Computer Science

# CSC380: Principles of Data Science

## Probability Primer 5

Alon Efrat

Today:
- Expectation
- Variance
- Covariance
- Correlation

Credit:
- Jason Pacheco,
- Kwang-Sung Jun,
- Chicheng Zhang
- Xinchen yu

1

1

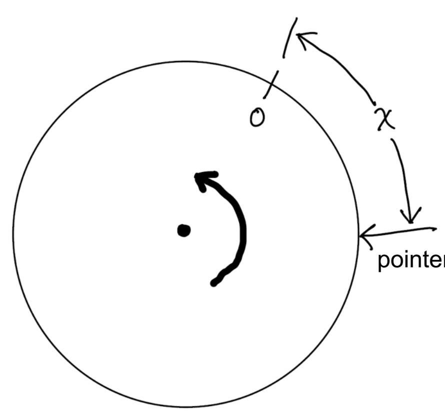
## Review: Continuous Probability

2

**Experiment** Spin continuous wheel and measure X displacement from 0

Say the circumference is 1.

Outcome space $\Omega$ is all points (real numbers) in $(0,1]$




**Question** Assuming uniform distribution, what is $P(X = x)$?

A much better question: What is P(X$\leq x$).
(that is, that the wheel reached at most x.
Examples P(X$\leq$ $0.5)$

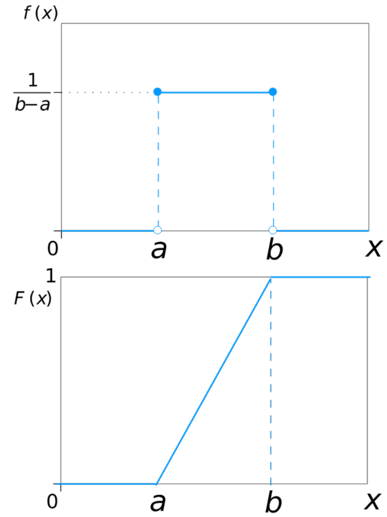
2

# Review: Continuous Random Variable

3

- Probability can be assigned to intervals
- Define CDF: $F(x) := P(X \leq x)$
- Then, PDF: $f(x) := p(X = x) := F'(x)$ // the slope at $F(x)$
- $P(X \in [a, b])$ = F(b) – F(a) // area under the PDF curve

**Uniform[0,1]**




**<u>Another viewpoint</u>**

- A continuous distribution is defined by PDF $f(x)$ whose area under the curve is 1
- Then, we can compute $P(X \in [a, b])$ by computing the area under the curve on [a,b].

Note:

$$P(X \in [a, b]) = P(X \in (a, b]) = P\big(X \in [a, b)\big) = P\big(X \in (a, b)\big)$$

3

# Review: Continuous Random Variable

4

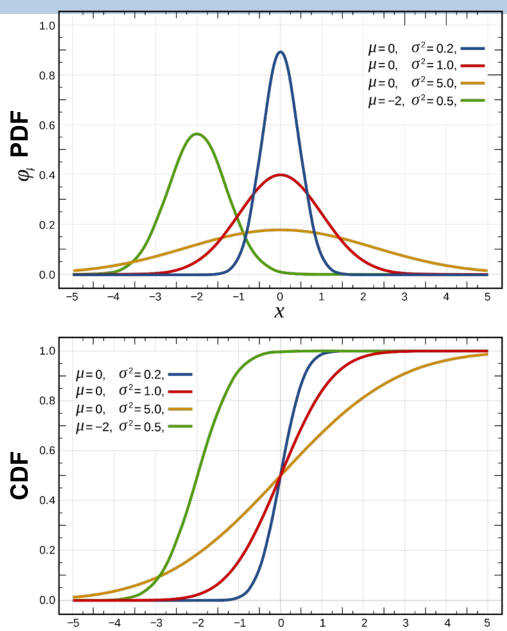
**Gaussian** (a.k.a. Normal) distribution with mean mean (location) $\mu$ and variance (scale) $\sigma^2$ parameters,

