Reprinted From

A Modern Course on
STATISTICAL DISTRIBUTIONS IN SCIENTIFIC WORK

Edited by
G. P. PATIL, S. KOTZ, AND J. K. ORD

Published by
D. Reidel Publishing Company
38 Papeterspad, P. O. Box 17, Dordrecht, Holland
306 Dartmouth Street, Boston, Massachusetts 02116
STATISTICAL DISTRIBUTIONS IN SCIENTIFIC WORK
Based on the NATO Advanced Study Institute
A Modern Course on Statistical Distributions in Scientific Work
and the International Conference on Characterizations
of Statistical Distributions with Applications

Held at
The University of Calgary, Calgary, Alberta, Canada
July 29-August 10, 1974

Sponsored by
International Statistical Institute
The Pennsylvania State University
The University of Calgary
Indian Statistical Institute

With the Support of
North Atlantic Treaty Organization
National Research Council of Canada
United States Army Research Office

DIRECTOR
C. P. Patil

SCIENTIFIC DIRECTORS
C. P. Patil, S. Kotz, J. K. Ord

JOINT DIRECTORS
E. G. Enns (Local Chairman), J. K. Wani, P. C. Consul

ADVISORS
T. Cacoullos, J. B. Douglas, A. Hald
W. L. Harkness, N. L. Johnson, A. M. Kagan

G. D. Kemp, E. Lukacs, L. J. Martin
I. Kotlarski, W. Molenaar, P. A. F. Moran
I. Olkin, L. R. Shenton, J. E. Mosimann
C. R. Rao, B. A. Sprott, J. Tiago de Oliveira
H. Teicher
CHANCE MECHANISMS IN COMPUTER GENERATION OF RANDOM VARIABLES

G.P. Patil, M.T. Boswell, D.S. Friday
The Pennsylvania State University

SUMMARY

This paper is a first attempt at developing a unifying theory for various procedures used in the generation of independent, univariate, random variables on a computer. Some generalizations and new interpretations of existing methods are presented. The emphasis is on chance mechanisms. Considerations of inference are irrelevant in that the user of random numbers is interested only in the accuracy of their distributions and in the computational efficiency with which they are generated.

We do not discuss all of the distributions which are commonly simulated, but we do attempt to isolate all techniques used in the literature. Methods considered in this paper generate theoretically exact distributions. Approximations in these methods occur only because of limitations of the computer on which they are implemented.

KEYWORDS: chance mechanisms, characterizations, computer generation of random variables, mixtures, negative mixtures, random variable decomposition, rejection techniques, simulation, waiting time techniques.

1. INTRODUCTION

Throughout this paper X will represent the random variable (r.v.) which we are interested in generating, and Y will represent the r.v. we are capable of generating. The only exceptions

are where \( Y \) has a uniform distribution on the interval \((0, 1)\) in which case the symbol \( U \) is used, and where \( Y \) is a fair Bernoulli trial \((0 \text{ or } 1)\) in which case the symbol \( B \) is used. We assume that we can generate a sequence of independent, fair, \((0,1)\), Bernoulli trials. In reality this is a purely deterministic sequence which passes many of the statistical tests for randomness.

We begin with a development of the main techniques which will be shown to fall in two general classes: (i) transformation of variables, (Section 2); and (ii) mixtures of distributions, (Sections 3, 5 and 7). Rejection techniques are discussed in Section 4 and a new technique involving r.v. decomposition is presented in Section 6. Sections 5 and 7 show the equivalence of the rejection and decomposition techniques respectively, to certain types of mixtures. In Section 8 waiting time techniques are considered, and in Section 9 we present examples using several techniques. Some new characterization theorems related to r.v. generation are in Section 10.

2. TRANSFORMATIONS OF VARIABLES

Suppose we can generate a r.v. \( Y \) with distribution \( F_Y(y) \) and are interested in generating a r.v. \( X \) with distribution \( F_X(x) \). Then we say r.v. \( X \) is generated by a transformation, \( x = h(y) \), if \( h(Y) \) has the desired distribution \( F_X \).

