

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

Review: Continuous Random Variable

2

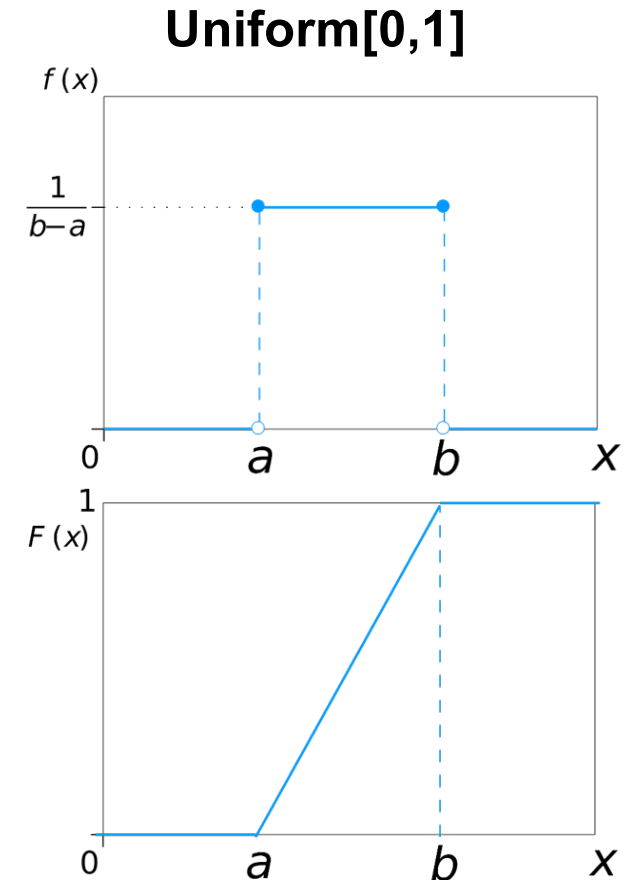
- Probability can be assigned to intervals
- Define CDF: $F(x) := P(X \leq x)$
- Then, PDF: $f(x) := p(X = x) := F'(x)$ // the slope at $F(x)$
- $P(X \in [a, b]) = F(b) - F(a)$ // area under the PDF curve

Another viewpoint

- 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:

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



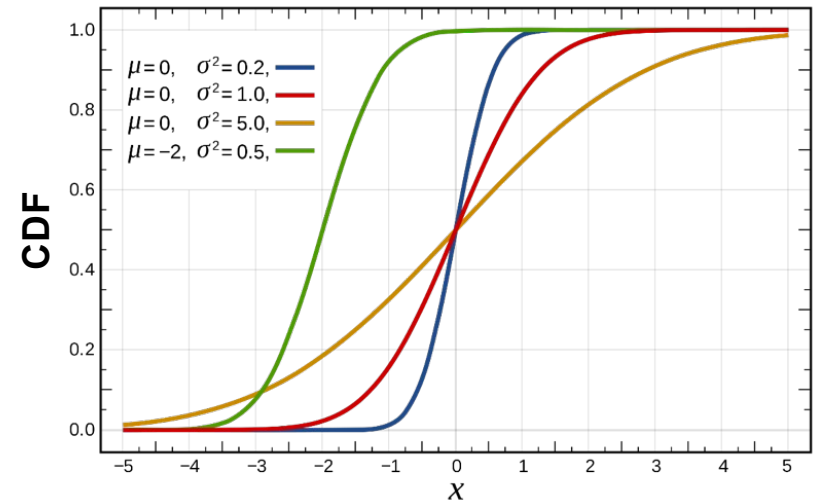
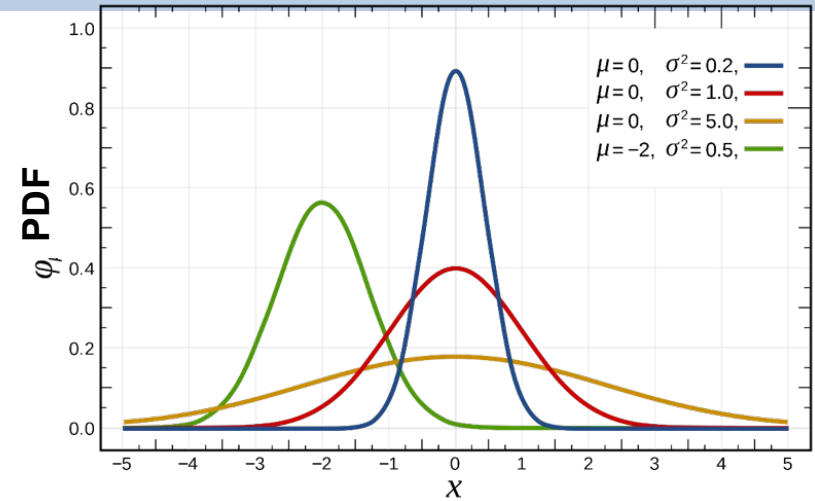
Review: Continuous Random Variable

