Lecture-5

Random Variables and Probability Distributions

Learning Objectives

After careful study of this lecture, you should be able to do the following:

1. Determine probabilities from probability mass functions and the reverse
2. Determine probabilities and probability mass functions from cumulative distribution functions and the reverse
3. Calculate means and variances for discrete random variables
4. Understand the assumptions for some common discrete probability distributions
5. Select an appropriate discrete probability distribution to calculate probabilities in specific applications
6. Calculate probabilities and determine means and variances for some common discrete probability distributions

Random Variables:

A random variable whose values are any definite numbers or quantities that arise as a result of chance factors such that they cannot exactly be predicted in advance, is called a random variable.

For example, the sample space giving a detailed description of each possible outcome when three electronic components are tested may be written

S = {NNN, NND, NDN, DNN, NDD, DND, DDN, DDD}, where N denotes non defective and D denotes defective. One is naturally concerned with the number of defectives that occur. Thus, each point in the sample space will be assigned a numerical value of 0, 1, 2, or 3. These values are, of course, random quantities determined by the outcome of the experiment. They may be viewed as values assumed by the random variable X, the number of defective items when three electronic components are tested.

Example 5-1:

Two balls are drawn in succession without replacement from a #urn containing 4 red balls and 3 black balls. The possible outcomes and the values y of the random variable Y, where Y is the number of red balls, are

Sample Space	y
RR	2
RB	1
BR	1
BB	0

A random variable may be classified as either discrete or continuous depending upon the specific numerical values it can assume.

Discrete random variable: A random variable defined over a discrete sample space (i.e. that may only take on a finite or countable number of different isolated values) is referred to as a discrete random variable.

Examples:

1. The number of telephone calls received in a telephone booth during one day;
2. Number of correct answers in 100-MCq type examination;
3. Number of defective bulbs produced during a day's rum;
4. Size of shoes produced by Bata Shoe Company;
5. Stock of coins of different denominations in Bangladesh Bank.

Continuous random variable:

A random variable defined over a continuous sample space (i.e. which may take any value in a certain interval or collection of intervals) is referred to as a continuous random variable.

Examples:

1. Time taken to serve a customer;
2. Weight of a six-month old baby;
3. Volume of a 3 c.c. bottle of medicine;
4. Length of a steel reinforcement rod;
5. Temperature recorded by the meteorological office.

Example 5-2:

The time to refresh the flash is tested in three cell-phone cameras. The probability that a camera passes the test is 0.8, and the cameras perform independently. See Table 5-1 for the sample space for the experiment and associated probabilities. For example, because the cameras are independent, the probability that the first and second cameras pass the test and the third one fails, denoted as ppf, is P(ppf) = (0.8)(0.8)(0.2) = 0.128.

The random variable X denotes the number of cameras that pass the test. The last column of the table shows the values of X assigned to each outcome of the experiment.

Camera 1	Camera 2	Camera 3	Probability	X
Pass	Pass	Pass	0.512	3
Fail	Pass	Pass	0.128	2
Pass	Fail	Pass	0.128	2
Fail	Fail	Pass	0.032	1
Pass	Pass	Fail	0.128	2
Fail	Pass	Fail	0.032	1
Pass	Fail	Fail	0.032	1
Fail	Fail	Fail	0.008	0

Table 5-1

Probability function:

The set of ordered pairs (x, f(x)) is a probability function, probability mass function, or probability distribution of the discrete random variable X if, for each possible outcome x,

1. f (x) ≥ 0,

2.$\sum_x f(x) = 1$,

3. P(X = x) = f(x).

Example5.3:

Digital Channel there is a chance that a bit transmitted through a digital transmission channel is received in error. Let X equal the number of bits in error in the next four bits transmitted. The possible values for X are {0, 1, 2, 3, and 4}. Based on a model for the errors that is presented in the following section, probabilities for these values will be determined. Suppose that the probabilities are P (X = 0) = 0.6516; P (X = 1) = 0.2916; P (X = 2) = 0.0486; P(X = 3) = 0.0036; P (X = 4) = 0.0001.

The probability distribution of X is specified by the possible values along with the probability of each. A graphical description of the probability distribution of X is shown in Figure 5-1. Check that the probabilities sum to 1.

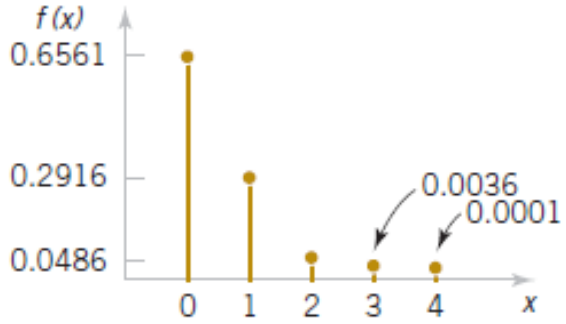


FIGURE 5-1 Probability distribution for bits in error.

The cumulative distribution function of a discrete random variable X, denoted as F (x), is $F(x) = P(X \leq x) = \sum_{x_i \leq x} f(x_i)$

Example5.4:

Digital Channel there is a chance that a bit transmitted through a digital transmission channel is received in error. Let X equal the number of bits in error in the next four bits transmitted. The possible values for X are {0, 1, 2, 3, and 4}. Based on a model for the errors, probabilities for these values will be determined. Suppose that the probabilities are P (X = 0) = 0.6516; P (X = 1) = 0.2916; P (X = 2) = 0.0486; P(X = 3) = 0.0036; P (X = 4) = 0.0001.

We might be interested in the probability that three or fewer bits are in error. This question can be expressed as P(X ≤ 3). The event that {X ≤ 3} is the union of the events {X = 0}, {X = 1}, {X = 2}, and {X = 3}. Clearly, these three events are mutually exclusive. Therefore, P(X ≤ 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) = 0.6561 + 0.2916 + 0.0486 +0.0036 = 0.9999.

This approach can also be used to determine P(X = 3) = P(X ≤ 3) − P(X ≤ 2) = 0.0036.

