Chapter 3

Discrete Random Variables and Probability Distributions

Chapter 3: Discrete Random Variables and Probability Distributions

LEARNING OBJECTIVES

1. Discrete random variables
2. Probability Distributions and Probability mass function
3. Cumulative distribution function
4. Mean and Variance
5. Discrete Uniform Distribution
6. Binomial Distribution
7. Geometric and Negative Binomial Distributions
8. Hypergeometric Distribution
9. Poisson Distribution

Random variable

Definition

A **random variable** is a function that assigns a real number to each outcome in the sample space of a random experiment.

A random variable is denoted by an uppercase letter such as X. After an experiment is conducted, the measured value of the random variable is denoted by a lowercase letter such as x 70 milliamperes.

Random Variables

Example 3.1:

Two balls are drawn in succession without replacement from an 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

Random Variables

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 Variables

Example 3.3:

Consider the simple condition in which components are arriving from the production line and they are stipulated to be defective or not defective. Define the random variable X by

$$X = \begin{cases} 1, & \text{if the component is defective,} \\ 0, & \text{if the component is not defective.} \end{cases}$$

Clearly the assignment of 1 or 0 is arbitrary though quite convenient. This will become clear in later chapters. The random variable for which 0 and 1 are chosen to describe the two possible values is called a **Bernoulli random variable**.

Random Variables

Example 3.4:

Statisticians use **sampling plans** to either accept or reject batches or lots of material. Suppose one of these sampling plans involves sampling independently 10 items from a lot of 100 items in which 12 are defective.

Let X be the random variable defined as the number of items found defective in the sample of 10. In this case, the random variable takes on the values 0, 1, 2, . . . , 9, 10.

Random Variables

Example 3.5:

Suppose a sampling plan involves sampling items from a process until a defective is observed. The evaluation of the process will depend on how many consecutive items are observed. In that regard, let X be a random variable defined by **the number of items observed before a defective is found**. With N a nondefective and D a defective,

sample spaces are $S = \{D\}$ given $X = 1$,

$S = \{ND\}$ given $X = 2$,

$S = \{NND\}$ given $X = 3$.

Random Variables

Example 3.6:

Interest centers around the proportion of people who respond to a certain mail order solicitation. Let X be that proportion. X is a random variable that takes on all values x for which $0 \leq x \leq 1$.

Random Variables

Example 3.7:

Let X be the random variable defined by **the waiting time**, in hours, between successive speeders spotted by a radar unit. The random variable X takes on all values x for which $x \geq 0$.

Random Variables

Definition

A **discrete** random variable is a random variable with a finite (or countably infinite) range.

A **continuous** random variable is a random variable with an interval (either finite or infinite) of real numbers for its range.

Discrete Random Variables

Definition

A **discrete** random variable is a random variable with a finite or countably infinite range.

Example

1. Roll a die twice: Let X be the number of times 4 comes up then X could be 0, 1, or 2 times.

2. Toss a coin 5 times: Let X be the number of heads then $X = 0$, 1, 2, 3, 4, or 5.

3. X = The number of stocks in the Dow Jones Industrial Average that have share price increases on a given day then X is a discrete random variable because whose share price increases can be counted.

Discrete Random Variables

Determining a Discrete Random Variable

Let X be a discrete random variable with possible outcomes $x_1, x_2, \ldots, x_n$.

1. Find the probability of each possible outcome.

2. Check that each probability is between 0 and 1 and that the sum is 1.

3. Summarizing results in following table:

X	x_1	x_2	….	x_n
$P(x)$	p_1	p_2	….	p_n

Discrete Random Variables

Example

Let the random variable X denote the number of heads in three tosses of a fair coin. Determine the probability distribution of X.

Probability mass function (PMF)

Definition

For a discrete random variable X with possible values x_1, x_2, …, x_n , a **probability mass function** is a function such that

(1) $f(x_i) \geq 0$

(2) $\sum_{i=1}^{n} f(x_i) = 1$

(3) $f(x_i) = P(X = x_i)$

In above example, we have

$f(0) = 1/8$, $f(1) = 3/8$, $f(2) = 3/8$ and $f(3) = 1/8$.

