Testing independence of simulation subruns: a note on the power of the Von Neumann test

J.P.C. KLEIJNEN
Department of Business and Economics, Tilburg University, Tilburg, The Netherlands

R. van der VEN and B. SANDERS
Dr. Neher Laboratory, PTT, Leidschendam, The Netherlands

Received 9 April 1981

In simulation experiments confidence intervals may be based on subruns. To test whether subruns are independent the Von Neumann statistic is often used. This note shows that the power of this statistic may be very small. The practical recommendation is to apply the statistic only if at least 100 subruns are available.

The simulation literature shows a continuing interest in the use of subruns for the derivation of confidence intervals; see [2,4,5,6]. In such an approach the total run is divided into, say, n subruns, and the subrun-averages are tested for independence. A popular statistic to test for independence is the Von Neumann ratio; see [7]. The purpose of the present note is to emphasize that the power (complement of β-error) of this test statistic is small for less than, say, 100 observations, i.e., if $n < 100$, then there is a sizable chance that the experimenter erroneously accepts the independence hypothesis. Note that this power issue is indirectly addressed in [4, pp. 518–519].

The Von Neumann statistic is defined as

$$Q = \sum_{i=1}^{n-1} (x_{i+1} - x_i)^2 / \sum_{i=1}^{n} (x_i - \bar{x})^2. \tag{1}$$

If H_0 denotes the hypothesis of independent x, then

$$E(Q \mid H_0) = 2. \tag{2}$$

If moreover x is normally distributed, then

$$\sigma^2(Q \mid H_0) = 4(n-2)/(n-1)(n+1). \tag{3}$$

Hence while the expected value remains 2 for any n the standard deviation is quite large for small n, for instance, 0.57 for $n = 10$ and 0.20 for $n = 100$. For $n > 20$ the distribution of Q may be approximated by a normal distribution.

We wish to concentrate on the power of the above test as a function of the sample size n. Therefore we assume that x is indeed normal. We characterize the dependence among the x through the first-order autocorrelation coefficient ρ_1. We quantify the power of the test in two ways, namely analytically introducing some additional assumptions, and empirically using Monte Carlo simulation.

Analytically it is well known that

$$E(Q) = 2 - \frac{2}{n} - 2E(\hat{\rho}_1). \tag{4}$$

If we further assume that under the alternative hypothesis H_1, the statistic Q remains normally distributed with the variance, say σ^2, shown in (3), then (5) results:

$$E\{Q \mid H_1\} = E\{Q \mid H_0\} - 1.96\,\sigma - D_\beta \sigma \tag{5}$$

where D_β stands for the distance between the mean and the β% point of the H_1 distribution. Substituting (3) and (4) into (5) enables us to compute the relation between ρ_1 and β for different n values (provided we neglect the bias of $\hat{\rho}_1$; see below). This results in Table 1; see also fig. 1. This table means that, e.g., a correlation between successive observations of 0.92 results in acceptance of H_0 in 5% of the applications in which 10 subruns are available.

The analytical derivation can be checked through a Monte Carlo simulation. This simulation generates normally distributed variables x with prescribed ρ_1; see [3, p. 234]. The experiment is replicated 100 times for each n value; ρ_1 varies between 0.001 and 0.5. This experiment yields the dashed curves in Fig. 1. The simulation confirms the analytical result except in the situation of $n = 10$ subruns. This deviation is probably explained by the bias of $\hat{\rho}_1$ for instance $E\{\hat{\rho}_1 \mid H_0\} \approx -1/n$. The exact bias under H_1 is complicated

North-Holland Publishing Company
European Journal of Operational Research 9 (1982) 92–93

0377-2217/82/0000-0000/$02.75

Table 1
ρ_1 as a function of number of subruns n and β-error ($\alpha = 0.05$)

Sample size	β (in%)							
	5	10	15	25	50	75	90	97.5
10	0.92	0.82	0.75	0.65	0.46	0.26	0.09	0
50	0.48	0.43	0.40	0.35	0.25	0.16	0.07	0
100	0.35	0.31	0.29	0.25	0.18	0.12	0.06	0

and depends on the structure of the x-process; see [1].

In conclusion, it may be good practice to compute confidence intervals in simulation from only 10 to 20 subruns; see [6]. However, first it must be decided whether the batches are independent. The Von Neumann test yields reliable results only if, say, 100 subruns are available!

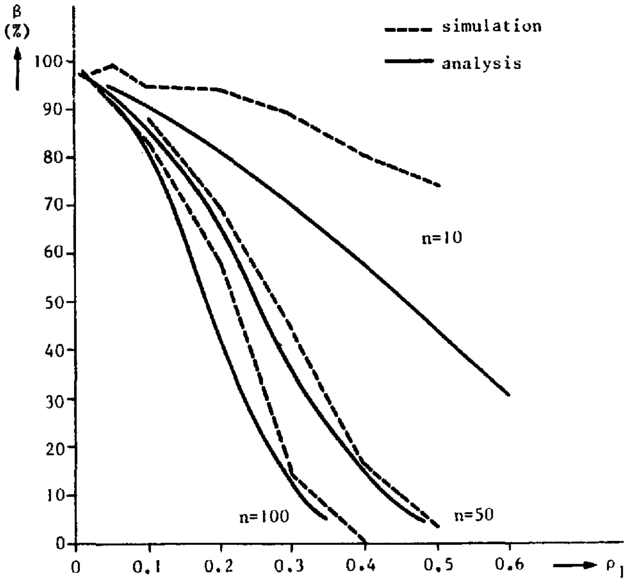

Fig. 1. Relationship between β error, sample size n and correlation ρ_1.

References

[1] O.D. Anderson, Serial dependence properties of ARUMA models, Obtainable from the author (Address: O.D. Anderson, 9 Ingham Grove, Lenton Gardens, Nottinghma NG7 2LQ, England).

[2] R.W. Conway, Some tactical problems in digital simulation, Management Sci. 10 (1) (1963) 47–61.

[3] G.S. Fishman, Concepts and Methods in Discrete Event Digital Simulation (Wiley, New York, 1973).

[4] G.S. Fishman, Grouping observations in digital simulation, Management Sci. 24 (5) (1978) 510–521.

[5] A.M. Law and J.S. Carson, A sequential procedure for determining the length of a steady-state simulation, Operations Res. 27 (5) (1979) 1011–1025.

[6] B. Schmeiser, Batch size effects in the analysis of simulation output, School of Industrial Engineering, Purdue University, June 1980.

[7] J. Von Neumann, Distribution of the ratio of the mean square successive difference to the variance, Ann. Math. Statist. 12 (1941) 367–395.