BASIC STATISTIC

Hoang-Nguyen Vu

Objectives

Introduction

- ❖ Random Variables
- ❖ Probability Distribution
- ❖ Mathematical Expectations (Mean, Median, Variance)

Application

- ❖ Covariance & Correlation

Outline

SECTION 1
Discrete Random Variables

SECTION 2
Mathematical Expectations

SECTION 3
Covariance & Correlation

Discrete Random Variables

A Random Variable

A random variable X is a function X: $\Omega \rightarrow \mathbb{R}$, maps an outcome s $\in \Omega$ to a number on the real line X(s) $\in \mathbb{R}$

A continuous random variable X is a function X(s): $\Omega \rightarrow \mathbb{R}$, maps an outcome s from **an uncountably infinite** to a number on the real line X(s) $\in \mathbb{R}$

A discrete random variable X is a function X(s): $\Omega \rightarrow \mathbb{R}$, maps an outcome from a finite or countably infinite sample space to a number on the real line X(s) $\in \mathbb{R}$

Discrete Random Variables

A Continuous Random Variable

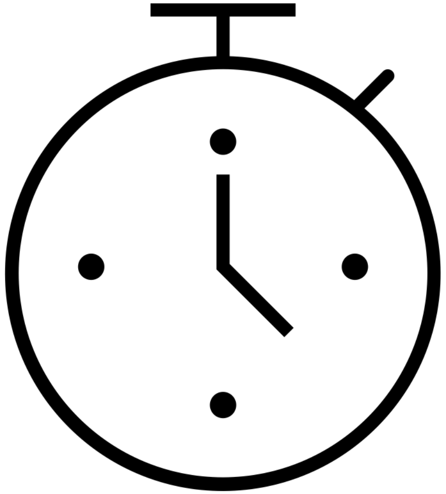
A continuous random variable X is a function X(s): Ω → ℝ, maps an outcome s from **an uncountably infinite** to a number on the real line X(s) ∈ ℝ

Rotate a pointer about a pivot in a plane (a clock)

Outcome: the angle where stops: $2\pi\theta, \theta \in (0,1]$

$\Omega = (0\ ,\ 1]$

A continuos random variable: $X(\theta) = \theta$



Discrete Random Variables

A Discrete Random Variable

A discrete random variable X is a function $X(s): \Omega \rightarrow \mathbb{R}$, maps an outcome from a finite or countably infinite sample space to a number on the real line $X(s) \in \mathbb{R}$

Toss a coin 3 times in sequence

$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$

- $X(s)$ = the number of Heads in the sequence

  $X(HTH) = 2 \qquad X(THT) = 1 \qquad \ldots$

Discrete Random Variables

A Discrete Random Variable

A discrete random variable X is a function $X(s): \Omega \rightarrow \mathbb{R}$, maps an outcome from a finite or countably infinite sample space to a number on the real line $X(s) \in \mathbb{R}$

Toss a coin 3 times in sequence

Ω = {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}

- $X(s)$ = the number of Heads in the sequence

  $X(HTH) = 2 \quad X(THT) = 1 \quad \ldots$

- $X(s) \begin{cases} \text{The index of the first H} \\ 0 \text{ if the sequence has no H} \end{cases}$

  $X(TTH) = 3 \quad X(TTT) = 0 \quad \ldots$

AI VIET NAM
@aivietnam.edu.vn

Discrete Random Variables

A Discrete Random Variable

A discrete random variable X is a function $X(s): \Omega \rightarrow \mathbb{R}$, maps an outcome from a finite or countably infinite sample space to a number on the real line $X(s) \in \mathbb{R}$

Toss a coin 3 times in sequence

$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$

- $X(s)$ = the number of Heads in the sequence

  $X(s) = 2$

  => corresponds to the event $s = \{HHT, HTH, THH\}$

  $1 < X(s) \leq 3$

  => $s = \{HHH, HHT, HTH, THH\}$

Discrete Random Variables

A Discrete Random Variable

Probability measure on random variables

Toss a coin 3 times in sequence

Ω = {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}

- X(s) = the number of Heads in the sequence

  X(s) = 2 => corresponds to the event s = {HHT, HTH, THH}
  => P(X=2) = P({HHT, HTH, THH}) = 3/8
  $1 < X(s) \leq 3$ => s = {HHH, HHT, HTH, THH}
  => $P(1 < X \leq 3)$ = P({HHH, HHT, HTH, THH}) = 4/8

