香港中文大學（深圳）
The Chinese University of Hong Kong, Shenzhen

DDA2001: Introduction to Data Science

Lecture 8: Continuous Random Variable

Zicheng Wang

Recap of Continuous Random Variable

Continuous R.V.

- A continuous random variable can take any value within its range (an interval of a union of multiple intervals of real numbers).
- We cannot list all the possible values and their probabilities as in the discrete case.

How to describe the probability?

- If P(E) = 0, then E is a zero-probability event.
- If E is empty, then E is impossible.

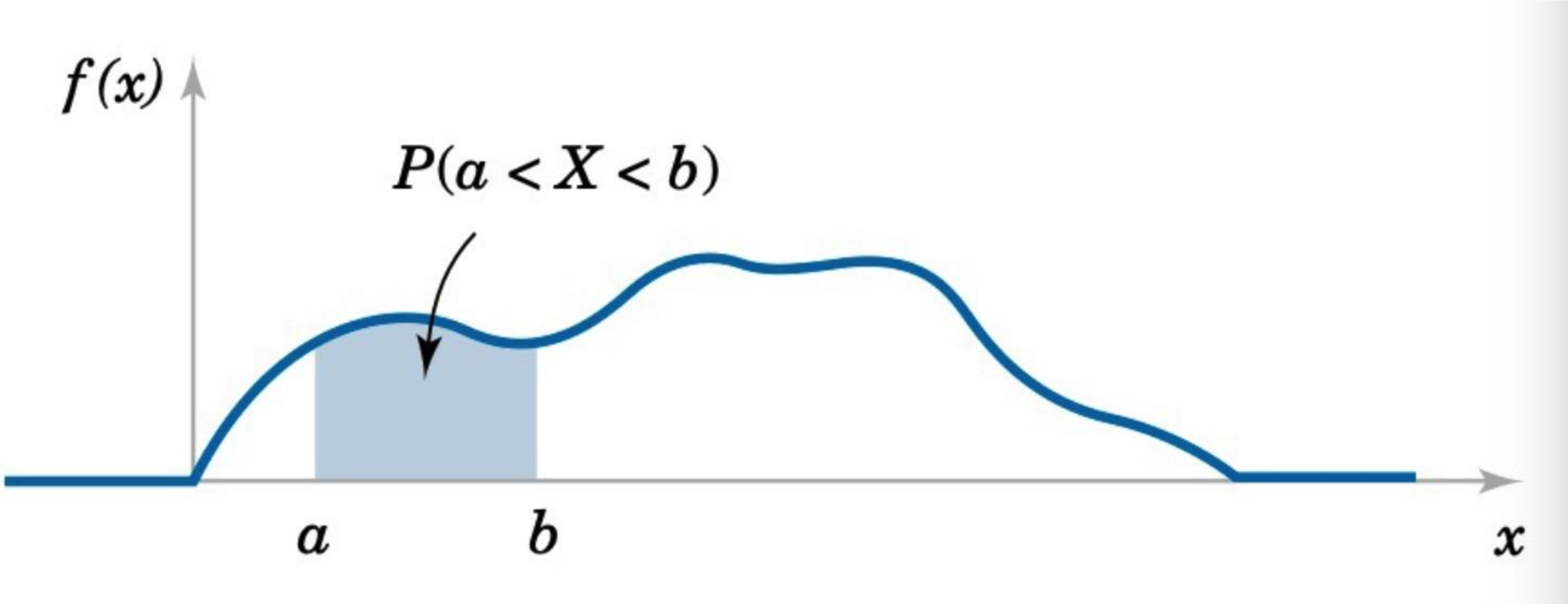
- For a continuous RV X, P(X=x) = 0 but {x} is not an impossible event.

- We will not use the probability mass function (pmf), namely P(X=x).

- Instead, we introduce a function f(ω), called the probability **density** function (pdf).
  - f(ω) > 0, if $\omega \in S$
  - f(ω) = 0, if $\omega \notin S$
  - $\int_{-\infty}^{\infty} f(x)dx = 1.$

Probability of $X \in [a, b]$

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$



Properties of PDF

- For *x* that is not in the sample space, *f(x)=0*
- A large value of *f(x)* means that the values around *x* is more likely to be observed. (remember this implication)
- As a pdf, f(x) can be larger than 1, while as a pmf, f(x) cannot be larger than 1.
  - f(ω) = 2 , if $\omega \in [0, 0.5]$
  - f(ω) = 0, if $\omega \notin [0, 0.5]$

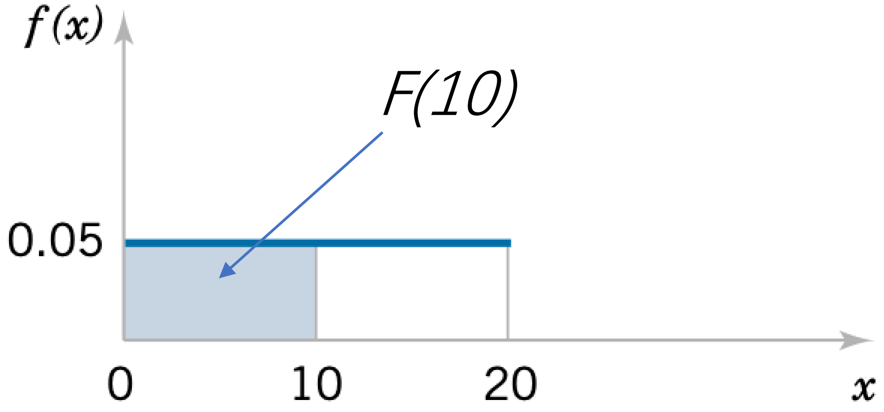
CDF

- Recall: the CDF of a discrete random variable X is

$$F(x) = P(X \leq x) = \sum_{\tilde{x} \leq x} f(\tilde{x})$$

- CDF for continuous random variable is defined as:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)du$$



CDF

- Recall: the CDF of a discrete random variable X is

$$F(x) = P(X \leq x) = \sum_{\tilde{x} \leq x} f(\tilde{x})$$

- CDF for continuous random variable is defined as:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)du$$

✓ $0 \leq F(x) \leq 1$

✓ If $x \leq y$, then $F(x) \leq F(y)$

For both discrete and continuous RVs

Mean and Variance

- Discrete:
  - ✓ Probability mass function.
- Continuous
  - ✓ Probability density function.

$\boxed{\textbf{Summation} \leftrightarrow \textbf{Integration}}$

- Mean

$$\mathrm{E}[X] = \sum x\, f(x)$$

- Variance

$$\mathrm{Var}[X] = \sum (x - \mathrm{E}[X])^2\, f(x)$$

- Mean

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x f(x) dx$$

- Variance

$$\mathrm{Var}[X] = \int_{-\infty}^{\infty} (x - E(X))^2 f(x) dx$$

Expectation of g(X)

$$\mathrm{E}[\mathrm{g}(X)] = \int_{-\infty}^{\infty} g(x) f(x) dx$$

Uniform Distribution

- With the same probability, X takes a value within [a, b], where b>a.

Discrete version: toss a coin, roll a dice.

- What's the pdf?

