Chapter 3. Random Variables and Probability Distributions

1. Concept of a Random Variable

It is often important to **allocate a numerical description of each outcome** in statistical experiment.

1. Concept of a Random Variable

It is often important to **allocate a numerical description of each outcome** in statistical experiment.

Example: three electronic components are tested

$$S = \{NNN, NND, NDN, DNN, NDD, DND, DDN, DDD\}$$

1. Concept of a Random Variable

It is often important to **allocate a numerical description of each outcome** in statistical experiment.

Example: three electronic components are tested

$$S = \{NNN, NND, NDN, DNN, NDD, DND, DDN, DDD\}$$

Consider the number of defectives,

1. Concept of a Random Variable

It is often important to **allocate a numerical description of each outcome** in statistical experiment.

Example: three electronic components are tested

$$S = \{NNN, NND, NDN, DNN, NDD, DND, DDN, DDD\}$$

Consider the number of defectives,each sample point will be assigned a value of $0, 1, 2$, or 3.

1. Concept of a Random Variable

It is often important to **allocate a numerical description of each outcome** in statistical experiment.

Example: three electronic components are tested

$$S = \{NNN, NND, NDN, DNN, NDD, DND, DDN, DDD\}$$

Consider the number of defectives,each sample point will be assigned a value of $0, 1, 2$, or 3.

The number of defective items is a **random quantity** determined by the outcome of the experiment, so called **random variable** X.

Random Variable

Definition 3.1

A **random variable** is a function that associates a real number with each element in the sample space.

Random Variable

Definition 3.1

A **random variable** is a function that associates a real number with each element in the sample space.

X: a random variable

x: one of its values

Random Variable

Definition 3.1

A **random variable** is a function that associates a real number with each element in the sample space.

X: a random variable

x: one of its values

Each possible value of X represents an event

Random Variable

Definition 3.1

A **random variable** is a function that associates a real number with each element in the sample space.

X: a random variable

x: one of its values

Each possible value of X represents an event

electronic component test:

$\{X = 2\}$

Random Variable

Definition 3.1

A **random variable** is a function that associates a real number with each element in the sample space.

X: a random variable

x: one of its values

Each possible value of X represents an event

electronic component test:

$\{X = 2\} = \{DDN, DND, NDD\} = E.$

Random Variable

Example 3.1 Two balls are drawn in succession without replacement from an urn containing 4 red balls and 3 black balls. Y is the number of red balls. The possible outcomes and the values y of the random variable Y?

Random Variable

Example 3.2 A stockroom clerk returns three safety helmets at random to three steel mill employees who had previously checked them. If Smith, Jones, and Brown, in that order, receive one of the three hats, list the sample points for the possible orders of returning the helmets,and find the value m of the random variable M that represents the number of correct matches.

Random Variable

The sample space contains a finite number of elements in Example 3.1 and 3.2.

Random Variable

The sample space contains a finite number of elements in Example 3.1 and 3.2.

another example: a die is thrown until a 5 occurs,

F: the occurrence of a 5

N: the nonoccurrence of a 5

Random Variable

The sample space contains a finite number of elements in Example 3.1 and 3.2.

another example: a die is thrown until a 5 occurs,
F: the occurrence of a 5
N: the nonoccurrence of a 5

obtain a sample space with an unending sequence of elements

$$S = \{F, NF, NNF, NNNF, \dots\}$$

Random Variable

The sample space contains a finite number of elements in Example 3.1 and 3.2.

another example: a die is thrown until a 5 occurs,
F: the occurrence of a 5
N: the nonoccurrence of a 5

obtain a sample space with an unending sequence of elements

$$S = \{F, NF, NNF, NNNF, \dots\}$$

the number of elements can be equated to the number of **whole numbers**;

Random Variable

The sample space contains a finite number of elements in Example 3.1 and 3.2.

another example: a die is thrown until a 5 occurs,

F: the occurrence of a 5

N: the nonoccurrence of a 5

obtain a sample space with an unending sequence of elements

$$S = \{F, NF, NNF, NNNF, \dots\}$$

the number of elements can be equated to the number of **whole numbers**; **can be counted**

Random Variable

Definition 3.2 If a sample space contains a finite number of possibilities or an unending sequence with as many elements as there are whole numbers, it is called a **discrete sample space**.