Discrete Random Variables

Example

A shipment of 20 similar laptop computers 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.

Discrete Random Variables

Example

If a car agency sells 50% of its inventory of a certain foreign car equipped with side airbags, find a formula for the probability distribution of the number of cars with side airbags among the next 4 cars sold by the agency.

Exercises

Marketing estimates that a new instrument for the analysis of soil samples will be very successful, moderately successful, or unsuccessful with probabilities 0.3, 0.6, and 0.1, respectively. The yearly revenue associated with a very successful, moderately successful, or unsuccessful product is $10 million, $5 million, and $1 million, respectively. Let the random variable X denote the yearly revenue of the product. Determine the probability mass function of X.

Exercises

An assembly consists of three mechanical components. Suppose that the probabilities that the first, second, and third components meet specifications are 0.95, 0.98, and 0.99, respectively. Assume that the components are independent. Determine the probability mass function of the number of components in the assembly that meet specifications.

Exercises

The distribution of the time until a Web site changes is important to Web crawlers that search engines use to maintain current information about Web sites. The distribution of the time until change (in days) of a Web site is approximated in the following table.

Days until Changes	Probability
1.5	0.05
3.0	0.25
4.5	0.35
5.0	0.20
7.0	0.15

Calculate the probability mass function of the days until change.

Cumulative distribution function (CDF)

Definition

The **cumulative distribution function** of a discrete random variable *X*, denoted as *F*(*x*), is

$$F(x) = P(X \le x) = \sum_{x_i \le x} f(x_i)$$

For a discrete random variable *X*, F(x) satisfies the following properties.

(1) $0 \le F(x) \le 1$

(2) If $x \le y$, then $F(x) \le F(y)$

Cumulative distribution function

Example

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*?

The first we find the probability mass function of *X*.

X	0	1	2
***f*(*x*)**	0.886	0.111	0.003

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

Example 3.2: Cumulative distribution function

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.*** Find the cumulative distribution function of the random variable M.

Sample Space	m
SJB	3
SBJ	1
BJS	1
JSB	1
JBS	0
BSJ	0

Cumulative distribution function

Example

Determine the probability mass function of X from the following cumulative distribution function:

$$F(x) = \begin{cases} 0 & x < -2 \\ 0.2 & -2 \le x < 0 \\ 0.7 & 0 \le x < 2 \\ 1 & 2 \le x \end{cases}$$

Example — Cumulative distribution function

If a car agency sells 50% of its inventory of a certain foreign car equipped with side airbags, Find the cumulative distribution function of the number of cars with side airbags among the next 4 cars sold by the agency. Using $F(x)$, verify that $f(2) = 3/8$.

Mean and Variance

Definition

The **mean** or **expected value** of the discrete random variable X, denoted as μ or $E(X)$ is

$$\mu = E(X) = \sum_i x_i f(x_i)$$

The **variance (phương sai)** of X, denoted as σ^2 or $V(X)$ is

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

$$= \sum_i x_i^2 f(x_i) - \mu^2$$

The **standard deviation (độ lệch chuẩn)** of X is σ.

Mean and Variance

Example

The number of messages sent per hour over a computer network has the following distribution:

X	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 sent per hour.

Example — Mean and Variance

A lot containing 7 components is sampled by a quality inspector; the lot contains 4 good components and 3 defective components. A sample of 3 is taken by the inspector. Find the expected value and variance of the number of good components in this sample.

Example

Mean and Variance

A salesperson for a medical device company has two appointments on a given day. At the first appointment, he believes that he has a 70% chance to make the deal, from which he can earn $1000 commission if successful. On the other hand, he thinks he only has a 40% chance to make the deal at the second appointment, from which, if successful, he can make $1500. What is his expected commission based on his own probability belief? Assume that the appointment results are independent of each other.

