Stochastic Methods for Multiple Integrals over Unbounded Regions *

Alan Genz †
Department of Mathematics
Washington State University
Pullman, WA 99164-3113
alangenz@wsu.edu

Abstract

Recent work in the development of stochastic methods for multiple integrals over unbounded regions is reviewed and generalized. This includes randomization of deterministic rules, and new stochastic rules for integrals with multivariate Normal weight. Stochastic spherical-radial rules will also be discussed. These rules use a spherical-radial transformation of the infinite integration region and combine stochastic rules for the infinite radial interval with stochastic rules for the spherical surface. Example problems taken from Bayesian statistical analysis and computational finance are used to illustrate the use of the different methods.

1 Introduction

An important applications problem is to numerically compute integrals in the form

$$I(f) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} w(\mathbf{x}) f(\mathbf{x}) dx_1 dx_2 \cdots dx_n.$$

with $\mathbf{x} = (x_1, x_2, ..., x_n)^t$. This paper focuses on the case where w is the multivariate Normal weight $w(\mathbf{x}) = (\sqrt{2\pi})^{-n} e^{-\mathbf{x}^T\mathbf{x}/2}$, but all of the methods described also have generalizations for the multivariate Student-t weight with ν degree of freedom $w(\mathbf{x}) = \Gamma(\frac{\nu+n}{2})/(\Gamma(\frac{\nu}{2})(\sqrt{\nu\pi})^n)(1 + \mathbf{x}^T\mathbf{x}/\nu)^{-(\nu+n)/2}$. A large set of applications problems comes from the area of statistics where these integrals often arise in the form

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(\boldsymbol{\theta}\ ) p(\boldsymbol{\theta}\ ) d\theta_1 d\theta_2 \cdots d\theta_n.$$

The function $p(\boldsymbol{\theta}\ )$ is often an unnormalized unimodal posterior density function and $g(\boldsymbol{\theta}\ )$ is some function for which an approximate expected value is needed. A common assumption is that $p(\boldsymbol{\theta}\ )$ is

*Revised version published in *Mathematics and Computers in Simulation* **47** (1998), 287–298.
†Supported by NATO Collaborative Research Grant CRG 940139 and Flemish FWO and US NSF grants.

approximately multivariate Normal ($\boldsymbol{\theta} \sim N_n(\boldsymbol{\mu}, \Sigma)$) or multivariate Student-t ($\boldsymbol{\theta} \sim t_{(n,\nu)}(\boldsymbol{\mu}, \Sigma)$). In these cases, a standardizing transformation in the form $\boldsymbol{\theta} = \boldsymbol{\mu} + C\mathbf{x}$ can be determined (possibly using numerical optimization), where $\boldsymbol{\mu}$ is the point where $\log(p(\boldsymbol{\theta}))$ is maximized, Σ is the inverse of the negative of the Hessian matrix for $\log(p(\boldsymbol{\theta}))$ at $\boldsymbol{\mu}$, and C is the lower triangular Cholesky factor for Σ ($\Sigma = CC^t$). The transformed integrals then take the $I(f)$ form with $f(\mathbf{x}) = |C|g(\boldsymbol{\mu} + C\mathbf{x})p(\boldsymbol{\mu} + C\mathbf{x})/w(\mathbf{x})$. There has been much recent interest (see for example, Joy, Boyle and Tan (1996)) in integrals in the $I(f)$ form that come from computational finance applications. In this case the Normal weight, which arises from the modeling of a Wiener process, is very common.

This type of integration problem has traditionally been handled using Monte Carlo algorithms (see the book by Davis and Rabinowitz, 1984, and the more recent paper by Evans and Swartz, 1995). A simple Monte Carlo algorithm for estimating $I(f)$ might use

$$I(f) \approx I_N = \frac{1}{N} \sum_{i=1}^{N} f(\mathbf{x}_i),$$

with the points $\mathbf{x}_i$ randomly chosen with probability density proportional to $w(\mathbf{x})$. This Monte Carlo algorithm, which is an importance sampling algorithm for the original problem of estimating $I(f)$, is often effective, but in cases where the resulting $f(\mathbf{x})$ is not approximately constant, the algorithm can have low accuracy and slow convergence. However, an important feature of simple Monte Carlo algorithms is the availability of practical and robust error estimates. If we let σ denote the standard error for the sample, then

$$\sigma = \Big(\sum_{i=1}^{N} \frac{(f(\mathbf{x}_i) - I_N)^2}{N(N-1)} \Big)^{\frac{1}{2}},$$

and $\text{Prob}(|I(f) - I_N| < \alpha\sigma) \approx \int_{-\alpha}^{\alpha} \frac{e^{-t^2/2}}{\sqrt{2\pi}} dt$.