Example 2.1. Suppose that we wish to generate a binomial r.v. \( X \) with parameters \( n \) and \( p \), and suppose we can generate random vectors \( \mathbf{Y} = (B_{ij}, \ldots, B_{im}), i = 1, 2, \ldots, k \) of independent r.v.'s with \( \Pr[B_{ij} = 0] = \Pr[B_{ij} = 1] = 1/2, j = 1, 2, \ldots, m \). For some integer \( k > 1 \) let

\[
c_1 2^{k-2} + \ldots + c_{k-1} 2^0 = (r-1)/2
\]

be a binary expansion where \( r \) is an odd integer less than \( 2^k \), and observe that the representation may have zeros as leading coefficients. Define \( \mathbf{Y} = B_{i1} \cdots + B_{im} \), and define in terms of the coefficients \( c_s \)

\[
Y_{i0} \cdots Y_{j1} \cdots \cdots Y_{jm} = \begin{cases} 
(B_{i1} \cup B_{j1}, \ldots, B_{im} \cup B_{jm}) & \text{if } c_s = 1 \\
(B_{i1} \cap B_{j1}, \ldots, B_{im} \cap B_{jm}) & \text{if } c_s = 0 
\end{cases}
\]

where \( s = 1, \ldots, k-1; i = 1, 2, \ldots, k; j = 1, 2, \ldots, m. \)
Then \( X = \pm Y = \pm Y_1 \) (\( Y_2 \), \( Y_3 \), \( \ldots \)) has the binomial distribution with parameters \( n = m \) and \( p = r/2^k \). Since \( (r/2^k) : r = 1, 3, \ldots, r < 2^k, k = 2, 3, \ldots \) is a dense subset of the interval \((0,1)\), the value of \( p \) is limited only by the accuracy of the computer.

Many transformations involve the inverse distribution function. To generate a r.v. \( X \) with distribution \( F_X(x) \) we use the fact [Hogg & Craig, 1970, p. 349] that \( F(x) \) is uniformly distributed on \((0,1)\). Then \( X = F^{-1}(u) \) has the desired distribution.

Example 2.2. Let \( \{p_i, i=1, \ldots, k\} \) be any discrete distribution (possibly truncated to \( k \) outcomes) on the sample space \( \{x_i, i=1, \ldots, k\} \) where \( p_i = \Pr(X=x_i) \). If we can generate a uniform r.v. \( U \) on \((0,1)\) then \( X \) has the desired distribution, where

\[
X = x_i \quad \text{if} \quad \sum_{i=0}^{i-1} p_j \leq U < \sum_{i=0}^{i} p_j, \quad i=1, \ldots, k.
\]

Note that in this case \( F^{-1} \) is not well-defined but the same concept applies.

Example 2.3. Some continuous distributions and their inverse transformations are: (see also Example 2.7)

(a) Exponential, \( X = -\lambda \ln(1-U) \)
(b) Cauchy, \( X = \tan(\pi(U-\frac{1}{2})) \)
(c) Logistic, \( X = \ln(U/(1-U)) \)

where \( U \) is distributed uniformly on \((0,1)\).

Transformations which are not inverse distribution functions are also used in generating random variables.

Example 2.4. If we can generate a standard normal r.v. \( Y \) then \( X = Y^2 \) is chi-square with \( 1 \) degree of freedom.

Example 2.5. The polar (Box-Muller, 1958) transformation is

\[
X_1 = (-2 \ln U_1)^{1/2} \cos(2\pi U_2)
\]

\[
X_2 = (-2 \ln U_1)^{1/2} \sin(2\pi U_2).
\]

Given a pair \( (U_1, U_2) \) of independent uniform r.v.'s on \((0,1)\) the transformation results in a pair \( (X_1, X_2) \) of independent standard normal r.v.'s. A modification (Marsaglia, 1962 a) which is computationally more efficient will be discussed in Example 8.1.
In addition to transformations from a given r.v. to the desired r.v. there are transformations which we call equivalent mechanisms. The concept of equivalent mechanisms involves replacing the obvious method of generating a r.v. with a less obvious but equivalent method which is computationally more efficient. Equivalent mechanisms may use mixture techniques, transformations or any combination of the two.

Example 2.6. Suppose we want to generate a r.v. \( X \) with distribution \( F_X(x) = x^2 ; 0 < x < 1 \). The inverse distribution transformation is \( X = \sqrt{U} \) where \( U \) is uniform on \((0,1)\). We can generate \( X \) by the equivalent mechanism \( X = \max(U_1, U_2) \) more efficiently. (Knuth 1969, p. 103).

Example 2.7. In Example 2.3 (a) a subtraction can be avoided by observing that \( U \) and \( 1-U \) are both uniform on \((0,1)\), and in Example 2.3 (b) a subtraction can be avoided by observing that \( \tan\left[\pi(U-1/2)\right] \) and \( \tan(\pi U) \) have the same distribution.

3. MIXTURES OF DISTRIBUTIONS

A mixture \( F_X(x) = \sum_{i=1}^{\infty} F_i(x) \) of a family \( F_i, i=1,... \) of distributions functions is a weighted average of the \( F_i \)'s with weights \( p_i \geq 0, \sum p_i = 1 \). A random variable \( X \) with distribution \( F_X(x) \) can then be generated by choosing one of the distributions \( F_i(x) \) with probability \( p_i \), generating \( X_i \) from the chosen distribution and setting \( X = X_i \). This method is often more efficient than generating \( X \) directly from \( F_X(x) \) since some of the \( F_i \)'s may be very easy distributions from which to generate random variables. If we can make the \( p_i \)'s for these distributions large, then most of the times that we generate \( X \), it will come from one of these distributions.

