SIMULATING RELIABILITY DISTRIBUTIONS IN APL

David C. Trindade
IBM General Technology Division
Essex Junction, Vermont 05452

Abstract

The theory and methods for generating in APL (pseudo-) random numbers from the uniform, normal, lognormal, exponential, and Weibull distributions are presented. Procedures for comparing simulation results to parent populations are discussed. Applications to reliability studies are emphasized.

1. INTRODUCTION

The ability to simulate random numbers from any population is a very valuable tool for the analyst. For the reliability engineer in particular, simulation provides a powerful means to verify the adequacy and sensitivity of the experimental design and methods of analysis. In areas where no theory exists, simulation may provide the only means of investigating and understanding the fundamental reliability concerns [6].

APL is well suited for simulation studies because of the flexibility, power, and simplicity of this language. This paper presents methods for generating random numbers in APL from each of the following distributions: uniform, normal, lognormal, exponential, and Weibull. Some familiarity with APL notation and some simple statistical concepts is assumed. The distinction between a population and a sample should be understood.

2. RANDOM NUMBERS

To generate any distribution one needs a source of random numbers. In APL random numbers are obtained by use of the monadic "roll" function *?X*, where *X* is any integer number. The numbers are called pseudo-random because a specific algorithm [3] exists for their calculation. *?X* produces a pseudo-random integer between the values 1 (or 0 depending on the index origin ⎕*IO*) and *X*. We assume that ⎕*IO*←1 in all procedures that follow. If *X* is a vector, then *?X* gives a vector of random numbers, each number an integer between 1 and *X*.

3. UNIFORM DISTRIBUTION

The uniform distribution is a continuous distribution with probability density function for the random variable T given by [5] $f(t) = 1/(\theta_2 - \theta_1)$, $\theta_1 \le t \le \theta_2$, and zero elsewhere, where θ_1 and θ_2 are the parameters specifying the range of T. The rectangular shape of this distribution is shown in Figure 1. We note that f(t) is constant between θ_1 and θ_2. The cumulative distribution function (CDF) of T, denoted by F(t), for the uniform case is given by $F(t) = (t - \theta_1)/(\theta_2 - \theta_1)$. Thus, F(t) is linear in t in the range $\theta_1 \le t \le \theta_2$. The uniform distribution has mean $(\theta_1 + \theta_2)/2$ and variance $(\theta_2 - \theta_1)^2/12$.[5]

Since the uniform distribution is continuous and the random numbers arising from the roll function are discrete, an approximation of a continuous variable by a discrete set of numbers is necessary. Fortunately the approximation can be made quite precise.

To illustrate with an example which we shall use later, let us generate N random numbers, where N is any integer greater than zero, in the interval zero to one, that is (0,1). Let us further require that the fineness of the discrete separation of the numbers generated be 10^{-8}. Then, the APL command to accomplish these requirements is:

```
U←(?Nρ1E8)÷1E8  .
```

To obtain a set of numbers uniformly distributed in the interval (θ_1, θ_2), one need only multiply *U* by the range $(\theta_2 - \theta_1)$ and add θ_1. For example, the expression

```
TU←10+40×U
```

would generate a set of random numbers uniformly distributed in the range 10 to 50. The numbers so obtained are unordered.

Presented in Hilo, Hawaii and Published in the Proceedings of the American Institute for Decision Sciences, 1980

4. NORMAL DISTRIBUTION

The normal distribution is a continuous symmetric distribution with density function [5]

$$f(t) = (\sigma\sqrt{2\pi})^{-1}\exp-\{(t-\mu)^2/2\sigma^2\},$$

where $\sigma>0$, $-\infty<\mu<\infty$, $-\infty<t<\infty$. Here, μ and σ are the mean and standard deviation respectively of the normal distribution. A graph of the normal density curve is shown in Figure 2. The cumulative distribution function cannot be written in closed form and requires the evaluation of an integral. Fortunately, any normal distribution can be reduced to a standard normal distribution by the transformation [5] $z = (t - \mu)/\sigma$, where z has the standard normal distribution with mean zero and variance one. Thus, the evaluation of the integral can be represented by only one set of tables.

