COMPUTER SCIENCES
DEPARTMENT

University of Wisconsin-
Madison

GENERALIZED ANTITHETIC TRANSFORMATIONS
FOR MONTE CARLO SAMPLING

by

John H. Halton

Computer Sciences Technical Report #408

December 1980

GENERALIZED ANTITHETIC TRANSFORMATIONS FOR MONTE CARLO SAMPLING

John H. Halton

Computer Sciences Department
The University of Wisconsin
1210 West Dayton Street
Madison, Wisconsin 53706

SYNOPTIC ABSTRACT

The method of "antithetic variates" for Monte Carlo sampling was invented and named by Hammersley and Morton (1956) and has been generalized by Halton and Handscomb (1957) and Laurent (1961). Given only that a Monte Carlo estimator possesses derivatives up to a certain order, in the sample space, transformations of the estimator are supplied (independent of the particular estimator used), which reduce the variance of the resulting estimates in a very marked degree. In the present paper, the explicit forms of these transformations are derived. It is demonstrated that, contrary to common belief, the transformations of Halton and Handscomb are more efficient than those proposed by Laurent.

Key Words and Phrases: numerical integration; Monte Carlo; statistical sampling; variance reduction; antithetic transformation.

1. INTRODUCTION.

We consider the evaluation of an integral of the form

$$\theta = \int_0^1 f(x)\ dx \tag{1}$$

by Monte Carlo sampling [see Halton (1970).] If ξ denotes a *canonical random variable* (that is, a random variable distributed with uniform probability density in the unit interval $U = [0, 1)$), then *crude Monte Carlo* consists in sampling the (*primary*) estimator $f(\xi)$, which, by repeated independent trials, yields the (*secondary*) estimate

$$\psi_k(\xi_1, \xi_2, \ldots, \xi_k) = \frac{1}{k} \sum_{i=1}^{k} f(\xi_i); \tag{2}$$

and, since the expectations are

$$\mathrm{E}[\psi_k] = \mathrm{E}[f] = \theta, \tag{3}$$

Kolmogorov's form of the Strong Law of Large Numbers [see, e.g., Gnedenko (1963) p. 245, or Kingman and Taylor (1966) p. 344] shows that, if (as we shall suppose) θ exists and is finite,

$$\psi_k \to \theta \text{ (almost surely) as } k \to \infty. \tag{4}$$

If we further suppose that the integral

$$\int_0^1 [f(x)]^2\ dx = \theta^2 + \mathrm{var}[f] \tag{5}$$

exists and is finite, then the variance

$$\mathrm{var}[\psi_k] = \mathrm{var}[f]/k \to 0 \text{ as } k \to \infty. \tag{6}$$

Now Chebyshev's inequality [see, e.g., Feller (1968) p. 233, Gnedenko (1963) p. 225, Kingman and Taylor (1966) p. 288, or Loève (1977) p. 11] shows us that, for any $a > 0$,

$$\mathrm{Prob}\left(|\psi_k - \theta| \geqslant a \sqrt{\mathrm{var}[\psi_k]}\right) \leqslant \frac{1}{a^2}; \tag{7}$$

and the Central Limit Theorem [see, e.g., Feller (1968) p. 244, Gnedenko (1963) p. 293, or Kingman and Taylor (1966) p. 348] says that

$$\text{Prob}\left(|\psi_k - \theta| \geqslant a\sqrt{\text{var}[\psi_k]}\right) \rightarrow \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \int_a^\infty e^{-x^2/2}\, dx \qquad (8)$$

as $k \to \infty$. Both results indicate that the distribution of ψ_k about θ is scaled by the dimension $\sqrt{\text{var}[\psi_k]} = \sqrt{(\text{var}[f]/k)}$, which is therefore used as a measure of the error $|\psi_k - \theta|$ to be expected. The slow decrease of this quantity with increasing k has spurred much effort to reduce the variance var$[f]$ of the primary estimator f by transforming it in some way. The new primary estimator may be any integrable function g on $\overline{U} = [0, 1]$, for which

$$\int_0^1 g(x)\, dx = \int_0^1 f(x)\, dx = \theta. \qquad (9)$$

Much ingenuity has been devoted to devising techniques for transforming functions f into functions g so as to reduce var$[g]$ appreciably while preserving the integral θ. One approach seeks an approximation φ to f on $\overline{U}$, which is easy to integrate. One can then either define

$$g(x) = f(x) - \varphi(x) + \int_0^1 \varphi(y)\, dy \qquad (10)$$

and sample $g(\xi)$ as before —— this is called *correlated sampling* —— or define

$$g(x) = \frac{f(x)}{\varphi(x)} \int_0^1 \varphi(y)\, dy \qquad (11)$$

and sample $g(\eta)$, where η is distributed with probability density

$$\varphi(\eta)/\int_0^1 \varphi(y)\, dy \qquad (12)$$

—— this is called *importance sampling* —— [for further discussion and references, see Halton (1970).] The approach which we shall consider here seeks instead to construct general transformations applicable to a broad range of functions to reduce their variance.

It is clear that, if the problem is that of evaluating an integral

$$\theta = \int_a^b h(y)\, dy, \qquad (13)$$

then the simple transformation $y = a + (b - a)x$ yields (1) with

$$f(x) = (b - a)\, h[a + (b - a)x]. \qquad (14)$$

In the case of an integral over an infinite range, it is easy to transform the range into $\overline{U}$; but it is not obvious that this can be done without affecting the differentiability of the integrand at the ends of the range: we shall see later that this is important to the validity of the method presented here.

2. THE METHOD OF ANTITHETIC VARIATES.

This method was invented and named by Hammersley and Morton (1956). The question of what general classes of antithetic transformations are best was discussed by Hammersley and Mauldon (1956) and Handscomb (1958). The extension of the method to multi-dimensional integrals was examined by Morton (1957) and Halton and Handscomb (1957). It is the one-dimensional treatment in Halton and Handscomb (1957), with reference to the work of Laurent (1961) and Handscomb (1964), which concerns us here.

Halton and Handscomb (1957) consider the class of transformations

$$g(x) = \mathbb{A}f(x) = \sum_{j=1}^{t} \kappa_j\, f(\lambda_j + \mu_j x), \qquad (15)$$

$g(\xi)$ being sampled repeatedly and independently to yield secondary estimates

$$\gamma_k(\xi_1, \xi_2, \ldots, \xi_k) = \sum_{i=1}^{k} g(\xi_i). \qquad (16)$$

In order that

$$\gamma_k \to \theta \text{ (a.s.) as } k \to \infty, \qquad (17)$$

with θ finite, it is necessary and sufficient that

$$\mathrm{E}[\gamma_k] = \mathrm{E}[g] = \mathrm{E}[\mathbb{A}f] = \theta; \qquad (18)$$

and this is achieved when

$$\sum_{j=1}^{t} \kappa_j \int_0^1 f(\lambda_j + \mu_j x)\, dx = \sum_{j=1}^{t} \frac{\kappa_j}{\mu_j} \int_{\lambda_j}^{\lambda_j+\mu_j} f(y)\, dy = \theta. \qquad (19)$$

In particular, these authors consider the transformations (which satisfy (15) and (19))

$$\mathfrak{U}f(x) = \frac{1}{2}\,[f(x) + f(1-x)] \qquad (20)$$

and

$$\mathfrak{U}_n f(x) = \frac{1}{n}\sum_{j=0}^{n-1} f\left(\frac{x+j}{n}\right); \qquad (21)$$

and they prove that

$$\begin{aligned}\operatorname{var}[\mathfrak{U}_n f] &= \sum_{r,s\geqslant 0} \frac{(-1)^r\,\Delta_r\,\Delta_s\,B_{r+s+2}}{(r+s+2)!\quad n^{r+s+2}} \\ &= \frac{\Delta_0^{\,2}}{12n^2} + \frac{\Delta_1^{\,2} - 2\Delta_0\Delta_2}{720n^4} + \frac{\Delta_2^{\,2} - 2\Delta_1\Delta_3 + 2\Delta_0\Delta_4}{30240n^6} + \ldots, \qquad (22)\end{aligned}$$

where the B_t are Bernoulli numbers [see Abramowitz and Stegun (1964) §23, whose notation we follow; or Hardy and Wright (1962) pp. 90 and 245, Jahnke and Emde (1945) p. 272, or Titchmarsh (1951) p. 20: Abramowitz and Stegun (ibid., pp. 804 and 807) have

$$\frac{z}{e^z - 1} = \sum_{t=0}^{\infty} B_t\,\frac{z^t}{t!} \qquad (23)$$

and

$$B_{2t} = 2\,(-1)^{t-1}\,\frac{(2t)!}{(2\pi)^{2t}}\,\zeta(2t), \qquad (24)$$

(which agrees with Hardy and Wright's β_{2t}); while all of the other references use a positive "B_t", which is our $(-1)^{t-1}\,B_{2t}$ for $t =$ 1, 2, ... (note that our $B_{2t+1} = 0$ for $t = 1, 2, \ldots$) —— the ζ in (24) is the Riemann Zeta Function,

$$\zeta(s) = \sum_{r=1}^{\infty} \frac{1}{r^s}: \qquad (25)$$

see, once again, any of the above references]. Also,

$$\Delta_j = \Delta_j f = f^{(j)}(1) - f^{(j)}(0), \qquad (26)$$

where $f^{(j)}(x)$ denotes the j-th derivative of f at x, with $f^{(0)} = f$.

It is clear from (22) that

$$\operatorname{var}[\mathfrak{U}_n f] = O(n^{-2M}) \qquad (27)$$

if we can arrange that

$$\Delta_j = 0 \quad \text{for} \quad j = 0, 1, 2, \ldots, M-2. \qquad (28)$$

The authors now observe that

$$\Delta_j \mathfrak{D} f = \begin{cases} 0 & \text{if } j \text{ is even} \\ \Delta_j f & \text{if } j \text{ is odd} \end{cases} \tag{29}$$

and

$$\Delta_j \mathfrak{U}_n f = \frac{1}{n^{j+1}} \Delta_j f \quad \text{for all } j \geqslant 0; \tag{30}$$

so that condition (28) may be achieved by successive applications of transformations of the forms $\mathfrak{T}_a$ [see Hammersley and Morton (1956): this will not be pursued here], $\mathfrak{D}$, and

$$\mathfrak{C}_n^{(j)} f(x) = \frac{n^{j+1}\, \mathfrak{U}_n f(x) - f(x)}{n^{j+1} - 1}, \tag{31}$$

because of (29) and since, by (30),

$$\Delta_M \mathfrak{C}_n^{(j)} f = \left(\frac{n^{j-M} - 1}{n^{j+1} - 1} \right) \Delta_M f, \tag{32}$$

and, in particular,

$$\Delta_j \mathfrak{C}_n^{(j)} f = 0. \tag{33}$$

[It is also true that $\Delta_0 \mathfrak{T}_a f = 0$.] We further have that

$$E[\mathfrak{T}_a f] = E[\mathfrak{D} f] = E[\mathfrak{U}_n f] = E[\mathfrak{C}_n^{(j)} f] = E[f] = \theta, \tag{34}$$

so that the estimators will remain *unbiased*. For example, their transformation

$$\mathfrak{F}_M = \mathfrak{D}\, \mathfrak{C}_2^{(1)}\, \mathfrak{C}_2^{(3)} \ldots \mathfrak{C}_2^{(2h-1)}, \tag{35}$$

where $M = 2h + 2$, satisfies (28) [we see that $\Delta_j \mathfrak{F}_M f = 0$, by (29), for $j = 0, 2, \ldots, 2h$, and, by (33), for $j = 1, 3, \ldots, 2h - 1$: (27) now follows.] Thus, by (22) and (29) - (33), we see that, as $n \to \infty$,

$$\operatorname{var}[\mathfrak{U}_n \mathfrak{F}_M f] \sim \frac{|B_{2M}|}{(2M)!\, n^{2M}} (\Delta_{M-1} \mathfrak{F}_M f)^2, \tag{36}$$

and now, by (29), (32), and (35), since M is even,

$$\begin{aligned} \Delta_{M-1} \mathfrak{F}_M f &= \frac{(2^{2-M} - 1)(2^{4-M} - 1) \ldots (2^{-4} - 1)(2^{-2} - 1)}{(2^2 - 1)(2^4 - 1) \ldots (2^{M-4} - 1)(2^{M-2} - 1)} \Delta_{M-1} f \\ &= (-1)^h\, 2^{2-M}\, 2^{4-M} \ldots 2^{-4}\, 2^{-2}\, \Delta_{M-1} f \\ &= (-1)^h\, 2^{-2(1+2+\ldots+h)}\, \Delta_{M-1} f = (-1)^h\, 2^{-h(h+1)}\, \Delta_{M-1} f \end{aligned}$$

and (36) yields that

$$\operatorname{var}[\mathfrak{U}_n \mathfrak{F}_M f] \sim \frac{|B_{2M}| \quad (\Delta_{M-1} f)^2}{(2M)! \quad 2^{M(M-2)/2}} n^{-2M}. \tag{37}$$

[Note: in reviewing earlier articles referred to here, some slips and typographical errors were found and have been corrected here: for example, the last denominator in (22) is given as $35280n^6$ by Hammersley and Morton (1956), as was noted in Halton and Handscomb (1957); but they in turn have $M = 2h + 1$ in (35) - (37), rather than $M = 2h + 2$, and even so, the power of 2 is given incorrectly as $(M + 1)(M + 3)/2$.]

