Management Science/ Operations Research
D.F. Shanno Editor

Computer Generation of Gamma Random Variates with Non-integral Shape Parameters

N.D. Wallace
Southern Illinois University at Edwardsville

When the shape parameter, α, is integral, generating gamma random variables with a digital computer is straightforward. There is no simple method for generating gamma random variates with non-integral shape parameters. A common procedure is to approximately generate such random variables by use of the so-called probability switch method. Another procedure, which is exact, is due to Jöhnk. This paper presents a rejection method for exactly generating gamma random variables when α is greater than 1. The efficiency of the rejection method is shown to be better than the efficiency of Jöhnk's method. The paper concludes that when α is non-integral the following mix cf procedures yields the best combination of accuracy and efficiency: (1) when α is less than 1, use Jöhnk's method; (2) when 1 is less than α and α is less than 5, use the rejection method; (3) when α is greater than 5, use the probability switch method.

Key Words and Phrases: simulation, gamma random variables, probability distribution, random numbers

CR Categories: 5.5, 8.1

Author's address: Department of Mathematical Studies, Southern Illinois University at Edwardsville, Edwardsville, IL 62025.

[1] It is understood that when one says a method exactly generates random variables on a computer, that the exactness is limited by the computer used, and the randcmness of the underlying pseudo-random number generator, but not by the method itself.

[2] In this paper the phrase uniform random variable will be taken to be a random variable R, uniformly distributed between 0 and 1. The notation r_i will be taken to mean the ith value of a uniform random variable.

This paper compares three methods for computer generation of gamma random variables with non-integral shape parameters. The gamma distribution in generalized form can be written as

$$g(y) = (\beta^{\alpha}(y-\xi)^{\alpha-1} \exp(-\beta(y-\xi))/\Gamma(\alpha)), \quad y \geq \xi, \quad \alpha, \beta, > 0. \tag{1}$$

Note that it is sufficient to generate gamma variables from the distribution

$$f(x) = (x^{\alpha-1}e^{-x}/\Gamma(\alpha)), \quad x \geq 0, \quad \alpha > 0, \tag{2}$$

and make a simple change of variables

$$y = (x/\beta) + \xi \tag{3}$$

to obtain the more general distribution. Thus, all methods will be compared by generating variables from (2).

If the shape parameter α is an integer, there is a simple procedure for generating gamma variables [1]. However, if α is non-integral, there are considerable difficulties with the generating procedure. For some some time, no exact method was known and approximate methods were used. The most common method, which is used by the SIMSCRIPT II [2] and GERT [3] simulation languages, is the so-called probability switch method [1]. This method proceeds as follows:

1. Let m be the largest integer less than α (denoted by $m = [\alpha]$).
2. Let $q = \alpha - m$. With probability q, generate gamma variables with shape parameter $m + 1$.
3. With probability $(1 - q)$, generate gamma variables with shape parameter m.

This mixture of gamma variables with integral shape parameters which bracket the true value of α will approximate the desired gamma distribution. This method will only work when $\alpha \geq 1$.

Another method has been developed by Jöhnk which will exactly[1] generate gamma variables with non-integral shape parameters [4, 5]. The algorithm given below is a slight modification of Jöhnk's algorithm (see [5, p. 3]), but it achieves the same result and will be referred to as Jöhnk's method.

Beta-variable Algorithm $B(a, b+1)$

Step 1. $i = 1$.

Step 2. Generate a uniform random number[2] r_i and set $x = r_i^{1/a}$.

Step 3. Generate a uniform random number r_{i+1} and $y = r_{i+1}^{1/b}$.

Step 4. If $x + y \leq 1$, go to 6; otherwise go to 5.

Step 5. $i \rightarrow i + 2$, go to 2.

Step 6. Done. x is a beta variate from $B(a,b+1)$.

Gamma-variable Algorithm with non-integral shape parameters α

Step 1. Set $Z = -\log \prod_{i=1}^{[\alpha]} r_i$. The r_i are uniform random numbers. $z = 0$ if $\alpha < 1$.

Step 2. Generate a random number x from $B(\alpha-[\alpha], 2-\alpha+[\alpha])$, using the above beta algorithm.

Step 3. Generate uniform random numbers r_1 and r_2, and set $y = -\log(r_1 r_2)$.

Step 4. $(Z+xy)$ is the desired gamma variable.

