POLYTECHNIC UNIVERSITY
Department of Computer and Information Science

GENERATING NON-UNIFORM RANDOM DEVIATES

K. Ming Leung

Abstract: Methods for generating random deviates that obey non-uniform probability distributions are discussed. Examples of these methods include the inverse function method, the superposition method, and the rejection method.

Directory

- **Table of Contents**
- **Begin Article**

Last Revision Date: February 17, 2004 Version 2.0

Table of Contents

1. Introduction

With a fixed number of trials N used in a Monte Carlo simulation, the probable error in estimating the mean is proportional to the standard deviation, σ, and therefore can be decreased by decreasing the variance. The variance is independent of N, but depends on the way in which the simulation is carried out, as we saw earlier when we computed the area of a circle. Techniques for reducing the variance of a simulation are therefore important in reducing the error. They are referred to as variance reduction (or importance sampling) techniques. These techniques require judicial uses of non-uniform random deviates, as we will discuss at length in the next chapter. The probability distribution densities of these non-uniform deviates are not constants. We will discuss how to generate them from the uniform random deviates in this chapter. [1, 2]

We consider a continuous random variable X whose values, x, obey a probability density function $f(x)$. By definition, $f(x)$ must be non-negative and must have unit area. The (cumulative) probability distribution function, $F(x)$, is defined in terms of the probability density by

$$F(x) = \int_{-\infty}^{x} f(x)\ dx, \tag{1}$$

and therefore the probability density function is given in terms of the probability distribution function by differentiation

$$f(x) = \frac{dF(x)}{dx}. \tag{2}$$

The task is to use the uniform deviates to generate random deviates that are distributed according to any given probability density function $f(x)$.

2. Modeling a Discrete Random Variable

We will first consider modeling discrete random variables here. Let X be a discrete random variable whose values x are distributed according to the following table:

x	x_1	x_2	$\cdots$	x_n
$p(x)$	p_1	p_2	$\cdots$	p_n

with a fixed n. Normalization condition requires that $\sum_{k=1}^{n} p_k = 1$.

To generate discrete random deviates with the above distribution, we break up the interval $[0, 1]$ into n segments with lengths p_1, p_2, $\ldots$, p_n. A uniform deviate u is then picked. If u lies inside the k-th interval, that is $p_1 + p_2 + \ldots + p_{k-1} < u < p_1 + p_2 + \ldots + p_{k-1} + p_k$, where we define p_0 as 0, then we choose x_k for the value of x.

In most computer languages, the above procedure can be implemented using n nested-if-else statements:

```
if $u < p_1$
    $x = x_1$
else if $u < p_1 + p_2$
    $x = x_2$
else if ...
    ...
else if $u < p_1 + \ldots + p_{n-1}$
    $x = x_{n-1}$
else
    $x = x_n$
```

Note that the precise ordering of these intervals does not matter at all. However, we should put the larger intervals first to improve the efficiency in the search. If n is very large, one may also want to consider other alternate ways of searching for the right interval.[10]

For the special case of equal probabilities: $p_k = 1/n$, for all k, we pick the value x_k if $p_1+p_2+\ldots+p_{k-1} < u < p_1+p_2+\ldots+p_{k-1}+p_k$. But $p_1+p_2+\ldots+p_{k-1} = (k-1)/n$ and $p_1+p_2+\ldots+p_{k-1}+p_k = k/n$, and so this condition is the same as $k-1 < nu < k$. If we use the notation $[x]$ to denote the integer part of x, then the condition becomes $[nu] = k-1$. Therefore if the chosen uniform deviate is such that $[nu]+1 = k$ then we should pick $x = x_k$.

Some computer languages, such as C and C++, has built-in generators that produce random integers, I, varying between 0 and a large global integer constant `RAND_MAX`. In that case we can simply compute k as $1 + I$ modulo n.

3. The Inverse Function Method

The Inverse function method provides a general scheme for generating non-uniform random deviates from the uniform random deviates. The method involves finding a certain indefinite integral and inverting it to find its inverse function. We cannot expect to be able to perform these steps analytically for any arbitrary probability distribution. Fortunately there are a number of rather common forms of density functions that are amenable to this method.

