CHAPTER FOUR

RANDOM NUMBERS

4.1 Preliminaries

There is - at least at present - no way to attain randomness. This is not a new insight by any means. Von Neumann (1951) wrote eloquently about randomness in the following excerpt:

> "Any one who considers arithmetical methods of producing random digits is, of course, in a state of sin. For, as has been pointed out several times, there is no such thing as a random number."

However, some random number sequences are better than others. We would like our random numbers to:

1) Be uniformly distributed.

2) Not be correlated with one another.

3) Be relatively easy to generate.

Before we show one of the fundamental and commonly used random number generators (RNGs), we need to present some elementary number theory.

Number Theory

Number theory is the systematic study of the natural numbers. ***Divisible*** means that there is no remainder after division. For example $6 \div 3 = 2$. Therefore, 6 is divisible by 3 - written $3 \mid 6$.

The ***remainder*** when a number x is divided by n has n possibilities, $0, 1, 2, \ldots, (n-1)$. If 11 is divided by 5, the remainder is 1. If x is divided by 5, the remainders could be 0, 1, 2, 3, or 4.

We use these ideas in the following language of number theory.

$11 \equiv 1 \pmod 5$ This is stated as "11 is congruent to 1 moduluo (or mod) 5." We could interpret divisibility and modulus in two ways which include the following:

1) $5\ \overset{2}{\overline{\big)11}}$ remainder 1 or

2) $5 \mid 11-1$ or $5 \mid 10$ remainder 0

In the relation $11 \equiv 1 \pmod 5$, 5 is called the modulus and 1 is called the remainder. In general, $a \equiv b \pmod m$ means that $m \mid a-b$.

Midsquare Random Number Generator

Von Neumann and Metropolis in the 1940s developed the ***midsquare method*** to generate random numbers in the interval (0, 1). The method proposed is summarized as follows:

1) Start with any four digit positive integer N_o, called the ***seed***.

2) Take $(N_o)^2$. If $(N_o)^2$ does not have eight digits, add zeroes to the left to make it an eight digit number.

3) Take the middle four digits of $(N_o)^2$. This number is N_1

4) Let $N_1 = a_1a_2a_3a_4$. The first random number is $,a_1a_2a_3a_4$

5) Repeat the process where N_2 is the middle four digits of $(N_1)^2$, etc.

The midsquare method looked good in that the numbers generated appeared random. However, the whole sequence $N_1, N_2, \ldots$ is predictable given N_0, and worse - the results rapidly approach zero.*

*Law, A. and Kelton, W. *Simulation Modeling and Analysis*, McGraw Hill, New York, 1991, pp. 420-424.

For example, let $N_0 = 1256$. $N_0^2 = 1{,}577{,}536$. Let $N_o^2 = 01\underline{5775}36$. $N_1 = 5775$. The first random number $R_1 = .5775$. $(N_1)^2 = 33{,}\underline{350{,}6}25$. $N_2 = 3506$. The second random number generated is $R_2 = .3506$. $(N_2)^2 = 12\underline{{,}292{,}0}36$. Our next random number is $R_3 = .2920$ and so on.

EXERCISES 4.1

1) Prove that if $a \mid b$, then $a \mid bc$.

2) If $a \mid b$ and $a \mid c$, show that $a \mid (b + c)$.

3) Is the converse of the proof in question #2 true or false?

4) Find the first 10 random numbers using the midsquare method in the seed $N_0 = 1239$.

5) For the linear congruence $2x \equiv b \pmod{5}$, find b if $x = 0, x = 3$.

4.2 Linear Congruential Random Number Generators

The majority of random number generators are ***linear congruential generators*** (LCGs). We use the formula: $N_i \equiv (a\ N_{i-1} + c) \pmod{m}$, where m is called the modulus, a the multiplier, c the increment, and N_0 the seed. These numbers are whole numbers (0, 1, 2, ...). To obtain the random numbers on [0, 1], take N_i/m as the ith random number.

