Bayesian Analysis (2014) **9**, Number 3, pp. 551–556

Comment on Article by Finegold and Drton

François Caron* and Luke Bornn†

The authors should be commended for their creation of a flexible yet computationally efficient approach to robust graphical models. The framework proposed by the authors opens up many opportunities for generalizations and extensions. In this comment, we highlight several such ideas which we feel might be useful to practitioners.

Extensions to other scale mixtures of Gaussians and to skew distributions. The models proposed by the authors can be easily extended beyond the Student-t case to other classes of scale mixtures of Gaussians. In particular, the class of generalized hyperbolic distributions (Barndorff-Nielsen and Shephard 2001) generalizes the Student-t and leads to tractable inference.

Let $\mathrm{GiGauss}(\nu, \delta, \gamma)$ be the generalized inverse Gaussian distribution of pdf

$$\frac{(\gamma/\delta)^{\nu}}{2K_{\nu}(\delta\gamma)} x^{\nu-1} \exp\left(-\frac{1}{2}(\delta^2 x^{-1} + \gamma^2 x)\right), \; x > 0. \tag{1}$$

This distribution admits as special cases the gamma, inverse gamma and inverse Gaussian distributions. The classical (multivariate) generalized hyperbolic (GH) distribution can be constructed in the following way:

$$\boldsymbol{X} \sim \mathcal{N}_p(0, \boldsymbol{\Psi}), \quad \boldsymbol{Y} = \boldsymbol{\mu} + \boldsymbol{X}/\sqrt{\tau}$$

with $1/\tau \sim \mathrm{GiGauss}(\nu, \delta, \gamma)$. Its marginal density is

$$\frac{(\gamma/\delta)^{\nu}}{(2\pi)^{p/2}\gamma^{\nu-p/2}K_{\nu}(\delta\gamma)} q_y(\boldsymbol{\mu}, \boldsymbol{\Psi})^{\nu-p/2} K_{\nu-p/2}(\gamma q_y(\boldsymbol{\mu}, \boldsymbol{\Psi}))$$

where $K_\nu(x)$ is the modified Bessel function of the third kind, and $q_y(\boldsymbol{\mu}, \boldsymbol{\Psi}) = (\boldsymbol{y} - \boldsymbol{\mu})^T \boldsymbol{\Psi}^{-1}(\boldsymbol{y} - \boldsymbol{\mu})$. This multivariate GH distribution admits as a special case the (classical) multivariate Student-t ($\nu < 0$ and $\gamma = 0$), as well as multivariate versions of the Laplace (Kyung et al. 2010; Bornn et al. 2010), Cauchy, normal-gamma and normal inverse Gaussian distributions. The three parameters of the GH distributions allow for more flexibility, in particular for modeling the tails, as shown in Figure 1.

This model has conjugacy properties, as the conditional distribution of τ given the other variables is still generalized inverse Gaussian (e.g. Caron et al. 2012). Alternative GH and Dirichlet GH can be constructed in the same way:

$$\boldsymbol{X} \sim \mathcal{N}_p(0, \boldsymbol{\Psi})$$

- Classical GH distribution

$$\boldsymbol{Y} = \boldsymbol{\mu} + \boldsymbol{X}/\sqrt{\tau} \quad \text{with} \quad 1/\tau \sim \mathrm{GiGauss}(\nu, \delta, \gamma)$$

*University of Oxford caron@stats.ox.ac.uk
†Harvard University bornn@stat.harvard.edu

 DOI:10.1214/14-BA908

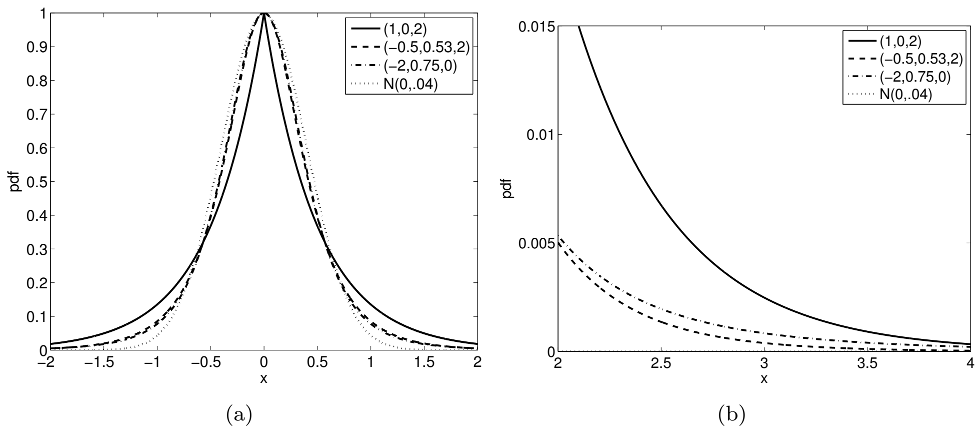


Figure 1: [Figure reproduced from (Caron et al. 2012) with permission] Probability density functions of the generalized hyperbolic distribution for several values of (ν, δ, γ) correspond to a Laplace, Normal inverse Gaussian and Student t distribution. The pdf of the normal distribution is also shown for comparison. (a) Behavior around 0 and (b) Tail behavior.

