Chapter 3: Discrete Probability Distributions

Lianming Wang

University of South Carolina

January 4, 2023

Section 3.1: Two types of random variables

- A **random variable** is a variable that assumes numerical values associated with the random outcomes of an experiment
- **Discrete random variables:** can only assume a countable number of values.
  - The number of students admitted to medical schools
  - The number of customers waiting to be served in a restaurant at a particular time
- **Continuous random variables:** can assume any values in an interval.
  - The length of time for completing a statistical problem
  - The weight of a chicken sandwich bought at Chick-fil-A
  - The amount of gasoline filled at a gas station

Section 3.2: Probability distribution function for a Discrete Random Variable

The **probability distribution** of a discrete random variable is a graph, table, or formula that specifies the probability associated with each possible value that the random variable can assume.

- Define $p(x) = P(X = x)$ for each x.
- For all value x, $0 \leq p(x) \leq 1$.
- $\sum_x p(x) = 1$, where the summation is over all possible values of X.

Exercise 1. The number x of messages sent per hour over a computer network has the following distribution:

x	10	11	12	13	14	15
$p(x)$	0.08	0.15	0.30	0.20	0.20	0.07

a). What are the possible values for the time of messages sent from the network in an hour according to the distribution?

b). Find the probability that exactly 11 messages will be sent by the network in the next hour.

c). Find the most probable number of messages that will be sent by the network in the next hour.

Section 3.3: Cumulative distribution function

The cumulative distribution function $F(x)$ of random variable X (discrete) is defined as

$$F(x) = P(X \leq x) = \sum_{y \leq x} p(y).$$

- $0 \leq F(x) \leq 1$ for any x.
- $F(x)$ is a nondecreasing function of x.
- $F(x)$ is a step function.
- $F(x)$ is a right-continuous function.

Exercise 1 continued. The number x of messages sent per hour over a computer network has the following distribution:

x	10	11	12	13	14	15
$p(x)$	0.08	0.15	0.30	0.20	0.20	0.07
$F(x)$						

a). Complete the table.

b). Find the probability that no more than 12 messages will be sent by the network in the next hour.

c). Find the probability that more than 11 messages will be sent by the network in the next hour.

Section 3.4: Mean and Variance of Discrete Random Variables

- The expected value (or mean) of a discrete random variable is the weighted average of its possible values.

$$\mu = E(X) = \sum xp(x).$$

- The variance of a discrete random variable is the expected value of the squared distance from the mean.

$$\sigma^2 = E(X-\mu)^2 = \sum x^2 p(x) - \mu^2.$$

- σ is called the standard deviation of the distribution.
- Note: all these three are population parameters, not based on sample.

Exercise 2. A dust mite allergy level that exceeds 2 micrograms per gram (ug/g) of dust has been associated with the development of allergies. Consider a random sample of 4 homes, and let X denote the number of homes that with a dust mite level that exceeds 2 ug/g. The probability distribution of X is given in the following table.

x	0	1	2	3	4
$p(x)$	0.09	0.30	0.37	0.20	0.04

a. Find the probability $P(X \geq 2)$.

b. Find the expected value μ of X.

c. Find the standard deviation σ of X.

d. Find the probability that X falls within interval $(\mu - 2\sigma, \mu + 2\sigma)$.

Some useful formula

- Expected value of a function of a discrete random variable

$$E[h(X)] = \sum_x h(x)p(x).$$

- For a linear function,
  - $E(aX + b) = aE(X) + b$
  - $Var(aX + b) = a^2 Var(X)$
- For two independent random variables X and Y,
  - $E(aX + bY) = aE(X) + bE(Y)$
  - $Var(aX + bY) = a^2 Var(X) + b^2 Var(Y)$
    Note that this is not true if X and Y are dependent.

Sections 3.6: The Binomial Random Variable

The Binomial experiment

- The experiment consists of n identical trials.
- There are only two possible outcomes on each trial. We denote one outcome by S (for Success) and the other by F (for Failure).
- The probability of S remains the same from trial to trial. $P(S)=p$ and $P(F)=1-p=q$
- The trials are independent.

