stichting
mathematisch
centrum

MC

AFDELING MATHEMATISCHE STATISTIEK SN 8/79 AUGUSTUS
(DEPARTMENT OF MATHEMATICAL STATISTICS)

A.J. VAN ES & C. VAN PUTTEN

THE STATAL RANDOM NUMBER GENERATOR
Statal Report 1

2e boerhaavestraat 49 amsterdam

Printed at the Mathematical Centre, 49, 2e Boerhaavestraat, Amsterdam.

The Mathematical Centre, founded the 11-th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications. It is sponsored by the Netherlands Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O).

1980 Mathematics subject classification: Primary 65C10
Secondary 62910, 62M10

The STATAL random number generator

by

A.J. van Es & C. van Putten

ABSTRACT

A pseudo random number generator is a deterministic device to simulate random samples from a uniform distribution. This report gives a description of the construction of and tests on randomness for the linear congruential pseudo random number generator implemented in the statistical package STATAL.

KEY WORDS & PHRASES: *Random numbers, tests on randomness.*

CONTENTS

1. INTRODUCTION

Before discussing the main issue of this report, i.e. the construction of a random*) number generator, we would like to mention two important statistical applications of random number generators.

The first one is *sampling*. Often it is impractical or even impossible to examine all objects in a population. Therefore, one often has to be content with examining a random sample of the population. With a random number generator we can, in an effective way, create (large) samples from populations.

As a second application we mention *simulation*. Sometimes a statistician has to consider a random variable with a distribution function which cannot be calculated analytically or satisfactory approximated using the methods of numerical analysis. For example: the asymptotic properties of a statistic based on a sample might be known, while the statistician is still very interested in how far these asymptotic properties hold for his, finite, sample. One way of tackling this problem is to generate a vast number of samples, using a random number generator, and then to examine the empirical distribution function of the statistic based on these samples.

The main issue of this report will be the construction of a random number generator which generates samples which reasonably can be considered to be samples from an $H((0,1])$ distribution, i.e. a rectangular distribution over the set of real numbers x satisfying $0 < x \le 1$. It will be clear that (especially in simulation) often samples are needed from other distributions. This problem can theoretically be solved by using the following property:

If $\underline{x}$ *is a random variable having an* $H((0,1])$ *distribution and* F *is a distribution function, then* $\underline{y} = F^{-1}(\underline{x})$ *is a random variable having distribution function* F*, where* F^{-1} *is a suitably chosen inverse of* F.

Apart from this method, sometimes more efficient methods are available for generating samples from non $H((0,1])$ distributions, using a random number generator. Some of these methods will be discussed in STATAL Report 2.

*) Strictly speaking one should say "pseudo random".

During the construction of the random number generator, the following criteria were used:

(i) The generated samples have the desired *statistical properties*. Test statistics with known distribution functions under the null hypothesis that a given set of numbers, generated by a random number generator, is a sample from an H((0,1]) distribution, should be looked at.

(ii) *Speed*. Since random numbers are used in great quantities, an effective algorithm is desirable.

(iii) If necessary, it has to be possible to *reproduce* the generated random numbers, in order to check calculations, etc.

One should be aware that the construction of a random number generator often depends on the computer on which it is to be used. In our case this is the SARA computer, consisting of a CDC Cyber 73-28 and a CDC Cyber 173-8.

Our main reference in this report is KNUTH [3].

To conclude with we would like to thank F.J. Burger, T. Jonker and H. Kruizinga for giving us permission to use their computer programs and H.J. Bos and D.T. Winter for helping us with the COMPASS source text of the generator.

2. THE LINEAR CONGRUENTIAL METHOD OF GENERATING RANDOM NUMBERS

There are many methods of generating random numbers. The one we have chosen, the linear congruential method, will be discussed in this chapter.

According to the linear congruential method a sequence of integers is generated by the following algorithm:

choose an integer x_1 $(0 \le x_1 < m)$
determine $x_2, x_3, \ldots$ by

$$x_{i+1} = (ax_i + c) \bmod m \quad (i \ge 1), \tag{2.1}$$

where a, c and m are nonnegative integers
(a is called the multiplier and m the modulus).

Evidently by this algorithm we get sequences of integers x_i, satisfying $0 \le x_i < m$. We get sequences of numbers u_i, belonging to the interval $[0,1)$, by deviding x_i by m,

$$(2.2) \qquad u_i = x_i/m,$$

and hence

$$(2.3) \qquad u_{i+1} = \frac{1}{m}((a(m\ u_i) + c) \bmod m) \qquad (i \ge 1).$$

By now this way of constructing random numbers is classical. More recently POHL [5] has introduced a promising alternative method based on an entirely different idea.

In Chapter 3 we want to use theoretical results based on formula (2.3); therefore, we have to be aware of the fact that no rounding errors should occur in the computations of (2.3). But the numbers u are no integers and real arithmetic on a computer involves rounding errors, so we have to choose specific a, c and m to guarantee that errors will not occur in (2.3). Our random number generator is written in COMPASS (the CDC assembly language) and the modulus m has been chosen to be 2^{48}, 48 being the number of bits used for the mantissa of the representation of a real number in the computer. As a consequence, multiplication by m reduces to adding 48 to the exponential part of a bit representation. In the same way division by m reduces to substracting 48. Some special properties of COMPASS are used to ensure that the other calculations in (2.3) are exact. (For the source text see Appendix VI.)

There are three reasons for choosing this (linear congruential) method.

(i) We expect this method to be faster than alternative methods using more than one previously generated number to calculate the next one.

(ii) Arithmetic modulo 2^{48} can be implemented effectively in COMPASS.

(iii) As mentioned above, there are some important theoretical results on linear congruential random number generators (see Chapter 2).

Apart from these positive points there are some negative points too. Most of the criticism is concentrated on the behaviour of n-tuples of consecutive numbers (cf. MARSAGLIA [4]). However, we think that these negative aspects may be weakened by taking only a fixed number of the most significant bits of the generated numbers and anyway no alternative equally fast generators were known to us when we started the construction.

It is clear that the linear congruential generator generates numbers in $[0,1)$. Since we want to be able to compute the logarithm of the generated numbers we would like them to be elements of $(0,1]$. This is effected by adding $1/m$ to the generated numbers after their computation according to (2.3).

Our main task in the following will be to choose suitable constants a and c ($m = 2^{48}$ will be argumented more extensively in Chapter 3) and then to examine the statistical properties of the resulting random number generator.

3. THE CHOICE OF THE CONSTANTS a, c AND m

3.1. The choice of m

A realization of a random variable having an $H((0,1])$ distribution is a real number in the interval $(0,1]$. Since only a finite number of real values can be represented in a digital computer it is impossible to use a computer to generate samples from an $H((0,1])$ distribution. Therefore, we have to be content with an approximation of the continuous distribution by a discrete one. The obvious choice for such a discrete distribution is a homogeneous distribution on a set of equidistant numbers in $(0,1]$ (i.e. each number has equal probability and the distances between consecutive numbers are equal).

The formulas (2.1) and (2.2) show that the linear congruential random number generator can only generate numbers in the set $R_m = \{0,1/m,2/m,\ldots,(m-1)/m\}$. Since our aim is to approximate a continuous distribution we want m to be the largest integer for which all elements of the set R_m have different representations in the computer.

The representation of a real number on the CDC 73 and 173 is $\text{sign}\times\text{mantissa}\times 2^{\text{exponent}}$, where the mantissa and exponent are integers. There are 48 bits available to represent the mantissa; therefore, two numbers of which the 48 most significant bits are equal will be identified in the computer. To avoid this identification of numbers in the set R_m, we have chosen m to be 2^{48}.

With this choice of m we have a set of 2^{48} equidistant numbers with different representations, but there is no guarantee that all elements of R_m will eventually occur in a linear congruential sequence, i.e. a sequence of numbers defined by formula (2.3). However, we can make choices of a and c to obtain a maximal set of possible values, using the following theorem (based on formulas (2.1) and (2.2)).

