Multivariate Gaussian Simulation Outside Arbitrary Ellipsoids

Nick Ellis and Ranjan Maitra*

Abstract

Methods for simulation from multivariate Gaussian distributions restricted to be from outside an arbitrary ellipsoidal region are often needed in applications. A standard rejection algorithm that draws a sample from a multivariate Gaussian distribution and accepts it if it is outside the ellipsoid is often employed: however, this is computationally inefficient if the probability of that ellipsoid under the multivariate normal distribution is substantial. We provide a two-stage rejection sampling scheme for drawing samples from such a truncated distribution. Experiments show that the added complexity of the two-stage approach results in the standard algorithm being more efficient for small ellipsoids (i.e. with small rejection probability). However, as the size of the ellipsoid increases, the efficiency of the two-stage approach relative to the standard algorithm increases indefinitely. The relative efficiency also increases as the number of dimensions increases, as the centers of the ellipsoid and the multivariate Gaussian distribution come closer, and as the shape of the ellipsoid becomes more spherical. We provide results of simulation experiments conducted to quantify the relative efficiency over a range of parameter settings.

1 Introduction

The need to simulate from the extreme regions of a multivariate Gaussian distribution arises in a variety of applications. An example is in the context of environmental risk assessment where one may need to simulate from an extreme event to make inference on certain parameters (Hefferman and Tawn, 2004). Other application areas include environmental impact assessment (de Haan and de Ronde, 1998), contaminant modelling (Lockwood and Schervish, 2005) and strategies for financial management (Poon *et al*, 2004). However, the primary motivation for our interest in this problem, comes from the context of outlier detection within

*Nick Ellis is Natural Resource Modeler at CSIRO Marine and Atmospheric Research, 233 Middle St, Cleveland, QLD 4163, Australia. Ranjan Maitra is Associate Professor in the Department of Statistics and Statistical Laboratory, Iowa State University, Ames, IA 50011-1210, USA.

a proposed refinement of the multi-stage clustering procedure for massive datasets defined in Maitra (2001). The methodology adopted there is first to cluster a random sample of the dataset using some clustering technique, and then to identify representativeness of the clusters in the rest of the dataset using a likelihood-ratio test under the assumption that the clusters are from homogeneous multivariate Gaussian distributions. The rejected observations are then resampled, clustered, and the groups tested again for representativeness. The process is repeated until no further sampling is possible.

One refinement to the above methodology stems from the fact that, at each stage, the sample contains observations from groups that are too scarcely represented to be recognized from the sample as a separate cluster. These observations should be identified as outliers to the cluster they get assigned to, and should be removed from these groups before inference is performed on those groups. One way to test for the presence of outliers in Gaussian populations is to compute the multivariate sample kurtosis measure T of Mardia (1970, 1974, 1975) and use that to detect outliers (Schwager and Margolin, 1982), with the rejection region decided by simulation.

In the first stage of the clustering algorithm, this approach is straightforward, since the rejection region can be estimated using the quantiles of T, obtained from samples drawn from a standard multivariate Gaussian distribution. However, in subsequent stages, the observations in identified groups are no longer multivariate Gaussian. Rather, they are observations from a multivariate normal distribution restricted to be outside a union of ellipsoids, being the rejection regions of the previous stages. To detect outliers, one could use the same test statistic T, but now the sampling distribution of T must be estimated using simulations sampled outside these ellipsoids. While most of these ellipsoids will perhaps be far away from the support of the Gaussian distribution restricted to the cluster, there will possibly be a few ellipsoids with substantial overlap. Such a scenario presents a need for multivariate Gaussian simulation from outside arbitrary ellipsoids.

There has been a lot of interest in this and related problems. Many authors in particular have provided algorithms to compute the probability of a multivariate normally distributed random vector over different kinds of regions. For instance, Bohrer and Schervish (1981), building upon the work of Milton (1972), provided algorithms for calculating multivariate normal probabilities over rectangular regions. Schervish (1984) provided a faster algorithm for calculating such probabilities along with their error bounds. More recently, Lohr (1993) presented an algorithm to compute the multivariate normal probabilities inside general-shaped regions, of which the ellipsoid is a special case. On the other hand, interest in multi-Gaussian simulation from extreme regions is much more recent. Hefferman and Tawn (2004) developed the theory for simulation from a general class of multivariate distributions restricted to component-wise extreme regions, and Lockwood and Schervish (2005) discussed strategies for MCMC sampling on component-wise censored data.

In this paper, we specialize to the case of sampling from a multivariate Gaussian distribution from extreme regions defined to be in the form of the complement to an arbitrary ellipsoid.

Without loss of generality, we can assume a standard multivariate Gaussian distribution, since, under transformation to standard coordinates, the ellipsoid is mapped to another ellipsoid. Therefore we aim to simulate from the following p-variate density, given by

$$f(\boldsymbol{x}) \propto \exp\{-\frac{\boldsymbol{x}'\boldsymbol{x}}{2}\}\mathbf{1}[(\boldsymbol{x}-\boldsymbol{\mu})'\Gamma(\boldsymbol{x}-\boldsymbol{\mu}) > a].$$

Simulation from the above distribution can be done, using crude rejection sampling. However, when the ellipsoid has probability close to one under the standard multivariate normal distribution, this approach can be extremely inefficient with most realizations being discarded rather than being accepted. The worst-case scenario is when $\Gamma \approx \boldsymbol{I}$, $\boldsymbol{\mu} \approx 0$ and a is large. When $a = \chi^2_{p;0.99}$, for instance, only about one percent of the proposals will be accepted. This is unacceptably inefficient for computational purposes. We therefore define rejection algorithms which will account for two alternative scenarios. In the first case, we address the above mentioned worst-case scenario. In the second, we extend this situation to include more general cases. These two cases are presented in the following section. We next briefly describe the available C programs and software and detail performance evaluations on a range of simulation experiments. Finally, we conclude with a brief summary and outline questions for further research.

2 Theory and Methods

2.1 Standard Multivariate Simulation Outside a Zero-Centered Sphere

We start with the following result:

Theorem 2.1 *Let $\boldsymbol{X} \sim N_p(0, \boldsymbol{I})$ and let Y be independent of $\boldsymbol{X}$ and distributed as:*

$$f(y) \propto \exp\{-\frac{y}{2}\}y^{\frac{p}{2}-1}\mathbf{1}[y > d], \tag{1}$$

the density of the χ^2_p-distribution restricted to the part of the positive half of the real line above $d \geq 0$. Then, $\boldsymbol{Z} = \sqrt{Y}\frac{\boldsymbol{X}}{\|\boldsymbol{X}\|}$ has the density given by

$$f(\boldsymbol{z}) \propto \exp\{-\frac{\boldsymbol{z}'\boldsymbol{z}}{2}\}\mathbf{1}[\boldsymbol{z}'\boldsymbol{z} > d]. \tag{2}$$

Proof: The proof follows from first principles of transformation of variables $(\boldsymbol{X}, Y)$ to $(\boldsymbol{Z}, U)$, with $\boldsymbol{Z}$ as above and $U = \|\boldsymbol{X}\|/\sqrt{Y}$. Then the joint density of the transformed variables is

$$f(\boldsymbol{z}, u) \propto u^{p-1}\exp\{-(u^2+1)\frac{\|\boldsymbol{z}\|^2}{2}\}\|\boldsymbol{z}\|^p\mathbf{1}(\|\boldsymbol{z}\|^2 > d)$$

from where the result follows, on integrating out u over its range (0 through ∞). □

The above result means that if we have an efficient scheme for sampling from the right tail of a χ^2-distribution, then we can simulate the standard multivariate distribution outside a given zero-centered sphere quite easily. We now address the issue of sampling from the density given by (1). Devroye (1986) has shown one way to sample using the smallest dominating exponential of the tail distribution (p 425, exercise 7). However, another alternative, which we adopt, is to use the adaptive rejection sampling scheme of Gilks and Wild (1992), which we expect to be more efficient. To do this, note that the logarithm of the density is given by

$$h(y) = \text{constant} - \frac{y}{2} + \left(\frac{p}{2} - 1\right) \log y,$$

as long as $y \in [d, \infty)$. It is easy to see that $h''(y) = -(\frac{p}{2} - 1)/y^2$ which means that for $p \geq 2$, h is concave in the domain, and hence the conditions for adaptive rejection sampling apply. Sampling from (2) is then straightforward.

