Estimation of multidimensional integrals: is Monte Carlo the best method?

1993 J. Phys. A: Math. Gen. 26 943

(http://iopscience.iop.org/0305-4470/26/4/022)

J. Phys. A: Math. Gen. 26 (1993) 943–953. Printed in the UK

Estimation of multidimensional integrals: is Monte Carlo the best method?

E J Janse van Rensburg† and G M Torrie‡

† Department of Mathematics and Statistics, York University, 4700 Keele street, North York, Ontario M3J 1P3, Canada

‡ Department of Mathematics and Computer Science, Royal Military College, Kingston, Ontario K7K 5L0, Canada

Received 21 June 1992, in final form 23 October 1992

Abstract. The effectiveness of a deterministic algorithm recently proposed by Woźniakowski for the numerical estimation of high-dimensional integrals based on so-called optimal sampling points is studied by using it to compute virial coefficients for the hard-sphere fluid. Although the algorithm is inherently more efficient than the familiar Monte Carlo method for sufficiently large samples, in practice this advantage diminishes rapidly with increasing dimensionality. For the virial coefficient problem, we find the new algorithm to be noticeably more efficient than crude Monte Carlo for dimensions less than 10 and of comparable efficiency for larger dimensionality up to 14.

1. Introduction

Monte Carlo methods (Hammersley and Handscomb 1964) have proven to be powerful techniques for the estimation of multi-dimensional integrals. However, it is known that random sampling from a distribution (inherent in the Monte Carlo process) is not the optimal design for the numerical evaluation of multi-dimensional integrals, and alternative methods which are applicable in limited circumstances have been examined in the mathematical and scientific literature (Haselgrove 1961, Conroy 1967, see Niederreiter 1978 for hundreds of references). One such potentially useful alternative based on so-called optimal sampling points has recently been proposed by Woźniakowski (1991). In this paper we examine the utility of this algorithm in a practical application of the sort encountered in the statistical mechanics of classical fluids.

The problem of numerical integration is abstractly presented by the following: Find the best numerical approximation, I_n, to

$$I = \int_\Omega f(x)\,\mathrm{d}x \tag{1.1}$$

with the least computational effort, where the integrand f is to be integrated over the domain Ω, which is usually assumed to be compact, and (in this paper) is assumed to be a subset of $\mathcal{R}^d$, (d-dimensional real space). In a crude Monte Carlo scheme,

0305-4470/93/040943+11$07.50

I_n is obtained by computing f at n random points in Ω. Let $\{r_i\}_{i=0}^{n-1}$ be n random points in Ω; then (if the r_i are distributed uniformly)

$$I_n = |\Omega| \sum_{i=0}^{n-1} f(r_i)/n \tag{1.2}$$

where $|\Omega| = \int_\Omega \mathrm{d}x$. A very important issue in (1.2) is the possibility of computing a confidence interval on the approximation. Usually, the root of the mean square error, σ, is calculated. Standard probability theory provides a straightforward estimate for σ (for example, Dowdy and Wearden 1983),

$$\sigma^2 = \left[|\Omega| \sum_{i=0}^{n-1} f^2(r_i)/n - \left(|\Omega| \sum_{i=0}^{n-1} f(r_i)/n \right)^2 \right] \Big/ (n-1) . \tag{1.3}$$

An important observation (which follows immediately from (1.3)) is that

$$\sigma = \mathrm{O}\left(\frac{1}{\sqrt{n}} \right) . \tag{1.4}$$

Hence, using random sampling one can approximate (1.1) at a *cost* of $\mathrm{O}(\sigma^{-2})$.

In the case of the optimal sampling points proposed by Woźniakowski (1991), by contrast, the cost of estimating an integral to within an average case error ϵ is $\Theta(\epsilon^{-1}(\log \epsilon^{-1})^{(d-1)/2})$. As a consequence of the logarithmic term, then, for small enough ϵ this cost will be less than that of a Monte Carlo estimation with a comparable error. The issue from a practical standpoint is whether this inherent advantage is realized for the sort of precision (typically, 0.01%) normally sought for statistical mechanical calculations. As a case-study, we re-examine the estimation of the virial coefficients of hard discs and hard spheres. In the following section we review the basis of the algorithm of Woźniakowski and consider the calculation of confidence intervals. In section 3 we describe the use of this algorithm to compute the virial coefficients and compare our results to the earlier ones obtained by Ree and Hoover (1964, 1967) by Monte Carlo methods.

