

Chapter 5: Discrete Random Variables

Section 5.1 Random Variables (UE 2.1)

Note: This is a combination of Section 5.1 and *Undergraduate Econometrics* (UE) 2.1.

Recall in our discussion on probability we started out with some random experiment that gave rise to our set of all possible outcomes S . We developed some rules for calculating probabilities about various events. Often the events can be expressed in terms of a "random variable" taking on certain outcomes. Loosely, this random variable will represent the value of the variable or characteristic of interest, but *before we look*. Before we look, the value of the variable is not known and could be any of the possible values with various probabilities, hence the name of a "random" variable.

Definition:

- A **random variable** is the numerical outcome of a random experiment or phenomenon.

The value of a random variable is unknown until it is observed, and it is not perfectly predictable.

We will consider **two types** of random variable:

- **discrete** random variables: take a finite number of values.
- **continuous** random variables: take any value in some interval.

We study discrete random variables in Chapter 5 and continuous random variables in Chapter 6.

Some general notes are:

- random variables will be denoted by capital letters, such as _____
- outcomes of random variables are represented with small letters, such as _____

Section 5.2 Probability Functions (UE 2.2.1)

A **discrete** random variable, X , is a random variable with a finite number of possible outcomes. The probability distribution of a discrete random variable is the list of outcomes and their probabilities:

Value of X	x_1	x_2	...	x_n
Probability	p_1	p_2	...	p_n

The **probability function** for a discrete random variable X is

$$f(x) = P(X = x).$$

A probability function has two properties:

1. $0 \leq f(x) \leq 1$;
2. $\sum f(x) = 1$.

If X takes n values x_1, x_2, \dots, x_n , then $\sum f(x) = f(x_1) + f(x_2) + \dots + f(x_n) = 1$.

- A **probability stick graph** (or bar graph) can be used to display the distribution for a discrete random variable.
- The x-axis represents the outcomes.
- The y-axis is the probabilities of the outcomes.

Example: Toys

The probability distribution for the discrete random variable $X =$ number of toys played with by children.

#toys	0	1	2	3	4	5
Prob	0.03	0.16	0.30	0.23	0.17	

a. Graph the probability function for X .

b. What is the probability that a child will play with at least 3 toys?

c. Given the child has played with at least 3 toys, what is the probability that he/she will play with all 5 toys?

UE 2.3 Expected Values Involving a Single Random Variable

We consider computing means, variances and standard deviations of random variables.

Example 1: Suppose that there are 1000 people: 500 males and 500 females. The average age of the males is 24 and the average age of the females is 26. What is the mean age of the 1000 people?

Example 2: Suppose that there are 1000 people: 900 males and 100 females. The average age of the males is 24 and the average age of the females is 26. What is the mean age of the 1000 people?

Definition: **Mean of a discrete random variable**

Suppose that X is a discrete random variable whose distribution is:

Value of X	x_1	x_2	...	x_n
Probability	$f(x_1)$	$f(x_2)$...	$f(x_n)$

The **mean** (also called the **expected value**), denoted by $E(X)$ or μ_X , is

$$E(X) = \sum_x x f(x) = \sum_{i=1}^n x_i f(x_i) =$$

Example: Toys

The probability distribution for the discrete random variable $X =$ number of toys played with by children.

#toys	0	1	2	3	4	5
Prob	0.03	0.16	0.30	0.23	0.17	0.11

a. What is the mean number of toys played with?

Expectation of a Function of a Random Variable

Let X be a discrete r.v. with probability function $f(x)$.

1. If $Y = g(X)$ is another r.v., then the expected value of Y is

$$E[Y] = E[g(X)] = \sum_x g(x) f(x)$$

2. If $Y = g(X) = g_1(X) + g_2(X)$, then the expected value of Y is

$$E[Y] = E[g(X)] = \sum_x g(x) f(x) =$$

In words, the expected value of a sum of (functions of) r.v.'s is always the sum of the expected values.

Note: In general, $E[g(X)] \neq g[E(X)]$.

Example: X = number of toys played with by children.

b. What is the expected value of X^2 ?

Rules for mean: If X is a r.v. and a and c are constants, then

1. $E(c) =$
2. $E(cX) =$
3. $E(a + cX) =$

Variance and Standard Deviation of a r.v.