- Let X be the number of successes in a binomial experiment with n trials. Then X is a Binomial random variable. Denoted by $X \sim B(n, p)$.
- The possible values for X are $0, 1, \cdots, n$.
- For each $x = 0, 1, \cdots, n$,

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

- Mean $\mu = np$ and variance $\sigma^2 = np(1-p)$.

- Cumulative distribution function $F(x) = P(X \le x) = \sum_{k \le x} p(k)$.
- Know how to use Binomial table. Suppose $X \sim B(10, 0.4)$.

  $P(X \le 5)$

  $P(X = 3)$

  $P(1 \le X \le 5)$

  $P(2 < X < 8)$

  $P(X \ge 4)$

Exercise 3.

1.It is known that 40% of people prefer cola drinks to iced tea. If 20 people are selected at random, what is the probability that

a. exactly 10 of them will prefer cola drinks to iced tea?

b. more than 14 of them will prefer cola drinks to iced tea?

c. between 4 and 13 (inclusive) of them will prefer cola drinks to iced tea?

d. What is the expected number of people who prefer cola drinks to iced tea?

e. What is the variance of the number of people who prefer cola drinks to iced tea? What is the standard deviation?

Exercise 4. A multiple-choice test consists of 10 items, each with five choices, only one of which is correct. If you guess on each item, what is the probability you will

a. get at least 4 items correct?

b. get at most 2 items correct?

c. get exactly three items correct?

Exercise 5. Suppose you invest money into 20 independent ventures. Assume you know the probability that each venture is a success is 0.6.

a. What is the expected number of ventures that will be a success?

b. Find the standard deviation of the number of successful ventures.

c. Find the interval $\mu \pm 2\sigma$ and calculate the probability that the number of successful ventures falls within that interval.

Examples:

- The number of traffic accidents per month at a busy intersection (e.g., 106-107 on highway I-26)
- The number of noticeable surface defects found by quality inspectors on a new automobile
- The number of parts per million (ppm) of some toxin found the water from a manufacturing plant
- The number of unscheduled admission per day to a hospital
- The number of death claims received per day by an insurance company

Characteristics of a Poisson random variable

- The experiment consists of counting the number of the times a certain event occurs during a given unit of time or in a given area or volume (or weight, distance, or any other unit of measurement).
- The probability that an event occurs in a given unit of time, area, or volume is the same for all units.
- The number of events that occur in a one unit of time, area, or volume is independent of the number that occur in other units.
- The mean (expected) number of events in each unit is the same, denoted by λ.

Poisson distribution: $X \sim P(\lambda)$

- The probability mass function $p(x) = P(X = x)$ is given by

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

  for $x = 0, 1, 2, \cdots$

- λ is the mean number of the events
- $e = 2.71828$ a constant.
- Mean $\mu = \lambda$ and variance $\sigma^2 = \lambda$.

Exercise 6. U.S. Airlines average about 1.6 fatalities per month. Assume that the probability distribution for X, the number of fatalities per month, can be approximated by a Poisson distribution.

a). What is the probability that no fatalities will occur during any given month?

b). What is the probabililty that one fatality will occur during any given month?

c). Find the expected number and standard deviation of X

Exercise 7. The number of cracks in a section of interstate highway that are significant enough to require repair is assumed to follow a Poisson distribution with a mean of two cracks per mile. A one-mile section of the interstate highway is randomly chosen for inspection.

a). What is the probability that exactly two cracks are found in that section of the interstate hightway?

b). What is the probability that at least 1 crack are found in that section of the interstate hightway?

c). What is the probability that at most three cracks are found in that section of the interstate hightway?

A generalization: Poisson process

- In the previous definition of Possion distribution, we focus on 1 unit.
- If we are interested in the number Y of events occurring in t units, then $Y \sim P(\lambda t)$.
- Mean $\mu_Y = \lambda t$ and variance $\sigma_Y^2 = \lambda t$.

Exercise 8. The number of cracks in a section of interstate highway that are significant enough to require repair is assumed to follow a Poisson distribution with a mean of two cracks per mile.

a). What is the probability that there are no cracks that require repar in 5 miles of highway?

b). What is the probability that at least 1 crack requires repair in 5 mile of highway?

c). What is the probability that at least 1 crack requires repair in 0.5 mile of highway?