Published in final edited form as:

Stat Appl Genet Mol Biol. 2015 August ; 14(4): 317–332. doi:10.1515/sagmb-2014-0063.

Exact Likelihood-free Markov Chain Monte Carlo for Elliptically Contoured Distributions

Patrick Muchmore and

Division of Biostatistics, Department of Preventive Medicine, University of Southern California,
2001 N Soto Street Los Angeles, CA 90089-9237, USA

Paul Marjoram

Division of Biostatistics, Department of Preventive Medicine, University of Southern California,
2001 N Soto Street Los Angeles, CA 90089-9237, USA

Patrick Muchmore: muchmore@usc.edu

Abstract

Recent results in Markov chain Monte Carlo (MCMC) show that a chain based on an unbiased estimator of the likelihood can have a stationary distribution identical to that of a chain based on exact likelihood calculations. In this paper we develop such an estimator for elliptically contoured distributions, a large family of distributions that includes and generalizes the multivariate normal. We then show how this estimator, combined with pseudorandom realizations of an elliptically contoured distribution, can be used to run MCMC in a way that replicates the stationary distribution of a likelihood based chain, but does not require explicit likelihood calculations. Because many elliptically contoured distributions do not have closed form densities, our simulation based approach enables exact MCMC based inference in a range of cases where previously it was impossible.

1 Introduction

In this paper we show how simulations can be used in lieu of likelihood calculations to derive a Markov chain Monte Carlo (MCMC) algorithm which enables exact posterior sampling of the location and scale parameters for almost the entire class of elliptically contoured distributions (ECDs). This class includes the multivariate normal, but it also includes a number of distributions with no closed form density function. Members of this class are often employed as models of unexplained residual deviation, or “error”. The practical application of these distributions is typically restricted to either the normal or the elliptically contoured Student’s t , primarily for reasons of analytic convenience. Our method allows tractable analysis to be performed using a much wider range of elliptically contoured distributions, without requiring knowledge about, or even the existence of, a closed form density function.

We assume a Bayesian analysis paradigm, which can be succinctly summarized by the proportionality

$$f(\theta|x) \propto f(x|\theta)f(\theta), \quad (1.1)$$

relating the posterior $f(\theta|x)$ to the product of the likelihood $f(x|\theta)$ and the prior $f(\theta)$. A number of problems may arise when using equation (1.1). Analytically, it may be not possible to find a closed form expression for the posterior density. Conjugate distributions may simplify analytic calculations, but insisting on conjugacy severely restricts the choice of distribution. Even when the posterior is known analytically, a major computational difficulty is frequently encountered when θ has more than a few dimensions. In this case simply evaluating the posterior at a representative sample of points can require a prohibitive number of function evaluations (the so-called “curse of dimensionality”). Markov chain Monte Carlo based inference techniques can potentially address both the analytical and computational difficulties listed. Although MCMC may eliminate the need to find an analytic closed form posterior density, explicitly evaluating the likelihood is still required. If the likelihood function is unknown or intractable then most MCMC methods are inapplicable. This limitation is not unique to MCMC, as an unknown or intractable likelihood will preclude most likelihood based inference methods, almost by definition.

In an attempt to get around the problem of an intractable likelihood Marjoram, Molitor, Plagnol, and Tavaré (2003) introduced “Markov chain Monte Carlo without likelihoods”, which was an early MCMC based example of a larger family of methods collectively known as approximate Bayesian computation (ABC). ABC methods come in many forms, but broadly speaking they are applicable to distributions from which one can easily generate pseudorandom realizations despite the unknown or intractable likelihood. Marjoram et al. (2003) showed that accepting proposals when the observed and simulated data are identical enables exact posterior inference. Frequently in practice, however, the probability of an exact match is effectively, if not actually, zero.

Because exact matching tends to be infeasible, two simplifications are typically employed. The first is to reduce the dimension of the data using summary statistics, and the second is to accept proposals when the simulated data is sufficiently “close” to the observed data. In Marjoram et al. (2003) this is implemented by choosing a fixed tolerance ε and accepting proposals whenever the Euclidean distance between simulated and observed summary statistics is smaller than ε ; however, calibrating ε in this algorithm can be very difficult. Around the same time, another ABC algorithm, colloquially referred to as ABC regression, was introduced by Beaumont, Zhang, and Balding (2002). Their method, which is not based on MCMC, has proven popular at least in part because there is no need to calibrate the tolerance ε . Rather, ABC regression more heavily weights points that produce simulated data closer to the observed. Often, a very large percentage of the points are assigned a weight of zero, which is analogous to rejecting points based on a quantile cutoff for the simulated distances. The points with non-zero weights are modeled as a function of the corresponding distances. In Beaumont et al. (2002) this model takes the form of a local linear regression based on an Epanechnikov kernel.

In addition to regression and MCMC based methods, ABC algorithms based on sequential Monte Carlo (SMC) techniques have also been developed, beginning with Sisson, Fan, and

Tanaka (2007). An advantage of SMC methods is they allow for trivial parallelization, which can facilitate speedups of 1–2 orders of magnitude.

The SMC algorithm of Sisson et al. (2007) permits the user to pre-specify a decreasing sequence of tolerances ε_i . More recently, techniques for automatically estimating a sequence of tolerances have been developed by Del Moral, Doucet, and Jasra (2012) and Drovandi and Pettitt (2011). SMC based approaches have also employed adaptive perturbation kernels to improve efficiency. Early efforts, such as in Beaumont, Cornuet, Marin, and Robert (2009), employed independent normal perturbation kernels for each parameter component. With each new population, the variance of the kernel is adjusted based on the distribution of the previous populations. Filippi, Barnes, Cornebise, and Stumpf (2013) extended this approach to multivariate normal perturbation kernels, which can account for correlated parameters. Filippi et al. (2013) also consider both globally and locally adaptive kernels, and they derive optimality criteria for balancing the similarity of the kernel to the target with the ABC acceptance rate.

Despite these significant advances in ABC methodology, fundamental questions about the effects of choosing $\varepsilon = 0$, and the use of summary statistics in place of the true data, remain unanswered, see e.g. Robert, Cornuet, Marin, and Pillai (2011). In this paper, we show how pseudorandom realizations from a large family of distributions can be used to calculate an unbiased estimate of the likelihood. When this estimator is combined with recent results in MCMC, it yields a chain that has the same stationary distribution as one based on exact density calculations. Therefore, for these distributions it is possible to perform exact posterior sampling without evaluating the density function. The family of distributions our method is applicable to consists of almost the entire class ECDs. As many distributions in this class do not have closed form densities, our simulation based approach effectively bypasses a major impediment to likelihood based MCMC for this family. Moreover, the density estimator we describe is the same for all distributions in this class, so our algorithm can be applied to many different distributions with minimal modification.