Example5.5: Suppose that a day's production of 850 manufactured parts contains 50 parts that do not conform to customer requirements. Two parts are selected at random, without replacement, from the batch. Let the random variable X equal the number of nonconforming parts in the sample. What is the cumulative distribution function of X? (Sampling without replacement)

The question can be answered by first finding the probability mass function of X.

$$P(X = 0) = \frac{800}{850} \times \frac{799}{849} = 0.886$$

$$P(X = 1) = 2 \times \frac{800}{850} \times \frac{50}{849} = 0.111$$

$$P(X=2)=\frac{50}{850}\times\frac{49}{849}=0.003$$

Therefore,

$$F(0)=P(X\le 0)=0.886$$

$$F(1)=P(X\le 1)=0.886+0.111=0.997$$

$$F(2)=P(X\le 2)=1$$

and

$$F(x)=\begin{cases}0; & x<0\\ 0.886; & 0\le x<1\\ 0.997; & 1\le x<2\\ 1; & 2\le x\end{cases}$$

The cumulative distribution function for this example is graphed in Fig. 5-2.

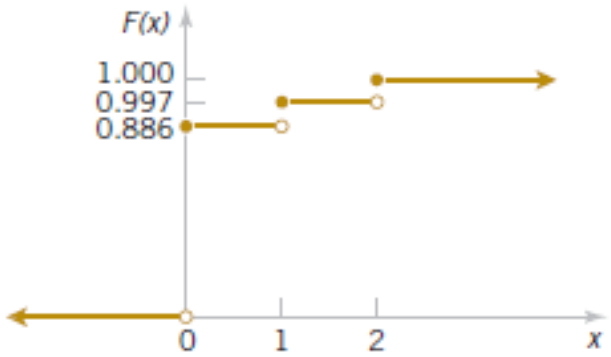


FIGURE 5-2 Cumulative distribution function for the example 5.5

Mean and Variance of a Discrete Random Variable:

Two numbers are often used to summarize a probability distribution for a random variable X. The mean is a measure of the center or middle of the probability distribution, and the variance is a measure of the dispersion, or variability in the distribution. These two measures do not uniquely identify a probability distribution. That is, two different distributions can have the same mean and variance. Still, these measures are simple, useful summaries of the probability distribution of X.

The mean or expected value of the discrete random variable X, denoted by μ or E(X), is $\mu=E(X)=\sum_x xf(x)$

The variance of X, denoted as σ^2 or V(X) is

$$\sigma^2=V(X)=E(X-\mu)^2=\sum_x(x-\mu)^2f(x)=\sum_x x^2f(x)-\mu^2$$

Figure 5-3 illustrates probability distributions with equal means but different variances.

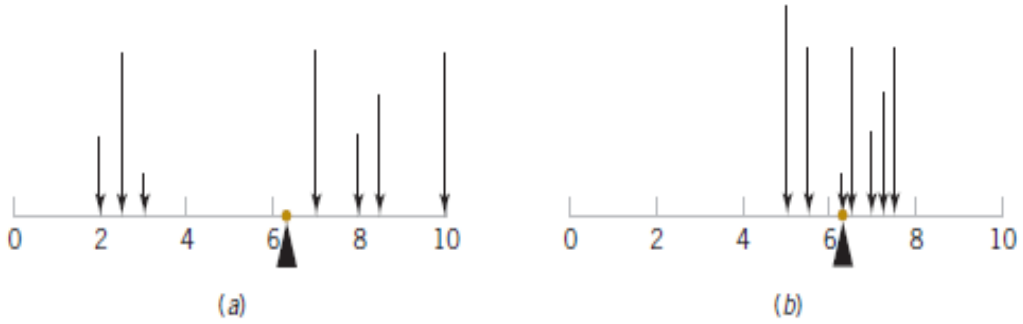


FIGURE 5-3: A probability distribution can be viewed as a loading with the mean equal to the balance point. Parts (a) and (b) illustrate equal means, but part (a) illustrates a larger variance.

Figure 5-4 illustrates that two probability distributions can differ even though they have identical means and variances.

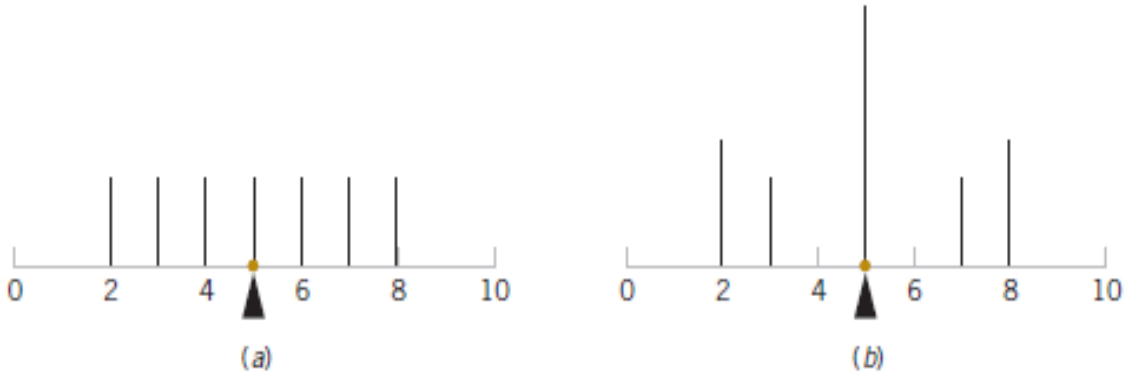


FIGURE 5-4: The probability distributions in parts (a) and (b) differ even though they have equal means and equal variances

Example5.6:

Digital Channel there is a chance that a bit transmitted through a digital transmission channel is received in error. Let X equal the number of bits in error in the next four bits transmitted. The possible values for X are {0, 1, 2, 3, and 4}. Based on a model for the errors that is presented in the following section, probabilities for these values will be determined. Suppose that the probabilities are P (X = 0) = 0.6516; P (X = 1) = 0.2916; P (X = 2) = 0.0486; P(X = 3) = 0.0036; P (X = 4) = 0.0001. Calculate mean and variance

Solution:

Now $\mu = E(x) = 0 \times f(0) + 1 \times f(1) + 2f(2) + 3f(3) + 4f(4) = 0 \times 0.6561 + 1 \times 0.2916 + 2 \times 0.0486 + 3 \times 0.0036 + 4 \times 0.0001 = 0.4$

To calculate variance V(x), a table is convenient

x	x-0.4	$(x-0.4)^2$	f(x)	$(x-0.4)^2 f(x)$
0	-0.4	0.16	0.6561	0.104976
1	0.6	0.36	0.2916	0.104976
2	1.6	2.56	0.0486	0.124416
3	2.6	6.76	0.0036	0.024336
4	3.6	12.96	0.0001	0.001296

$$V(X) = \sigma^2 = \sum_{i=1}^{5} (x_i - 0.4)^2 f(x_i) = 0.36$$

Example5.7:

Marketing Two new product designs are to be compared on the basis of revenue potential. Marketing believes that the revenue from design A can be predicted quite accurately to be $3 million. The revenue potential of design B is more difficult to assess. Marketing concludes that there is a probability of 0.3 that the revenue from design B will be $7 million, but there is a 0.7 probability that the revenue will be only $2 million. Which design do you prefer?

Solution:

Let X denote the revenue from design A. Because there is no uncertainty in the revenue from design A, we can model the distribution of the random variable X as $3 million with probability 1. Therefore, E (X) = $3 million.

Let Y denote the revenue from design B. The expected value of Y in millions of dollars is $E(Y) = \$7(0.3) + \$2(0.7) = \$3.5$.

Because E (Y) exceeds E (X), we might prefer design B.

However, the variability of the result from design B is larger. That is, $\sigma^2 = (7 - 3.5)^2 \times (0.3) + (2 - 3.5)^2 \times (0.7) = 5.25$ millions of dollars squared

Because the units of standard deviation are the same as the units of the random variable, the standard deviation σ is easier to interpret. Here $\sigma = \sqrt{5.25} = 2.29$ millions of dollars and σ is large relative to μ.

Example5.8:

The number of e-mail messages received per hour has the following distribution:

x = number of messages	10	11	12	13	14	15
$f(x)$	0.08	0.15	0.30	0.20	0.20	0.07

Determine the mean and standard deviation of the number of messages received per hour.

Solution:

$$E(X) = 10 \times 0.08 + 11 \times 0.15 + 12 \times 0.30 + 13 \times 0.20 + 14 \times 0.20 + 15 \times 0.07 = 12.5$$

$$V(X) = 10^2 \times 0.08 + 11^2 \times 0.15 + 12^2 \times 0.30 + 13^2 \times 0.20 + 14^2 \times 0.20 + 15^2 \times 0.07 = 1.85$$

$$\sigma = \sqrt{V(X)} = \sqrt{1.85} = 1.36$$

Discrete Probability Distributions:

No matter whether a discrete probability distribution is represented graphically by a histogram, in tabular form, or by means of a formula, the behavior of a random variable is described. Often, the observations generated by different statistical experiments have the same general type of behavior. Consequently, discrete random variables associated with these experiments can be described by essentially the same probability distribution and therefore can be represented by a single formula. In fact, one needs only a handful of important probability distributions to describe many of the discrete random variables encountered in practice.

Such a handful of distributions describe several real-life random phenomena. For instance, in a study involving testing the effectiveness of a new drug, the number of preserved patients among all the patients who use the drug approximately follows a binomial distribution. In an industrial example, when a sample of items selected from a batch of production is tested, the number of defective items in the sample usually can be modeled as a hyper geometric random variable. In a statistical quality control problem, the experimenter will signal a shift of the process mean when observational data exceed certain limits. The number of samples required to produce a false alarm follows a geometric distribution which is a special case of the negative binomial distribution. On the other hand, the number of white cells from a fixed amount of an individual's blood sample is usually random and may be described by a Poisson distribution.

1. Binomial Distribution:

The Bernoulli Process:

Rigorously speaking, the Bernoulli process must possess the following properties:

1. The experiment consists of repeated trials.
2. Each trial results in an outcome that may be classified as a success or a failure.
3. The probability of success, denoted by p, remains constant from trial to trial.

4. The repeated trials are independent.

The number X of successes in n Bernoulli trials is called a binomial random variable. The probability distribution of this discrete random variable is called the binomial distribution, and its values will be denoted by b(x; n, p) since they depend on the number of trials (n = 1, 2,…..) and the probability of a success (0 < p <1) on a given trial.

Thus, the probability distribution of X is

$$b(x;n,p) = P(x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad x = 0, 1, 2, \cdots n$$

Practical Interpretation: Binomial random variables are used to model many physical systems and probabilities for all such models can be obtained from the binomial probability mass function.

