SEQUENTIAL MONTE CARLO

By J. H. HALTON

Received 5 October 1960

ABSTRACT. This paper defines the concept of sequential Monte Carlo and outlines the principal modes of approach which may be expected to yield useful sequential processes. Three workable sequential processes, derived from a non-sequential method of J. von Neumann and S. M. Ulam for solving systems of linear algebraic equations, are described and analysed in detail.

1. *Monte Carlo processes.* We define a Monte Carlo process of numerical solution as one consisting of a finite number of operations, involving the use of one or more random numbers (in the wider sense, including pseudo-random and quasi-random numbers),* together with the parameters of the problem, and leading to an estimate of its solution. The process is viewed as a statistical (stochastic, random) one, whose repetition, using a random number sequence, leads to a sequence of primary estimates, which may be combined to give (improved) secondary estimates of the required solution.

If $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ is a sequence of random points uniformly and independently distributed in a unit hypercube $\mathscr{C}$ of suitable dimensionality, it can be shown† that the primary estimates can always be written as a sequence of functions $f_1(\xi_1), f_2(\xi_2), \ldots, f_n(\xi_n), \ldots$, and the secondary estimates will be functions $F_1 = f_1$, $F_2(f_1, f_2), \ldots, F_n(f_1, f_2, \ldots, f_n), \ldots$. It is clearly particularly convenient to have $F_n = F_n(f_n, F_{n-1})$, so that only the value of the last secondary estimate need be retained at any stage.

We shall limit the discussion to functions belonging to L^2, to ensure the existence of their means, variances, and covariances, defined in the usual manner, in any Lebesgue subset $\mathscr{E}$ of the unit hypercube $\mathscr{C}$. A function $f(\xi)$ will be termed an *unbiased estimator* of a number θ if its mean, taken over $\mathscr{C}$, equals θ. Two estimators $f(\xi)$ and $g(\xi)$ are said to be *uncorrelated* if their covariance in $\mathscr{C}$ vanishes. Note that they need not therefore be independent.

In order to illustrate the processes and concepts, we consider the problem of evaluating the integral

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

The simplest Monte Carlo process for this is to sample $f(\xi_1), f(\xi_2), \ldots, f(\xi_n), \ldots$ for our primary estimates, where each ξ_n is uniformly distributed in $(0, 1)$. The secondary

* Number sequences which, although generated by determinate mathematical processes, yet have, for a considerable length at least, all the properties of random sequences, are termed *pseudo-random.* Sequences which are determinately generated and which do not satisfy all the statistical tests of randomness, but which yield unbiased estimates when used in certain classes of Monte Carlo processes, are called *quasi-random.* (See (6).)

† See Lévy (14), p. 121.

https://doi.org/10.1017/S0305004100036227 Published online by Cambridge University Press

estimates could then be the averages, $F_n = \sum_{i=1}^{n} f(\xi_i)/n$. The mean of the estimator $f(\xi)$ is then precisely the integral (1), so that there is no bias; and the variance is

$$\int_0^1 [f(x)-\theta]^2\,dx.$$

The successive primary estimates are uncorrelated (and independent).

It is clear, however, that we may sample, for instance, a sequence of functions $f_1(\eta_1), f_2(\eta_2), \ldots, f_n(\eta_n), \ldots$, where

$$f_n(y) = e_n(y) + \frac{f(y)}{g_n(y)}; \tag{2}$$

and $f_n(\eta_n)$ will give an unbiased estimate of θ, for each n, if $f(y)/g_n(y)$ is bounded and integrable in $(0, 1)$,

$$g_n(y) \geqslant 0, \quad \int_0^1 g_n(y)\,dy = 1, \quad \int_0^1 e_n(y)\,g_n(y)\,dy = 0, \tag{3}$$

and the numbers η_n are random variables with the distribution functions

$$q_n(t) = \int_0^t g_n(y)\,dy. \tag{4}$$

Thus we may take $\eta_n = Q_n(\xi_n)$, where ξ_n is uniform in $(0, 1)$ and $Q_n(x)$ is the upper bound of all t such that $q_n(t) < x$.

It follows that, to use the estimators $f_n(\eta_n)$, we must be able to evaluate the integrals involved in (3) and (4).

2. *The reduction of variance.* A great deal of effort has gone into devising methods for reducing the variances of the primary estimates in Monte Carlo processes by a suitable choice of the estimator functions.* Consider, for instance, the use of the functions (2) in estimating the value of the integral (1). First, put $e_n(y) = 0$ and let $g_n(y)$ be an approximation $g(y)$ to a multiple of $f(y)$, as far as is compatible with (3). Let us put

$$f(y) = \theta g(y)\{1+h(y)\}; \tag{5}$$

then (1) to (5) give us that

$$g(y) \geqslant 0 \quad \text{and} \quad \int_0^1 g(y)\,dy = 1, \tag{6}$$

$$f_n(y) = f(y)/g(y) = \theta\{1+h(y)\}, \tag{7}$$

$$\text{mean}\,(f/g) = \int_0^1 \{f(y)/g(y)\}\,g(y)\,dy = \theta \tag{8}$$

and

$$\text{var}\,(f/g) = \int_0^1 \{\theta h(y)\}^2\,g(y)\,dy. \tag{9}$$

The resultant process is called *unbiased importance sampling*, and is well known. The basic idea behind it is simple. If $g(y)$ is almost proportional to $f(y)$, then f/g varies little, and its variance is thus small (see (9)). The conditions (6) ensure that $q_n(t)$ is a proper distribution function, and the process is then always unbiased (see (8)).

* See, for example, Kahn (12); Fieller and Hartley (3); Hammersley, Mauldon, and Morton (8, 9, 10); and Halton and Handscomb (7, 11).

https://doi.org/10.1017/S0305004100036227 Published online by Cambridge University Press

Note, however, that it is these conditions which limit the use of the method, for we can only make $h(y)$ small if $f(y)$ is almost everywhere non-negative in $(0, 1)$, and we must be able to integrate $g(y)$ explicitly.*

Suppose, next, that we relax the condition that the estimator f/g should be unbiased. Keeping the definition of $h(y)$ in (5), we let η_n be sampled with a distribution function

$$\Phi(t) = \int_0^t \phi(y)\, dy, \tag{10}$$

where $\phi(y)$ is an arbitrary, non-negative, integrable function in $(0, 1)$, for which $\Phi(1) = 1$. Then

$$\text{mean}_\phi\,(f/g) = \int_0^1 \{f(y)/g(y)\}\,\phi(y)\, dy = \theta + \int_0^1 \theta h(y)\,\phi(y)\, dy \tag{11}$$

and

$$\text{var}_\phi\,(f/g) = \int_0^1 \{\theta h(y)\}^2\,\phi(y)\, dy - \left(\int_0^1 \theta h(y)\,\phi(y)\, dy\right)^2. \tag{12}$$

As was to be expected, it is still true that, if $h(y)$ can be made small, both the *bias*, $\{\text{mean}_\phi\,(f/g) - \theta\}$, and the variance, $\text{var}_\phi\,(f/g)$, of the estimator will be small.

This process will be called *biased importance sampling*. The reason why it has not hitherto had any prominence is that, while simple averaging over n independent primary estimates will reduce the variance by a factor n^{-1}, the bias is fixed once and for all by the choice of $g(y)$ and $\phi(y)$.

Returning to (2), let us finally consider the effect of putting $g_n(y) = 1$ and choosing $e_n(y)$ to be an approximation $e(y)$ to a constant minus $f(y)$. Let us write

$$e(y) = \theta - f(y) + h(y). \tag{13}$$

Then (1) to (4) give us that

$$\int_0^1 e(y)\, dy = \int_0^1 h(y)\, dy = 0, \tag{14}$$

$$f_n(y) = e(y) + f(y) = \theta + h(y), \tag{15}$$

$$\text{mean}\,(e+f) = \int_0^1 \{e(y) + f(y)\}\, dy = \theta \tag{16}$$

and

$$\text{var}\,(e+f) = \int_0^1 \{h(y)\}^2\, dy. \tag{17}$$

Thus, if $h(y)$ is small, so is $\text{var}\,(e+f)$, and the correlation coefficient of e and f is nearly -1. This process is called *correlated*, or *control variate*, *sampling*. Note that, by (4), since $g_n(y) = 1$, η_n is uniformly distributed in $(0, 1)$; the process is then unbiased.

We observe that, in all the processes described above, it appears to be possible to choose the estimators so as to reduce their variances to zero; however, the formulae obtained imply a knowledge of θ, the answer which we are seeking, so that we are limited to using the best available approximations. Nevertheless, it is clear that, as

* The possibility of interpreting (4) when $g(y)$ is negative will be dealt with elsewhere.

https://doi.org/10.1017/S0305004100036227 Published online by Cambridge University Press

we take successive samples and so improve our knowledge of the situation, we should be able to reduce the variances of subsequent estimates by adjusting our estimator functions. We shall return to this idea in § 4.

3. *Sequential Monte Carlo.* We shall define sequential Monte Carlo processes as those in which the sampling is done in stages, and in which the sampling scheme or estimator function (or both), at each stage, depend upon the results of one or more previous stages. The most convenient arrangement is to have $f_n = f_n(\xi, F_{n-1})$.

We shall say that a sequential process is *proper* if $\mathrm{var}\,(f_n)$ decreases monotonically with n, and that it is *convergent* if (apart from round-off errors arising in the numerical computation),

$$\mathrm{var}\,(f_n) \to 0 \quad \text{as} \quad n \to \infty. \tag{18}$$

We see that this is equivalent to stating that f_n converges in (second degree) mean to θ, which implies that $f_n \to \theta$ in measure (that is, *in probability*) over all random sequences.* This means that, for any positive ϵ, the probability that $|f_n - \theta| < \epsilon$ tends to unity as $n \to \infty$.