Uniform Distribution

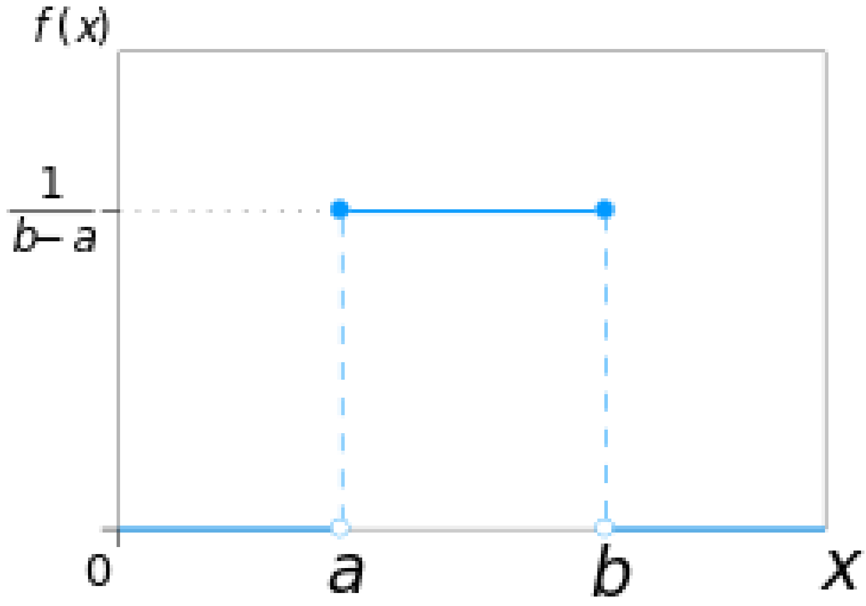
- With the same probability, X takes a value within [a, b], where b>a.

- What's the pdf?

- f(x) = c for $x \in [a, b]$ and f(x) = 0 for $x \notin [a, b]$
- As $\int_{-\infty}^{\infty} f(x)dx = c(b-a) = 1,$ we have
$$c = \frac{1}{b-a}$$

Uniform Distribution

- With the same probability, X takes a value within [a, b]
- $X \sim Uniform(a, b)$



Mean= $(a + b)/2$
Variance=$(b - a)^2/12$

Applications

- Given $X \sim Uniform(0,2)$
- What's the value of $E[2\, e^{X^2+\cos(X)}]$?

Applications

- Given $X \sim Uniform(0,2)$
- What's the value of $E[2\, e^{X^2+\cos(X)}]$?

- f(x) = ½ for $x \in [0,2]$
- $E[2\, e^{X^2+\cos(X)}] = \int_0^2 2\, e^{x^2+\cos(x)} f(x)dx = \int_0^2 e^{x^2+\cos(x)} dx$

How to approximate $\int_0^2 e^{x^2+\cos(x)}dx$?

Given $X \sim Uniform(0,2)$, $E[2\, e^{X^2+\cos(X)}] = \int_0^2 e^{x^2+\cos(x)} dx$

- Draw N samples of $X \sim Uniform(0,2)$: $X_1, X_2, X_3, \ldots, X_N$

- Calculate $\frac{\sum_i 2\, e^{X_i^2+\cos(X_i)}}{N}$

Why? Expectation can be approximated by long-run average.

General Case

- How to calculate $\int_a^b h(x)dx$?

  - Draw N samples of $\mathrm{X} \sim Uniform(a, b)$: $X_1, X_2, X_3, \ldots, X_N$
  - Calculate $\frac{\Sigma_i (\boldsymbol{b}-\boldsymbol{a})\, h(X_i)}{N}$

- $E[h(x)]$ only gives you the average "height" of $h(x)$
- In order to get $\int_a^b h(x)dx$, which is the area, we need to multiply $E[h(x)]$ by $(b-a)$

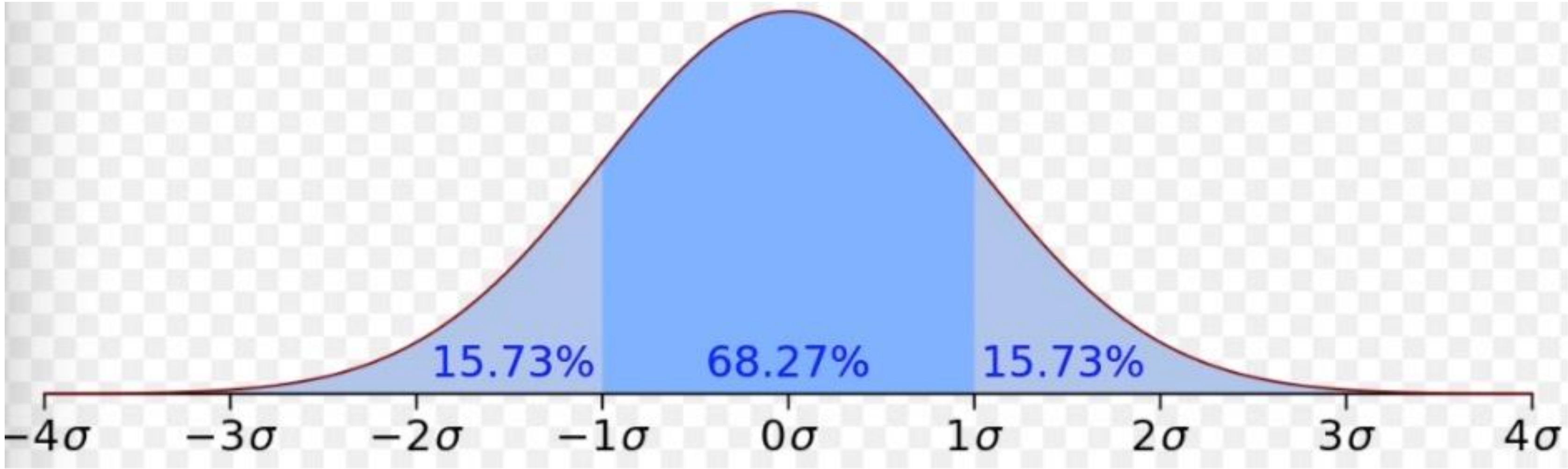
- Let $X \sim Uniform(a, b)$
- f(x) = 1/(b-a) for $x \in [a, b]$
- $E[(b-a)h(x)] = \int_a^b (b-a)h(x)f(x)dx = \int_a^b h(x)dx$