- Alternative GH distribution

$$Y_j = \mu_j + X_j/\sqrt{\tau_j} \quad \text{with} \quad 1/\tau_j \overset{iid}{\sim} \text{GiGauss}(\nu, \delta, \gamma)$$

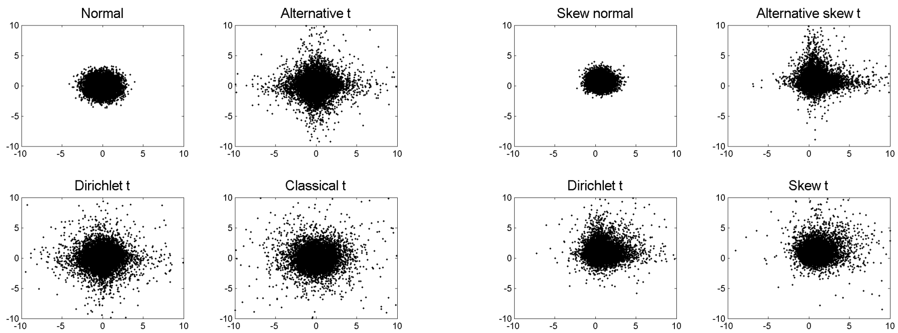
- Dirichlet GH distribution

$$Y_j = \mu_j + X_j/\sqrt{\tau_j} \quad \text{with} \quad 1/\tau_j|P \overset{iid}{\sim} P, \quad P \sim \text{DP}(\alpha, \text{GiGauss}(\nu, \delta, \gamma))$$

Another line of research would be to extend these models to handle asymmetry in the data, a topic of great interest, see e.g. Frühwirth-Schnatter and Pyne (2010). The recent works of Capitanio et al. (2003) or Zareifard et al. (2013) on skew normal graphical models could be extended to the family of Student-t distributions proposed in the paper of Finegold and Drton. Example draws from a classical skew-t, alternative skew-t and skew Dirichlet-t bivariate distributions are given in Figure 2.

Alternative models for the precision parameters τ_j. The classical, alternative, and Dirichlet-t formulations presented in the paper are some of many possible specifications for the prior on the precision parameters τ_j. We now highlight some other possible models, plotting simulations of them along the way in Figure 3. Aside from the alternative model which draws independent gamma random variates (Figure 3, row 1) and the Dirichlet-t model (Figure 3, row 2) presented in the paper, one could also specify a spike and slab model, whereby apriori

$$\tau_{ij} \sim \begin{cases} \delta_1, & \text{if } b_{ij} = 1 \\ \Gamma(\nu/2, \nu/2), & \text{if } b_{ij} = 0 \end{cases}$$



(a) t (b) skew t

Figure 2: (a) Multivariate t (b) Skew multivariate t

where b_{ij} are independent Bernoulli draws (Figure 3, row 3). This allows one to induce most of the τ_{ij} to be equal to one, corresponding to a normal noise model. The remaining contaminated observations and/or variables would fall under the independent Gamma (alternative Student-t) regime.

If one wishes to induce correlation within an observation or variable, we feel it is more appropriate to correlate the presence or absence of contamination rather than the actual level of contamination as represented by the value of τ_{ij}. In most cases, contamination is either present or not, and if it is present, the degree to which it affects the observation will be variable-dependent. Because of this, we feel it is natural to induce dependence in b_{ij}, rather than τ_{ij}. This could be done using an autologistic model, a probit model (Figure 3, row 4), or any of a variety of methods for creating correlated Bernoulli draws.

Alternatives also exist for correlating the τ_{ij} while maintaining gamma marginals. Specifically, one could use a copula to maintain the correct Student-t structure marginally while inducing correlation across variables or observations (Figure 3, row 5). Closer to the authors' method, one could alternatively build more complex DP-based models; see e.g. MacEachern (2000).

Alternative priors over decomposable graphs. The authors assume a simple model on decomposable graphs that does not try to induce any structure in the graph. Alternative priors over decomposable graphs, where one can control the clustering of the nodes into cliques and separators, could also be considered in the same framework (Bornn and Caron 2011; Byrne and Dawid 2014).