We conclude the paper with two examples. The first is a simple model consisting of correlated bivariate observations. With this model we illustrate three cases, corresponding to three different distributional assumptions. The second example is a nonlinear regression model of the interaction between insulin and glucose in the human metabolism. Using real data, we show that for this model our approach leads to inferences that are almost indistinguishable from the corresponding likelihood based method, while ABC regression leads to posterior inferences that are clearly different from the other two methods. We then introduce a generalized error distribution which can be easily simulated, but which is not amenable to analytic calculations. We also show how this generalized error distribution can be used to test the assumption of normality, and we conclude that, for our data, assuming independent normal errors may not be appropriate.

2 Elliptically Contoured Distributions and Normal Scale Mixtures

Historically, the multivariate normal distribution has played a central role in the statistical modeling of a wide range of phenomena. However, in many applications there is substantial

empirical evidence strongly suggesting one or more observations will exhibit large deviations from their average values with a far higher probability than a normal model would predict, e.g. Newman (2005) and Clauset, Shalizi, and Newman (2009). The definition of ECDs does not imply the existence of moments of any order, so there are members of the family with substantially more mass concentrated in the tails compared to the normal distribution. Thus, for data exhibiting this type of extreme behavior, ECDs with “heavy tails” offer a plausible alternative model.

The existence of a closed form density for an ECD is the exception, not the rule, and our method enables exact posterior sampling for almost the entire family. The crux of our approach is a collection of invariance results which show that, under some conditions, exact results derived for any member of the elliptically contoured family remain valid for every member of the family. For example, the variance ratio underlying the F -test has a sampling distribution which is known exactly for the normal distribution. Kelker (1970) showed that the exact sampling distribution derived in the normal case is also the exact sampling distribution for more general elliptically contoured distributions.

In Kelker (1970), a p -dimensional random vector X is a member of the class of elliptically contoured distributions $EC_p(\mu, \Sigma, \varphi)$ if and only if the characteristic function of X is of the form $\varphi_X(t) = E(\exp(it^\top X)) = \exp(it^\top \mu) \psi(t^\top \Sigma t)$, where $\psi: [0, \infty) \rightarrow \mathbb{R}$. An equivalent, and more intuitive, definition appeared in Cambanis, Huang, and Simons (1981), which we state under the assumption that Σ is of full rank. A p -dimensional vector X is elliptically contoured if and only if

$$X \stackrel{d}{=} \mu + R \sum^{\frac{1}{2}} U^{(p)}, \quad (2.1)$$

where $U^{(p)}$ is the uniform distribution on the unit sphere in $\mathbb{R}^p$, R is an independent non-negative univariate random variable, and $\mu \in \mathbb{R}^p$ and $\sum^{\frac{1}{2}} \in \mathbb{R}^{p \times p} | \sum^{\frac{1}{2}} (\sum^{\frac{1}{2}})^\top = \Sigma$ are constant. If the mean exists it is equal to μ , and if the variance matrix exists it is proportional to Σ . The notation $\stackrel{d}{=}$ denotes equality in distribution; the derivation of some algebraic properties of this operator can be found in Anderson and Fang (1982).

A particularly useful subset of the class of ECDs is the class of normal scale mixtures. For our purposes, a random variable X will be called a normal scale mixture if it is defined by

$$X \stackrel{d}{=} \mu + WY \quad (2.2)$$

where $\mu \in \mathbb{R}^p$ is constant, Y is a p -dimensional multivariate normal with mean zero and full rank variance matrix Σ , and W is an independent non-negative univariate random variable satisfying $P(W = 0) = 0$. The model (2.2) encompasses a wide range of limiting behaviors determined by the distribution of W . As shown by Fang and Zhang (1990), specific choices of W will result in X having a distribution that is a multivariate generalization of symmetric univariate examples such as Student's t , Laplace (double exponential), or the α -stable distributions. In the case $W \equiv 1$ the model reduces to a multivariate normal. Gupta, Varga,

and Bodnar (2013, corollary 2.6) give a proof of the fact that the class of ECDs contains the class of normal scale mixtures.

Producing a simulated realization of a normal scale mixture in the general case requires just slightly more computation than the normal case. One simply generates a single realization of W and multiplies that by the independently simulated multivariate normal Y . Despite the ease with which they can be simulated, more often than not the density function of a normal scale mixture cannot be expressed in closed form. This has frequently limited the practical application of these models to cases where a closed form density exists, such as the multivariate Student's t , and precluded the application of other cases, such as the multivariate Laplace and α -stable distributions, which in general do not have closed form density functions.

One of the major reasons ECDs have attracted significant attention is because there are a number of statistics, $H(X)$, whose sampling distribution is invariant for large subsets of $X \in EC_p$. Thus, if the analytical form of the statistic can be derived for any X in the subset, it is immediately known for every X in the subset. In the vast majority of examples this means the analytical form is known in the case X is multivariate normal and then extended to otherwise intractable distributions in EC_p . There are numerous ways to establish the invariance of a statistic $H(X)$. A convenient method is given by Fang and Zhang (1990, corollary 2 of theorem 5.1.1) which states any statistic $H(X)$ with the property that

$$H(X) \stackrel{d}{=} H\left(\left(XX^\top\right)^{-\frac{1}{2}}X\right) \quad (2.3)$$

will be invariant for the subset of ECDs that also satisfy the condition $P(|XX^\top| = 0) = 0$. This condition will be satisfied when Σ is full rank, $P(R = 0) = 0$ (or in the normal scale mixture notation $P(W = 0) = 0$), and $n > p$.

3 Univariate Normal Case

Our basic approach is to search for unbiased estimators of density functions, and then to use these estimators to calculate unbiased density estimates based on simulated samples.

Ultimately, we will show how to calculate such an unbiased estimate for almost every elliptically contoured distribution. However, to motivate the method in the general case we will first outline the derivation in the case of data x consisting of a single observation from a univariate normal random variable $X \sim N(\mu, \sigma^2)$. The density function of X is

$$f(x|\mu, \sigma^2) = c(\sigma^2)^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \frac{(x - \mu)^2}{\sigma^2}\right) \quad (3.1)$$

where $c = (2\pi)^{-\frac{1}{2}}$. Given an observed value of x , equation (3.1) is also the likelihood for the parameters (μ, σ^2) . The tractability of (3.1) makes it possible to employ a wide range of likelihood based inference methods for (μ, σ^2) , and it also means that it is easy to generate large numbers of simulated realizations x_i .