The linear congruential sequence $(u_n)_{n=1}^{\infty}$ *has a period* m *(i.e., all elements of* R_m *occur in* $(u_n)_{n=1}^{\infty}$*) if and only if*

(i) c *is relatively prime to* m;

(ii) a-1 *is a multiple of* p *for every prime* p *dividing* m;

(iii) a-1 *is a multiple of* 4 *if* m *is a multiple of* 4.

(Cf. KNUTH [3,p.15].)

As a consequence of the choice $m = 2^{48}$ this reduces to

(3.1)
(i) c is odd;
(ii) a-1 is a multiple of 4.

3.2. The choice of a

3.2.1. Conditions for a

Apart from condition (3.1(ii)) there are other conditions for a (KNUTH [3]).

(3.2)
(i) $\sqrt{m} < a < m-\sqrt{m}$ (preferably $m/100 < a$)
(values of a satisfying (i) will probably give better values for serial correlations);
(ii) the digits in the binary representation of a should not have a regular, simple pattern.

Another condition for the multiplier a (in combination with m) is whether it "passes the spectral test".

3.2.2. The spectral test

Let $(x_k)_{k=1}^{\infty}$ be a sequence of integers, defined by formula (2.2) $(0 \le x_k < m)$.

$$F(t_1,\ldots,t_n) \overset{\text{def}}{=} \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} \delta_{x_k,t_1} \cdots \delta_{x_{k+n-1},t_n}$$

$$(0 \le t_\ell < m,\ \ell = 1,\ldots,n),$$

$$\delta_{i,j} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

$F(t_1,\ldots,t_n)$ is the limiting density of the number of appearances of the n-tuple $(t_1,\ldots,t_n)$.

Using a finite Fourier transform and formula (2.1), a function $f(s_1,\ldots,s_n)$ is obtained (see KNUTH [3]), satisfying

$$(3.3) \qquad F(t_1,\ldots,t_n) =$$

$$\frac{1}{m^n} + \frac{1}{m^n} \sum_{(s_1,\ldots,s_n)\in I_{a,m,n}} \exp\left(2\pi i\left(\frac{s_1}{m}t_1+\ldots+\frac{s_n}{m}t_n\right)\right) f(s_1,\ldots,s_n),$$

where

$$I_{a,m,n} \overset{\text{def}}{=} \{(s_1,\ldots,s_n) \mid 0 \le s_\ell < m \ (\ell = 1,\ldots,n),\ (s_1,\ldots,s_n) \ne (0,\ldots,0),\ s_1 + s_2 a + \ldots + s_n a^{n-1} = 0 \bmod m\}$$

(a is the multiplier of the linear congruential sequence), and

$$|f(s_1,\ldots,s_n)| = 1.$$

If $(x_n)_{n=1}^{\infty}$ would have been a genuine random sequence of integers, the corresponding limiting density $F^*(t_1,\ldots,t_n)$ would have been a constant, equal to $1/m^n$ (there are m^n possible n-tuples). Therefore, we may get an idea of the randomness of the n-tuples by comparing $F(t_1,\ldots,t_n)$ and $F^*(t_1,\ldots,t_n)$. KNUTH [3,pp.82-96] gives the following criterion for judging a multiplier a (in combination with a modulus m), but his arguments are rather intuitive.

Determine

$$v_n \overset{\text{def}}{=} \min_{(s_1,\ldots,s_n)\in I_{a,m,n}} \sqrt{s_1^2+\ldots+s_n^2} \qquad (n = 1,2,\ldots,6);$$

then

$$c_n \overset{\text{def}}{=} \frac{\pi^{n/2}\, v_n^n}{\Gamma(n/2)m} \qquad (n = 1,2,\ldots,6).$$

We say that a multiplier a passes the spectral test, if $c_n \geq 0.1$ for $n = 2,3,4$ ($c_1 = 2$ for all a and m). It passes the spectral test with "flying colours" if $c_n \geq 1$ for $n = 2,3,4$.

3.2.3. The resulting choice of a

Our choice of a is $5^{17} = 762939453125 \cong 7.63\times10^{11}$. The binary representation of 5^{17} is:

1011 0001 1010 0010 1011 1100 0010 1110 1100 0101

Clearly, $5^{17}-1$ is a multiple of 4, so condition (3.1(ii)) is satisfied. The digits in the binary representation do not seem to have a simple, regular pattern, so condition (3.2(ii)) is satisfied too.

$$\begin{aligned} m - \sqrt{m} &\cong 2.81\times10^{14} \\ m/100 &\cong 2.81\times10^{12} \\ \sqrt{m} &\cong 1.68\times10^{7}. \end{aligned}$$

Clearly, $\sqrt{m} < a < m-\sqrt{m}$; so condition (3.2(i)) is satisfied.

The results of the spectral test for $a = 5^{17}$ and $m = 2^{48}$ are:

$$c_2 = 1.69 \quad c_3 = 1.59 \quad c_4 = 2.33 \quad c_5 = 0.625 \quad c_6 = 0.813.$$

a passes the spectral test with "flying colours".

(These results of the spectral test are approximations. For a discussion on their reliability see Appendix II.)

3.3. The choice of c

According to KNUTH [3] there is a connection between c and the serial correlation calculated over the full range of 2^{48} possible consecutive random numbers generated by a linear congruential generator. KNUTH suggests c should be chosen somewhere near $m \times (1/2 - 1/6\sqrt{3})$ to minimize this serial correlation. Although we have not spotted the effect of this special choice of c, we shall still follow his suggestion.

Our choice of c is 59482661568303. Since c is odd, condition (3.1(i)) is satisfied.

4. STATISTICAL TESTS

In this chapter the quality of the pseudo random number generator is investigated. Before being able to do so, "quality" has to be made more precise in a mathematical sense. This may be done by posing the following problem:

> "Is it possible to distinguish between the pseudo random number generator and a (genuine) random number generator on account of only sequences of their realizations, without any other information available?"

An answer to this question is all that is relevant for the statistical utility of a pseudo random number generator.

To examine the problem we consider results $u_1, \ldots, u_n$ of the pseudo random number generator to be realizations of random variables $\underline{u}_1, \ldots, \underline{u}_n$ and test the null hypothesis:

(4.1) H_0: $\underline{u}_1, \ldots, \underline{u}_n$ is a random sample from an $H((0,1])$ distribution.

In the following a description of 8 tests to test (4.1) is given and their results for the pseudo random number generator are summarized. These tests are the Kolmogorov-Smirnov test, the frequency test, the serial test, the gap test, the partition test, the coupon collector's test, the permutation test and the run test.

Like in all statistical tests, not rejecting the null hypothesis doesn't imply we accept H_0 to be true. In the present case, applying tests to samples generated by a pseudo random number generator, only negative aspects of the generator may be discovered and if not, a tentative answer might be given to the question above with interpretation:

> "The quality of the pseudo random number generator doesn't seem to be bad."

From now on, in this chapter $x_1,\dots,x_n$ denotes a sample, generated by the pseudo random number generator, according to (2.3).

4.1. The Kolmogorov-Smirnov test

This test compares the empirical distribution function $\hat{\underline{F}}$ and the theoretical distribution function F; in this case that of the H((0,1]) distribution.

The test statistic is:

$$\max_{i=1,\dots,n} \operatorname{abs}(\hat{F}(\underline{x}_i) - F(\underline{x}_i)).$$

The specific property of the random number generator tested by this test is whether the generated numbers $x_1,\dots,x_n$ are equally distributed over (0,1]. (For a more detailed description of the Kolmogorov-Smirnov test see e.g. [2].)