Example 3.1. Consider the standard normal distribution \( F_X(x) \) (or any distribution symmetric about zero). Since it is symmetric about zero we need only simulate half of it and then randomly attach a sign to the result. Let \( F_X^+(x) \) be the distribution for \( x > 0 \) and \( F_X^-(x) \) be the distribution for \( x < 0 \) then \( F_X(x) = (1/2)F_X^+(x) + (1/2)F_X^-(x) = (1/2)F^+(x) + (1/2)F^-(x) \).

Example 3.2. In some simulation problems the underlying model may be a contaminated population. Suppose \( X_1 \) has distribution \( F_1(x) \) and \( X_2 \) has distribution \( F_2(x) \) and we know that the desired r.v. \( X \) comes from \( F_1(x) \) fraction \( \rho \) of the time and from \( F_2(x) \) fraction \((1-\rho) \) of the time. Then we know \( F(x) = \rho F_1(x) + (1-\rho) F_2(x) \) and can generate \( X \) using this mixture.

The above examples are some obvious applications of mixture techniques. Some techniques which do not seem to involve mixtures
are in fact special cases of mixtures and will now be discussed.

4. REJECTION TECHNIQUES

Suppose we can generate a r.v. $Y$ with distribution function $F_Y(y)$ and density function $f_Y(y)$, and we want to generate a r.v. $X$ with distribution function $F_X(x)$. This method consists of two parts, generating an observation of $Y$, and then deciding whether or not to set $X$ equal to this sample value or to reject the value and generate a new value of $Y$. Let $a_y$ be the probability of accepting $Y$ (i.e. setting $X = y$) when $Y = y$.

In order for this method to work it must be possible for $Y$ to assume every value that $X$ can assume. Let the support of a r.v. $X$ be the set of all values $x$ for which $f_X(x) > 0$. Then the support of $X$ must be a subset of the support of $Y$. Further observe that unless $f_Y = f_X$ there are values $y$ for which $f_X(y) > f_Y(y)$ and values $y$ for which $f_X(y) < f_Y(y)$.

Example 4.1. To generate a $(0,1)$-truncated exponential r.v. $X$ first generate a value $Y$ of an exponential r.v. and then set $X = y$, if $0 < y < 1$, or reject $Y$ if $y \geq 1$. That is $a_y = 1$, if $0 < y < 1$, $a_y = 0$, otherwise.

Let $A$ be the event that sampling stops on a given trial. Then decomposing the event $X < x$ into the union of mutually exclusive events depending on the number of trials until acceptance gives

$$\Pr[X < x] = \Pr[A, Y < x|0 \text{ rejections}] \Pr^0[A^c] + \Pr[A, Y < x|1 \text{ rejection}] \Pr^1[A^c] + \ldots$$

$$= \Pr[A, Y < x] (\Pr^0[A^c] + \Pr^1[A^c] + \ldots) = \Pr[A, Y < x]/\Pr[A].$$

Now (4.1) is true whether $X$ is discrete or absolutely continuous. In the continuous case $\Pr(A) = \int_{-\infty}^{\infty} a_y f_Y(y) \, dy$ and $\Pr[A, Y < x] = \int_{-\infty}^{x} a_y f_Y(y) \, dy$. Thus $f_X(x) = \int_{-\infty}^{x} a_y f_Y(y) \, dy/\Pr(A)$ and $f_X(x) = a_x f_Y(x)/\Pr(A)$. Therefore

$$a_x = \Pr(A) f_X(x)/f_Y(x)$$

the acceptance probability, is proportional to the likelihood ratio of the observations. Similarly (4.2) is true in the discrete case. An optimal design is to choose $\Pr(A) < 1$ as large as possible so $a_x < 1$ for all $x$. Note $\Pr(A)$ is never one unless $f_Y \equiv f_X$, an uninteresting case.
The acceptance or rejection of $Y$ can often be accomplished by generating a second value of $Y$, say $Y_2$, and given the first value of $Y$ say $Y_1$ we decide with $\Pr(Y_2 \in A_y) = a_y$ if an event $A_y$ has occurred or not and act accordingly.

