# Probability

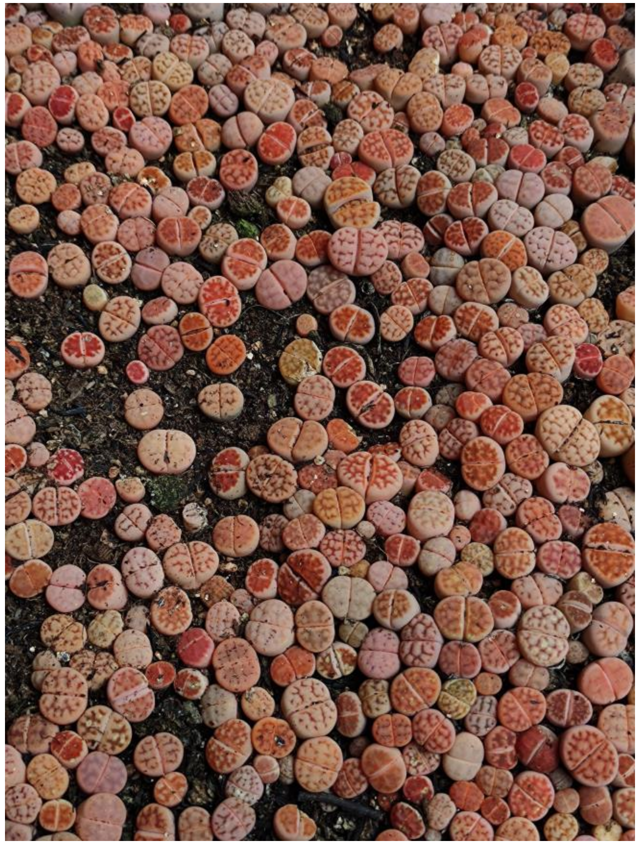
Bootcamp
13 Aug, 2024

## I am Zijian

2nd year, still rotating… I rotated with Tatiana, Jonathan, and Jonathan(currently)
General interest in memory, learning, and how to extract sth from neural activity(latent variable decoding)

Please feel free to ask any questions

Now, my major hobby is collecting shells. Previously, I grew succulent plants and climbed mountains, but I give them up now because of the rainy weather and getting lazier and lazier.



# Overview

1. Probability basics
2. Gaussian and Poisson distribution
3. Bayes’ theorem

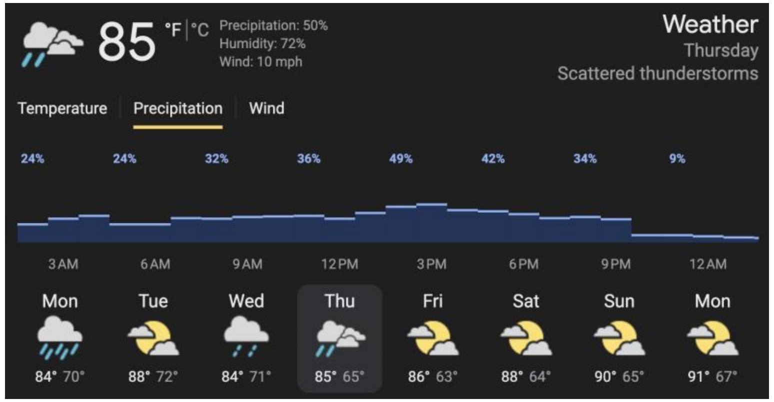
# Part 1
# Basic of probability

- Interpretations of probability
- PDF & PMF, normalization
- Joint, marginal, condition probability

# Approaches(interpretations) of probability

None of these can guarantee perfect estimation!

| | Definition | Pros | Cons |
|---|---|---|---|
| Classical probability | $P(A) = \dfrac{\text{Number of favorable outcomes for event A}}{\text{Total number of outcomes}}$ | Simplest, Clear math | Too-idealize |
| Frequentist probability | $P(A) = \lim_{n \to \infty} \dfrac{\text{Number of times event A occurs}}{n}$ | Objective, Empirical-basis, Model free | Inconsistent, not always practical |
| Subjective probability | Based on subjective experience, knowledge··· | Flexible, Widest-applicative | Subjective, Low reliability |



# Accurate probability estimation is hard!


Flipping the coins 350,757 times to reveal that same side landing has ~1% more probability

Table 1: By-person summary of the probability of a same side landing.

| Person | Same side | Flips | Coins | Proportion [95% CI] | Joined |
|---|---|---|---|---|---|
| XiaoyiL | 780 | 1600 | 2 | 0.487 [0.463, 0.512] | Marathon-MSc |
| JoyceP | 1126 | 2300 | 3 | 0.490 [0.469, 0.510] | Marathon-MSc |
| AndreeaZ | 2204 | 4477 | 4 | 0.492 [0.478, 0.507] | Marathon-MSc |
| KaleemU | 7056 | 14324 | 8 | 0.493 [0.484, 0.501] | Bc Thesis |
| FelipeFV | 4957 | 10015 | 3 | 0.495 [0.485, 0.505] | Internet |
| ArneJ | 1937 | 3900 | 4 | 0.497 [0.481, 0.512] | Marathon-MSc |
| AmirS | 7458 | 15012 | 6 | 0.497 [0.489, 0.505] | Bc Thesis |
| ChrisGI | 4971 | 10005 | 5 | 0.497 [0.487, 0.507] | Marathon-Manheim |
| FrederikA | 5219 | 10500 | 5 | 0.497 [0.487, 0.507] | Internet |
| FranziskaN | 5368 | 10757 | 3 | 0.499 [0.490, 0.508] | Internet |
| JasonN | 3352 | 6700 | 7 | 0.500 [0.488, 0.512] | Marathon-PhD |
| RietvanB | 1801 | 3600 | 4 | 0.500 [0.484, 0.517] | Marathon-PhD |
| PierreG | 7506 | 15000 | 9 | 0.500 [0.492, 0.508] | Bc Thesis |
| KarolineH | 2761 | 5500 | 5 | 0.502 [0.489, 0.515] | Marathon-PhD |
| SjoerdT | 2510 | 5000 | 5 | 0.502 [0.488, 0.516] | Marathon-MSc |
| SaraS | 5022 | 10000 | 3 | 0.502 [0.492, 0.512] | Marathon-Manheim |
| HenrikG | 8649 | 17182 | 8 | 0.503 [0.496, 0.511] | Marathon |
| IrmaT | 353 | 701 | 1 | 0.504 [0.467, 0.540] | Bc Thesis |
| KatharinaK | 2220 | 4400 | 5 | 0.504 [0.490, 0.519] | Marathon-PhD |
| JillR | 3261 | 6463 | 2 | 0.505 [0.492, 0.517] | Marathon |
| FrantisekB | 10148 | 20100 | 11 | 0.505 [0.498, 0.512] | Marathon |
| IngeborgR | 4340 | 8596 | 1 | 0.505 [0.494, 0.515] | Marathon |
| VincentO | 2475 | 4900 | 5 | 0.505 [0.491, 0.519] | Marathon-MSc |
| EricJW | 2071 | 4100 | 5 | 0.505 [0.490, 0.520] | Marathon-MSc |
| MalteZ | 5559 | 11000 | 7 | 0.505 [0.496, 0.515] | Marathon-Manheim |
| TheresaL | 1769 | 3500 | 4 | 0.505 [0.489, 0.522] | Marathon-MSc |
| DavidV | 7586 | 14999 | 5 | 0.506 [0.498, 0.514] | Bc Thesis |
| AntonZ | 5069 | 10004 | 2 | 0.507 [0.497, 0.516] | Marathon-Manheim |
| MagdaM | 2510 | 4944 | 6 | 0.508 [0.494, 0.522] | Marathon-MSc |
| ThomasB | 2540 | 5000 | 5 | 0.508 [0.494, 0.522] | Marathon-PhD |
| JonasP | 5080 | 9996 | 5 | 0.508 [0.498, 0.518] | Marathon |
| BohanF | 1118 | 2200 | 3 | 0.508 [0.487, 0.529] | Marathon-MSc |
| HannahA | 1525 | 3000 | 4 | 0.508 [0.490, 0.526] | Marathon-MSc |
| AdrianK | 1749 | 3400 | 3 | 0.514 [0.498, 0.531] | Marathon-MSc |
| AaronL | 3815 | 7400 | 5 | 0.515 [0.504, 0.527] | Marathon-MSc |

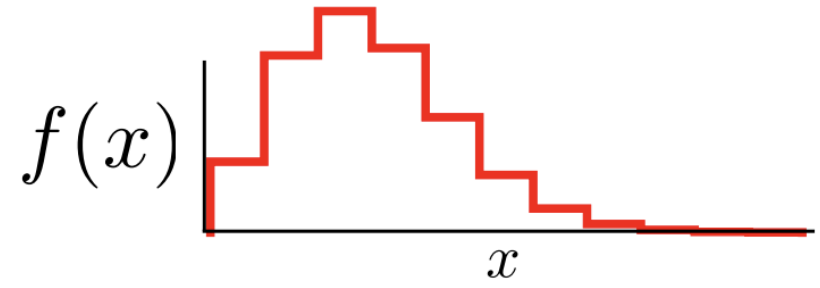
# Two types of random variables

**continuous** **discrete**

# discrete probability distribution

takes finite (or countably infinite) number of values, eg $x \in \mathbb{N}$

**probability mass function (pmf):**

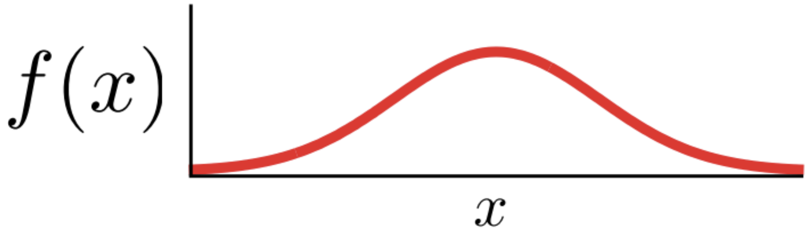



- $f(x_i) \geq 0$ for all i — non-negative
- $\sum_{i=1}^{N} f(x_i) = 1$ — sums to 1
- $P(x = a) = f(a)$ — gives probability of observing a particular value of x