The primary purpose of this paper is to describe integration rules for $I(f)$ that have higher accuracy and better convergence properties than the simple Monte Carlo methods, but also have practical and robust error estimates. The basic idea is to randomize some deterministic integration rules that have higher degrees of accuracy than simple Monte Carlo rules. This randomization also allows the construction of simple Monte Carlo error estimates for the new stochastic rules. The radomization of two families of deterministic rules will be described. The first family, developed by Genz and Kiester (1996), consists of an efficient sequence of fully symmetric interpolatory rules for $I(f)$. The randomization uses a generalization of a method described by Haber (1969). These rules are available for high degree approximations to $I(f)$, but computations using these rules are currently feasible for only $1 \leq n \leq 15$, approximately. The second family consists of rules constructed by products of rules for the hyper-spherical surface and rules for the appropriately weighted interval $(-\infty, \infty)$. Randomization of these rules was first described by Genz and Monahan (1996) and Monahan and Genz (1997). This paper describes these new stochastic rules with some generalizations, and considers the use of these rules for high-dimensional problems where $1 \leq n \leq 1000$.

2 Interpolatory Stochastic Rules

2.1 Background

The deterministic fully symmetric interpolatory rules $Q^{(m,n)}(f)$ developed by Genz (1986) take the form

$$Q^{(m,n)}(f) = \sum_{|\mathbf{p}| \le m} w_{\mathbf{p}} f[\boldsymbol{\lambda}_{\mathbf{p}}].$$

Here $\boldsymbol{\lambda}_{\mathbf{p}} = (\lambda_{p1}, \lambda_{p_2}, ..., \lambda_{p_n})$, $|\mathbf{p}| = p_1 + p_2 + \cdots + p_n$ and the normalized fully symmetric sums $f[\boldsymbol{\lambda}_{\mathbf{p}}]$ are defined by

$$f[\boldsymbol{\lambda}_{\mathbf{p}}] = 2^{-\{\mathbf{p}\}} \sum_{\mathbf{s}} f(s_1 \lambda_{p_1}, s_2 \lambda_{p_2}, ... s_n \lambda_{p_n}),$$

where $\{\mathbf{p}\}$ is the number of nonzero components in $\mathbf{p}$, and the sum is taken over all of the sign combinations that occur when $s_i = \pm i$, for those i with $\lambda_i \neq 0$. The $\lambda' s$ (*generators*) are assumed to be distinct with $\lambda_0 = 0$, and the weights $w_{\mathbf{p}}$ are given by

$$w_{\mathbf{p}} = \sum_{|\mathbf{k}| \le m - |\mathbf{p}|} \prod_{i=1}^{n} \frac{a_{k_i + p_i}}{\prod_{j=0, \neq p_i}^{k_i + p_i} (\lambda_{p_i}^2 - \lambda_j^2)},$$

where

$$a_i = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} \prod_{j=0}^{i-1} (x^2 - \lambda_j^2) dx,$$

for $i > 0$, with $a_0 = 1$. Given these definitions and assumptions, the theory for fully symmetric interpolatory rules developed by Genz (1986) shows that $Q^{(m,n)}(f)$ is an approximation to $I(f)$ (assuming multivariate Normal weight) that is exact for all polynomials with degree $2m+1$ or less. Genz and Kiester (1996) showed how to choose the generators so that the some of the weights $w_{\mathbf{p}}$ are zero, thereby reducing significantly the cost for computing $Q^{(m,n)}(f)$ for larger values of m and n.

In order to construct randomized $Q^{(m,n)}(f)$ rules, the randomization method of Haber (1969) for interpolatory rules is first described. Let $P_d(\mathbf{x})$ be a degree d interpolating polynomial for $f(\mathbf{x})$ on some set of points $\{\mathbf{x}_k\}$, $k = 1, 2, ..., \nu_d$. Then $I(P_d)$, which can be written as a weighted sum of f values at the interpolation points, is a degree d interpolatory rule for $I(f)$. Consider the modified rule

$$r_d(f, \mathbf{y}) = f(\mathbf{y}) - P_d(\mathbf{y}) + I(P_d).$$

If f is a polynomial of degree d or less, then $f(\mathbf{y}) = P_d(\mathbf{y})$ for all $\mathbf{y}$, so $r_d(f, \mathbf{y})) = I(P_d) = I(f)$, is also an integration rule for $I(f)$ of polynomial degree d. The expected value of $r_d(f, \mathbf{y})$ is simply $I(f) - I(P_d) + I(P_d) = I(f)$. If a random set of points $\{\mathbf{y}_k\}$ is chosen from the integration region $H = (-\infty, \infty)^m$, with density $w(\mathbf{x})$, the sum

$$R_d(f) = \frac{1}{N} \sum_{k=1}^{N} r_d(f, \mathbf{y}_k),$$

defines a degree d "stochastic" rule. This rule is an unbiased estimator for $I(f)$. An unbiased robust degree d error estimate $R_d(f)$ is provided by the Monte Carlo standard error

$$\hat{E}_d(f) = \Big(\frac{1}{N(N-1)} \sum_{k=1}^{N} (r_d(f, \mathbf{y}_k) - R_d(f))^2\Big)^{\frac{1}{2}}.$$

This method for randomizing a polynomial rule is a type of "control variates" method (see Davis and Rabinowitz, 1984, p. 389) for reducing variance. For a smooth function, the interpolating polynomial $P_d(\mathbf{y})$ can provide a good model for $f(\mathbf{y})$, so $\hat{E}_d(f)$ is expected to be much smaller than the standard error for a simple Monte Carlo method.

