Designation: D 5124 – 96

Standard Practice for Testing and Use of a Random Number Generator in Lumber and Wood Products Simulation[1]

This standard is issued under the fixed designation D 5124; the number immediately following the designation indicates the year of original adoption or, in the case of revision, the year of last revision. A number in parentheses indicates the year of last reapproval. A superscript epsilon (ϵ) indicates an editorial change since the last revision or reapproval.

1. Scope

1.1 This practice gives a minimum testing procedure of computer generation routines for the standard uniform distribution. Random observations from the standard uniform distribution, R_U, range from zero to one with every value between zero and one having an equal chance of occurrence.

1.2 The tests described in this practice only support the basic use of random number generators, not their use in complex or extremely precise simulations.

1.3 Simulation details for the normal, lognormal, 2-parameter Weibull and 3-parameter Weibull probability distributions are presented.

1.4 *This standard does not purport to address all of the safety concerns, if any, associated with its use. It is the responsibility of the user of this standard to establish appropriate safety and health practices and determine the applicability of regulatory limitations prior to use.* See specific warning statement in 5.5.3.

2. Referenced Documents

2.1 *ASTM Standards:*

E 456 Terminology Relating to Quality and Statistics[2]

3. Terminology

3.1 *Definitions:*

3.1.1 *period*—the number of R_U deviates the computer generates before the sequence is repeated.

3.1.2 *seed value*—a number required to start the computer generation of random numbers. Depending upon the computer system, the seed value is internally provided or it must be user specified. Consult the documentation for the specific random number generator used.

3.1.3 *serial correlation*—the statistical correlation between ordered observations. See 5.2.2.

3.1.4 *standard normal deviate,* R_N —a computer generated random observation from the normal probability distribution having a mean equal to zero and standard deviation equal to one.

3.1.5 *standard uniform deviate,* R_U —a random observation from the *standard uniform distribution.*

3.1.6 *standard uniform distribution*—the probability distribution defined on the interval 0 to 1, with every value between 0 and 1 having an equal chance of occurrence.

3.1.7 *trial*—a computer experiment, and in this standard the generation and statistical test of one set of random numbers.

4. Significance and Use

4.1 Computer simulation is known to be a very powerful analytical tool for both practitioners and researchers in the area of wood products and their applications in structural engineering. Complex structural systems can be analyzed by computer with the computer generating the system components, given the probability distribution of each component. Frequently the components are single boards for which a compatible set of strength and stiffness properties are needed. However, the entire structural simulation process is dependent upon the adequacy of the standard uniform number generator required to generate random observations from prescribed probability distribution functions.

4.2 The technological capabilities and wide availability of microcomputers has encouraged their increased use for simulation studies. Tests of random number generators in commonly available microcomputers have disclosed serious deficiencies **(1)**.[3] Adequacy may be a function of intended end-use. This practice is concerned with generation of sets of random numbers, as may be required for simulations of large populations of material properties for simulation of complex structures. For more demanding applications, the use of packaged and pretested random number generators is encouraged.

5. Uniformity of Generated Numbers

5.1 *Test of the Mean*—The mean of the standard uniform distribution is ½. Generate 100 sets of 1000 random uniform numbers and conduct the following statistical test on each set.

$$Z = \frac{\bar{X} - 0.50}{0.009129} \tag{1}$$

where:

Z = test statistic,

[1] This practice is under the jurisdiction of ASTM Committee D-7 on Wood and is the direct responsibility of Subcommittee D07.05 on Wood Assemblies.

Current edition approved April 10, 1996. Published June 1996. Originally published as D 5124 – 91. Last previous edition D 5124 – 91.

[2] *Annual Book of ASTM Standards*, Vol 14.02.

[3] The boldface numbers in parentheses refer to the list of references at the end of this standard.

$\bar{X} = \Sigma R_U / 1000$,

the standard deviation is assumed to be $\sqrt{1/12}$, and the summation over 1000 values is implied.

If the absolute value of Z exceeds 1.28 for more than 10 % and less than 30 % of the trials, the random number generator passes. If the random number generator fails the test using 100 sets, then the number of sets can be increased or the random number generator can be rejected.