Question

- P(X=1)
- P(X=2)
- P(X<3)
- $P(X \leq -1)$
- $P(X \leq 3)$
- $P(1 < X \leq 3)$

Discrete Random Variables

Probability Mass Function

Probability mass function

$$p_X(x) = P(X = x)$$

Toss a coin 3 times in sequence

Ω = {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}

X(s) = the number of Heads in the sequence

- p(0) = 1/8
- p(1) = 3/8
- p(2) = 3/8
- p(3) = 1/8

AI VIET NAM
@aivietnam.edu.vn

Discrete Random Variables

Probability Distribution Function

Probability mass function
$$p_X(x) = P(X = x)$$

(Cumulative) Probability distribution function
$$F_X(x) = P(X \leq x)$$

Toss a coin 3 times in sequence

Ω = {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}

X(s) = the number of Heads in the sequence

- $p(0) = 1/8$
- $p(1) = 3/8$
- $p(2) = 3/8$
- $p(3) = 1/8$

- $F(-1) = P(X \leq -1) = 0/8$
- $F(0) = P(X \leq 0) = 1/8$
- $F(1) = P(X \leq 1) = 4/8$
- $F(2) = P(X \leq 2) = 7/8$

- $F(3) = P(X \leq 3) = 1$
- $F(4) = P(X \leq 4) = 1$

AI VIET NAM
@aivietnam.edu.vn

Discrete Random Variables

Probability Distribution Function

(Cumulative) probability distribution function

$$F_X(x) = P(X \le x)$$

Toss a coin 3 times in sequence

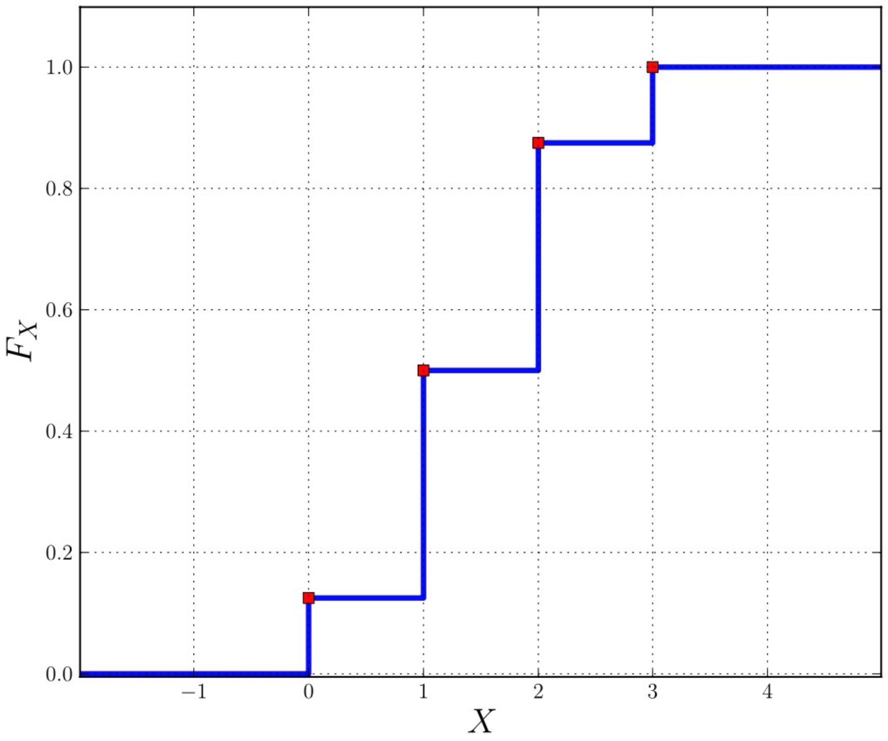
Ω = {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}

X(s) = the number of Heads in the sequence

- $F(-1) = P(X \le -1) = 0/8$
- $F(0) = P(X \le 0) = 1/8$
- $F(1) = P(X \le 1) = 4/8$
- $F(2) = P(X \le 2) = 7/8$
- $F(3) = P(X \le 3) = 1$
- $F(4) = P(X \le 4) = 1$

$P(0 < X \le 2) = P(X = 1) + P(X = 2) = F(2) - F(0)$

The graph of the probability distribution function



Discrete Random Variables

Practice

Throw a die twice. X is sum of two numbers. P[X=7]?

Throw a die twice in sequence

$\Omega = \{(1,1), (1,2),\ldots, (2,1), (2,2),\ldots, (6,1), (6,2),\ldots,(6,6)\}$ => $n(\Omega)=6*6=36$

X = sum of two numbers

$P[X=7] = P[\{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\}]$

$P[X=7] = P[(1,6)] + P[(2,5)] + P[(3,4)] + P[(4, 3)] + P[(5,2)] + P[(6,1)]$

$$P[X = 7] = \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} = \frac{1}{6}$$

Discrete Random Variables

Practice

Throw a die twice. X is sum of two numbers. F[3]?