It can easily be shown, by a transformation of variables, that Step 4 does indeed yield a gamma random variable with the proper shape parameter. Also note that if $\alpha < 1$, $Z = 0$, and Steps 2 through 4 yield the proper gamma random variable. This statement is confirmed in Appendix B. Thus, Jöhnk's method is exact for all values of α.

Berman [5] compared the efficiency of Jöhnk's method with the probability switch method. He found that for values of α from 1.5 to 10.5 in steps of 1.0, Jöhnk's method required from 2.1 to 3.1 times as long as the probability switch method. The actual time required for Jöhnk's method was less than 2 cpu sec greater than the probability switch method for generating 1000 random variates on an IBM 360/65. This amount of time increase is not unimportant when considered over a period of time, if the method is used in a number of large simulation runs.

Berman's recommendation was that, except when computational efficiency is much more important than accuracy, Jöhnk's method be used for $1.0 < \alpha < 5.0$. The probability switch method should be used for $\alpha > 5.0$. The accuracy of the probability switch method is quite good for $\alpha > 5.0$. (Note that when α is an integer, both methods reduce to the same thing.)

An alternate method will now be derived, based on the rejection technique [6].

General Rejection Method

Let $n(x)$ and $m(y)$ be probability density functions, and let $T(x)$ be an arbitrary function. Then:

Step 1. Select, at random, an x out of $n(x)$.

Step 2. Select, independently, a y out of $m(y)$.

Step 3. If $y \leq T(x)$, accept x. Otherwise repeat Steps 1 and 2.

Justification: Let $f(x)$ be the desired probability density for the accepted values of the random variable x. $f(x)$ is, of course, a conditional distribution, but the condition that x is an accepted value will not be specifically indicated by notation.

The probability of selecting an x with a value in the interval $(x,x+dx)$ is $n(x)\,dx$. The probability of accepting this selected value is

probability $y \leq T(x) = M[T(x)]$

where $M(y)$ is the cumulative distribution function for the random variable Y. Thus, the probability of selecting an x in the interval $(x,x+dx)$ and accepting this value is:

$$M[T(x)]n(x)\,dx \tag{4}$$

The probability of accepting any value of x on any given trial is

$$E = \int_{-\infty}^{\infty} M[T(x)]n(x)\,dx. \tag{5}$$

The rejection process is a sequence of Bernoulli trials, and E represents the expected fraction of variables accepted in a sequence of trials. Thus, E is called the efficiency of the test.

Applying the above results, it can be seen that the probability that an accepted value of X is in the interval $(x, x+dx)$ is:

$$f(x)\,dx = (M[T(x)]n(x)\,dx/E). \tag{6}$$

It is clear that if a rejection technique is to be useful it must be efficient. Therefore, it should be easy to sample from the distributions $n(x)$ and $m(y)$, and E must be as close to one as possible. In addition

$$M[T(x)]n(x)$$

must be a function that is nonzero over the range of x that is desired for $f(x)$. It is rather difficult to establish procedures for designing a rejection technique, based on the general procedure just described, that is most efficient for any given desired distribution $f(x)$. There are simply too many degrees of freedom in the choice of procedure.[3] A somewhat simpler procedure can be derived by letting Y be a uniform random variable. Thus, in the notation of the previous rejection procedure $M[T(x)] = T(x)$, where $T(x)$ bounded so that $T(x) \leq 1$. In order to make the design of the rejection procedure simple, $T(x)$ is defined as:

$$T(x) = f(x)/Kn(x) \tag{7}$$

where K is greater than or equal to the maximum value of $f(x)/n(x)$. Since $f(x)$ and $n(x)$ are nonzero for the same values of x, it can easily be seen that $K \geq 1$. Otherwise,

$$f(x)/n(x) \leq K < 1 \tag{8}$$

for all x. This would imply

$$\int_{-\infty}^{\infty} f(x)\,dx < \int_{-\infty}^{\infty} n(x)\,dx. \tag{9}$$

Equation (9) is a contradiction, since $f(x)$ and $n(x)$ are both probability density functions. The simplified rejection procedure can now be described as follows:

Step 1. Select at random an x from the probability density function $n(x)$.

Step 2. Generate a uniform random number r.

Step 3. If $r \leq f(x)/Kn(x)$, accept x. Otherwise repeat Steps 1 and 2.

