IST1990 Probability And Statistics

- Lecture 7

- Discrete Distributions:
- Bernoulli,
- Binomial,
- Geometric.

Some Discrete Probability Distributions

In many cases, the random observations generated by different statistical experiments have the same general type of behavior. Therefore, discrete random variables associated with these experiments can be described by essentially the same probability distribution and can be represented by a single formula.

In this lecture, we present these commonly used discrete distributions with various examples.

Discrete Probability Distribution in Engineering Applications

The four most used discrete probability distributions in business operations are:

- Binomial Distribution
- Geometric Distribution
- Hypergeometric Distribution
- Poisson Distribution

This lecture will focus on the applications of these four probability distributions in engineering with particular interest to quality engineering.

Bernoulli Random Variables

Definition: Bernoulli Random Variable

Let X be a random variable taking only two possible values $\{0, 1\}$. Let $p = P(X = 1)$. Then X is said to be a **Bernoulli** random variable with parameter p, or equivalently $X \sim \text{Ber}(p)$.

Examples:

- Flipping a "fair" coin: $X \sim \text{Ber}(1/2)$

Let $X \sim \text{Ber}(p)$. Then

$$\mathbb{E}[X] = 0 \times P(X = 0) + 1 \times P(X = 1) = p$$

$$\mathbb{E}[X^2] = 0^2 \times P(X = 0) + 1^2 \times P(X = 1) = p$$

$$\text{V}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = p - p^2 = p(1 - p)$$

Bernoulli Trials

Independent Bernoulli random variables are important blocks to build more complicated random variables...

Consider the following examples:

- Flip a coin 10 times. Let X be the number of heads obtained.
- The packets received during a UDP connection arrive corrupted 5% of the times. Suppose you send 100 packets and let X be the number of packets received in error.
- In the next 30 births at an hospital, let X be the number of female babies.

These are all examples of binomial random variables (provided you assume each of the Bernoulli trials are independent).

Repeated Bernoulli Trials

Example: Suppose we send 3 packets through a communication channel. Assume

- Each packet is received with probability 0.9.
- The events {packet 1 is received}, {packet 2 is received}, and {packet 3 is received} are all **independent**.
- Let X be the total number of received packets.

$$X_i \sim \text{Ber}(p) \rightarrow X_i = 1 \text{ iif } \{\text{packet i was received}\}, i = 1, 2, 3$$

$$X = X_1 + X_2 + X_3$$

x_1	x_2	x_3	$P(X_1 = x_1, X_2 = x_2, X_3 = x_3)$	$x = x_1 + x_2 + x_3$
0	0	0	0.001	0
0	0	1	0.009	1
0	1	0	0.009	1
0	1	1	0.081	2
1	0	0	0.009	1
1	0	1	0.081	2
1	1	0	0.081	2
1	1	1	0.729	3

Binomial Distribution

x_1	x_2	x_3	$P(X_1 = x_1, X_2 = x_2, X_3 = x_3)$	$x = x_1 + x_2 + x_3$
0	0	0	0.001	0
0	0	1	0.009	1
0	1	0	0.009	1
0	1	1	0.081	2
1	0	0	0.009	1
1	0	1	0.081	2
1	1	0	0.081	2
1	1	1	0.729	3

Therefore

$$P(X = x) = \begin{cases} 0.001 & \text{if } x = 0 \\ 0.027 & \text{if } x = 1 \\ 0.243 & \text{if } x = 2 \\ 0.729 & \text{if } x = 3 \end{cases}$$

This is what is called a Binomial Random Variable!!!

Binomial Distribution

Consider the following random experiments and random variables:

1. Flip a coin 10 times. Let X number of heads obtained.

2. A worn machine tool produces 1% defective parts. Let X number of defective parts in the next 25 parts produced.

3. Each sample of air has a 10% chance of containing a particular rare molecule. Let X the number of air samples that contain the rare molecule in the next 18 samples analyzed.

4. Of all bits transmitted through a digital transmission channel, 10% are received in error. Let X the number of bits in error in the next five bits transmitted.

5. A multiple choice test contains 10 questions, each with four choices, and you guess at each question. Let X the number of questions answered correctly.

6. In the next 20 births at a hospital, let X the number of female births.

Binomial Distribution

Definition: Binomial Random Variable

Consider a random experiment consisting of $n \in \mathbb{N}$ **independent** Bernoulli trials, where each trial can result in either success or failure.