Throw a die twice in sequence

$\Omega = \{(1,1), (1,2), \ldots, (2,1), (2,2), \ldots, (6,1), (6,2), \ldots, (6,6)\}$ => $n(\Omega)=6*6=36$

X = sum of two numbers

$$F(3) = P[X \leq 3] = P[2] + P[3] = P[(1,1)] + P[\{(1,2), (2,1)\}]$$

$$F(3) = P[X \leq 3] = P[(1,1)] + P[(1,2)] + P[(2,1)]$$

$$F(3) = P[X \leq 3] = \frac{1}{36} + \frac{1}{36} + \frac{1}{36} = \frac{1}{12}$$

Discrete Random Variables

Practice

We extract 3 balls from a box containing an equal number of white and black balls, and we put the ball back into the box after each extraction. How many white balls can occur and what are the associated probabilities?

X = Number of white balls

Sample space	X	Pro	Sample space	X	Pro
WWW	3	1/8	WBB	1	1/8
WWB	2	1/8	BWB	1	1/8
WBW	2	1/8	BBW	1	1/8
BWW	2	1/8	BBB	0	1/8

X	0	1	2	3
P(X)	1/8	3/8	3/8	1/8

Outline

SECTION 1
Discrete Random Variables

SECTION 2
Mathematical Expectations

SECTION 3
Covariance & Correlation

AI VIET NAM
@aivietnam.edu.vn

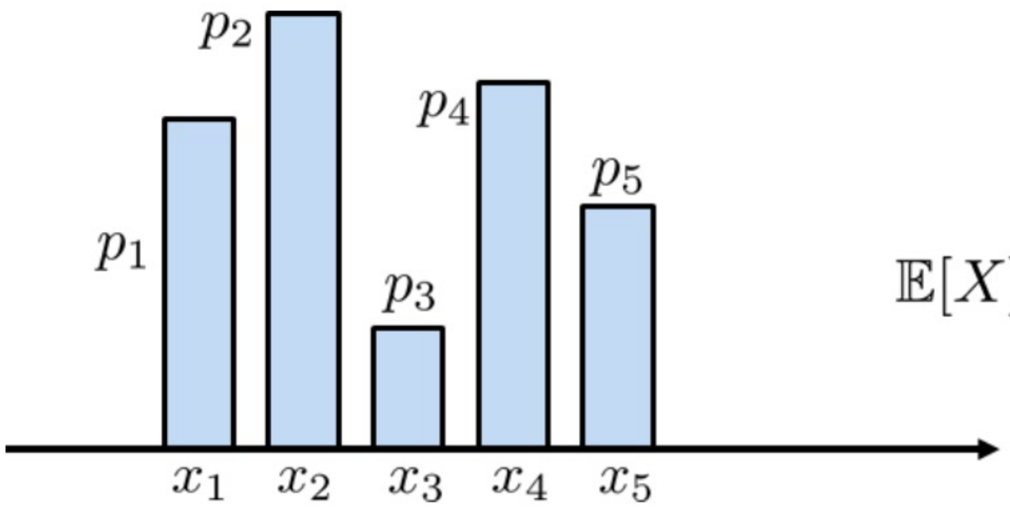
Mathematical Expectations

Expectation

The expected value of a discrete random variable X is

$$E[X] = \sum_k x_k \,.\, P(X = x_k) = \sum_k x_k \,.\, p_X(x_k)$$

The **E[X]** represents the weighted average value of **X**
E**[X]** is also called the mean of **X**



$$\mathbb{E}[X] = p_1 x_1 + \ldots + p_5 x_5$$

Mathematical Expectations

Expectation

The expected value of a discrete random variable X is

$$E[X] = \sum_k x_k \, . \, P(X = x_k) = \sum_k x_k \, . \, p_X(x_k)$$

Rolling a die

$\Omega = \{1, 2, 3, 4, 5, 6\}$

x	1	2	3	4	5	6
p(x)	1/6	1/6	1/6	1/6	1/6	1/6

$$E[X] = 1 * \frac{1}{6} + 2 * \frac{1}{6} + 3 * \frac{1}{6} + 4 * \frac{1}{6} + 5 * \frac{1}{6} + 6 * \frac{1}{6} = \frac{7}{2}$$

Prove the following:

- $E[\alpha X] = \alpha E[X]$
- $E[\alpha X + b] = \alpha E[X] + b$

Mathematical Expectations

Expectation

The expected value of a discrete random variable X is

$$E[X] = \sum_k x_k \,.\, P(X = x_k) = \sum_k x_k \,.\, p_X(x_k)$$

Toss a coin 3 times in sequence

Ω = {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}

X(s) = the number of Heads in the sequence

- $p(0) = 1/8$
- $p(1) = 3/8$
- $p(2) = 3/8$
- $p(3) = 1/8$

x	0	1	2	3
p(x)	1/8	3/8	3/8	1/8

$$E[X] = 0 * \frac{1}{8} + 1 * \frac{3}{8} + 2 * \frac{3}{8} + 3 * \frac{1}{8} = \frac{3}{2}$$

Mathematical Expectations

Variance

Mean of X:
μ = E[X]