These stochastic rules proposed by Haber are based on simple interpolatory rules for $I(f)$. However, the most efficient polynomial rules for multivariate integration are usually fully symmetric rules, and the next subsection describes how Haber's method can be generalized for fully symmetric interpolatory rules.

2.2 Fully Symmetric Rules

The $Q^{(m,n)}(f)$ rules are constructed by integration of the even power Lagrange interpolation polynomials $M_m(f, \mathbf{x})$ defined by

$$M_m(f, \mathbf{x}) = \sum_{|\mathbf{p}| \le m} f[\boldsymbol{\lambda}_\mathbf{p}] M_{m,\mathbf{p}}(\mathbf{x}),$$

with

$$M_{m,\mathbf{p}}(\mathbf{x}) = \sum_{|\mathbf{k}| \le m - |\mathbf{p}|} \prod_{i=1}^{n} \frac{\prod_{j=0}^{k_i+p_i-1} (x_i^2 - \lambda_j^2)}{\prod_{j=0, \neq p_i}^{k_i+p_i} (\lambda_{p_i}^2 - \lambda_j^2)}.$$

The polynomials $M_m(f, \mathbf{x})$ satisfy the even monomial reproduction property $M_m(f, \mathbf{x}) = f(\mathbf{x})$, for all $\mathbf{x}$ whenever $f(\mathbf{x}) = \mathbf{x}^{2\mathbf{k}}$, and $|\mathbf{k}| \le m$. They also satisfy the property $M_m(f, \mathbf{x}) = f[\mathbf{x}] = 0$, for all $\mathbf{x}$ whenever $f(\mathbf{x}) = \mathbf{x}^{\mathbf{k}}$ and at least one component of $\mathbf{k}$ is odd. The weights $w_\mathbf{p}$ given previously are determined using $w_\mathbf{p} = I(M_{m,\mathbf{p}}(\mathbf{x}))$, so $Q^{(m,n)}(f) = I(M_m(f, \mathbf{x}))$.

Now define

$$r_m(f, \mathbf{y}) = f[\mathbf{y}] - M_m(f, \mathbf{y}) + I(M_m(f, \mathbf{y})).$$

Notice that if $f(\mathbf{x}) = \mathbf{x}^{2\mathbf{k}}$ and $|\mathbf{k}| \le m$, then $r_m(f, \mathbf{y}) = I(M_m(\mathbf{y}))$; and if $f(\mathbf{x}) = \mathbf{x}^{\mathbf{k}}$ with at least one component of $\mathbf{k}$ odd, then $r_m(f, \mathbf{y}) = I(M_m(\mathbf{y})) = 0$. These results show that $r_m(f, \mathbf{y})$ is an integration rule for $I(f)$ of polynomial degree $2m + 1$. The expected value for $r_m(f, \mathbf{y})$ is

$$I(r_m(f, \mathbf{y})) = I(f[\mathbf{y}]) - I(M_m(\mathbf{y})) + I(I(M_m(\mathbf{y}))) = I(f)$$

so, if N random points $\{\mathbf{y}_k\}$ are chosen from H with density $w(\mathbf{x})$, the sum

$$R_m(f) = \frac{1}{N} \sum_{k=1}^{N} r_m(f, \mathbf{y}_k)$$

is an unbiased degree $2m+1$ stochastic rule for $I(f)$, and an unbiased error estimate for $R_m(f)$ is provided by the Monte Carlo standard error

$$\hat{E}_m(f) = \Big(\frac{1}{N(N-1)} \sum_{k=1}^{N} (r_m(f, \mathbf{y}_k) - R_m(f))^2\Big)^{\frac{1}{2}}.$$

The primary extra computational cost for $R_m(f)$, compared to the cost of computing $Q^{(m,n)}(f)$, is the extra $2^n N$ f values needed for the N values of $f[\mathbf{y}]$, each of which require 2^n f values. For this reason, using these stochastic rules with currently available computing equipment becomes infeasible for large values of n (e.g. $n > 15$). For the larger values of n, it is of course possible to use Haber's original randomization with rules that are not fully symmetric. For a given polynomial degree the computational cost for a simple interpolatory rule is much higher than for a fully symmetric rule, but the randomization cost for the extra f values is only $O(N)$, instead of $O(2^n N)$. A simple interpolatory rule of degree d requires $\binom{n+d}{d}$ f values, so it is certainly feasible to use these rules for larger n values but small d values (e.g. $d \leq 3$), but this has not been carefully investigated, although the multivariate Lagrange interpolation theory developed by Genz (1982) could be used for the construction and implementation of simple interpolatory rules for $I(f)$. Another method for avoiding the 2^n integrand values cost for the $f[\mathbf{y}]$ sums is to replace the sums $f[\mathbf{y}]$ for larger n values with random samples of the terms in the sums, using fewer than 2^n terms. Rules with sampled $f[\mathbf{y}]$ sums would no longer have the polynomial exactness properties of the original $r_m(f, \mathbf{y})$ rules but the deviations from exactness might not be significant compared to the other approximation errors. This sampling strategy has not yet been carefully investigated. A completely different method for constructing feasible rules for large n is described in the next section; this section is concluded with an example.