The following tests (except the run test) are based on the *chi-square test*. Say we have n independent observations $\underline{x}_1,\dots,\underline{x}_n$. Suppose there are k possible disjoint events $A_1,\dots,A_k$ and $P(\underline{x}_i \in \bigcup_{j=1}^{k} A_j) = 1$ $(i = 1,\dots,n)$. Let $\underline{z}_j$ $(j = 1,\dots,k)$ be the number of occurrences of event A_j in $\underline{x}_1,\dots,\underline{x}_n$.

$$p_j \overset{\text{def}}{=} P(\underline{x}_i \in A_j) \quad (i = 1,\dots,n;\ j = 1,\dots,k).$$

Then the statistic

$$\sum_{j=1}^{k} \frac{(\underline{z}_j - np_j)^2}{np_j}$$

approximately (as n tends to infinity) has a chi-square distribution with k-1 degrees of freedom.

A reasonable approximation may be expected if $np_j \geq 5$ $(j = 1,\dots,k)$, i.e., the expected number of occurrences of each event should at least be equal to 5.

4.2. The frequency test

The interval (0,1] is divided into k parts $((j-1)/k,j/k]$ of equal length $1/k$ $(j = 1,\dots,k)$. We say that event A_j $(j = 1,\dots,k)$ occurs when observing $\underline{x}$ from an $H((0,1])$ distribution iff $\underline{x} \in ((j-1)/k,j/k]$. Clearly p_j, the probability of occurrence of A_j under the null hypothesis (4.1), is:

$$p_j = 1/k \qquad (j = 1,2,\dots,k).$$

Now apply the chi-square test to the sequence $\underline{x}_1,\dots,\underline{x}_n$. The resulting test statistic approximately has a chi-square distribution with k-1 degrees of freedom.

The specific property of the random number generator tested by this test is whether the numbers $\underline{x}_1,\dots,\underline{x}_n$ are equally distributed over the k intervals mentioned above.

Instead of using the sample $\underline{x}_1,\dots,\underline{x}_n$ directly, we shall often use a sequence $\underline{y}_1,\dots,\underline{y}_n$, defined by:

$$(4.2) \qquad \underline{y}_i \overset{\text{def}}{=} \begin{cases} [d\underline{x}_i] & \text{if } \underline{x}_i \in (0,1) \\ 0 & \text{if } \underline{x}_i = 1 \end{cases} \qquad (i = 1,\dots,n),$$

where d is a certain fixed positive integer. ([x] is the entier of x.)

Under the null hypothesis (4.1) $\underline{y}_1,\dots,\underline{y}_n$ are independent and $P(\underline{y}_1 = m) = 1/d$ $(m = 0,1,\dots,d-1;\ i = 1,2,\dots,n)$.

4.3. The serial test

Starting from a sample $\underline{x}_1,\dots,\underline{x}_m$ $(m = 2n)$ a sequence of n consecutive pairs $(\underline{y}_1,\underline{y}_2),(\underline{y}_3,\underline{y}_4),\dots,(\underline{y}_{2n-1},\underline{y}_{2n})$ is formed. Under the null hypothesis (4.1) these pairs are independent and

$$P((\underline{y}_{2j-1},\underline{y}_{2j}) = (s,t)) = 1/d^2 \qquad (s = 0,1,\dots,d-1;\ t = 0,1,\dots d-1).$$

Now apply the chi-square test to the sequence $(\underline{y}_1,\underline{y}_2),\dots,(\underline{y}_{2n-1},\underline{y}_{2n})$. The d^2 events correspond to the d^2 possible pairs and therefore

$$p_j = 1/d^2 \qquad (j = 1,\dots,d^2).$$

The resulting test statistic approximately has a chi-square distribution with d^2-1 degrees of freedom. Specific properties of the random number generator tested by this test are whether the $\underline{y}_i$ are equally distributed over the set $\{0,1,\ldots,d-1\}$ and whether $\underline{y}_{2i-1}$ and $\underline{y}_{2i}$ are independent.

4.4. The gap test

Let α and β be real numbers satisfying $0 \leq \alpha < \beta \leq 1$. We now consider lengths of consecutive subsequences $\underline{x}_m, \underline{x}_{m+1}, \ldots, \underline{x}_{m+r}$ of $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ with $\underline{x}_{m+r} \in (\alpha,\beta)$ and $\underline{x}_i \notin (\alpha,\beta)$ for $i = m,m+1,\ldots,m+r-1$ (such a subsequence will be called a gap of length r). We do not want to have too many possible lengths, since the probability of occurrence of large lengths of gaps is very small and a justified application of the chi-square test would require a very large sample. Therefore, we choose a positive integer t and all subsequences, which would have had a length larger than t, are broken off at their t-th element and given length t. The next element of $\underline{x}_1, \underline{x}_2, \ldots$ will be the first element of the next subsequence.

Now, starting from the sequence $\underline{x}_1, \ldots, \underline{x}_m$ $(m = nt)$, we form the sequence $\underline{\ell}_1, \ldots, \underline{\ell}_n$ $(0 \leq \underline{\ell}_i \leq t,\ \underline{\ell}_i \in \mathbb{N})$ of lengths of consecutive gaps. Under the null hypothesis (4.1) the $\underline{\ell}_i$ are independent and for $i = 1,\ldots,n$:

$$\begin{aligned} p_0 &= P(\underline{\ell}_i=0) = p \overset{\text{def}}{=} \beta-\alpha \\ p_r &= P(\underline{\ell}_i=r) = p(1-p)^r \qquad (r = 1,\ldots,t-1) \\ p_t &= P(\underline{\ell}_i=t) = (1-p)^t. \end{aligned}$$

Now apply the chi-square test to $\underline{\ell}_1, \ldots, \underline{\ell}_n$. The events are the t+1 possible lengths with probabilities as mentioned above. The resulting test statistic has a chi-square distribution with t degrees of freedom.

The property of the random number generater tested by the gap test is whether the generated numbers tend to avoid or prefer a given interval (α,β) in a way that cannot be expected from a random sequence.

4.5. The partition test

We consider n groups of k successive elements of the sequence $\underline{y}_1, \ldots, \underline{y}_m$ $(m = nk)$ defined by (4.2) for a certain positive integer d.

$\underline{m}_i \overset{\text{def}}{=}$ the number of different integers in the k-tuple $(\underline{y}_{(i-1)k+1},\dots,\underline{y}_{ik})$ $(i = 1,\dots,n)$.

Under the null hypothesis (4.1) the $\underline{m}_i$ are independent and

$$P(\underline{m}_i = r) = \frac{d(d-1)\dots(d-r+1)}{d^k} \left[{k \atop r}\right], \qquad r = 1,\dots,k,$$

where $\left[{k \atop r}\right]$ is a Stirling number, defined by

$$(4.3) \qquad \left[{k \atop 1}\right] = \left[{k \atop k}\right] = 1 \quad \text{and} \quad \left[{k \atop r}\right] = r\left[{k-1 \atop r}\right] + \left[{k-1 \atop r-1}\right] \qquad (r = 2,\dots,k-1).$$

$\left[{k \atop r}\right]$ is the number of ways a set of k elements can be divided into r nonempty subsets. (For a table of $\left[{k \atop r}\right]$ $(k,r = 1,2,\dots,10)$ see Appendix V.)

Now apply the chi-square test to $\underline{m}_1,\dots,\underline{m}_n$. The k events are the occurrences of the k possible values of the $\underline{m}_i$ and the probabilities are as mentioned above. The resulting test statistic approximately has a chi-square distribution with k-1 degrees of freedom.

4.6. The coupon collector's test

Having chosen a positive integer t, we consider the sequence $\underline{y}_1,\underline{y}_2,\dots,\underline{y}_m$ $(m = nt)$, defined by (4.2) for a certain positive integer d. From this sequence we form the sequence $\underline{\ell}_1,\dots,\underline{\ell}_n$ of lengths of consecutive subsequences, required to obtain a complete set of integers $0,1,\dots,d-1$. Like the gap test sequences which would have had lengths larger than t will be given length t.

Now apply the chi-square test. The events are the occurrences of the $t-d+1$ possible values of $\underline{\ell}_i$ $(d,d+1,\dots,t)$. Under the null hypothesis (4.1) the $\underline{\ell}_i$ are independent and the probabilities of occurrence of the events are

$$p_j = \frac{d!}{d^j}\left[{j-1 \atop d-1}\right] \qquad (j = d,d+1,\dots,t-1),$$

$$p_t = 1 - \frac{d!}{d^{t-1}}\left[{t-1 \atop d}\right],$$

where [ ] again are the Stirling numbers.

