

RESEARCH SUPPORT  
UNIVERSITY OF MICHIGAN BUSINESS SCHOOL

DECEMBER 1996

**UNIFORM SAMPLING IN THE HYPERSPHERE VIA  
LATENT VARIABLES AND THE GIBBS SAMPLER**

**WORKING PAPER #9612-36**

**CRAIG CUMBUS  
KIMBERLY-CLARK CORPORATION  
AND  
PAUL DAMIEN  
UNIVERSITY OF MICHIGAN BUSINESS SCHOOL  
AND  
STEPHEN WALKER  
IMPERIAL COLLEGE, LONDON, UK**



# Uniform Sampling in the Hypersphere via Latent Variables and the Gibbs Sampler

Craig Cumbus

Kimberly-Clark Corporation, Atlanta, GA, USA

Paul Damien

School of Business, University of Michigan, Ann Arbor, MI, USA

Stephen Walker

Department of Mathematics, Imperial College, London, UK.

## ABSTRACT

A simple and general algorithm for generating random vectors from the unit hypersphere is developed. The algorithm is shown to be efficient when compared to methods of rejection from a circumscribing hypercube and more viable when the dimension is large. It is seen that such random vectors can be obtained purely via simple transformations of uniform random variables, or by combining such transformations with Gibbs sampling schemes obtained by appropriate introduction of latent variables.

Key words: Gibbs sampler, Unit hypersphere, Inverse transform, Latent variables.

## I. Introduction

While there are at least a few efficient ways (Devroye, 1986) to generate random vectors uniformly *on* a unit hypersphere of dimension  $d$ ,  $C_d$ , apparently, the only known “efficient” way to generate random vectors uniformly *in*  $C_d$  is via rejection from  $(-1, 1)^d$ . However, this method is terribly inefficient for  $d$  larger than, say 5; for example, the acceptance rate is less than 0.0025 for  $d = 10$ . Additionally, methods such as the Metropolis-Hastings algorithm are difficult to implement in this situation as well; for an excellent discussion of the Metropolis-Hastings algorithm and related methods, see, for example, Chib and Greenberg, 1995. Here we describe a fairly efficient and convenient method, especially for sampling large batches, of generating random vectors from the unit hypersphere.

The basic idea is that a random vector in hyperspherical coordinates can be simulated relatively easily from  $d$  univariate probability densities, and this random vector can be transformed into a random vector in rectangular coordinates that is uniformly distributed in  $C_d$ .

It may be of interest to note that uniformity is not lost under a linear transformation (Devroye, 1986), so, in particular, it is possible to transform a sample from  $C_d$  into a sample from any hyperellipsoid.

In Section II we elaborate on the efficiency of the rejection method and re-derive the formula for the volume of a hypersphere. In Section III we utilize some of the results of Section II to develop the theory behind the algorithm. Illustrative analyses are given in Section IV, followed by conclusions in Section V.

## II. Sampling via Rejection Algorithms

### Problems with Rejection Algorithms

If  $\mathbf{x}$  is uniformly distributed in the hypercube  $(-1, 1)^d$ , and the  $\mathcal{L}^2$  norm,  $\|\mathbf{x}\|$ , of  $\mathbf{x}$  is less than 1, then  $\mathbf{X}$  is distributed uniformly in the unit hypersphere,  $C_d$ . Furthermore, if we denote volume of  $C_d$  by  $V(C_d)$ , then the acceptance rate,  $a(d)$ , of the corresponding rejection method is given by (Devroye, 1986)

$$a(d) = \frac{V(C_d)}{2^d} = \frac{\pi^{d/2}}{2^d \Gamma(\frac{d}{2} + 1)}.$$

For example,  $a(2) \approx 0.785$ ,  $a(3) \approx 0.524$ ,  $a(4) \approx 0.308$ ,  $a(5) \approx 0.164$ ,  $a(10) \approx 0.0025$ , and  $a(20) \approx 2.46 \times 10^{-8}$ . Alternatively, the mean number of variates,  $n(d)$ , required to generate one random vector in  $C_d$  is given by

$$n(d) = \frac{d}{a(d)} = \frac{d 2^d \Gamma(\frac{d}{2} + 1)}{\pi^{d/2}}.$$

For example, on average, about 4015 uniform variates are required to accept one random vector in  $C_{10}$ .