It is also interesting to consider the stronger *convergence with probability one.* Suppose now that the secondary estimates are obtained by a linear formula,

$$F_n = v_n f_n + (1 - v_n) F_{n-1}, \quad v_1 = 1, \quad 0 \leqslant v_n \leqslant 1; \tag{19}$$

so that

$$F_n = \sum_{m=1}^{n} w_m^{(n)} f_m, \tag{20}$$

where

$$w_n^{(n)} = v_n, \quad w_m^{(n)} = v_m \prod_{r=m+1}^{n} (1 - v_r) \quad (m < n), \quad 0 \leqslant w_m^{(n)} \leqslant 1. \tag{21}$$

We then obtain

THEOREM 1. *If the functions $f_n(\xi)$ are unbiased estimators of θ, combined according to* (19) *to give secondary estimates F_n, if $u_n = \max_m (w_m^{(n)}) \to 0$ as $n \to \infty$, and if there is a t such that both $\sum_{s=1}^{\infty} v_s s^{-t}$ converges and $\mathrm{var}\,(f_n) = O(s^{-r})$, where $r > 2t+1$; then F_n converges to θ, as $n \to \infty$, for almost every random sequence, that is, with probability one.*

Proof. Let $\mathscr{E}_s$ be the set of all ξ such that $|f_s - \theta| \geqslant \gamma s^{-t} > 0$. Then, if $\mathrm{var}\,(f_s) = O(s^{-r})$, there exists a finite constant B such that

$$\gamma^2 s^{-2t} p(\mathscr{E}_s) \leqslant \int_{\mathscr{E}_s} (f_s - \theta)^2 \, dp \leqslant \int_{\mathscr{C}} (f_s - \theta)^2 \, dp = \mathrm{var}\,(f_s) < B s^{-r},$$

where p is the probability measure and $p(\mathscr{C}) = 1$, that is, $\mathscr{C}$ is the unit hypercube which is the set of all permitted ξ. The primary estimates of the process are $f_1(\xi_1)$, $f_2(\xi_2)$, ..., $f_s(\xi_s)$, ..., where the random points $\xi_1, \xi_2, \ldots, \xi_s, \ldots$ are independently distributed in $\mathscr{C}$. Let $\mathscr{F}_m = \bigtimes_{s=m+1}^{\infty} (\mathscr{C} - \mathscr{E}_s)$, so that $\mathscr{F}_m$ is the set of all sequences

* See Halmos(5), p. 103, Theorem A. The proof can be directly modified to apply to the quadratic mean.

https://doi.org/10.1017/S0305004100036227 Published online by Cambridge University Press

$\xi_{m+1}, \xi_{m+2}, \xi_{m+3}, \ldots$, such that $|f_s(\xi_s)-\theta| < \gamma s^{-t}$ for all $s > m$. For any arbitrary $\gamma' > 0$, if $r > 2t+1$ and m is sufficiently large,*

$$\begin{aligned} p(\mathscr{F}_m) &= \prod_{s=m+1}^{\infty} p(\mathscr{C}-\mathscr{E}_s) > \prod_{s=m+1}^{\infty} (1-B\gamma^{-2}s^{-r+2t}) \\ &> \prod_{s=m+1}^{\infty} \exp(-2B\gamma^{-2}s^{-r+2t}) = \exp\left\{-2B\gamma^{-2}\sum_{s=m+1}^{\infty} s^{-r+2t}\right\} \\ &\geqslant \exp\left\{-2B\gamma^{-2}\int_m^{\infty} s^{-r+2t}\,ds\right\} = \exp\left\{-\frac{2B\gamma^{-2}}{r-2t-1}m^{-r+2t+1}\right\} \\ &> 1-\gamma'. \end{aligned}$$

Now, the series of positive terms $\sum_{s=1}^{\infty} v_s s^{-t}$ converges; so, for any arbitrary $\gamma'' > 0$, we can find an m' such that $\sum_{s=n+1}^{n'} v_s s^{-t} < \gamma''/\gamma$ for all $n' > n \geqslant m'$. Thus, for all random sequences in $\mathscr{F}_m$, if we put $m'' = \max(m, m')$,

$$\left|\sum_{s=n+1}^{n'} v_s(f_s(\xi_s)-\theta)\right| \leqslant \sum_{s=n+1}^{n'} v_s|f_s-\theta| < \gamma\sum_{s=n+1}^{n'} v_s s^{-t} < \gamma'' \quad \text{for all } n' > n \geqslant m''.$$

Since γ'' is arbitrarily small (and $\mathscr{F}_m$ does not depend on it), this is Cauchy's condition for the (uniform) absolute convergence of $\sum_{s=1}^{\infty} v_s(f_s-\theta)$ in $\mathscr{F}_m$. But γ' is also arbitrarily small. This proves that $\sum_{s=1}^{\infty} v_s(f_s-\theta)$ converges for almost all random sequences $\xi_1, \xi_2, \ldots, \xi_n, \ldots$. Let the sum be c for one such sequence. Then, for any positive δ, there is an m_0 such that

$$\left|\sum_{s=1}^{m} v_s(f_s-\theta)-c\right| < \delta \quad \text{whenever} \quad m \geqslant m_0.$$

Again, if $u_n \to 0$ as $n \to \infty$, for any $\delta' > 0$ and given m_0, we can find an n_0 ($> m_0$) such that $0 \leqslant u_n < \delta'$ whenever $n \geqslant n_0$. Hence

$$\begin{aligned} \left|\sum_{m=1}^{m_0} w_m^{(n)}\left\{\sum_{s=1}^{m} v_s(f_s-\theta)-c\right\}\right| &\leqslant u_n m_0 \sup_m \left|\sum_{s=1}^{m} v_s(f_s-\theta)-c\right| \\ &= u_n m_0 M < \delta' m_0 M \quad \text{whenever} \quad n \geqslant n_0, \end{aligned}$$

where M is finite and independent of n. Therefore, since $\sum_{m=1}^{n} w_m^{(n)} = 1$,

$$\begin{aligned} \left|\sum_{m=1}^{n} w_m^{(n)}\sum_{s=1}^{m} v_s(f_s-\theta)-c\right| &= \left|\sum_{m=1}^{n} w_m^{(n)}\left\{\sum_{s=1}^{m} v_s(f_s-\theta)-c\right\}\right| \\ &\leqslant \left|\sum_{m=1}^{m_0} w_m^{(n)}\left\{\sum_{s=1}^{m} v_s(f_s-\theta)-c\right\}\right| + \sum_{m=m_0+1}^{n} w_m^{(n)}\left|\sum_{s=1}^{m} v_s(f_s-\theta)-c\right| \\ &< \delta' m_0 M + \delta\sum_{m=m_0+1}^{n} w_m^{(n)} \\ &< \delta' m_0 M + \delta \quad \text{whenever} \quad n \geqslant n_0. \end{aligned}$$

† See Halmos (5), p. 191.

https://doi.org/10.1017/S0305004100036227 Published online by Cambridge University Press

Thus, given any positive ϵ, if we choose $\delta = \frac{1}{4}\epsilon$, which fixes m_0, and then

$$\delta' = \tfrac{1}{2}\epsilon/(m_0 M + \delta + c),$$

which fixes n_0, then

$$\begin{aligned}
|F_n - \theta| &= \left| \sum_{m=1}^{n} w_m^{(n)}(f_m - \theta) \right| \\
&= \left| \sum_{m=1}^{n} \frac{w_m^{(n)}}{v_m} \left\{ \sum_{s=1}^{m} v_s(f_s - \theta) - \sum_{s=1}^{m-1} v_s(f_s - \theta) \right\} \right| \\
&= \left| \sum_{s=1}^{n} v_s(f_s - \theta) + \sum_{m=2}^{n} \left(\frac{w_{m-1}^{(n)}}{v_{m-1}} - \frac{w_m^{(n)}}{v_m} \right) \sum_{s=1}^{m-1} v_s(f_s - \theta) \right| \\
&= \left| \sum_{s=1}^{n} v_s(f_s - \theta) - \sum_{m=2}^{n} v_m \left\{ \prod_{r=m+1}^{n} (1 - v_r) \right\} \sum_{s=1}^{m-1} v_s(f_s - \theta) \right| \\
&= \left| \sum_{s=1}^{n} v_s(f_s - \theta) - \sum_{m=1}^{n} w_m^{(n)} \sum_{s=1}^{m} v_s(f_s - \theta) + \sum_{m=1}^{n} w_1^{(n)} v_m(f_m - \theta) \right| \\
&\leqslant \left| \sum_{s=1}^{n} v_s(f_s - \theta) - c \right| + \left| \sum_{m=1}^{n} w_m^{(n)} \sum_{s=1}^{m} v_s(f_s - \theta) - c \right| + u_n \left| \sum_{m=1}^{n} v_m(f_m - \theta) - c \right| + u_n |c| \\
&< \delta + (\delta' m_0 M + \delta) + \delta'\delta + \delta' c = \epsilon \quad \text{whenever} \quad n \geqslant n_0.
\end{aligned}$$

Thus we have shown that $F_n \to \theta$ whenever $\sum_{n=1}^{\infty} v_n(f_n - \theta)$ converges, and the Theorem follows.

4. *Three sequential processes.* Most of the previous work on sequential sampling is limited to the analysis of sequential procedures for deciding when to stop the sampling process.*

Since the aim of sequential Monte Carlo is the progressive reduction of variance, we naturally turn to the methods described in § 2 in search of ideas.

Consider first unbiased importance sampling, as applied to the evaluation of the integral (1). We may construct a sequential scheme as follows. At stage n of the process the values of $f(\eta_i)$ at $(n-1)$ points $\eta_1, \eta_2, \ldots, \eta_{n-1}$ will have been obtained. We construct a function $G_n(y)$, such that $G_n(\eta_i)$ approximates to $f(\eta_i)$ for $i = 1, 2, \ldots, (n-1)$, and such that $\int_0^t G_n(y)\,dy$ is known. We define

$$g_n(y) = G_n(y) \Big/ \int_0^1 G_n(y')\,dy', \tag{22}$$

so that $g_n(y)$ satisfies the conditions (6). Thus, if we use the estimator function

$$f_n(\eta_n) = f(\eta_n)/g_n(\eta_n) = \theta\{1 + h_n(\eta_n)\}, \tag{23}$$

where η_n has the distribution $q_n(t)$ in (4), then, as in (8) and (9), mean $(f_n) = \theta$ and

$$\text{var}\,(f_n) = \theta^2 \int_0^1 \{h_n(y)\}^2 g_n(y)\,dy. \tag{24}$$

* See, for instance, the work of Wald (18, 19, 20) and others (2, 13, 23, 24). The only work on a sequential Monte Carlo method, in the sense defined here, which I have been able to find, is in a paper by Marshall (15). He considers only a two-stage process.

https://doi.org/10.1017/S0305004100036227 Published online by Cambridge University Press

The scheme described will be a proper sequential process if var (f_n) decreases monotonically with n; and a sufficient condition for convergence is, by (18) and (24), that

$$\sup_{0 \leqslant y \leqslant 1} |h_n(y)| \to 0 \quad \text{as} \quad n \to \infty. \tag{25}$$

As our secondary estimates, we may take an unbiased average of the form (19), or we may take $F_n = \int_0^1 G_n(y)\,dy$, but this will generally be biased (although the bias will tend to zero as $n \to \infty$).

Consider next biased importance sampling. The construction of $G_n(y)$ and the form of the estimator function defined in (22) and (23) remain unchanged. The function $q_n(t)$ is replaced by a distribution function

$$\Phi_n(t) = \int_0^t \phi_n(y)\,dy, \tag{26}$$

as in (10). The equivalents of (11) and (12) now hold;

$$\text{bias}_{\phi_n}(f_n) = \theta \int_0^1 h_n(y)\,\phi_n(y)\,dy, \tag{27}$$

$$\text{var}_{\phi_n}(f_n) = \theta^2 \left[\int_0^1 \{h_n(y)\}^2 \phi_n(y)\,dy - \left(\int_0^1 h_n(y)\,\phi_n(y)\,dy \right)^2 \right]. \tag{28}$$

Again, if (25) holds, both the bias and the variance of f_n will tend to zero as $n \to \infty$.* Thus we see that, in a sequential process, the fact that the estimator functions are biased is not very important, since the total error converges in probability to zero.