To use simulated values x_i for inference on (μ, σ^2) we will build upon a result first demonstrated in Beaumont (2003), which provided an example illustrating how a Markov chain Monte Carlo sampler could still have an exact stationary distribution while using an approximation to the true value of the likelihood in the Hastings ratio. Conditions under which exact sampling is possible were thoroughly explored in Andrieu and Roberts (2009), who termed the approach “pseudo marginal” Monte Carlo calculation. Essentially, a sufficient condition for exact sampling is for the likelihood estimate to be non-negative and unbiased. For a fixed pair (μ, σ^2) we can generate a simulated sample $\{x_i\}_{i=1}^n$, where the X_i are independent random variables each with $N(\mu, \sigma^2)$ distribution. If we can use the simulated values to calculate a non-negative and unbiased estimate of (3.1), we can take advantage of the pseudo marginal framework to perform exact inference on (μ, σ^2) .

In the univariate case, Sathe and Varde (1969) provide the uniform minimum variance unbiased estimate (UMVUE) $\hat{f}(x)$ of a density at a point x based on a sample of size n for a number of distributions, including the normal with unknown mean and variance. Their approach builds on the work of Basu (1964), which we summarize following the derivation of Shao (2010, chapter 3). Assume we have a sample $X = (X_1, \dots, X_n)$ of n independent and identically distributed (iid) random variables $X_i \sim N(\mu, \sigma^2)$. We begin by choosing an arbitrary index $j \in [1, \dots, n]$ and noting the indicator $I(X_j \leq x_j)$ provides an unbiased estimator of the probability $P(X_j \leq x_j)$. In the normal case the statistic $T(X) = (\bar{X}, S^2)$, with

$\bar{X} = \frac{1}{n} \sum X_i$ and $S^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2$, is complete and sufficient for (μ, σ^2) . By the Rao-Blackwell theorem conditioning on T gives us the UMVUE estimator

$$P(X_j \leq x_j | T) = P\left(\frac{X_j - \bar{X}}{S} \leq \frac{x_j - \bar{X}}{S} | T\right).$$

The statistic $Z(X) = \frac{X_j - \bar{X}}{S}$ can be interpreted as the standardized deviation of any single sample point from the sample mean. It was first considered by Thompson (1935); a more extensive discussion appears in Cramér (1946), who derives the explicit density function

$$f_Z(z) = c \left(1 - \frac{z^2}{n-1}\right)^{\frac{n-4}{2}} \quad (3.2)$$

if $z^2 \leq n-1$, $f_Z(z) = 0$ otherwise, and $c = \Gamma\left(\frac{n-1}{2}\right) \left[\Gamma\left(\frac{n-2}{2}\right) \sqrt{(n-1)\pi}\right]^{-1}$ is constant.

The distribution of Z is independent of both μ and σ^2 , and functions that are independent of the parameters, such as $Z(X)$, are known as ancillary statistics. Because $Z(X)$ is ancillary, by Basu's theorem it is also independent of the statistic $T(X)$. Therefore, we can drop the

condition, i.e. $P(X_j \leq x_j | T = (\bar{x}, s^2)) = P\left(Z \leq \frac{x_j - \bar{x}}{s} | T = (\bar{x}, s^2)\right) = P\left(Z \leq \frac{x_j - \bar{x}}{s}\right)$, and the UMVUE of $P(X_j \leq x_j)$ is

$$P(X_j \leq x_j | T) = \int_{-\infty}^{\frac{x_j - \bar{X}}{S}} fZ(z) dz. \quad (3.3)$$

Taking the derivative with respect to x_j on both sides of (3.3) shows the conditional density

of X_j given $\bar{X} = \bar{x}$ and $S^2 = s^2$ is $\frac{1}{s} fZ\left(\frac{x_j - \bar{x}}{s}\right)$, so the marginal density of X_j is

$$fX_j(x_j) = \int \frac{1}{s} fZ\left(\frac{x_j - \bar{x}}{s}\right) f_T(t) dt = E_T \left[\frac{1}{S} fZ\left(\frac{x_j - \bar{X}}{S}\right) \right]. \quad (3.4)$$

The quantity $\frac{1}{S} fZ\left(\frac{x_j - \bar{X}}{S}\right)$ is a function of a complete and sufficient statistic, and (3.4) shows it is an unbiased estimator of fX_j . Hence, for an iid sample of size n , the UMVUE of a normal density at the point x with (μ, σ^2) unknown is

$$\hat{f}(x | \mu, \sigma^2) = \frac{1}{s} fZ\left(\frac{x - \bar{x}}{s}\right) = c(s^2)^{-\frac{1}{2}} \left(1 - \frac{1}{n-1} \frac{(x - \bar{x})^2}{s^2}\right)^{\frac{n-4}{2}} \quad (3.5)$$

if $\frac{(x - \bar{x})^2}{s^2} \leq n-1$, $\hat{f}(x | \mu, \sigma^2) = 0$ otherwise, and c is the same as in (3.2).

Because (3.5) is a non-negative and unbiased estimator of (3.1), in the pseudo marginal framework it can be used instead of the density function in the Hastings ratio. Substituting (3.5) for (3.1) will result in a chain that marginally for (μ, σ^2) has the same stationary distribution as one based on the exact likelihood. Therefore, instead of evaluating the density function to run standard MCMC, we present a novel method based on using simulated realizations x_i to calculate an unbiased estimate of the likelihood. Our new simulation based MCMC method, which samples from the same stationary distribution as a likelihood based approach, proceeds as follows:

Algorithm 1 [Univariate Normal Pseudo Marginal MCMC]

0 From a prior distribution $\pi(\mu, \sigma^2)$ select initial values for μ and σ^2 , generate a simulated sample $\{x_i\}_{i=1}^n$ where $X_i \sim N(\mu, \sigma^2)$, and estimate the likelihood $f(x | \mu, \sigma^2)$ in (3.1) using $\hat{f}(x | \mu, \sigma^2)$ in (3.5).

1 Propose new parameters using a transition kernel $(\mu, \sigma^2 \rightarrow \mu', \sigma'^2)$.

2 Generate a simulated sample $\{x'_i\}_{i=1}^n$ where $X'_i \sim N(\mu', \sigma'^2)$, and estimate the likelihood $f(x | \mu', \sigma'^2)$ in (3.1) using $\hat{f}(x | \mu', \sigma'^2)$ in (3.5).