[3] Note in Appendix B that the beta-variable algorithm of Jöhnk's method is an example of this type of rejection technique.

The efficiency of this rejection procedure is:

$$E = \int_{-\infty}^{\infty} (f(x)/Kn(x))n(x)\,dx = \int_{-\infty}^{\infty} f(x)/K = 1/K \quad (10)$$

Thus, the efficiency is the minimum value of $n(x)/f(x)$. This suggests that $n(x)$ should be chosen to be "similar" to $f(x)$ so that the minimum of $n(x)/f(x)$ is close to 1. However, $n(x)$ must be easy to sample from by use of a computer, or the purpose of choosing $n(x)$ "similar" to $f(x)$ will be defeated.

The simplified rejection procedure will now be applied to the generation of random variables from the gamma distribution with a non-integral shape parameter $\alpha > 1$.

The distribution $n(x)$ is chosen to be

$$n(x) = (px^{m-1}e^{-x}/(m-1)!) + ((1-p)x^m e^{-x}/m!), \quad x \geq 0 \quad (11)$$

where $m = [\alpha]$ and $0 \leq p \leq 1$. Thus samples from the distribution $n(x)$ can be simulated by (see [1])

$$x = -\ln\left(\textstyle\prod_{i=1}^{m} r_i\right) \quad (12)$$

with probability p, and

$$x = -\ln\left(\textstyle\prod_{i=1}^{m+1} r_i\right) \quad (13)$$

with probability $1 - p$. The set $\{r_i\}$ is a set of uniform random numbers. Thus,

$$f(x)/n(x) = ((x^{\alpha-1}e^{-x}/\Gamma(\alpha))/((px^{m-1}e^{-x}/(m-1)!) + ((1-p)x^m e^{-x}/m!))). \quad (14)$$

Let $q = \alpha - m$. Then (14) may be written

$$f(x)/n(x) = ((m-1)!x^q/[p+(x/m)(1-p)]\Gamma(m+q)). \quad (15)$$

It is clear that for $x \geq 0$, $f(x)/n(x)$ is a concave function. The maximum of $f(x)/n(x)$ occurs at

$$x = m\left(\frac{p}{1-p}\right)\left(\frac{q}{1-q}\right). \quad (16)$$

Substituting (16) into (15) yields

$$\max_x (f(x)/n(x)) = (m^q(m-1)!q^q(1-q)^{1-q}/\Gamma(m+q)p^{1-q}(1-p)q) = a(m,p,q). \quad (17)$$

Since the question of what to choose for p has been left open, p is now chosen in order to minimize $a(m, p, q)$. This will yield the most efficient rejection procedure. It is clear that $a(m, p, q)$ is convex for a fixed q and m and for $0 \leq p \leq 1$. The minimum occurs at $p = 1 - q$. Thus,

$$e(m,q) \triangleq \min_p [\max_x (f(x)/n(x))] = a(m,1-q,q) = (m-1)!m^q/\Gamma(m+q). \quad (18)$$

Equation (18) yields the smallest value for K, and therefore the most efficient rejection procedure. Substituting (18) for K and combining this result with (15) yields

$$f(x)/Kn(x) = (x/m)^q/(1+[(x/m)-1]q). \quad (19)$$

A summary of the revised rejection procedure is as follows:

1. Given $\alpha > 1$, compute $q = \alpha - m$, where $m = [\alpha]$.
2. With probability q, compute

$x = -\ln\left(\prod_{i=1}^{m+1} r_i\right)$

with probability $1 - q$, compute

$x = -\ln\left(\prod_{i=1}^{m} r_i\right)$

where the set $\{r_i\}$ is a set of uniform random numbers.

3. Generate another uniform random number r. If

$r \leq ((x/m)^q/(1+[(x/m)-1]q))$

accept x. Otherwise reject x and go back to step 2.

The efficiency of the rejection procedure is $1/e(m,q)$. It is shown in Appendix A that $\min_{m,q} [1/e(m,q)] \cong .886$.

Thus, the minimum efficiency of the rejection procedure is 88.6 percent. This minimum occurs at $m = 1$, and rapidly increases toward 100 percent as m gets large. Table II shows the efficiency versus m for values up to $m = 5$. At $m = 5$, the efficiency is 98 percent. This means that a random variable is rarely rejected for $\alpha > 5$, and the rejection procedure is essentially the probability switch method. Since the rejection procedure is exact, this explains why the probability switch method yields such good results for $\alpha > 5$, and is the recommended procedure by Berman [5]. However, it will be shown that the described rejection procedure is more efficient than Jöhnks method for all values of $\alpha > 1$, and is therefore the recommended procedure for all values of $\alpha > 1$.