Variance (The average weighted **square distance** from the mean) of X:
$\mathrm{Var}(X) = \mathrm{E}[(X-\mu)^2] = \sum_k (x_k - \mu)^2 p(x_k)$

Toss a coin 2 times in sequence

Ω = {HH, TH, HT, TT}

X(s) = the number of Heads in the sequence

- p(0) = 1/4
- p(1) = 2/4
- p(2) = 1/4

x	0	1	2
p(x)	1/4	2/4	1/4

$$E[X] = 0 * \frac{1}{4} + 1 * \frac{2}{4} + 2 * \frac{1}{4} = 1$$

$$Var(X) = (0-1)^2 * \frac{1}{4} + (1-1)^2 * \frac{2}{4} + (2-1)^2 * \frac{1}{4} = 0.5$$

Mathematical Expectations

Variance

Mean of X:
μ = E[X]

Variance (The average weighted **square distance** from the mean) of X:
$\mathrm{Var}(X) = \mathrm{E}[(X-\mu)^2] = \sum_k (x_k-\mu)^2 p(x_k) = E[X^2]-\mu^2$

Toss a coin 2 times in sequence

Ω = {HH, TH, HT, TT}

X(s) = the number of Heads in the sequence

- p(0) = 1/4
- p(1) = 2/4
- p(2) = 1/4

x	0	1	2
p(x)	1/4	2/4	1/4

$$E[X] = 0*\frac{1}{4} + 1*\frac{2}{4} + 2*\frac{1}{4} = 1$$

$$E[X^2] = 0*\frac{1}{4} + 1*\frac{2}{4} + 4*\frac{1}{4} = 1.5$$

$$\boldsymbol{Var(X) = 1.5 - 1 = 0.5}$$

Mathematical Expectations

Standard Deviation

Standard Deviation (The average weighted **distance** from the mean) of X:

$$\sigma(X) = \sqrt{Var(X)} = \sqrt{E[(X-\mu)^2]} = \sqrt{E[X^2] - \mu^2}$$

Toss a coin 2 times in sequence

$$\Omega = \{HH, TH, HT, TT\}$$

X(s) = the number of Heads in the sequence

- p(0) = 1/4
- p(1) = 2/4
- p(2) = 1/4

x	0	1	2
p(x)	1/4	2/4	1/4

$$E[X] = 0 * \frac{1}{4} + 1 * \frac{2}{4} + 2 * \frac{1}{4} = 1$$

$$\mathrm{Var}(X) = 1.5 - 1 = 0.5$$

$$\sigma(X) = \sqrt{1.5 - 1} = \sqrt{0.5} = 0.707$$

Mathematical Expectations

Standard Deviation

Standard Deviation (The average weighted **distance** from the mean) of X:

$$\sigma(X) = \sqrt{Var(X)} = \sqrt{\mathrm{E}[(X-\mu)^2]} = \sqrt{E[X^2]-\mu^2}$$

Rolling a die

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

x	1	2	3	4	5	6
p(x)	1/6	1/6	1/6	1/6	1/6	1/6

$$E[X] = \frac{7}{2}$$

$$E[X^2] = \sum_{k=1}^{6}\left[k^2 * \frac{1}{6}\right] = \frac{(1+4+9+16+25+36)}{6} = \frac{91}{6}$$

Mathematical Expectations

Standard Deviation

Standard Deviation (The average weighted **distance** from the mean) of X:

$$\sigma(X) = \sqrt{Var(X)} = \sqrt{E[(X-\mu)^2]} = \sqrt{E[X^2] - \mu^2}$$

Rolling a die

$\Omega = \{1, 2, 3, 4, 5, 6\}$

x	1	2	3	4	5	6
p(x)	1/6	1/6	1/6	1/6	1/6	1/6

$$E[X] = \frac{7}{2} \qquad E[X^2] = \frac{91}{6} \qquad \mathrm{Var(X)} = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12} \qquad \sigma(X) = \sqrt{\frac{35}{12}} \approx 1.7$$

Mathematical Expectations

Standard Deviation

X = {1, 3, 4, 4}

$$E[X] = \mu = \sum_{k} x_k \,.\, p_X(x_k) \qquad Var(X) = \sum_{k} (x_k - \mu)^2 p(x_k) \qquad \sigma(X) = \sqrt{Var(X)}$$

x	1	3	4
times	1	1	2
p(x)	1/4	1/4	1/2

$$E[X] = \mu = 1 * \frac{1}{4} + 3 * \frac{1}{4} + 4 * \frac{1}{2} = 3$$

$$Var(X) = \frac{1}{4} * (1-3)^2 + \frac{1}{4} * (3-3)^2 + \frac{1}{2} * (4-3)^2 = 1.5$$

$$\sigma(X) = \sqrt{Var(X)} = \sqrt{1.5} \approx 1.22$$

Mathematical Expectations

Standard Deviation

X = {1, 3, 4, 4}

$$E[X] = \mu = \sum_{k} x_k \, . \, p_X(x_k) \qquad Var(X) = \sum_{k} (x_k - \mu)^2 p(x_k) \qquad \sigma(X) = \sqrt{Var(X)}$$

