REVIEW OF PROBABILITIES

Narges Norouzi

TOPICS

- Motivation
- Notations and definitions
- Review of probabilities

THE CONCEPT OF PROBABILITY

- Actual repeated experiments
  - Example: You record the color of 1000 cars driving by 57 of them are green. You estimate the probability of a car being green as $\frac{57}{1000} = 0.057$
- In idealized conceptions of repeated process
  - Example: You consider the behavior of an unbiased six-sided die. The expected probability of rolling a 5 is $\frac{1}{6} = 0.1667$
  - Example: You need a model for how people's heights are distributed. You choose a normal distribution (bell-shaped curve) to represent the expected relative probabilities

NOTATION

- $a \in A$: Set membership: a is a member of set A
- $|B|$: Cardinality: number of items in set B
- $||v||$: Norm: length of a vector v
- $\sum$: Summation
- $\int$: Integral

NOTATION

- $\mathbf{x}, \mathbf{y}, \mathbf{z}$: Vector (bold, lower case)
- $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{X}$: Matrix (bold, upper case)
- $y = f(x)$: Function (map)
- $\frac{dy}{dx}$: Derivative of y with respect to single variable x
- $\frac{\partial y}{\partial x_i}$: Partial derivative of y with respect to element i of vector x

AXIOMS OF PROBABILITY

- Non-negativity:

  for any event $E \in F, \; p(E) \geq 0$

- All possible outcomes, where $\boldsymbol{\Omega}$ is sample space:

  $p(\boldsymbol{\Omega}) = 1$

- Additivity of disjoint events:

  for all events $E, G \in F$ where $E \cap G = \emptyset$,

  $$p(E \cup G) = p(E) + p(G)$$

TYPES OF PROBABILITY SPACES

Define $|\boldsymbol{\Omega}| =$ number of possible outcomes

- Discrete space $|\boldsymbol{\Omega}|$ is finite
  - Analysis involves summations: $\sum$
- Continuous space $|\boldsymbol{\Omega}|$ is infinite
  - Analysis involves integrals: $\int$

EXAMPLE: DISCRETE PROBABILITY SPACE

Single roll of a six-sided die

- 6 possible outcomes: $O = \{1, 2, 3, 4, 5, 6\}$
- If die is fair, then probability of outcomes are equal

$$p[i] = \frac{1}{6} \text{ for } i \in \{1, 2, 3, 4, 5, 6\}$$

- What is the probability of event $E = (outcome\ is\ odd)$?

EXAMPLE 2

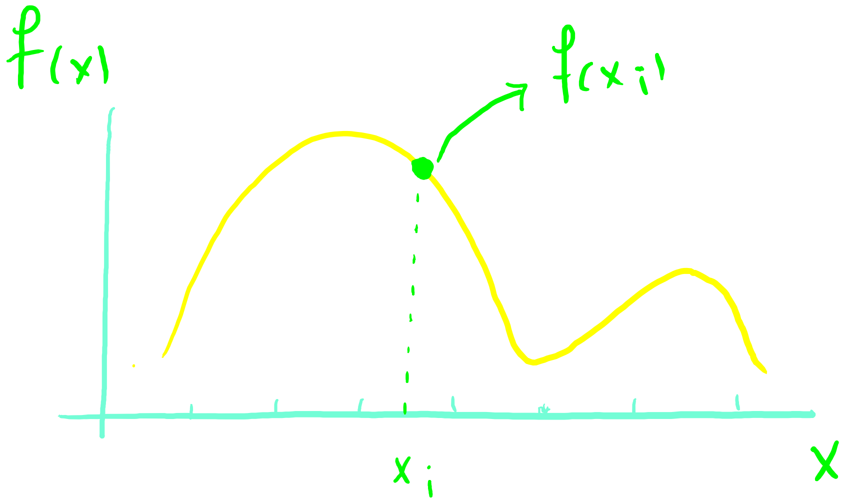
Consider 3 consecutive flips of a coin

- How many possible outcomes?

- What is the probability that we see exactly two heads?

CONTINUOUS PROBABILITY SPACE

- Infinite number of possible outcomes
- Probabilities of outcomes are not equal, and are described by a continuous function.

