CS 107, Probability, Spring 2019
Lecture 40

Michael Poghosyan

AUA

03 May 2019

Content

- Markov and Chebyshev Inequalities
- Covariance of R.V.s

Properties of a Variance and a SD

Another important property is the calculation of the Variance of a sum of several r.v.s. We will give the general formula later, but one important case is the following:

Properties of a Variance and a SD

Another important property is the calculation of the Variance of a sum of several r.v.s. We will give the general formula later, but one important case is the following:

Variance of a Sum of Independent RVs

Assume X and Y are Independent r.v.s, $X \perp\!\!\!\perp Y$. Then

$$Var(X+Y) = Var(X) + Var(Y).$$

Properties of a Variance and a SD

Another important property is the calculation of the Variance of a sum of several r.v.s. We will give the general formula later, but one important case is the following:

Variance of a Sum of Independent RVs

Assume X and Y are Independent r.v.s, $X \perp\!\!\!\perp Y$. Then

$$Var(X+Y) = Var(X) + Var(Y).$$

Note: Important is to remember that

Properties of a Variance and a SD

Another important property is the calculation of the Variance of a sum of several r.v.s. We will give the general formula later, but one important case is the following:

Variance of a Sum of Independent RVs

Assume X and Y are Independent r.v.s, $X \perp\!\!\!\perp Y$. Then

$$Var(X + Y) = Var(X) + Var(Y).$$

Note: Important is to remember that

- The property $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$ works for ANY r.v. X and Y;

Properties of a Variance and a SD

Another important property is the calculation of the Variance of a sum of several r.v.s. We will give the general formula later, but one important case is the following:

Variance of a Sum of Independent RVs

Assume X and Y are Independent r.v.s, $X \perp\!\!\!\perp Y$. Then

$$Var(X + Y) = Var(X) + Var(Y).$$

Note: Important is to remember that

- The property $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$ works for ANY r.v. X and Y;
- The property $Var(X + Y) = Var(X) + Var(Y)$ is NOT TRUE for any r.v.s. It is true, in particular, when $X \perp\!\!\!\perp Y$.

Properties of a Variance and a SD

Another important property is the calculation of the Variance of a sum of several r.v.s. We will give the general formula later, but one important case is the following:

Variance of a Sum of Independent RVs

Assume X and Y are Independent r.v.s, $X \perp\!\!\!\perp Y$. Then

$$Var(X+Y) = Var(X) + Var(Y).$$

Note: Important is to remember that

- The property $\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$ works for ANY r.v. X and Y;
- The property $Var(X+Y) = Var(X) + Var(Y)$ is NOT TRUE for any r.v.s. It is true, in particular, when $X \perp\!\!\!\perp Y$. Later we will see that this property holds only for *uncorrelated* r.v.s.

Properties of a Variance

Note: It is an interesting and remarkable fact, that the above property of a Variance for Independent R.V. Can be interpreted as an analogue of the Pythagorean Theorem. Here we interpret X and Y as vectors, $SD(X)$ as the length of a vector X - so $Var(X)$ is the square of the length, and the independence is interpreted as orthogonality (perpendicularity): and this is the reason that we are using the notation $X \perp\!\!\!\perp Y$ for the independence!
So if $a = SD(X)$, $b = SD(Y)$, $c = SD(X+Y)$, and $X \perp\!\!\!\perp Y$, then

$$c^2 = a^2 + b^2.$$

Example

Example: Assume $X, Y \sim Bernoulli(0.5)$ and $X \perp\!\!\!\perp Y$. Calculate

$$Var(2X - 3Y + 5).$$

Solution: OTB
Example: Is it true that if $X \perp\!\!\!\perp Y$, then

$$Var(X - Y) = Var(X) - Var(Y),$$

in general? **Solution:** OTB

Generalization for n Independent R.V.s

The above property works also for more than n independent r.v.s:

Generalization for n Independent R.V.s

The above property works also for more than n independent r.v.s:

Variance of a Sum of Independent RVs

Assume $X_1, X_2, ..., X_n$ are Independent (i.e., Mutually Independent) r.v.s. Then

$$Var(X_1 + X_2 + ... + X_n) = Var(X_1) + Var(X_2) + ... + Var(X_n).$$

Generalization for n Independent R.V.s

The above property works also for more than n independent r.v.s:

Variance of a Sum of Independent RVs

Assume $X_1, X_2, ..., X_n$ are Independent (i.e., Mutually Independent) r.v.s. Then

$$Var(X_1 + X_2 + ... + X_n) = Var(X_1) + Var(X_2) + ... + Var(X_n).$$

Later we will see that this property holds in a more general case, when X_k-s are **uncorrelated**, and we will learn how to calculate the Variance of a sum if the terms are not Independent.

Markov and Chebyshev Inequalities: Intro

Of course, if we know the CDF or PD(M)F of a r.v. X, then we can calculate the Probabilities like $\mathbb{P}(X \in [a, b])$.
Now, assume we do not know a complete information about X. Say, we know the Mean of a r.v. X. Or we know the Mean and the Variance of X. Can we calculate probabilities concerning X?

Markov and Chebyshev Inequalities: Intro

Of course, if we know the CDF or PD(M)F of a r.v. X, then we can calculate the Probabilities like $\mathbb{P}(X \in [a, b])$.
Now, assume we do not know a complete information about X. Say, we know the Mean of a r.v. X. Or we know the Mean and the Variance of X. Can we calculate probabilities concerning X? Of course, no. Knowing only $\mathbb{E}(X)$ and $Var(X)$ is not giving us the possibility to calculate Probabilities like $\mathbb{P}(X \in [a, b])$.

Markov and Chebyshev Inequalities: Intro

Of course, if we know the CDF or PD(M)F of a r.v. X, then we can calculate the Probabilities like $\mathbb{P}(X \in [a, b])$.
Now, assume we do not know a complete information about X. Say, we know the Mean of a r.v. X. Or we know the Mean and the Variance of X. Can we calculate probabilities concerning X? Of course, no. Knowing only $\mathbb{E}(X)$ and $Var(X)$ is not giving us the possibility to calculate Probabilities like $\mathbb{P}(X \in [a, b])$. This is because, in the general case, only $\mathbb{E}(X)$ and $Var(X)$ are not giving us the complete information about X.

Markov and Chebyshev Inequalities: Intro

Of course, if we know the CDF or PD(M)F of a r.v. X, then we can calculate the Probabilities like $\mathbb{P}(X \in [a, b])$.

Now, assume we do not know a complete information about X. Say, we know the Mean of a r.v. X. Or we know the Mean and the Variance of X. Can we calculate probabilities concerning X?

Of course, no. Knowing only $\mathbb{E}(X)$ and $Var(X)$ is not giving us the possibility to calculate Probabilities like $\mathbb{P}(X \in [a, b])$. This is because, in the general case, only $\mathbb{E}(X)$ and $Var(X)$ are not giving us the complete information about X.

Fortunately, it turns out that we can say something, we can estimate some Probabilities.

Markov and Chebyshev Inequalities: Intro

Of course, if we know the CDF or PD(M)F of a r.v. X, then we can calculate the Probabilities like $\mathbb{P}(X \in [a, b])$.

Now, assume we do not know a complete information about X. Say, we know the Mean of a r.v. X. Or we know the Mean and the Variance of X. Can we calculate probabilities concerning X? Of course, no. Knowing only $\mathbb{E}(X)$ and $Var(X)$ is not giving us the possibility to calculate Probabilities like $\mathbb{P}(X \in [a, b])$. This is because, in the general case, only $\mathbb{E}(X)$ and $Var(X)$ are not giving us the complete information about X.

Fortunately, it turns out that we can say something, we can estimate some Probabilities. This is because of Markov and Chabyshev Inequalities.