3

Gaussian (a.k.a. Normal) distribution with mean mean (location) μ and variance (scale) σ^2 parameters,

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

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

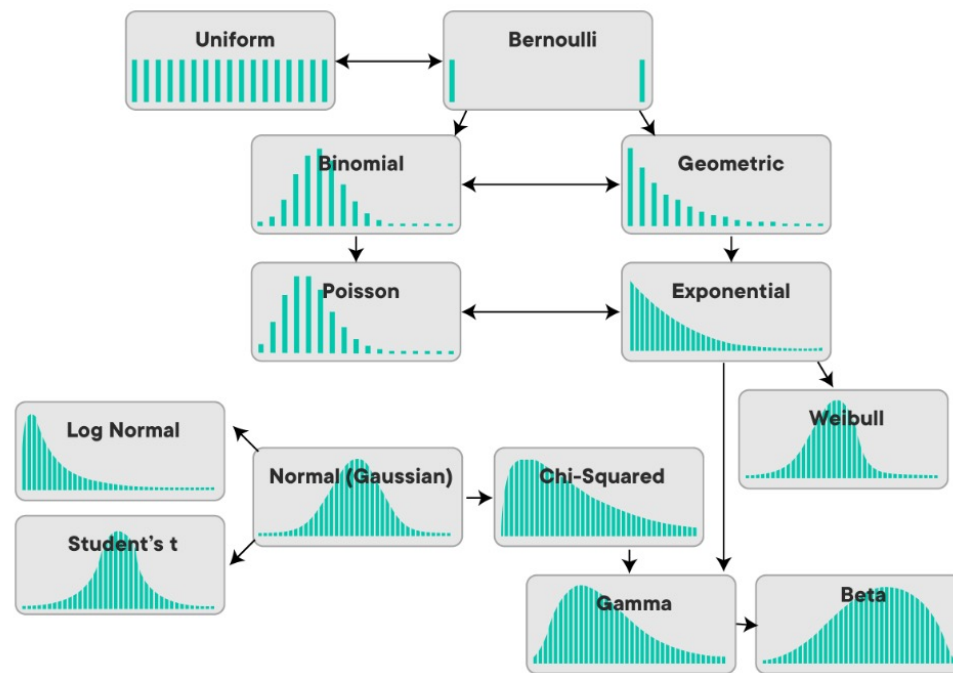


- Expectation
- Variance
- Covariance
- Correlation

Moments of Random Variables

5

Q: How to describe characteristics of different distributions?



