4

Probability

The **probability** of an event is a ratio of the number of ways the event can occur to the total number of possible outcomes of the random experiment. For example, in a single coin toss the probability of a head is

$$P(\text{HEADS}) = (\text{\# of ways a HEAD can occur})/(\text{total \# of outcomes}) = 1/2$$

For this simple example, it is easy to count the total number of outcomes.

This chapter describes the application of probability theory to more complex random experiments. The concept of random variables is introduced because of its importance to probability theory. A **random variable** is any function that maps all the events of a probabilistic experiment (the sample space) to the x-axis. Each random variable has its own characteristic **probability density function** and **probability distribution function**. These functions are very useful in determining probabilities and are also defined in this chapter.

Each probability density function has **mean** and **variance**. The mean is the average or expected value of the random variable. The variance determines the spread of events about the mean. The **standard deviation** is the square root of the variance. These concepts are defined in more detail in the statistics chapter.

4.1 COMBINATORIAL ANALYSIS

Many applications of probability involve experiments that are much more complex than the coin toss. In many games of chance, the number of outcomes is purposely large to make the game interesting. For these cases, it may be difficult to count each and every outcome. For many problems, **combinatorial analysis** can be used to provide closed form expressions of the number of outcomes without having to enumerate them.

4.1.1 Permutations and Factorial

A **permutation** is any possible arrangement of a set of distinct items. The order of the arrangement of the items is important. For example, *abcd* is different from permutation *dcba*.

The number of different ways of arranging (permuting) n distinct objects is

$$n! = (n) * (n-1) * (n-2) * \ldots 1$$

where n is an integer.

The expression $n!$ is called *n-factorial.* Zero factorial is defined to be unity (i. e., $0! = 1$).

The number of different orderings of r objects from a total of n, ${}_nP_r$, is given by:

$${}_nP_r = n!/(n-r)!$$

where $r < n$ and both n and r are integers. Note that if $r = n$ then ${}_nP_n = n!$.

See Functions: permut(), fact().

4.1.2 Combinations

Often only the number of occurrences of a specific outcome is important, not the order. For example, while a poker player usually hopes to be dealt four aces, the order in which the aces are received is not important. The number of *combinations* refers to the different ways of choosing r objects from a total of n, without regard to their order. The number of such combinations is usually denoted by ${}_nC_r$.

$${}_nC_r = n!/(r! * (n-r)!)$$

where both n and r are integers and $r \leq n$.

There are several useful identities involving the combinations function. Perhaps the most important of these is

$${}_nC_r = {}_nC_{n-r}$$

This formula is intuitively clear, since when we explicitly decide to choose r objects at a time, we are implicitly deciding to exclude $n-r$ objects at a time.

See Function: combin().

4.1.3 Binomial Expansion and Pascal's Triangle

Sometimes the combinations expressions, ${}_nC_r$, are referred to as binomial coefficients because of their relation to the binomial expansion:

$$(x+y)^n = {}_nC_0 x^n y^0 + {}_nC_1 x^{n-1} y^1 + \ldots + {}_nC_n x^0 y^n$$

For different powers of n, the coefficients of this expansion form a pattern that is called **Pascal's Triangle**, as shown in the table.

n	Pascal's Triangle and the Binomial Coefficients
0	1
1	1 1
2	1 2 1
3	1 3 3 1
4	1 4 6 4 1
5	1 5 10 10 5 1

4.2 DISCRETE RANDOM VARIABLES

Discrete random variables are useful in describing events that can have only a finite or countable (e. g., integer) number of outcomes. For a discrete random variable to be properly defined, the sum of the probabilities of all the discrete events must be one. Collectively, the probabilities are referred to as the **probability density function,** $f(k)$, since their envelope describes the distributional characteristics of the random variable.

The **cumulative probability distribution function,** $F(k_0)$, of a discrete random variable is defined as the sum of the density $f(k)$ over k, where $k \leq k_0$. For a given random variable, these two functions are all that is needed to calculate the probability of any outcome.

4.2.1 Binomial Distribution

The **binomial** distribution is frequently used to predict the outcome of repetitive sampling from a population. The underlying assumption behind the binomial distribution is that after each sample is selected, it is returned to the population before the next sampling occurs. Hence the phrase, "sampling with replacement" is often used in describing binomial distributed events.