2. Numerical integration and Hammersley points

2.1. *Hammersley points*

The sampling points that form the basis of the method are related to classical Hammersley points (Niederreiter 1978, Roth 1980, Woźniakowski 1991). Let $p_1, p_2, \ldots, p_{d-1}$ be the first $d-1$ prime numbers. Any integer $k \geqslant 0$ can be expressed as $\sum_{i=0}^{\lceil \log k \rceil} c_i p_j^i$, where $c_i \in [0, p_j - 1]$ are integers. The radical inverse function is defined as $\phi_{p_j}(k) = \sum_{i=0}^{\lceil \log k \rceil} c_i p_j^{-1-i}$. In these expressions, $\lceil x \rceil$ is the smallest integer larger than x, and the $\log k$ is taken to base p_j. Let

$$u_k = (\phi_{p_1}(k), \phi_{p_2}(k), \ldots, \phi_{p_{d-1}}(k)) . \tag{2.1}$$

Then the sequence $\{u_k\}_{k=0}^{n-1}$ is a sequence of $(d-1)$-dimensional points which are related to *shifted Hammersley points* by

$$\{z_k(t)\}_{k=0}^{n-1} = \{(k+t)/n, u_k\}_{k=0}^{n-1} \qquad \text{such that} \qquad 0 \leqslant k+t < n. \tag{2.2}$$

If $t = 0$, then (2.2) defines the *Hammersley points*. For example, suppose that $d = 3$ and $k = 6$. Then $6 = \sum_{i=0}^{2} c_i 2^i$, where $c_0 = 0$ and $c_1 = c_2 = 1$, and $6 = \sum_{i=0}^{2} d_i 3^i$, where $d_0 = d_2 = 0$ and $d_1 = 2$. The radical inverses are $\phi_2(6) = 2^{-2} + 2^{-3} = \frac{3}{8}$ and $\phi_3(6) = 2 \times 3^{-2} = \frac{2}{9}$, and if $n > 6$, then $z_6(t) = ((6+t)/n, \frac{3}{8}, \frac{2}{9})$. Let Ω in (1.1) be the unit hypercube and suppose that f belongs to the class of real continuous functions equipped with the classical Wiener sheet measure; then an approximation to I is given by $I_n^H = \sum_{k=0}^{n-1} f(1 - z_k(t))/n$. Woźniakowski proved the following theorem:

Theorem (Woźniakowski 1991). The average case error of I_n^H, ϵ, is related to n by

$$n = \Theta\left(\epsilon^{-1}(\log \epsilon^{-1})^{(d-1)/2}\right).$$

Observe that the final definition of $\{z_k(t)\}_{k=0}^{n}$ depends on n, and that the shift t in (2.2) is not specified. In this paper we examine the estimation of integrals using instead the potentially infinite sequence defined by

$$\{\tilde{z}_k(t)\}_{k=1}^{\infty} = \left\{\left(\phi_{p_1}(k), \phi_{p_2}(k), \ldots, \phi_{p_d}(k)\right)\right\}_{k=1}^{\infty}. \tag{2.3}$$

In this case, the average case $\text{cost}(I_n^H, \epsilon)$ (which is proportional to n, the number of sample points thrown) of approximating I to within an average case error which is at most ϵ is given by (Woźniakowski 1991)

$$\text{cost}(I_n^H, \epsilon) = O\left(\epsilon^{-1}(\log \epsilon^{-1})^d\right). \tag{2.4}$$

That is, one loses some efficiency due to a larger power of the logarithmic factor as compared to Woźniakowski's theorem above.

