Computer Science

CSC380: Principles of Data Science

Probability Primer 5

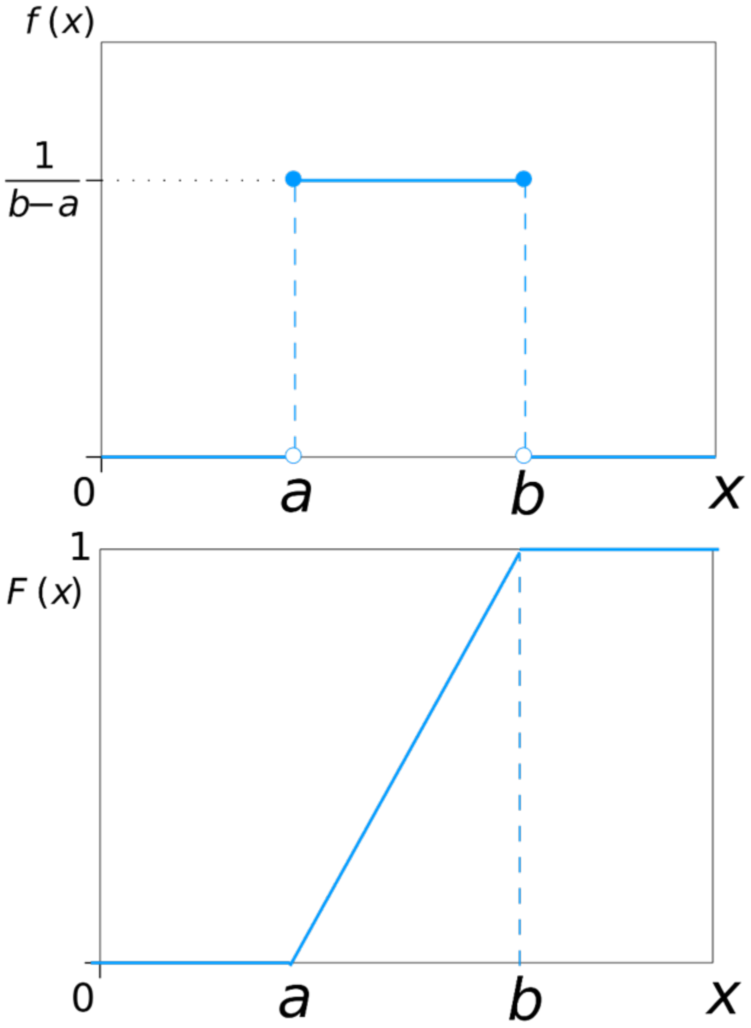
Xinchen Yu

Acknowledgement: Built on Jason Pacheco, Kwang-Sung Jun, Chicheng Zhang's slides

Review: Continuous Random Variable

- 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]



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:

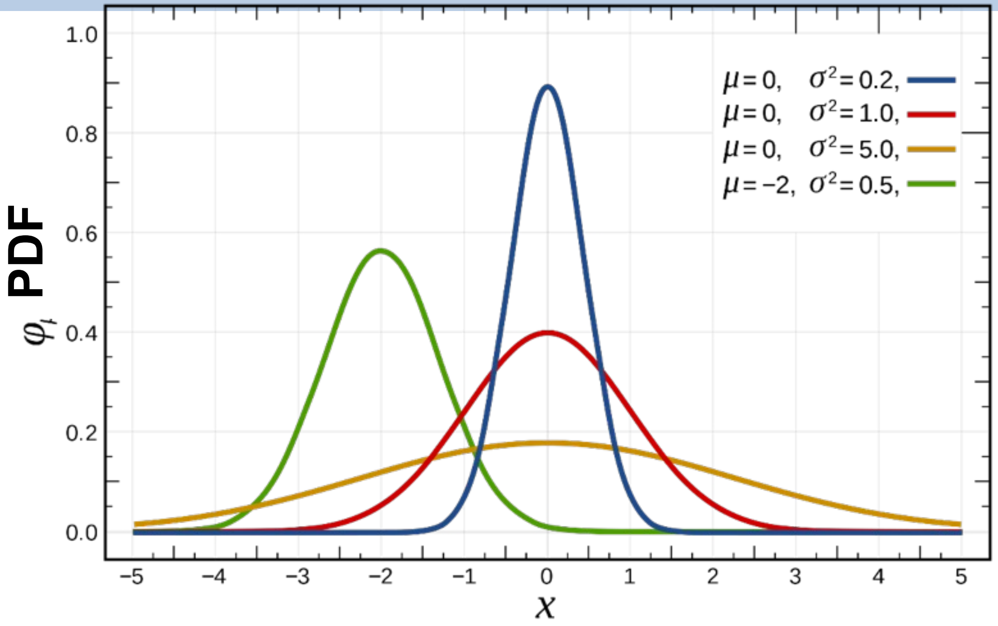
$$P(X \in [a, b]) = P(X \in (a, b]) = P\big(X \in [a, b)\big) = P\big(X \in (a, b)\big)$$

