Statistics for Data Science -1

Lecture 9.5: Properties of variance

Usha Mohan

Indian Institute of Technology Madras

Learning objectives

Learning objectives

1. Define what is a random variable.

Learning objectives

1. Define what is a random variable.
2. Types of random variables: discrete and continuous.

Learning objectives

1. Define what is a random variable.
2. Types of random variables: discrete and continuous.
3. Probability mass function, graph, and examples.

Learning objectives

1. Define what is a random variable.
2. Types of random variables: discrete and continuous.
3. Probability mass function, graph, and examples.
4. Cumulative distribution function, graphs, and examples.

Learning objectives

1. Define what is a random variable.
2. Types of random variables: discrete and continuous.
3. Probability mass function, graph, and examples.
4. Cumulative distribution function, graphs, and examples.
5. Expectation and variance of a random variable.

Variance of a function of a random variable

Variance of a function of a random variable

Proposition

Let X be a random variable, let c be a constant, then

Variance of a function of a random variable

Proposition

Let X be a random variable, let c be a constant, then

- $Var(cX) = c^2 Var(X)$

Variance of a function of a random variable

Proposition

Let X be a random variable, let c be a constant, then

- $Var(cX) = c^2 Var(X)$
- $Var(X + c) = Var(X)$

Variance of a function of a random variable

Proposition

Let X be a random variable, let c be a constant, then

- $Var(cX) = c^2 Var(X)$
- $Var(X + c) = Var(X)$

Corollary

If a and b are constants, $V(aX + b) = a^2 Var(X)$

Variance of a function of a random variable

Proposition

Let X be a random variable, let c be a constant, then

- $Var(cX) = c^2 Var(X)$
- $Var(X + c) = Var(X)$

Corollary

If a and b are constants, $V(aX + b) = a^2 Var(X)$

Proof.

We know $E(aX + b) = a\mu + b$. Hence,
$Var(aX + b) = E\,(aX + b - a\mu - b)^2 = E\left(a^2(X - \mu)^2\right) =$
$a^2 E(X - \mu)^2 = a^2 Var(X)$ $\square$

Variance of sum of two random variables

- The expected value of the sum of random variables is equal to the sum of the individual expected values. i.e if X and Y be two random variables. Then, $E(X + Y) = E(X) + E(Y)$.
- What can be said about the Variance of sum of two random variables?

Variance of sum of two random variables

- The expected value of the sum of random variables is equal to the sum of the individual expected values. i.e if X and Y be two random variables. Then, $E(X+Y) = E(X) + E(Y)$.
- What can be said about the Variance of sum of two random variables?
- $Var(X+X) =$

Variance of sum of two random variables

- The expected value of the sum of random variables is equal to the sum of the individual expected values. i.e if X and Y be two random variables. Then, $E(X+Y) = E(X) + E(Y)$.
- What can be said about the Variance of sum of two random variables?
- $Var(X+X) = Var(2X) =$

Variance of sum of two random variables

- The expected value of the sum of random variables is equal to the sum of the individual expected values. i.e if X and Y be two random variables. Then, $E(X+Y) = E(X) + E(Y)$.
- What can be said about the Variance of sum of two random variables?
- $Var(X+X) = Var(2X) = 4Var(X) \neq$

Variance of sum of two random variables

- The expected value of the sum of random variables is equal to the sum of the individual expected values. i.e if X and Y be two random variables. Then, $E(X+Y)=E(X)+E(Y)$.
- What can be said about the Variance of sum of two random variables?
- $Var(X+X)=Var(2X)=4Var(X)\neq Var(X)+Var(X)$

Independent random variables

Independent random variables

Definition

Random variables X and Y are independent if knowing the value of one of them does not change the probabilities of the other.

Independent random variables

Definition

Random variables X and Y are independent if knowing the value of one of them does not change the probabilities of the other.

Example:

- Roll a dice twice. $S = \{(1, 1), \ldots, (6, 6)\}$

Independent random variables

Definition

Random variables X and Y are independent if knowing the value of one of them does not change the probabilities of the other.

Example:

- Roll a dice twice. $S = \{(1,1), \ldots, (6,6)\}$
- X is the outcome of the first dice. Y be the out come of the second dice.

Independent random variables

Definition

Random variables X and Y are independent if knowing the value of one of them does not change the probabilities of the other.

Example:

- Roll a dice twice. $S = \{(1,1), \ldots, (6,6)\}$
- X is the outcome of the first dice. Y be the out come of the second dice.
- Knowing $X = i$ does not change the probability of Y taking any value $1, 2, \ldots, 6$.

Independent random variables

Definition

Random variables X and Y are independent if knowing the value of one of them does not change the probabilities of the other.

Example:

- Roll a dice twice. $S = \{(1,1), \ldots, (6,6)\}$
- X is the outcome of the first dice. Y be the out come of the second dice.
- Knowing $X = i$ does not change the probability of Y taking any value $1, 2, \ldots, 6$.
- X and Y are independent random variables.

Variance of sum of independent random variables

Variance of sum of independent random variables

Result

Let X and Y be independent random variables. Then

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

Example: Rolling a dice twice

Example: Rolling a dice twice

- Let X be the outcome of a fair dice. Let Y be the outcome of another fair dice.

