Scale Mixture Models with Applications to Bayesian Inference

Zhaohui S. Qin[*], Paul Damien[†] and Stephen Walker[**]

[*]*Department of Statistics, Harvard University*
[†]*University of Michigan Business School*
[**]*University of Bath*

Abstract. Scale mixtures of uniform distributions are used to model non-normal data in time series and econometrics in a Bayesian framework. Heteroscedastic and skewed data models are also tackled using scale mixture of uniform distributions.

This paper uses scale mixture methodology to develop models that result in a full Bayesian analysis of non-normal data. The benefits are two-fold: it provides a flexible approach to model skewed and heavy-tailed data, as well as situations where heteroscedasticity or autocorrelation is present; the natural hierarchical structure of the scale mixture representation results in a very simple-to-implement Markov chain Monte Carlo (MCMC) scheme that provides a full Bayesian analysis.

The scale mixture idea appeared as early as the 1970's. The scale mixture of normal family was introduced in [1]. It was first used to sample symmetric distributions that have a normal component; see [2] for recent representations of scale mixtures of normal distributions.

Scale Mixture of Normals Given latent variable λ, a random variable X with location and scale parameters θ and σ is conditionally distributed as normal $X|\lambda \sim N(\theta, \alpha(\lambda)\sigma^2)$, where $\alpha(.)$ is a positive function on $\Re$ and $\lambda \sim \pi(\lambda)$; $\pi(.)$ is a probability density function either discrete or continuous. The distribution of X is referred to as a scale mixture of normals (SMN), with mixing parameter λ and scale mixing density $\pi(.)$.

The class of models defined above is very large and useful. Yet, such representations of normal distributions can be restrictive in many situations where the Gaussian assumption may not be the appropriate choice: for example, contexts such as analysis of financial data may clearly require assuming a non-normal likelihood. Thus distributions generated above are very much close to normals, differing only at the first two moments. To overcome this difficulty, and mimicking somewhat the idea in SMN, we have the following:

Scale Mixture of Uniforms Suppose that a random variable X has unimodal and skewed distribution with mode μ, then it may be written using the following representation: $X|V = v \sim U(\mu - \sigma_1 g(v), \mu + \sigma_2 g(v))$, $\sigma_1 > 0$, $\sigma_2 > 0$ and $\sigma_1 \neq \sigma_2$; $g(.)$ is positive and invertible; V has distribution defined on $\Re$ with density $f_V(.)$. The distribution of X is termed as a scale mixture of uniforms (SMU).

Here are some facts which can be easily established.

If $X|V = v \sim U(\mu - \sigma\sqrt{v}, \mu + \sigma\sqrt{v})$ and $V \sim Gamma(3/2, 1/2)$ then $X \sim N(\mu, \sigma^2)$.

If f_X is a unimodal, symmetric density about 0 and $f'_X(x)$ exists for all x, then the following hold: $f_X(x) = -\int_{v>g^{-1}(x)} f'_X(g(v))g'(v)dv$, then $f_V(v) = -f'_X(g(v))g(v)g'(v)$.

From the above development we have the following:

Remark 1. The class of unimodal, symmetric distributions can be written as scale mixtures of uniform distributions.

Remark 2. All densities having the quadratic form $f((\frac{x-\mu}{\sigma})^2)$ are covered under the SMU Definition. Many well known distribution belongs to this category; as examples, normal, student t, Cauchy. Also, non-quadratic form models are encapsulated in the scale mixture of uniforms family.

Example: Heteroscedastic regressions The assumption of constant variance is usually inappropriate in many business applications. Traditionally, there are two approaches dealing with non-constant variance: weighted least squares is appropriate when the form of the non-constant variance is either known exactly or there is some known parametric form. Alternatively, one can transform y to $h(y)$ where $h()$ is chosen so that var $h(y)$ is constant. Here we discuss how to model both the mean and variance simultaneously under our scale mixture of uniform framework. The basic model is given by:

$$E[X_i] = Z_i\beta, \qquad \log\text{var}[X_i] = 2W_i\theta, \qquad i = 1, \ldots, n$$

$$Z_i = (1, Z_{i1}, \ldots, Z_{iJ}), \qquad \beta = (\beta_0, \beta_1, \ldots, \beta_J),$$
$$W_i = (1, W_{i1}, \ldots, W_{iK}), \qquad \theta = (\theta_0, \theta_1, \ldots, \theta_K)$$

If for some k, $\sum_{i=1}^{n} W_{ik} < 0$, we use $-W_{ik}$ and $-\theta$ to replace W_{ik} and θ. Another change is that we use $\lambda_k = e^{-\theta_k}$. λ is often referred to as "precision", and it is conventional to let it have an inverse gamma distribution. The SMU representation is the follows:

$$X_i|[V_i = v_i] \sim U\left(Z_i\beta - \frac{\sqrt{v_i}}{\prod \lambda_k^{W_k}}, Z_i\beta + \frac{\sqrt{v_i}}{\prod \lambda_k^{W_k}}\right),$$

and $V_i \sim_{\text{iid}} f_V(.)$. The condition for the above representation is that we constrain $E[V] = 3$. It's easy to find such V, for instance, let $V \sim Gamma(3/2, 1/2)$.

For the implementation of a Gibbs sampler we need all the full conditional distributions. There will be a parameter associated with f_V but the full conditional for this parameter will be based on the V_i being iid from f_V and so should not pose any problem. All the full conditionals for v_i, β_j and λ_k are truncated version of some standard distributions which can be sampled using well established algorithms.

REFERENCES

1. Andrews, D. F., and Mallows, C. L., *Journal of the Royal Statistical Society, Seires B*, **36**, 99–102 (1974).
2. Karim, A. M., and Paruolo, P., *Econometrica*, **65**, 671–680 (1996).