Consider, in particular, the linear transformations $\mathfrak{D}$, $\mathfrak{U}_n$, $\mathfrak{C}_n^{(j)}$, and $\mathfrak{F}_M$ defined above, and define linear combinations and products of transformations $\mathfrak{P}$ and $\mathfrak{Q}$ by

$$\left.\begin{aligned} (\lambda\mathfrak{P} + \mu\mathfrak{Q})f(x) &= \lambda\mathfrak{P}f(x) + \mu\mathfrak{Q}f(x) \\ \text{and} \qquad (\mathfrak{P}\mathfrak{Q})f(x) &= \mathfrak{P}(\mathfrak{Q}f)(x). \end{aligned}\right\} \tag{38}$$

[This was already tacitly assumed in (35); and, for example,

$$\mathfrak{C}_n^{(j)} = \frac{n^{j+1}\,\mathfrak{U}_n - \mathfrak{U}_1}{n^{j+1} - 1}, \tag{39}$$

by (31), and we note that $\mathfrak{U}_1$ is the *identity* transformation, by (21).] It is easy to verify the properties of closure and the associative and distributive laws which make the set of all polynomial expressions in $\mathfrak{D}$ and $\mathfrak{U}_n$ an *algebra* (that is, simultaneously a vector space and a ring with unity), since there are *null* and *identity* transformations, given respectively by

$$\mathfrak{N}f(x) = 0 \quad \text{and} \quad \mathfrak{I}f(x) = \mathfrak{U}_1 f(x) = f(x), \tag{40}$$

and for which $\mathfrak{P} + \mathfrak{N} = \mathfrak{P}$ and $\mathfrak{I}\,\mathfrak{P} = \mathfrak{P}$, for all transformations $\mathfrak{P}$; and each $\mathfrak{P}$ has a *negative* defined by

$$(-\mathfrak{P})f(x) = -[\mathfrak{P}f(x)], \tag{41}$$

for which $\mathfrak{P} + (-\mathfrak{P}) = \mathfrak{N}$. In addition, the algebra is *commutative*, since, by (20) and (21),

$$\mathfrak{D}\mathfrak{U}_n f(x) = \frac{1}{2n} \sum_{j=0}^{n-1} \left[f\left(\frac{x + j}{n}\right) + f\left(\frac{n - j - x}{n}\right)\right] = \mathfrak{U}_n \mathfrak{D} f(x), \tag{42}$$

and $$\mathfrak{U}_m\mathfrak{U}_n f(x) = \frac{1}{m}\sum_{i=0}^{m-1}\frac{1}{n}\sum_{j=0}^{n-1} f\left(\left(\frac{x+j}{n}+i\right)/m\right) = \frac{1}{mn}\sum_{s=0}^{mn-1} f\left(\frac{x+s}{mn}\right)$$
$$= \mathfrak{U}_{mn}f(x) = \mathfrak{U}_m\mathfrak{U}_n f(x); \qquad (43)$$

and finally, the algebra contains all of the $\mathfrak{C}_n^{(j)}$ and $\mathfrak{F}_M$, by (35) and (39).

We observe that the evaluation of $\mathfrak{U}_n\mathfrak{F}_M f(\xi)$ takes $2n(2^{h+1} - 1) = 2n(2^{M/2} - 1)$ evaluations of the function f at different points. [This is because, by (20), $\mathfrak{D}$ multiplies the number of evaluations by two; by (21), $\mathfrak{U}_n$ multiplies it by n; and, by (35) and (39), $\mathfrak{F}_M$ is the result of applying $\mathfrak{D}$ to a linear combination of $\mathfrak{U}_1$, $\mathfrak{U}_2$, ..., $\mathfrak{U}_g$ with $g = 2^h$; so that the total number of evaluations is $2n(1 + 2 + 4 + \ldots + 2^h = 2n(2^{h+1} - 1)$.] For example,

$$\mathfrak{U}_5\mathfrak{F}_6 f(\xi) = \mathfrak{U}_5\mathfrak{D}\ \mathfrak{C}_2^{(1)}\ \mathfrak{C}_2^{(3)} f(\xi) = \mathfrak{U}_5\mathfrak{D}\ \frac{2^2\mathfrak{U}_2 - \mathfrak{U}_1}{2^2 - 1}\times\frac{2^4\mathfrak{U}_2 - \mathfrak{U}_1}{2^4 - 1} f(\xi)$$
$$= \mathfrak{U}_5\mathfrak{D}\ \left(\frac{1}{45}\right)[64\mathfrak{U}_4 - 20\mathfrak{U}_2 + \mathfrak{U}_1]f(\xi)$$
$$= \mathfrak{U}_5\mathfrak{D}\ \left(\frac{1}{45}\right)\Big(16[f(\tfrac{\xi}{4}) + f(\tfrac{\xi+1}{4}) + f(\tfrac{\xi+2}{4}) + f(\tfrac{\xi+3}{4})]$$
$$- 10[f(\tfrac{\xi}{2}) + f(\tfrac{\xi+1}{2})] + f(\xi)\Big)$$
$$= \mathfrak{U}_5\ \left(\frac{1}{90}\right)\Big(16[f(\tfrac{\xi}{4}) + f(\tfrac{\xi+1}{4}) + f(\tfrac{\xi+2}{4}) + f(\tfrac{\xi+3}{4})$$
$$+ f(\tfrac{1-\xi}{4}) + f(\tfrac{2-\xi}{4}) + f(\tfrac{3-\xi}{4}) + f(\tfrac{4-\xi}{4})]$$
$$- 10[f(\tfrac{\xi}{2}) + f(\tfrac{\xi+1}{2}) + f(\tfrac{1-\xi}{2}) + f(\tfrac{2-\xi}{2})]$$
$$+ [f(\xi) + f(1-\xi)]\Big)$$
$$= \left(\frac{1}{450}\right)\Big(16[f(\tfrac{\xi}{20}) + f(\tfrac{\xi+1}{20}) + \ldots + f(\tfrac{\xi+19}{20}) + f(\tfrac{1-\xi}{20})$$
$$+ f(\tfrac{2-\xi}{20}) + \ldots + f(\tfrac{20-\xi}{20})] - 10[f(\tfrac{\xi}{10}) +$$
$$f(\tfrac{\xi+1}{10}) + \ldots + f(\tfrac{\xi+9}{10}) + f(\tfrac{1-\xi}{10}) + f(\tfrac{2-\xi}{10}) +$$
$$f(\tfrac{10-\xi}{10})] + [f(\tfrac{\xi}{5}) + \ldots + f(\tfrac{5-\xi}{5})]\Big), \qquad (44)$$

which has $2\times5(2^3-1) = 70$ terms.

3. LAURENT'S TRANSFORMATIONS.

Laurent (1961) discusses the papers of Hammersley and Morton (1956) and Halton and Handscomb (1957), and makes the important observation that the condition (28) may be obtained by using transformations much simpler than those of the earlier authors. He proposes transformations of the forms

$$\mathfrak{H}_M f = \sum_{p=1}^{M} \alpha_p \mathfrak{U}_p f \quad \text{and} \quad \mathfrak{K}_{2N} f = \sum_{q=1}^{N} \beta_q \mathfrak{U}_q \mathfrak{D} f. \tag{45}$$

The condition that

$$\mathrm{E}[\mathfrak{H}_M f] = \mathrm{E}[\mathfrak{K}_{2N} f] = \mathrm{E}[f] = \theta \tag{46}$$

entails, by (34), that

$$\sum_{p=1}^{M} \alpha_p = 1 \quad \text{and} \quad \sum_{q=1}^{N} \beta_q = 1. \tag{47}$$

By analogy with (27) and (36), we seek to arrange that

$$\mathrm{var}[\mathfrak{U}_n \mathfrak{H}_M f] = O(n^{-2M}) \quad \text{and} \quad \mathrm{var}[\mathfrak{U}_n \mathfrak{K}_{2N} f] = O(n^{-4N}). \tag{48}$$

These conditions are guaranteed if we can make

$$\left.\begin{aligned} \Delta_i \mathfrak{H}_M f &= 0 \quad \text{for} \quad i = 0, 1, 2, \ldots, M-2, \\ \Delta_j \mathfrak{K}_{2N} f &= 0 \quad \text{for} \quad j = 0, 1, 2, \ldots, 2N-2, \end{aligned}\right\} \tag{49}$$

respectively [compare (28)]; and this reduces, by (29), (30), and (45), to

$$\left.\begin{aligned} &\sum_{p=1}^{M} \frac{\alpha_p}{p^{r-1}} = 0 \quad \text{for} \quad r = 2, 3, 4, \ldots, M, \\ &\text{and} \\ &\sum_{q=1}^{N} \frac{\beta_q}{q^{2(s-1)}} = 0 \quad \text{for} \quad s = 2, 3, 4, \ldots, N; \end{aligned}\right\} \tag{50}$$

where we have put $r = i + 2$ and $s = (j + 3)/2$ (taking $j = 1, 3, \ldots, 2N - 3$ only, since even j are taken care of by the $\mathfrak{D}$ operation.) [In Laurent's paper, (18) is given incorrectly; and our index notation varies from his (we have chosen to be internally consistent); but no matter.] The importance of Laurent's contribution lies in the fact that; while $\mathfrak{F}_{2N}$ yields a variance (37) which is $O(n^{-4N})$ at a cost of $2n(2^N - 1)$ function-evaluations; the same asymptotic behavior is achieved by his transformations $\mathbb{H}_{2N}$ and $\mathbb{K}_{2N}$, with only $nN(2N + 1)$ and $nN(N + 1)$ function-evaluations, respectively, as is easily verified.

Laurent concludes his paper by exhibiting the first three or four transformations of each type, with the corresponding numbers of function-evaluations they require. His final remark, "Le procédé s'étend facilement à l'évaluation d'intégrales multiples" —— "The process is easily extended to the evaluation of multiple integrals" —— is rather optimistic, in view of the discussion of this very point by Halton and Handscomb (1957), but is literally (if perhaps not computationally) correct.

4. EXPLICIT FORMULATION OF THE TRANSFORMATIONS.

Define the $(H \times H)$ matrix

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ a_1 & a_2 & a_3 & \dots & a_H \\ a_1^2 & a_2^2 & a_3^2 & \dots & a_H^2 \\ \dots & \dots & \dots & \dots & \dots \\ a_1^{H-1} & a_2^{H-1} & a_3^{H-1} & \dots & a_H^{H-1} \end{pmatrix}$$

$$= \mathbf{A}(a_1, a_2, a_3, \dots, a_H). \tag{51}$$

Then the equations (47) and (50) for the α_r take the form

$$\mathbf{A}\,\mathbf{x} = \mathbf{e} \tag{52}$$

with $\mathbf{e}$ the column-vector with elements 1, 0, 0, ..., 0 [that is, $(\mathbf{e})_r = \delta_{r1}$], with $H = M$ and $a_r = 1/r$ $(r = 1, 2, \dots, M)$, and with $\mathbf{x}$ the column-vector with elements $(\mathbf{x})_r = \alpha_r$; and similarly, the equations for the β_s take the same form (52) with the same vector $\mathbf{e}$, with $H = N$ and $a_s = 1/s^2$ $(s = 1, 2, \dots, N)$, and with $(\mathbf{x})_s = \beta_s$.

