Uniform Scale Mixture Models with Applications to Variance Regression

Z. STEVE QIN
University of Michigan, Ann Arbor, MI, USA.

PAUL DAMIEN
University of Michigan, Ann Arbor, MI, USA.

STEPHEN WALKER
University of Bath, UK.

SUMMARY

Scale mixtures of uniform distributions are used to model non-normal data in both univariate and multivariate settings. In addition to providing greater modelling flexibility, the use of scale mixtures of uniforms also results in straightforward computational strategies, particularly in a Bayesian analysis where Monte Carlo methods are used.

Key words: Scale mixtures, Gibbs sampling, Bayesian inference, variance regression.

1 Introduction

Markov chain Monte Carlo (MCMC) methods, such as the Gibbs sampler (Gelfand and Smith, 1990; Smith and Roberts, 1993) have made Bayesian analysis of complex models relatively straightforward. Within the framework of MCMC, various methods have been developed to make computations simpler; see, for example, Polson (1996), Damien et al. (1999).

It has also been recognised that statistical models can be generalised, in a variety of ways, with Bayesian inference remaining tractable, due to MCMC. This paper is concerned with generalising the popular scale mixture of normal (SMN) model, Andrews and Mallows (1974), using the idea of a scale mixture of uniform (SMU) model. We develop a richer class of model with uniform distributions replacing normal distributions. Surprisingly perhaps, it turns out that the MCMC is easier to implement for the scale mixture of uniform model compared to the original scale mixture of normal model.

The important issues are as follows. Scale mixtures of normals only provide densities with heavier tails compared to the normal density. On the other hand, scale mixtures of uniforms coincide with the class of unimodal and symmetric densities. Another point is that we are able to consider variance regression models in a manner which is essentially no more complex than the standard mean regression model.

The layout of the paper is as follows. In Section 2 we provide definitions and facts relating to scale mixtures of uniforms. In Section 3 we demonstrate the usefulness of scale mixtures of uniforms for building and estimating a mean and variance regression model, and include both simulated and real examples. Section 4 provides a comparative analysis between SMU and SMN models. Section 5 considers the extension to multivariate models and finally we end with a brief discussion in Section 6.

2 Background

We start by considering scale mixture of normals (Andrews and Mallows, 1974).

Definition 2.1 Scale mixtures of normals
A random variable Y defined on $\mathbb{R}$ has a scale mixture of normal distribution if the density for Y can be written in the form

$$p_Y(y) = \int_{\lambda>0} N(y|\theta, \alpha(\lambda)\sigma^2)\,\pi(\lambda)\,d\lambda,$$

where $\alpha(\cdot)$ is a positive function on and $\pi(\lambda)$ is a probability density function.

The class of models defined above is very large and useful. However, a restrictive property of $p_Y(\cdot)$ is that it only provides densities with tails heavier than normal, i.e. leptokurtic. To deal with this we can introduce the idea of a scale mixture of uniform density:

Definition 2.2 Scale mixture of uniforms
A random variable Y defined on $\mathbb{R}$ has a scale mixture of uniform representation if it may be written using the following representation:

$$Y|[V = v] \sim Un(\mu - \sigma v, \mu + \sigma v),$$

where $\sigma > 0$, and V has density $f_V(\cdot)$, defined on $\mathbb{R}_+$.

We now collect together some facts which can be easily established.

Fact 2.1 *If $V \sim Ga(3/2, 1/2)$ then $Y \sim N(\mu, \sigma^2)$.*

Fact 2.2 *The class of unimodal, symmetric densities on $\mathbb{R}$ coincides with the scale mixture of uniform densities.*

Fact 2.3 *If Y has a scale mixture of normal representation, i.e.,*

$$Y|\lambda \sim N(\mu, \sigma^2/\lambda),$$

and

$$\lambda \sim \pi(\cdot),$$

then the corresponding representation via scale mixture of uniforms is given by

$$Y|[V = v] \sim Un\left(\mu - \sigma\sqrt{v}, \mu + \sigma\sqrt{v}\right),$$

$$V \sim f_V(v),$$

where

$$f_V(v) = \int p(v|\lambda)\,\pi(\lambda)\,d\lambda$$

and $p(v|\lambda)$ is the density function of $Ga(v|3/2, \lambda/2)$.

