evaluation of integrals*, The Mathematical Intelligencer **28** (2006), no. 3, 65–68.

[7] Maurice C Bryson and Mark E Johnson, *Constructing and simulating multivariate distributions using Khintchine's theorem*, Journal of Statistical Computation and Simulation **16** (1982), no. 2, 129–137.

[8] Yogendra P Chaubey, Govind S Mudholkar, and MC Jones, *Reciprocal symmetry, unimodality and Khintchine's theorem*, Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, vol. 466, The Royal Society, 2010, pp. 2079–2096.

[9] Tilmann Gneiting, *Normal scale mixtures and dual probability densities*, Journal of Statistical Computation and Simulation **59** (1997), no. 4, 375–384.

[10] Chris Hans, *Comment on Article by Polson and Scott*, Bayesian Analysis **6** (2011), no. 1, 37–41.

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

[12] MC Jones, *On Khintchine's theorem and its place in random variate generation*, The American Statistician **16** (2002), no. 4, 304–307.

[13] Paul Lévy, *Sur certains processus stochastiques homogènes*, Compositio mathematica **7** (1940), 283–339.

[14] Jason A. Palmer, Ken Kreutz-Delgado, and Scott Makeig, *AMICA: An adaptive mixture of independent component analyzers with shared components*, Tech. report, San Diego, CA: Technical report, Swartz Center for Computational Neuroscience, 2011.

[15] Natesh S Pillai and Xiao-Li Meng, *An unexpected encounter with Cauchy and Lévy*, Annals of Statistics (to appear) (2016).

[16] Nicholas G. Polson and James G. Scott, *On the Half-Cauchy Prior for a Global Scale Parameter*, Bayesian Analysis **7** (2012), no. 4, 887–902 (en).

[17] ———, *Data augmentation for non-Gaussian regression models using variance-mean mixtures*, Biometrika **100** (2013), no. 2, 459–471.

[18] Nicholas G. Polson, James G. Scott, and Jesse Windle, *Bayesian inference for logistic models using Pólya–Gamma latent variables*, Journal of the American Statistical Association **108** (2013), no. 504, 1339–1349.

[19] ———, *The Bayesian bridge*, Journal of the Royal Statistical Society. Series B: Statistical Methodology **76** (2014), no. 4, 713–733.

[20] Nicholas G. Polson and Steven L. Scott, *Data augmentation for support vector machines*, Bayesian Analysis **6** (2011), no. 1, 1–23.

[21] V. Seshadri, *Halphen's laws*, Encyclopedia of Statistical Sciences, John Wiley and Sons, Inc., Hoboken, New Jersey, 2004.

[22] Kai Zhang, Lawrence D Brown, Edward George, and Linda Zhao, *Uniform Correlation Mixture of Bivariate Normal Distributions and Hypercubically Contoured Densities That Are Marginally Normal*, The American Statistician **68** (2014), no. 3, 183–187.