The matrix $\mathbf{A}$ is well-known [see, e.g., Lang (1971) p. 179, Mirsky (1955) p. 17, or Shilov (1971) p. 15], and its determinant is the *Vandermonde determinant*:

$$\det \mathbf{A}\,(a_1, a_2, \ldots, a_H) = \prod_{r=2}^{H} \prod_{s=1}^{r-1} (a_r - a_s); \tag{53}$$

and by *Cramer's Rule* [see, e.g., Lang (1971) p. 192, Mirsky (1955) p. 134 —— he points out that Cramer's 1750 rule was known fifty years earlier to Leibnitz! —— or Shilov (1971) p. 18], the solution of (52) takes the form

$$(\mathbf{x})_t = \frac{\det \mathbf{A}\,(a_1, a_2, \ldots, a_{t-1}, 0, a_{t+1}, \ldots, a_H)}{\det \mathbf{A}\,(a_1, a_2, \ldots, a_{t-1}, a_t, a_{t+1}, \ldots, a_H)}$$

$$= \left[\prod_{r=2}^{H} \prod_{s=1}^{r-1} (a_r - a_s)\right]_{a_t=0} \Bigg/ \prod_{r=2}^{H} \prod_{s=1}^{r-1} (a_r - a_s). \tag{54}$$

By collecting and cancelling factors, we obtain from (54) that

$$(\mathbf{x})_t = \left\{\prod_{s=1}^{t-1} (-a_s) \prod_{r=t+1}^{H} a_r\right\} \Bigg/ \left\{\prod_{s=1}^{t-1} (a_t - a_s) \prod_{r=t+1}^{H} (a_r - a_t)\right\}. \tag{55}$$

We now return to the equations determining the α_r and the β_s. For the former, we take $a_r = 1/r$ $(r = 1, 2, \ldots, M)$, to yield from (55), after we multiply the numerator and denominator by the factor

$$t(2t)(3t)\ldots[(t-1)t][(t+1)t]\ldots(Mt), \tag{56}$$

that

$$\alpha_t = (-1)^{t-1}\, t^{M-1}/[(-1)^{t-1}\,(t-1)!\,(-1)^{M-t}\,(M-t)!],$$

or

$$\alpha_r = \frac{(-1)^{M-r}\, r^{M-1}}{(r-1)!\,(M-r)!}. \tag{57}$$

For the latter coefficients, we take $a_s = 1/s^2$ $(s = 1, 2, \ldots, N)$, to yield similarly [by multiplying by the square of (56)] that

$$\beta_t = (-1)^{t-1}\, t^{2(N-1)}/\{(-1)^{t-1}\,(t^2-1^2)(t^2-2^2)\ldots[t^2-(t-1)^2] \times (-1)^{N-t}\,[(t+1)^2-t^2]\ldots(M^2-t^2)\}$$

$$= (-1)^{N-t}\, t^{2(N-1)}/[(t-1)!\,(t+1)\ldots(2t-1)\,(N-t)! \times (2t+1)\ldots(N+t)],$$

or

$$\beta_s = \frac{2\,(-1)^{N-s}\, s^{2N}}{(N-s)!\,(N+s)!}. \tag{58}$$

The formulae (57) and (58) give us explicit means of computing the transformations $\mathfrak{H}_M$ and $\mathfrak{K}_{2N}$ of Laurent, as needed. [It should be mentioned that these formulae agree with Laurent's few computed cases —— in our notation, these are $\mathfrak{H}_2$, $\mathfrak{H}_3$, $\mathfrak{H}_4$, and $\mathfrak{K}_2$, $\mathfrak{K}_4$, $\mathfrak{K}_6$, and $\mathfrak{K}_8$.]

Computer implementation of (57) and (58) readily yields the following values: we write a_{Mr} for the coefficient a_r when the transformation is $\mathfrak{H}_M$, and similarly, β_{Ns} for β_s in the transformation $\mathfrak{K}_{2N}$ [see (45).]

$a_{11} = 1$ $M = 1$

$a_{21} = -1$ $M = 2$
$a_{22} = 2$

$a_{31} = 1/2 = .5$ $M = 3$
$a_{32} = -4$
$a_{33} = 9/2 = 4.5$

$a_{41} = -1/6 = -.16666666666667$ $M = 4$
$a_{42} = 4$
$a_{43} = -27/2 = -13.5$
$a_{44} = 32/3 = 10.666666666667$

$a_{51} = 1/24 = 4.1666666666667 \times 10^{-2}$ $M = 5$
$a_{52} = -8/3 = -2.6666666666667$
$a_{53} = 81/4 = 20.25$
$a_{54} = -128/3 = -42.666666666667$
$a_{55} = 625/24 = 26.041666666667$

$a_{61} = -1/120 = -8.3333333333333 \times 10^{-3}$ $M = 6$
$a_{62} = 4/3 = 1.3333333333333$
$a_{63} = -81/4 = -20.25$
$a_{64} = 256/3 = 85.333333333333$
$a_{65} = -3125/24 = -130.20833333333$
$a_{66} = 324/5 = 64.8$

$a_{71} = 1/720 = 1.3888888888889 \times 10^{-3}$ $M = 7$
$a_{72} = -8/15 = -.53333333333333$
$a_{73} = 243/16 = 15.1875$
$a_{74} = -1024/9 = -113.77777777778$
$a_{75} = 15625/48 = 325.52083333333$
$a_{76} = -1944/5 = -388.8$
$a_{77} = 117649/720 = 163.40138888889$

$a_{81} = -1/5040 = -1.984126984127 \times 10^{-4}$ $M = 8$
$a_{82} = 8/45 = .17777777777778$
$a_{83} = -729/80 = -9.1125$
$a_{84} = 1024/9 = 113.77777777778$
$a_{85} = -78125/144 = -542.53472222222$
$a_{86} = 5832/5 = 1166.4$
$a_{87} = -823543/720 = -1143.8097222222$
$a_{88} = 131072/315 = 416.10158730159$

$a_{91} = 1/40320 = 2.4801587301587 \times 10^{-5}$ $M = 9$
$a_{92} = -16/315 = -5.0793650793651 \times 10^{-2}$
$a_{93} = 729/160 = 4.55625$
$a_{94} = -4096/45 = -91.022222222222$
$a_{95} = 390625/576 = 678.16840277778$
$a_{96} = -11664/5 = -2332.8$
$a_{97} = 5764801/1440 = 4003.3340277778$
$a_{98} = -1048576/315 = -3328.8126984127$
$a_{99} = 4782969/4480 = 1067.6270089286$

and similarly:

$\beta_{11} = 1$ $N = 1$

$\beta_{21} = -1/3 = -.33333333333333$ $N = 2$
$\beta_{22} = 4/3 = 1.3333333333333$

$\beta_{31} = 1/24 = 4.1666666666667 \times 10^{-2}$ $N = 3$
$\beta_{32} = -16/15 = -1.0666666666667$
$\beta_{33} = 81/40 = 2.025$

$\beta_{41} = -1/360 = -2.7777777777778 \times 10^{-3}$ $N = 4$

$\beta_{42} = 16/45 = .35555555555556$

$\beta_{43} = -729/280 = -2.6035714285714$

$\beta_{44} = 1024/315 = 3.2507936507937$

$\beta_{51} = 1/8640 = 1.1574074074074 \times 10^{-4}$ $N = 5$

$\beta_{52} = -64/945 = -6.7724867724868 \times 10^{-2}$

$\beta_{53} = 6561/4480 = 1.4645089285714$

$\beta_{54} = -16384/2835 = -5.779188712522$

$\beta_{55} = 390625/72576 = 5.3822889109347$

$\beta_{61} = -1/302400 = -3.3068783068783 \times 10^{-6}$ $N = 6$

$\beta_{62} = 8/945 = 8.4656084656085 \times 10^{-3}$

$\beta_{63} = -2187/4480 = -.48816964285714$

$\beta_{64} = 65536/14175 = 4.6233509700176$

$\beta_{65} = -9765625/798336 = -12.232474797579$

$\beta_{66} = 17496/1925 = 9.0888311688312$

$\beta_{71} = 1/14515200 = 6.8893298059965 \times 10^{-8}$ $N = 7$

$\beta_{72} = -32/42525 = -7.5249853027631 \times 10^{-4}$

$\beta_{73} = 19683/179200 = .10983816964286$

$\beta_{74} = -1048576/467775 = -2.2416247127358$

$\beta_{75} = 244140625/19160064 = 12.742161247478$

$\beta_{76} = -629856/25025 = -25.169070929071$

$\beta_{77} = 13841287201/889574400 = 15.559448654323$

$\beta_{81} = -1/914457600 = -1.0935444136502 \times 10^{-9}$ $N = 8$

$\beta_{82} = 32/637875 = 5.0166568685087 \times 10^{-5}$

$\beta_{83} = -177147/9856000 = -1.7973518668831 \times 10^{-2}$

$\beta_{84} = 1048576/1403325 = .74720823757861$

$\beta_{85} = -6103515625/747242496 = -8.1680520817167$

$\beta_{86} = 5668704/175175 = 32.360234051663$

$\beta_{87} = -678223072849/13343616000 = -50.827532270788$

$\beta_{88} = 17179869184/638512875 = 26.906065416457$

$\beta_{91} = 1/73156608000 = 1.3669305170628 \times 10^{-11}$ $\quad N = 9$

$\beta_{92} = -128/49116375 = -2.6060555161084 \times 10^{-6}$

$\beta_{93} = 177147/78848000 = 2.2466898336039 \times 10^{-3}$

$\beta_{94} = -16777216/91216125 = -.18392818155781$

$\beta_{95} = 152587890625/41845579776 = 3.646451822195$

$\beta_{96} = -22674816/875875 = -25.88818724133$

$\beta_{97} = 33232930569601/426995712000 = 77.829658789644$

$\beta_{98} = -1099511627776/10854718875 = -101.29342274431$

$\beta_{99} = 22876792454961/487911424000 = 46.887183471566$

[The numerators and denominators listed above are integers, and care has been taken that they are precisely correct. The resulting values of the coefficients have been obtained by a single division, correct to 14 significant decimal digits, minimizing round-off errors. As M and (even more so) N increase beyond this point, the numerators and denominators tend to increasing size ($\beta_{10,9}$ and $\beta_{10,10}$ have numerators of more than 14 decimal digits) and the floating-point calculations of the coefficients tend to accumulate increasing round-off errors — so the reader is advised to proceed with caution, if meaningful results are to be obtained. The underlying equations (47) and (50) were checked for the coefficients listed above, yielding sums departing from the correct right-hand sides by no more than 4×10^{-11} for the a_{Mr}, 2×10^{-12} for the β_{Ns}.]