Example 4.2. von Neumann (1951) observed that we can always generate a perfectly fair Bernoulli trial if we are given pairs of Bernoulli trials with any probability $p$ of a success. Let $\Pr(Y_1=1) = p$, $\Pr(Y_1=0) = 1-p$; $i=1,2,\ldots$ and the $Y_i$'s be independent. Observe a pair $(Y_1, Y_2)$. Reject the pair if $(Y_1, Y_2) = (1,1)$ or $(0,0)$ and set $X = Y_1$ if $(Y_1, Y_2) = (0,1)$ or $(1,0)$. This is equivalent to generating a value $Y_1 = y_1$ and setting $X = y_1$ if $Y_2 \neq y_1$. Thus $a_0 = \Pr(Y_2=1) = p$, $a_1 = \Pr(Y_2 = 0) = 1-p$, and $\Pr(A) = 2p(1-p)$. From (4.2)

$$a_0 = p = 2p(1-p) \frac{\Pr(X=0)}{\Pr(Y_1=0)} \text{ or } \Pr(X=0) = 1/2.$$

Example 4.3. Suppose we want to generate $X$ on $(0,1)$ using a r.v. $U$ which is distributed uniformly on $(0,1)$. Then (4.2) is $a_X = a \cdot f_X(x)$ where $a = \Pr(A)$ is chosen as large as possible so that $a \cdot f_X(x) \leq 1$ for $0 < x < 1$. Let $(u_1, u_2)$ be a pair of observations of independent uniform r.v.'s $(U_1, U_2)$. Then we set $X = u_1$ if $u_2 < \Pr(A)f_X(u_1)$, otherwise we reject the pair and generate another sample $(u_1, u_2)$. To be specific, let $f_X(x) = 2(1-x)$; $0 < x < 1$. Then $a = 1/2$, and we set $X = u_1$, if $u_2 \leq 1-u_1$ or equivalently if $u_1 \leq 1 - u_2$.

The efficiency of this procedure can, however, be improved by using equivalent mechanisms. Observe that setting $X = u_1$ if $u_1 < u_2$ is equivalent to the above since $U_2$ and $1-U_2$ have the same distribution. Also observe that sampling until $u_1 < u_2$ produces a pair of r.v.'s $(U_1, U_2)$ with the same distribution as letting $U_1$ be the minimum and $U_2$ the maximum of a pair of independent uniform r.v.'s. Thus the equivalent method is to generate a pair of uniform r.v.'s and set $X$ equal to the smaller one. What was initially a rejection technique now becomes a mixture technique.

Example 4.4. Suppose we wish to generate a continuous r.v. $X$ with a density $f_X(x)$ which is nearly linear on $(0,1)$ and with $f_X(1) = 0$. It is then reasonable to generate $Y$ with $f_Y(y) = 2(1-y)$; $0 \leq y < 1$ as the minimum of two uniforms (as in Example 4.3) and use a third uniform $U_1$ to accept or reject $Y$. Let $X = y$ if $u_2 \leq a f_X(y)/f_Y(y) = a f_X(y)/2(1-y)$ where $a \leq 1$ is made as large as possible subject to $a f_X(y) \leq 2(1-y)$ for all $y$ in $(0,1)$. This should result in a few rejections since $f_Y(y)/2(1-y)$ will be nearly 1 and $a = \Pr(A)$ is the probability that any particular trial is accepted. Note that the above method can be translated to generate a r.v. $X$ on any interval.
5. REJECTION TECHNIQUE AS A "NEGATIVE MIXTURE"

We now study the rejection techniques and their relation to mixture techniques. To generate $X$ from observations of $Y$ by the rejection technique the likelihood ratio $f_X(x)/f_Y(x)$ must be bounded. This implies the support of $f_X$ must be a subset of the support of $f_Y$. Since $f_X$ and $f_Y$ are both densities $\sup(f_X(x)/f_Y(x)) = \infty$ unless $f_X \equiv f_Y$, an uninteresting case. Thus $f_X(x)/b f_Y(x) < 1$.

Let $f(x) = [bf_Y(x) - f_X(x)]/(b-1)$; clearly $f(x)$ is a density function. Then $f_X(x) = b f_Y(x) - (b-1) f(x)$. We say that $f_X$ is a negative mixture of $f_Y$ with $f$. Observe that $f_X$ is a negative mixture of $f_Y$ with $f$ whenever $f_Y$ is a mixture of $f_X$ and $f$. Similarly if $f_Y$ is a negative mixture of $f_Y$ with $f$ as above with $b>1$ then $f_X(x)/bf_Y(x) < 1$, and the rejection technique is applicable with the acceptance probability $a = Pr(A) = (1/b) < 1$. The closer $Pr(A)$ is to 1 the more efficient the rejection technique is. As observed above, if $f_X$ is a negative mixture of $f_Y$ with $f$ then $f_Y(x) = af_Y(x) + (1-a) f(x)$.

