6. Discrete Random Variables and Probability Distributions

Random variables and probability distributions

- A random variable is a function that assigns a numerical value to each outcome in a sample space.
- A random variable reflects the aspect of a random experiment that is of interest to us.
- There are two types of random variables:

  1. Discrete random variable

  2. Continuous random variable

A random variable is *discrete* if it can assume only a countable (or finite) number of values. A random variable is *continuous* if it can assume an uncountable number of values, for example, any value from a certain interval.

Discrete probability distribution

- A table, formula, or graph that lists all possible values a discrete random variable can assume, together with associated probabilities, is called *a discrete probability distribution.*
- To calculate $P(X = x)$, the probability that the random variable X assumes the value x, add the probabilities of all the outcomes for which X is equal to x.

Discrete probability distribution: example

Example 1. Find the probability distribution of the random variable describing the number of heads that turn up when a coin is flipped twice.

Solution.

The possible values are 0, 1, and 2. Therefore, we get

$$P(X=0)=P(TT)=1/4$$
$$P(X=1)=P(TH)+P(HT)=1/2$$
$$P(X=2)=P(HH)=1/4$$

Therefore, the probability table for X is given by

x	0	1	2
$P(X=x)$	$1/4$	$1/2$	$1/4$

Requirements of discrete probability distribution

If a random variable X can take values $x_1, x_2, \ldots$ with probabilities

$$p(x_i) = P(X = x_i)$$

then the following *must* be true:

1. $0 \le p(x_i) \le 1$ for all x_i

2. $p(x_1) + p(x_2) + \cdots = \sum_{all\ x_i} p(x_i) = 1$

The probability distribution can be used to calculate probabilities of different events.

Probabilities as relative frequencies

In practice, often probabilities are estimated from relative frequencies.

Probabilities as relative frequencies: example

Example 2. The numbers of cars a dealer is selling daily were recorded in the last 100 days. This data was summarized as follows:

Daily sales	Frequency
0	5%
1	15%
2	35%
3	25%
4	20%

1. Construct the probability distribution table for the number of cars sold daily.

2. Find the probability of selling more than 2 cars a day.

Probabilities as relative frequencies: example

Solution

1. Let X be the number of cars sold during a randomly selected day. Based on the 100 day observation, the estimated probability distribution table is given by

x	0	1	2	3	4
$P(X=x)$	$.05$	$.15$	$.35$	$.25$	$.20$

2. The probability of selling more than two cars is

$$P(X>2)=P(X=3)+P(X=4)=.25+.20=.45$$

Expected value

Given a discrete random variable X that takes values x_i with probabilities $p(x_i) = P(X = x_i)$, the expected value of X is

$$E(X) = \sum_{all\ x_i} x_i p(x_i) = x_1 p(x_1) + x_2 p(x_2) + \cdots$$

The expected value of a random variable X is the *weighted average* of the possible values it can assume, where the weights are the corresponding probabilities of each x_i.

Laws of expected value

Properties of the expected value:

- $E(c) = c$, if c is a constant
- $E(aX) = aE(X)$, if a is a constant
- $E(X + Y) = E(X) + E(Y)$
- $E(XY) = E(X)E(Y)$, *if* random variables X and Y are *independent*

Independent random variables

Definition: **Random variables X and Y are *independent* iff**

$$P(X = x \; and \; Y = y) = P(X = x)P(Y = y)$$

for *all possible* x and y.

Variance

Let X be a discrete random variable with possible values x_i that occur with probabilities $p(x_i) = P(X = x_i)$. The variance of X is defined to be

$$Var(X) = \sum_{all\ x_i} (x_i - E(X))^2 p(x_i)$$

$$= (x_1 - E(X))^2 p(x_i) + (x_2 - E(X))^2 p(x_2) + \cdots$$

The variance is the weighted average of the squared deviations of the values of X from their mean $E(X)$, where the weights are the corresponding probabilities of each x_i.

Standard deviation

The standard deviation of a random variable X, usually denoted by σ, is the positive square root of the variance of X, that is,

$$\sigma = \sqrt{Var(X)}$$

The standard deviation gives us the average deviation of values of random variables from the expected value in terms of original units.

Variance and mean: example

Example 3. The total number of cars to be sold on a randomly selected day, X, is described by the following probability distribution:

x	0	1	2	3	4
$P(X = x)$	$.05$	$.15$	$.35$	$.25$	$.20$

Determine the expected value and standard deviation of random variable X, the number of cars sold.