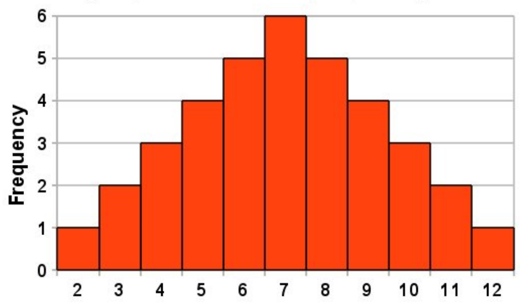


PROBABILITY FUNCTIONS

DISCRETE

Example: sum of two fair dice

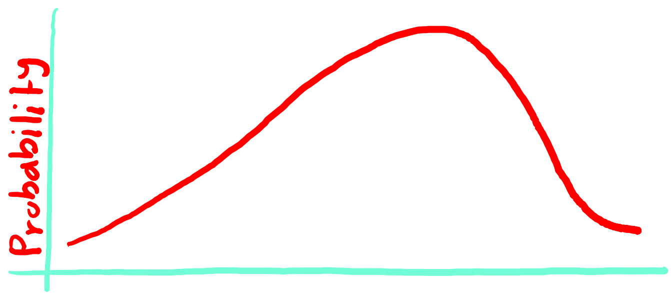
Probability mass function (pmf)



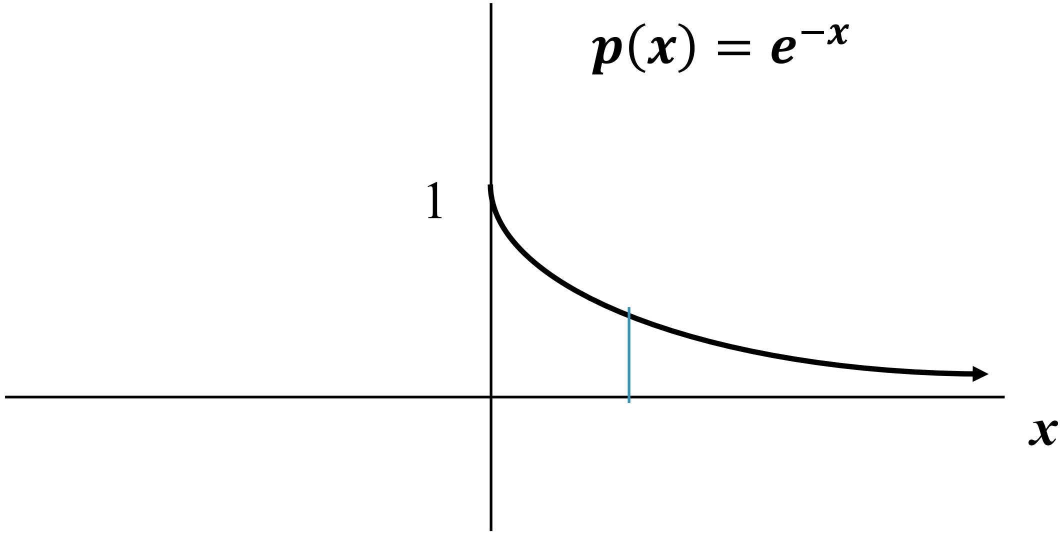
CONTINUOUS

Example: waiting time for doctor appointment

Probability density function (pdf)

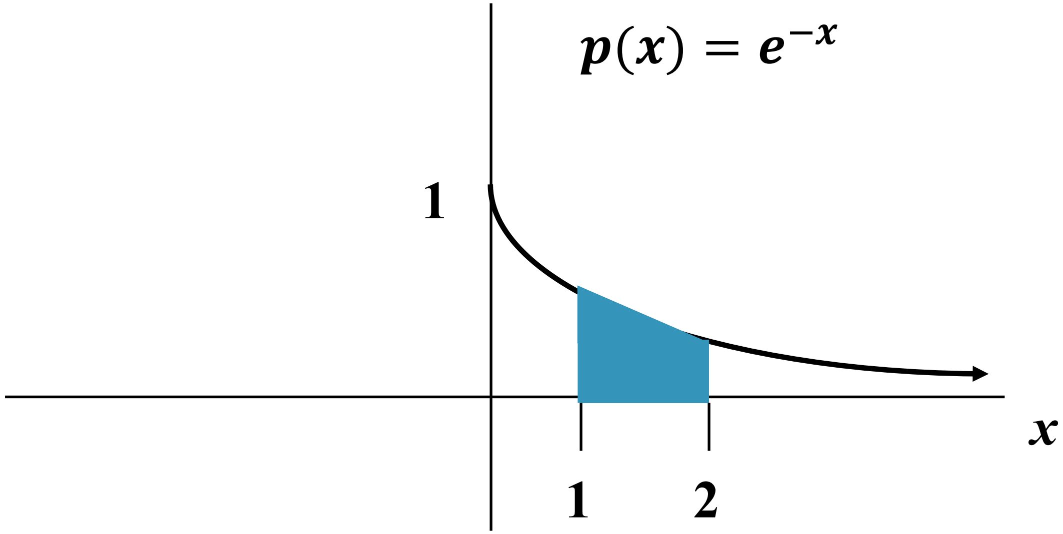


CONTINUOUS CASE:
"PROBABILITY DENSITY FUNCTION" (PDF)



The probability that x is any exact particular value (such as 1.9976) is 0; we can only assign probabilities to possible ranges of x.