2.2 Standard Multivariate Gaussian Simulation Outside Arbitrary Ellipsoids

2.2.1 Case I: Origin inside the Ellipsoid

We now address the case for simulation from a standard multivariate Gaussian distribution restricted to outside an arbitrary ellipsoid that includes the origin. Our approach is to find, respectively, the largest origin-centered sphere contained in the ellipsoid and the smallest origin-centered sphere containing the ellipsoid. While the density of a standard multi-Gaussian outside the largest enclosed sphere is the outer envelope of a possible rejection algorithm, the density outside the exterior smallest enclosing sphere can be used as the squeezing function of Ripley (1987). We next provide methods to approximate the radii of the largest enclosed and the smallest enclosing spheres.

If the ellipsoid is centered at zero *i.e.* $\boldsymbol{\mu} = \mathbf{0} \equiv (0, 0, \ldots, 0)'$, the smallest and the largest spheres, centered at zero and touching the ellipsoids have squared radii given by $d/\lambda_{\max}$ and $d/\lambda_{\min}$ where $\lambda_{\min}$ and $\lambda_{\max}$ are the smallest and the largest eigenvalues of $\boldsymbol{\Gamma}$, respectively. We now focus our discussion on the cases where the ellipsoid is centered away from zero.

As we show below, there are up to $2p$ so-called *osculating* spheres centered at zero that touch the ellipsoid. Figure 1(*left*) shows an example of 4 osculating circles for the two-dimensional case of an ellipse. We find the radii of these spheres using a combination of techniques from linear algebra, multi-variable calculus and numerical methods. Define a quadratic form $\psi(\boldsymbol{x}) = (\boldsymbol{x} - \boldsymbol{\mu})'\Gamma(\boldsymbol{x} - \boldsymbol{\mu})$ which describes a family of ellipsoids centered at $\boldsymbol{\mu}$. We consider the particular ellipsoid satisfying $\psi(\boldsymbol{x}) = d$, and note that the condition for $\mathbf{0}$ to be inside the ellipsoid is $\psi(\mathbf{0}) < d$. We wish to find the position $\boldsymbol{x}_c$ of the point of contact of the osculating sphere and the ellipsoid. This is the point where the normal of the sphere is parallel to the

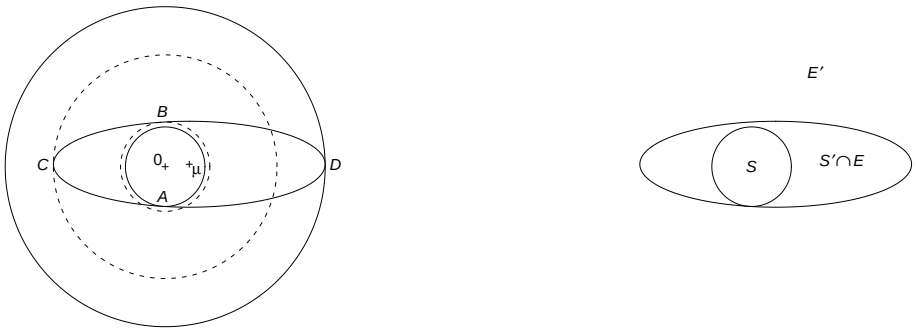

Figure 1
(*left*) An example of an ellipse centered at $\boldsymbol{\mu}$ with 4 osculating circles centered at $\mathbf{0}$ touching at A, B, C, and D. (*right*) Definition of the regions S, S', E and E' for the two-dimensional case shown at left.

normal of the ellipsoid. The normal of the sphere is parallel to $\boldsymbol{x}_c$, and the normal of the ellipsoid is parallel to the gradient vector $\nabla\psi = 2\Gamma(\boldsymbol{x} - \boldsymbol{\mu})$, therefore $\boldsymbol{x}_c$ is a solution to the equation

$$\Gamma(\boldsymbol{x} - \boldsymbol{\mu}) = \lambda \boldsymbol{x} \tag{3}$$

for some $\lambda > 0$. The matrix Γ is symmetric positive definite. Letting $\lambda_1, \lambda_2, \ldots, \lambda_p$ represent its eigenvalues (in decreasing order of magnitude) and $\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2, \ldots, \boldsymbol{\zeta}_p$ as the corresponding eigenvectors, the matrix Γ has a spectral decomposition given by $\Gamma = \sum_{i=1}^{p} \lambda_i \boldsymbol{\zeta}_i \boldsymbol{\zeta}_i'$. Note that the eigenvectors $\{\boldsymbol{\zeta}_i; i = 1, 2, \ldots, p\}$ form a basis of R^p so that we can write $\boldsymbol{x} = \sum_{i=1}^{p} \alpha_i \boldsymbol{\zeta}_i$ and $\boldsymbol{\mu} = \sum_{i=1}^{p} \beta_i \boldsymbol{\zeta}_i$. Then (3) implies that $\sum_{i=1}^{p} \lambda_i(\alpha_i - \beta_i)\boldsymbol{\zeta}_i = \lambda \sum_{i=1}^{p} \alpha_i \boldsymbol{\zeta}_i$ and therefore we have $\lambda_i(\alpha_i - \beta_i) = \lambda\alpha_i$ for all $i = 1, 2, \ldots, p$. This implies that each α_i can be written in terms of the knowns and the unknown λ: $\alpha_i = \lambda_i\beta_i/(\lambda_i - \lambda)$. Further, the equation of the ellipse, $\psi(\boldsymbol{x}_c) = d$, yields the additional constraint: $\sum_{i=1}^{p} \lambda_i(\alpha_i - \beta_i)^2 = d$ which provides the following equation in the unknown λ:

$$\sum_{i=1}^{p} \frac{\lambda^2 \lambda_i \beta_i^2}{(\lambda_i - \lambda)^2} = d. \tag{4}$$

Equation (4) may be cast in a slightly more convenient form by using reciprocals $\nu = 1/\lambda$

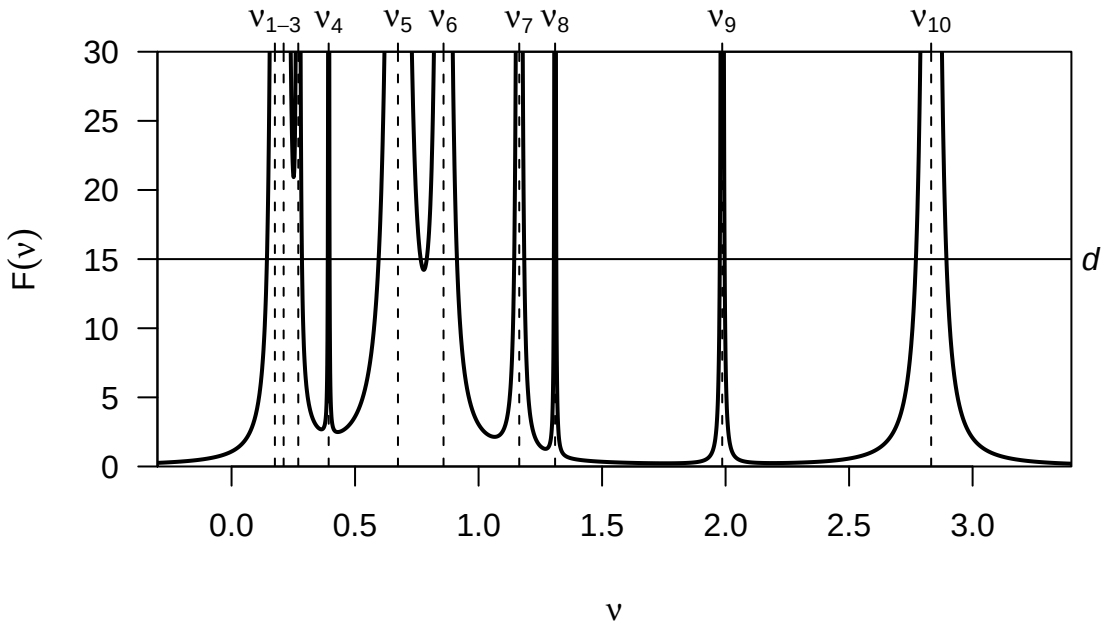

Figure 2

Graph of $F(\nu)$ versus ν for a 10-dimensional ellipsoid. The horizontal dashed line denotes the value $d = 15$, and the β_i take the values (–0.24, 0.31, 0.09, 0.02, –0.36, –0.21, –0.07, –0.02, –0.03, –0.14). The positions of the inverse eigenvalues ν_i are shown by vertical dashed lines. For $\nu_1 < \nu < \nu_2$ the function $F(\nu)$ exceeds the upper limit of the Figure.