Random Variable

The outcomes of some statistical experiments may be **neither finite nor countable.**

Random Variable

The outcomes of some statistical experiments may be **neither finite nor countable.**

Example: measure the distances that a certain make of automobile will travel over a prescribed test course on 5 liters of gasoline

Random Variable

The outcomes of some statistical experiments may be **neither finite nor countable.**

Example: measure the distances that a certain make of automobile will travel over a prescribed test course on 5 liters of gasoline

distance: a variable measured to any degree of accuracy

Random Variable

The outcomes of some statistical experiments may be **neither finite nor countable.**

Example: measure the distances that a certain make of automobile will travel over a prescribed test course on 5 liters of gasoline

distance: a variable measured to any degree of accuracy

we have infinite number of possible distances in the sample space, cannot be equated to the number of **whole numbers**.

Random Variable

Definition 3.3

If a sample space contains an infinite number of possibilities equal to the number of points on a line segment, it is called a **continuous sample space**

Random Variable

A random variable is called a **discrete random variable** if its set of possible outcomes is countable.

Random Variable

A random variable is called a **discrete random variable** if its set of possible outcomes is countable.

• Y in Example 3.1 and M in Example 3.2 are discrete random variables.

Random Variable

A random variable is called a **discrete random variable** if its set of possible outcomes is countable.

• Y in Example 3.1 and M in Example 3.2 are discrete random variables.

When a random variable can take on values on a continuous scale, it is called a **continuous random variable.**

Random Variable

A random variable is called a **discrete random variable** if its set of possible outcomes is countable.

• Y in Example 3.1 and M in Example 3.2 are discrete random variables.

When a random variable can take on values on a continuous scale, it is called a **continuous random variable.**

• The measured distance that a certain make of automobile will travel over a test course on 5 liters of gasoline is a continuous random varibale.

Random Variable

continuous random variables represent **measured data**: all possible heights, weights, temperatures, distance, or life periods.

Random Variable

continuous random variables represent **measured data**:
all possible heights, weights, temperatures, distance, or life periods.

discrete random variables represent **count data**:
the number of defectives in a sample of k items, or the number of highway fatalities per year in a given state.

2. Discrete Probability Distribution

A discrete random variable assumes each of its values with a certain probability

2. Discrete Probability Distribution

A discrete random variable assumes each of its values with a certain probability

assume equal weights for the elements in **Example 3.2**, what's the probability that no employee gets back his right helmet.

2. Discrete Probability Distribution

A discrete random variable assumes each of its values with a certain probability

assume equal weights for the elements in **Example 3.2**, what's the probability that no employee gets back his right helmet.

$P(M = 0) =$

2. Discrete Probability Distribution

A discrete random variable assumes each of its values with a certain probability

assume equal weights for the elements in **Example 3.2**, what's the probability that no employee gets back his right helmet.

$P(M = 0) = 1/3$.

2. Discrete Probability Distribution

A discrete random variable assumes each of its values with a certain probability

assume equal weights for the elements in **Example 3.2**, what's the probability that no employee gets back his right helmet.

$P(M = 0) =1/3$.

The possible values m of M and their probabilities are

0	1	3
$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{6}$

Probability Mass Function

It is convenient to represent all the probabilities of a random variable X by a formula.

Probability Mass Function

It is convenient to represent all the probabilities of a random variable X by a formula.

write $p(x) = P(X = x)$

Probability Mass Function

It is convenient to represent all the probabilities of a random variable X by a formula.

write $p(x) = P(X = x)$

The set of ordered pairs $(x, p(x))$ is called the **probability function** or **probability distribution** of the discrete random variable X.

Probability Mass Function

Definition 3.4

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

1. $p(x) \geq 0$

2. $\sum_x p(x) = 1$

3. $P(X = x) = p(x)$

Probability Mass Function

Example 3.3 A shipment of 8 similar microcomputers to a retail outlet contains 3 that are defective. If a school makes a random purchase of 2 of these computers, find the probability distribution for the number of defectives.

Probability Mass Function

Example 3.3 A shipment of 8 similar microcomputers to a retail outlet contains 3 that are defective. If a school makes a random purchase of 2 of these computers, find the probability distribution for the number of defectives.

Solution

X: the possible numbers of defective computers

x can be any of the numbers 0, 1, and 2.

Probability Mass Function

Example 3.3 A shipment of 8 similar microcomputers to a retail outlet contains 3 that are defective. If a school makes a random purchase of 2 of these computers, find the probability distribution for the number of defectives.