As in obtaining (37) from (35) and (36), we now take (22) with (49) to yield that

$$\operatorname{var}[\mathfrak{U}_n \mathfrak{H}_M f] \sim \frac{|B_{2M}|}{(2M)!\, n^{2M}} (\Delta_{M-1} \mathfrak{H}_M f)^2 \tag{59}$$

and

$$\operatorname{var}[\mathfrak{U}_n \mathfrak{K}_{2N} f] \sim \frac{|B_{4N}|}{(4N)!\, n^{4N}} (\Delta_{2N-1} \mathfrak{K}_{2N} f)^2; \tag{60}$$

and then use (29), (30), and (45) to show that, by (57),

$$\Delta_{M-1} \mathfrak{H}_M f = \sum_{p=1}^{M} a_p \, \Delta_{M-1} \mathfrak{U}_p f = \sum_{p=1}^{M} \frac{a_p}{p^M} \Delta_{M-1} f = \sum_{p=1}^{M} \frac{(-1)^{M-p}}{p!\,(M-p)!} \Delta_{M-1} f$$

$$= \frac{1}{M!} [(1-1)^M - (-1)^M] \, \Delta_{M-1} f = \frac{(-1)^{M-1}}{M!} \Delta_{M-1} f, \tag{61}$$

and by (58),

$$\Delta_{2N-1}\mathfrak{K}_{2N}f = \sum_{q=1}^{N} \beta_q \, \Delta_{2N-1}\mathfrak{U}_q\mathfrak{D}f = \sum_{q=1}^{N} \frac{\beta_q}{q^{2N}} \Delta_{2N-1}f$$

$$= \sum_{q=1}^{N} \frac{2\ (-1)^{N-q}}{(N-q)!\ (N+q)!} \Delta_{2N-1}f$$

$$= [\frac{1}{0!(2N)!} - \frac{1}{1!(2N-1)!} + \ldots - \frac{(-1)^N}{(N-1)!(N+1)!}$$

$$- \frac{(-1)^N}{(N+1)!(N-1)!} + \ldots - \frac{1}{(2N-1)!1!} + \frac{1}{(2N)!0!}] \ \Delta_{2N-1}f$$

$$= \frac{1}{(2N)!} [(1-1)^{2N} - \binom{2N}{N} (-1)^N] \ \Delta_{2N-1}f$$

$$= \frac{(-1)^{N-1}}{(N!)^2} \Delta_{2N-1}f; \tag{62}$$

whence we finally obtain that

$$\text{var}[\mathfrak{U}_n\mathfrak{H}_M f] \sim \frac{|B_{2M}| \ (\Delta_{M-1}f)^2}{(2M)! \ (M!)^2} n^{-2M} \tag{63}$$

and

$$\text{var}[\mathfrak{U}_n\mathfrak{K}_{2N}f] \sim \frac{|B_{4N}| \ (\Delta_{2N-1}f)^2}{(4N)! \ (N!)^4} n^{-4N}, \tag{64}$$

asymptotically as $n \to \infty$.

5. HANDSCOMB'S BOUND.

We see from (63) and (64) that there exist constants $\tilde{H}_M(f)$ and $\tilde{K}_{2N}(f)$, such that, if the function f is differentiable w times [with $w = M - 1$ for (65) and $w = 2N - 1$ for (66)], then

$$\text{var}[\mathfrak{U}_n\mathfrak{H}_M f] \leqslant \tilde{H}_M(f) \ n^{-2M} \tag{65}$$

and

$$\text{var}[\mathfrak{U}_n\mathfrak{K}_{2N}f] \leqslant \tilde{K}_{2N}(f) \ n^{-4N}. \tag{66}$$

Handscomb (1964) discovered an upper bound for the variance of $\mathfrak{U}_n g$: if we write

$$C_w = \frac{4}{\pi} \sum_{j=0}^{\infty} \frac{(-1)^{jw}}{(2j+1)^w}, \tag{67}$$

and if the integral

$$J_w^{\ 2} = \int_0^1 |g^{(w-1)}(x)|^2 \ dx \tag{68}$$

exists and is finite, while $\Delta_j g = 0$ for $j = 0, 1, \ldots, w - 1$ (which entails that g is differentiable $w - 1$ times); then

$$\mathrm{var}[\mathfrak{U}_n g] \leqslant C_w^{\;2}\, J_w^{\;2}/(2\pi n)^{2w-2}. \tag{69}$$

Just as we got (65) and (66) from (63) and (64), we may derive from (22) that there is a constant $\tilde{C}(g)$, such that

$$\mathrm{var}[\mathfrak{U}_n g] \leqslant \tilde{C}(g)\; n^{-2w-2}, \tag{70}$$

provided that g is differentiable w times. Thus the addition of one more degree of differentiability to g yields a bound for the variance which behaves as n^{-2w-2} rather than n^{-2w+2}.

Unfortunately, while it was an easy matter to relate the factors $\Delta_{M-1}\mathfrak{H}_M f$ and $\Delta_{2N-1}\mathfrak{K}_{2N} f$ occurring in (59) and (60) to $\Delta_w f$; it is not clear how one may relate the integrals (68) for $g = \mathfrak{H}_M f$ and $g = \mathfrak{K}_{2N} f$ to the corresponding integrals for $g = f$. Thus we can only deduce from (69) that, if f is differentiable only $w - 1$ times, then there will be constants $\breve{H}_M(f)$ and $\breve{K}_{2N}(f)$, such that

$$\mathrm{var}[\mathfrak{U}_n \mathfrak{H}_M f] \leqslant \breve{H}_M(f)\; n^{-2M+4} \tag{71}$$

and

$$\mathrm{var}[\mathfrak{U}_n \mathfrak{K}_{2N} f] \leqslant \breve{K}_{2N}(f)\; n^{-4N+4}. \tag{72}$$

[Perhaps this is all that can be obtained from (69), in any case; since $J_w^{\;2}$ may well be harder to compute than the original integral θ in (1)!]

6. EFFICIENCY.

Comparing (63) with (64), we see that, as $n \to \infty$,

$$\mathrm{var}[\mathfrak{U}_n \mathfrak{H}_{2N} f]/\mathrm{var}[\mathfrak{U}_n \mathfrak{K}_{2N} f] \sim (N!)^4/[(2N)!]^2; \tag{73}$$

and we know that the ratio of the number of required function-evaluations is

$$\mathrm{work}[\mathfrak{U}_n \mathfrak{H}_{2N} f]/\mathrm{work}[\mathfrak{U}_n \mathfrak{K}_{2N} f] = (2N+1)/(N+1); \tag{74}$$

so $\mathfrak{H}_{2N}$ is clearly preferable to $\mathfrak{K}_{2N}$, for any N, at least if n is large.

Comparing (37) with (63), we get similarly that

$$\mathrm{var}[\mathfrak{U}_n\mathfrak{H}_{2N}f]/\mathrm{var}[\mathfrak{U}_n\mathfrak{F}_{2N}f] \sim 4^{N(N-1)}/[(2N)!]^2, \qquad (75)$$

while $$\mathrm{work}[\mathfrak{U}_n\mathfrak{H}_{2N}f]/\mathrm{work}[\mathfrak{U}_n\mathfrak{F}_{2N}f] = N(2N+1)/[2(2^N-1)]; \qquad (76)$$

so that, contrarily to the impression which has prevailed since Laurent published his paper (1961), the transformation of Halton and Handscomb (1957) appears to be more efficient, if we measure efficiency by 1/(var × work). [The argument runs as follows: we measure efficiency as the reciprocal of the work required to achieve a given variance. If the variance of the estimator from a single random sample is 'var' and the variance we wish to achieve is V, then we must average the estimator over var/V independent samples; and if the work required to compute one sample value of the estimator is 'work', then var/V samples use var × work/V; so the efficiency can be measured by 1/(var × work), since V is merely an arbitrary scale-factor. Now, from (75) and (76),

$$\frac{1/\{\mathrm{var}[\mathfrak{U}_n\mathfrak{F}_{2N}f]\ \mathrm{work}[\mathfrak{U}_n\mathfrak{F}_{2N}f]\}}{1/\{\mathrm{var}[\mathfrak{U}_n\mathfrak{H}_{2N}f]\ \mathrm{work}[\mathfrak{U}_n\mathfrak{H}_{2N}f]\}} \sim \frac{N\ (2N+1)\ 4^{N(N-1)}}{2(2^N-1)\ [(2N)!]^2}. \qquad (77)$$

Stirling's formula [see, e.g., Abramowitz and Stegun (1964) §6, Jahnke and Emde (1945) p. 10, or Whittaker and Watson (1927) p. 251] states that, as $z \to \infty$,

$$z! \sim \left(\frac{z}{\mathrm{e}}\right)^z \sqrt{(2\pi z)}. \qquad (78)$$

It now follows that, for large enough N, the ratio (77) behaves like

$$4^{N^2}\ N^{-4N+1}\ (\mathrm{e}^4/2^7)^N\ (4\pi)^{-1} \to \infty; \qquad (79)$$

since the logarithms of the first two (dominant) factors are respectively $N^2 \log 4$ and $(-4N+1) \log N$, and the first outweighs the second. Thus, at least for sufficiently large N, the transformation $\mathfrak{F}_{2N}$ is more efficient than $\mathfrak{H}_{2N}$.]

7. HALTON AND HANDSCOMB'S TRANSFORMATIONS.

On the basis of the last result, it is clearly of value to re-examine the transformations of Halton and Handscomb (1957).

Let us write, by analogy with (45),

$$\mathfrak{E}_M f = \sum_{p=1}^{M} \lambda_p \mathfrak{U}_{2^{p-1}} f \quad \text{and} \quad \mathfrak{F}_{2N} f = \sum_{q=1}^{N} \mu_q \mathfrak{U}_{2^{q-1}} \mathfrak{D} f. \tag{80}$$

Impose conditions analogous to (46) and (49):

$$\mathrm{E}[\mathfrak{E}_M f] = \mathrm{E}[\mathfrak{F}_{2N} f] = \mathrm{E}[f] = \theta, \tag{81}$$

$$\left.\begin{aligned} \Delta_i \mathfrak{E}_M f &= 0 \quad \text{for} \quad i = 0, 1, 2, \ldots, M-2, \\ \Delta_j \mathfrak{F}_{2N} f &= 0 \quad \text{for} \quad j = 0, 1, 2, \ldots, 2N-2. \end{aligned}\right\} \tag{82}$$

Then we obtain the equations [analogous to (47) and (50)]

$$\left.\begin{aligned} &\sum_{p=1}^{M} \frac{\lambda_p}{2^{(p-1)(r-1)}} = \delta_{r1} \quad \text{for } r = 1, 2, \ldots, M, \\ \text{and} \quad &\sum_{q=1}^{N} \frac{\mu_q}{4^{(q-1)(s-1)}} = \delta_{s1} \quad \text{for } s = 1, 2, \ldots, N. \end{aligned}\right\} \tag{83}$$

Comparison of (83) with (51) shows that the first set of equations corresponds to $H = M$ and $a_r = 2^{-r+1}$ ($r = 1, 2, \ldots, M$), with $\mathbf{x}$ the column vector with elements $(\mathbf{x})_r = \lambda_r$; while the second set corresponds to $H = N$ and $a_s = 4^{-s+1}$ ($s = 1, 2, \ldots, N$), with $(\mathbf{x})_s = \mu_s$. In both cases, $(\mathbf{e})_r = \delta_{r1}$. Since the solution of (51) is given by (55), we may apply the same technique as in obtaining (57); but, now, factors of the form $1/t$ are replaced by 2^{-t+1} or by 4^{-t+1}. We may simplify the results by defining the notation

$$(z^k-1)(z^{k-1}-1)\ldots(z^2-1)(z-1) = k_z\!\downarrow. \tag{84}$$

Then the products $(2^{t-1}-1)(2^{t-1}-2)(2^{t-1}-2^2)\ldots(2^{t-1}-2^{t-2})$ and $(2^t-2^{t-1})(2^{t+1}-2^{t-1})\ldots(2^{M-1}-2^{t-1})$ become respectively $2^0 2^1 2^2 \ldots 2^{t-2}\ (t-1)_2\!\downarrow$ and $2^{(t-1)(M-t)}\ (M-t)_2\!\downarrow$; and we get that

$$\lambda_r = \frac{(-1)^{M-r}\ 2^{r(r-1)/2}}{(r-1)_2\!\downarrow\ (M-r)_2\!\downarrow}\ ; \tag{85}$$

and, similarly,

$$\mu_s = \frac{(-1)^{N-s}\ 4^{s(s-1)/2}}{(s-1)_4\!\downarrow\ (N-s)_4\!\downarrow}\ . \tag{86}$$

One interesting consequence of (84) in (85) and (86) is that, while (57) and (58) had a considerable cancellation of factors in reducing the integer numerators and denominators to their lowest terms, the numerators of the fractions (85) and (86) are all pure powers of 2, while the denominators are all odd.