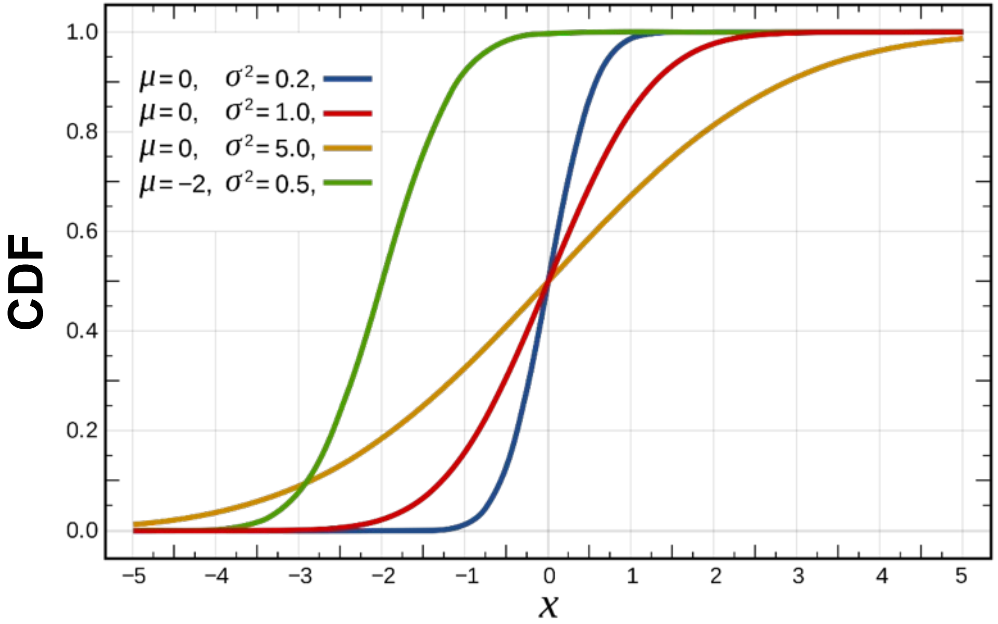
Review: Continuous Random Variable

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

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

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



