THE IVERSE PROBABILITY METHOD FOR GENERATING RANDOM VARIABLES*

Ewa Paszek

This work is produced by OpenStax-CNX and licensed under the
Creative Commons Attribution License 2.0†

Abstract

This course is a short series of lectures on Introductory Statistics. Topics covered are listed in the Table of Contents. The notes were prepared by Ewa Paszek and Marek Kimmel. The development of this course has been supported by NSF 0203396 grant.

1 THE IVERSE PROBABILITY METHOD FOR GENERATING RANDOM VARIABLES

Once the generation of the uniform random variable[1] is established, it can be used to generate other types of random variables.

1.1 The Continuous Case

THEOREM I

Let **X** have a continuous distribution $F_X(x)$, so that $F_X^{-1}(\alpha)$ exists for $0 < \alpha < 1$ (and is hopefully countable). Then the random variable $F_X^{-1}(U)$ has distribution $F_X(x)$, **U** is uniformly distributed on (0,1).

PROOF

$$P\left(F_X^{-1}(U) \leq x\right) = P\left(F_X\left(F_X^{-1}(U)\right) \leq F_X(x)\right). \tag{1}$$

Because $F_X(x)$ is monotone. Thus,

$$P\left(F_X^{-1}(U) \leq x\right) = P\left(U \leq F_X(x)\right) = F_X(x). \tag{2}$$

The last step follows because **U** is uniformly distributed on (0,1). Diagrammatically, we have that $(X \leq x)$ if and only if $[U \leq F_X(x)]$, an event of probability $F_X(x)$.

As long as we can invert the distribution function $F_X(x)$ to get the inverse distribution function $F_X^{-1}(\alpha)$, the theorem assures us we can start with a pseudo-random uniform variable **U** and turn into a random variable $F_X^{-1}(U)$, which has the required distribution $F_X(x)$.

*Version 1.3: Oct 8, 2007 4:24 pm -0500

†http://creativecommons.org/licenses/by/2.0/

[1] "PSEUDO-NUMBERS": Section UNIFORM PSEUDO-RANDOM VARIABLE GENERATION <http://cnx.org/content/m13103/latest/#sec_1>

Example 1
The Exponential Distribution

Consider the exponential distribution defined as

$$\alpha = F_X(x) = \{ \begin{array}{l} 1 - e^{-\lambda x}, \lambda > 0, x \geq 0, \\ 0, x < 0. \end{array} \tag{3}$$

Then **f** or the inverse distribution function we have

$$x = -\frac{1}{\lambda} \ln(1-\alpha) = F^{-1}(\alpha). \tag{4}$$

Thus if **U** is uniformly distributed on 0 to 1, then $X = -\frac{1}{\lambda}\ln(1-U)$ has the distribution of an exponential random variable with parameter λ. We say, for convenience, that **X** is exponential (λ).

NOTE: If **U** is uniform (0,1), then so is (1-**U**), and the pair **U** and (1-**U**) are interchangeable in terms of distribution. Hence, $X' = -\frac{1}{\lambda}\ln(U)$ is exponential. However, the two variables **X** and **X'** are correlated and are known as **an antithetic pair**.

Example 2
Normal and Gamma Distributions

For both these cases there is no simple functional form for the inverse distribution $F_X^{-1}(\alpha)$, but because of the importance of the Normal and Gamma distribution models, a great deal of effort has been expended in deriving good approximations.

The Normal distribution is defined through its density,

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right]. \tag{5}$$

So that,

$$F_X(x) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma}} \exp\left[\frac{-(x-u)^2}{2\sigma^2}\right] dv. \tag{6}$$

The normal distribution function $F_X(x)$ is also often denoted $\Phi(x)$, when the parameter **u** and σ are set to 0 to 1, respectively. The distribution has no closed-form inverse, $F_X^{-1}(\alpha)$, but the inverse is needed do often that $\Phi^{-1}(\alpha)$, like logarithms or exponentials, is a system function.

The inverse of the Gamma distribution function, which is given by

$$F_X(x) = \frac{1}{\Gamma(k)} \int_0^{kx/u} v^{k-1} e^{-v} dv, x \geq 0, k > 0, u > 0. \tag{7}$$

Is more difficult to compute because its shape changes radically with the value of **k**. It is however available on most computers as a numerically reliable function.

Example 3
The Normal and Gamma Distributions

A commonly used symmetric distribution, which has a shape very much like that of the Normal distribution, is the standardized logistic distribution.

$$F_X(x) = \frac{e^x}{1+e^x} = \frac{1}{1+e^{-x}}, -\infty < x < \infty, \tag{8}$$

with probability density function

$$F_X(x) = \frac{e^x}{1+e^x} = \frac{1}{1+e^{-x}}, -\infty < x < \infty. \tag{9}$$

NOTE: $F_X(-\infty) = e^{-\infty}/(1+e^{-\infty}) = 0$ and $F_X(\infty) = 1$ by using the second form for $F_X(x)$.

