Chapter 1a

Appendix - Recap of Essential Math Concepts

Shaikh Tanvir Hossain

October 1, 2022

East West University

Outline

- In probability theory everything starts from a random experiment. Roughly an *experiment* is a process which we can perform in an identical enviroement as many times as want, we know all possible outcomes of the experiment but we never know which outcome will be realized.

- Typical examples of the experiment are, tossing a coin, tossing 2 coins, throwing a dice, taking a random sample from a population, etc.

- If we put all possible outcomes of an expemtiment in a set, then we say it's a sample space of the experiment.

Definition 1.1 (Sample Space)
The set of all possible outcomes of an experiment is called the sample space of an experiment.

- For example
  - ✓ for the experiment, tossing single coin, the sample space is $S = \{H, T\}$
  - ✓ for the experiment tossing two coins the sample space is $S = \{HH, HT, TH, TT\}$
  - ✓ for the experiment throwing a dice the sample space is $S = \{1, 2, 3, 4, 5, 6\}$
  - ✓ for the experiment throwing two dice the sample space is $S = \{(1, 2), (1, 3), (1, 4), (1, 5),$ $(1, 6), (2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6), (3, 1), (3, 2), (3, 3), \ldots, (6, 6)\}$, which is a set if 36 possible pairs.

Probability Theory Recap

- Once we have a sample space of an experiment, we can take a subset of the sample space, and any subset of the sample space is called an *event*, for example, if $S = \{HH, HT, TH, TT\}$, then $A = \{HH\}$ is a subset so it's an event, also $B = \{HH\}$ is a subset so it's also an event.

- We will see in a minute that probabilities are defined on events, or sets. So probability is a set function, this means the input or probability is always a set.

- Now, we can define a set of which will have *all possible events*, the sample space and null event and then we will get a huge set which looks like,

$$\mathcal{P} = \Big\{\{\}, \{HH\}, \{HT\}, \{TH\}, \{TT\}\}, \{HH, HT\}, \{HH, TH\},$$
$$\{HH, TT\}\}, \{HT, TH\}, \{HT, TT\}\}, \{TH, TT\}\}, \{HH, HT, TH\}, \{HH, HT, TT\}\},$$
$$\{HH, TH, TT\}\}, \{HT, TH, TT\}\}, \{HH, HT, TH, TT\}\}\Big\}$$

- And then intuitively it makes sense that in a random experiment we should be able to calculate probabilities for all events in $\mathcal{P}$.

- Note that $\mathcal{P}$ is *set of sets* or you can say a *set of events*, another fancy name of this in this case is σ-*algebra*, but you can ignore this name now.

- When you think about probability, possibly you think about the probability is a count of something happening when we do it many times, for some example if I toss coin 100 times, and I get 50 times head, then probability of appearing head is 0.5. This is called the *frequentist interpretation of probability*.
- If we define probability in this way, then although for this example this seems ok, but for very large sample sizes we might get into some troubles to define probability systematically (there are actually some paradoxical results.)
- Russian mathematician Andrey Kolmogorov in 1933 proposed an *axiomatic* definition of probability, and now this has become the standard definition of probability you will find almost everywhere.
- What is very nice about this is it solves all the paradoxes and it can also explain the frequentist interpretation of probability.
- What is this *axiomatic definition*, it simply tells us *what is probability*?
- Here is the axiomatic definition of probability

$P(A)$ is a set function such that

- *Axiom -1* - For every event A, $P(A) \geq 0$.
- *Axiom -2* - If S is the sample space then we will have $P(S) = 1$.
- *Axiom -3* - If we have events $A_1, A_2, A_3, \dots$, which are pairwise-disjoint and countable then

$$P(A_1 \cup A_2 \cup A_3 \dots) = P(A_1) + P(A_2) + P(A_3) + \dots$$

- The first should make sense.
- For the second one note that S is the set of all possible outcomes. So when we think about S, we think about at least one of the outcomes from the experiment will happen, now this will always happen, so $P(S) = 1$ should make sense as well.
- For the third one, you can think, if we first add some sets (meaning we take union of two sets) and then calculate the probabilities, this is same as calculating probabilities for each of these sets first and then adding the probabilities
- Throughout the course we will see the use of this definition many times.
- If this definition looks very abstract you can just take it now and don't worry too much about it, later you will understand how we are using it.
- The important things to take away from here are - probability is a set function, probability is a non-negative number, probability of the entire sample space is 1 and we