and $\nu_i = 1/\lambda_i$. Then equation (4) becomes

$$F(\nu) \equiv \sum_{i=1}^{p} \frac{\nu_i \beta_i^2}{(\nu - \nu_i)^2} = d. \tag{5}$$

The function $F(\nu)$ resembles a long marquee with tent poles at the ν_i. An example with $p = 10$ is shown in Figure 1. For all ν, except at the poles, $F(\nu)$ is positive, continuous and convex ($F''(\nu) > 0$) and $F(\nu) \to 0$ as $\nu \to \pm\infty$. It follows that the equation $F(\nu) = d$ has exactly one solution for $\nu < \nu_1$, exactly one solution for $\nu > \nu_p$, and either 0 or 2 (real) solutions between each pair of adjacent poles. There are therefore at most $2p$ real roots in all.

Each root yields an osculating sphere of squared radius $R(\nu)$, where

$$R(\nu) \equiv \sum_{i=1}^{p} \alpha_i^2 = \nu^2 \sum_{i=1}^{p} \frac{\beta_i^2}{(\nu - \nu_i)^2}. \tag{6}$$

In general, an osculating sphere that is tangent to the ellipsoid at the point $\boldsymbol{x}_c$ may intersect the ellipsoid elsewhere away from $\boldsymbol{x}_c$, so that the sphere is partly inside and partly outside the ellipse. The exceptions are the osculating spheres of minimum radius r_m (which is the largest zero-centered sphere enclosed by the ellipsoid) and of maximum radius r_M (which is the smallest zero-centered sphere enclosing the ellipsoid). In fact the minimum and maximum radii correspond to the minimum and maximum roots ν of equation (5), as we now show.

Lemma 2.2 *Let* $0 \le z_1 < z_2 < \ldots < z_n < \infty$ *be the distinct roots of* $F(\nu) \equiv \sum_{i=1}^{p} \nu_i \beta_i^2/(\nu - \nu_i)^2 = d$. *Then* $R(z_1) < R(z_2) < \ldots < R(z_n)$, *where* $R(\nu) = \nu^2 \sum_{i=1}^{p} \beta_i^2/(\nu - \nu_i)^2$.

Proof: From (5) and (6) it follows that

$$R(\nu) = \nu F(\nu) + \nu \sum_{i=1}^{p} \frac{\beta_i^2}{(\nu - \nu_i)}.$$

For any two distinct roots $0 < z_l < z_k$ of $F(\nu) = d$, we find

$$\begin{aligned} R(z_k) - R(z_l) &= (z_k - z_l) \left(d - \sum_{i=1}^{p} \frac{\nu_i \beta_i^2}{(z_k - \nu_i)(z_l - \nu_i)} \right) \\ &= \frac{(z_k - z_l)}{2} \left(\sum_{i=1}^{p} \frac{\nu_i \beta_i^2}{(z_k - \nu_i)^2} + \sum_{i=1}^{p} \frac{\nu_i \beta_i^2}{(z_l - \nu_i)^2} - 2 \sum_{i=1}^{p} \frac{\nu_i \beta_i^2}{(z_k - \nu_i)(z_l - \nu_i)} \right) \\ &= \frac{(z_k - z_l)^3}{2} \sum_{i=1}^{p} \frac{\nu_i \beta_i^2}{(z_k - \nu_i)^2 (z_l - \nu_i)^2} \\ &> 0. \end{aligned}$$

□

Therefore the radii of the osculating spheres increase monotonically with the corresponding roots of equation (5). In particular, r_m corresponds to the single root smaller than ν_1, and r_M to the single root larger than ν_p. These two roots are easily obtained by simple search methods.

When $\beta_j = 0$ for some j we have a degenerate case that requires special handling. Let us define $F_{-j}(\nu)$ as the function $F(\nu)$ with the j-th term omitted from the sum in (5), and similarly, $R_{-j}(\nu)$ as the function $R(\nu)$ with the j-th term omitted from the sum in (6). If $F_{-j}(\nu_j) < d$ then in the limit $\beta_j \to 0$ there is a root with multiplicity 2 at $\nu = \nu_j$. One can see this from Figure 1: there are two roots just above and below ν_9, which has a correspondingly small $\beta_9 = 0.03$, and the same can be seen for the 4th and 8th eigenvalues ($\beta_4 = 0.02$, $\beta_8 = -0.02$). As $\beta_j \to 0$, these roots satisfy

$$\nu = \nu_j \pm \frac{\beta_j \sqrt{\nu_j}}{\sqrt{d - F_{-j}(\nu_j)}} + O(\beta_j^2).$$

Hence, in the limit the roots coalesce to a double root $\nu = \nu_j$, giving $\alpha_j = \pm\sqrt{\nu_j(d - F_{-j}(\nu_j))}$ and $R(\nu) = R_{-j}(\nu_j) + \nu_j(d - F_{-j}(\nu_j))$. The remaining roots ($\nu \neq \nu_j$) have $\alpha_j = 0$ and are found by solving the reduced equation $F_{-j}(\nu) = d$. Since it is only necessary to find the smallest and largest roots, this degenerate case only needs to be handled when $j = 1$ or $j = p$.

Once the largest enclosed and the smallest enclosing spheres are identified, we come up with the algorithm, in the spirit of Theorem 2.1. For, we can generate, using the methods of Gilks and Wild (1992) and Gilks (1992), realization Y from the restricted χ^2_p distribution, restricted to the part of the real line greater than r_m^2. Independent of that, we generate $\boldsymbol{X}$ from a standard p-variate normal distribution. Let $\boldsymbol{Z} = \sqrt{Y}\frac{\boldsymbol{X}}{\|\boldsymbol{X}\|}$. Then, if $Y > r_M^2$, return $\boldsymbol{Z}$ as a realization, otherwise we need to evaluate whether $\boldsymbol{Z}$ is outside the ellipse and decide on a rejection or acceptance of the realization accordingly. Formally, we accept the realization $\boldsymbol{Z}$ if $(\boldsymbol{Z} - \boldsymbol{\mu})'\boldsymbol{\Gamma}(\boldsymbol{Z} - \boldsymbol{\mu}) > d$, otherwise we reject and return to sampling Y and $\boldsymbol{X}$ all over again.

2.2.2 Case II: Origin outside the Ellipsoid

For the case when the origin is either on the ellipsoid or outside, we propose finding, using the above, the radius r_m of the largest sphere such that every point inside the sphere is closer to the origin than any point inside the ellipsoid, and the radius r_M of the smallest sphere such that every point outside the sphere is farther from the origin than any point inside the ellipsoid. Then the proposal is to generate $\boldsymbol{X} \sim N(\boldsymbol{0}, \boldsymbol{I})$ and, independent of $\boldsymbol{Z}$, $Y \sim \chi^2_p$. Let $\boldsymbol{Z} = \sqrt{Y}\frac{\boldsymbol{X}}{\|\boldsymbol{X}\|}$. If $Y < r_m^2$ or $Y > r_M^2$, we return $\boldsymbol{Z}$ as the realization otherwise we check whether $\boldsymbol{Z}$ is outside the ellipse, and accept $\boldsymbol{Z}$ as our realization or reject $\boldsymbol{Z}$ and start all over again.

