

## Probability Models.S2

### Discrete Random Variables

---

Results of an experiment involving uncertainty are described by one or more random variables. This section considers the discrete random variable, while the continuous case is the subject of the next section. Here the probability distribution is specified by a nonzero probability assigned to each possible value of the random variable.

For a particular decision situation, the analyst must assign a distribution to each random variable. One method is to perform repeated replications of the experiment. Statistical analysis provides estimates of the probability of each possible occurrence. Another, and often more practical method, is to identify the distribution to be one of the *named distributions*. It is much easier to estimate the parameters of a named distribution, than to estimate the entire set of probabilities. The section provides a catalog of some of the important named distributions and examples of their use.

The section also presents formulas for the moments and probabilities for the general discrete distribution and the special case named distributions. For applications, it is not always necessary to know the formulas or even how to use them. Easily available computer programs readily provide numerical results for given parameters<sup>1</sup>. The formulas, however, provide a higher level of knowledge regarding the distributions and are frequently very helpful for estimating parameters.

#### Describing Probability Distributions

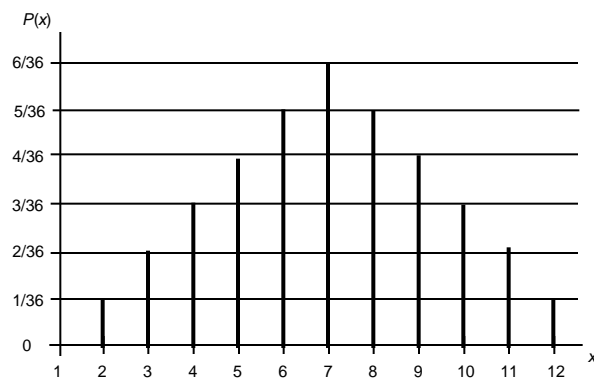


Figure 3. Distribution functions for total on dice (Triangular Distribution)

A discrete probability distribution function is completely described by the set of possible values the random variable can take and by the probabilities assigned to each value. The notation  $P_X(k)$  for  $k = 0, 1, \dots$ , indicates that the random variable takes on the value of any nonnegative integer. Providing values of  $P_X(k)$  for each value of  $k$  completely specifies the distribution. The triangular distribution of Fig. 3 is used as an

illustration.

Certain quantities can be computed from the p.d.f. that describe simple characteristics of the distribution. These are called moments. The most common is the mean,  $\mu$ , the first moment about the origin. We give the general definition for the mean along with its computation for the triangular distribution in Table 3.

---

<sup>1</sup> The Random Variables Add-in allows the user to define random variables and assign them named distributions. Functions are available to compute moments and probabilities. Random variables may be simulated to model more complex situations.

This and following definitions assume that the random variable is defined over all nonnegative integers. The definitions can be easily expanded to cover other domains. The mean provides a measure of the *center* of the distribution. For a symmetric distribution, the mean is the middle value of the range of the random variable. Another term often used for the mean is the *expected value of  $x$* , written  $E(x)$ .

The variance,  $\sigma^2$ , is a measure of the *spread* of the distribution about the mean. It is the second moment about the mean. Where the random variable has high probabilities near the mean, the variance is small, while large probabilities away from the mean imply that the variance is large. The standard deviation,  $\sigma$ , is simply the positive square root of the variance and also describes the spread of the distribution. Discussions more often use the standard deviation because it has the same dimension as the mean.

Two additional measures that describe features of a distribution are the skewness and kurtosis<sup>2</sup> with general measures  $\beta_1$  and  $\beta_2$  given in Table 3. A positive skewness indicates that the distribution is concentrated to the left with a long thin tail pointing to the right, and a negative skewness has the concentration to the right and the tail pointed to the left. Kurtosis measures the peakedness of the distribution. We illustrate the skewness measure in the several examples of this section. The example in Fig. 3 has a skewness of 0 because the distribution is symmetric.