For example, let N_0 (the seed) = 5. Let a = 2, c = 4, and m = 7.

$$N_1 \equiv (2\ (5) + 4) \pmod{7}$$

$$N_1 \equiv 14 \pmod{7}$$

$$N_1 \equiv 0$$

This is an exceptionally weak random number generator since

$$N_2 \equiv (2\ (0) + 4) \pmod{7},\ N_2 \equiv 4 \pmod{7} \rightarrow N_2 = 4$$

Similarly, $N_3 \equiv (2\,(4) + 4) \pmod{7}$, $N_3 \equiv 12 \pmod{7} \rightarrow N_3 = 5$. $N_4 = 0$, and the sequence repeats.

Our sequence, $N_0, N_1, N_2, \ldots$ is 5, 0, 4, 5, 0, 4, . . . The sequence of "random numbers" is 5/7, 0/7, 4/7, 5/7, 0/7, 4/7, . . . This is a particularly poor candidate for randomness, but all LCGs fall short. The looping behavior always occurs, but sometimes we can achieve a full cycle of m distinct "random numbers" before the cycle repeats. The theorem that follows the definition of ***relatively prime*** helps in selecting LCGs that have a full period.

A definition that is preliminary to the theorem is ***relatively prime***. Two integers are relatively prime if the only divisors that they have in common is 1. For example, 7 and 15 are both divisible only by one; they are relatively prime. The integers 8 and 10 are not relatively prime since both are divisible by 2. The following theorem uses this concept and is helpful in selecting effective pseudo-random number generators.

Theorem: A linear congruential random number generator has a full period if:

1) The integers m and c are relatively prime.

2) If q is a prime, and $q \mid m \rightarrow q \mid a-1$.

3) If $4 \mid m \rightarrow 4 \mid a-1$.

For example, consider our equation for obtaining LCGs - $N_i \equiv (a\,N_{i-1} + c) \pmod{m}$. Let $a = 6$, $m = 5$, $c = 7$. The three conditions of the theorem hold. We leave the verification as an exercise.

Let $N_o = 3$ be the seed.

$N_1 \equiv (6\,(3) + 7) \pmod{5}$

$N_1 \equiv 25 \pmod{5} \rightarrow N_1 = 0$

$N_2 \equiv (6\,(0) + 7) \pmod{5} \rightarrow N_2 = 2$

$N_3 \equiv (6\,(2) + 7) \pmod{5} \rightarrow N_3 = 4$

$N_4 \equiv (6\,(4) + 7) \pmod{5} \rightarrow N_4 = 1$

$N_5 \equiv (6\,(1) + 7) \pmod{5} \rightarrow N_5 = 3$, and the cycle repeats.

The cycle is 0, 2, 4, 1, 3, 0, 2, 4, 1, 3, . . .

The associated random variables are 0/5, 2/5, 4/5, 1/5, 3/5, 0/5, 2/5, 4/5, 1/5, 3/5, . . .

Irrational Numbers

Several years ago this author took several of his college students to listen to Persi Diaconis (Stanford University) lecture on randomness. Dr. Diaconis dismissed all audience suggestions on ways to attain perfect randomness except when this author broached the irrationals as a possible source of random number sequences. The irrationals, such as $\sqrt{2}$, cannot be derived from modular congruence relations from polynomial functions with rational coefficients. They are truly a quantitative leap from the rationals. The proof of this recently proved theorem relies upon linear congruence, arithmetic and divisibility, and is presented as follows:

Theorem: Let $\sqrt{2} = 1.b_1b_2b_3\ldots$. There is no polynomial function $f(x) = a_0 + a_1x + a_2x^2 + \ldots + a_nx^n$, with $a_i \in$ integers, where $f(x) \equiv b_k \pmod{m}$, where $k = 1, 2, 3, \ldots$ and $m \in$ integers.

Proof: Assume such a function exists. Therefore, $f(1) \equiv b_1$, $f(2) \equiv b_2, \ldots f(k) \equiv b_k \pmod{m}$. From elementary number theory we know that $f(a + k) \equiv f(a) \pmod{k}$ for polynomial functions $f(x)$. Therefore, $\sqrt{2}$ repeats its decimal expansion every k digits. $\sqrt{2} = 1.b_1b_2\ldots b_k\ \overline{b_1b_2\ldots b_k}$. This is a contradiction. Therefore, no such polynomial function exists.

