Stochastic Integration Rules for Infinite Regions *

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

John Monahan
Department of Statistics
North Carolina State University
Raleigh, NC 27695-8203
monahan@stat.ncsu.edu

Abstract

Stochastic integration rules are derived for infinite integration intervals, generalizing rules developed by Siegel and O'Brien (1985) for finite intervals. Then random orthogonal transformations of rules for integrals over the surface of the unit m-sphere are used to produce stochastic rules for these integrals. The two types of rules are combined to produce stochastic rules for multidimensional integrals over infinite regions with Normal or Student-t weights. Example results are presented to illustrate the effectiveness of the new rules.

Key Words: Monte Carlo, multiple integrals, numerical integration, statistical computation.

AMS Subject Classifications: 65C05, 65C10, 65D30, 65D32.

1 Introduction

A common problem in applied science and statistics is to numerically compute integrals in the form

$$E(g) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} g(\boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta},$$

with $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_m)^t$. For statistics applications the function $p(\boldsymbol{\theta})$ may be an unnormalized unimodal posterior density function and $g(\boldsymbol{\theta})$ is some function for which an approximate expected value is needed. We are interested in problems where $p(\boldsymbol{\theta})$ is approximately multivariate normal ($\boldsymbol{\theta} \sim N_m(\boldsymbol{\mu}, \Sigma)$ or multivariate Student-t ($\boldsymbol{\theta} \sim t_m(\boldsymbol{\mu}, \Sigma)$). 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 form

$$I(f) = |C| \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} w(||\mathbf{x}||) f(\mathbf{x}) d\mathbf{x},$$

where $||\mathbf{x}|| = \sqrt{\mathbf{x}^t\mathbf{x}}$, $f(\mathbf{x}) = g(\mu + C\mathbf{x})p(\mu + C\mathbf{x})/w(||\mathbf{x}||)$, and $w(||\mathbf{x}||) = e^{-\mathbf{x}^t\mathbf{x}/2}$ (multivariate normal), or $w(||\mathbf{x}||) = (1 + \frac{\mathbf{x}^t\mathbf{x}}{\nu})^{-(m+\nu)/2}$ (multivariate Student-t). If the approximation to $p(\boldsymbol{\theta})$ is good, then $f(\mathbf{x})$

*Accepted for publication in *SIAM Journal on Scientific Computing.*
†Partially supported by NSF grant DMS-9211640.

can be accurately approximated by a low degree polynomial in $\mathbf{x}$, and this motivates our construction of stochastic multidimensional polynomial integrating rules for integrals $I(f)$.

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, 1992). 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 $E(g)$, 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_E denote the standard error for the sample, then

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

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

The new methods that we will describe can be considered a refinement of this Monte-Carlo with importance sampling algorithm. Simple Monte-Carlo with importance sampling results are exact whenever the importance modified integrand is constant, but our methods will be exact whenever the importance modified integrand is a low degree polynomial. Our methods will also provide a robust error estimate from the sample standard error. The new one-dimensional integration rules that we develop are generalizations of the rules derived for the interval [-1,1], with weight $w(r) = 1$, by Siegel and O'Brien (1985). Their work extends earlier work by Hammersley and Handscomb (1964), who also considered the construction of stochastic integration rules for finite intervals. Our work is also partly based on work by Haber (1969), who introduced the word "stochastic" for generalized Monte-Carlo rules.

Our development of stochastic multidimensional integration rules requires an additional change of variables to a radial-spherical coordinate system. We 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 $m \geq 1$,

$$\begin{aligned} I(f) &= \int\limits_{\mathbf{z}^t\mathbf{z}=1} \int_0^\infty w(r) r^{m-1} f(r\mathbf{z}) dr d\mathbf{z} \\ &= \frac{1}{2} \int\limits_{\mathbf{z}^t\mathbf{z}=1} \int_{-\infty}^\infty w(r)|r|^{m-1} f(r\mathbf{z}) dr d\mathbf{z}. \end{aligned}$$

The numerical approximations to $I(f)$ that we propose to use will be products of stochastic integration rules for the radial interval $(-\infty, \infty)$ with weight $w(r)|r|^{m-1}$, and stochastic rules (of the same polynomial degree) for the surface of the unit m-sphere. Averages of properly chosen samples of these rules will 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. Our development was partially motivated by the work of Deák (1990), who used a transformation to a spherical coordinate system combined with random orthogonal transformations to develop a method for computing multivariate normal probabilities, but he did not consider using higher degree rules.

2 Stochastic Radial Rules

The basic radial integration rules that we use are combinations of the symmetric sums $C(\rho) = (h(\rho) + h(-\rho))/2$. A radial rule $R(h)$ takes the form

$$R(h) = \sum_{i=0}^{n} w_i C(\rho_i). \tag{1}$$

Given points $\{\rho_i\}$, the weights $\{w_i\}$ will be determined so that R has polynomial degree $2n+1$. The points $\{\rho_i\}$ will be randomly chosen so that R is an unbiased estimate for $T(h) = \int_{-\infty}^{\infty} |r|^{m-1} w(r)h(r)dr$.

For fixed points $\{\rho_i\}$, the selection of the weights is a standard integration rule construction problem. If we want a degree d rule, it is sufficient that $R(h) = T(h)$ whenever $h(r) = r^k$ for k = 0, 1, ..., d. When k is an odd integer, the equation is automatically satisfied because both R and the integration operator are symmetric. Define $P(h, r)$ by

$$P(h,r) = \sum_{i=0}^{n} C(\rho_i) \prod_{j=0,\neq i}^{n} \frac{r^2 - \rho_j^2}{\rho_i^2 - \rho_j^2}.$$

Now $P(h, r)$ is a even degree Lagrange interpolating polynomial for h, so it follows from standard interpolation theory, that $P(h, r) = h$, whenever $h = r^{2k}$ and $0 \leq k \leq n$. Therefore $T(r^{2k}) = T(P(r^{2k}, r))$, and the weights $\{w_i\}$ that we need to make R degree $2n+1$ are just integrals of the even degree Lagrange basis functions. We have the following theorem:

Theorem 1 *If the points $\{\rho_i\}$ are distinct non-negative real numbers and the weights $\{w_i\}$ are defined by*

$$w_i = T\Big(\prod_{j=0,\neq i}^{n} \frac{r^2 - \rho_j^2}{\rho_i^2 - \rho_j^2}\Big), \tag{2}$$

for $i = 0, 1, ..., n$, then R is a degree $2n+1$ integration rule for T.