# continuous probability distribution

takes values in a continuous space, e.g., $x \in \mathbb{R}$

**probability density function (pdf)**:

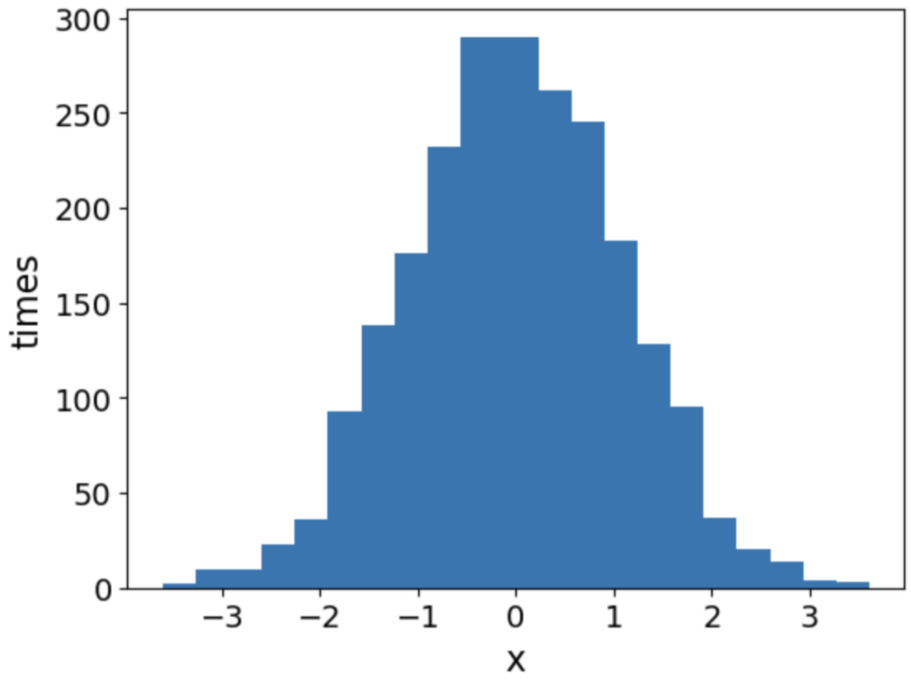



- $f(x) \geq 0$ for all x — non-negative
- $\int_{-\infty}^{\infty} f(x)\,dx = 1$ — integrates to 1

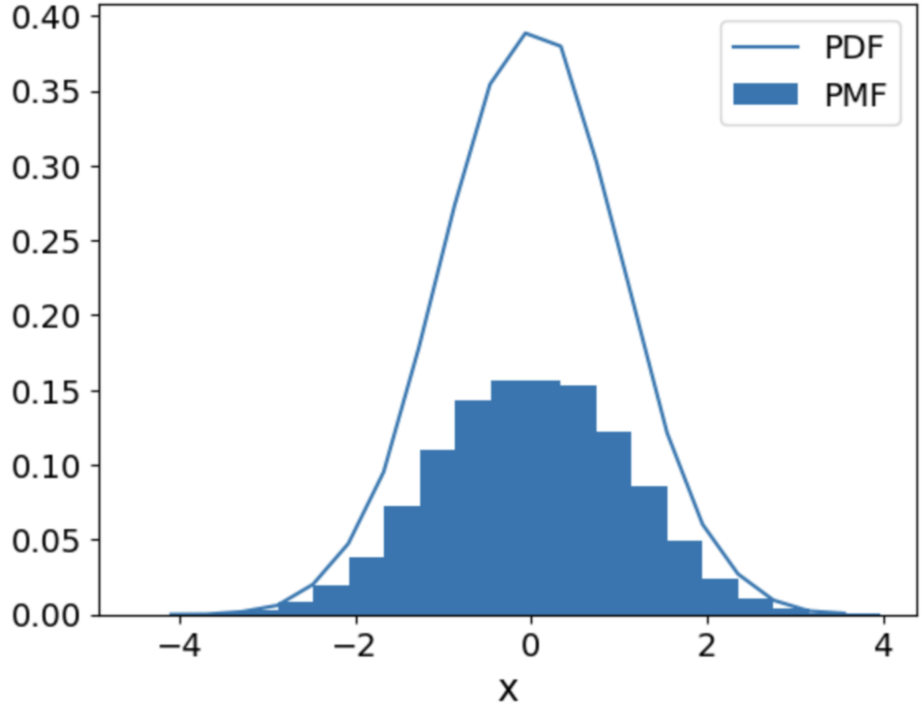
- $P(x = a) = 0$
- $P(a < x < b) = \int_a^b f(x)\,dx$

gives probability of x falling within some interval

# Estimation of PMF and PDF from data samples




Sample counts




```
matplotlib.pyplot.hist(x, bins=None, range=None, density=False,
weights=None, cumulative=False, bottom=None, histtype='bar', align='mid',
orientation='vertical', rwidth=None, log=False, color=None, label=None,
stacked=False, *, data=None, **kwargs) #
```

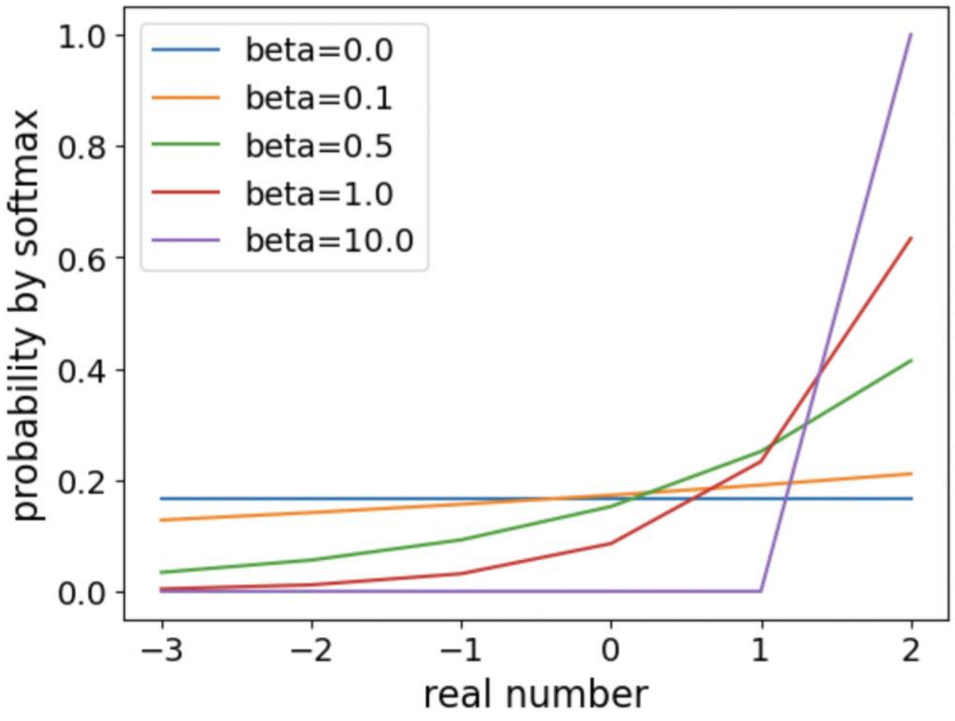
[source]

# Application of normalization

## Softmax

$$\sigma(\mathbf{z})_i = \frac{e^{\beta z_i}}{\boxed{\sum_{j=1}^{K} e^{\beta z_j}}}$$

Make sure it sums up to 1

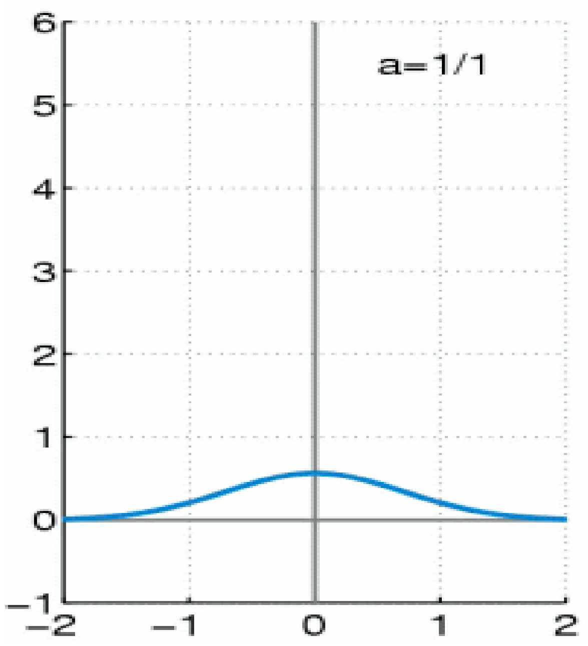



Real numbers → Softmax → PMF

# Dirac Delta function

"Impluse function"

Describes continuous variable that only happens at only one point



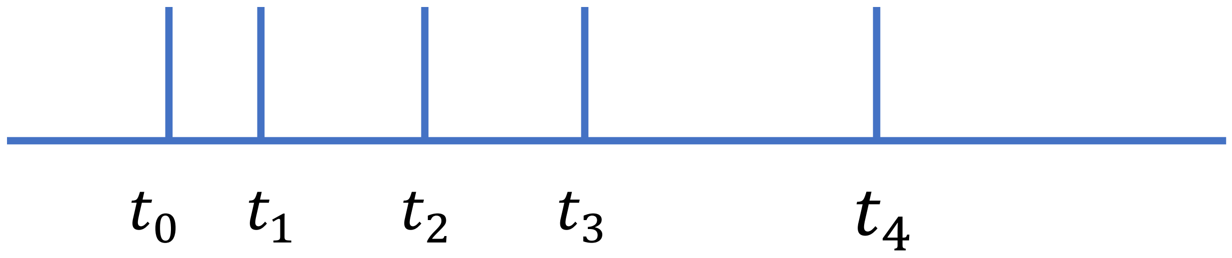

Requires $\int P(x)dx = 1$ and $P(x) = 0$ for all $x \neq 0$

$$\delta(x) = \lim_{a \to 0} \frac{1}{\sqrt{2\pi a^2}} e^{-\left(\frac{x}{2a^2}\right)^2}$$

Infinity high peak!