- Once we defined the probability, the next important topic for us is *a random variable*.
- The idea of the random variable comes because maybe we are not interested directly in the sample space but rather other kinds of information from an experiment.
- Random variable helps us to systematically do this, a random is a function where the input is coming from the sample space and the output will be a real number.
- The proper definition of random variable is beyond the score of this course, but simply you can think that when we write $X(s)$ is random variable, it is actually a function where s is coming from the sample space, and the value of the random variable is a real number.

Definition 1.2 (Random Variable (Informal))
A random variable is a function for which the domain (the input space) is the sample space and output is a real number.

- You have already seen many examples of random variables, for example, when $S = \{H, T\}$, then we can define a random variable with $X(H) = 1$, and $X(T) = 0$. So this random variable takes value 0 and 1
- When $S = \{(H, H), (H, T), (T, H), (T, T)\}$, then you can think about a random variable which counts the total number of heads, then X takes value 0, 1 and 2.

$$X(\omega) = \begin{cases} 0, & \text{if } s = (T, T) \\ 1, & \text{if } s \in \{(H, T), (T, H)\} \end{cases}$$

- If X is a (discrete) random variable taking values $10, 20$, and 30 with probabilities $P(X = 10) = 1/4$, $P(X = 20) = 1/4$ and $P(X = 30) = 2/4$, then expectation of X is

$$\begin{aligned}\mathbb{E}(X) &= (10 \times P(X = 10)) + (20 \times P(X = 20)) + (30 \times P(X = 30)) \\ &= (10 \times 1/4) + (20 \times 1/4) + (30 \times 2/4) \\ &= 90/4 = 22.5\end{aligned}$$

- We can write these probabilities as a function $f(10) = 1/4$, $f(20) = 1/4$, $f(30) = 2/4$, this function is called *probability density function* (in short we write PDF, also sometimes we call it just density) of X.
- Here $f(x)$ is the density function of X, when we have two or more random variables together we will write $f_X(x)$ to denote the density function of X.
- So the formula for the expectation can be written as

$$\mathbb{E}(X) = (10 \times f(10)) + (20 \times f(20)) + (30 \times f(30)) = 22.5$$

- So knowing PDF of X is same as knowing the distribution of X
- So we can say in general if X is a (discrete) random variable which takes values $x_1, x_2 \dots, x_k$, then the expectation of X will be

- *Very important point: Do not confuse* $x_1, x_2 \ldots, x_k$ with sample values when we do sampling and get $x_1, x_2 \ldots, x_n$. There is *no sampling* going on here.
- Here $x_1, x_2 \ldots, x_k$ are different values that the random variable X can take.
- For our example $k = 3$, and X takes 3 values, so we have $x_1 = 10$, $x_2 = 20$ and $x_3 = 30$.
- Now a question may arise how did we get these probabilities $1/4$, $1/4$ and $2/4$, which are $.25$, $.25$ and $.50$.
- One way to think about this is, these are *population probabilities.* Suppose you have the whole population of Bangladesh as a population and then if X is income this means 25% of the population have income 10 thousands, 25% of the population have income 20 thousands and remaining 50% have 30 thousands. These are all 100% population.
- This is why these probabilities have to sum to 1 (does it make sense to you?)
- Also note that, *Expectation* is a population object because to calculate expectation we need to know the true population probabilities, this is why we call this also *population mean.*
- We also use μ other than $\mathbb{E}(X)$ to denote the expectation. But these are same things.