Markov and Chebyshev Inequalities: Intro

Of course, if we know the CDF or PD(M)F of a r.v. X, then we can calculate the Probabilities like $\mathbb{P}(X \in [a, b])$.
Now, assume we do not know a complete information about X. Say, we know the Mean of a r.v. X. Or we know the Mean and the Variance of X. Can we calculate probabilities concerning X? Of course, no. Knowing only $\mathbb{E}(X)$ and $Var(X)$ is not giving us the possibility to calculate Probabilities like $\mathbb{P}(X \in [a, b])$. This is because, in the general case, only $\mathbb{E}(X)$ and $Var(X)$ are not giving us the complete information about X.
Fortunately, it turns out that we can say something, we can estimate some Probabilities. This is because of Markov and Chabyshev Inequalities.
These two inequalities are examples of Concentration Inequalities, see `https://en.wikipedia.org/wiki/Concentration_inequality`

Markov and Chebyshev Inequalities

First, we give the Markov's Inequality:

Markov and Chebyshev Inequalities

First, we give the Markov's Inequality:

Markov Inequality

Assume X is a r.v. with $X \geq 0$ almost surely, and let $a > 0$. Then

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}.$$

Markov and Chebyshev Inequalities

First, we give the Markov's Inequality:

Markov Inequality

Assume X is a r.v. with $X \geq 0$ almost surely, and let $a > 0$. Then

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}.$$

Note: It is important that $X \geq 0$. In the general case, the Markov's Inequality looks like: $\mathbb{P}(|X| \geq a) \leq \dfrac{\mathbb{E}(|X|)}{a}$.

Markov and Chebyshev Inequalities

First, we give the Markov's Inequality:

Markov Inequality

Assume X is a r.v. with $X \geq 0$ almost surely, and let $a > 0$. Then
$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}.$$

Note: It is important that $X \geq 0$. In the general case, the Markov's Inequality looks like: $\mathbb{P}(|X| \geq a) \leq \dfrac{\mathbb{E}(|X|)}{a}$.

Note: We can rewrite the Markov's Inequality as
$$\mathbb{P}(X \geq a \cdot \mathbb{E}(X)) \leq \frac{1}{a}.$$

Markov and Chebyshev Inequalities

First, we give the Markov's Inequality:

Markov Inequality

Assume X is a r.v. with $X \geq 0$ almost surely, and let $a > 0$. Then

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}.$$

Note: It is important that $X \geq 0$. In the general case, the Markov's Inequality looks like: $\mathbb{P}(|X| \geq a) \leq \dfrac{\mathbb{E}(|X|)}{a}$.

Note: We can rewrite the Markov's Inequality as

$$\mathbb{P}(X \geq a \cdot \mathbb{E}(X)) \leq \frac{1}{a}.$$

We interpret this as for a non-negative r.v., the probability that it takes values much larger than the Expected Value is small.

Markov and Chebyshev Inequalities

So the Markov's Inequality gives us a tool to estimate the probability $\mathbb{P}(X \geq a)$ if we know only the Expected Value $\mathbb{E}(X)$.

Markov and Chebyshev Inequalities

So the Markov's Inequality gives us a tool to estimate the probability $\mathbb{P}(X \geq a)$ if we know only the Expected Value $\mathbb{E}(X)$.
Example: Let X be the wage of a (randomly chosen working) person in Armenia. We know that $\mathbb{E}(X) = 172,056$AMD.

Markov and Chebyshev Inequalities

So the Markov's Inequality gives us a tool to estimate the probability $\mathbb{P}(X \geq a)$ if we know only the Expected Value $\mathbb{E}(X)$.

Example: Let X be the wage of a (randomly chosen working) person in Armenia. We know that $\mathbb{E}(X) = 172,056$AMD. Then the proportion of persons receiving more than the 10 times the average wage is not more than

Markov and Chebyshev Inequalities

So the Markov's Inequality gives us a tool to estimate the probability $\mathbb{P}(X \geq a)$ if we know only the Expected Value $\mathbb{E}(X)$.

Example: Let X be the wage of a (randomly chosen working) person in Armenia. We know that $\mathbb{E}(X) = 172,056$AMD. Then the proportion of persons receiving more than the 10 times the average wage is not more than $1/10$.

Markov and Chebyshev Inequalities

So the Markov's Inequality gives us a tool to estimate the probability $\mathbb{P}(X \geq a)$ if we know only the Expected Value $\mathbb{E}(X)$.