Example: Rolling a dice twice

- Let X be the outcome of a fair dice. Let Y be the outcome of another fair dice.
- We know $E(X) = E(Y) = 3.5$

Example: Rolling a dice twice

- Let X be the outcome of a fair dice. Let Y be the outcome of another fair dice.
- We know $E(X) = E(Y) = 3.5$
- $X + Y$ is the sum of outcomes of both the dice rolled together.

Example: Rolling a dice twice

- Let X be the outcome of a fair dice. Let Y be the outcome of another fair dice.
- We know $E(X) = E(Y) = 3.5$
- $X + Y$ is the sum of outcomes of both the dice rolled together. Then, we know $E(X + Y) = E(X) + E(Y) = 3.5 + 3.5 = 7$.
- We also know $Var(X) = Var(Y) = 2.917$

Example: Rolling a dice twice

- Let X be the outcome of a fair dice. Let Y be the outcome of another fair dice.
- We know $E(X) = E(Y) = 3.5$
- $X + Y$ is the sum of outcomes of both the dice rolled together. Then, we know $E(X + Y) = E(X) + E(Y) = 3.5 + 3.5 = 7$.
- We also know $Var(X) = Var(Y) = 2.917$
- X and Y are independent, hence, $Var(X + Y) = 2.917 + 2.917 \approx 5.83$, which is the same as what we obtained earlier applying the computational formula.

Hypergeometric random variable

Hypergeometric random variable

- Suppose that a sample of size n is to be chosen randomly (without replacement) from a box containing N balls, of which m are red and $N - m$ are blue.

Hypergeometric random variable

- Suppose that a sample of size n is to be chosen randomly (without replacement) from a box containing N balls, of which m are red and $N-m$ are blue.
- Let X denote the number of red balls selected, then

$$P(X=i)=\frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}}, i=0,1,2,\ldots,n$$

Hypergeometric random variable

- Suppose that a sample of size n is to be chosen randomly (without replacement) from a box containing N balls, of which m are red and $N - m$ are blue.
- Let X denote the number of red balls selected, then

$$P(X = i) = \frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}}, i = 0, 1, 2, \ldots, n$$

- X is a said to be a hypergeometric variable for some values of n, m, and N

Hypergeometric random variable

- Suppose that a sample of size n is to be chosen randomly (without replacement) from a box containing N balls, of which m are red and $N - m$ are blue.
- Let X denote the number of red balls selected, then

$$P(X = i) = \frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}}, i = 0, 1, 2, \ldots, n$$

- X is a said to be a hypergeometric variable for some values of n, m, and N
- $E(X) = \frac{nm}{N}$

Hypergeometric random variable

- Suppose that a sample of size n is to be chosen randomly (without replacement) from a box containing N balls, of which m are red and $N-m$ are blue.
- Let X denote the number of red balls selected, then

$$P(X=i)=\frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}}, i=0,1,2,\ldots,n$$

- X is a said to be a hypergeometric variable for some values of $n, m,$ and N
- $E(X)=\frac{nm}{N}$
- It can be verified that $Var(X)=\frac{nm}{N}\left[\frac{(n-1)(m-1)}{(N-1)}+1-\frac{nm}{N}\right]$

Variance of sum of many independent random variables

Variance of sum of many independent random variables

- The result that the variance of the sum of independent random variables is equal to the sum of the variances holds for not only two but any number of random variables.

Variance of sum of many independent random variables

- The result that the variance of the sum of independent random variables is equal to the sum of the variances holds for not only two but any number of random variables.
- Let $X_1, X_2, \ldots, X_k$ be k discrete random variables. Then,

$$Var\left(\sum_{i=1}^{k} X_i\right) = \sum_{i=1}^{k} Var(X_i)$$

Example: Tossing a coin three times

Example: Tossing a coin three times

- Toss a fair coin i times.
- Let X_i be a random variable which equals 1 if the outcome is a head, 0 otherwise.

Example: Tossing a coin three times

- Toss a fair coin i times.
- Let X_i be a random variable which equals 1 if the outcome is a head, 0 otherwise.
- $E(X_i) = 0.5, Var(X_i) = 0.25$

Example: Tossing a coin three times

- Toss a fair coin i times.
- Let X_i be a random variable which equals 1 if the outcome is a head, 0 otherwise.
- $E(X_i) = 0.5, Var(X_i) = 0.25$
- $X_1 + X_2 + \ldots + X_n$ is the total number of heads in n tosses of the coin.

Example: Tossing a coin three times

- Toss a fair coin i times.
- Let X_i be a random variable which equals 1 if the outcome is a head, 0 otherwise.
- $E(X_i) = 0.5, Var(X_i) = 0.25$
- $X_1 + X_2 + \ldots + X_n$ is the total number of heads in n tosses of the coin.
- $Var(X_1 + X_2 + \ldots + X_n) = \sum_{i=1}^{n} Var(X_i) = 0.5 \times n$
- For $n = 3, X_1 + X_2 + X_3$ is equal to the number of heads in three tosses of a coin.

$$Var(X_1 + X_2 + X_3) = 3 \times 0.25 = 0.75$$

This is the same as variance of number of heads in three tosses of a coin.

Section summary

- Properties of variance.
- Variance of sum of independent random variables.