The computational efficiency of this algorithm may be enhanced by using the Cauchy-Schwarz inequality, i.e.,  $\|\mathbf{X}\| < \sum |X_i|$ , in a squeezing step (e.g., Devroye, 1986) to avoid

some of the more computationally expensive multiplications in favor of addition and sign changes. Unfortunately, the acceptance rate remains unchanged so that this savings is proportionately small for large  $d$ .

Since the computationally expensive part of this algorithm lies in the rejection step, minimizing the number of rejection steps is critical for large  $d$ . An obvious way to accomplish this is to accept from large batches of vectors. Ideally, one would want to batch generate at least  $m/a(d)$  vectors from  $(-1, 1)^d$ , where  $m$  is the number of variates from  $C_d$  that are still needed after the previous rejection step to reach the desired sample size. For example, if a sample size of 5000 was desired from  $C_d$ , then, on average,  $5000/a(d)$  random vectors would be required to retain the desired number after one rejection. If, after one rejection, 100 vectors are still needed, on average, this would require the generation of another  $100/a(d)$  random vectors to achieve this after the next rejection. Clearly, a limitation of this approach would be the availability of computing memory.

That being said, rejection methods are very inefficient and become unviable for large  $d$ .

### A Derivation of the Volume of the Unit Hypersphere

The volume of a  $d$  dimensional hypersphere can be obtained via repeated application of the Dirichlet integrals (see pages 132-133, Jeffreys, 1985). However, in order to obtain intermediate results for use in deriving the results of the next section, we re-derive the formula for  $V(C_d)$  as follows.

Suppose  $\mathbf{x} \in R^d$  such that  $\|\mathbf{x}\| = 1$ . Define the transformation  $[r, \boldsymbol{\theta}] \rightarrow \mathbf{x}$ , where  $r \in R$ ,  $\boldsymbol{\theta} \in R^{d-1}$ , by

$$\begin{aligned} x_1 &= r \cos(\theta_1) \prod_{j=2}^{d-1} \sin(\theta_j) \\ x_2 &= r \sin(\theta_1) \prod_{j=2}^{d-1} \sin(\theta_j) \\ x_3 &= r \cos(\theta_2) \prod_{j=3}^{d-1} \sin(\theta_j) \\ &\vdots \\ x_i &= r \cos(\theta_{i-1}) \prod_{j=i}^{d-1} \sin(\theta_j) \\ &\vdots \\ x_d &= r \cos(\theta_{d-1}). \end{aligned}$$

$\theta_i$  maybe be thought of as the angle that  $\mathbf{x}$  makes with the positive  $(i + 1)$ st coordinate axis. Note that this transformation is regular on the set for which  $r > 1$ ,  $0 < \theta_1 < 2\pi$ , and

$0 < \theta_i < \pi$ ,  $i > 1$ . The absolute value of the Jacobian of this transformation is

$$r^{d-1} \prod_{i=2}^{d-1} \sin^{i-1}(\theta_i) dr \prod_{i=2}^{d-1} d\theta_i.$$

Therefore, the volume inside the  $d$ -dimensional unit hypersphere,  $C_d$ , can be written as

$$V(C_d) = \int_0^\pi \cdots \int_0^\pi \int_0^{2\pi} \int_0^1 r^{d-1} \prod_{i=2}^{d-1} \sin^{i-1}(\theta_i) dr \prod_{i=1}^{d-1} d\theta_i.$$