Solution

X: the possible numbers of defective computers

x can be any of the numbers 0, 1, and 2.

$$p(0) = P(X = 0) = \frac{C_3^0 C_5^2}{C_8^2} = \frac{10}{28},$$

Probability Mass Function

$$p(1) = P(X = 1) = \frac{C_3^1 C_5^1}{C_8^2} = \frac{15}{28},$$

$$p(2) = P(X = 2) = \frac{C_3^2 C_5^0}{C_8^2} = \frac{3}{28}.$$

Cumulative Function

There are many problem where we may wish to compute the probability that the observed value of a random variable X will be less than or equal to some real number x.

Cumulative Function

There are many problem where we may wish to compute the probability that the observed value of a random variable X will be less than or equal to some real number x.

Writing $F(x) = P(X \leq x)$ for every real number x.

Cumulative Function

There are many problem where we may wish to compute the probability that the observed value of a random variable X will be less than or equal to some real number x.

Writing $F(x) = P(X \leq x)$ for every real number x.

Definition 3.5

The **cumulative distribution** $F(x)$ of a discrete random variable X with probability distribution $p(x)$ is

$$F(x) = P(X \leq x) = \sum_{t \leq x} p(t), \quad \text{for } -\infty < x < \infty.$$

Cumulative Function

For the random variable M, the number of correct matches in **Example 3.2**, we have

$$F(2) = P(M \leq 2) = p(0) + p(1) = \frac{1}{3} + \frac{1}{2} = \frac{5}{6}.$$

Cumulative Function

For the random variable M, the number of correct matches in **Example 3.2**, we have

$$F(2) = P(M \leq 2) = p(0) + p(1) = \frac{1}{3} + \frac{1}{2} = \frac{5}{6}.$$

The cumulative distribution of M is

$$F(m) = \begin{cases} 0 & \text{for } m < 0 \\ 1/3 & \text{for } 0 \leq m < 1 \\ 5/6 & \text{for } 1 \leq m < 3 \\ 1 & \text{for } m \geq 3. \end{cases}$$

Cumulative Function

Remark. the cumulative distribution is defined not only for the values assumed by given random variable but for all real numbers.

Cumulative Function

Example 3.5 The probability distribution of X is

0	1	2	3	4
$\frac{1}{16}$	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{1}{16}$

Find the cumulative distribution of the random variable X.

Cumulative Function

Example 3.5 The probability distribution of X is

0	1	2	3	4
$\frac{1}{16}$	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{1}{16}$

Find the cumulative distribution of the random variable X.

$$F(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1/16 & \text{for } 0 \leq x < 1 \\ 5/16 & \text{for } 1 \leq x < 2 \\ 11/16 & \text{for } 2 \leq x < 3 \\ 15/16 & \text{for } 3 \leq x < 4 \\ 1 & \text{for } x \geq 4. \end{cases}$$

Cumulative Function

Certain probability distribution are applicable to more than one physical situation.

Cumulative Function

Certain probability distribution are applicable to more than one physical situation.

The probability distribution of Example 3.5 can apply to different experimental situations.

Cumulative Function

Certain probability distribution are applicable to more than one physical situation.

The probability distribution of Example 3.5 can apply to different experimental situations.

Example 1: the distribution of Y, the number of heads when a coin is tossed 4 times

Cumulative Function

Certain probability distribution are applicable to more than one physical situation.

The probability distribution of Example 3.5 can apply to different experimental situations.

Example 1: the distribution of Y, the number of heads when a coin is tossed 4 times
Example 2: the distribution of W, the number of read cards that occur when 4 cards are drawn at random from a deck in succession with each card replaced and the deck shuffled before the next drawing.

graphs

It is helpful to look at a probability distribution in graphic form.

graphs

It is helpful to look at a probability distribution in graphic form.

bar chart;

histogram;

cumulative distribution.

graphs

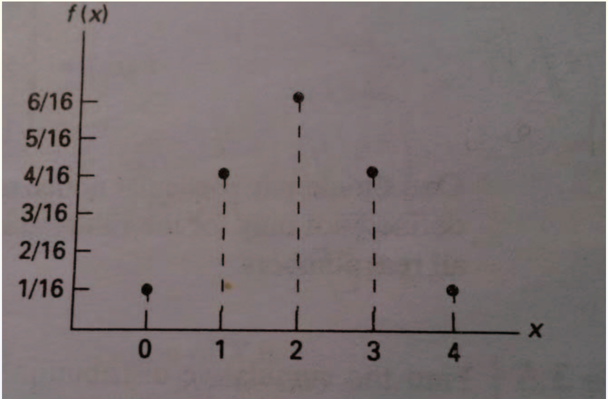

Figure 3.1 Bar chart.