2.2. Estimating confidence intervals

The assignment of confidence intervals is an essential part of the process of estimating multi-dimensional integrals. In the case of Monte Carlo integration there is a simple way to estimate confidence intervals based on (1.4). Assume that $\sigma = C/\sqrt{n}$; the proportionality factor C is obtained as follows. The total sample size of $n = Nm$ points is partitioned into N blocks of size m and the N block averages $\{I_m^{\text{MC}}(i)\}_{i=1}^{N}$ are treated as independent estimates normally distributed around the sample mean $I_n^{\text{MC}} = \sum_{i=1}^{N} I_m^{\text{MC}}(i)/N$. The 66% confidence interval in the sample mean, σ_{MC}, is then extrapolated from the variance of the distribution of block averages, $\text{Var}(I_m^{\text{MC}})$, by $\sigma_{\text{MC}} = \sqrt{\text{Var}(I_m^{\text{MC}})/N}$. For crude Monte Carlo estimation the block averages are completely uncorrelated and σ_{MC} must be independent of m.

In the case of Hammersley point integration based on the sample points defined by (2.3) the average case error is related to the sample size by (2.4). If we are to estimate the confidence interval in the spirit of a block analysis as set out above for a Monte Carlo integration, then we need the following hypothesis.

Hypothesis. Let σ_H be a confidence interval on an approximation I_n^H (obtained by sampling Hammersley points). Then there exists a constant C such that

$$n = C(\log \sigma_H^{-1})^d / \sigma_H \tag{2.5}$$

where n is the number of points sampled in the estimation.

Should we be able to calculate C in (2.5), then we can solve for σ_H as a function of n using a standard numerical technique. This approach is similar to the situation for Monte Carlo integration, as explained above, except that (2.5) replaces (1.4). A particularly simple way of measuring σ_H after m points are sampled is to perform a block analysis as set out in the case of Monte Carlo integration. The implicit assumption in this scheme is that each sequence of m Hammersley points, if m is large enough, provides an uncorrelated measurement of the integral, and that these measurements are normally distributed. The confidence interval is then extrapolated to a sample size of n in exactly the same fashion as in the Monte Carlo integration, but using (2.5).

3. Numerical results

The virial series for the pressure p of an imperfect gas is a power series in the *number density* $\rho \equiv n/V$,

$$\frac{pV}{nkT} = 1 + B_2\rho + B_3\rho^2 + \cdots \tag{3.1}$$

where n is the number of atoms in the volume V at temperature T, k is Boltzmann's constant and the virial coefficients B_i are sums of integrals over the coordinates of i particles. The calculation is organized using graphical expansion methods and we follow here Ree and Hoover's modification of the original Ursell–Mayer formalism (Hill 1987, Kilpatrick 1971, Ree and Hoover 1964)

$$B_i = \frac{1-i}{i!} \lim_{V\to\infty} \frac{1}{V} \int \mathrm{d}\mathbf{r}_1 \mathrm{d}\mathbf{r}_2 \ldots \mathrm{d}\mathbf{r}_i V_i \tag{3.2}$$

where

$$V_i = \sum_{G_i}{}^{*} a(G_i) \prod_{\substack{k>l \\ k,l\in G_i}} (I(k,l) f(\mathbf{r}_k,\mathbf{r}_l) + (1 - I(k,l)) \hat{f}(\mathbf{r}_k,\mathbf{r}_l)) . \tag{3.3}$$

The summation in (3.3) is over irreducible graphs with i vertices, G_i; I is an indicator which is 1 if the edge (k,l) is in the edge-set of G_i and 0 otherwise, f is the Mayer f-function defined by $f(\mathbf{r}_k,\mathbf{r}_l) = \exp(-\phi(\mathbf{r}_k,\mathbf{r}_l)/kT) - 1$ where ϕ is the pair-potential, and $\hat{f} = f + 1$. $a(G_i)$ is a symmetry factor which depends on the number of possible labellings of G_i and the details of the reformulation described by Ree and Hoover (1964).

