Chapter 3: Discrete Random Variables and Probability Distributions

LEARNING OBJECTIVES

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

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.

Determining a Discrete Random Variable

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

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

Example

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

The sample space:

$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}.$

The events: $[X = 0] = \{TTT\}$ $\quad [X = 1] = \{HTT, THT, TTH\}$

$[X = 2] = \{HHT, HTH, THH\}$ $\quad [X = 3] = \{HHH\}$

X	0	1	2	3
P(x)	1/8	3/8	3/8	1/8

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 $f(3) = 1/8$.

Cumulative distribution function

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.

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

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

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

Therefore,

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

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

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

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

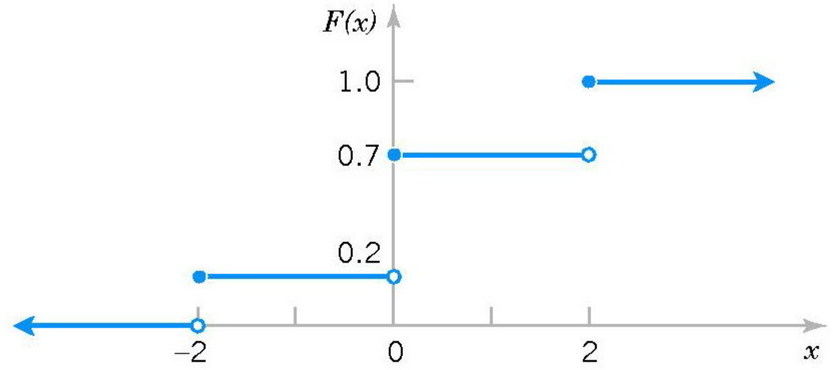
$$\Longrightarrow F(x)=\begin{cases}0 & \text{if } x<0\\ 0.886 & \text{if } 0\le x<1\\ 0.997 & \text{if } 1\le x<2\\ 1 & \text{if } x\ge 2\end{cases}$$

Cumulative distribution function

Example

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

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



$f(-2) = 0.2 - 0 = 0.2$

$f(0) = 0.7 - 0.2 = 0.5$

$f(2) = 1.0 - 0.7 = 0.3$

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** 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** of X is σ.

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.

$$E(X) = 10(0.08) + 11(0.15) + \cdots + 15(0.07) = 12.5$$

$$V(X) = 10^2(0.08) + 11^2(0.15) + \cdots + 15^2(0.07) - 12.5^2 = 1.85$$

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

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

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

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

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,

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

Ex:

Let X be the number on a die roll, between 1 and 6. Find E(X) and V(X)?

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 $a \le b$. The mean and variance of X

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

Binomial Distribution

The components in the system are regarded as independent trials. We have a system with 100 independent components. The probability that a component will fail during a prescribed period of time is 0.005. What is the probability that the system is operating at the end of the period, which requires that no component fail?

