Bayesian Analysis (2014) **9**, Number 3, pp. 557–560

# Comment on Article by Finegold and Drton

Babak Shahbaba *

**Abstract.** Scale mixtures of normals have been discussed extensively in the literature as heavy-tailed alternatives to the normal distribution for robust modeling. They have been used either as error models to handle outliers or as prior distributions to provide more reasonable shrinkage of model parameters. The proposed method by Finegold and Drton goes beyond the existing literature both in terms of application (graphical models) and methodology (Dirichlet $t$) for outlier handling. While this approach can be applied to many other problems, in this discussion I will focus on its application in Bayesian modeling of high throughput biological data.

**Keywords:** Robust Bayesian modeling, Scale mixtures of normals, High dimensional problems

## 1 Background

I would like to congratulate the authors for their excellent paper, which nicely combines ideas from graphical modeling, robust Bayesian inference, and Bayesian nonparametrics. Although the proposed method is presented in the context of Bayesian graphical modeling, its application of course is not limited to these models. In this discussion, I mainly focus on what the authors call Dirichlet $t$-distributions. This is a generalization of standard $t$-distribution, which itself can be presented as a scale mixture of normals. In general, the random variable $Y = X\sigma$ has a scale mixture of normal distribution if $X$ has a standard normal distribution, and $\sigma > 0$ has some distribution with a continuous or discrete density $h(\sigma)$ (Andrews and Mallows 1974; West 1984). When $\sigma^2$ has Inv-Gamma$(\nu/2, \nu/2)$ distribution, $Y$ has a $t$-distribution with $\nu$ degrees of freedom. The distribution of $Y$ will become Laplace or horseshoe (Carvalho et al. 2010) if instead of Inv-Gamma we use exponential or half-Cauchy respectively.

Scale mixtures of normals are commonly used as heavy-tailed alternatives to the normal distribution for a better handling of outliers as discussed in this current paper by Finegold and Drton. However, they are also used as prior distributions to provide more reasonable shrinkage of model parameters (e.g., Gelman 2006; Carvalho et al. 2010). Indeed, West (1984) used them both as a model for the error term and as the prior for parameters in regression models. Finegold and Drton explore an interesting extension of these models, i.e., Dirichlet $t$-distributions, for handling outliers. In the next section, I will discuss the application of Dirichlet $t$ as a prior for model parameters in high dimensional problems.

*Department of Statistics, University of California, Irvine, babaks@uci.edu

 DOI:10.1214/14-BA899

## 2 Dirichlet $t$ Prior in High Dimensional Problems

High throughput biological studies typically involve simultaneous assessments of many factors (e.g., genes, proteins). In general, it is expected that only a small number of these factors are directly associated with the outcome of interest (e.g., disease); that is, most of the factors have very small effects (i.e., close to zero). In such problems, the $t$-distribution might be a more reasonable prior, compared to the normal distribution, for the effects: it properly shrinks small effects to zero while allowing for some large effects in the tails.

Shen (2013) examines the shrinkage effect of the $t$ prior in genome-wide association studies (GWAS) and shows that the following alternative specification performs very well compared to normal and Student-$t$ models:

$$\begin{aligned} \gamma_i &= \xi_i \sqrt{\nu/u_i} \\ \boldsymbol{\xi} &\sim N(0, \Sigma) \\ u_i &\sim \chi^2_\nu. \end{aligned}$$

This is similar to what the authors in this paper call the "alternative $t$-distribution." Here, $\gamma_i$ is the effect of $i^{th}$ Single Nucleotide Polymorphism (SNP), and $u_i$ is a SNP-specific scale parameter. Note that there is a separate scale parameter for each SNP, as opposed to using a common scale parameter, $u$, for all SNPs. It would be interesting to see whether a Dirichlet $t$ prior for SNP effects could further improve the performance of this model.