- Recall that so far we assumed our random variable X is a discrete random variable (it takes only countable number of values $x_1 \ldots, x_k$)
- But it is possible that the random variable X is continuous, in this case, we will write expectation with integration, so we will write $\mathbb{E}(X) = \int_{-\infty}^{\infty} xf(x)dx$, where $f(x)$ is the density function for a continuous random variable.
- So now below we write the two formulas together,

$$\mathbb{E}(X) = \sum_{i=1}^{k} x_i f(x_i), \text{ when } X \text{ is discrete and it takes values } x_1, x_2, x_3, \ldots, x_k$$

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} xf(x)dx, \text{ when } X \text{ is continuous}$$

- In general we will either write $\mathbb{E}(X)$ or μ denote expectation. Depending on different examples X might be discrete or continuous.
- Lastly we would like to point out that expectation is a measure of *central tendency*, it gives information about the center of the *population*.

- So population mean or expectation, is a measure of central location in the population.
- Like mean there is another measure called *variance* which gives us a measure of the spread in the population, the formula is

$$\mathrm{Var}(X) = \mathbb{E}\left(\left(X - \mathbb{E}\left(X\right)\right)^2\right)$$

- Note that this is also expectation, but not for X, rather for $(X - \mathbb{E}(X))^2$.
- For discrete random variable X taking values $x_1, x_2, \ldots, x_k$ we can calculate this if we know the probability density function of $f(x)$, the formula will be

$$\mathrm{Var}(X) = \sum_{i=1}^{k} (x_i - \mu)^2 f(x_i)$$

- Similarly when X is continuous the same formula will be slightly modified with integration.
- Here we give the formula for the *population* variance both for discrete and continuous case,

$\mathrm{Var}(X) = \sum_{i=1}^{k} (x_i - \mu)^2 f(x_i)$ when X is discrete and it takes values $x_1, x_2, x_3, \ldots, x_k$

$\mathrm{Var}(X) = \int^{\infty} (x - \mu)^2 f(x)$ when X is continuous

- So far we did not bring sample into the picture, we are only talking about population mean $\mathbb{E}(X)$ and population variance $\mathrm{Var}(X)$.
- Now suppose we have a random sample $X_1, X_2, \ldots, X_n$, then how does sample mean and sample variance look like? (you already know the answers!)

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \text{sample mean}$$

$$S^2 = \frac{1}{n}\sum_{i=1}^{n} (X_i - \bar{X})^2, \quad \text{sample variance}$$

- Recall sample mean $\bar{X}$ and sample variance S^2 are also *estimators* of the population mean $\mathbb{E}(X)$ and population variance $\mathrm{Var}(X)$.
- Now if you notice there is a pattern on *writing estimators* of different population quantities.
- Actually there is a nice trick that is - "*replace expectation with averages*"
- This means if we need an estimator of $\mathbb{E}(X)$, then we have $\frac{1}{n}\sum_{i=1}^{n} X_i$
- Similarly if we need an estimator of $\mathrm{Var}(X) = \mathbb{E}\left(\left(X - \mathbb{E}\left(X\right)\right)^2\right)$, then we have $\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$

- Note probability density function (or PDF) helps us to calculate the probabilities when X takes different values, so this helps us to know the distribution of X.
- For the discrete case, the idea of probability density function is easy to understand, for example for our example at page 28, we have $f(10) = 1/4$, $f(20) = 1/4$, $f(30) = 2/4$. So it's function where we give input as the values of the random variable and output will be a probability.
- When X is continuous, the density function plays its role differently. Here we cannot calculate probabilities like $P(X = 10) = 1/4$, actually for any value this will be always 0.
- For a continuous random variable we use the density function to calculate probabilities, for example, if X is continuous, we can calculate $P(9 < X < 11)$.
- And we take the help of density function to calculate this, so $P(9 < X < 11) = \int_9^{11} f(x)dx$, where $f(x)$ is the density function of X.
- A typical example of a continuous random variable is when X is a normally distributed random variable.