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(x-\mu)^2}{2\sigma^2})$$

Compactly, $X \sim \mathcal{N}(\mu, \sigma^2)$

**Probability Density function**

Cumulative Distribution Function




4

# Outline

5

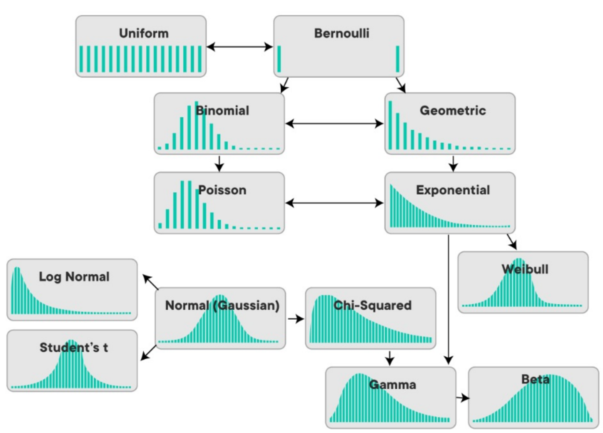
- Expectation
- Variance
- Covariance
- Correlation

5

# Moments of Random Variables

6

Q: How to describe characteristics of different distributions?




6

# Moments of Random Variables

7

Properties of a RV are characterized by its distribution / PMF / PDF
But there are "summary" numbers capturing important characteristics
This is called "**moments**".

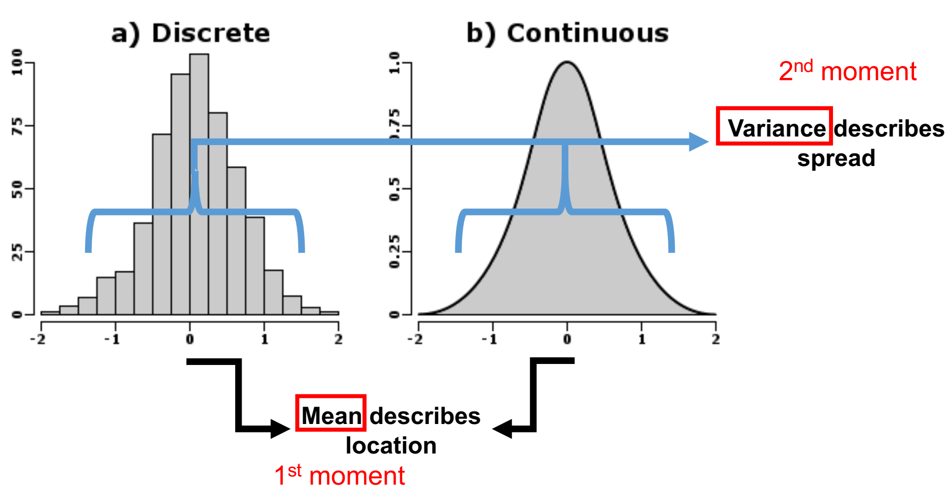
| Moment ordinal | Moment | | | Cumulant | |
|---|---|---|---|---|---|
| | Raw | Central | Standardized | Raw | Normalized |
| 1 | Mean | 0 | 0 | Mean | N/A |
| 2 | – | Variance | 1 | Variance | 1 |

(Wikipedia)

7

# Moments of Random Variables

8




*Moments characterize properties of the distribution "shape"*

8

# Expectation





## Expectation: a game-theoretic viewpoint



- Consider the following game:
  - Flip an unfair coin $\mathrm{X}$ with PMF
    - If $\mathrm{X} = 1$, you receive $\$1$
    - If $\mathrm{X} = -1$, you lose $\$1$

| outcome | prob. |
| --- | --- |
| $\mathrm{X} = 1$ | 0.7 |
| $\mathrm{X} = -1$ | 0.3 |

- How much are you willing to pay to play the game?
  - As long as you pay $\leq \$0.4$ per game, your wealth will not decrease in the long run
  - 'value of the game' = $\$0.4$