Most important:
Normal distribution

- X can be any real number
- Parameters: μ and σ

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

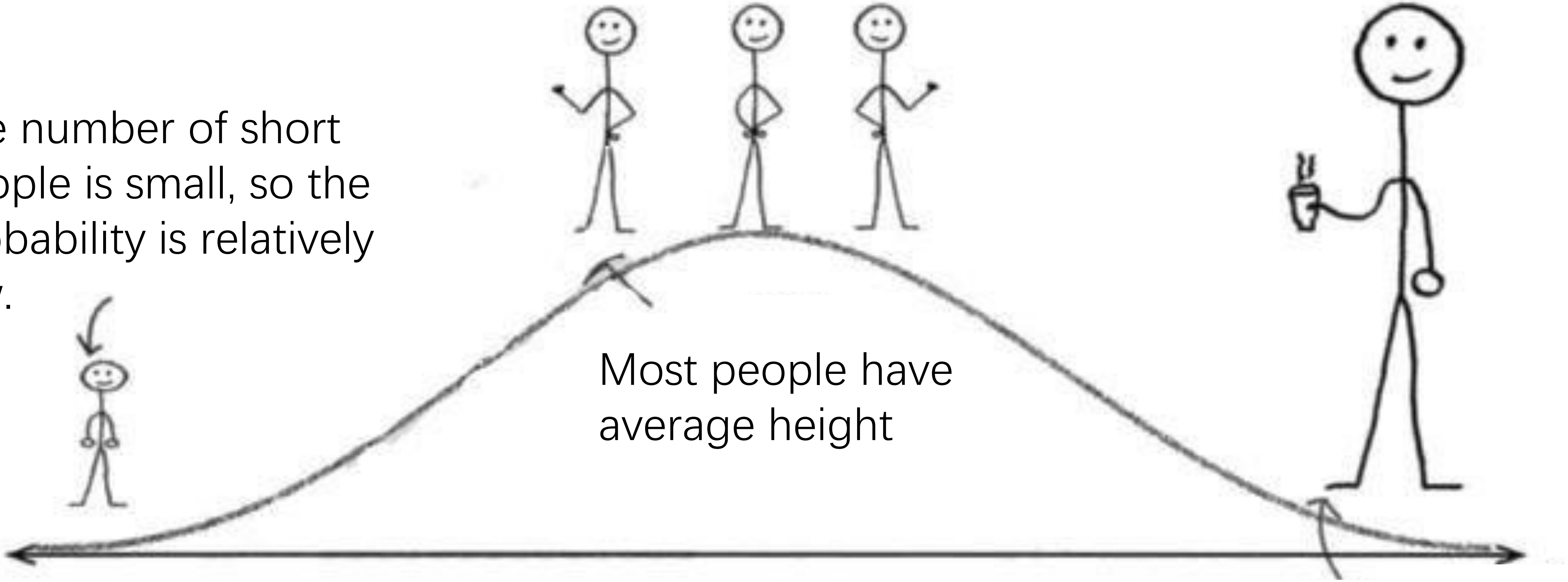
- $X \sim Normal(\mu, \sigma)$



Why we have this distribution?

Normal Distribution: examples

The number of short people is small, so the probability is relatively low.



Most people have average height

There are not many tall people

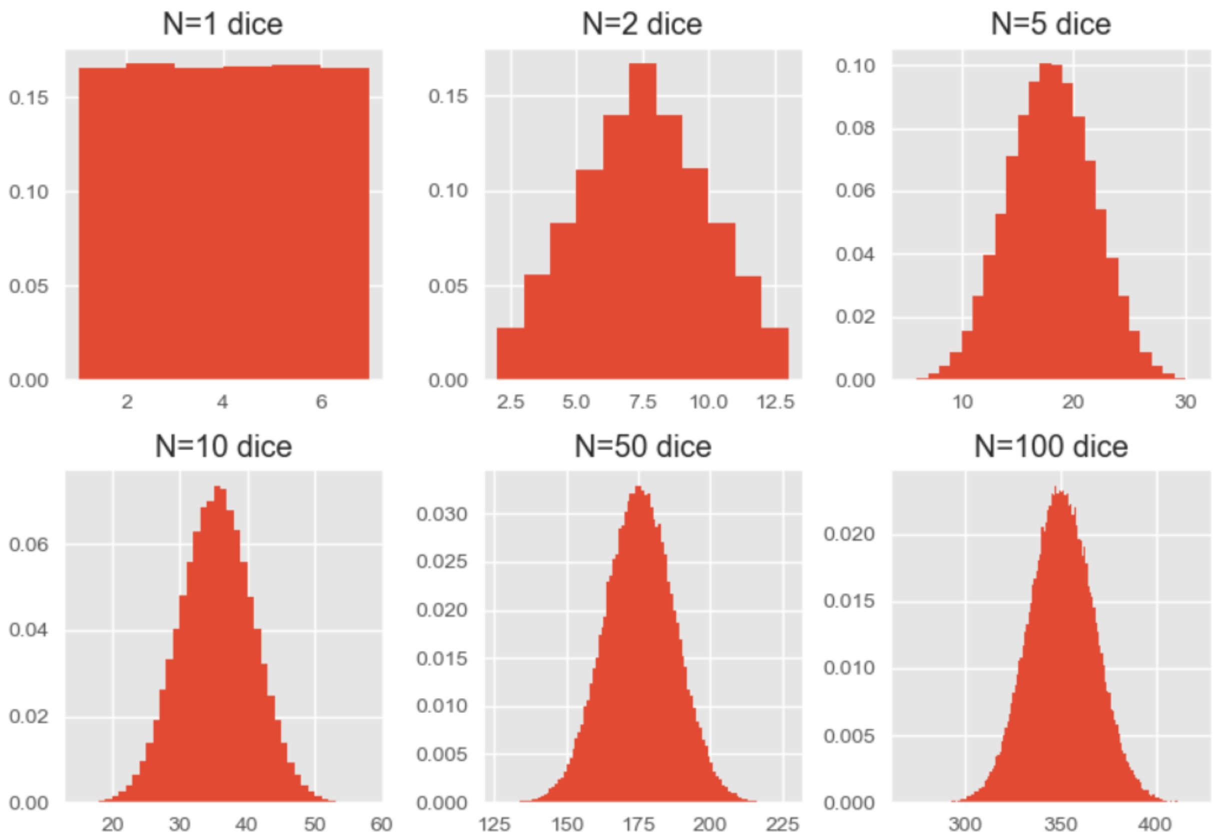
Human Being’s Height

Normal Distribution: examples

- A large number of layers
- When a ball goes through each layer, it randomly goes left or right