Can you write down an alternative form of the Delta function?
(hint: the limit of a uniform distribution)

## Spike train




$$S(t) = \sum_i \delta(t - t_i)$$

$\int_{t_a}^{t_b} S(t)dt =$ the spike counts during $[t_a, t_b]$

## A quick recap of joint, marginal, and conditional probability

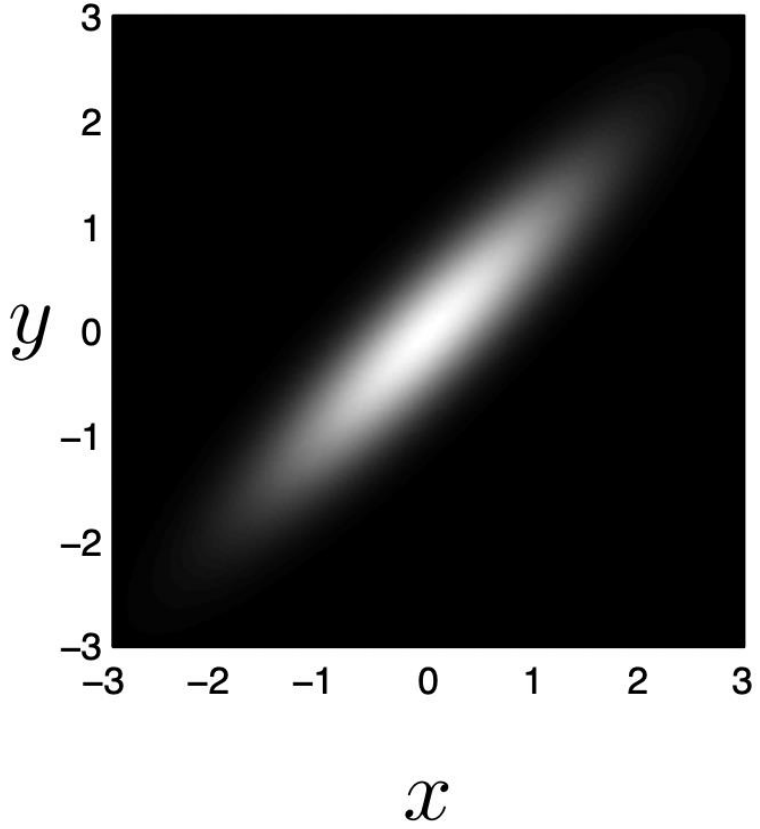
P(x,y)

| y \ x | 1 | 2 | 3 |
|---|---|---|---|
| 3 | 0.1 | 0 | 0.1 |
| 2 | 0 | 0.3 | 0 |
| 1 | 0.2 | 0.1 | 0.2 |

1. Compute the marginal P(x)
2. Compute the marginal P(y)
3. Compute the conditional P(y | x = 2)
4. Compute the conditional P(x | y = 1)
5. What is the most probable value for y?
6. What is the conditional P(x | y > 1)?
7. What is the conditional P(x | x = y)?
8. What is the conditional expectation $\mathbb{E}[x \mid y = 3]$

# joint distribution

$$P(x, y)$$

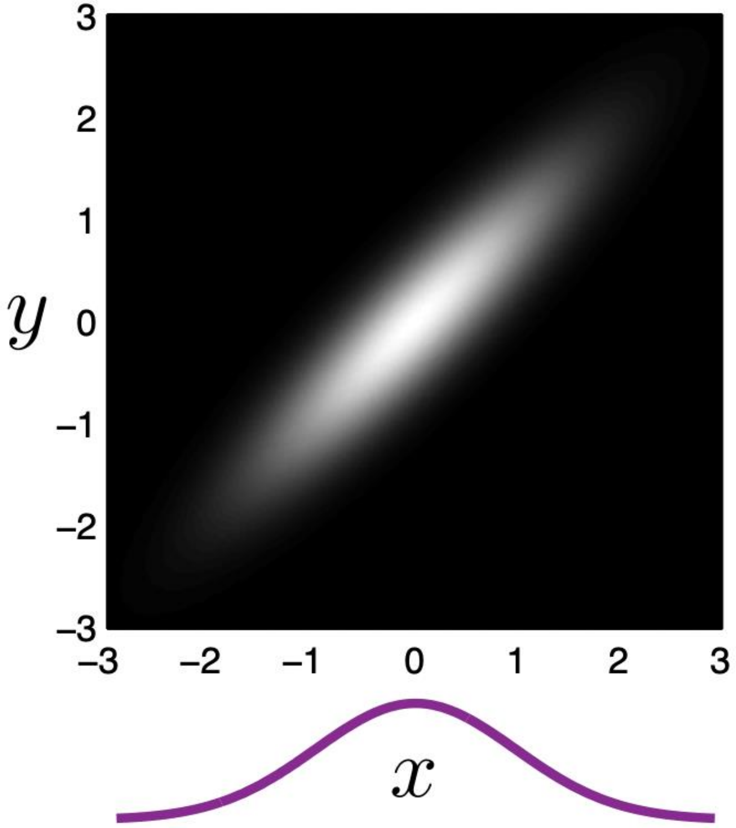



- positive
- sums to 1

$$\iint P(x, y)\, dx\, dy = 1$$

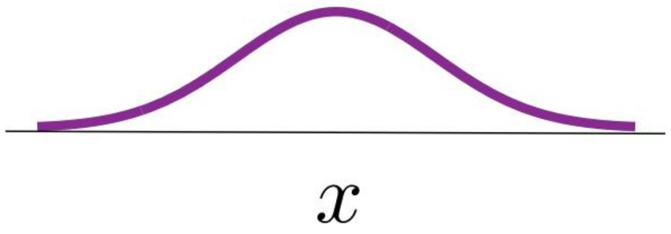
# marginalization (“integration”)

$$P(x,y)$$




“stacking”