Assume the probability of success for each trial is equal to $0 \leq p \leq 1$, and remains the same throughout the experiment.

The random variable X that equals the number of trials that result in success is a **Binomial Random Variable** with parameters n and p. The probability mass function of X is given by

$$P(X = k) = f(k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, 1, \ldots, n\ .$$

We have seen $\binom{n}{k} = C_k^n$ before, the number of possible combinations of k out of n elements.

Binomial Distribution

The name of this distribution comes from the similarity with the *binomial expansion*: For any $a, b \in \mathbb{R}$

$$(a+b)^n = \sum_{i=0}^{n} \binom{n}{i} a^i b^{n-i} \ .$$

This immediately shows that the distribution is indeed valid, as

$$\sum_{k=0}^{n} f(k) = 1 \ .$$

The computation of the mean and variance of X can also be done directly from the p.m.f., but it is rather tedious and complicated... But there is a much easier way...

Binomial Distribution

Note that

$$X = X_1 + X_2 + \cdots + X_n \ ,$$

where $X_i \sim \text{Ber}(p)$ are independent Bernoulli trials.

As stated earlier, for independent random variables $X_1, \ldots, X_n$ we have

$$\mathbb{E}[X_1 + X_2 + \cdots + X_n] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \cdots + \mathbb{E}[X_n] \ ,$$

and

$$\mathbb{V}(X_1 + X_2 + \cdots + X_n) = \mathbb{V}(X_1) + \mathbb{V}(X_2) + \cdots + \mathbb{V}(X_n) \ .$$

Let $X \sim \text{Bin}(n, p)$ be a binomial random variable with parameters n and p. Then

$$\mathbb{E}[X] = np, \qquad \text{and} \qquad \mathbb{V}(X) = np(1-p)$$

Binomial Distribution 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. Find the probability that in the next 18 samples, exactly 2 contain the pollutant.

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,

$$P(X = 2) = \binom{18}{2}(0.1)^2(0.9)^{16}$$

Now $\binom{18}{2} = 18!/[2!\ 16!] = 18(17)/2 = 153$. Therefore,

$$P(X = 2) = 153(0.1)^2(0.9)^{16} = 0.284$$

Determine the probability that at least four samples contain the pollutant. The requested probability is

$$P(X \geq 4) = \sum_{x=4}^{18} \binom{18}{x} (0.1)^x (0.9)^{18-x}$$

However, it is easier to use the complementary event,

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

Determine the probability that $3 \leq X < 7$. Now

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

Binomial Distribution Example

Samples from a certain water supply have a 10% chance of containing an organic pollutant. Suppose you collect 20 samples over time, and these are taking in such a way that you can assume these are independent. Let X be the number of contaminated samples.