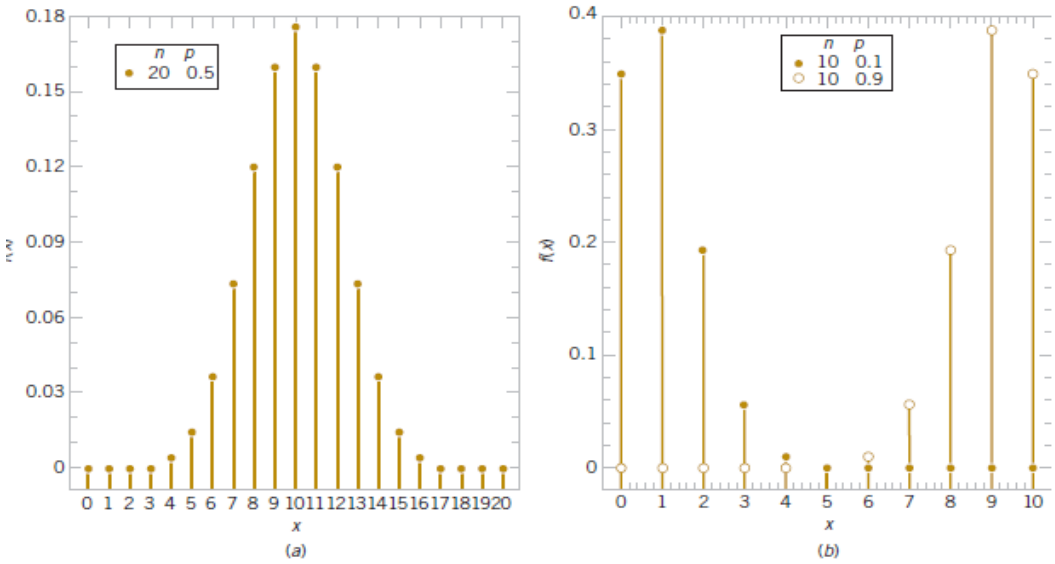


FIGURE 5-5: Binomial distributions for selected values of n and p

Areas of Application

- An industrial engineer is strongly interested in the "proportion defective" in an industrial process. Often, quality control measures and sampling schemes for processes are based on the binomial distribution.
- This distribution applies to any industrial situation where an outcome of a process is bilateral and the results of the process are independent, with the probability of success being constant from trial to trial.
- The binomial distribution is also used extensively for medical and military applications. In both fields, a success and failure result is important. For example,

"cure" or "no cure" is important in pharmaceutical work, and "hit" or "miss" is often the interpretation of the result of firing a guided projectile.

Properties of binomial distribution:

1. Binomial distribution has two parameters- n and p $(or\ q)$
2. Mean $=np$
3. Variance $=npq$
4. Standard deviation $= \sqrt{npq}$
5. Skewness, $\beta_1 = \frac{(q-p)^2}{npq}$
6. Kurtosis, $\beta_2 = 3 + \frac{1-6pq}{npq}$
7. Binomial distribution the symmetrical if $p = q = 0.5$

Example5.9:

Organic Pollution each sample of water has a 10% chance of containing a particular organic pollutant. Assume that the samples are independent with regard to the presence of the pollutant. Determine

(i) The probability that in the next 18 samples, exactly 2 contain the pollutant;

(ii) The probability that at least four samples contain the pollutant,

(iii) The probability that 3 ≤ X < 7.

Solution:

(i) Let X = the number of samples that contain the pollutant in the next 18 samples analyzed. Then X is a binomial random variable with p = 0.1 and n = 18. Therefore,

$$P(X = 2) = \binom{18}{2}(0.1)^2(1-0.1)^{18-2} = 0.284$$

(ii) The requested probability that at least four samples contain the pollutant is

$$P(X \geq 4) = 1 - P(X < 4) = 1 - \sum_{x=0}^{3}\binom{18}{x}(0.1)^x(0.9)^{18-x} = 1 - [0.150 + 0.300 + 0.284 + 0.168] = 0.098$$

(iii) The requested probability that 3 ≤ X < 7 is

$$P(3 \leq X < 7) = \sum_{x=3}^{6}\binom{18}{x}(0.1)^x \times (0.9)^{18-x} = 0.168 + 0.070 + 0.022 + 0.005 = 0.265.$$

Example 5.10:

A box contains 100 transistors, 20 of which are defective, 10 are selected for inspection. Indicate what is the probability that.

(i) all 10 are defective.

(ii) all 10 are non-defective.
(iii) at least one is defective, and
(iv) at the most 3 are defective.

Solution:

Let X represent the number of defective transistors selected. Then the possible values of X are $0, 1, 2, \dots, 10$. Now,

$$p = \frac{20}{100} = \frac{1}{5}, q = 1 - \frac{1}{5}$$

The probability of X defective transistors is $p(x) = {}^{10}C_x \left(\frac{1}{5}\right)^x \left(\frac{4}{5}\right)^{10-x}$

(i) Property that all 10 are defective is.

$$p(10) = {}^{10}C_{10} \left(\frac{1}{5}\right)^{10} \left(\frac{4}{5}\right)^{10-10} = \frac{1}{5^{10}}$$

(ii) Probability that all 10 are non-defective is.

$$= 1 - \text{P (all are defective)}$$

$$= 1 - \frac{1}{5^{10}}$$

(iii) Probability that at least one is defective is.

$$= p(1) + p(2) + \cdots + p(10)$$

$$\text{Or} = 1 - p(0) = 1 - {}^{10}C_0 \left(\frac{1}{5}\right)^0 \left(\frac{4}{5}\right)^{10-0} = 1 - \left(\frac{4}{5}\right)^{10}$$

(iv) Probability of at the most 3 defective items is.

$$p(x < 3) = p(X = 0) + p(X = 1) + p(X = 2) + p(X = 3)$$

$$= p(0) + p(1) + p(2) + p(3) = 0.859$$

Example 5.11: The probability that a patient recovers from a rare blood disease is 0.4. If 15 people are known to have contracted this disease, what is the probability that (a) at least 10 survive, (b) from 3 to 8 survive, and (c) exactly 5 survive?

Solution:

Let X be the number of people who survive.

(a) $P(X \geq 10) = 1 - P(X < 10) = 1 - \sum_{x=0}^{9} b(x; 15, 0.4) = 1 - 0.9662 = 0.0338$

(b) $P(3 \leq X \leq 8) = \sum_{x=3}^{8} b(x; 15, \ 0.4) = \sum_{x=0}^{8} b(x; 15, \ 0.4) - \sum_{x=0}^{2} b(x; 15, 0.4) = 0.9050 - 0.0271 = 0.8779.$

(c) $P(X = 5) = b(5; 15, 0.4) = \sum_{x=0}^{5} b(x; 15, 0.4) - \sum_{x=0}^{4} b(x; 15, 0.4) = 0.4032 - 0.2173 = 0.1859.$

Example 5.12: The probability that a patient recovers from a delicate heart operation is 0.10. If 15 people are known to have this operation what is the probability that

a) at least 5 survive.
b) from 4 to 8 survive and
c) exactly four survive?