In the next section we demonstrate how, using scale mixtures of uniforms, we can analyse straightforwardly a mean/variance regression model. We illustrate with some examples.

3 Application to variance regression

In this section we consider the basic model given by:

$$\begin{aligned} E[Y_i] &= X_i\beta, \\ \log \mathrm{var}[Y_i] &= 2Z_i\theta, \end{aligned} \tag{1}$$

for $i = 1, \ldots, n$, where

$$X_i = (1, X_{i1}, \ldots, X_{iJ}), \qquad \beta = (\beta_0, \beta_1, \ldots, \beta_J),$$

$$Z_i = (1, Z_{i1}, \ldots, Z_{iK}), \qquad \theta = (\theta_0, \theta_1, \ldots, \theta_K)$$

are covariate and regression vectors, respectively. Many methods have been proposed in the literature for parameter estimation. See, for example, Carroll and Ruppert (1988).

Here we demonstrate the advantage of using the scale mixture of uniform family in the analysis of a variance regression model. For convenience, we reparameterise $\lambda_k = e^{-\theta_k}$. A specific model which results in the specifications (1) is given by

$$Y_i|[V_i = v_i] \sim \mathrm{Un}\left(X_i\beta - \frac{\sqrt{v_i}}{\prod_k \lambda_k^{Z_{ik}}}, X_i\beta + \frac{\sqrt{v_i}}{\prod_k \lambda_k^{Z_{ik}}} \right),$$

and

$$V_i \sim_{\mathrm{iid}} f_V(\cdot).$$

The condition for the variance is satisfied provided we constrain $\mathrm{E}\, V = 3$. For example, for normal errors we can let $V \sim \mathrm{Ga}(3/2, 1/2)$. For alternative levels of kurtosis we can take $V \sim \mathrm{Ga}(3\alpha/2, \alpha/2)$, for $\alpha > 0$. Here $\alpha > 1$ leads to tails heavier than normal and $\alpha < 1$ leads to lighter tails than normal.

In a Bayesian context, this model is straightforward to study via a Gibbs sampler. This follows since all the full conditional densities required to implement the Gibbs sampler are of standard type. Let $\pi(\cdot)$ represent the prior for (β, λ) which comprises the product $\prod_{l=0}^{J} \pi(\beta_l) \prod_{l=0}^{K} \pi(\lambda_l)$. Consequently, the full conditional densities are as follows:

- $$\pi(v_i|\cdots) \propto \exp(-v_i/2) I \left\{ v_i > (Y_i - X_i\beta)^2 \prod_k \lambda_k^{2Z_{ik}} \right\}$$

- $$\pi(\beta_l|\cdots) \propto \pi(\beta_l) I\left\{\beta_l \in \left(\max_j\{A_j\}, \min_j\{B_j\}\right)\right\},$$

  where

  $$A_j = \min_{i:X_{ij}\neq 0}\left\{\frac{Y_i - \sqrt{v_i}/\prod_k \lambda_k^{Z_{ik}} - \gamma_{ij}}{X_{ij}}, \frac{Y_i + \sqrt{v_i}/\prod_k \lambda_k^{Z_{ik}} + \gamma_{ij}}{X_{ij}}\right\},$$

  $$B_j = \max_{i:X_{ij}\neq 0}\left\{\frac{Y_i - \sqrt{v_i}/\prod_k \lambda_k^{Z_{ik}} - \gamma_{ij}}{X_{ij}}, \frac{Y_i + \sqrt{v_i}/\prod_k \lambda_k^{Z_{ik}} + \gamma_{ij}}{X_{ij}}\right\},$$

  and $\gamma_{ij} = \sum_{l\neq j} X_{il}\beta_l$.