Integrating (see, e.g., Selby, 1973),

$$\begin{aligned} V(C_d) &= \frac{2\pi}{d} \prod_{i=2}^{d-1} \int_0^\pi \sin^{i-1}(\theta_i) d\theta_i = \frac{2\pi}{d} \prod_{i=2}^{d-1} \left[ 2 \cdot \int_0^{\pi/2} \sin^{i-1}(\theta_i) d\theta_i \right] \\ &= \frac{2\pi}{d} \prod_{i=2}^{d-1} \sqrt{\pi} \frac{\Gamma(\frac{i}{2})}{\Gamma(\frac{i+1}{2})} = \frac{2\pi}{d} \cdot \frac{\pi^{\frac{d-1}{2}}}{\Gamma(\frac{d}{2})} \\ &= \frac{\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2} + 1)}. \end{aligned}$$

### III. Sampling via Transformations and the Gibbs Sampler

In this section, we prove that there exist  $d$  univariate densities such that, under the transformation defined in the previous section, a random vector of independent variables from these densities becomes a random vector uniformly distributed in  $C_d$ . We then proceed to provide viable methods by which one can simulate from these  $d$  univariate densities.

**Theorem:** *Let  $R, \Theta_1, \dots, \Theta_d$  be independently distributed such that they have the following density functions:*

$$\begin{aligned} f_R(r) &= dr^{d-1} \cdot I(0 < r < 1) \\ f_{\Theta_1}(\theta_1) &= \frac{1}{2\pi} \cdot I(0 < \theta_1 < 2\pi), \\ f_{\Theta_2}(\theta_2) &= \frac{2}{\pi} \sin(\theta_2) \cdot I(0 < \theta_2 < \pi), \\ &\vdots \\ f_{\Theta_i}(\theta_i) &= \frac{\sin^{i-1}(\theta_i)}{\int_0^\pi \sin^{i-1}(\theta_i) d\theta_i} \cdot I(0 < \theta_i < \pi). \\ &\vdots \\ f_{\Theta_{d-1}}(\theta_{d-1}) &= \frac{\sin^{d-2}(\theta_{d-1})}{\int_0^\pi \sin^{d-2}(\theta_{d-1}) d\theta_{d-1}} \cdot I(0 < \theta_{d-1} < \pi). \end{aligned}$$

Under the transformation defined in the previous section,  $\mathbf{X} = (X_1, \dots, X_d)$  is distributed uniformly in  $C_d$ .

**Proof:**

Note that the inverse transformation is defined by

$$\begin{aligned} r &= \sqrt{\sum_{i=1}^d x_i^2} = \|\mathbf{x}\| \\ \theta_1 &= \tan^{-1}(x_2/x_1) \\ \theta_2 &= \cos^{-1}\left(x_3/\sqrt{x_1^2 + x_2^2 + x_3^2}\right) = \cos^{-1}(x_3/\|\mathbf{x}_3\|) \\ &\vdots \\ \theta_i &= \cos^{-1}\left(x_{i+1}/\sqrt{\sum_{j=1}^i x_j^2}\right) = \cos^{-1}(x_{i+1}/\|\mathbf{x}_{i+1}\|) \\ &\vdots \\ \theta_{d-1} &= \cos^{-1}\left(x_d/\sqrt{\sum_{i=1}^d x_i^2}\right) = \cos^{-1}(x_d/\|\mathbf{x}\|). \end{aligned}$$

This is obvious if one thinks of  $x_i$  as the projection of  $\mathbf{x}$  onto the positive  $i$ th axis, whence  $\cos(\theta_{i-1}) = x_i/\|\mathbf{x}\|$ .

Because the original transformation is regular on the support under consideration, from the previous section, we see that the absolute value of the Jacobian of the inverse transformation is just

$$\frac{1}{r(\mathbf{X})^{d-1} \prod_{i=2}^{d-1} \sin^{i-1}(\theta_i(\mathbf{X}))}.$$