Comparison of Efficiency of Rejection Procedure with Jöhnk's Method

It is shown in Appendix B that the beta-variable algorithm in Jöhnk's method has a minimum efficiency at values of $q = \alpha - m = 1/2$. Since the rejection procedure in this paper has a minimum efficiency at $q \cong 1/2$, the two methods will be compared most realistically for values of $\alpha = m + 1/2$. Table I gives values of the ratio of average time required to generate 1000 random variates by Jöhnk's method (t_j) and by the proposed rejection procedure (t_r). For comparison purposes the ratio t_j/t_s is also shown, where t_s is the average time required to generate 1000 random variates by the probability switch method. The computer results were obtained using a GE 235 computer. The actual time differences will vary with the computer, but the ratio is not likely to change much. For example, the ratios t_j/t_s in Table I are very close to those cited by Berman [5], even though Berman used an IBM 360/65.

Table I. Ratios of Generation Times for Jöhnk's Method (t_j) the Probability Switch Method (t_s), and the Rejection Method (t_r)

α	1.5	2.5	3.5	4.5
t_j/t_r	1.63	1.61	1.58	1.56
t_j/t_s	3.08	2.89	2.74	2.61

Table II. Values of q^* and Efficiency, E, Versus m

m	q^*	$1/e(m,q^*) = E$
1	.462	.886
2	.479	.940
3	.491	.960
4	.495	.970
5	.498	.980

Conclusions

The rejection procedure proposed in this paper cannot be used for $\alpha < 1$. A rejection procedure can be devised which will work for $\alpha < 1$, but the author has been unable to devise one that is more efficient than Jöhnk's method. Thus, the following scheme is recommended.

1. Use Jöhnk's procedure for $\alpha < 1$.
2. Use the rejection procedure of this paper for $1 < \alpha < 5$.
3. Use the probability switch method for $\alpha > 5$.

It is a simple matter to write a Fortran program that incorporates all procedures as well as handling the case where α is an integer. A flowchart for such a program is given in Appendix C.

Appendix A

At $q = 1$ and $q = 0$, it is clear from eq. (18) that $e(m,q) = 1$. Since $e(m,q) \geq 1$ for all values of q and m, it follows that for a fixed m $e(m,q)$ has a maximum between $q = 0$ and $q = 1$. We find this maximum by solving the equation

$$\frac{\partial e(m,q)}{\partial q} = 0 = \frac{\partial}{\partial q}\left[\frac{(m-1)!m^q}{\Gamma(q+m)}\right]. \tag{A1}$$

Performing the above differentiation results in

$$\Psi(m+q) = \ln m \tag{A2}$$

where $\Psi(\cdot)$ is the Psi function [7]. The solution of (A2), subject to $0 \leq q \leq 1$, maximizes $e(m,q)$ for fixed m. This optimal solution will be denoted by $q^*(m)$.

Eq. (A1) was solved numerically for $m = 1, 2, 3, 4, 5$. The solutions and resultant efficiency are given in Table II. It is of interest to find an asymptotic solution of (A1) and the resultant symptotic value of $e(m,q^*)$.

An asymptotic formula for $\Psi(m+q)$ can be written (see [7])

$$\begin{aligned}\Psi(m+q) \sim \ln (m+q) &- (1/2(m+q)) \\ &- (1/12(m+q)^2) \\ &+ (1/120(m+q)^4) \quad (1/252(m+q)^6) + \cdots\end{aligned} \tag{A3}$$

