One way to visualize the period of the random number generator is to use it to generate a plot of the displacement x of a random walker as a function of the number of steps N. When the period of the random number is reached, the plot will begin to repeat itself. Generate such a plot using (7.59) for $a = 899$, $c = 0$, and $m = 32768$, and for $a = 16807$, $c = 0$, and $m = 2^{31} - 1$ with $x_0 = 12$. What are the periods of the corresponding random number generators? Obtain similar plots using different values for the parameters a, c, and m. Why is the seed value $x_0 = 0$ forbidden for the choice $c = 0$? Do some combinations of a, c, and m give longer periods than others?

(b) *Uniformity*. A random number sequence should contain numbers distributed in the unit interval with equal probability. The simplest test of uniformity is to divide this interval into M equal size subintervals or bins. For example, consider the first $N = 10^4$ numbers generated by (7.59) with $a = 106$, $c = 1283$, and $m = 6075$ (see Press et al.). Place each number into one of $M = 100$ bins. Is the number of entries in each bin approximately equal? What happens if you increase N?

(c) *Chi-square test*. Is the distribution of numbers in the bins of part (b) consistent with the laws of statistics? The most common test of this consistency is the *chi-square* or χ^2 test. Let y_i be the observed number in bin i and E_i be the expected value. The chi-square statistic is

$$\chi^2 = \sum_{i=1}^{M} \frac{(y_i - E_i)^2}{E_i}. \tag{7.61}$$

For the example in part (b) with $N = 10^4$ and $M = 100$, we have $E_i = 100$. The magnitude of the number χ^2 is a measure of the agreement between the observed and expected distributions; χ^2 should not be too big or too small. In general, the individual terms in the sum (7.61) are expected to be order one, and because there are M terms in the sum, we expect $\chi^2 \leq M$. As an example, we did five independent runs of a random number generator with $N = 10^4$ and $M = 100$ and found $\chi^2 \approx 92$, 124, 85, 91, and 99. These values of χ^2 are consistent with this expectation. Although we usually want χ^2 to be as small as possible, we would be suspicious if $\chi^2 \approx 0$, because such a small value suggests that N is a multiple of the period of the generator and that each value in the sequence appears an equal number of times.

(d) *Filling sites*. Although a random number sequence might be distributed in the unit interval with equal probability, the consecutive numbers might be correlated in some way. One test of this correlation is to fill a square lattice of L^2 sites at random. Consider an array $n(x, y)$ that is initially empty, where $1 \leq x_i, y_i \leq L$. A site is selected randomly by choosing its two coordinates x_i and y_i from two consecutive numbers in the sequence. If the site is empty, it is filled and $n(x_i, y_i) = 1$; otherwise it is not changed. This procedure is repeated t times, where t is the number of Monte Carlo steps per site. That is, the time is increased by $1/L^2$ each time a pair of random numbers is generated. Because this process is analogous to the decay of radioactive nuclei, we expect that the fraction of empty lattice sites should decay as e^{-t}. Determine the fraction of unfilled sites using the random number generator that you have been using for $L = 10$, 15, and 20. Are your results consistent with the expected fraction? Repeat the same test using (7.59) with $a = 65{,}549$, $c = 0$, and $m = 231$. The existence of triplet correlations can be determined by a similar test on a simple cubic lattice by choosing the three coordinates x_i, y_i, and z_i from three consecutive random numbers.

(e) *Parking lot test*. Fill sites as in part (d) and draw the sites that have been filled. Do the filled sites look random, or are there stripes of filled sites? Try $a = 65{,}549$, $c = 0$, and $m = 231$.

(f) *Hidden correlations*. Another way of checking for correlations is to plot x_{i+k} versus x_i. If there are any obvious patterns in the plot, then there is something wrong with the generator. Use the generator (7.59) with $a = 16{,}807$, $c = 0$, and $m = 2^{31} - 1$. Can you detect any structure in the plotted points for $k = 1$ to $k = 5$? Test the random number generator that you have been using. Do you see any evidence of lattice structure, for example, equidistant parallel lines? Is the logistic map $x_{n+1} = 4x_n(1 - x_n)$ a suitable random number generator?

(g) *Short-term correlations*. Another measure of short term correlations is the autocorrelation function

$$C(k) = \frac{\langle x_{i+k} x_i \rangle - \langle x_i \rangle^2}{\langle x_i x_i \rangle - \langle x_i \rangle \langle x_i \rangle}, \tag{7.62}$$

where x_i is the ith term in the sequence. We have used the fact that $\langle x_{i+k} \rangle = \langle x_i \rangle$; that is, the choice of the origin of the sequence is irrelevant. The quantity $\langle x_{i+k} x_i \rangle$ is found for a particular choice of k by forming all the possible products of $x_{i+k} x_i$ and dividing by the number of products. If x_{i+k} and x_i are not correlated, then $\langle x_{i+k} x_i \rangle = \langle x_{i+k} \rangle \langle x_i \rangle$ and $C(k) = 0$. Is $C(k)$ identically zero for any finite sequence? Compute $C(k)$ for $a = 106$, $c = 1283$, and $m = 6075$.

(h) *Random walk*. A test based on the properties of random walks has been proposed by Vattulainen et al. Assume that a walker begins at the origin of the xy-plane and walks for N steps. Average over M walkers and count the number of walks that end in each quadrant q_i. Use the χ^2 test (7.61) with $y_i \to q_i$, $M = 4$, and $E_i = M/4$. If $\chi^2 > 7.815$ (a 5% probability if the random number generator is perfect), we say that the run fails. The random number generator fails if two out of three independent runs fail. The probability of a perfect generator failing two out of three runs is approximately $3 \times 0.95 \times (0.05)^2 \approx 0.007$. Test several random number generators. ■

Problem 7.36 Improving random number generators

One way to reduce sequential correlation and to lengthen the period is to mix or *shuffle* the random numbers produced by a random number generator. A standard procedure is to begin with a list of N random numbers (between 0 and 1) using a given generator `rng`. The number N is arbitrary but should be less than the period of `rng`. Also generate one more random number r_{extra}. Then for each desired random number, use the following procedure:

(i) Calculate the integer k given by `(int)(N*r`$_{\text{extra}}$`)`. Use the kth random number r_k from your list as the desired random number.
(ii) Set r_{extra} equal to the random number r_k chosen in step (i).
(iii) Generate a new random number r from `rng` and use it to replace the number chosen in step (i), that is, $r_k = r$.