Chapter 4: Discrete Probability Distributions

Department of Mathematics
Izmir University of Economics

Week 5-6
2014-2015

Introduction

In this chapter we will focus on

- *random variables* and the classification of random variables,
- probability and cumulative probability distributions for discrete random variables,
- expected value, variance, and standard deviation of a discrete random variable, and
- some special discrete distributions.

Random Variables

Definition:

A *random variable* X is a function which assigns a **unique numerical value** to each outcome of a random experiment. ($X : S \to \mathbb{R}$)

Example. Consider the experiment of flipping a coin three times and let X be the number of heads. Find the values of the random variable X.

Solution.

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

$$X(HHH) = 3, X(THH) = X(HTH) = X(HHT) = 2,$$
$$X(TTH) = X(THT) = X(HTT) = 1, X(TTT) = 0.$$

X can take values 0, 1, 2, and 3.

Random Variables

Example. Consider the experiment of flipping a coin repeatedly until the first occurrence of a head and let Y be the number of flips. Find the values of the random variable Y.

Solution.

The sample space is $S = \left\{ H, TH, TTH, TTTH, \ldots, \underbrace{TT\ldots T}_{k}H, \ldots \right\}$.

$$Y(H) = 1, Y(TH) = 2, Y(TTH) = 3, \ldots, Y\left(\underbrace{TT\ldots T}_{k}H\right) = k + 1, \ldots$$

Y can take values 1, 2, 3 $\ldots$.

Random Variables

It is very important to distinguish between a random variable and the possible values that it can take. We use capital letters such as X to denote the random variable and the corresponding lower-case letter, x, to denote a possible value.

Random variables can be classified into two types: discrete and continuous. In this chapter we'll only consider discrete random variables.

Random Variables

Definition:

A random variable is a *discrete random variable* if it can take on no more than countable (finitely many or countably infinite) number of values.

Some examples of discrete random variables are:

1. The number of sales resulting from 10 customers.
2. The number of defective items in a sample of 20 items from a large shipment.
3. The number of customers arriving at a checkout counter in an hour.
4. The number of errors detected in a corporation's accounts.
5. The number of claims on a medical insurance policy in a particular year.

Probability Distributions for Discrete Random Variables

Definition:

The probability distribution function, $P(x)$, of a discrete random variable X represents the probability that X takes the value x, as a function of x. That is,

$$P(x) = P\{X = x\} \text{ for all values of } x.$$

Example. Consider the experiment of flipping a coin three times and let X be the number of heads. Find $P(x)$ for all values of x and show the result in a table, a graph, and a function representation.

Properties of Probability Distribution for Discrete Random Variables
Let X be a discrete random variable with probability distribution $P(x)$. Then

1. $0 \leq P(x) \leq 1$ for any value x, and
2. $\sum_x P(x) = 1$, where summation is over all possible values of x.

Example. Check whether the following functions can serve as probability distribution functions of appropriate random variables:

a) $f(x) = \frac{\binom{3}{x}}{8}, x = 0, 1, 2, 3,$

b) $f(x) = \frac{x+2}{12}, x = 1, 2, 3,$

c) $f(x) = \frac{x^2-1}{25}, x = 0, 1, 2, 3, 4.$

Cumulative Probability Distributions for Discrete Random Variables

Definition:

The cumulative probability distribution function, $F(x)$, of a discrete random variable X represents the probability that X does not exceed the value x_0, as a function of x_0. That is,

$$F(x_0) = P\{X \leq x_0\},$$

where the function is evaluated at all values of x_0.

Let X be a random variable with probability distribution $P(x)$ and cumulative probability distribution $F(x_0)$. Then

$$F(x_0) = \sum_{x \leq x_0} P(x),$$

where the summation is over all possible values of x less than or equal to x_0.

Properties of Cumulative Probability Distribution for Discrete Random Variables

Let X be a discrete random variable with cumulative probability distribution $F(x_0)$. Then

1. $0 \leq F(x_0) \leq 1$ for every number x_0, and
2. If x_0 and x_1 are two numbers with $x_0 < x_1$, then $F(x_0) \leq F(x_1)$.