Solution:

Let X denote the number of people that survive after operation. Then the binomial distribution of the random variable X is $b(x; 15,0.10)$

$$\text{i.e., } f(x) = \binom{15}{x}(0.10)^x(1-0.10)^{15-x}, x = 0, 1, \dots, 15$$

a) $P(x \geq 5) = 1 - P(x \leq 4) = 1 - F(4) = 1 - 0.9873 = 0.0127$
b) $P(4 \leq x \leq 8) = P(x \leq 8) - P(x \leq 3) = F(8) - F(3) = 1 - 0.9444 = 0.0556$
c) $P(x = 4) = P(x \leq 4) - P(x \leq 3) = F(4) - F(3) = 0.9873 - 0.9444 = 0.0429$

Example 5.13: It is conjectured that an impurity exists in 30% of all drinking wells in a certain rural community. In order to gain some insight into the true extent of the problem, it is determined that some testing is necessary. It is too expensive to test all of the wells in the area, so 10 are randomly selected for testing.

(a) Using the binomial distribution, what is the probability that exactly 3 wells have the impurity, assuming that the conjecture is correct?

(b) What is the probability that more than 3 wells are impure?

Solution:

(a) We require

$$b(3; 10, 0.3) = \sum_{x=0}^{3} b(x; 10, 0.3) - \sum_{x=0}^{2} b(x; 10, 0.3) = 0.6496 - 0.3828 = 0.2668$$

(b) $P(X > 3) = 1 - \sum_{x=0}^{3} b(x; 10,\ 0.3) = 1 - 0.6496 = 0.3504$

Example 5.14: A traffic control officer reports that 75% of the trucks passing through a check post are from within Dhaka city. What is the probability that at least three of next five trucks for from out of the city?

Solution:

Let X be the number of trucks that pass through are from out of Dhaka city. The probability of such an event is then $P = 1 - 0.75 = 0.25 = \frac{1}{4}$

Hence, $P(x \geq 3) = \sum_{x=3}^{5} b\left(x; 5, \frac{1}{4}\right) = b\left(3; 5, \frac{1}{4}\right) + b\left(4; 5, \frac{1}{4}\right) + b\left(5; 5, \frac{1}{4}\right)$

$= {}^5C_3\left(\frac{1}{4}\right)^3\left(\frac{3}{4}\right)^2 + {}^5C_4\left(\frac{1}{4}\right)^4\left(\frac{3}{4}\right)^1 + {}^5C_5\left(\frac{1}{4}\right)^5\left(\frac{3}{4}\right)^0$

$= \frac{90}{1024} + \frac{15}{1024} + \frac{1}{1024} = \frac{106}{1024} = 0.1035$

Example 5.15: Twenty percent of the TVS produced in an industry are defective. If 4 TVS are put in a box for marketing. In how many boxes do you expect to have

(i) one defective TVS
(ii) two defective TVS
(iii) at most two defective TVS in a consignment of 2000 such boxes

Solution:

Here $P = 20\% = \frac{1}{5}, q = 1 - p = \frac{4}{5}$

If X stands for the number of defective TVS, then X can assume values $0,1,2,3,4$. Hence,

i) $P(X = 1) = \binom{4}{1}\left(\frac{1}{5}\right)\left(\frac{4}{5}\right)^3 = \frac{256}{625} = 0.4096$

Hence the no of boxes having one defective TVS is

$$N \times P(X = 1) = 2000 \times 0.4096 = 819$$

ii) $P(X = 2) = \binom{4}{2}\left(\frac{1}{5}\right)^2\left(\frac{4}{5}\right)^2 = \frac{96}{625} = 0.1536$

Hence the no of boxes having two defective TVS is

$$N \times P(X = 2) = 2000 \times 0.1536 = 307$$

iii) $P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$

$$= \binom{4}{0}\left(\frac{1}{5}\right)^0\left(\frac{4}{5}\right)^4 + \binom{4}{1}\left(\frac{1}{5}\right)^1\left(\frac{4}{5}\right)^3 + \binom{4}{2}\left(\frac{1}{5}\right)^2\left(\frac{4}{5}\right)^2$$

$$= \frac{608}{625}$$

$$= 0.9728$$

Hence the no of boxes having at most two defective TVS is,

$$N \times P(X \leq 2) = 2000 \times 0.9728 = 1946$$

2. **Negative Binomial Distribution:**

Let us consider an experiment where the properties are the same as those listed for a binomial experiment, with the exception that the trials will be repeated until a fixed number of successes occur. Therefore, instead of the probability of x successes in n trials, where n is fixed, we are now interested in the probability that the k^{th} success occurs on the x^{th} trial. Experiments of this kind are called negative binomial experiments.

As an illustration, consider the use of a drug that is known to be effective in 60% of the cases where it is used. The drug will be considered a success if it is effective in bringing some degree of relief to the patient. We are interested in finding the probability that the fifth patient to experience relief is the seventh patient to receive the drug during a given week. Designating a success by S and a failure by F, a possible order of achieving the desired result is SFSSSFS, which occurs with probability

$(0.6)(0.4)(0.6)(0.6)(0.6)(0.4)(0.6) = (0.6)^5(0.4)^2$.

We could list all possible orders by rearranging the F's and S's except for the last outcome, which must be the fifth success. The total number of possible orders is equal to the number of partitions of the first six trials into two groups with 2 failures assigned to the one group and 4 successes assigned to the other group. This can be done in $\binom{6}{4} = 15$ mutually exclusive ways. Hence, if X represents the outcome on which the fifth success occurs, then

$$P(X = 7) = \binom{6}{4}(0.6)^5(0.4)^2 = 0.1866.$$

Negative Binomial Random Variable:

The number X of trials required to produce k successes in a negative binomial experiment is called a negative binomial random variable, and its probability distribution is called the negative binomial distribution. Since its probabilities depend on the number of successes desired and the probability of a success on a given trial, we shall denote them by $b^*(x; k, p)$.

If repeated independent trials can result in a success with probability p and a failure with probability q = 1 − p, then the probability distribution of the random variable X, the number of the trial on which the kth success occurs, is

$$b^*(x; k, p) = \binom{x-1}{k-1} p^k q^{x-k}, \quad x = k, k+1, k+2, \cdots$$

Example 5.16:

In an NBA (National Basketball Association) championship series, the team that wins four games out of seven is the winner. Suppose that teams A and B face each other in the championship games and that team A has probability 0.55 of winning a game over team B.