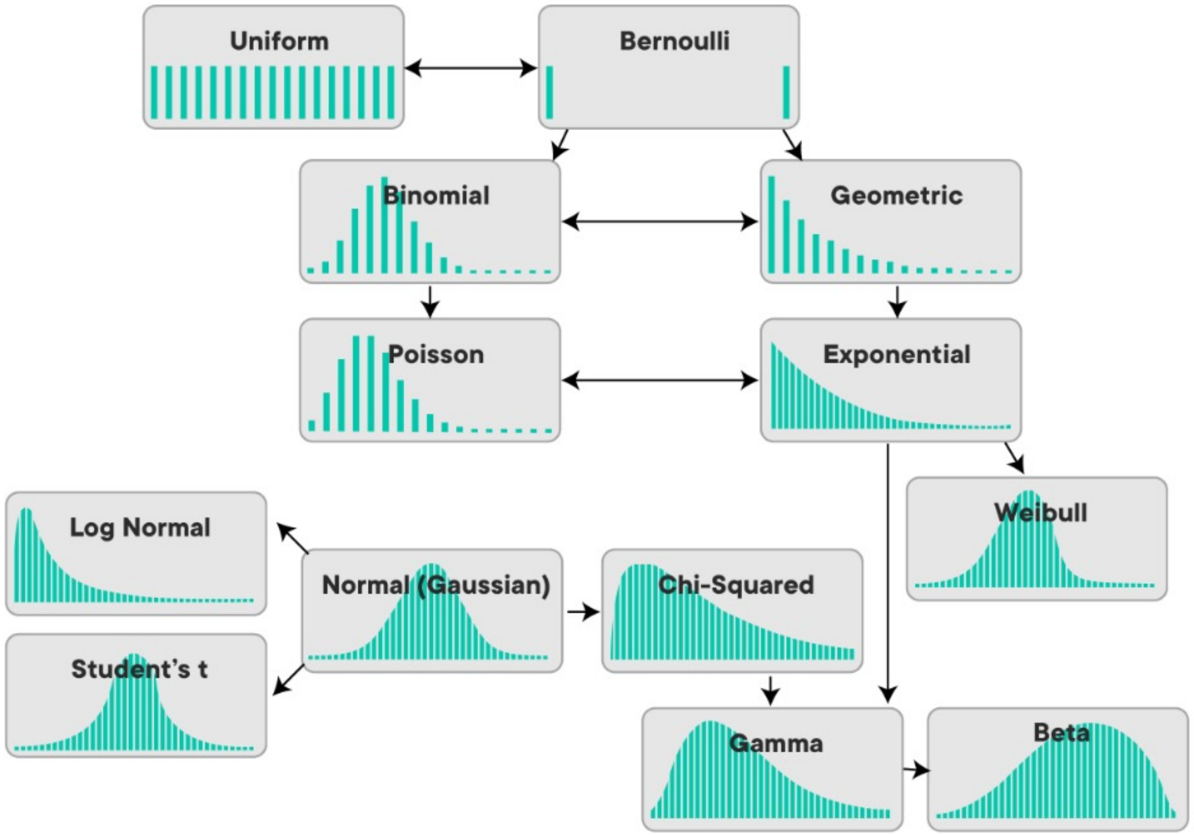


Outline

- Expectation
- Variance
- Covariance
- Correlation

Moments of Random Variables

Q: How to describe characteristics of different distributions?