2.3 An Example

This example is a seven dimensional proportional hazards model problem used by Dellaportas and Wright (1992). The unnormalized posterior probability density function is given by

$$p(\rho, \boldsymbol{\beta}\ ) = \prod_{i=1}^{48} \rho t_i^{\rho-1} e^{\mathbf{z}_i^T \boldsymbol{\beta}} \ \prod_{i=1}^{65} e^{-t_i^{\rho} e^{\mathbf{z}_i^T \boldsymbol{\beta}}}$$

with $\rho \in [0, \infty)$ and $\boldsymbol{\beta}\ \in (-\infty, \infty)^6$. The $\mathbf{z}_i$ vectors are data vectors from the original problem. After a transformation of ρ using $\theta_1 = \log(\rho)$, $e^{\theta_1} p(e^{\theta_1}, \boldsymbol{\beta}\ )$ is modeled with a multivariate Normal approximation (the extra e_1^{θ} factor arising because $d\rho = e^{\theta_1} d\theta_1$ for the transformed integral). To complete the modeling, $g(\boldsymbol{\theta}\ )$ is defined to be $e^{\theta_1} p(e^{\theta_1}, (\theta_2, \ldots, \theta_7))$ with the assumption that $g(\boldsymbol{\theta}\ ) \sim e^{-(\boldsymbol{\theta}\ -\boldsymbol{\mu}\ )^T \Sigma^{-1} (\boldsymbol{\theta}\ -\boldsymbol{\mu}\ )/2}$. Then the point $\boldsymbol{\mu}$ (the "mode"), where $\log(g(\boldsymbol{\theta}\ ))$ is maximimized, is determined numerically, and the Hessian matrix H for $\log(g(\boldsymbol{\theta}\ ))$ at $\boldsymbol{\mu}$ is also approximated numerically. Finally, the Cholesky factor C of the modal covariance matrix $\Sigma = -H^{-1}$ is computed, so that the final transformed integrand is $f(\mathbf{x}) = K e^{\mathbf{x}^T \mathbf{x}/2} g(\boldsymbol{\mu}\ + C\mathbf{x})$. The scaling parameter $K = e^{207.191}$, where $-207.191 = \max(g(\boldsymbol{\theta}\ ))$, is used to avoid underflow during the floating point computations. If the Normal model is good, then $f(\mathbf{x}) \approx 1$ and the final integral should be easy.

In the following table results are given for the $Q^{(m,n)}(f)$ rules for $N = 20$ samples for each degree value.

Table 1: Proportional Hazards Model Problem Results I

2m+1	f Vals.	$I(f)$	Std. Err.	$I(x_1 f)/I(f)$	Std. Err.	$I(x_2 f)/I(f)$	Std. Err.
9	1317	.00015165	(?)	0.13669	(?)	-3.98542	(?)
9	3877	.00015540	(.00000293)	0.13660	(.00266)	-3.99503	(.08822)
11	3837	.00015228	(?)	0.13513	(?)	-3.98060	(?)
11	6397	.00015135	(.00000048)	0.13452	(.00072)	-3.97523	(.01531)
13	9941	.00015201	(?)	0.13560	(?)	-3.98298	(?)
13	12501	.00015147	(.00000056)	0.13455	(.00103)	-3.97414	(.01904)
15	22725	.00015256	(?)	0.13436	(?)	-3.97876	(?)
15	25285	.00015241	(.00000013)	0.13401	(.00031)	-3.97859	(.00357)
17	48401	.00015204	(?)	0.13567	(?)	-3.98254	(?)
17	50961	.00015192	(.00000003)	0.13559	(.00008)	-3.98234	(.00103)

The columns of integral estimates were obtained using randomized $Q^{(m,n)}(f)$ rules with the Genz and Kiester generators. Exact values for these integrals are not known. The quantities in the fifth and seventh columns, denoted by $I(x_1 f)/I(f)$ and $I(x_2 f)/I(f)$, respectively, are estimates for the expected values for variables $x_1 = \log(\rho)$ and $x_2 = \beta_1$. For each row in the table, these estimates were determined by dividing the estimate for $I(x_i f)$ by the estimate for $I(f)$, for $i = 1, 2$. The numbers in the standard error (abbreviation: Std. Err.) columns six and eight were also scaled by $I(f)$ estimates. Each $2m + 1$ value has two rows associated with it. The first row for each $2m + 1$ value shows results without randomization, and the second shows results with randomization. The cost of the randomization for each $2m + 1$ value was $20 \times 2^7 = 2560$ f values. Without randomization, a standard method for estimating errors is to compare successive integral estimates for different degrees, which often requires untestable assumptions about the integrand. The question mark (?) entries in the alternate rows indicate this uncertainty in the errors for the unrandomized results. The standard errors given in the other rows could be used to provide confidence intervals for the expected errors with a confidence level of approximately 68%. Higher confidence levels could be determined using an appropriate scaling of the standard errors. The slight increase in standard errors between the results for the degree 11 and degree 13 randomized rules is due to the randomization.