Each individual sampling is called a **Bernoulli trial.** A Bernoulli trial consists of two mutually exclusive outcomes. One (the success) has a probability, p, and the other (the failure) has a probability of $1 - p$. These probabilities remain the same for all of the n repetitions (e. g., 0.5 for heads or tails in a coin flip).

The binomial density function represents the probability of i successes out of n trials:

$$b(n, p, i) = {}_nC_i p^i (1-p)^{n-i}$$

where $i = 0, 1, 2, \ldots, n$.

It can be shown that the mean u and variance s^2 of the binomial density are

$$u = np \quad \text{and} \quad s^2 = np(1-p)$$

The outcome of a multiple coin toss experiment is a binomial random variable. For example, consider the number of heads after 10 coin tosses. The binomial probability density function for this case is plotted in Figure 4.1. The cumulative binomial distribution function can be used to compute the probabilities of consecutive Bernoulli trials. After n trials, the probability that a binomial random variable, X, is less than or equal to $k0$ is

$$P(X \leq k0) = \sum_{i=0}^{k0} \binom{n}{i} p^i (1-p)^{n-i}$$

where $\binom{n}{i} = {}_nC_i$, and $k0 = 0, 1, \ldots n$.

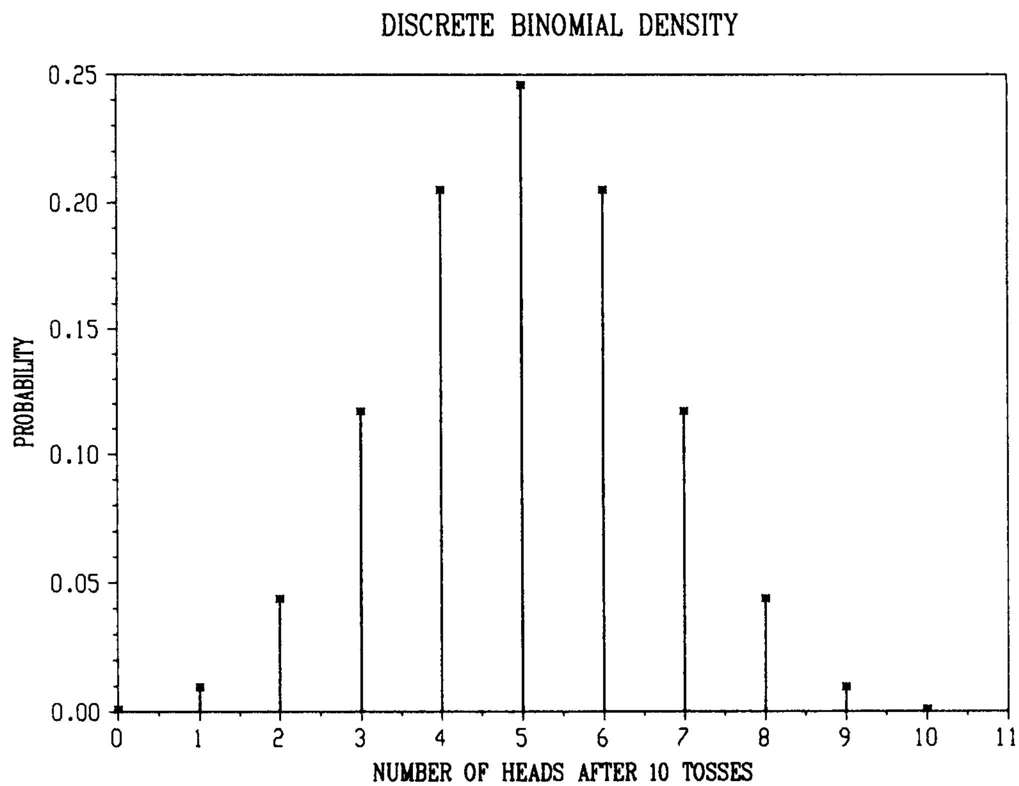

Figure 4.1 Binomial Probability Density

The cumulative distribution is computed by summing (up to $k0$) the individual probabilities shown in Figure 4.1. For example, out of ten flips of a coin the probability that four or less are heads is

$$b(10, 0.5, 0) + b(10, 0.5, 1) + \ldots + b(10, 0.5, 4) = 0.377$$

See Function: binomdst().