Let Y be a random variable whose values, y, are the uniform deviates in the interval $[0, 1]$. We want to find a transformation from y to x in such a way that the values of y are distributed according to the probability density function $f(x)$ of interest. We can obtain such a transformation by first considering the probability $P(y \leq Y \leq y+dy)$ for finding Y having a value within the interval $[y, y+dy]$. Since y has a uniform distribution, it is clear that $P(y \leq Y \leq y + dy) = dy$, the width of that interval. Next we consider $P(x \leq X \leq x + dx)$, which by definition is given by

$$P(x \leq X \leq x + dx) = \int_{x}^{x+dx} f(x)\ dx = f(x)\ dx. \tag{3}$$

The second equality can be obtained by differentiating the upper limit of the integral. However we must have

$$P(y \leq Y \leq y + dy) = P(x \leq X \leq x + dx), \tag{4}$$

because each side of the equation represents exactly the same probability although they are expressed in terms of different random variables. Therefore we have the result[8]

$$dy = f(x)\ dx. \tag{5}$$

Integrating both sides of this equation, yields a relationship expressing y as a function of x:

$$y = F(x). \tag{6}$$

The constant of integration here must be zero since both y and $F(x)$ must vary between 0 and 1.[9] Finally we must invert this relationship to obtain an expression for x in terms of y:

$$x = F^{-1}(y), \tag{7}$$

where F^{-1} is the inverse function of F.

Using this equation with uniform deviates, y, we can generate random deviates, x, that are guaranteed to be distributed according to the given probability density function $f(x)$. However the above procedure requires one first to find $F(x)$, but that involves performing an indefinite integral over $f(x)$. In addition, one has to find the inverse function F^{-1} from F. These steps can be performed analytically only for a few rather simple but sometimes useful probability density functions. In those cases the method is very simple and very efficient. For more general forms of $f(x)$, we will have to resort to numerical means.[5] In addition, other methods for generating non-uniform deviates often rely indirectly on the use of the inverse function method.

3.1. Algorithm of the Inverse Function Method

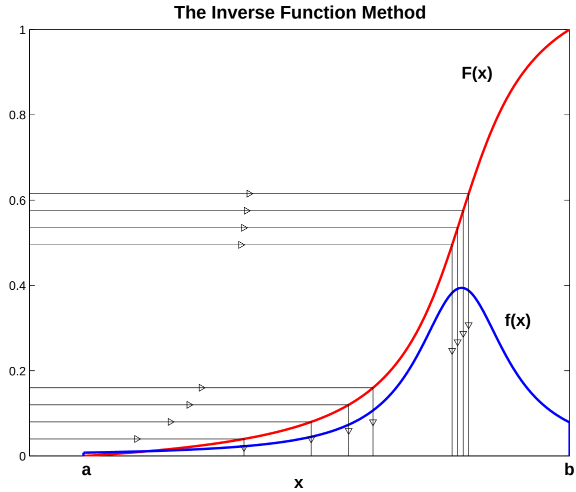
The inverse function method for a given density function, $f(x)$ is:

1. Find the indefinite integral of $f(x)$ to obtain $F(x)$.
2. Find the inverse function of F to obtain F^{-1}.
3. Values $x = F^{-1}(u)$ are distributed according to probability density function $f(x)$ if the u are uniform deviates.

3.2. Graphical Explanation of the Inverse Function Method

It is actually very easy to understand how the inverse function method works without using any mathematics. Recall that given a probability density function $f(x)$, the cumulative distribution function $F(x)$ gives the accumulated area under the $f(x)$ curve from $-\infty$ up to x.

The curve of $F(x)$ as a function of x is always monotonic increasing, however it remains rather flat in a region where $f(x)$ is small since there is little area to add up. On the other hand in a region where $f(x)$ has a peak, $F(x)$ picks up a lot of area quickly for the same amount of increase in x, and so it increases rapidly with increasing x, as shown in the figure.



For each uniform deviate, u, the inverse function method produces a value for x according to the formula

$$x = F^{-1}(u). \tag{8}$$

Graphically this means that we locate the point u on the vertical axis and find out where this point comes from on the horizontal axis according to the function $F(x)$. Now imagine randomly throwing points between 0 and 1 on the vertical axis covering it uniformly and finding where they land on the horizontal axis. It is clear that there are many more points hitting the curve $F(x)$ where it rises rapidly than in the place where it remains rather flat. Therefore more points will end up on the horizontal axis where the probability density function $f(x)$ is large compared with the places where it is small. This explains why the resulting points x are distributed according to $f(x)$.

3.3. Examples of the Inverse Function Method

We will use the the Inverse function method to generate some common non-uniform deviates. Some of these random deviates will be of use later in the treatment of variance reduction techniques in Monte Carlo simulations.

