Chapter 5 Discrete Probability Distributions

The observations generated by different statistical experiments have the same general type of behavior. 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. The following are the probability distributions that will be covered in this chapter:

- Discrete Uniform Distribution
- Binomial Distribution
- Multinomial Distribution
- Hypergeometric Distribution
- Geometric Distribution
- Poisson Distribution and Poisson Process

The Discrete Uniform Distribution

**Definition.** If the random variable $X$ assumes the values $x_1, x_2, \ldots, x_k$ with equal probabilities, then the discrete uniform distribution is given by

$$f(x; k) = \frac{1}{k}, \quad x = x_1, x_2, \ldots, x_k.$$

**Example 1.** When a die is tossed, each element of the sample space $S = \{1, 2, 3, 4, 5, 6\}$ occurs with probability $1/6$. Therefore, we have a uniform distribution with

$$f(x; 6) = \frac{1}{6}, \quad x = 1, 2, 3, 4, 5, 6.$$

**Theorem.** The mean and variance of the discrete uniform distribution $f(x; k)$ are

$$\mu = \frac{\sum_{i=1}^{k} x_i}{k} \quad \text{and} \quad \sigma^2 = \frac{\sum_{i=1}^{k} (x_i - \mu)^2}{k}.$$

The Binomial Distribution

Perhaps the most commonly used discrete probability distribution is the binomial distribution. An experiment which follows a binomial distribution will satisfy the following requirements (think of repeatedly flipping a coin as you read these):

1. The experiment consists of $n$ identical trials, where $n$ is fixed in advance.
2. Each trial has two possible outcomes, $S$ or $F$, which we denote `success" and `failure" and code as 1 and 0, respectively.
3. The trials are independent, so the outcome of one trial has no effect on the outcome of another.
4. The probability of success, $p$ is constant from one trial to another.
Definition. The random variable $X$ of a binomial distribution counts the number of successes in $n$ trials. The probability that $X$ is a certain value $x$ is given by the formula

$$P(X = x) = b(x; n, p) = \binom{n}{x} p^x (1 - p)^{n-x}$$

where $0 \leq p \leq 1$ and $x = 0, 1, 2, \ldots, n$. Recall that the quantity $\binom{n}{x}$, "$n$ choose $x$," above is

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

Mean and Variance of a Binomial Random Variable. We could use the formulas previously given to compute the mean and variance of $X$. However, for the binomial distribution these will always be equal to

$$E(X) = \mu = np \quad \text{and} \quad Var(X) = \sigma^2 = npq$$

Note: A particularly important example of the use of the binomial distribution is when sampling with replacement (this implies that $p$ is constant).

Example 1. Suppose we have 10 balls in a bowl, 3 of the balls are red and 7 of them are blue. Define success $S$ as drawing a red ball. If we sample with replacement, $P(S) = 0.3$ for every trial. Let's say $n = 20$, then $X \sim b(x; 20, 0.3)$ and we can figure out any probability we want. For example,

$$P(X = 5) = \binom{20}{5} 0.3^5 (1 - 0.3)^{20-5} = 0.1789.$$  

The mean and variance are

$$E(X) = 20(0.3) = 6, \quad \text{Var}(X) = 20(0.3)(0.7) = 4.2$$

We are interested in problems where it is necessary to find $P(X < r)$ or $P(a \leq X \leq b)$. Fortunately binomial sums

$$B(r; n, p) = \sum_{x=0}^{r} b(x; n, p)$$

are available and are given in Table A.1 of the appendix for $n = 1, 2, \ldots, 20$. 

Areas of Applications

In statistics, the so-called binomial distribution describes the possible number of times that a particular event will occur in a sequence of observations. The event is coded binary, it may or may not occur. The binomial distribution is used when a researcher is interested in the occurrence of an event, not in its magnitude. For instance, in a clinical trial, a patient may survive or die. The researcher studies the number of survivors, and not how long the patient survives after treatment. Another example is whether a person is ambitious or not. Here, the binomial distribution describes the number of ambitious persons, and not how ambitious they are.

The binomial distribution is specified by the number of observations, n, and the probability of occurrence, which is denoted by p.

A classic example that is used often to illustrate concepts of probability theory, is the tossing of a coin. If a coin is tossed 4 times, then we may obtain 0, 1, 2, 3, or 4 heads. We may also obtain 4, 3, 2, 1, or 0 tails, but these outcomes are equivalent to 0, 1, 2, 3, or 4 heads. The likelihood of obtaining 0, 1, 2, 3, or 4 heads is, respectively, 1/16, 4/16, 6/16, 4/16, and 1/16. In the figure on this page the distribution is shown with p = 1/2. Thus, in the example discussed here, one is likely to obtain 2 heads in 4 tosses, since this outcome has the highest probability.