4.2.2 Hypergeometric Distribution

The **hypergeometric** distribution is frequently used to compute the probability of outcomes of repetitive sampling from a small population. However, unlike the binomial distribution, the samples are **not** returned to the population. Hence, the sampling is done **without** replacement.

There are two mutually exclusive classes for each random sampling (e. g., an ace or "not an ace"). The total number of elements in one class is $r1$. The total population size is r. Thus, the number of elements in the other class is $r - r1$.

The hypergeometric density function represents the probability of i successes out of n trials:

$$h(n, r, r1, i) = {}_{r1}C_i * {}_{(r-r1)}C_{(n-i)} / {}_rC_n$$

where $i = 0, 1, 2, \ldots, n$.

It can be shown that the mean u and variance s^2 of the hypergeometric density are

$$u = n * r1/r$$
$$s^2 = n * r1 * (r - r1) * (r - n)/[r^2 * (r - 1)]$$

The hypergeometric distribution can be used to compute probabilities of a poker hand dealt from a fair deck of cards. Let $n = 5$ be the number of cards in a poker hand, and $r1 = 4$ be the number of aces in a deck. The total number of cards in the deck is $r = 52$. The hypergeometric probability density function for this example is shown in Figure 4.2. For example, the odds of being dealt exactly four aces is about 0.000018 (although the odds are much better in Hollywood).

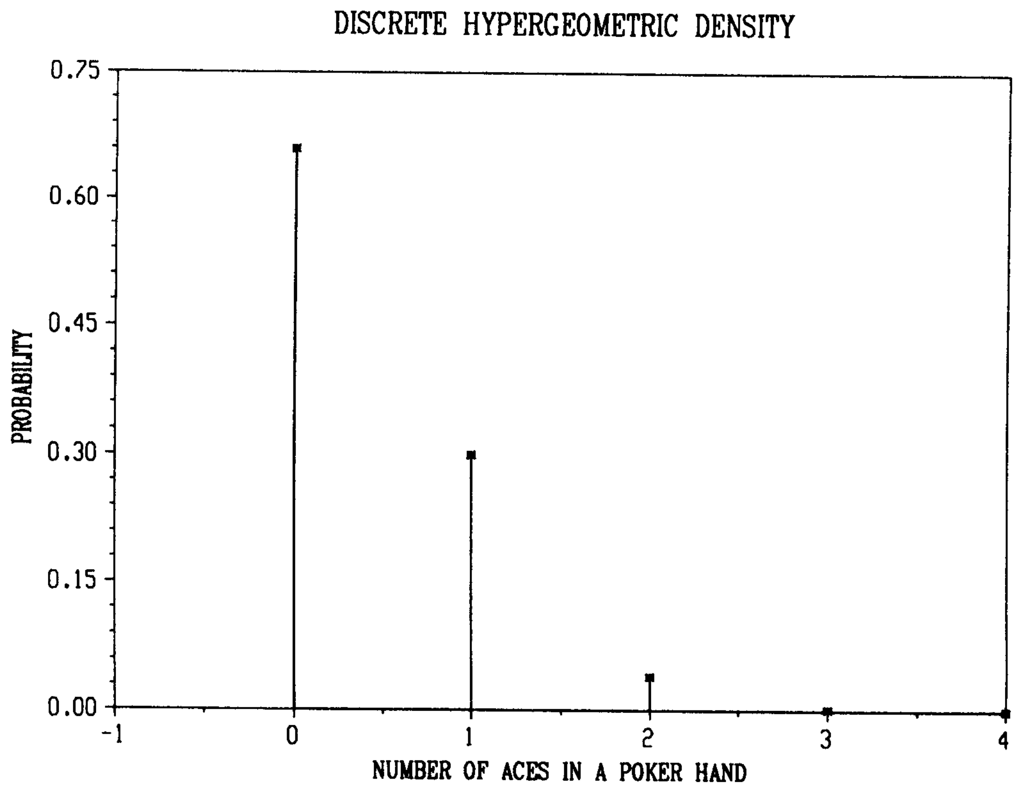

Figure 4.2 Hypergeometric Probability Density

The cumulative distribution is computed by summing (up to $k0$) the individual probabilities shown in Figure 4.2. After n trials, the probability that a hypergeometric random variable, X, is less than or equal to $k0$ is

$$P(X \le k0) = \sum_{i=0}^{k0} \frac{\binom{r1}{i}\binom{r-r1}{n-i}}{\binom{r}{n}}$$

where $k0 = 0, 1, \ldots n$.