● Uniform Distributions

Our first example is a case in which x is restricted in the interval $[a, b]$ and the distribution density function $f(x) = c =$ constant within the interval, otherwise $f(x) = 0$. Normalization of $f(x)$ requires that $c = 1/(b-a)$.

First we have to calculate $F(x)$

$$F(x) = \int_{-\infty}^{x} f(x)\ dx = \frac{1}{b-a}\int_{a}^{x}\ dx = \frac{x-a}{b-a}. \tag{9}$$

Setting the value of a uniform deviate u equal to this $F(x)$ and inverting the result to find x as a function of u gives

$$x = (b-a)u + a, \tag{10}$$

which we obtained before when the uniform deviates are transformed from the fundamental interval $[0, 1]$ to the more general interval $[a, b]$.

• Exponential Distribution

Random deviates that are distributed according to an exponential distribution density function

$$f(x) = ae^{-ax}, \tag{11}$$

for $x \geq 0$, and $f(x) = 0$ for $x < 0$, can be generated using the inverse function method as well. The probability distribution function can be calculated:

$$F(x) = a \int_0^x e^{-ax} = 1 - e^{-ax}. \tag{12}$$

Setting this equal to a uniform deviate x and solving x in terms of u gives

$$x = -\frac{1}{a} \ln(1 - u). \tag{13}$$

Because the numbers u are uniform deviates then so are the numbers $1 - u$, we can rewrite the above equation simply as:

$$x = -\frac{1}{a} \ln(u). \tag{14}$$

- **Distribution Density Functions** $f(x) = (m+1)x^m$

Distribution density function given by $f(x) = \alpha x^m$ for x in $[0,1]$, and $f(x) = 0$ otherwise. We will determine the normalization constant α below.

$$F(x) = \alpha \int_0^x x^m \, dx = \alpha \left[\frac{x^{(m+1)}}{m+1} \right]_0^x = \alpha \frac{x^{(m+1)}}{m+1}. \qquad (15)$$

In order to satisfy the normalization condition $F(1) = 1$ we must have $\alpha = m + 1$.

We pick a uniform deviate u and set

$$u = F(x) = x^{m+1}. \qquad (16)$$

This relation is then solved for x in terms of u:

$$x = u^{1/(m+1)}. \qquad (17)$$

This formula enables us to generate random deviates x that are distributed with the prescribed density function.

Also note that for this distribution, the exact mean can be calculated:

$$E(X) = (m+1) \int_0^1 x^{m+1} \, dx = \frac{m+1}{m+2}, \tag{18}$$

and so is the exact value of the mean of X^2:

$$E(X^2) = (m+1) \int_0^1 x^{m+2} \, dx = \frac{m+1}{m+3}. \tag{19}$$

Therefore the exact variance is given by

$$V(X) = E(X^2) - (E(X))^2 = \frac{m+1}{(m+3)(m+2)^2}. \tag{20}$$

We can use these exact values for the mean and variance to check the properties of random deviates generated by the inverse function method, as well as other methods.

4. Superposition Method

The superposition (or also known as the composition) method [1, 4] can be applied to a probability distribution function $F(x)$ that can be written as a superposition of two or more probability distribution functions, $F_1(x), F_2(x), \ldots, F_m(x)$ so that

$$F(x) = \sum_{k=1}^{m} c_k F_k(x), \tag{21}$$

where all $c_k > 0$ and $\sum_{k=1}^{m} c_k = 1$. The method is useful if random variables with probability distribution function $F_k(x)$ can all be easily modeled, for example, using the inverse functions $F_k^{-1}(u)$, where u is a uniform deviate.

The generation of random deviates that are distributed according to probability function $F(x)$ relies on the use of a discrete random integer variable, Q, whose values, q, obey the following distribution

q	1	2	$\cdots$	m
$p(q)$	c_1	c_2	$\cdots$	c_m

in other words

$$P(Q = k) = c_k. \tag{22}$$

In this method, a random deviate u_1 is first used to select randomly an integer from 1 to m as the value of Q. If the chosen integer is k, then the inverse function for $F_k(x)$ is used to produce a random deviate

$$x = F_k^{-1}(u_2), \tag{23}$$

where u_2 is another uniform deviate. The resulting random deviates x will then be distributed according to the probability distribution function $F(x)$.