We find the osculating spheres in almost exactly the same way as for Case I. The only change is that now $F(0) > d$, so that $\nu_{\min}$ lies in the interval $(-\infty, 0)$. This negative value arises because the normal of the sphere is now anti-parallel to the normal of the ellipsoid.

3 Performance Evaluations

3.1 Programs and Available Software

We provide ANSI/ISO C99-compliant C functions to obtain pseudo-random realizations from the p-variate standard multi-normal distribution restricted outside a given ellipsoid. Because of the origin of our interest in this problem, we specify the ellipsoid in terms of $\{\boldsymbol{\mu}, \boldsymbol{\Sigma}, a\}$ where $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}^{-1}$. This can then be viewed as realizations from the multinormal distribution outside the ellipsoid of concentration for another multi-normal distribution.

Our C function is called `rtmvnorm` (random sampling from a truncated multivariate normal) and is specified as follows:

```
int rtmvnorm(int nsample, int p, double *x, double *ltsigma,
             double *mu, double d, int *niter)
```

The C function `rtmvnorm` is integer-valued and takes in six arguments. The first argument `nsample` specifies the desired sample size while the second (`p`) specifies the dimensionality of the desired realizations. The third argument `x` is a pointer to a one-dimensional C double-precision array of length `nsample`×`p` for which space has been previously allocated in the calling program. This is where the sample realizations will be returned, row-wise, on a successful call to this function. The argument `ltsigma` is a similar pointer to a double-precision one-dimensional array containing the elements of $\mathbf{\Sigma}$, row-wise, in packed lower-triangular format, while `mu` is another pointer to a one-dimensional double-precision array containing the p coordinates of $\boldsymbol{\mu}$ above. The argument `d` contains the d in the definition of the ellipse above. Finally, the argument `niter` simply points to the number of simulations in the argument. The function returns a value of C variable type `int`. A return value of zero indicates a successful draw of the sample from the target distribution. Further details on the function and associated functions are provided at the journal's supplementary materials website.

Finally, we note that the program `rtmvnorm` applies to a standard multivariate normal distribution. To sample for the general case of a multivariate distribution one would need to transform $\mathbf{\Gamma}$ and $\boldsymbol{\mu}$ to coordinates with respect to which the distribution is standard normal, sample using `rmtvnorm`, then transform the samples back.

3.2 The Experimental Setup

We now present performance evaluations of our sampling method over a range of cases. All computations were done on a Dell Precision 650 workstation, having two 3.06GHz Intel(r) Xeon(tm) processors and running the Fedora Core 5 Linux 2.6.17-1.2174_FC5smp kernel and using the GCC 4.1.1 suite of compilers.

3.2.1 Factors affecting performance relative to the standard approach

Before turning to the empirical results from the simulations, we can anticipate the factors that affect performance. They are

- the overhead in calculating the osculating sphere,
- the relative cost of calculating a single candidate sample, and

- the volume of the multivariate normal density covered by the osculating sphere.

Let t_{over} be the CPU time used in the overhead calculations (finding the eigenvalues of $\mathbf{\Gamma}$ and solving for the radii of the osculating spheres), and let $t_{\text{std},1}$ and $t_{\text{2S},1}$ be the CPU times to calculate a single candidate sample with the standard and two-stage methods respectively. Since the two-stage method requires making a standard multivariate sample $\boldsymbol{X}$ as well as sampling from the χ^2_p-distribution, we have $t_{\text{2S},1} > t_{\text{std},1}$. Let $P(E')$ and $P(S')$ be the probabilities of a standard multivariate sample lying outside the ellipse E and the sphere S respectively. (See Figure 1 (*right*) for an example of these regions.) The expected CPU time to obtain N samples from E' is

$$t_{\text{std},N} = N t_{\text{std},1} / P(E') \tag{7}$$

for the standard method and

$$t_{\text{2S},N} = N t_{\text{2S},1} P(S') / P(E') + t_{\text{over}} \tag{8}$$

for the two-stage method. Therefore the two-stage method will only be faster than the standard method if

$$P(S') < t_{\text{std},1} / t_{\text{2S},1} \tag{9}$$

and

$$N > t_{\text{over}} P(E') / (t_{\text{std},1} - t_{\text{2S},1} P(S')). \tag{10}$$

We obtain estimates for t_{over}, $t_{\text{2S},1}$ and $t_{\text{std},1}$ from the simulations. We also explore how $P(S')$ and $P(E')$ depend on the choices of p, d, $\mathbf{\Gamma}$ and $\boldsymbol{\mu}$.

3.2.2 The simulation suites

We ran a suite of simulations for various settings of p, d, $\mathbf{\Gamma}$ and $\boldsymbol{\mu}$ and for different sample sizes N. For each simulation we obtained the number of candidate samples before rejection N_{cand} and the user time t (as reported in the `tms_utime` field of the C function `times`). For the two-stage method we also recorded the radius r_m of the smallest osculating sphere. Each simulation was replicated 10 times. The input settings were as follows:

- The dimension p was set to 2, 10, 50 and 100.
- The size d of the ellipsoid was set to $p + d_1\sqrt{2p}$, with d_1 set to 1, 3 and 5. The motivation for this choice was that the size should be related to the mean and standard deviation of the χ^2_p-distribution which are p and $\sqrt{2p}$, respectively.
- The shape of the ellipsoid was controlled by a parameter s which was set to 0.5, 0.7 and 0.9. The matrix $\mathbf{\Gamma}$ was generated by randomly selecting its p log-eigenvalues uniformly on $(-\log s, \log s)$ and choosing random but orthogonally constrained orientations for

the eigenvectors. The eigenvalues were scaled to have product equal to 1 so that the ellipsoid given by $d = 1$ has unit volume. Thus the size of the ellipsoid is determined only by d and the shape by s. For $s = 0.9$ the ellipsoid is nearly spherical and for $s = 0.5$ the ellipsoid is fairly elongated in certain random directions.

- The position $\boldsymbol{\mu}$ of the ellipsoid center is given by $(\sqrt{\delta d/\boldsymbol{\Gamma}_{1,1}}, 0, 0, \ldots, 0)$ where $0 \leq \delta < 1$. The condition on δ ensures that the origin lies inside the ellipsoid. We set $\delta = \sqrt{2/p}\delta_1$ with δ_1 taking values 0, 0.1 and 0.5. The direction of $\boldsymbol{\mu}$ may be fixed because the orientation of the ellipsoid is completely random. The purpose of the $\sqrt{2/p}$ term is to prevent $P(S')$ growing too rapidly with p; initial experiments showed that, if a constant were used in place of this term, $P(S')$ quickly approached 1 as p increased. For $\delta_1 = 0$, the ellipsoid is centered at the origin and osculating spheres do not need to be found.

- In most cases N was set at 10,000 and 50,000. However, in cases where $P(E')$ was anticipated to be very small, N was reduced to 50 and 100 for the standard method. In addition, we ran simulations with $N = 0$ for the purpose of estimating the overhead time.

This suite of simulations, which we call the main suite, is designed to establish the relative performance of the two methods over a broad range of parameter space. We also design a second suite that is targeted to the region of parameter space in which the two stage method should be superior. This is the region where $P(S')$ is sufficiently small. We can target this region by making the settings of s depend on p.