If the irrationals are a source of interest to the reader, you could turn your eventual research interests to the foundations of mathematics and to firming the foundations or to the infinite decimal expansion as a future source of improved random numbers.

Multiplicative Random Number Generators

Multiplicative linear congruential random number generators are derived by deleting c in the previous formula. The revised formula is: $N_i \equiv a N_{i-1} \pmod{m}$. This is a special case of the general formula: $N_i \equiv (a N_{i-1} + c) \pmod{m}$ if $c = 0$. If $c > 0$, we call these linear congruential RNGs by the name ***mixed generators***. According to Law and Kelton (1991), the improved performance anticipated in (recently developed) mixed generators has not been shown.

The multiplicative generators cannot achieve full period because m always divides $c = 0$ [condition 1 of the last theorem]. For computational efficiency, we select $m = 2^a$ to avoid actual division. The computer uses integral overflow for large m, e.g., $m \geq 2^{31}$. The largest integer y that can be represented is $2^{31} - 1$, and if we try to store any integer with $h > 31$ binary numbers, we will lose the left-most $(h - 31)$ binary digits. The remaining number is $y \pmod{2^a}$. Thus we can achieve modulo division by overflow for $m = 2^a$.

There have been problems in selecting appropriate values $m = 2^a$. For example, the very poor and flawed random number generator RANDU used $m = 2^{31}$, $a = 2^{16} + 3 = 65{,}539$, $c = 0$, and demonstrated a severe problem with a lack of uniformity on the unit cube. Uniformity on the unit cube* means that if the cube were divided into smaller cubes, e.g., 1000 cubes with edge of 1/10, an equal number of random numbers would appear in each of the 1000 cubes. Of course, by equal distribution in the 1000 smaller cubes, we rely upon a goodness of fit test like the χ^2 to compare observed to expected frequency where expected equals $1/1000 \cdot N$ (N representing a large sample of three digit pseudo-random numbers). Although no random number generator yields truly random numbers, some generators are better than others. It took several years to uncover the serious problems with RANDU.

*Law, A. and Kelton, W. *Simulation Modeling and Analysis*, 2nd ed., McGraw Hill, New York, 1991, p. 444.

EXERCISES 4.2

1. Find $N_1, N_2, N_3, \ldots$ for the following multiplicative RNG.

   a) $N_i \equiv 5\,(N_{i-1}) \pmod{3}$, $N_o = 1$

   b) $N_i \equiv 12\,(N_{i-1}) \pmod{16}$, $N_o = 3$

   c) $N_i \equiv 4\,(N_{i-1}) \pmod{13}$, $N_o = 1$

2) Find $N_1, N_2, N_3, \ldots$ for the following linear congruential RNG.

   a) $N_i \equiv (2\,N_{i-1} + 3) \pmod{5}$

   b) $N_i \equiv (3\,N_{i-1} + 5) \pmod{16}$

   c) $N_i \equiv (N_i + 11) \pmod{11}$

3) Find N_{100} for (a) of the first exercise.

4) For the random number generator, $N_i \equiv (5\,N_{i-1} + d) \pmod{8}$, find a d necessary for N_i to achieve a full period.

5) Knuth (1981) showed that if $m = 2^b$, then the period is at most 2^{b-2}.* What is the maximum number of integers $0, 1, \ldots, (m - 1)$ that can be obtained for the N_is?

6) Use an advanced numerical procedure to generate the decimal expansion of an irrational, e.g., $\sqrt{2}$, beyond machine storage. Design a method to select digits of this pseudo-random number. Analyze the output and modify your algorithm to produce an improved generator. Test for uniformity and for autocorrelation. (To test for autocorrelation review the section on correlation from Chapter One and compute r. Test whether $r = 0$ using the t test that is mentioned in this text.)

*Knuth, D. *The Art of Computer Programming*, Vol. 2, 2nd ed., Addison-Wesley, Reading Mass. (1981).