We now describe how to choose the points $\{\rho_i\}$ so that R is an unbiased estimate for $T(h)$. In order to accomplish this, we need to find a joint probability density function $p(\rho_0, \rho_1, ..., \rho_n)$ that satisfies

$$E\{R\} \equiv \int_0^\infty \int_0^\infty ... \int_0^\infty R(h)p(\rho_0, \rho_1, ..., \rho_n) d\rho_0 d\rho_1 ... d\rho_n = T(h),$$

for any integrable h. We will explicitly show how to do this when $n = 0, 1$ and 2, and conjecture the general form for p for $n \geq 3$. We will let $T_k = T(|r|^k)$ and use the fact that

$$2\int_0^\infty r^{m+k-1} w(r) dr = T_k. \tag{3}$$

The case $n = 0$ is straightforward. Define $R^1(\rho) = T_0 C(\rho)$, and choose $\rho \geq 0$ randomly with density $2\rho^{m-1}w(\rho)/T_0$. Then we have

$$\begin{aligned} E\{R^1\} &= \int_0^\infty C(\rho) 2\rho^{m-1} w(\rho) d\rho \\ &= \int_0^\infty \rho^{m-1} w(\rho)(h(\rho) + h(-\rho)) d\rho \\ &= \int_{-\infty}^{\infty} |\rho|^{m-1} w(|\rho|) h(\rho) d\rho \\ &= T(h). \end{aligned}$$

For $n = 1$, we set $\rho_0 = 0$. A degree three rule for $T(h)$ is

$$\begin{aligned} R^3(\rho) &= C(0)T((\rho^2 - r^2)/\rho^2) + C(\rho)T(r^2/\rho^2) \\ &= C(0)(\rho^2 T_0 - T_2)/\rho^2 + C(\rho)T_2/\rho^2. \end{aligned}$$

If we choose $\rho \geq 0$ randomly with density $\frac{2\rho^{m+1}w(\rho)}{T_2}$ then

$$\begin{aligned} E\{R^3\} &= \int_0^\infty R^3(\rho)\frac{2\rho^{m+1}w(\rho)}{T_2}d\rho \\ &= 2\int_0^\infty C(0)(\rho^2 T_0 - T_2)\frac{\rho^{m-1}w(\rho)}{T_2}d\rho \ + \ 2\int_0^\infty \rho^{m-1}w(\rho)C(\rho)d\rho \\ &= 2C(0)\int_0^\infty \rho^{m-1}w(\rho)(\rho^2\frac{T_0}{T_2} - 1)d\rho \ + \ \int_{-\infty}^\infty |\rho|^{m-1}w(|\rho|)h(\rho)d\rho \\ &= 2C(0)(\int_0^\infty \rho^{m+1}w(\rho)\frac{T_0}{T_2}d\rho \ - \ \int_0^\infty \rho^{m-1}w(\rho)d\rho) \ + \ T(h) \\ &= T(h). \end{aligned}$$

A degree five rule for $T(h)$ is

$$R^5(\rho,\delta) = C(0)T(\frac{(r^2-\rho^2)(r^2-\delta^2)}{\rho^2\delta^2}) + C(\rho)T(\frac{r^2(r^2-\delta^2)}{\rho^2(\rho^2-\delta^2)}) + C(\delta)T(\frac{r^2(r^2-\rho^2)}{\delta^2(\delta^2-\rho^2)}).$$

We will choose $\rho \geq 0$ and $\delta \geq 0$ randomly with joint density

$$p(\rho,\delta) = K\rho^{m+1}w(\rho)\delta^{m+1}w(\delta)(\rho-\delta)^2(\rho+\delta),$$

where K is determined by the condition $\int_0^\infty\int_0^\infty p(\rho,\delta)d\rho d\delta = 1$. We now need to show that $E\{R^5\} = T(h)$. There are three terms in R^5 to consider, so we start with the first one, and we find

$$\begin{aligned} & E\{C(0)T(\frac{(r^2-\rho^2)(r^2-\delta^2)}{\rho^2\delta^2})\} \\ &= C(0)K\int_0^\infty\int_0^\infty \rho^{m+1}w(\rho)\delta^{m+1}w(\delta)(\rho-\delta)^2(\rho+\delta)T(\frac{r^4-(\rho^2+\delta^2)r^2+\rho^2\delta^2}{\rho^2\delta^2})d\rho d\delta \\ &= C(0)(\ K\int_0^\infty\int_0^\infty (\rho\delta)^{m-1}w(\rho)w(\delta)(\rho-\delta)^2(\rho+\delta)(T_4 - T_2(\rho^2+\delta^2) + T_0\rho^2\delta^2))d\rho d\delta \\ &= C(0)K(\ T_4(T_3T_0 - T_2T_1) - T_2(T_5T_0 - T_4T_1) + T_0(T_5T_2 - T_4T_3)\ )/2 \\ &= 0. \end{aligned}$$

