Proceedings of the 1996 Winter Simulation Conference
ed. J. M. Charnes, D. J. Morrice, D. T. Brunner, and J. J. Swain

THE HIT-AND-RUN SAMPLER: A GLOBALLY REACHING MARKOV CHAIN SAMPLER FOR GENERATING ARBITRARY MULTIVARIATE DISTRIBUTIONS

Robert L. Smith

Department of Industrial and Operations Engineering
The University of Michigan
Ann Arbor, Michigan 48109, U.S.A.

ABSTRACT

The problem of efficiently generating general multivariate densities via a Monte Carlo procedure has experienced dramatic progress in recent years through the device of a Markov chain sampler. This procedure produces a sequence of random deviates corresponding to a random walk over the support of the target distribution. Under certain regularity conditions, the corresponding Markov chain converges in distribution to the target distribution. Thus the sample of points so generated can serve as a statistical sample of points drawn from the target distribution. A random walk that can globally reach across the support of the distribution in one step is called a Hit-and-Run sampler. Hit-and-Run Markov chain samplers offer the promise of faster convergence to the target distribution than conventional small step random walks. Applications to optimization are considered.

1 INTRODUCTION

The Monte Carlo problem of efficiently generating univariate distributions can be attacked from a variety of directions including the simple device of evaluating the inverse of the random variables cdf at a uniform [0,1] deviate. However, the multivariate case is more problematic with no corresponding transform method for the general case (Schmeiser (1980)). There are transform methods for special cases, including multivariate normals and uniform distributions over simplices and ellipsoids (Rubinstein (1982)), but no constant Jacobian analytic function is known that maps from the unit cube to the epigraph of a general multivariate density function. Nonetheless, the ability to generate general multivariate deviates would have many important applications. The simulation of complex systems can involve generating multivariate densities that are not only not normal but may not belong to a named family of distributions. An interesting and potentially very significant application from the field of optimization is to generate samples from multivariate distributions that have a high likelihood of being close to a global optimum.

The basic idea is very intuitive. Consider the problem of minimizing a continuous objective function f over a unit cube C in R^n,

$$\max_{x \in C \subseteq R^n} f(x)$$

If we were to generate a point X uniformly distributed in the epigraph of f on C, i.e., in $S = \{(x, f(x)) | x \in C\}$, then points $x \in C$ with higher objective function values $f(x)$ would be more likely to be generated than those with lower values. In fact, the pdf of X would be a scalar multiple of f. We could increase this bias toward higher values of f by selecting X uniform from the epigraph of a transformed version of f that made higher values of f correspondingly more likely. For example, consider the epigraph of the function $e^{f(x)/T}$ where $T > 0$. It is clear that as $T \to \infty$, the distribution of X would converge to the singular distribution of unit mass at the global optimum $x^* = \text{argmax}\{f(x) | x \in C\}$. This observation in fact forms the basis for an extension of simulated annealing to the continuous case (Romeijn and Smith (1994a,1994b)). However since the objective function f is arbitrary apart from being continuous, the epigraph of the function $g(x) = e^{f(x)/T}$ can also be quite general and in particular non-convex.

A straightforward approach to generating a uniform point from such a general bounded region $S \subseteq R^n$ would be to enclose it in a sphere say and then use Von Neumann's rejection method (Hammersley and Handscomb (1964)) by generating a sequence of independent uniform deviates in the sphere until a point falls in S. Unfortunately, the number of such trial points grows exponentially fast in dimension n even for nice regions S such as a cube. For example, the expected number of points that must be generated within a circumscribed sphere around a cube S

grows from 1.5 for $n = 2$ to 10^{30} for dimension $n = 100$. (The n-dimensional content of an n-dimensional sphere of radius r is $2r^n\pi^{n/2}(n\Gamma(n/2))^{-1}$, so that the volume of the sphere enclosing an n-dimensional unit cube is $2(n\pi/4)^{n/2}(n\Gamma(n/2))^{-1}$).

Lacking either an efficient transformation or rejection method of generating a uniform deviate in a general region S, people have turned to the device of a so-called Markov chain sampler. Probably the first instance of such a Monte Carlo procedure was the celebrated Metropolis Method (Metropolis et al (1953)). The basic idea is to begin with any point in the region S and initiate a random walk within S from this starting point. Since the distribution of each successor point depends only upon the current point, the sequence of points generated form a homogeneous continuous state Markov chain. The generator that produces a successor point from the current point is so constructed that the limiting distribution of the Markov chain is the target distribution over S, e.g., the uniform distribution.