Example: Let X be the wage of a (randomly chosen working) person in Armenia. We know that $\mathbb{E}(X) = 172,056$AMD. Then the proportion of persons receiving more than the 10 times the average wage is not more than $1/10$.

Proof: Indeed, by the Markov's Inequality,

$$\mathbb{P}(X \geq 10\mathbb{E}(X)) \leq \frac{\mathbb{E}(X)}{10\mathbb{E}(X)} = \frac{1}{10}.$$

Markov and Chebyshev Inequalities

Now, we give the Chebyshev's Inequality.

Markov and Chebyshev Inequalities

Now, we give the Chebyshev's Inequality. We assume that we know about X not only its Expectation $\mathbb{E}(X)$, but also the Variance $Var(X)$.

Markov and Chebyshev Inequalities

Now, we give the Chebyshev's Inequality. We assume that we know about X not only its Expectation $\mathbb{E}(X)$, but also the Variance $Var(X)$. So, having more information, we can estimate Probabilities more precisely:

Markov and Chebyshev Inequalities

Now, we give the Chebyshev's Inequality. We assume that we know about X not only its Expectation $\mathbb{E}(X)$, but also the Variance $Var(X)$. So, having more information, we can estimate Probabilities more precisely:

Chebyshev's Inequality

Assume X is a r.v., and let $a > 0$. Then

$$\mathbb{P}\Big(|X - \mathbb{E}(X)| \geq a\Big) \leq \frac{Var(X)}{a^2}.$$

Markov and Chebyshev Inequalities

Now, we give the Chebyshev's Inequality. We assume that we know about X not only its Expectation $\mathbb{E}(X)$, but also the Variance $Var(X)$. So, having more information, we can estimate Probabilities more precisely:

Chebyshev's Inequality

Assume X is a r.v., and let $a > 0$. Then

$$\mathbb{P}\Big(|X - \mathbb{E}(X)| \geq a\Big) \leq \frac{Var(X)}{a^2}.$$

Note: We can rewrite the Chebyshev's Inequality as

$$\mathbb{P}\Big(|X - \mathbb{E}(X)| \geq a \cdot SD(X)\Big) \leq \frac{1}{a^2}.$$

Markov and Chebyshev Inequalities

Chebyshev's Inequality says that we can estimate the probability that X takes values away from its Expected Value by its Variance.

Markov and Chebyshev Inequalities

Chebyshev's Inequality says that we can estimate the probability that X takes values away from its Expected Value by its Variance.
Say, if the Variance of X is small, then the probability that X will be far from $\mathbb{E}(X)$ is small.

Markov and Chebyshev Inequalities

Chebyshev's Inequality says that we can estimate the probability that X takes values away from its Expected Value by its Variance.
Say, if the Variance of X is small, then the probability that X will be far from $\mathbb{E}(X)$ is small.
E.g., the probability that X will be more than $3SD(X)$-away from $\mathbb{E}(X)$ is:

$$\mathbb{P}\Big(|X - \mathbb{E}(X)| \geq 3 \cdot SD(X)\Big) \leq \frac{1}{3^2} = \frac{1}{9}.$$

Markov and Chebyshev Inequalities

Chebyshev's Inequality says that we can estimate the probability that X takes values away from its Expected Value by its Variance.
Say, if the Variance of X is small, then the probability that X will be far from $\mathbb{E}(X)$ is small.
E.g., the probability that X will be more than $3SD(X)$-away from $\mathbb{E}(X)$ is:

$$\mathbb{P}\Big(|X - \mathbb{E}(X)| \geq 3 \cdot SD(X)\Big) \leq \frac{1}{3^2} = \frac{1}{9}.$$

And this works for ANY r.v. X!

Example

Example: Assume X is a r.v. with $\mathbb{E}(X) = 9$ and $Var(X) = 0.2$. Estimate

Example

Example: Assume X is a r.v. with $\mathbb{E}(X) = 9$ and $Var(X) = 0.2$. Estimate

- $\mathbb{P}(8 < X < 10)$;

Example

Example: Assume X is a r.v. with $\mathbb{E}(X) = 9$ and $Var(X) = 0.2$. Estimate

- $\mathbb{P}(8 < X < 10)$;
- $\mathbb{P}(8.1 < X < 10)$.