Turning finally to correlated sampling, we define $G_n(y)$ as before, and use an estimator function

$$f_n(\xi_n) = f(\xi_n) + e_n(\xi_n) = f(\xi_n) + \int_0^1 G_n(y)\,dy - G_n(\xi_n) = \theta + h_n(\xi_n). \tag{29}$$

Then, as in (16) and (17), mean $(f_n) = \theta$ and

$$\text{var}(f_n) = \int_0^1 \{h_n(y)\}^2\,dy \leqslant \int_0^1 \{f(y) - G_n(y)\}^2\,dy. \tag{30}$$

Thus, if

$$\sup_{0 \leqslant y \leqslant 1} |f(y) - G_n(y)| \to 0 \quad \text{as} \quad n \to \infty, \tag{31}$$

the process is convergent by (18) and (30).

5. *The Neumann–Ulam process.* Consider a system of N linear algebraic equations, represented by the matrix equation

$$\mathbf{Ax} = \mathbf{b}. \tag{32}$$

We take $\mathbf{A}$ to be symmetric, without loss of generality.† Let

$$\mathbf{H} = \mathbf{I} - c\mathbf{A}^2 \quad \text{and} \quad \mathbf{a} = c\mathbf{Ab}. \tag{33}$$

Then (32) becomes

$$\mathbf{x} = \mathbf{a} + \mathbf{Hx}, \tag{34}$$

* This contrasts with the *fixed bias* which is encountered in the non-sequential biased importance sampling process described in § 2.

† If $\mathbf{A}$ is not symmetric, premultiply (32) by the transposed matrix $\mathbf{A}'$.

https://doi.org/10.1017/S0305004100036227 Published online by Cambridge University Press

and a suitable choice of c ensures that all the eigenvalues of $\mathbf{H}$ are of modulus less than one.* The solution of (34) is then given by the convergent Neumann vector series

$$\mathbf{x} = (\mathbf{I}-\mathbf{H})^{-1}\mathbf{a} = (\mathbf{I}+\mathbf{H}+\mathbf{H}^2+\ldots+\mathbf{H}^n+\ldots)\mathbf{a}. \tag{35}$$

We note that, if we write

$$\mathbf{y} = \mathbf{x}-\mathbf{X} \quad \text{and} \quad \mathbf{d} = \mathbf{a}+\mathbf{H}\mathbf{X}-\mathbf{X}, \tag{36}$$

then (33) transforms into an equation of the same form,

$$\mathbf{y} = \mathbf{d}+\mathbf{H}\mathbf{y}. \tag{37}$$

The elements x_i, a_i, and H_{ij} in (34) are defined over the index set $\{1, 2, \ldots, N\}$. It is clear that we may add an index value 0, producing 'augmented' vectors $\hat{\mathbf{x}}$, $\hat{\mathbf{a}}$, and matrix $\hat{\mathbf{H}}$, with the definitions

$$\hat{x}_0 = 0, \quad \hat{a}_0 = 0, \quad \hat{H}_{0j} = \hat{H}_{i0} = 0, \quad \hat{H}_{00} = 1, \tag{38}$$

and then we have an augmented equation $\hat{\mathbf{x}} = \hat{\mathbf{a}}+\mathbf{H}\hat{\mathbf{x}}$, just like (34), whose solution is (35), together with $\hat{x}_0 = 0$.

Forsythe and Leibler (4) have described an (apparently unpublished) Monte Carlo method, devised by J. von Neumann and S. M. Ulam, for estimating the value of the vector series (35).†

We consider the augmented index set $\{0, 1, 2, \ldots, N\}$ and the sequences of index values $\Gamma_{ni} \equiv \{i = i_0, i_1, \ldots, i_n\}$. We look upon Γ_{ni} as the beginning of an infinite *random walk* Γ_i on the index set, with the probability for the nth step $p_{i_{n-1}i_n}$ depending on its end-points (this is a Markov chain). With each random walk, we associate two numerical sequences; first,

$$X_{0i} = 1, \quad X_{mi} = X_{m-1,i}\{H_{i_{m-1}i_m}/p_{i_{m-1}i_m}\}, \tag{39}$$

so that

$$X_{ni} = \prod_{m=1}^{n}\{H_{i_{m-1}i_m}/p_{i_{m-1}i_m}\}; \tag{40}$$

and then the derived sequence,

$$Y_{ni} = \sum_{m=0}^{n} X_{mi}a_{i_m}. \tag{41}$$

Note that, by (38), once the walk Γ_i visits the index value 0, the subsequent X_{ni} all vanish, and the Y_{ni} are therefore unchanged thereafter. Thus the walk effectively terminates on reaching index value 0.

LEMMA 1. *Over all random walks Γ_i whose initial point is index value i, Y_{ni} converges in probability to a number Y_i, and*

$$\text{mean}\,(Y_i) = x_i. \tag{42}$$

* If the eigenvalues of $\mathbf{A}$ are α_i, and if $\bar{\alpha} > \max_i |\alpha_i|$, then take $c = \bar{\alpha}^{-2}$. In the absence of more precise information, we can always put $\bar{\alpha} = \max_i \left(\sum_{j=1}^{N} |A_{ij}|\right)$ (or slightly more, if all the A_{ij} in each column are of the same sign and all the sums $\sum_{j=1}^{N} |A_{ij}|$ are equal).

† The method has been applied to matrix problems (see, for example, Opler (16) and Wasow (22)), integral equations (see, for example, Cutkosky (1) and Page (17)), and elliptic differential and difference equations (see, for example, Wasow (21)).

https://doi.org/10.1017/S0305004100036227 Published online by Cambridge University Press

Proof. The probability of the sequence Γ_{ni} is $\mathscr{P}(\Gamma_{ni}) = \prod_{m=1}^{n} p_{i_{m-1}i_m}$. Thus, by (40) and (41),

$$\begin{aligned}\text{mean}\,(Y_{ni}) &= \sum_{m=0}^{n} \text{mean}\,(X_{mi}a_{i_m}) = \sum_{m=0}^{n} \sum_{\Gamma_{ni}} \mathscr{P}(\Gamma_{ni})\, X_{mi}a_{i_m} \\ &= \sum_{m=0}^{n} \sum_{\text{Each } i_r=0}^{N} \left(\prod_{r=1}^{m} H_{i_{r-1}i_r}\right) a_{i_m} \left(\prod_{r=m+1}^{n} p_{i_{r-1}i_r}\right) = \sum_{m=0}^{n} (\mathrm{H}^m \mathrm{a})_i,\end{aligned}$$

and hence, by (35), $\lim_{n\to\infty}\{\text{mean}\,(Y_{ni})\} = x_i$. Now, given i, consider the Y_{ni} as a sequence of functions defined on the space $\mathscr{W}_i$ of all random walks Γ_{ni}, the infinite-dimensional product-space of identical sets $(0, 1, 2, \ldots, N)$. Clearly the Y_{ni} are all integrable simple functions over $\mathscr{W}_i$, with respect to probability, their integrals being given by the mean values calculated above.* Also, since the series (35) converges, mean $(|Y_{mi} - Y_{ni}|) \to 0$ as m and $n \to \infty$. Thus the sequence of Y_{ni} is mean-convergent (or mean-fundamental). Hence there exists an integrable function Y_i on $\mathscr{W}_i$ to which the Y_{ni} converge in mean as $n \to \infty$, so that mean $(Y_i) = \lim_{n\to\infty}\{\text{mean}\,(Y_{ni})\}$.† Also, if the Y_{ni} converge in mean to Y_i, they converge to Y_i in measure, that is, in probability.‡ This completes the Lemma.

LEMMA 2. *Over all random walks Γ_i whose initial point is index value i,* var (Y_i) *exists and is finite, and satisfies the equation*

$$\text{var}\,(Y_i) = \sum_{j=1}^{N} (H_{ij}^2/p_{ij})\{\text{var}\,(Y_j) + x_j^2\} - \left(\sum_{j=1}^{N} H_{ij}x_j\right)^2. \tag{43}$$

Proof. With $\Gamma_i \equiv \{i = i_0, j = i_1, i_2, i_3, \ldots\}$ write $\Gamma_j' \equiv \{j = i_1, i_2, i_3, \ldots\}$ for the random walk starting with the second step of Γ_i. Then $\mathscr{P}(\Gamma_{ni}) = p_{ij}\mathscr{P}(\Gamma'_{n-1,j})$, and

$$Y_{ni} = a_i + (H_{ij}/p_{ij})\, Y_{n-1,j}. \tag{44}$$

Now, just like the Y_{ni}, the Y_{ni}^2 are integrable simple functions over $\mathscr{W}_i$, and by the mean-value theorem,§ we see that

$$\text{mean}\,(|Y_{mi}^2 - Y_{ni}^2|) < \text{mean}\,(|Y_{mi} - Y_{ni}|) \max_{\Gamma_i} (|Y_{mi} + Y_{ni}|) \to 0$$

as m and $n \to \infty$. Thus, arguing as in the proof of Lemma 1, we see that the Y_{ni}^2 converge in mean and in measure to an integrable function in $\mathscr{W}_i$, which must therefore equal Y_i^2, for almost every sequence in $\mathscr{W}_i$. It follows that var (Y_i) exists and is finite. Also,

$$\begin{aligned}\text{var}\,(Y_i) &= \text{mean}\,(Y_i^2) - x_i^2 \\ &= \lim_{n\to\infty}\{\text{mean}\,(Y_{ni}^2)\} - x_i^2 \\ &= \lim_{n\to\infty}\{a_i^2 + 2a_i\,\text{mean}\,[(H_{ij}/p_{ij})\, Y_{n-1,j}] + \text{mean}\,[(H_{ij}/p_{ij})^2\, Y_{n-1,j}^2]\} - x_i^2 \\ &= \lim_{n\to\infty}\left\{a_i^2 + 2a_i \sum_{j=1}^{N} H_{ij}\,\text{mean}\,(Y_{n-1,j}) + \sum_{j=1}^{N} (H_{ij}^2/p_{ij})\,\text{mean}\,(Y_{n-1,j}^2)\right\} - x_i^2 \\ &= a_i^2 + 2a_i \sum_{j=1}^{N} H_{ij}x_j + \sum_{j=1}^{N} (H_{ij}^2/p_{ij})\,\text{mean}\,(Y_j^2) - x_i^2.\end{aligned}$$

The Lemma now follows, by (34).