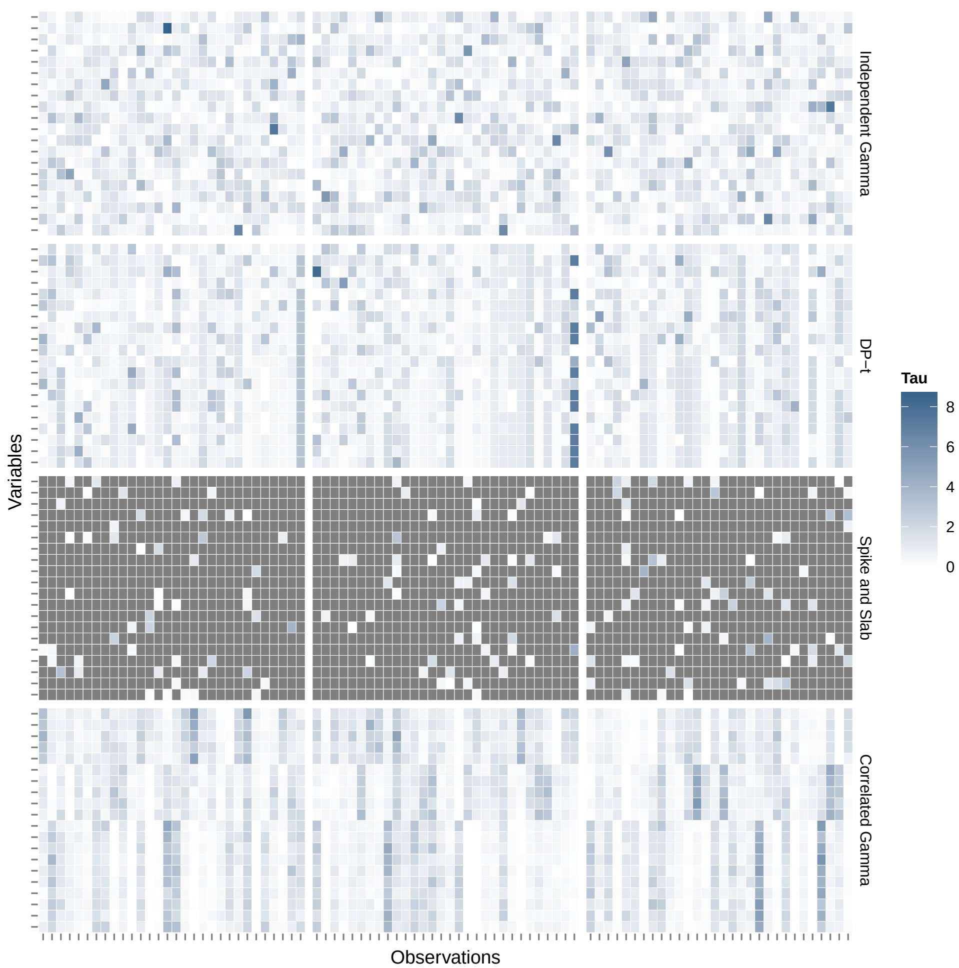


Figure 3: Prior simulations from different model specifications for τ_j. Each of the 5 rows is a different model, and the 3 large columns show separate draws. Within each row/column block, the $p = 20$ rows represent variables and the $n = 30$ columns represent observations. Row 1 is indendent samples from a $\Gamma(1, 1)$ distribution. Row 2 extends this using the Dirichlet-t formulation proposed by Finegold and Drton, using $\alpha = 5$. Row 3 shows independent draws from the spike and slab model with $P(b_{ij} = 1) = 0.9$, and a slab distribution $\Gamma(1, 1)$. Row 4 correlates b_{ij} by drawing them from a probit model where the underlying Gaussian distribution (truncated at 0) has correlation 0.8 within observations and 0.8 within variables. Lastly, row 5 has $\Gamma(1, 1)$ marginals tied together with a pairwise Gaussian copula with parameter $\rho = 0.9$.

References

Barndorff-Nielsen, O. and Shephard, N. (2001). "Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics." *Journal of the Royal Statistical Society B*, 63: 167–241. 551

Bornn, L. and Caron, F. (2011). "Bayesian clustering in decomposable graphs." *Bayesian Analysis*, 6(4): 829–846. 553

Bornn, L., Gottardo, R., and Doucet, A. (2010). "Grouping Priors and the Bayesian Elastic Net." Technical report, Department of Statistics, University of British Columbia.
URL http://arxiv.org/abs/1001.4083 551

Byrne, S. and Dawid, A. P. (2014). "Structural Markov graph laws for Bayesian model uncertainty." *arXiv preprint arXiv:1403.5689*. 553

Capitanio, A., Azzalini, A., and Stanghellini, E. (2003). "Graphical models for skew-normal variates." *Scandinavian Journal of Statistics*, 30(1): 129–144. 552

Caron, F., Bornn, L., and Doucet, A. (2012). "Sparsity-promoting Bayesian dynamic linear models." Technical report, arXiv preprint arXiv:1203.0106. 551, 552

Frühwirth-Schnatter, S. and Pyne, S. (2010). "Bayesian inference for finite mixtures of univariate and multivariate skew-normal and skew-t distributions." *Biostatistics*, 11(2): 317–336. 552

Kyung, M., Gill, J., Ghosh, M., and Casella, G. (2010). "Penalized Regression, Standard Errors, and Bayesian Lassos." *Bayesian Analysis*, 5: 369–412. 551

MacEachern, S. N. (2000). "Dependent Dirichlet processes." Technical report, Ohio State University. 553

Zareifard, H., Rue, H., Khaledi, M. J., and Lindgren, F. (2013). "A skew Gaussian decomposable graphical model." Technical report, arXiv preprint arXiv:1309.5192. 552