NOTE 1—The assumption of standard deviation being equal to $\sqrt{1/12}$ may be examined with a Chi-Square test where

$$s = \sqrt{\frac{(\Sigma R_U^{\,2} - 1000\,\bar{X}^2)}{999}} \quad (2)$$

where:

$\bar{X}$ = estimated mean

s = estimated standard deviation of the 1000 R_U values, and

the summation over 1000 values is implied.

A significant difference between s and $\sqrt{1/12}$, suggests a non-random generator.

5.2 *Test for Patterns in Pairs*—The purpose of this visual test is to evaluate the tendency of pairs of deviates to form patterns when plotted. Generate 2000 pairs of standard uniform deviates. Plot each pair of deviates on an x-y Cartesian coordinate system. Inspect the resulting plot for signs of patterns, such as "strips." Fig. 1 is one example of "stripes" generated by a BASIC function on a personal computer. In more than two dimensions, all generated random numbers fall mainly on parallel hyperplanes, a fact discovered by Marsaglia **(3)**.

5.2.1 The following shuffling technique is an effective remedy for the general problem of "stripes" and random numbers falling on planes. Fill a 100-element array with standard uniform deviates. Select a deviate from the array using the integer portion of the product of a random deviate and 100. Replace the selected deviate with a new uniform deviate. Repeat the process until the desired number of deviates has been generated. The plot of Fig. 2 resulted from using the shuffling technique on the random number generator which produced Fig. 1.

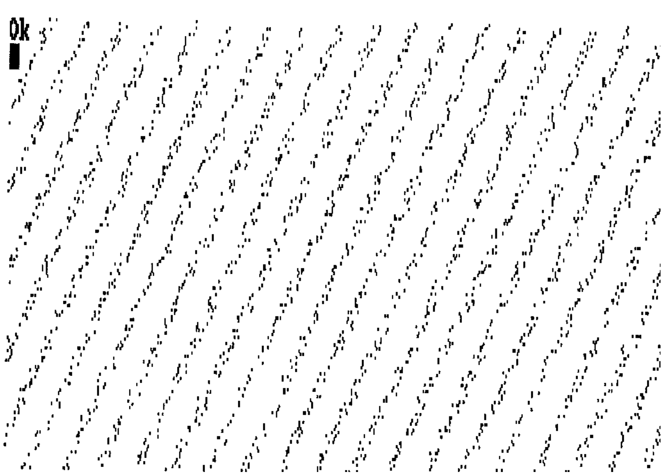
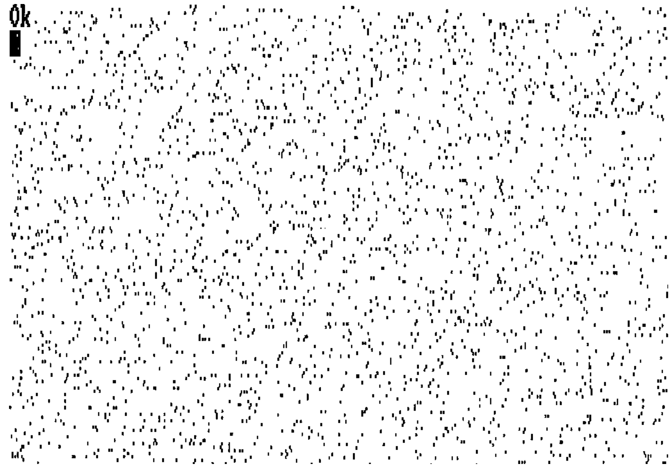


FIG. 1 Plotted Pairs of Random Numbers Showing "Stripes"



NOTE 1—The plot resulted from using the shuffling technique on the generator which produced Fig. 1.

FIG. 2 Plotted Pairs of Random Numbers with no Detectable Patterns

5.2.2 Unless the R_U generator is extensively tested by stringent tests **(4, 5, 6)** a shuffling procedure comparable to that described in 5.2.5 should be used.

5.3 *Visual Test for Uniform Distribution Conformance*:

5.3.1 The purpose of the visual test for distribution conformance is to detect some odd behavior of the random number generator beyond what might be detected by the method in 5.4. It is impossible to predict the various shapes of the histograms which might indicate a problem with the generator. However, a few examples given here may alert the user of the general form of a problem.

