onsidération de la
.lté des Sciences de

ing. *Decis. Sci.* 13,

ms. *Mgmt Sci.* 30,

. 227–229.
goal programming
and V. CHANKONG,

ods in multicriteria

:ms. *Mgmt Sci.* 22,

i8–494. Wiley, New

J. Opl Res. Soc. Vol. 41, No. 12, pp. 1133–1142, 1990
Printed in Great Britain.

0160-5682/90 $3.00 + 0.10

Descriptive Sampling: A Better Approach to Monte Carlo Simulation

EDUARDO SALIBY*
Graduate Business School, University of Rio de Janeiro, Brazil

Although simple random sampling is the standard sampling procedure in Monte Carlo simulation, such practice is questioned in this paper. In any Monte Carlo application, sampled distributions are assumed to be known. Using simple random sampling, sample histograms or, equivalently, sample moments will vary at random, thus producing an imprecise description of the known input distribution, and consequently increasing the variance of simulation estimates. This problem can be avoided with descriptive sampling, here proposed as a more appropriate approach in Monte Carlo simulation than simple random sampling. Descriptive sampling is based on a deterministic and purposive selection of the sample values—in order to conform as closely as possible to the sampled distribution—and the random permutation of these values. As such, it represents a fundamental conceptual change in Monte Carlo sampling, departing from the 'principle' that sample values must be randomly generated in order to describe random behaviour. The basis of this new idea, examples of its use and empirical results are presented.

Key words: Monte Carlo, sampling, simulation

INTRODUCTION

Up to the present, a basic idea in Monte Carlo simulation has been that input samples should be randomly generated in order to describe a random behaviour. As a consequence, simple random sampling turns out to be the standard sampling procedure in simulation, in spite of the fact that, owing to sampling errors, it produces poor results, i.e. estimates with high variability.

The low precision of simulation estimates was identified in the early days of Monte Carlo practice, leading to the proposal of variance reduction techniques,[1] or simply VRT. Examples of VRT are antithetic variates,[2] stratified sampling,[3] importance sampling,[4] common random numbers (also known as correlated sampling),[5] control variates[6] and, a more recent development, Latin hypercube sampling.[7] Apart from control variates, the above VRTs look for a more controlled sample selection, but without breaking with the Monte Carlo 'principle' or 'paradigm' that sample values should vary at random in order to imitate a random behaviour. With control variates the approach is different: the sampling process remains unchanged, but observed sample deviations are taken into account in order to adjust the estimates. More about variance reduction techniques can be found in James[8] or Law and Kelton.[9]

In our view, this 'imitation paradigm' is false. In fact, the need for random sampling from a population derives from the lack of knowledge about this population, and is intended to avoid sampling bias. However, in any Monte Carlo application, samples are, by assumption, drawn from known distributions. In this case, the purpose of sampling is not to make inferences about the sampled population, but only to describe random variable behaviour. Using simple random sampling for such a purpose leads to an unnecessarily imprecise description of the sampled distribution, thus inflating the variance of simulation estimates.

As a better and more appropriate alternative for Monte Carlo sampling, we are proposing the use of descriptive sampling instead of traditional simple random sampling. The main characteristic of this new approach is that descriptive sampling is based on a deterministic and purposive selection of the input sample values. This sample-values selection aims to achieve the closest fit with the represented distribution, instead of letting the sample histogram vary at random. Only the sample sequence remains random.

Descriptive sampling is easy to implement and, without any relevant increase in processing time, generally produces more precise estimates. The only additional requirement for using descriptive sampling is to know, in advance, the approximate input sample size.

Although descriptive sampling contributes to simulation practice, leading to more precise estimates, this work is mainly conceptual. The basic point of discussion is to show that, contrary to

* This paper was written when the author was at the London School of Economics and Political Science.

common belief, there is no need for a random selection of sample values in a Monte Carlo study, or equivalently, that there is nothing wrong with a deterministic selection of such values. Once this point is accepted, it becomes evident that a deterministic selection of sample values is the most appropriate approach to be followed in any Monte Carlo application, including simulation.

THE PROBLEM

Initially, a simulation problem formulation is presented:

(a) There is a model describing the system behaviour.

(b) This model, as illustrated in Figure 1, transforms a set of input random variables (XA, XB, ...) into a set of output random variables (ZA, ZB, ...), also known as the response variables.