THE PROBABILITY OF x FALLING WITHIN 1 TO 2:



$$P(1 \le x \le 2) = \int_1^2 e^{-x} = -e^{-x} \Big|_1^2 = -e^{-2} - -e^{-1} = -.135 + .368 = .23$$

CUMULATIVE DISTRIBUTION

- Cumulative distribution function $F(x)$ is the probability of getting at most x.

When playing two dice the sum of outcomes lies between $2 - 12$. Using cumulative distribution we can easily find probabilities for different events:

$$p(X < 7) = \frac{15}{36}$$

$$p(X > 9) = 1 - \frac{30}{36} = \frac{6}{36}$$

$$p(4 < X < 9) = \frac{26}{36} - \frac{6}{36} = \frac{20}{36}$$

x	F(x)
2	1/36
3	3/36
4	6/36
5	10/36
6	15/36
7	21/36
8	26/36
9	30/36
10	33/36
11	35/36
12	36/36

CUMULATIVE DISTRIBUTION FUNCTION

- As in the discrete case, we can specify the "cumulative distribution function" (CDF)

- The CDF in $p(x) = e^{-x}$

$$p(x \leq A) = \int_0^A e^{-x} = -e^{-x} \Big|_0^A = -e^{-A} - -e^{0} = -e^{-A} + 1 = 1 - e^{-A}$$

PRACTICE PROBLEM

Which of the following are probability functions?

a) $f(x) = \frac{1}{4}$ for $x = 9, 10, 11, 12$

b) $f(x) = \frac{3-x}{2}$ for $x = 1, 2, 3, 4$

c) $f(x) = \frac{x^2+x+1}{25}$ for $x = 0, 1, 2, 3$

PROBABILITY DISTRIBUTION

Joint Probability distribution

$$Probability\ of\ X = x\ \&\ Y = y$$
$$Prob(X = x, Y = y)$$
$$p(x, y)$$

Conditional probability Distribution

$$probability\ of\ X = x\ given\ Y = y$$
$$prob(X = x|Y = y)$$
$$p(x|y) = \frac{p(x, y)}{p(y)}$$

EXAMPLE

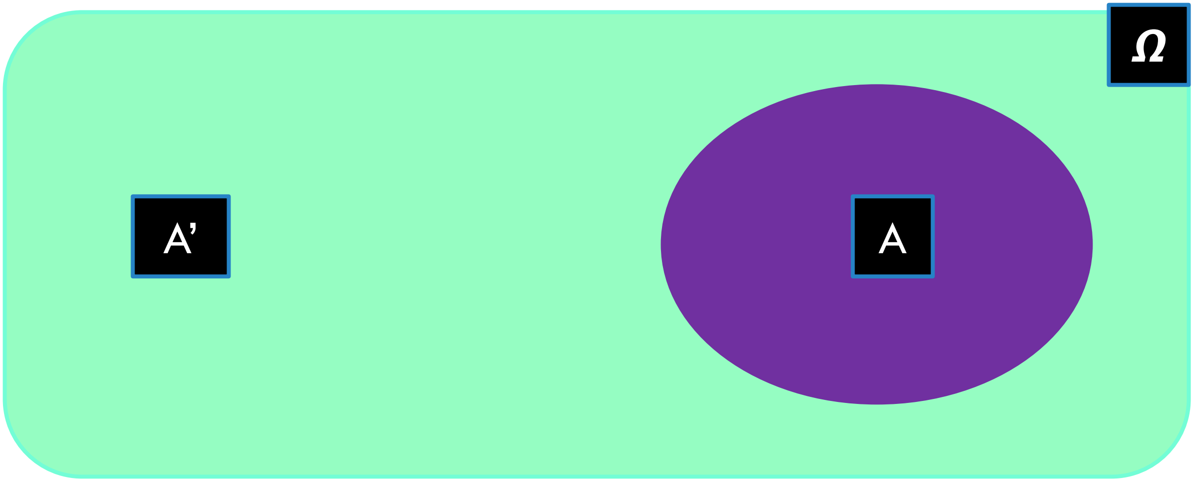
Suppose that we know that a die throw was odd.