3 Stochastic Spherical-Radial Rules

3.1 Background

In this section a change of variables to a spherical-radial coordinate system is used. Let $\mathbf{x} = r\mathbf{z}$, with $\mathbf{z}^t\mathbf{z} = 1$, so that $\mathbf{x}^t\mathbf{x} = r^2$, for $r \in [0, \infty)$. Then, for $n \geq 1$,

$$I(f) = \int_{\mathbf{z}^t\mathbf{z}=1} \int_0^\infty w(r) r^{n-1} f(r\mathbf{z}) dr d\mathbf{z}$$

$$= \frac{1}{2} \int_{\mathbf{z}^t\mathbf{z}=1} \int_{-\infty}^{\infty} w(r)|r|^{n-1} f(r\mathbf{z}) dr d\mathbf{z},$$

with $w(r) = (\sqrt{2\pi})^{-n} e^{-r^2/2}$. Integration rules for $I(f)$ can be constructed as products of rules for the radial interval $(-\infty, \infty)$ with weight $w(r)|r|^{n-1}$, and rules for the surface of the unit n-sphere, U_n. Averages of properly chosen samples of these rules can provide unbiased estimates for $I(f)$, and standard errors for the samples can be used to provide robust error estimates for the $I(f)$ estimates.

3.2 Radial Rules

The radial rules used by Genz and Monahan (1996) have the form

$$R(h) = \sum_{i=0}^{m} w_i (h(\rho_i) + h(-\rho_i))/2,$$

with $\rho_0 = 0$. These rules are designed to approximate the weighted infinite range integrals $T(h) = \int_{-\infty}^{\infty} |r|^{n-1} w(r) h(r) dr$. The weights are standard symmetric interpolatory weights defined by $w_i = T(\prod_{j=0,\neq i}^{n} \frac{r^2-\rho_j^2}{\rho_i^2-\rho_j^2})$, so that $R(h)$ has polynomial degree $2m+1$. Results in the Genz and Monahan paper generalize a method developed by Siegel and O'Brien (1985) to show how averages of samples of the rules $R(h)$ can be used to provide unbiased estimates for $T(h)$. For $m = 1$, the point ρ_1 needs to be chosen randomly with probability density proportional to $\rho_1^{n+1} e^{-\rho_1^2/2} \sim \text{Chi}(n+2)$ in order to provide unbiased degree three rules for $T(h)$. For $m = 2$, the points ρ_1 and ρ_2 need to be chosen randomly with joint probability density proportional to $(\rho_1\rho_2)^{n+1} e^{-(\rho_1^2+\rho_2^2)/2} (\rho_1 - \rho_2)^2 (\rho_1 + \rho_2)$, in order to provide unbiased degree five rules for $T(h)$. Genz and Monahan show that if r is randomly chosen from a $\text{Chi}(2n+7)$ density and q from a $\text{Beta}(n+2, \frac{3}{2})$ density, then $\rho_1 = r \sin(\frac{\sin^{-1}(q)}{2})$ and $\rho_2 = r \cos(\frac{\sin^{-1}(q)}{2})$ have the correct joint distribution. Higher degree rules can also be constructed, but the joint probability density functions for the ρ parameters for unbiased rules become more complicated and difficult to sample from.

A simpler method for randomizing a rule for $T(h)$, which avoids the ρ parameter sampling problems, is to use a control variates method. Given any symmetric $(2m+1)$-point rule $Q(h)$ for $T(h)$, a degree $2m$ interpolating polynomial $P_{2m}(r)$ for h can be determined using the values of h at the points for $Q(h)$. If t is chosen randomly from $(0, \infty)$ with density $t^{n-1} w(t)$, then the rule

$$r(h, t) = V((h(t) + h(-t))/2 - P_{2m}(t)) + Q(h),$$

where $V = T(1)$, has polynomial degree $2m+1$, and the stochastic rule

$$R(h) = \frac{1}{N} \sum_{k=1}^{N} r(h, t_k)$$

provides an unbiased degree $2m+1$ estimate for $T(h)$. An optimal $(2m+1)$-point polynomial rule for $T(h)$ is a Gauss rule with weight $|t|^{n-1} w(t)$, and such a rule would have polynomial degree

$4m+1$, but, unfortunately, a stochastic rule $R(h)$ constructed from this Gauss rule has polynomial degree $2m+1$ only, because the difference $((h(t)+h(-t))/2 - P_{2m}(t)$ is identically zero only when h is a polynomial of degree at most $2m+1$.

