

Global-Local Mixtures

BY ANINDYA BHADRA

Department of Statistics, Purdue University.
 bhadra@purdue.edu.

JYOTISHKA DATTA

Department of Statistical Science, Duke University.
 jd298@stat.duke.edu

NICHOLAS G. POLSON AND BRANDON WILLARD

The University of Chicago Booth School of Business.
 ngp@chicagobooth.edu AND bwillard@uchicago.edu

SUMMARY

We show how global-local mixtures can be generated using integral identities known as the Schlömilch and the Liouville transformation. These identities lead to simpler proofs of many known normal-scale mixture distributions such as the Laplace or Lasso, the logit and the quantile. We also show how they lead to new probability distributions as global-local scale mixtures of appropriate baseline densities.

Some key words: Global-local mixture, Scale mixture, Stable laws.

1. INTRODUCTION

Many statistical problems involve regularization penalties, that can be viewed as global-local mixture prior distributions (Polson & Scott, 2011; Hans, 2011). A random variable $X = (x_1, \dots, x_p)^T$ is said to have a global-local mixture distribution if we can write its marginal probability density function as:

$$p(x_i) = \int_0^\infty p(x_i | \tau) p(\tau) d\tau \text{ where,}$$

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

Here τ is a global mixing parameter and λ_i 's are the local ones. There is great interest in being able to analytically calculate the marginal distributions of a global-local mixture - primarily as a log-penalty in sparse Bayesian estimation. Our approach here is to show how two simple integral identities, Schlömilch and Liouville integrals, derive useful 'closed-form' identities very quickly. Moreover, it provides a framework for generating new random variables via the transformation $Y = aX - bX^{-1}$ for hyper-parameters a and b . It also provides a transformation framework that links together different global-local mixtures.

We start by stating the two key integral identities that we shall use throughout the paper, the Cauchy-Schlömilch transformation:

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

and the Liouville transformation:

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

See (Boros et al., 2006; Baker, 2008; Jones, 2014) for a detailed discussion of these identities.

Proof. Equation (1) follows from the simple transformation $t = b/ax$:

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

Now, adding we get:

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

Then, transforming $y = b/t - at$ that implies $dy = -a(1 + b/at^2)$, we get (1), namely, $I = (2a)^{-1} \int_0^\infty f(y^2) dy$. \square

The proof for the Liouville transformation follows a very similar proof which we omit for the sake of brevity.

A useful generalization of the Cauchy-Schlömilch transformation is:

$$\int_0^\infty f \{ [x - g(x)]^2 \} dx = \int_0^\infty f(y^2) dy \quad (3)$$

Where $g(x) = g^{-1}(x)$ is a self-inverse function such as $g(x) = b/x$ or $g(x) = -a^{-1} \log\{1 - \exp(ax)\}$, see (Jones, 2010, 2014) for a detailed discussion.

2. GLOBAL LOCAL SCALE MIXTURE

2.1. Lasso as a Normal Scale Mixture

It is well known that the double exponential or Laplace distribution arises as a normal scale mixtures with a Gamma mixing density (Andrews & Mallows, 1974). A simple transformation proof follows from the Cauchy-Schlömilch formula where, $f(x) = e^{-x}$ in (1). The normal integral identity, $\int_0^\infty f(y^2) dy = \int_0^\infty e^{-y^2} dy = \sqrt{\pi}/2$ becomes $\int_0^\infty f(ax - bx^{-1}) dx$:

$$\int_0^\infty e^{-ax^2 - b/x^2} dx = \frac{1}{2} \sqrt{\frac{\pi}{a}} e^{-2\sqrt{ab}}$$

This can also be written with $x = b^{1/2}t$ and $c = ab$ to get the Andrews & Mallows (1974) result for double exponential:

$$\int_0^\infty \exp\{-ct^2 - 1/t^2\} dt = \frac{1}{2} (\pi/c)^{1/2} \exp(-2\sqrt{c}) \quad (4)$$

The advantage of this short proof is that the Laplace density can be viewed as a transformation of the normal.

Remark 1. (Need to check this carefully!)