- So far we have looked at the probability density function of X, and there was only one random variable.
- Things get a little bit different when we have multiple random variables, for example when we have two random variable X and Y
- In this case there is a *joint distribution* which specifies the probability of happening both X and Y together.
- The joint density function then looks like below

$$f(x, y) = P(X = x \text{ and } Y = y)$$

- We call this *discrete joint probability density function* and it gives the (joint) probability that X takes the value of x and Y takes the value of y.
- The following table gives the joint PDF of the discrete variables X and Y

		X			
		–2	**0**	**2**	**3**
Y	**3**	0.27	0.08	0.16	0
	6	0	0.04	0.10	0.35

- Now when we know the joint distribution of X and Y, we actually have lots of information.
- From the joint PDF we can calculate the marginal PDFs.
- Marginal PDF means PDF of a single random variable.
- For example from this joint PDF we can figure out the marginal PDF of X, we will denote this by $f_X(x)$ and the marginal PDF of Y, we will denote this by $f_Y(y)$
- Here are the formulas to calculate the marginal PDF of X and Y

$$f_X(x) = \sum_y f(x,y), \text{ marginal PDF of } X$$

$$f_Y(y) = \sum_x f(x,y), \text{ marginal PDF of } Y$$

- Let's figure out the marginal PDF from the joint PDF. The marginal PDF of X will be

$$f_X(-2) = \sum_y f(x,y) = 0.27 + 0 = 0.27, \quad f_X(0) = \sum_y f(x,y) = 0.08 + 0.04 = 0.12,$$

$$f_X(2) = \sum_y f(x,y) = 0.16 + 0.10 = 0.26, \quad f_X(3) = \sum_y f(x,y) = 0 + 0.35 = 0.35$$

- So we have $f_X(-2) = 0.27$, $f_X(0) = 0.12$, $f_X(2) = 0.26$, and $f_X(3) = 0.35$. This is

- I am skipping the marginal PDF of Y, but if you do this then you will get $f_Y(3) = 0.51$ and $f_Y(6) = 0.49$
- Similarly for the continuous case, we will have to integrate (but don't worry we will not do integration)

$$f_X(x) = \int f(x, y)dy, \text{ marginal PDF of } X$$

$$f_Y(y) = \int f(x, y)dx, \text{ marginal PDF of } Y$$

- Another object we can calculate from the joint distribution is *conditional distribution*. The idea of the conditional distribution is we have to *condition* on a value.
- For X and Y, we can condition on either of them,

$$f(x \mid y) = P(X = x \mid Y = y) \text{ condtional PDF of } X \text{ on } Y = y$$
$$f(y \mid x) = P(Y = y \mid X = x) \text{ condtional PDF of } Y \text{ on } X = x$$

- Now how do you calculate conditional PDF from the joint distribution?
- We need to use a formula which is derived using the concept *conditional probability*, here are formulas to derive

$$f(x \mid y) = \frac{f(x, y)}{f_Y(y)}, \text{ condtional PDF of } X \text{ on } Y = y$$
$$f(y \mid x) = \frac{f(x, y)}{f_X(x)}, \text{ condtional PDF of } Y \text{ on } X = x$$

- Let's do an example, fix $X = 2$, then we need to calculate the conditional PDF of Y given $X = 2$, this means we need to calculate $f(y|X = 2)$
- This means we need to calculate $f(3|X = 2)$ and $f(6|X = 2)$

- We can use the joint PDF, then using the formula $\frac{f(x,y)}{f_X(x)}$, we get

$$f(3|X=2) = \frac{f(2,3)}{f_X(2)} = \frac{0.16}{0.26} = 0.62 \text{ and}$$
$$f(6|X=2) = \frac{f(2,6)}{f_X(2)} = \frac{0.10}{0.26} = 0.38$$