The numerical difficulties posed by the estimation of the integrals arising in (3.2) are typical of quadrature problems in classical statistical mechanics. The dimensionality is high, the integrand varies greatly in magnitude over the coordinate space with the overall result involving cancellation of opposing contributions, and the integrand is sufficiently complex that improvements over the uniformly random sampling of crude Monte Carlo may be hard to devise.

3.1. Computational details

In its most elemental form a crude Monte Carlo estimate of B_i would consist of the generation of a succession of sets of i random vertices, uniformly distributed in a suitably large volume V, and the evaluation for each such vertex set of each of the graphs G_i contributing to (3.3). In this paper, however, we consider only hard discs and spheres,

$$\phi(\mathbf{r}_k, \mathbf{r}_l) = \begin{cases} \infty & |\mathbf{r}_k - \mathbf{r}_l| < 1 \\ 0 & \text{otherwise} \end{cases} \tag{3.4}$$

so that each f (and $\hat{f}$) is a step-function $f(|\mathbf{r}_k - \mathbf{r}_l|) = \theta(|\mathbf{r}_k - \mathbf{r}_l|) - 1$ where

$$\theta(|\mathbf{r}|) = \begin{cases} 0 & |\mathbf{r}| < 1 \\ 1 & |\mathbf{r}| \geqslant 1\,. \end{cases} \tag{3.5}$$

As a result, only configurations in which each vertex lies within unit distance of at least one other need be considered since otherwise the integrand is zero for all G_i. Following Ree and Hoover (1964, 1967) we enforce this by sampling points from a product space of discs (two dimensions) or spheres (three dimensions). In (3.2) one integration (over $\mathbf{r}_i$) is unnecessary if we fix the origin at the centre of the particle labelled i. Then, the first point is sampled within unit distance of the origin, the second from within unit distance of the first, and so forth. For each configuration generated in this manner the remaining interparticle separations are examined to determine the connectivity of the corresponding graph. For example, the graph of figure 1 corresponds to the integral

$$\begin{aligned} \int_\Omega f(0,\mathbf{r}_1)\hat{f}(0,\mathbf{r}_2)f(0,\mathbf{r}_3)f(\mathbf{r}_1,\mathbf{r}_2)\hat{f}(\mathbf{r}_1,\mathbf{r}_3)f(\mathbf{r}_2,\mathbf{r}_3)\mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2\mathrm{d}\mathbf{r}_3 \\ = \int_\Omega \theta(|\mathbf{r}_1|)(1-\theta(|\mathbf{r}_2|))\theta(|\mathbf{r}_3|)\theta(|\mathbf{r}_2-\mathbf{r}_1|)(1-\theta(|\mathbf{r}_3-\mathbf{r}_1|)) \\ \times \theta(|\mathbf{r}_3-\mathbf{r}_2|)\mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2\mathrm{d}\mathbf{r}_3\,. \end{aligned} \tag{3.6}$$

For $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3 \in \mathcal{R}^2$ the corresponding contribution to the fourth virial coefficient for hard disks will be proportional to π^3 times the fraction of occurrences of this particular graph in the sample. In order to make a rapid identification of the graph for a particular configuration we have used the following hash table addressing scheme for the irreducible graphs.

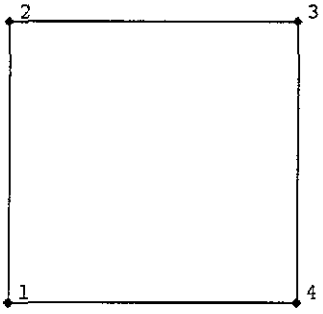

Figure 1. The graph 101101.