Therefore the joint density function of  $X$  is

$$\begin{aligned} f(\mathbf{x}) &= \frac{1}{\|\mathbf{x}\|^{d-1} \prod_{i=3}^d (\|\mathbf{x}_i\|/\|\mathbf{x}\|)^{i-2}} \cdot d\|\mathbf{x}\|^{d-1} I(0 < \|\mathbf{x}\| < 1) \\ &\quad \cdot \frac{1}{2\pi} I(0 < \tan^{-1}\left(\frac{x_1}{x_2}\right) < 2\pi) \cdot \prod_{i=3}^{d-1} \frac{(\|\mathbf{x}_i\|/\|\mathbf{x}\|)^{i-2}}{\int_0^\pi \sin^{i-1}(\theta_i) d\theta_i} I(0 < \cos^{-1}\left(\frac{x_i}{\|\mathbf{x}\|}\right) < \pi), \\ &= \frac{d}{2\pi \prod_{i=3}^{d-1} \int_0^\pi \sin^{i-1}(\theta_i) d\theta_i} I(0 < \|\mathbf{x}\| < 1) \end{aligned}$$

so that, by the previous section,

$$f(\mathbf{X}) = \frac{1}{V(C_d)} I(0 < \|\mathbf{x}\| < 1),$$

whence  $\mathbf{X}$  is uniformly distributed in  $C_d$ . ■

Clearly,  $\Theta_1 \sim U(0, 2\pi)$ , i.e.,  $\Theta_1$  is uniformly distributed on  $(0, 2\pi)$ . Note that one can readily obtain and invert the cdf of  $R$ , whence it follows that, if  $U \sim U(0, 1)$ , then

$U^{\frac{1}{d}} \sim R$ . Similarly, if  $U \sim U(0, 1)$ , then  $\cos^{-1}(1 - 2U) \sim \Theta_2$ . Thus, for  $d \leq 3$  a uniform random vector in  $C_d$  can be obtained directly from  $d$  univariate uniform random variables by means of simple transformations.

Unfortunately, for  $n \geq 2$ , it is not possible to invert the distribution function when  $f(\theta) \propto \sin^n(\theta) \cdot I(0 < \theta < \pi)$ . If the desired sample size is small, a viable option is to generate such random variables by, say, rejection from suitable envelope functions (Devroye, 1986), e.g., truncated normals or betas on  $(0, \pi)$  with mean  $\pi/2$  and suitable variance.

Since the  $\sin^n(\theta)$  is an invertible function for all  $n$ , it is possible to introduce a latent variable (Damien and Walker, 1996)  $Y$  such that the conditional densities of the joint density of  $Y$  and  $\Theta$  are uniform and the marginal density of  $\Theta$  is proportional to  $\sin^n(\theta)$ . Therefore, if the desired sample size merits the cost of convergence, we may sample quite efficiently and conveniently from densities of this form via Gibbs sampling (Smith and Roberts, 1993) as follows:

Given  $f(\theta) \propto \sin^n(\theta) \cdot I(0 < \theta < \pi)$ ,  $n \geq 2$ , introduce the latent variable  $Y$  such the joint density of  $\Theta$  and  $Y$  is given by

$$f(\theta, y) \propto I(0 < y < \sin^n(\theta)) \cdot I(0 < \theta < \pi).$$

Note that the marginal density of  $\theta$  is the target density, and conditional densities, given by

$$\begin{aligned} [Y|\Theta = \theta] &\sim U(0 < y < \sin^n(\theta)) \\ [\Theta|Y = y] &\sim U(\sin^{-1}(y^{1/n}) < \theta < \pi - \sin^{-1}(y^{1/n})), \end{aligned}$$

are just uniform densities.

## IV. Illustrative Analyses

The algorithm developed in the previous section and a straight rejection algorithm were implemented in SAS/IML and run for various  $d$  on an HP 9000/735 workstation. A final sample size of 5000 random vectors ( $5000d$  random variates) was generated in each case.

In the case of the rejection method, the sample matrix was initialized, without loss of generality, as a  $5000 \times d$  matrix of uniform variates from  $(-1, 1)$ , with one for one replacement upon rejection. That is, if  $k$  of the 5000 vectors were rejected, then  $k \cdot d$  more variates were generated from  $(-1, 1)$ , and so on until all 5000 vectors satisfied the acceptance criterion.