* See Halmos(5), p. 95.
† See Halmos(5), p. 107, Theorem B.
‡ See Halmos(5), p. 103, Theorem A.
§ See Halmos(5), p. 114.

https://doi.org/10.1017/S0305004100036227 Published online by Cambridge University Press

Now, we see that, if

$$p_{ij} = H_{ij}x_j \Big/ \sum_{k=1}^{N} H_{ik}x_k = H_{ij}x_j/(x_i - a_i), \tag{45}$$

then equation (43) becomes homogeneous in the $\mathrm{var}\,(Y_i)$. If we define a matrix $\mathbf{K}$ with elements

$$K_{ij} = \left(\sum_{k=1}^{N} H_{ik}x_k\right) H_{ij}/x_j = (x_i - a_i)\,H_{ij}/x_j, \tag{46}$$

we obtain

THEOREM 2. *If we adopt the probability distribution* (45) *and if no eigenvalue of the matrix* $\mathbf{K}$ *equals one, then all the* $\mathrm{var}\,(Y_i)$ *vanish; that is,* $Y_i = x_i$ *for almost all random walks* Γ_i. *If, further, for every* i,

$$|1 - (a_i/x_i)| < 1, \tag{47}$$

then all the Y_i *are identically equal to* x_i, *for every walk* Γ_i. *This will always be true if all the elements* a_i *and* H_{ij} *are non-negative.*

Proof. If (45) holds, then (43) takes the form $\mathbf{z} = \mathbf{Kz}$, so that $\mathbf{z} = \mathbf{0}$, or $\mathrm{var}\,(Y_i) = 0$ $(i = 1, 2, \ldots, N)$ unless unity is an eigenvalue of $\mathbf{K}$. Thus $(Y_i - x_i)^2$ vanishes for almost every random walk.*

Now define $\zeta_i = 1 - (a_i/x_i)$. Then, by (40), (41), and (45),

$$Y_{ni} = \sum_{m=0}^{n} \left\{\prod_{r=1}^{m} (x_{i_{r-1}} - a_{i_{r-1}})/x_{i_r}\right\} a_{i_m} = x_i \sum_{m=0}^{n} \left(\prod_{r=1}^{m} \zeta_{i_{r-1}}\right)(1 - \zeta_{i_m})$$

$$= x_i\left(1 - \prod_{m=0}^{n} \zeta_{i_m}\right).$$

If (47) holds, that is, $|\zeta_i| < 1$, then, in the limit, $Y_i = x_i$, whatever walk Γ_i is chosen. If the a_i and H_{ij} are all non-negative, so are all the x_i, by (35); hence the ζ_i are also non-negative, by (34), giving $0 \leqslant \zeta_i \leqslant 1$, and completing the Theorem.

6. *The first sequential method.* It is shown in Theorem 2 that the probability distribution (45) leads to a zero variance. It will be seen that this is equivalent to unbiased importance sampling for each of the terms in the series (35). Let the secondary estimate for x_i in stage s be $Z_i^{(s)}$. Then we choose a probability distribution,†

$$p_{i0}^{(s)} = \varpi^{(s)}, \quad p_{ij}^{(s)} = (1 - \varpi^{(s)})\,H_{ij}Z_j^{(s-1)} \Big/ \sum_{k=1}^{N} H_{ik}Z_k^{(s-1)}, \tag{48}$$

where the $\varpi^{(s)}$ is inserted to ensure that the walk terminates.

LEMMA 3. *For any probability distribution in which all* $p_{i0} = \varpi$, *if a random walk* Γ *terminates at the* L*-th step, then*

$$\mathrm{mean}\,(L) = 1/\varpi. \tag{49}$$

Proof.

$$\mathrm{Mean}\,(L) = \sum_{L=1}^{\infty}{}' \mathscr{P}(L)\,L = \sum_{L=1}^{\infty} (1 - \varpi)^{L-1}\varpi L = -\varpi \frac{\partial}{\partial \varpi}\left\{\sum_{L=0}^{\infty}{}' (1 - \varpi)^L\right\}$$

$$= -\varpi \frac{\partial}{\partial \varpi}(\varpi^{-1})$$

and the Lemma follows.

* See Halmos (5), p. 104, Theorem B.

† Note that this presumes that all the $H_{ij}Z_j^{(s-1)}$ are of the same sign, for each i.

https://doi.org/10.1017/S0305004100036227 Published online by Cambridge University Press

We shall suppose that the secondary estimates $Z_i^{(s)}$ are obtained by the averaging process defined in (19), (20), and (21). Each stage uses independent walks, so that the $Y_i^{(s)}$ are certainly uncorrelated. Thus

$$\operatorname{var}(Z_i^{(s)}) = v_s^2 \operatorname{var}(Y_i^{(s)}) + (1-v_s)^2 \operatorname{var}(Z_i^{(s-1)}) = \sum_{r=1}^{s} w_r^{(s)} \operatorname{var}(Y_i^{(r)}). \tag{50}$$

Hence, writing $$U^{(s)} = \max_i \{\operatorname{var}(Y_i^{(s)})\}, \quad V^{(s)} = \max_i \{\operatorname{var}(Z_i^{(s)})\}, \tag{51}$$

we obtain

THEOREM 3. *If the errors in $Z_i^{(s-1)}$ are small quantities, so that their higher powers may be neglected, and if*

$$\varpi^{(s)} + \max_i \left(\sum_{j=1}^{N} K_{ij} \right) < 1, \tag{52}$$

then

$$U^{(s)} \leqslant E_1^{(s)} V^{(s-1)} + E_2^{(s)}, \tag{53}$$

where

$$E_1^{(s)} = 2 \max_i \left(\sum_{j=1}^{N} K_{ij} \right) \Big/ \left\{ 1 - \varpi^{(s)} - \max_i \left(\sum_{j=1}^{N} K_{ij} \right) \right\} \tag{54}$$

and

$$E_2^{(s)} = \varpi^{(s)} \max_i \{(x_i - a_i)^2\} \Big/ \left\{ 1 - \varpi^{(s)} - \max_i \left(\sum_{j=1}^{N} K_{ij} \right) \right\}. \tag{55}$$

Proof. Substituting (48) in the equation (43), and writing $Z_j^{(s-1)} = x_j + \epsilon_j^{(s-1)}$, we get

$$\operatorname{var}(Y_i^{(s)}) = (1-\varpi^{(s)})^{-1} \left\{ \sum_{k=1}^{N} H_{ik}(x_k + \epsilon_k^{(s-1)}) \right\} \left\{ \sum_{j=1}^{N} H_{ij}(x_j + \epsilon_j^{(s-1)})^{-1} \operatorname{var}(Y_j^{(s)}) \right\}$$

$$+ (1-\varpi^{(s)})^{-1} \left\{ \sum_{k=1}^{N} H_{ik}(x_k + \epsilon_k^{(s-1)}) \right\} \left\{ \sum_{j=1}^{N} H_{ij}(x_j + \epsilon_j^{(s-1)})^{-1} x_j^2 \right\} - \left(\sum_{j=1}^{N} H_{ij} x_j \right)^2.$$

Now

$$\operatorname{mean}(\epsilon_j^{(s-1)}) = 0,$$

$$\operatorname{mean}(\epsilon_j^{(s-1)2}) = \operatorname{var}(Z_j^{(s-1)}),$$

and

$$\operatorname{mean}(\epsilon_j^{(s-1)} \epsilon_k^{(s-1)}) = \operatorname{cov}(Z_j^{(s-1)}, Z_k^{(s-1)}).$$

Thus, averaging over the variation of $Z_j^{(s-1)}$ and ignoring higher powers of the $\epsilon_j^{(s-1)}$, we get that

$$(1-\varpi^{(s)}) \operatorname{var}(Y_i^{(s)}) = \left(\sum_{k=1}^{N} H_{ik} x_k \right) \left\{ \sum_{j=1}^{N} H_{ij} x_j^{-1} \operatorname{var}(Y_j^{(s)}) \right\} + \varpi^{(s)} \left(\sum_{j=1}^{N} H_{ij} x_j \right)^2$$

$$+ \sum_{j=1}^{N} \sum_{k=1}^{N} H_{ij} H_{ik} \{ x_j^{-1} x_k \operatorname{var}(Z_j^{(s-1)}) - \operatorname{cov}(Z_j^{(s-1)}, Z_k^{(s-1)}) \}.$$

If we are to use the p_{ij} of (48) as probabilities, we must suppose that, for any i, all the $H_{ij} x_j$ are of the same sign. Also

$$\begin{aligned} 0 &\leqslant \operatorname{mean} \{ (x_j^{-1} Z_j^{(s-1)} - x_k^{-1} Z_k^{(s-1)})^2 \} \\ &\leqslant 2 \operatorname{mean} \{ (x_j^{-1} Z_j^{(s-1)} - 1)^2 + (x_k^{-1} Z_k^{(s-1)} - 1)^2 \} \\ &= 2 \operatorname{var}(x_j^{-1} Z_j^{(s-1)}) + 2 \operatorname{var}(x_k^{-1} Z_k^{(s-1)}), \end{aligned}$$

so that the last sum above is positive, being, by (46),

$$\frac{1}{2} \sum_{j=1}^{N} \sum_{k=1}^{N} H_{ij} x_j H_{ik} x_k \operatorname{mean} \{ (x_j^{-1} Z_j^{(s-1)} - x_k^{-1} Z_k^{(s-1)})^2 \}$$

$$\leqslant 2 \left(\sum_{k=1}^{N} H_{ik} x_k \right) \left\{ \sum_{j=1}^{N} H_{ij} x_j^{-1} \operatorname{var}(Z_j^{(s-1)}) \right\} = 2 \sum_{j=1}^{N} K_{ij} \operatorname{var}(Z_j^{(s-1)}).$$

https://doi.org/10.1017/S0305004100036227 Published online by Cambridge University Press

Thus $$(1-\varpi^{(s)})\operatorname{var}(Y_i^{(s)}) \leqslant \sum_{j=1}^{N} K_{ij}\{\operatorname{var}(Y_j^{(s)}) + 2\operatorname{var}(Z_j^{(s-1)})\} + \varpi^{(s)}(x_i-a_i)^2$$
$$\leqslant \max_i\left(\sum_{j=1}^{N} K_{ij}\right)(U^{(s)} + 2V^{(s-1)}) + \varpi^{(s)}\max_i\{(x_i-a_i)^2\},$$

and, since the right-hand side is independent of i, if (52) holds, the Theorem follows.

