MATH 460: Probability & Statistics — September 2018

Chapter 1: Basic Probability & Chapter 5: Sampling Theory

Class Notes — **Dr. Basilio**

Southern California Institute of Technology

Monday 9.10.2018

* * *

Guiding Question(s)

(1) What does it mean for events to be "certain" or "uncertain"?

(2) How can we systematically study uncertainty?

(3) What does it mean to be "average"?

(4) How can we apply this to our everyday lives?

Concept of Probability

Definition 1: Concept of Probability

In any random experiment there is always uncertainty as to whether a particular event will or will not occur. As a measure of the **chance**, or **probability**, with which we can expect the event to occur, it is convenient to assign a number between 0 and 1. If we are sure or certain that the event will occur, we say that its probability is 100%, or 1, but if we are sure that the event will not occur, we say that its probability is zero. If, for example, the probability is $\frac{1}{4}$, we would say that there is a 25% chance it will occur and a 75% chance that it will not occur. Equivalently, we can say that the odds against its occurrence are 75% to 25%, or 3 to 1.

There are two important procedures by means of which we can estimate the probability of an event.

- **CLASSICAL APPROACH:** If an event can occur in h different ways out of a total number of n possible ways, all of which are equally likely, then the probability of the event is $\frac{h}{n}$.
- **FREQUENCY APPROACH:** If after n repetitions of an experiment, where n is very large, an event is observed to occur in h of these, then the probability of the event is $\frac{h}{n}$. This is also called the **empirical probability** of the event.

Example 1: Probability

(a) CLASSICAL APPROACH:
Suppose we want to know the probability that a head will turn up in a single toss of a coin. Since there are two equally likely ways in which the coin can come up?namely, heads and tails (assuming it does not roll away or stand on its edge)?and of these two ways a head can arise in only one way, we reason that the required probability is $\frac{1}{2}$. In arriving at this, we assume that the coin is fair, i.e., not loaded in any way.

(b) FREQUENCY APPROACH:
If we toss a coin 1000 times and find that it comes up heads 532 times, we estimate the probability of a head coming up to be $\frac{532}{1000}$, or 0.532.

Activity 1: Frequency Approach

Pair up into a group of 2 (or 3) students. One student will flip a coin and the other will record the results.

(a) Flip a coin ten times. What is the empirical probability from your experiment?

(b) Flip a coin 50 times. What is the empirical probability from your experiment?

Axioms of Probability

Both the classical and frequency approaches have serious drawbacks, the first because the words "equally likely" are vague and the second because the "large number" involved is vague. Also, it would be very impractical to have to conduct long experiments to determine probabilities. Because of these difficulties, mathematicians have been led to an axiomatic approach to probability.

Definition 2: Concept of Probability

Suppose we have a sample space S.
To each event A, we associate a real number $P(A)$. Then P is called a **probability function**, and $P(A)$ the **probability of the event** A, if the following axioms are satisfied:

- **Axiom 1** For every event A,
$$P(A) \geq 0 \tag{1}$$
- **Axiom 2** For the sample space S,
$$P(S) = 1 \tag{2}$$
- **Axiom 3** If two events A and B are mutually exclusive,
$$P(A \cup B) = P(A) + P(B) \tag{3}$$

Some observations:

- For any event A, the probability is between 0 and 1: $0 \leq P(A) \leq 1$. Of course, $P(A) = 0$ are for impossible events and $P(A) = 1$ are for events that are certain to happen.
- Because the sample space S consists of all possible outcomes for a chance experiment, in the long run an outcome that is in S must occur 100% of the time. This means that $P(S) = 1$.
- Axiom 3 can be applied to any number of mutually exclusive events. If $A_1, A_2, A_3, \ldots$ are all mutually exclusive events (that is $A_i \cap A_j = \emptyset$ for any two indices i and j) then
$$P(A_1 \cup A_2 \cup A_3 \cup \cdots) = P(A_1) + P(A_2) + P(A_3) + \cdots \tag{4}$$
For example, if we roll a fair dice and A_1 is the event of rolling a 1, A_2 is the event of rolling a 2, etc then there are 6 mutually exclusive events and the sample space is $S = A_1 \cup A_2 \cup A_3 \cup \cdots \cup A_6$. Then
$$1 = P(S) = P(A_1 \cup A_2 \cup A_3 \cup \cdots \cup A_6) = P(A_1) + P(A_2) + P(A_3) + \cdots + P(A_6). \tag{5}$$