In a graph G with i vertices labelled 1 through i each possible edge can be labelled by the ordered pair (k,l), $k < l$ of vertices at its endpoints. We order the $i(i-1)/2$ possible edges as $(1,2),(1,3),\ldots,(1,i),(2,3),(2,4),\ldots,(i-1,i)$. There is now a natural representation of G in a string of $i(i-1)/2$ binary digits; the digit j takes the value 1 if the edge it represents is present in the graph, else it is zero. For example, the graph of figure 1 has the representation 101101 and, in an obvious notation, we may abbreviate the corresponding integrand in (3.6) as $f_{101101}(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3)$. (Of course, each *unlabelled* graph will have as many distinct representations in this scheme as it has distinct labellings. For instance, the graph in figure 1 has a second labelling with the representation 110011.) This representation of graphs as strings of binary digits allows us to assign to each graph a unique *numeral* namely, the positive integer corresponding to its binary string. This numeral may then be used to compute an address for the graph in a hash table (Knuth 1973) of manageable size, with collisions resolved using linear probing. In this way we first construct a list of all the irreducible graphs with i vertices in a hash table. For each configuration generated in the sampling we may determine whether it is irreducible (i.e. contributes to B_i) by computing its numeral and then querying the table in O(1) CPU-time.

3.2. *Computing confidence intervals*

The calculation of confidence intervals and their dependence on sample size is essential for the comparison of the two integration methods. We present here numerical evidence in support of the block average approach hypothesized in section 2.2. Figure 2 shows the distribution of block averages for the integral (3.6) corresponding to the graph in figure 1, based on a total sample size of 1.5×10^7 configurations and a block-size of 10^4. Assuming the distribution to be normal, we can compute the variance and the confidence interval of the mean can be calculated by extrapolating under the hypothesis (2.5). The extrapolation will be justified if it is independent of the block size. The results are shown in table 1 (upper section) for a

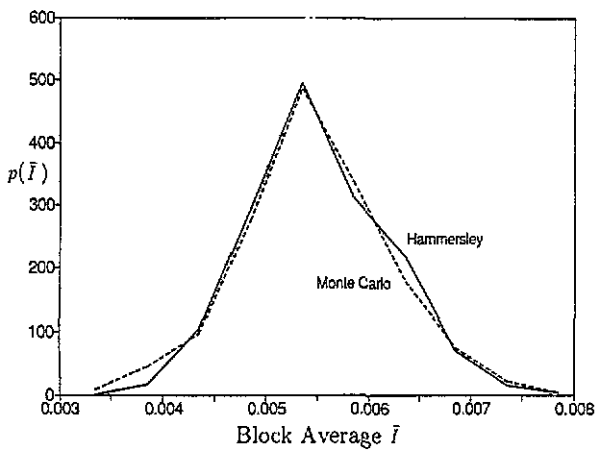

Figure 2. Distributions of the block averages of the integral corresponding to the graph 101101. The full curve is the distribution according to Hammersley points; the broken curve corresponds to Monte Carlo sampling. The variance in these distributions are assumed to indicate the confidence interval in a single block average. The block size was taken to be 10 000 in this case.

Table 1. The dependence of the confidence interval on the block size. The integral corresponds to the graph 101101 (upper section) and 101101100101111 (lower section). The sample size was 2.5×10^6 (upper section) and 5×10^6 (lower section).

Block size	Hammersley	Monte Carlo
1000	0.00076	0.0015
2000	0.00075	0.0015
5000	0.00081	0.0015
10000	0.00079	0.0015
25000	0.00082	0.0015
50000	0.00081	0.0015
1000	0.00014	0.00057
2000	0.00014	0.00057
5000	0.00014	0.00057
10000	0.00014	0.00057
25000	0.00014	0.00057
50000	0.00016	0.00057

Table 2. Virial coefficients in two (upper section) and three (lower section) dimensions.

Coefficient		σ_H	σ_{MC}	*Dimensions*	*Sample size*
$B_4/(B_2^3)$	0.532 15	0.000 04	0.000 30	6	2×10^7
$B_6/(B_2^5)$	0.198 82	0.000 07	0.000 15	10	10^8
$B_7/(B_2^6)$	0.114 90	0.000 09	0.000 20	12	10^8
$B_6/(B_2^5)$	0.039 19	0.000 07	0.000 15	15	10^8
$B_7/(B_2^6)$	0.013 11	0.000 08	0.000 20	18	10^8

wide range of block sizes along with the confidence limit for a Monte Carlo sample of the identical size. (For the present case of hard particles the crude Monte Carlo sampling is just a Bernoulli process. As Kratky (1977) observed in his reassessment of the work of Ree and Hoover, the confidence limit can then be computed directly from the mean itself without explicitly calculating block averages. That is the procedure we have used here.) The lower section of table 1 shows the results of a similar test applied to the higher-dimensional integral

$$\int_\Omega f_{101101100101111}(r_1, r_2, r_3, r_4, r_5)\mathrm{d}r_1\mathrm{d}r_2\mathrm{d}r_3\mathrm{d}r_4\mathrm{d}r_5 .$$

