MATH 460: Probability & Statistics January_2019

Ch 2 Probability Distribution Functions ∪ Ch 3 Expected Values

Ch 4 Special Distributions

Class 3 Notes **Dr. Basilio**

Mon Jan_14 ∪ Tues Jan_15

* * *

Guiding Question(s)

(1) What are random variables?

(2) What's the difference between discrete and continuous variables?

(3) What's a Probability Distribution?

(4) What's the Expected Value of a random variable?

(5) How can we mathematically describe a "Fair Game"?

(6) What's a Binomial Distribution?

Chapter 5: Sampling Theory

A bit more practice with topics from Class 2:

Activity 1: 1-Var Stats

Let $S = \{123, 100, 111, 124, 132, 154, 132, 160\}$ be our data set. Find:

(a) Mean, Median, and Mode

(b) Standard Deviation

(c) What does the standard deviation mean in this case?

Activity 2: Five-Number-Summary

(a) Find the five number summary, and draw a Box-Whisker plot for $S = \{42, 20, 31, 10, 5, 3, 2, 1, 67, 53, 44\}$.

(b) Find the standard deviation for the set from part (a).

Chapter 2: Random Variables and Probability Distributions

Discrete vs Continuous Variables

Definition 1: Discrete-vs-Continuous-Variable

- **Variable:** a function defined on the sample space. That is, given any event A from a sample space S, a variable assigns a number to the event A. Using function notation, we write this as $X(A)$.
- **Discrete variable:** a variable that can attain only specific (whole number) values. Example: think the values of the roll of a dice.
- **Continous variable:** a variable can attain infinitely many values over a certain span or range. Example: the height of a person.
- **RANDOM variable:** a variable defined on a sample space that is comprised of a random process or experiment–that is, an experiment where you don't know what the outcome is until it is completed. Example: flipping a coin is a random experiment.

Example 1: Random Variables

(a) Let $S = \{HH, HT, TH, TT\}$ the sample space of flipping a coin twice. Let X be the random variable that assigns the number of heads that comes up. Then $X(\{HH\}) = 2$, $X(\{HT\}) = 1$, $X(\{TH\}) = 1$, $X(\{TT\}) = 0$.
Notice X is a discrete random variable.

(b) Let X be the random variable that assigns the number facing up when rolling a fair dice. Then $X(\{\text{ rolling a } 5\}) = 5$.
Notice X is a discrete random variable.

(c) Let X be the random variable that assigns the number of inches of rain collected at NASA headquarters.
Notice X is a continuous random variable.

It might appear that a random variable is the same as an output. Although they are not the same, what a random variable does is *quantify the outcome*!

Probability Distributions

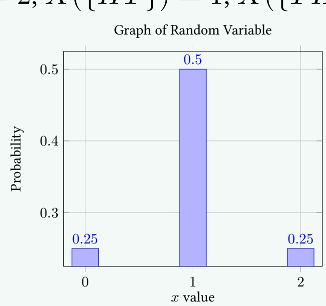
Definition 2: Discrete-vs-Continuous-Variable

- **Probability Distribution:** a description that gives the probability for each value of a random variable. Often expressed as a table, formula, or graph.

Example 2: Probability Distribution

Let $S = \{HH, HT, TH, TT\}$ the sample space of flipping a coin twice. Let X be the random variable that assigns the number of heads that comes up. Then $X(\{HH\}) = 2$, $X(\{HT\}) = 1$, $X(\{TH\}) = 1$, $X(\{TT\}) = 0$.
Notice X is a discrete random variable.

x	$P(X = x)$
0	1/4 = 0.25
1	1/2 = 0.5
2	1/4 =0.25



Activity 3: Probability Distribution

Find the probability distribution for rolling a dice. Let X be the random variable of rolling a dice. Plot a bar graph for the probability distribution.

x	$P(X = x)$
1	
2	
3	
4	
5	
6	

Activity 3 shows a **uniform distribution** since each outcome is equally likely.

Activity 4: Probability Distribution

Suppose that a dice is to be tossed twice, and let the random variable X denote the sum of the two tosses. Find the probability distribution for X. Plot a bar graph for the probability distribution.

x	$P(X = x)$
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	

Activity 5: Probability Distribution

An urn holds 4 red marbles and 6 black marbles. If 2 marbles are to be drawn at random without replacement and X denotes the number of red marbles, find the probability distribution for X.

Hint: $S = \{RR, RB, BR, BB\}$.