6. DECOMPOSITION OF RANDOM VARIABLES

In general a r.v. $X = Y_1 + \ldots + Y_k$ is decomposed into component r.v.'s $Y_i$, $i=1, \ldots, k$. It may be possible that generating the component r.v.'s and summing them is more efficient computationally than generating $X$ directly. We denote by lattice decomposition the special case wherein a continuous r.v. $X = Y + Z$ is decomposed into a lattice r.v. $Y$ and a continuous r.v. $Z$ taking values on an interval corresponding to the gaps between the lattice points.

Example 6.1. Let $Y = \lceil X \rceil$ be the integer part and $Z = (X)$ be the fractional part of the random variable $X$. If $X$ has distribution $F_X(x)$ then it can be shown that

$$F_X(x) = F_X(y + z) = F_Y(y - 1) + F_Z(z | y) f_Y(y) \quad (6.1)$$

where $F_Y$ is the distribution of the lattice r.v. $Y$, $f_Y$ its density and $F_Z$, the conditional distribution of $Z$ given $Y = y$.

Solving for $F_Y$ and $F_Z$ in terms of $F_X$ gives

$$F_Y(y) = F_X(y + 1) \quad (6.2)$$

$$F_Z(z | y) = (F_X(y + z) - F_X(y))/(F_X(y + 1) - F_X(y)). \quad (6.3)$$

To illustrate, suppose $X$ is an exponential r.v. with parameter $\lambda$. Then $F_X(x) = 1 - e^{-\lambda x}$ and by (6.2) and (6.3)
\[ F_Y(y) = 1 - e^{-\lambda (y+1)} \]

\[ F_Z(z|y) = \frac{(1 - e^{-\lambda (y+z)} - (1 - e^{-\lambda y})/(1 - e^{-\lambda (y+1)} - (1 - e^{-\lambda y}))}{(1 - e^{-\lambda z})/(1 - e^{-\lambda})} = F_Z(z) \]

We observe that for an exponential r.v. X the lattice and continuous components, Y and Z, of the decomposition are statistically independent. Further Y has a geometric distribution with parameter \( P = (1 - e^{-\lambda}) \) and Z has a truncated exponential distribution with parameter \( \lambda \). Therefore we can generate Z independently of the value generated for Y.

7. DECOMPOSITION OF RANDOM VARIABLES AS A MIXTURE

Writing the density of X in terms of the component r.v. densities we have \( f_X(x) = f_Y(x+y) = f_{Y,Z}(y,z) = f_Z(z|y) f_Y(y) \). We now generalize the greatest integer function to the greatest lattice point function. Define \( X_L = \lfloor X \rfloor \) to be the value of the largest lattice point smaller than or equal to \( x \) and \( z_L = (X)_L = X - \lfloor X \rfloor \) to be the remainder. The r.v. \( X_L \) has non-zero density only on the interval \( (0,\lambda) \) and is dependent on r.v. \( Y_L \) which takes values \( 0, 1, 2, \ldots \). We define \( X_i, i = 0,1,\ldots \), to be a family of r.v.'s which have non-zero density only on the intervals \( (i\lambda, (i+1)\lambda) \); \( i = 0,1,\ldots \) respectively. Then the density can be expressed in terms of the \( X_i \)'s, as follows:

\[ f_X(x) = f_X(x_1 + z_L) = f_Z(z_L|y_L) f_Y(y_L) f_X(x_1) \]

\[ = f_X(x_1|y_L) f_Y(y_L) = f_{X_1}(x_1|y_L) f_Y(y_L). \]

If we write \( f_Y(y_L) \) as \( p_L = \Pr(Y_L = i\lambda); i=0,1,\ldots \) and note that only one \( f_X(x|i\lambda) = f_i(x) \) is non-zero for any given value of \( x \) then we can write the density of \( X \) as the mixture \( f_X(x) = \sum_{i=0}^{\infty} p_i f_i(x) \). We note that the \( X_i \)'s have disjoint supports and further that the union of these supports is the range of \( X \). This decomposition can be generalized to irregular spacing.

Decomposition of a r.v. \( X \) is therefore equivalent in the above sense to the class of mixtures on a family of densities whose supports partition the support of the density of \( X \). Mixtures and \( r.v. \) decompositions however still remain as distinct concepts and tools for gaining different perspectives of particular statistical problems.
8. WAITING TIME TECHNIQUES

The rejection technique was seen in Section 4 to consist of two parts, generating an observation \( Y = y \) and then deciding whether to set \( X = y \) or to reject \( y \) and generate a new value of \( Y \). The waiting time technique consists of generating observations \( Y_1, Y_2, \ldots \) until a particular event occurs and then setting \( X \) equal to some function of the observations \( Y_1, \ldots, Y_n \). This extension of the rejection technique is used in Examples 8.1 and 8.2.