## Mean = Expectation = Expected Value

11

**Definition** *The* ***expectation*** *of a discrete RV* $X$*, denoted by* $\mathbf{E}[X]$*, is:*

(with PMF)

$$\mathbf{E}[X] = \sum_x x \cdot p(X = x)$$

Summation over all values in domain of X

Probability Mass Function

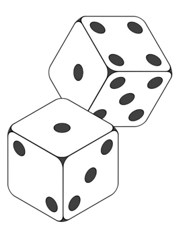
- **Effectively, a weighted average**: *each outcome weighted by probability of occurring*

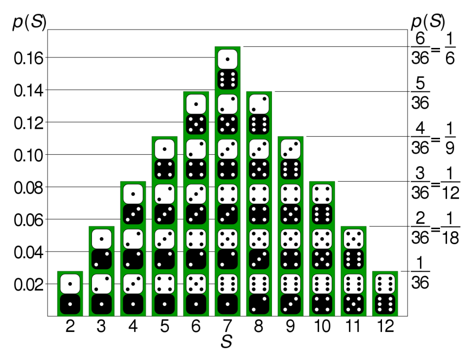
11

## Expected Value

Let X = sum of two dice, probability of S on different values:

$P(X = 2) = 1/36$
$P(X = 3) = 2/36$
$P(X = 4) = 3/36$
...
$P(X = 12) = 1/36$






Q: $\mathbf{E}[X]$ ?

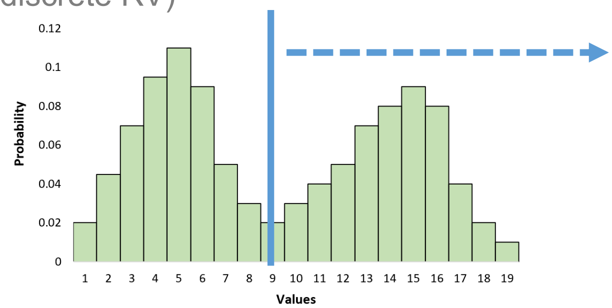
$$2 \cdot \frac{1}{36} + 3 \cdot \frac{2}{36} + \ldots + 6 \cdot \frac{5}{36} + 7 \cdot \frac{6}{36} + 8 \cdot \frac{5}{36} + \ldots + 12 \cdot \frac{1}{36} = 7$$

12

# Expected Value

13

(discrete RV)




*Expected value is not always a high probability event…*

*…in fact, it may not even be a feasible value…*

**Example** Let $X$ be the outcome of a fair die, then:

$$\mathbf{E}[X] = \frac{1}{6} \cdot (1 + 2 + 3 + 4 + 5 + 6) = 3.5$$

Can't actually roll 3.5

13

# Expected Value

14

**Theorem (Linearity of Expectations)** *For any finite collection of discrete RVs $X_1, X_2, \ldots, X_N$ with finite expectations,*

$$\mathbf{E}\left[\sum_{i=1}^{N} X_i\right] = \sum_{i=1}^{N} \mathbf{E}[X_i]$$

E.g. for two RVs X and Y
$\mathbf{E}[X+Y] = \mathbf{E}[X] + \mathbf{E}[Y]$

you do not need an independence!

**Example** Throw two fair dice. What is the expected sum? Let X and Y be the outcome of the first and second die, respectively.

$$\mathbf{E}[X+Y] = \mathbf{E}[X] + \mathbf{E}[Y] = 3.5 + 3.5 = 7$$

14

## Expected Value

15

**Proof:** $E[X+Y] = E[X] + E[Y]$

$$\sum_{i=1}^{3}\sum_{j=1}^{2} a_{ij} = \sum_{i=1}^{3}(a_{i1}+a_{i2}) = (a_{11}+a_{12}) + (a_{21}+a_{22}) + (a_{31}+a_{32}).$$

$$\mathbf{E}[X+Y] = \sum_i \sum_j (i+j)p(X=i, Y=j)$$

**Sum is linear operator**

$$= \sum_i \sum_j i \cdot p(X=i, Y=j) + \sum_i \sum_j j \cdot p(X=i, Y=j)$$

**Sum is linear operator**

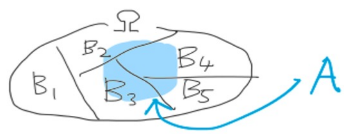
$$= \sum_i i \underline{\sum_j p(X=i, Y=j)} + \sum_j j \sum_i p(X=i, Y=j)$$