5.3.2 *Histogram Preparation*—Fig. 3 is a histogram of 1000 generated standard uniform numbers. The theoretical density function is a horizontal dashed line crossing the ordinate at 1.0. The interval width is 0.1. The values of the ordinates for each interval were calculated as follows:

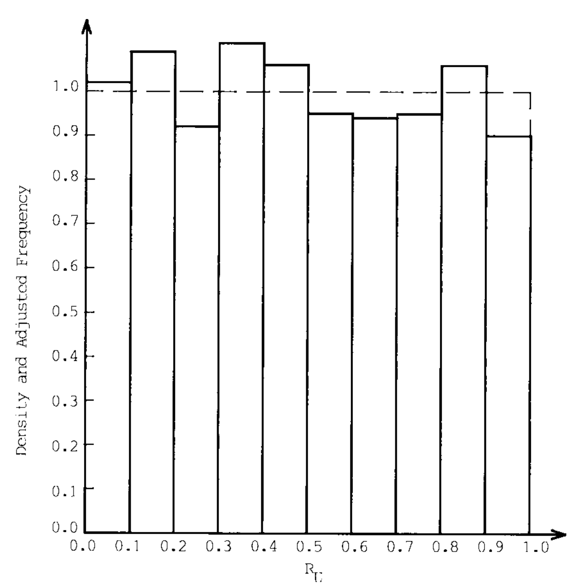


FIG. 3 Histogram of Random Numbers with Theoretical Density Function Superimposed

$$f_i = \frac{N_i}{W_I \times T} \quad (3)$$

where:

f_i = adjusted relative frequency,
N_i = number observed in interval i,
W_I = interval width, and
T = total number generated.

Since the interval width, W_I, in this case equalled 0.1 and 1000, values were generated as follows:

$$f_i = \frac{N_i}{0.1 \times 1000} \quad (4)$$

$$f_i = \frac{N_i}{100}$$

NOTE 2—If different sample sizes are used, bias may exist in making visual interpretations from histograms. One way to lessen this bias is to apply the Sturgess Rule **(7)** to determine the number of cells for the histograms.

$$N_c = 1 + 3.3(\log_{10} N_g) \quad (5)$$

where:

N_c = number of histogram cells, and
N_g = number of generated numbers.

5.3.3 *Histogram Evaluation*—The histogram of Fig. 3 has a very typical appearance for a sample as large as 1000. If one would increase the sample size, less variation in f_i is expected. On the contrary, by decreasing the sample size to perhaps 50, tremendous variation in f_i can be expected. A problem would be evidenced, if for a sample size of 1000, one of the following occurs: (*1*) if f_i equalled zero or near zero for one class interval, (*2*) if one class interval had an f_i value 50 % greater than any other interval, or (*3*) if there is any noticeable trend in the f_i value such as an increase in f_i from left to right, a decrease, or whatever. The f_i values should vary about 1.0 in a random fashion. The data must span the entire range from 0 to 1.

5.4 *Formal Test of Distribution Conformance*:

5.4.1 The Kolmogorov-Smirnov (KS) goodness-of-fit test given in Ref **(5)** should be used to test the conformance of the random numbers to the standard uniform distribution. The KS test should be conducted on 100 sets of generated random number data each containing 1000 observations.

5.4.2 *Kolmogorov-Smirnov Test*—Generate the R_U numbers and store in an array. Rank the data from smallest to largest. Calculate the following:

$$D_n^+ = \max\left[\frac{i}{N} - X_i\right] \quad (i = 1, N) \quad (6)$$

$$D_n^- = \max\left[X_i - \frac{i-1}{N}\right] \quad i = 1, N$$

$$D_n = \max[D_n^+, D_n^-]$$

where:

N = sample size, (1000),
X_i = i^{th} value of the ranked array, and
D_n = Kolmogorov-Smirnov (K-S) test statistic.

For the test in 5.4, N equals 1000. X_1 is the smallest value of the ranked array, X_2 is the second smallest and so on. D_n as calculated is the largest vertical distance between the sample density function and the hypothesized distribution, in this case the standard uniform distribution. If D_n is greater than ($1.07/\sqrt{N}$) for more than 10 % and less than 30 % of the trials, the random number generator passes. If the generator fails the tests using 100 sets, then the number of sets can be increased or the generator can be rejected.