Mean: $\mu = \frac{1}{n} \sum_{k} x_k$

Variance: $Var(X) = \frac{1}{n} \sum_{k} (x_k - \mu)^2$

Standard Deviation: $\sigma(X) = \sqrt{Var(X)}$

$$\mu = \frac{1}{4}(1 + 3 + 4 + 4) = 3$$

$$Var(X) = \frac{1}{4}[(1-3)^2 + (3-3)^2 + (4-3)^2 + (4-3)^2] = 1.5$$

$$\sigma(X) = \sqrt{Var(X)} = \sqrt{1.5} \approx 1.22$$

Mathematical Expectations

Standard Deviation

X = {1, 3, 4, 4}

$$E[X] = \mu = \sum_k x_k \,.\, p_X(x_k) \qquad Var(X) = \sum_k (x_k - \mu)^2 p(x_k) \qquad \sigma(X) = \sqrt{Var(X)}$$

$$\text{Mean: } \mu = \frac{1}{n}\sum_k x_k$$

$$\text{Variance: } Var(X) = \frac{1}{n}\sum_k (x_k - \mu)^2$$

$$\text{Standard Deviation: } \sigma(X) = \sqrt{Var(X)}$$

```
data = np.array([1, 3, 4, 4])
print(data)

print("Mean: ", np.mean(data))
print("Std: ", np.std(data))
print("Variance: ", np.var(data))
```

```
[1 3 4 4]
Mean:  3.0
Std:  1.224744871391589
Variance:  1.5
```

Mathematical Expectations

Practice

X = {1, 2, 8, 5}

$$E[X] = \mu = \sum_{k} x_k \,.\, p_X(x_k) \qquad Var(X) = \sum_{k} (x_k - \mu)^2 p(x_k) \qquad \sigma(X) = \sqrt{Var(X)}$$

Mean: $\mu = \frac{1}{n}\sum_{k} x_k$

Variance: $Var(X) = \frac{1}{n}\sum_{k} (x_k - \mu)^2$

Standard Deviation: $\sigma(X) = \sqrt{Var(X)}$

$$\mu = \frac{1}{4}(1 + 2 + 8 + 5) = 4$$

$$Var(X) = \frac{1}{4}[(1-4)^2 + (2-4)^2 + (8-4)^2 + (5-4)^2] = 7.5$$

$$\sigma(X) = \sqrt{Var(X)} = \sqrt{7.5} \approx 2.74$$

Mathematical Expectations

Practice

Find the variance and standard deviation of the following probability distribution table?

X	0	1	2	3
P(X)	0.1	0.2	0.4	0.3

$$\mathrm{E[X]} = \mu = \sum_{\mathrm{k}} \mathrm{x_k} \,.\, \mathrm{p_X(x_k)}$$

$$\mathrm{Var(X)} = \sum_k (x_k - \mu)^2 p(x_k)$$

$$\sigma(\mathrm{X}) = \sqrt{Var(X)}$$

$$E[X] = 0 * 0.1 + 1 * 0.2 + 2 * 0.4 + 3 * 0.3 = 1.9$$

$$\mathrm{Var(X)} = (0 - 1.9)^2 * 0.1 + (1 - 1.9)^2 * 0.2 + (2 - 1.9)^2 * 0.4 + (3 - 1.9)^2 * 0.3 = 0.89$$

$$\sigma(\mathrm{X}) = \sqrt{0.89} = 0.94$$

AI VIET NAM
@aivietnam.edu.vn

Mathematical Expectations

Practice

Flip a coin with probability p to get a head. Let X be a random variable denoting the outcome. Find E[X] and Var[X] based on the PMF of X is

$$p_X(0) = 1 - p, \qquad p_X(1) = p$$

$$\mathrm{E[X]} = \mu = \sum_{\mathrm{k}} \mathrm{x_k} \,.\, \mathrm{p_X(x_k)}$$

$$\mathrm{Var(X)} = E[X^2] - \mu^2$$

$$\sigma(\mathrm{X}) = \sqrt{Var(X)}$$

$$E[X]$$
$$= 0 * p_X(0) + 1 * p_X(1)$$
$$= 0 * (1 - p) + 1 * p = p$$

$$\mathrm{Var(X)} = \mathrm{E[X^2] - E[X]^2}$$
$$= 0^2 * \mathrm{p_X}(0) + 1^2 * \mathrm{p_X}(1) - \mathrm{E[X]}^2$$
$$= 0 * (1 - \mathrm{p}) + 1 * \mathrm{p} - \mathrm{p}^2 = \mathrm{p}(1 - \mathrm{p})$$