**Law of Total Probability**

$$= \sum_i i \cdot \underline{p(X=i)} + \sum_j j \cdot p(Y=j)$$

**By definition of Expectation**

$$= \mathbf{E}[X] + \mathbf{E}[Y]$$




**Sum of Summations**

$$\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i = \sum_{i=1}^{n}(x_i + y_i)$$

**Scaling of Summations**

$$\lambda \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} \lambda x_i$$

15

## Expected Value

16

**Theorem** For any random variable X and constant c,

$$\mathrm{E}[cX] = c\mathrm{E}[X]$$
$$\mathrm{E}[cX + k] = c\mathrm{E}[X] + k$$
$$\mathrm{E}[\mathrm{k}] = \mathrm{k}$$

Caveat: $k$ has to be a constant, not a random variable!

**Example** Throw two fair dice twice, X: outcome of 1st die, Y: outcome of 2nd die. The expected sum:

$$\mathbf{E}[2(X+Y)] = \mathbf{E}[2X] + \mathbf{E}[2Y]$$
$$= 2\mathbf{E}[X] + 2\mathbf{E}[Y]$$
$$= 2 \cdot 3.5 + 2 \cdot 3.5 = 14$$

16

# Conditional Expected Value

17

**Definition** *The* *<u>conditional expectation</u>* *of a discrete RV $X$, given $Y$ is:*

$$\mathbf{E}[X \mid Y = y] = \sum_x x\, p(X = x \mid Y = y)$$ cf. $\mathbf{E}[X] = \sum_x x \cdot p(X = x)$

**Example** Roll two fair dice. $X_1$: first die outcome, $Y$: sum of two dice is 5

quiz candidate

$$\mathbf{E}[X_1 \mid Y = 5] = \sum_{x=1}^{4} x\, p(X_1 = x \mid Y = 5)$$

$$= \sum_{x=1}^{4} x \frac{p(X_1 = x, Y = 5)}{p(Y = 5)} = \sum_{x=1}^{4} x \frac{1/36}{4/36} = \frac{5}{2}$$

*Conditional expectation follows properties of expectation (linearity, etc.)*

17

# Variance

18

18

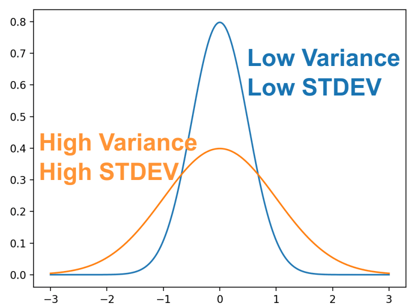
## Variance and Standard Deviation

19

**Definition** The variance of a RV $X$ is defined as,

$$\mathbf{Var}[X] = \mathbf{E}[(X - \mathbf{E}[X])^2] \qquad = E[(X-\mu)^2], \quad \text{when } \mu = E[X]$$

The standard deviation (STDEV) is $\sigma[X] = \sqrt{\mathbf{Var}[X]}$.




- *Describes the "spread" of a distribution*
- *Describes uncertainty of outcome*
- *STDEV is in original units (more intuitive), variance is in units²*
- *Variance is more mathematically useful than STDEV*

19

## Variance

20

**Example** Let X be the outcome of a fair SEVEN-sided die.

The expected value is E(X)=(1+2+3..+7)/7=28/7=4 $\boxed{\mathbf{Var}[X] = \mathbf{E}[(X - \mathbf{E}[X])^2]}$

The values of X-E(X) are {-3,-2,-1,0,1,2,3}

Y:=$(X-E(X))^2$ =(9,4,1,0,1,4,9)

Var(X)=E(Y)= *9 p(Y=9)+4 p(Y=4)+p(Y=1)+ 0 p(Y=0)*= 4

and STDEV(X)=$\sqrt{4}$=2

The STDEV is 2, so we should expect outcomes to vary around the mean of 4 by $\pm$ 2