- $$\pi(\lambda_k|\cdots) \propto \pi(\lambda_k)\lambda_k^{\sum_i Z_{ik}} I\left\{-\log\lambda_k \in \left(\max_{i:Z_{ik}<0}\{E_{ik}\}, \min_{i:Z_{ik}>0}\{E_{ik}\}\right)\right\},$$

  where

  $$E_{ik} = \left(\frac{\sqrt{v_i}}{|Y_i - X_i\beta|\prod_{l\neq k}\lambda_l^{Z_{il}}}\right)^{1/Z_{ik}},$$

  If $Z_{ik} > 0$ for all i then $\max_{i:Z_{ik}<0}\{E_{ik}\} = \infty$. If $Z_{ik} < 0$ for all i then $\min_{i:Z_{ik}<0}\{E_{ik}\} = -\infty$.

Hence, all the full conditional densities are sampled using standard techniques; see, for example, Devroye (1986), Damien and Walker (2000). Next, we consider some examples.

3.1 Example 1

A simulated dataset contains 50 cases, generated as follows :

$$y_i = \beta_0 + \beta_1 x_i + \sigma_i\epsilon_i,$$

$$\log\sigma_i = \theta_0 + \theta_1 z_i, \quad i = 1, \ldots, 50,$$

with $\epsilon_i \sim N(0,1)$ independently. The regression parameters were taken to be $\beta_0 = 10$, $\beta_1 = 5$, $\theta_0 = -3$ and $\theta_1 = 1$. The 'standard' ordinary least squares estimates for β_0 and β_1 are 6.75 and 5.46, respectively.

Table 1: Posterior summaries of model parameters β_0 β_1, θ_0 and θ_1 for the simulated dataset.

	β_0	β_1	θ_0	θ_1
mean	10.149	4.960	-2.384	0.890
std. dev	0.095	0.019	0.151	0.037

In this example, and the others to follow, we used non-informative prior distributions for the regression parameters; that is, we took

$$\beta_j \sim N(0, \sigma_j^2),$$

$$\lambda_k \sim \mathrm{Ga}(a_k, a_k),$$

with large σ_j^2 and with a_k very small. Using the scale mixture of uniform model, the resulting posterior estimates, obtained using the Gibbs sampler, are summarized in Table 1. The true values appear to be well approximated by the sample based statistics.

3.2 Example 2

The dataset used in this example is taken from the original epitaxial layer growth experiment of Kackar and Shoemaker (1986), as reported in Shoemaker, Tsui and Wu (1991). One of the initial steps in fabricating integrated circuit (IC) devices is to grow an epitaxial layer on polished silicon wafers. The experimenters needed to find process factors that can be set to minimize the epitaxial layer nonuniformity while maintaining average thickness as closely as possible to nominal. Here we consider a simplified version of this experiment, four experimental factors, susceptor-rotation method, nozzle position, deposition temperature and deposition time (labeled A, B, C and D) are to be investigated at the two levels, $-$ and $+$.

A four factor, full factorial design of 16 runs with 6 replications was adopted. We are interested in both the location and dispersion effects. Traditionally, these two analyses will be performed separately, via linear regression. Since the assumption of equal variance is violated, such a dichotomous

Table 2: Posterior summaries of model parameters β_0 β_1, θ_0 and θ_1 for the layer growth experiment example.