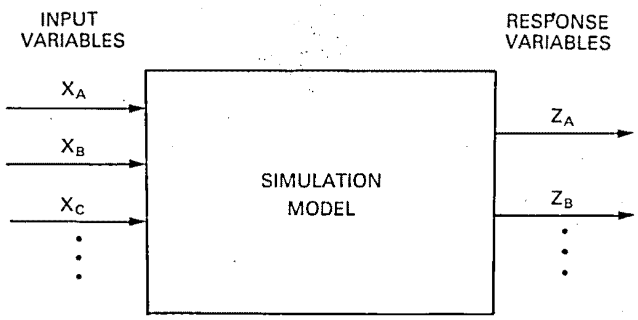

FIG. 1. *Simulation model representation: a set of input random variables is transformed into a set of output or response random variables.*

(c) The distributions of input variables are assumed known, while the response variables' distributions are unknown. For instance, the purpose of the study is to determine the response variables' distributions, or at least their main parameters.

(d) When the problem is simulated, input random variables are replaced by samples. Consequently, response variables are also replaced by samples.

(e) A simulation run, as understood here, is defined as the computing effort that produces one set of values for each response variable, so that each sample originates one and only one estimate for each response parameter under study. Notice that, when a terminating system is simulated, each trial produces one observed value for each response variable. In such cases, a simulation estimate is computed from a set of $N \geqslant 1$ response values, so that a simulation run is defined by N trials. The distinction between a simulation run, which produces one estimate, and a simulation trial, which produces one response value, is irrelevant when using simple random sampling, but it is very important when using descriptive sampling.

Although response parameters are constants, each simulation run produces a different set of estimates. This estimation error, which should be minimized, is the price paid for the Monte Carlo sampling approach.

Two examples clarify the above concepts.

Example 1: the newsboy problem

Daily, a newsboy buys $Q = 80$ issues of a newspaper at \$3.00 each. The selling price is \$8.00. At the end of the day, all remaining issues are lost (there is no residual value). Daily demand is independent and identically distributed according to the following distribution:

Daily demand	Probability
50	0.10
60	0.12
70	0.15
80	0.20
90	0.18
100	0.15
110	0.10

The simulation purpose is to study the daily profit distribution, estimating the mean and standard deviation.

In this very simple problem,

(a) there is one input variable, the daily demand (D);
(b) there is one response variable, the daily profit (P);
(c) it is a terminating system, where each trial is defined by one simulated day;
(d) two estimates are under study:

$$\bar{P} = \sum_{i=1}^{N} P_i/N \quad \text{and} \quad S_P = \left[\sum_{i=1}^{N} (P - \bar{P})^2/(N-1) \right]^{1/2};$$

(e) therefore, each run is defined by N simulated days.

Example 2: M/M/1 queue

This problem concerns the steady-state queue-size (including customers in service) distribution for an M/M/1 queue. Here again, two estimates are under study: the mean and standard deviation of the queue-size distribution.

In this problem,

(a) there are two input variables: interarrival times and service times;
(b) there is one response variable: the queue size;
(c) it is a non-terminating system, with a simulation run defined by a long time-period, T—the concept of simulation trial does not apply here;
(d) both estimates are integrated over time; they are identified as $\overline{\text{SIZE}}$ and S_{SIZE}.

Simulation problem formulation

Assuming, for simplicity, that only one response variable is under study and that only one input random variable drives the simulation, a simulation run can be seen as a set of function evaluations

$$Y_j = F_j(X_1, \ldots, X_n), \qquad j = 1, \ldots, L, \tag{1}$$

where Y_j is the estimator for parameter $\theta_j, j = 1, \ldots, L$, concerning the response variable distribution, and F_j is a deterministic function that relates an input sample X_i, $i = 1, \ldots, n$, to the corresponding estimator Y_j, $j = 1, \ldots, L$.

When the input sample size tends to infinity, two asymptotic properties are required:

(a) unbiassedness, so that $E(Y_j) \simeq \theta_j$, $j = 1, \ldots, L$, and
(b) consistency, so that $\text{Var}(Y_j) \simeq 0$, $j = 1, \ldots, L$.