Since time is always positive, the normal distribution should be used only for those situations in which the mean is at least three standard deviation units above zero. This stipulation assures a very low probability (less than 0.0013) of negative values.

The simplest method for the simulation of the normal distribution employs the central limit theorem of statistics [5]. This highly important theorem basically states that for any population with a finite variance σ^2 and mean μ, the distribution of the sample mean approaches the normal distribution with variance σ^2/n and mean μ as the sample size n increases. The sample mean is just the sum of the observations divided by the number of observations k.

Let U have a uniform distribution defined in the interval (0,1); U is called a unit rectangular variate. Then U has mean 1/2 and variance 1/12. According to the central limit theorem, the distribution of means

$$\bar{U} = \left(\sum_1^k U_i\right)/k$$

approaches the normal distribution for large k. Hence, the quantity $(\bar{U}-1/2)/\sqrt{1/12k}$ approaches the standard normal for large k. Generally, it is sufficient [2] to take k=12. Let $Z = (\bar{U} - 1/2)/\sqrt{1/12k}$. Then, any normal variate T with mean μ and standard deviation σ can be obtained from the equation $T = \mu + \sigma Z$. The procedure for generating a normal variate is then quite simple. Noting that Z can be written as

$$Z = \left(\sum_1^k U_i - k/2\right)\sqrt{12/k}\ ,$$

an APL expression to generate *N* pseudo-random numbers from a normal population with mean *M* and standard deviation *S* is, therefore,

```
M+S×((+/(N,K)ρ(?(N×K)ρ1E8)÷1E8)-K÷2)
   ×(12÷K)*0.5  .
```

The value of *K* should be specified as at least 12. However, if values near the tails of the distribution are of primary importance, then higher values of *K* would be advised, e.g., *K* specified as 25.

5. LOGNORMAL DISTRIBUTION

The lognormal distribution is a nonsymmetric continuous distribution that has density function [1]

$$f(t) = (t\sigma\sqrt{2\pi})^{-1}\exp-\{[\ell n(t/m)]^2/2\sigma^2\},$$

in natural base, where $\sigma>0, 0<m<\infty$, and $0<t<\infty$. Here, m and σ are the median and shape parameter respectively of the lognormal distribution. A graph of several lognormal density curves is shown in Figure 3. Note that the lognormal distribution is defined for only positive values of t, a property well suited to reliability work. As with the normal distribution, the lognormal CDF cannot be obtained in closed form. However, if T is the lognormal variate, then $\ell n\ T$ is normally distributed with mean $\mu = \ell n\ m$ and standard deviation σ [4]. Hence, the lognormal CDF can be determined by reference to standard normal tables. The lognormal distribution has mean $m\times\exp\{\sigma^2/2\}$ and variance $m^2w(w-1)$ where $w = \exp\{\sigma^2\}$ [2].

We invoke the relationship of the lognormal distribution to the standard normal distribution to generate lognormal pseudo-random numbers. Thus, if Z has the standard normal distribution (mean 0, standard deviation 1), then $T = m\ \exp\{\sigma Z\}$ has the lognormal distribution [2] with median m and shape parameter σ.

A simple APL program to generate *N* lognormal pseudo-random numbers from a population with median *M* and shape parameter *S* is, therefore,

```
M×*S×((+/(N,K)ρ(?(N×K)ρ1E8)÷1E8)-K÷2)
   ×(12÷K)*0.5  .
```

As for the normal, *K* should be 12 or greater, especially if values in the tails of the distribution are of interest.