Moments of Random Variables

6

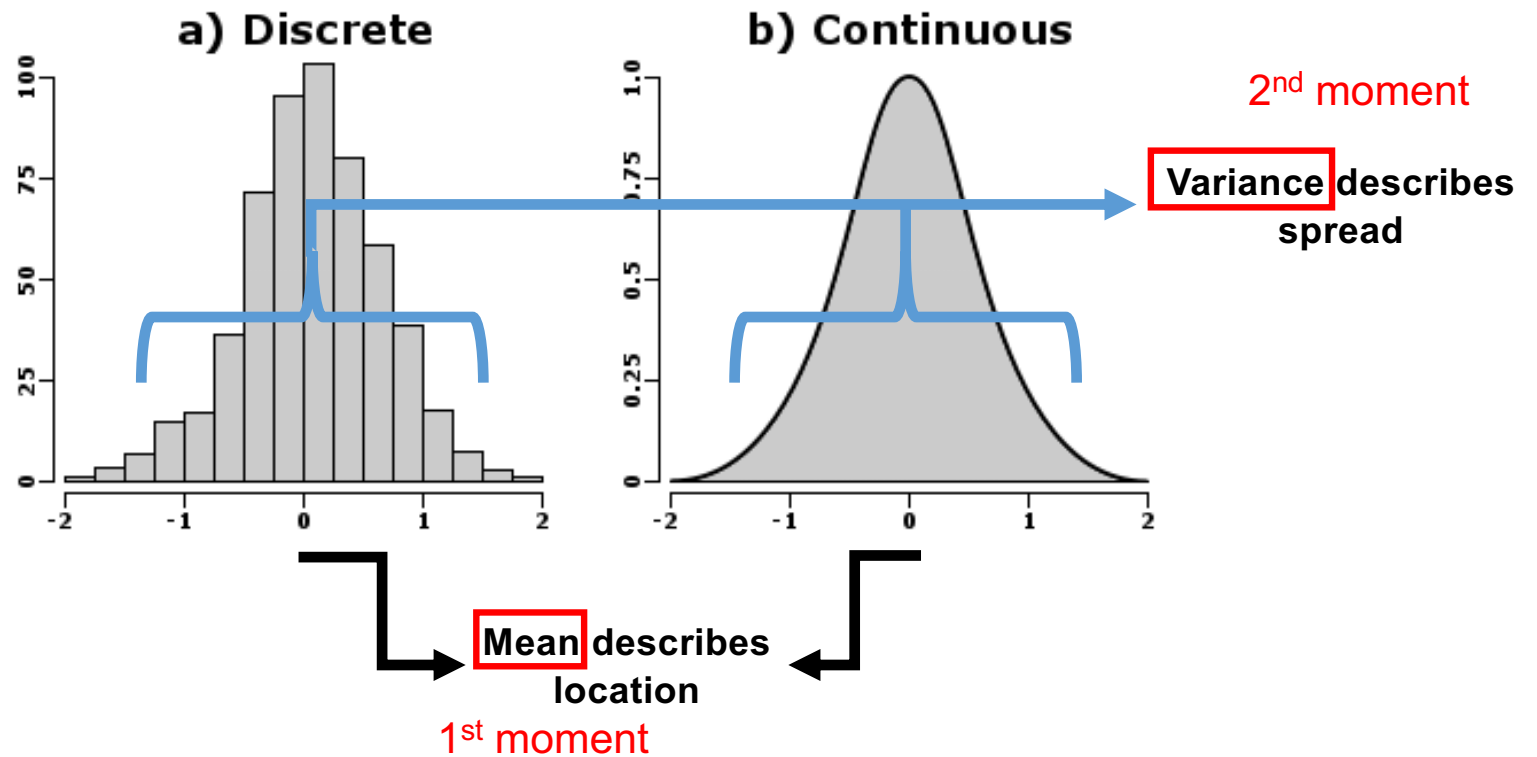
Properties of a RV are characterized by its distribution / PMF / PDF
But there are “summary” numbers capturing important characteristics
This is called “**moments**”.

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

(Wikipedia)

Moments of Random Variables

7



Moments characterize properties of the distribution "shape"

Expectation

Expectation: a game-theoretic viewpoint

9

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

outcome	prob.
$X = 1$	0.7
$X = -1$	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