Example. In a geography assignment the grade obtained is the random variable X. It has been found that students have these probabilities of getting a specific grade: $A : 0.18$, $B : 0.32$, $C : 0.25$, $D : 0.07$, $E : 0.03$, and $F : 0.15$. Based on this,

- **a)** Calculate the cumulative probability distribution of X and show the result in a table and a graph representation.
- **b)** Calculate the probability of getting a higher grade than B.
- **c)** Calculate the probability of getting a lower grade than C.

Expected Value of a Discrete Random Variable

Definition:

The expected value, $E(X)$, of a discrete random variable X is defined by

$$E(X) = \sum_{x} xP(x),$$

where the summation is over all possible values of x.

The expected value of a random variable is also called its mean and is denoted by μ.

Example. A review textbooks in a segment of the business area found that 81% of all pages of texts were error free, 17% of all pages contained one error, and the remaining 2% contained two errors.Find the mean number of errors per page.

Example. The police chief of a city knows that the probabilities for 0, 1, 2, 3, 4, or 5 car thefts on any given day are respectively 0.21, 0.37, 0.25, 0.13, 0.03, and 0.01. How many car thefts can he expect per day?

Variance and Standard Deviation of a Discrete Random Variable

Definition:

The variance, $Var(X)$, of a discrete random variable X is denoted by σ^2 and is given by

$$\sigma^2 = Var(X) = E\left((X-\mu)^2\right) = \sum_x (x-\mu)^2 P(x),$$

where the summation is over all possible values of x.

The variance of a discrete random variable X can also be expressed as

$$\sigma^2 = E\left(X^2\right) - (E(X))^2 = \sum_x x^2 P(x) - \mu^2.$$

The standard deviation, σ, is the positive square root of the variance.

Example. An automobile dealer calculates the proportion of new cars sold that have been returned a various number of times for the correction of defects during the warranty period. The results are shown in the following table

Number of returns	0	1	2	3	4
Proportion	0.28	0.36	0.23	0.09	0.04

a) Graph the probability distribution.

b) Calculate the cumulative probability distribution.

c) Find the mean of the number of returns of an automobile for corrections for defects during the warranty period.

d) Find the variance of the number of returns of an automobile for corrections for defects during the warranty period.

Expected Value of Functions of Discrete Random Variables

Definition:

Let X be a discrete random variable with probability distribution $P(x)$ and let $g(X)$ be some function of X. Then the expected value, $E[g(X)]$, of that function is defined as follows:

$$E[g(X)] = \sum_{x} g(x)\,P(x),$$

where the summation is over all possible values of x.

Expected Value of Functions of Discrete Random Variables

Properties for Linear Functions of a Random Variable
Let X be a random variable with mean μ_X and variance σ_X^2 and let a and b be any constants. Define the random variable Y as $a + bX$. Then, the mean and variance of Y are

$$\mu_Y = E(Y) = E(a + bX) = E(a) + E(bX) = a + bE(X) = a + b\mu_X$$

and

$$\sigma_Y^2 = Var(Y) = E\left(Y^2\right) - \mu_Y^2 = E\left((a + bX)^2\right) - (a + b\mu_X)^2 = \cdots = b^2\sigma_X^2$$

so that the standard deviation of Y is

$$\sigma_Y = |b|\sigma_X.$$

Example. A contractor is interested in the total cost of a project on which she intends to bid. She estimates that the materials will cost \$25000 and that her labor will be \$900 per day. If the project takes X days to complete, the total labor cost will be \$900X and the total cost of the project (in dollars) will be

$$C = 25000 + 900X.$$

Using her experience the contractor form probabilities of likely completion times for the project as

Completion time (in days)	10	11	12	13	14
Probability	0.1	0.3	0.3	0.2	0.1

a) Find the mean and variance for the completion time X.

b) Find the mean, variance, and standard deviation for total cost C.

Example. In a game of chance we win 5 TL if we roll at least one 4 or a sum of 7 when a pair of dice is used and lose 3 TL otherwise.

a) Find the expected gain or loss.

b) If we pay 30 TL before playing the game 10 times what is the expected total prize?