The usual identity for lasso is the following proved by Lévy (1940).

$$\int_0^\infty \frac{a}{\sqrt{2\pi x^3}} e^{-a^2/2x} e^{-\lambda x} dx = e^{-a\sqrt{2\lambda}}$$

We can also write the above identity as:

$$E \left[\exp\{-\lambda^2/4\gamma_{\frac{1}{2}, \frac{1}{2}}\} \right] = \exp(-\lambda),$$

where $\gamma_{\frac{1}{2}, \frac{1}{2}}$ is a Gamma random variable.

2.2. Logit and Quantile as Normal Scale Mixtures

Bayesian approaches to logistic modeling involves the Pólya–Gamma distribution (Polson et al., 2013). The two key integral identities for the hyperbolic-GIG (Barndorff-Nielsen & Halgreen, 1977) and Z-Pólyamixtures (Polson et al., 2013; Barndorff-Nielsen et al., 1982) are

$$\frac{\alpha^2 - \kappa^2}{2\alpha} e^{-\alpha|\theta-\mu| + \kappa(\theta-\mu)} = \int_0^\infty \phi(\mu + \kappa\lambda, \lambda) p_{GIG}(1, 0, \sqrt{\alpha^2 - \kappa^2})(\lambda) d\lambda \quad (5)$$

$$\frac{1}{B(\alpha, \kappa)} \frac{e^{\alpha(\theta-\mu)}}{(1 + e^{\theta-\mu})^{\alpha+\kappa}} = \int_0^\infty \phi(\mu + \kappa\lambda, \lambda) p_{Polya}(\alpha, \kappa)(\lambda) d\lambda \quad (6)$$

where $\phi(\mu + \kappa\lambda, \lambda)$ denotes the normal density. The corresponding mixture distributions are Generalized Inverse-Gaussian and the Polya-Z denoted by p_{GIG} and p_{Polya} respectively. Rather surprisingly, the logit and quantile identities (5) and (6) can also be derived from the integral identity (1). We give the proofs below:

LEMMA 2.1. *The integral identities (5) and (6) admit a global-local mixture representation via the Cauchy-Schlömilch transformation in (1).*

Proof. To prove (5), let $f(x) = e^{-x^2/2}$, $a = \alpha$, $b = |\theta - \phi|$ in (1), then we have:

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

Letting $\nu = y^2$ and re-arranging the constant terms, we get the relation:

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

Now, multiplying both sides by $2^{-1}(\alpha^2 - \kappa^2) \exp\{\kappa(\theta - \mu)\}$, re-arranging terms to complete squares, so that:

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

The mixing distribution on the right hand side is the exponential distribution with rate parameter $(\alpha^2 - \kappa^2)/2$, which is a special case of the Generalized inverse Gaussian or GIG distribution. Etienne Halphen circa 1941 (Seshadri, 1997) introduced the GIG and Barndorff-Nielsen & Halgreen (1977) studied its properties. The density with parameters $(\lambda, \delta, \gamma)$ has the form:

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

where K_λ is the modified Bessel function of the second kind. When δ or γ is zero (in our case, $\delta = 0$), the normalizing constant is to be replaced by the corresponding limits as $K_\lambda(u) \sim \Gamma(|\lambda|)2^{|\lambda|-1}u^{|\lambda|}$ when $\lambda \neq 0$. It turns out that if $\delta = 0$, GIG is identical to a Gamma distribution with density

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

Next we present a simple proof for the Pólya–Gamma mixture in (6). First, write κ for $a - b/2$:

$$\frac{(e^\psi)^a}{(1 + e^\psi)^b} = 2^{-b} e^{\kappa\omega} \int_0^\infty e^{-\omega\psi^2/2} p(\omega) d\omega \quad (7)$$

where $\omega \sim PG(b, 0)$, the Pólya–Gamma distribution with density given by:

$$p(x \mid b, 0) = \frac{2^{(b-1)}}{\Gamma(b)} \sum_{n=0}^\infty (-1)^n \frac{\Gamma(n+b)}{\Gamma(n+1)} \frac{2n+b}{\sqrt{2\pi}x^3} e^{-\frac{(2n+b)^2}{8x}}$$

The logit function corresponds to $a = 0, b = 1$ in (7). The identity then simplifies to:

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

70 We will now prove that the right hand side equals the left hand side in (8). To do this, write the left-hand side as:

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

Now, using the change of variable $\omega = t^{-2}$, we get

$$\begin{aligned} 75 \quad I &= e^{-\psi/2} \sum_{n=0}^\infty (-1)^n \frac{(2n+1)}{\sqrt{2\pi}} \int_0^\infty e^{-\left(\frac{\psi^2}{2t^2} + \frac{(2n+1)^2 t^2}{8}\right)} dt \\ &= e^{-\psi/2} \sum_{n=0}^\infty (-1)^n \frac{(2n+1)}{\sqrt{2\pi}} \int_0^\infty e^{-\frac{1}{2}\left(\frac{(2n+1)t}{2} - \frac{\psi}{t}\right)^2} e^{-\frac{(2n+1)}{2}\psi} dt \\ &= \sum_{n=0}^\infty (-1)^n \left[e^{-(n+1)\psi} \frac{(2n+1)}{\sqrt{2\pi}} \left\{ \int_0^\infty e^{-\frac{1}{2}\left(\frac{(2n+1)t}{2} - \frac{\psi}{t}\right)^2} dt \right\} \right]. \end{aligned}$$

Finally, apply our first integral identity to the integral within curly brackets to obtain:

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

Putting things back together, we arrive at the logit function:

$$I = \sum_{n=0}^\infty (-1)^n e^{-(n+1)\psi} = \frac{1}{1 + e^\psi}.$$

An alternative proof using Laplace transformation is in Polson et al. (2013).

Remark 2. In the special case $\alpha = \kappa$, Equation (5) and the limiting result yields the following identity:

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

Hence it is as if the latent variable has a marginal improper uniform prior $p(\lambda) = 1, \forall \lambda \in \mathbb{R}$. We can also write the result as:

$$\int_0^\infty \phi(b \mid -a\lambda, c\lambda) d\lambda = a^{-1} \exp(-2 \max(ab/c, 0)) \quad (9)$$

This is useful in support vector machines Polson & Scott (2011).

Remark 3. In a discussion of Polson & Scott (2011), Hans (2011) pointed out that the elastic-net regression can be recast as a normal mean variance mixture problem where the mixing density belongs to a ‘orthant-normal’ family of distributions. The orthant-normal prior on a single regression coefficient β given hyper-parameters λ_1 and λ_2 is given by:

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

Then Hans (2011) proves the following theorem:

THEOREM 2.1. *Under the orthant normal prior on β given above and prior on λ_2 with the density*

$$p(\lambda_2 \mid \lambda_1, \sigma^2) = \frac{\lambda_1^2}{2\sigma^2\lambda_2^2} \Phi\left(\frac{-\lambda_1}{2\sigma\sqrt{\lambda_2}}\right), \lambda_2 > 0$$

the induced marginal prior on β is a double-exponential, in particular, $\beta \mid \lambda_1, \sigma^2 \sim DE(\lambda_1/\sigma^2)$.

This result comes as a corollary of (9).

3. TRANSFORMATIONS OF SCALE DISTRIBUTION

Jones (2014) characterizes the densities of the form $f(x) = 2g\{t(x)\}$ and provided conditions for $t(\cdot)$ for $f(\cdot)$ to be a density. This result allows us to generate new distributions starting from a baseline ‘simple’ distribution and transforming the scale by a suitable function. The main result is quoted below:

PROPOSITION 3.1. *Let $\Pi : \mathcal{D} \rightarrow \mathcal{S}_f$ be a piecewise differentiable monotone increasing function with inverse t , where $\mathcal{D} \supset \mathcal{S}_g \ni 0$. Suppose that*

$$\Pi(y) - \Pi(-y) = y, \text{ for all } y \in \mathcal{D}$$

Then if $g(x)$ is a density on \mathcal{S}_g symmetric about zero, $f(x) = 2g\{t(x)\} \equiv 2g\{\Pi^{-1}(x)\}$ is a density on \mathcal{S}_f .

Curiously, an equivalent formulation of $t(\cdot)$ is that $t(x)$ should have the form $x - s(x)$ where $s : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ is an onto monotone decreasing function that is self-inverse, i.e. $s\{s(x)\} = x$. This together with the Cauchy-Schlömilch transformation (3) guarantees that the left hand side is a density on the appropriate domain, and provides for a way to represent the resulting $g\{t(x)\}$ as a global-local scale mixture!