Mean and Variance

Mean of a Function of a Discrete Random Variable

If X is a discrete random variable with probability mass function $f(x)$ then for any function $h(x)$

$$E[h(x)] = \sum_i h(x_i) f(x_i)$$

Corollary

$$E(aX + b) = aE(X) + b$$

$$V(aX + b) = a^2 V(X)$$

Example — Mean and Variance

Suppose that the number of cars X that pass through a car wash between 4:00 P.M. and 5:00 P.M. on any sunny Friday has the following probability distribution:

x	4	5	6	7	8	9
$P(X=x)$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{6}$

Let $g(X) = 2X-1$ represent the amount of money, in dollars, paid to the attendant by the manager. Find the attendant's expected earnings for this particular time period.

Example

Mean and Variance

Let X be a random variable with probability distribution as follows:

x	0	1	2	3
$f(x)$	$\frac{1}{3}$	$\frac{1}{2}$	0	$\frac{1}{6}$

Find the expected value of $Y = (X - 1)^2$.

Example

Mean and Variance

Calculate the variance of $g(X) = 2X + 3$, where X is a random variable with probability distribution

x	0	1	2	3
$f(x)$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{1}{8}$

3.5. Discrete Uniform Distribution

Discrete Uniform Distribution

Definition

A random variable X has a **discrete uniform distribution** if each of the n values in its range, say, $x_1, x_2, \ldots, x_n$ has equal probability. Then,

$$f(x_i) = 1/n$$

Mean and Variance

Suppose X is a discrete uniform random variable on the consecutive integers $a, a+1, \ldots, b$ for $a \leq b$. The mean and variance of X

$$\mu = E(X) = (a + b)/2 \qquad \sigma^2 = \frac{(b-a+1)^2 - 1}{12}$$

Discrete Uniform Distribution

EXAMPLE 3.10 Serial Number

The first digit of a part's serial number is equally likely to be any one of the digits 0 through 9. If one part is selected randomly from a large batch and X is the first digit of the serial number, X has a discrete uniform distribution with probability 0.1 for each value in $R = \{0, 1, 2, \ldots, 9\}$.

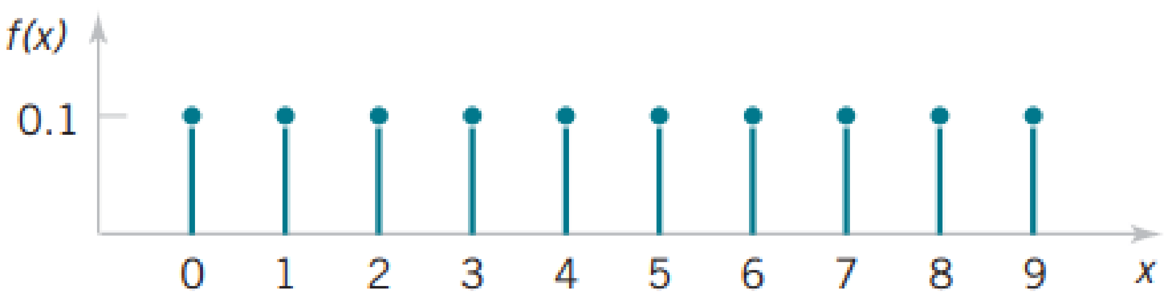


FIGURE 3.6

Probability mass function for a discrete uniform random variable.

Discrete Uniform Distribution

EXAMPLE 3.11 Number of Voice Lines

Let the random variable X denote the number of 48 voice lines that are used at a particular time. Assume that X is a discrete uniform random variable with a range of 0 to 48.

Discrete Uniform Distribution

EXAMPLE.

Let the random variable Y denote the proportion of the 48 voice lines used at a particular time, and X denote the number of lines used at a particular time.

Discrete Uniform Distribution

EXAMPLE.

Suppose that the discrete uniform random variable $Y = 5X$ where X has range 1, 2, … , 6. Find E(Y) and V(Y).