We previously explored a similar idea in the context of gene expression data analysis by using the following model (Shahbaba and Johnson 2012):

$$\begin{aligned} y_{ij} \,|\, \alpha_i, \beta_i &\sim N(\alpha_i + \beta_i x_{ij}, \sigma_i^2) \\ \beta_i \,|\, \tau_i^2 &\sim N(0, \tau_i^2) \\ \tau_i^2 | G &\sim G \\ G &\sim \mathcal{D}(G_0, \gamma). \end{aligned}$$

Here, $y_{ij}$ denotes the $j^{th}$ observed gene expression value for gene $i$, $x_{ij}$ is a binary indicator for disease status (0=healthy, 1=diseased), and $\beta_i$ is the corresponding gene effect. Note that in our model, the hyperparameter $\tau_i^2$ (as opposed to $\sigma_i^2$) has an unknown distribution $G$ with a Dirichlet process prior, $\mathcal{D}(G_0, \gamma)$. The model was simplified for gene-level summary statistics, $z_i$,

$$\begin{aligned} z_i | \tau_i^2 &\sim N(0, \tau_i^2) \\ \tau_i^2 | G &\sim G \\ G &\sim \mathcal{D}(G_0, \gamma). \end{aligned}$$

We proposed this model as an alternative to the method of Guindani et al. (2009). Although we found that our method performs well especially when the underlying model is misspecified, we could not clearly describe its properties as Finegold and Drton did in their paper. An open problem is to provide a theoretical basis for improved performance under Dirichlet-$t$ priors.

## 3 Extension to Other Scale Mixtures of Normals

In our proposed model for gene expression data, we did not limit our choice of $G_0$ to conjugate priors. By using other distributions instead of Inv-Gamma, it is possible to create alternative prior distributions besides Dirichlet $t$. Figure 1 shows the result of using a half-Cauchy (Gelman 2006; Carvalho et al. 2010) distribution as $G_0$. It would be interesting to examine the properties of such distributions as priors in high dimensional problems.

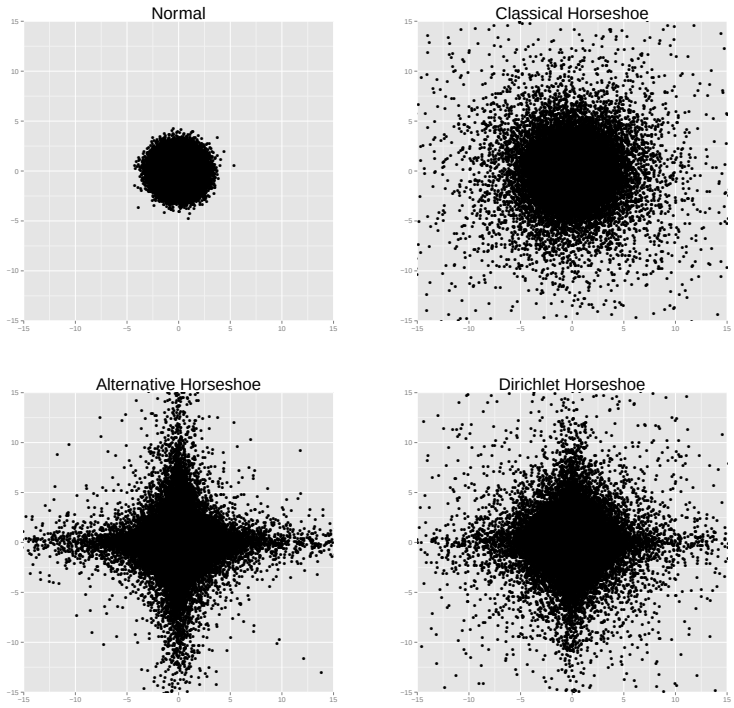


Figure 1: Using a half-Cauchy distribution instead of Inv-Gamma as the baseline distribution in the method of Finegold and Drton.

## References

Andrews, D. F. and Mallows, C. L. (1974). "Scale Mixtures of Normal Distributions." *Journal of the Royal Statistical Society. Series B (Methodological)*, 36(1): 99–102. 557

Carvalho, C. M., Polson, N. G., and Scott, J. G. (2010). "The horseshoe estimator for sparse signals." *Biometrika*, 97(2): 465–480. 557, 559

Gelman, A. (2006). "Prior distributions for variance parameters in hierarchical models." *Bayesian Analysis*, 1: 515–533. 557, 559

Guindani, M., Müller, P., and Zhang, S. (2009). "A Bayesian discovery procedure." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(5): 905–925. 558

Shahbaba, B. and Johnson, W. O. (2012). "Bayesian nonparametric variable selection as an exploratory tool for discovering differentially expressed genes." *Statistics in Medicine*, 32(12): 2114–26. 558

Shen, J. (2013). "Bayesian hierarchical models with spatial prior distributions in genome-wide association studies." Ph.D. thesis, Department of Statistics, UC Irvine. 558

West, M. (1984). "Outlier models and prior distributions in Bayesian linear regression." *Journal of the Royal Statistical Society, Series B*, 46(3): 431–439. 557