For the second term we find

$$\begin{aligned} & E\{C(\rho)T(\frac{r^2(r^2-\delta^2)}{\rho^2(\rho^2-\delta^2)})\} \\ &= K\int_0^\infty\int_0^\infty \rho^{m-1}w(\rho)\delta^{m+1}w(\delta)(\rho-\delta)^2(\rho+\delta)C(\rho)\frac{T_4-\delta^2T_2}{(\rho^2-\delta^2)})d\rho d\delta \\ &= K\int_0^\infty\int_0^\infty \rho^{m-1}w(\rho)\delta^{m-1}w(\delta)\delta^2(\rho-\delta)C(\rho)(T_4-\delta^2T_2)d\rho d\delta \\ &= K\int_0^\infty \rho^{m-1}w(\rho)C(\rho)\int_0^\infty \delta^{m-1}w(\delta)\delta^2(\rho-\delta)(T_4-\delta^2T_2)d\delta d\rho \\ &= K\int_0^\infty \rho^{m-1}w(\rho)C(\rho)(\frac{T_2T_4-T_4T_2}{2} - \frac{T_3T_4-T_5T_2}{2})d\rho \\ &= K\frac{T_5T_2-T_4T_3}{2}\int_0^\infty \rho^{m-1}w(\rho)C(\rho)d\rho \\ &= K(T_5T_2-T_4T_3)T(h)/4 \end{aligned}$$

Now

$$\begin{aligned}
&\int_0^\infty \int_0^\infty \rho^{m+1} w(\rho)\delta^{m+1} w(\delta)(\rho-\delta)^2(\rho+\delta) d\rho d\delta \\
&= \int_0^\infty \int_0^\infty \rho^{m-1} w(\rho)\delta^{m-1} w(\delta)\rho^2\delta^2(\rho^2-\delta^2)(\rho-\delta) d\rho d\delta \\
&= \int_0^\infty \int_0^\infty \rho^{m-1} w(\rho)\delta^{m-1} w(\delta)\rho^2\delta^2(\rho^3-\rho^2\delta-\rho\delta^2+\delta^3) d\rho d\delta \\
&= (T_5T_2 - T_4T_3)/2,
\end{aligned}$$

so $K = 2/(T_5T_2 - T_4T_3)$, and therefore

$$E\{C(\rho)T(\frac{r^2(r^2-\delta^2)}{\rho^2(\rho^2-\delta^2)})\} = T(h)/2.$$

Because R^5 is symmetric in ρ and δ, the last term in R^5 also has expected value $T(h)/2$, so we have shown that $E\{R^5\} = T(h)$. We summarize our results in this section with Proposition 1.

Proposition 1 *If $w(r) = w(\text{-}r)$, $\rho_0 = 0$ and the points $\{\rho_i\}$, $0 < i \leq n$, for the rules R^{2n+1} given by (1) with weights given by (2), are chosen with probability density proportional to*

$$p(\rho_1, ..., \rho_n) = \prod_{i=1}^{n} \rho_i^{m+1} w(\rho_i) \prod_{j=1}^{i-1} (\rho_i - \rho_j)^2(\rho_i + \rho_j),$$

then R is an unbiased degree $2n+1$ integration rule for $T(h)$.

We have proved this for $n = 1$ and $n = 2$. The form for the probability density for $n > 2$ is a conjectured natural generalization of the Siegel and O'Brien Theorem 5.1 (1985). Because of practical problems associated with generating random $\boldsymbol{\rho}$'s from this density when $n > 2$ we focus on the $n = 1$ and $n = 2$ cases.

3 Stochastic Spherical Integration Rules

The spherical surface integrals will be approximated by averages of random rotations of appropriately chosen rules for the spherical surface. Let

$$S(s) = \sum_{j=1}^{N} \tilde{w}_j s(\mathbf{z}_j),$$

with $\mathbf{z}_j^t\mathbf{z}_j = 1$ for all j, be an integration rule that approximates an integral of a function $s(\mathbf{z})$ over the surface U_m of the unit m-sphere defined by $\mathbf{z}^t\mathbf{z} = 1$. If Q is an $m \times m$ orthogonal matrix then

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

is also an integration rule for s over U_m, because $||Q\mathbf{z}|| = ||\mathbf{z}||$. Furthermore, if S has polynomial degree d, then so does S_Q, because $s(Qz)$ has the same degree as $s(z)$. If Q is chosen uniformly (see Stewart, 1980) and S has polynomial degree d, then S_Q is an unbiased random degree d rule for U_m.

There are many choices that could be used for S. We consider rules given in the book by Stroud (1971, pages 294-296) and the review paper by Mysovskikh (1980, pages 236-237). The rules that we will combine with radial rules have degree 1, 3 or 5, and we now list them. A simple degree 1 rule is

$$S^1(s) = |U_m|(s(-\mathbf{z}) + s(+\mathbf{z}))/2,$$

where $|U_m| = 2\pi^{m/2}/\Gamma(m/2)$ is the surface content of U_m, and $\mathbf{z}$ is any point on U_m. A simple degree 3 rule is

$$S^3(s) = \frac{|U_m|}{2m}\sum_{j=1}^{m}(s(-\mathbf{e}_j)+s(+\mathbf{e}_j)),$$

where $\mathbf{e}_j = (0,...,0,1,0,...,0)^t$, with the "1" in the jth position. This rule uses $2m$ values of $s(\mathbf{z})$. A different degree 3 rule (Mysovskikh, 1980) is

$$\hat{S}^3(s) = \frac{|U_m|}{2(m+1)}\sum_{j=1}^{m+1}(s(-\mathbf{v}_j)+s(+\mathbf{v}_j)),$$

where $\mathbf{v}_j$ is the jth vertex of a regular m-simplex with vertices on U_m. The degree 3 rule $\hat{S}^3$ is slightly more expensive to use than S^3, but it leads to an efficient general degree 5 rule (Mysovskikh, 1980)

$$\hat{S}^5(s) = |U_m| \quad (\quad \frac{(7-m)m}{2(m+1)^2(m+2)}\sum_{j=1}^{m+1}(s(-\mathbf{v}_j)+s(+\mathbf{v}_j))$$
$$+\frac{2(m-1)^2}{m(m+1)^2(m+2)}\sum_{j=1}^{m(m+1)/2}(s(-\mathbf{y}_j)+s(+\mathbf{y}_j)) \ ).$$

The points $\mathbf{y}_j$ are determined by taking the midpoints of edges of the m-simplex with vertices $\mathbf{v}_j$, and projecting those midpoints onto the surface of U_m. $\hat{S}^5$ requires only $(m+1)(m+2)$ values of $s(\mathbf{z})$. A degree five rule which extends S^3 (Stroud, 1971, page 294) is

$$S^5(s) = |U_m|(\frac{(4-m)}{2m(m+2)}\sum_{j=1}^{m}(s(-\mathbf{e}_j)+s(+\mathbf{e}_j)) + \frac{1}{m(m+2)}\sum_{j=1}^{2m(m-1)} s(\mathbf{u}_j) \ ),$$

where $\mathbf{u}_j$ is one of the $2m(m-1)$ points in the fully symmetric set that is determined by all possible permutations and sign changes of the coordinates of the point $(r,r,0,...,0)^t$, with $r = 1/\sqrt{2}$.

