MATH 460: Probability & Statistics — September 2018

Chapter 5: Sampling Theory & Chapter 2: Probability Distribution Functions

Class Notes — **Dr. Basilio**

Southern California Institute of Technology

Wednesday 9.12.2018

* * *

Chapter 5: Sampling Theory

Measurements of Central Tendency

Definition 1: 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 <u>middle</u> 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**.

Measurement of Dispersion

Definition 2: 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:

$$\sigma = \sqrt{\frac{\sum (x - \bar{x})^2}{n}} \tag{1}$$

- 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

Activity 1: Standard Deviation

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

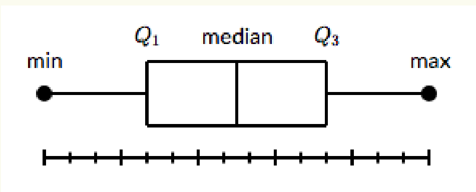
(a) Mean, Median, and Mode

(b) Standard Deviation

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

Definition 3: 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.
  - **FIRST QUARTILE** Q_1: it is a median of the data smaller than the median.
  - **THIRD QUARTILE** Q_3: it is a median of the data larger than the median.
- **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 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 problem 4.

Chapter 2: Random Variables

Discrete vs Continuous Variables

Definition 4: Discrete-vs-Continuous-Variable

- **Variable:** a function defined on the sample space. That is, given any event A from a sample space S, a random variable assigns a number to even A. Using function notation, we write this as $X(A)$.
- **Discrete variable:** a variable that can attain only specific 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 random process or experiment–that is, 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 $X(\{HH\}) = 2$, $X(\{HT\}) = 1$, $X(\{TH\}) = 1$, $X(\{TT\}) = 0$.

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

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

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*!
The random variables from Example 1 (a) and (b) are both discrete. The random variable from Example 1 (c) is continuous.

Chapter 4: Probability Distribution Functions

Binomial Distribution

Definition 5: Binomial-Distribution

For a Binomial distribution it is important that we have a random process or experiment and we will run the experiment many times but each trail must be an **independent trail** which means previous trails do not have any influence on future trails.

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

$$P(X = x) = \binom{n}{x} p^x q^{n-x} \tag{2}$$

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

Activity 3: 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 $x = 0$:

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

So the probability of no five turning up if a die is tossed 3 times is 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 $x = 1$:

$$\begin{aligned} P(X=0) &= \binom{3}{1}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^{3-1} \\ &= \left(\frac{3!}{1!(3-1)!}\right)\cdot\frac{1}{6}\cdot\left(\frac{5}{6}\right)^2 \\ &= 3\cdot\frac{1}{6}\cdot\frac{25}{36} \\ &= \frac{25}{72} = 0.347\bar{2} = 0.3472 \end{aligned}$$

So the probability of one five turning up if a die is tossed 3 times is 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 $x = 3$:

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

So the probability of three fives turning up if a die is tossed 3 times is 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

$$\text{binomcdf}(n, p, x) = P(X \leq x) \tag{3}$$

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

Activity 4: Binomcdf-probability

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

Solution: Here $n = 6$ and $p = 0.25$. 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:"

$$P(X \geq 4) = 1 - P(X < 4) = 1 - \text{binomcdf}(6, 0.25, 3) = 1 - 0.96240234375 = 0.03759765625$$

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

Activity 5: Binomial-Distribution-Probability

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 6: Binomial-Distribution-Probability

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