In both cases the data clearly support our use of an extrapolation based on block averages to estimate confidence intervals for the Hammersley point integrations.

3.3. The sixth and seventh virial coefficients

The definitive calculation of virial coefficients for fluids of hard discs and hard spheres remains that of Ree and Hoover (1964, 1967) whose work provides a useful benchmark against which to test Hammersley point integration. As an initial test we computed B_4 for hard discs, a six-dimensional integral. The results for a sample of 2×10^7 points are listed in table 2 (upper section). σ_H is the confidence interval on the Hammersley sampling and σ_{MC} is the theoretical Monte Carlo confidence interval referred to above, computed for a hypothetical Monte Carlo sample of equal size. For this relatively low-dimensional integral there is quite a significant gain in

efficiency in the Hammersley sampling scheme over the Monte Carlo sampling. The value obtained for B_4 is consistent with the earlier, less precise estimate of Ree and Hoover (1964), $B_4/(B_2^3) = 0.5327 \pm 0.0005$. B_6 and B_7 were similarly obtained (over 1×10^8 Hammersley points); these results are also listed in table 2 (upper section). They are again consistent with the previous estimates of Ree and Hoover who reported $B_6/(B_2^5) = 0.1992 \pm 0.0008$ and $B_7/(B_2^6) = 0.1141 \pm 0.0005$. A comparison of the third and fourth columns of table 2 (upper section) shows the Hammersley point integration to be superior to Monte Carlo in every case but that the gain in efficiency diminishes as the dimensionality increases.

The sixth and seventh virial coefficients were also calculated in three dimensions. These results are listed in table 2 (lower section). Again, they are consistent with those obtained by Ree and Hoover (1964, 1967), $B_6/(B_2^5) = 0.0386 \pm 0.0004$ and $B_7/(B_2^6) = 0.0138 \pm 0.0006$. (As pointed out by Kratky (1977) the confidence limit for B_7 originally reported by Ree and Hoover is too low. We have followed his practice here in assigning the appropriate value for a Monte Carlo sample of the size reported by Ree and Hoover.) For the sample size we have used, 10^8 points, Hammersley point integration appears to retain its superiority over Monte Carlo integration in all cases.

3.4. Discussion

For other sample sizes the relative efficiencies of the two integration methods will behave differently because of the different functional dependencies on sample size in (1.4) and (2.5). The comparison will also be strongly dependent on the dimensionality of the integral through its appearance in (2.5). Consider first the low dimensionality case, B_4 for hard discs, $d = 6$. In figure 3 we have plotted the extrapolated the confidence intervals for B_4 for both Monte Carlo and Hammersley sampling, using (1.4) and (2.5). The logarithm of σ is plotted using a vertical scale on which each unit corresponds to a difference of one significant figure in the confidence interval and the sample size is indicated, also logarithmically, on the horizontal scale. The two filled squares on the curves represent the confidence limits for our actual sample

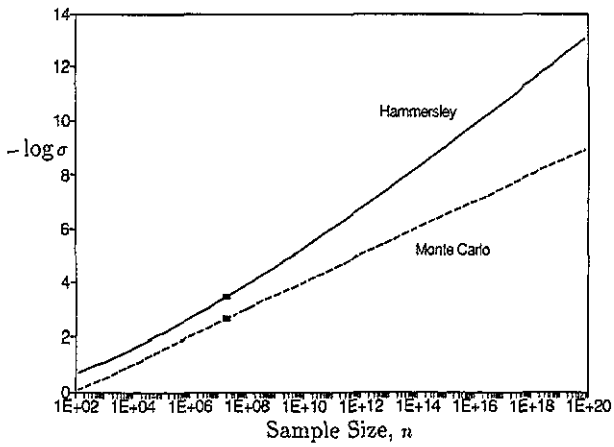

Figure 3. The confidence interval extrapolated as a function of the sample size n for $B_4/(B_2^3)$ for hard discs. The full curve corresponds to Hammersley sampling, and the broken curve to Monte Carlo sampling.