7) Consider $f(x) = x^3 + x - 2$. Solve the following congruence relation for b.

a) $11 \equiv b \pmod{10}$

b) $f(11) \equiv b \pmod{10}$

c) $25 \equiv b \pmod{10}$

d) $f(25) \equiv b \pmod{10}$

e) $f(35) \equiv b \pmod{10}$

This last example illustrates an important result from number theory. The property that this shows is the following:

If $a \equiv b \pmod{m}$ and $f(x) = a_0x^n + a_1x^{n-1} + \ldots + a_n$ is a polynomial function with integral coefficients for a_i, then $f(a) \equiv f(b) \pmod{m}$.

4.3 Mathematical Induction

One of the most elegant and powerful techniques in mathematics is ***mathematical induction***. To get an idea of induction (as it is familiarly called), imagine a set of dominoes, back to back. If one falls, the next falls. We drop the first domino, and this causes the second to drop and so on.

Induction proofs usually prove a statement over the natural numbers (or perhaps extended to whole numbers or integers). For example, consider the sum of the arithmetic progression.

$$s_1 = 2, \qquad s_2 = 2 + 4 = 6, \qquad s_3 = 2 + 4 + 6 = 12,$$

$$s_4 = 2 + 4 + 6 + 8 = 20, \qquad s_5 = 30, \qquad s_6 = 42.$$

$s_n = \frac{n}{2}(a + \ell)$ is the general formula where a = first term, ℓ = last term, and d = the common difference, which is constant and not part of the general formula. For our example, $s_n = \frac{n}{2}(a + \ell) = n(n + 1)$. Let us prove this by induction.

To show $s_n = \frac{n}{2}(a + \ell)$ is a valid formula for $s_n = 2 + 4 + 6 + 8 + \ldots + 2n$, we use three steps:

1) Show that the statement is true for $n = 1$.

2) Assume the statement true for n.

3) Show that from assuming the statement true for n, the statement is necessarily true for $(n + 1)$.

Thus, we have shown that the statement is true for $n = 1, 2, 3, \ldots$, namely all natural numbers.

Let us return to the proof that $s_n = \frac{n}{2}(a + \ell)$ for our example.

1) $s_n = 2 + 4 + 6 + 8 + \ldots + 2n$

a) Test if statement (1) is true for $n = 1$.

$n = 1 \rightarrow s_1 = 2,\ (1) = 2n = 2.$ This is true.

b) Next, assume the statement is true for n. This means the following:

$$s_n = 2 + 6 + 6 + \ldots + 2n = \frac{n}{2}(a + \ell) = \frac{n}{2}(2 + 2n) = n\,(n + 1).$$

c) We must show that from assuming (b) we can show the statement is true for $(n + 1)$. This means the following: $s_{n+1} = \frac{(n + 1)}{2}(a + \ell) =$

$$\frac{(n + 1)}{2}(2 + 2\,(n + 1)) = (n + 1)\,(1 + n + 1) = (n + 1)\,(n + 2)$$

To prove this, start with (b) which we assumed. $s_n = n\,(n + 1)$.

Add the $(n + 1)$st term to both sides. $s_n + 2\,(n + 1) = s_{n+1} =$

$n\,(n + 1) + 2\,(n + 1)$. Factor the right side of the equation.

$\therefore$ $s_{n+1} = (n + 1)\,(n + 2)$. This completes our proof using induction.

Consider a second example from elementary calculus. To show $\frac{d^n\,(x\,e^x)}{d\,x^n} = (x + n)\,e^x$

follow the three steps.

1) For $n = 1$, $\frac{d\,(x\,e^x)}{dx} = x\ e^x + e^x = (x + 1)\ e^x$

The formula holds for $n = 1$.

2) Assume $\frac{d^n\,(x\,e^x)}{d\,x^n} = (x + n)\ e^x$

3) $\frac{d^{n+1}\,(x\,e^x)}{d\,x^n} = \frac{d}{dx}\left[\frac{d^n}{dx^n}(x\,e^x)\right]$

$$\frac{d^{n+1}\,(x\,e^x)}{d\,x^n} = (x + n)\ e^x + e^x\ (1)$$

$$= (x + n + 1)\ e^x$$

Therefore, (3) holds. We conclude that the above formula works $\forall n = 1, 2, 3, \ldots$

Now let us return to our linear congruential random number generator and use mathematical induction to show that each N_i is determined by the parameters m, a, c, and N_o in the congruence relation. Hence, the numbers that we derive from the congruence relation are far from random.