It also follows from (50) and (53) that

$$V^{(s)} \leqslant v_s^2 U^{(s)} + (1-v_s)^2 V^{(s-1)} \leqslant \{v_s^2 E_1^{(s)} + (1-v_s)^2\} V^{(s-1)} + v_s^2 E_2^{(s)}, \tag{56}$$

and that
$$U^{(s)} \leqslant E_1^{(s)} \sum_{r=1}^{s-1} w_r^{(s-1)2} U^{(r)} + E_2^{(s)}. \tag{57}$$

LEMMA 4. *If*
$$w_r^{(s)} = r^p \Big/ \sum_{t=1}^{s} t^p, \quad v_s = w_s^{(s)} = s^p \Big/ \sum_{r=1}^{s} r^p, \tag{58}$$

and p is a non-negative integer, then

$$v_s = (p+1)s^{-1} - \tfrac{1}{2}(p+1)^2 s^{-2} + \tfrac{1}{12}(p+1)^2(2p+3)s^{-3} - \ldots, \tag{59}$$

whenever the series converges absolutely; and this will certainly be the case if $s \geqslant p+1$.

Proof. Since we can write, with suitable coefficients c_m,

$$r^p = c_{p+1}\{r^{p+1} - (r-1)^{p+1}\} + c_p\{r^p - (r-1)^p\} + \ldots + c_1\{r - (r-1)\},$$

it is easily verified that

$$\sum_{r=1}^{s} r^p = \sum_{m=1}^{p+1} c_m s^m = (p+1)^{-1}s^{p+1} + \tfrac{1}{2}s^p + \tfrac{1}{12}ps^{p-1} + \ldots = (p+1)^{-1}s^{p+1} + R_p(s) \tag{60}$$

is a polynomial of degree $(p+1)$ in s. Also, since $p \geqslant 0$, x^p is monotonically increasing with x, so that

$$(p+1)^{-1}s^{p+1} = \int_0^s x^p\,dx \leqslant \sum_{r=1}^{s} r^p \leqslant \int_1^s x^p\,dx + s^p = (p+1)^{-1}s^{p+1} + s^p - (p+1)^{-1}.$$

Thus* $$0 \leqslant R_p(s) \leqslant s^p - (p+1)^{-1} < s^p; \tag{61}$$

and, if $$(p+1)^{-1}s^{p+1} > R_p(s), \tag{62}$$

we may write, by (58) and (60),

$$v_s = s^p \Big/ \sum_{r=1}^{s} r^p = (p+1)s^{-1}\{1 + (p+1)s^{-p-1}R_p(s)\}^{-1}$$
$$= (p+1)s^{-1} - (p+1)^2 s^{-p-2}R_p(s) + (p+1)^3 s^{-2p-3}[R_p(s)]^2 - \ldots$$

and this geometric series is absolutely convergent. We may therefore substitute the polynomial $\sum_{m=1}^{p} c_m s^m$ for $R_p(s)$ in each term of the series, yielding the absolutely convergent series (59).† Finally, by (61), the condition (62) shows that (59) will certainly converge absolutely if $(p+1)^{-1}s^{p+1} \geqslant s^p$, completing the Lemma.

* If $p = 0$, $R_p(s) = 0$.

† Conversely, whenever the series (59) is absolutely convergent, it can be rearranged into the geometric series in $R_p(s)$, which must also be absolutely convergent (to the same limit). Thus (62) is a *necessary and sufficient* condition for the absolute convergence of (59).

https://doi.org/10.1017/S0305004100036227 Published online by Cambridge University Press

THEOREM 4. *For the weights* (58), *if* $\varpi^{(s)} = O(s^{-q})$, *and if* $E_1 = \max(E_1^{(s)})$ *and* $E_2 = \max_s(E_2^{(s)} s^q)$; *then there is a constant* A *such that*

$$U^{(s)} \leqslant AE_1 s^{-2(p+1)}\{1+o(1)\} \quad [+E_2 s^{-q}\{1+o(1)\} \text{ if } q < 2(p+1)] \tag{63}$$

and

$$V^{(s)} \leqslant As^{-2(p+1)}\{1+o(1)\}+T, \tag{64}$$

where

$$T = \begin{cases} 0 & \text{if} \quad q \geqslant 2(p+1), \\ E_2(p+1)^2 s^{-q-1}\log s\{1+o(1)\} & \text{if} \quad q = 2p+1, \\ E_2(p+1)^2(2p+1-q)^{-1}s^{-q-1}\{1+O(s^{-1})\} & \text{if} \quad 2(p+1) > q \neq 2p+1. \end{cases}$$

Proof. Consider the equation

$$\phi(s) = E_1 \sum_{r=1}^{s-1} w_r^{(s-1)^2}\phi(r) + E_3 s^{-q}, \tag{65}$$

where $E_3 = E_2\phi(1)/U^{(1)}$; then comparison with (57) shows, inductively, that, for all $s \geqslant 1$,

$$U^{(s)} \leqslant \{U^{(1)}/\phi(1)\}\phi(s) = (E_2/E_3)\phi(s). \tag{66}$$

By (58) and (65),

$$\phi(s+1)-(1-v_s)^2\phi(s) = E_1\left\{v_s^2\phi(s)+\sum_{r=1}^{s-1}[w_r^{(s)^2}-(1-v_s)^2 w_r^{(s-1)^2}]\phi(r)\right\} + E_3\{(1+s^{-1})^{-q}-(1-v_s)^2\}s^{-q},$$

or, since $w_r^{(s)}/w_r^{(s-1)} = 1-v_s$,

$$\phi(s+1) = \phi(s)\{1-v_s[2-(E_1+1)v_s]\}+E_3\{(1+s^{-1})^{-q}-(1-v_s)^2\}s^{-q}. \tag{67}$$

The solution of the corresponding homogeneous equation is clearly proportional to

$$\phi_0(s) = \prod_{r=s_0}^{s-1}\{1-v_r[2-(E_1+1)v_r]\} \quad (s > s_0). \tag{68}$$

Let us suppose that

$$\phi(s) = \phi_0(s)\psi(s);$$

then, by (67) and (68), we get the equation

$$\psi(s+1)-\psi(s) = E_3\{(1+s^{-1})^{-q}-(1-v_s)^2\}s^{-q}/\phi_0(s+1), \tag{69}$$

whose solution is

$$\psi(s) = \psi(s_0)+E_3\sum_{r=s_0}^{s-1}\{[(1+r^{-1})^{-q}-(1-v_r)^2]r^{-q}/\phi_0(r+1)\} \quad (s > s_0). \tag{70}$$

Now, by (58) and (61), $v_s \leqslant (p+1)s^{-1}$; so that, if

$$s_0 > \tfrac{1}{2}(p+1)\{1+\max(E_1,1)\}, \quad 0 < v_r[2-(E_1+1)v_r] < 1,$$

whenever $r \geqslant s_0$. Also, by Lemma 4,

$$v_r[2-(E_1+1)v_r] = 2(p+1)r^{-1}-(E_1+2)(p+1)^2 r^{-2}+O(r^{-3}); \tag{71}$$

so that, by (68),

$$\begin{aligned}\log\phi_0(s) &= \sum_{r=s_0}^{s-1}\log\{1-v_r[2-(E_1+1)v_r]\} \\ &= -\sum_{r=s_0}^{s-1}\sum_{m=1}^{\infty}\{v_r[2-(E_1+1)v_r]\}^m/m \\ &= -\sum_{r=s_0}^{s-1}\sum_{m=1}^{\infty}\{2(p+1)r^{-1}-(E_1+2)(p+1)^2 r^{-2}+O(r^{-3})\}^m/m \\ &= -\sum_{r=s_0}^{s-1}\{2(p+1)r^{-1}-E_1(p+1)^2 r^{-2}+O(r^{-3})\}.\end{aligned}$$

https://doi.org/10.1017/S0305004100036227 Published online by Cambridge University Press

But, as is easily verified,

$$\sum_{r=s_0}^{s-1} r^{-1} = [\log r - \tfrac{1}{2}r^{-1} + O(r^{-2})]_{s_0}^{s}, \quad \sum_{r=s_0}^{s-1} r^{-2} = [-r^{-1} + O(r^{-2})]_{s_0}^{s},$$

so that $$\log \phi_0(s) = [-2(p+1)\log r - \{E_1(p+1)-1\}(p+1)r^{-1} + O(r^{-2})]_{s_0}^{s}$$

giving $$\phi_0(s) = \left(\frac{s}{s_0}\right)^{-2(p+1)} \left\{\frac{1-[E_1(p+1)-1](p+1)s^{-1}+O(s^{-2})}{1-[E_1(p+1)-1](p+1)s_0^{-1}+O(s_0^{-2})}\right\}. \tag{72}$$

Returning now to (70), and using (59), (71), and (72), we obtain

$$\begin{aligned}\phi(s) - \psi(s_0)\phi_0(s) &= \{\psi(s) - \psi(s_0)\}\phi_0(s) \\ &= E_3 s^{-2(p+1)}\{1-[E_1(p+1)-1](p+1)s^{-1}+O(s^{-2})\} \\ &\quad \times \sum_{r=s_0}^{s-1} r^{2p+1-q}\{(2p+2-q) + \tfrac{1}{2}[q^2 - (2E_1(p+1)^2 + (2p+1))q \\ &\qquad + 4E_1(p+1)^3]r^{-1} + O(r^{-2})\}. \end{aligned} \tag{73}$$

We note that, unless $\alpha = -1$,

$$\sum_{r=s_0}^{s-1} r^{\alpha} = [(\alpha+1)^{-1} r^{\alpha+1} - \tfrac{1}{2}r^{\alpha} + O(r^{\alpha-1})]_{s_0}^{s}.$$

Therefore the sum in (73) is

$$\begin{aligned}&[\{r^{2p+2-q} - \tfrac{1}{2}(2p+2-q)r^{2p+1-q}\} \\ &\quad + \tfrac{1}{2}[q^2 - (2E_1(p+1)^2 + (2p+1))q + 4E_1(p+1)^3]\{(2p+1-q)^{-1}r^{2p+1-q}\} + O(r^{2p-q})]_{s_0}^{s};\end{aligned}$$

except that, if $q = 2p+2$, the first expression in curly brackets is omitted; if $q = 2p+1$, the second expression in curly brackets becomes $\log r$; and if $q = 2p$, $O(r^{2p-q})$ becomes $O(\log r)$. By (72), $\psi(s_0)\phi_0(s)$ is a multiple of $s^{-2(p+1)}\{1+O(s^{-1})\}$. Thus, if $q \geqslant 2(p+1)$, we see that there is an A such that

$$\phi(s) = (AE_1E_3/E_2)s^{-2(p+1)}\{1+o(1)\}, \tag{74}$$

and the $o(1)$ strengthens to $O(s^{-1})$ when $q = 2(p+1)$ or $q \geqslant 2p+3$. If $q < 2(p+1)$, however, there will be an A such that

$$\phi(s) = (AE_1E_3/E_2)s^{-2(p+1)}\{1+O(s^{-1})\} + E_3 s^{-q}\{1+g(s)\}, \tag{75}$$

where $$g(s) = \begin{cases} E_1(p+1)^2 s^{-1}\log s + O(s^{-1}) & \text{if} \quad q = 2p+1, \\ E_1(p+1)^2(2p+1-q)^{-1}s^{-1} + O(s^{-2}) & \text{if} \quad q \neq 2p+1. \end{cases} \tag{76}$$

Hence, by (66), we obtain (63).* Now, finally, by (50), (51), (65), and (66),

$$V^{(s)} \leqslant \sum_{r=1}^{s} w_r^{(s)2} U^{(r)} \leqslant (E_2/E_3)\sum_{r=1}^{s} w_r^{(s)2}\phi(r) = (E_2/E_1E_3)\{\phi(s+1) - E_3 s^{-q}\}. \tag{77}$$

Thus, by (74), (75), and (76), we obtain (64),* completing the theorem.