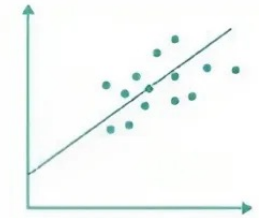
QUIZ TIME

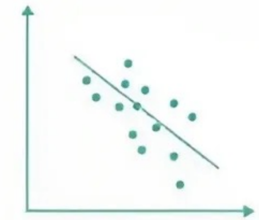
Outline

SECTION 1
Discrete Random Variables

SECTION 2
Mathematical Expectations

SECTION 3
Covariance & Correlation


Covariance


Correlation

Covariance & Correlation

Covariance

X, Y: random variables

$E[X] = \mu_X; \ E[Y] = \mu_Y$

Covariance of X and Y:

Sample

$$\mathrm{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = \frac{\sum(x - \mu_X)(y - \mu_Y)}{n - 1}$$

X = {1, 3, 4, 4}

$$E[X] = \mu = \frac{1}{4}(1 + 3 + 4 + 4) = 3$$

$$\mathrm{Var}(X) = \frac{1}{4}[(1 - 3)^2 + (3 - 3)^2 + (4 - 3)^2 + (4 - 3)^2] = 1.5$$

$$\sigma(X) = \sqrt{\mathrm{Var}(X)} = \sqrt{1.5} \approx 1.22$$

Y = {1, 2, 3, 2}

$$E[Y] = \mu = \frac{1}{4}(1 + 2 + 3 + 2) = 2$$

$$\mathrm{Var}(Y) = \frac{1}{4}[(1 - 2)^2 + (2 - 2)^2 + (3 - 2)^2 + (2 - 2)^2] = 0.5$$

$$\sigma(Y) = \sqrt{\mathrm{Var}(Y)} = \sqrt{0.5} \approx 0.707$$

Covariance & Correlation

Covariance

X, Y: random variables
$E[X] = \mu_X;\ E[Y] = \mu_Y$

Covariance of X and Y:

$$\mathrm{Cov}(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] = \frac{\sum(x-\mu_X)(y-\mu_Y)}{n-1}$$

$\mathbf{X = \{1, 3, 4, 4\}}$

$E[X] = \mu_X = 3$

$\mathbf{Y = \{1, 2, 3, 2\}}$

$E[Y] = \mu_Y = 2$

$$\mathrm{Cov}(X,Y) = \frac{\sum(x-\mu_X)(y-\mu_Y)}{n-1}$$

$$= \frac{(1-3)(1-2) + (3-3)(2-2) + (4-3)(3-2) + (4-3)(2-2)}{4-1}$$

$$= \frac{2+1}{3} = 1$$

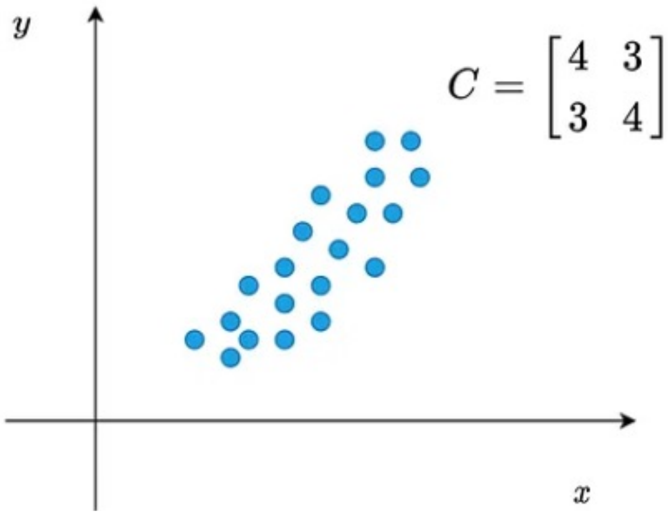
Covariance & Correlation

Covariance

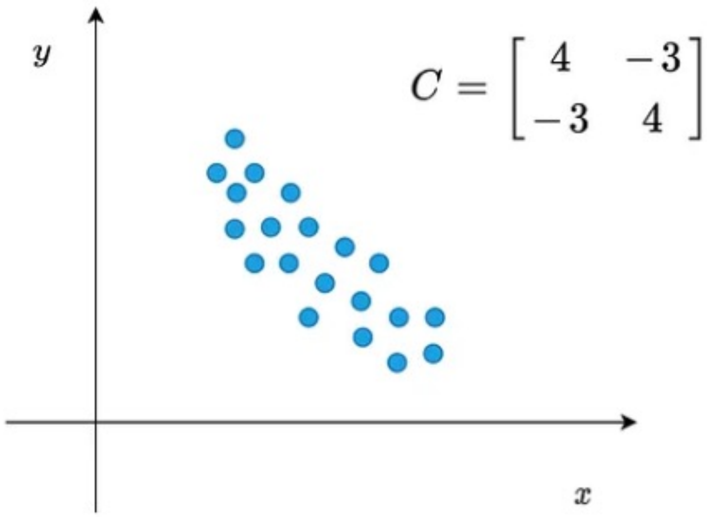
Cov(X,Y)
measures "concordance" or
"coherence" of X and Y