Binomial Distribution

Definition

A random experiment consists of n Bernoulli trials such that

(1) The trials are independent

(2) Each trial results in only two possible outcomes, labeled as "success" and "failure"

(3) The probability of a success in each trial, denoted as p, remains constant

The random variable X that equals the number of trials that result in a success has a **binomial random variable** with parameters $0<p<1$ and n=1,2,…The probability mass function of X is

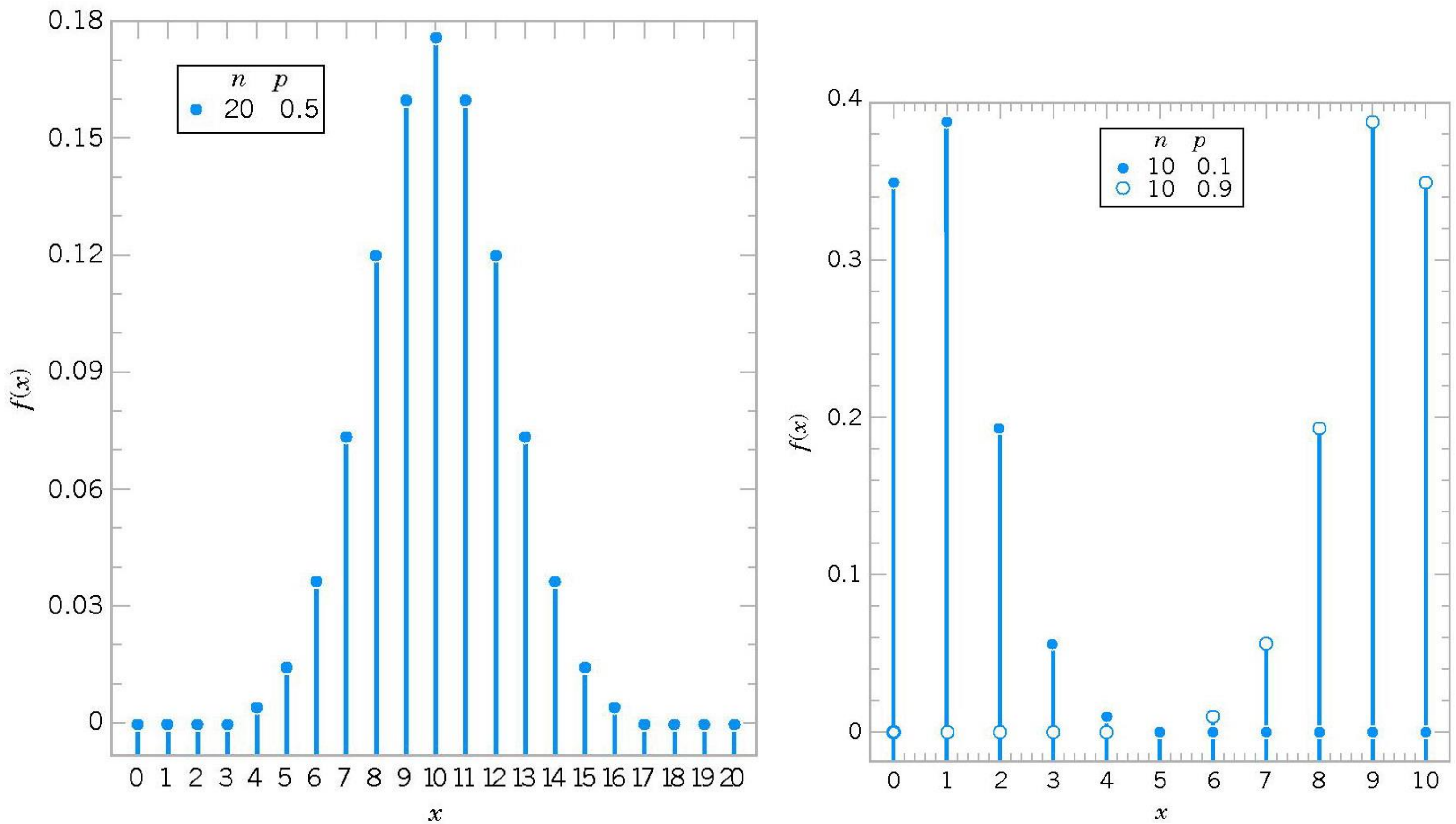
$$f(x) = C_n^x p^x (1-p)^{n-x} \qquad x = 0, 1, 2, \ldots, n.$$

Binomial Dis.

Example

The probability that a certain kind of component will survive a shock test is 3/4. Find the probability that exactly 2 of the next 4 components tested survive.

