ISIMS - UNIVERSITY OF SFAX
LSI ADBD - BACHELOR 2
2023/2024

***

DESCRIPTIVE AND INFERENTIAL STATISTICS

***

LECTURE IV

DISCRETE PROBABILITY DISTRIBUTIONS

Lecture Content

PROBABILITY DISTRIBUTION FOR A DISCRETE VARIABLE

Introduction

Recall from chapter I that numerical variables can either be:

- *discrete*: integer values that represent count of something, or
- *Continuous*: values that arise from a measuring process.

Definition: A probability distribution for a discrete variable is a mutually exclusive list of all possible numerical outcomes for that variable and a probability of occurrence associated with each outcome.

Formaly, if x_i, $i \in \mathbb{N}$ is the set of possible values of the random variables X, and $p_i = P(X = x_i)$ the probability of occurrence of outcome i, then the probability distribution of X is the set of couples

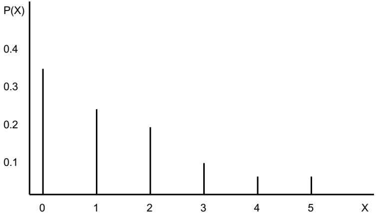
$$(x_i, p_i)_{i\in\mathbb{N}}$$

.

Discrete Probability Distribution

Example 1: Probability Distribution of the number of interruptions per day is given in the table below, with a graphical representation.

Interruptions Per Day In Computer Network x_i	Probability $P(x_i)$
0	0.35
1	0.25
2	0.20
3	0.10
4	0.05
5	0.05



Example 2: A six-sided die has faces for 1, 2, 3, 4, 5, and 6. Therefore for one roll of a fair six-sided die, the set of all possible values are the values 1 through 6. The probability distribution of the outcome is $P(X = k) = \frac{1}{6}, \quad \forall k \in \{1, 2, 3, 4, 5, 6\}$.

Discrete probability Distribution

Exercise 1: A player tosses a fair six-sided die. If he obtains 1, 2 or 3 he wins the equivalent amount in euros. Otherwise he loses 2 euros. Let X be the outcome in euros. Give the probability distribution of X.

Exercise 2: In this game, he outcomes are still the equivalent in euros if he obtains 1, 2 or 3. However if the player obtains greater than 3, he will toss again the die. If he obtains at least 3 in the second toss, he wins 3 euros, otherwise he loses 5 euros. Let Y be the outcome of this new game. Give the probability distribution of Y.

Exercise 3: In a given country, the probability of graduating from High School is p. Give the probability distribution of the associated dummy variable.

Cumulative Distribution Function

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

$$\forall k \in \mathbb{R}, \quad F(x) = P(X \leq x) = \sum_{x_i \leq x} P(X = x_i)$$

Properties: $F(x)$ satisfies the following properties:

1. $0 \leq F(x) \leq 1$
2. $F(-\infty) = 0$ and $F(+\infty) = 1$
3. If $x < y$, then $F(x) < F(y)$

Cumulative Distribution Function

Example 4: The Cumulative Distribution Function for the variable X in example 1 is,

$$F(x) = \begin{cases} 0 & \textit{if} \quad x < 0 \\ 0.35 & \textit{if} \quad 0 \leq x < 1 \\ 0.60 & \textit{if} \quad 1 \leq x < 2 \\ 0.80 & \textit{if} \quad 2 \leq x < 3 \\ 0.90 & \textit{if} \quad 3 \leq x < 4 \\ 0.95 & \textit{if} \quad 4 \leq x < 5 \\ 1 & \textit{if} \quad x \geq 5 \end{cases}$$

Example 5: The Cumulative Distribution Function for the variable Y in example 2 is,

$$F(y) = \begin{cases} 0 & \textit{if} \quad y < 1 \\ 1/6 & \textit{if} \quad 1 \leq y < 2 \\ 2/6 & \textit{if} \quad 2 \leq y < 3 \\ 3/6 & \textit{if} \quad 3 \leq y < 4 \\ 4/6 & \textit{if} \quad 4 \leq y < 5 \\ 5/6 & \textit{if} \quad 5 \leq y < 6 \\ 1 & \textit{if} \quad y \geq 6 \end{cases}$$

Exercise 4: For each of the above examples, represent the graph of F.
Exercise 5: For each of the exercises 1, 2 and 3, find the CDF and represent its graph.

The Expected Value of a Discrete Variable

The **expected value** of a discrete variable is the population mean, μ, of its probability distribution.

$$\mu = E(X) = \sum_{i=1}^{N} x_i p_i$$

where,
x_i = the ith value of the discrete variable X.
$p_i = P(X = x_i)$= probability of occurrence of the ith value of X.
N = number of values of the discrete variable X.