Theorem 1: Theorems of Probability

For any event A,

(a) $P(A) + P(A') = 1$
This one makes sense since in words it says: any event either occurs or doesn't occur!

(b) $P(A') = 1 - P(A)$
This follows from Property (a) by subtracting $P(A)$ from both sides and noting that $A' = S - A$.

(c) If A and B are two events (not necessarily disjoint), then

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{6}$$

Activity 2: Probability

Suppose A and B are two disjoint events in a sample space S and that $P(A) = .22$; $P(B) = .33$. Calculate the following probabilities.

(a) $P(A \cup B)$

(b) $P(A \cap B)$

(c) $P(A' \cup B)$

(d) $P(A' \cap B)$

(e) $P(A - B)$

Activity 3: Probability

A ball is drawn at random from a box containing 6 red balls, 4 white balls, and 5 blue balls. Determine the probability that it is (a) red, (b) white, (c) blue, (d) not red, (e) red or white.

Activity 4: Probability

Determine the probability for the following events.

(a) Roll a 7 or 11 from a pair of fair 6 sided dice.

(b) A non-defective television found next if out of 2,300 televisions already examined, 15 were defective.

(c) At least 1 tails appears in 5 tosses of a fair coin.

(d) The probability of drawing an ace or a club in a standard deck.

Definition 3: Conditional Probability

Let A and B be two events such that $P(A) > 0$. Denote by $P(B|A)$ the **probability of B given that A has occurred**. Since A is known to have occurred, it becomes the new sample space replacing the original S. From this we are led to the definition

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \tag{7}$$

or, equivalently,

$$P(B \cap A) = P(A) \cdot P(B|A) \tag{8}$$

In words, (11) says that the probability that both A and B occur is equal to the probability that A occurs times the probability that B occurs given that A has occurred.

Example 2: Conditional Probability

Two cards are drawn from a well-shuffled ordinary deck of 52 cards. Find the probability that they are both aces if the first card is (a) replaced, (b) not replaced.

Let A be the event "ace on first draw" and B be the even "ace on second draw." Then we are looking for $P(A \cap B)$.
(a) Since there are 4 aces in 52 cards, $P(A) = 4/52$. If the card is replaced (and the deck is re-shuffled) then $P(B|A) = P(B) = 4/52$ as well. Thus, using the formula

$$P(A \cap B) = P(A)P(B|A) = P(A)P(B) = \frac{4}{52} \cdot \frac{4}{52} = \frac{1}{169}.$$

(b) When the ace is not replaced, then there are 3 aces left and $P(B|A) = 3/51$ and

$$P(A \cap B) = P(A)P(B|A) = P(A)P(B) = \frac{4}{52} \cdot \frac{3}{51} = \frac{1}{121}.$$

Activity 5: Conditional Probability

(a) Five marbles are picked at random out of a jar containing 10 red marbles, 15 white marbles, 20 blue marbles, 25 orange marbles, and 30 purple marbles. What is the probability of picking one of each color, assuming you pick a marble one at a time?

(b) Drawing a king and a spades cards for first two draws from a well-shuffled 52 card deck.

Chapter 5: Sampling Theory

Organizing and Visualizing Data

Definition 4: Data

- **FREQUENCY DISTRIBUTION:** It is useful to organize or group the raw data. We arrange the data into classes or categories and determine the number of individuals belonging to each class, called the **class frequency**. When arranged as a table it is called a **frequency distribution** or **frequency table**.