20

## Variance

21

$E[2XE[X]] = 2E[XE[X]] = 2E[X]E[X]$

**Lemma** An equivalent form of variance is:

$E[X]$ is a constant

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

**Proof**

$$\mathbf{E}[(X - \mathbf{E}[X])^2] = \mathbf{E}[X^2 - \boxed{2X\mathbf{E}[X]} + \mathbf{E}[X]^2] \quad \textbf{(Expand it)}$$

$$= \mathbf{E}[X^2] - 2\mathbf{E}[X]\mathbf{E}[X] + \mathbf{E}[X]^2 \quad \textbf{(Linearity of expectations)}$$

$$= \mathbf{E}[X^2] - 2\mathbf{E}[X]^2 + \mathbf{E}[X]^2 \quad \textbf{(Algebra)}$$

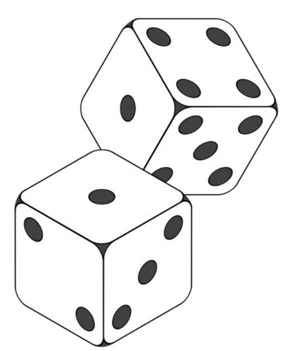
$$= \mathbf{E}[X^2] - \mathbf{E}[X]^2 \quad \textbf{(Algebra)}$$

21

## Variance

22

**Example** General form of variance for a fair **n-sided** fair die,



22

# Variance

23

- If c is a constant, $Var[cX] = c^2 Var[X]$
  - Exercise: try to convince yourself why this is true
  - Hint: use $\mathbf{E}[cX] = c\mathbf{E}[X]$

$\boxed{\mathbf{Var}[X] = \mathbf{E}[X^2] - (\mathbf{E}[X])^2}$

$$Var[cX] = E[(cX)^2] - (E[cX])^2$$

$$= E[c^2X^2] - (cE[X])^2$$

$$= c^2E[X^2] - c^2E[X]^2$$

$$= c^2(E[X^2] - E[X]^2)$$

23

# Moments of Useful Discrete Distributions

24

**Bernoulli** *A.k.a. the* ***coinflip*** *distribution on* *binary* *RVs* $X \in \{0, 1\}$

$$p(X) = \pi^X (1-\pi)^{(1-X)}$$

$\boxed{\mathbf{Var}[X] = \mathbf{E}[X^2] - (\mathbf{E}[X])^2}$

*Where* $\pi$ *is the probability of* ***success*** *(i.e., heads), and also the mean*

$$\mathbf{E}[X] = \pi \cdot 1 + (1-\pi) \cdot 0 = \pi \qquad \mathbf{Var}[X] = \pi(1-\pi)$$

$$E[X^2] = \pi \cdot 1^2 + (1-\pi) \cdot 0^2 = \pi$$

$$Var[X] = \pi - \pi^2$$

24

## Covariance – warm up

25

Assume $E[X] = E[Y] = 0$

We define in this case

$$Cov(X,Y) = E[X \cdot Y] = \sum_i \sum_j x_i \cdot y_j \cdot p(X = x_i, Y = y_j)$$

25

## Covariance

26

**Definition** *The covariance of two RVs $X$ and $Y$ is defined as,*

$$\mathbf{Cov}(X,Y) = \mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])]$$

**Question** *What is Cov(X,X)?*

**Answer** *Cov(X,X) = Var(X)*

In the case of discrete vars

$$Cov(X,Y) = \sum_i \sum_j (x_i - \mu_x)(y_i - \mu_i) \cdot p(X = x_i, Y = y_i)$$

$$Where \quad \mu_x = E(X), \quad \mu_y = E(Y)$$

26

## Covariance

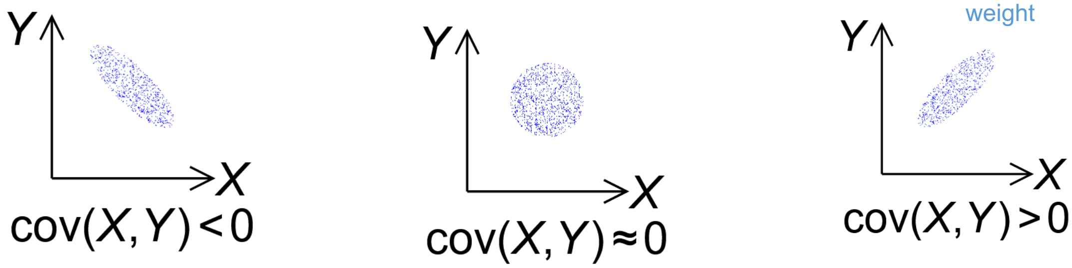
27

**Definition** *The covariance of two RVs $X$ and $Y$ is defined as,*

$$\mathbf{Cov}(X,Y) = \mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])]$$