Table 3. Descriptive Measures

Measure	General Definition	Sum of the Dice
Mean	$\mu = \sum_{k=0} k P_X(k)$	$\mu = 2(1/36) + 3(2/36) + \dots + 12(1/36) = 7.$
Variance	$\sigma^2 = \sum_{k=0} (k - \mu)^2 P_X(k)$	$\sigma^2 = (2 - 7)^2(1/36) + (3 - 7)^2(2/36) + \dots + (12 - 7)^2(1/36) = 5.833.$
Standard Deviation	$\sigma = \sqrt{\sigma^2}$	$\sigma = \sqrt{5.833} = 2.415.$
Skewness	$\beta_1 = \frac{(\mu_3)^2}{\sigma^6}$ $\mu_3 = \sum_{k=0} (k - \mu)^3 P_X(k)$	$\beta_1 = 0$
Kurtosis	$\beta_2 = \frac{\mu_4}{\sigma^4}$ $\mu_4 = \sum_{k=0} (k - \mu)^4 P_X(k)$	$\beta_2 = 2.365$

<sup>2</sup> Barnes, J. Wesley, Statistical Analysis for Engineers and Scientists, McGraw Hill, 1994.

## Named Discrete Distributions

It is useful for modeling purposes to know about the named discrete distributions. When an experiment on which a random variable is based satisfies the logical conditions associated with a named distribution, the distribution for the random variable is immediately determined. Then we can use the distribution without extensive experimentation to answer decision questions about the situation.

### *Bernoulli Distribution*

**Ex. 3.** Consider again the Craps game. If on the first roll of the dice you throw a number other than 2, 3, 7, 11, or 12, the number you do throw is your point. The rules say you must roll the dice again and continue to roll until you throw your point and win, or a 7, and lose. Say your point is 4. Based on your probability model you determine that on any given roll following the first:

$$P(\text{win}) = P(x = 4) = 3/36.$$

$$P(\text{lose}) = P(x = 7) = 6/36.$$

$$P(\text{roll again}) = 1 - P(\text{win}) - P(\text{lose}) = 27/36 = 3/4.$$

For each roll, the game either terminates with probability 1/4, or you must roll again with probability 3/4.

*Bernoulli Distribution*  
 Parameter:  $0 < p < 1$   
 $P(1) = p$  and  $P(0) = 1 - p$   
 $\mu = p$ ,  $\sigma^2 = p(1 - p)$

An experiment that has two outcomes is called a *Bernoulli trial*. For the example we take the two outcomes as “roll again” and “terminate”, and arbitrarily assign the value 0 to the roll again outcome and value 1 to the terminate outcome. The parameter associated with the probability

distribution is the probability that the variable assumes the value 1 indicated by  $p$ . Given the value of  $p$ , the entire distribution is specified. For the example

$$P(\text{terminate}) = P(1) = 1/4 \text{ and } P(\text{roll again}) = P(0) = 3/4.$$

The simple *Bernoulli* distribution illustrated with this example is the first of several *named distributions* presented in this chapter. These distributions are useful because they model a variety of situations.

### *Geometric Distribution*

**Ex. 4.** If you don't win or lose on the first roll, you might wonder how long the game will last. Assume you roll a point of 4. Now you begin the second phase of the game and define the random variable as the number of rolls prior to the last roll. That number may be 0, 1, 2, .... The number is random variable described by the geometric distribution.

This random variable has an infinity of possible values in that there is no upper limit to the number of rolls conceivable required. There is only one way a

particular value,  $k$ , of the random variable can occur. There must be  $k$  roll again results followed by one termination. The probability of this occurrence describes the probability distribution function.

*Geometric Distribution*  
 Parameter:  $0 < p < 1$   
 $P(k) = p (1 - p)^k$  for  $k = 0, 1, 2 \dots$   
 $\mu = \frac{1 - p}{p}$ ,  $\sigma^2 = \frac{1 - p}{p^2}$

The geometric distribution has the single parameter  $p$ , with the requirement that ( $0 < p < 1$ ). The logical condition for this distribution is that the separate trials be independent, that is, the outcome of one trial does not affect the probability of a later trial. This is certainly true

for sequential throws of the dice.

For the example the parameter is  $p = P(\text{terminate}) = 1/4$ . Table 4 shows the probability distribution for this case, and Fig. 4 shows a plot of the distribution. The game may take quite a few rolls to complete with a greater than 10% chance that more than 7 are required. The moments for the example are:

$$\mu = 3, \sigma^2 = 12, \sigma = 3.46, \beta_1 = 4.08, \beta_2 = 9.08$$

The positive skewness is clearly indicated in the plot.