If $|q| \leq 1$, then as m gets large we may write

$$\ln (m+q) = \ln m + (q/m) - (q^2/2m^2) + \cdots \tag{A4}$$

$$(1/2(m+q)^2) = (1/2m) - (q/2m^2) + \cdots \tag{A5}$$

$$(1/12(m+q)^2) = (1/12m^2) + \cdots \tag{A6}$$

Thus, from (A2)–(A6)

$$\ln m \cong \ln m + ((q-1/2)/m) + ((-q^2/2)+(q/2) -(1/12))1/m^2. \tag{A7}$$

The solution of (A7) yields

$$q^* = (1/2) - (1/24m). \tag{A8}$$

Now consider

$$e(m,q^*) = ((m-1)!m^{q^*}/\Gamma(m+q^*)). \tag{A9}$$

In general, an asymptotic formula for $\Gamma(Z)$ is (see [7])

$$\begin{aligned}\Gamma(Z) \sim e^{-Z}Z^{Z-\frac{1}{2}}(2\pi)^{\frac{1}{2}}[1&+(1/12Z)+(1/288Z^2) \\ -(139/51840Z^3)&-(571/248230Z^4)+\cdots].\end{aligned} \tag{A10}$$

It is clear that the expression in brackets in (A7) is closer to 1 when $Z = m + q^*$ than it is when $Z = m$. Since $(m-1)! = \Gamma(m)$, it follows that

$$e(m,q^*) = (e^{q^*}m^{m-\frac{1}{2}}m^{q^*}/(m+q^*)^{m+q^*-\frac{1}{2}}). \tag{A11}$$

Thus,

$$\begin{aligned}\ln [e(m,q^*)] &\cong q^* + (m+q^*-1/2) \ln m \\ &\quad - (m+q^*-1/2) \ln (m+q^*) \\ &\cong q^* - (m+q^*-1/2)[(q^*/m)-((q^*)^2/2m^2)].\end{aligned} \tag{A12}$$

Substituting (A8) into (A12), we obtain

$$\ln [e(m,q^*)] \cong (1/8m).$$

Therefore,

$$e(m,q^*) \cong 1 + (1/8m). \tag{A13}$$

Although (A8) and (A13) are asymptotic solutions, it can be seen by comparison with the results in Table II that they are surprisingly accurate, even for $m = 1$. Thus, q^* rapidly approaches 1/2, and $e(m,q^*)$ rapidly approaches 1 as m increases.

Appendix B

The rejection procedure used in Jöhnk's algorithm for generating beta random variables can be described as a general rejection procedure with

$$n(x) = ax^{a-1}, \quad 0 \le x \le 1, \tag{B1}$$
$$m(y) = by^{b-1}, \quad 0 \le y \le 1, \tag{B2}$$
$$T(x) = 1 - x, \tag{B3}$$
$$M[T(x)] = (1-x)^b. \tag{B4}$$

Thus

$$Ef(x)\,dx = a(1-x)^b x^{a-1}\,dx$$
$$E = \int_0^1 a(1-x)^b x^{a-1}\,dx$$
$$E = (a\Gamma(b+1)\Gamma(a)/\Gamma(a+b+1)), \tag{B5}$$

and

$$f(x) = (\Gamma(a+b+1)(1-x)^b x^{a-1}/\Gamma(a)\Gamma(b+1)). \tag{B6}$$

Step 3 of Jöhnk's method for generating gamma random variables will generate random variables with the probability density function

$$g(y) = ye^{-y}, \quad y \ge 0. \tag{B7}$$

THEOREM. *Let x have the distribution given by (B6) with $a = \alpha$, $b = 1 - \alpha$. Let y have the distribution given by (B7). Then $v = xy$ has the distribution*

$$v^{\alpha-1}e^{-v}/\Gamma(\alpha), \quad v \ge 0.$$

PROOF. Let $u = y$, $v = xy$. Then $x = v/u$, $y = u$. The Jacobian of this transformation is

$$J = \begin{vmatrix} -v/u^2 & 1/u \\ 1 & 0 \end{vmatrix} = -1/u.$$

The joint distribution of (u,v) is therefore given by

$$h(u,v)\,du\,dv = (\Gamma(a+b+1)/\Gamma(a)\Gamma(b+1)) \quad (1-v/u)^b (v/u)^{a-1} e^{-u}\,du\,dv.$$

Since $a = \alpha$, $b = 1 - \alpha$,

$$h(u,v)\,du\,dv = ((u-v)^{1-\alpha} v^{\alpha-1} e^{-u}/\Gamma(\alpha)\Gamma(2-\alpha))\,du\,dv, \quad 0 \le v \le u.$$

The marginal distribution for v is computed as follows:

$$\begin{aligned} \omega(v) &= \textstyle\int_v^\infty h(u,v)\,su, \\ &= (v^{\alpha-1}/\Gamma(\alpha)\Gamma(2-\alpha)) \textstyle\int_v^\infty (u-v)^{1-\alpha} e^{-u}\,du, \\ &= (v^{\alpha-1}e^{-v}/\Gamma(\alpha)), \qquad v \ge 0. \ \square \end{aligned}$$