It is clear that the solutions (85) and (86) of the equations (83) are *unique*. Now, the transformation $\mathfrak{F}_{2N}$ defined in (35) is easily seen, by (39) and (43), to be of the form defined in the second equation of (80); and it also satisfies (81) and the second condition (82) [by (34), (29), and (33).] By uniqueness, the formulae (86) give the expansion (80) of (35), for $M = 2N$.

A similar argument shows that the formulae (85) give the expansion (80) of the transformation defined by

$$\mathfrak{E}_M = \mathfrak{C}_2^{(0)}\ \mathfrak{C}_2^{(1)}\ \mathfrak{C}_2^{(2)}\ \ldots\ \mathfrak{C}_2^{(M-2)}. \tag{87}$$

[Note that, while Halton and Handscomb's $\mathfrak{F}$ transformations are defined exactly as in (35) —— with an index defined smaller by 1 than ours —— their $\mathfrak{E}$ transformations are defined with the factor $\mathfrak{C}_2^{(0)}$ replaced by $\mathfrak{T}_a$, which would be inconvenient for us here.]

Computer implementation of (85) and (86) is again straightforward [though the same warnings about the accumulation of round-off errors in the naive programming of the formulae must be heeded.] We write λ_{Mr} for λ_r in $\mathfrak{E}_M$, and μ_{Ns} for μ_s in $\mathfrak{F}_{2N}$, much as we did earlier with a_{Mr} and β_{Ns}.

$\lambda_{11} = 1$ $\qquad M = 1$

$\lambda_{21} = -1$ $\qquad M = 2$

$\lambda_{22} = 2$

$\lambda_{31} = 1/3 = .33333333333333$ $\qquad M = 3$

$\lambda_{32} = -2$

$\lambda_{33} = 8/3 = 2.6666666666667$

$\lambda_{41} = -1/21 = -4.7619047619048 \times 10^{-2}$ $\qquad M = 4$

$\lambda_{42} = 2/3 = .66666666666667$

$\lambda_{43} = -8/3 = -2.6666666666667$

$\lambda_{44} = 64/21 = 3.047619047619$

$\lambda_{51} = 1/315 = 3.1746031746032 \times 10^{-3}$ $M = 5$
$\lambda_{52} = -2/21 = -9.5238095238095 \times 10^{-2}$
$\lambda_{53} = 8/9 = .88888888888889$
$\lambda_{54} = -64/21 = -3.047619047619$
$\lambda_{55} = 1024/315 = 3.2507936507937$

$\lambda_{61} = -1/9765 = -1.0240655401946 \times 10^{-4}$ $M = 6$
$\lambda_{62} = 2/315 = 6.3492063492063 \times 10^{-3}$
$\lambda_{63} = -8/63 = -.12698412698413$
$\lambda_{64} = 64/63 = 1.015873015873$
$\lambda_{65} = -1024/315 = -3.2507936507937$
$\lambda_{66} = 32768/9765 = 3.3556579621096$

$\lambda_{71} = 1/615195 = 1.6255008574517 \times 10^{-6}$ $M = 7$
$\lambda_{72} = -2/9765 = -2.0481310803891 \times 10^{-4}$
$\lambda_{73} = 8/945 = 8.4656084656085 \times 10^{-3}$
$\lambda_{74} = -64/441 = -.14512471655329$
$\lambda_{75} = 1024/945 = 1.0835978835979$
$\lambda_{76} = -32768/9765 = -3.3556579621096$
$\lambda_{77} = 2097152/615195 = 3.4089223742066$

and similarly:

$\mu_{11} = 1$ $N = 1$

$\mu_{21} = -1/3 = -.33333333333333$ $N = 2$
$\mu_{22} = 4/3 = 1.3333333333333$

$\mu_{31} = 1/45 = 2.2222222222222 \times 10^{-2}$ $N = 3$
$\mu_{32} = -4/9 = -.44444444444444$
$\mu_{33} = 64/45 = 1.4222222222222$

$\mu_{41} = -1/2835 = -3.5273368606702 \times 10^{-4}$ $N = 4$
$\mu_{42} = 4/135 = .02962962962963$
$\mu_{43} = -64/135 = -.47407407407407$
$\mu_{44} = 4096/2835 = 1.4447971781305$

$\mu_{51} = 1/722925 = 1.3832693571256 \times 10^{-6}$ $\quad N = 5$
$\mu_{52} = -4/8505 = -4.7031158142269 \times 10^{-4}$
$\mu_{53} = 64/2025 = 3.1604938271605 \times 10^{-2}$
$\mu_{54} = -4096/8505 = -.48159905937684$
$\mu_{55} = 1048576/72295 = 1.4504630494173$

$\mu_{61} = -1/739552275 = -1.3521694595558 \times 10^{-9}$ $\quad N = 6$
$\mu_{62} = 4/2168775 = 1.8443591428341 \times 10^{-6}$
$\mu_{63} = -64/127575 = -5.0166568685087 \times 10^{-4}$
$\mu_{64} = 4096/127575 = 3.2106603958456 \times 10^{-2}$
$\mu_{65} = -1048576/2168775 = -.4834876831391$
$\mu_{66} = 1073741824/739552275 = 1.4518809018605$

$\mu_{71} = 1/3028466566125 = 3.3020011222363 \times 10^{-13}$ $\quad N = 7$
$\mu_{72} = -4/2218656825 = -1.802892612741 \times 10^{-9}$
$\mu_{73} = 64/32531625 = 1.967316419023 \times 10^{-6}$
$\mu_{74} = -4096/8037225 = -5.096286342612 \times 10^{-4}$
$\mu_{75} = 1048576/32531625 = 3.2232512209273 \times 10^{-2}$
$\mu_{76} = -1073741824/2218656825 = -.48396030062017$
$\mu_{77} = 4398046511104/3028466566125 = 1.4522354515313$

[Again, numerators and denominators above are precisely-determined integers, and the coefficients have been obtained by a single division, correct to 14 significant decimal digits. Here, μ_{81} and μ_{88} have denominators of more than 14 digits; so we terminated the listing at $M = N = 7$. The equations (83) were checked for these coefficients, yielding sums departing from the correct right-hand sides by no more than 10^{-13} for the λ_{Mr}, 2×10^{-14} for the μ_{Ns}.]

We may now proceed for the transformations $\mathfrak{E}_M$ and $\mathfrak{F}_{2N}$ just as we did for $\mathfrak{H}_M$ and $\mathfrak{K}_{2N}$ in §4 and §6. First, we note that (22), (33), and (87) yield [like (36), (59), and (60)]

$$\mathrm{var}[\mathfrak{U}_n \mathfrak{E}_M f] \sim \frac{|B_{2M}|}{(2M)!\ n^{2M}} (\Delta_{M-1} \mathfrak{E}_M f)^2 \qquad (88)$$

as $n \to \infty$; and then that (32) and (87) yield

$$\Delta_{M-1}\mathfrak{E}_M f = \frac{(2^{1-M}-1)(2^{2-M}-1)\ldots(2^{-1}-1)}{(2-1)(2^2-1)\ldots(2^{M-1}-1)}\,\Delta_{M-1}f$$
$$= (-1)^{M-1}\,2^{-[1+2+\ldots+(M-1)]}\,\Delta_{M-1}f$$
$$= (-1)^{M-1}\,2^{-M(M-1)/2}\,\Delta_{M-1}f;$$

so that (88) becomes [like (37), (63), and (64)]

$$\mathrm{var}[\mathfrak{U}_n\mathfrak{E}_M f] \sim \frac{|B_{2M}|}{(2M)!}\,\frac{(\Delta_{M-1}f)^2}{2^{M(M-1)}}\,n^{-2M}. \tag{89}$$

Thus, by (37) and (89) [like (73)], we have that, as $n \to \infty$,

$$\mathrm{var}[\mathfrak{U}_n\mathfrak{E}_{2N}f]/\mathrm{var}[\mathfrak{U}_n\mathfrak{F}_{2N}f] \sim \frac{2^{2N(N-1)}}{2^{2N(2N-1)}} = 2^{-2N^2}; \tag{90}$$

and clearly [like (74)], since $2^{2N}-1 = (2^N-1)(2^N+1)$,

$$\mathrm{work}[\mathfrak{U}_n\mathfrak{E}_{2N}f]/\mathrm{work}[\mathfrak{U}_n\mathfrak{F}_{2N}f] = (2^N+1)/2. \tag{91}$$

Thus, the efficiency-ratio of the two transformations is

$$\frac{1/\{\mathrm{var}[\mathfrak{U}_n\mathfrak{F}_{2N}f]\ \mathrm{work}[\mathfrak{U}_n\mathfrak{F}_{2N}f]\}}{1/\{\mathrm{var}[\mathfrak{U}_n\mathfrak{E}_{2N}f]\ \mathrm{work}[\mathfrak{U}_n\mathfrak{E}_{2N}f]\}} \sim (2^N+1)/2^{2N^2+1}, \tag{92}$$

and we may conclude that, at least if n is large, $\mathfrak{E}_{2N}$ is preferable to $\mathfrak{F}_{2N}$ for any N. [This result is analogous to that for $\mathfrak{H}_{2N}$ and $\mathfrak{K}_{2N}$.]

To summarize these comparisons, we have seen that

(a) *For sufficiently large n,* $\mathfrak{U}_n\mathfrak{E}_{2N}$ *is more efficient than* $\mathfrak{U}_n\mathfrak{F}_{2N}$, *and* $\mathfrak{U}_n\mathfrak{H}_{2N}$ *is more efficient than* $\mathfrak{U}_n\mathfrak{K}_{2N}$, *for all N;*

and (b) *For sufficiently large n and sufficiently large N,* $\mathfrak{U}_n\mathfrak{F}_{2N}$ *is more efficient than* $\mathfrak{U}_n\mathfrak{H}_{2N}$.

[See (90) and (91), (73) and (74), and (75) and (76), respectively; noting that the result (b) holds only asymptotically, for large N, by virtue of (77) - (79). Indeed, if we write $\rho(N)$ for the right-hand side of (77), it is easy to verify that $\rho(1) = .375$, $\rho(2) \approx .046$, $\rho(3) \approx .01185$, $\rho(4) \approx .01238$, $\rho(5) \approx .074$, $\rho(6) \approx 3.11$, and $\rho(7) \approx 1052.1$.]

Finally, a third comparison results from these considerations. If we divide the efficiency-ratio (77) by that in (92), we obtain

$$\frac{1/\{\mathrm{var}[\mathfrak{U}_n\mathfrak{E}_{2N}f]\ \mathrm{work}[\mathfrak{U}_n\mathfrak{E}_{2N}f]\}}{1/\{\mathrm{var}[\mathfrak{U}_n\mathfrak{H}_{2N}f]\ \mathrm{work}[\mathfrak{U}_n\mathfrak{H}_{2N}f]\}} \sim \frac{N\ (2N+1)\ 4^{N(2N-1)}}{(4^N-1)\ [(2N)!]^2} = \bar{\rho}(N); \tag{93}$$

and we see that $\bar{\rho}(1) = 1$, $\bar{\rho}(2) \approx 4.74$, $\bar{\rho}(3) \approx 690.42$, and $\bar{\rho}(N) > 1$ for all $N \geqslant 2$. Thus,

(c) *For sufficiently large n,* $\mathfrak{U}_n\mathfrak{E}_{2N}$ *is more efficient than* $\mathfrak{U}_n\mathfrak{H}_{2N}$, *for all* $N \geqslant 2$, *with equal efficiency for* $N = 1$.

[The proof that $\rho(N)$ and $\bar{\rho}(N)$ are monotone-non-decreasing beyond the stated values of N is straightforward but tedious.]