We have defined the set of "pseudo-random" numbers by the relation:

$N_i \equiv (a N_{i-1} + c) \pmod{m}$

Let us use mathematical induction to prove that linear congruential random numbers are not random. Indeed we will prove that:

For recursive random number

(1) generator $N_i \equiv (a N_{i-1} + c) \pmod{m}$

(2) that $N_i \equiv (a^i N_o + c \frac{(a^i - 1)}{a - 1}) \pmod{m}$

Before we prove the theorem, start with examples:

Let $a = 6, m = 5, c = 7, N_o = 3$

This example lets us achieve a full period of "pseudo" random numbers. This is because

1) m, c are relatively prime.

2) If q is a prime and $q \quad m \rightarrow q \quad a - 1$

$5 = q, 5 \quad m \rightarrow 5 \quad 6 - 1$

3) If $4 \quad m \rightarrow 4 \rightarrow a - 1$

Since 4 does not divide m, 3 does not apply.

Substitute. $N_1 \equiv (6(3) + 7) \pmod{5}$ in 1) above. $N_1 = 0$

$N_0 = 0$

$N_2 \equiv (6(0) + 7) \pmod{5}$, $N_2 = 2$

Continue. $N_3 = 4$, $N_4 = 1$, $N_5 = 3$. The cycle is full - 30241.

Before we prove (2), let us start with examples. We will prove

(2) $$N_i \equiv (a^i N_o + c(\frac{a^i - 1}{a - 1})) \pmod m$$

Substitute

Let i = 3.

$$N_3 \equiv (6^3 (3) + 7 (\frac{6^3 - 1}{6 - 1})) \pmod 5$$

$$N_3 \equiv 949 \pmod 5$$

$$N_3 = 4$$

$$N_4 \equiv (6^4 (3) + 7 (\frac{6^4 - 1}{6 - 1})) \pmod 5$$

$$N_4 \equiv 5701 \pmod 5$$

$$N_4 = 1$$

Now we are ready to prove (2) by mathematical induction.

(1) Show true for $i = 1$

$$N_1 \equiv a N_0 + c \frac{(a^1 - 1)}{a-1} \pmod m$$

$$N_1 \equiv a N_0 + c \pmod m \quad \cdot \text{ (given)}$$

Assume true for i

(2) $$N_i \equiv a^i N_0 + c \frac{(a^i - 1)}{a-1} \pmod m$$

(3) Must show true for (i + 1)

Use $N_i \equiv a N_{i+1} + c \pmod m$

$$\begin{aligned}
N_{i+1} &\equiv a \left[a^{i} N_0 + c \ \frac{(a^{i} - 1)}{a-1} \right] + c \pmod m \\
&\equiv a^{i+1} N_0 + ac \frac{(a^{i} - 1)}{a-1} + c \pmod m \\
&\equiv a^{i+1} N_0 + c \frac{(a^{i+1} - a)}{a-1} + \frac{c\,(a-1)}{(a-1)} \pmod m \\
&\equiv a^{i+1} N_0 + c \frac{(a^{i+1} - a)}{a-1} + \frac{ca - c}{(a-1)} \pmod m \\
&\equiv a^{i+1} N_0 + \frac{ca^{i+1} - ca + ca - c}{a-1} \pmod m \\
&\equiv a^{i+1} N_0 + c \frac{(a^{i+1} - 1)}{a-1} \pmod m
\end{aligned}$$

EXERCISES 4.3

1) Prove using induction that $1 + 2 + \ldots + n = \frac{n(n+1)}{2}$

2) Prove by induction that if a set x contains n elements, then x has 2^n subsets (counting itself and the empty set).