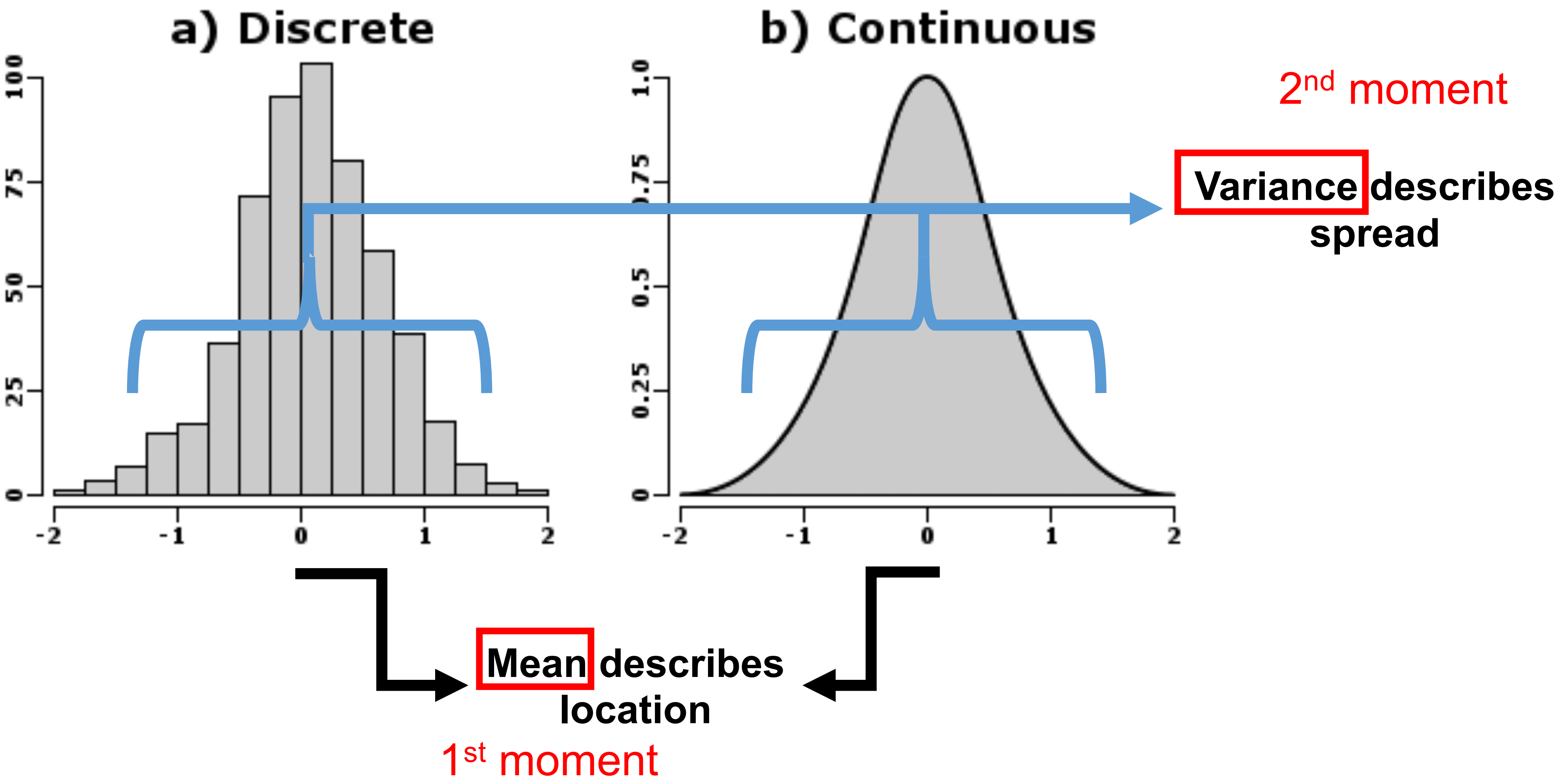
Moments of Random Variables

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



Moments characterize properties of the distribution "shape"

Expectation

Expectation: a game-theoretic viewpoint

- Consider the following game:
  - Flip an unfair coin 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

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

(with PMF)

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

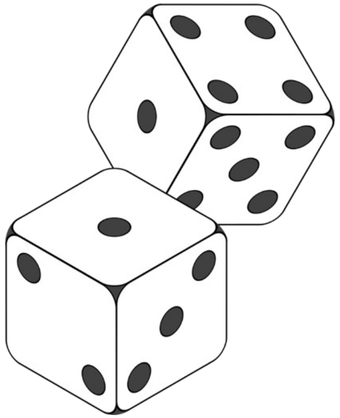
Summation over all values in domain of X