Binomial Distribution

Consider a random experiment that can give rise to just two possible mutually exclusive and collectively exhaustive outcomes, which for convenience we label "success" and "failure". Let p denote the probability of success and $1 - p$ the probability of failure. Define the random variable X so that X takes the value 1 if the outcome of the experiment is a success and 0 if it is a failure, that is,

$$X = \begin{cases} 1 & \text{if the outcome of the experiment is a success,} \\ 0 & \text{otherwise.} \end{cases}$$

The probability distribution of X is then

$$P(0) = P\{X = 0\} = 1 - p \qquad P(1) = P\{X = 1\} = p.$$

This distribution is known as the *Bernoulli distribution*.

Binomial Distribution

Example. (Mean and variance of a Bernoulli random variable)
Let X be a Bernoulli random variable. Find μ_X and σ_X^2.
Solution.

$$\begin{aligned}\mu_X &= E(X) = \sum_{x=0}^{1} xP(x) = 0P(0) + 1P(1) \\ &= 0(1-p) + 1(p) = p\end{aligned}$$

$$\begin{aligned}\sigma_X^2 &= Var(X) = E\left(X^2\right) - \mu_X^2 \\ &= \sum_{x=0}^{1} x^2 P(x) - p^2 = 0^2 P(0) + 1^2 P(1) - p^2 \\ &= 0(1-p) + 1(p) - p^2 = p - p^2 = p(1-p)\end{aligned}$$

Binomial Distribution

An important generalization of the Bernoulli distribution concerns the case where a random experiment with two possible outcomes is repeated several times and the repetitions are independent.

Suppose that a random experiment can result in two possible mutually exclusive and collectively exhaustive outcomes, "success" and "failure", and that p is the probability of success in a single trial. If n independent trials are carried out, the distribution of the number of the successes, X, is called the *binomial distribution*.

Binomial Distribution

The probability distribution of the binomial random variable X is given by

$$\begin{aligned} P(x) &= P\{X = x\} = P\{x \text{ successes occur in } n \text{ independent trials}\} \\ &= \binom{n}{x} p^x (1-p)^{n-x} = \frac{n!}{x!\,(n-x)!} p^x (1-p)^{n-x}, \quad x = 0, 1, \ldots, n. \end{aligned}$$

The mean and variance of a binomial random variable X are

$$\mu_X = E(X) = np \qquad \text{and} \qquad \sigma_X^2 = Var(X) = np(1-p).$$

Binomial Distribution

In a binomial distribution application,

1. there are several trials, each of which has only to outcomes: yes/no, on/off, success/failure,
2. the probability of the outcome is the same for each trial,
3. the probability of the outcome on one trial does not affect the probability on other trials.

Binomial Distribution

Example. Suppose that a real estate agent has 5 contacts and she believes that for each contact the probability of making a sale is 0.40.

a) Find the probability that she makes at most 1 sale.

b) Find the probability that she makes between 2 and 4 sales.

c) Graph the probability distribution function.

Binomial Distribution

Example. A study conducted at a certain college shows that 65% of the school's graduates obtain a job in their fields within a year after graduation. Find the probabilities that within a year after graduation of 14 randomly selected graduate of that college

a) at least 6 will find a job in their fields,

b) at most 3 will find a job in their fields,

c) between 5 and 8 will find a job in their fields.

Binomial Distribution

Example. A student takes an eight-question multiple choice exam. Each question has five choices for answers, only one of which is correct. The student forgot to study for the exam and guesses each question. Let X be the number of correct answers. Find $P\{X = 6\}$.

Binomial Distribution

Example. If the probability of a set of a tennis match will go into tie-breaker is 0.18, what is the probability that two of three sets will go into tie-breaker? Find the expected value and variance of the number of sets that go into tie-breaker in a three sets tennis match.

Binomial Distribution

Example. What is the expected number of correct answers in a multiple choice exam consisting of 20 questions where each question has 4 choices and all questions are answered only by guessing?

Poisson Distribution

The random variable X is said to follow a Poisson distribution if it has the probability distribution

$$P(x) = P\{X = x\} = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots,$$

where $\lambda (> 0)$ is the expected number of occurrences per unit time and $e \approx 2.7182$.