Table 5-2
Heights of 100 Male Students at XYZ University

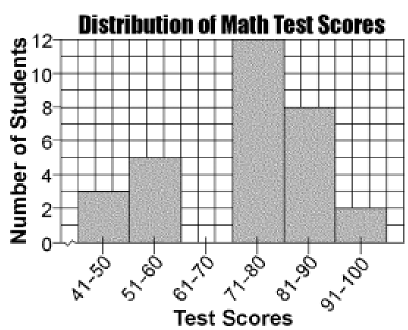
Height (inches)	Number of Students
60–62	5
63–65	18
66–68	42
69–71	27
72–74	8
TOTAL	100

- **RELATIVE FREQUENCY DISTRIBUTION:** If we recorded the relative frequency or **percentage** rather than the number in each class/category, the result would be a **relative, or percentage, frequency distribution**.

- **BAR GRAPHS and HISTOGRAMS:**

Activity 6: Frequency Distributions

The graph below shows the distribution of scores of 30 students on a mathematics test.



Complete the frequency table below using the data in the frequency histogram shown.

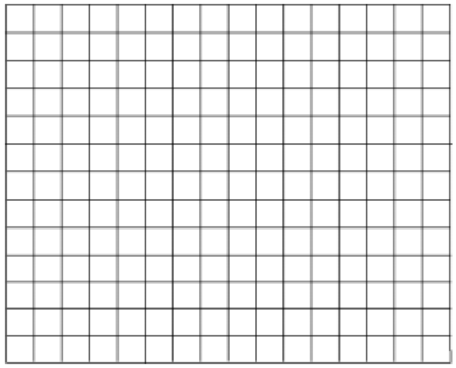
Test Scores	Frequency
91–100	
81–90	
71–80	
61–70	
51–60	
41–50	

(a)

Activity 7: Frequency Distributions

The scores on a mathematics test were 70, 55, 61, 80, 85, 72, 65, 40, 74, 68, and 84. Complete the accompanying table, and use the table to construct a frequency histogram for these scores.

Score	Tally	Frequency
40-49		
50-59		
60-69		
70-79		
80-89		

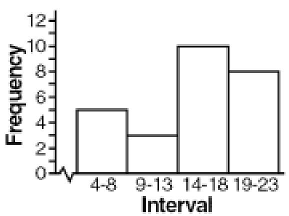


(a)

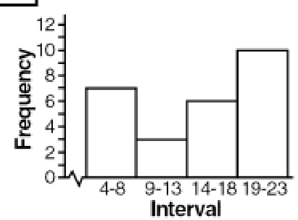
Which one of the following histograms represents the data in the table below?

Interval	Frequency
4-8	8
9-13	3
14-18	10
19-23	5

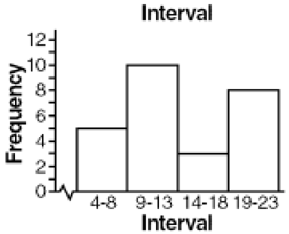
A)



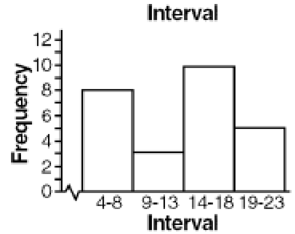
C)



B)



D)



(b)

Measurements of Central Tendency

Definition 5: Mean-Median-Mode-Range

- **MEAN:** the mean is computed by adding all of the numbers in the data together and dividing by the number elements contained in the data set.
- **MEDIAN:** The median of a data set is dependent on whether the number of elements in the data set is odd or even. First re-order the data set from the smallest to the largest then if the number of elements are odd, then the Median is the element in the middle of the data set. If the number of elements are even, then the median is the average of the two middle terms.
- **MODE:** The mode for a data set is the element that occurs the most often. It is not uncommon for a data set to have more than one mode. This happens when two or more elements occur with equal frequency in the data set. A data set with two modes is called **bimodal**. A data set with three modes is called **trimodal**.