Note that by differentiating Eq.(21) with respect to x, we have

$$\begin{aligned} f(x) &= \frac{d}{dx}F(x) = \frac{d}{dx}\sum_{k=1}^{m} c_k F_k(x) \qquad (24) \\ &= \sum_{k=1}^{m} c_k \frac{d}{dx}F_k(x) = \sum_{k=1}^{m} c_k f_k(x). \end{aligned}$$

4.1. Algorithm for the Superposition Method

The algorithm for the Superposition Method is:

1. Randomly pick an integer from 1 to m according to the probability given in the table for the discrete variable Q, for example using the method we discussed earlier. Denote this chosen integer by k.
2. Choose a random deviate u_2.
3. A random deviate x is then given using the inverse function method by $F_k^{-1}(u_2)$.

4.2. Proof of the Superposition Method

We assume that the probability distribution function $F(x)$ can be written as a superposition of m probability distribution functions,

$F_1(x), F_2(x), \ldots, F_m(x)$ so that

$$F(x) = \sum_{k=1}^{m} c_k F_k(x), \tag{25}$$

where all $c_k > 0$ and $\sum_{k=1}^{m} c_k = 1$. Let u be a uniform deviate, and x be any real value. We want to consider the probability that the discrete random variable Q takes on an integer value such that $u < F_Q(x)$. Because of the monotonicity of the probability distribution functions and their inverse functions, that probability is the same as $P(F_Q^{-1}(u) < x)$. We can express it in terms of a sum of conditional probabilities:

$$P(F_Q^{-1}(u) < x) = \sum_{k=1}^{m} P(F_Q^{-1}(u) < x \mid Q = k) P(Q = k), \tag{26}$$

because Q takes on the values $1, 2, \ldots, m$ with corresponding probabilities $P(Q = k)$. Note that

$$P(F_Q^{-1}(u) < x \mid Q = k) = P(F_k^{-1}(u) < x) \tag{27}$$

and

$$P(F_k^{-1}(u) < x) = P(u < F_k(x)) = F_k(x). \tag{28}$$

This last equality holds because u is a random deviate between 0 and 1, and $F_k(x)$ is a probability.

4.3. Example of the Superposition Method

When low energy photons (light) are scattered by slowly moving electrons, the scattering angle θ is a random variable. The scattering angle measures the change in the angle between the direction of the incident photon and that of the scattered photon. The cosine of the angle $X = \cos\theta$ has values y obeying Rayleigh's law:

$$f(x) = \frac{3}{8}(1 + x^2), \qquad \text{for} - 1 \leq x \leq 1. \tag{29}$$

If we use the inverse function method here then we need to calculate

$$\begin{aligned} F(x) &= \int_{-1}^{x} \frac{3}{8}(1 + x^2)\ dx = \frac{3}{8}\left[x + \frac{x^3}{3}\right]_{-1}^{x} \qquad (30) \\ &= \frac{1}{8}(x^3 + 3x + 4). \end{aligned}$$

However finding the inverse function F^{-1} requires the solving of a cubic equation.

Instead we will use the superposition method. We need to decompose $p(y)$ into a linear combination of the two probability density functions $f_1(x) = \alpha_1$ and $f_2(x) = \alpha_2 x^2$, where α_1 and α_2 are constants to be determined from normalization. Normalization of $f_1(x)$ means that

$$\int_{-1}^{1} \alpha_1 \; dx = 2\alpha_1 = 1, \tag{31}$$

and so $\alpha_1 = \frac{1}{2}$ and $f_1(x) = \frac{1}{2}$. Similarly normalization of $f_2(x)$ means that

$$\int_{-1}^{1} \alpha_2 x^2 \; dx = \alpha_2 \left[\frac{x^3}{3} \right]_{-1}^{1} = \frac{2}{3}\alpha_2, \tag{32}$$

and so $\alpha_2 = \frac{3}{2}$ and $f_2(x) = \frac{3}{2}x^2$. We want to write

$$f(x) = c_1 f_1(x) + c_2 f_2(x), \tag{33}$$

and so

$$\frac{3}{8}(1 + x^2) = \frac{c_1}{2} + \frac{3}{2}c_2 x^2. \tag{34}$$

Equating terms having the same powers of x gives $c_1 = \frac{3}{4}$ and $c_2 = \frac{1}{4}$. Coefficients obtained this way always obey the normalization condition $\sum_{k=1}^{m} c_k = 1$ since each individual probability density function in Eq.(33) is normalized.