4 Stochastic Spherical-Radial Integration Rules

In this section we combine stochastic radial rules with stochastic spherical rules to produce random rules for $I(f)$. There are many ways that this could be done. A natural approach is to form a stochastic product rule $SR_{Q,\boldsymbol{\rho}}(f)$ from a spherical surface rule S and a radial rule R. Such a rule takes the form

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

If S and R both have degree d, then $SR_{Q,\boldsymbol{\rho}}(f)$ will also have degree d (Stroud, 1971, Theorem 2.3-1). If Q is a uniformly random orthogonal matrix and $\boldsymbol{\rho}$ is random chosen with the correct density for R, then $SR_{Q,\boldsymbol{\rho}}(f)$ will be an unbiased estimate for $I(f)$. We have the following theorem:

Theorem 2 *If $\boldsymbol{\rho}$ is random with density given by Proposition 1, S has degree $2n+1$ and Q is an $m\times m$ uniform random orthogonal matrix, then*

$$SR_{Q,\rho}(f) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}...\int_{-\infty}^{\infty} w(||x||)f(x)dx$$

whenever f is a degree $2n+1$ polynomial, and

$$E\{SR_{Q,\rho}(f)\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}...\int_{-\infty}^{\infty} w(||x||)f(x)dx$$

for any integrable f.

We give three examples of SR rules. A degree one rule constructed from S^1 and R^1 is

$$SR^1_\rho(f) = \frac{1}{2}|U_m|T_0\frac{f(-\rho\mathbf{z}) + f(+\rho\mathbf{z})}{2}.$$

Here Q is unnecessary, because uniform random vectors $\mathbf{z}$ from U_m give unbiased rules. A degree three rule constructed from S^3 and R^3 is

$$SR^3_{Q,\rho}(f) = \frac{|U_m|}{m}\sum_{j=1}^{m}(\ w_0 f(\mathbf{0}) + w_1\frac{f(-\rho Q\mathbf{z}_j) + f(+\rho Q\mathbf{z}_j)}{2}\ ).$$

A degree five rule constructed from $\hat{S}^5$ and R^5 is

$$\begin{aligned}\hat{S}R^5_{Q,\rho,\delta}(f) \quad = \quad & \tilde{w}_1\sum_{j=1}^{m+1}(w_0 f(\mathbf{0}) + w_1\frac{f(-\rho Q\mathbf{v}_j) + f(+\rho Q\mathbf{v}_j)}{2} + w_2\frac{f(-\delta Q\mathbf{v}_j) + f(\delta Q\mathbf{v}_j)}{2}) \\ & +\tilde{w}_2\sum_{j=1}^{m(m+1)/2}(w_0 f(\mathbf{0}) + w_1\frac{f(-\rho Q\mathbf{y}_j) + f(+\rho Q\mathbf{y}_j)}{2} + w_2\frac{f(-\delta Q\mathbf{y}_j) + f(\delta Q\mathbf{y}_j)}{2}),\end{aligned}$$

with $\tilde{w}_1 = |U_m|\frac{(7-m)m}{2(m+1)^2(m+2)}$ and $\tilde{w}_2 = |U_m|\frac{2(m-1)^2}{m(m+1)^2(m+2)}$.

$SR^1_\rho(f)$, $SR^3_{Q,\rho}$ and $\hat{S}R^5_{Q,\rho,\delta}$ require 2, $2m+1$ and $2(m+1)(m+2)+1$ f values, respectively. A sample of one of these rules can be generated, and the sample average used to estimate $I(f)$. The standard error for the sample can be used to provide an error estimate. For comparison purposes with the examples in Section 6, we will use $SR^0(f)$ to denote the one point rule $f(\mathbf{z})$, with the components of $\mathbf{z}$ chosen from Normal(0,1). $SR^0(f)$ is just the simple Monte-Carlo rule for $I(f)$ with multivariate normal weight.

5 Implementation Details and Algorithms

In this section we focus on integrals of the form

$$I(f) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}...\int_{-\infty}^{\infty} w(||\mathbf{x}||)f(\mathbf{x})d\mathbf{x},$$

where $w(||\mathbf{x}||) = (2\pi)^{-m/2}e^{-\mathbf{x}^t\mathbf{x}/2}$. For integrals of this type, we have determined explicit formulas for the radial rule weights, along with explicit methods for generating the random radial rule points. We will also discuss the multivariate Student-t weight $w(||\mathbf{x}||) = \frac{\Gamma(\frac{\nu+m}{2})}{\Gamma(\frac{\nu}{2})\sqrt{(\nu\pi)^m}}(1+\frac{\mathbf{x}^t\mathbf{x}}{\nu})^{-(m+\nu)/2}$.