Notation

In the present paper, N refers to the number of trials that defines one run for a terminating system; alternatively, T is the run-length for a non-terminating system. n refers to the corresponding input sample size for each run. Finally, although there are other methods for estimating the variance of the estimates, the use of replicated runs will be assumed. M is the number of replicated runs in a simulation experiment.

DESCRIPTIVE SAMPLING

From expression (1), it follows that the variability of simulation estimates is only due to the input sample variability. In each run, a different input sample is drawn, thus producing a different set of estimates. As expected, there is a relationship between the input samples and the corresponding estimates. This relationship has been previously studied,[10] and is now summarized.

Two sources of variability: set and sequence

As far as a simulation is concerned, any input sample is determined by two general features:

(a) the set of values, defined by all sample values taken together, but without considering their particular sequence, and

(b) the particular sequence in which those values occur.

Both features are closely related to the two basic probability concepts: the set of values is expected to display a pattern of relative frequencies in good agreement with the sampled distribution; the sequence is expected to display a lack of order of such values or, roughly speaking, a 'pattern of randomness'. When using simple random sampling, both set and sequence are allowed to vary. Once they vary, simulation estimates will also vary. Therefore, both features can be seen as the two sources of variability of simulation estimates.

The estimates' variability

Breaking down an input sample into a set and a sequence, the variability of simulation estimates, when using simple random sampling, was studied.[10,11] The main conclusions were as follows:

(a) The set variability associated with the input sample has a specific and usually important influence on the estimates' variances. The relative contribution of the set variability to the estimate variance is often 50% or more. Another important feature of the set relative contribution is that it was found to be nearly constant, irrespective of run-length. By increasing the run-length, the input sample is closer to the represented distribution, but the relative contribution of the set variability remains the same.

(b) A regression model, named the linear response model (LRM),[10,12] explains the set effect in terms of the observed deviations between input sample moments and the corresponding theoretical values. The linear response model is mathematically equivalent to the method of control variates.[12] As with control variates, the variability explained by the LRM is given by the R2 regression coefficient of determination, computed from a set of M independent replications using simple random sampling. This R2 coefficient provides an estimate for the set effect.

(c) While the set influence appeared to be a common feature in most simulation problems, a sequence effect was only detected in some cases, like the M/M/1 queue simulation.[11] Unlike the set effect that can be explained by the linear response model, no general model relating sequence properties to the estimates was found.

Knowing the role played by set variability in the estimate variance, this information could enhance the use of variance reduction techniques. However, a different approach was followed to avoid this variability.

Understanding set variability

Using simple random sampling, the input values—as well as the corresponding moments—will vary between different simulation runs. As a consequence, a curious paradox arises: although the input distribution is known, estimates are based on samples that differ from this distribution! This means that the model's assumptions are not precisely followed and that the estimates are, in fact, related to different running conditions, depending on the particular set of input values generated. This departure from the known distribution is due to the random selection of the sample values. This random selection inflates the variance of the estimates and causes the set effect.

Questioning the use of simple random sampling in simulation

The use of simple random sampling in Monte Carlo simulation, which leads to set variability, derives from the 'imitation' paradigm. This paradigm says that to describe random behaviour, a sample has to imitate this behaviour, being fully random as well. However, this paradigm is questionable.

According to Stuart,[13] the need for a random selection when sampling is to prevent sample bias; otherwise, owing to a lack of knowledge about the sampled population, there could exist an unsuspected relationship between the selection criterion and the response variable. But this is not the case in Monte Carlo sampling. In this case, once the sampled distributions are assumed known, simple random sampling is not used to prevent sample bias, but only to describe random behaviour. Paradoxically, however, there is no need to use simple random sampling for such a purpose.

In order to describe a random behaviour, the two basic probability concepts
(a) a pattern of frequencies and
(b) a 'pattern of randomness'
are to be represented. The pattern of frequencies is described by the set of values. Once the sampled distribution is assumed known, the sampling cumulative distribution function should be as close as possible to the corresponding probability distribution function. This can be done with a deterministic and purposive selection of the sample values.

The pattern of randomness is a sequence property. The only way to represent randomness is by randomly permutating the sample values. Notice that, using simple random sampling, a random permutation of the randomly generated set values is automatically produced.