For example, the odds of being dealt one or less ace is

$$h(n, r, r1, 0) + h(n, r, r1, 1) = 0.9583.$$

See Function: hyperdst().

4.2.3 Poisson Distribution

The **Poisson** distribution can be used to predict the number of discrete random events that occur in a fixed time interval. This distribution has many applications in queueing theory. The events are assumed to be independent from occurrences in prior intervals. The rate at which events occur is assumed to be constant. The expected number of arrivals λ_T is the product of the average rate of arrivals of the process with the interval of interest, T.

A Poisson density function with parameter λ_T represents the probability of i arrivals in T units of time:

$$p(\lambda_T, i) = (\lambda_T)^i e^{-\lambda_T} / i!$$

where $i = 0, 1, 2, \ldots$.

For the Poisson density, it can be shown that the mean u and variance s^2 are equal:

$$u = s^2 = \lambda_T$$

The Poisson density can be used to forecast automobile traffic flow. If cars are arriving at a stop sign at a rate of one per minute, and the interval of interest is $T = 20$ minutes, then the parameter of the process is $\lambda_T = 20$. The Poisson probability density function for this example is shown in Figure 4.3. For example, the probability that exactly 20 cars arrive in 20 minutes is 0.0888.

The cumulative Poisson distribution is computed by summing (up to $k0$) the individual probabilities shown in Figure 4.3. After n trials, the probability that a Poisson random variable, X, is less than or equal to $k0$ is

$$P(X \le k0) = \sum_{i=0}^{k0} \frac{(\lambda_T)^i e^{-\lambda_T}}{i!}$$

where $k0 = 0, 1, 2, \ldots$.

For example, the probability that 21 or less cars arrive in 20 minutes is

$$p(20, 0) + p(20, 1) + p(20, 2) + \ldots + p(20, 21) = 0.6437$$

See Functions: poissdst(), poisson().

4.3 CONTINUOUS RANDOM VARIABLES

Many statistical events are well modeled by continuous random variables. For these cases, the probability densities are continuous functions. The domain of the proba-