We first consider the rule SR^1_ρ. In the case $w(||\mathbf{x}||) = (2\pi)^{-m/2}e^{-\mathbf{x}^t\mathbf{x}/2}$, we have $w(r) = (2\pi)^{-m/2}e^{-r^2/2}$, so $T_0 = \pi^{-m/2}\Gamma(m/2)$, and $|U_m|T_0 = 2$. Therefore

$$SR^1_\rho(f) = \frac{f(-\rho\mathbf{z}) + f(+\rho\mathbf{z})}{2}.$$

The probability density for ρ is proportional to $\rho^{m-1}e^{-\rho^2/2}$, a Chi density with m degrees of freedom. It is a standard statistical procedure to generate a random ρ with this density (Monahan, 1987). A standard procedure for generating uniformly random vectors $\mathbf{z}$ from U_m, consists of first generating $\mathbf{x}$ with components x_i random from Normal(0,1) and setting $\mathbf{z} = \mathbf{x}/||\mathbf{x}||$. However, this combined procedure for generating random vectors $\rho\mathbf{z}$ must be equivalent to just generating random $\mathbf{z}$ from $[-\infty, \infty]^m$ with density $w(||\mathbf{z}||)$. Therefore, all we need to do is generate the components z_i random from Normal(0,1), and this is a simpler procedure. We propose the following algorithm for random degree one rules:

Degree One Spherical-Radial Rule Integration Algorithm

1. **Input** ϵ, m, f and N_{max}.
2. Set $N = 0$, $I = 0$, $V = 0$.
3. **Repeat**
   (a) Set $N = N + 1$.
   (b) Generate a random $\mathbf{x}$ with $x_i \sim$ Normal(0,1).
   (c) Set $SR = \frac{f(-\mathbf{x})+f(+\mathbf{x})}{2}$, $D = (SR - I)/N$, $I = I + D$ and $V = (N-2)V/N + D^2$.

   Until $V < \epsilon^2$ or $N = N_{max}$.
4. **Output** $I \approx I(f)$, $\sigma_E = \sqrt{V}$ and N.

The input ϵ is an error tolerance, the input N_{max} provides a limit on the time for the algorithm, and the output σ_E is the standard error for the integral estimate I. The algorithm computes I and V using a modified version of a stable one-pass algorithm (Chan and Lewis, 1979). The unscaled sample standard error σ_E will usually be an error bound with approximately 68% certainty. Users of this algorithm who desire a higher degree of confidence can scale σ_E appropriately. For example, a scale factor of 2 increases the certainty level to approximately 95%.

The error estimates obtained by scaling σ_E with this algorithm (and the other algorithms in this section) should be used with caution for low N values. These error estimates are based on the use of the Central Limit Theorem to infer that the sample averages SR are approximately Normal. A careful implementation of the algorithms in this section could include an N_{min} parameter and/or use a larger scale factor for σ_E for small N values. For large N, a scaled σ_E should provide a robust, statistically sound error estimate, as long as the multivariate normal model adequately represents the tails in the posterior density. Posterior densities with thicker tails are often more efficiently and reliably handled using a multivariate Student-t model. One technique for monitoring this is discussed by Monahan and Genz (1996).

If we consider the Student-t weight, we can see that the density for ρ is proportional to $r^{m-1}(1 + \frac{r^2}{\nu})^{-(m+\nu)/2}$ and a change of variable shows this to be proportional to a Beta$(\frac{m}{2}, \frac{\nu}{2})$ probability density (see Devroye, 1986, for generating methods), so the random ρ's and the uniformly random vectors $\mathbf{z}$ from U_m, needed for SR^1 can easily be generated. We can also show $|U_m|T_0 = 2$, so the formula for $SR^1(f)$ is the same as the formula for the multivariate Normal case. By making appropriate changes to line 3(b) and 3(c) of the previous algorithm, a modified algorithm could be produced.

Next, we consider the rule $SR^3_{Q,\rho}$. Integration by parts with $w(r) = (2\pi)^{-m/2}e^{-r^2/2}$, shows $T_2 = mT_0$, so

$$R^3(\rho) = T_0(\ C(0)(1 - \frac{m}{\rho^2}) + C(\rho)\frac{m}{\rho^2}\ ),$$

and therefore

$$\begin{aligned} SR^3_{Q,\rho}(f) &= \frac{1}{m}\sum_{j=1}^{m}(f(\mathbf{0})(1 - \frac{m}{\rho^2}) + \frac{m(f(-\rho Q\mathbf{e}_j) + f(+\rho Q\mathbf{e}_j))}{2\rho^2}) \\ &= f(\mathbf{0})(1 - \frac{m}{\rho^2}) + \sum_{j=1}^{m}\frac{f(-\rho Q\mathbf{e}_j) + f(+\rho Q\mathbf{e}_j)}{2\rho^2}. \end{aligned}$$

The probability density for ρ is proportional to $\rho^{m+1}e^{-\rho^2/2}$, a Chi density with $m+2$ degrees of freedom. We propose the following algorithm for stochastic degree three rules:

Degree Three Spherical-Radial Rule Integration Algorithm

1. **Input** ϵ, m, f and N_{max}.
2. Set $N = 0$, $I = 0$, $V = 0$ and compute $F_0 = f(\mathbf{0})$
3. **Repeat**
   (a) Set $N = N + 1$ and $SR = 0$.
   (b) Generate a uniformly random orthogonal $m \times m$ matrix Q.
   (c) Generate a random $\rho \sim \mathrm{Chi}(m+2)$.
   (d) For $j = 1, 2, ..., m$ set
   $$SR = SR + f(-\rho Q\mathbf{e}_j) + f(+\rho Q\mathbf{e}_j)$$
   (e) Set $SR = F_0(1 - \frac{m}{\rho^2}) + SR/(2\rho^2)$, $D = (SR - I)/N$,
   $I = I + D$ and $V = (N-2)V/N + D^2$.

   Until $V < \epsilon^2$ or $N = N_{max}$.
4. **Output** $I \approx I(f)$, $\sigma_E = \sqrt{V}$ and N.

The random orthogonal matrices Q can be generated using a product of appropriately chosen random reflections (see Stewart, 1980). Other methods are discussed by Devroye (1986, p. 607).