3) Show by mathematical induction that $\sqrt{n} < \sqrt[n]{n!}$, n = 1, 2, 3, ..., e.g., $\sqrt{10} \cong 3.16$

$$\sqrt[10]{10!} = (3628800)^{1/10} \cong 4.53$$

$$\sqrt{5} = \cong 2.24$$

$$\sqrt[5]{5!} = (120)^{1/5} \cong 2.60$$

4) *Strong induction* is a different form for the same process as regular induction. It has only two steps:

   1) Prove the statement true for $n = 1$.

   2) Prove that if the statement is true for $n = 1, n = 2, \ldots n = k$, then the statement is true for $n = k + 1$.

   Show using strong induction that if s_n denotes the nth Fibonacci number, then $s_n \leq \frac{(1+\sqrt{5})^{n-1}}{2}$, $n = 1, 2, 3, \ldots$ The Fibonacci numbers are defined by the sequence,

   $$s_{n+1} = s_n + s_{n-1}$$ or by the sequence 1, 1, 2, 3, 5, 8, 13, 21, . . .

   [Hint: Use the fact that $1 \leq \frac{1+\sqrt{5}}{2}$ and observe that $\frac{1+\sqrt{5}}{2}$ is a solution of the quadratic equation $y^2 = y + 1$.]

5) Prove the formula for the sum of a geometric series by induction:

$$1 + r + r^2 + \ldots + r^n = \sum_{i=0}^{n} r^i = \frac{1 - r^{n+1}}{1 - r}$$

6) Show using induction that:

a) $2^n > 2n + 1 \quad \forall\, n \geq 3$

b) $2^n \geq n + 1 \quad \forall\, n = 1, 2, 3, \ldots$

4.4 Tests of Randomness

There are virtually an unlimited number of tests that could be used to examine randomness. For example, if we thought $\sqrt{2}$ were random, we could test whether 10% of its decimal expansion (to perhaps 10,000 places) were each of the digits $0, 1, 2, \ldots, 9$. If we found that $\sqrt{2}$ were uniform in the distribution of these ten digits for a sample of 10,000 places, we could then ask to sample 100,000, then a million, and so on. As a result of the difficulty (if not impossibility) of achieving perfect randomness, one researcher in random number generation described his quest as "mystical." We present two standard tests for randomness - a chi-square test for uniformity and a test for correlation.

Chi-Square Test for Uniformity

Consider random numbers $x_1, x_2, \ldots, x_n$ that we hope are uniformly distributed on $[0, 1]$. Divide $[0, 1]$ into at least 100 intervals and take a large sample of x_i so that each interval has at least 5, e.g., $n \geq 500$ if we have 100 intervals. Use the chi-square test,

$\chi^2 = \sum \frac{(0 - E)^2}{E}$, $(k - 1)$ df. For large n, χ^2 will have the chi-square distribution. Our null hypothesis is that the x_i are uniformly distributed on $[0, 1]$.

Irrational Numbers

As we discussed, Persi Diaconis neatly dispelled preconceptions of randomness by showing that coin flips, lotteries and radioactive decay were not good candidates for random number generation. To my delight, Dr. Diaconis supported my hypothesis that the infinite decimal expansions of irrational numbers were promising candidated for future random number generation.

To illustrate, suppose that we take $\sqrt{k}$ (an arbitrary irrational, where k is not a perfect square) to 5000 decimal places. To test for uniformity on [0, 1], we divide [0, 1] into 100 partitions, i.e., [0.0, .01], [.01, .02], . . . We count the observed as the number of random numbers x_i in each partition. Since there are 100 partitions - each equally likely - the expected number for each category is 5000 (1/100) = 50. Use the χ^2 formula:

$$\text{computed } \chi^2 = \frac{\sum (O - E)^2}{E}, \ (k - 1) \text{ degrees of freedom, } k = 100.$$

Then we compare the critical χ^2 .05, 99df with our computed χ^2 value. Consider an imaginary computed χ^2 value = 93. Look up critical χ^2 .05, 99 = 124. Since 93 < 124, we conclude with 95% confidence that our random numbers are uniformly distributed. They have passed the first of many tests of randomness.

Next we could generalize the chi-square test of uniformity to several dimensions. We could take pairs of digits of $\sqrt{k}$ and test 2500 pairs for uniformity on the unit square, partitioning the unit square into 100 equal squares with side 1/10. Or we could take triples of $\sqrt{k}$ and generalize to 1000 cubes with side 1/10. The procedure is cumbersome for paper and pencil, but not for a high speed computer.