Table 4 . Geometric Distribution

Number	0	1	2	3	4	5	6	7
Probability	0.25	0.188	0.141	0.105	0.079	0.059	0.044	0.033
Cumulative	0.25	0.438	0.578	0.684	0.763	0.822	0.867	0.9

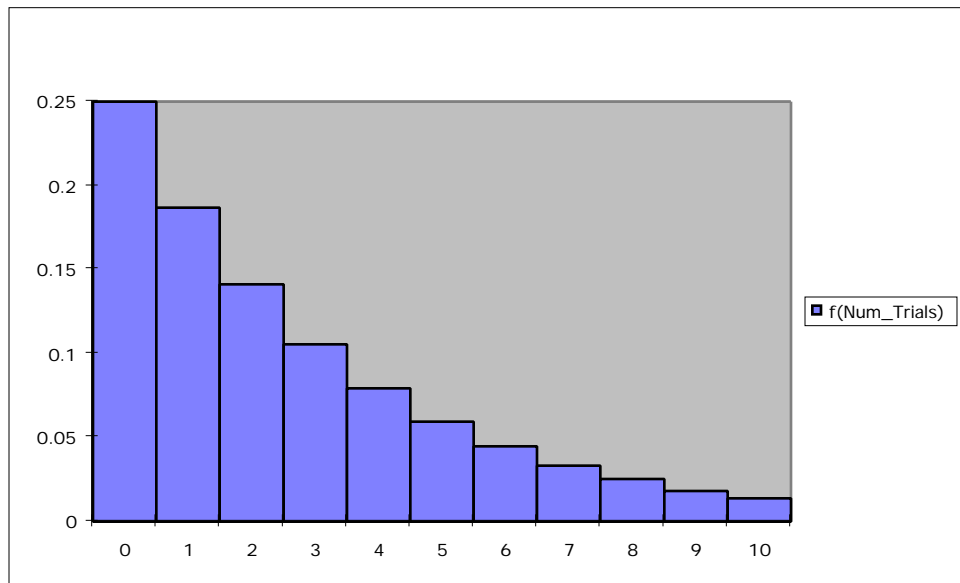


Figure 4. Plot of the geometric distribution.

Using this probability model, it is easy to compute the mean and variance of the number of rolls to termination for each of the point numbers as given below.

Point	4	5	6	8	9	10
$p$	0.25	0.278	0.306	0.306	0.278	0.25
$\mu$	3	2.6	2.273	2.273	2.6	3
$\sigma^2$	12	9.36	7.438	7.438	9.36	12

*Negative Binomial Distribution*

**Ex. 5.** In the game of craps, you decide to play until you lose 5 games. You wonder how many games you will play with this termination rule. Recall that the probability of losing any one game is 0.5071. The games are a series of independent Bernoulli trials, and the random variable is the number of wins until the fifth loss. This is a situation described by the negative binomial distribution.

*Negative Binomial Distribution*  
 Parameters:  $0 < p < 1$ ,  
 $r$  1 and integer  

$$P_x(k) = \binom{r+k-1}{r-1} p^r (1-p)^k$$
 for  $k = 0, 1, 2, \dots$   

$$\mu = \frac{r(1-p)}{p}, \sigma^2 = \frac{r(1-p)}{p^2}$$

For this distribution we first identify the result of success. In this case, we perversely identify success as a “loss”.  $p$  is the probability of a success, 0.5071 for the example. The random variable is the number of trials that result in 0 before the  $r$  th 1 is observed.  $r = 5$  for this case.

The distribution for the example is shown in the table. It is important to remember that the random variable is not the total number of trials, but the number of failed trials before the  $r$ th success. In Table 5, the entry for 0 describes the probability that the first five plays were losses and there were no wins. The geometric distribution is a special case when  $r = 1$ .

Table 5 . Negative Binomial Distribution

Number	0	1	2	3	4	5	6	7
Probability	0.034	0.083	0.122	0.141	0.139	0.123	0.101	0.078
Cumulative	0.034	0.116	0.238	0.379	0.517	0.64	0.741	0.82

*Binomial Distribution*

**Ex. 6.** The reliability of a computer is defined as the probability of successful operation throughout the mission. A study determines that the reliability for a given mission as 0.9. Because the mission is very important and computer failure is extremely serious, we provide five identical computers for this mission. The computers operate independently and the failure or success of one does not affect the probability of failure or success of the others. Our job is to compute the probability of mission success, or system reliability, under the following three operating rules:

- a. All five of the computers must work for mission success.
- b. At least three out of five must work for mission success.
- c. At least one computer must work for mission success.