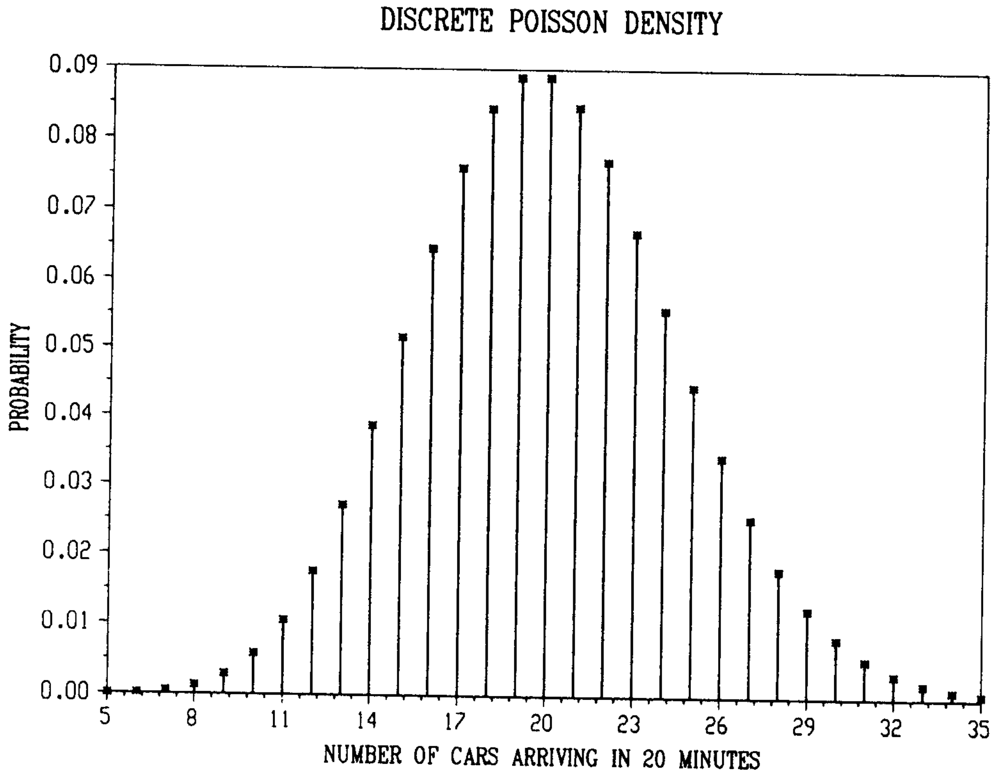

Figure 4.3 Poisson Probability Density

bility density function is the set of possible random outcomes (x values), sometimes called the **sample space**.

An outcome of a continuous random experiment can be any real value. For a continuous random variable to be properly defined, the area under its probability density function must be one. That is, the integral of the density function over the entire sample space must be unity.

The **cumulative probability distribution function** $F(x_0)$ of a continuous random variable is defined as the integral of the density $f(x)$ over x, where $x \leq x_0$.

This section describes some of the properties of the two most commonly used continuous random variables: the **normal** and the **uniform** distributions.

4.3.1 Normal Distribution

The **normal** distribution (the "bell curve") is the most frequently used function in applied probability. This distribution is also called the Gaussian distribution. In its **standard** form, the normal distribution function is given by

$$p0 = P(X \leq x_0) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x_0} e^{-x^2/2} dx$$

This form is particularly convenient because of the widespread availability of the standard normal tables. These tables list approximations to the above integral which facilitate probabilistic computations.

It can be shown that the mean of the standard normal density is zero and the variance is one. However, the real usefulness of the normal distribution is in modeling random events of arbitrary mean u and variance s^2. This is allowed by a simple transformation of variables:

$$X' = s * X + u$$

where X is the standard normal random variable and X' is the normal random variable with mean u and variance s^2. A useful shorthand notation in referring to normal variables is to say that X' is $N(u, s^2)$ and X is $N(0, 1)$.

The normal probability density function for the random variable X' is plotted in Figure 4.4. Because this is a valid probability density, the area under this curve is one. Approximately 68 percent of the area lies within one standard deviation (s) of the mean (u). This is the area in between the dashed lines of Figure 4.4.