(a) What is the probability that team A will win the series in 6 games?

(b) What is the probability that team A will win the series?

(c) If teams A and B were facing each other in a regional playoff series, which is decided by winning three out of five games, what is the probability that team A would win the series?

Solution:

(a) $b^*(6;4,0.55) = \binom{5}{3}(0.55)^4(1-0.55)^{6-4} = 0.1853$

(b) $P(\text{team A wins the championship series}) = b^*(4;4,0.55) + b^*(5;4,0.55) + b^*(6;4,0.55) + b^*(7;4,0.55) = 0.0915 + 0.1647 + 0.1853 + 0.1668 = 0.6083$

(c) $P(\text{team A wins the playoff series}) = b^*(3;3,0.55) + b^*(4;3,0.55) + b^*(5;4,0.55) = 0.1664 + 0.2246 + 0.2021 = 0.5931.$

3. Poisson distribution:

Experiments yielding numerical values of a random variable X, the number of outcomes occurring during a given time interval or in a specified region, are called Poisson experiments. The given time interval may be of any length, such as a minute, a day, a week, a month, or even a year.

For example, a Poisson experiment can generate observations for the random variable X representing the number of telephone calls received per hour by an office, the number of days school is closed due to snow during the winter, or the number of games postponed due to rain during a baseball season. The specified region could be a line segment, an area, a volume, or perhaps a piece of material. In such instances, X might represent the number of field mice per acre, the number of bacteria in a given culture, or the number of typing errors per page.

A Poisson experiment is derived from the Poisson process and possesses the following properties.

Properties of the Poisson Process

1. The number of outcomes occurring in one time interval or specified region of space is independent of the number that occur in any other disjoint time interval or region. In this sense we say that the Poisson process has no memory.
2. The probability that a single outcome will occur during a very short time interval or in a small region is proportional to the length of the time interval or the size of the region and does not depend on the number of outcomes occurring outside this time interval or region.
3. The probability that more than one outcome will occur in such a short time interval or fall in such a small region is negligible.

The number X of outcomes occurring during a Poisson experiment is called a Poisson random variable, and its probability distribution is called the Poisson distribution.

The mean number of outcomes is computed from μ = λt, where t is the specific "time," "distance," "area," or "volume" of interest. Since the probabilities depend on λ, the rate of occurrence of outcomes, we shall denote them by p(x; λt).

The following formula is used for computing Poisson probabilities.

The probability distribution of the Poisson random variable X, representing the number of outcomes occurring in a given time interval or specified region denoted by t, is

$$p(x;\lambda t) = \frac{e^{-\lambda t}(\lambda t)^x}{x!}, x = 0, 1, 2, \cdots$$

where λ is the average number of outcomes per unit time, distance, area, or volume and e = 2.71828.

The Poisson distribution is a discrete distribution with a parameter λ.

The various constants are:

1. Mean $= \lambda = P$
2. Standard deviation $= \sqrt{\lambda}$
3. Skewness, $\beta_1 = \frac{1}{\lambda}$
4. Kurtosis, $\beta_2 = 3 + \frac{1}{\lambda}$
5. Variance $= \lambda$

Example 5.17:

During a laboratory experiment, the average number of radioactive particles passing through a counter in 1 millisecond is 4. What is the probability that 6 particles enter the counter in a given millisecond?

Solution:

Using the Poisson distribution with x = 6 and $\lambda t = 4$ and referring table A.2, we have

$$p(6;4) = \frac{e^{-4}4^6}{6!} = \sum_{x=0}^{6} p(x;4) - \sum_{x=0}^{5} p(x;4) = 0.8893 - 0.7851 = 0.1042$$

Example 5.18:

Ten is the average number of oil tankers arriving each day at a certain port. The facilities at the port can handle at most 15 tankers per day. What is the probability that on a given day tankers have to be turned away?

Solution:

Let X be the number of tankers arriving each day. Then, using Table A.2, we have

$P(X > 15) = 1 - P(X \leq 15) = 1 - \sum_{x=0}^{15} p(x;10) = 1 - 0.9513 = 0.0487$

Example 5.19: Telephone calls enter a switchboard on the average of two every 3-minute interval. What is the probability of five or more call arriving in a 9 minute?

Solution:

Let X denote the number of calls in a 9-minute period. We see the at the average number of calls in a 9-minute interval is $\lambda = \frac{9}{3} \times 2 = 6$. So the distribution of X is Poisson with $\lambda = 6$ and

$f(x) = \frac{e^{-6}6^x}{x!}, x = 0, 1, 2, \ldots,$

$P(X \geq 5) = 1 - P(x \leq 4) = 1 - \sum_{x=0}^{4} \frac{e^{-6}6^x}{x!} = 1 - 0.285$

$= 1 - E(4) = 0.715.$

Example 5.20:

The number of fireballs found by a radar system in any 30-second interval under specified conditions averages 1.81. Assume the fireballs appear randomly and independently.

a) What is the probability that no fireballs are found in a one-minute interval?

b) What is the probability of observing at least five but not more than eight fireballs in two minutes of observation?

Solution:

a) $\lambda = (1.81) / (0.50 \text{ minute}) = 3.62 / \text{minute}$.

For a one-minute interval, $\mu = \lambda t = 3.62$.

P [none in one minute] = $e^{-\lambda t} = e^{-3.62} = 0.0268$.

b) For two minutes, $\mu = \lambda t = (3.62)(2) = 7.24$.

P [R = r] = $\{(\lambda t)^r e^{-\lambda t}\} / r!$

Then P [R = 5] = $\{(7.24)^5 e^{-7.24}\} / 5! = 0.1189$.

From equation, P [R = r+1] = {λt / r+1} P [R = r], we get

P [R = 6] = {7.24/ 6} (0.1189) = 0.1435,

P [R=7] = {7.24/ 7} (0.1435) = 0.1484,

and P [R=8] = {7.24/ 8} (0.1484) = 0.1343.