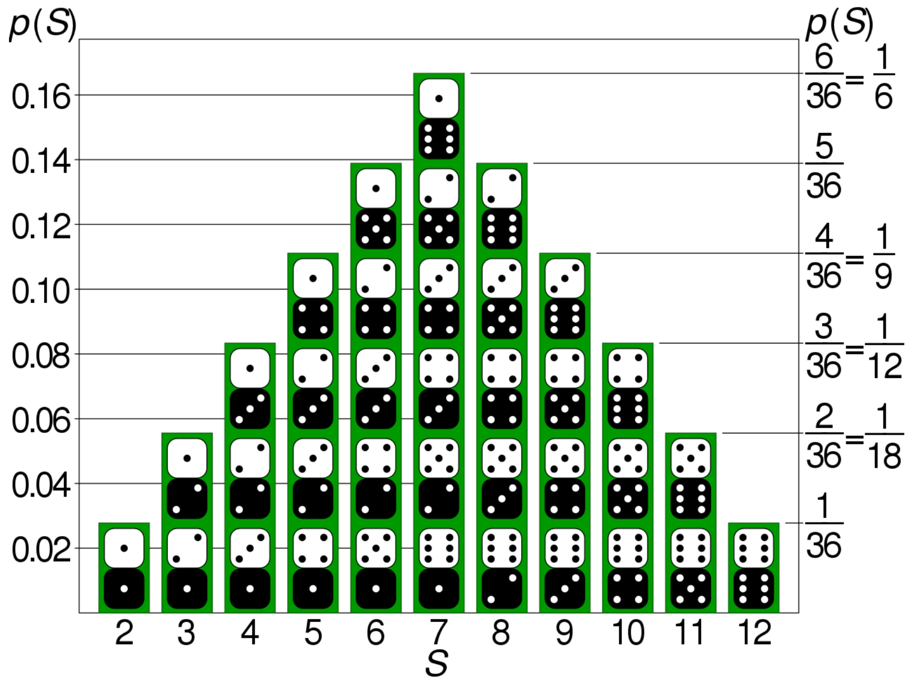
- **Effectively, a weighted average**: *each outcome weighted by probability of occurring*

Expected Value

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

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



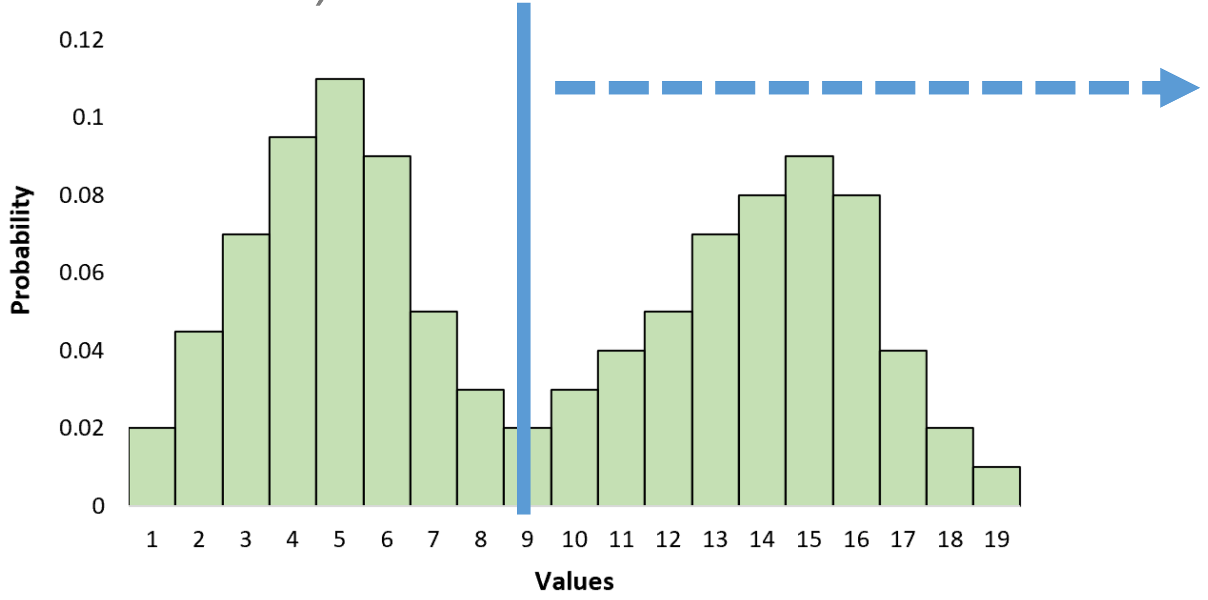


Q: $\mathbf{E}[X]$?