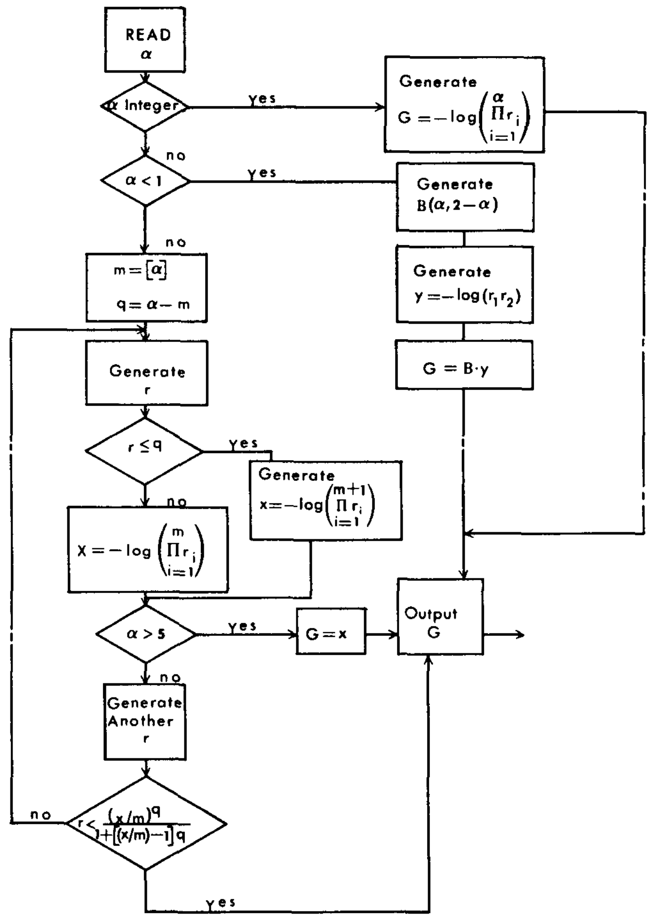
Thus, Jöhnk's method generates variables with the desired distribution for $\alpha < 1$. It is easy to generalize the above proof for the case where $\alpha > 1$, but in this paper it is recommended that Jöhnk's algorithm not be used when $\alpha > 1$.

It is desirable to consider the minimum efficiency of Jöhnk's procedure for generating beta random variables. Consider (B5) with $a = q$, $b = 1 - q$, $0 \le q \le 1$. When $q = 0$ or $q = 1$, $E = \Gamma(2)\Gamma(1) = 1$. Since $E \le 1$, E has a minimum for $0 < q < 1$. It can easily be shown that $\partial E/\partial q = 0$ implies that

$$\Psi(1+q) = \Psi(2-q). \tag{B8}$$

The solution of (B8) is $q = 1/2$. When $q = 1/2$, $E = [\Gamma(3/2)]^2 = \pi/4 \approx .785$.

Appendix C. Flowchart for Gamma Random Variable Generator



Received May 1974; revised July 1974

References

1. Naylor, T.H., Balintfy, J.L., Burdick, D.S., and Chu, K. Computer Simulation Techniques. Wiley, New York, 1968.
2. Kiviat, P.J., Shukiar, H.J., Urman, J.B., and Villanueva, R. The Simscript II programming language: IBM 360 implementation. The Rand Corp., RM-5777-PR, July 1969.
3. Walters, L.J., and Vasilek, M.W. A stochastic network approach to test and checkout. Fourth Conf. on Applications of Simulation Proceedings, Dec. 1970, pp. 113–123.
4. Johnk, M.D. Erzeugung Von Betavesteilten und Gammavesteilten Zufellszahlen. *Metrica 8*, 1, (1964), 5–15.
5. Berman, M.B. Generating gamma distributed variates for computer simulation models. The Rand Corp., R-641-PR, Feb. 1971.
6. Kahn, H. Applications of Monte Carlo. The Rand Corp., RM-1337-AEC, Apr. 1956.
7. Abramowitz, M., and Stegun, I.A. Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables. U.S. Dept. of Commerce, NBS Applied Mathematics Series No. 55, June 1964.