The efficiency considerations summarized above, in (a), (b), and (c), would lead us to conclude that, for sufficiently large n, we should always apply the transformation $\mathfrak{U}_n\mathfrak{E}_{2N}$, where $\mathfrak{E}_{2N}$ is defined as in (80) and (87), rather than the $\mathfrak{F}$, $\mathfrak{H}$, or $\mathfrak{K}$ transformations. However, the situation is somewhat complicated by two further considerations. First, the computation of the coefficients λ_r and μ_s tends (as has already been pointed out) to be somewhat (though not enormously) more complicated than that of the α_r and β_s, and thus greater round-off errors may be anticipated. Secondly, for each random sample $\mathfrak{U}_n\mathfrak{E}_{2N}f(\xi)$, we need $n(4^N-1)$ function-evaluations; while the comparable number of function-evaluations for the simplest transformation, $\mathfrak{U}_n\mathfrak{K}_{2N}$, is $nN(N+1)$. Thus, when we derive efficiency as 1/(var × work) [see the remarks between (76) and (77) in §6], we should really use $V/(\lceil \mathrm{var}/V \rceil \times \mathrm{work})$, where $\lceil x \rceil$ denotes the *roof* function [the integer-supremum; i.e., the least integer not less than x.] This means that, if V is *small*, the efficiency decreases markedly, in reality. [Another way of demonstrating this is to point out that, if $\mathrm{var}[\mathfrak{U}_n\mathfrak{E}_{2N}f] = v_1$ and $\mathrm{var}[\mathfrak{U}_n\mathfrak{K}_{2N}f] = v_2 > v_1$ and if we wish to achieve a variance $V = v_2$, then the $\mathfrak{E}$ transformation requires work $n(4^N-1)$ (yielding more accuracy than we require), while the $\mathfrak{K}$ transformation requires only $nN(N+1)$ of work to achieve our aim.]

8. EXAMPLES.

To illustrate and gauge the behavior of the various transformations described above, a number of examples have been run. The functions integrated are the following. (In each case, they were scaled, so that $\theta = 1$.)

$$\left.\begin{aligned} f_1 &= 7\ z^6, \\ f_2 &= 20\ z^{19}, \\ f_3 &= c_3\ e^{20z}, \\ f_4 &= c_4\ \cos(4\ z), \\ f_5 &= c_5\ z^2\ \cos(20\ z^3) \\ f_6 &= c_6\ \{2\ z\ [1 + \log(1 + z^2)] + 3\ z^2\ e^{z^3}\}; \end{aligned}\right\} \quad (94)$$

where

$$\left.\begin{aligned} c_3 &= 20/(e^{20} - 1) \approx 4.1223072533738 \times 10^{-8}, \\ c_4 &= 4/\sin 4 \approx -5.2853948352436, \\ c_5 &= 60/\sin 20 \approx 65.721356184479, \\ c_6 &= 2\ \log 2 + e - 1 \approx .32210515668989. \end{aligned}\right\} \quad (95)$$

Each of these functions was integrated by *crude Monte Carlo*, using an estimator of the form ψ_k [see (2)], namely,

$$\Psi_{jk} = \Psi_{jk}(\xi_1, \xi_2, \ldots, \xi_k) = \frac{1}{k} \sum_{r=1}^{k} f_j(\xi_r); \quad (96)$$

and with the use of the *antithetic transformations*

$$\mathfrak{A}_{1nM} = \mathfrak{U}_n \mathfrak{E}_M, \quad \mathfrak{A}_{2nM} = \mathfrak{U}_n \mathfrak{F}_M, \quad \mathfrak{A}_{3nM} = \mathfrak{U}_n \mathfrak{H}_M, \quad \mathfrak{A}_{4nM} = \mathfrak{U}_n \mathfrak{K}_M, \quad (97)$$

using an estimator of the form γ_h [see (15) and (16)], namely,

$$\Upsilon_{ijknM} = \Upsilon_{ijknM}(\xi_1, \xi_2, \ldots, \xi_h) = \frac{1}{h_{iknM}} \sum_{r=1}^{h_{iknM}} \mathfrak{A}_{inM} f_j(\xi_r), \quad (98)$$

where

$$h = h_{iknM} = \max\left\{2, \left\lfloor \frac{k}{W_{inM}} + \frac{1}{2} \right\rfloor\right\}, \quad (99)$$

with W_{inM} denoting the number of function-evaluations entailed in $A_{inM}f$, namely,

$$\left.\begin{aligned} W_{1nM} &= n(2^M - 1), & W_{2nM} &= 2n(2^{M/2} - 1), \\ W_{3nM} &= \tfrac{1}{2}nM(M + 1), & W_{4nM} &= \tfrac{1}{2}nM(\tfrac{M}{2} + 1), \end{aligned}\right\} \quad (100)$$

and $\lfloor x \rfloor$ denoting the *floor* function [the integer infimum; i.e., the greatest integer not greater than x.]

Two computational experiments were performed. Each consisted of a number of runs; and in each run, all the estimates were obtained with the same sequence of pseudo-random numbers. In the first experiment, the efficiencies of the transformations were compared. The actual *errors* of the estimates were computed by subtracting $\theta = 1$ from them, and the corresponding *efficiencies* were measured by

$$\mathrm{eff}_E[\text{estimate}] = 1/\{(\text{error})^2 \times \text{work}\}. \quad (101)$$

In addition, the *sample variances* of the estimates were computed by the formula

$$\mathrm{svar}[\frac{1}{t} \sum_{r=1}^{t} x_r] = \frac{1}{t(t - 1)} \sum_{s=2}^{t} \frac{s}{s - 1} (\frac{1}{s} \sum_{r=1}^{s} x_r - x_s)^2, \quad (102)$$

for the sample variance of an *average* $\frac{1}{t} \Sigma_{r=1}^{t} x_r$ of *statistically independent, identically distributed* random variables x_r. [It is easily verified that, if

$$E[x_r] = \mu < \infty \quad \text{and} \quad \mathrm{var}[x_r] = \sigma^2 < \infty, \quad (103)$$

then

$$E[\frac{1}{t} \Sigma_{r=1}^{t} x_r] = \mu \quad (104)$$

and

$$E[\mathrm{svar}(\frac{1}{t} \Sigma_{r=1}^{t} x_r)] = \frac{1}{t(t-1)} \sum_{s=2}^{t} \frac{s}{s-1} E[(\frac{1}{s} \sum_{r=1}^{s} x_r - x_s)^2]$$

$$= \frac{1}{t(t-1)} \sum_{s=2}^{t} \frac{s}{s-1} \left\{\frac{1}{s^2} E[\sum_{r=1}^{s} \sum_{r'=1}^{s} x_r x_{r'}] - \frac{2}{s} E[x_s \sum_{r=1}^{s} x_r] + E[x_s^2]\right\}$$

$$= \frac{1}{t(t-1)} \sum_{s=2}^{t} \left\{\frac{1}{s-1} (\sigma^2 + \mu^2) + \mu^2 - \frac{2}{s-1} (\sigma^2 + \mu^2) - 2\mu^2 + \frac{s}{s-1} (\sigma^2 + \mu^2)\right\}$$

$$= \frac{1}{t(t-1)} \sum_{s=2}^{t} \sigma^2 = \frac{\sigma^2}{t} = \mathrm{var}[\frac{1}{t} \sum_{r=1}^{t} x_r]. \quad (105)$$

Indeed, it is easy to see that

$$\text{svar}[\frac{1}{t}\sum_{r=1}^{t} x_r] = \frac{1}{t(t-1)}\sum_{s=2}^{t}\left\{(\frac{1}{s-1}-\frac{1}{s})(\sum_{r=1}^{s} x_r)^2 - \frac{2}{s-1}x_s\sum_{s=1}^{s} x_r + \frac{s}{s-1}x_s^{\,2}\right\} = \frac{1}{t(t-1)}\sum_{s=2}^{t}\left\{\frac{1}{s-1}(\sum_{r=1}^{s-1} x_r)^2 - \frac{1}{s}(\sum_{r=1}^{s} x_r)^2 + x_s^{\,2}\right\}$$

$$= \frac{1}{t(t-1)}\left\{x_1^{\,2} - \frac{1}{t}(\sum_{r=1}^{t} x_r)^2 + \sum_{s=2}^{t} x_s^{\,2}\right\}$$

$$= \frac{1}{t(t-1)}\left\{\sum_{r=1}^{t} x_r^{\,2} - \frac{1}{t}(\sum_{r=1}^{t} x_r)^2\right\}, \tag{106}$$

the usual formula for the unbiased estimator of the sample mean variance.] The advantage of the formula (102) over (106) is computational: the latter accumulates two large sums and then subtracts them, losing many significant digits of accuracy in the process; while the former adds together many contributions (all positive), whose accuracy is much less impaired. In computing either formula, one accumulates the sum

$$R_t = \sum_{r=1}^{t} x_r; \tag{107}$$

then, for (106), one also accumulates

$$S_t = \sum_{r=1}^{t} x_r^{\,2}, \tag{108}$$

while, for (102) one accumulates

$$T_t = \sum_{s=2}^{t} \frac{s}{s-1}(\frac{1}{s}R_s - x_s)^2. \tag{109}$$

A second measure of efficiency was computed by

$$\text{eff}_V[\text{estimate}] = 1/\{\text{ svar } \times \text{work}\}. \tag{110}$$

In each run of either experiment, a particular pseudo-random sequence is selected and the values of the parameters k and n are chosen. For every f_j, the estimates Ψ_{jk} and a selection of Υ_{ijknM} are then computed; and, for each of these estimates, the error, the sample variance, svar, and the two efficiency measures, eff_E and eff_V, are obtained.

The routines central to the programs used are listed in §9 below, in an algorithmic language similar to BASIC or FORTRAN.

In the first experiment, the efficiencies of the four transformations were compared, using the parameters $k = 600$, $n = 10$, and $M = 2, 4, 6$, and 8, for each of the six test-functions. The *logarithmic efficiency-ratios*,

$$\left.\begin{aligned} K^{(4)}_{\mathrm{E}ijknM} &= \log_{10}\left(\frac{\mathrm{eff}_{\mathrm{E}}[\Upsilon_{ijknM}]}{\mathrm{eff}_{\mathrm{E}}[\Upsilon_{4jknM}]}\right) \quad (i = 1, 2, 3), \\ K^{(4)}_{\mathrm{V}ijknM} &= \log_{10}\left(\frac{\mathrm{eff}_{\mathrm{V}}[\Upsilon_{ijknM}]}{\mathrm{eff}_{\mathrm{V}}[\Upsilon_{4jknM}]}\right) \quad (i = 1, 2, 3), \\ K^{(4)}_{\mathrm{E}0jknM} &= \log_{10}\left(\frac{\mathrm{eff}_{\mathrm{E}}[\Upsilon_{4jknM}]}{\mathrm{eff}_{\mathrm{E}}[\Psi_{jk}]}\right), \\ K^{(4)}_{\mathrm{V}0jknM} &= \log_{10}\left(\frac{\mathrm{eff}_{\mathrm{V}}[\Upsilon_{4jknM}]}{\mathrm{eff}_{\mathrm{V}}[\Psi_{jk}]}\right), \end{aligned}\right\} \qquad (111)$$

are given in Table 1, below. We note that the coefficients λ_{8r} (for the $\mathfrak{E}$ transformation) were not available; so the $\mathfrak{E}_8$ transformation was omitted. We observe, also, that

$$\mathfrak{E}_2 = \mathfrak{H}_2, \quad \mathfrak{F}_2 = \mathfrak{K}_2, \quad \text{and} \quad \mathfrak{F}_4 = \mathfrak{K}_4; \qquad (112)$$

so that

$$K^{(4)}_{\star 1jkn2} = K^{(4)}_{\star 3jkn2}, \quad \text{and} \quad K^{(4)}_{\star 2jknM} = 0 \ (M = 2, 4), \qquad (113)$$

for $\star$ = E, V.

TABLE 1: Comparison of the transformations.