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

Expected Value

(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

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.

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

Expected Value

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

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

Expected Value

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:

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

Conditional Expected Value

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.)

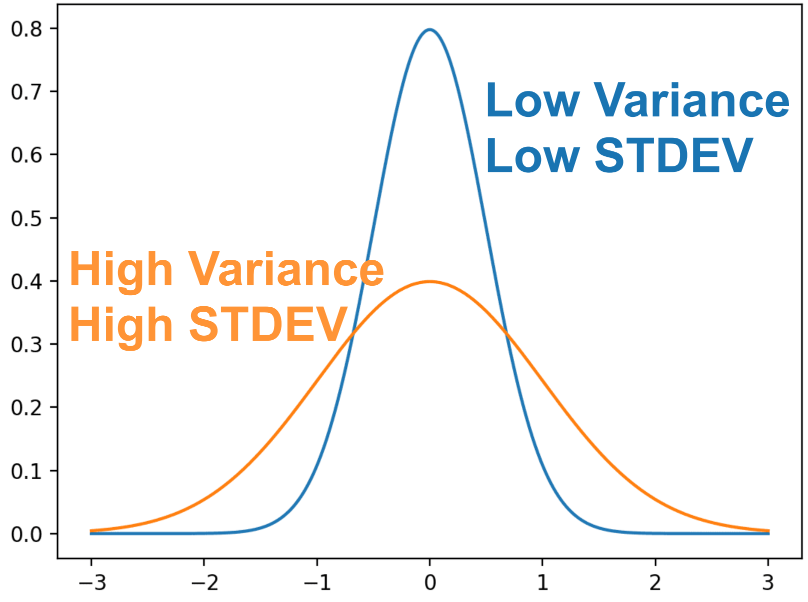
Variance

Variance

Definition The variance of a RV X is defined as,

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

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 units^2*
- *Variance is more mathematically useful than STDEV*

Variance

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

The variance is then,

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

$$\mathrm{Var}(X) = \sum_{i=1}^{6} \frac{1}{6}\left(i - \frac{7}{2}\right)^2 \longleftarrow E\left[\left(X - \frac{7}{2}\right)^2\right]$$

$$= \frac{1}{6}\left((-5/2)^2 + (-3/2)^2 + (-1/2)^2 + (1/2)^2 + (3/2)^2 + (5/2)^2\right)$$

$$= \frac{35}{12} \approx 2.92.$$

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

Variance

$E\left[2XE[X]\right] = 2E\left[XE[X]\right] = 2E[X]E[X]$

$E[X]$ is a constant

Lemma An equivalent form of variance is:

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

Proof

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

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

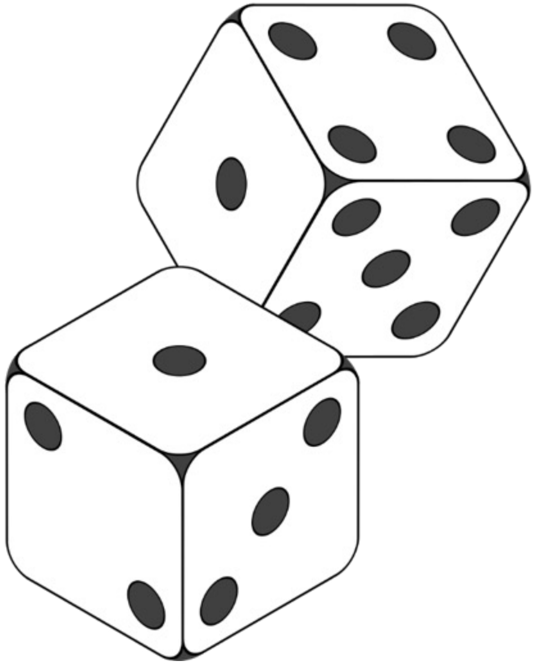
$$= \mathbf{E}[X^2] - 2\mathbf{E}[X]^2 + \mathbf{E}[X]^2 \quad \textbf{(Algebra)}$$