3 Accept the proposed move to μ', σ'^2 with probability $h((\mu, \sigma^2), (\mu', \sigma'^2))$,

$$h((\mu, \sigma^2), (\mu', \sigma'^2)) = \min \left[1, \frac{\hat{f}(x | \mu', \sigma'^2) \pi(\mu', \sigma'^2) q(\mu', \sigma'^2 \rightarrow \mu, \sigma^2)}{\hat{f}(x | \mu, \sigma^2) \pi(\mu, \sigma^2) q(\mu, \sigma^2 \rightarrow \mu', \sigma'^2)} \right].$$

4

If the move is accepted set $(\mu, \sigma^2) = (\mu', \sigma'^2)$ and $\hat{f}(x | \mu, \sigma^2) = \hat{f}(x | \mu', \sigma'^2)$ then return to 1. If the move is not accepted simply return to 1.

4 Multivariate Normal Case

The same general approach can also be applied to a multivariate normal distribution. Again we start by assuming we have a single data point $x \in \mathbb{R}^p$, an observed value of a p -dimensional multivariate normal random variable X with mean μ and positive definite variance matrix Σ . Generating pseudorandom realizations of X in the multivariate case is a straightforward extension of the univariate case. Specifically, given a sample z of p univariate independent standard normal random variables, a sample from X can be generated by calculating $x = \mu + Az$ where $AA^\top = \Sigma$, see e.g. Genz and Bretz (2009).

The matrix A is not unique and, as Genz and Bretz (2009) note, possible choices can be found using eigen, singular value, or Cholesky decompositions of Σ . All three methods are readily available in R. Independent simulated realizations of X can easily be generated in parallel using the R package “rlecuyer” from Sevcikova and Rossini (2012).

For the multivariate normal, derivations of UMVUE estimators equivalent to (3.5) can be found in Lumel'skii and Sapozhnikov (1969), Ghurye and Olkin (1969), and Eaton and Morris (1970). Following Eaton and Morris (1970), let $X = (X_1, \dots, X_n)$ be a $p \times n(p+2)$ random matrix with uncorrelated columns X_i , each of which is p -dimensional $N(\mu, \Sigma)$. We continue to assume that Σ is positive definite. As in the univariate case, their derivation of the multivariate unbiased estimator is predicated on identifying a complete and sufficient statistic $T(X)$, along with an ancillary statistic $Z(X)$ whose distribution can be calculated.

In the multivariate case a complete and sufficient statistic is $T(X) = (\bar{X}, S)$, where

$\bar{X} = \frac{1}{n} X e_p$, $S = (X - \bar{X} e_p^\top)(X - \bar{X} e_p^\top)^\top$, and e_p is a length p vector of ones. The ancillary statistic $Z(X)$ used by Eaton and Morris (1970) is

$$Z(X) = B^{-1}(X - \bar{X} e_p), \quad (4.1)$$

where B is the unique matrix in $G_{T,p}^+$ ($p \times p$ lower triangular matrices with positive diagonal) such that $S = BB^\top$, the Cholesky decomposition of S .

Eaton and Morris (1970) use the concept of invariance to motivate their results, and one example of this invariance is also their primary reason for choosing $B \in G_{T,p}^+$. Specifically, if $A \in G_{T,p}^+$ is used to transform X into AX , then B will be transformed into AB . To calculate

$Z(AX)$ we use the fact that we can rewrite $(X - \bar{X} e_p^\top)$ as $X(I - P_e)$, where $P_e = \frac{1}{n} e_p e_p^\top$. Therefore, $Z(X) = B^{-1}X(I - P_e)$, and

$$\begin{aligned}
 Z(AX) &= (AB)^{-1}AX(I - P_e) \\
 &= B^{-1}X(I - P_e) \\
 &= Z(X),
 \end{aligned} \tag{4.2}$$

which implies the statistic Z is invariant to transformations of the data X by matrices

$$A \in G_{T,p}^+$$

Once the statistics $T(X)$ and $Z(X)$ have been identified, the remaining task is to calculate the density of the ancillary statistic Z . The derivation in the multivariate case is similar to the univariate case, and details can be found in Eaton and Morris (1970). Based on the results of this derivation we see that in the multivariate case the UMVUE of the normal density based on a sample of size n is

$$\hat{f}(x|\mu, \Sigma) = c|S|^{-\frac{1}{2}} \left(1 - \frac{n}{n-1} (x - \bar{X})^\top S^{-1} (x - \bar{X}) \right)^{\frac{n-p-3}{2}} \tag{4.3}$$

if $(x - \bar{X})^\top S^{-1} (x - \bar{X}) \leq \frac{n-1}{n}$, and $\hat{f}(x|\mu, \Sigma) = 0$ otherwise. The normalizing constant is $c = \left(\frac{n}{n-1} \right)^{\frac{p}{2}} \Gamma\left(\frac{n-1}{2}\right) \left[\pi^{\frac{p}{2}} \Gamma\left(\frac{n-p-1}{2}\right) \right]^{-1}$.

The ability to generate pseudorandom samples from a $N(\mu, \Sigma)$ distribution, combined with the unbiasedness of (4.3), leads immediately to a multivariate extension of the univariate pseudo marginal approach introduced in algorithm 1.

Algorithm 2 [Multivariate Elliptical Pseudo Marginal MCMC]

- 0 From a prior distribution $\pi(\mu, \Sigma)$ select initial values for μ and Σ , simulate the $p \times n$ matrix X , and estimate the likelihood $f(x|\mu, \Sigma)$ using $\hat{f}(x|\mu, \Sigma)$ in (4.3).
- 1 Propose new parameters using a transition kernel $q(\mu, \Sigma \rightarrow \mu', \Sigma')$.
- 2 With μ' and Σ' simulate the $p \times n$ matrix X' and estimate the proposed likelihood $f(x|\mu', \Sigma')$ using $\hat{f}(x|\mu', \Sigma')$ in (4.3).
- 3 Accept the proposed move to μ', Σ' with probability

$$h((\mu, \Sigma), (\mu', \Sigma')) = \min \left[1, \frac{\hat{f}(x|\mu', \Sigma') \pi(\mu', \Sigma') q(\mu', \Sigma' \rightarrow \mu, \Sigma)}{\hat{f}(x|\mu, \Sigma) \pi(\mu, \Sigma) q(\mu, \Sigma \rightarrow \mu', \Sigma')} \right]$$
- 4 If the move is accepted set $(\mu, \Sigma) = (\mu', \Sigma')$ and $\hat{f}(x|\mu, \Sigma) = \hat{f}(x|\mu', \Sigma')$ then return to 1. If the move is not accepted simply return to 1.

As in the univariate case, our new algorithm allows one to sample from the posterior distribution of μ and Σ for a multivariate normal, given observed data x , without evaluating the multivariate normal density function.