5.5 *Correlations Among Generated Numbers*:

5.5.1 The computer generated values of R_U must appear to be random and independent. The word "appear" is used since the numbers are actually being generated by a mathematical algorithm and all such algorithms have a cycle. Provided the numbers have the appropriate distribution function (as tested in 5.3 and 5.4) and the numbers are not serially correlated, then the generated numbers are most useful for simulation purposes. Since the generated numbers are not truly random they are often called "pseudo random."

5.5.2 *Period*—Some personal computer brands have a uniform number generator with an extremely short period depending upon the seed. Some machines repeat the same sequence of numbers after approximately 200 numbers. Depending upon the simulation application, the user must determine if the period of the machine is adequate. Reference **(1)** is useful for evaluating the period of various random number generators.

5.5.3 *Test for Lag-1 Serial Correlation*—Lag-1 serial correlation is a measure of association between the X_i observation and the following X_{i+1}. Lag-2 serial correlation is a measure of association between X_i and X_{i+2} or all pairs of observations separated by one observation. In theory, it is possible to have any lag-k serial correlation. For random number generators, it is necessary for all lag-k to be zero for k less than the period. For k equal the period, lag-k serial correlation equals 1.0. The following statistical test from Ref. **(2)** is for lag-1 serial correlation and it is recommended as a minimum test for statistical independence.

NOTE 3—**Warning:** Random number generators that pass the tests in this standard can display very bad behavior in more than two dimensions. There are existing random number generators that can pass the tests in this standard but whose values fall on a small number of hyperplanes.

5.5.3.1 Let X_i be an array of generated R_U values. X_1 being the first generated, X_2 the second and so on. Generate 1000 values of X_i. Calculate:

$$r(1) = \frac{\Sigma X_i X_{i+1} - (\Sigma X_i)^2/1000}{\Sigma X_i^2 - (\Sigma X_i)^2/1000} \quad (7)$$

where:

r (1) = lag one serial correlation, and Σ denotes an implied summation from 1 to 1000.

X_{i+1} must be replaced by X_1 when i equals 1000; take $X_{1001} = X_1$. If the calculated r (1) falls outside of the following limits for $-0.042 < r(1) < 0.040$ more than 10 % and less than 30 % of the trials, the random number generator passes. If the random number generator fails the test using 100 sets, then the number of sets can be increased or the random number generator can be rejected. (Assuming there is no lag-1 serial correlation, 20 % of the calculated r (1) values would be expected to fall outside of the specified range as the number trials increased indefinitely.)

NOTE 4—Serial correlations greater than lag-1 may affect modeling procedures. It is the responsibility of the investigator to assess, in an appropriate manner, the significance of these correlations.

5.6 *Selection of R_U Generator*—Provided the R_U generator passes the tests and provisions in 5.1-5.4, it can be considered useful for purposes of computer simulation. The tests in 5.1-5.4 are considered as minimum for qualification; an individual user may want to increase the number of trials.

5.7 *Rejection of R_U Generator*—In the tests of 5.1-5.4, there is a chance of falsely rejecting a good generator. For this reason, one may choose to repeat all tests (using different seed values) if a given generator failed on the first series of tests.

5.8 *Generation of R_U* —In BASIC programs the generator may produce different results depending on whether the program is compiled or interpreted. On some systems considerable differences have been observed, between the modes, because of differences in how the generator is seeded. In any case, the results from both methods of program execution should be checked when using BASIC. For a comprehensive discussion on the various methods of generating R_U, Chapter 6 of Ref **(5)** is recommended.

6. Simulation from Selected Distributions

6.1 Random values from a prescribed distribution will be noted by y' which is often referred to as a random deviate. This section assumes that the parameters of the various probability distributions have been estimated by the various methods available and are now known quantities.

6.2 *Simulation from the Normal Distribution*—In general, when simulating lumber and wood product properties from the normal distribution, truncation is required since the normal distribution is defined from minus infinity to plus infinity. With simulation it is possible to generate extremely small or negative values. Therefore it is the responsibility of the user to discard all values below a user specified minimum. The definition of the minimum is a difficult problem. A normal distribution is not usually preferred over other distributions because of the truncation issue.

6.2.1 *Normal Density Function*—The normal density function is given by:

$$f_x(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \quad (8)$$

where:
μ = mean of x, and
σ = standard deviation.