We observe that in Section 6, \( X \) is decomposed into the integral part \( Y \) and the fractional part \( Z \), where \( f_Y(y) = F_X(y+1) \). It follows that \( f_Y(y) = F_X(y+1) - F_X(y) \). We observe that for a process operating in time, involving inverse sampling, the probability that the stopping time is in the \( n \)th period \( (n,n+1] \) is \( G(n+1) - G(n) \), \( n = 0,1, \ldots \) where \( G(n) \) is the probability of the process continuing beyond the \( n \)th period. Thus the integral part \( Y \) can be realized as a suitable waiting time. In Example 8.1 the integral part \( Y \) of the exponential r.v. \( X \) has been noted to have the geometric distribution, which arises as a distribution for the waiting time in classical inverse binomial sampling.

Example 8.1. Marsaglia (1961) generates an exponential r.v. \( X \) by generating first a discrete part \( Y \) and then a continuous part \( Z \). A sequence of Poisson r.v.'s with parameter 1 is generated; the sequence stops with the first non-zero r.v. which we will index by \( M \). We observe that the integral part \( Y = M-1 \) is therefore generated by a waiting time technique, and \( M \) has a geometric distribution with parameter \( p = 1-e^{-l} \). Marsaglia uses the inverse transformation of the geometric distribution as an equivalent mechanism for generating \( M \). This method is based on the fact that the time until the first Poisson event has an exponential distribution. We note that \( Z = \min(U_1, \ldots, U_M) \) where \( N \) is the first non-zero Poisson r.v., and that although the above method produces an exponential r.v. with parameter 1 a r.v. \( X \) with arbitrary parameter \( \lambda > 0 \) can be generated by multiplying by the appropriate constant.

Example 8.2. To generate a gamma r.v. \( X \) with integer parameters \( \alpha = k \) and \( \beta = 1 \), Shibuya (1961) generates a sequence of \( M \) Poisson r.v.'s stopping when their sum first exceeds \( k \). Similar to Example 8.1 the integer part of \( X \) is \( Y = M - 1 \) which is a waiting time random variable. This method is based upon the fact that the time until the \( k \)th Poisson event has a gamma distribution. The number of Poisson trials to obtain a total of \( k \) or more events is \( M \). We note that a gamma r.v. with arbitrary parameter \( \beta \) can be generated by multiplying the above r.v. \( X \) by the appropriate constant.
9. METHODS USING COMBINATIONS OF TECHNIQUES

We now present further examples of methods of generating random variables which combine several of the above techniques.

Example 9.1. In Example 2.5 the polar method for generating two normal r.v.'s was presented. Marsaglia (1962a) modified this method using rejection techniques to eliminate the sin and cos functions. The equivalent technique is to generate pairs of independent uniform (0,1) r.v.'s \((U_1, U_2)\) until \(W=U_1^2 + U_2^2 \leq 1\). Then set

\[
X_1 = U_1 \left( -\frac{2 \ln W}{W} \right)^{1/2} \quad \text{and} \quad X_2 = U_2 \left( -\frac{2 \ln W}{W} \right)^{1/2}.
\]

This results in \(X_1\) and \(X_2\) being independent standard normal random variables.

Example 9.2. Marsaglia's rectangular-Wedge-tail method (1962b) is one of the most efficient methods of generating a normal r.v. on a computer. It involves several of the mixture techniques. The normal density is decomposed into two symmetric halves (as in Example 3.1) and the positive half is then decomposed initially into a center region \(0 \leq x \leq 3\) and a tail region \(x > 3\). The center region r.v. is decomposed further into a lattice r.v. on the lattice \((0,1/4,\ldots,3)\) and a continuous r.v. on the interval \((0,1/4)\). This continuous r.v., in turn is generated as a mixture of two uniform (rectangular) and one wedge distribution. The uniform corresponding to the largest rectangle is chosen to occur with probability a multiple of 1/256 in order to use binary properties of the computer. The other uniform corresponds to the remaining rectangle. The wedge is generated by the method of nearly linear densities as in Example 4.4 and the tail \(X > 3\) by a modification of the polar method.

This method is essentially a mixture on 74 different distributions (37 for the half normal). It is extremely efficient because the binary properties are exploited.

Example 9.3. Forsythe (1972), Ahrens and Dieter (1973) and Dieter and Ahrens (1973) have developed a more versatile method for generating normal random variables. The half-normal r.v. \(X\) is partitioned into

\[
f(x) = a e^{-\left(x^2 - a^2\right)/2}
\]

on the intervals \([a_i, a_{i+1}]\) \(i = 1, 2, \ldots\) which are generated by rejection techniques. Various cases of this method give one a choice between moderate efficiency with simple programs or high efficiency with more lengthy programs. The resulting densities are generated in a manner similar to those used in Example 9.2.
10. CHARACTERIZATION THEOREMS

We now consider some characterization theorems motivated by mechanisms for the generation of r.v.'s. Patil (1963) discusses the linear exponential family of distributions with density function \( f_X(x) = a(x) e^{\beta x} \) where \( x \) takes values in some interval. A special case of this is what we call a linear exponential family with constant coefficient given by density \( f_X(x) = e^{\beta x} / (e^{\beta d} - 1) \) if \( 0 < x < d \) (\( d \) may be \( \infty \) in which case \( x \) has a negative exponential distribution).