Covariance of X and Y:

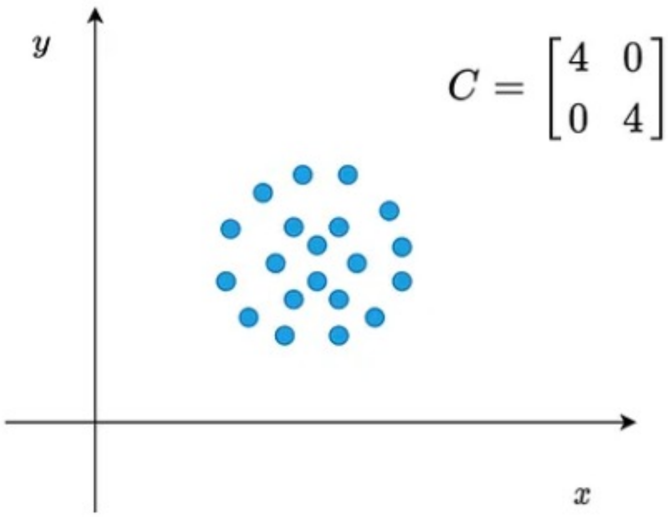
$$\mathrm{Cov}(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] = \frac{\sum(x-\mu_X)(y-\mu_Y)}{n-1}$$



Positive Covariance – Cov(X, Y) > 0



Negative Covariance – Cov(X, Y) < 0



Zero Covariance – Cov(X, Y) = 0

Covariance & Correlation

Correlation

A statistical measure the quantifies the strength and direction of a linear relationshop between two random variables

$$\mathbf{X = \{1, 3, 4, 4\}}$$

$$E[X] = \mu_X = 3$$

$$\mathbf{Y = \{1, 2, 3, 2\}}$$

$$E[Y] = \mu_Y = 2$$

Correlation of X and Y:

$$\mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{var}(X)}\sqrt{\mathrm{var}(Y)}}$$

$$= \frac{n(\sum_i x_i y_i) - (\sum_i x_i)(\sum_i y_i)}{\sqrt{n\sum_i x_i^2 - (\sum_i x_i)^2}\sqrt{n\sum_i y_i^2 - (\sum_i y_i)^2}}$$

$$\mathrm{Corr}(X, Y) = \frac{27n - 96}{\sqrt{42n - 144}\sqrt{18n - 64}} = \frac{12}{\sqrt{24}\sqrt{8}} \approx 0.866$$