Descriptive sampling proposal

Descriptive sampling intends to remove, or at least to minimize, the set variability in Monte Carlo sampling. It is based on a deterministic selection of the input values and their random permutation. Symbolically, it follows that

$$\text{descriptive sampling} = \text{deterministic set} \times \text{random sequence},$$

while

$$\text{simple random sampling} = \text{random set} \times \text{random sequence}.$$

The only additional requirement for using descriptive sampling is to know, in advance, the input sample size. In cases where the input sample size is not determined, an approximated value for the sample size is enough. However, it is important to relate the input sample to a full simulation run, but not to a single trial.

USING DESCRIPTIVE SAMPLING

Although it is not as straightforward as simple random sampling, descriptive sampling is quite easy to program and use. More important, its use can be fully automated, without requiring any extra programming effort by the user, other than specifying the input sample sizes. The generation of a descriptive sample consists of two steps: the generation of a descriptive set of values and the random permutation of those values.

Set values generation

Using descriptive sampling, set values are generated in advance and stored in memory for later use. With a single modification, the general method of the inverse transform can also produce descriptive sampling values, given by

$$xd_i = F^{-1}[(i - 0.5)/n], \qquad i = 1, \ldots, n. \tag{2}$$

TABLE 1. *Descriptive set of values for a negative exponential distribution with mean $E(X) = 1$ and $n = 10$; inverse: $X = -Ln(1 - R)$*

i	$(i - 0.5)/n$	$xd_i = -Ln[1 - (i - 0.5)/n]$
1	0.05	0.051
2	0.15	0.163
3	0.25	0.288
4	0.35	0.431
5	0.45	0.598
6	0.55	0.798
7	0.65	1.050
8	0.75	1.386
9	0.85	1.897
10	0.95	2.996
Mean	0.50	0.966

Illustrating the method, Table 1 presents a set of $n = 10$ descriptive values for a negative exponential distribution with a mean of 1. Although $n = 10$ is a very small sample size for a simulation run, Figure 2 shows that a good agreement with the sampled distribution was achieved.

In cases where the inverse of the distribution function is not analytically available, approximations are required. Also notice that the same library of routines for random sampling generation, if based on the inverse transform, can be used for the generation of descriptive values.

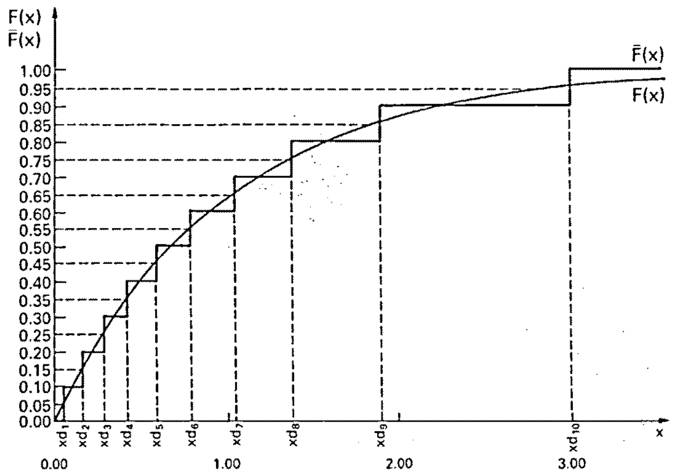

FIG. 2. *Distribution function [$F(x)$] and cumulative distribution function for a descriptive sample of size $n = 10$ [$\bar{F}(x)$], for a negative exponential random variable with mean $E(X) = 1.0$.*

Random permutation

While the set values can be generated only once for all replicated runs in a simulation experiment, each run is based on a different random permutation. The set values can all be shuffled before carrying out each run, but it is more convenient to shuffle them during the run, drawing a set element whenever it is required. In practice, this sequential process is done by sampling the pre-defined set of descriptive values without replacement. A data structure and an algorithm are suggested.

Data structure

For each input random variable, define a record with the following structure and content:

n: integer defining the sample size;

XD: array[1 ... n] of real, containing the set values;

ip: integer pointing to the first available XD element to be drawn. If $ip = 1$, no element has been drawn yet. If $ip = n + 1$, a full set of values has already been drawn.

Algorithm

(a) Initialization. Before running the simulation, define n, generate the set values XD and let $ip := 1$.

(b) Sampling without replacement during the run. Whenever a descriptive sample value is required,

(b1) if $ip > n$, then let $ip := 1$;

(b2) randomly generate an integer iaux $\in [ip, n]$;

(b3) interchange $XD[ip]$ with XD[iaux];

(b4) let $ip := ip + 1$.