The resulting test statistic approximately has a chi-square distribution with t-d degrees of freedom.

4.7. The permutation test

We consider n consecutive subsequences of length t of the sequence $\underline{x}_1,\ldots,\underline{x}_m$ (m = nt). Each subsequence has one of t! possible rankings, so we can form a sequence $\underline{m}_1,\ldots,\underline{m}_n$, where $\underline{m}_j$ indicates the ranking of the j-th subsequence (j = 1,...,n). Under the null hypothesis (4.1) the $\underline{m}_i$ are independent and each ranking has probability

$$p_j = 1/t! \quad (j = 1,\ldots,t!).$$

Now apply the chi-square test. The events are the occurrences of the t! rankings and the probabilities are as mentioned above.

The resulting test statistic approximately has a chi-square distribution with t!-1 degrees of freedom.

4.8. The run test

There are two kinds of run tests. One kind considers the lengths of "runs up" in the sequence $\underline{x}_1,\ldots,\underline{x}_n$, i.e. increasing subsequences. The other kind considers "runs down", i.e. decreasing subsequences.

Since there is a known dependency between lengths of consecutive runs, we cannot apply the chi-square test. However, by a suitable transformation of the number of occurrences of runs of a given length smaller than or equal to some integer t (to runs which would have had length larger than t, length t is given) one can obtain a test statistic, approximately having a chi-square distribution with t degrees of freedom under the null hypothesis (4.1) (cf. KNUTH [3,pp.60-63]).

4.9. Results

Each test has been performed 40 times and since the tests are performed on disjoint subsequences of a sequence generated by the generator, the statistics may be considered to be independent under the null hypothesis (4.1).

A description with more details of the tests performed and the values of the calculated statistics is given in Appendix III.

The following methods are used to interpret the results.

(i) For each test the significant values are counted ($\alpha = 0.05$).

(ii) For each test, except the Kolmogorov-Smirnov test, the sum of the 40 statistics is calculated. These sums and their upper tail probabilities are given in the table below. (Normal approximations are used.)

(iii) The sum of all statistics, except those of the Kolmogorov-Smirnov test, is calculated. This sum and its upper tail probability is given below. (A normal approximation is used.)

(iv) Fisher combinations (see Appendix III). For each test the statistic and its upper tail probability are given in the table below.

(v) Fisher combination of all tests. The statistic and its upper tail probability are given below.

By the results mentioned in Table 4.9 there does not seem to be a reason to suspect the generator of having serious faults.

As an additional test of the statistical independency of the generated numbers, Pearson correlations and serial correlations have been computed. Again these results do not give reason to suspect the generator. For a detailed description see Appendix IV.

Table 4.9 Results

test	df	number of significant values	sum of test statistics	upper tail probability of the sum of test statistics	Fisher combination	upper tail probability of the Fisher combination
Kolmogorov-Smirnov test	-	1	-	-	87.36	0.27
Frequency test	50	0	1947.90	0.80	70.65	0.76
Serial test	99	5	4050.00	0.16	107.45	0.02
Gap test (0,½)	7	5	298.95	0.21	94.70	0.13
Gap test (¼,¾)	7	2	257.03	0.83	69.43	0.79
Gap test (½,1)	7	1	255.44	0.85	66.92	0.85
Partition test	3	1	108.57	0.77	70.87	0.76
Coupon collector's test	5	0	193.18	0.63	72.40	0.71
Permutation test	23	4	964.10	0.16	93.00	0.15
Run test (up)	6	2	245.26	0.41	83.71	0.37
Run test (down)	6	4	271.65	0.07	98.88	0.07

df = degrees of freedom of the statistic of one single test.

The total sum (iii) is 8592.08. The upper tail probability is 0.29.

The value of the statistic of the Fisher combination of all tests is 915.36. The upper tail probability is 0.20.

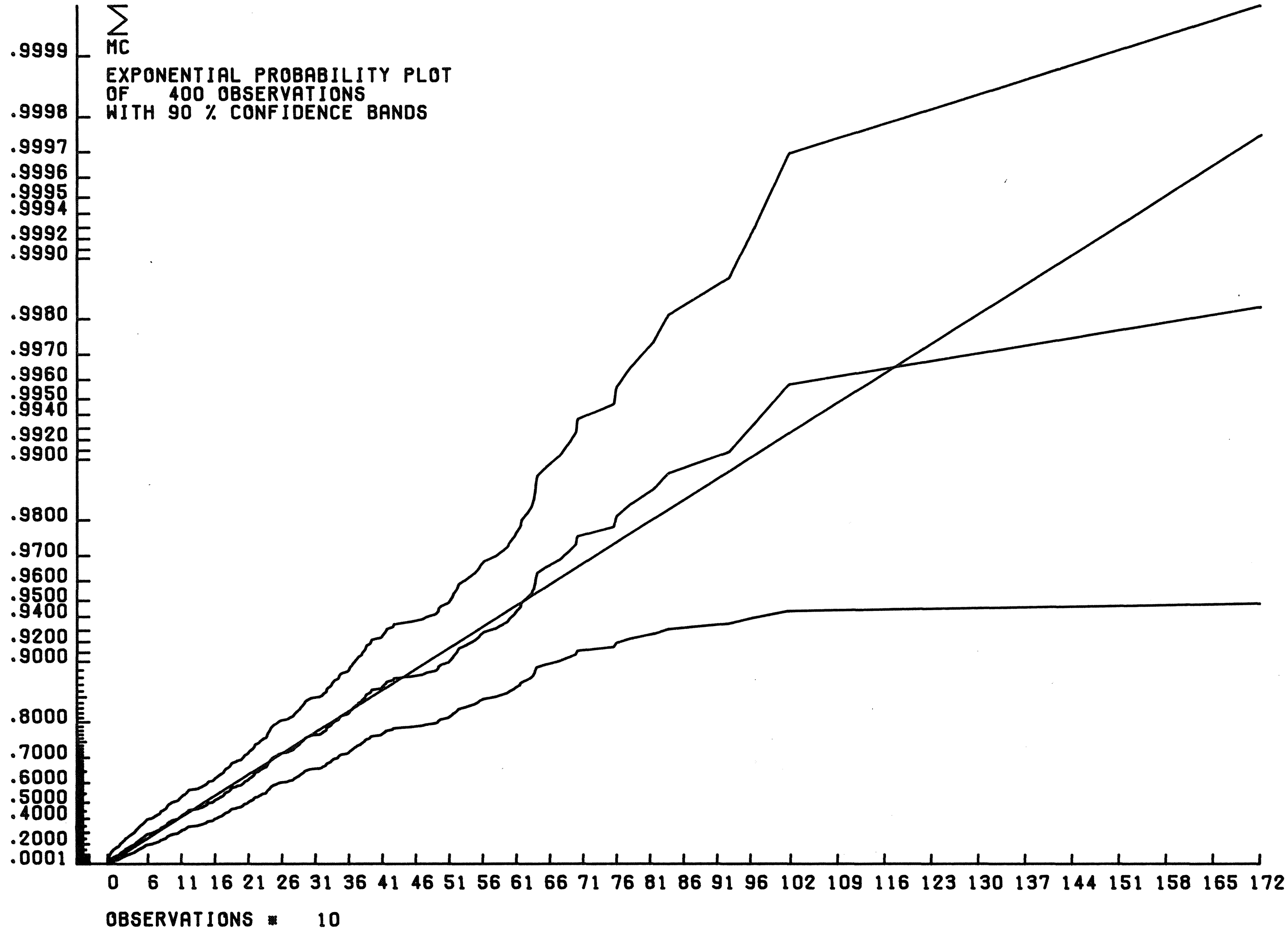


Figure 4.9 Plot of the terms -2 log($t_{i,j}$) of the Fisher combination of all tests (defined in Appendix III.2)

REFERENCES

[1] ANDERSON, T.W. & A.M. WALKER (1964), *On the asymptotic distribution of the autocorrelations of a sample from a linear stochastic process,* Ann. of Math. Stat. 35, 1296-1303.