Consider the largest possible enclosed sphere S for an ellipsoid; it is the concentric sphere with radius $\sqrt{\nu_1 d}$, having $P(S) = P(\chi^2_p < \nu_1 d)$. In our simulations ν_1 is close to s. We arrive at bounds on s using a rough argument that $\nu_1 d$ (and therefore sd) should be no smaller than $p - \sqrt{2p}$, which is 1 standard deviation below the mean of the χ^2_p distribution. Since in our simulations $d \approx p$, s should therefore lie between $1 - \sqrt{2/p}$ and 1. Hence for our targeted suite of simulations:

- Let $s = (1 - \sqrt{2/p}) + s_1\sqrt{2/p}$, with s_1 set to 0.1, 0.5 and 0.9.

3.3 Results of the simulations

3.3.1 The overhead time

The overhead time for a single simulation was computed by averaging the time over $2,000,000/p$ repetitions, implemented by embedding the overhead code in a `for` loop. The entire main simulation suite was run for the two-stage method with $N = 0$, and the results analyzed for dependence on the parameter settings. We found that t_{over} was independent of d_1, but

depended strongly on p and and weakly on δ and s. Table 1 shows the average times for each (p, δ_1, s) combination. For $\delta = 0$, when the root finding calculation is omitted, the value of t_{over} is slightly less than for cases with $\delta > 0$. This shows that the cost of the root finding is small compared to the cost of the eigenvalue decomposition, becoming relatively smaller as p increases. The variation in cost with δ and shape s appears random, and may be attributable to the randomness of the ellipsoids. The dependence on p is approximately quadratic. We also calculated overhead times for the targeted simulation suite. The results were very similar and are not shown.

Table 1
Estimated overhead time t_{over} (ms) for the two-stage method in the main simulation suite.

δ_1	s	$p = 2$	$p = 10$	$p = 50$	$p = 100$
0	0.5	0.0078	0.0506	1.63	9.42
0	0.7	0.0078	0.0493	1.68	9.40
0	0.9	0.0077	0.0498	1.71	9.36
0.1	0.5	0.0121	0.0639	1.71	9.40
0.1	0.7	0.0118	0.0630	1.77	9.42
0.1	0.9	0.0118	0.0632	1.78	9.48
0.5	0.5	0.0120	0.0638	1.70	9.46
0.5	0.7	0.0119	0.0625	1.75	9.49
0.5	0.9	0.0118	0.0635	1.76	9.51

3.3.2 The time per candidate sample

We analyzed the times t reported from the simulation suite with $N > 0$ to estimate $t_{\text{std},1}$ and $t_{\text{2S},1}$. An estimator of these times is $(t - t_{\text{over}})/N_{\text{cand}}$; we explored the dependency of this quantity to the parameter settings and found that $t_{\text{std},1}$ depended only on p and that $t_{\text{2S},1} - t_{\text{std},1}$ was a positive constant. The values of the estimates are shown in Table 2. The extra time spent by the two-stage method in sampling the truncated χ^2_p-distribution was 2.8μs. The increase with p is very roughly linear.

Table 2
Estimated time per sample $t_{\text{std},1}$ and $t_{\text{2S},1}$ (μs) for each dimension.

method	$p = 2$	$p = 10$	$p = 50$	$p = 100$
STD	0.477	2.27	13.4	30.7
2S	3.622	5.41	16.6	33.9

3.3.3 The probability $P(S')$ outside the osculating sphere

The two-stage method is most effective when $P(S')$ is small. Figure 3 shows the values of $P(S')$ generated by the suite of simulations. The most favorable case is for $\delta = 0$ when the ellipsoid is centered on the origin and the osculating sphere is largest, thus excluding a large portion of the rejection region and making $P(S')$ small. As $\boldsymbol{\mu}$ moves away from the origin, $P(S')$ increases. This occurs more quickly for larger p. Also $P(S')$ increases as the ellipsoid becomes more elongated (s decreases). For larger ellipsoids $P(S')$ decreases exponentially with d.

For the targeted simulation suite, $P(S')$ still depends strongly on d_1 and δ_1 but is only weakly dependent on p and s_1 (Figure 6). This was the purpose of parametrizing s in terms of p and s_1. In particular, $P(S')$ is bounded away from 1 over most of this region of parameter space.

3.3.4 The acceptance probability

The acceptance probability P_{acc} is the probability that a candidate sample is accepted, i.e. lies in the region E'. This is a measure of the efficiency of the sampling scheme. Theoretically, $P_{\text{acc}} = P(E')$ for the standard method and $P_{\text{acc}} = P(E')/P(S')$ for the two-stage method. An estimator of P_{acc} is N/N_{cand}. The weighted average of N/N_{cand} over replicates (weighted by N) is shown in Figure 4 for all settings of the main simulation suite. By comparison with Figure 3, we see that P_{acc} is raised in proportion to $1/P(S')$. In fact the ratio of these estimates of P_{acc} agrees very closely with the theoretical value, thus providing a check that the two-stage method is sampling correctly.

3.3.5 The time per sample

The potential efficiency gains are translated into actual computation costs in Figure 5, which shows the average time per sample for all parameter settings with $N = 50,000$. The domain of parameter space that favors the two-stage method is quite complicated. For $p = 100$, the two-stage method is more effective for the more spherical ellipsoid. The improvement is marginal for the highly elongated case because P_{acc} is nearly 1 in any case (see Figure 4). For $p = 2$, the relatively high value of $t_{2\text{S},1}$ counts against the two-stage method, except for some cases with $s = 0.9$ where P_{acc} differs strongly between the two methods. For $p = 10$ the relative performance of the two methods is mixed. In all cases the two-stage method becomes more effective as d_1 increases. The improvement can be quite dramatic, e.g. the 3 orders of magnitude speed-up for $p = 100$, $s = 0.9$, and in fact the ratio of computation times changes exponentially.

For the targeted simulation suite (Figure 7), the two-stage method tends to perform better for $p \geq 10$ and $\delta_1 < 0.5$. For cases $\delta_1 = 0.5$ the performance of the two methods is comparable.

For $p = 2$ the two-stage method is generally less efficient except for large, nearly circular, and nearly concentric ellipses.

4 Discussion

In this paper, we present an approach to simulate from a multivariate Gaussian density which is restricted to outside an ellipsoidal region. Although the methodology presented here is in the context of standard multivariate Gaussian densities, it is general enough to apply to more general scenarios by means of an appropriate affine transformation on the realizations. We present software, written in C, for the purpose. Performance evaluations are reported and indicate that while the approach outlined by us brings in several layers of sophistication, it comes with an initial overhead. For sampling outside large ellipsoids, this overhead is more than matched by cost savings in rejection sampling and therefore, our method is very practical in such situations.

Our simulations specialized to the case of a standard multivariate normal. As we have stated, the general multivariate normal case would be handled by transforming to standard coordinates, which would require an eigenvalue decomposition of the variance-covariance matrix. The overhead time would then be approximately doubled, since we have seen that t_over is mainly due to the eigenvalue decomposition for the ellipsoid, and the time per sample would be slightly increased due to the transformation to and from standard coordinates. This last cost is also incurred by the standard rejection method.

There are a few possibilities to improve the algorithm. If $\boldsymbol{\mu}$ were close to $\mathbf{0}$ (i.e. $|\boldsymbol{\mu}| \ll \sqrt{\nu_1 d}$) then we could dispense with the root finding altogether by setting $r_m = \sqrt{\nu_1 d} - |\boldsymbol{\mu}|$ and $r_M = \sqrt{\nu_p d} + |\boldsymbol{\mu}|$. These would be slightly sub-optimal choices but they would still guarantee correct sampling.

The efficiency of the two-stage approach depends strongly on the speed of the truncated χ^2 sampler. If a more efficient algorithm than ARS could be found, this would widen considerably our algorithm's region of efficacy.

An important related statistical issue is that of parameter estimation and inference in such a setup. Suppose that we have data that are realizations from multivariate Gaussian densities restricted to be outside an ellipsoidal region, and that one or more of the parameters governing the distribution or the ellipsoid are unknown. Parameter estimation can be obtained in the usual way, for instance, using likelihood methods, and the properties of the estimator can be studied via the parametric bootstrap, using our suggested simulation methodology.