If we consider the Student-t weight case, then integration by parts shows that $T_2 = \frac{m\nu}{\nu-2}T_0$, and we therefore require $\nu > 2$. In this case, SR^3 becomes

$$SR^3_{Q,\rho}(f) = f(\mathbf{0})(1 - \frac{m\nu}{(\nu-2)\rho^2}) + \frac{\nu}{\nu-2}\sum_{j=1}^{m}\frac{f(-\rho Q\mathbf{e}_j) + f(+\rho Q\mathbf{e}_j)}{2\rho^2},$$

Further analysis shows that $r^{m+1}(1+\frac{r^2}{\nu})^{-(m+\nu)/2}$ is proportional to a Beta$(\frac{m+2}{2}, \frac{\nu-2}{2})$ probability density, so the random ρ's for these SR^3 can easily be generated, and by making appropriate changes to lines 3(c) and 3(e) of the previous algorithm, a modified algorithm could be produced.

Finally, we consider the rule $\hat{SR}^5_{Q,\rho}$. For the weight $w(r) = (2\pi)^{-m/2}e^{-r^2/2}$, we find $T_4 = (m+2)mT_0$, so

$$R^5(\rho) = T_0(\ C(0)(1 - \frac{m(\rho^2+\delta^2-(m+2))}{\rho^2\delta^2}) + C(\rho)\frac{m(m+2-\delta^2)}{\rho^2(\rho^2-\delta^2)} + C(\delta)\frac{m(m+2-\rho^2)}{\delta^2(\delta^2-\rho^2)}\ ),$$

and a little algebra shows

$$\begin{aligned}\hat{SR}^5_{Q,\rho,\delta}(f) = \ & f(\mathbf{0})(1 - \tfrac{m(\rho^2+\delta^2-(m+2))}{\rho^2\delta^2}) \\ & + \tfrac{(7-m)m^2}{2(m+1)^2(m+2)}\sum_{j=1}^{m+1}\Big(\frac{(m+2-\delta^2)(f(-\rho Q\mathbf{v}_j)+f(+\rho Q\mathbf{v}_j))}{\rho^2(\rho^2-\delta^2)} \\ & \qquad + \frac{(m+2-\rho^2)(f(-\delta Q\mathbf{v}_j)+f(+\delta Q\mathbf{v}_j))}{\delta^2(\delta^2-\rho^2)}\Big) \\ & + \tfrac{2(m-1)^2}{(m+1)^2(m+2)}\sum_{j=1}^{m(m+1)/2}\Big(\frac{(m+2-\delta^2)(f(-\rho Q\mathbf{y}_j)+f(+\rho Q\mathbf{y}_j))}{\rho^2(\rho^2-\delta^2)} \\ & \qquad + \frac{(m+2-\rho^2)(f(-\delta Q\mathbf{y}_j)+f(+\delta Q\mathbf{y}_j))}{\delta^2(\delta^2-\rho^2)}\Big)\end{aligned}$$

In order to develop an algorithm for $\hat{S}R^5_{Q,\rho}$, we need a set of regular m-simplex unit vertices $\{\mathbf{v}_j\}$. We use the set given in Stroud (1971, page 345, correcting a minor misprint), where $v_{i,j} = 0$ for $0 < j < i < m+1$, $v_{i,i} = (\frac{(m+1)(m-i+1)}{m(m-i+2)})^{\frac{1}{2}}$ for $i = 1, 2, ..., m$, and $v_{i,j} = -(\frac{m+1}{(m-i+1)m(m-i+2)})^{\frac{1}{2}}$ for $0 < i < j \leq m+1$.

The joint probability density for (ρ, δ) is proportional to $(\rho\delta)^{m+1}e^{-(\rho^2+\delta^2)/2}(\rho-\delta)^2(\rho+\delta)$, which is not a standard probability density, but there is a transformation to standard densities. Consider the integral

$$P = \int_0^\infty \int_0^\infty (\rho\delta)^{m+1}e^{-(\rho^2+\delta^2)/2}(\rho-\delta)^2(\rho+\delta)h(\rho,\delta)d\rho d\delta,$$

and make the change of variables $\rho = r\sin(\theta)$, $\delta = r\cos(\theta)$. Then

$$\begin{aligned} P = & \int_0^\infty r^{2m+6}e^{-r^2/2}\int_0^{\pi/2}(sin(\theta)cos(\theta))^{m+1} && (sin(\theta)-cos(\theta))^2(sin(\theta)+cos(\theta)) \\ & && h(r\sin(\theta), r\cos(\theta))\ d\theta\ dr \\ = & \ 2^{-(m+1)}\int_0^\infty r^{2m+6}e^{-r^2/2}\int_0^{\pi/2} sin(2\theta)^{m+1} && (1-sin(2\theta))\sqrt{1+sin(2\theta)} \\ & && h(r\sin(\theta), r\cos(\theta))\ d\theta\ dr. \end{aligned}$$

Finally, let $q = \sin(2\theta)$, so that $d\theta = dq/(2\sqrt{(1-q)(1+q)})$ and then

$$\begin{aligned} P \ = \ & 2^{-(m+2)}\int_0^\infty r^{2m+6}e^{-r^2/2}(\int_0^1 \sqrt{1-q}\ q^{m+1}h(r\sin(\frac{\sin^{-1}(q)}{2}), r\cos(\frac{\sin^{-1}(q)}{2}))dq \\ & + \int_0^1 \sqrt{1-q}\ q^{m+1}h(r\sin(\sin^{-1}(q)+\frac{\pi}{4}), r\cos(\sin^{-1}(q)+\frac{\pi}{4}))dq\ )dr. \end{aligned}$$

The function $q^{m+1}\sqrt{1-q}$ is proportional to a standard Beta$(m+2, \frac{3}{2})$ probability density. The first inner integral has the resulting $\rho < \delta$ and the second has $\rho > \delta$. Because these cases are both equally likely and $\hat{S}R^5$ is symmetric in ρ and δ, there is no loss of generality in always using $\rho < \delta$. Therefore, we choose r from a Chi$(2m+7)$ density and q from a Beta$(m+2, \frac{3}{2})$ density, and then $\rho = r\sin(\frac{\sin^{-1}(q)}{2})$ and $\delta = r\cos(\frac{\sin^{-1}(q)}{2})$ will be distributed with joint probability density proportional to $(\rho\delta)^{m+1}e^{-(\rho^2+\delta^2)/2}(\rho-\delta)^2(\rho+\delta)$. We note here that the same changes of variables could also be used to provide a practical method for generating the corresponding ρ and δ for the Siegel and O'Brien (1985) finite interval rules. This question was not addressed in their paper.