[2] CONOVER, W.J. (1971), *Practical nonparametric statistics,* J. Wiley, New York.

[3] KNUTH, D.E. (1969), *The art of computer programming,* Volume 2, Seminumerical algorithms, Addison Wesley, Reading Mass.

[4] MARSAGLIA, G. (1972), *The structure of linear congruential sequences,* in: Applications of number theory to numerical analysis, ed. S. Zaremba, Academic Press, New York, 249-286.

[5] POHL, P. (1976), *Description of MCV, a pseudo random number generator,* Scand. Act. J., 1-14.

[6] WITTING, H. & G. NÖLLE (1970), *Angewandte mathematische Statistik,* B.G. Teubner, Stuttgart.

APPENDIX I

AN EXAMPLE OF THE SPECTRAL TEST

In this appendix we consider the linear congruential random number generator with a = 5, c = 1 and m = 8 and for this generator we shall calculate c_2. Choose

$$\begin{aligned}
x_1 &= 0 \Rightarrow \\
x_2 &= (5x_1+1) \bmod 8 = 1 \bmod 8 = 1 \\
x_3 &= (5x_2+1) \bmod 8 = 6 \bmod 8 = 6 \\
x_4 &= (5x_3+1) \bmod 8 = 31 \bmod 8 = 7 \\
x_5 &= (5x_4+1) \bmod 8 = 36 \bmod 8 = 4 \\
x_6 &= (5x_5+1) \bmod 8 = 21 \bmod 8 = 5 \\
x_7 &= (5x_6+1) \bmod 8 = 26 \bmod 8 = 2 \\
x_8 &= (5x_7+1) \bmod 8 = 11 \bmod 8 = 3 \\
x_9 &= (5x_8+1) \bmod 8 = 16 \bmod 8 = 0 = x_1 \\
x_{10} &= x_2
\end{aligned}$$

etc.

From this we see that the following consecutive pairs (t_1,t_2), $0 \le t_1,t_2 \le 8$ appear:

$$(0,1),(1,6),(6,7),(7,4),(4,5),(5,2),(2,3),(3,0) \quad \text{(8 pairs)}$$

$$I_{5,8,2} = \{(s_1,s_2) \mid 0 \le s_1,s_2 < 8,\ (s_1,s_2) \ne (0,0),\ s_1+5s_2 = 0 \bmod 8\}.$$

In order to calculate c_2 we have to determine the elements (s_1,s_2) of this set. Suppose $s_1+5s_2 = 0 \bmod 8$.

$$\begin{aligned}
s_1 &= 0 \Rightarrow s_2 = 0; \\
s_1 &= 1 \Rightarrow 5s_2+1 = 0 \bmod 8 \Rightarrow s_2 = 3; \\
s_1 &= 2 \Rightarrow 5s_2+2 = 0 \bmod 8 \Rightarrow s_2 = 6; \\
s_1 &= 3 \Rightarrow 5s_2+3 = 0 \bmod 8 \Rightarrow s_2 = 1; \\
s_1 &= 4 \Rightarrow 5s_2+4 = 0 \bmod 8 \Rightarrow s_2 = 4; \\
s_1 &= 5 \Rightarrow 5s_2+5 = 0 \bmod 8 \Rightarrow s_2 = 7; \\
s_1 &= 6 \Rightarrow 5s_2+6 = 0 \bmod 8 \Rightarrow s_2 = 2; \\
s_1 &= 7 \Rightarrow 5s_2+7 = 0 \bmod 8 \Rightarrow s_2 = 5.
\end{aligned}$$

$I_{5,8,2}$ consists of the following pairs:

$$(1,3),(2,6),(3,1),(4,4),(5,7),(6,2),(7,5) \qquad \text{(7 pairs)}$$

$$\Rightarrow v_2 = \min_{(s_1,s_2)\in I_{5,8,2}} \sqrt{s_1^2+s_2^2} = \sqrt{10}$$

$$\Rightarrow c_2 = \frac{\pi^{2/2}\times(\sqrt{10})^2}{\Gamma(2/2)\times 8} = \frac{5}{4}\pi \approx 3.93.$$

APPENDIX II

DISCUSSION ON ROUNDING ERRORS IN THE COMPUTER PROGRAM FOR THE SPECTRAL TEST

If, like in our case, m is the so called wordsize of the computer, then the algorithm of the spectral test needs multiple precision integer arithmetic. KNUTH [3] claims that experience shows that triple precision is adequate.

However, the program available for the spectral test, used only double precision real arithmetic. Therefore, we had reason to believe that our results were not exact and we decided to examine the effect of rounding errors to the results of the program.

A procedure has been added to effect that all double precision real calculations were done with an a priori chosen precision (less than double). With this modified version of the program we performed spectral tests for three combinations of a and m with precision running down from 96 bits (double precision). The results are presented in the tables below.

The three generators tested are: gen I - $a = 26353589$
$m = 2^{26}$
gen II - $a = 5^{17}$
$m = 2^{42}$
gen III - $a = 5^{17}$
$m = 2^{48}$

(gen III is our ultimate choice.)

Table II Values of $c_2,\ldots,c_6$ for different precisions and constants

	no. of bits	c_2	c_3	c_4	c_5	c_6
gen I	96	0.27820779	2.61798268	3.43431251	0.05012644	0.71314257
	90	0.27820779	2.61798268	3.43431251	0.05012644	0.71314257
	85	0.27820779	2.61798268	3.43431251	0.05012644	0.71314258
	80	0.27820779	2.61798268	3.43431250	0.05012644	0.71314636
	75	0.27820779	2.61798265	3.43431246	0.05012644	0.71314257
	70	0.27820773	2.61798179	3.43431107	0.05012644	0.04929749
	65	0.27820596	2.61795411	3.43426651	0.05012659	0.71314270
	60	0.27814910	2.61706849	3.43284083	0.05013149	0.54322148
	55	0.27632987	2.58878142	3.38737520	0.05028889	0.04699253
	48	neg. arg. SQRT				
gen II	96	1.48054214	1.66539457	1.69033297	0.35349367	0.43202142
	94	1.48051416	1.66538968	1.69022004	0.35329656	0.43191698
	92	1.48039553	1.66537031	1.68937165	0.35259968	0.43165423
	90	1.47992285	1.66529383	1.68599399	0.34984391	0.43007613
	88	1.47806497	1.66499048	1.67295451	0.33902830	0.35483813
	86	1.47063344	1.66377840	1.62134986	0.29778615	0.40240655
	84	1.44090731	1.52073337	1.21204256	0.16361628	neg. arg. SQRT
gen III	96	1.69222718	1.58578335	2.33342210	0.62506645	0.81349145
	95	1.69221725	1.38298518	2.12657848	neg. arg. SQRT	

In each of the three cases an error message occurred when the rounding errors became large enough to cause an argument of the square root procedure to be negative (error message: "negative argument SQRT").

The general idea we get from the first two tables is that the results tend to increase with increasing precision of the arithmetic. If $m = 2^k$ then KNUTH claims, as mentioned above, that a precision of 3k bits is usually adequate. Table II) seems to confirm this claim. However, double precision (2k bits) is hardly enough to make the program perform properly (i.e. without ending prematurely with an arithmetic error). In the case of our generator (gen III) the 96 bits of the double precision arithmetic in the original program seem to produce rather unreliable results, but since it is likely they are smaller than the exact values of $c_2,c_3,\ldots,c_6$ and since we are interested in large values of the c's, we may still conclude that our generator passes the spectral test.

APPENDIX III

RESULTS OF STATISTICAL TESTS

1. The parameters of the tests and numerical results

For each test the parameters and probabilities are given below (their meaning is the same as in the description of the tests in Chapter 4).
If the test statistic approximately has a chi-square distribution, then the degrees of freedom, df, are mentioned.