Then P [at least five but not more than eight meteors in two minutes]

= P [5 or 6 or 7 or 8 meteors in two minutes]

= 0.1189+0.1435+0.1484+0.1343 = 0.545

Example 5.21: The average number of accidents occurring in a week during the summer months at a particular intersection is 2.00. Assume that the requirements of the Poisson distribution are satisfied.

a) What is the probability of no accidents in any particular week?

b) What is the probability that there will be exactly one accident in a week?

c) What is the probability of exactly two accidents in a week?

d) What is the probability of finding not more than two accidents in a week?

e) What is the probability of finding more than two accidents in a week?

f) What is the probability of exactly two accidents in a particular two-week interval?

Solution:

Here λ = 2.00/week, t = 1 week, so λt = 2.00.

a) P [A = 0] = $e^{-\lambda t} = e^{-2.00}$ = 0.135.

b) P [exactly one accident in a week] = P [A = 1] = (λt) $e^{-\lambda t} = 2.00\ e^{-2.00}$ = 0.271.

c) P [exactly two accidents in a week] = P [A = 2] = $\{(\lambda t)^2\ e^{-\lambda t}\}/\ 2! = \{(2.00)^2\ e^{-2.00}\}/2!$ = 0.271.

d) P [not more than two accidents in a week] = P [A ≤ 2]

= P [A = 0] + P [A = 1] + P [A = 2] = 0.135 + 0.271 + 0.271 = 0.677

e) P [more than two accidents in a week] = P [A > 2] = 1– P [A ≤ 2] = 1 – 0.677

= 0.323.

f) Now we still have λ = 2.00/week, but t = 2 weeks, so λt = 4.00

Then P [exactly two accidents in a two-week interval] = $\{(\lambda t)^2\ e^{-\lambda t}\}/\ 2!$

= $\{(4.00)^2\ e^{-4.00}\}/\ 2!$ = 0.147.

Example 5.22:

The demand for a particular type of pump at an isolated mine is random and independent of previous occurrences, but the average demand in a week (7 days) is for 2.8 pumps. Further supplies are ordered each Tuesday morning and arrive on the weekly plane on Friday morning. Last Tuesday morning only one pump was in stock, so the stores man ordered six more to come in Friday morning.

a) Find the probability that one pump will still be in stock on Friday morning when new stock arrives.

b) Find the probability that stock will be exhausted and there will be unsatisfied demand for at least one pump by Friday morning.

c) Find the probability that one pump will still be in stock this Friday morning and at least five will be in stock next Tuesday morning.

Solution:

First, we have to recognize that the Poisson distribution will apply.

λ = 2.8 / 7 days = 0.4 / day.

a) From Tuesday morning to Friday morning is three days.

Then $\lambda t = (0.4 / day) (3 days) = 1.2$.

P [no demand in three days] = $e^{-\lambda t} = e^{-1.2} = 0.3012$.

Then P [one pump will still be in stock Friday morning when new stock arrives] = 0.301.

b) P [demand for two or more pumps in three days]

= 1 – P [demand for zero or one pump in three days]

= 1 – P [demand for no pumps in three days] – P [demand for one pump in three days]

= 1 – 0.3012 – {(0.3012) (1.2)}/1 (using P [R = r+1] = {λt / r+1} P [R = r])

= 0.3374.

Then P [unsatisfied demand for at least one pump by Friday morning] = 0.337.

c) From part (a), P [one pump will still be in stock this Friday morning] = 0.3012.

From Friday morning to Tuesday morning is four days, so

$(\lambda t) = (0.4 /day) (4 days) = 1.6$.

After the new stock arrives, we will have 1 + 6 = 7 pumps in stock Friday morning.

If we have at least five in stock Tuesday morning, the demand in four days is ≤ 2

pumps.

P [demand for 0 pumps in 4 days] = $e^{-1.6} = 0.2019$.

P [demand for 1 pump in 4 days] = $((e^{-1.6}) (1.6)/1) = 0.3230$.

P [demand for 2 pumps in 4 days] = $\{(e^{-1.6}) (1.6)^2\} / 2 = 0.2584$.

Then P [demand for 2 or fewer pumps in 4 days] = 0.7834.

Then P [at least 5 will be in stock next Tuesday morning | one pump in stock Friday morning] = 0.7834. Note that this is a conditional probability.

Theorem: Both the mean and the variance of the Poisson distribution p(x; λt) are λt

Nature of the Poisson Probability Function Like so many discrete and continuous distributions, the form of the Poisson distribution becomes more and more symmetric, even bell-shaped, as the mean grows large. Figure 5.6 illustrates this, showing plots of the probability function for μ = 0.1, μ = 2, and μ = 5. Note the nearness to symmetry when μ becomes as large as 5.

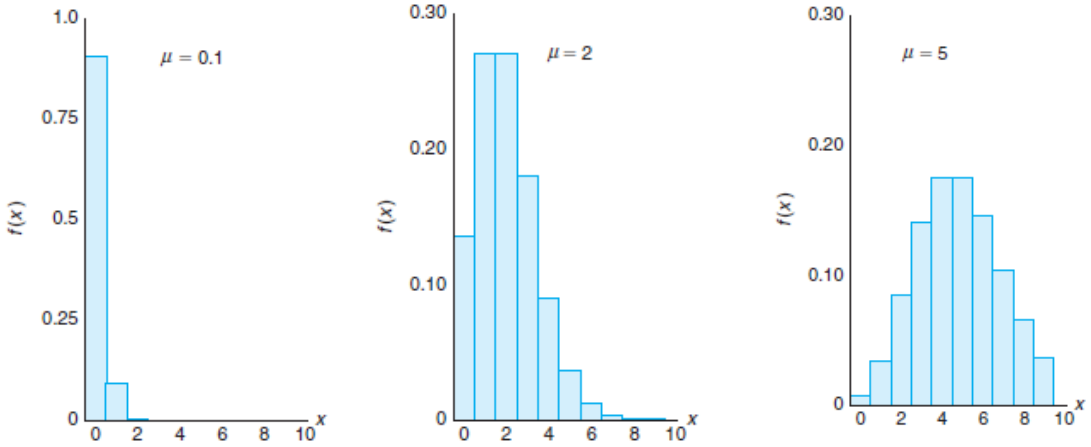


Figure 5.6: Poisson density functions for different means.