Exercise 6: For each of the previous examples calculate the expected value of the corresponding discrete random variable.

Variance and Standard Deviation of a Discrete Variable

The **variance of a discrete variable** is,

$$\sigma^2 = \sum_{i=1}^{N} [x_i - E(X)]^2 p_i$$

The **standard deviation of a discrete variable** is,

$$\sigma = \sqrt{\sigma^2} = \sqrt{\sum_{i=1}^{N} [x_i - E(X)]^2 p_i}$$

Exercise 7: Calculate the variance and the standard deviation of the discrete random variable number of interruptions per day.

Discrete Probability Distributions - Exercise

Exercise 8: In a semester in college, data was collected of the number of student absences and was correlated with whether students passed the class or not. The data was collected as follows:

Number of absences	Frequency
0	150
1	120
2	67
3	50
4	40
5	10
6	5
7	2
8	1

1. Calculate the expected number of absences per semester.
2. Calculate the standard deviation.

Discrete Probability Distributions - Exercise

Exercise 9: You are trying to develop a strategy for investing in two different stocks. The anticipated annual return for a $1,000 investment in each stock under four different economic conditions has the following probability distribution:

		Returns	
Probability	**Economic Condition**	**Stock X**	**Stock Y**
0.1	Recession	- 50	- 100
0.3	Slow growth	20	50
0.4	Moderate growth	100	130
0.2	Fast growth	150	200

Compute the

1. expected return for stock X and for stock Y.
2. standard deviation for stock X and for stock Y
3. Would you invest in stock X or stock Y ?

BINOMIAL DISTRIBUTION

The Binomial Distribution

Consider the following random experiments and random variables:

- Flip a coin 10 times. Let X = number of heads obtained.
- A worn machine tool produces 1% defective parts. Let X = number of defective parts in the next 25 parts produced.
- 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.
- In 20 births at a hospital, let X = the number of female births.
- Of all patients suffering a particular illness, 35% experience improvement from a particular medication. In the next 100 patients administered the medication, let X = the number of patients who experience improvement.

The Binomial Distribution

Each of these random experiments can be thought of as consisting of a series of repeated, random trials: 10 flips of the coin in experiment 1, the production of 25 parts in experiment 2, and so forth.

- Each trial has only two possible outcomes: either a *success* or a *failure*. These two terms are just labels. We can just as well use A and B or 0 or 1.
- These trials are called **Bernoulli trials**, and have two properties:
  1. independent.
  2. probability of success in each trial is constant and equals π
  3. probability of failure is $(1 - \pi)$

The Binomial Distribution - Formal Definition

Consider a random experiment consisting 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 π, remains constant.

The random variable X that equals the number of trials resulting in a success is a binomial random variable with parameters $0 < \pi < 1$ and $n = 1, 2, \ldots$. The probability distribution of X is given by:

$$P(X = k) = \frac{n!}{k!(n-k)!}\pi^k(1-\pi)^{n-k}$$

The term $\frac{n!}{k!(n-k)!}$ is the number of combinations of the x events of interest from the n observations possible.

The Binomial Distribution - Expectation and Variance

Theorem: If X is a binomial random variable with parameters π and n, the mean (expected value) and variance of X are given by,

$$\mu = E(X) = n\pi \qquad \text{and} \qquad \sigma^2 = V(X) = n\pi(1-\pi)$$

Example 6: The increase or decrease in the price of a stock between the beginning and the end of a trading days assumed to be an equally likely random event. To calculate the probability that a stock will show an increase in its closing price on three days out of five, use the formula.

With $n = 5$ and $\pi = 0.5$ we have, $P(X = 3) = \frac{5!}{3! \times 2!} 0.5^3 0.5^2$

Or simply, instead of your calculator use the table of the,

BINOMIAL DISTRIBUTION CUMULATIVE PROBABILITIES

$$P(X = 3) = P(X \leq 3) - P(X \leq 2) = 0.8125 - 0.5 = 0.3125$$

The Binomial Distribution - Exercises

Exercise 10: Let X be a random variable following a binomial distribution with parameters $n = 20$ and $p = 0.1$.

1. Calculate $P(X = 5)$, $P(X \leq 2)$, $P(X < 4)$, $P(X = 1.5)$, $P(3 \leq X \leq 4)$, $P(2 < X \leq 8)$
2. Find the values of x such that $P(X \geq x) \leq 0.75$
3. Let Y be a binomial random variable with parameters $n = 20$ and $p = 0.9$. Calculate $P(Y = 16)$.