	β_0	β_1	θ_0	θ_1
mean	14.415	0.430	-1.389	0.616
std. dev	0.014	0.013	0.047	0.060

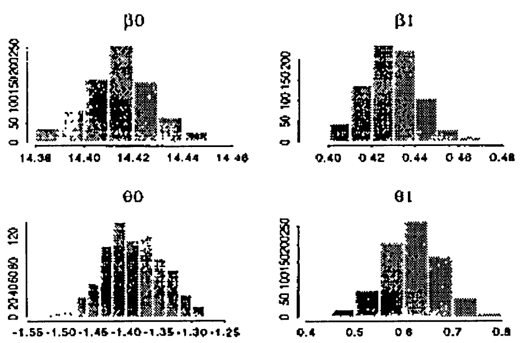

Figure 1: Posterior distributions of model parameters β_0 β_1, θ_0 and θ_1 for the layer growth experiment.

approach may be inappropriate. Using screening techniques, such as half normal plots, one can easily identify that factor D is the most influential factor for location effect and factor A is the most influential one for dispersion effect.

Using the scale mixture models described earlier in this paper, the location and dispersion effects are modelled simultaneously. The posterior statistics are summarized in Table 2, and the histograms appear in Figure 1. The sample based summaries provide a full analysis for the variance regression model.

Table 3: Posterior summaries of model parameters β_0 β_1, β_2, θ_0, θ_1 and θ_2 for the tensile strength experiment.

	β_0	β_1	β_2	θ_0	θ_1	θ_2
mean	42.903	0.994	1.473	-1.128	-0.108	0.786
std. dev	0.204	0.066	0.176	0.224	0.215	0.161

3.3 Example 3

In Example 2, it is easy to estimate variance under each setting, since there are replications. However, in the absence of replication, it is difficult to obtain estimates for the dispersion effects. We consider such an illustration.

Box and Meyer (1986) presented an interesting analysis of dispersion effects in a fractional-factorial experiment. The experiment concerns the tensile strength of welds in an off-line welding experiment performed by the National Railway Corporation of Japan (Taguchi and Wu, 1980). This experiment was also studied by Carroll and Ruppert (1988) for dispersion effects. Box and Meyer found that the mean could be adequately explained by two factors, B and C. Here we also want to find out the effect of these two factors on the dispersion. Using the model described earlier, a full Bayesian analysis of the dispersion effects is obtained. The posterior statistics are summarized in Table 3, and the histograms appear in Figure 2.

4 A Comparison of SMN and SMU models

In this section, we will demonstrate the differences between the two scale mixture methodologies: Scale mixture of normals and scale mixture of uniforms.

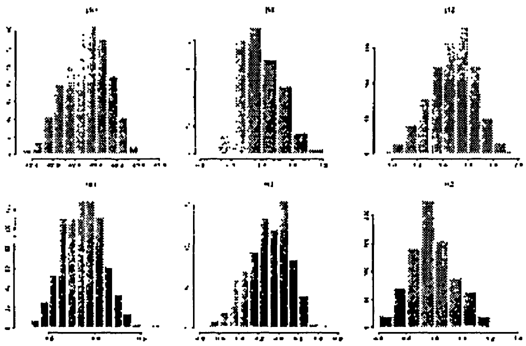

Figure 2: Posterior distributions of model parameters β_0 β_1, β_2, θ_0, θ_1 and θ_2 for the tensile strength experiment.

4.1 Regression model with Cauchy errors

We revisit the linear regression model under Cauchy errors with location and scale parameters 0 and σ, given by

$$y_i = \alpha + \beta x_i + \epsilon_i, \qquad i = 1, \dots, n.$$

Scale mixture of normals
The scale mixture of normals representation is then given by:

$$y_i | \xi_i \sim N(\alpha + \beta x_i, \sigma^2 / \xi_i),$$

and

$$\xi_i \sim \text{Ga}(1/2, 1/2).$$

Assuming the non-informative prior,

$$\pi(\alpha, \beta, \lambda) \propto \lambda^{1/2},$$

the full conditionals are given by:

- For ξ_i's:

$$\xi_j | \cdots \sim \text{Ex}\left(\frac{1 + \lambda_j (y_j - \alpha - \beta x_j)^2}{2}\right),$$