Variance and mean - example

First, let us calculate the expected value:

$$\boldsymbol{E(X) = 0 \times .05 + 1 \times .15 + 2 \times .35 + 3 \times .25 + 4 \times .20 = 2.4}$$

Second, the variance is equal to

$$\boldsymbol{\begin{aligned} Var(X) = {} & (0 - 2.4)^2 \times .05 + (1 - 2.4)^2 \times .15 \\ & + (2 - 2.4)^2 \times .35 + (3 - 2.4)^2 \times .25 \\ & + (4 - 2.4)^2 \times .20 = 1.24 \end{aligned}}$$

Finally,

$$\boldsymbol{\sigma = \sqrt{1.24} \approx 1.114}$$

That, on average we expect to sell $\mathbf{2.4}$ cars $\boldsymbol{\pm 1.1}$

Properties of the variance

Properties of the variance:

- $Var(c) = 0$, where c is a constant
- $Var(aX) = a^2 Var(X)$
- $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$, where
$$Cov(X, Y) = E(XY) - E(X)E(Y)$$
- If X and Y are independent, then $Cov(X, Y) = 0$ and, as a consequence,
$$Var(X + Y) = Var(X) + Var(Y)$$
- If $Var(X) = 0$, then X is a number, not a random variable.

An expected value of $f(X)$

An expected value of $f(X)$ is given by

$$E\big(f(X)\big) = \sum_{all\,x_i} f(x_i)p(x_i)$$

This allows us to write a concise expression for the variance of X:

$$Var(X) = E\big(X - E(X)\big)^2$$

Moreover, using the properties of the expected value one can show also that

$$Var(X) = E\big(X^2\big) - \big(E(X)\big)^2$$

Discrete probability distribution: exercises

Exercise 1. Ten thousand Instant Money lottery tickets were sold. One ticket has a face value of $1000, 5 tickets have a face value of $500 each, 20 tickets are worth $100 each, 500 are worth $1 each, and the rest are losers. Let $X =$ face value of a ticket that you buy.

1. Find the probability distribution for X.
2. Calculate the mean and variance of X.

Discrete probability distribution: exercises

Exercise 2. A high school class decides to raise some money by conducting a raffle. The students plan to sell 2000 tickets at $1 apiece. They will give one prize of $100, two prizes of $50, and three prizes of $25. If you plan to purchase one ticket, what are your expected net winnings (expected return)?

Discrete probability distribution: exercises

Exercise 3. A marble is drawn at random and without replacement from a bowl containing four red and three green marbles until a red marble is picked. Let the random variable X denote the number of marbles drawn. Find the probability distribution of X.

Bernoulli trial

The *Bernoulli trial* can result in only one out of two outcomes.

Typical cases where the Bernoulli trial applies:

- A coin flipped results in heads or tails
- An election candidate wins or loses
- An employee is male or female
- A car uses 87 octane gasoline, or another gasoline

Binomial experiment

- There are n Bernoulli trials (n is finite and *fixed*).
- Each trial can result in a *success* or a *failure*.
- The probability p of success is *the same* for all n trials.
- All the trials of the experiment are *independent*.

Binomial random variable

The binomial random variable **counts the number of successes in n trials of the binomial experiment. By definition, this is a discrete random variable. The list of possible values is: $0, 1, \dots, n$**

But what is $P(X = x)$ for $x \in \{0, 1, \dots, n\}$?

Calculating the binomial probability

One can show that the binomial probability for $x \in \{0, 1, \dots, n\}$ is given by the following formula:

$$P(X = x) = C_x^n \times p^x \times (1 - p)^{n-x}$$

where C_x^n (sometimes it is denoted by $C(n, x)$) is a special number called the *number of combinations*. This number gives us the number of different ways of choosing x objects from a collection of n objects. Using *multiplication principle,* we can show that

$$C_x^n = \frac{n!}{x!\,(n-x)!}$$

Recall: $n! = 1 \times 2 \times \cdots \times n$, and by convention $0! = 1$.

Number of combinations or “n-choose-x”

Example 4

Suppose that we have a group of 4 people, say A, B, C, and D. How many different pairs can we select from this group?

Solution

The answer is “4-choose-2”:

$$C_2^4 = \frac{4!}{2!\,2!} = \frac{1 \times 2 \times 3 \times 4}{1 \times 2 \times 1 \times 2} = 6$$

Indeed, we have 6 pairs: (AB), (AC), (AD), (BC), (BD), and (CD).