- So the conditional PDF Y conditioning on $X=2$, we have $f(3|X=2)=0.62$, $f(6|X=2)=0.38$
- Check that whether you can calculate the conditional PDF $f(y|X=0)$. This means you have to figure out $f(3|X=0)=?$, $f(6|X=0)=?$
- Check that whether you can calculate conditional PDF $f(x|Y=6)$, This means you have to calculate $f(-2|Y=6)=?$, $f(0|Y=6)=?$ $f(2|Y=6)=?$, $f(3|Y=6)=?$

- With the conditional PDF, we can actually calculate both *conditional expectation* and *conditional variance*. For example the conditional expectation of Y on $X = x$ (when X and Y are discrete) is

$$\begin{aligned}\mathbb{E}(Y \mid X = x) &= \sum_y y f(y \mid X = x), \text{ if } Y \text{ is discrete} \\ &= \int y f(y \mid X = x) dy, \text{ if } Y \text{ is continuous}\end{aligned}$$

- Let's look at one example for the discrete case. Let's use the joint distribution in page 34, we can calculate $\mathbb{E}(Y \mid X = 2)$ (Recall we already calculated $f(3|X = 2) = 0.62$, $f(6|X = 2) = 0.38$)

$$\begin{aligned}\mathbb{E}(Y \mid X = 2) &= \sum_y y f(y \mid X = 2) \\ &= (3 \times f(3 \mid X = 2)) + (6 \times f(6 \mid X = 2)) \\ &= (3 \times 0.62) + (6 \times 0.38) \\ &= 4.15\end{aligned}$$

- Note conditional expectation works just like expectation, except we are using conditional density.

◇ Like conditional mean we can also calculate conditional variance,

$$\mathrm{Var}(Y \mid X = x) = \mathbb{E}\left[(Y - \mathbb{E}\left(Y \mid X = x\right))^{2} \mid X = x\right]$$

◇ Again for discrete case we will sum and for continuous we will integrate,

$$\begin{aligned}\mathrm{Var}(Y \mid X = x) &= \sum_{y} \left[y - \mathbb{E}\left(Y \mid X = x\right)\right]^{2} f(y \mid X = x) \\ &= \int \left[y - \mathbb{E}\left(Y \mid X = x\right)\right]^{2} f(y \mid X = x) dy\end{aligned}$$

◇ Let's look at an example for the discrete case with table in page 24, we would like to calculate $\mathrm{var}(Y \mid X = 2)$. Recall we already calculated $f(3|X = 2) = 0.62$, $f(6|X = 2) = 0.38$,

$$\begin{aligned}\mathrm{Var}(Y \mid X = 2) &= \sum_{y} \left[y - \mathbb{E}\left(Y \mid X = 2\right)\right]^{2} f(Y \mid X = 2) \\ &= (3 - 4.15)^2(0.62) + (6 - 4.15)^2(0.38) \\ &= 2.13\end{aligned}$$

◇ So we can calculate conditional mean and variance if we have the joint distribution.

- If we know the joint distribution then we can also calculate two more important objects, they are *covariance* and *correlation*.
- Let X and Y be two rv's with means $\mathbb{E}(X)$ and $\mathbb{E}(Y)$, respectively. Then the covariance between the two variables is defined as

$$\operatorname{cov}(X,Y) = \mathbb{E}\left[(X - \mathbb{E}(X))\,(Y - \mathbb{E}(Y))\right] = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

- Let's think about the definition. If X and Y tend to move in the same direction, then $X - \mathbb{E}(X)$ and $Y - \mathbb{E}(Y)$ will tend to be either both positive or both negative, so $(X - \mathbb{E}(X))\,(Y - \mathbb{E}(Y))$ will be positive on average, giving a positive covariance.
- If X and Y tend to move in opposite directions, then $X - \mathbb{E}(X)$ and $Y - \mathbb{E}(Y)$ will tend to have opposite signs, giving a negative covariance.
- For our joint PDF at page 34, we can calculate *population covariance*
- Our joint PDF is given as

$$f(-2,3) = 0.27, \quad f(0,3) = .08, \quad f(2,3) = 0.16, \quad f(3,3) = 0,$$
$$f(-2,6) = 0, \quad f(0,6) = .04, \quad f(2,6) = .10, \quad f(3,6) = 0.35$$