Consider an experiment that involves  $n$  independent Bernoulli trials. Associate with each outcome, the random variable that is the sum of the of the  $n$  Bernoulli

*Binomial Distribution*

Parameters:  $0 < p < 1$ ,  $n \geq 1$  and integer

$$P_x(k) = \binom{n}{k} p^k (1 - p)^{n-k} \text{ for } k = 0, 1, 2 \dots n.$$

$$\binom{n}{k} = \frac{n!}{(n - k)! k!}.$$

$$\mu = np, \sigma^2 = np(1 - p)$$

random variables. This is called the binomial random variable. The variable has  $n+1$  possible values ranging from 0 to  $n$ . Its p.d.f. is the binomial distribution. The binomial distribution has two parameters  $p$  and  $n$ .

In the case given, the success or failure of each computer is a

Bernoulli random variable with 1 representing success and 0 representing failure. The probability of success is  $p$ , and we assume that the computers are independent with respect to failure.

The number of working computers,  $x$ , is the random variable of interest, and the binomial distribution, with parameters  $n = 5$  and  $p = 0.9$ , is the appropriate p.d.f.. With these parameters the probability of  $k$  successful computers is computed and the results are shown in the table. The moments of the distribution are:

$$\mu = 4.5, \sigma^2 = 0.45, \sigma = 0.67, \beta_1 = -1.42, \beta_2 = 4.02$$

The negative skewness is illustrated in the Fig. 5.

Number	0	1	2	3	4	5
Probability	0.00001	0.00045	0.0081	0.0729	0.32805	0.59049
Cumulative	0.00001	0.00046	0.00856	0.08146	0.40951	1

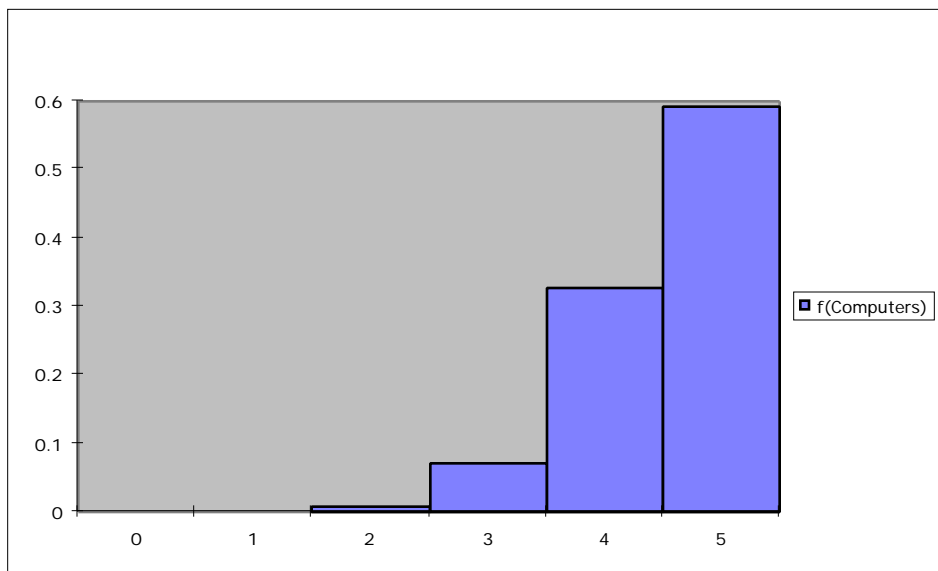


Figure 5. Plot of the binomial distribution

The probabilities of mission success under the three conditions listed earlier are

a.  $P(x = 5) = 0.59049$ .

b.  $P(x \leq 3) = P(3) + P(4) + P(5) = 0.99144$ .

c.  $P(x \geq 1) = 1 - P(0) = 0.99999$ .

These results show the value of redundancy for increasing reliability. In case *a*, none of the computers is redundant since all are required for successful operation. In case *b*, we say that two computers are redundant since only three are required. In case *c*, four are redundant. The reliability of the system obviously increases as redundancy is increased.