Definition 3: Expectation

- A very important concept in probability and statistics is that of the mathematical expectation, expected value, or briefly the **expectation**, of a random variable. Expected value uses probability to tell us what outcomes to expect in the long run.
- For a discrete random variable X having the possible values $x_1, x_2, \ldots, x_n$, the expectation of X is defined as

$$E(X) = x_1 P(X = x_1) + x_n P(X = x_n) + \cdots + x_n P(X = x_n) = \sum_{i=1}^{n} x_i P(X = x_i) \tag{1}$$

- As a special case of (2), where the probabilities are all equal, we have

$$E(X) = \frac{x_1 + x_2 + \cdots + x_n}{n} \tag{2}$$

  which is called the **arithmetic mean**, or simply, the **mean** of the $x_1, x_2, \ldots, x_n$.
- The expectation of X is very often called the mean of X and is denoted by μ_X, or simply μ, when the particular random variable is understood.
- The mean, or expectation, of X gives a single value that acts as a representative or average of the values of X, and for this reason it is often called a **measure of central tendency**.
- **FAIR GAME:** When the expected value of a game (random variable) equals 0. This is interpreted as: in the long run, you can expect to win and lose the same amount of money.

Example 3: Expectation

(a) Alex is a basketball player who makes 50% of his 2-point shots and 20% of his 3-point shots. If $X =$ making a 2-point shot, then his expectation in the long term is $E(X) = 2 \cdot 0.5 = 1$ point since the outcome is $x = 2$ and the probability is $p = 0.5$. If $X =$ making a 3-point shot, then his expectation in the long term is $E(X) = 3 \cdot 0.2 = 0.6$ points since the outcome is $x = 3$ and the probability is $p = 0.2$. This is confusing to many people at first since you can't score 0.6 points! But this is because we are taking an average!

(b) If marks of five students is given to be 65, 76, 88, 34, and 90, then the expected mark for a random student is

$$E(X) = \frac{65 + 76 + 88 + 34 + 90}{5} = 70.6$$

Activity 6: Expectation

Suppose that a game is to be played with a single die assumed fair. In this game a player wins $20 if a 2 turns up, $40 if a 4 turns up; loses $30 if a 6 turns up; while the player neither wins nor loses if any other face turns up.

(a) State what the random variable X is

(b) Find all the outcomes $x_1, \ldots, x_6$

(c) Find all the probabilities for each respective outcome

(d) Find the expected sum of money to be won (or lost).

(e) In a fair game, what do you think is a reasonable buy-in is in order to play the game?

Activity 7: Expectation

A game is played where a player rolls a six sided die and if the result is an even number, they win 4 times the number in dollars, but if the result is odd, they lose 6 times the number in dollars. Find the expected winnings (or losings).

(a) Find the expected winnings (or losings).

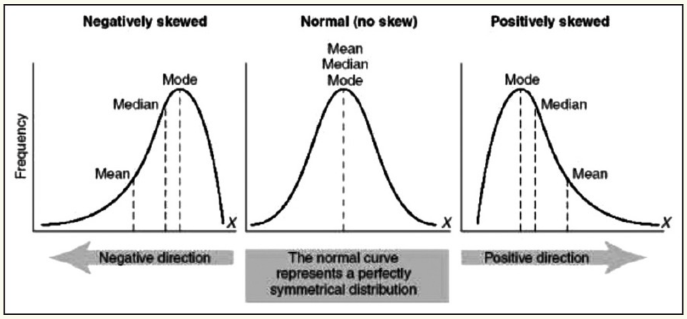
(b) Even if the game is free, should you play?

Skewness

Definition 4: Skewness

Skewness is asymmetry in a statistical distribution, in which the histogram (or curve) appears distorted or skewed either to the left or to the right.

- **Positively Skewed:** the "tail" of the distribution is to the right of the mean, or the mean is greater than the median and mode ("hump")
- **Negatively Skewed:** the "tail" of the distribution is to the left of the mean, or the mean is smaller than the median and mode ("hump")



Activity 8: Frequency-Skewness

The following is a list of prices (in dollars) of birthday cards found in various drug stores:

1.45	2.20	0.75	1.23	1.25
1.25	3.09	1.99	2.00	0.78
1.32	2.25	3.15	3.85	0.52
0.99	1.38	1.75	1.22	1.75

(a) Organize this data with intervals of 50 cents (i.e. .50-0.99, 1.00-0.49, and so on) using create a frequency distribution table.

(b) Draw a Histogram of the data. State the skewness of the data.

Binomial Distribution

Definition 5: Binomial-Probability-Distribution

For a **Binomial probability distribution** it is important that we have a random process or experiment and must satisfy:

1. The experiment has a fixed number of trails
2. The trials must be **independent** (that is, the outcome of any individual trial doesn't effect the probabilities of other trials)
3. Each trail must have all outcomes classified into exactly two categories: *success* and *failure*
4. The probability of success remains the same in all trails

Let $P(X = x)$ denote the **probability of exactly x successful trails out n in a binomial probability distribution**, then

$$\boxed{P(X = x) = \binom{n}{x} p^x q^{n-x} =_n C_x p^x q^{n-x}} \tag{3}$$

where

- n be the total number of trails run in the experiment
- X be a random variable of a single "successful" trail
- p be the probability of the successful trail X
- q be the probability of trail X failing. (NOTE: $p + q = 1$, or $q = 1 - p$)
- x be the number of successful trials of X. So notice that x can take values from 0 up to n, i.e. $x = 0, 1, 2, 3, \ldots, n$.

Recall that: $\binom{n}{x} =_n C_x = \frac{n!}{x!(n-x)!}$.