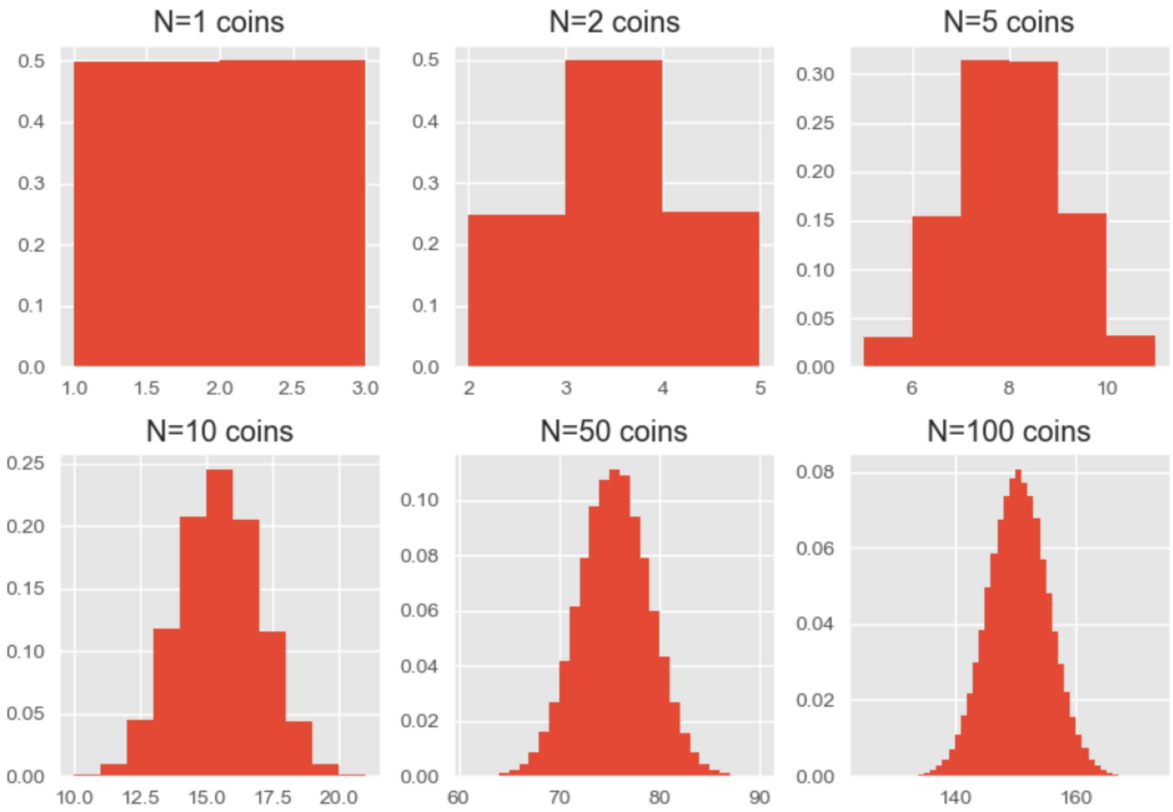
An example

- Toss a die N times
- Let X be the sum
- A demo: N = 1,2,⋯.,100, see the pmf of X



An example

- Flip a fair coin N times
- Let X be the sum (head: 1; tail: 2)
- A demo: N = 1,2,⋯.,100, see the pmf of X



Normal Distribution

Central limit theorem

Lindeberg–Lévy CLT. Suppose $\{X_1, \ldots, X_n\}$ is a sequence of i.i.d. random variables with $\mathbb{E}[X_i] = \mu$ and $\operatorname{Var}[X_i] = \sigma^2 < \infty$. Then as n approaches infinity, the random variables $\sqrt{n}(\bar{X}_n - \mu)$ converge in distribution to a normal $\mathcal{N}(0, \sigma^2)$:

$$\sqrt{n}\left(\bar{X}_n - \mu\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2\right).$$

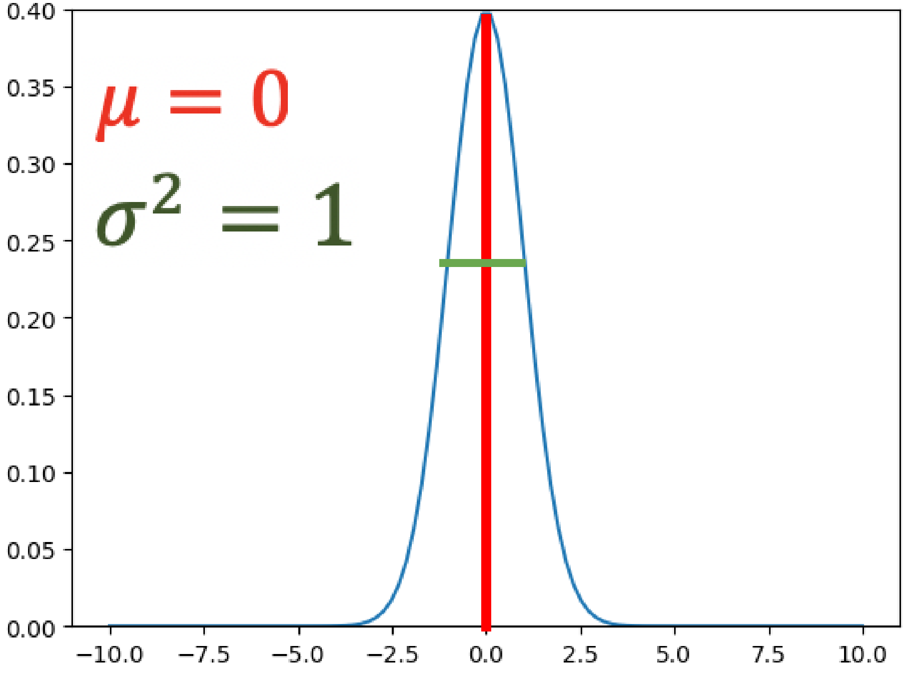
no need to grasp!!!

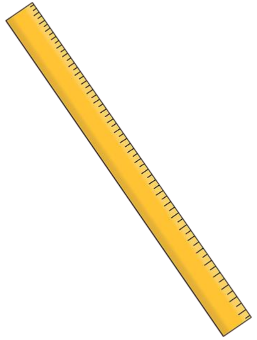
What is the meaning of the parameters?