Generating the matrix X' of simulated data in step 2 will be computationally intensive for large values of p . However, as the columns of X are independent, generating the n independent vectors X_i can be trivially parallelized, and the Cholesky decomposition of Σ' need only be calculated once. High performance linear algebra algorithms for large scale distributed computing are an active field of research, and libraries such as PLASMA and MAGMA, described in Agullo et al. (2009), can perform such decompositions on matrices with tens of thousands of rows. An R interface to these packages is also available, see Smith (2010). In situations where a sparseness assumption is appropriate for Σ , significant additional computational shortcuts will be possible via specialized algorithms such as CHOLMOD from Chen et al. (2008).

5 Elliptically Contoured Case

In this section we show how the results of section 4 can be extended to the class of elliptically contoured distributions. This is possible because a small variation of (4.2) shows that the invariance property of $Z(X)$, defined in (4.1), remains true for this larger class with minor additional assumptions. Additionally, the statistic $T(X)$ is complete and sufficient for the location and scale parameters of an elliptically contoured distribution. Thus, the derivations leading to the estimator (4.2) and algorithm 2 are applicable to any ECD satisfying the necessary conditions to ensure the invariance of $Z(X)$. These conditions are given formally below, but roughly speaking we require a full rank scale matrix Σ , a large enough sample size to estimate Σ , and a scaling variable R that is equal to zero with probability zero.

The matrix X' generated in step 2 of algorithm 2 consists of uncorrelated multivariate normal columns, which means it is a matrix elliptically contoured distribution, or $X' \sim EC_{p,n}(\mu' \otimes e_n, \Sigma' \otimes I_n, \phi)$. Just as in the normal case, for elliptically contoured distributions $T(X) = (\bar{X}, S)$ is complete and sufficient for (μ, Σ) , see for example Fang and Zhang (1990, theorem 4.3.1). To generalize the normal case we also need to know the distribution of the statistic $Z(X)$ in the elliptically contoured case. Similar to the result (4.2), we can easily check the condition (2.3)

$$\begin{aligned} Z((XX^\top)^{-\frac{1}{2}}X) &\stackrel{d}{=} ((XX^\top)^{-\frac{1}{2}}B)^{-1}(XX^\top)^{-\frac{1}{2}}X(I - P_e) \\ &\stackrel{d}{=} B^{-1}X(I - P_e) \\ &\stackrel{d}{=} Z(X). \end{aligned} \quad (5.1)$$

We assume Σ has full rank, $n > p$, and the scaling variable R satisfies $P(R = 0) = 0$, as in (2.3). The first equality therefore follows because we can find a square root matrix

$$\begin{aligned} (XX^\top)^{\frac{1}{2}} &\in G_{T,p}^+ \text{ via the Cholesky decomposition, and using it to transform } X \text{ into} \\ (XX^\top)^{-\frac{1}{2}}X &\text{ will transform } B \text{ into } (XX^\top)^{-\frac{1}{2}}B. \end{aligned}$$

Eaton and Morris (1970) show $Z(X)$ is independent of μ and Σ for the class of multivariate normal distributions. With some minor assumptions $Z(X)$ also satisfies (2.3), so the distribution of $Z(X)$ is invariant for almost the entire class of elliptically contoured

distributions. This invariance property, combined with the result of Fang and Zhang (1990) showing $T(X) = (\bar{X}, S^2)$ is complete and sufficient for any $X \in EC_p(\mu, \Sigma, \phi)$, means the result (4.3) derived from an assumption of normality will hold true for this much larger class. Therefore, algorithm 2 can be used to generate exact posterior samples from μ and Σ for almost any elliptically contoured distribution.

An important feature of algorithm 2 is that once the simulated sample data $\{x_i\}_{i=1}^n$ has been generated, the calculation of the density estimator (4.3) is exactly the same for every elliptically contoured distribution satisfying the conditions of (2.3). This means the estimator (4.3) is robust with respect to distributional assumptions; as long as it is fed simulated data from any of a large class of distributions it will enable exact posterior sampling of the parameters (μ, Σ) . Therefore, the only way the implementation of algorithm 2 ever fundamentally changes is by changing the distribution sampled from to generate the simulated data matrix X' in step 2.

As discussed in section 2, sampling from an elliptically contoured distribution is relatively easy, and in the special case of a normal scale mixture even easier still. The primary computation is to simulate a multivariate normal Y ; the only other computation required is to generate a single realization of the univariate variable W and multiply that by Y . Therefore, the order of complexity in running algorithm 2 for any p -dimensional normal scale mixture with parameters (μ, Σ) is the same, and in particular will be equal to the complexity in the p -dimensional normal case.

6 A Simple Example

To demonstrate algorithm 2 we apply it to the simple model $X \sim EC_2(\mu, \Sigma, \phi)$, where the data

consists of a single bivariate observation $x = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. To facilitate plotting we will restrict our

model to two univariate parameters by assuming $\mu = \begin{pmatrix} \nu \\ \nu \end{pmatrix}$, where $\nu \in \mathbb{R}$, and

$\Sigma = \begin{pmatrix} 1 & \sigma \\ \sigma & 1 \end{pmatrix}$, with the restriction $-0.9 < \sigma < 0.9$ to ensure Σ is positive definite. In the plots that follow ν is the “location” parameter and σ is the “scale” parameter, and we consider three different elliptically contoured distributions. These examples are meant to serve two purposes: primarily, the intent is to illustrate the utility of the normal scale mixture representation (2.2) for elliptically contoured distributions. This representation enables posterior sampling from a wide range of distributions with very minor changes to the implementation. The first two cases simply replicate likelihood based results, while the third illustrates the ease with which our algorithm can produce posterior samples for a distribution with no closed form density, which would be intractable to existing methods.

The second purpose of these examples is to illustrate the behavior of our method in practice. Although our simulation based chains possess the same stationary distribution as their likelihood based counterparts, this does not guarantee comparable results in a finite number of steps. The first two cases we consider have closed form densities, so it is straightforward

to implement both simulation and likelihood based samplers. We see that, as expected, the two approaches generate very similar results.

In the first example we assume that the distribution of X is multivariate normal. In terms of the representation (2.2), this corresponds to the case $W \equiv 1$. As our baseline for comparison we ran a typical likelihood based version of algorithm 2, using the exact multivariate normal density to calculate the Hastings ratio. The results from this chain are shown in the plot on the left side of figure 1. We then ran algorithm 2, and the results of this run are illustrated in the plot on the right side of figure 1. In both cases the prior distributions were $\nu \sim U(-2,4)$ and $\sigma \sim U(-0.9,0.9)$, while the other settings (initial conditions, transition kernels, number of proposals, and sampling frequency) were identical. Our theoretical result is that the two chains will have identical stationary distributions, and figure 1 demonstrates this result.