A further question that remains is whether the strategy can be extended from ellipsoidal regions to more general regions. The eigenvalue approach is very well suited to the ellipsoidal region. Rejection sampling for non-ellipsoidal regions would probably require a different approach. For instance, the largest sphere inside a polyhedral volume could be found by

quadratic programming techniques. Thus, we see that while this paper provides a promising approach to simulation from the extremes of multivariate Gaussian densities, there are still a few issues remaining that require further statistical attention.

Acknowledgements

The research of the second author was supported in part by the National Science Foundation, under its CAREER award DMS-0437555. We are indebted to Peter Baker, Bill Eddy, Geoff Laslett, Francis Pantus, Rouben Rostamian, Bill Venables and two anonymous referees for suggestions that improved earlier versions of this manuscript.

References

[1] Anderson, E., Bai Z., Bischof, C., Blackford, L. S., Demmel, J., Dongarra, J., Du Croz J., Greenbaum, A., Hammarling, S., McKenney, A., Sorensen, D. (1999) *LAPACK Users' Guide. Third Edition*. Society for Industrial and Applied Mathematics, Philadelphia, PA.

[2] Bohrer, R. E. and Schervish, M. J. (1981). An error-bounded algorithm for normal probabilities of rectangular regions. *Technometrics*, **23** 297-300.

[3] de Haan, L. and de Ronde, J. (1998). Sea and wind: multivariate extremes at work. *Extremes*, **1**, 7-45.

[4] Devroye, L. (1986). *Non-Uniform Random Variate Generation*. Springer-Verlag, New York.

[5] Gilks W. R. (1992). Derivative-free adaptive rejection sampling for Gibbs sampling. In *Bayesian Statistics 4 (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, Eds.)*, Oxford University Press, London.

[6] Gilks W. R. and Wild (1992). Adaptive rejection sampling for Gibbs sampling. *Applied Statistics*, **41 2** 337-348.

[7] Hefferman, J. E. and Tawn, J. A. (2004) A conditional approach for multivariate extreme values (with discussion). *Journal of the Royal Statistical Society Series B* **66 3** 497–546.

[8] J. R. Lockwood and Mark J. Schervish (2005). MCMC strategies for computing Bayesian predictive densities for censored multivariate data. *Journal of Computational and Graphical Statistics*, **14 2** 395-414.

[9] Lohr, S. L. (1993). Algorithm AS 285: Multivariate Normal probabilities of star-shaped regions, *Journal of Applied Statistics*, **42 3**, 576–582.

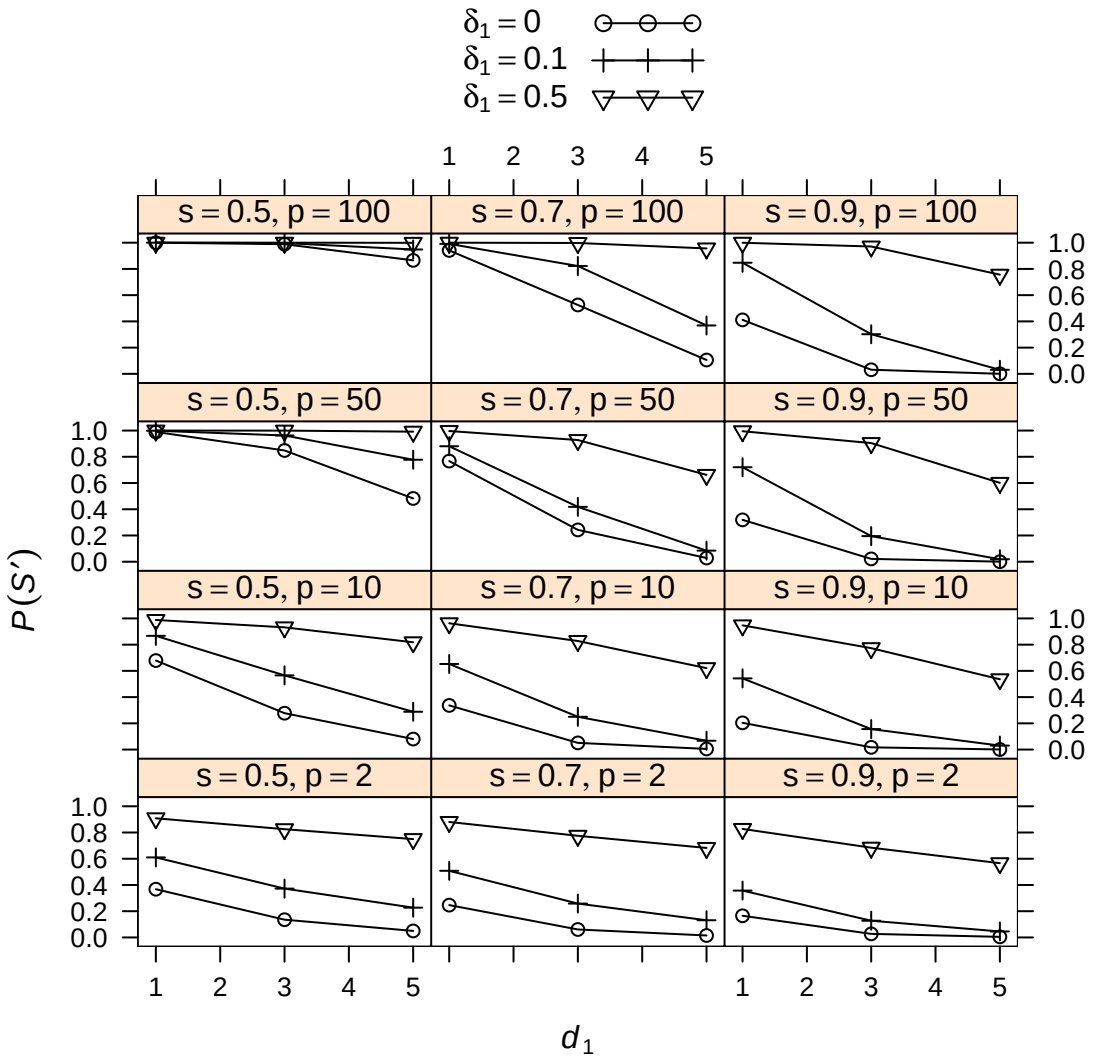

Figure 3

Value of $P(S')$ for all simulations. Each panel shows $P(S')$ versus d_1 for a particular combination of s and p. Different values of δ are denoted by a different symbol. Symbols are joined by a solid line for clarity.