Other situations in which binomial distributions arise are quality control, public opinion surveys, medical research, and insurance problems.

Example 2. 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 2 survive?
(b) between 3 and 7 survive?
(c) exactly 5 survive?

Solution. Probability of success= p = 0.4, and the probability of failure= q = 0.6.

\( n = 15 \) and \( X \) : no. of surviving patients

(a) \( P(X \geq 2) = 1 - P(X \leq 1) = 1 - [P(X = 0) + P(X = 1)] = 1 - B(1; 15, 0.4) = 1 - 0.0052 = 0.9948 \)

(b) \( P(3 < X < 7) = P(4 \leq X \leq 6) = b(4; 15, 0.4) + b(5; 15, 0.4) + b(6; 15, 0.4) \)

\( = B(6; 15, 0.4) - B(3; 15, 0.4) = 0.6098 - 0.0905 = 0.5193 \)
Example 3. A traffic control engineer reports that 75% of the vehicles passing through a checkpoint are from within the state. What is the probability that fewer than 4 of the next 9 vehicles are from out of the state?

Solution. Probability of success = $p = 0.25$, and the probability of failure = $q = 1 - 0.25 = 0.75$.

$n = 9$ and $X$ : no. of vehicles passing through the checkpoint

$P(X < 4) = P(X \leq 3) = B(3; 9, 0.25) = 0.8343$

Example 4. Assuming that 6 in 10 automobile accidents are due mainly to speed violation,

(a) find the probability that among 8 automobile accidents 6 will be due mainly to a speed violation.

(b) Find the mean and variance of the number of automobile accidents for 8 automobile accidents.

Solution. Probability of success = $p = 6/10$, and the probability of failure = $q = 1 - 0.25 = 0.75$.

$n = 8$ and $X$ : no. of automobile accidents

(a) $P(X = 6) = b(6; 8, 6/10) = \binom{8}{6} \left(\frac{6}{10}\right)^6 \left(1 - \frac{6}{10}\right)^{8-6} = 0.2090$

(b) The mean of the number of automobile accidents is $\mu = E(X) = np = 8 * \frac{6}{10} = 4.8$.

The variance of the no. of auto. accidents is $\sigma^2 = Var(X) = npq = 8 * \frac{6}{10} * \frac{4}{10} = 1.92$.

Example 5. A coin is tossed 5 times. What is the probability that

(a) at most 2 heads will occur?

(b) At least 2 heads will occur?

Example 6. If the probability of a defective bolt is 0.1, find

(a) the mean

(b) the standard deviation for the distribution of defective bolts in a total of 400.

Example 7. One prominent physician claims that 70% of those with lung cancer are heavy smokers. If his assertion is correct, find the probability that out of 10 such patients recently admitted to a hospital

(a) more than half are heavy smokers

(b) exactly 4 are heavy smokers

(c) less than 2 are non-smokers
Example 8. It is estimated that 80% of people in Cyprus take paracetamol for headache treatment. Out of 5 having headache, find the probability that exactly 2 will take paracetamol.

Example 9. The probability that an entering college student will graduate is 0.4. Determine the probability that out of 5 students
   (a) none
   (b) at least 1
   (c) all will graduate.

The Multinomial Distribution

If a given trial can result in the \(k\) outcomes \(E_1, E_2, \ldots, E_k\) with probabilities \(p_1, p_2, \ldots, p_k\) then the probability distribution of the random variables \(X_1, X_2, \ldots, X_k\) representing the number of occurrences for \(E_1, E_2, \ldots, E_k\) in \(n\) independent trials is

\[
P(x_1, x_2, \ldots, x_k; p_1, p_2, \ldots, p_k, n) = \binom{n}{x_1, x_2, \ldots, x_k} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}
\]

with \(\sum_{i=1}^{k} x_i = n\) and \(\sum_{i=1}^{k} p_i = 1\).

Example 1. If a pair of dice are tossed 6 times, what is the probability of obtaining a total of 7 or 11 twice, a matching pair once and any other combinations 3 times?