Some Additions

Above we considered two important numerical (partial) characteristics of a r.v. X: its Expectation $\mathbb{E}(X)$ and Variance $Var(X)$.

Some Additions

Above we considered two important numerical (partial) characteristics of a r.v. X: its Expectation $\mathbb{E}(X)$ and Variance $Var(X)$. In Probability Theory and Statistics, one considers also other, higher-order characteristics:

Some Additions

Above we considered two important numerical (partial) characteristics of a r.v. X: its Expectation $\mathbb{E}(X)$ and Variance $Var(X)$. In Probability Theory and Statistics, one considers also other, higher-order characteristics:

- k-th order Moments: $\mu_k = \mu_k(X) = \mathbb{E}(X^k)$, provided that $\mathbb{E}(|X|^k) < +\infty$;

Some Additions

Above we considered two important numerical (partial) characteristics of a r.v. X: its Expectation $\mathbb{E}(X)$ and Variance $Var(X)$. In Probability Theory and Statistics, one considers also other, higher-order characteristics:

- k-th order Moments: $\mu_k = \mu_k(X) = \mathbb{E}(X^k)$, provided that $\mathbb{E}(|X|^k) < +\infty$;
- k-th order Central Moments:

$$\mu_k^{central} = \mathbb{E}\Big((X - \mathbb{E}(X))^k\Big);$$

Some Additions

Above we considered two important numerical (partial) characteristics of a r.v. X: its Expectation $\mathbb{E}(X)$ and Variance $Var(X)$. In Probability Theory and Statistics, one considers also other, higher-order characteristics:

- k-th order Moments: $\mu_k = \mu_k(X) = \mathbb{E}(X^k)$, provided that $\mathbb{E}(|X|^k) < +\infty$;
- k-th order Central Moments:

$$\mu_k^{central} = \mathbb{E}\Big((X - \mathbb{E}(X))^k\Big);$$

- k-th order Normalized or Standardized Moments:

$$\mu_k^{standardized} = \mathbb{E}\left(\left(\frac{X - \mathbb{E}(X)}{SD(X)}\right)^k\right).$$

Some Additions

Notes:

Some Additions

Notes:

- $\mathbb{E}(X)$ is the first-order Moment of X;

Some Additions

Notes:

- $\mathbb{E}(X)$ is the first-order Moment of X;
- $\mathbb{E}(X^2)$ is the second-order Moment of X;

Some Additions

Notes:

- $\mathbb{E}(X)$ is the first-order Moment of X;
- $\mathbb{E}(X^2)$ is the second-order Moment of X; Knowing this and $\mathbb{E}(X)$ is equivalent to knowing the Variance, because of the relationship $Var(X) = \mathbb{E}(X^2) - \big(\mathbb{E}(X)\big)^2$;

Some Additions

Notes:

- $\mathbb{E}(X)$ is the first-order Moment of X;
- $\mathbb{E}(X^2)$ is the second-order Moment of X; Knowing this and $\mathbb{E}(X)$ is equivalent to knowing the Variance, because of the relationship $Var(X) = \mathbb{E}(X^2) - \big(\mathbb{E}(X)\big)^2$;
- The more Moments of X we know, the more information we have about the Distribution of X;

Some Additions

Notes:

- $\mathbb{E}(X)$ is the first-order Moment of X;
- $\mathbb{E}(X^2)$ is the second-order Moment of X; Knowing this and $\mathbb{E}(X)$ is equivalent to knowing the Variance, because of the relationship $Var(X) = \mathbb{E}(X^2) - \big(\mathbb{E}(X)\big)^2$;
- The more Moments of X we know, the more information we have about the Distribution of X;
- $Var(X)$ is the second order Central Moment of X;

Notes:

Some Additions

Notes:

- The 3-rd order Standardized Moment is called the Skewness of X,

$$Skewness(X) = \mathbb{E}\left(\left(\frac{X - \mathbb{E}(X)}{SD(X)}\right)^3\right)$$

Notes:

- The 3-rd order Standardized Moment is called the Skewness of X,

$$Skewness(X) = \mathbb{E}\left(\left(\frac{X - \mathbb{E}(X)}{SD(X)}\right)^3\right)$$

It measures the **asymmetry** of the Distribution of X.