* (74), (75), and (76) hold for all $s > s_0$. Any discrepancy for $s \leqslant s_0$ is bounded, and so can be included in the $o(1)$ term.

https://doi.org/10.1017/S0305004100036227 Published online by Cambridge University Press

We can now derive

THEOREM 5. *For the weights* (58), *if* $\varpi^{(s)} = O(s^{-q})$ *and* $q > 0$, *then the process is convergent, and if* $q > 1$, *the scores* $Z_i^{(s)}$ *converge to* x_i *with probability one.*

Proof. By Theorem 4, if $q > 0$, (63) shows that $\text{var}(Y_i^{(s)}) \to 0$ as $s \to \infty$, so that the process is convergent, by the definition (18), since the $Y_i^{(s)}$ are the primary estimates. Further, by (58) and (61), $v_s \leqslant (p+1)s^{-1}$, so that $u_s = \max_r (w_r^{(s)}) = v_s \to 0$ as $s \to \infty$ and $\sum_{s=1}^{\infty} v_s s^{-t}$ converges for any $t > 0$; while, by (63), if

$$r = \min(q, 2p+2), \quad \text{var}(Y_i^{(s)}) = O(s^{-r}).$$

Thus, if $q > 1$, $r > 1$, and we may take any t satisfying $0 < t < \frac{1}{2}(r-1)$, so that all the conditions of Theorem 1 are satisfied, and it follows that the secondary estimates $Z_i^{(s)}$ converge to x_i with probability one, completing the Theorem.

Suppose that we chose to work with relative errors, writing

$$U_{\mathscr{R}}^{(s)} = \max_i \{x_i^{-2} \text{var}(Y_i^{(s)})\}, \quad V_{\mathscr{R}}^{(s)} = \max_i \{x_i^{-2} \text{var}(Z_i^{(s)})\}, \tag{78}$$

instead of (51). Then Theorem 3 would become, by a parallel proof,*

THEOREM $3_{\mathscr{R}}$. *If the errors in* $Z_i^{(s-1)}$ *are small quantities, so that their higher powers may be neglected, and if*

$$\rho^2 = \max_i \{[1-(a_i/x_i)]^2\} < 1 - \varpi^{(s)}, \tag{79}$$

then

$$U_{\mathscr{R}}^{(s)} \leqslant F_1^{(s)} V_{\mathscr{R}}^{(s-1)} + F_2^{(s)}, \tag{80}$$

where

$$F_1^{(s)} = 2\rho^2/(1-\varpi^{(s)}-\rho^2), \quad F_2^{(s)} = \varpi^{(s)}\rho^2/(1-\varpi^{(s)}-\rho^2). \tag{81}$$

(50) and (80) then yield results (56) and (57) with E replaced by F, and we may derive

THEOREM $4_{\mathscr{R}}$. *The statement of Theorem* 4 *holds with* E *replaced by* F *throughout.*

7. *The second sequential method.* Just as the first sequential method is an application of unbiased importance sampling, the second uses biased importance sampling. The probability distributions $p_{ij}^{(s)}$ for the random walks are now arbitrary, and may be chosen for simplicity and convenience. The scores, however, are obtained, in the sth stage, by putting

$$p_{ij} = H_{ij} Z_j^{(s-1)} \Big/ \sum_{k=1}^{N} H_{ik} Z_k^{(s-1)} \tag{82}$$

in (39) and (40); so that, by (41),

$$Y_{ni}^{(s)} = \sum_{m=0}^{n} \left\{ \prod_{r=1}^{m} \left(\sum_{k=1}^{N} H_{i_{r-1}k} Z_k^{(s-1)} / Z_{i_r}^{(s-1)} \right) \right\} a_{i_m}, \tag{83}$$

where, if $Z_j^{(s-1)} = 0$, we must ensure that $p_{ij}^{(s)}$ (the probability, that is) vanishes for all i. Note that, since the p_{ij} used for the scores do not terminate the random walks, no $\varpi^{(s)}$ need be used, and a nearer approach to the ideal form (45) may be achieved.

* Note that $\mathbf{K}$ is replaced by the matrix $\mathbf{K}_{\mathscr{R}} = \Delta^{-1}\mathbf{K}\Delta$, where Δ is the diagonal matrix with elements $\Delta_{ij} = x_i^2 \delta_{ij}$, so that the eigenvalues of $\mathbf{K}_{\mathscr{R}}$ are the same as those of $\mathbf{K}$.

https://doi.org/10.1017/S0305004100036227 Published online by Cambridge University Press

The scores (83) are equivalent to the unbiased Neumann–Ulam scores for solving an equation

$$\mathbf{x}^{(s)} = \mathbf{a} + \mathbf{H}^{(s)}\mathbf{x}^{(s)}, \tag{84}$$

of the form (34), where

$$H_{ij}^{(s)} = p_{ij}^{(s)} \sum_{k=1}^{N} H_{ik} Z_k^{(s-1)}/Z_j^{(s-1)}. \tag{85}$$

Thus, by (34) and (84),

$$\text{bias}\,(Y_i^{(s)}) = x_i^{(s)} - x_i = \sum_{j=1}^{N} H_{ij}^{(s)}\,\text{bias}\,(Y_j^{(s)}) + \sum_{j=1}^{N} (H_{ij}^{(s)} - H_{ij})\,x_j, \tag{86}$$

and, by (43),

$$\text{var}\,(Y_i^{(s)}) = \sum_{j=1}^{N} (H_{ij}^{(s)2}/p_{ij}^{(s)})\{\text{var}\,(Y_j^{(s)}) + x_j^{(s)2}\} - \left(\sum_{j=1}^{N} H_{ij}^{(s)} x_j^{(s)}\right)^2. \tag{87}$$

LEMMA 5. *If* $\mathbf{Z}^{(s-1)} = \mathbf{x}$ *and no eigenvalue of* $\mathbf{H}^{(s)}$ *equals one, then* $\mathbf{x}^{(s)} = \mathbf{x}$, *or* $\text{bias}\,(Y_i^{(s)}) = 0$ *for all* i, *and* $\text{var}\,(Y_i^{(s)}) = 0$ *for all* i. *If, further,* (47) *holds for every* i, *then* $Y_i^{(s)} = x_i$ *for all* i *and every random walk.*

Proof. If $Z_i^{(s-1)} = x_i$ for all i, then, by (85), $H_{ij}^{(s)} = p_{ij}^{(s)}(x_i - a_i)/x_j$, and, by (86),* $\sum_{j=1}^{N} (H_{ij}^{(s)} - H_{ij})\,x_j = 0$, so that, if no eigenvalue of $\mathbf{H}^{(s)}$ equals one, $\mathbf{x}^{(s)} - \mathbf{x} = \mathbf{0}$ is the only solution, or $\text{bias}\,(Y_i^{(s)}) = 0$ for all i. Again, by (87),* since $\mathbf{x}^{(s)} = \mathbf{x}$,

$$\sum_{j=1}^{N} H_{ij}^{(s)2} x_j^2/p_{ij}^{(s)} = (x_i - a_i)^2 = \left(\sum_{j=1}^{N} H_{ij}^{(s)} x_j\right)^2,$$

so that $\text{var}\,(Y_i^{(s)}) = 0$ is the only solution of (89) under the conditions of the Lemma. Finally, by (83), $Y_{ni}^{(s)}$ takes the same form as in Theorem 2, so that, as in the proof of that theorem, if (47) holds, $Y_{ni}^{(s)} \to x_i$ as $n \to \infty$ for any random walk. This completes the Lemma.

Let us write

$$B_{\mathscr{R}}^{(s)} = \max_i |x_i^{-1}\,\text{bias}\,(Y_i^{(s)})|, \quad D_{\mathscr{R}}^{(s)} = \max_i |x_1^{-1}\,\text{bias}\,(Z_i^{(s)})|; \tag{88}$$

then we may derive the equivalent of Theorem 3,

THEOREM $6_{\mathscr{R}}$. *If the errors in* $Z_i^{(s-1)}$ *are small quantities, so that their higher powers may be neglected, and if* (47) *holds, then*

$$B_{\mathscr{R}}^{(s)} \leqslant C^{(s)} D^{(s-1)}, \tag{89}$$

where

$$C^{(s)} = \max_i \left\{ \sum_{j=1}^{N} \left| x_i^{-1} H_{ij} x_j - \left(1 - \frac{a_i}{x_i}\right) p_{ij}^{(s)} \right| \right\} \Big/ (1-\rho), \tag{90}$$

and

$$U_{\mathscr{R}}^{(s)} \leqslant F_1' V_{\mathscr{R}}^{(s)} + F_2'^{(s)} D_{\mathscr{R}}^{(s-1)2}, \tag{91}$$

where

$$F_1' = 2\rho^2/(1-\rho^2), \quad F_2'^{(s)} = (C^{(s)} + 1)^2 \rho^2/(1-\rho^2). \tag{92}$$

Proof. Proceeding as the proof of Theorem 3, we write $Z_j^{(s-1)} = x_j + \epsilon_j^{(s-1)}$, so that now mean $(\epsilon_j^{(s-1)}) = \text{bias}\,(Z_j^{(s-1)})$. Thus (85) and (86) yield, on averaging over the variation of $Z_j^{(s-1)}$ and ignoring higher powers of $\epsilon_j^{(s-1)}$, that

$$\text{bias}\,(Y_i^{(s)}) = (x_i - a_i) \sum_{j=1}^{N} p_{ij}^{(s)} x_j^{-1}\,\text{bias}\,(Y_j^{(s)}) + \sum_{j=1}^{N} \{H_{ij} - (x_i - a_i)\,p_{ij}^{(s)} x_j^{-1}\}\,\text{bias}\,(Z_j^{(s-1)}).$$

* Noting (34) and that $\sum_{j=0}^{N} p_{ij}^{(s)} = 1$, for all i.

https://doi.org/10.1017/S0305004100036227 Published online by Cambridge University Press

Thus, by (79) and (88), if (47) holds, we obtain (89) and (90). Applying the same procedure to (87), we obtain that

$$\begin{aligned}\operatorname{var}(Y_i^{(s)}) = (x_i-a_i)^2 \sum_{j=1}^{N} p_{ij}^{(s)} x_j^{-2} \operatorname{var}(Y_j^{(s)}) + (x_i-a_i)^2 \Bigg(\sum_{j=1}^{N} p_{ij}^{(s)} x_j^{-2} \{\operatorname{var}(Z_j^{(s-1)}) \\ + [\operatorname{bias}(Y_j^{(s)}) - \operatorname{bias}(Z_j^{(s-1)})]^2\} - \sum_{j=1}^{N}\sum_{k=1}^{N} p_{ij}^{(s)} p_{ik}^{(s)} x_j^{-1} x_k^{-1} \{\operatorname{cov}(Z_j^{(s-1)}, Z_k^{(s-1)}) \\ + [\operatorname{bias}(Y_j^{(s)}) - \operatorname{bias}(Z_j^{(s-1)})][\operatorname{bias}(Y_k^{(s)}) - \operatorname{bias}(Z_k^{(s-1)})]\} \Bigg).\end{aligned}$$

As in the proof of Theorem 3, we can show that the expression in large parentheses above is non-negative and not greater than twice the first of its two sums. Hence, by (78), using (89), we get (91) and (92), completing the Theorem.