6. EXPONENTIAL DISTRIBUTION

The exponential distribution is a continuous distribution with probability density [4] $f(t) = (1/\theta)\exp\{-t/\theta\}$, for

$t>0$, and $\theta>0$. The parameter θ is called the mean life of the exponential distribution. The reciprocal $1/\theta$ is the hazard function [4] for the exponential distribution. Note the hazard function is constant. A graph of the exponential curve for $\theta = 1$ appears in Figure 4. The CDF is given by $F(t) = 1 - \exp\{-t/\theta\}$. The variance of the exponential distribution is also equal to the mean life θ [4].

To simulate the exponential distribution one uses the following property of the cumulative distribution function: If F is <u>any</u> CDF with an inverse F^{-1}, then for U uniformly distributed over (0,1), the substitution of U for F in the inverse expression denoted by $F^{-1}(U)$ will generate a random variable distributed according to F. The importance of the above property is that random observations from any desired distribution can be generated. First, one obtains pseudo-random numbers from the interval (0,1). Next, one applies the inverse transformation of the desired distribution to these uniform numbers. The transformed numbers will then be distributed according to the selected CDF.

For the exponentially distributed variate T, the inverse equation is

$$F^{-1}(U) = -\theta \ell n\ (1 - U).$$

Hence, an APL expression to generate pseudo-random numbers from a population exponentially distributed with mean life *ML* is

```
E←-ML×⍟(?Nρ1E8)÷1E8   .
```

Note U instead of 1-U is used for the calculation since both represent a uniformly distributed variate in the interval (0,1).

7. WEIBULL DISTRIBUTION

The Weibull distribution is a continuous distribution with probability density function [4]

$$f(t) = (\beta t^{\beta-1}/\alpha^{\beta})\exp\{-(t/\alpha)^{\beta}\} \quad ,$$

where $\alpha,\beta,t>0$. Here α and β are called the scale and shape parameters respectively. A graph of several Weibull densities is shown in Figure 5. The CDF is given by $F(t) = 1 - \exp\{-(t/\alpha)^{\beta}\}$. We see that $\beta = 1$ corresponds to the exponential distribution. Furthermore, the Weibull hazard rate is given by $\beta t^{\beta-1}/\alpha^{\beta}$. For $\beta = 1$, the rate is constant; for $\beta < 1$ the rate decreases in time; for $\beta > 1$ the rate increases in time. The Weibull distribution has mean $\alpha\Gamma[(\beta + 1)/\beta]$, where Γ is the gamma function [2], and variance $\alpha^2(\Gamma[(\beta+2)/\beta]-\{\Gamma[(\beta+1)/\beta]\}^2)$.

To generate Weibull distributed numbers, we use the inverse transformation method. The inverse expression is $F^{-1}(U) = \alpha[-\ell n(1 - U)]^{1/\beta}$. Hence, an APL expression to generate *N* Weibull distributed pseudo-random numbers from a population with scale parameter *A* and shape parameter *B* is

```
W←A×(-⍟(?Nρ1E8)÷1E8)*÷B   .
```

8. PLOTTING DISTRIBUTIONS

The empirical CDF can be defined for any real t as the proportion of random values that are less than or equal to t. Hence, by use of the empirical CDF, any of the previously discussed distributions can be plotted. To illustrate, suppose we have generated N pseudo-random lognormal numbers, denoted by the symbol *L*. The numbers are ordered by executing `L←L[⍋L]`. The empirical CDF will be provided by the APL expression `F←(⍳N)÷N`. Then, the CDF is plotted as *F* versus *L*.

If desired, the theoretical CDF from which the random sample is drawn can be overlaid on the sample plot for comparison. The population CDF would be obtained by substituting the pseudo-random numbers generated for the argument t in the equation for F(t). For example, let *W* be the pseudo-random numbers drawn from a Weibull population with scale factor *A* and shape factor *B*. First order by executing `W←W[⍋W]`. Then the theoretical CDF is given by

```
FT←1-*-(W÷A)*B.
```

Then, for `F←(⍳N)÷N`, the comparative plot of both CDF's is accomplished by plotting *FT* and *F* versus *W*. An example of such a plot is shown in Figure 6. For testing goodness of fit, Kolmogorov-Smirnov tests [1,4] would be appropriate.