Measures the linear relationship between X and Y

Example: height vs weight




Negative relationship

Positive relationship

If we are given a set of points $\{(x_1, y_1) \ldots (x_n, y_n)\}$ and pick from with uniform probability 1/n, then

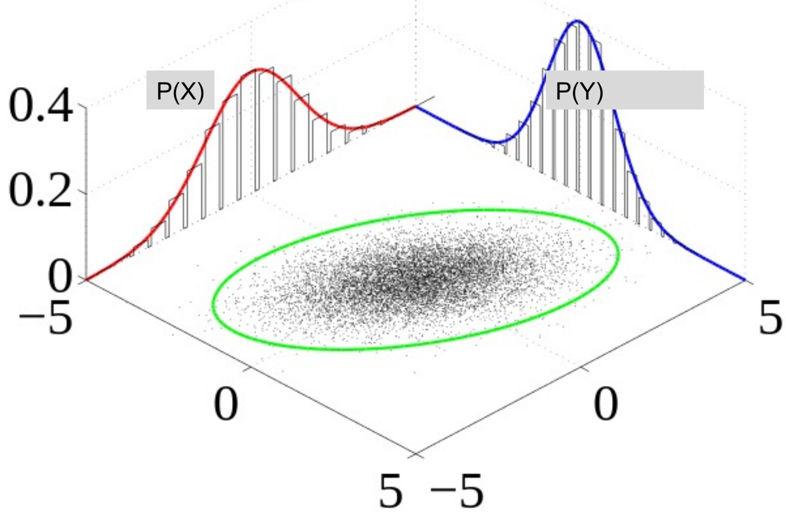
$$\mathrm{cov}(X,Y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - E(X))(y_i - E(Y)).$$

27

## Joint Probability Mass distribution p(X,Y)

28

- Credit: Wiki




28

## Properties



Credit: wikipedia

$$\begin{aligned}
\operatorname{cov}(X, a) &= 0 \\
\operatorname{cov}(X, X) &= \operatorname{var}(X) \\
\operatorname{cov}(X, Y) &= \operatorname{cov}(Y, X) \\
\operatorname{cov}(aX, bY) &= ab \operatorname{cov}(X, Y) \\
\operatorname{cov}(X + a, Y + b) &= \operatorname{cov}(X, Y)
\end{aligned}$$

$$|\operatorname{cov}(X, Y)| \leq \sqrt{\sigma^2(X)\sigma^2(Y)}$$



## Covariance



- A shortcut to compute covariance.
- $Cov(X, Y) = E[(X - E[X])(Y - E[Y])]$

  $= E\big[XY - X \cdot E[Y] - Y \cdot E[X] + E[X]E[Y]\big]$

  $= E[XY] - E[X]E[Y] - E[Y]E[X] + E[X]E[Y]$

  $= E[XY] - E[X]E[Y]$
- Safety check: $Cov(X, X) = E[XX] - E[X]E[X] = Var(X)$

# Covariance

31

**Lemma** *For any two RVs $X$ and $Y$,* $\boxed{\mathbf{Var}[X] = \mathbf{E}[(X - \mathbf{E}[X])^2]}$

$$\mathbf{Var}[X+Y] = \mathbf{Var}[X] + \mathbf{Var}[Y] + 2\mathbf{Cov}(X,Y)$$

=> *variance is* <u>*not a linear operator*</u>.