- Using this we can calculate the population covariance either of the two formulas, let's apply the second one which is

$$\mathrm{cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

- First we need to calculate the marginal expectation $\mathbb{E}(X)$ and $\mathbb{E}(Y)$, to do this we need two marginal densities $f_X(x)$ and $f_Y(y)$
- We already know this, recall $f_X(-2) = 0.27$, $f_X(0) = 0.12$, $f_X(2) = 0.26$, and $f_X(3) = 0.35$
- $f_Y(3) = 0.51$ and $f_Y(6) = 0.49$
- $\mathbb{E}(X) = (-2 \times 0.27) + (0 \times 0.12) + (2 \times 0.26) + (3 \times 0.35) = 1.03$
- $\mathbb{E}(Y) = (3 \times 0.51) + (6 \times 0.49) = 4.47$
- Now to calculate covariance we need $\mathbb{E}(XY)$, this can be calculated using the joint density

$$
\begin{aligned}
\mathbb{E}(XY) &= \sum_x \sum_y xyf(x,y) \\
&= ((-2\times 3)\times f(-2,3)) + ((0\times 3)\times f(0,3)) + \\
&= ((2\times 3)\times f(2,3)) + ((3\times 3)\times f(3,3)) + \\
&= ((-2\times 6)\times f(-2,6)) + ((0\times 6)\times f(0,6)) + \\
&= ((2\times 6)\times f(2,6)) + ((3\times 6)\times f(3,6)) \\
&= (-2\times 3)\times(0.27) + (0\times 3)\times(0.08) + (2\times 3)\times(0.16) + (3\times 3)\times(0) + \\
&= (-2\times 6)\times(0) + (0\times 6)\times(0.04) + (2\times 6)\times(0.10) + (3\times 6)\times(0.35) \\
&= 6.84
\end{aligned}
$$

- Note that $\sum_x \sum_y$, it's just mean we are summing over all x and y combinations.
- Then

$$
\begin{aligned}
\text{cov}(X,Y) &= E(XY) - \mathbb{E}(X)\mathbb{E}(Y) \\
&= 6.84 - (1.03)(4.47) \\
&= 2.24
\end{aligned}
$$

- The above formula is for the population covariance, for the sample counterpart we will replace the expectation with the sample average (we used $\widehat{\text{cov}}(X, Y)$ to denote it's a sample covariance)

$$\widehat{\text{cov}}(X, Y) = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}_i\right)\left(Y_i - \bar{Y}_i\right) = \frac{1}{n}\sum_{i=1}^{n} X_i Y_i - \bar{X}\bar{Y}$$

- $\widehat{\text{cov}}(X, Y)$ is the estimator of $\text{cov}(X, Y)$.
- Now let's try to understand covariance.
- Actually in some way the covariance between two random variables shows the dependence/independence between them.
- Recall, the concept of independent random variables, intuitively two random variables are independent, if there is no *probabilistic dependence* between them.
- This means their joint probability can be broken down towards the marginal one.
- If X and Y are discrete *independent* random variables, this means