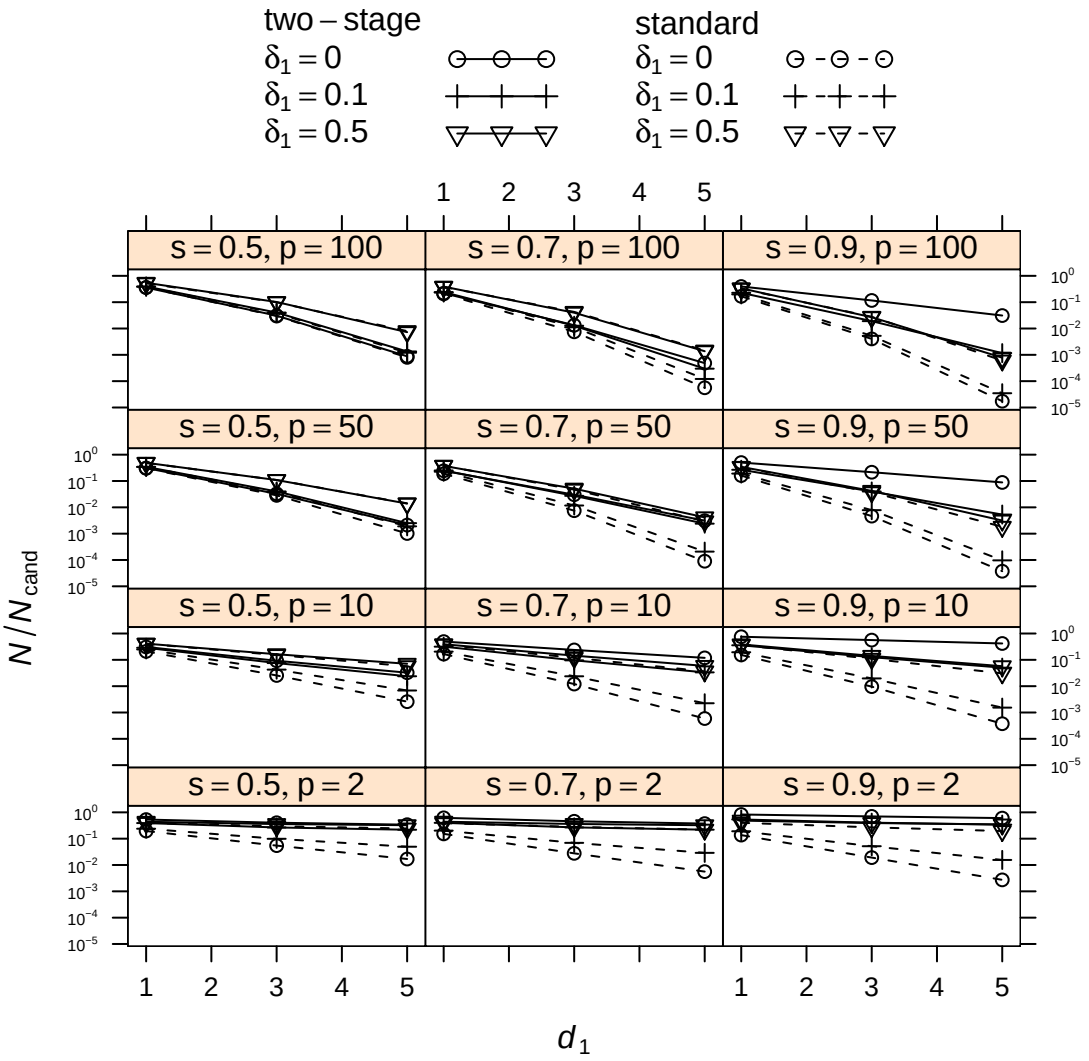

Figure 4

Comparison of acceptance probabilities for the two methods. Each panel shows $N/\bar{N}_{\text{cand}}$ versus d_1 for a particular combination of s and p with vertical axis on a logarithmic scale. Different values of δ are denoted by a different symbol. Symbols for the two-stage method are joined by a solid line and those for the standard method are joined by a dashed line.

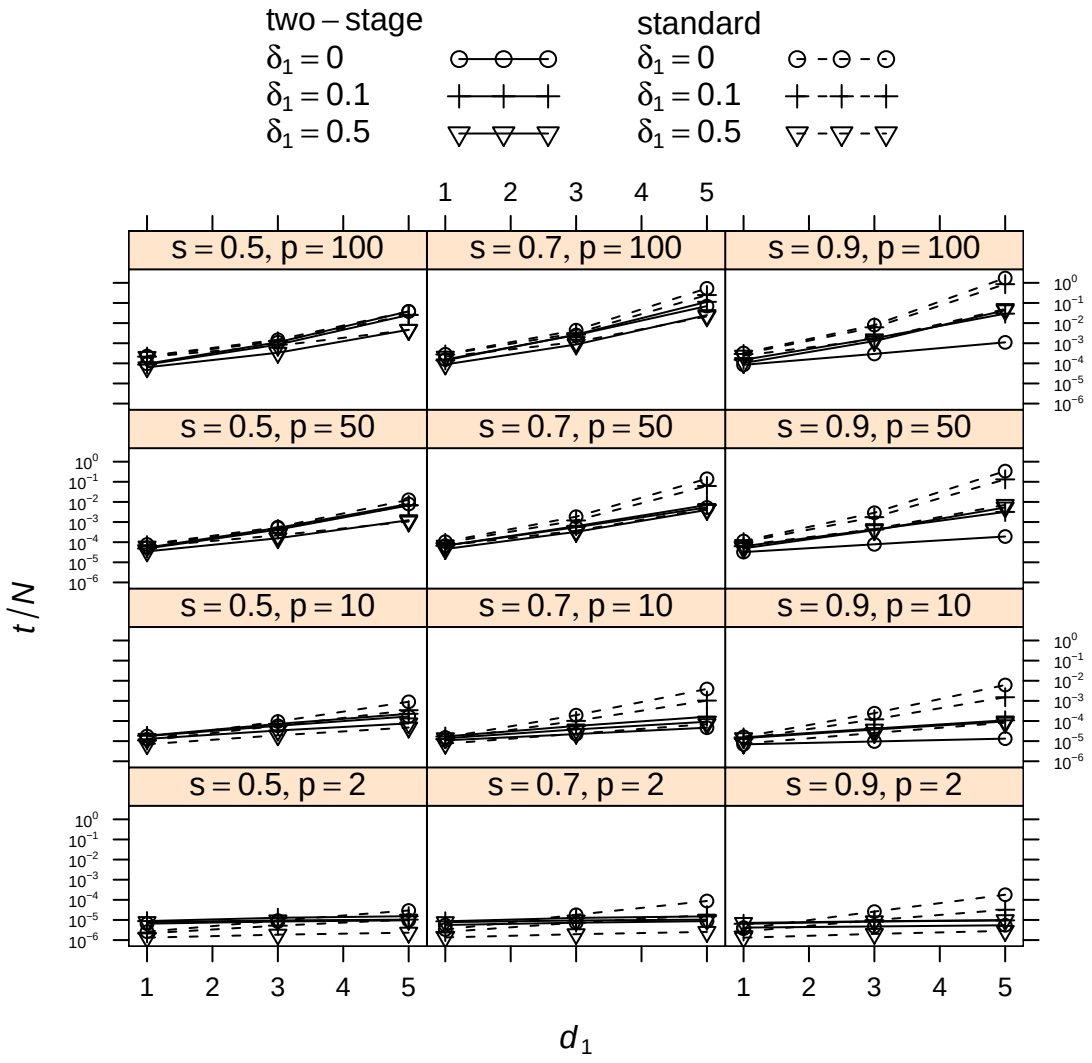

Figure 5

Comparison of time per sample for the two methods. Each panel shows $\bar{t}/N$ versus d_1 for a particular combination of s and p with vertical axis on a logarithmic scale. Different values of δ are denoted by a different symbol. Symbols for the two-stage method are joined by a solid line and those for the standard method are joined by a dashed line.