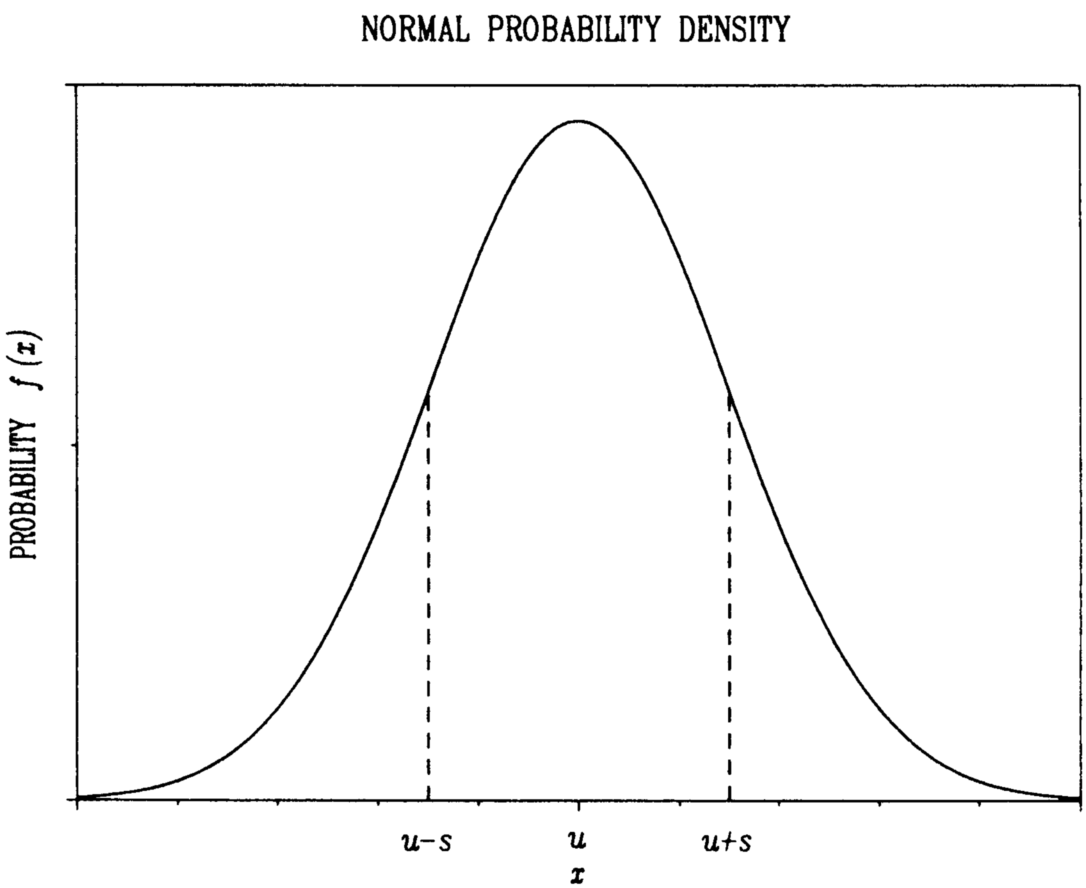

Figure 4.4 Normal Probability Density

The **Central Limit Theorem** is of fundamental importance to probability and statistics. This theorem simply states that if a random variable can be represented as a sum of any finite number of independent random variables, then under most conditions, it is normally distributed.

More formally, the **Central Limit Theorem** is as follows:

Let $X_1, X_2, \ldots, X_n, \ldots$ be a sequence of independent random variables each with mean u_i and variance s_i^2, $i = 1, 2, \ldots$. Let $Y = X_1 + X_2 + \ldots + X_n$. Under some general conditions that are beyond the scope of this development, an approximation to the standard normal distribution is given by

$$Z = (Y - U_n)/S_n$$

where $U_n = u_1 + u_2 + \ldots + u_n$

and $S_n = (s_1^2 + s_2^2 + \ldots + s_n^2)^{0.5}$

Note that as n (the number of random variables that are summed) increases, the distribution of Z more closely approximates $N(0, 1)$. This theorem emphasizes the great importance of the normal distribution, since it gives a mathematical justification of why the normal distribution works well in so many statistical applications.

Approximating Normal Probabilities. The use of the normal distribution usually requires consulting tables of values. Although this may be adequate for a few hand calculations, these tables are cumbersome and are difficult to incorporate into computer programs. The smooth nature of the normal curve allows a closed form alternative to tables via polynomial approximation.

Consider again the following equation of the standard normal distribution:

$$P(X \leq x0) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x0} e^{-x^2/2} dx$$

There are two approximations concerning this equation that are useful. The first approximates the cumulative normal probability $p0$ for a given x_0. The second approximation determines the specific value of x_0 required for a given cumulative probability $p0$. Note that the second case is really the inverse of the first.

Approximating the Normal Distribution. The goal is to estimate the cumulative normal probability $P(X \leq x0)$ for a given $x0$.

For $x0 \geq 0$, the probability is approximated with the following truncated power series:

$$P(X \leq x0) = 1 - f(x0) * (b_1 t + b_2 t^2 + b_3 t^3 + b_4 t^4 + b_5 t^5)$$

where $t = (1 + 0.2316419 * x0)^{-1}$ and $f(x0) = 0.39894228 * e^{-x0^2/2}$
and $b_1 = 0.31938153,\quad b_2 = -0.356563782$
$b_3 = 1.781477937,\quad b_4 = -1.821255978$
$b_5 = 1.330274429$

For $x0 < 0$, the complementary power series is used:

$$P(X \leq x0) = f(x0) * (b_1 t + b_2 t^2 + b_3 t^3 + b_4 t^4 + b_5 t^5)$$

The error of these approximations is less than $7.5e - 8$.

See Functions: nprob(), normal().

Approximating the Inverse Normal Distribution. This is the inverse of the previous approximation. The goal is to estimate the value of x_0 for a given cumulative normal probability, $p0 = P(X \leq x_0)$.