10

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

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

Expected Value

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

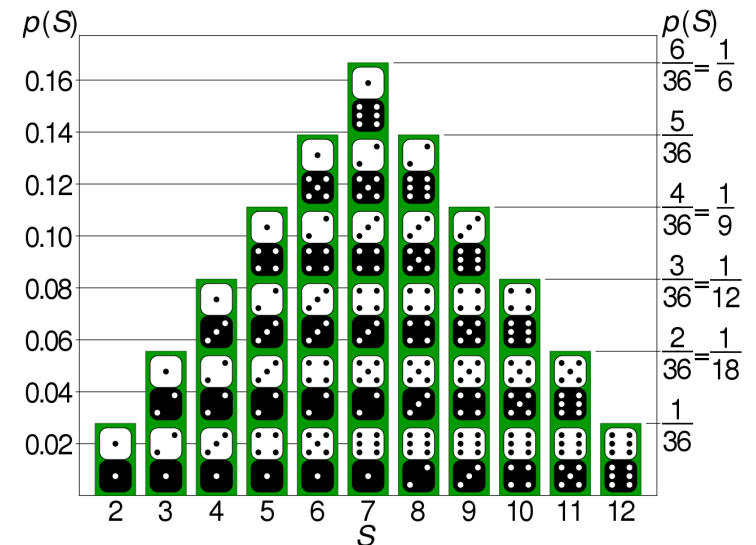
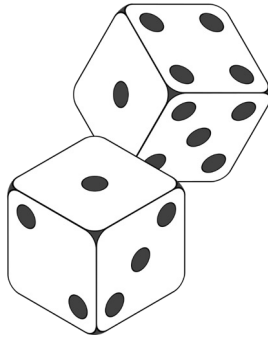
$$P(X = 2) = 1/36$$

$$P(X = 3) = 2/36$$

$$P(X = 4) = 3/36$$

...

$$P(X = 12) = 1/36$$



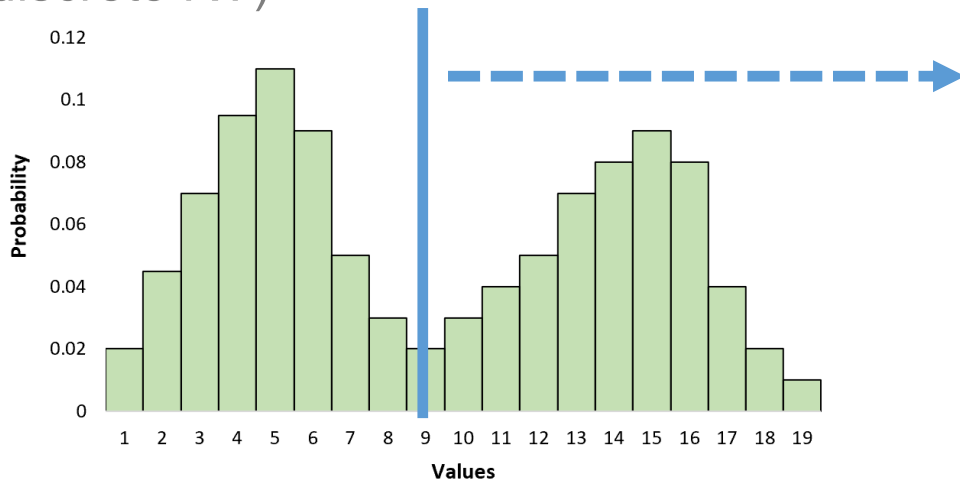
Q: $E[X]$?

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

Expected Value

12

(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

Expected Value

13

Theorem (Linearity of Expectations) *For any finite collection of discrete RVs X_1, X_2, \dots, 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.

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

Expected Value

14

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

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

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

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

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

Law of Total Probability

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

By definition of Expectation

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

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

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

Expected Value

15

Theorem For any random variable X and constant c ,

$$\begin{aligned}E[cX] &= cE[X] \\E[cX + k] &= cE[X] + k \\E[k] &= k\end{aligned}$$

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:

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

Conditional Expected Value

16

Definition The conditional expectation of a discrete RV X , given Y is:

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

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

$$\begin{aligned} \mathbf{E}[X_1 \mid Y = 5] &= \sum_{x=1}^4 x p(X_1 = x \mid Y = 5) && \text{quiz candidate} \\ &= \sum_{x=1}^4 x \frac{p(X_1 = x, Y = 5)}{p(Y = 5)} = \sum_{x=1}^4 x \frac{1/36}{4/36} = \frac{5}{2} \end{aligned}$$

