Chapter 3: Distributions of Random Variables (Part 1)

Day 6 topics:
Section 3.1: Random variables
Section 3.2: Binomial distribution

3.1. **Random variables.**

Definition 3.1. *A* ***random variable (r.v.)*** *assigns numerical values to the outcome of a random phenomenon.*

Notation:
A random variable is usually denoted with a capital letter such as X, Y, or Z.

Example 3.2. ***Data points***
Suppose you have a dataset of size 3 ($n = 3$) with $x_1 = 5, x_2 = 3$, and $x_3 = 6$.

- *Each data point is the outcome of a random phenomenon*
- *Each data point is a numerical value*
- *The data points are examples of values of a sequence of random variables X_1, X_2, and X_3*
- *For datasets, we almost always assume the data points came from random variables that are independent and have the same* ***distribution****.*
- *To calculate the likelihood of data, we need to know the distribution of the random variable that models the data.*
- *First, let's remind ourselves how to calculate them mean and variance of a dataset:*
  - *What is the mean of the data points?*
  - *What are the variance and standard deviation of the data points?*

Example 3.3. ***Rolling a die***
Suppose you roll a fair die. Let the random variable (r.v.) X be the outcome of the roll, i.e. the value of the face showing on the die.

(1) *What is the probability distribution of the r.v. X?*

(2) *What is the expected outcome of the r.v. X?*

(3) *Now suppose the 6-sided die is not fair. How would we calculate the expected outcome?*

x	$\mathbb{P}(X = x)$
1	0.10
2	0.20
3	0.05
4	0.05
5	0.25
6	0.35

(From Textbook § 2.1.5)

Definition 3.4. *A* ***probability distribution*** *consists of all disjoint outcomes and their associated probabilities.*

Rules for a probability distribution
A probability distribution is a list of all possible outcomes and their associated probabilities that satisfies three rules:

(1) The outcomes listed must be disjoint.
(2) Each probability must be between 0 and 1.
(3) The probabilities must total to 1.

Probability distributions are usually either **discrete** or **continuous**, depending on whether the random variable is discrete or continuous.

(Back to Textbook § 3.1.1)

Definition 3.5. *A* ***discrete*** *r.v. X takes on a finite number of values or countably infinite number of possible values.*

Definition 3.6. *A* ***continuous*** *r.v. X can take on any real value in an interval of values or unions of intervals.*

§ 3.1.2 Expectation

- We call the mean of a random variable its **expected value**
- The expected value is calculated as a weighted average

Definition 3.7. ***Expected value*** *of a discrete random variable*
If X takes on outcomes x_1, ..., x_k with probabilities $P(X = x_1)$, ..., $P(X = x_k)$, the expected value of X is the sum of each outcome multiplied by its corresponding probability:

§ 3.1.3 Variability of random variables

- Just like with data, the variability of a r.v. is described with its variance or standard deviation.
- Squared deviations from the mean are weighted by their respective probabilities

Definition 3.8. ***Variance*** *of a discrete random variable*
If X takes on outcomes x_1, ..., x_k with probabilities $P(X = x_1), \ldots, P(X = x_k)$ and expected value $\mu = E(X)$, then the variance of X, denoted by Var(X) or σ^2, is

Definition 3.9. ***Standard deviation*** *of a discrete random variable*
The standard deviation of X, labeled $SD(X)$ or σ, is

Example 3.10. ***Rolling a fair die: variance***

Suppose you roll a fair 6-sided die. Let the random variable (r.v.) X be the outcome of the roll, i.e. the value of the face showing on the die.

Find the variance and standard deviation of X.

x	$\mathbb{P}(X = x)$
1	*1/6*
2	*1/6*
3	*1/6*
4	*1/6*
5	*1/6*
6	*1/6*

Example 3.11. ***Vaccinated people testing positive for Covid-19***
About 25% of people that test positive for Covid-19 are vaccinated for Covid-19. Define the r.v. X to be 1 if someone that tests positive is vaccinated and 0 if they are not vaccinated.

(1) *Make a table for the probability distribution for the r.v. X*

(2) *What is the expected value of X?*

(3) *What is the variance of X?*

§ 3.1.4 Linear combinations of random variables

Definition 3.12. ***Linear combinations*** *of random variables.*
If X and Y are random variables and a and b are constants, then

is a linear combination of the random variables.