6.2.2 *Simulation from Normal Distribution*—Observations from the normal distribution are given by:

$$y' = \mu + \sigma \times R_N \quad (9)$$

where:
y' = values less than a specified truncation point are discarded, a new value of y' should be calculated.
R_N = standard normal deviate which can be calculated from generated R_U observations. Using two values of R_U, denoted R_{U1} and R_{U2}, then by the following relationships:

$$R_{N1} = \cos(2\pi R_{U2}) \times \sqrt{-2L_n(R_{U1})} \quad (10)$$

$$R_{N2} = \sin(2\pi R_{U2}) \times \sqrt{-2L_n(R_{U1})}$$

A pair of statistically independent standard normal variates R_{N1} and R_{N2} result **(8)**. L_n is the natural logarithm. While one could be discarded, it would not be a wise use of computational time. It is recommended that the two values of R_N be calculated in one step, two more, and so on until the necessary number is obtained.

6.3 *Lognormal Distribution*—The lognormal density function is given by:

$$f_x(x) = \frac{1}{\xi x\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{L_n(x)-\lambda}{\xi}\right)^2\right] \quad (11)$$

where:
L_n = natural logarithm,
λ = mean of the logarithms, and
ξ = standard deviation of the logarithms.

6.3.1 *Simulation from the Lognormal Distribution*—Observations from lognormal distribution are given by:

$$y' = \exp(\lambda + R_N * \xi) \quad (12)$$

where:
R_N is calculated as in 6.2.2.

6.4 *The 2-Parameter Weibull*—The 2-parameter Weibull density function is given by:

$$f_x(x) = \frac{\eta}{\sigma}\left(\frac{x}{\sigma}\right)^{\eta-1} \exp\left(-\left(\frac{x}{\sigma}\right)^{\eta}\right) \quad (13)$$

where:
σ = scale parameter and
η = shape parameter.

6.4.1 *Simulation from the 2-Parameter Weibull*—Observations from the 2-parameter Weibull are given by:

$$y' = \sigma(-L_n(RU))^{1/\eta} \quad (14)$$

where:
L_n = natural logarithm.

6.5 *The 3-Parameter Weibull*—The 3-parameter Weibull density function is given by:

$$f_x(x) = \frac{\eta}{\sigma}\left(\frac{x-\mu}{\sigma}\right)^{\eta-1} \exp\left(-\left(\frac{x-\mu}{\sigma}\right)^{\eta}\right) \quad (15)$$

where:
σ = scale parameter,
η = shape parameter, and
μ = location parameter.

6.5.1 *Simulation from the 3-Parameter Weibull*—Observations from the 3-parameter Weibull are given by:

$$y' = \mu + \sigma(-L_n(RU))^{1/\eta} \quad (16)$$

where:
L_n = natural logarithm.

6.6 Random observations from other continuous probability distributions can be generated by the methods of 5.2 in Ref **(8)** or Chapter 7 of Ref **(5)**.

REFERENCES

(**1**) Modianos, D. T., Scott, R. C., and Cornwell, L. W. "Random Number Generation on Microcomputers," *Interfaces* 14(4): 81-87. 1984.

(**2**) Haan, C. T. *Statistical Methods in Hydrology*, Iowa State Press, Ames, Iowa, 1977, p. 378.

(**3**) Marsaglia, G. "Random Numbers Fall Mainly in the Planes," National Academy of Sciences, 1968 Proceedings 61: 25-28.

(**4**) Kennedy, Wm. J. Jr., and Gentle, James E., *Statistical Computing*, Statistics: Textbooks Monograph Series, Vol 33, Section 6.6, Marcel Dekker, Inc., NY, 1980.

(**5**) Law, A. M., and Kelton, W. D., *Simulation Modeling and Analysis*, McGraw Hill, New York, NY., 2nd ed., 1991, p. 759.

(**6**) Marsaglia, G. "A Current View of Random Number Generators," *Computer Science and Statistics: Proceedings of the Sixteenth Symposium on the Interface*, L. Billard, ed., North Holland, 1985. p 3 to 10.

(**7**) King, James R., *Probability Charts for Decision Making,* Team, Tamworth, NH, 1981, Chapter 2, p. 7.

(**8**) Ang, A. H-S, and Tang, W. H.," Probability Concepts in Engineering Planning and Design," Volume II *Decision, Risk and Reliability*, John Wiley & Sons, Inc., New York. 1984, p. 562.