of 2×10^7 points and the vertical distance between them corresponds to the difference in precision of just under one significant digit shown in table 2 (upper section). This advantage of the Hammersley point integration would be expected to increase with larger sample sizes as indicated by the increasing separation between the two curves in the figure. For 10^{14} points there would be a gain of about 2 significant figures, at 10^{20} points about 4 significant figures.

The curvature (on the log scale) of the Hammersely sampling error function (2.5) in figure 3 suggests, however, that more complex behaviour can occur, and this indeed turns out to be the case for the higher dimensional integrals, B_6 and B_7 for hard spheres. Figure 4 is the log–log plot of the extrapolated confidence interval versus the sample size for B_6. Again, the filled squares indicate the confidence limits for the sample size we actually used, 10^8 points, and show the superiority of Hammersley sampling. However, the curves indicate that this advantage would be lost if the sample size were *increased* to 10^{10} points and would be regained only with a sample size of about 10^{18} points. In the interval between, our data show a slight advantage for the Monte Carlo method, but the difference is not big enough to be definitive. The message, however, is clear. There would be no advantage in using Hammersley points rather than Monte Carlo to improve the estimate of B_6 beyond that in table 2 (lower section). An examination of our data shows that a large contribution to the confidence interval comes from the complete graph K_6 (in our notation, this is the graph 111111111111111) (figure 5). We find for this graph the value $-0.05895 \pm \frac{0.00005}{0.00006}$ where the Hammersley point confidence limit is shown above and the Monte Carlo confidence limit below. Already one observes that there is little difference between the confidence intervals for the two methods. The cycle graph C_6 (in our notation 100011000100101, figure 5) also contributes a large part to the virial coefficient, and our data indicate that this graph is computed just as effectively by Monte Carlo. We find a value of $0.01095 \pm \frac{0.00004}{0.00004}$.

A similar trend is observed if we consider the seventh virial coefficient of hard spheres instead. Again, the graphs K_7 and C_7 contributed most to the confidence interval. We find for K_7 a value of $-0.02316 \pm \frac{0.00004}{0.00005}$ and for C_7 a value of $-0.00813 \pm \frac{0.00007}{0.00005}$. That is, C_7 is computed more effectively by Monte Carlo.

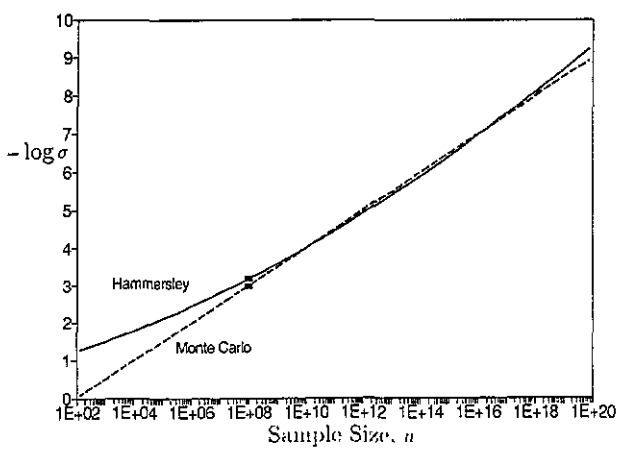

Figure 4. The confidence interval extrapolated as a function of the sample size n for $B_6/(B_2^5)$ for hard spheres.