Solution. \(E_1\): a total of 7 or 11 occurs
\(E_2\): a matching pair occurs
\(E_3\): neither a matching pair nor a total of 7 or 11 occurs

\[
P\left(\frac{2,1,3,2}{9,6,1,11,6}\right) = \binom{6}{2,1,3} \left(\frac{2}{9}\right)^2 \left(\frac{1}{6}\right)^3 = 0.1127.
\]

The Hypergeometric Distribution

The hypergeometric distribution arises when a random selection (without repetition) is made among objects of two distinct types. Typical examples:

- Choose a team of 8 from a group of 10 boys and 7 girls
- Choose a committee of five from the legislature consisting of 52 Democrats and 48 Republicans
**Definition.** The hypergeometric distribution is described by three parameters: \( N \), the total number of objects; \( k \), the number of objects of the first type; and \( n \) the number of objects to be chosen. The probability distribution of the hypergeometric random variable \( X \) is

\[
P(X = x) = h(x; N; n; k) = \frac{\binom{k}{x} \binom{N-k}{n-x}}{\binom{N}{n}}, \quad x = 0, 1, 2, 3, \ldots, k
\]

**Theorem.** The mean and variance of the hypergeometric distribution \( h(x; N; n; k) \) are

\[
\mu = \frac{nk}{N} \quad \text{and} \quad \sigma^2 = \frac{N-n}{N-1} \frac{k}{N} \left( 1 - \frac{k}{N} \right). 
\]

**Example 1.** From a lot of 10 missiles, 4 are selected at random and fired. If the lot contains 3 defective missiles that will not fire, what is the probability that

(a) all 4 will fire?
(b) at most 2 will not fire?

**Solution.** \( N = 10, \ n = 4, \ k = 3 \).

(a) Let \( X \) : number of non-defective missiles,

\[
P(X = 4) = \frac{\binom{7}{4} \binom{3}{0}}{\binom{10}{4}} = \frac{35}{210} = 0.1666
\]

(b) Let \( Y \) : number of defective missiles,

\[
P(Y \leq 2) = \frac{\binom{3}{0} \binom{7}{4}}{\binom{10}{4}} + \frac{\binom{3}{1} \binom{7}{3}}{\binom{10}{4}} + \frac{\binom{3}{2} \binom{7}{2}}{\binom{10}{4}} = \frac{35}{210} + \frac{105}{210} + \frac{63}{210} = \frac{203}{210}.
\]

**Example 2.** A random committee of size 3 is selected from 4 doctors and 2 nurses. Write a formula for the probability distribution of the random variable \( X \) representing the number of doctors on the committee. Find \( P(2 \leq X \leq 3) \).

**Example 3.** A committee of size 5 is to be selected at random from 3 chemists and 5 physicists. Find the probability distribution for the number of chemists on the committee.
The Relationship to Binomial Distribution (Binomial Approximation to Hypergeometric Distribution)

There is an interesting relationship between the hypergeometric and binomial distribution. If \( n \) is small compared to \( N \), the nature of the \( N \) items changes very little in each draw. Thus the quantity \( \frac{k}{N} \) plays the role of the binomial parameter \( p \).

The mean and the variance then become

\[
\mu = np = \frac{nk}{N} \quad \text{and} \quad \sigma^2 = npq = \frac{N-n}{N-1} \cdot n \cdot \frac{k}{N} \left(1 - \frac{k}{N}\right).
\]

Example 1. A manufacturer of automobile tires reports that among a shipment of 5000 sent to a local distributor, 1000 are blemished. If one purchases 10 of these tires at random, what is the probability that exactly 3 are blemished?

Solution. \( N = 5000, \ n = 10, \ k = 1000 \).

Let \( X \) be the number of blemished tires, then

\[
P(X = 3) = h(3; 5000, 10, 1000) \approx b\left(3; 10, \frac{1}{5}\right) = \binom{10}{3} \left(\frac{1}{5}\right)^3 \left(\frac{4}{5}\right)^7 = 0.201312.
\]

The Geometric Distribution

Geometric Random Variables

This is an example of a waiting time problem, that is, we wait until a certain event occurs. We make the assumptions used in the binomial distribution so we assume that:

1. on each trial of our experiment, the result is one of two outcomes, "success" or "failure",

2. the trials are independent,

3. the probability of success at any trial is \( p \) so that \( P(s) = p \) and consequently, \( P(f) = 1 - p = q \).

4. the random variable, \( X \), denotes the number of successes in \( n \) trials.

While in the binomial we have fixed number of trials and a variable number of successes, in the geometric distribution we wait for a single success, but the number of trials is variable.