If one wishes to plot and compare a theoretical density function to the simulation output, one can make a histogram plot of the pseudo-random numbers and compare the histogram to the population density function evaluated at selected points, for example, at the midpoints of the interval. This procedure would also facilitate application of the chi-squared [1,4] goodness of fit test to the data. An example of a histogram comparison is given in Figure 7, for the simulation of 200 numbers from a standard normal population.

9. REFERENCES

1. Bury, K.V., "Statistical Models in Applied Science," New York: John Wiley & Sons, Inc., 1975.

2. Hastings, N.A.J. and Peacock, J.B., "Statistical Distributions," London: Butterworths, 1974.
3. IBM, "APL Language GC26-3847-2," San Jose: IBM Corporation, 1976.
4. Mann, N.R., Schafer, R.E., and Singpurwalla, N.D., "Methods for Statistical Analysis of Reliability and Life Data," New York: John Wiley & Sons, Inc., 1974.
5. Mendenhall, W. and Schaeffer, R.L., "Mathematical Statistics with Applications," North Scituate: Duxbury Press, 1973.
6. Trindade, D.C. and Haugh, L., "Nonparametric Estimation of Lifetime Distribution Via the Renewal Method," Chicago: Presented at American Statistical Association Meeting, 1977.

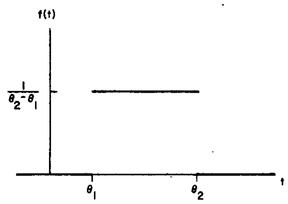

Fig. 1. The Uniform Probability Density Function.

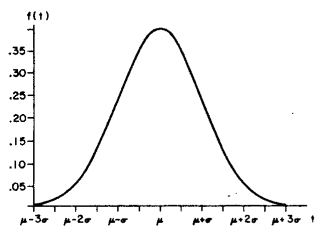

Fig. 2. The Normal Probability Density Function.

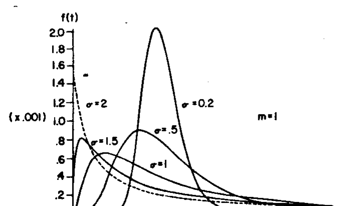

Fig. 3. The Lognormal Density Function.

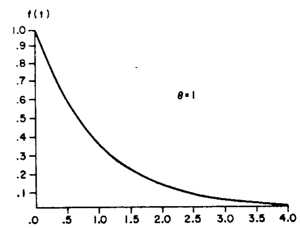

Fig. 4. The Exponential Probability Density Function.

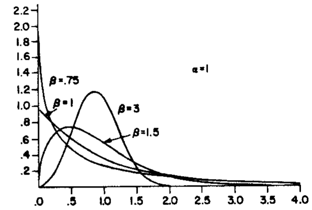

Fig. 5. The Weibull Probability Density Function.

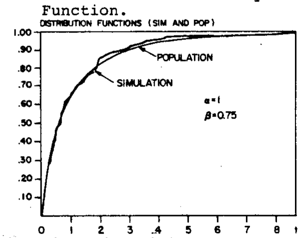

Fig. 6. Plot of Empirical CDF and Population CDF for Simulation Sample of N=100 from Weibull Distribution.

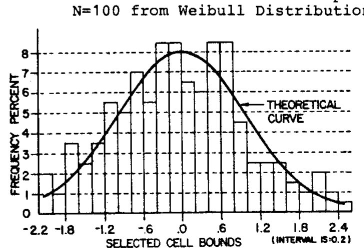

Fig. 7. Comparison of Simulation Results for N=200 to Standard Normal Density Function.