2 HIT-AND-RUN SAMPLERS

An important class of Markov chain samplers is the so-called Hit-and-Run sampler. They differ from other samplers, such as the neighborhood walk of discrete simulated annealing, by their ability to execute movements that can span the entire region S in a single step. This ability to globally reach across a region has led researchers to conjecture that their rate of convergence is superior to neighborhood random walks.

When the number of iterations of the walk required to obtain a variational distance from the target distribution of at most ϵ is a polynomial function of the size of the problem (e.g., the dimension of the region), then the sampler is called rapidly mixing. Recently, Lovasz and Simonovits (1993) have demonstrated a neighborhood walk that achieves rapid mixing for a class of "well rounded" convex regions. The walk proceeds by generating successor points within the intersection of the region S with a small ball around the current point. Surprisingly however, the polynomiality of Hit-and-Run remains an open question.

2.1 Uniform Target Distributions

We now turn to describing a series of Hit-and-Run samplers that have in practice experienced excellent rates of convergence to their target distributions. The first is the first instance of a Hit-and-Run sampler which was termed at the time a symmetric mixing algorithm (Smith (1984); see also Smith (1980), Boneh and Golan (1979)). It is only required that the region S be an open subset of R^n (and that the target distribution be the uniform distribution over S). In particular, S may be non-convex and indeed disconnected. However, since S is open, for every point $x \in S$ there must exist a small ball around it that lies totally within S.

Hit-and-Run Sampler (Uniform Target Distribution)

1. Select a starting point $x_0 \in S$ and set $i = 0$.
2. Generate a random direction d_{i+1} uniformly distributed over a direction set $D \subseteq R^n$. Find the line set $L = S \cap \{x | x = x_i + \lambda d_{i+1}, \lambda \text{ a real scalar}\}$ and generate a random point x_{i+1} uniformly distributed over L.
3. If $i = N$, stop. Otherwise, set $i = 1 + 1$ and return to 2.

Since the region S is arbitrary apart from being open, the step of generating x_{i+1} uniform over L can be computationally very demanding. To avoid this, in practice one would execute step 2 by having first enclosed the region S in a hypercube. Then after generating a uniform direction d over D (easily done by simply normalizing a vector of n independent normal [0,1] components), the end points y^1 and y^2 of the bidirectional line emanating from the current iterate x and intersecting the enclosing cube are determined. Then a rejection method is employed by generating one-dimensional uniform points on the line segment from y^1 to y^2 until one falls within L. Since this subproblem is one-dimensional, a rejection method is typically very efficient for this task. The remaining $(n-1)$-dimensional effort lies in choosing d uniform over D, an easy task for D a sphere for example. In this latter case the procedure is called the Hypersphere Directions Hit-and-Run Sampler.

The reason that the Markov chain corresponding to the iterates $X_0, X_1, X_2, \ldots$ converges in distribution to a uniform distribution over S is easily seen from the fact that 1) it is possible to go from any point in S to a neighborhood of any other point in one step, and 2) the uniform distribution is a stationary distribution of the chain. The latter observation is clear if we observe that from the directional symmetry of the direction distribution and next iterate chosen, the probability of transitioning from a point x to a neighborhood of a point y in S is the same as the probability of transitioning from the point y to a neighborhood of the point x, for all x and y in S (see Smith (1984) for a formal proof of this result). Smith (1984) reports on empirical tests that suggest

rapid convergence to a uniform distribution for relatively large dimensional regions. An analytic bound on the error from uniform after a fixed number of iterations is also provided. The bound is smallest for the case where the region is a sphere since the bound is decreasing in the ratio of the volume of the region to the volume of the smallest circumscribed sphere. Thus elongated ellipsoids can be expected to require more iterations to achieve a given closeness to uniformity. This is also intuitively clear since most of the directions generated from a current iterate will result in small movements from the current point (thus the conductance is small). Since the uniform distribution remains the limiting distribution regardless of how directions are generated (Smith (1984)), it can prove profitable to alter the distribution employed to better fit the region. Kaufman and Smith (1996) show how to select a direction distribution to accelerate convergence and in particular show how to achieve a sphere-case rate of convergence for any ellipsoid region. Incidentally, if we choose the direction distribution set D to be a hypercube, we obtain the so-called Coordinate Directions Hit-and-Run Sampler, a version of the Gibbs Sampler except that the sequence of coordinate directions chosen is not deterministic. See Chen and Schmeiser (1993) for a computational comparison of the two approaches.