A) Kolmogorov-Smirnov test.
$n = 100$.

B) Frequency test.
$n = 1000$, $k = 51$, $df = 50$
probabilities: $p_1 = p_2 = \ldots = p_{51} = 1/k = 1/51 \approx 0.01961$.

C) Serial test.
$n = 1000$, $d = 10$, $df = 99$
probabilities: $p_1 = p_2 = \ldots = p_{100} = 0.01$.

D,E,F) Gap test.
$n = 1000$, $t = 7$, $df = 7$
The gap test has been performed for three pairs of (α,β):

D) $(\alpha,\beta) = (0,\frac{1}{2})$
E) $(\alpha,\beta) = (\frac{1}{4},\frac{3}{4})$
F) $(\alpha,\beta) = (\frac{1}{2},1)$.

Since for each pair $\beta-\alpha = \frac{1}{2}$ the probabilities are:

$p_0 = 0.5$	$p_4 = 0.03125$
$p_1 = 0.25$	$p_5 = 0.015625$
$p_2 = 0.125$	$p_6 = 0.0078125$
$p_3 = 0.0625$	$p_7 = 0.0078125$.

G) Partition test.
$n = 1000$, $d = 5$, $k = 4$, $df = 3$
probabilities: $p_1 = 0.008$ $\quad p_3 = 0.576$
$p_2 = 0.224$ $\quad p_4 = 0.192$.

H) Coupon collector's test.

$n = 500$, $d = 5$, $t = 10$, $df = 5$

probabilities: $p_5 = 0.0384$ $\quad p_8 = 0.1075$
$p_6 = 0.0768$ $\quad p_9 = 0.1045$
$p_7 = 0.0998$ $\quad p_{10} = 0.5729$.

I) Permutation test.

$n = 1000$, $t = 4$, $df = 23$

probabilities: $p_1 = p_2 = \ldots = p_{24} = 1/24 \approx 0.0417$.

j,K) Run test.

$n = 5000$, $df = 6$

J) runs up

K) runs down.

Each of the tests has been performed 40 times. To avoid dependency of the test statistics (under the null hypothesis (4.1)) care has been taken that each of these 11×40 tests are performed on disjoint subsequences of a sequence generated by our random number generator.

The results are given in the tables below.

Table III.1 Upper tail probabilities of the Kolmogorov-Smirnov tests (A)

0.08	0.64	0.08	0.82	0.55	0.15	0.06	0.14
0.03	0.11	0.89	0.89	0.78	0.96	0.30	0.96
0.90	0.46	0.79	0.17	0.59	0.36	0.55	0.65
0.73	0.07	0.45	0.97	0.15	0.95	0.41	0.40
0.69	0.90	0.45	0.91	0.07	0.20	0.02	0.52

Table III.2 The values of the test statistics for the tests B,C,D,E,F,G,H,I,J,K

B	C	D	E	F	G	H	I	J	K
56.4	101.4	6.40	14.02	2.05	1.53	3.97	26.9	6.88	4.89
60.3	66.0	5.79	5.67	9.93	4.63	5.13	20.0	2.39	2.06
59.0	124.8	9.12	10.50	7.67	2.31	5.10	34.7	14.67	14.72
42.6	70.8	2.70	7.27	2.49	8.23	6.01	37.0	1.36	4.76
54.5	82.8	14.30	1.29	12.93	1.90	1.60	27.8	15.00	9.98
54.2	124.6	2.65	8.30	8.21	0.57	6.25	34.9	11.48	4.11
52.9	120.8	4.08	2.19	5.00	1.15	3.36	18.0	8.11	7.21
64.1	82.4	10.77	14.25	6.24	1.41	11.07	32.2	9.87	8.80
36.6	90.8	5.17	6.56	8.19	2.92	1.63	36.0	0.84	2.72
47.7	103.4	2.69	6.75	10.81	0.23	7.46	8.8	3.11	3.42
32.9	113.0	10.51	1.98	6.50	3.25	6.81	25.3	6.70	18.01
56.7	134.8	10.01	3.85	4.05	2.33	2.30	21.1	2.95	2.61
54.0	157.0	4.95	0.61	4.86	3.84	2.59	18.1	6.02	7.24
60.8	101.2	1.00	9.87	3.72	1.45	2.50	23.5	3.39	1.82
37.7	79.8	2.60	4.87	5.22	2.42	7.79	27.7	4.01	6.76
50.6	95.2	5.15	4.37	8.59	3.28	8.10	23.6	9.17	11.66
60.1	115.8	14.16	9.78	2.10	1.10	8.19	16.3	5.52	6.23
31.4	112.6	4.37	5.65	4.86	3.73	2.69	35.4	8.62	6.49
39.9	85.6	10.43	5.32	3.37	0.93	2.33	15.8	11.87	5.80
43.7	95.2	1.50	3.94	5.69	5.23	4.88	24.4	4.66	7.18
53.7	105.0	3.78	9.31	2.51	7.29	4.71	12.3	2.48	12.82
51.6	101.4	5.40	5.92	3.91	2.11	5.81	15.3	2.91	2.95
64.9	66.8	5.83	3.89	5.29	3.38	1.28	16.9	8.01	3.21
45.7	124.2	12.84	3.51	5.11	0.61	1.93	27.1	8.09	2.52
46.2	70.2	12.51	10.93	8.40	1.61	3.08	14.4	5.54	11.99
44.5	105.6	3.97	7.67	1.59	1.18	3.75	21.4	3.43	3.99
50.0	78.0	7.62	3.59	1.86	2.05	5.12	20.1	5.85	7.28
38.0	100.4	1.36	7.43	4.63	0.19	6.53	21.0	4.19	13.34
55.9	104.4	5.13	3.06	9.16	6.42	6.63	29.7	9.52	11.32
45.1	119.6	4.51	9.48	5.94	0.19	3.74	37.3	11.11	11.17
55.5	103.8	19.03	7.08	7.62	1.85	8.76	26.2	3.32	5.71
40.1	80.4	11.87	2.18	8.38	4.56	6.09	25.3	6.00	6.98
33.1	112.0	17.35	3.01	12.95	4.70	6.03	26.3	3.05	6.15
58.6	111.8	10.20	10.18	6.53	7.39	6.76	25.2	6.74	6.38
41.8	94.2	6.79	3.72	17.03	1.51	2.91	20.2	9.89	5.46
31.9	110.8	3.08	5.09	4.59	4.10	2.37	26.1	2.00	2.35
52.8	100.6	15.11	6.40	3.59	4.60	6.67	20.0	6.74	4.95
51.3	103.6	6.92	7.10	7.91	0.02	4.14	30.8	2.26	0.65
61.6	99.2	6.34	15.46	9.27	0.37	1.92	24.1	2.57	7.17
29.5	100.0	10.96	4.71	6.69	2.00	5.19	16.9	4.94	8.79

(significant values at 5% level are underlined).

2. The interpretation of the numerical results

Let $\underline{x}_{i,j}$ denote the i-th statistic of test j $(i = 1,\ldots,40;$ $j = A,B,\ldots,K)$. Then under the null hypothesis the $\underline{x}_{i,j}$ are independent and $\underline{x}_{i,j}$ approximately has a chi-square distribution for $j = B,C,\ldots,K$ with df_j degrees of freedom.

(i) Significant values.

The tests A,B,...,K are one-sided tests (reject H_0 if the value of the statistic is too large). For each test the number of significant values is counted $(\alpha = 0.05)$.

(ii) Sums per test.

$$\underline{a}_j \overset{\text{def}}{=} \sum_{i=1}^{40} \underline{x}_{i,j}.$$

For $j = B,C,D,\ldots,K$ $\underline{a}_j$ approximately has a chi-square distribution with $40 \times df_j$ degrees of freedom. For these tests we can use a normal approximation for $\underline{a}_j$.