The Poisson distribution has only one parameter, μ, whereas the binomial distribution has two parameters, n and p. Probabilities according to the Poisson distribution are easier to calculate with a pocket calculator than for the binomial distribution, especially for very large values of n and very small values of p.

Approximation of Binomial Distribution by a Poisson distribution:

In the case of the binomial, if n is quite large and p is small, the conditions begin to simulate the continuous space or time implications of the Poisson process.

Indeed, if n is large and p is close to 0, the Poisson distribution can be used, with μ = np, to approximate binomial probabilities. If p is close to 1, we can still use the Poisson distribution to approximate binomial probabilities by interchanging what we have defined to be a success and a failure, thereby changing p to a value close to 0.

Let X be a binomial random variable with probability distribution b(x; n, p). When $n \to \infty, p \to 0$ and $np \to \mu$ remains constant, $b(x; n, p) \to p(x; \mu)$ $when$ $n \to \infty$.

Example 5.20: In a certain industrial facility, accidents occur infrequently. It is known that the probability of an accident on any given day is 0.005 and accidents are independent of each other.

(a) What is the probability that in any given period of 400 days there will be an accident on one day?

(b) What is the probability that there are at most three days with an accident?

Solution:

Let X be a binomial random variable with n = 400 and p = 0.005. Thus, np = 2.

Using the Poisson approximation,

(a) $P(X = 1) = 1 - e^{-2}2^1 = 0.271$

(b) $P(X \leq 3) = \sum_{x=0}^{3} \frac{e^{-2}2^1}{x!} = 0.857$

Example 5.21: In a manufacturing process where glass products are made, defects or bubbles occur, occasionally rendering the piece undesirable for marketing. It is known that, on average, 1 in every 1000 of these items produced has one or more bubbles. What is the probability that a random sample of 8000 will yield fewer than 7 items possessing bubbles?

Solution:

This is essentially a binomial experiment with n = 8000 and p = 0.001. Since p is very close to 0 and n is quite large, we shall approximate with the Poisson distribution using μ = (8000) x (0.001) = 8.

Hence, if X represents the number of bubbles, we have

$$P(X < 7) = \sum_{x=0}^{6} b(x; 8000, 0.001) \approx p(x; 8) = 0.3134$$

Example 5.22: 5% of the tools produced by a certain process are defective. Find the probability that in a sample of 40 tools chosen at random, exactly three will be defective. Calculate a) using the binomial distribution, and b) using the Poisson distribution as an approximation.

Solution:

a) For the binomial distribution with n = 40, p = 0.05,

Pr [D = 3] = $^{40}C_3$ $(0.05)^3(0.95)^{37}$ = 0.185

b) For the Poisson distribution, μ = (n)(p) = (40) (0.05) = 2.00.

Pr [D = 3] = $\{(2.00)^3\ e^{-2.00}\}$ / (3) (2)(1) = 0.180

For Practice

1. The probability that a thirty-year-old man will survive a fixed length of time is 0.995. The probability that he will die during this time is therefore 1– 0.995 = 0.005. An insurance company will sell him a $20,000 life insurance policy for this length of time for a premium of $200.00. What is the expected gain for the insurance company?

2. On the basis of past experience, the probability that a certain electrical component will be satisfactory is 0.98. The components are sampled item by item from continuous production. In a sample of five components, what are the probabilities of finding (a) zero, (b) exactly one, (c) exactly two, (d) two or more defectives?

3. A company is considering drilling four oil wells. The probability of success for each well is 0.40, independent of the results for any other well. The cost of each well is $200,000. Each well that is successful will be worth $600,000.

a) What is the probability that one or more wells will be successful?

b) What is the expected number of successes?

c) What is the expected gain?

d) What will be the gain if only one well is successful?

e) Considering all possible results, what is the probability of a loss rather than a gain?

f) What is the standard deviation of the number of successes?

4. A boiler containing eight welds is manufactured in a small shop. When the boiler is completed, each weld is checked by an inspector. If more than one weld is defective on a single boiler, the person who made that boiler is reported to the foreman.

a) If 9.0% of all welds made by Joe Smith are defective, what percentage of all boilers made by him will have more than one defective weld?

b) Over a long period of time how many times will Joe Smith be reported to the foreman for each 15 boilers he makes?

c) If Joe makes 15 boilers in a shift, what is the probability that he will be reported for more than two of these 15 boilers?

5. A small hotel has rooms on only four floors, with four smoke detectors on each floor. Because of improper maintenance, the probability that any one detector is functioning is only 0.55. The probabilities that smoke detectors are functioning are randomly and independently distributed.

a) What is the probability that exactly one smoke detector is working on the top floor?

b) What is the probability that there is exactly one detector working on each of two floors and there are two detectors working on each of the other two floors?

c) What is the probability that there will be no functioning smoke detectors on one particular floor? What is the probability that there will be at least one functioning smoke detector on that floor?

d) What is the probability that on at least one of the four floors there will be no functioning smoke detectors?

e) What is the probability that there will be at least 15 functioning smoke detectors in the hotel at any one time?

6. The probability that a certain type of IC chip will fail after installation is 0.06. A memory board for a computer contains twelve such chips. The operation will be satisfactory if ten or more of the chips on the board do not fail.

a) What is the probability that a memory board operates satisfactorily?

b) If there are five such memory boards in a given computer, what is the probability that at least four of them operate satisfactorily?

c) State any assumptions.

7. The number of meteors found by a radar system in any 30-second interval under specified conditions averages 1.81. Assume the meteors appear randomly and independently.

a) What is the probability that no meteors are found in a one-minute interval?

b) What is the probability of observing at least five but not more than eight meteors in two minutes of observation?

8. 5% of the tools produced by a certain process are defective. Find the probability that in a sample of 40 tools chosen at random, exactly three will be defective. Calculate a) using the binomial distribution, and b) using the Poisson distribution as an approximation.