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

Variance

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

$$
\begin{aligned}
\operatorname{Var}(X) &= \mathrm{E}\left(X^2\right) - (\mathrm{E}(X))^2 \\
&= \frac{1}{n}\sum_{i=1}^{n} i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} i\right)^2 \\
&= \frac{(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2 \\
&= \frac{n^2-1}{12}.
\end{aligned}
$$



Variance

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

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

$$
\begin{aligned}
Var[cX] &= E[(cX)^2] - (E[cX])^2 \\
&= E[c^2X^2] - (cE[X])^2 \\
&= c^2E[X^2] - c^2E[X]^2 \\
&= c^2(E[X^2] - E[X]^2)
\end{aligned}
$$

Moments of Useful Discrete Distributions

Bernoulli *A.k.a. the* ***coinflip*** *distribution on <u>binary</u> 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 π *is the probability of* ***success*** *(i.e., heads), and also the mean*

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

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

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

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

Covariance

Definition *The <u>covariance</u> 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)?*

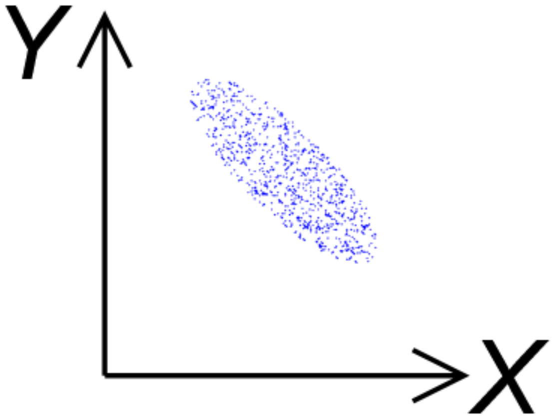
Answer *Cov(X,X) = Var(X)*

Covariance

Definition *The* <u>*covariance*</u> *of two RVs* X *and* Y *is defined as,*

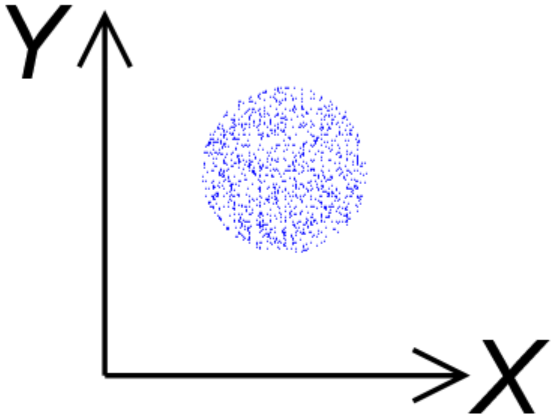
$$\mathbf{Cov}(X,Y) = \mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])]$$

Measures the <u>linear relationship</u> between X and Y

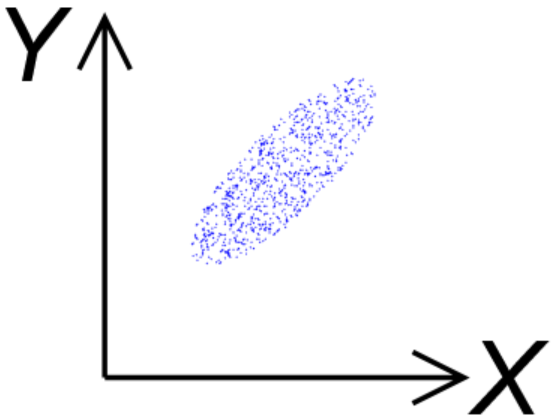


$\mathrm{cov}(X,Y) < 0$

Negative relationship



$\mathrm{cov}(X,Y) \approx 0$



$\mathrm{cov}(X,Y) > 0$

Positive relationship

Example: height vs weight

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

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

Lemma *For any two RVs* X *and* Y*,*

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

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

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

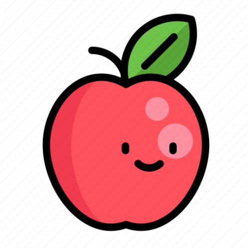
(Linearity of expt.) $= \mathbf{E}[(X + Y - \mathbf{E}[X] - \mathbf{E}[Y])^2]$