(iii) The total sum.

$$\underline{a} \overset{\text{def}}{=} \sum_{j=B}^{K} \sum_{i=1}^{40} \underline{x}_{i,j} = \underline{a}_B + \underline{a}_C + \ldots + \underline{a}_K.$$

$\underline{a}$ approximately has a chi-square distribution with $40 \times \sum_{j=B}^{K} df_j$ degrees of freedom, so we can use a normal approximation for $\underline{a}$.

(iv) Fisher combination per test.

$$t_{i,j}(x_{i,j}) \overset{\text{def}}{=} P(\underline{x}_{i,j} \geq x_{i,j}) \qquad (i = 1,\ldots,40;\ j = A,B,\ldots,K)$$

$$\underline{t}_{i,j} \overset{\text{def}}{=} t_{i,j}(\underline{x}_{i,j}).$$

$\underline{t}_{i,j}$ has a Hom((0,1]) distribution and the $\underline{t}_{i,j}$ are independent.

$\Rightarrow$ $-2\log(\underline{t}_{i,j})$ has an exponential distribution with parameter ½, but this is the same as a chi-square distribution with 2 degrees of freedom.

$\Rightarrow$ $\underline{f}_j \overset{\text{def}}{=} \sum_{i=1}^{40} -2\log(\underline{t}_{i,j})$ $(j = A,B,\ldots,K)$ has a chi-square distribution with 80 degrees of freedom.

(v) Fisher combination of all tests.

$$\underline{f} \overset{\text{def}}{=} \sum_{j=A}^{K} \underline{f}_j$$ has a chi-square distribution with 880 degrees of freedom.

APPENDIX IV

CORRELATIONS

In order to examine the independency of the generated numbers, Pearson and serial correlations have been computed.

PEARSON CORRELATIONS

For k = 2,3,4,5 the correlation matrices based on the data matrices $(z_{i,j}^{(k)})$, $i = 1,2,\dots,1000$; $j = 1,2,\dots,k$ have been computed, where $z_{i,j}^{(k)} = x_m$, $m = (i-1)k + j$ and $x_1, x_2, \dots$ is a sequence of numbers generated by our generator.

The data matrix (k fixed) is:

$$(z_{i,j}^{(k)}) = \begin{pmatrix} x_1 & x_2 & x_3 & \dots & x_k \\ x_{k+1} & x_{k+2} & x_{k+3} & \dots & x_{2k} \\ \cdot & & & & \\ \cdot & & & & \\ \cdot & & & & \\ x_{999k+1} & & & & x_{1000k} \end{pmatrix}$$

The sequence (x_i) which has been used is disjoint from the sequences used for the computations of the statistics in Appendix III, so the results can be considered to be independent.

If $\underline{r}_n$ is the Pearson correlation coefficient based on a sample $(\underline{x}_1,\underline{y}_1),\dots,(\underline{x}_n,\underline{y}_n)$, then, if the correlation of $\underline{x}_i$ and $\underline{y}_i$ is zero, $\sqrt{n}\underline{r}_n$ is known to have approximately a standard normal distribution (for large n) (cf. WITTING, NÖLLE [6,p.49]). So in this case (n = 1000) correlations which are in absolute value greater than $1.96\times\sqrt{0.001} \approx 0.06198$ might be considered to be significant ($\alpha = 0.05$) when testing the hypothesis that the correlation of $\underline{x}_i$ and $\underline{y}_i$ is zero. For a reasonable test of this hypothesis one would have to consider the complete correlation matrix instead of each correlation separately. However, since only one of the correlations computed is significant, we do not have reason to suspect the generator of having serious faults.

SERIAL CORRELATIONS

Define the serial correlation of order k:

$$\underline{s}_k \overset{\text{def}}{=} \frac{n \sum_{i=1}^{n} \underline{x}_i \, \underline{x}_{(i+k) \bmod n} - \left(\sum_{i=1}^{n} \underline{x}_i\right)^2}{n \left(\sum_{i=1}^{n} \underline{x}_i^2\right) - \left(\sum_{i=1}^{n} \underline{x}_i\right)^2} .$$

The serial correlations of order 1 to 5 have been computed on disjoint sequences of 100 random numbers. We cannot guarantee, however, that these sequences are disjoint from the sequences used in the computations of the statistics in Chapter 4 or from those used for the computation of the Pearson correlations.

The statistics $\sqrt{n}\,\underline{s}_1,\ldots,\sqrt{n}\,\underline{s}_m$ based on a sample $\underline{x}_1,\ldots,\underline{x}_n$ under certain conditions which hold in case the $\underline{x}_i$ are Hom((0,1]) distributed and independent, are known to have approximately (for large n) a multivariate normal distribution with covariance matrix equal to the identity matrix (cf. ANDERSON & WALKER [1]).

In case n = 100, values of the serial correlations which are in absolute value greater than $1.96\times\sqrt{0.01} = 0.196$ might be considered to be significant ($\alpha = 0.05$) when testing the hypothesis that the correlation of $\underline{x}_i$ and $\underline{x}_{i+k}$ is zero for $k = 1,\ldots,m$.

Since only two of the computed serial correlations are significant we do not have reason to suspect the generator of having serious faults.

RESULTS

The computed Pearson correlation matrices (significant values at the 5% level are underlined) are:

k = 2:	1				
	0.024	1			

k = 3:	1			
	0.003	1		
	0.000	0.002	1	

k = 4:	1				
	0.011	1			
	-0.003	0.003	1		
	-0.018	0.037	-0.033	1	

k = 5:	1				
	-0.014	1			
	0.016	0.004	1		
	0.036	0.004	0.017	1	
	0.031	0.041	0.014	-0.045	1

k = 2:	1	
	0.013	1

k = 3:	1		
	-0.054	1	
	0.022	0.064	1

k = 4:	1			
	-0.006	1		
	-0.013	-0.006	1	
	-0.011	-0.032	0.002	1

k = 5:	1				
	0.006	1			
	0.054	0.040	1		
	-0.054	-0.004	-0.053	1	
	0.015	0.034	0.014	0.032	1

The computed serial correlations are:

order:	1	2	3	4	5
	-0.219	-0.062	0.005	0.045	0.026
	-0.175	-0.021	0.013	-0.014	0.145
	-0.031	-0.130	0.158	-0.033	0.050
	-0.052	-0.173	0.090	-0.078	0.061
	0.151	-0.390	-0.076	0.004	0.046
	-0.054	-0.028	-0.001	-0.129	0.106
	-0.183	0.056	0.134	-0.015	-0.031
	0.141	0.028	-0.193	-0.019	0.031
	0.038	0.028	-0.169	-0.081	-0.075
	-0.123	-0.179	0.132	-0.152	0.047

APPENDIX V

TABLE OF THE STIRLING NUMBER $\left[\begin{smallmatrix}k\\r\end{smallmatrix}\right]$ (k,r = 1,...,10)

k \ r	1	2	3	4	5	6	7	8	9	10
1	1									
2	1	1								
3	1	3	1							
4	1	7	6	1						
5	1	15	25	10	1					
6	1	31	90	65	15	1				
7	1	63	301	350	140	21	1			
8	1	127	966	1701	1050	266	28	1		
9	1	255	3025	7770	6951	2646	462	36	1	
10	1	511	9330	34105	42525	22827	5880	750	45	1

APPENDIX VI

SOURCE TEXT

The procedure is written in COMPASS and can be used in an ALGOL 60 program as a code procedure (code number = 41308).