A perfect model for this situation is tossing a coin, loaded so that the probability of a head on a single toss is \( p \) (and the probability of a tail at a single toss is \( 1-p=q \)), until a head appears for the first time.
**Definition.** 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 first success occurs, is

\[
g(x; p) = pq^{x-1}, \quad x = 1, 2, 3, \ldots
\]

**Example 1.** The probability that a student pilot passes the written test for a private pilot’s license is 0.7, find the probability that the student will pass the test on the fourth try.

\[
g(x; p) = g(4; 0.7) = (0.7)(1 - 0.7)^{4-1} = (0.7)(0.3)^3 = 0.0189.
\]

**Example 2.** Find the probability that a person flipping a coin will get the first head on the third flip.

\[
g(x; p) = g(3; 0.5) = (0.5)(1 - 0.5)^{3-1} = (0.5)(0.5)^2 = 0.125.
\]

**Theorem.** The mean and variance of a random variable following the Geometric Distribution are

\[
\mu = \frac{1}{p} \quad \text{and} \quad \sigma^2 = \frac{1-p}{p^2}.
\]

**Example 3.** A shooter has the constant probability of 0.4 for hitting the target.

(a) What is the probability that the shooter will hit on the fifth try?

(b) What is the expected number of shootings for the shooter to hit the target?

**Solution.**

(a) \[ g(x; p) = g(5; 0.4) = (0.4)(1 - 0.4)^{5-1} = (0.4)(0.6)^4 = 0.05184 \]

(b) \[ E(X) = \mu = \frac{1}{0.4} = \frac{10}{4} = 2.5 \]

**The Poisson Distribution**

The Poisson distribution is most commonly used to model the number of random occurrences of some phenomenon in a specified unit of space or time. For example,

- The number of phone calls received by a telephone operator in a 10-minute period.
- The number of flaws in a bolt of fabric.
- The number of typos per page made by a secretary.
**Definition.** For a Poisson random variable, the probability that $X$ is some value $x$ is given by the formula

$$P(X = x) = f(x; \lambda t) = \frac{e^{-\lambda t} (\lambda t)^x}{x!} \quad x = 0, 1, 2, ...$$

where $\lambda$ is the average number of occurrences per unit time or region denoted by $t$. For the Poisson distribution,

$$E(X) = \lambda t \quad \text{and} \quad Var(X) = \lambda t.$$

**Example 1.** The number of false fire alarms in a suburb of Houston averages 2.1 per day. Assuming that a Poisson distribution is appropriate, the probability that 4 false alarms will occur on a given day is given by

$$P(X = 4) = \frac{2.1^4 e^{-2.1}}{4!} = 0.0992.$$

**Example 2.** 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?

**Example 3.** A secretary makes 2 errors per page, on average. What is the probability that on the next page he or she will make

(a) 4 or more errors?

(b) no errors?

**Example 4.** The number of customers arriving per hour at a certain automobile service facility is assumed to follow a Poisson distribution with $\lambda = 7$.

(a) Compute the probability that more than 10 customers will arrive in a 3-hour period.

(b) What is the mean number of arrivals during a 4-hour period?

**Example 5.** A restaurant chef prepares tossed salad containing, on average, 5 vegetables. Find the probability that the salad contains more than 5 vegetables

(a) on a given day

(b) on 3 of the next 4 days

(c) for the first time in April on April 5.
Poisson Approximation to Binomial Distribution

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

\[
b(x; n, p) \to p(x; \mu).
\]

**Example 1.** The probability that a person will die from a certain respiratory infection is 0.002. Find the probability that fewer than 5 of the next 2000 so infected will die.

\( X \): number of infected people who will die

Given \( n = 2000, \ p = 0.002, \ \mu = np = 2000 \times 0.002 = 4 \),

\[
P(X < 5) = P(X \leq 4) = \sum_{x=1}^{4} b(x; 2000, 0.002) \approx \sum_{x=1}^{4} p(x; 4) = P(4; 4) = 0.6288.
\]

**Example 2.** It is known that 5% of the books bound by a certain bindery have defective bindings. Find the probability that 2 of 100 books bound by this bindery will have defective bindings using

(a) the formula for the binomial distribution

(b) the Poisson approximation to the Binomial distribution.

**Solution.** (a) Given \( p = 0.05, n = 100 \), and \( X \), number of defective bindings,

\[
P(X = 2) = b(2; 100, 0.05) = \binom{100}{2} (0.05)^2 (0.95)^{98} = 0.081
\]

(b) \( P(X = 2) = b(2; 100, 0.05) \approx p(2; 5) = \frac{e^{-5} 5^2}{2!} = 0.084 \) where \( \lambda = np = 100 \times 0.05 = 5 \).