Binomial Dis.



Binomial distributions for selected values of *n* and *p*.

Binomial Dis.

Mean and Variance

$\mu = E(X) = np$ $\sigma^2 = V(X) = np(1-p)$

Example

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.

(a) Find the probability that in the next 18 samples, exactly 2 contain the pollutant.

(b) Determine the probability that at least four samples contain the pollutant.

(c) Determine the probability that $3 \leq X < 7$.

Binomial Dis.

Example

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,

(c) exactly 5 survive?

Binomial Dis.

Example

A large chain retailer purchases a certain kind of electronic device from a manufacturer. The manufacturer indicates that the defective rate of the device is 3%.

(a) The inspector randomly picks 20 items from a shipment. What is the probability that there will be at least one defective item among these 20?

(b) Suppose that the retailer receives 10 shipments in a month and the inspector randomly tests 20 devices per shipment. What is the probability that there will be exactly 3 shipments each containing at least one defective device among the 20 that are selected and tested from the shipment?

Binomial Dis.

EXAMPLE 3-16 Digital Channel

The chance that a bit transmitted through a digital transmission channel is received in error is 0.1. Also, assume that the transmission trials are independent. Let *X* the number of bits in error in the next four bits transmitted.

a) Determine P(X=2).
b) Find E(X), V(X).

Geometric and Negative Binomial Distributions

Geometric and Negative Binomial Distributions

Geometric Distribution (X~Geo(p))

In a series of Bernoulli trials (independent trials with constant probability p of a success), the random variable X that equals the number of trials until the first success is a **geometric random variable** with parameter $0 < p < 1$ and

$$f(x) = (1 - p)^{x-1} p \qquad x = 1, 2, \ldots$$

Geometric and Negative Binomial Distributions

Example

At a "busy time," a telephone exchange is very near capacity, so callers have difficulty placing their calls. It may be of interest to know the number of attempts necessary in order to make a connection. Suppose that we let $p = 0.05$ be the probability of a connection during a busy time. We are interested in knowing the probability that 5 attempts are necessary for a successful call.

Geometric Distribution

EXAMPLE 3.18 Wafer Contamination

The probability that a wafer contains a large particle of contamination is 0.01. If it is assumed that the wafers are independent, what is the probability that exactly 125 wafers need to be analyzed before a large particle is detected?

Comparison/Contrast with binomial distribution:

Similar They both have independent Bernoulli trials.

Different

For a r.v. with a binomial distribution, we *know* how many trials we will have, there will be n trials, $X \sim B(n; p)$ and $X = \{0; 1; 2; \ldots; n\}$.

For a r.v. with a geometric distribution, we *do not know* how many trials we will have, $X \sim Geo(p)$ and $X =\{1; 2; 3; \ldots; n\}$. We stop only when we get a success.

Geometric Distribution

Mean and Variance

If X is a geometric random variable with parameter p,

$$\mu = E(X) = \frac{1}{p}$$

$$\sigma^2 = V(X) = \frac{1-p}{p^2}$$

EXAMPLE

For a certain manufacturing process, it is known that, on the average, 1 in every 100 items is defective.

a) What is the probability that the fifth item inspected is the first defective item found?

b) Find $E(X)$ and $V(X)$ and the standard deviation of X.

Example (Weld strength)

A test of weld strength involves loading welded joints until a fracture occurs. For a certain type of weld, 80% of the fractures occur in the weld itself, while the other 20% occur in the beam. A number of welds are tested and the tests are independent. Let X be the number of test at which the first **beam** fracture is observed.

1. Find $P(X \geq 3)$ (i.e. Find the probability that the first beam fracture happens on the third trial or later.)
2. Find $E(X)$ and $V(X)$ and the standard deviation of X

Negative Binomial Distribution

In a series of Bernoulli trials (independent trials with constant probability p of a success), the random variable X that equals the number of trials until r successes occur is a **negative binomial random variable** with parameters $0 < p < 1$ and $r = 1, 2, 3, \ldots$, and

$$f(x, r, p) = \binom{x-1}{r-1}(1-p)^{x-r} p^r, x = r, r+1, r+2, \ldots$$

Negative Binomial Distribution

Mean and Variance

If X is a negative binomial random variable with parameters p and r,

$$\mu = E(X) = \frac{r}{p}$$

$$\sigma^2 = V(X) = \frac{r(1-p)}{p^2}$$

Negative Binomial Distribution

Example (Weld strength, cont.)