This makes it obvious that even irrationals won't pass every test for randomness. They have to (or almost certainly have to) fail for some dimension
n = 1, 2, 3, . . . Of course, at the .05 level of significance, we can expect to reject a true null hypothesis 5% of the time.

Law and Kelton (1991) subjected RANDU, a random number generator, to the χ^2 tests for n = 1, 2, and 3. RANDU's generator is defined by:

$$N_i = 65,539 \ N_{i-1} \pmod{2^{31}}$$

They found that the uniformity on [0, 1] and the random square was acceptable. However, the χ^2 value for three dimensions proved unacceptably high, making the RANDU unacceptable for research.*

Correlation

We have developed correlation - one of the most important concepts in statistics - in earlier chapters. In the context of random numbers, it is desirable that the sequence of "pseudo-random" numbers $x_1, x_2, \ldots, x_n$ have zero correlation. We first compute the correlation between successive x_i, i.e., (x_n, x_{n+1}). Next we compute the correlation using covariance, x_i with gaps 2, 3, etc. Lag 2 would look like $(x_1, x_3), (x_2, x_4), (x_3, x_5), \ldots$ Lag 3 would be represented $(x_1, x_4), (x_2, x_5), \ldots$ Lag j would correlate $(x_i, x_{i+j}), (x_{i+j}, x_{i+2j}), \ldots$.

To illustrate, consider the decimal expansion of $1/7 = .142857142857\ldots..$ If we were to consider the autocorrelation of the decimal expansion of 1/7 for the first five digits, lag 2, we would have the following pairs of numbers:

$(x_1, x_3) = (1, 2)$

$(x_2, x_4) = (4, 8)$

$(x_3, x_5) = (2, 5)$

$(x_4, x_6) = (8, 7)$

$(x_5, x_7) = (5, 1)$

This notion of correlating the digits of a decimal expansion with itself is called **autocorrelation**

*Law, A. and Kelton, D. *Simulation Modeling and Analysis*, McGraw Hill, New York, 1991, pp. 445-446.

To test for zero correlation follow the outlined steps (Law and Kelton, 1991). Take random numbers on the interval (0, 1).

1) Define the correlation as $p_j = c_j/c_o$

2) $c_j = \text{cov}(x_i, x_{i+j}) = E(x_i, x_{i+j}) = E(x_i) E(x_{i+j})$

3) $E(\mu_i) = ½$ (uniformity assumption on [0, 1])

$\text{var}(\mu_i) = 1/12$ (variance of uniform distribution)

4) $c_j = E(\mu_i \mu_{i+j}) - (1/2)^2$

5) $c_o = 1/12$

6) $p_j = c_j / c_o = 12 E(\mu_i \mu_{i+j}) - 3$

7) Obtain estimate of $\hat{p}$ by estimating $E(\mu_i \mu_{i+j})$ from $\mu_i, \mu_{i+j}, \mu_{i+2j}, \ldots$

$$P_j = \frac{12}{h+1} \sum_{k=0}^{h} \mu_{1+kj} \, \mu_{1+(k+1)j} - 3 \text{ , where}$$

$$h = \frac{(n-1)}{j} - 1$$

8) $\text{var}(\hat{p}_j) = \dfrac{13h + 7}{(h+1)^2}$ [Banks and Carson, 1984]

9) Under H_o: $P_j = 0$, for large n

$$A_j = \frac{\hat{p}_j}{\sqrt{\text{var}(p_j)}}$$ has a standard normal distribution

and we reject H_o if $| A_j | > z_{1-\alpha/2}$

10) Carry out the test for several values of j - 1, 2, 3 as a minimum. Recall that RANDU passed χ^2 uniformity tests for j = 1 and 2, but failed miserably for j = 3.