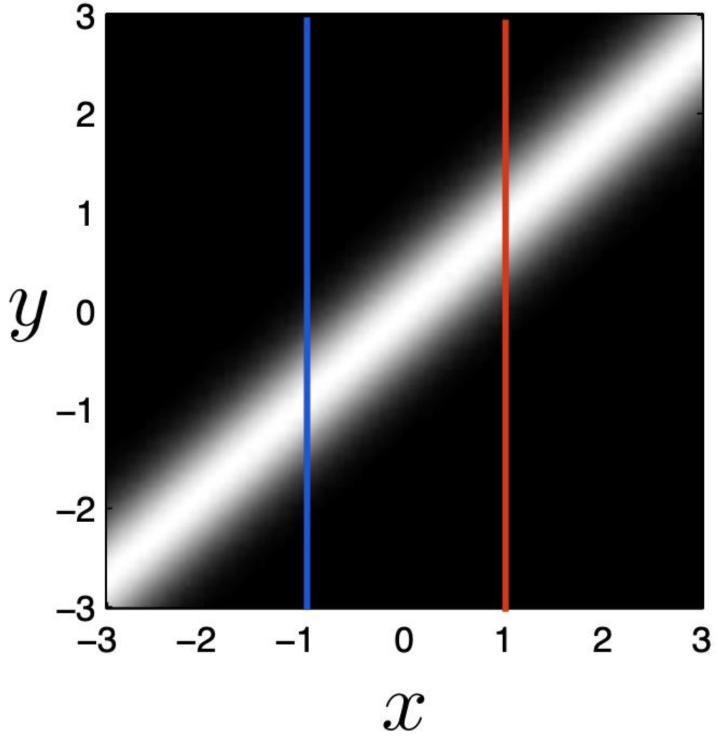
$$P(x) = \int P(x,y)dy$$



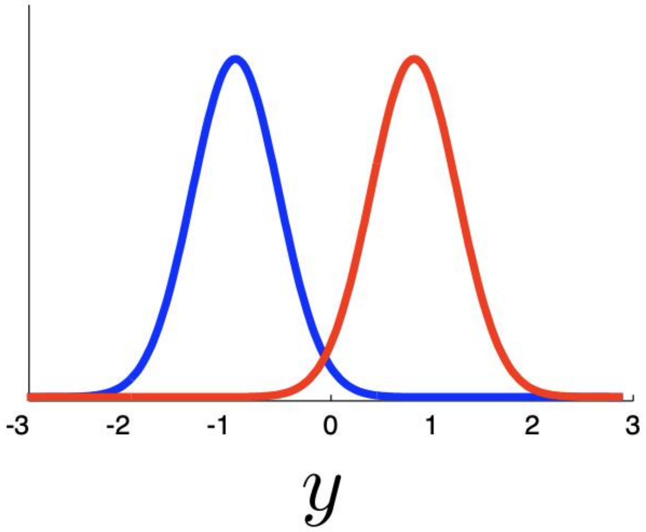

**Very very frustrating in high-dimension**

# conditional densities

$$P(y|x)$$

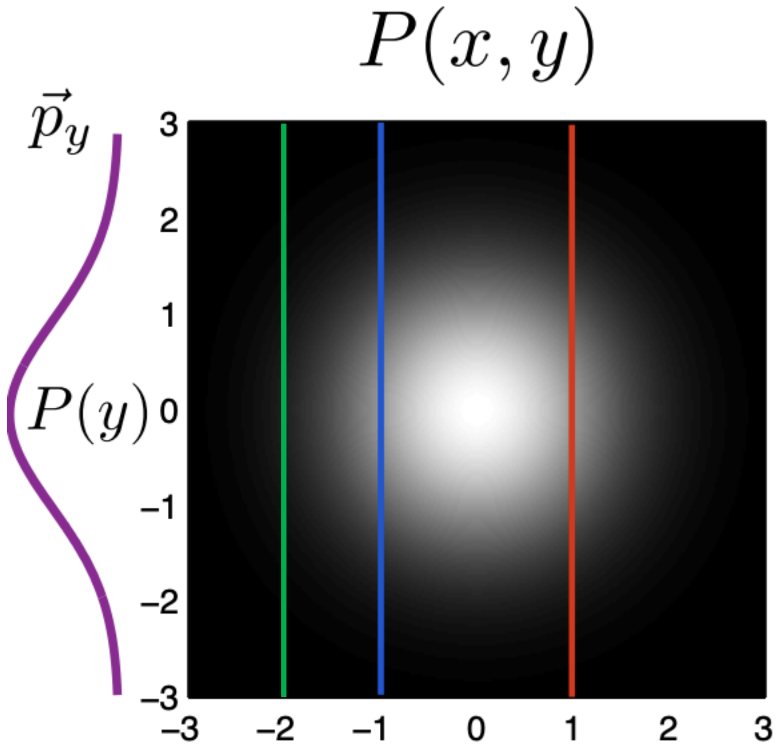



$$P(y|x) = \frac{P(x,y)}{P(x)}$$




**"Slicing and renormalizing"**

# independence



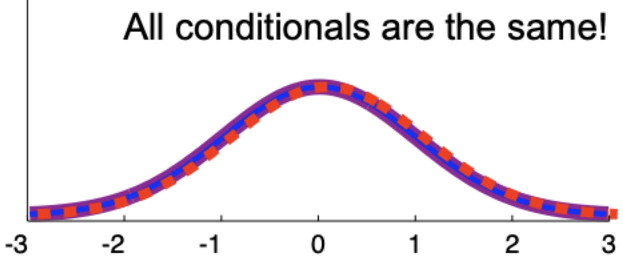

**Original definition:**

$$P(x,y) = P(x)P(y)$$

**Equivalent definition:**

$$P(y|x) = P(y) \quad \text{for all x}$$

Prove it!



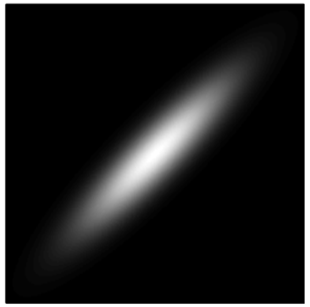

Why does slicing in different parts of the sphere result in the same marginal distribution?

# Correlation vs. Dependence

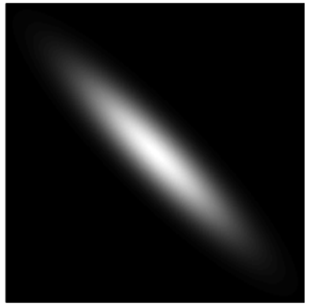
## 1. Correlation coefficient

$$\mathrm{corr}(x, y) = \frac{\mathbb{E}[(x - \bar{x})(y - \bar{y})]}{\sqrt{\mathrm{var}(x)\mathrm{var}(y)}}$$

positive correlation



negative correlation



Linear relationship
between x and y

Q: Can you draw a distribution that is uncorrelated but dependent?
Q: if two random variables are independent, are they correlated?

## Part 2
## Gaussian and Poisson distribution

- PDF of Gaussian
- Central limit theorem
- Standardization
- Multivariate Gaussian

- Definition of Poisson distribution
- Deviation from the binomial distribution

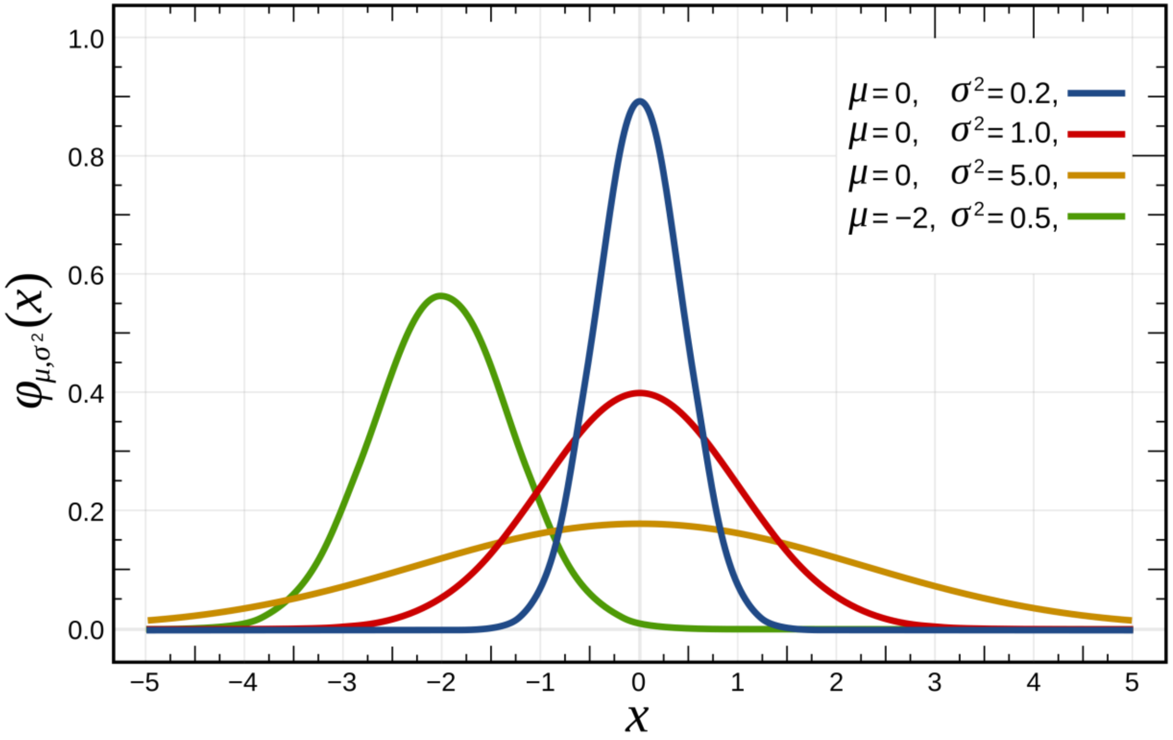
## Gaussian distribution

**Worth memorizing it!**

Mean

$$f(x) = \frac{1}{\sqrt{2\pi\boxed{\sigma^2}}} e^{-\frac{(x-\boxed{\mu})^2}{2\sigma^2}}$$