The inverse is obtained by setting $\alpha = \frac{e^x}{1+e^x}$. Then, $\alpha + \alpha e^x = e^x$ or $\alpha = e^x(1-\alpha)$.

Therefore,

$$x = F_X^{-1}(\alpha) = \ln\alpha - \ln(1-\alpha).$$

And the random variable is generated, using the inverse probability integral method. As follows $X = \ln U - \ln(1-U)$.

1.2 The Discrete Case

Let **X** have a discrete distribution $F_X(x)$ that is, $F_X(x)$ jumps at points $x_k = 0, 1, 2, \ldots$. Usually we have the case that $x_k = k$, so that **X** is an integer value.

Let the probability function be denoted by

$$p_k = P(X = x_k), k = 0, 1, \ldots. \tag{10}$$

The probability distribution function is then,

$$F_X(x_k) = P(X \le x_k) = \sum_{j \le k} p_j, k = 0, 1, \ldots, \tag{11}$$

and the reliability or survivor function is

$$R_X(x_k) = 1 - F_X(x_k) = P(X > x_k), k = 0, 1, \ldots. \tag{12}$$

The survivor function is sometimes easier to work with than the distribution function, and in fields such as reliability, it is habitually used. The inverse probability integral transform method of generating discrete random variables is based on the following theorem.

THEOREM

Let **U** be uniformly distributed in the interval (0,1). Set $X = x_k$ whenever $F_X(x_{k-1}) < U \le F_X(x_k)$, for $k = 0, 1, 2, \ldots$ with $F_X(x_{-1}) = 0$. Then **X** has probability function p_k.

PROOF

By definition of the procedure,

$X = x_k$ if and only if $F_X(x_{k-1}) < U \le F_X(x_k)$.

Therefore,

$$P(X = x_k) = PF_X((x_{k-1}) < U \le F_X(x_k)) = F_X(x_k) - F(x_{k-1}) = p_k. \tag{13}$$

By the definition of the distribution function of a uniform (0,1) random variable.

Thus the inverse probability integral transform algorithm for generating **X** is to find x_k such that $U \le F_X(x_k)$ and $U > F_X(x_{k-1})$ and then set $X = x_k$.

In the discrete case, there is never any problem of numerically computing the inverse distribution function, but the search to find the values $F_X = (x_k)$ and $F_X(x_{k-1})$ between which **U** lies can be time-consuming, generally, sophisticated search procedures are required. In implementing this procedure, we try to minimize the number of times one compares **U** to $F_X = (x_k)$. If we want to generate many of **X**, and $F_X = (x_k)$ is not easily computable, we may also want to store $F_X = (x_k)$ for all **k** rather than recomputed it. Then we have to worry about minimizing the total memory to store values of $F_X = (x_k)$.

Example 4
The Binary Random Variable

To generate a binary-valued random variable **X** that is 1 with probability **p** and 0 with probability 1-**p**, the algorithm is:

- If $U \leq p$, set **X**=1.
- Else set **X**=0.

Example 5
The Discrete Uniform Random Variable

Let **X** take on integer values between and including the integers **a** and **b**, where $a \leq b$, with equal probabilities. Since there are $(b-a+1)$ distinct values for **X**, the probability of getting any one of these values is, by definition, $1/(b-a+1)$. If we start with a continuous uniform (0,1) random number **U**, then the discrete inverse probability integral transform shows that

X= integer part of $[(b-a+1)U+a]$.

NOTE: The continuous random variable $[(b-a+1)U+a]$ is uniformly distributed in the open interval $(a, b+1)$.

Example 6
The Geometric Distribution

Let **X** take values on zero and the positive integers with a geometric distribution. Thus,

$$P(X=k) = p_k = (1-\rho)\rho^k, k = 0, 1, 2, ..., 0 < \rho < 1, \tag{14}$$

and

$$P(X \leq k) = F_X(k) = 1 - \rho^{k+1}, k = 0, 1, 2, ..., 0 < \rho < 1. \tag{15}$$

To generate geometrically distributed random variables then, you can proceed successively according to the following algorithm:

- Compute $F_X(0) = 1-\rho$. Generate **U**.
- If $U \leq F_X(0)$ set **X**=0 and exit.
- Otherwise compute $F_X(1) = 1-\rho^2$.
- If $U \leq F_X(1)$ set **X**=1, and exit.
- Otherwise compute $F_X(2)$, and so on.