3.3 Spherical Rules

The stochastic spherical-radial rules derived by Genz and Monahan (1996) are based on spherical surface rules of the form

$$S_Q(s) = \sum_{j=1}^{p} \tilde{w}_j s(Q\mathbf{z}_j).$$

Here Q is an $n \times n$ orthogonal matrix, and the $\tilde{w}_j$'s and $\mathbf{z}_j$'s are the respective weights and points for some integration rule for a function $s(\mathbf{z})$ over U_n. Efficient spherical surface rules for degree as high as seven are given in the books by Stroud (1972) and Mysovskikh (1981), and the review paper by Mysovskikh (1980). Grabner and Tichy (1993) describe some spherical rules with equal weights. If Q is a random orthogonal matrix chosen with Haar distribution from the set of all matrices in the orthogonal group (see Stewart, 1980) and the weights and points for S_Q are weights and points for a rule of degree d, then averages of rules $S_{Q_i}(s)$ provide unbiased degree d estimates for integrals over U_n.

3.4 Spherical-Radial Rules

A stochastic spherical-radial rule can easily be constructed using a product of a spherical surface rule and symmetric radial rule in the form

$$SR_{Q,\boldsymbol{\rho}}\ (f) = \frac{1}{2}\sum_{j=1}^{p} \tilde{w}_j \sum_{i=1}^{m} w_i (f(-\rho_i Q\mathbf{z}_j) + f(\rho_i Q\mathbf{z}_j))/2.$$

If the spherical surface and radial rules both have degree d, and the Q matrix and $\boldsymbol{\rho}$ parameters are randomly chosen with appropriate densities, then $SR_{Q,\boldsymbol{\rho}}\ (f)$ is an unbiased degree d estimate for $I(f)$ and averages of samples of these rules are stochastic rules for $I(f)$. Error estimates can be easily constructed using standard errors for these samples.

For large n, a possibly significant overhead cost for the SR rules is the generation of the random orthogonal matrices. Using the algorithm given by Stewart (1980), it can be shown that the cost for generating one such matrix Q is approximately $4n^3/3$ floating point operations (flops) plus the cost of generating $n^2/2$ Normal(0,1) random numbers. For the Genz and Monahan (1996) degree three SR rules the columns of Q are used for the evaluation points for $O(n)$ integrand values, so the overhead cost per integrand value is $O(n^2)$ flops plus the cost of generating $n/4$ Normal(0,1) random numbers. Once an integrand evaluation point is available, the expected cost for the evaluation of f is at least $O(n)$, because there are n components for the input variable for the integrand. However, for many application problems this could be much higher (e.g. $O(n^2)$). Therefore, if the $O(n)$ integrand evaluation cost is measured in flops, the constant in $O(n)$ is expected to be large, so that the $O(n^2)$ flops for the generation of the evaluation point for that integrand evaluation should not be significant unless n is very large. For the Genz and Monahan (1996) degree five SR rules the

number of f values for each random Q is $O(n^2)$, so the Q overhead cost per evaluation point drops to approximately $O(n)$ and this is not significant compared to the total f value cost for typical applications problems.

3.5 Examples

The first example is the seven dimensional problem that was used as an example for the randomized fully symmetric interpolatory rules. The results in the following table were obtained using stochastic rules with different degrees. In this case the higher degree rule results are not always better.

Table 2: Proportional Hazards Model Problem Results II

d	f Vals.	$I(f)$	Std. Err.	$I(x_1 f)/I(f)$	Std. Err.	$I(x_2 f)/I(f)$	Std. Err.
0	25000	.00015312	(.00000126)	.13296	(.00150)	-3.96961	(.02983)
0	50000	.00015134	(.00000071)	.13480	(.00115)	-3.97691	(.01921)
1	25000	.00014910	(.00000061)	.13640	(.00116)	-3.97775	(.01838)
1	50000	.00015235	(.00000102)	.13481	(.00139)	-3.98131	(.02456)
2	24993	.00015217	(.00000052)	.13539	(.00069)	-3.98157	(.01426)
2	49993	.00015186	(.00000045)	.13595	(.00061)	-3.98316	(.01275)
3	24993	.00015215	(.00000070)	.13569	(.00064)	-3.98278	(.01924)
3	49985	.00015187	(.00000052)	.13650	(.00089)	-3.98455	(.01601)
4	24993	.00015293	(.00000137)	.13650	(.00219)	-3.98937	(.04321)
4	49985	.00015169	(.00000056)	.13625	(.00062)	-3.98387	(.01489)
5	24913	.00015143	(.00000082)	.13551	(.00082)	-3.97950	(.02222)
5	49969	.00015305	(.00000083)	.13612	(.00077)	-3.98758	(.02285)