**Proof** $\mathbf{Var}[X+Y] = \mathbf{E}[(X+Y-\mathbf{E}[X+Y])^2]$

**(Linearity of expt.)** $= \mathbf{E}[(X+Y-\mathbf{E}[X]-\mathbf{E}[Y])^2]$

**(Distributive property)** $= \mathbf{E}[(X-\mathbf{E}[X])^2 + (Y-\mathbf{E}[Y])^2 + 2(X-\mathbf{E}[X])(Y-\mathbf{E}[Y])]$

**(Linearity of expt.)** $= \mathbf{E}[(X-\mathbf{E}[X])^2] + \mathbf{E}[(Y-\mathbf{E}[Y])^2] + 2\mathbf{E}[(X-\mathbf{E}[X])(Y-\mathbf{E}[Y])]$

**(Definition of Var / Cov)** $= \mathbf{Var}[X] + \mathbf{Var}[Y] + 2\mathbf{Cov}(X,Y)$

31

# Covariance

$Cov(X,Y) = E[XY] - E[X]E[Y]$

| |  |  |
|---|---|---|
| Person_1 | 1 | 1 |
| Person_2 | 3 | 0 |
| Person_3 | -1 | -1 |
| Expectation | E[A] | E[B] |

$$\begin{bmatrix} \begin{matrix} Cov(A,A) \\ Var(A) \end{matrix} & Cov(A,B) \\ Cov(B,A) & \begin{matrix} Cov(B,B) \\ Var(B) \end{matrix} \end{bmatrix}$$

$E[A] = \frac{1}{3}\cdot 1 + \frac{1}{3}\cdot 3 + \frac{1}{3}\cdot(-1) = 1, \quad E[B] = 0$

$Cov(A,B) = Cov(B,A)$
$= E[AB] - E[A]E[B]$
$= E[AB] - 0$
$= \frac{1}{3}\cdot 1 + \frac{1}{3}\cdot 0 + \frac{1}{3}\cdot 1 = \frac{2}{3}$

$Cov(A,A)$
$= E[A^2] - (E[A])^2$
$= \left(\frac{1}{3}\cdot 1 + \frac{1}{3}\cdot 9 + \frac{1}{3}\cdot 1\right) - 1$
$= \frac{8}{3}$

$Cov(B,B)$
$= E[B^2] - (E[B])^2$
$= \left(\frac{1}{3}\cdot 1 + \frac{1}{3}\cdot 0 + \frac{1}{3}\cdot 1\right) - 0$
$= \frac{2}{3}$

32

32

## Correlation

33

**Definition** *The correlation of two RVs $X$ and $Y$ is given by,*

$$\mathbf{Corr}(X,Y) = \frac{\mathbf{Cov}(X,Y)}{\sigma_X \sigma_Y} \quad \textit{where} \quad \sigma_X = \sqrt{\mathbf{Var}(X)}$$

Normalized version of covariance!
⇒ Always between -1 and 1

Useful when you are interested in how X and Y are related, independent of the individual variability.
⇒ $Cov(cX, dY) \neq Cov(X, Y)$ **but** $Corr(cX, dY) = Corr(X, Y)$

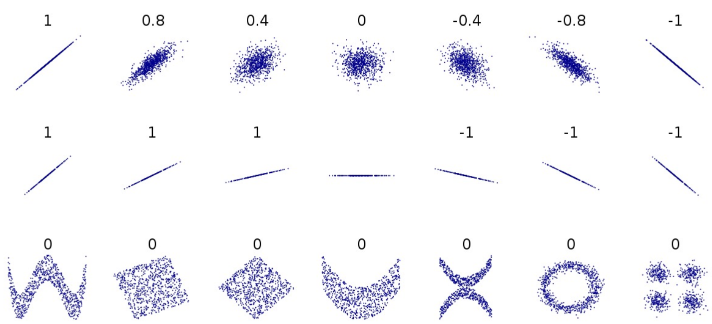
33

## Correlation

34

**Definition** *The correlation of two RVs $X$ and $Y$ is given by,*

$$\mathbf{Corr}(X,Y) = \frac{\mathbf{Cov}(X,Y)}{\sigma_X \sigma_Y} \quad \textit{where} \quad \sigma_X = \sqrt{\mathbf{Var}(X)}$$




*Like covariance, only expresses linear relationships!*

34