Find the probability that the 3rd beam fracture (*success*) occurs on the 5th trial.

Negative Binomial Distribution

Example: 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?

Negative Binomial Distribution

Example (Weld strength, cont.)

Let X represent the number of trials until 3 beam fractures occur.

Then, X follows a negative binomial distribution with parameters $p = 0.2$ and $r = 3$.

Find the expected value of X and the variance of X.

Negative Binomial Distribution

Example (Participants in study)

A pediatrician wishes to recruit 5 couples, each of whom is expecting their first child, to participate in a new natural childbirth regimen.
Let $p = P$(a randomly selected couple agrees to participate).

If $p = 0.15$, what is the probability that 15 couples must be asked before 5 are found who agree to participate? That is, with $S =$ agrees to participate, what is the probability that 10 Fs occur before the fifth S?

Hypergeometric Distribution

Hypergeometric Distribution

A set of N objects contains

K objects classified as successes

N - K objects classified as failures

A sample of size n objects is selected randomly (without replacement) from the N objects where $K \le N$ and $n \le N$.

The random variable X that equals the number of successes in the sample is a **hypergeometric random variable** and

$$f(x) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}, \qquad x = \max\{0, n+K-N\}\; to \min\{K, n\}$$

Hypergeometric Distribution

EXAMPLE. If a lot has 25 items and 6 of them are defectives, what is the probability that a sample of size 10 has none that are defective?

Hypergeometric Distribution

EXAMPLE Lots of 40 components each are deemed unacceptable if they contain 3 or more defectives. The procedure for sampling a lot is to select 5 components at random and to reject the lot if a defective is found. What is the probability that exactly 1 defective is found in the sample if there are 3 defectives in the entire lot?

Hypergeometric Distribution

Mean and Variance

If X is a hypergeometric random variable with parameters N, K, and n, then

$$\mu = E(X) = np$$

$$\sigma^2 = V(X) = npq\left(\frac{N-n}{N-1}\right)$$

where $p = K/N$.

Hypergeometric Distribution

EXAMPLE 3.24 Parts from Suppliers

A batch of parts contains 100 from a local supplier of circuit boards and 200 from a supplier in the next state. If four parts are selected randomly and without replacement. Let X equal the number of parts in the sample from the local supplier.

a) What is the probability they are all from the local supplier?

b) What is the probability that two or more parts in the sample are from the local supplier?

c) What is the probability that at least one part in the sample is from the local supplier?

d) E(X), V(X)?

Poisson Dis.

Definition

Given an interval of real numbers, assume events occur at random throughout the interval. If the interval can be partitioned into subintervals of small enough length such that

(1) the probability of more than one event in a subinterval is zero,

(2) the probability of one event in a subinterval is the same for all subintervals and proportional to the length of the subinterval, and

(3) the event in each subinterval is independent of other subintervals, the random experiment is called a **Poisson process.**

The random variable X that equals the number of events in the interval is a **Poisson random variable** with parameter $\lambda > 0$, and the probability mass function of X is

$$f(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$$

Poisson Distribution

Example

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?

Using the Poisson distribution with $x = 6$ and $\lambda = 4$. We have

$$P(X = 6) = \frac{e^{-4} 4^6}{6!} = 0.1042$$

Poisson Distribution

Example

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?

Poisson Distribution

Mean and Variance

If X is a Poisson random variable with parameter λ, then

$$\mu = E(X) = \lambda \qquad \sigma^2 = V(X) = \lambda$$

Example

For the case of the thin copper wire, suppose that the number of flaws follows a Poisson distribution with a mean of 2.3 flaws per millimeter.

a) Determine the probability of exactly 2 flaws in 1 millimeter of wire.

b) Determine the probability of 10 flaws in 5 mm of wire.

c) Determine the probability of at least 1 flaw in 2 millimeters of wire.