Exercise 11: A multiple-choice test contains 25 questions, each with four answers. Assume that a student just guesses on each question.

1. What is the probability that the student answers more than 20 questions correctly?
2. What is the probability that the student answers fewer than 5 questions correctly?

The Binomial Distribution - Exercises

Exercise 12: An article in *QSR Magazine* reports, about Wendy's fast-food-chain drives, that the percentage of orders filled correctly was approximately 86.9%. Suppose that you go to the drive-through window at Wendy's with your two friends and place an order.

1. What are the probabilities that all three, that none of the three, and that at least two of the three orders will be filled correctly?
2. What are the mean and standard deviation of the binomial distribution for the number of orders filled correctly?

Exercise 13: Let X be a random variable following a binomial distribution $Bin(n, p)$.
Y is another random variable defined by $Y = X$ if $X \neq 0$ and Y takes randomly a value in $\{0, 1, ..., n\}$ if $X = 0$.
Give the distribution of Y and calculate $E(Y)$.

POISSON DISTRIBUTION

The Poisson Distribution

Poisson distribution is used to calculate probabilities of the numbers of occurrences of an event in situations if the following properties hold:

- We are interested in counting the number of times a particular event occurs in a fixed interval of time or space, often referred to as an *area of opportunity*.
- The probability that an event occurs in a given area of opportunity is the same for all the areas of opportunity.
- The number of events that occur in one area of opportunity is independent of the number of events that occurs in any other area of opportunity.
- The number that two or more events will occur approaches zero as the area opportunity becomes smaller.

The Poisson Distribution

Example of variables that follow the Poisson distribution are:

- The number of networks failure in a day
- The number of passengers in a given train trip
- The number of fleas on the body of a dog.
- The number of customers arriving during the lunch hour at a bank

The Poisson distribution has one parameter called λ. The Poisson distribution gives the formula for computing the probability of $(X = x)$ events,

$$P(X = x|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$$

Theorem: If X is a random variable following a Poisson distribution with parameter λ, then,

$$\mu = E(X) = \lambda \qquad \text{and} \qquad \sigma^2 = V(X) = \lambda$$

The Poisson Distribution

Example: Suppose the mean number of customers who arrive per minute at a bank during the noon-to-1 P.M. hour is equal to 3.

- The probability that in a given minute, exactly two customers will arrive is

$$P(X = 2|\lambda = 3) = \frac{e^{-3}3^2}{2!} = 0.224$$

- The probability that more than two customers will arrive in a given minute is,

$$\begin{aligned} P(X > 2|\lambda = 3) &= 1 - P(X \leq 2) \\ &= 1 - [P(X = 0) + P(X = 1) + P(X = 2)] \\ &= 1 - \left[\frac{e^{-3}3^0}{0!} + \frac{e^{-3}3^1}{1!} + \frac{e^{-3}3^2}{2!}\right] \\ &= 0.5768 \end{aligned}$$

Or simply, instead of your calculator use the table of the,

POISSON DISTRIBUTION CUMULATIVE PROBABILITIES

$$P(X \leq 2) = 0.4232$$

The Poisson Distribution - Exercises

Exercise 14: A toll-free phone number is available from 9 A.M. to 9 P.M. for your customers to register complaints about a product purchased from your company. Past history indicates that a mean of 0.8 calls is received per minute.

1. What properties must be true about the situation described here in order to use the Poisson distribution to calculate probabilities concerning the number of phone calls received in a one-minute period?
2. Assuming that this situation matches the properties discussed in (1), what is the probability that during a one-minute period,
   - 2.1 zero phone calls will be received ?
   - 2.2 three or more phone calls will be received ?
   - 2.3 what is the minimum number of phone calls that will be received in a one-minute period 99.99% of the time ?

The Poisson Distribution - Exercises

Exercise 15: The article "An Association Between Fine Particles and Asthma Emergency Department Visits for Children in Seattle" [Environmental Health Perspectives June, 1999 107(6)] used Poisson models for the number of asthma emergency department (ED) visits per day. For the zip codes studied, the mean ED visits were 1.8 per day. Determine the following:

1. Probability of more than five visits in a day.
2. Probability of fewer than five visits in a week.
3. Instead of a mean of 1.8 per day, determine the mean visits per day such that the probability of more than five visits is 0.1.

Exercise 16: Let X be a random variable following a Poisson distribution with parameter $\lambda = 5$.

1. Calculate $P(X = 6)$, $P(X < 4)$, $P(X \geq 5)$, $P(\pi/2 < X < \pi)$
2. Find the value of x such that $P(X < x) \geq 0.95$