The next example is a multivariate elliptical Student's t distribution, with density function given in Kotz and Nadarajah (2004) as

$$f(x|\mu, \Sigma) = c |\Sigma|^{-\frac{1}{2}} \left[1 + \frac{1}{\nu} (x - \mu)^\top \Sigma^{-1} (x - \mu) \right]^{-\frac{\nu+p}{2}}, \quad (6.1)$$

where $c = \Gamma\left(\frac{\nu+p}{2}\right) \left[(\pi\nu)^{\frac{p}{2}} \Gamma\left(\frac{\nu}{2}\right) \right]^{-1}$ and ν is the degrees of freedom. For our example we set $\nu = 1$, or in other words X is multivariate elliptical Cauchy. Since $E(X)$ does not exist when $\nu = 1$, in this example μ is not the mean and Σ is not proportional to the variance. As before we first ran a standard likelihood based MCMC sampler using the exact density function (6.1); the joint parameter sample for this run is shown in the left hand plot of figure 2.

Running our simulation based algorithm requires generating a sample of simulated data from the appropriate distribution, in this case Student's t . Kotz and Nadarajah (2004) show that the multivariate elliptical Student's t distribution arises from the model (2.2) if one sets

the scale factor to $W = \sqrt{\frac{\nu}{C^2}}$, where C^2 is chi-squared with ν degrees of freedom. In practical terms, this result makes it trivial to modify an implementation of the normal algorithm for the multivariate elliptical Student's t . An implementation of a pseudorandom number generator for the multivariate Student's t based on this scaling factor is available in R via the function `rmvt` from the “mvtnorm” package by Genz et al. (2014). Figure 2 again demonstrates the equivalence of our method and the likelihood-based approach.

The last distribution we consider is the multivariate elliptically contoured Laplace distribution. From Kotz, Kozubowski, and Podgorski (2001) the density function can be expressed as

$$f(x|\mu, \Sigma) = c |\Sigma|^{-\frac{1}{2}} \left(\frac{1}{2} (x - \mu)^\top \Sigma^{-1} (x - \mu) \right)^{\frac{\nu}{2}} K_\nu \left(\sqrt{2(x - \mu)^\top \Sigma^{-1} (x - \mu)} \right), \quad (6.2)$$

where $v = \frac{2-p}{p}$, $c = \frac{2}{(2\pi)^{p/2}}$ and $K_v(\cdot)$ is a modified Bessel function of the second kind. For our purposes the most important feature of $K_v(\cdot)$ is that it lacks a general closed form expression. This means the multivariate elliptical Laplace density does not in general have a closed form either, and therefore it is not generally possible to perform exact likelihood based MCMC using (6.2).

To run an approximate version of likelihood based MCMC we used the function `dmvl` from the R package “LaplacesDemon” developed by Statisticat, LLC (2014), which uses the results of Wang, Zhang, and Zhao (2008) to approximate the Laplace density function. The posterior sample generated by running MCMC based on this approximation of the likelihood is shown in the plot on the left of figure 3.

For the simulation-based pseudo marginal MCMC we are again able to leverage the representation in terms of normal mixtures expressed in (2.2). In particular, as noted in Kotz et al. (2001), the appropriate form for W to generate a sample from an elliptically contoured Laplace distribution is $W = \sqrt{E}$, where E is standard exponential with parameter $\lambda = 1$. A pseudorandom number generator based on the normal scale mixture representation is implemented by the function `rmvl` from the package developed by Statisticat, LLC (2014). This function was used to generate the simulated values fed to algorithm 2, and the results from the simulation based chain are displayed in the plot on the right in figure 3.

Although the plots in figure 3 have the same general location and shape, inspection of the scales reveals the sample based on the approximate density function yields a peak almost twice as tall as the simulation based pseudo marginal sampler. With the taller peak in the left plot comes a more rapid rate of decay as the location parameter deviates from one. In both cases the MAP estimators are roughly the same, but the credible intervals these examples generate, while overlapping, are substantially different. As before all of the MCMC parameters were identical for the two chains. In order to keep the plots legible, for both of these chains the location prior was changed to $\nu \sim U(0,2)$. Therefore, not only are these two posterior samples different from each other, they are also significantly different from either a normal or Cauchy distributional assumption, as in figures 1 and 2.

7 A Biological Example

We now apply our method to a non-linear regression problem arising from a model of the glucose/insulin interaction in the human metabolism, described by Pacini and Bergman (1986). This model, known as MINMOD, involves three quantities of interest: $G(t)$, $I(t)$, and $X(t)$. The first two are the concentrations of glucose and insulin [respectively] in the bloodstream as functions of time t , while $X(t)$ describes the effect of insulin on net glucose disappearance. These three quantities are related by the differential equation

$$\frac{dG}{dt} = -(p_1 + X(t)) \cdot G(t) + p_1 \cdot G_b \quad G(0) = G_0 \quad (7.1)$$

$$\frac{dX}{dt} = -p_2 \cdot X(t) + p_3 \cdot (I(t) - I_b X(0)) = 0 \quad (7.2)$$

$$\frac{dI}{dt} = -n \cdot I(t) + \gamma \cdot (G(t) - h) \cdot t \quad I(0) = I_0 \quad (7.3)$$

G_b and I_b are baseline concentrations, while $p_1, p_2, p_3, G_0, n, \gamma, h$, and I_0 are parameters. Typically one of G, I is treated as a known input, and in the following example we treat $I(t)$ as such. Simulating the model consists in part of solving equations (7.1) and (7.2), and accordingly our interest lies in the posterior distribution of the 4 parameters p_1, p_2, p_3 , and G_0 . The stochastic component of the model arises from the assumption that the glucose measurements are subject to a normally distributed error with mean zero and variance proportional to the true value. That is, at time t the measured value $G_t = G(t) + \varepsilon_{G(t)}$.

Three different methods, likelihood based MCMC, our simulation based MCMC, and ABC regression, were used to estimate posteriors for p_1, p_2, p_3 , and G_0 . The data consists of 16 measured time points each for glucose and insulin, previously reported in Watanabe et al. (2007). For all three methods the prior distributions were $p_1 \sim U(0, 2^{-1})$, $p_2 \sim U(0, 1)$, $p_3 \sim U(0, 10^{-2})$, $G_0 \sim U(0, 500)$, and for the latter two methods an identical number of simulations were performed. The ABC regression implementation was based on R code available from the website of the first author of Beaumont et al. (2002), which was modified to support multiple parameters. The data was log transformed as suggested in Beaumont et al. (2002), but no summary statistics were employed, as all 16 glucose observations were used as predictors.