### Poisson Distribution

**Ex. 7.** A traffic engineer is interested in the traffic intensity at a particular street corner during the 1 - 2 a.m. time period. Using a mechanical counting device, the number of vehicles passing the corner is counted during the one hour interval for several days of the week. Although the numbers observed are highly variable, the average number is 50 vehicles. The engineer wants a probability model to answer a variety of questions regarding the traffic.

#### Poisson Distribution

Parameter:  $\lambda > 0$

$$P_x(k) = \frac{e^{-\lambda} \lambda^k}{k!} \text{ for } k = 0, 1, 2, \dots$$

$$\mu = \lambda, \sigma^2 = \lambda$$

This situation fits the logical requirements of the *Poisson Distribution*. Consider arrivals that occur randomly but independently in time. Let the average arrival rate be equal to  $\lambda$  per unit of time and let the time interval be  $t$ . Then one would expect the number of arrivals during the interval to be

$= \lambda t$ . The actual number of arrivals occurring in the interval is a random variable governed by the Poisson distribution.

The parameter of the distribution is the dimensionless quantity  $\lambda t$ , which is the mean number of arrivals in the interval. We call an arrival process that gives rise to this kind of distribution a *Poisson process*. The distribution is very important in queueing theory and is discussed more fully in Chapter 10.

To use the distribution for the example, we must only assume that vehicles arrive independently and that the average arrival rate is constant. This does not mean that the vehicles pass in a steady stream with a fixed interval between cars. Rather, with the assumption of randomness, vehicle arrivals are extremely variable; the rate of 50 per hour is an average. Using the distribution, we can model the probabilities for any interval of time, however consider a one minute period. The random variable is the number of vehicles passing during the one minute period. The parameter of the distribution is

$$\lambda t = (50/\text{hour}) (1/60 \text{ hour}) = 5/6 = 0.833.$$

The probability distribution for this example is computed with the formulas given for the Poisson distribution and shown in Table 6. Descriptive measures are

$$\mu = 0.833, \sigma^2 = 0.833, \sigma = 0.913, \beta_1 = 1.2, \beta_2 = 4.2$$

The mean and variance of the Poisson distribution are always equal.

Table 6. The Poisson Distribution with  $\lambda = 0.833$ .

Number	0	1	2	3	4	5
Probability	0.4346	0.3622	0.1509	0.0419	0.0087	0.0015
Cumulative	0.4346	0.7968	0.9477	0.9896	0.9983	0.9998

From the distribution, various probability statements can be made.

- $P(\text{no cars pass the corner}) = P(0) = 0.4346$
- $P(\text{at least two cars pass}) = P(2) + P(3) + \dots$

This expression can be evaluated if enough terms are added together however an easier, more accurate way is

- $P(\text{at least two cars pass}) = 1 - P(\text{no more than one car passes})$   
 $= 1 - P(0) - P(1) = 1 - F(1) = 1 - 0.7968 = 0.2032.$

### Hypergeometric Distribution

**Ex. 8.** You are dealt a hand of 5 cards from a standard deck of 52 cards. Before looking at the hand you wonder about the number of aces among the five cards. This is a case for the hypergeometric distribution.

*Hypergeometric Distribution*  
 Parameters:  $N, a, n$   
 all positive and integer  
 $0 \leq a \leq N, 0 \leq n \leq N$   
 $0 \leq k \leq n$   
 $P(k) = \frac{\binom{a}{k} \binom{N-a}{n-k}}{\binom{N}{n}}$  for  $k = 0, 1, \dots, n$ .  
 $\mu = \frac{na}{N}, \sigma^2 = n \frac{a}{N} \frac{N-a}{N} \frac{N-n}{N-1}$

The situation involves  $N$  items.  $a$  of the items are assigned the label 1 and  $N - a$  are assigned the label 0. Select at random  $n$  items from the  $N$  available without replacement. The number of items labeled 1 is the random variable of interest. For the example, there is a fixed number from which to draw ( $N = 52$  cards), the number of aces in the deck ( $a = 4$  aces) and the sample is 5 cards ( $n = 5$ ).

$P(k)$  is the probability that  $k$  of the items selected have the label 1. There are three parameters  $a, N$  and  $n$ . All parameters are integers.  $0 \leq a \leq N, 0 \leq n \leq N$ .