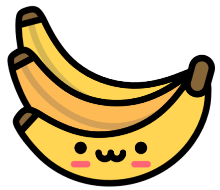
(Distributive property) $= \mathbf{E}[(X - \mathbf{E}[X])^2 + (Y - \mathbf{E}[Y])^2 + 2(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])]$

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

Covariance





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

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

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

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

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

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

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

Correlation

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

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

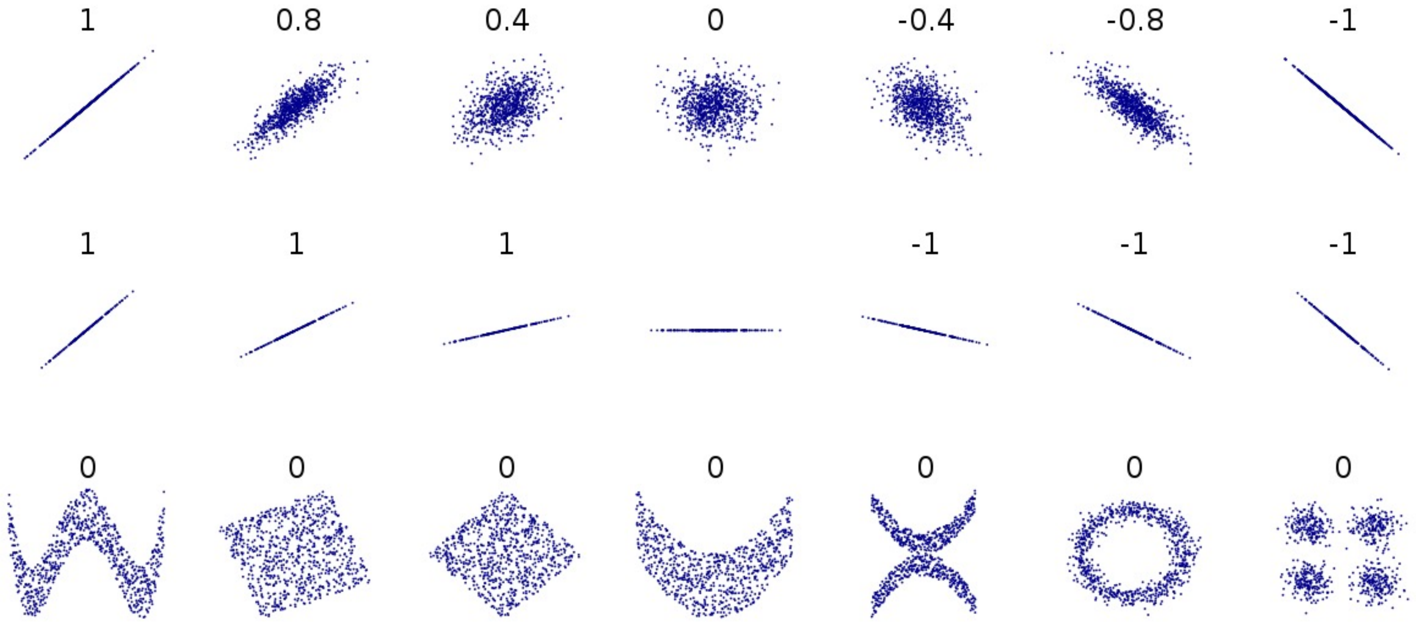
Normalized version of covariance!
⇒ Always between -1 and 1

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** $Corr(cX, dY) = Corr(X, Y)$

Correlation

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!