Binomial distribution: example

Example 5

Suppose you toss a fair die 4 times. What is the probability that the face 6 will show up at least twice?

Solution

Let us say that if the face 6 shows up we have a success, otherwise, it is a failure. We have 4 trials. The probability of a success in one trial is $1/6$. Since the die has no memory what happened to it in the other trials, all the trials are independent. That is, we have a binomial experiment with $n = 4$ trials and probability of a success $p = 1/6$.

Let X be the number of times when the face 6 will show up in our 4 tosses. It has binomial distribution with $n = 4$ and $p = 1/6$.

Binomial distribution: example

Therefore,

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

Now, every individual probability can be found with help of the binomial formula:

$$P(X = 2) = C_2^4 \left(\frac{1}{6}\right)^2 \left(1 - \frac{1}{6}\right)^{4-2} = 6\left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^2 \approx .11574$$

$$P(X = 3) = C_3^4 \left(\frac{1}{6}\right)^3 \left(1 - \frac{1}{6}\right)^{4-3} = 4\left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^1 \approx .01543$$

$$P(X = 4) = C_4^4 \left(\frac{1}{6}\right)^4 \left(1 - \frac{1}{6}\right)^{4-4} = 1\left(\frac{1}{6}\right)^4 \left(\frac{5}{6}\right)^0 \approx .00077$$

That is, $P(X \geq 2) \approx .13194$

Mean and variance of binomial random variable

If X has a binomial distribution with probability of success p and number of trials n, then by definition the expected value of X is given by

$$E(X) = 0C_0^n p^0 (1-p)^n + 1C_0^n p^1 (1-p)^{n-1} + \ldots + nC_n^n p^n (1-p)^0$$

However, one can show that this long expression can be simplified to just

$$E(X) = np$$

Similarly, for the variance we have

$$Var(X) = np(1-p)$$

Binomial random variable: exercises

Exercise 3. A fair die is tossed three times. If we observe the face 6 we call it success. What is the probability of having exactly two successes?

Binomial random variable: exercises

Exercise 4. Forty percent of the students at a large university are in favor of a ban on drinking in the dorms. Suppose 15 students are to be randomly selected. Find the probability that

1. seven favor the ban;
2. fewer than two favor the ban.
3. What is expected number of students in the sample that favor the ban? Variance?

Binomial random variable: exercises

Exercise 5. Sixty percent of all students at a university are female. A committee of five students is selected at random. Only one is a woman. Find the probability that no more than one woman is selected. What might be your conclusion about the way the committee was chosen?

Binomial random variable: exercises

Exercise 6. A jury has 12 jurors. A vote of at least 10 of 12 for "guilty" is necessary for a defendant to be convicted of a crime. Assume that each juror acts independently of the others and that the probability that anyone juror makes the correct decision on a defendant is .80. If the defendant is guilty, what is the probability that the jury makes the correct decision?

Poisson distribution (optional)

The Poisson experiment typically fits cases of rare events that occur over a fixed amount of time or within a specified region

Typical cases:

- The number of errors a typist makes per page
- The number of customers entering a service station per hour
- The number of telephone calls received by a switchboard per hour

Poisson experiment (optional)

Properties of the Poisson experiment:

- The number of successes (events) that occur in a certain time interval is independent of the number of successes that occur in another non-overlapping time interval.
- The average number of a success in a certain time interval is 1) the same for all time intervals of the same size and 2) proportional to the length of the interval
- The probability that two or more successes will occur in an interval approaches zero as the interval becomes smaller.

The Poisson random variable (optional)

The Poisson variable indicates the number of successes that occur during a given time interval or in a specific region in a Poisson experiment

Probability distribution of the Poisson random variable (optional)

Probability distribution of Poisson random variable is given by

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

where $x = 0, 1, 2, \ldots$

The expected value and variance are given by

$$E(X) = \lambda$$

$$Var(X) = \lambda$$

Poisson approximation of the binomial distribution (optional)

When n is very large, the binomial formula might be difficult to use. Instead approximations (via Poisson or normal) are employed.

In particular if p is very small ($p <.05$), but n is large ($np > 5$), we can approximate the binomial probabilities using Poisson distribution. More specifically, we have the following approximation:

$$P(X = x) \approx P(Y = x)$$

where X has binomial distribution with parameters n and p, and Y has Poisson with parameter $\lambda = np$.