Comments

(a) ip divides vector XD into two parts: the first part, up to $ip - 1$ (empty, if $ip = 1$), contains the set values already drawn. The remaining part (empty, if $ip = n + 1$) contains the set values not yet drawn.

(b) Once a cycle of $nd = n$ draws is completed, the random permutation process will restart.
(c) At any time, in order to restart the shuffling process, it is only necessary to reset ip to 1.
(d) A similar approach, using the same data structure and generating set values in advance, can also be used for random sampling generation. It is only necessary to modify the above algorithm for sampling with replacement, instead of sampling without replacement.

The newsboy problem with descriptive sampling

Defining a simulation run by, say, $N = 100$ daily periods, the input sample size is also defined by $n = 100$ values. Here there is no need explicitly to use the inverse transform, since the set values, given by vector $D[i]$, $i = 1, \ldots, 100$, are in the same proportion as in the demand distribution. Table 2 shows how those values are defined.

TABLE 2. *Set values definition for using descriptive sampling in the newsboy problem* ($n = 100$)

Demand	Probability	Set frequency	Set values
50	0.10	10	$D[1..10] = 50$
60	0.12	12	$D[11..22] = 60$
70	0.15	15	$D[23..37] = 70$
80	0.20	20	$D[38..57] = 80$
90	0.18	18	$D[58..75] = 90$
100	0.15	15	$D[76..90] = 100$
110	0.10	10	$D[91..100] = 110$
Total	1.00	100	

The newsboy problem is a very particular and simple case where estimates are not affected by the input sample sequence. Therefore, there is no need in this case to permutate the input values nor to replicate simulation runs.

The M/M/1 queue with descriptive sampling

Two different values for the traffic intensity are considered here: $\rho = 0.75$ and $\rho = 0.90$. At the beginning of every run ($T = 0$), the system is empty and the first customer arrives. To avoid problems that might arise owing to a short run-length, a simulation run lasted until the 10000th service was started. Thus, precisely $ns = 10000$ service times were generated in each run. On the other hand, $na \geqslant 10000$ interarrival times were generated, with na varying between different runs.

Descriptive values for both input variables were generated in advance. For the service times, set values were defined by

$$TS_i = -MS \times Ln[(i - 0.5)/ns], \qquad i = 1, \ldots, ns,$$

where MS is the mean service time.

For the interarrival times, we have that $na \geqslant ns$. Given that $(na - ns)$ is the number of customers remaining in the system when the run finishes, it is unlikely that $(na - ns)$ will exceed, say, 50. Hence, we also used $n = ns = 10000$ for the set-values generation, so that

$$TA_i = -MA \times Ln[(i - 0.5)/ns], \qquad i = 1, \ldots, ns,$$

where MA is the mean interarrival time. Since $(na - ns)$ varies between runs, full set-control for the interarrival times cannot be achieved. However, owing to the large sample-sizes involved, the remaining set variability was negligible. Therefore, each run is based on practically the same set of input values for both variables, but in different random permutations.

EMPIRICAL RESULTS

Descriptive sampling has been successfully applied in a wide range of simulation problems.[12] A sample of such tests is reported here. For each simulation problem, $M = 100$ independent runs were carried out using both simple random sampling (SRS) and descriptive sampling (DS). Summarizing each experiment, mean and variance of the estimates were computed, together with the observed variance reduction and the 95% confidence interval for the overall mean estimate.

Table 3 presents the results for the newsboy problem, the M/M/1 queue and some other problems studied.[12]

TABLE 3. *Summary of results comparing simple random sampling (SRS) with descriptive sampling (DS) for different simulation problems*