- For α:

$$\alpha|\cdots \sim \mathrm{N}\left(\frac{\sum_{i=1}^n \xi_i(y_i-\beta x_i)}{\sum_{i=1}^n \xi_i}, \frac{1}{\lambda\sum_{i=1}^n \xi_i}\right),$$

- For β:

$$\beta|\cdots \sim \mathrm{N}\left(\frac{\sum_{i=1}^n \xi_i x_i(y_i-\alpha)}{\sum_{i=1}^n \xi_i x_i^2}, \frac{1}{\lambda\sum_{i=1}^n \xi_i x_i^2}\right),$$

- For λ:

$$\lambda|\cdots \sim \mathrm{Ga}\left(\frac{n+3}{2}, \frac{1}{2}\sum_{i=1}^n \xi_i(y_i-\alpha-\beta x_i)^2\right).$$

Based on the development in Section 2, after some algebra, we obtain the following posterior conditional distributions for the parameters of the model using the SMU representation.

Without loss of generality, in canonical notation, $\pi(\alpha)$ and $\pi(\beta)$ denote prior distributions for α, β respectively. We assume, without loss of generality,

$$\pi(\lambda) \sim Ga(\alpha_\lambda, \beta_\lambda);$$

i.e., the prior for λ is a gamma density.

It is straightforward to derive the posterior conditional distributions for the parameters, and the auxiliary variable. These are given by:

- For v_i:

$$v_i|\cdots \propto \frac{1}{(1+v_i)^2} I[\lambda(y_i-\alpha-\beta x_i)^2, \infty),$$

- For α:

$$\alpha|\cdots \propto \pi(\alpha)\prod_i^n I\left[y_i-\beta x_i-\sqrt{\frac{v_i}{\lambda}}, y_i-\beta x_i+\sqrt{\frac{v_i}{\lambda}}\right],$$

$$= \pi(\alpha) I\left[\max_i\left\{y_i-\beta x_i-\sqrt{\frac{v_i}{\lambda}}\right\}, \min_i\left\{y_i-\beta x_i+\sqrt{\frac{v_i}{\lambda}}\right\}\right];$$

- For β:

$$\beta|\cdots \propto \pi(\beta)\prod_i^n I\left[\frac{y_i-\alpha-\sqrt{\frac{v_i}{\lambda}}}{x_i}, \frac{y_i-\alpha+\sqrt{\frac{v_i}{\lambda}}}{x_i}\right],$$

$$= \pi(\beta) I \left[\max_i \left\{ \frac{y_i - \alpha - \sqrt{\frac{v_i}{\lambda}}}{x_i} \right\}, \min_i \left\{ \frac{y_i - \alpha + \sqrt{\frac{v_i}{\lambda}}}{x_i} \right\} \right],$$

- For λ:

$$\lambda | \cdots \propto \pi(\lambda) \prod_i P(y_i, v_i, \mu | \lambda)$$

$$\propto Gamma(\alpha_\lambda + n/2, \beta_\lambda) I \left(0, \min_i \left\{ \frac{v_i}{(y_i - \alpha - \beta x_i)^2} \right\} \right].$$

  a truncated gamma distribution. Here $\pi(\alpha)$ and $\pi(\beta)$, in canonical notation, are prior distributions for α and β.

Setting α and β to be 10 and 3, respectively, and σ, the scale parameter in the Cauchy distribution, to be 3, 100 pairs of (x_i, y_i) were generated. The least squares estimates for $\alpha = 6.6772$ and $\beta = 6.5973$. Using the Gibbs sampler detailed above, posterior summaries under the SMN and SMU models are provided below.