Example 4: Binomial-Distribution-Probability

A die is tossed 3 times. What is the probability of

(a) No fives turning up?

Solution: Let $n = 3$. Let $X = \{$ rolling a 5 $\}$. The probability of X being successful is $p = 1/6$ and $q = 5/6$ is probability of X failing. We want X to be successful zero times! so $x = 0$:

$$\begin{aligned} P(X = 0) &= \binom{3}{0} \left(\frac{1}{6}\right)^0 \left(\frac{5}{6}\right)^{3-0} \\ &= \left(\frac{3!}{0!(3-0)!}\right) \cdot 1 \cdot \left(\frac{5}{6}\right)^3 \\ &= \left(\frac{5}{6}\right)^3 = \frac{125}{216} \approx 0.5787037 \approx 0.579 \end{aligned}$$

So the probability of no five turning up if a die is tossed 3 times is approximately 57.9%. □

(b) 1 five turning up?

Solution: Let $n = 3$. Let $X = \{$ rolling a 5 $\}$. The probability of X being successful is $p = 1/6$ and $q = 5/6$ is probability of X failing. We want X to be successful once so $x = 1$:

$$\begin{aligned} P(X=1) &= \binom{3}{1}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^{3-1} \\ &= \left(\frac{3!}{1!(3-1)!}\right)\cdot\frac{1}{6}\cdot\left(\frac{5}{6}\right)^2 \\ &= 3\cdot\frac{1}{6}\cdot\frac{25}{36} \\ &= \frac{25}{72} \approx 0.347\bar{2} \approx 0.3472 \end{aligned}$$

So the probability of one five turning up if a die is tossed 3 times is approximately 34.7%. □

(c) 3 fives turning up?

Solution: n, X, p, and q are exactly as in part (a) and (b), but this time we want $x = 3$:

$$\begin{aligned} P(X=3) &= \binom{3}{3}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^{3-3} \\ &= \frac{1}{216} \approx 0.00462963 \approx 0.0046 \end{aligned}$$

So the probability of three fives turning up if a die is tossed 3 times is approximately 0.5%. □

Definition 6: Binompdf-vs-Binomcdf

USING CALCULATOR TI83: binompdf(n,p,x)

$\boxed{DIST}$ key in yellow ($\boxed{2nd}$ > $\boxed{VARS}$) > Scroll to 10 “binompdf” or scroll to A “binomcdf”

- Binompdf is when we want exactly x trails to be successful so this is binomial distribution pdf. Thus this is 1-valued random variables.
- Binomcdf is when we want multiple values of x to be true. It is defined as

$$\text{binomcdf}(n, p, x) = P(X \leq x) \tag{4}$$

Notice the sneaky “$\leq$” less than or equal to sign in the binomcdf. This means:

$$\begin{aligned} \text{binomcdf}(n, p, x) &= P(X \leq x) = P(X = 0, 1, 2, \ldots, x) \\ &= P(X = 0) + P(X = 1) + P(X = 2) + \cdots + P(X = x) \end{aligned}$$

This can help us with “at most” and “at least” type of problems.

Example 5: Binomcdf-probability

What is the probability of at least four successful trials out of a total of 6 trails in a random experiment, with probability of success of a single trial being 25%?

Solution: Here $n = 6$ and $p = 0.25$. Notice we must convert the percentage to a decimal. We don't need to know what X is. We want at least four successful trials, so $x = 4$ is when exactly 4 trials are successful. When $x = 5$ is when exactly 5 trials are successful, and $x = 6$ is when exactly 6 trials are successful.

One way to find the answer is: $P(X = 4) + P(X = 5) + P(X = 6)$. This is alot to type into the calculator.

ANOTHER WAY (FASTER): we can use binomcdf! Because binomcdf(6,0.25,3) calculates $P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)$ quickly we can use this to find the remaining probability with the "1 minus trick:"

$$\begin{aligned}
\underbrace{P(X \geq 4)}_{\text{at least 4 successful trials}} &= 1 - P(X < 4) \qquad (\text{because } P(X < 4) + P(X \geq 4) = 1) \\
&= 1 - P(X \leq 3) \qquad (\text{because } P(X < 4) = P(X \geq 3)) \\
&= 1 - \text{binomcdf}(6, 0.25, 3) = 1 - 0.96240234375 = 0.03759765625
\end{aligned}$$

So the probability of at least 4 successful trials is approximately 3.76%. □

Activity 9: Binomial-Distribution-Probability

For each part, please label the n, X, and p in addition to your work and answer. Leave answers as decimals and round to three decimal places.

Find the probability that in tossing a fair coin three times, there will appear

(a) three heads

(b) two tails and a head

(c) at least one head

(d) not more than one tail

Activity 10: Binomial-Distribution-Probability

For each part, please label the n, X, and p in addition to your work and answer. Give answers as percentages and round to one decimal place.

Find the probability that in five tosses of a fair die, a 3 will appear

(a) twice

(b) at most once

(c) at least two times