Problem	Estimate	SRS mean	DS mean	SRS variance	DS variance	Variance reduction	95% confidence interval SRS mean	DS mean
Newsboy	$\overline{\text{PROF}}$	344.76	344.80	59.790	0.0000	100%	(343.23, 346.29)	(344.80, 344.80)
	S_{PROF}	82.24	82.13	38.689	0.0000	100%	(81.01, 83.47)	(82.13, 82.13)
M/M/1 queue ($\rho = 0.75$)	$\overline{\text{SIZE}}$	2.999	2.989	0.0440	0.0149	66%	(2.957, 3.041)	(2.965, 3.013)
	S_{SIZE}	3.407	3.410	0.1018	0.0722	29%	(3.344, 3.470)	(3.357, 3.463)
M/M/1 queue ($\rho = 0.90$)	$\overline{\text{SIZE}}$	9.123	8.773	3.0990	1.2505	60%	(8.774, 9.472)	(8.551, 8.995)
	S_{SIZE}	9.132	8.679	5.4622	3.1842	42%	(8.669, 9.595)	(8.325, 9.033)
PERT	$\overline{\text{DT}}$	11.214	11.201	0.0652	0.0244	63%	(11.163, 11.265)	(11.170, 11.232)
	S_{DT}	1.863	1.899	0.0322	0.0289	10%	(1.827, 1.899)	(1.865, 1.933)
Inventory	$\overline{\text{COST}}$	78.530	78.526	2.021	0.956	52%	(78.248, 78.812)	(78.332, 78.720)
	S_{COST}	69.497	69.499	6.550	4.411	33%	(68.990, 70.004)	(69.083, 69.915)
Risk analysis	$\overline{\text{PV}}$	392.82	390.44	3031.6	86.5	97%	(381.90, 403.74)	(388.60, 392.28)
	S_{PV}	576.57	586.69	1996.3	1099.6	45%	(567.71, 585.43)	(580.11, 593.27)
M/G/2 queue	$\overline{\text{WT}}$	1.400	1.420	0.02773	0.01930	30%	(1.367, 1.433)	(1.393, 1.447)
M/G/3 queue	$\overline{\text{WT}}$	0.150	0.152	0.00016	0.00009	44%	(0.148, 0.152)	(0.150, 0.154)

Comments on the results

(a) For all problems considered, except for the M/M/1 queue with $\rho = 0.90$, which is discussed later, the estimates means are nearly the same for both sampling methods. This suggests that, in general, descriptive sampling estimates are not biased or that such bias, if present, is small.

(b) Descriptive sampling estimates were more precise. Variance reductions over simple random sampling were, in general, quite high. Therefore, DS's confidence intervals for overall estimates are narrower than the corresponding SRS ones. This means that, using DS instead of SRS, the same interval width can be achieved with fewer runs.

(c) As reported elsewhere,[12] there is a good agreement between the observed variance reductions with DS and the values predicted by the linear response model. This fact gives strong support to the proposed idea.

(d) The newsboy problem deserves special attention. Since estimates do not vary with sample sequence, and once the set values are fixed, there is no variability left. In this case, also confirming the proposed idea, the simulation with descriptive sampling leads to the problem's exact solution.

(e) It is worth mentioning that the only cases where descriptive sampling produced biased results were for steady-state studies of highly congested queueing systems. However, such results were caused by the inadequate use of too short run-lengths; increasing them, the problem disappeared. The M/M/1 queue with $\rho = 0.90$ is one such case, where a run-length defined by 10000 customers being served is still quite small. This follows because, as traffic intensity increases towards unity, the simulation run-length required to produce reliable results grows almost explosively.[14] However, such situations deserve further investigation.

Other empirical results, for more complex systems, are also available.[12] All results confirm the higher statistical efficiency of descriptive sampling. At present, however, to validate the proposed theory, independent studies are required.

CONCLUSIONS

Although a simple idea, descriptive sampling represents a fundamental conceptual change in Monte Carlo sampling, and, as a consequence, it causes heated reactions among simulation specialists and statisticians. This is because the so-called 'imitation principle', until now a basic issue in Monte Carlo, is abandoned. Two objections commonly presented against descriptive sampling are that the results may be biased and that we are not allowing the input sample to vary, even when input distribution parameters are not precisely known. Both objections are now discussed.

Concerning the bias issue, although it is still open to study, empirical tests show that in cases where descriptive sampling produces biased estimates, the bias magnitude is small when compared with the inherent variability of simulation estimates. Therefore, this bias will be irrelevant, except when the simulation run is too short. This becomes clear if we notice that when we increase the input sample size, both descriptive sampling and simple random sampling converge to the same limiting case: the input random variable distribution.