To illustrate the computations associated with "pseudo-random number" generator's test of covariance, consider the RNG that we introduced earlier in the chapter:

$$N_i \equiv (6\,N_{i-1} + 7) \pmod 5$$

The generator is terrible, but the computations associated with our simplified covariance test are clear and rather easy. The "random numbers" are 0, 2, 4, 1, 3, 0, 2, 4, 1, 3, . . . Our pseudo-random numbers on [0, 1], obtained by dividing by 5, are 0, 2/5, 4/5, 1/5, 3/5, 0, 2/5, 4/5, 1/5, 3/5, . . . Consider the computations associated with lag $j = 2$. Let $n = 21$.* The value $h = \frac{21 - 1}{2} - 1 = 9$. We have already computed $\mu_1, \mu_2, \mu_3, \ldots = 0, 2/5, 4/5, 1/5, 3/5, 0, 2/5, 4/5, 1/5, 3/5, \ldots$ From the formula from step (7).

$$\hat{p}_j = 12/10 \sum_{k=0}^{9} \mu_{1+2k}\,\mu_{1+(k+1)2} - 3$$

Table

$\mu_1 = 0$

$\mu_2 = 2/5$

$\mu_3 = 4/5$

$\mu_4 = 1/5$

$\mu_5 = 3/5$

$\mu_6 = 0$

$\mu_7 = 2/5$

$\mu_8 = 4/5$

$\mu_9 = 1/5$

$\mu_{10} = 3/5$

$\mu_{11} = 0$

$\mu_{12} = 2/5$

$\mu_{13} = 4/5$

$\mu_{14} = 1/5$

$\mu_{15} = 3/5$

$\mu_{16} = 0$

$\mu_{17} = 2/5$

$\mu_{18} = 4/5$

$\mu_{19} = 1/5$

$\mu_{20} = 3/5$

$\mu_{21} = 0$

*[In practice, n should be several thousand and all statistics should be calculated by computer.]

$$\hat{p}_j = 6/5 \; [\overset{k=0}{\mu_1\mu_3} + \overset{k=1}{\mu_3\mu_5} + \overset{k=2}{\mu_5\mu_7} + \overset{k=3}{\mu_7\mu_9} + \ldots + \overset{k=9}{\mu_{19}\mu_{21}}\,] - 3$$

$$\hat{p}_j = 6/5\,[0(4/5) + 4/5(3/5) + 3/5(2/5) + 2/5(1/5) + (1/5)0 + 0(4/5) +$$

$$4/5(3/5) + (3/5)2/5 + 2/5(1/5) + (1/5)0\,] - 3$$

$$\hat{p}_j = 6/5\,[40/25] - 3 = -1.08$$

$$\text{var}\,(\hat{p}_j) = \frac{13\,(9) + 7}{(10)^2} = 1.24$$

$$A_j = \hat{p}_j \;/\; \sqrt{\text{var}\,(p_j)} = -1.08/\sqrt{1.24} = -1.08/1.11 = -.97$$

We next look up in the table ($\alpha = .05$) critical $z_{1-\alpha/2} = 1.96$. Since 1 - .971 = .97 < 1.96, we accept H_o, namely that there is not sufficient evidence to reject a zero (lag 2) correlation. Of course, for a more suitable example the computer would compute for large n all of the statistics that we calculated by hand. This example simply was chosen for instructive purposes with values that could be computed easily with a calculator.

EXERCISES 4.4

1) A possible candidate for randomness produced the following frequency of digits 0, 1, 2, ..., 9. Perform a chi-square test of uniformity.

0	1	2	3	4	5	6	7	8	9
502	497	499	521	488	531	497	503	456	996

2) Define $N_j \equiv 9\,N_{i-1} \pmod{16}$, $N_o = 3$. Take 1000 random numbers $N_o, N_1, \ldots N_{999}$. Draw a table of frequencies like the above chart and list the frequencies 0, 1, 2, . . . , 9. Perform a chi-square goodness of fit test. For $n = 20$, obtain the correlation r between a_n and a_{n+1}. What can we conclude about N_i?

3) For a "pseudo-random number" generator, lag j, we estimated $\hat{p}_j = .1574$ and observed $h = 1665$. At the $\alpha = .05$ level, can we conclude that the autocorrelation is zero?

4) For the pseudo-random number in example 2, perform the correlation test for $n = 50$, lag 2.