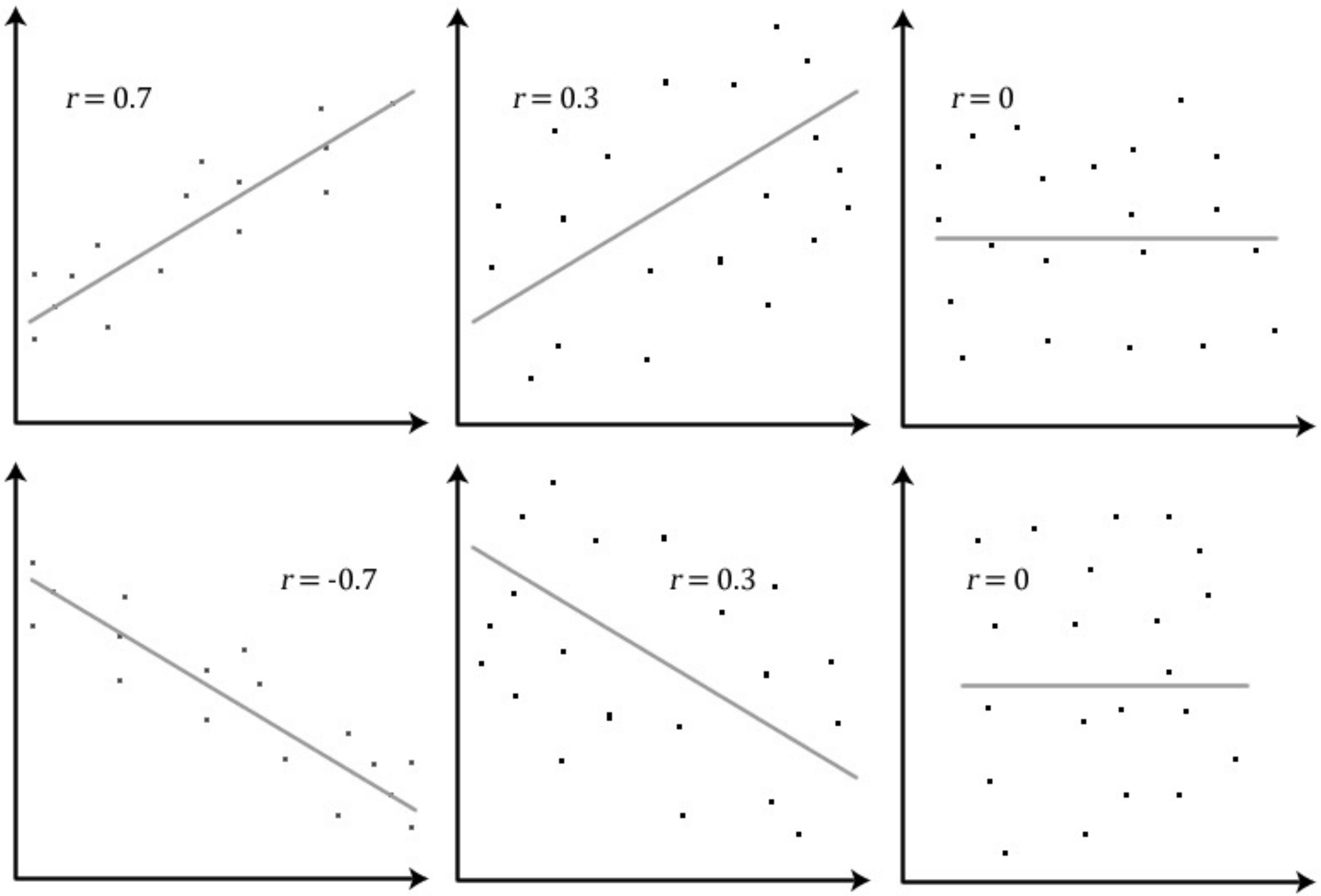
Covariance & Correlation

Correlation

Correlation of X and Y:

$$\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{var}(X)}\sqrt{\mathrm{var}(Y)}}$$

$$Corr(X,Y) \in [-1,1]$$



Covariance & Correlation

Correlation

Correlation of X and Y:

$$\mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{var}(X)}\sqrt{\mathrm{var}(Y)}}$$

$$Corr(X, Y) \in [-1,1]$$

```python
# Sample data for two variables
var1 = np.array([1, 3, 4, 4])
var2 = np.array([1, 2, 3, 2])

# Compute covariance
print("Covariance: \n", np.cov(var1, var2))

# Compute correlation coefficient
print("Correlation: \n", np.corrcoef(var1, var2))
```

```
Covariance: 
 [[2.         1.        ]
 [1.         0.66666667]]
Correlation: 
 [[1.        0.8660254]
 [0.8660254 1.       ]]
```

Covariance & Correlation

Practice

$$\mathbf{X = \{1, 1, 2, 4\}}$$

$$\mathbf{Y = \{2, 3, 4, 3\}}$$

$$E[X] = \mu_X = 2$$

$$E[Y] = \mu_Y = 3$$

$$\mathrm{Cov}(X, Y) = \frac{\sum(x - \mu_X)(y - \mu_Y)}{n - 1}$$

$$= \frac{(1-2)(2-3) + (1-2)(3-3) + (2-2)(4-3) + (4-2)(3-3)}{4 - 1} = \frac{1}{3}$$

$$\mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

$$= \frac{n(\sum_i x_i y_i) - (\sum_i x_i)(\sum_i y_i)}{\sqrt{n\sum_i x_i^2 - (\sum_i x_i)^2}\sqrt{n\sum_i y_i^2 - (\sum_i y_i)^2}}$$

$$= \frac{25n - 96}{\sqrt{22n - 64}\sqrt{38n - 144}} = \frac{4}{\sqrt{24}\sqrt{8}} \approx 0.288$$

AI VIET NAM
@aivietnam.edu.vn

Summary

Introduction

- Random Variable
- Discrete Random Variable
- Continuous Random Variable
- Probability Mass Function
- Probability Distribution Function
- Mean

$$\mu = \frac{1}{n}\sum_{k} x_k$$

- Variance

$$\mathrm{Var}(X) = \frac{1}{n}\sum_{k} (x_k - \mu)^2$$

- Standard Deviation

$$\sigma(X) = \sqrt{\mathrm{Var}(X)}$$

Application

- Covariance & Correlation

$$\mathrm{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$$
$$= \frac{\sum(x - \mu_X)(y - \mu_Y)}{n - 1}$$

$$\mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{var}(X)}\sqrt{\mathrm{var}(Y)}}$$
$$= \frac{n(\sum_i x_i y_i) - (\sum_i x_i)(\sum_i y_i)}{\sqrt{n\sum_i x_i^2 - (\sum_i x_i)^2}\sqrt{n\sum_i y_i^2 - (\sum_i y_i)^2}}$$

Thanks!

Any questions?