graphs

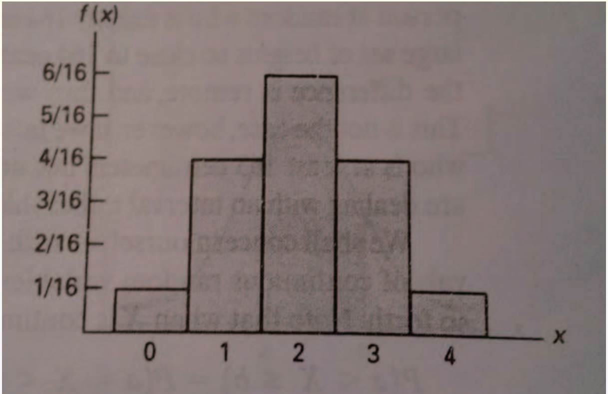

Figure 3.2 Probability histogram.

graphs

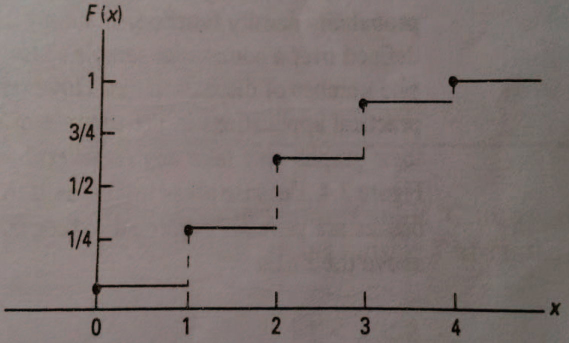

Figure 3.3 Discrete cumulative distribution.

3. Continuous Probability Distribution

A continuous random variable has **a probability of zero** of assuming exactly any of its values.

3. Continuous Probability Distribution

A continuous random variable has **a probability of zero** of assuming exactly any of its values. Consequently, its probability distribution cannot be given in tabular form.

3. Continuous Probability Distribution

A continuous random variable has **a probability of zero** of assuming exactly any of its values. Consequently, its probability distribution cannot be given in tabular form.

Example: the heights of all people over 21 years of age (random variable)

3. Continuous Probability Distribution

A continuous random variable has **a probability of zero** of assuming exactly any of its values. Consequently, its probability distribution cannot be given in tabular form.

Example: the heights of all people over 21 years of age (random variable)

Between 163.5 and 164.5 centimeters, or even 163.99 and 164.01 centimeters, there are an **infinite** number of heights, one of which is 164 centimeters.

Continuous Probability Distribution

The probability of selecting a person at random who is exactly 164 centimeters tall and not one of the infinitely large set of heights so close to 164 centimeters is remote.

Continuous Probability Distribution

The probability of selecting a person at random who is exactly 164 centimeters tall and not one of the infinitely large set of heights so close to 164 centimeters is remote.

We assign a probability of zero to a point, but this is not the case for an **interval**.

Continuous Probability Distribution

The probability of selecting a person at random who is exactly 164 centimeters tall and not one of the infinitely large set of heights so close to 164 centimeters is remote.

We assign a probability of zero to a point, but this is not the case for an **interval**. We will deal with an interval rather than a point value, such as $P(a < X < b), P(W \geq c)$.

Continuous Probability Distribution

The probability of selecting a person at random who is exactly 164 centimeters tall and not one of the infinitely large set of heights so close to 164 centimeters is remote.

We assign a probability of zero to a point, but this is not the case for an **interval**. We will deal with an interval rather than a point value, such as $P(a < X < b), P(W \geq c)$.

$$P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b)$$

where X is continuous.

Continuous Probability Distribution

The probability of selecting a person at random who is exactly 164 centimeters tall and not one of the infinitely large set of heights so close to 164 centimeters is remote.

We assign a probability of zero to a point, but this is not the case for an **interval**. We will deal with an interval rather than a point value, such as $P(a < X < b), P(W \geq c)$.

$$P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b)$$

where X is continuous. It does not matter whether we include an endpoint of the interval or not. **This is not true when X is discrete.**

Continuous Probability Distribution

Although the probability distribution of a continuous random variable cannot be presented in tabular form, it can be stated as a formula.