M	j	$K^{(4)}_{\mathrm{E}0}$	$K^{(4)}_{\mathrm{V}0}$	$K^{(4)}_{\mathrm{E}1}$	$K^{(4)}_{\mathrm{V}1}$	$K^{(4)}_{\mathrm{E}2}$	$K^{(4)}_{\mathrm{V}2}$	$K^{(4)}_{\mathrm{E}3}$	$K^{(4)}_{\mathrm{V}3}$
2	1	2.99	2.76	1.32	.47	0	0	1.32	.47
		1.71	2.71	.52	.46	0	0	.52	.46
	2	1.28	1.43	1.42	.47	0	0	1.42	.47
		2.26	1.45	.37	.47	0	0	.37	.47
	3	1.25	1.40	1.43	.47	0	0	1.43	.47
		2.20	1.41	.37	.47	0	0	.37	.47

M	j	$K_{E0}^{(4)}$	$K_{V0}^{(4)}$	$K_{E1}^{(4)}$	$K_{V1}^{(4)}$	$K_{E2}^{(4)}$	$K_{V2}^{(4)}$	$K_{E3}^{(4)}$	$K_{V3}^{(4)}$
2	4	5.35	4.28	1.36	.46	0	0	1.36	.46
		5.12	4.21	.46	.45	0	0	.46	.45
	5	2.03	1.03	.29	-.90	0	0	.29	-.90
		2.08	.92	-.87	-.78	0	0	-.87	-.78
	6	3.87	3.21	1.32	.47	0	0	1.32	.47
		3.00	3.15	.52	.46	0	0	.52	.46
4	1	7.42	8.15	1.58	2.63	0	0	1.17	1.30
		7.18	8.04	1.32	1.83	0	0	.77	1.30
	2	3.34	4.46	1.59	2.77	0	0	1.37	1.28
		5.38	4.42	1.45	1.89	0	0	.90	1.44
	3	3.09	4.21	1.61	2.65	0	0	1.48	1.28
		5.14	4.17	1.52	1.92	0	0	.99	1.50
	4	9.97	9.87	1.56	2.77	0	0	1.21	1.27
		10.77	9.73	1.33	1.83	0	0	.78	1.32
	5	3.03	3.11	.45	-.83	0	0	-1.32	-1.61
		3.83	2.93	-.48	-.10	0	0	-1.26	-1.50
	6	7.81	8.11	1.58	2.76	0	0	1.22	1.30
		7.98	7.98	1.35	1.85	0	0	.80	1.34
6	1	14.67	15.28	4.83	7.28	.02	1.16	2.32	4.19
		13.48	15.24	5.01	7.56	.58	.31	2.35	2.10
	2	7.37	8.37	4.88	6.40	.01	1.15	2.09	3.07
		8.47	8.41	5.00	7.63	.58	.31	3.00	2.16
	3	6.70	7.70	4.93	6.05	.01	1.15	2.29	2.85
		7.79	7.74	5.05	7.26	.58	.31	3.09	2.27
	4	16.37	16.14	4.82	6.51	.02	1.16	1.81	3.13
		16.22	16.08	4.73	7.98	.58	.31	3.18	2.08
	5	3.16	3.05	4.02	4.75	.07	1.27	.36	1.85
		3.37	2.96	4.20	5.87	.64	.38	1.59	.70
	6	13.05	13.22	4.87	6.90	.02	1.16	2.02	3.36
		12.26	13.17	4.99	8.17	.58	.31	2.90	2.12
8	1	21.50	23.27			2.29	.33	-7.36	-5.75
		20.48	23.52			∞	∞	-7.36	-4.68
	2	11.76	15.12			1.65	2.28	1.42	1.04
		13.89	13.00			.73	5.23	.73	7.24
	3	10.42	13.78			1.65	2.27	2.58	5.03
		12.55	11.66			.72	5.22	2.39	5.57
	4	23.32	24.97			.19	-1.20	-6.53	-7.14
		22.97	22.71			1.77	2.72	-6.06	-4.36

M	j	$K_{E0}^{(4)}$	$K_{V0}^{(4)}$	$K_{E1}^{(4)}$	$K_{V1}^{(4)}$	$K_{E2}^{(4)}$	$K_{V2}^{(4)}$	$K_{E3}^{(4)}$	$K_{V3}^{(4)}$
8	5	4.51	6.77			1.75	2.37	2.24	2.14
		5.78	4.52			.83	5.30	1.32	5.64
	6	18.12	20.61			1.67	1.36	-3.35	-3.07
		18.39	18.45			.73	6.76	-4.18	.68

Each entry in Table 1 is double, corresponding to two runs of the experiment, with different pseudo-random sequences. Table 2 gives the values of h_{iknM}, computed according to (99), and the corresponding numbers of function-evaluations, $h_{iknM}W_{inM}$.

TABLE 2: Number of function-evaluations.

i =	1	2	3	4	1	2	3	4
M	h_{iknM}				$h_{iknM}\ W_{inM}$			
2	20	30	20	30	600	600	600	600
4	4	10	6	10	600	600	600	600
6	2	4	3	5	1260	560	630	600
8		2	2	3		600	720	600

Since, on the one hand, actual errors may not reflect the standard deviation of a random variable; and, on the other hand, when h is small (e.g., 2, 3, 4, 5, 6, above), the sample variances may be quite inaccurately estimated; it is clear that considerable variations may be expected, as was observed. Using the asymptotic formulae (73), (74), (77), and (92), we may compute theoretical values for the logarithmic efficiency-ratios: these are listed in Table 3.

TABLE 3: Theoretical logarithmic efficiency-ratios.

M	$K_1^{(4)}$	$K_2^{(4)}$	$K_3^{(4)}$
2	.426	0	.426
4	2.010	0	1.334
6	5.198	.433	2.359
8		1.528	3.435

It will be observed that agreement is really quite good; and it is clearly indicated that the relative efficiencies of the four antithetic transformations $\mathfrak{E}$, $\mathfrak{F}$, $\mathfrak{H}$, and $\mathfrak{K}$ are as asserted in rules (a), (b), and (c) of §7. The entries under $K_{\star 0}^{(4)}$ clearly confirm the superiority of all four antithetic transformations, as compared with crude Monte Carlo. It is a relatively easy matter to compute the theoretical logarithmic efficiency-ratios K_{0jknM} for j = 1, 2, 3, and 4 (it was not attempted for j = 5 and 6!) The results are given in Table 4 below.

TABLE 4: Theoretical logarithmic efficiency-ratios.

N	f_1	f_2	f_3	f_4
2	1.867	-.570	-.602	1.981
4	7.218	2.411	2.146	7.525
6	14.463	6.453	5.744	13.913
8	∞	11.312	9.919	20.890

Again, agreement is as good as can be expected, and the general trends are clear.

We reiterate that a K-value of, say, 5 means that the antithetic transformation yields an efficiency 100,000 times bigger than does crude Monte Carlo: that is, the same standard deviation is obtainable with 1/100,000 of the number of function-evaluations!

In the second experiment, consideration was restricted to the transformations $\mathfrak{E}_M$ (for M = 1, 2, 3, 4, 5, 6, and 7.) Only the functions f_1, f_2, f_3, and f_4 were tested; because of the relative simplicity of their theoretical properties, with which the experimental results could be compared. In this experiment, the values of the parameters k and n, and the pseudo-random sequence, were varied. First, as we already observe in Table 1, the clear superiority of Υ_{1jknM} over Ψ_{jk} is seen in all the results, with the ratio of the efficiencies, both eff_E and eff_V, increasing sharply with increasing M. (The efficiency of the antithetic estimate exceeded that of the crude estimate in 430 comparisons, out of 448 made. All exceptions occurred among the 224 cases with $k = 100$ —— in the other 224 cases, $k \geqslant 250$ —— 13 times for $M = 1$, 4 times for $M = 2$, and once for $M = 3$. This is clearly a matter of random fluctuation.) Since Table 1 already illustrates this, the additional results are omitted here, for brevity [the author will be glad to supply them in detail to any interested reader.]

The principal purpose of the second experiment was to test the accuracy of the asymptotic formula (89) and of the efficiency measures used. By (89) and (100), we see that

$$\text{eff}[\mathfrak{U}_n \mathfrak{E}_M f_j] \sim \left\{ \frac{(2M)!}{|B_{2M}|} \frac{2^{M^2-M}}{(2^M - 1)} \right\} \frac{n^{2M-1}}{(\Delta_{M-1} f_j)^2}. \qquad (114)$$

Now, we know that $|B_2| = 1/6$, $|B_4| = 1/30$, $|B_6| = 1/42$, $|B_8| = 1/30$, $|B_{10}| = 5/66$, $|B_{12}| = 691/2730$, and $|B_{14}| = 7/6$ [see Abramowitz and Stegun (1964) p. 810, or Jahnke and Emde (1945) p. 272 —— the reader is reminded of the notational remarks in relation to (23) and (24).] Thus we may compute the coefficients in the curly bracket {--} of (114): these are given in Table 5.

TABLE 5: Coefficients in relation (114).

M	coefficient
1	12
2	960
3	276480
4	330301440
5	1620224457166.5*
6	32253799671384000*
7	2587722246048300000000*

* Note: these numbers are *not* integers.

For the functions f_j ($j = 1, 2, 3, 4$), we may directly calculate the differences Δ_{M-1}. Using (94) and (95), it is easy to verify that

$$\left.\begin{aligned} f_1^{(M-1)}(z) &= 7\times6\times5\times\ldots\times(8-M)\; z^{7-M},\\ f_2^{(M-1)}(z) &= 20\times19\times\ldots\times(21-M)\; z^{20-M},\\ f_3^{(M-1)}(z) &= 20^M\, e^{20z}/(e^{20}-1),\\ f_4^{(M-1)}(z) &= \begin{cases}(-1)^{(M-1)/2}\, 4^M \cos(4z)/\sin 4 & \text{if } M \text{ is odd},\\ (-1)^{M/2}\, 4^M \sin(4z)/\sin 4 & \text{if } M \text{ is even};\end{cases}\end{aligned}\right\} \quad (115)$$

whence we see that

$$\left.\begin{aligned} \Delta_{M-1}f_1 &= 7\times6\times\ldots\times(8-M) \quad \text{if } M\leqslant 6, \;= 0 \quad \text{if } M=7,\\ \Delta_{M-1}f_2 &= 20\times19\times\ldots\times(21-M),\\ \Delta_{M-1}f_3 &= 20^M,\\ \Delta_{M-1}f_4 &= \begin{cases}(-1)^{(M-1)/2}\, 4^M\, (\cos 4 - 1)/\sin 4 & \text{if } M \text{ is odd},\\ (-1)^{M/2}\, 4^M & \text{if } M \text{ is even}.\end{cases}\end{aligned}\right\} \quad (116)$$

Using this information, it is possible to compute the right-hand side of (114); while the Monte Carlo estimates obtained in the experiment yield values of eff_E and eff_V. Table 6 presents the logarithmic ratio of the efficiencies to the asymptotic formula.

TABLE 6: Logarithmic ratios of estimates of efficiency to formula on right of (114).