Some Additions

Notes:

- The 3-rd order Standardized Moment is called the Skewness of X,

$$Skewness(X) = \mathbb{E}\left(\left(\frac{X - \mathbb{E}(X)}{SD(X)}\right)^3\right)$$

  It measures the **asymmetry** of the Distribution of X.

- The 4-rd order Standardized Moment is called the Kurtosis of X,

$$Kurtosis(X) = \mathbb{E}\left(\left(\frac{X - \mathbb{E}(X)}{SD(X)}\right)^4\right)$$

Some Additions

Notes:

- The 3-rd order Standardized Moment is called the Skewness of X,

$$Skewness(X) = \mathbb{E}\left(\left(\frac{X-\mathbb{E}(X)}{SD(X)}\right)^3\right)$$

  It measures the **asymmetry** of the Distribution of X.

- The 4-rd order Standardized Moment is called the Kurtosis of X,

$$Kurtosis(X) = \mathbb{E}\left(\left(\frac{X-\mathbb{E}(X)}{SD(X)}\right)^4\right)$$

  It measures the **tailedness** of the Distribution of X.

Some Additions

Notes:

- The generalization of the Markov's inequality for higher-order moments is:

$$\mathbb{P}(|X| > a) = \mathbb{P}(|X|^k > a^k)$$

Some Additions

Notes:

- The generalization of the Markov's inequality for higher-order moments is:

$$\mathbb{P}(|X| > a) = \mathbb{P}(|X|^k > a^k) \le \frac{\mathbb{E}(|X|^k)}{a^k}, \quad \forall k \in \mathbb{N}, \forall a > 0.$$

Covariance of Random Variables

Covariance of r.v.s, intro

Recall that we can describe the Joint Distribution of X and Y, if we know their Joint CDF and/or their Joint PD(M)F.

Covariance of r.v.s, intro

Recall that we can describe the Joint Distribution of X and Y, if we know their Joint CDF and/or their Joint PD(M)F.
But, usually, having the Joint Distribution of X and Y is a Royal Gift - usually, we do not have this complete information. So we want to give, at least, some partial information. Of course, for some information about the Distributions of individual X and Y, we can give their Expectations $\mathbb{E}(X), \mathbb{E}(Y)$ and their Variances $Var(X), Var(Y)$.

Covariance of r.v.s, intro

Recall that we can describe the Joint Distribution of X and Y, if we know their Joint CDF and/or their Joint PD(M)F.
But, usually, having the Joint Distribution of X and Y is a Royal Gift - usually, we do not have this complete information. So we want to give, at least, some partial information. Of course, for some information about the Distributions of individual X and Y, we can give their Expectations $\mathbb{E}(X), \mathbb{E}(Y)$ and their Variances $Var(X), Var(Y)$.
But here the most important thing is to describe, at least, partially, **the relationship between X and Y.** And this is what we want to do below.

Example:

Example: Assume X is taking the values $1, 2, 3$ with equal Probabilities, and Y is taking the values $2, 4, 6$ with equal Probabilities. What can be said about the relationship between X and Y?

Example:

Example: Assume X is taking the values $1, 2, 3$ with equal Probabilities, and Y is taking the values $2, 4, 6$ with equal Probabilities. What can be said about the relationship between X and Y?
Of course, nothing!

Example:

Example: Assume X is taking the values $1, 2, 3$ with equal Probabilities, and Y is taking the values $2, 4, 6$ with equal Probabilities. What can be said about the relationship between X and Y?
Of course, nothing! Say, we can have that if $X = 1$, then $Y = 2$, if $X = 2$, then $Y = 4$, and if $X = 3$, then $Y = 6$.

Example:

Example: Assume X is taking the values $1, 2, 3$ with equal Probabilities, and Y is taking the values $2, 4, 6$ with equal Probabilities. What can be said about the relationship between X and Y?