The **variance**, denoted by $\text{var}(X)$ or σ_X^2 or σ^2 , is the expected value of $g(X) = [X - E(X)]^2 = (X - \mu)^2$,

$$\text{var}(X) = \sigma^2 = E[(X - \mu)^2] = \sum_x (x - \mu)^2 f(x)$$

The **standard deviation**, $\text{sd}(X)$ or σ_X , is the square root of variance,

A useful formula for variance

$$\text{var}(X) = E(X^2) - [E(X)]^2 = E(X^2) - \mu^2$$

Proof.

Example: X = number of toys played with by children.

c. What are the variance and standard deviation for the number of toys played with?

Rules for the variance and s.d.: If X is a r.v. and a and c are constants, then

1. $\text{var}(c) =$
2. $\text{var}(cX) =$
3. $\text{var}(a + cX) =$

Note: Compare the rules for sample mean and sample s.d. with linear transformations (in Chapter 2).

Section 5.3 The Binomial Distribution

Background:

- n = fixed number of observations
- n observations are _____
- two outcomes: _____ and _____
- probability of success on any outcome =

Then X = the number of *successes* in a sample of size n has the _____ distribution with parameters n and p . Denoted by

Think about it:

What are the possible values for X ? Is X a discrete or continuous random variable?

We will learn how to recognize the binomial setting, study the binomial distribution and how to find binomial probabilities.

Example: Is this a binomial setting?

- a. Observe the gender of the next 50 children born at a local hospital;

$$X = \text{number of girls}$$

- b. Select one car from each hour's production for inspection

$$X = \text{number of finish defects in paint}$$

Note about the same probability p : if we sample without replacement, then removing say a success item on the first selection will truly change the probability p of a success for the next selected item — but it is negligible if the population is large relative to the sample.

Rule of Thumb: population at least 10 times as large as the sample => ok!

The Binomial Formula

We will develop the formula together using our probability knowledge.

Example:

According to government data, 25% of employed women have never been married.

If 10 employed women are selected at random, what is the probability that ...

... all 10 women will have never been married?

... all 10 women will have been married (0 have never been married)?

... just 1 woman will have never been married?

This is the **binomial distribution**:

$$P(X = k) =$$

for $k = 0, 1, 2, \dots, n$

where

Example: $n = 10$ (employed women) and $p = 0.25$ (proportion never married)

Think of it $\binom{n}{k}$ as: suppose you had n friends, how many ways could you invite k to dinner? The ones on the "ends" are easy to do without the formula ...

1. $\binom{10}{0} =$

2. $\binom{10}{10} =$

3. $\binom{10}{1} =$

4. $\binom{10}{9} =$

5. What is the probability of getting exactly 1 success out of 10 independent trials?

6. What is the probability of getting exactly 2 success out of 10 independent trials?

7. What is the probability of getting at least 1 success out of 10 independent trials?

The Easier Way: Use the Binomial Distribution Table (Appendix A2)

For $n = 2, 3, \dots, 12, 15, 20$ and $p = .01, .05, .10, \dots, .95, .99$.

Try question 5 above: $n = 10, p = 0.25, k = 1$ probability =

Try question 6 above: $n = 10, p = 0.25, k = 2$ probability =

Could actually give the **full binomial probability distribution** for $n = 10, p = 0.25$

X	0	1	2	3	4	5	6	7	8	9	10
$prob$.056	.188	.282	.250	.146	.058	.016	.003	.000	.000	.000

Could graph it too!

Recall we have that 25% of employed women have never been married. If we were to take a random sample of 10 employed women, how many would you **expect** to have never been married?

You just computed the **mean of a binomial distribution**.

More formally: If X has the binomial distribution $B(n, p)$, then

Mean of X is given by

Standard Deviation of X is given by

In general, you can use table to do Exercises, but try at least 1 or 2 with formula (what if the n and/or p is not in the table?) If n is small, will have to do with formula, if n is large, can do an approximation (in Chapter 6).

Example: Lie Detector Tests – given to truthful people have a probability of 0.2 of indicating a person is lying. Suppose we ask 12 job applicants about thefts from previous employers and that all 12 answer truthfully.

a. What is the probability that the test says at least 1 person is being deceptive?

b. What is the mean and standard deviation for the number of people classified as deceptive?