The second example in this section has been used by Caflisch and Morokoff (1996) to illustrate the use of a quasi-Monte Carlo for a computational finance problem. The integrand, which has a multivariate Normal weight, is

$$P(\mathbf{x}) = \sum_{k=1}^{n} \frac{((1 - w_k(\mathbf{x})) + w_k(\mathbf{x})c_k) \prod_{j=1}^{k-1}(1 - w_j(\mathbf{x}))}{\prod_{j=0}^{k-1}(1 + i_k(\mathbf{x}))},$$

where $i_k(\mathbf{x}) = i_0 K_0^k e^{\sigma(x_1+x_2+\cdots+x_k)}$, $K_0 = e^{-\sigma^2/2}$, $w_k(\mathbf{x}) = K_1 + K_2 \tan^{-1}(K_3 i_k(\mathbf{x}) + K_4)$ and $c_k = \sum_{j=0}^{n-k}(1+i_0)^{-1}$, with an initial interest rate i_0 and constants $(K_1, K_2, K_3, K_4, \sigma)$ given. With this integrand $I(P)$ is the present value of a security backed by mortgages of length n months with a fixed monthly interest rate i_0. A related integral $I(A)$, which determines the average life for the mortgages, was $A(\mathbf{x})$ defined by

$$A(\mathbf{x}) = \sum_{k=1}^{n} k w_k(\mathbf{x}) \prod_{j=1}^{k-1}(1 - w_j(\mathbf{x})).$$

Two types of problem were considered: a "nearly linear" problem, and a "nonlinear" problem. These problem types depend on the choice of the constants K_1, K_2, K_3, K_4. The constants chosen

for the "nearly linear" problem provided a problem that was easier than the "nonlinear" problem (see Caflisch and Morokoff, 1996, for further discussion). This can be seen from the numerical results, where a one-point Gauss-Hermite rule result, which is provided for comparison purposes, is accurate to more than three decimal digits for the "nearly linear" case but is much less accurate for the "nonlinear" case. In both cases $i_0 = 0.007$, $\sigma = 0.02$, and the time length was thirty years ($n = 360$ months). The following results were obtained using the SR rules constructed by Genz and Monahan (1996).

Table 3: Nearly Linear Results, $n = 360$, $(K_1, K_2, K_3, K_4) = (0.01, -0.005, 10, 0.5)$

Degree	Present Value	Rel. Std. Err.	Average Life	Rel. Std. Err.	f Vals.
1	131.96705124	(?)	100.95445646	(?)	1
0	131.64356185	(7.59E-04)	100.93227103	(3.33E-05)	4000
0	131.77671967	(5.47E-04)	100.93507159	(2.31E-05)	8000
0	131.80541015	(3.90E-04)	100.93464903	(1.64E-05)	16000
0	131.69346344	(2.73E-04)	100.93429738	(1.16E-05)	32000
0	131.80329755	(1.93E-04)	100.93217315	(8.19E-06)	64000
1	131.78717265	(1.98E-05)	100.93214659	(1.15E-05)	4000
1	131.78782169	(1.45E-05)	100.93329874	(7.68E-06)	8000
1	131.78553453	(1.02E-05)	100.93334875	(5.71E-06)	16000
1	131.78638077	(7.15E-06)	100.93321769	(3.94E-06)	32000
1	131.78652790	(5.06E-06)	100.93399299	(2.76E-06)	64000
3	131.78708507	(9.99E-07)	100.93339140	(4.04E-07)	3611
3	131.78697558	(8.74E-07)	100.93343541	(2.69E-07)	7943
3	131.78698070	(4.87E-07)	100.93340938	(2.61E-07)	15885
3	131.78692483	(3.48E-07)	100.93344686	(1.77E-07)	31769
3	131.78699398	(2.25E-07)	100.93341537	(1.01E-07)	63537
5	131.78702325	(3.56E-08)	100.93340794	(3.97E-09)	522729
5	131.78702811	(1.66E-08)	100.93340822	(7.37E-10)	1045457
5	131.78702918	(1.43E-08)	100.93340820	(1.57E-09)	2090913

For each set of results the initial degree one values are the values for the one-point Gauss-Hermite product rule ($I(f) \approx (\sqrt{2\pi})^{-360} f(\mathbf{0})$), the degree zero values are simple Monte Carlo values obtained using the method described in the introduction. The other degree one values are simple Monte Carlo with antithetic variates values (half as many random $\mathbf{x}_i$ points are used but for each $\mathbf{x}_i$, $(f(\mathbf{x}_i) + f(-\mathbf{x}_i))/2$ is used in the sum instead of $f(\mathbf{x}_i)$). The results show that the SR rules can be feasibly implemented for high dimensional problems, and that they can provide much more accurate results than simple Monte Carlo methods for comparable work. The degree five SR rule results are given for larger sample sizes than the degree three rules because the degree five rules require $2(n+1)(n+2)+1$ f values for the first sample and $2(n+1)(n+2)$ f values for subsequent samples, with a minimum of two samples needed to produce an error estimate. With $n = 360$, 522729 f values are needed to start. The accuracies given here are comparable to those obtained by Caflisch and Morokoff using quasi-Monte Carlo methods.