Another bias concern is when rare events play an important role in the simulation. It appears that, in such cases, the descriptive sampling control is too rigid since this event will either occur in all runs or not occur at all. Again, here it is a matter of run-definition, since an event may be rare in a simulation trial but not in a simulation run, where it has to occur at least once. For example, if E is a rare event with probability $p(E)$ and if there are $(k \geqslant 1)$ cases that might produce E in a simulation trial, then $k.p(E)$ is the expected number of E occurrences in a simulation trial. Notice that while the number of E occurrences in a trial is random, the number of E occurrences in a simulation run is fixed and equal to the nearest integer to $N.k.p(E)$, where N is the number of trials in a simulation run. It is up to the analyst to make $N.k.p(E)$ close to an appropriate integer. The previous approach refers to a terminating system, but a similar approach applies to a non-terminating system. In this case, the number of E occurrences in a simulation run is now given by $T.p(E)$.

The second objection against descriptive sampling, that the input sample is not allowed to vary, derives from the erroneous belief that sampling errors can cope with uncertainties about the input distribution parameters. However, such uncertainties cannot be simply compensated for by the introduction of sampling errors, in the same way that a blurred picture cannot become clear by unfocussing its negative. Using descriptive sampling, and not allowing sample parameters to vary, we are closely following model assumptions, as required. And once model assumptions are closely observed, why should the estimates be biassed?

There are many other points to be discussed about descriptive sampling, such as the role played by variance-reduction techniques, procedures for descriptive sampling generation and its similarities with related approaches, like bootstrap[15] and Latin hypercube sampling.[7] These we leave for another occassion.

However, two final points deserve our attention. First, it is worth mentioning that descriptive sampling can be easily tested. The conversion of a simulation program previously written using simple random sampling to descriptive sampling is quite straightforward. The few program modifications to be done are easily located. Second, it is curious that the 'imitation principle' has remained unquestioned since the original Monte Carlo applications. In our view, simulation sampling concerns the more general problem of numerically representing the idea of probability. The new approach to this problem, as suggested by descriptive sampling, may also throw new light into other areas where probability has been applied.

Acknowledgements—I would like to thank John Crookes, my former supervisor at the University of Lancaster, where this research was begun. Special thanks are also due to Dr Ray Paul from the London School of Economics; his help and co-operation were essential in completing this paper. William D. Gallagher also helped me with the English. Thanks are also due to the British Council and IBM Brasil for the financial support given to my visiting programme at the LSE.

REFERENCES

1. J. M. Hammersley and D. C. Handscomb (1964) *Monte Carlo Methods*. Chapman and Hall, London.
2. K. D. Tocher (1963) *The Art of Simulation*. English Universities Press, London.
3. C. E. Clark (1960) The utility of statistics of random numbers. *Opns Res.* **8**, 185–195.
4. C. E. Clark (1961) Importance sampling in Monte Carlo analyses. *Opns Res.* **9**, 603–620.
5. J. P. C. Kleijnen (1974) *Statistical Techniques in Simulation*, Part I. Marcel Dekker, New York.
6. S. S. Lavenberg and P. D. Welch (1981) A perspective on the use of control variables to increase the efficiency of Monte Carlo simulations. *Mgmt Sci.* **27**, 322–335.
7. M. D. McKay, R. J. Beckman and W. J. Conover (1979) A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* **21**, 239–245.
8. B. A. P. James (1985) Variance reduction techniques. *J. Opl Res. Soc.* **36**, 525–530.

9. A. M. LAW and W. D. KELTON (1982) *Simulation Modeling and Analysis*. McGraw-Hill, New York.
10. E. SALIBY (1980) A reappraisal of some simulation fundamentals. Ph.D Thesis, University of Lancaster.
11. E. SALIBY (1990) Understanding the variability of simulation results: an empirical study. *J. Opl Res. Soc.* **41**, 319–327.
12. E. SALIBY (1989) *Rethinking Simulation: Descriptive Sampling*. Atlas/EDUFRJ, Sao Paulo. (In Portuguese.)
13. A. STUART (1976) *Basic Ideas of Scientific Sampling*, 2nd edition. Griffin, London.
14. W. WHITT (1989) Planning queueing simulations. *Mgmt Sci.* **35**, 1341–1365.
15. B. EFRON (1982) The jack-knife, the bootstrap and other resampling plans. In *Regional Conference Series in Applied Mathematics*, No. 38. SIAM, Philadelphia.