We propose the following algorithm for stochastic degree five rules:

Degree Five Spherical-Radial Rule Integration Algorithm

1. **Input** ϵ, m, f and N_{max}.
2. Set $N = 0$, $I = 0$, $V = 0$ and $F_0 = f(\mathbf{0})$, and compute the m-simplex vertices $\{\mathbf{v}_j\}$.
3. **Repeat**
   (a) Set $N = N + 1$.
   (b) Generate a uniformly random orthogonal $m \times m$ matrix Q and set $\{\tilde{\mathbf{v}}_j\} = \{Q\mathbf{v}_j\}$.
   (c) Generate a random $r \sim$ Chi$(2m+7)$ and a random $q \sim$ Beta$(m+2, \frac{3}{2})$,
   and set $\rho = r\sin(\frac{\sin^{-1}(q)}{2})$, $\delta = r\cos(\frac{\sin^{-1}(q)}{2})$, $F_v = 0$ and $F_y = 0$.
   (d) For $j = 1, 2, ..., m+1$, set

$$F_v = F_v + \frac{(m+2-\delta^2)(f(-\rho\tilde{\mathbf{v}}_j)+f(+\rho\tilde{\mathbf{v}}_j))}{\rho^2(\rho^2-\delta^2)} + \frac{(m+2-\rho^2)(f(-\delta\tilde{\mathbf{v}}_j)+f(+\delta\tilde{\mathbf{v}}_j))}{\delta^2(\delta^2-\rho^2)}.$$

- For $i = 1, 2, ..., j-1$, compute $\mathbf{y} = (\tilde{\mathbf{v}}_j + \tilde{\mathbf{v}}_i)/||\tilde{\mathbf{v}}_j + \tilde{\mathbf{v}}_i||_2$, and set

$$F_y = F_y + \frac{(m+2-\delta^2)(f(-\rho\mathbf{y}) + f(+\rho\mathbf{y}))}{\rho^2(\rho^2-\delta^2)} + \frac{(m+2-\rho^2)(f(-\delta\mathbf{y}) + f(+\delta\mathbf{y}))}{\delta^2(\delta^2-\rho^2)}.$$

(e) Set

$$SR = F_0\left(1 - \frac{m(\rho^2+\delta^2-(m+2))}{\rho^2\delta^2}\right) + \frac{F_v(7-m)m^2 + 4F_y(m-1)^2}{2(m+1)^2(m+2)},$$

$D = (SR - I)/N$, $I = I + D$ and $V = (N-2)V/N + D^2$.

Until $V < \epsilon^2$ or $N = N_{max}$.

4. **Output** $I \approx I(f)$, $\sigma_E = \sqrt{V}$ and N.

If we consider the Student-t weight, then it can be shown that $T_4 = \frac{m(m+2)\nu^2}{(\nu-2)(\nu-4)}T_0$, and we must have $\nu > 4$. In this case, we could also produce a formula for $\hat{S}R^5$. However, we have not found any easy method for generating the random ρ's and δ's needed for R^5, and so we do not consider this further. Anyway, for large ν, $\frac{\Gamma(\frac{\nu+m}{2})}{\Gamma(\frac{\nu}{2})\sqrt{(\nu\pi)^m}}(1+\frac{r^2}{\nu})^{-(m+\nu)/2} \approx (2\pi)^{-m/2}e^{-r^2/2}$, so the rules that we have already developed for the multivariate Normal weight should be effective.

A possibly significant overhead cost for the SR^3 and $\hat{S}R^5$ 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 $4m^3/3$ floating point operations (flops) plus the cost of generating $m^2/2$ Normal(0,1) random numbers. For SR^3 rules the columns of Q are used for the evaluation points for $2m$ integrand values, so the overhead cost per integrand value is $2m^2/3$ flops plus the cost of generating $m/4$ Normal(0,1) random numbers. Once an integrand evaluation point is available, we expect the cost for the evaluation of the integrand to be at least $O(m)$, because there are m components for the input variable for the integrand. However, with application problems in statistics, the posterior density is often a complicated expression made up of a combination of standard elementary functions evaluated using the input variable components combined with the problem data (see the second example in the next section). Therefore, if the $O(m)$ integrand evaluation cost is measured in flops, we expect the constant in $O(m)$ to be very large, so that the $2m^2/3$ flops for the generation of the evaluation point for that integrand evaluation should not be significant unless m is very large. For $\hat{S}R^5$ rules the Q overhead cost per evaluation point drops to approximately $2m/3$ flops (plus the cost of $m/4$ Normal variates), and this is not significant compared to the integrand evaluation cost for typical statistics integration problems. We also note here that we need $m/2$ and m Normal variates, respectively, for the rules SR^1 and SR^0, per integrand evaluation, so the Normal variate overhead is higher for the two lowest degree rules. Overall, except for very simple integrands or large m values, we do not expect the overhead costs for the four rules to be significant compared to the integrand evaluation cost.

6 Examples

We begin with a simple example, where

$$I(f_1) = (2\pi)^{-4}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}...\int_{-\infty}^{\infty} e^{-\mathbf{x}^t\mathbf{x}/2}\sqrt{1+e^{\sum_{i=1}^{8} x_i/i}}\,dx_1dx_2...dx_8.$$

The following table of results we obtained using the SR rules:

Table 1: $I(f_1)$ Test Results from SR Rules

f_1	SR^0 Rules		SR^1 Rules		$\hat{S}R^3$ Rules		$\hat{S}R^5$ Rules	
Values	I	σ_E	I	σ_E	I	σ_E	I	σ_E
1000	1.66030	0.02238	1.62878	0.01586	1.63449	0.00190	1.63365	0.00021
4000	1.62666	0.01085	1.63157	0.00720	1.63219	0.00059	1.63379	0.00011
16000	1.62950	0.00545	1.63383	0.00373	1.63348	0.00035	1.63352	0.00005

These results are as expected, with much smaller standard errors for the higher degree rules.