For $p0 \leq 0$, the value is estimated with the following truncated power series:

$$x_0 = -t + (c_0 t + c_1 t^2 + c_2 t^3)/(1 + d_1 t + d_2 t^2 + d_3 t^3)$$

where $t = (-2 * \ln(p0))^{0.5}$
and $c_0 = 2.515517,\quad d_1 = 1.432788$
$c_1 = 0.802853,\quad d_2 = 0.189269$
$c_2 = 0.010328,\quad d_3 = 0.001308$

For $p0 > 0.5$, the following complementary power series is used:

$$x_0 = t - (c_0 t + c_1 t^2 + c_2 t^3)/(1 + d_1 t + d_2 t^2 + d_3 t^3)$$

where $t = (-2 * \ln(1 - p0))^{0.5}$.

The error of this approximation is less than $4.5e - 4$.

See Functions: invprob(), normal().

4.3.2 Uniform Distribution

The **uniform** random variable is one of the most important random variables in probability and statistics. Many probability distributions can be generated with computations that involve the uniform distribution. Like the normal distribution, the uniform distribution has a standard form:

$$\begin{aligned} u(x) &= 1 \quad && \text{for } 0 < x < 1 \\ &= 0 && \text{otherwise} \end{aligned}$$

It can be shown that the mean of the standard uniform density is 1/2 and the variance is 1/12. However, the standard form can easily be transformed to a uniform random variable with arbitrary mean u and variance s^2:

$$X' = (b - a)(X - 1/2) + (a + b)/2$$

where X is the standard uniform random variable and X' is the uniform random variable with mean u and variance s^2:

$$u = (a+b)/2$$
$$s^2 = (b-a)^2/12$$

This transformation can be used to generate uniform white noise samples of any mean and variance.

The uniform probability density function for the random variable X' is plotted in Figure 4.5.

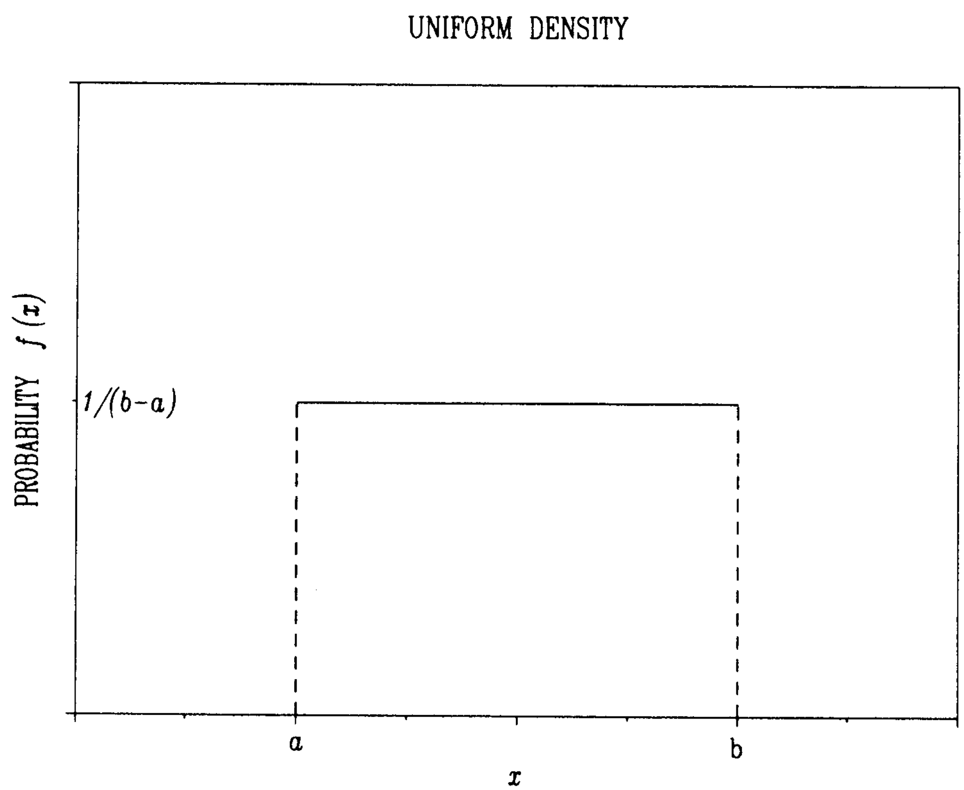

Figure 4.5 Uniform Probability Density

A useful shorthand notation in referring to uniform variables is to say that X' is $U(a,b)$ and X is $U(0,1)$.