Turning to applications, in addition to their direct use for the Monte Carlo generation of uniform samples, these Hit-and-Run Samplers with uniform target distributions have also been employed as random probes to discover the properties of various regions in R^n. Berbee et al (1987) explored their use in identifying redundant constraints in a linear program. See also Caron, Hlynka and McDonald (1993). The idea here is that the bidirectional line emanating from the current iterate x_i will hit two points on the boundary of the polyhedral feasible region of the LP, thus identifying two non-redundant constraints. By repeating for each iterate, we play a coupon-collectors problem of non-redundant constraint identification until all non-redundant constraints have been identified with a given probability. By then removing the redundant constraints, the resulting smaller LP may be more efficiently solved for an optimal solution (if the resulting solution is feasible for the original LP, it must be optimal as well). The construction of a stopping rule that guarantees that every non-redundant constraint has been identified with a certain probability is complicated by the fact that the hit point on the boundary of the polyhedron will not in general be uniformly distributed on that surface (see Caron and MacDonald (1989) for how to address the dependency of the two hit points). In fact, the distribution will in general be quite irregular and complexly dependent on the geometry even though the interior iterates $X_0, X_1, X_2, \ldots$ will asymptotically be uniform within the polyhedron. The problem is that the probability of hitting a neighborhood of a facet depends on the angle the facet makes with the bidirectional line emanating from the current iterate: the more orthogonal the line meeting the facet, the better the chance. Boender et al (1991) alter the Hit-and-Run Sampler to reject hit points in proportion to the cosine of that angle to remove this bias and thereby produce hit points that are asymptotically uniform on the boundary of the polyhedron. Since the volume of the surface of a polyhedron is zero, rejection methods cannot be used here; nor can transformational methods since the number of simplices in a decomposition of the surface is exponentially large in the dimension of the polyhedron. The resulting Hit-and-Run sampler, somewhat whimsically called the Shake-and-Bake algorithm, is arguably the only practical way to generate uniform samples on the surface of large dimensional polyhedrons.

2.2 General Target Distributions

Thus far we have only discussed Hit-and-Run Samplers for generating *uniform* distributions on regions $S \subseteq R^n$. The problem of devising a Markov chain sampler to generate essentially arbitrary distributions f over $S \subseteq R^n$ is addressed in Belisle, Romeijn, and Smith (1993). They propose the following sampler:

Hit-and-Run Sampler (General Target Distribution)

1. Select a starting point $x_0 \in S$ and set $i = 0$.
2. Generate a direction d_{i+1} in D with distribution ν. Find the line set $L = S \cap \{x | x = x_i + \lambda d_{i+1}, \lambda \text{ a real scalar}\}$ and generate a point $x_{i+1} = x_i + \lambda_i d_{i+1} \in L$ with λ_i having the density
$$f_i(\lambda) = \frac{f(x_i + \lambda d_{i+1})}{\int f(x_i + r d_{i+1}) dr}.$$
3. If $i = N$, stop. Otherwise return to 2.

The sampler for arbitrary densities f over S then proceeds as in uniform Hit-and-Run except that the next iterate x_{i+1} is chosen according to the conditionalized density of f in the direction d_{i+1}. The direction distribution ν needs to span R^n but is otherwise arbitrary. In Belisle, Romeijn, and Smith (1993), it is shown that for S a bounded open set in R^n and f the pdf of the target distribution π, the sequence of

iterates $X_0, X_1, X_2, \ldots$ converges in total variation to the target distribution π.

Romeijn and Smith (1994a) extend this sampler to a Metropolis type sampler that also converges to arbitrary target distributions. They utilize the sampler within a simulated annealing type optimal search procedure to generate densities that come from a temperature parametrized family of Boltzman distributions to converge to a global optimal solution. See also Zabinsky et al (1993) for a more direct use of Hit-and-Run within an optimization framework.

3 CONCLUSION

The field of Markov Chain Samplers is currently quite active both from the theoretical and practical points of view. At least some of the interest stems from the recent results that random procedures can achieve polynomial performance for tasks such as approximating the volume of convex bodies that it has been demonstrated no deterministic procedure can ever achieve (Dyer and Frieze (1991), Dyer, Frieze, and Kannan (1991)). However in a larger context the problem of generating general multivariate distributions is a fundamental problem of simulation whose importance hardly needs justification.

REFERENCES

Bélisle, C.J.P., H.E. Romeijn and R.L. Smith, "Hit-and-Run Algorithms for Generating Multivariate Distributions," *Mathematics of Operations Research* 18, pp. 255–266,1993.

Berbee, H.C.P., C.G.E. Boender, A.H.G. Rinnooy Kan, C.L. Scheffer, R.L. Smith, and J. Telgen "Hit-and-Run Algorithms for the Identification of Nonredundant Linear Inequalities," *Mathematical Programming* 37, pp. 184–207,1987.