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

Variance

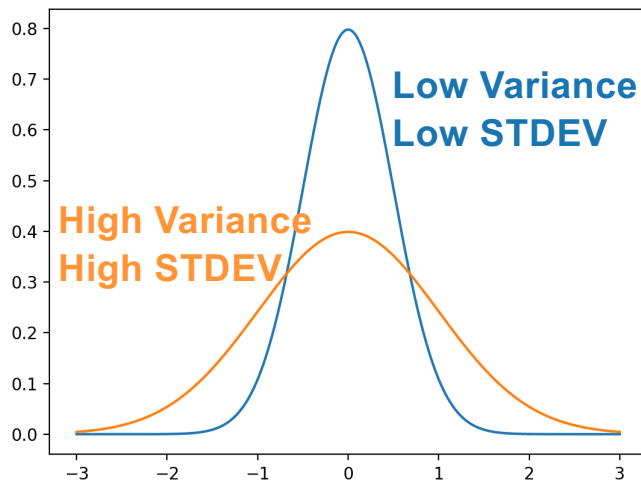
Variance

18

Definition The variance of a RV X is defined as,

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

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



- 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


Variance

19

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

The variance is then,

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


$$E \left[\left(X - \frac{7}{2} \right)^2 \right]$$

The STDEV is $\sqrt{\text{Var}(X)} \approx 1.71$, which suggests we should expect outcomes to vary around the mean of 3.5 by ± 1.71

Variance

20

Lemma An equivalent form of variance is:

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

$E[X]$ is a constant

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

Proof

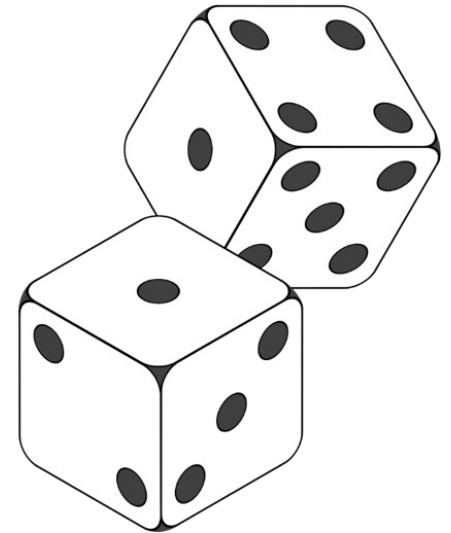
$$E[(X - E[X])^2] = E[X^2 - 2XE[X] + E[X]^2] \quad \text{(Expand it)}$$

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

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

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

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



Variance

22

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

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

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

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

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

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

Moments of Useful Discrete Distributions

23

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

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

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

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

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

$$\text{Var}[X] = \pi(1 - \pi)$$

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

$$\text{Var}[X] = \pi - \pi^2$$



Covariance

24

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

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

Question What is $\text{Cov}(X, X)$?

Answer $\text{Cov}(X, X) = \text{Var}(X)$

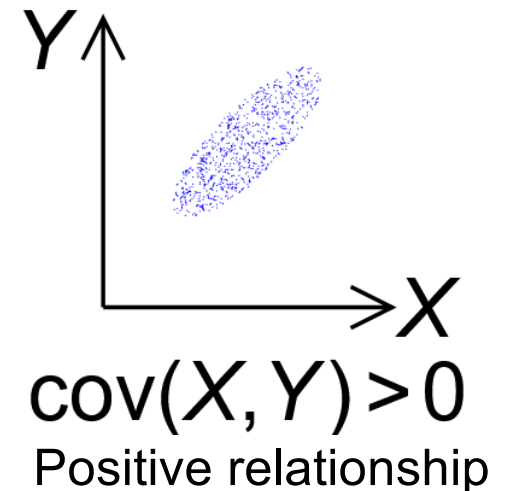
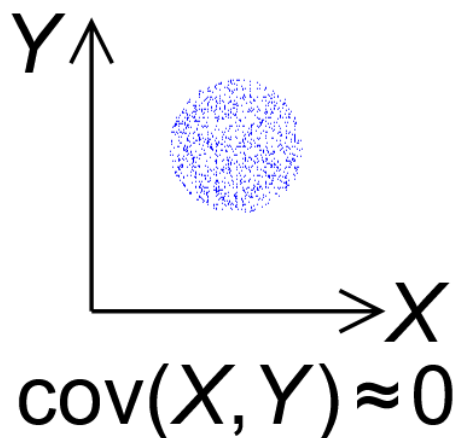
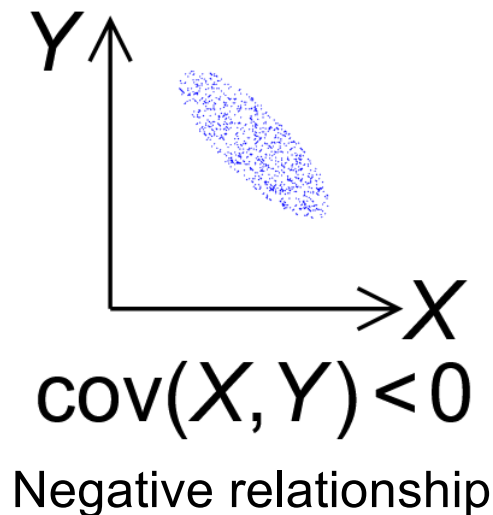
Covariance