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

$$105 \quad t(2) : \Pi_T(y) = (1/2)(y + \sqrt{4b + y^2}) \Rightarrow \Pi_T^{-1}(x) = t_T(x) = x - (b/x) \quad (10)$$

$$\text{Logistic} : \Pi_L(y) = a^{-1} \log(1 + e^{ay}) \Rightarrow \Pi_L^{-1}(x) = t_L(x) = a^{-1} \log(e^{ax} - 1) \quad (11)$$

Now, 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.

3.1. Symmetric and R-symmetric distributions

For absolutely continuous random variables supported on the positive half-line \mathbb{R}^+ , the density function $f(\cdot)$ has reciprocal symmetry (R-symmetry) if $f(\theta y) = f(\theta/y)$ for all $y > 0$ and some $\theta > 0$ (Mudholkar and Wang, 2007). It follows from the integral identity (1) that if $f(x)$, $x \geq 0$ is a density function, so is $g(x) = 2af(|ax - b/x|)$, $x > 0$, and $f(\cdot)$ and $g(\cdot)$ are called the mother and the daughter pdf. As Chaubey et al. (2010) pointed out, there is an one-to-one correspondence between f and g , and furthermore if $f(x)$ is the pdf of a symmetric real-valued random variable Y , 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. Chaubey et al. (2010) provides 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:

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

110

4. DISCUSSION

115

Apart from providing simple proofs for well-known normal scale mixture representations, the main motivation behind studying properties of the integral identities is twofold: we can generate new distributions by a suitable scale transformation $f(x) = 2g\{t(x)\}$ of a ‘simple’ baseline function f under some conditions, and for $t(x) = x - 1/x$, we get R-symmetric densities on \mathbb{R}^+ from well-known symmetric densities and vice-versa, and given a density $f(x)$ we can create a new global-local scale family $f(ax - bx^{-1})$, by reallocating its probability mass.

120

125

130

Khintchine’s theorem for unimodality of univariate distributions provides a useful tool for generating univariate and multivariate random variables (Bryson & Johnson, 1982). 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 (2010, 2012) discuss how Khintchine’s theorem allows us to construct both univariate and multivariate distributions, the latter with special dependence structure. Jones (2014) develops an extended Khintchine’s theorem that further lets one generate random variable with unimodal density of the form $2g(t(x))$ discussed before in Proposition 3.1. The Cauchy-Schlömilch transformation not only guarantees an ‘astonishingly simple’ normalizing constant for $f(\cdot)$, it also establishes the wide class of unimodal densities as global-local scale mixtures. As we have discussed before, the global-local scale mixtures with conditional Gaussianity hold a special place in Statistical literature, as these models can be rapidly fit using an expectation-maximization algorithm as pointed out by Polson & Scott (2013). This has been extended in a 2011 unpublished note by Palmer, Kreutz-Delgado and Makeig who have provided

a 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 this note with conjectures that two remarkable identities arise as corollaries of the Cauchy-Schlömilch transformation. 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:

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

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

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

where $s_{\alpha/2}$ is the density of the positive stable distribution with index $\alpha/2$ (Feller, 1971). We conjecture that these two results follow as upshots of the Cauchy-Schlömilch formula (1).

REFERENCES

- 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. & HALGREEN, C. (1977). Infinite divisibility of the hyperbolic and generalized inverse gaussian distributions. *Probability Theory and Related Fields* **38**, 309–311.
- 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.
- 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.
- HANS, C. (2011). Comment on article by polson and scott. *Bayesian Analysis* **6**, 37–41.
- JONES, M. (2010). Distributions generated by transformation of scale using an extended cauchy-schlömilch transformation. *Sankhya A* **72**, 359–375.
- JONES, M. (2012). On khintchine's theorem and its place in random variate generation. *The American Statistician* .
- JONES, M. (2014). Generating distributions by transformation of scale. *Statistica Sinica* **24**, 749–772.
- LÉVY, P. (1940). Sur certains processus stochastiques homogènes. *Compositio mathematica* **7**, 283–339.
- POLSON, N. G. & SCOTT, J. G. (2013). Data augmentation for non-gaussian regression models using variance-mean mixtures. *Biometrika* , ass081.
- 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, S. L. (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.