$$P(X = x, Y = y) = P(X = x)P(Y = y)$$

- Note that this is same as writing

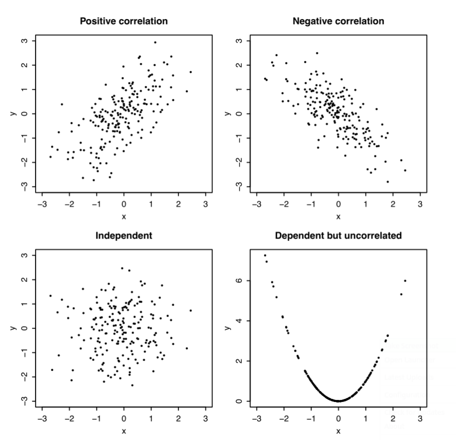
$$f(x, y) = f_X(x)f_Y(y)$$

- Using this we can easily show that when X and Y are independent then

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$$

- can you do this?
- This means when two random variables X and Y are independent, $\text{cov}(X, Y) = E(XY) - \mathbb{E}(X)\mathbb{E}(Y) = \mathbb{E}(X)\mathbb{E}(Y) - \mathbb{E}(X)\mathbb{E}(Y) = 0$
- So independent means 0 covariance in the population.

- If we have a random sample of X and Y where there is 0 covariance in the population, we can also plot the data and this pattern will be visible,
- Following is a plot from different covariance structures in the population.



- If we take the covariance of two random variables X and Y, and we divide them by their variances, then we get what is known as *correlation*.
- Usually for the correlation, we use the notation $\rho_{X,Y}$,

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sqrt{(\text{var}\,(X))\,(\text{var}\,(Y))}} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y}$$

- where σ_X and σ_Y are the standard deviation of X and Y respectively.
- So correlation is nothing new, it's just covariance divided by the standard deviations of the two variables.
- What is the benefit of correlation?
- The difference between the covariance and correlation is, the covariance can be any number, but the correlation will be always between 0 and 1.
- Note that $\rho_{X,Y}$, is the population correlation, for the sample correlation we just have to use the estimators of each of these quantities, so we get

$$r_{X,Y} = \frac{\widehat{\text{cov}}(X,Y)}{S_X S_Y}$$

- Where S_X and S_Y are sample standard deviations of X and Y respectively.

- There are some rules to calculate Expectation and Variance, these are easy to apply (the proofs are also easy but I am skipping it!)

Theorem 1.3 (Rules for Expectation)

✓ 1. The expected value of a constant is the constant itself. Thus, if b is a constant, $\mathbb{E}(b) = b$.

✓ 2. If a and b are constants,

$$\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$$

This can be generalized. If $X_1, X_2, \ldots, X_n$ are n random variables and $a_1, a_2, \ldots, a_n$ and b are constants, then
$\mathbb{E}\left(a_1 X_1 + a_2 X_2 + \cdots + a_n X_n + b\right) = a_1 \mathbb{E}\left(X_1\right) + a_2 \mathbb{E}\left(X_2\right) + \cdots + a_n \mathbb{E}\left(X_n\right) + b$

✓ 3. If X and Y are independent random variables, then

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$$

That is, the *expectation of the product XY is the product of the (individual) expectations of X and Y.*

✓ 4. If X is a random variable with PDF $f(x)$ and if $g(X)$ is any function of X, then

$$E[g(X)] = \sum_x g(X) f(x) \qquad \text{if } X \text{ is discrete}$$

- Similarly there are some rules to calculate Variance.

Theorem 1.4 (Rules for Variance)

✓ 1. The variance of a constant b is zero, so $\operatorname{Var}(b) = 0$

✓ 2. If a and b are constants, then

$$\operatorname{Var}(aX + b) = a^2 \operatorname{var}(X)$$

✓ 3. If X and Y are *independent* random variables, then

$$\operatorname{var}(X + Y) = \operatorname{var}(X) + \operatorname{var}(Y)$$
$$\operatorname{var}(X - Y) = \operatorname{var}(X) + \operatorname{var}(Y)$$

This can be generalized to more than two independent variables.

✓ 4. If X and Y are independent random variables, and a and b are constants, then

$$\operatorname{var}(aX + bY) = a^2 \operatorname{var}(X) + b^2 \operatorname{var}(Y)$$

✓ 5. $\mathbb{E}\left((X - \mu)^2\right) = \mathbb{E}\left(X^2\right) - (\mathbb{E}(X))^2$, this is an alternative formula for variance.