The mean and variance of a Poisson random variable X are

$$\mu_X = E(X) = \lambda \qquad \text{and} \qquad \sigma_X^2 = Var(X) = \lambda.$$

Poisson Distribution

We use Poisson distribution to determine the probability of a random variable that is characterized as the number of occurrences or successes of a certain event in a given continuous interval. Some examples of these random variables are:

1. The number of failures in a large computer system during a given day.
2. The number of replacement orders for a part received by a firm in a given month.
3. The number of delivery trucks to arrive at a central warehouse in an hour.
4. The number of customers to arrive for flights during each 10-minute interval from 3.00 pm to 6.00 pm on weekdays.
5. The number of pine trees per unit area in a mixed forest.

Poisson Distribution

Assume that an interval is divided into a very large number of equal subintervals so that the probability of the occurrence of an event in any subinterval is very small. Then, we can use Poisson distribution if the following are true:

1. The probability of the occurrence of an event is constant for all subintervals.
2. There can be no more than one occurrence in each subinterval.
3. Occurrences are independent, that is, an occurrence in one subinterval does not influence the probability of an occurrence in another subinterval.

Poisson Distribution

Example. Find

a) $P\{X = 7|\lambda = 3.5\}$ and

b) $P\{X \leq 2|\lambda = 5\}$.

Poisson Distribution

Example. Customers arrive at a photocopying machine at an average rate of 2 every five minutes. Assuming that the arrivals are independent, find the probability that more than two customers arrive in a 5-minute interval.

Poisson Distribution

Example. An instructor receives an average 4.2 emails from students the day before a final exam. What is the probability of receiving at least 3 emails on such a day?

Poisson Distribution

Note: The sum of k Poisson random variables with respective means $\lambda_1, \lambda_2, \ldots, \lambda_k$ is a Poisson random variable with mean $\lambda_1 + \lambda_2 + \cdots + \lambda_k$.

Poisson Distribution

Example. A computer center manager reports that his computer system experienced three component failures during the past 100 days.

a) What is the probability of no failures in a given day?

b) What is the probability of one or more component failures in a given day?

c) What is the probability of at least two failures in a 3-day period?

Poisson Approximation to the Binomial Distribution

Let X be the number of successes resulting from n independent trials each with probability of success p. Then the distribution of X is binomial with mean np. If the number of trials (n) is large and the probability of success (p) is small so that np is of moderate size (preferably $np \leq 7$), this distribution can be approximated by the Poisson distribution with $\lambda = np$. The probability distribution of the approximating distribution is

$$P(x) = \frac{e^{-np}(np)^x}{x!}, \qquad x = 0, 1, 2, \ldots$$

Poisson Approximation to the Binomial Distribution

Example. A corporation has 250 PCs. The probability that any of them will require repair in a week is 0.01. Find the probability that fewer than 4 of them will require repair in a particular week.

Hypergeometric Distribution

Suppose that a random sample of n objects is chosen from a group of N objects, s of which are successes. The distribution of the number of successes, X, in the sample is called the hypergeometric distribution. The probability distribution of the random variable X is

$$P(x) = P\{X = x\} = \frac{\binom{s}{x}\binom{N-s}{n-x}}{\binom{N}{n}} = \frac{\frac{s!}{x!(s-x)!}\frac{(N-s)!}{(n-x)!(N-s-n+x)!}}{\frac{N!}{n!(N-n)!}},$$

where $\max(0, n-N+s) \leq x \leq \min(n, s)$ and its mean is $\mu_X = E(X) = n\frac{s}{N}$.

Hypergeometric Distribution

Example. Let 4 of the tape recorders in a lot which contains a total of 16 are defective. Suppose that we randomly select 3 of tape recorders from this lot. Let X be the number of defective tape recorders in the selected 3.

a) Find $P\{X = 1\}$.

b) Graph the probability distribution of X.

c) Find $E(X)$.

Hypergeometric Distribution

Example. A committee of 8 members is to be formed from a group of 8 men and 8 women. If the choice of the committee members is made randomly, what is the probability that precisely half of these members will be women?