In the case of the transformations/Gibbs algorithm, 5000 random variates were simulated independently for each component density. Wherever the Gibbs sampler was used, after experimenting with various simulations, convergence diagnostics not reported here showed a burn-in of 50 would suffice. Note that for  $d = 2$  and  $d = 3$  this method is one for one, hence  $5000d$  uniform random variates were required. If  $d > 3$ , it follows that  $15,000 + 10,100d$  uniform random variates were required, since each iteration of the Gibbs sampler requires two random variates. (Thus, for large  $d$  this algorithm requires approximately  $2d$  random uniform variates for each desired random vector. These values were tabulated in Table 1 for various values of  $d$ , and contrasted with the total expected number (Section II) of variates required to obtain an equivalent sample via the rejection



method. The experimentally observed run times were also obtained, whenever feasible, and estimated when not.

**Table 1:** Comparison of run times and number of variates involved for the rejection and transformations/Gibbs methods for various values of  $d$ .

		Straight Rejection		Transformations/Gibbs	
d	number of variates desired	run time (hh:mm:ss.ss)	ave. number of variates required	run time (hh:mm:ss.ss)	number of variates required
2	$1.0 \times 10^4$	0.53	$1.27 \times 10^4$	0.21	$1.00 \times 10^4$
3	$1.5 \times 10^4$	1.33	$2.86 \times 10^4$	0.42	$1.50 \times 10^4$
4	$2.0 \times 10^4$	2.98	$6.58 \times 10^4$	1.70	$2.51 \times 10^4$
5	$2.5 \times 10^4$	9.02	$1.52 \times 10^5$	3.04	$3.53 \times 10^4$
6	$3.0 \times 10^4$	20.55	$3.72 \times 10^5$	5.62	$4.53 \times 10^4$
7	$3.5 \times 10^4$	44.21	$9.48 \times 10^5$	6.92	$5.54 \times 10^4$
8	$4.0 \times 10^4$	2:07.57	$2.52 \times 10^6$	8.37	$6.55 \times 10^4$
9	$4.5 \times 10^4$	6:03.66	$6.98 \times 10^6$	9.86	$7.56 \times 10^4$
10	$5.0 \times 10^4$	21:22.35	$2.01 \times 10^7$	11.46	$8.57 \times 10^4$
15	$7.5 \times 10^4$	114 hrs *	$6.44 \times 10^9$	19.23	$1.36 \times 10^5$
20	$1.0 \times 10^5$	3000 days *	$4.06 \times 10^{12}$	27.23	$1.87 \times 10^5$
30	$1.5 \times 10^5$	$1.5 \times 10^7$ yrs *	$7.35 \times 10^{18}$	44.72	$2.88 \times 10^5$
50	$2.5 \times 10^5$	$3.3 \times 10^{21}$ yrs *	$1.63 \times 10^{33}$	1:18.49	$4.90 \times 10^4$
100	$5.0 \times 10^5$	$5.4 \times 10^{63}$ yrs *	$2.67 \times 10^{75}$	2:53.45	$9.95 \times 10^5$

\* Linearly extrapolated from actual run times as a function of expected number of required variates.

## V. Conclusions

In this paper a simple algorithm to generate random variates from a hypersphere was developed. The algorithm was implemented for various  $d$ , the dimension of the hypersphere. It was shown that the algorithm is more efficient than rejection sampling for any  $d$ , especially so for large  $d$ , in which case rejection sampling becomes impractical.

## Acknowledgements

The authors wish to thank Adrian Smith, Imperial College, London, for helpful discussions.

## References

- Chib, S. and Greenberg, E., "Understanding the Metropolis-Hastings algorithm," *The American Statistician*, **49**, 327-335, 1995.
- Damien P., and Walker, S.G., "Sampling probability densities via uniform random variables and a Gibbs sampler," (Submitted for publication) 1996.
- Devroye, L., *Non-Uniform Random Variate Generation.*, Springer, New York, 1986.

Jeffreys, H., *Theory of Probability*, Oxford University Press, New York, 1985.

Selby, Samuel M., *Standard Mathematical Tables*, CRC Press, Cleveland, Ohio, 1973.

Smith, A.F.M., and Roberts, G.O., "Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods," *JRSS-B*, **55**, 3-24, 1993.