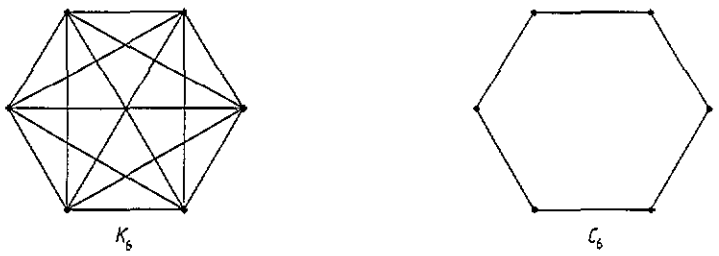

Figure 5. The complete graph K_6 and the cycle graph C_6. Both these graphs contribute significantly to the confidence interval in B_6.

4. Conclusions

In this paper we set out to study the effectiveness of a technique based on 'optimal sampling' using Hammersley points in place of the familiar Monte Carlo method to evaluate high-dimensional integrals. Using the evaluation of virial coefficients for the hard disc and hard sphere fluids as test cases, we found Hammersley sampling proved notably more effective than simple Monte Carlo integration provided that the dimensionality was not too high ($d < 10$) and that this advantage increased with sample size (cf figure 3). The method is, however, handicapped by the logarithmic correction term in (2.5) which degrades its performance with increasing dimensionality. This handicap could be mitigated somewhat by using instead the shifted Hammersley points of (2.2). In that case, one hypothesizes that the confidence interval would be related to the sample size n by $n = C(\log \sigma_H^{-1})^{(d-1)/2}/\sigma_H$. If so, this would double the range of dimensions where the sampling would prove effective. Hence, Monte Carlo remains a very competitive method for the range of dimensions and numbers of sampling points used in the calculation of B_6 and B_7 for hard spheres, values which are representative of problems in statistical physics. In addition, the simple crude Monte Carlo method used here as the benchmark can often be greatly improved for particular problems through the use of appropriate variance reduction techniques (Hammersley and Handscomb 1964), such as the method of important sampling, which has found wide applications in the integration of integrals of very high dimensionality. This additional flexibility inherent to the Monte Carlo method can be very effective in reducing the cost of an integration, and makes the Monte Carlo method probably the superior technique in most applications.

Overall, however, the promise held out by the stronger decay with sample size of the error estimate in (2.5) has not been fulfilled in a very satisfying way in practice. To realize the advantage implicit in this functional dependence when the dimensionality is high one must use sample sizes which are astronomical. For the 15-dimensional integral in figure 4, no advantage is noted until the sample sizes are of $O(10^{22})$; if we sample 10^7 points per hour (which is reasonable on a fast workstation), we shall need $O(10^{18})$ seconds to observe an advantage in using Hammersley sampling. This is about the age of universe (and in that case, we will know $B_6/(B_2^5)$ to about 10 significant digits).

Acknowledgments

We have benefitted from numerous discussions of this work with John Valleau who

offered many helpful comments and suggestions. This research has been supported by the Defence Research Board of Canada.

References

Conroy H 1967 *J. Chem. Phys.* **47** 5307
Dowdy S and Wearden S 1983 *Statistics for Research* (New York: Wiley)
Hammersley J M and Handscomb D C 1964 *Monte Carlo Methods* (London: Methuen)
Haselgrove C B 1961 *Math. Comput.* **15** 323
Hill T L 1987 *Statistical Mechanics: Principles and Selected Applications* (New York: Dover)
Kilpatrick J E 1971 *Adv. Chem. Phys.* **20** 39
Knuth D E 1973 *The Art of Computer Programming* vol 3 (Reading MA: Addison-Wesley)
Kratky K W 1977 *Physica* **87A** 584
Metropolis N, Rosenbluth A W, Rosenbluth M N, Teller A H and Teller E 1953 *J. Chem. Phys.* **21** 1087
Niederreiter H 1978 *Bull. Am. Math. Soc.* **84** 957
Ree F H and Hoover G H 1964 *J. Chem. Phys.* **40** 939
—— 1967 *J. Chem. Phys.* **46** 4181
Roth K F 1980 *Acto. Arith.* **37** 67
Woźniakowski H 1991 *Bull Am. Math. Soc.* **24** 185