Mean and Variance

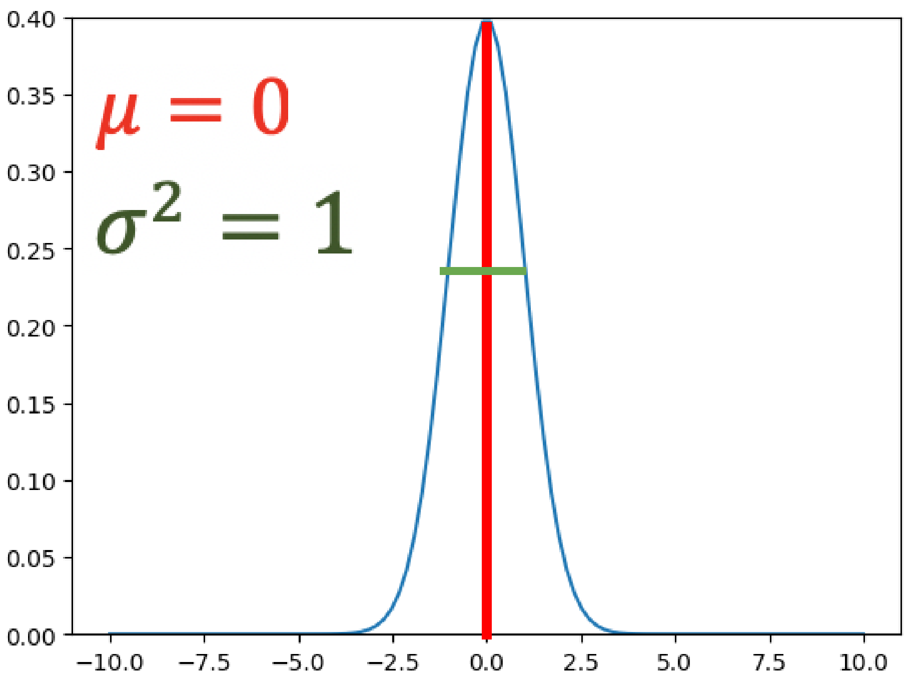
- Mean: μ
- Variance: σ^2

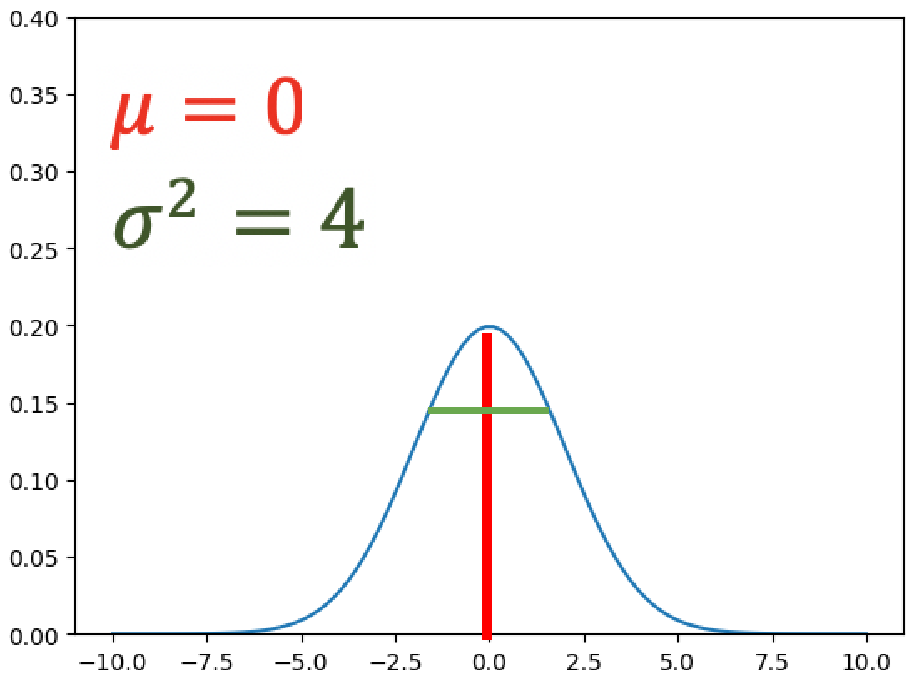
The Normal PDF

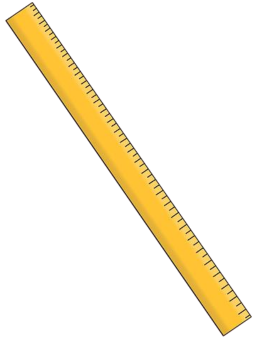




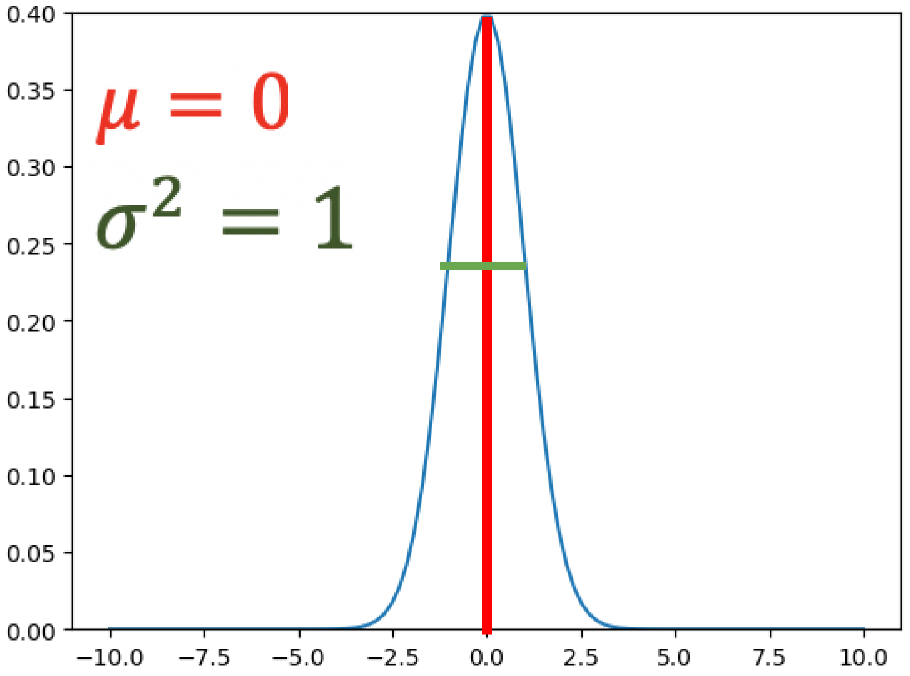
The Normal PDF

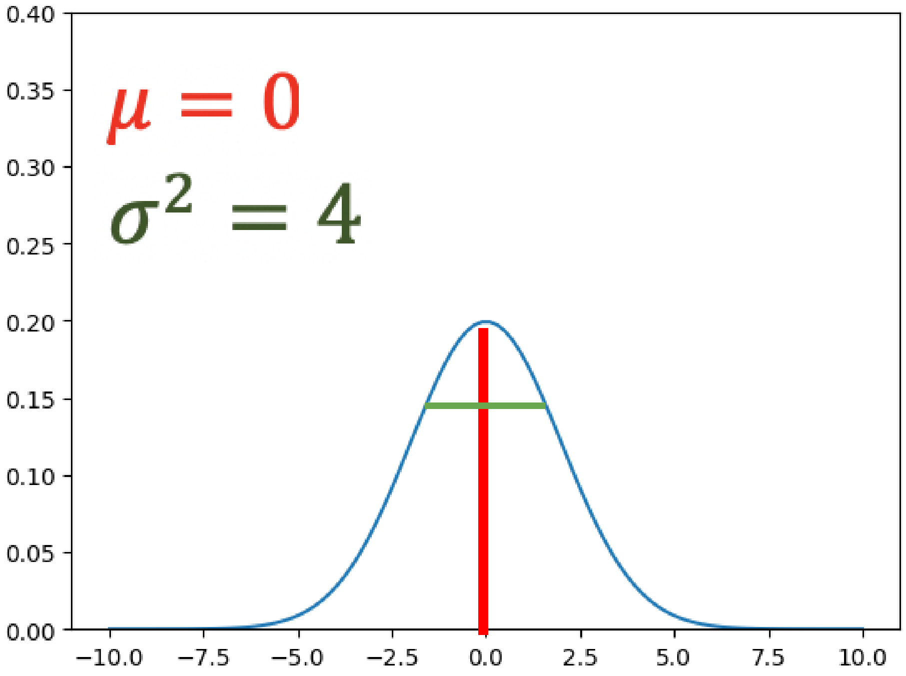


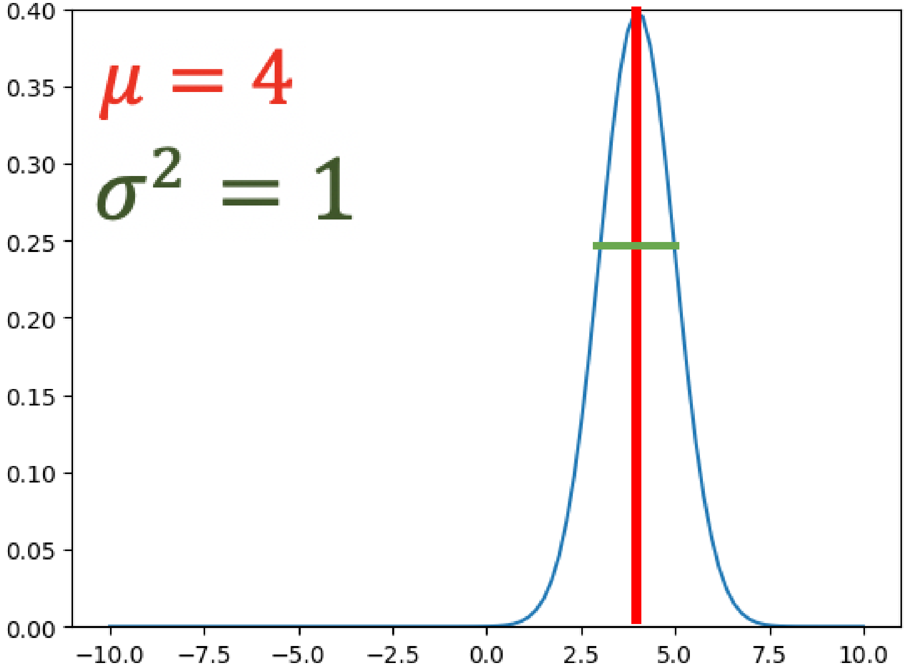


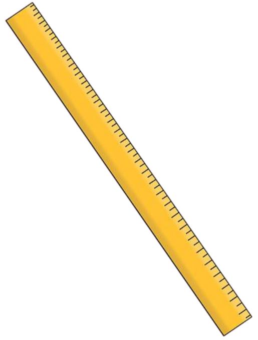