Continuous Probability Distribution

Although the probability distribution of a continuous random variable cannot be presented in tabular form, it can be stated as a formula.

refer to **histogram**

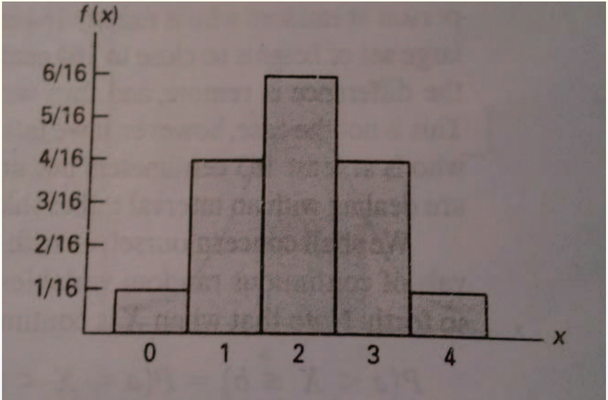

Figure 3.2 Probability histogram.

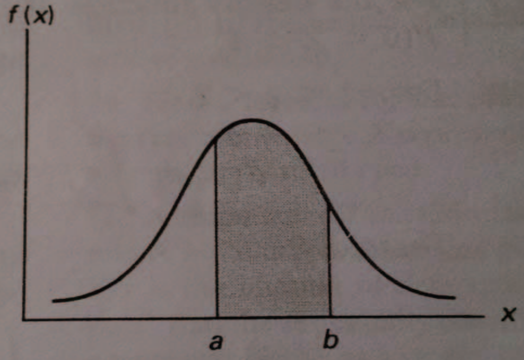

Figure 3.5 $P(a < X < b)$.

Continuous Probability Distribution

Definition 3.6 The function $f(x)$ is a **probability density function** for the continuous random variable X, defined over the set of real numbers R, if

1. $f(x) \geq 0$, for all $x \in R$.

2. $\int_{-\infty}^{\infty} f(x)dx = 1$.

3. $P(a < X < b) = \int_{a}^{b} f(x)dx$.

Continuous Probability Distribution

Example 3.6 Suppose that the error in the reaction temperature, in ^{o}C, for a controlled laboratory experiment is a continuous random variable X having the probability density function

$$f(x) = \begin{cases} x^2/3, & -1 < x < 2 \\ 0, & \text{elsewhere,} \end{cases}$$

(a) Verify condition 2 of Definition 3.6.

(b) Find $P(0 < X \leq 1)$.

Continuous Probability Distribution

Example 3.6 Suppose that the error in the reaction temperature, in ^{o}C, for a controlled laboratory experiment is a continuous random variable X having the probability density function

$$f(x) = \begin{cases} x^2/3, & -1 < x < 2 \\ 0, & \text{elsewhere}, \end{cases}$$

(a) Verify condition 2 of Definition 3.6.

(b) Find $P(0 < X \leq 1)$.

Solution: (a) . . . . . .

(b) $P(0 < X \leq 1) = 1/9.$

Continuous Probability Distribution

Definition 3.7 The **cumulative distribution** $F(x)$ of a continuous random variable X with density function $f(x)$ is

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt \qquad \text{for} - \infty < x < \infty.$$

Continuous Probability Distribution

Definition 3.7 The **cumulative distribution** $F(x)$ of a continuous random variable X with density function $f(x)$ is

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt \qquad \text{for } -\infty < x < \infty.$$

immediate consequences:

$P(a < X \leq b) = F(b) - F(a)$ and $f(x) = \frac{dF(x)}{dx}$.

Continuous Probability Distribution

Example 3.7 For the density function of Example 3.6 find $F(x)$, and use it to evaluate $P(0 < X \leq 1)$.

$$f(x) = \begin{cases} x^2/3, & -1 < x < 2 \\ 0, & \text{elsewhere}, \end{cases}$$

Continuous Probability Distribution

$$F(x) = \begin{cases} 0, & x \leq -1 \\ \dfrac{x^3+1}{9}, & -1 \leq x < 2 \\ 1, & x \geq 2. \end{cases}$$

$$P(0 < x \leq 1) = F(1) - F(0) = 1/9.$$