What is the probability that a "three" has been thrown.

COMPLEMENT RULE

Given: event A, which can occur or not

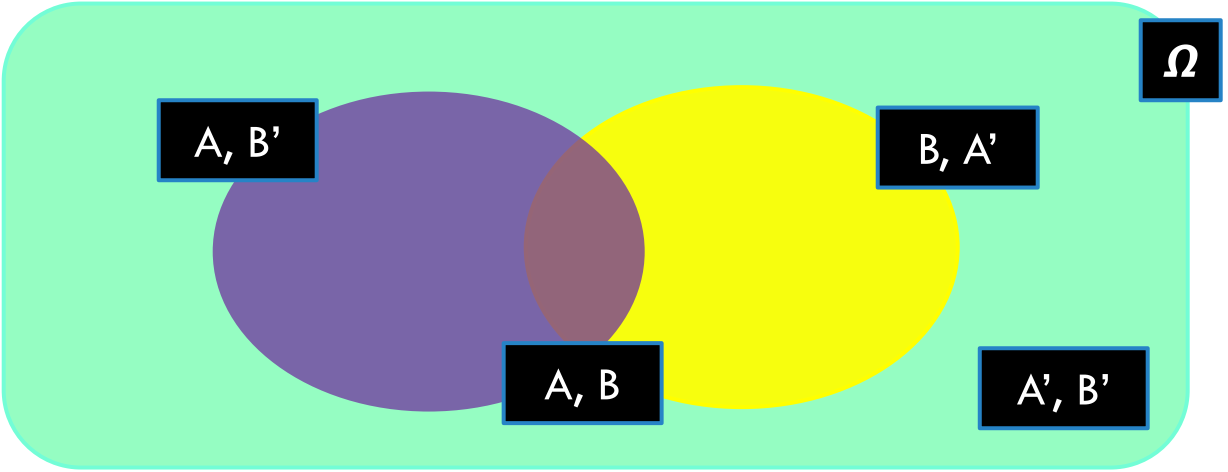
$$P(A') = 1 - p(A)$$



RULE OF TOTAL PROBABILITY

Given: events A and B, which can co-occur (or not)

$$P(A) = p(A, B) + p(A, B')$$



OTHER RULES OF PROBABILITY

- Sum Rule (marginalization/summing out):

$$p(x) = \sum_{y} p(x, y)$$

$$p(x_1) = \sum_{x_2} \dots \sum_{x_N} p(x_1, x_2, \dots, x_N)$$

- Product/Chain Rule

$$p(x, y) = p(y|x)p(x)$$

$$p(x_1, x_2, \dots, x_N) = p(x_1)p(x_2|x_1) \dots . p(x_N|x_1, x_2, \dots, x_{N-1})$$

EXAMPLE OF PRODUCT RULE

- Calculate the probability that a man has gray hair and is over 65 if:
  - Probability of a chosen random man being over 65: 20%
  - Probability of having gray hair given that a randomly chosen man is over 65: 80%

INDEPENDENCE

- Two random variables are said to be independent if and only if their joint distribution factors

$$p(A,B) = p(A).p(B) = p(A|B)p(B) = p(B|A)p(A)$$
$$p(A|B) = p(A)$$
$$p(B|A) = p(B)$$

- Two random variables are conditionally independent given a third if they are independent after conditioning on the third

$$p(A,B|C) = p(A|C).p(B|C) = p(A|B,C)p(B|C) = p(B|A,C)p(A|C)$$

LET'S TEST SOME EVENTS

- Outcomes on multiple rolls of a die
- Outcomes on multiple flips of a coin
- Height of two related individuals
- Probability of getting a king on successive draws from a deck of cards, without replacement
- Probability of getting a king on successive draws from a deck of cards, with replacement

BAYES' RULE

A way to find conditional probabilities for one variable when conditional probability for another variable is known.

$$p(B|A) = \frac{p(A|B).p(B)}{p(A)}$$

Where

$$p(A) = p(A, B) + p(A, B')$$

BAYES' RULE

$\propto$ Likelihood $\times$ prior probability

Posterior probability

$$p(B|A) = \frac{p(A|B).p(B)}{p(A)}$$

EXAMPLE OF BAYES' RULE

A desk lamp produced by a company was found to be defective (*D*). There are three factories (*A*, *B*, C) where such desk lamps are manufactured. This is what we know about the company's desk lamp production and the possible source of defects:

Factory	% of total production	Probability of defective lamps
A	$0.35 = p(A)$	$0.015 = p(D \mid A)$
B	$0.35 = p(B)$	$0.010 = p(D \mid B)$
C	$0.30 = p(C)$	$0.020 = p(D \mid C)$

If a randomly selected lamp is defective, what is the probability that the lamp was manufactured in factory C?