$$X \sim \text{Bin}(20, 0.1)$$

$$P(X = 2) = \binom{20}{2} 0.1^2 (1 - 0.1)^{20-2} \approx 0.285$$

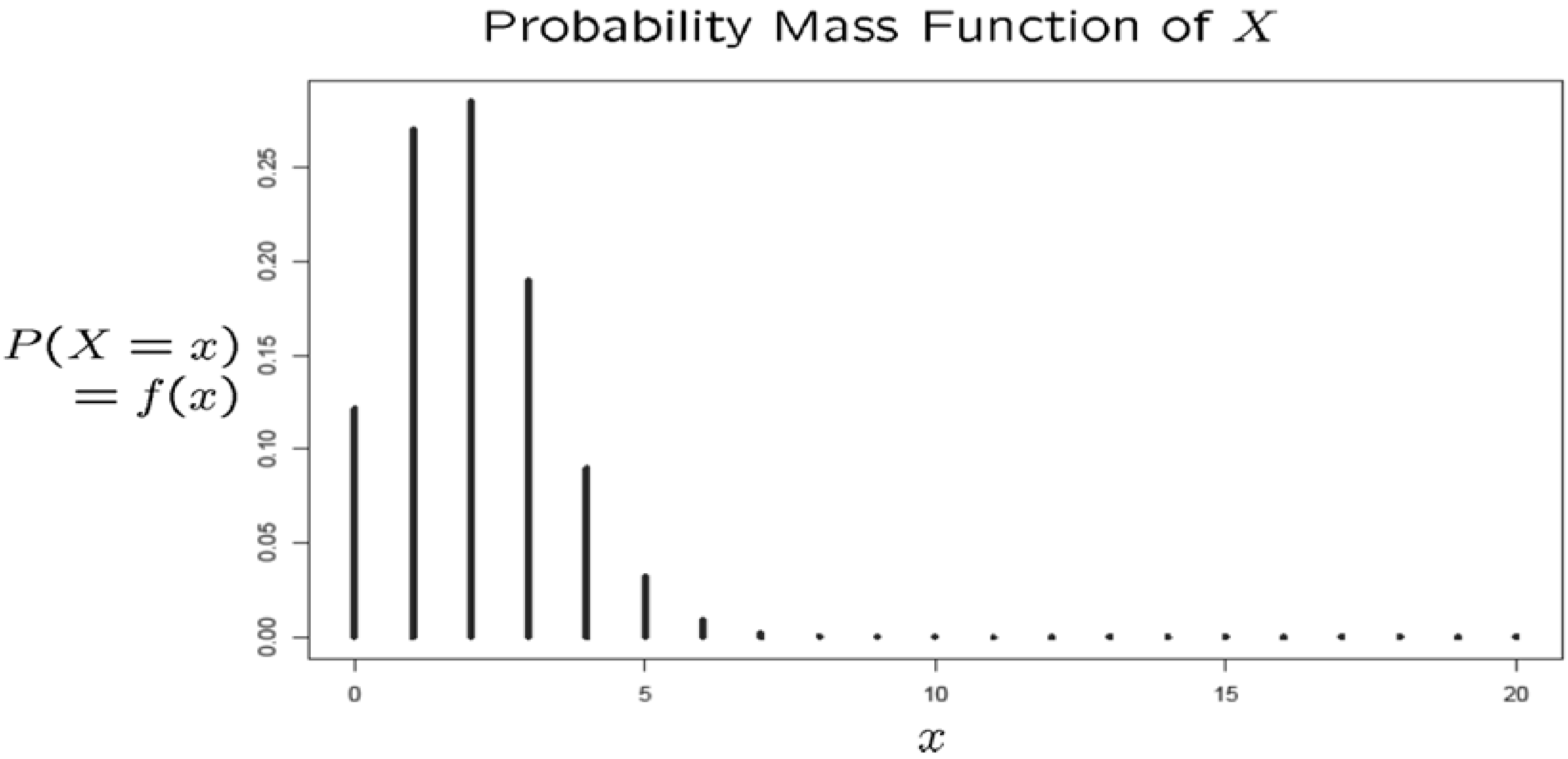
$$\begin{aligned} P(X \geq 2) = 1 - P(X < 2) &= 1 - P(X = 0) - P(X = 1) \\ &= 1 - 0.1^0\ 0.9^{20} - 200.10.9^{20-1} = 0.608 \end{aligned}$$

$$P(X \leq 7) = (\text{use table or computer}) = 0.9996$$

$$\begin{aligned} P(3 \leq X \leq 6) = P(X \leq 6) - P(X < 3) &= P(X \leq 6) - P(X \leq 2) \\ &= 0.3207 \end{aligned}$$

$$P(3 < X < 7) = 0.9976 - 0.8670 = 0.1306$$

$$\mathbb{E}[X] = 20 \times 0.1 = 2 \qquad \text{V}(X) = 20 \times 0.1(1 - 0.1) = 1.8$$



Probability Mass Function of X
$P(X = x)$
$= f(x)$
0.00
0.05
0.10
0.15
0.20
0.25
0
5
10
15
20
x

Homework:

Because not all airline passengers show up for their reserved seat, an airline sells 125 tickets for a flight that holds only 120 passengers. The probability that a passenger does not show up is 0.10, and the passengers behave independently.

(a) What is the probability that every passenger who shows up can take the flight?

(b) What is the probability that the flight departs with empty seats?

Let X denote the passengers with tickets that do not show up for the flight. Then, X is binomial with n = 125 and p = 0.1.

$$a)P(X \geq 5) = 1 - P(X \leq 4)$$

$$= 1 - \left[\binom{125}{0}0.1^0(0.9)^{125} + \binom{125}{1}0.1^1(0.9)^{124} + \binom{125}{2}0.1^2(0.9)^{123} + \binom{125}{3}0.1^3(0.9)^{122} + \binom{125}{4}0.1^4(0.9)^{121}\right] = 0.9961$$

$$b)P(X > 5) = 1 - P(X \leq 5) = 0.9886$$

Example. A coin is tossed 6 times.

a) What is the probability that 2 heads will occur?