We have used the relative form here, as it seems more convenient. An absolute form of Theorem 6, in $U^{(s)}$ and $V^{(s)}$, could easily be obtained. Note that F_1' is not dependent on s.

By (19), (20), (21), (50), (89), and (91), we obtain that

$$D_{\mathscr{R}}^{(s)} \leqslant (v_s C^{(s)} + 1 - v_s) D_{\mathscr{R}}^{(s-1)}, \quad B_{\mathscr{R}}^{(s)} \leqslant C^{(s)} \sum_{r=1}^{s-1} w_r^{(s-1)} B_{\mathscr{R}}^{(r)}; \tag{93}$$

and that

$$V_{\mathscr{R}}^{(s)} \leqslant \{v_s^2 F_1' + (1-v_s)^2\} V_{\mathscr{R}}^{(s-1)} + v_s^2 F_2'^{(s)} D_{\mathscr{R}}^{(s-1)^2}, \tag{94}$$

and

$$U_{\mathscr{R}}^{(s)} \leqslant F_1' \sum_{r=1}^{s-1} w_r^{(s-1)^2} U_{\mathscr{R}}^{(s)} + F_2'^{(s)} D_{\mathscr{R}}^{(s-1)^2}, \tag{95}$$

corresponding to (56) and (57).

We may now proceed to the equivalent of Theorem 4,

THEOREM $7_{\mathscr{R}}$. *For the weights* (58), *if*

$$C = \max_s (C^{(s)}) \quad \textit{and} \quad F_2' = \max_s (F_2'^{(s)}) = (C+1)^2 \rho^2/(1-\rho^2),$$

then there are constants A_1 *and* A_2, *such that*

$$B_{\mathscr{R}}^{(s)} \leqslant A_1 C s^{(C-1)(p+1)} \{1 + O(s^{-1})\}, \quad D_{\mathscr{R}}^{(s)} \leqslant A_1 s^{(C-1)(p+1)} \{1 + O(s^{-1})\}, \tag{96}$$

$$U_{\mathscr{R}}^{(s)} \leqslant A_2 F_1' s^{-2(p+1)} \{1 + o(1)\} + F_2' A_1^2 s^{2(C-1)(p+1)} \{1 + o(1)\}, \tag{97}$$

and

$$V_{\mathscr{R}}^{(s)} \leqslant A_2 s^{-2(p+1)} \{1 + o(1)\} + T_{\mathscr{R}}, \tag{98}$$

where

$$T_{\mathscr{R}} = \begin{cases} F_2' A_1^2 (p+1)^2 s^{2(C-1)(p+1)-1} \log s \{1 + o(1)\} & \textit{if} \quad C = \tfrac{1}{2}(p+1)^{-1}, \\ F_2' A_1^2 (p+1)^2 [2C(p+1)-1]^{-1} s^{2(C-1)(p+1)-1} \{1 + O(s^{-1})\} & \textit{otherwise.} \end{cases}$$

Proof. Consider the equation (compare (65))

$$\chi(s) = C \sum_{r=1}^{s-1} w_r^{(s-1)} \chi(r), \tag{99}$$

https://doi.org/10.1017/S0305004100036227 Published online by Cambridge University Press

which, by comparison with (93), shows, inductively, that for all $s \geqslant 1$, if $\chi(1) = B_{\mathscr{R}}^{(1)}$, $B_{\mathscr{R}}^{(s)} \leqslant \chi(s)$. By (99) (compare (67)) $\chi(s+1) - (1-v_s)\chi(s) = Cv_s\chi(s)$. The solution is thus

$$\begin{aligned} B_{\mathscr{R}}^{(1)} \prod_{r=s_0}^{s-1} \{1 - v_r(1-C)\} &= B_{\mathscr{R}}^{(1)} \exp\left(-\sum_{r=s_0}^{s-1}\sum_{m=1}^{\infty} [v_r(1-C)]^m/m\right) \\ &= B_{\mathscr{R}}^{(1)} \exp\left(\sum_{r=s_0}^{s-1} \{(p+1)(C-1)r^{-1} - \tfrac{1}{2}(p+1)^2 C(C-1)r^{-2} + O(r^{-3})\}\right) \\ &= B_{\mathscr{R}}^{(1)} \exp\left([(p+1)(C-1)\log r + \tfrac{1}{2}(p+1)(C-1)\{C(p+1)-1\}r^{-1} + O(r^{-2})]_{s_0}^{s}\right), \end{aligned}$$

by Lemma 4, whence the first inequality of (96) follows. Now, by (20) and (99),

$$D_{\mathscr{R}}^{(s)} \leqslant \sum_{r=1}^{s-1} w_r^{(s-1)} B_{\mathscr{R}}^{(r)} \leqslant \sum_{r=1}^{s-1} w_r^{(s-1)} \chi(r) = \chi(s)/C,$$

whence we get the second inequality of (96). Again, if we compare (94) and (95) with (56) and (57), we see that the results (97) and (98) should be the same as (63) and (64), with F_1' for E_1, $F_2'A_1^2$ for E_2, and $2(1-C)(p+1)$ for q. Since, by (47) and (90), $C \geqslant 0$, $q < 2(p+1)$, so that the additional terms in (63) and (64) must always be included. This completes the Theorem.

The equivalent of Theorem 5, and its consequence, is

THEOREM 8. *For the weights* (58), *if* $C < 1$, *the process is convergent, and if*

$$C < 1 - \tfrac{1}{2}(p+1)^{-1},$$

the scores $Z_i^{(s)}$ *converge to* x_i *with probability one.*

8. *The third sequential method.* Unlike the first and second sequential methods, the third does not depend on Theorem 2 (or its equivalent, Lemma 5). It uses instead the result embodied in (36) and (37), to give the following procedure, which may be looked upon as a form of sequential correlated sampling. The probability distribution $p_{ij}^{(s)}$ is arbitrary, and the score used is

$$Y_{ni}^{(s)} = \sum_{m=0}^{n} \left\{ \prod_{r=1}^{m} (H_{i_{r-1}i_r}/p_{i_{r-1}i_r}^{(s)}) \right\} d_{i_m}^{(s)}, \tag{100}$$

where

$$\mathbf{d}^{(s)} = \mathbf{a} + \mathbf{H}\mathbf{Z}^{(s-1)} - \mathbf{Z}^{(s-1)}, \tag{101}$$

and

$$\mathbf{Z}^{(s)} = \mathbf{Z}^{(s-1)} + \mathbf{Y}^{(s)} = \sum_{r=1}^{s} \mathbf{Y}^{(r)}. \tag{102}$$

By (36), (37), and (42),

$$\text{mean}\,(Y_i^{(s)}) = x_i - Z_i^{(s-1)}, \quad \text{mean}\,(Z_i^{(s)}) = x_i. \tag{103}$$

Thus the $Y_i^{(s)}$ are, at each stage, unbiased estimators of the correction $(x_i - Z_i^{(s-1)})$, and $Z_i^{(s)}$ is an unbiased estimator of x_i.

THEOREM 9. *If*

$$\sigma^{(s)} = \max_i \left\{ \sum_{j=1}^{N} (H_{ij}^2/p_{ij}^{(s)}) \right\} < 1, \tag{104}$$

then

$$V^{(s)} \leqslant G^{(s)} V^{(s-1)}, \tag{105}$$

where

$$G^{(s)} = \sigma^{(s)}/(1-\sigma^{(s)}). \tag{106}$$

https://doi.org/10.1017/S0305004100036227 Published online by Cambridge University Press

Proof. By Lemma 2,

$$\operatorname{var}(Y_i^{(s)}) = \sum_{j=1}^{N} (H_{ij}^2/p_{ij}^{(s)}) \{\operatorname{var}(Y_j^{(s)}) + (x_j - Z_j^{(s-1)})^2\} - \left\{ \sum_{j=1}^{N} H_{ij}(x_j - Z_j^{(s-1)}) \right\}^2.$$

But $$\operatorname{var}(Y_i^{(s)}) = \operatorname{mean}\{(Y_i^{(s)} - x_i + Z_i^{(s-1)})^2\} = \operatorname{var}(Z_i^{(s)})$$

and $$\operatorname{mean}\{(x_j - Z_j^{(s-1)})^2\} = \operatorname{var}(Z_j^{(s-1)}).$$

Thus, averaging, we get that

$$\operatorname{var}(Y_i^{(s)}) \leqslant \sum_{j=1}^{N} (H_{ij}^2/p_{ij}^{(s)}) \operatorname{var}(Y_j^{(s)}) + \sum_{j=1}^{N} (H_{ij}^2/p_{ij}^{(s)}) \operatorname{var}(Z_j^{(s-1)}),$$

whence the Theorem follows.

LEMMA 6. *If* (104) *holds and* $G = \max_s (G^{(s)})$,

$$V^{(s)} \leqslant G^{s-1} V^{(1)}, \tag{107}$$

and the process will converge (geometrically) if

$$G < 1 \quad \text{or} \quad \sigma = \max_s (\sigma^{(s)}) < \tfrac{1}{2}. \tag{108}$$

Proof. This Lemma is an immediate consequence of Theorem 9.

THEOREM 10. *If* (108) *holds, the secondary estimator*

$$\mathfrak{Z}_i^{(s)} = \frac{1}{s} \sum_{r=1}^{s} Z_i^{(r)} \tag{109}$$

for x_i is unbiased, and converges to x_i with probability one as $s \to \infty$.