Next we integrate these probability density functions to find the probability distribution functions:

$$F_1(x) = \int_{-1}^{x} \frac{1}{2}\ dx = \frac{1}{2}(x+1). \tag{35}$$

$$F_2(x) = \int_{-1}^{x} \frac{3}{2}x^2\ dx = \frac{1}{2}(x^3+1). \tag{36}$$

If u_2 is a uniform deviate and we set $u_2 = F_1(x) = (x+1)/2$, we get $x = 2u_2 - 1$. On the other hand if we set $u_2 = F_2(x) = (x^3+1)/2$, we get $x = (2u_2 - 1)^{1/3}$. Therefore to generate random deviates according to the probability density function $f(x)$, we pick two uniform deviates, u_1 and u_2, and let

$$x = \begin{cases} 2u_2 - 1, & \text{if } u_1 < \frac{3}{4} \\ (2u_2 - 1)^{1/3}, & \text{if } u_1 > \frac{3}{4} \end{cases} \tag{37}$$

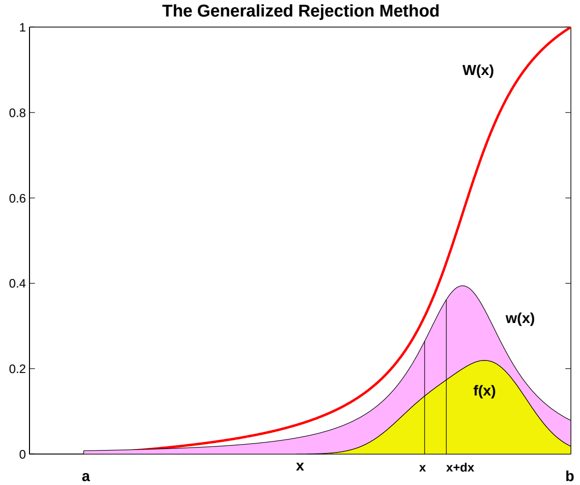
5. The Generalized Rejection Method

We discuss here the generalized rejection method for generating random deviates distributed with any given probability density function $f(x)$. The method is based on the following important observation. In the graph of $f(x)$ versus x, if we can generate points covering up the area under the curve uniformly, then the x-coordinates of these points will have values distributed with the probability density function $f(x)$.

For an arbitrary form of $f(x)$, generating these points is non-trivial. In the generalized rejection method, a comparison function $w(x)$ is chosen such that $w(x) \geq f(x)$ for all x within the domain of interest. We also want to choose this comparison function so that the indefinite integral

$$W(x) = \int_{-\infty}^{x} f(x)\, dx \tag{38}$$

can be calculated analytically, and is analytically invertible to obtain W^{-1}. Note that since $f(x)$ is normalized, therefore $w(x)$ is not normalized. In fact the area underneath $w(x)$, which we denote by A, must be larger than 1.



To generate random deviates x with the probability density function $f(x)$, we first use uniform deviates u to produce random numbers Au that are uniformly distributed in the interval $[0, A]$ along the vertical axis. The corresponding values of x on the horizontal axis are found using $x = W^{-1}(Au)$. For each of the values of x, a value of y is picked randomly and uniformly between 0 and $w(x)$. Clearly the points whose coordinates are given by (x, y) are uniformly distributed in the area under $w(x)$.

Next we must reject those points whose values of y lie above $f(x)$. The numbers obtained from the x-coordinates of the points that are retained will then be distributed with probability density $f(x)$.

The ratio of the number of points retained to the total number of points used to generate them in this method is called the efficiency, e. The value of e is therefore given by

$$e = \frac{\int_{-\infty}^{\infty} f(x)\ dx}{\int_{-\infty}^{\infty} w(x)\ dx} = \frac{1}{A}. \tag{39}$$

Consequently in order for the method to be efficient A should be only slightly larger than unity. That means that the comparison function $w(x)$ should only be slightly larger than $f(x)$ within the domain of interest.

The efficiency is unity for the special case in which we choose $w(x)$

equal to $f(x)$. No point is then rejected. However the method then becomes exactly the inverse function method, that means that we need to be able to find $F(x)$ and to invert it.

A simple but clearly not an optimal choice is to use a constant function given by $w(x) = \max f(x) = M$ in the domain. This method becomes the original rejection method of von Neumann, and is basically the same as the Hit-Or-Miss method. The efficiency of von Neumann's rejection method is $e = 1/((b - a)M)$.

5.1. Algorithm for the Generalized Rejection Method

The procedure for the generalized rejection method is summarized below.