Table 4: Comparison of SMN and SMU on the accuracy of estimated parameters for Cauchy distribution

para	True Value	SMN		SMU	
		mean	std dev	mean	std dev
α	10	10.100	0.629	9.770	0.598
β	3	3.119	0.374	3.393	0.282
σ	3	2.948	0.380	2.914	0.332

As expected, in both examples, the two representations yield comparable results. This is partly because a non-informative prior was employed. The effect of this prior is that the resulting full conditionals in the Gibbs sampler under the SMN framework reduces to standard forms, sampling from which is straightforward. However, in general, this is not true; take, for example, in the Cauchy illustration, a normal prior for the regression parameters. On the other hand, no matter what prior is used, under SMU, the resulting posterior full conditionals take the form of a truncated version of the prior. It was this in mind a canonical notation for the prior distributions was used in the earlier development.

The SMU representation, typically, converges more slowly than SMN because of the nature of the resulting Markov chain. This, of course, is hardly problematic in an era where fast computers and sophisticated software is a norm. It then seems appropriate to consider models that *result* in algorithms that are trivial to code – an ubiquitous feature of the SMU enterprise. Also, at least as important, the SMU *includes* the SMN as a special case, thus making it a more flexible and practically appealing family of models.

5 Multivariate distributions

In this section, we briefly describe how the SMN concept can be extended to multivariate distributions. In one dimension, we generated a scale mixture of uniform density using intervals with random lengths. In two or higher dimensions, intervals are replaced by ellipses, ellipsoids or hyper-ellipsoids, and lengths replaced by radii. Hence, multivariate scale mixtures of uniform distributions can be generated via uniform distributions on ellipses, ellipsoids or hyper-ellipsoids, with random radii.

Let a denote a d-dimensional vector $(a_1, a_2, \ldots, a_d)$, A denotes a $d \times d$ positive definite matrix and $r > 0$. Let $E_d(a, A, r)$ denote the d dimensional ellipse, ellipsoid or hyper-ellipsoid which satisfies

$$(x - a)'A(x - a) \leq r^2.$$

Definition 5.1 Multivariate scale mixture of uniforms
Suppose for a vector $Y = (Y_1, Y_2, \ldots, Y_m)$, we have

$$Y|[V = v] \sim Un\{E_m(\mu, \Sigma, v)\},$$

and

$$V \sim f_V(.),$$

where Σ is positive definite and V is positive, then Y has a multivariate scale mixture of uniform density, with $f_V(\cdot)$ being the generating density.

For example, the density function of a multivariate normal distribution with mean μ and covariance matrix Σ is given by

$$f(x) = (2\pi)^{-\frac{d}{2}}|\Sigma|^{-\frac{1}{2}}e^{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)} \qquad (x \in \mathrm{R}^d).$$

The multivariate scale mixture of uniform representation is given by: $(Y_1, Y_2)|[V = v]$ is uniformly distributed in $E_2(\mu, \Sigma^{-1}, \sqrt{v})$ and

$$v \sim \mathrm{Ga}(2, 1/2).$$

Generalising, we have for a m-dimensional vector $Y = (Y_1, \ldots, Y_m)$ if $Y|[V = v]$ is uniformly distributed in $E_m(\mu, \Sigma^{-1}, \sqrt{v})$ and

$$v \sim \mathrm{Ga}(1 + m/2, 1/2),$$

then Y will have an m dimensional multivariate normal distribution with mean μ and covariance matrix Σ.

This result is similar to a result of Johnson and Ramberg (1977). It is clear that this result can be extended to a very general "symmetric" multivariate distribution family: by symmetric, we mean the density function f satisfies

$$f(x_1, x_2, \ldots, x_n) = f(-x_1, -x_2, \ldots, -x_n).$$

Tong (1990) describes Y to have an elliptically contoured distribution if its density function

$$f_{\mu,\Sigma}(x) = |\Sigma|^{-1/2}g\{(x-\mu)'\Sigma^{-1}(x-\mu)\}, \qquad x \in \mathbf{R}^m,$$

where $g : \mathbf{R} \to [0, \infty)$ is nonincreasing. It is clear that this family is included in our development above. Similarly, "elliptically symmetric distributions", Kelker (1970), is also a special case under our definition.