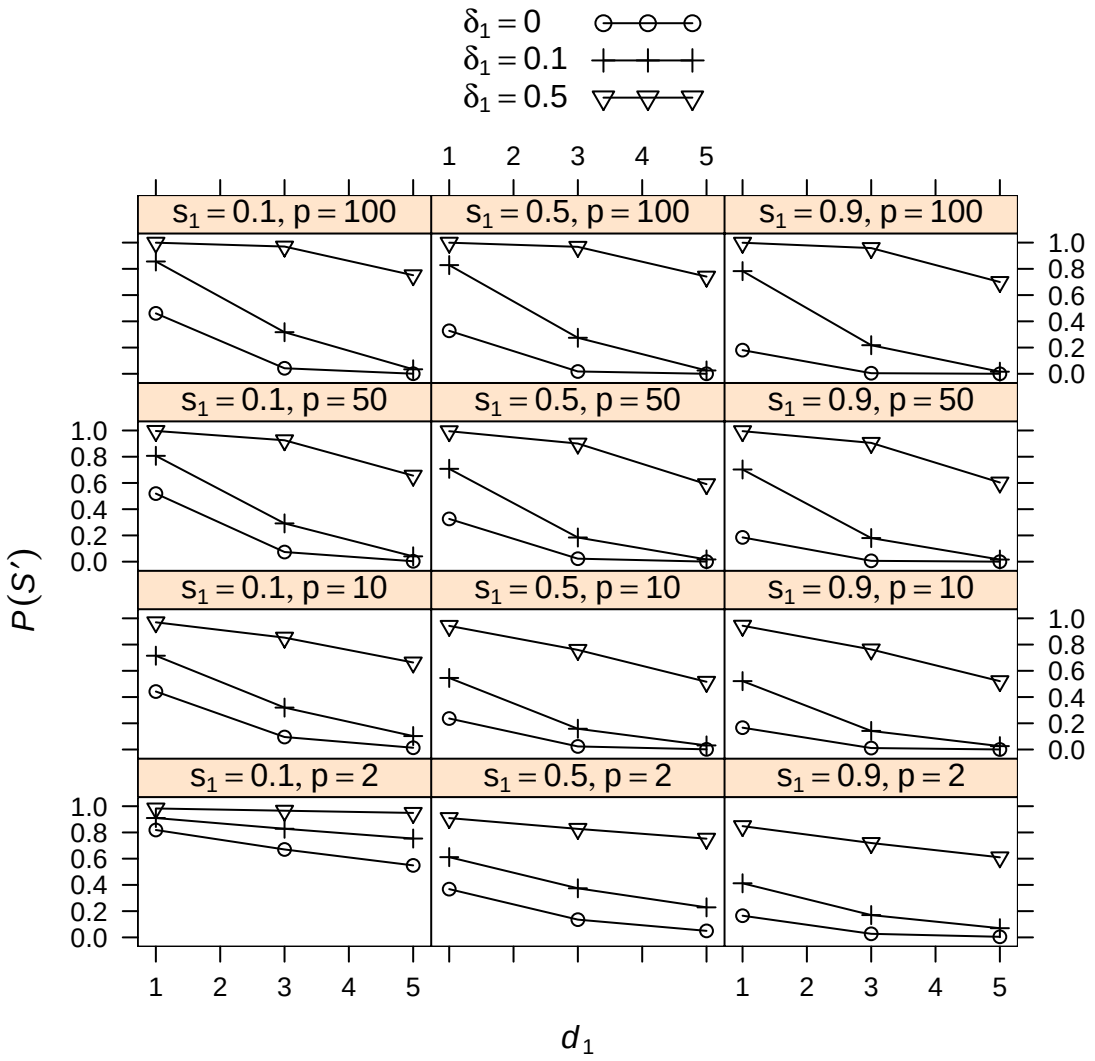

Figure 6

Value of $P(S')$ for all simulations in the targeted simulation suite. Each panel shows $P(S')$ versus d_1 for a particular combination of s_1 and p. Different values of δ are denoted by a different symbol. Symbols are joined by a solid line for clarity.

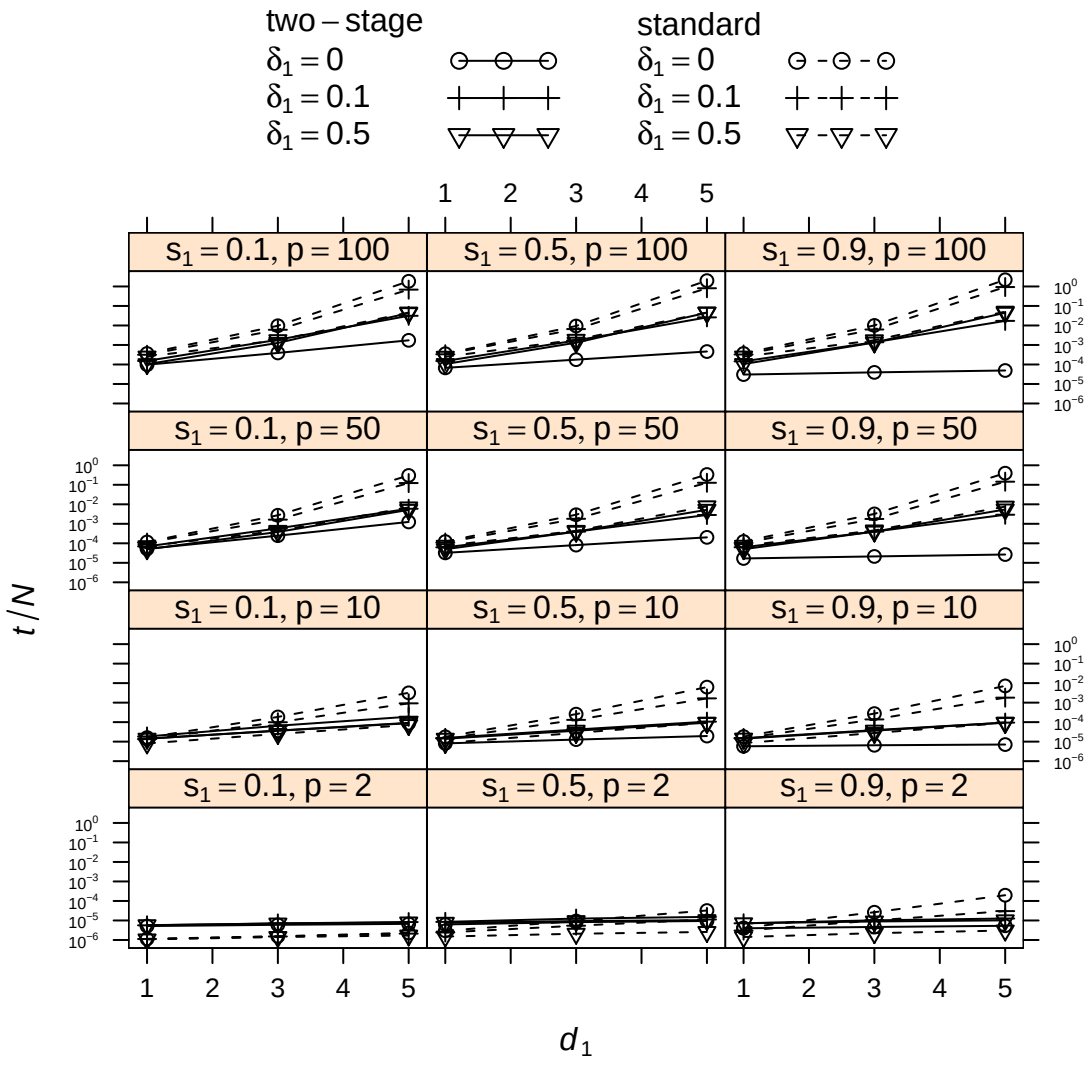

Figure 7

Comparison of time per sample for the two methods in the targeted simulation suite. Each panel shows $\bar{t}/N$ versus d_1 for a particular combination of s_1 and p with vertical axis on a logarithmic scale. Different values of δ are denoted by a different symbol. Symbols for the two-stage method are joined by a solid line and those for the standard method are joined by a dashed line.

[10] Maitra, R. (2001). Clustering massive datasets with applications in software metrics and tomography, *Technometrics*, **43 3**, 336–346.

[11] Mardia, K.V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, **57** 519–530.

[12] Mardia K. V. (1974). Applications of Some Measures of Multivariate Skewness and Kurtosis for Testing Normality and Robustness Studies, *Sankhya, Series A*, **36** 115–128.

[13] Milton R. C. (1972). Computer evaluation of the normal integral. *Technometrics*, **14** 881-89.

[14] Poon, S.-H., Rockinger, M. and Tawn, J. A. (2004) Extreme-value dependence in financial markets: diagnostics, models and financial implications. *Rev. Finan. Stud.*, 17, 581-610.

[15] Ripley, B. D. (1987). *Stochastic Simulation*. Wiley.

[16] Schervish, M. J. (1984). Algorithm AS 195: Multivariate normal probabilities with error bound (Corr: 85V34 p103-104) Applied Statistics, **33** 81-94.

[17] Schwager, S. J. and Margolin, B. H. Detection of multivariate normal outliers. *Annals of Statistics* **10 3** 943-954.