Variance




**Notation:** $X \sim \mathcal{N}(\mu, \sigma^2)$

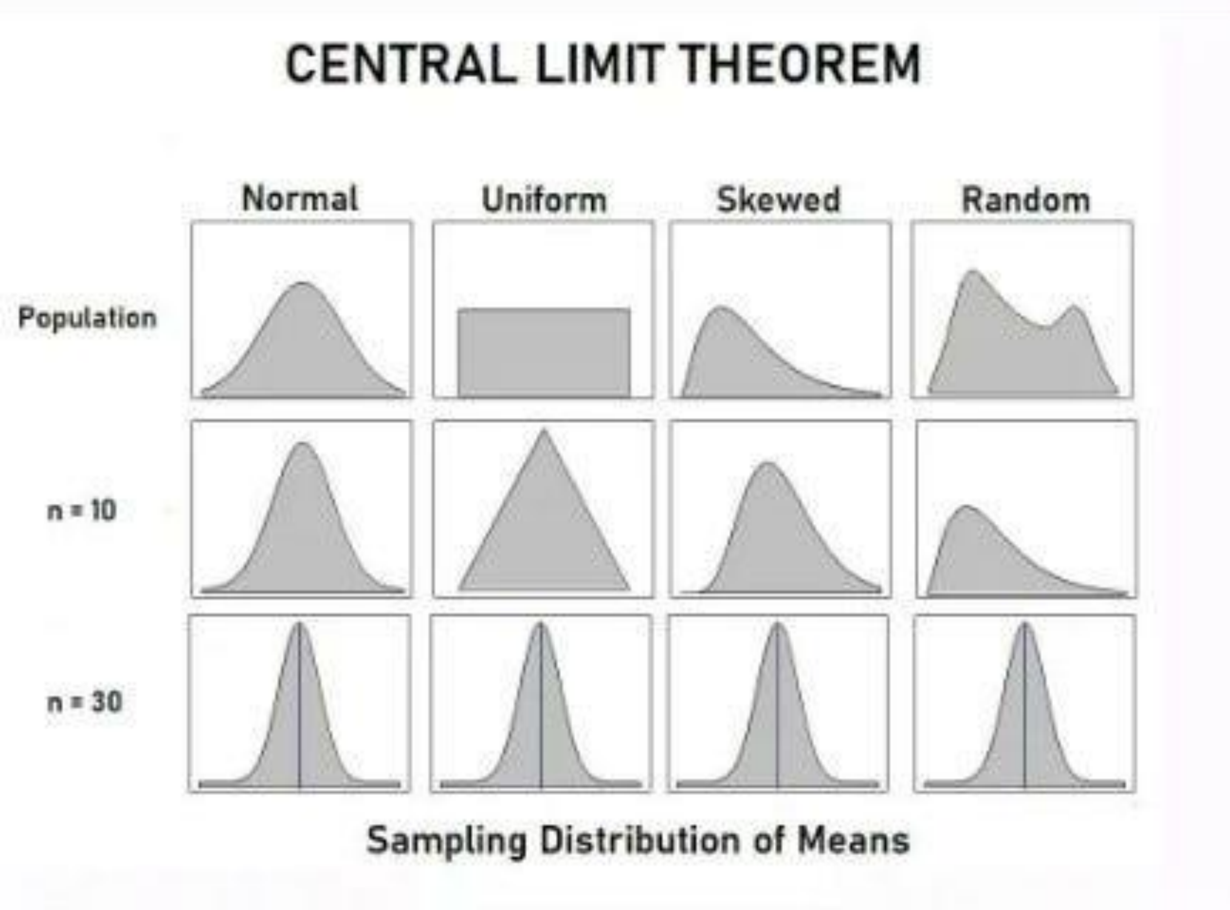
**Or** $P(x) = \mathcal{N}(x; \mu, \sigma^2)$

**Symmetry**

# Central limit theorem

i.i.d: identical and independent distributed

The sum of many i.i.d random variables is Gaussian, regardless of their original distribution



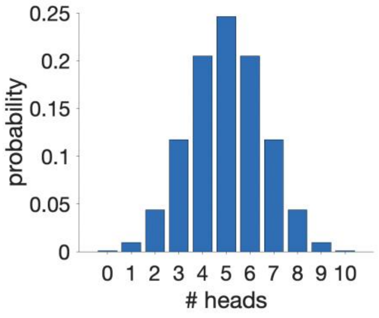

If the mean of the distribution is $\mu$ and the variance is $\sigma^2$, then the sum $\sim \mathcal{N}(n\mu, n\sigma^2)$

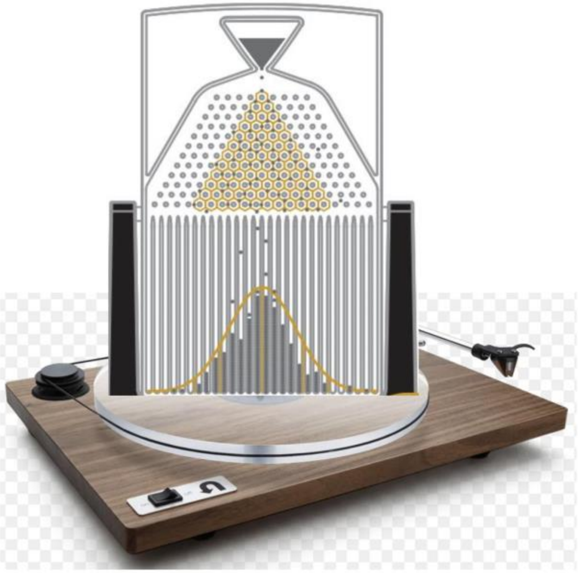
That is why Gaussian is everywhere!

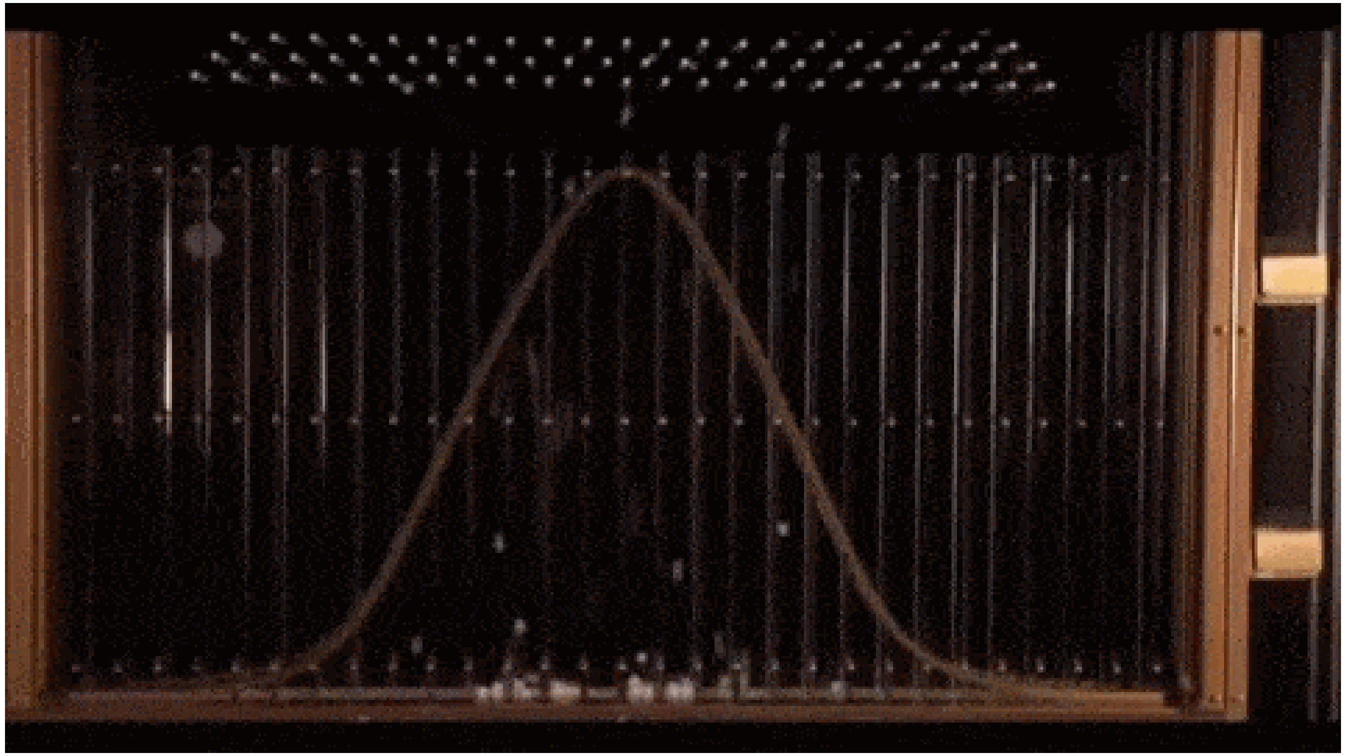
# Galton board

**binomial** $$P(k|\,n,p) = \binom{n}{k} p^k (1-p)^{n-k}$$

(sum of n coin flips)





$$n \to \infty, k \sim \mathcal{N}(np, np(1-p))$$

# Standardization of Gaussian distribution

Any Gaussian distribution can be expressed by a standard Gaussian random variable