Proof. By (103),

$$\operatorname{mean}(\mathfrak{Z}_i^{(s)}) = \frac{1}{s} \sum_{r=1}^{s} \operatorname{mean}(Z_i^{(r)}) = x_i;$$

that is, $\mathfrak{Z}_i^{(s)}$ is an unbiased estimator of x_i. We proceed as in the proof of Theorem 1, with v_s replaced by s^{-1}, f_s by $Z_i^{(s)}$, and θ by x_i. Then, by (107), $\gamma^2 s^{-2l} p(\mathscr{E}_s) \leqslant V^{(1)} G^{s-1}$, so that

$$p(\mathscr{F}_m) > \prod_{s=m+1}^{\infty} (1 - V^{(1)} \gamma^{-2} s^{2l} G^{s-1}) > \exp\left\{ -2 V^{(1)} \gamma^{-2} \sum_{s=m+1}^{\infty} s^{2l} G^{s-1} \right\}$$

$$\geqslant \exp\left\{ -2 V^{(1)} \gamma^{-2} G^{-2} \int_{m+1}^{\infty} s^{2l} G^s \, ds \right\} > 1 - \gamma',$$

for any arbitrary $\gamma' > 0$, if m is sufficiently large. (This follows from (108), for any t.) We may now choose $t > 0$, ensuring that $\sum_{s=1}^{\infty} s^{-t-1}$ converges, and complete the proof that $\sum_{s=1}^{\infty} \frac{1}{s} (Z_i^{(s)} - x_i)$ converges for almost all random sequences. Hence*

$$\frac{1}{s} \sum_{r=1}^{s} (Z_i^{(r)} - x_i) = (\mathfrak{Z}_i^{(s)} - x_i) \to 0 \quad \text{as} \quad s \to \infty$$

for almost all random sequences, proving the Theorem.

* By Halmos (5), p. 203, Theorem C.

https://doi.org/10.1017/S0305004100036227 Published online by Cambridge University Press

Finally, we have

LEMMA 7. *If* (108) *holds,*

$$\operatorname{var}(\mathfrak{Z}_i^{(s)}) < \{V^{(1)}/(1-G)\}\, s^{-2}. \tag{110}$$

Proof. By (107) and (109),

$$\operatorname{var}(\mathfrak{Z}_i^{(s)}) = s^{-2} \sum_{r=1}^{s} \operatorname{var}(Z_i^{(r)}) \leqslant s^{-2} \sum_{r=1}^{s} G^{r-1} V^{(1)} = s^{-2} V^{(1)} (1-G^s)/(1-G) < s^{-2} V^{(1)}/(1-G),$$

proving the Lemma.

9. *Experimental verification.* During 1958–60, the author wrote, tested, and ran a number of programmes for the Illiac and Mercury digital computers, to illustrate and test the sequential methods described above. The empirical quantity $\max_i |Z_i^{(s)} - x_i|$ was plotted against s and compared with the behaviour of the theoretical upper bounds for $\sqrt{V^{(s)}}$ embodied in Theorems 4 and 7, and Lemma 6. The agreement between theory and experiment was excellent throughout, in the sense that the asymptotic curves gave a good representation of the behaviour of the errors even for early stages of the processes, and that the theoretical upper bounds for errors were always well within a factor of two above the observed errors, for several examples tried. Deductions from the experimental results are listed below.

(i) A number of comparative runs with different types of pseudo- and quasi-random number sequences, including runs with an improper sequence having regular ordering,* showed that the third sequential method was relatively insensitive to non-uniformity of the sequence used, and that the only effect was a slight slowing down of the convergence, without the introduction of any bias.

(ii) Comparative runs with the non-sequential Neumann–Ulam method and the first, second, and third sequential methods indicated that the third method was the most rapidly convergent, not only asymptotically, but even for short runs of only a few stages. (In fact, the graphs of the first and second methods did not cross those of the third.) Since all three sequential methods require just about as much labour per stage, under the same conditions, it follows that the third sequential method will generally be best.

(iii) As had been anticipated, it was found that, for any accuracy text based on an internal estimate† of the variance, the ratio of actual to estimated errors decreased as the rapidity of convergence increased.‡

(iv) The principal advantage of the Neumann–Ulam method is that, unlike classical methods for the solution of sets of simultaneous linear equations, it enables

* The sequence used was $\phi_2(n)$ for $n = 1, 2, 3, \ldots$, as defined in (6). The ordering is very marked; e.g. if $\phi_2(n) < \frac{1}{2}$, $\phi_2(n+1) > \frac{1}{2}$.

† By this is meant an estimate independent of any analysis of the problem treated, but obtained from the empirical results.

‡ Since a random difference much smaller than the standard deviation, and so prematurely satisfying the test, becomes increasingly likely as the number of trials increases.

https://doi.org/10.1017/S0305004100036227 Published online by Cambridge University Press

us to estimate one component of the solution without having to do the same work as if we required the complete solution. When N is of the order of thousands or more, this is of great importance. The sequential methods described above relinquish this advantage, to gain a much faster rate of convergence. On balance, it turns out that the efficiency of the third sequential method is of the same order as that of a classical iterative method of solution.* However, we can retain the geometric convergence of the sequential method while reducing the work required if we know something of the shape of the full solution. This is often the case, for example, when the problem treated is the finite-difference form of a differential equation.

Test runs were made on a programme for solving the biharmonic equation for a square plate clamped at its edge and loaded at one eccentric point. The process used was the third sequential method, but scores were calculated only for the centre of the plate; the values of the displacements at all the other 1680 grid-points were estimated to be simply half the current estimate $Z_0^{(s)}$ of the central displacement. The initial estimate $Z_0^{(0)}$ was deliberately chosen to be very far from correct, and a clearly geometric convergence was obtained. In 112 walks, the error was reduced by a factor of 10^{-8}, even using such a crude estimator. The whole process, carried out on the Ferranti Mercury machine, took some 15 min. (By comparison, applying the non-sequential Neumann–Ulam method on the same machine for nearly an hour simply gave the answer zero, as no walk happened to pass through the point of loading!)

The process described above would clearly cease to converge geometrically when the residual error was of the same order as that involved in taking a simple step-function for the displacement. At this stage, a closer approximation could be adopted (e.g. a parabolic form-function, or a multiple step-function, involving the calculation of Monte Carlo scores at several—but still far less than N—points.) Alternatively, one could simply revert to the non-sequential method and rely on arithmetic convergence ($V^{(s)} = O(s^{-1})$) to give a final estimate of sufficient accuracy. The saving of labour, as against a straight application of the third sequential method, is manifest.

It is evident that considerable subtlety may advantageously be used in wringing the greatest accuracy from the least work, along the lines suggested above. We have not attempted to go into this in the present paper, since its aim is to present and establish the sequential Monte Carlo methods, rather than to investigate their practical development.

I wish to thank Dr J. M. Hammersley (Trinity College, Oxford) for helpful discussion of this work, and the U.K. Atomic Energy Authority for a Research Studentship at Oxford University. I am also grateful to the Directors and staffs of the computer departments of the A.E.R.E., Harwell, the University Computing Laboratory, Oxford, and the Digital Computer Laboratory, Graduate College, University of Illinois, Urbana, for their courtesy and assistance.

* In both cases, the convergence is geometric, and the work done per stage consists largely in multiplying a vector of order N by an $N \times N$ matrix.

https://doi.org/10.1017/S0305004100036227 Published online by Cambridge University Press

REFERENCES

(1) CUTKOSKY, R. E. A Monte Carlo method for solving a class of integral equations. *J. Res. Nat. Bur. Standards*, 47 (1951), 113–5.

(2) DE GROOT, M. H. Unbiased sequential estimation for binomial populations. *Ann. Math. Statist.* 30 (1959), 80–101.

(3) FIELLER, E. C. and HARTLEY, H. O. Sampling with control variables. *Biometrika*, 41 (1954), 494–501.

(4) FORSYTHE, G. E. and LEIBLER, R. A. Matrix inversion by a Monte Carlo method. *Math. Tables Aids Comput.* 4 (1950), 127–9.

(5) HALMOS, P. R. *Measure theory* (New York, 1956).

(6) HALTON, J. H. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numer. Math.* 2 (1960), 84–90. (Also *ibid.* p. 196.)

(7) HALTON, J. H. and HANDSCOMB, D. C. A method for increasing the efficiency of Monte Carlo integration. *J. Assoc. Comput. Mach.* 4 (1957), 329–40.

(8) HAMMERSLEY, J. M. Monte Carlo methods for solving multi-variables problem. *Ann. New York Acad. Sci.* 86 (1960), 844–74.

(9) HAMMERSLEY, J. M. and MAULDON, J. G. General principles of antithetic variates. *Proc. Cambridge Philos. Soc.* 52 (1956), 476–81.

(10) HAMMERSLEY, J. M. and MORTON, K. W. A new Monte Carlo technique: antithetic variates. *Proc. Cambridge Philos. Soc.* 52 (1956), 449–75.

(11) HANDSCOMB, D. C. Proof the antithetic variates theorem for $n > 2$. *Proc. Cambridge Philos. Soc.* 54 (1958), 300–1.

(12) KAHN, H. The use of different Monte Carlo sampling techniques. *University of Florida symposium on Monte Carlo methods*, pp. 146–90 (New York, 1956).

(13) KIEFER, J. and WEISS, L. Some properties of generalized sequential probability ratio tests. *Ann. Math. Statist.* 28 (1957), 57–74.

(14) LÉVY, P. *Théorie de l'addition des variables aléatoires* (Paris, 1937).

(15) MARSHALL, A. W. The use of multi-stage sampling schemes in Monte Carlo computation. *University of Florida symposium on Monte Carlo methods*, pp. 123–40 (New York, 1956).

(16) OPLER, A. Monte Carlo matrix calculations with punched card machines. *Math. Tables Aids Comput.* 5 (1951), 115–20.

(17) PAGE, E. S. The Monte Carlo solution of some integral equations. *Proc. Cambridge Philos. Soc.* 50 (1954), 414–25.

(18) STEIN, C. and WALD, A. Sequential confidence intervals for the mean of a normal distribution with known variance. *Ann. Math. Statist.* 18 (1947), 427–33.

(19) WALD, A. *Sequential analysis* (New York, 1950).

(20) WALD, A. and WOLFOWITZ, J. Bayes solutions of sequential decision problems. *Ann. Math. Statist.* 21 (1950), 82–9.

(21) WASOW, W. Random walks and the eigenvalues of elliptic difference equations. *J. Res. Nat. Bur. Standards*, 46 (1951), 65–73.

(22) WASOW, W. A note on the inversion of matrices by random walks. *Math. Tables Aids Comput.* 6 (1952), 78–81.

(23) WOLFOWITZ, J. On sequential binomial estimation. *Ann. Math. Statist.* 17 (1946), 489–93.

(24) WOLFOWITZ, J. The efficiency of sequential estimates and Wald's equation for sequential processes. *Ann. Math. Statist.* 18 (1947), 215–30.

BALLIOL COLLEGE
OXFORD

https://doi.org/10.1017/S0305004100036227 Published online by Cambridge University Press