Of course, nothing! Say, we can have that if $X = 1$, then $Y = 2$, if $X = 2$, then $Y = 4$, and if $X = 3$, then $Y = 6$. I.e., $Y = 2X$, so we have an exact linear relationship between X and Y.

Example:

Example: Assume X is taking the values $1, 2, 3$ with equal Probabilities, and Y is taking the values $2, 4, 6$ with equal Probabilities. What can be said about the relationship between X and Y?

Of course, nothing! Say, we can have that if $X = 1$, then $Y = 2$, if $X = 2$, then $Y = 4$, and if $X = 3$, then $Y = 6$. I.e., $Y = 2X$, so we have an exact linear relationship between X and Y.

Another example is that we can have that if $X = 1$, then $Y = 2$ or $Y = 4$ with equal probabilities, if $X = 2$, then $Y = 4$ or $Y = 6$ with equal probabilities and, in the case $X = 3$, we can have $Y = 2$ or $Y = 6$, equiprobably.

Example:

Example: Assume X is taking the values $1, 2, 3$ with equal Probabilities, and Y is taking the values $2, 4, 6$ with equal Probabilities. What can be said about the relationship between X and Y?

Of course, nothing! Say, we can have that if $X = 1$, then $Y = 2$, if $X = 2$, then $Y = 4$, and if $X = 3$, then $Y = 6$. I.e., $Y = 2X$, so we have an exact linear relationship between X and Y.

Another example is that we can have that if $X = 1$, then $Y = 2$ or $Y = 4$ with equal probabilities, if $X = 2$, then $Y = 4$ or $Y = 6$ with equal probabilities and, in the case $X = 3$, we can have $Y = 2$ or $Y = 6$, equiprobably.

So the above info say just that, in the third of cases, $X = 1$, in the third part of cases $X = 2$, and in the other cases $X = 3$, and the same for Y, but that third cases for X and Y can be very different!

Covariance of r.v.s

Now, we want to give some partial measure of "relationship" between r.v.s X and Y:

Covariance of r.v.s

Now, we want to give some partial measure of "relationship" between r.v.s X and Y:

Covariance of r.v.s

The **Covariance** of r.v.s X and Y is the following number:

$$Cov(X, Y) = \sigma_{XY} = \mathbb{E}\Big((X - \mathbb{E}(X)) \cdot (Y - \mathbb{E}(Y))\Big).$$

Covariance of r.v.s

Now, we want to give some partial measure of "relationship" between r.v.s X and Y:

Covariance of r.v.s

The **Covariance** of r.v.s X and Y is the following number:

$$Cov(X,Y) = \sigma_{XY} = \mathbb{E}\Big((X-\mathbb{E}(X))\cdot(Y-\mathbb{E}(Y))\Big).$$

Notes:

- Covariance = Co-Variance, measures How X and Y Co-Vary together

Covariance of r.v.s

Now, we want to give some partial measure of "relationship" between r.v.s X and Y:

Covariance of r.v.s

The **Covariance** of r.v.s X and Y is the following number:

$$Cov(X, Y) = \sigma_{XY} = \mathbb{E}\Big((X - \mathbb{E}(X)) \cdot (Y - \mathbb{E}(Y))\Big).$$

Notes:

- Covariance = Co-Variance, measures How X and Y Co-Vary together
- Covariance measures, in fact, the linear relationship between X and Y, it shows how much X and Y relate linearly to each other

Covariance of r.v.s

Now, we want to give some partial measure of "relationship" between r.v.s X and Y:

Covariance of r.v.s

The **Covariance** of r.v.s X and Y is the following number:

$$Cov(X, Y) = \sigma_{XY} = \mathbb{E}\Big((X - \mathbb{E}(X)) \cdot (Y - \mathbb{E}(Y))\Big).$$

Notes:

- Covariance = Co-Variance, measures How X and Y Co-Vary together
- Covariance measures, in fact, the linear relationship between X and Y, it shows how much X and Y relate linearly to each other
- We say that X and Y are **uncorrelated**, if $Cov(X, Y) = 0$