$$Z \sim \mathcal{N}(0,1)$$

For any random variable $X \sim \mathcal{N}(\mu, \sigma^2)$, we have

$$X = \mu + \sigma Z$$

Noise

$$X = \mu + \sigma Z$$

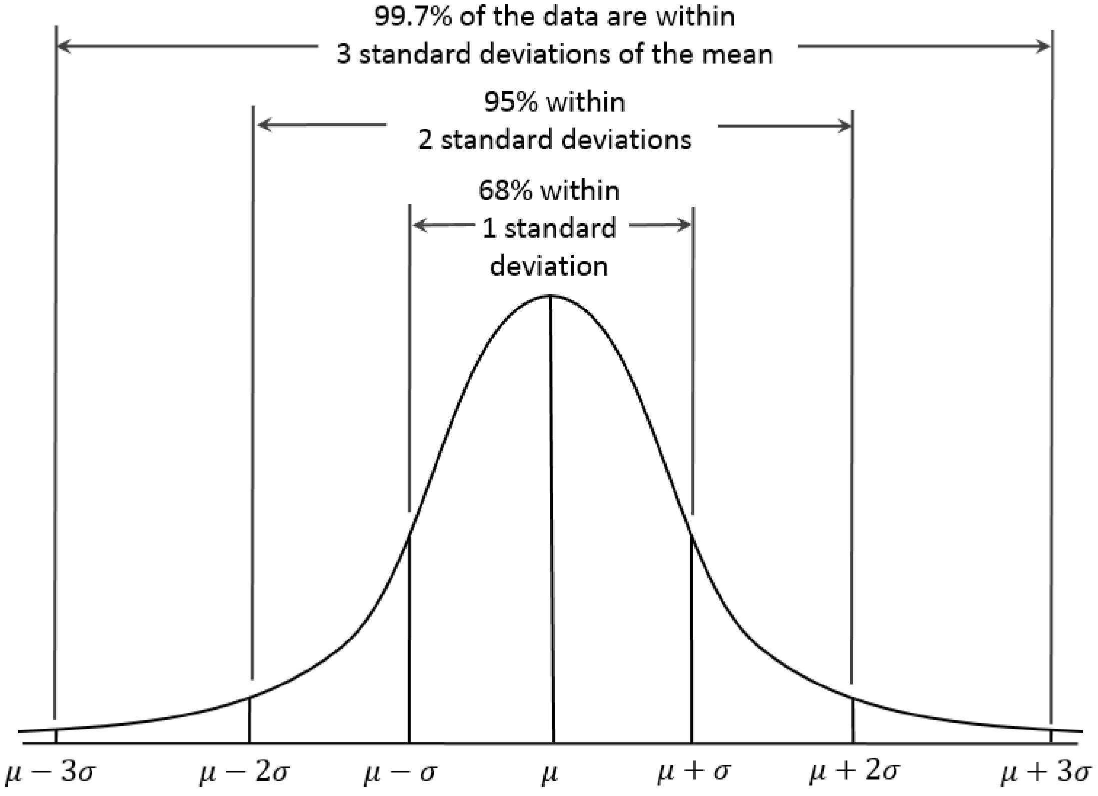
Signal

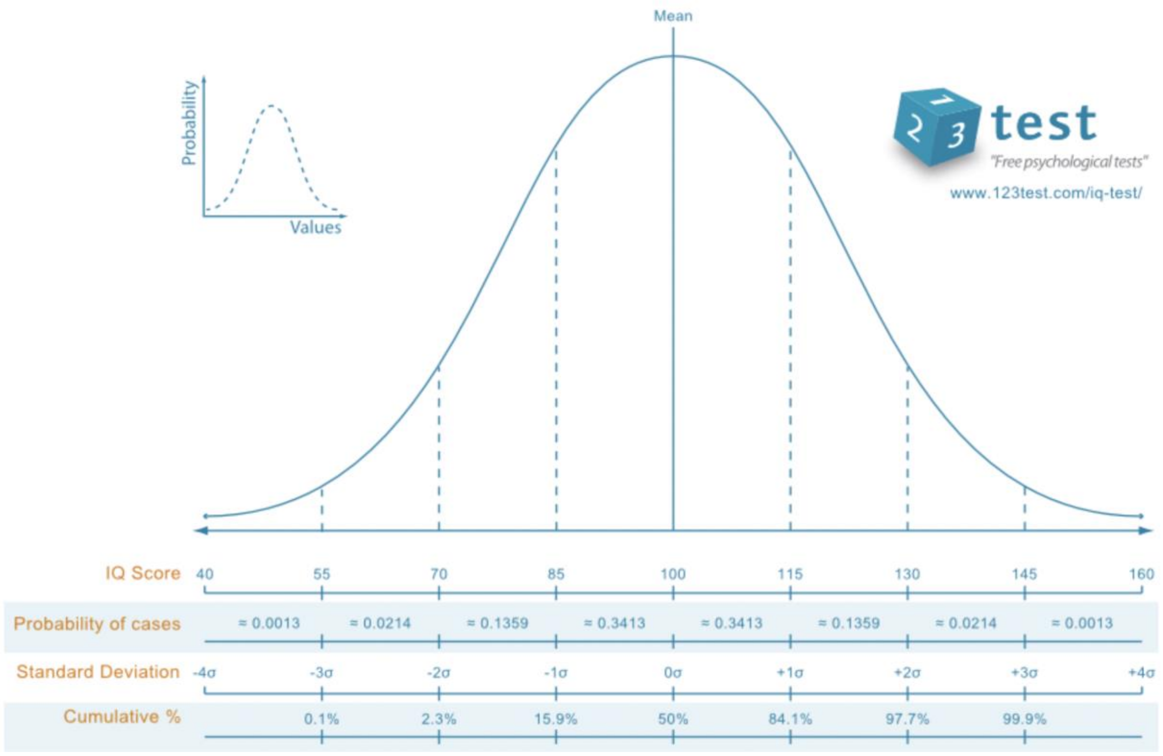
Reversely, any random variable can be standardized as a standard Gaussian

$$Z = \frac{X - \mu}{\sigma}$$

**Z-score!**

# IQ test




$$\mathrm{P}(X>\mu+\sigma)=P(\mu+\sigma Z>\mu+\sigma)=P(Z>1)$$

$$\mathrm{P}(X>\mu+2\sigma)=P(\mu+\sigma Z>\mu+2\sigma)=P(Z>2)$$

CDF

Every "proportion" problem in Gaussian can transformed into a standard Gaussian distribution problem

# Exercise

You will roll a 6 sided dice 10 times. Let $X$ be the total value of all 10 dice = $X_1 + X_2 + \cdots + X_{10}$. You win the game if $X \leq 25$ or $X \geq 45$. Use the central limit theorem to calculate the probability that you win.

Recall that $E[X_i] = 3.5$ and $\mathrm{Var}(X_i) = \frac{35}{12}$.

**Also recall that if the mean of the distribution is $\boldsymbol{\mu}$ and the variance is $\boldsymbol{\sigma}^2$, then the sum** $\sim \mathcal{N}(n\mu, n\sigma^2)$

For any random variable $X \sim \mathcal{N}(\mu, \sigma^2)$, we have $X = \mu + \sigma Z$

Cumulative density function

```
import math
def phi(x):
    """
    Cumulative distribution function
    the standard normal distribution
    """
    return (1.0 + math.erf(x / sqrt(2.0))) / 2.0
```

ANS behind

Bonus: Try to use another method to evaluate the probability of winning and compare it with the result from the central limit theorem.

# Connection to One-Sample t-Test

**“to determine whether the sample mean $\bar{x}$ is significantly larger than a threshold $\mu_0$”**

**Procedure:**

Collect n samples and calculate mean $\bar{x}$ and standard deviation $s$

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

When $n \to \infty$, $t$ is a standard Gaussian variable, so we estimate how is t likely to appear in this value.

**In practice, n is small, t's exact distribution is determined by n. The smaller the n, the heavier the tail.**

## Multivariate Gaussian distribution

$$\mathbf{X} = (X_1, \ldots, X_k)^{\mathrm{T}}$$

$$\mathbf{X} \ \sim \ \mathcal{N}(\boldsymbol{\mu}, \, \boldsymbol{\Sigma})$$

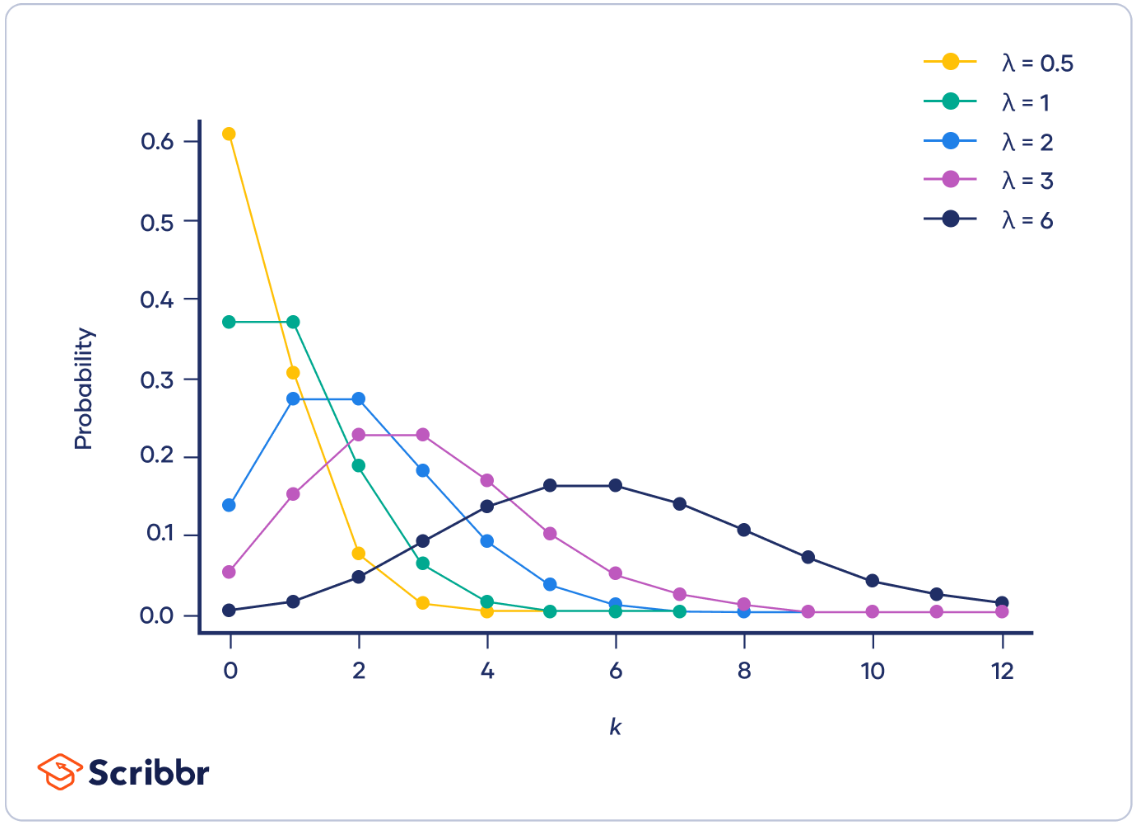
$$f_{\mathbf{X}}(x_1, \ldots, x_k) = \frac{\exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right)}{\sqrt{(2\pi)^k |\boldsymbol{\Sigma}|}}$$

# numpy.random.multivariate_normal

```
random.multivariate_normal(mean, cov, size=None, check_valid='warn', tol=1e-8)
```

# Poisson distribution

Definition: describes the number of events occurring within a fixed interval of time or space, given that these events occur with a known constant rate and independently of the time since the last event.