25

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

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

Measures the linear relationship between X and Y



Example: height vs weight

- A shortcut to compute covariance.
- $Cov(X, Y) = E[(X - E[X])(Y - E[Y])]$
$$= E[XY - X \cdot E[Y] - Y \cdot E[X] + E[X]E[Y]]$$
$$= E[XY] - E[X]E[Y] - E[Y]E[X] + E[X]E[Y]$$
$$= E[XY] - E[X]E[Y]$$
- Safety check: $Cov(X, X) = E[XX] - E[X]E[X] = Var(X)$

Covariance

27

Lemma For any two RVs X and Y ,

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

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

\Rightarrow variance is not a linear operator.

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

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

Covariance



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

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

$$\left[\begin{array}{c|c} Cov(A, A) & Cov(A, B) \\ \hline Cov(B, A) & Cov(B, B) \\ \hline \end{array} \right]$$

(Note: The top-left element in the matrix above is labeled as Var(A) in the original image)

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

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

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

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

Correlation

29

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

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

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.

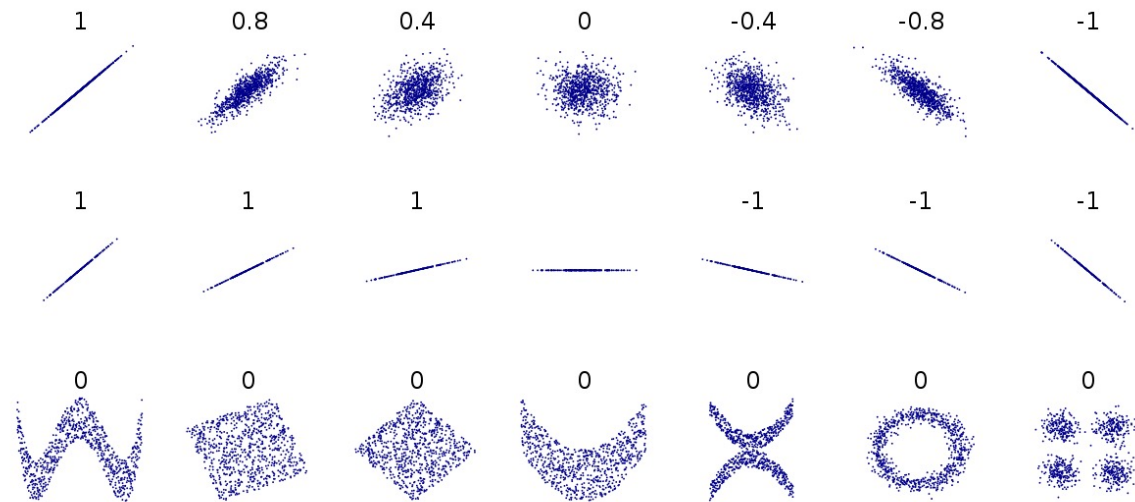
⇒ $\text{Cov}(cX, dY) \neq \text{Cov}(X, Y)$ **but** $\text{Corr}(cX, dY) = \text{Corr}(X, Y)$

Correlation

30

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

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



Like covariance, only expresses linear relationships!