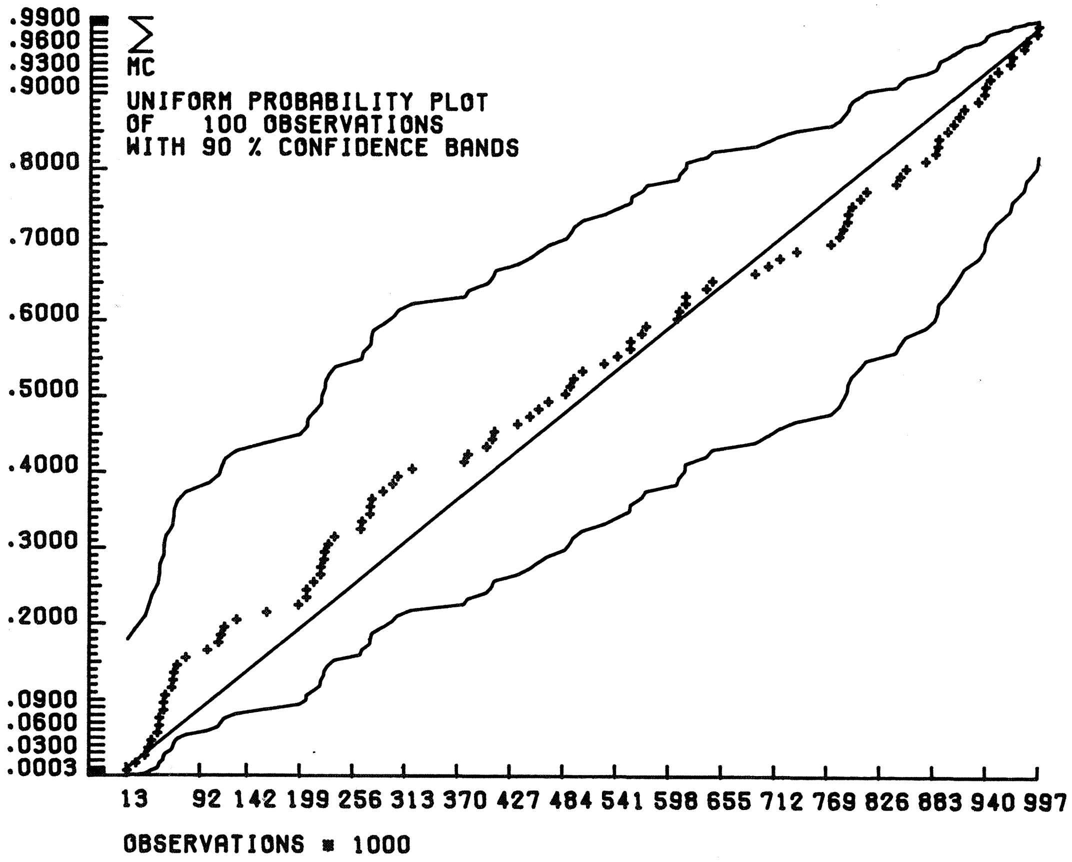
```
      IDENT    ASELECT
      SST
      CODE     41308
      SPEC     1,VS
      VALUE    1,X5
      UX5      B5
      SB5      B5+48
      LX5      B5
      SA4      =762939453125
      SA3      =59482661568303
      IX6      X5*X4
      IX5      X3+X6
      SB5      -48
      PX5      B5
      NX6      X5
      SA6      SAVE
      ASSIGN   1,X6,,CHECK
      SA5      SAVE
      SX3      1
      SB5      -48
      PX3      B5
      NX4      X3
      FX5      X4+X5
      RETURN
SAVE  BSS      1
      END
```

APPENDIX VII

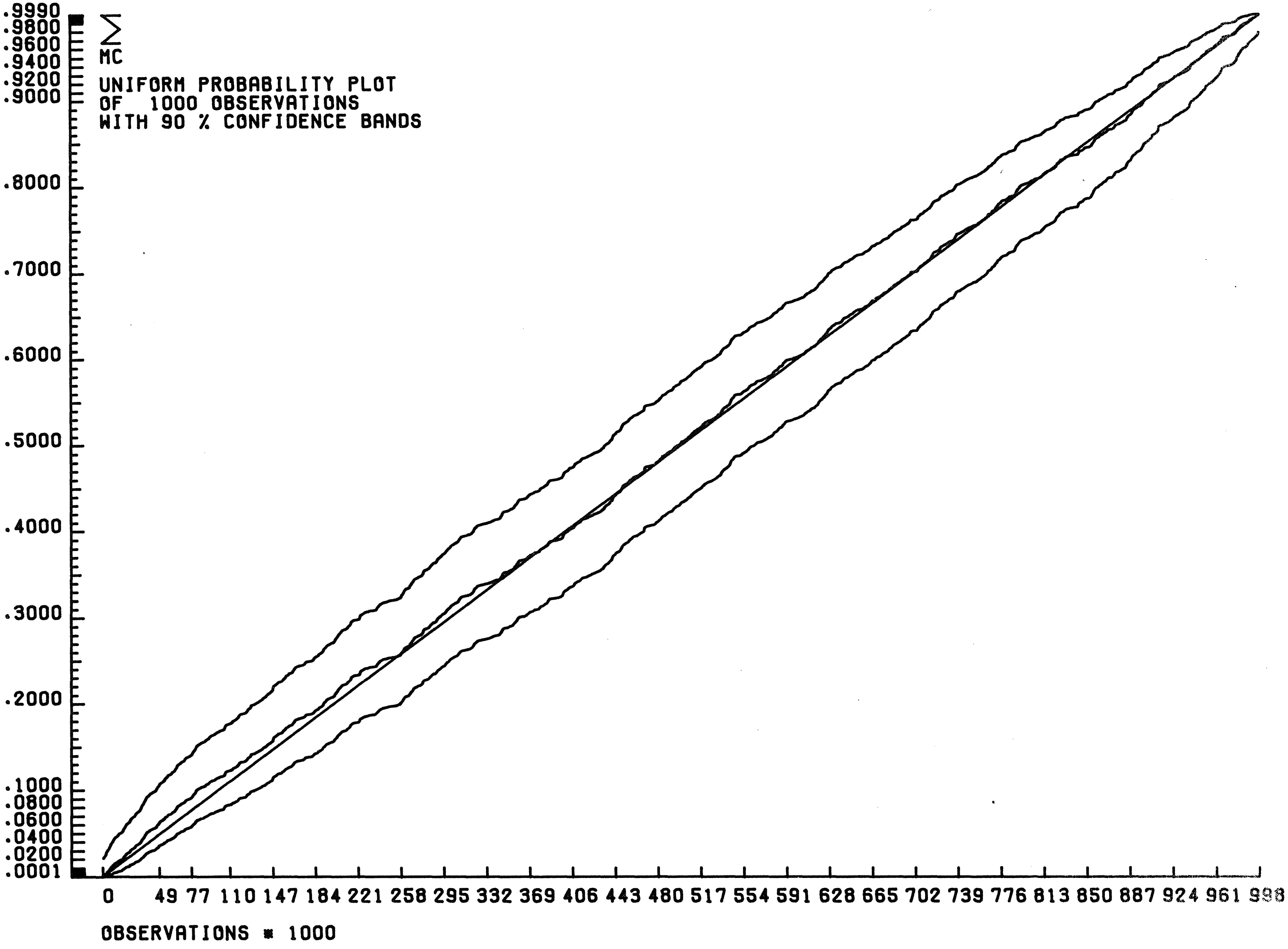
PROBABILITY PLOTS

The empirical distribution functions, based on two samples, have been plotted. The first sample consisted of 100 simulated observations; the second one of 1000 observations. Both samples had identical starting value 0.6789.

In the figures below 90% confidence bands are plotted. The theoretical distribution function is contained in the area enclosed by these bands with probability 0.90. Also an estimation of this theoretical distribution function (the straight line) is plotted. For the theoretical background of this the reader is referred to DOKSUM, K.A. (1977), *Some graphical methods in statistics. A review and some extensions,* Statistica Neerlandica <u>31</u>, 53-61.



<u>Figure VII.1</u>



.9990
.9800
.9600
.9400
.9200
.9000
.8000
.7000
.6000
.5000
.4000
.3000
.2000
.1000
.0800
.0600
.0400
.0200
.0001
∑
MC
UNIFORM PROBABILITY PLOT
OF 1000 OBSERVATIONS
WITH 90 % CONFIDENCE BANDS
0
49
77
110
147
184
221
258
295
332
369
406
443
480
517
554
591
628
665
702
739
776
813
850
887
924
961
998
OBSERVATIONS * 1000

Figure VII.2