The Normal PDF

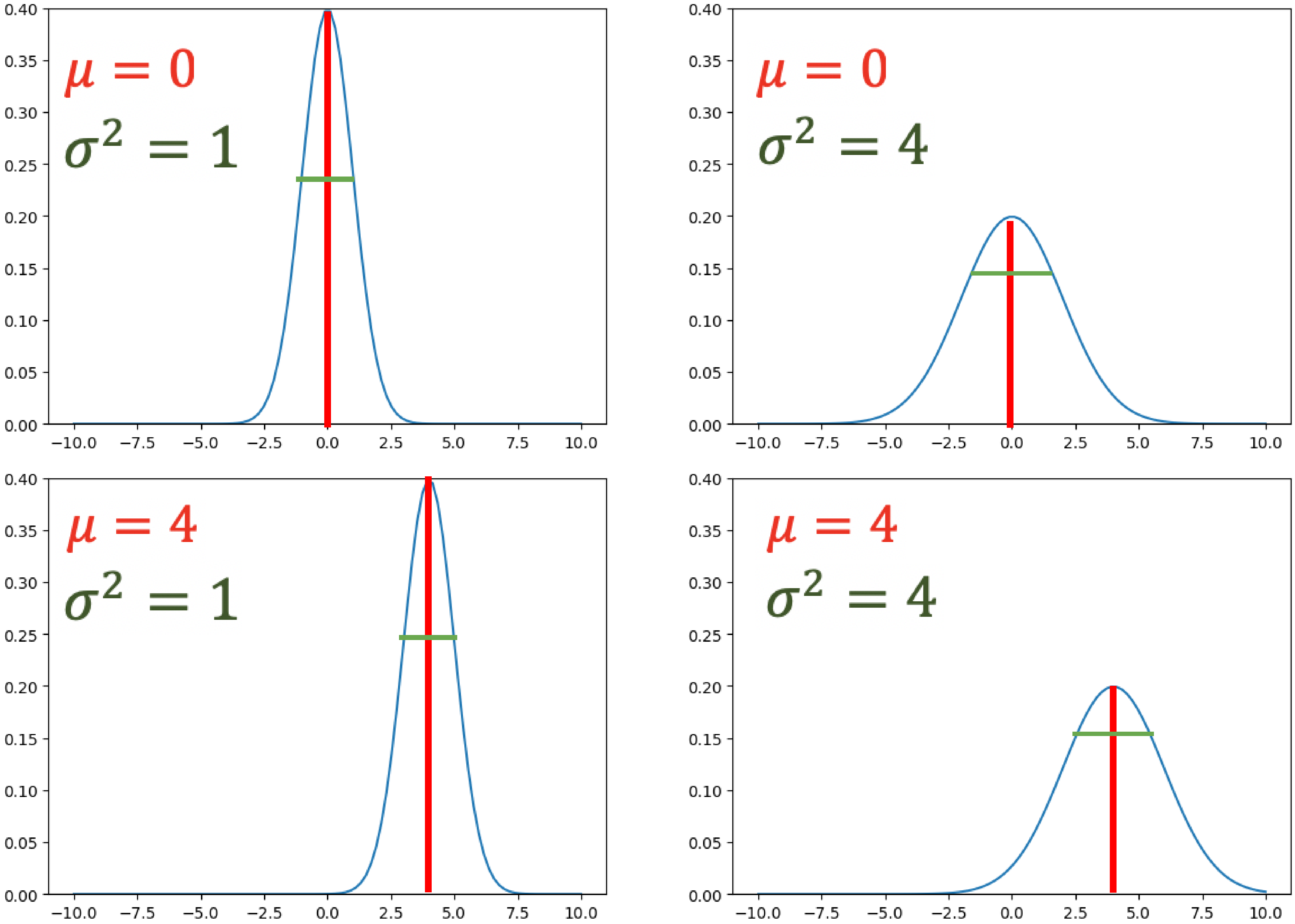


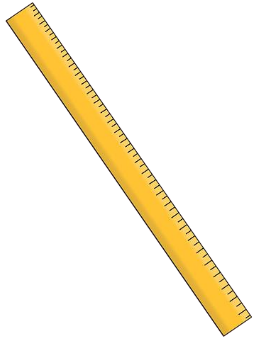






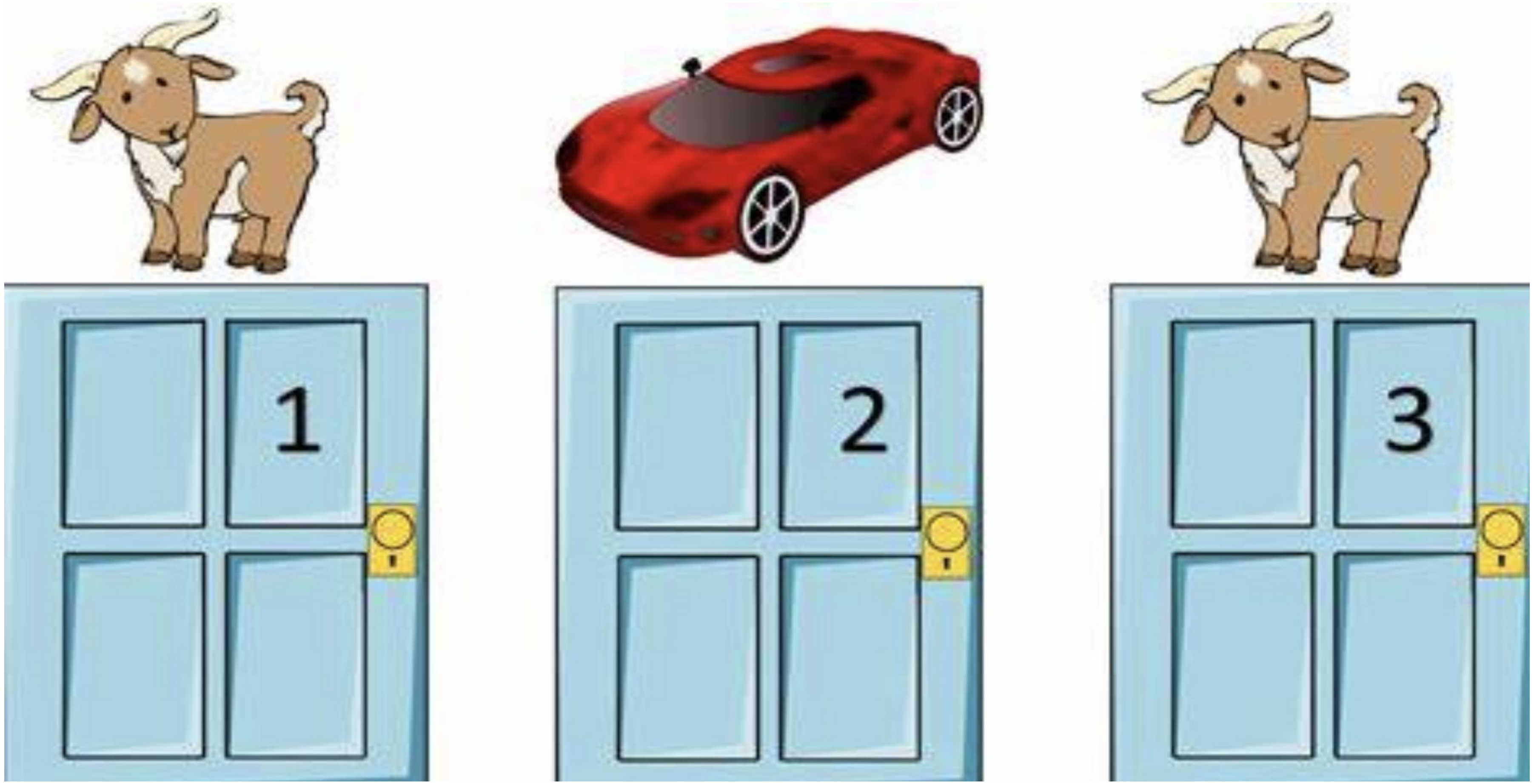
The Normal PDF





Advanced Concepts: Conditional Probability

Monty Hall problem



Monty Hall problem

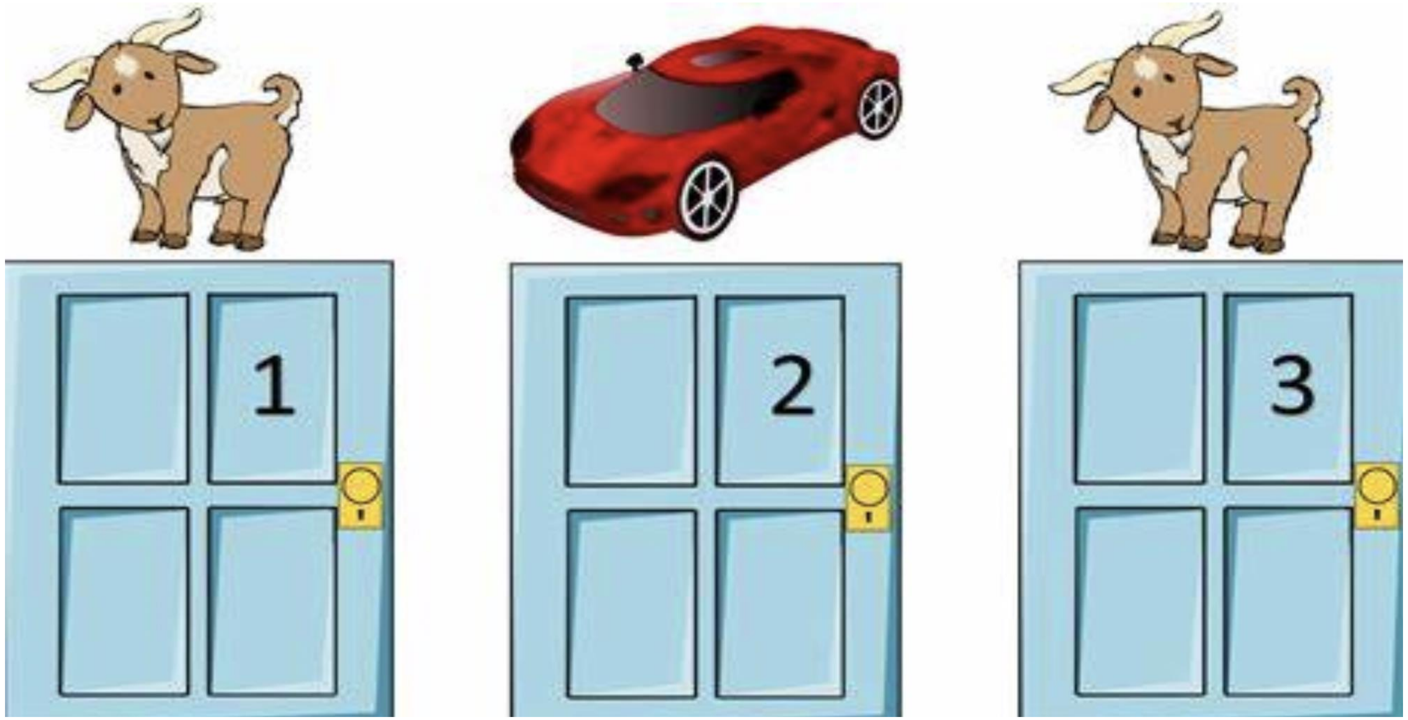
Suppose you're on a game show, and you're given the choice of three doors:

- Behind one door is a car;
- Behind the others, goats.

The car is randomly placed behind one door.

Which door should you choose to open?

- After you choose No. 1, I open a door (among the other two doors) with a goat behind,.
- Would you like to switch to the other door?
- Now you have additional information\event!