EXPECTATION & VARIANCE

EXAMPLE: THE LOTTERY

- The Lottery (also known as a tax on people who are bad at math…)
- A certain lottery works by picking 6 numbers from 1 to 49. It costs $\$1.00$ to play the lottery, and if you win, you win $\$2\ million$ after taxes.
- *If you play the lottery once, what are your expected winnings or losses?*

LOTTERY

Calculate the probability of winning in 1 try:

$$\frac{1}{\binom{49}{6}} = \frac{1}{\frac{49!}{43!6!}} = \frac{1}{13{,}983{,}816} = 7.2 \text{ x } 10^{-8}$$

"49 choose 6"

Out of 49 numbers, this is the number of distinct combinations of 6.

The probability function (note, sums to 1.0):

x	$p(x)$
-1	0.999999928
$+2\,million$	$7.2\ x\ 10^{-8}$

EXPECTED VALUE

The probability function (note, sums to 1.0):

x	$p(x)$
-1	0.999999928
$+\,2\,million$	$7.2\ x\ 10^{-8}$

Expected Value:

$$E(X) = p(win) * \$2{,}000{,}000 \quad + \quad p(lose) * -\$1.00$$
$$= \ 2.0\ x\ 10^{6} \ * \ 7.2\ x\ 10^{-8} + 0.999999928\ (-1)$$
$$= 0.144 \ - 0.999999928 \ = \ -\$0.86$$

Negative expected value is never good!
You shouldn't play if you expect to lose money!

EXPECTED VALUE AS A MATHEMATICAL OPERATOR

If c = a constant number (i.e., not a variable) and X and Y are any random variables…

- $E(c) = c$
- $E(cX) = cE(X)$
- $E(c + X) = c + E(X)$
- $E(X + Y) = E(X) + E(Y)$

EXPECTED VALUE ISN'T EVERYTHING THOUGH…

- Take the show "Deal or No Deal"
- Let's say you are down to two cases left. $\$0$ and $\$400{,}000$. The banker offers you $\$200{,}000$.
- So, Deal or No Deal?

DEAL OR NO DEAL…

- This could really be represented as a probability distribution and a non-random variable:

x	$p(x)$
$+0$	0.50
$+\$400{,}000$	0.50

x	$p(x)$
$+\$200{,}000$	1.0

EXPECTED VALUE DOESN'T HELP…

- This could really be represented as a probability distribution and a non-random variable:

x	$p(x)$
$+0$	0.50
$+\$400{,}000$	0.50

$$\mu = E(X) = \sum_{\text{all } x} x_i\, p(x_i) = 0\,(.50) + 400{,}000(.50) = 200{,}000$$

x	$p(x)$
$+\$200{,}000$	1.0

$$\mu = E(X) = 200{,}000$$

HOW TO DECIDE?

- Variance!
- If you take the deal, the variance/standard deviation is 0.
- If you don't take the deal, what is average deviation from the mean?

VARIANCE

"The average (expected) squared distance (or deviation) from the mean"

$$\sigma^2 = Var(x) = E[(x-\mu)^2] = \sum_{\text{all } x} (x_i - \mu)^2 p(x_i)$$

***We square because squaring has better properties than absolute value. Take square root to get back linear average distance from the mean (="standard deviation").*

SUMMARIZING PROBABILITY DISTRIBUTIONS

- It is often useful to give summaries of distributions without defining the whole distribution
- Mean or Expected value:

$$E[x] = \sum x.p[x] \quad \text{or} \quad E(x) = \int_x x.p(x)dx$$

- Variance:

$$\mathrm{var}(x) = \int_x \left(x - E(x)\right)^2 . p(x)dx = \mathrm{E}(\mathrm{x}^2) - \mathrm{E}(\mathrm{x})^2 \quad \text{or} \quad var[x] = E[x^2] - E[x]^2$$