Activity 8: Mean-Median-Mode-Range

(a) Consider the data set $S = \{2, 5, 9, 3, 5, 4, 7\}$. Compute the mean.

(b) Consider the data sets $A = \{2, 5, 9, 3, 5, 4, 7\}$ and $B = \{2, 5, 9, 3, 5, 4\}$. Compute the median of each data set. (Don't forget to re-order the data first!)

(c) Consider the data sets $A = \{2, 5, 9, 3, 5, 4, 7\}$, $B = \{2, 5, 2, 3, 5, 4, 7\}$, $C = \{2, 5, 2, 7, 5, 4, 7\}$. Compute the mode(s) of each data set.

(d) Consider the data set $S = \{2, 5, 9, 3, 5, 4, 7\}$. Compute the range of the data set.

Measurement of Dispersion

Definition 6: Standard-Deviation

A measure of how the values in a data set vary or deviate from the mean. Some notation to compute the standard deviation:

$$\sigma = \sqrt{\frac{\sum(x - \bar{x})^2}{n}} \tag{9}$$

- x: a value from a data set
- $\bar{x}$: mean
- n: number of values in a data set
- Σ: "to sum or add" (Capital Greek letter "Sigma")
- σ: standard deviation (lower-case Greek letter "sigma")

How to compute:
Step 1: Calculate mean
Step 2: Find the difference between the data value and the mean
Step 3: Square each difference
Step 4: Find the average (mean of these squares)
Step 5: Take the square root of the mean of the squares to find the standard deviation

Example 3: Standard Deviation

Data Set 1			
x	$\bar{x}$	$x - \bar{x}$	$(x - \bar{x})^2$
12.6	15	-2.4	5.76
15.1	15	0.1	0.01
11.2	15	-3.8	14.44
17.9	15	2.9	8.41
18.2	15	3.2	10.24
$\frac{\sum(x - \bar{x})^2}{n}$			7.772
Standard Deviation: $\sqrt{\frac{\sum(x-\bar{x})^2}{n}}$			≈ 2.79

Activity 9: Standard Deviation

(a)

Data Set 2			
x	$\bar{x}$	$x - \bar{x}$	$(x - \bar{x})^2$
13.4			
11.7			
18.3			
14.8			
14.3			
$\frac{\sum(x-\bar{x})^2}{n}$			
Standard Deviation: $\sqrt{\frac{\sum(x-\bar{x})^2}{n}}$			

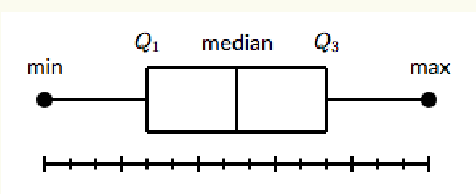
(b) Compare Data Set 1 and Data Set 2. Which set of data has a greater standard of deviation?

The data set with the larger standard of deviation has a larger more spread out range of values.

If many of the data values are close to the mean, then the data would have a relatively small standard deviation. This would tell you that the data is not very spread out.

Definition 7: Five-Number-Summary

- **RANGE:** The range for a data set is the difference between the largest value and smallest value contained in the data set. First reorder the data set from smallest to largest then subtract the first element from the last element.
- **FIVE NUMBER SUMMARY:** is the minimum, first quartile, median, third quartile, and maximum.
- **Box-Whisker Plot:** displays the five-number summary of a set of data. In a box plot, we draw a box from the first quartile to the third quartile. A vertical line goes through the box at the median. The whiskers go from each quartile to the minimum or maximum.



Activity 10: Five-Number-Summary

(a) Find the five number summary, and draw a Box-Whisker plot for $S = \{15, 25, 20, 29, 29, 36, 29, 15, 26, 28, 24, 25\}$.

(b) Find the standard deviation for the set from problem 4.