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

P(win when stay)=1/3

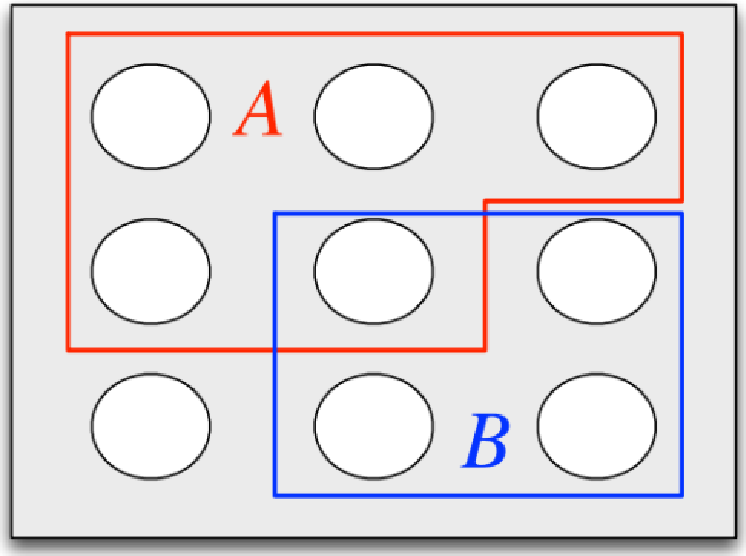
P(win when switch)=2/3

Conditional Probability

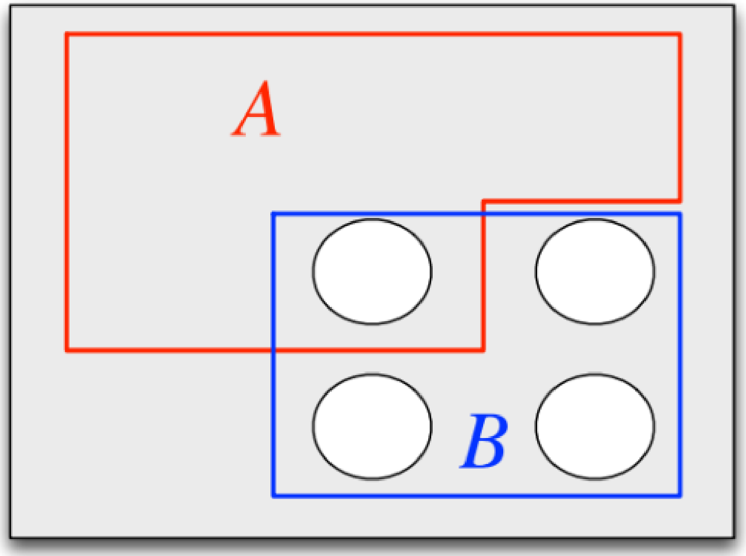
- Given the realization of event A, the probability of event B may change

- (Conditional probability) If A and B are events with P(B)>0, then the conditional probability of A given B, denoted by P(A|B), is defined as $P(A|B) = \frac{P(A \cap B)}{P(B)}$

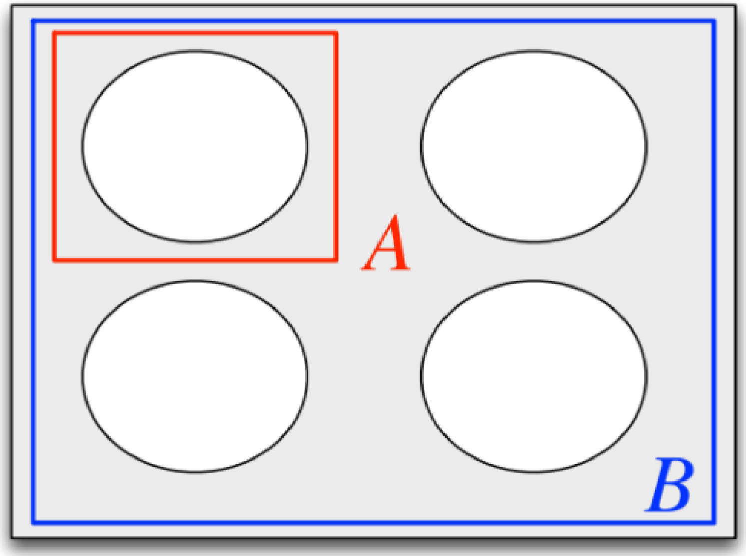
$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$



$P(A|B)=?$



$P(A|B)=?$



$P(A|B)=?$

- Given the realization of event A, the probability of event B may change

- (Conditional probability) If A and B are events with P(B)>0, then the conditional probability of A given B, denoted by P(A|B), is defined as $P(A|B) = \frac{P(A \cap B)}{P(B)}$

- (Independence) Two events A and B are called independent if and only if $P(A \cap B) = \mathrm{P}(A)P(B)$

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

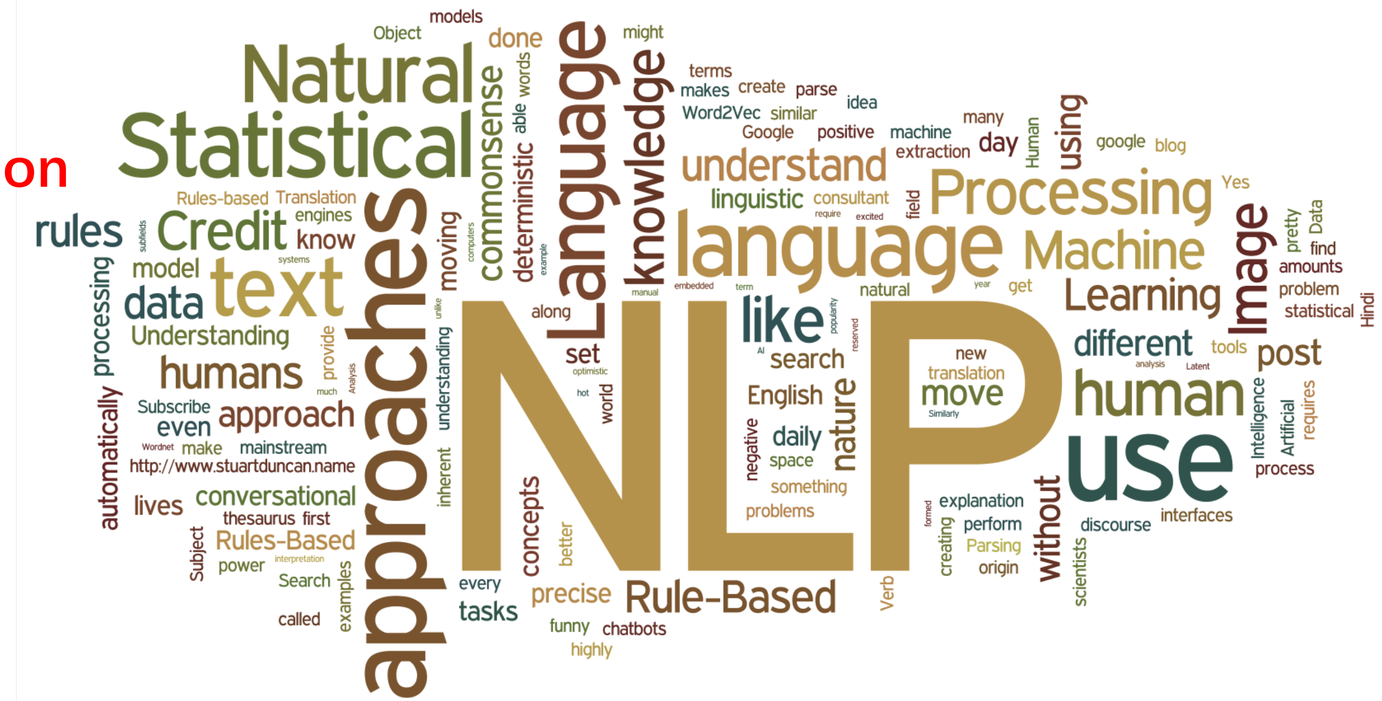
Conditional Probability in NLP

- H: mention “Interesting” in message
- D: mention “DDA2001” in message

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

- P(H) = .1
- P(D) =.1
- P(H ∩ D) = .08

Known Information from data



- P(H|D) = ??

Conditional Probability in NLP

- H: mention “Interesting” in message
- D: mention “DDA2001” in message

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

- P(H) = .1
- P(D) =.1
- P(H ∩ D) = .08

Known Information from data

- P(H|D) = 0.08/0.1 = 0.8