Binomial Distribution

Definition

A random experiment consists of *n* Bernoulli trials such that

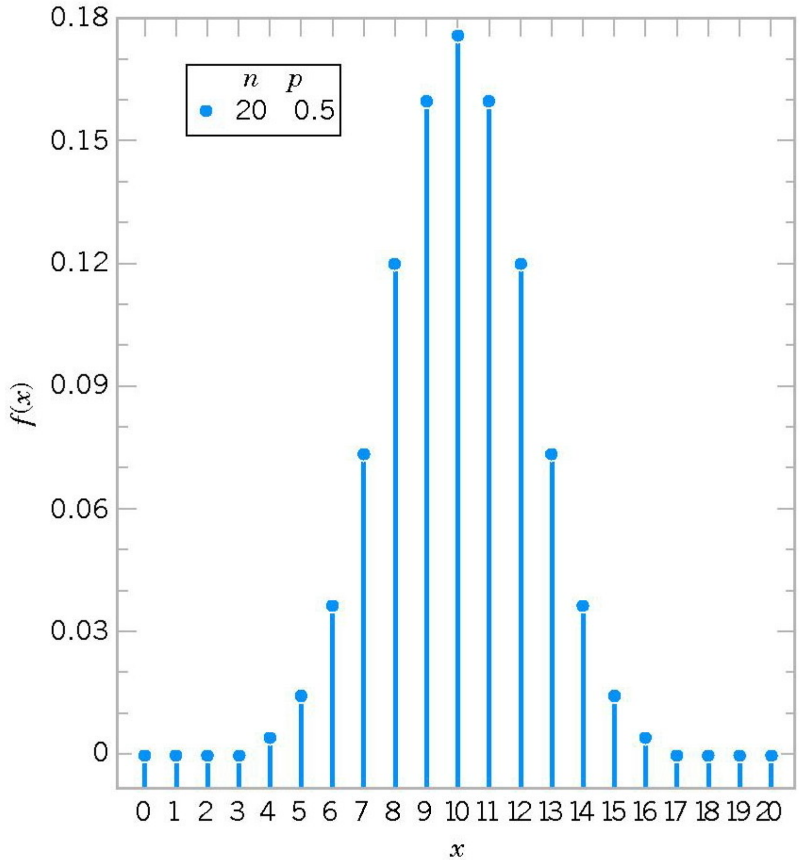
(1) The trials are independent

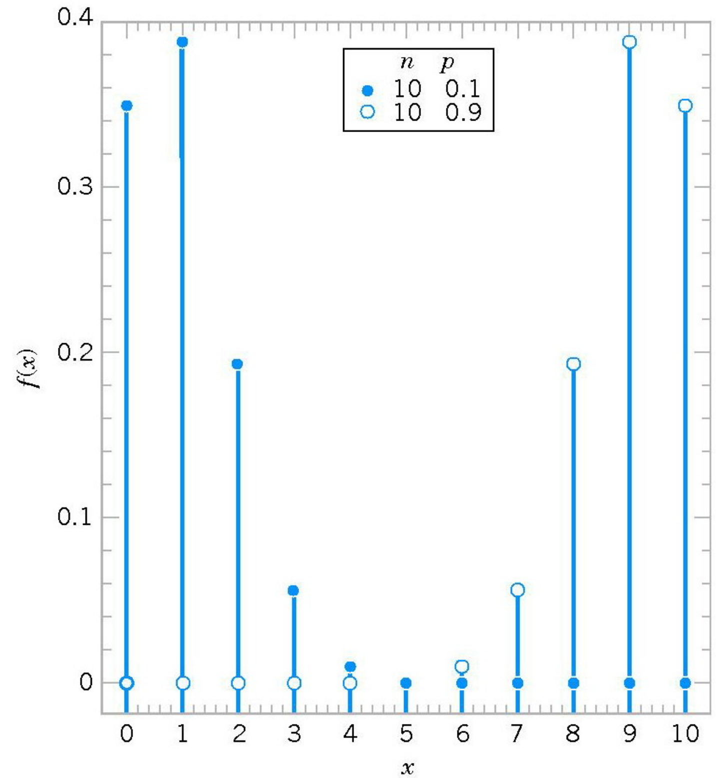
(2) Each trial results in only two possible outcomes, labeled as "success" and "failure"

(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

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





Binomial distributions for selected values of n and p.

D.R. variables
PMF & CDF
Mean & Variance
Uniform Dis.
Binomial Dis.
Poisson Dis.
Summary

Mean and Variance

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

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 $3 \leq X < 7$.

Solution. 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 $n = 18$.

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

(b) $$P(x\geq 4)=\sum_{x\geq 4}\binom{18}{x}0.1^x(1-0.1)^{18-x}=0.098$$

(c) $$P(3\leq x<7)=\sum_{x=3}^{6}\binom{18}{x}0.1^x(1-0.1)^{18-x}=0.265$$

Exercise

An office has three telephone lines. Assume that 3 lines are independent. At any given time, the probability that at least one line is in use is 0.8. Find the probability that, at any given time, all three are in use.

Poisson Distribution

Definition

In general, consider subintervals of small length Δt and assume that as Δt tends to zero,

1. The probability of more than one event in a subinterval tends to zero.
2. The probability of one event in a subinterval tends to $\lambda\Delta t$.
3. The event in each subinterval is independent of other subintervals.

A random experiment with these properties is called a **Poisson process**.

These assumptions imply that the subintervals can be thought of as approximate independent Bernoulli trials with the number of trials equal to $n = T/\Delta t$ and success probability $p = \lambda\Delta t = \lambda T/n$. This leads to the following result.

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

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

Mean and Variance

If X is a Poisson random variable over an interval of length T with parameter λ , then

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

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 at least 1 flaw in 2 millimeters of wire.

a) Let X denote the number of flaws in 1 millimeter of wire. Then, $\lambda T = 2.3$ flaws and

$$P(X=2)=\frac{e^{-2.3}2.3^2}{2!}=0.265$$

b) Let X denote the number of flaws in 2 millimeters of wire. Then, X has a Poisson distribution with

$$E(X) = 2 \text{ mm} \times 2.3 \text{ flaws/mm} = 4.6 \text{ flaws}$$

Therefore,

$$P(X \geq 1) = 1 - P(X=0) = 1 - e^{-4.6} = 0.9899$$

Exercise

8) Messages arrive at a computer at an average rate of 15 messages/second. The number of messages that arrive in 1 second is known to be a Poisson random variable.

a. Find the probability that no messages arrive in 1 second. **(answer: 3.06*10^-7);**

b. Find the probability that more than 10 messages arrive in a 4-second period.

7) Compute the probability of obtaining three defectives in a sample of size 10 taken without replacement from a box of twenty components containing four defectives. **Answer: 0.2477**

Definition

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

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

Mean and Variance

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

$$\mu = E(X) = np \qquad \text{and} \qquad \sigma^2 = V(X) = np(1-p)\left(\frac{N-n}{N-1}\right)$$

where $p = K/N$.

11) An urn contains 30 red and 20 green balls. A sample of 5 balls is selected at random, without replacement. Find the mean and standard deviation of the number of red balls in the sample. **Answer: 3; 1.05**

Geometric Distributions

12) From the past experience it is known that 3% of accounts in a large accounting population are in error.

a) What is the probability that 5 accounts are audited before an account in error is found? **Answers: 0.027**

b) What is the probability that the first account in error occurs in the first five accounts audited? **0.141**

Geometric Distributions

D.R. variables
PMF & CDF
Mean & Variance
Uniform Dis.
Binomial Dis.
Poisson Dis.
Summary

Definition

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

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

Mean and Variance

If X is a geometric random variable with parameter p,

$$\mu = E(X) = 1/p \quad \text{and} \quad \sigma^2 = V(X) = (1-p)/p^2$$

D.R. variables
PMF & CDF
Mean & Variance
Uniform Dis.
Binomial Dis.
Poisson Dis.
Summary

13)

a) Find the probability that a man flipping a coin gets the fourth head on the ninth flip.

b) How many times on average does that man have to flip to get four heads?

c) Find standard deviation of the number of flips until that man gets four heads.

Answer: 0.109; 8; 2.83

Negative Binomial Distributions

Definition

In a series of Bernoulli trials (independent trials with constant probability p of a suc- cess), 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

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

Negative Binomial Distributions

Mean and Variance

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

$$\mu = E(X) = r/p \qquad \text{and} \qquad \sigma^2 = V(X) = r(1-p)/p^2$$

Summary

We have studied:

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