n = 6, p = 0.5,

$$P(X = 2) = b(2,6,0.5) = \binom{6}{2}(0.5)^2(1-0.5)^{6-2} = 0.2344$$

b) What is the probability that at least 1 head occurs?

$$P(X \geq 1) = P(X=1) + P(X=2) + P(X=3) + P(X=4) + P(X=5) + P(X=6)$$

$$= \sum_{i=1}^{6} \binom{6}{i} 0.5^{i}(1-0.5)^{6-i} \qquad \textbf{or}$$

$$P(X \geq 1) = 1 - P(X<1) = 1 - P(X=0) = 1 - \left\{ \binom{6}{0} 0.5^{0}(1-0.5)^{6-0} \right\} = 0.984$$

c) Find the mean and variance of the number of heads.

$$E(X) = 6(0.5) = 3,$$

$$Var(X) = 6(0.5)(1-0.5) = 1.5.$$

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 2 survive?

b) between 3 and 7 survive?

c) exactly 5 survive?

n = 15, p = 0.4

(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) $P(3 < X < 7)$

$$= B(6;15,0.4) - B(3;15,0.4) = 0.6098 - 0.0905 = 0.5193$$

(c) $P(X = 5) = b(5;15,0.4) = \binom{15}{5}(0.4)^5(0.6)^{15-5} = 0.186$

Geometric Distribution

Let X be the number of times you need to send a packet over a network to ensure it was received.

$$\text{Sample Space}=\{1,2,3,...\}$$

Assume the errors that occur at each transmission are independent, and that the packet is transmitted successfully with probability $p = 0.9$.

$$\begin{aligned}
P(X=1) &= p = 0.9 \\
P(X=2) &= P(\text{1st trans. fails and 2nd succeeds}) \\
&= P(\text{2nd succeeds}|\text{1st fails})P(\text{1st fails}) \\
&= (1-p)\times p = 0.09 \\
P(X=k) &= (1-p)^{k-1}\times p = 0.9\times 0.1^{k-1}, \quad k=1,2,...
\end{aligned}$$

This is what is called a geometric distribution...

Geometric Distribution

Definition: Geometric Random Variable

Consider a series of independent Bernoulli trials with constant probability of success p. Let the random variable X denote the number of trials until the first success. Then X is called a **geometric random variable** with parameter p, and it has probability mass function given by

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

Let $X \sim \text{Geom}(p)$ be a geometrically distributed random variable with parameter p. Then

$$\mathbb{E}[X] = \frac{1}{p}, \qquad \text{and} \qquad \mathbb{V}(X) = \frac{1-p}{p^2}$$

The Geometric Distribution :

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

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

While in the binomial we have fixed number of trials (n) and the random variable is number of successes, in the geometric distribution we wait for a single success (fixed), but the number of trials is the random 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 is q = 1-p), until a head appears for the first time.

Geometric Distribution Example

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?

Let *X* denote the number of samples analyzed until a large particle is detected. Then *X* is a geometric random variable with *p* 0.01. The requested probability is

$$P(X = 125) = (0.99)^{124} 0.01 = 0.0029$$

Geometric Distribution Example

Let *X* be the number of times you need to send a packet over a network to ensure it was received.

$$\text{Sample Space} = \{1, 2, 3, \ldots\}$$

Assume the errors that occur at each transmission are independent, and that the packet is transmitted successfully with probability $p = 0.9$.

$$\mathbb{E}[X] = \frac{1}{0.9} \approx 1.111$$

$$V(X) = \frac{0.1}{0.9^2} \approx 0.1234$$

Geometric Probability Distribution Engineering Example

The probability for finding an error by an auditor in a production line is 0.01.

a) What is the probability that the first error is found at the 70th part audited?

b) What is the probability that more than 50 parts must be audited before the first error is found?

Solution:

(a)

$$P(k, p) = pq^{k-1}$$

$$P(70, 0.01) = 0.01 \times 0.99^{70-1} = 0.004998$$

The probability that the first error is found at the 70th part audited will be 0.004998.

(b)

$$P(k > n) = q^n$$

$$P(k > 50) = 0.99^{50} = 0.605$$

Example. The probability that a student passes the written test is 0.7, find the probability that the student will pass the test on the fourth try.

$$P = 0.7, \qquad g(x;p) = pq^{x-1}$$
$$x = 4;$$

$$g(4;0.7) = (0.7)(1-0.7)^{4-1} = (0.7)(0.3)^3 = 0.0189\,.$$

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

$$P = 0.5,$$
$$x = 3;$$

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