Marginal posterior density estimates based on equal size posterior samples from each of the three methods are shown in figure 4. Because this model assumes the error is normally distributed the likelihood based sample (dashed line) and our simulation based sample (green line) lead to marginal posterior density estimates that are almost directly on top of one another. On the other hand, the ABC regression posterior estimate (orange) is clearly different from the other two. The concordance between our method and the likelihood based approach corroborates our theoretical results, while using ABC regression for this model illustrates that another recent simulation based approach will yield reasonable, but clearly different, marginal posterior samples.

8 Generalizing the Biological Example

The analysis of section 7 is predicated on assuming the errors are independent and normally distributed. These assumptions, which are standard when performing inference for MINMOD, are made largely for reasons of analytic convenience. Our method allows us to analyze the model with more flexible error distributions, and here we present one possibility, which is a generalization of the multivariate normal. This distribution includes the normal as a testable special case, which allows the data, rather than assumptions, to determine if a normal error model is appropriate. We note that this is useful not just for this particular

example, but for any similar model where the assumption of normality does not have a rigorous justification.

As discussed in section 4, a standard procedure for generating a pseudorandom multivariate normal sample point is to first generate a vector z of p iid standard normal random variables,

which is transformed into a multivariate normal by calculating $\mu + \sum \frac{1}{2} z$. Because the multivariate normal is an elliptically contoured distribution, another way to generate a

simulated realization is implicitly given by the stochastic representation $\mu + R \sum \frac{1}{2} U^{(p)}$ of (2.1). For any elliptically contoured distribution $U^{(p)}$ is distributed uniformly on the surface of the p -dimensional unit sphere. Cambanis et al. (1981) show that if, and only if, R^2 has a

chi-square distribution with p degrees of freedom, computing $\mu + R \sum \frac{1}{2} U^{(p)}$ results in a multivariate normal distribution with mean μ and variance Σ . Based on this result, a natural way to generalize the multivariate normal is to let R^2 have a gamma distribution with shape

parameter k and scale parameter 2. If the shape parameter k is equal to $\frac{p}{2}$, R^2 is equivalent to a chi-square with $df=p$, and hence the model reduces to a multivariate normal. Any other value of k implies the corresponding elliptical distribution is non-normal. Therefore, testing the hypothesis $k=\frac{p}{2}$ will, in effect, also be testing the assumption of multivariate normality.

When R^2 has a gamma distribution with arbitrary $k \in (0, \infty)$, we are not aware of a general

expression for the density function of the random variable $X = \mu + R \sum \frac{1}{2} U^{(p)}$. However, generating a simulated realization of a $\Gamma(k, 2)$ random variable can be accomplished.

Therefore, applying our simulation based MCMC method to a model with arbitrary $k \in (0, \infty)$ is a straightforward extension of the normal case. Because k is no longer assumed to be known a priori, we can include it as a parameter in the MCMC to generalize the analysis of section 7. In figure 5a the previous, likelihood based, samples of the four MINMOD parameters are drawn as dashed lines. The green lines illustrate the posterior samples arising from the generalized model with $R^2 \sim \Gamma(k, 2)$, instead of chi-square p . As figure 5a shows, the generalized model results in posterior samples that are roughly twice as tall, and half as wide, compared to the results based on assuming normality.

Figure 5b is a plot of the posterior sample for the shape parameter k . Because the data consists of $p = 16$ points, when $k = 8$ the generalized model is equivalent to the normal model. In figure 5b the range of the middle 95% of the posterior sample is (1.02, 6.72), and the middle 99% range is (0.89, 8.44). These intervals also have an interpretation as a test of the assumption of multivariate normality. The 95% interval does not contain $k = 8$, while the 99% interval does. Thus, at a 5% level of significance, a null hypothesis of multivariate normality for this data would be rejected, while at a 1% significance level we do not reject the hypothesis of multivariate normality.

9 Conclusion

We have demonstrated a simulation based MCMC algorithm with a stationary distribution that enables exact posterior sampling of the location and scale parameters for almost the entire class of elliptically contoured distributions. This includes distributions with no closed form density, and therefore our approach enables exact sampling in cases where previously it was not possible. The example of section 8 demonstrates that even a simple, one parameter extension, of the multivariate normal can quickly lead to analytic obstacles which preclude employing a likelihood based analysis. However, these same cases are easily simulated, and hence easily handled by our method.

An interesting potential application for the methods we have developed arises from the special case of a (block) diagonal variance matrix Σ . In the normal case a variance matrix with a diagonal structure implies the variables are both uncorrelated and independent from one another. The number of parameters in a diagonal variance matrix grows linearly with p , while for an arbitrary variance matrix the number of parameters grows quadratically in p . Thus, in the normal case a large number of variables may make it necessary to impose restrictions on the number of non-independent variables simply to prevent an explosion in the number of non-zero terms in Σ that must be estimated.

For general elliptically contoured distributions a diagonal variance matrix also implies the variables are uncorrelated. However, it is only for the normal distribution that a diagonal variance matrix implies the variables are independent. In the normal case R^2 is a chi-square p random variable. This is the only choice of R^2 which makes uncorrelated and independent equivalent conditions; therefore, a diagonal variance matrix combined with any other choice for R^2 results in a model with uncorrelated, but dependent, variables. The nature of the dependence between the observations is determined by the distribution of R^2 . Thus, the combination of a diagonal matrix Σ and a more general distribution for R^2 , such as the gamma example of section 8, makes it possible to model dependencies in the data while simultaneously constraining the variance matrix to be diagonal. The number of parameters associated with the choice of R^2 will always be independent of the dimension of the data, and hence irrelevant to the scaling of the algorithm with respect to the number of observations. The gamma case shows how just one extra parameter can give rise to a distribution which subsumes the normal model. In general R can be almost any non-negative univariate random variable, so this is only one example of a very large number of otherwise intractable densities to which our simulation based MCMC algorithm can be applied for inference.

References

- Agullo E, Demmel J, Dongarra J, Hadri B, Kurzak J, Langou J, Ltaief H, Luszczek P, Tomov S. Numerical linear algebra on emerging architectures: The PLASMA and MAGMA projects. *Journal of Physics: Conference Series*, IOP Publishing. 2009; 180
- Anderson, TW.; Fang, K-T. Technical Report ADA115355. Department of Statistics; Stanford University: 1982. Distributions of quadratic forms and Cochran's theorem for elliptically contoured distributions and their applications.
- Andrieu C, Roberts GO. The pseudo-marginal approach for efficient Monte Carlo computations. *The Annals of Statistics*. 2009; 37:697–725.