PMF!

Always positive

Only one parameter $\lambda$

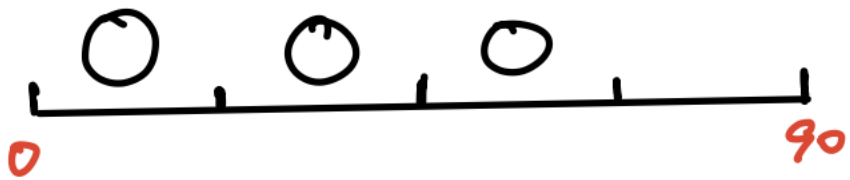
$$P(k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

How does this come out?

## Consider a 90-minute football game with k=3 goals

Now let's segment the game into N chunks, and assume each chunk has probability p to have a goal, 1-p without a goal.

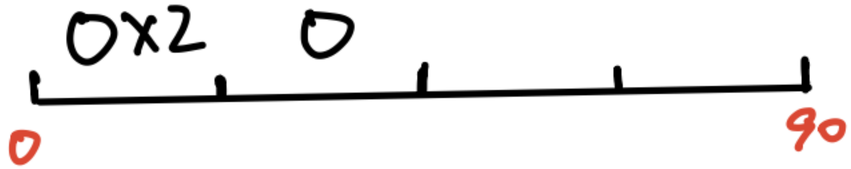
If N=4 and the game is not very dramatic




$$P(k=3)=\binom{4}{3}p^3(1-p)^1$$

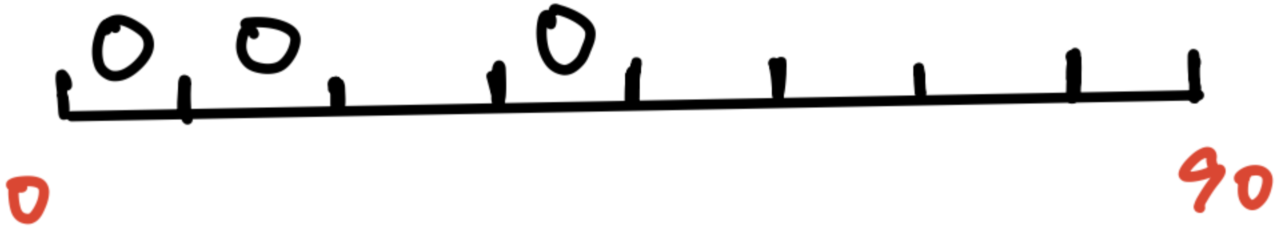
Binomial distribution

However, it is possible to have 2 or more goals in one chunk!



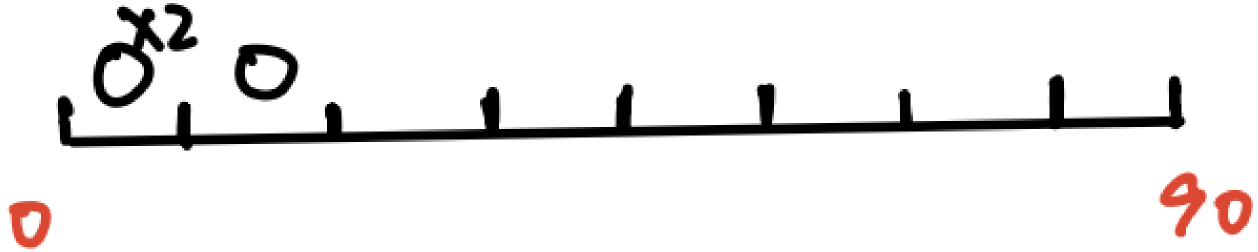

Binomial distribution fails!

Try to make more chunks




$$P(k=3) = \binom{8}{3} p^3(1-p)^5$$

Still may be over-crowded for one chunk…




So…

Infinite chunks!


$$P(k=3) = \lim_{N\to\infty} \binom{N}{3} p^3 (1-p)^{N-3}$$

For any goal number k $P(K) = \lim_{N\to\infty} \binom{N}{k} p^N (1-p)^{N-K}$

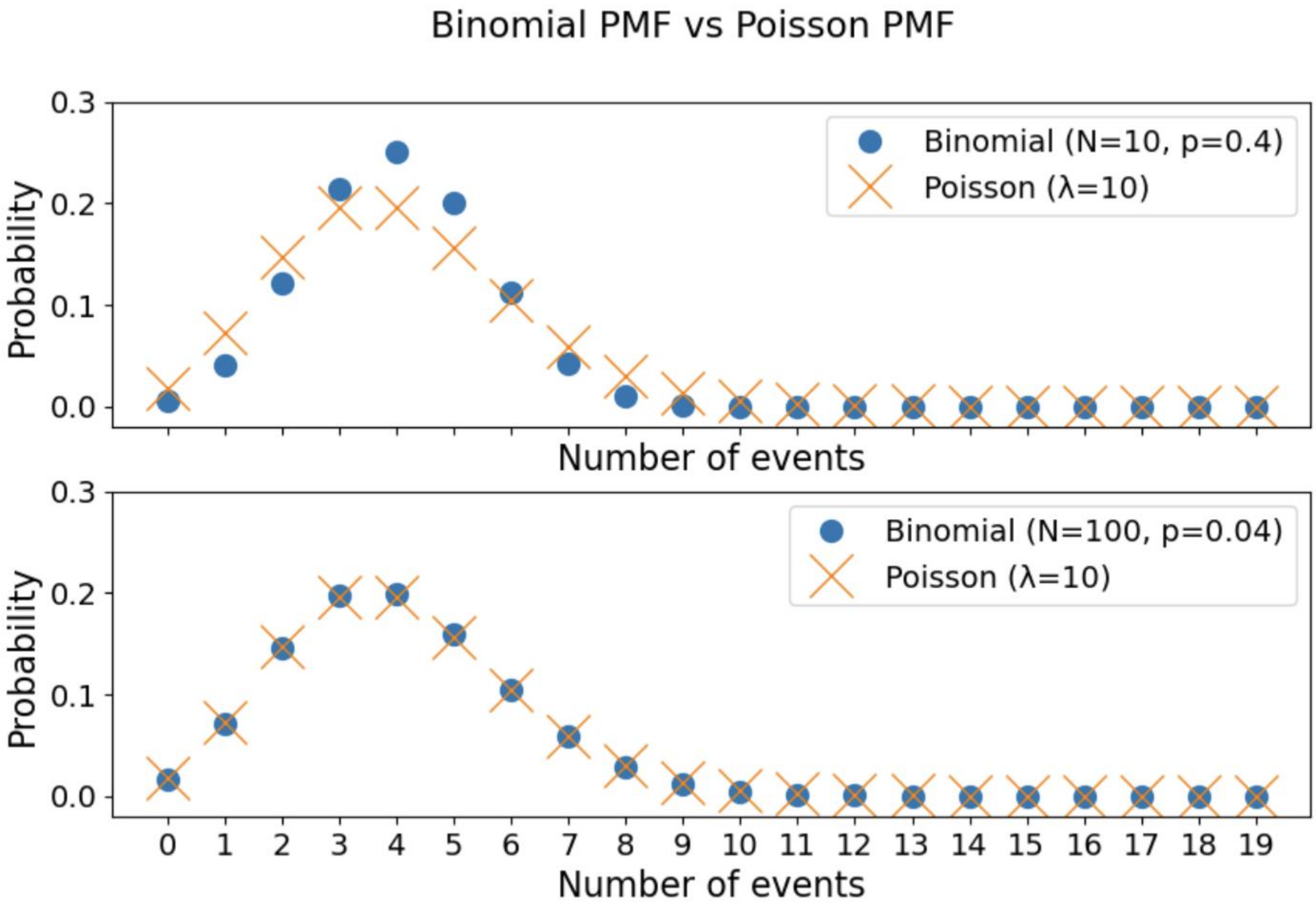
$\lambda = Np$ Rate (=3 in this case)

$p = \dfrac{\lambda}{N}$ Infinitely small

$$\lim_{N\to\infty} \binom{N}{k} \left(\frac{\lambda}{N}\right)^N \left(1-\frac{\lambda}{N}\right)^{N-k} = \frac{\lambda^k}{k!} e^{-\lambda}$$

Poisson distribution

Definition: describes the number of events occurring within a fixed interval of time or space, given that these events occur with a known constant rate and independently of the time since the last event.