Theorem 3.13. ***Expected value of a linear combination*** *of random variables.*
If X and Y are random variables and a and b are constants, then

Example 3.14. ***Expected money for rolling 3 dice***
Let the random variables X_1, X_2, X_3 be the values shown on 3 fair 6-sided dice rolls. Suppose you are given in dollars the amount of the first roll, plus twice the value of the second roll, plus 4 times the value of the third roll.
How much money do you expect to get?

Theorem 3.15. ***Variance of a linear combination*** *of random variables.*
If X and Y are INDEPENDENT random variables and a and b are constants, then

Example 3.16. ***Variance of money for rolling 3 dice***
Let the random variables X_1, X_2, X_3 be the values shown on 3 fair 6-sided dice rolls. Suppose you are given in dollars the amount of the first roll, plus twice the value of the second roll, plus 4 times the value of the third roll.
What are the variance and standard deviation of the amount you get from the 3 rolls?

Example 3.17. ***Vaccinated people testing positive for Covid-19 (revisited)***
About 25% of people that test positive for Covid-19 are vaccinated for Covid-19. Define the r.v. X_i to be 1 if someone that tests positive is vaccinated and 0 if they are not vaccinated.
Suppose 3 people have tested positive for Covid-19 (independently of each other). Let T denote the number of people that are vaccinated amongst the 3 that tested positive.

(1) *Using the r.v.'s X_i, write a mathematical equation for calculating T.*

(2) *What is the expected value of T?*

(3) *What is the variance of T?*

(4) *What is the probability distribution of T?*

3.2. **Binomial distribution.**

- Many situations involve modeling independent random events that have 2 possible outcomes (binary), such as
  - Repeatedly flipping a coin
  - Whether a person that tested positive with Covid-19 is vaccinated or not
- Repeated events are referred to as **trials**
- The 2 possible outcomes are referred to as **successes** and **failures**.
- We denote the probability of a success as p.
- We denote the probability of a failure as $q = 1 - p$.

3.2.1. *Bernoulli distribution.*

Definition 3.18. ***Bernoulli*** *random variable.*
If X is a random variable that takes value 1 with probability of success p and 0 with probability $1 - p$, then X is a Bernoulli random variable.

- We call the probability of success p the **parameter** of the Bernoulli distribution.
- Each value of p identifies a specific Bernoulli distribution out of the **family** of Bernoulli r.v.'s where p is any value between 0 and 1 (inclusive).
- If a r.v. X is modeled by a Bernoulli distribution, then we write in shorthand

Theorem 3.19. *Mean and SD of a Bernoulli r.v.*
If X is a Bernoulli r.v. with probability of success p, then

3.2.2. *Binomial distribution.*

Recall Example 3.17 from Day 7:

- About 25% of people that test positive for Covid-19 are vaccinated for Covid-19.
- Define the r.v. X_i to be 1 if someone that tests positive is vaccinated and 0 if they are not vaccinated.
- Suppose 3 people have tested positive for Covid-19 (independently of each other).
- Let T denote the number of people that are vaccinated amongst the 3 that tested positive.

The random variable T above is an example of a Binomial random variable.

In general, a random variable X is **Binomial** if the following hold:

(1) The trials are independent.
(2) The number of trials, n, is fixed.
(3) Each trial outcome can be classified as a *success* or *failure.*
(4) The probability of a success, p, is the same for each trial.
(5) The r.v. X is the total number of successes in the n trials.

Definition 3.20. *Distribution of a **Binomial** random variable.*
Let X be the total number of successes in n independent trials, each with probability p of a success.
Then probability of observing exactly k successes in n independent trials is

- The parameters of a binomial distribution are p and n.
- If a r.v. X is modeled by a binomial distribution, then we write in shorthand

Theorem 3.21. *Mean and SD of a Binomial r.v.*
If X is a binomial r.v. with probability of success p, then

Example 3.22. ***Vaccinated people testing positive for Covid-19 (revisited)*** *About 25% of people that test positive for Covid-19 are vaccinated for Covid-19. Suppose 10 people have tested positive for Covid-19 (independently of each other). Let X denote the number of people that are vaccinated amongst the 10 that tested positive.*

(1) *What is the expected value of X?*

(2) *What is the SD of X?*

(3) *What is the probability that exactly 4 of the 10 people that tested positive are vaccinated?*

(4) *What is the probability that at most 3 of the 10 people that tested positive are vaccinated?*

(5) *What is the probability that at least 5 of the 10 people that tested positive are vaccinated?*