Theorem 10.1. Let \( f_X(x) > 0 \) if \( 0 < x < 1 \) and 0 otherwise. Further assume \( f_X(x) \) is continuous on \([0,1]\). Then \( X \) has a linear exponential family distribution with constant coefficients if and only if \( Y(k) = [X]_k \) and \( Z(k) = (X)_k \) are independent for all \( k = n^{-1}, n^{-1}, \ldots \).

Proof. Let \( f_X(x) = e^{\beta x} / (e^{\beta d} - 1) \) if \( 0 < x < 1 \). Then \( f_{Y(k), Z(k)}(y(k), z(k)) = f_X(y(k) + z(k)) / (e^{\beta d} - 1) \) factors into a function \( y(k) \) times a function of \( z(k) \). Therefore \( Y(k) \) and \( Z(k) \) are independent.

Conversely, assume \( Y(k) \) and \( Z(k) \) are independent for all \( k = n^{-1}, n^{-1}, \ldots \). Let \( Y(k) = [X]_k \) and \( Z(k) = (X)_k \). Then \( f_X(x) = f_X(y(k) + z(k)) = f_Y(y(k)) f_Z(z(k)) \).

Substituting \( y(k) = 0 \), \( z(k) = 0 \) and both \( y(k) = 0 \) and \( z(k) = 0 \) into the above equation leads to \( f_X(x) = f_X(y(k)) f_X(z(k)) \) / \( f_X(0) \). Then \( f(y(k) + z(k)) = f(y(k)) f(z(k)) \) where \( g(x) = f_X(x) / f_X(0) \), for all \( k = n^{-1}, n^{-1}, \ldots \). This leads to Cauchy's equation \( g(y+z) = g(y) g(z) \) for all points \( y, z \) for which \( y = k/n, k=0,1,\ldots, n-1, \) \( z \) in \([0, n^{-1}]) \), \( n=1,2,\ldots \). Cauchy's equation can be shown to hold for arbitrary \( y, z \); \( 0 < y, 0 < z, y+z < 1 \) by an induction type argument as follows. First observe that \( g(1/2 + z) = g(1/2) g(z) \) if \( 0 < z < 1/2 \).

Then \( g(1/4 + z) = g(1/4) g(z) \) if \( 0 < z < 1/4 \).
Now \( g(1/4 + z) = g(1/2 + (z - 1/4)) = g(1/2) g(z - 1/4) = g(1/4) g(1/4) g(z - 1/4) = g(1/4) g(z) \) if \( 1/4 < z < 1/2 \). Of course \( g(x) \) is continuous since \( f(x) \) is continuous. Therefore \( g(1/4 + 1/4) = g(1/4) g(1/4) \), etc. Note continuity at a point for Cauchy's equation extends to continuity everywhere.

Theorem 10.2. Suppose \( f_X(x) > 0 \) if \( x > 0 \) and is zero elsewhere and \( f_X(x) \) is continuous.
Let \( Y(k) = [X]_k \), \( Z(k) = (X)_k \). Then \( Y(k) \) and \( Z(k) \) are independent for all \( k > 0 \) if and only if \( X \) has a negative exponential distribution.

Proof. Assume \( Y(k) \) and \( Z(k) \) are independent for all \( k > 0 \). Similar to the proof of Theorem 10.1 we obtain \( g(x) = g(y(k)) g(z(k)) \), where
\( g(x) = f_X(x)/f_X(0) \). Consider any two positive real numbers and let \( y \) be the larger and \( z \) be the smaller. Let \( x = y + z \) and \( z = y \), then \( y = |x|_b = y \) and \( z = x \).

We therefore have Cauchy's equation \( g(x+y) = g(x) g(y) \) for arbitrary \( x > 0 \) and \( y > 0 \). Since \( f_X(x) \) is a density then \( X \) has a negative exponential distribution. The converse is similar to the proof of Theorem 10.1.

In many of his papers Marsaglia uses individual random bits from a uniform r.v. to efficiently generate other r.v.'s. These bits are independent fair 0,1 trials. Marsaglia (1971) studies r.v.'s whose binary expansions have bits which are independent but not identically distributed. Necessary and sufficient conditions are given for the distributions to be discrete, singular, and absolutely continuous. We use a specific result in the following theorem.