The density function of n dimensional multivariate t distribution is as follows:

$$f_t(x, m, \mu, \Sigma) = \frac{\Gamma(\frac{m+t}{2})}{(t\pi)^{m/2}\Gamma(\frac{t}{2})|\Sigma|^{1/2}}\left\{1 + \frac{1}{t}(x-\mu)'\Sigma^{-1}(x-\mu)\right\}^{-(m+t)/2},$$

where $x, \mu \in \mathbf{R}^m$, Σ is a $m \times m$ positive definite matrix and t is the degrees of freedom.

The scale mixture representation for the multivariate t distribution is similar to the one for the multivariate normal except the distribution of v is given by:

$$f_V(v) \propto \frac{v}{(1+\frac{1}{t}v)^{-(m+t)/2}}.$$

The multivariate Cauchy is a special case of the multivariate t, having one degree of freedom.

6 Discussion

In this paper, we have developed a new class of models, namely, a scale mixture of uniform distributions, that enables analysis of data when normality assumptions are violated. The scale mixture representation provides a general and flexible approach to modelling; this was illustrated via examples in the context of variance regressions. An attractive feature of the approach is that the full conditional distributions in the resulting Gibbs sampler are all uniform; this makes coding a trivial task. Introducing auxiliary variables in the constructions described in this paper could likely lead to autocorrelation in the Markov chain. In a related work, Damien et al. (1999) provide a comparative study of the efficiency rates of Gibbs samplers constructed via the introduction of auxiliary variables.

References

Andrews, D.F. and Mallows, C.L. (1974). Scale mixtures of normal distributions, *Journal of the Royal Statistical Society, Series B* **36**, 99-102.

Box, G.E.P. and Meyer, G.D. (1986) Dispersion effects from Fractional Designs, *Technometrics.* **28**, 19-27.

Carroll, R.J. and Ruppert D. (1988), *Transformation and Weighting in Regression,* New York, London, Chapman and Hall.

Chew, V. (1968) Some useful alternatives to the normal distribution, *The American Statistician*, **22**, 22-24.

Damien, P., Wakefield, J.C. and Walker, S.G. (1999). Gibbs sampling for Bayesian non-conjugate and hierarchical models by using auxiliary variables, *Journal of the Royal Statistical Society, B.* **61**, 331-344.

Damien, P. and Walker, S.G. (2000). Sampling truncated normal, gamma and beta densities. *Journal of Graphical and Computational Statistics.* To appear.

Devroye, L. (1986). *Non-uniform Random Variate Generation.* Springer-Verlag, New York.

Gelfand, A.E. and Smith, A.F.M. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association,* **85**, 398-409.

Johnson, M.E. and Ramberg, J.S. (1977) Elliptically symmetric distributions: characterizations and random variate generation, *Proceedings of the American statistical Association, Statistical Computing Section,* 262-265.

Kackar, R.N. and Shoemaker, A.C. (1986) Robust design, a cost effective method for improving manufacturing process, AT&T Technical Journal, **65**, 39-50.

Kelker, D. (1970). Distribution theory of Spherical distributions and a location-scale parameter generalisation, *Sankyâ, A* **32**, 419-430.

Polson, N.G. (1996) Convergence of Markov chain Monte Carlo algorithms. In *Bayesian Statistics 5* (eds J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. smith). pp. 297-321, Oxford University Press, Oxford.

Shoemaker, A. C., Tsui, K. L., and Wu, C. F. J. (1991), Economical experimentation methods for robust design, *Technometrics,* **33**, 415-427.

Smith, A.F.M. and Roberts, G.O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods (with discussion), *Journal of the Royal Statistical Society, B* **55**, 3-23.

Taguchi, G. and Wu, Y. (1980), Introduction to off-line quality control, Nagoya, Japan: Central Japan quality control association.

Tong, Y.L. (1990) *The multivariate Normal Distribution*, New York: Springer-Verlag.