See Function: urand().

Normal Random Variable from a Uniform Distribution. A normal random variable of arbitrary mean and variance can be approximated with the uniform distribution. This approximation is very useful because of the widespread availability of uniform random number generators.

The method uses a special case of the previously described Central Limit Theorem. Consider the sum of a sequence of 12 independent random variables X_i, which are uniformly distributed over the interval $(0,1)$:

$$Y = X_1 + X_2 + X_3 + \ldots + X_{12}$$

Recall that the mean and variance of the standard uniform distribution were given by

$$u_x = 1/2, \quad \text{and} \quad s_x^2 = 1/12$$

It can be shown that the mean and variance of the random variable Y are given by

$$\begin{aligned} u_y &= 12 * (1/2) = 6 \\ s_y^2 &= 12 * (1/12) = 1 \end{aligned}$$

From the Central Limit Theorem, the standard normal random variable, Z, with zero mean and unity variance can be approximated by:

$$Z = (Y - 6)/s_y = Y - 6$$

Note that since the variance of Y is unity, the division by the standard deviation (also unity) in the above formula is avoided. This is the reason for summing exactly 12 uniform random variables. Summing more than 12 uniform random variables would improve the statistical approximation, but would require a division by the standard deviation, s_y.

The desired normal random variable, W, with mean u_w and standard deviation s_w is obtained with the simple transformation

$$W = Z * s_w + u_w = (Y - 6) * s_w + u_w$$

Thus, normal random numbers of any mean and variance can easily be approximated with a uniform random number generator.

See Functions: normal(), urand().

Poisson Random Variable from a Uniform Distribution. The Poisson distribution is very useful in modeling the number of discrete events that occur in a fixed time interval. The time, t, between events is exponentially distributed:

$$\begin{aligned} f(t) &= ae^{-at} && \text{for } t > 0 \text{ and } a > 0 \\ &= 0 && \text{otherwise} \end{aligned}$$

The mean and variance of the exponential random variable are

$$\begin{aligned} u_t &= 1/a \\ s_t^2 &= 1/a^2 \end{aligned}$$

The uniform distribution can be used to generate an exponential random variable of arbitrary mean and variance. In turn, the exponential approximation can be used to generate the desired Poisson random variable.

The most straightforward method of generating a Poisson random integer is to sum a sequence of K independent random variables Y_i, which are exponentially distributed with unity mean and variance (i. e., $a = 1$). K is incremented until the sum is equal to, or exceeds, the desired Poisson mean, $p0$. At this point, the value of K is the desired Poisson random integer. The sequence of exponentially distributed random variables, Y_i, is easily constructed from any standard uniform random number sequence X_i according to:

$$Y_i = -\log_e(X_i)$$

Note that the random variable Y_i/a is exponentially distributed with parameter a.

Consider the sum of K of these random variables:

$$Y = Y_1 + Y_2 + Y_3 + \ldots + Y_K$$

It is easy to show that the above sum of exponential random variables, Y, has a mean and variance of K, where K is approximately the desired Poisson parameter $p0$. This follows directly since:

1. The mean of the sum is the sum of the means.
2. The variance of the sum is the sum of the variances.

Recall that for the Poisson distribution, the variance must equal the mean. Thus, Y is the desired Poisson random variable since:

$$s_y^2 = u_y = K = p0$$

A slightly more efficient method of generating Poisson random variables uses a variation of the technique described above. This approach generates the Poisson random variable directly from the uniform distribution. K standard uniform random variables, X_i, are multiplied together. As K is incremented, the product declines until it is less than or equal to e^{-p0}. At this point, the value of K is the desired Poisson random integer. This approach results from raising both sides of the above expression for Y to the base of e:

$$e^{-Y} = X_1 * X_2 * X_3 * \ldots * X_K \geq e^{-p0}$$

where $Y_i = -\log_e(X_i)$ and $\exp(-Y_i) = X_i$.

Note that this approach avoids the natural logarithm and the need to generate the exponential random sequence.