**Lemma 10.1.** Suppose \( X \) has a linear exponential family distribution with constant coefficients on \((0,1)\) with density \( f_X(x) = \lambda e^{\beta x}/(e^{\beta} - 1), \) \( 0 < x < 1, -\infty < \beta < \infty \). Then \( X \) has the lack of memory property. That is \( f_{X-a} (y | a < x < a+d) \) does not depend on \( a \).

**Proof.** Now \( f_X(x | a < x < a+d) = \lambda e^{\beta (x-a)}/(e^{\beta d} - 1) \) if \( a < x < a+d \). Let \( Y = X-a \) then \( f_Y(y | a < x < a+d) = \lambda e^{\beta Y}/(e^{\beta d} - 1) \) if \( 0 < y < d \). Note if \( a+d > 1 \) a simple modification is necessary.

**Theorem 10.2.** Let the binary expansion of \( X \) be \( X = B_1 B_2 \ldots \sum_{k=1}^{\infty} B_k / 2^k \)
where \( X \) is a r.v. taking values in \((0,1)\). Further assume \( X \) has a distribution function which is differentiable at \( 2^{-k} \); \( k=0,1, \ldots \). Then \( X \) has a linear exponential family distribution with constant coefficients if and only if \( \{B_i : i=1,2, \ldots \} \) are independent.

**Proof.** Marsaglia (1971) proves the "if" part. We prove the converse. Let \( f_X(x) = \lambda e^{\beta x}/(e^{\beta} - 1) \) if \( 0 < x < 1 \). Observe the event \( [B_1 = b_1, \ldots, B_n = b_n] = \{a < X < a+2^{-n}\} \) for some \( a \). Further, \( X \) would be restricted to one half of this interval depending on the value \( B_n+1 \) takes. Thus \( P(B_n+1 = 0 | B_1 = b_1, \ldots, B_n = b_n) = P(\text{if } X < a+2^{-n}) - P(X < a+2^{-n-1}) \) where \( Y = X-a \). By Lemma 10.1 this does not depend on \( a \). Therefore \( \{B_i : i=1,2, \ldots \} \) are independent, which concludes the proof.

**Theorem 10.3.** Let \( X \geq 0 \) have distribution function \( F_X(x) \) which is differentiable at the points \( n+2^{-k} \) where \( n \) and \( k \) are non-negative integers. Let \( X = Y + Z \) be the decomposition into integer and fractional parts and let \( Z = \sum_{k=1}^{\infty} B_k / 2^k \) be the binary
expansion of $Z$. Then the binary digits, $b_i$, $i=1,2,\ldots$, are independent if and only if $X$ has a negative exponential distribution.

Proof. Assume $X$ has a negative exponential distribution then $Z$ has a $(0,1)$ truncated negative exponential distribution.

By Theorem 10.3 the binary expansion of $Z$ has independent bits. Conversely assume that the bits are independent then by Theorem 10.3 $Z$ has a linear exponential family distribution with constant coefficients. Since $f_X$ is differentiable on the integers $X$ has a negative exponential distribution.

REFERENCES


STATISTICAL DISTRIBUTIONS IN SCIENTIFIC WORK

Contents

VOLUME 1: MODELS AND STRUCTURES

Section

1. INAUGURAL ADDRESS - G. P. Patil
4. MOMENTS-RELATED PROBLEMS - M. S. Ramanujan, C. C. Heyde, W. L. Harkness

VOLUME 2: MODEL BUILDING AND MODEL SELECTION


5. MODELS IN THE ENVIRONMENTAL SCIENCES - M. F. Dacey, D. V. Gokhale, C. Ramachandran, W. G. Warren

6. A MODERN COURSE ON STATISTICAL DISTRIBUTIONS - Participants

VOLUME 3: CHARACTERIZATIONS AND APPLICATIONS

1. LINNIK MEMORIAL INAUGURAL LECTURE - C. R. Rao

2. MATHEMATICAL tools for CHARACTERIZATION PROBLEMS - E. Lukacs, J. Aczél, H. J. Rossberg


5. CHARACTERIZATIONS ON SPACES AND PROCESSES - I. I. Kotlarski, B. L. S. Prakasa Rao, R. Ahmad, E. Seneta


7. CHARACTERIZATIONS USING ENTROPY MEASURES AND RELATED PROBLEMS - D. V. Gokhale, R. Shimizu, A. W. Kemp, J. Aczél


FOR YOUR LIBRARY COPY

These Proceedings of a Modern Course on Statistical Distributions in Scientific Work provide carefully edited materials in three comprehensive volumes on related problems of contemporary interest and concern, such as, Models and Structures, Model Building and Model Selection, and Characterizations and Applications. Every quantitative scientist and statistician should find them very useful and worthwhile for both research and instruction involving any aspects of distributions. Order your copy today.