For our second example we use a seven dimensional proportional hazards model problem discussed by Dellaportas and Wright (1991, 1992). The posterior density 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 $\rho > 0$ and $\boldsymbol{\beta} \in (-\infty, \infty)^6$. After we first transform ρ using $x_1 = \log(\rho)$, we model $p(\boldsymbol{\theta})$ with a multivariate normal approximation. So we use

$$f_2(x) = K e^{\mathbf{x}^t \mathbf{x}/2} e^{x_1} p(\mu + C(e^{x_1}, x_2, ..., x_7)^t),$$

after computing the mode μ and C for $\log(\rho) + \log(p)$. We added a scaling constant $K = p(\mu)^{-1} \approx e^{207.19}$, to prevent problems with underflow. In the following table we show results from the use of SR rules to approximate $I(f_2)$, and expected vales for each of the integration variables. The constant S in the table is a normalizing constant. For each of the respective SR rules we used the computed value of $I(f_2)$ for S.

Table 2: $I(f_2)$ Test Results from SR Rules with 120,000 f_2 Values

	SR^0 Rules		SR^1 Rules		$\hat{S}R^3$ Rules		$\hat{S}R^5$ Rules	
Integrand	I	σ_E	I	σ_E	I	σ_E	I	σ_E
$10^4 f_2$	1.5130	0.0063	1.5140	0.0053	1.5230	0.0049	1.5160	0.0058
$log(\rho) f_2/S$	0.1355	0.0010	0.1360	0.0010	0.1359	0.0005	0.1357	0.0005
$\beta_1 f_2/S$	-3.9787	0.0176	-3.9821	0.0164	-3.9839	0.0137	-3.9829	0.0162
$\beta_2 f_2/S$	1.8195	0.0080	1.8208	0.0073	1.8234	0.0058	1.8200	0.0064
$\beta_3 f_2/S$	-0.1345	0.0007	-0.1352	0.0008	-0.1348	0.0005	-0.1344	0.0005
$\beta_4 f_2/S$	-0.0211	0.0001	-0.0212	0.0001	-0.0211	0.0001	-0.0211	0.0001
$\beta_5 f_2/S$	-0.0543	0.0021	-0.0526	0.0018	-0.0531	0.0007	-0.0529	0.0008
$\beta_6 f_2/S$	0.1290	0.0008	0.1299	0.0008	0.1292	0.0004	0.1294	0.0005

For this example, the $\hat{S}R^3$ and $\hat{S}R^5$ rule results have standard errors that are smaller than the SR^0 and SR^1 rule standard errors by factors that are on average about one half. Because the decrease in standard errors is inversely proportional to the square root of the number of samples, approximately four times as much integrand evaluation work would be needed for this problem when using the SR^0 and SR^1 rules to obtain errors comparable to the errors for the $\hat{S}R^3$ and $\hat{S}R^5$ rules. These results are not as good as those for the previous problem, but the higher degree SR rules are still approximately four times more efficient than the lower degree rules. The degree five rule was not better than the degree three rule for this problem. After the standardizing transformation, the problem is apparently close enough to multivariate normal, so that a rule with degree higher than three does not produce better results. We did not find any significant difference in running times needed by the four algorithms for the results in Table 2, and this supports our analysis of the relative importance of overhead costs for the different rules.

The two examples in this section are meant to illustrate the use of the algorithms given in this paper. Much more extensive testing is needed in order to carefully compare these algorithms with other methods available for numerical integration problems in applied statistics. For some of the testing work that has been recently done with these methods we refer the interested reader to the paper by Monahan and Genz (1996). Further testing work is still in progress.

7 Concluding Remarks

We have shown how to derive low degree stochastic integration rules for radial integrals with normal and Student-t weight functions. We have also shown how these new rules can be combined with stochastic rules for the surface of the sphere, to provide stochastic rules for infinite multivariate regions with multivariate normal and Student-t weight functions. Results from the examples suggest that averages of samples of these rules can provide more accurate integral estimates than simpler Monte-Carlo importance sampling methods. The standard errors from the samples provide robust error estimates for the new rules.

REFERENCES

Chan, T. F. and Lewis, J. G. (1979), Computing Standard Deviations: Accuracy, *Communications of the ACM* **22**, pp. 526-531.

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

Deák, I. (1990), *Random Number Generation and Simulation*, Akadémiai Kiadó, Budapest.

Dellaportas, P. and Wright, D. (1991), Positive Imbedded Integration in Bayesian Analysis, *Statistics and Computing*, 1, 1-12.

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, pp. 601-606.

Devroye, Luc (1986) *Non-Uniform Random Variate Generation*, Springer-Verlag, New York.

Engels, H. (1980), *Numerical Quadrature and Cubature*, Academic Press, New York.

Evans, M. and Swartz, T. (1992), Some Integration Strategies for Problems in Statistical Inference. *Computing Science and Statistics* **24**, pp. 310-317.

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

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

Johnson, N. L. and Kotz, S. (1970), *Continuous Univariate Distributions-I*, John Wiley and Sons, New York.

Monahan, J. F. (1987), An Algorithm for Generating Chi Random Variables, *ACM TOMS* **13**, pp. 168-171 (Correction **14**, p. 111).

Monahan, J. F. and Genz, A. (1996), A Comparison of Omnibus Methods for Bayesian Computation, to appear in *Computing Science and Statistics* **27**.

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

Siegel, A.F. and O'Brien, F. (1985), Unbiased Monte Carlo Integration Methods with Exactness for Low Order Polynomials, *SIAM. J. Sci. Stat. Comput.* **6**, pp. 169-181.

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

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