Table 4: Nonlinear Results, $n = 360$, $(K_1, K_2, K_3, K_4) = (0.04, 0.0222, -1500, 7)$

Degree	Present Value	Rel. Std. Err.	Average Life	Rel. Std. Err.	f Vals.
1	131.72003517	(?)	80.41606389	(?)	1
0	130.52966340	(5.37E-04)	76.88486720	(2.28E-03)	4000
0	130.74241971	(3.72E-04)	76.39663281	(1.65E-03)	8000
0	130.71072974	(2.59E-04)	76.60043200	(1.13E-03)	16000
0	130.66085695	(1.85E-04)	76.66036176	(8.04E-04)	32000
0	130.72181827	(1.31E-04)	76.52473741	(5.71E-04)	64000
1	130.70782444	(1.85E-04)	76.52039518	(8.45E-04)	4000
1	130.71396840	(1.26E-04)	76.53344528	(5.86E-04)	8000
1	130.71329258	(9.00E-05)	76.52751566	(4.12E-04)	16000
1	130.72415900	(6.33E-05)	76.54969549	(2.93E-04)	32000
1	130.71222885	(4.56E-05)	76.53109868	(2.09E-04)	64000
3	130.70842632	(2.24E-05)	76.49073517	(4.24E-04)	3611
3	130.71496584	(2.22E-05)	76.56024758	(4.00E-04)	7943
3	130.71225416	(9.17E-06)	76.54368578	(2.04E-04)	15885
3	130.71380030	(8.33E-06)	76.55793417	(1.70E-04)	31769
3	130.71253732	(5.94E-06)	76.54185151	(1.21E-04)	63537
5	130.71298042	(1.60E-06)	76.54720257	(1.05E-04)	522729
5	130.71192304	(1.95E-06)	76.52920789	(4.96E-05)	1045457
5	130.71226485	(2.85E-06)	76.53418023	(8.82E-05)	2090913

References

Caflisch, R. and Morokoff, W. (1996), Quasi-Monte Carlo Computation of a Finance Problem, in *Proceedings of the Workshop on Quasi-Monte Carlo Methods and Their Applications*, Statistics Research and Consultancy Unit, Hong Kong Baptist University, 15–30.

Davis, P.J., and Rabinowitz, P. (1984), *Methods of Numerical Integration*, Academic Press, New York.

Dellaportas, P. and Wright, D. (1992), A Numerical Integration Strategy in Bayesian Analysis, in *Bayesian Statistics 4*, Bernardo, J.M., Berger, J.O., David, A.P. and Smith, A.F.M. (Eds.), Oxford University Press, Oxford, 601–606.

Evans, M., and Swartz, T. (1995), Methods for Approximating Integrals in Statistics with Special Emphasis on Bayesian Integration Problems, *Statistical Science*, **10**, 254–272.

Grabner, P.J., Tichy, R.F. (1993), Spherical Designs, Discrepancy and Numerical Integration, *Math. Comp.* **60**, 327–336.

Joy, C., Boyle, P. and Tan, T. (1996), Quasi-Monte Carlo Methods in Numerical Finance, *Management Science* **42**, 926–938.

Genz, A. (1982), A Lagrange Extrapolation Algorithm for Sequences of Approximations to Multiple Integrals, *SIAM J. Sci. Stat. Comput.* **3**, pp. 160–172.

Genz, A. (1986), Fully Symmetric Interpolatory Rules for Multiple Integrals, *SIAM J. Numer. Anal.* **23**, 1273–1283.

Genz, A., and Monahan, J. (1996), Stochastic Integration Rules for Infinite Regions, to appear in *SIAM J. Sci. Comput.* **19**, 1998.

Haber, S. (1969), Stochastic Quadrature Formulas, *Math. Comp.* **23**, 751–764.

Hammersley, J.M. and Handscomb, D.C. (1964), *Monte Carlo Methods*, Chapman and Hall, London.

Monahan, J., and Genz, A. (1997), Spherical-Radial Integration Rules for Bayesian Computation, *Journal of the American Statistical Association*, **92**, 664–674.

Mysovskikh, I. P. (1980), The Approximation of Multiple Integrals by using Interpolatory Cubature Formulae, in *Quantitative Approximation*, R.A. DeVore and K. Scherer (Eds.), Academic Press, New York, 217–243.

Mysovskikh, I. P. (1981), *Interpolatory Cubature Formulas* (Russian), Izd 'Nauka', Moscow-Leningrad.

Siegel, A.F. and O'Brien, F. (1983), Unbiased Monte Carlo Integration Methods with Exactness for Low Order Polynomials, *SIAM. J. Sci. Stat. Comput.* **6**, 169–181.

Stewart, G.W. (1980), The Efficient Generation of Random Orthogonal Matrices with An Application to Condition Estimation, *SIAM J. Numer. Anal.* **17**, 403–409.

Stroud, A. H. (1971), *The Approximate Calculation of Multiple Integrals*, Prentice Hall, Englewood Cliffs, New Jersey.