- Basu A. Estimates of reliability for some distributions useful in life testing. *Technometrics*. 1964; 6:215–219.
- Beaumont MA. Estimation of population growth or decline in genetically monitored populations. *Genetics*. 2003; 164:1139–1160. [PubMed: 12871921]
- Beaumont MA, Cornuet J-M, Marin J-M, Robert CP. Adaptive approximate Bayesian computation. *Biometrika*. 2009 asp052.
- Beaumont MA, Zhang W, Balding DJ. Approximate Bayesian computation in population genetics. *Genetics*. 2002; 162:2025–2035. [PubMed: 12524368]
- Cambanis S, Huang S, Simons G. On the theory of elliptically contoured distributions. *Journal of Multivariate Analysis*. 1981; 11:368–385.
- Chen Y, Davis TA, Hager WW, Rajamanickam S. Algorithm 887: CHOLMOD, supernodal sparse Cholesky factorization and update/downdate. *ACM Transactions on Mathematical Software (TOMS)*. 2008; 35:22.
- Clauset A, Shalizi CR, Newman ME. Power-law distributions in empirical data. *SIAM review*. 2009; 51:661–703.
- Cramér, H. *Mathematical Methods of Statistics*. Princeton University Press; 1946.
- Del Moral P, Doucet A, Jasra A. An adaptive sequential monte carlo method for approximate Bayesian computation. *Statistics and Computing*. 2012; 22:1009–1020.
- Drovandi CC, Pettitt AN. Estimation of parameters for macroparasite population evolution using approximate Bayesian computation. *Biometrics*. 2011; 67:225–233. [PubMed: 20345496]
- Eaton ML, Morris CN. The application of invariance to unbiased estimation. *The Annals of Mathematical Statistics*. 1970:1708–1716.
- Fang, K.; Zhang, Y. *Generalized Multivariate Analysis*. Springer-Verlag; 1990.
- Filippi S, Barnes CP, Cornebise J, Stumpf MP. On optimality of kernels for approximate Bayesian computation using sequential monte carlo. *Statistical applications in genetics and molecular biology*. 2013; 12:87–107. [PubMed: 23502346]
- Genz, A.; Bretz, F. *Lecture Notes in Statistics*. Heidelberg: Springer-Verlag; 2009. *Computation of Multivariate Normal and t Probabilities*.
- Genz, A.; Bretz, F.; Miwa, T.; Mi, X.; Leisch, F.; Scheipl, F.; Hothorn, T. *mvtnorm: Multivariate Normal and t Distributions*. 2014. URL <http://CRAN.R-project.org/package=mvtnorm>, R package version 1.0-0
- Ghurye S, Olkin I. Unbiased estimation of some multivariate probability densities and related functions. *The Annals of Mathematical Statistics*. 1969:1261–1271.
- Gupta, A.; Varga, T.; Bodnar, T. *Elliptically Contoured Models in Statistics and Portfolio Theory*. 2nd. Springer-Verlag; 2013.
- Kelker D. Distribution theory of spherical distributions and a location-scale parameter generalization. *Sankhyā : The Indian Journal of Statistics, Series A*. 1970:419–430.
- Kotz, S.; Kozubowski, T.; Podgorski, K. *Progress in Mathematics Series*. Birkhäuser; Boston: 2001. *The Laplace Distribution and Generalizations: A Revisit With Applications to Communications, Economics, Engineering, and Finance*.
- Kotz, S.; Nadarajah, S. *Multivariate T-Distributions and Their Applications*. Cambridge University Press; 2004.
- Lumel'skii YP, Sapozhnikov P. Unbiased estimates of density functions. *Theory of Probability & Its Applications*. 1969; 14:357–364.
- Marjoram P, Molitor J, Plagnol V, Tavaré S. Markov chain Monte Carlo without likelihoods. *Proc Natl Acad Sci USA*. 2003; 100:15324–15328. [PubMed: 14663152]
- Newman ME. Power laws, Pareto distributions and Zipf's law. *Contemporary physics*. 2005; 46:323–351.
- Pacini G, Bergman RN. MINMOD: a computer program to calculate insulin sensitivity and pancreatic responsivity from the frequently sampled intravenous glucose tolerance test. *Comput Methods Programs Biomed*. 1986; 23:113–122. [PubMed: 3640682]

- Robert CP, Cornuet JM, Marin JM, Pillai NS. Lack of confidence in approximate Bayesian computation model choice. *Proc Natl Acad Sci USA*. 2011; 108:15112–15117. [PubMed: 21876135]
- Sathe Y, Varde S. On minimum variance unbiased estimation of reliability. *The Annals of Mathematical Statistics*. 1969; 40:710–714.
- Sevcikova, H.; Rossini, T. rlecuyer: R interface to RNG with multiple streams. 2012. URL <http://CRAN.R-project.org/package=rlecuyer>, R package version 0.3-3
- Shao, J. *Springer Texts in Statistics*. Springer; 2010. *Mathematical Statistics*.
- Sisson SA, Fan Y, Tanaka MM. Sequential Monte Carlo without likelihoods. *Proc Natl Acad Sci USA*. 2007; 104:1760–1765. [PubMed: 17264216]
- Smith, BJ. R package magma: Matrix Algebra on GPU and Multicore Architectures. 2010. URL <http://cran.r-project.org/package=magma>, R package version 0.2.2
- Statisticat, LLC. LaplacesDemon: Complete Environment for Bayesian Inference. 2014. URL <http://www.bayesian-inference.com/software>, R package version 14.06.23
- Thompson WR. On a criterion for the rejection of observations and the distribution of the ratio of deviation to sample standard deviation. *The Annals of Mathematical Statistics*. 1935; 6:214–219.
- Wang D, Zhang C, Zhao X. Multivariate Laplace filter: a heavy-tailed model for target tracking. *Pattern Recognition, 2008 ICPR 2008 19th International Conference on*, IEEE. 2008:1–4.
- Watanabe RM, Allayee H, Xiang AH, Trigo E, Hartiala J, Lawrence JM, Buchanan TA. Transcription factor 7-like 2 (TCF7L2) is associated with gestational diabetes mellitus and interacts with adiposity to alter insulin secretion in Mexican Americans. *Diabetes*. 2007; 56:1481–1485. [PubMed: 17317761]

Figure 1.
Exact normal likelihood and pseudo marginal samples

Figure 2.
Exact Student's t likelihood and pseudo marginal samples

Figure 3.
Approximate Laplace likelihood and pseudo marginal samples

Figure 4.
Marginal posterior densities using three different methods

Figure 5.
Marginal posterior densities for the generalized model