Binomial PMF vs Poisson PMF
Probability
0.3
0.2
0.1
0.0
Binomial (N=10, p=0.4)
Poisson (λ=10)
Number of events
Probability
0.3
0.2
0.1
0.0
Binomial (N=100, p=0.04)
Poisson (λ=10)
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
Number of events

$$\lambda = Np$$

## Quick questions:

1. The firing rate of a neuron is 10spk/s, what is the probability of this neuron spiking 20 times in a second?

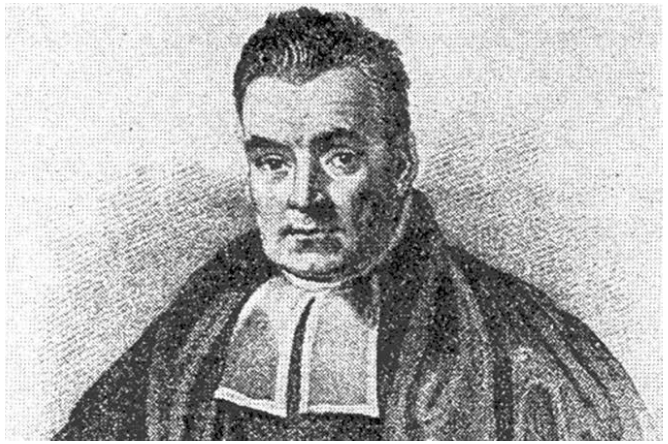
**2. In the last 100 years, there have been 93 earthquakes measuring 6.0 or more on the Richter scale. What is the probability of having 3 earthquakes in the same year that all measure 6.0 or more?**

# Part3: Bayes’ theorem

1. Content of Bayes’ theorem
2. Intuition of  Bayes’ theorem

## Bayes' theorem

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$



**Proof:**

$$P(A,B) = P(A|B)P(B)$$

$$P(A,B) = P(B|A)P(A)$$

# Interpretation

Likelihood Prior

Posterior

$$\underbrace{P(A|B)}_{\text{Posterior}} = \frac{\overbrace{P(B|A)}^{\text{Likelihood}}\overbrace{P(A)}^{\text{Prior}}}{\underbrace{P(B)}_{\text{Evidence}}}$$

Evidence

A: unknown
B: known

A: parameter of the model
B: model and data

A: latent state
B: behavior/neural data

A: neural activity
B: stimulus

...

Apply for everything!

# An intuition for Bayesian estimation

**Medical testing**

A random person goes to the doctor to get a medical test for a rare disease. On average, only 1% of the population has this disease. The test is pretty accurate: it gives a positive result for 99% of those who have the disease, and gives negative result for 90% of those who do not have the disease. What are the chances that this person has the disease if the test comes out positive?

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \Rightarrow P(has\ disease|+) = \frac{P(+|has\ disease)P(has\ disease)}{P(+)}$$

Here

A: unknown, if has the disease

B: known, test result

## Goal:

$$P(has\ disease|+) = \frac{P(+|has\ disease)P(has\ disease)}{P(+)}$$

$$P(+|has\ disease) = 0.99$$

"The test is pretty accurate: it gives a positive result for 99% of those who have the disease, and gives negative result for 90% of those who do not have the disease."

$$P(has\ disease) = 0.01$$

"On average, only 1% of the population has this disease."

How about $P(+)$?

$$P(+) = P(+|has\ disease)P(has\ disease) + P(+|no\ disease)P(no\ disease)$$

"The test is pretty accurate: it gives a positive result for 99% of those who have the disease, and gives negative result for 90% of those who do not have the disease."

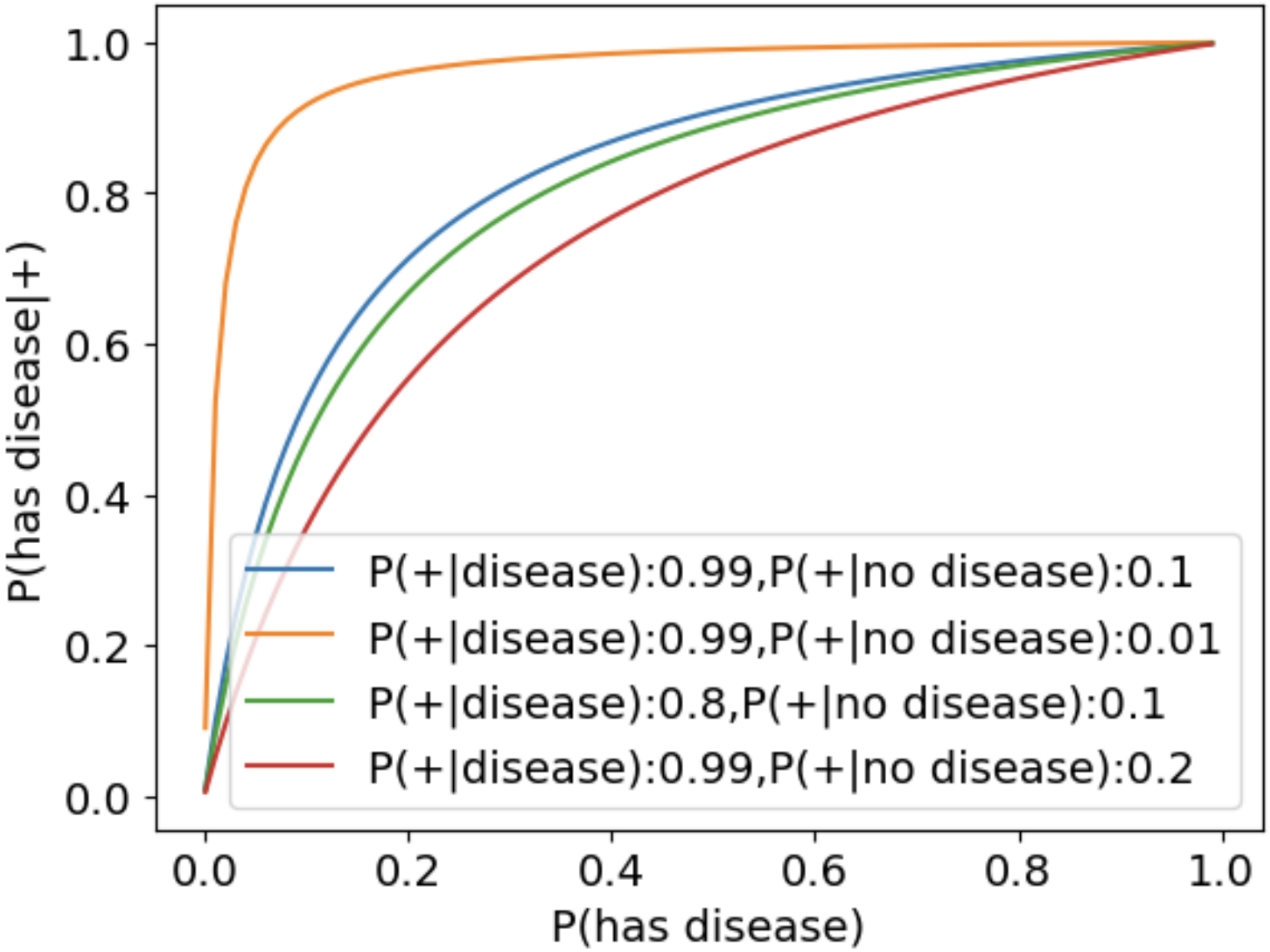
"On average, only 1% of the population has this disease."

$$P(+|no\ disease) = 0.1$$

$$P(no\ disease) = 0.99$$

$$P(+) = 0.99 * 0.01 + 0.1 * 0.99 = 0.1089$$

$$P(has\ disease|+) = \frac{P(+|has\ disease)P(has\ disease)}{P(+)} = \frac{0.99 * 0.01}{0.1089} \approx 9\%$$


1.0
0.8
0.6
0.4
0.2
0.0
P(has disease|+)
0.0
0.2
0.4
0.6
0.8
1.0
P(has disease)
P(+|disease):0.99,P(+|no disease):0.1
P(+|disease):0.99,P(+|no disease):0.01
P(+|disease):0.8,P(+|no disease):0.1
P(+|disease):0.99,P(+|no disease):0.2

# Another intuition for Bayesian estimation


Monty Hall problem

“Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say No. 1, and the host, who knows what's behind the doors, opens another door, say No. 3, which has a goat. He then says to you, "Do you want to pick door No. 2?" Is it to your advantage to switch your choice?”

## Exhaustion method

| Behind door 1 | Behind door 2 | Behind door 3 | Result if staying at door #1 | Result if switching to the door offered |
|---|---|---|---|---|
| Goat | Goat | **Car** | Wins goat | **Wins car** |
| Goat | **Car** | Goat | Wins goat | **Wins car** |
| **Car** | Goat | Goat | **Wins car** | Wins goat |

If staying, the chance to win is $\frac{1}{3}$

If switching, is $\frac{2}{3}$ !

**Tackled by Bayes' theorem**

**Equivalent to solve:**

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)} = \frac{P(E|H)P(H)}{P(E|H)P(H) + P(E|not\ H)P(not\ H)}$$

| Notations | H | E |
|---|---|---|
| | door 1 has a car behind it | The host has revealed a door with a goat behind it |

| Notations | H | E |
|---|---|---|
| | door 1 has a car behind it | The host has revealed a door with a goat behind it |

$$P(H) = \frac{1}{3} \qquad P(not\ H) = \frac{2}{3}$$

$$\boxed{P(E|H) = 1 \qquad P(E|not\ H) = 1}$$

**The host can always open a door with a goat behind it whatever door you choose**

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)} = \frac{P(E|H)P(H)}{P(E|H)P(H) + P(E|not\ H)P(not\ H)} = \frac{1}{3}$$

**Probability to win if stick to door 1**

So $P(not\ H|E) = \frac{2}{3}$ **Probability to win if switch**

# If the host just randomly open an remaining door

$$P(H) = \frac{1}{3}$$

$$P(not\ H) = \frac{2}{3}$$

$$P(E|H) = 1$$

$$P(E|not\ H) = \frac{1}{2}$$

| Notations | H | E |
|---|---|---|
| | door 1 has a car behind it | The host has revealed a door with a goat behind it |

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)} = \frac{P(E|H)P(H)}{P(E|H)P(H) + P(E|not\ H)P(not\ H)} = \frac{1}{2}$$