Boender, C.G.E., R.J. Caron, A.H.G. Rinnooy Kan, J.F. McDonald, H. Edwin Romeijn, R.L. Smith, J. Telgen, and A.C.F. Vorst,"Shake-and-Bake Algorithms for Generating Uniform Points on the Boundary of Bounded Polyhedra," *Operations Research* 39, pp. 935-953, 1991.

Boneh, A., and A. Golan, "Constraints' Redundancy and Feasible Region Boundedness by Random Feasible Point Generator (RFPG)," Third European Congress on Operations Research (EURO III), Amsterdam, 1979.

Caron, R. J. and J. F. McDonald, "A New Approach to the Analysis of Random Methods for Detecting Necessary Linear Inequality Constraints," *Mathematical Programming* 43, pp. 97-102, 1989.

Caron, R. J., Hlynka, M., McDonald, J. F., "On the Best Case Performance of Hit and Run Methods for Detecting Necessary Constraints," *Mathematical Programming* 54, pp. 233–249, 1992.

Chen, M. and B. Schmeiser, "Performance of the Gibbs, Hit-and-Run and Metropolis Samplers," *Journal of Computational and Graphical Statistics* 2, pp. 251–272, 1993.

Dyer, M and Frieze, A., "Computing the Volume of Convex Bodies: A Case Where Randomness Provably Helps." *Probabilistic combinatorics and its applications*, San Francisco, pp. 123–169, 1991.

Dyer, M, Frieze, A., and Kannan, R., "A Random Polynomial-time Algorithm for Approximating the Volume of Convex Bodies," *Journal of the ACM* 38, pp. 1–17. 1991.

Hammersley, J.M. and Handscomb, D.C., *Monte Carlo Methods*, John Wiley and Sons, NY, 1964.

Kaufman, D. and R. L. Smith, "Direction Choice For Accelerated Convergence In Hit-and-run Sampling," *Operations Research*, in press, 1996.

Lovasz, L. and Simonovits, M., "Random Walks in a Convex Body and an Improved Volume Algorithm," *Random Stuctures and Algorithms* 4, pp. 359-412, 1993.

Metropolis, N., A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller and E. Teller, "Equations of State Calculations by Fast Computing Machines," *The Journal of Chemical Physics* 21, pp. 1087–1092, 1953.

Romeijn, E. and R.L. Smith, "Simulated Annealing for Constrained Global Optimization," *Journal of Global Optimization* 5, pp. 101-126, 1994a.

Romeijn, E. and R.L. Smith, "Simulated Annealing and Adaptive Search in Global Optimization," *Probability in the Engineering and Informational Sciences* 8, pp. 571-590, 1994b.

Rubinstein, R.Y., "Generating Random Vectors Uniformly Distributed Inside and on the Surface of Different Regions," *European Journal of Operations Research* 10, pp. 205-209, 1982.

Schmeiser, B.W., "Random Variate Generation: A Survey". *Simulation with Discrete Models: State of the Art View*, T. I. Oren, C.M. Shub, P. F. Roth (eds), IEEE, pp. 79-104, 1980.

Smith, R.L., "Monte Carlo Procedures for Generating Random Feasible Solutions to Mathematical Programs," ORSA/TIMS Conference, Washington, May, 1980.

Smith, R.L. Efficient Monte Carlo Procedures for Generating Points Uniformly Distributed over Bounded Regions. *Operations Research* 32, pp. 1296–1308,1984.

Zabinsky, Z.B., R.L. Smith, J.F. McDonald, H.E. Romeijn, and D.E. Kaufman, "Improving Hit and

Run for Global Optimization," *Journal of Global Optimization* 3, pp. 171–192, 1993.

AUTHOR BIOGRAPHIES

ROBERT L. SMITH is Professor of Industrial and Operations Engineering at the University of Michigan, Ann Arbor. He received his Ph.D. in Engineering Science from the University of California at Berkeley where he was an NSF Fellow. He holds a bachelors degree in Physics from Harvey Mudd College and an MBA from Berkeley. He has been a visiting scholar at the Statistical Laboratory, University of Cambridge; the Econometric Institute, Erasmus University, Rotterdam, The Netherlands; the Laboratoire d'Automatique et d'Analyse du Systems du CNRS, Toulouse, France; the Faculty of Industrial Engineering and Management, Technion, Haifa, Israel; the Mathematical Sciences Research Institute, Berkeley, California; the Operations Research Group, Institute of Pure and Applied Mathematics, Rio de Janeiro, Brazil; and the Department of Industrial Engineering and Operations Research, University of California, Berkeley. His research is directed toward the modeling, analysis, and optimization of dynamical systems over time. He is a member of the Institute for Operations Research and Management Sciences and The Mathematical Programming Society, an Associate Editor of *Management Science*, and is the author of over sixty peer reviewed articles.