1. Set N, the total number of random deviates wanted, to a large integer.
2. Initialize an integer, N' to 0.
3. Go through the following loop until $N' = N$:
   (a) Get a uniform deviate u_1 and compute Au_1.
   (b) Let $x = W^{-1}(Au_1)$.
   (c) Get another uniform deviate u_2 and let $y = u_2 w(x)$.
   (d) If $y < f(x)$, retain the value of x and increment N' by 1, otherwise reject the point (x, y).

5.2. Example of the Generalized Rejection Method

We will apply the Generalized Rejection method to generate random deviates distributed with the probability density function

$$f(x) = \frac{v(x)}{x^\alpha}, \quad \text{for } 0 \le x \le 1, \tag{40}$$

where the function $v(x)$ has a maximum v_m and must be such that $f(x)$ is normalized to unity. The parameter α must be restricted so that $\alpha < 1$.

We can choose a comparison function

$$w(x) = \frac{v_m}{x^\alpha}, \tag{41}$$

which is clearly larger than $f(x)$ for x in $[0, 1]$. Moreover, we can analytically obtain

$$W(x) = \int_0^x \frac{v_m}{x^\alpha}\, dx = \frac{v_m}{1-\alpha} x^{1-\alpha}. \tag{42}$$

At $x = 1$, $W(1) = \frac{v_m}{1-\alpha} = A$, the area under $w(x)$. Therefore

$$W(x) = Ax^{1-\alpha}. \tag{43}$$

We pick a uniform deviate u_1 and let $Au_1 = W(x)$. Inverting this gives $x = u_1^{1/(1-\alpha)}$. We pick another uniform deviate u_2 and keep this value for x if

$$w(x)u_2 < f(x), \tag{44}$$

which is the same relation as

$$v_m u_2 < v(x), \tag{45}$$

or

$$v_m u_2 < v(u_1^{1/(1-\alpha)}). \tag{46}$$

Although the parameter α can be negative, the above example is most useful when $0 < \alpha < 1$ and $v(x)$ is non-zero at $x = 0$. Then $f(x)$ diverges at $x = 0$, but in such a way that it has unit area as required. The use of the comparison function $w(x) = v_m x^{-\alpha}$, which itself is divergent at 0, allows us to take care of the divergence of $f(x)$ at 0. The remaining part, described by $v(x)$, has a smoother behavior and therefore can be modeled more accurately.

References

[1] Ilya M. Sobol, *A Primer for the Monte Carlo Method*, (CRC Press, 1994). 3, 17

[2] D. E. Knuth, *The Art of Computer Programming: Vol. 2: Seminumerical Algorithms*, Second Edition, (Addison Wesley, 1981). 3

[3] J. Banks, J. S. Carson II, B. L. Nelson, and D. M. Nicol, *Discrete-Event System Simulation*, Third Edition, Ch. 8 (Prentice Hall, 2000).

[4] A. M. Law and W. D. Kelton, *Simulation Modeling and Analysis*, (McGraw-Hill, New York, 2000). 17

[5] P. Bratley, B. L. Fox and E. L. Schrage, *A Guide to Simulation*, (Springer-Verlag, New York, 1983).

[6] L. Devroye, *Non-Uniform Random Variate Generation*, (Springer-Verlag, New York, 1986). 9

[7] D. D. Wackerly, W. Mendenhall III, and R. L. Scheaffer, *Mathematical Statistics with Applications*, Sec. 8.3, (Wadssworth Publishing, 1996).

[8] This is actually a special form of the transformation law for probabilities. In general, if random variable Y has values y distributed with probablitiy $f_y(y)$, and if Y is transformed to a new random variable X via the relation $X = T(Y)$, where T is some kind of transformation function, then X will have values x distributed according to probability density function $f_x(x)$ which obeys the relation:

$$|f_x(x)\ dx| = |f_y(y)\ dy|$$

See, for example, D. D. Wackerly, W. Mendenhall III, and R. L. Scheaffer, *Mathematical Statistics with Applications*, p. 267, (Wadssworth Publishing, 1996).

[9] Of course we can also have chosen the opposite sign so that $-dy = f(x)\ dx$, which applies if x decreases while y increases, and vice versa. In that case there is an integration constant which must be chosen to give the result $1 - y = F(x)$. But since y is a uniform

deviate in the interval $[0, 1]$, so is $1 - y$, and so the resulting transformation between y and x is essential unchanged. 8

[10] D. E. Knuth, *The Art of Computer Programming: Vol. 3: Sorting and Searching*, Ch. 6, (Addison Wesley, 1973).

8

6