COVARIANCE: JOINT PROBABILITY

- The covariance measures the strength of the linear relationship between two variables
- The covariance:

$$E[(x-\mu_x)(y-\mu_y)]$$

$$\boxed{\sigma_{xy} = \sum_{i=1}^{N}(x_i-\mu_x)(y_i-\mu_y)\,P(x_i,y_i)}$$

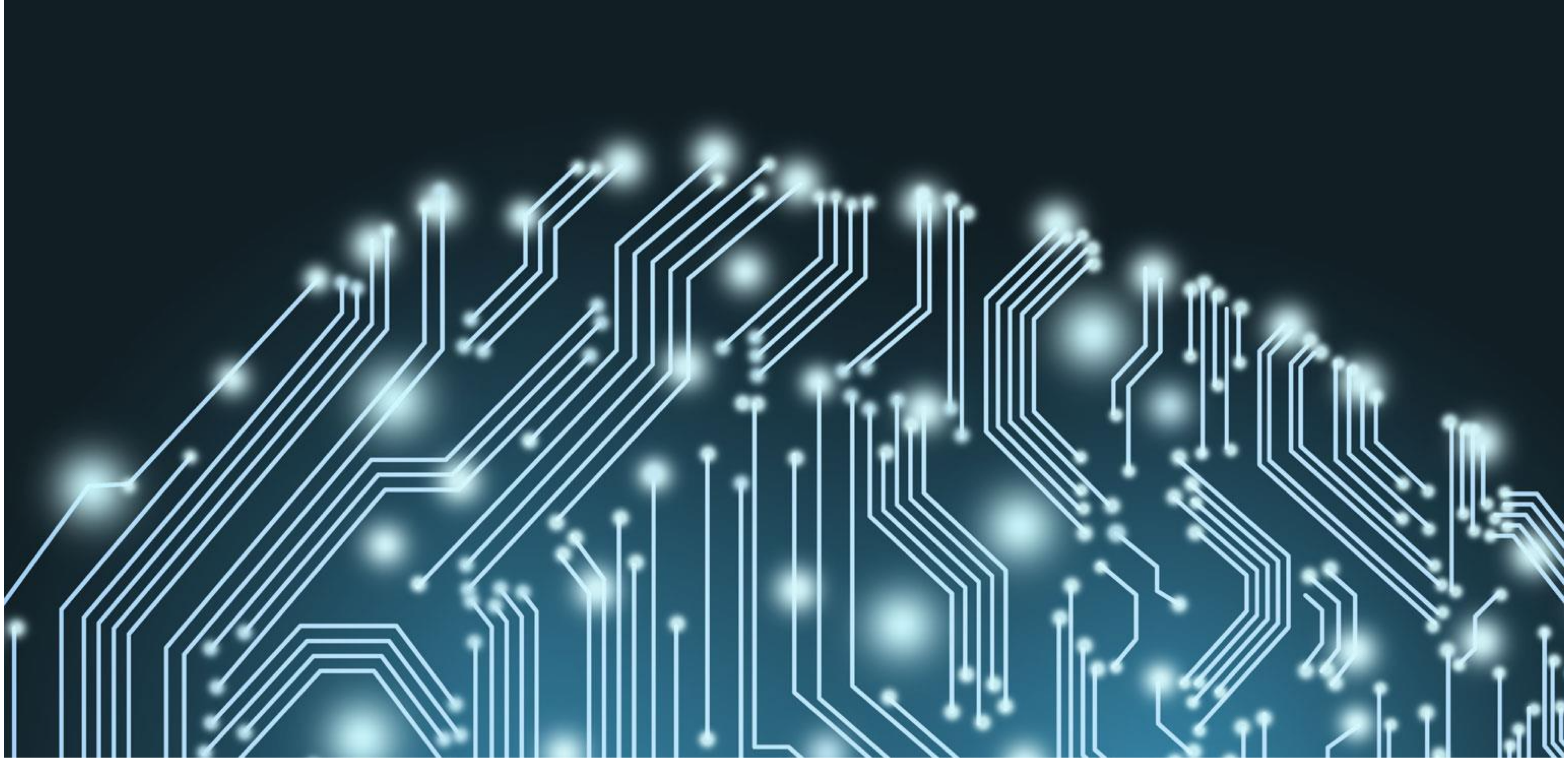
INTERPRETING COVARIANCE

- **Covariance** between two random variables:

$cov(X, Y) > 0$ X and Y are positively correlated

$cov(X, Y) < 0$ X and Y are inversely correlated

$cov(X, Y) = 0$ X and Y are independent (if relationship is linear)



QUESTIONS?