$P(k) = 0$  if  $k < 0$  or  $k > n$ . A combination expression is zero if the bottom number is greater than the top number. The probability distribution for the number of aces is in Table 7.

Table 7. The Hypergeometric Distribution with  $N = 52, n = 5, a = 4$ .

Number	0	1	2	3	4	5
Probability	0.6588	0.2995	0.0399	0.0017	0.00001	0
Cumulative	0.6588	0.9583	0.9982	1	1	1



*Triangular Distribution*

**Ex. 9.** A computer is shipped with a multiple number of some part. A company that assembles computers is interested in the distribution of the number of parts required. No statistics are available, however, design of the computer assures that every computer requires at least 1 part and the most that can be installed is six parts. A production supervisor estimates that most computers require two parts. The only information we have is the range of the random variable and its mode (the most likely number). A reasonable estimate for the distribution of parts is the triangular distribution.

*Triangular Distribution*  
 Parameters:  $a, b, m$  all integer  
 $a \quad m \quad b$   
 $P(k) = d(k - a + 1)$  for  $a \leq k \leq m$ ,  
 and  $P(k) = e(b - k + 1)$  for  $m \leq k \leq b$ .  
 Use the general formulas to compute moments.

In general, identify  $a$  as the lower limit to the range,  $b$  as the upper limit, and  $m$  as the mode. Construct the distribution with the constants  $d$  and  $e$  that satisfy the requirements:

$$\sum_{k=a}^b P(k) = 1 \text{ and } d(m - a + 1) = e(b - m + 1).$$

For the example situation,  $d = 0.143$  and  $e = 0.057$ , and the distribution is presented in Table 8.

Table 8. The Triangular Distribution with  $a = 1, b = 6, m = 2$ .

Number	1	2	3	4	5	6
Probability	0.143	0.286	0.229	0.171	0.114	0.057
Cumulative	0.143	0.429	0.657	0.829	0.943	1

*Uniform Distribution*

**Ex. 10.** We continue the example used for the triangular distribution, but now assume the production manager has no idea how many parts will be used in a computer. He knows only that the design limits the range of the random variable between 1 and 6. In this situation of complete uncertainty, the uniform distribution might be used.

*Uniform Distribution*  
 Parameters:  $a, b$  all integer  
 $P(k) = 1/(b - a + 1)$  for  $a \leq k \leq b$ .  

$$\mu = \frac{a + b}{2}$$
  
 Use general formula for  $\sigma^2$

This distribution assigns equal probabilities to all possible values of the random variable within the range  $a$  to  $b$ , inclusive. For the example situation:  $P(k) = 1/6$  for  $1 \leq k \leq 6$ .

## Modeling

When attempting to use the results of this section to model some situation, first determine if the random variable is discrete. Whenever the question concerns counting, only the integers are relevant and the variable is obviously discrete. To assign a probability distribution to the random variable, review the special cases to see if any are appropriate.

If the experiment only has two outcomes, the Bernoulli random variable is the obvious choice. If the experiment involves a sequence of independent observations each with two outcomes, then the binomial, geometric or negative binomial distributions may fit. A fixed number in the sequence suggests the binomial, while a question related to the first occurrence of one of the two outcomes suggests the geometric. The negative binomial is a generalization of the geometric where the random variable is the number of unsuccessful trials before the  $r$ th success.

The binomial and hypergeometric are appropriate when we are selecting a fixed number of items from a population. The binomial models the number of successes when the population is infinite or when the population is finite and the items are replaced after each trial. When items are not replaced the hypergeometric is the correct distribution.

The Poisson distribution is used when the question relates to counting the number of occurrences of some event in an interval of time (or some other measure). Key phrases that suggest the appropriateness of the Poisson are that the arrivals are "independent" or "at random".

Triangular and uniform distributions are often used when very little is known concerning the situation. The uniform requires only the range, while the triangular needs the additional knowledge of the mode. The distributions may also be logical consequences of the features of an experiment. For example, the uniform distribution models the number on the face of a single die, while the triangular models the sum of two dice.

If any of the special cases can logically be applied, then the parameters of the distribution must be determined. In practice they may be determined by the logic of a situation or estimation using *statistics* from historical data.

If none of the special cases fit, the random variable still has a probability distribution, however some other rational must be used to assign probabilities to possible outcomes.