M	k	n	h	j=1 E	j=1 V	j=2 E	j=2 V	j=3 E	j=3 V	j=4 E	j=4 V
1	100	4	25	-.11	.19	-.15	.31	-.13	.33	.25	.15[1]
				.17	-.08	.12	-.02	.13	-.01	.33	-.06[2]
				2.60	.00	2.36	.11	2.40	.12	3.69	-.02[3]
		10	10	.17	-.15	.05	-.14	.04	-.14	.29	-.14
	250	4	63	2.84	.05	1.33	.17	1.34	.18	1.07	.03
		10	25	2.93	-.00	2.52	.02	2.51	.02	3.40	-.01
	5000	10	500	-.06	.01	.09	.02	.11	.03	-.14	.02
	10000	20	500	-.09	.01	-.02	.01	-.02	.02	-.13	.02
2	100	4	8	.07	.02	-.17	.30	-.12	.37	-.28	-.15[1]
				.79	-.21	.50	.21	.54	.28	.63	-.29[2]
				-.03	.04	1.46	.15	1.93	.20	-.07	-.13[3]
		10	3	-.68	.83	-.51	.49	-.47	.46	-.68	.71

M	*k*	*n*	*h*	*j*=1		*j*=2		*j*=3		*j*=4	
				E	V	E	V	E	V	E	V
2	250	4	21	.99	-.02	1.40	.23	1.53	.29	.87	-.15
		10	8	-.13	.06	.09	.05	.15	.06	-.12	.02
	5000	10	167	.27	.01	-.04	.03	-.05	.05	.13	-.02
	10000	20	167	.37	.01	.15	.01	.12	.02	.29	.00
3	100	4	4	-.29	.43	.92	.45	1.43	.45	-.46	.03[1]
				.60	.69	.13	.01	.23	.06	2.78	.82[2]
				1.00	-.17	-.31	.22	-.28	.38	.52	-.15[3]
		10	2	2.37	-.01	.29	.06	.17	.10	1.20	-.01
	250	4	9	.97	-.10	.22	.12	.27	.22	2.09	-.14
		10	4	3.00	-.16	.34	-.14	.19	-.11	.86	-.13
	5000	10	71	1.56	-.01	.65	.03	.57	.06	4.16	-.03
	10000	20	71	1.91	-.02	1.10	-.00	1.00	.01	2.97	-.02
4	100	4	2	-.46	.91	-.34	1.80	-.13	1.51	-.77	1.43[1]
				-.54	3.12	.27	1.67	.84	1.73	-.66	1.85[2]
				-.57	1.96	.19	.49	.68	.52	-.67	.69[3]
		10	2	-.58	6.11	-.49	1.09	-.41	.91	-.60	1.75
	250	4	4	-.64	.79	.05	.34	.50	.34	-.77	.56
		10	2	-.58	6.11	-.49	1.09	-.41	.91	-.60	1.75
	5000	10	33	.32	-.01	.16	.02	.16	.05	.18	-.05
	10000	20	33	.35	-.01	.23	-.00	.21	.01	.28	-.02
5	100	4	2	-.08	.63	.46	.68	.18	.95	-.42	.81[1]
				2.00	1.36	-.29	2.12	-.25	2.93	.60	1.33[2]
				1.63	.20	-.28	.90	-.23	1.55	.75	.17[3]
	100	10	2	3.37	.21	.32	.30	.09	.38	1.22	.21
	250	4	2	1.63	.20	-.28	.90	-.23	1.55	.75	.17
		10	2	3.37	.21	.32	.30	.09	.38	1.22	.21
	5000	10	16	2.63	-.04	2.44	.00	2.05	.04	2.17	-.06
	10000	20	16	2.50	-.04	5.91	-.03	3.41	-.02	2.24	-.05
6	100	4	2	-.32	.81	-.39	1.63	-.05	1.21	-.78	1.88[1]
				-.58	3.33	.04	1.63	.95	1.67	-.70	1.80[2]
				-.55	2.52	.10	.53	1.03	.57	-.64	.65[3]
		10	2	-.58	2.00	-.47	1.17	-.36	.89	-.59	1.22
	250	4	2	-.55	2.52	.10	.53	1.03	.57	-.64	.65
		10	2	-.58	2.00	-.47	1.17	-.36	.89	-.59	1.22

M	k	n	h	j=1 E	j=1 V	j=2 E	j=2 V	j=3 E	j=3 V	j=4 E	j=4 V
6	5000	10	8	-.30	.05	-.10	-.01	.09	-.00	-.07	-.06
	10000	20	8	-1.50	-1.54	-.24	.02	-.18	.01	-2.80	-2.49
7	100	4	2	0	0	.65	.63	.11	1.02	-.33	.74[1]
				0	0	-.28	1.86	-.27	2.91	1.30	∞[2]
				0	1	-.25	.81	-.23	1.87	.45	-.06[3]
		10	2	1	0	.42	.33	.07	.45	-4.67	-3.80
	250	4	2	0	1	-.25	.81	-.23	1.87	.45	-.06
		10	2	1	0	.42	.33	.07	.45	-4.67	-3.80
	5000	10	4	1	0	.60	-.05	.14	.02	-4.22	-4.13
	10000	20	4	0	0	-.35	.30	.93	.10	-11.11	-9.34

[1, 2, 3] Note: These three lines represent runs made with different pseudo-random sequences. All other runs were made with the same sequence as lines marked ([3]).

These results show generally good agreement. The ideal entry would be $\log_{10} 1 = 0$. It will be observed that eff_V tends to be worse than eff_E when h is small, and better when h is large. As might be expected, the ratios improve as k, n, and M increase; since both the statistical estimates and the asymptotics improve. However, for large k and M, especially for odd values of M and for f_4, the effect of cumulative round-off errors is seen to worsen the ratios. It may be noted that only 145/448 ≈ 32% of the entries are negative: this is a slight indication that actual efficiencies may tend to be better than the formula (114) suggests.

9. PROGRAMS.

We conclude this paper with a listing of the routines which compute the Monte Carlo estimates and which generate the various antithetic transformations.

The language used below is rather like BASIC or FORTRAN: the usual arithmetic operations are used, and *variables* are denoted by strings of (lower-case or capital) Roman letters. We use ← for *assignment* [the instruction 'V ← *expression*' copies into the memory location denoted by the variable 'V' the value computed for the '*expression*' given]; '+', '-', '*', '/', and '↑' for *addition*, *subtraction*, *multiplication*, *division*, and *exponentiation*, respectively; and we adopt the usual *hierarchy of precedence* of these operations, with ↑ first, then * and /, then + and -, overridden, as usual, by *parentheses* (). MAX(x,y) computes the (algebraically) greater of the expressions denoted by x and y; FL(z) computes the *floor* function of z; SQRT(u) computes the (non-negative) *square root* of the expression u; f(j,v) computes the value of $f_j(v)$ for $0 \leqslant v \leqslant 1$ and j = 1, 2, 3, 4, 5, 6, as required; and RND computes the next number in a pseudo-random sequence.

The Monte Carlo routine follows:

```
value ← correct value θ [if known]
h ← number of random samples to be taken
R ← 0     [sum for estimate]
T ← 0     [sum for variance]
FOR s = 1 TO h
    x ← RND
    g ← estimator(i,x)
    R ← R + g
    IF  s > 1  THEN  T ← T + (s/(s-1))*(R/s - g)↑2
NEXT s
estimate ← R/h     [estimate of value θ]
error ← estimate - value    [if value is known]
svar ← T/(h*(h-1))     [estimate of variance of estimate]
stdev ← SQRT(svar)     [standard deviation]
IF  error ≠ 0  THEN  effE ← 1/(h*error↑2)  ELSE  effE ← 0
IF  svar > 0  THEN  effV ← 1/(h*svar)  ELSE  effV ← 0
                       [estimates of efficiency]
```

Comments are given in brackets []. In the case of crude Monte Carlo, i = 0 and estimator(i,x) = f(j,x). For the Σ transformation, i = 1 and the corresponding estimator routine is:

```
h ← MAX(2,FL((k/(n*(2↑M - 1))) + (1/2)))
```

and

```
estimator ← 0
FOR r = 1 TO M
    sum ← 0
    step ← 1/(n*2↑(r-1))
    y ← x*step
```

```
            FOR s = 1 TO 1/step
                sum ← sum + f(j,y)
                y ← y + step
            NEXT s
            estimator ← estimator + sum*step*lambda(M,r)
        NEXT r
```

Here, lambda(M,r) computes the coefficient λ_{Mr}. We note that, by (43), we have that $\mathfrak{U}_n\mathfrak{U}_{2^{p-1}} = \mathfrak{U}_{n2^{p-1}}$, and the terms of $\mathfrak{U}_n\mathfrak{E}_M$ are computed accordingly, from (80). For <u>the $\mathfrak{F}$ transformation,</u> i = 2 and the corresponding estimator routine is:

```
        h ← MAX(2,FL((k/(2*n*(2↑(M/2) - 1))) + (1/2)))
```

and

```
        estimator ← 0
        FOR r = 1 TO M/2
            sum ← 0
            step ← 1/(n*2↑(r-1))
            y ← x*step
            FOR s = 1 TO 1/step
                sum ← sum + f(j,y) + f(j,1-y)
                y ← y + step
            NEXT s
            estimator ← estimator + sum*step*mu(M,r)
        NEXT r
        estimator ← estimator/2
```

Here, mu(M,r) computes the coefficient μ_{Mr}. Again, we use the fact that $\mathfrak{U}_n\mathfrak{U}_{2^{p-1}} = \mathfrak{U}_{n2^{p-1}}$. For <u>the $\mathfrak{A}$ transformation,</u> i = 3 and the estimator routine is:

```
        h ← MAX(2,FL((2*k/(n*M*(M + 1))) + (1/2)))
```

and

```
        estimator ← 0
        FOR r = 1 TO M
            sum ← 0
            step ← 1/(n*r)
            y ← x*step
            FOR s = 1 TO 1/step
                sum ← sum + f(j,y)
                y ← y + step
            NEXT s
            estimator ← estimator + sum*step*alpha(M,r)
        NEXT r
```

Here alpha(M,r) computes the coefficient a_{Mr}; and we use the fact that $\mathfrak{U}_n\mathfrak{U}_p = \mathfrak{U}_{np}$. The same identity also applies to the case of the

𝔎 transformation, when i = 4 and the estimator routine is:

```
h ← MAX(2,FL((2*k/(n*M*(M/2 + 1))) + (1/2)))
```

and

```
estimator ← 0
FOR r = 1 TO M/2
    sum ← 0
    step ← 1/(n*r)
    y ← x*step
    FOR s = 1 TO 1/step
        sum ← sum + f(j,y) + f(j,1-y)
        y ← y + step
    NEXT s
    estimator ← estimator + sum*step*beta(M,r)
NEXT r
estimator ← estimator/2
```

Here, beta(M,r) computes the coefficient β_{Mr}. We observe that, as is indicated by (45) and (80), when i is *even* (for the 𝔉 and 𝔎 transformations), M must be even too.

10. REFERENCES.

M. Abramowitz and I. A. Stegun (1964). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* National Bureau of Standards (Applied Mathematics Series, No. 55)

B. V. Gnedenko (1963). *The Theory of Probability.* (Translated from the Russian by B. D. Seckler.) Chelsea Publishing Co., New York. Second Edition.

J. H. Halton (1970). A retrospective and prospective survey of the Monte Carlo method. *SIAM Review* (Society for Industrial and Applied Mathematics), vol. 12, pp. 1 - 63.

J. H. Halton and D. C. Handscomb (1957). A method for increasing the efficiency of Monte Carlo integration. *ACM Journal* (Association for Computing Machinery), vol. 4, pp. 329 - 340.

J. M. Hammersley and J. G. Mauldon (1956). General principles of antithetic variates. *Proceedings of the Cambridge Philosophical Society*, vol. 52, pp. 476 - 481.

J. M. Hammersley and K. W. Morton (1956). A new Monte Carlo technique: antithetic variates. *Proceedings of the Cambridge Philosophical Society*, vol. 52, pp. 449 - 475.

G. H. Hardy and E. M. Wright (1962). *An Introduction to the Theory of Numbers.* Clarendon Press, Oxford. Fourth Edition, corrected.

D. C. Handscomb (1958). Proof of the antithetic variate theorem for $n > 2$. *Proceedings of the Cambridge Philosophical Society*, vol. 54, pp. 300 - 301.

D. C. Handscomb (1964). A rigorous lower bound for the efficiency of Monte Carlo integration. *Proceedings of the Cambridge Philosophical Society*, vol. 60, pp. 357 - 358.

E. Jahnke and F. Emde (1945). *Tables of Functions with Formulae and Curves.* Dover Publications, New York. Fourth Edition.

S. Lang (1971). *Linear Algebra.* Addison-Wesley Publishing Co., Reading, Massachusetts. Second Edition.

P. J. Laurent (1961). Remarque sur l'évaluation d'intégrales par la méthode de Monte-Carlo [Remark on the evaluation of integrals by the Monte Carlo method.] *Comptes-Rendus de l'Académie des Sciences*, vol. 253, pp. 610 - 612.

L. Mirsky (1955). *An Introduction to Linear Algebra.* Clarendon Press, Oxford. Corrected, 1972.

K. W. Morton (1957). A generalization of the antithetic variate technique for evaluating integrals. *Journal of Mathematical Physics*, vol. 36, pp.289 - 293.

G. E. Shilov (1971). *Linear Algebra.* (Translated from the Russian by R. A. Silverman.) Prentice-Hall, New Jersey.

E. C. Titchmarsh (1951). *The Theory of the Riemann Zeta Function.* Clarendon Press, Oxford.

E. T. Whittaker and G. N. Watson (1927). *A Course of Modern Analysis.* Cambridge University Press, Cambridge. Fourth Edition.