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

DDA2001: Introduction to Data Science

Lecture 7: Continuous Random Variable

Zicheng Wang

Terminologies

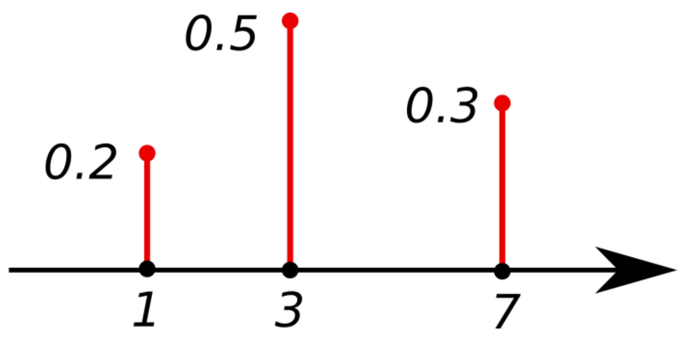
- **Random Experiment**: a repeatable procedure
- **Sample Space**: set of all possible outcomes S.
- **Probability function**, $P(\omega)$: how likely the outcome $\omega \in S$
  - Probability is between 0 and 1
  - Total probability of all possible outcomes is 1.

Sample space

- Discrete or continuous: countable (listable) or not?
- A sample space is **discrete** if it consists of a finite or countable infinite set of outcomes.
- A sample space is **continuous** if it contains an interval (or a union of multiple intervals) of real numbers.

Probability function

- Discrete:
  - ✓ Probability mass function.
  - ✓ P(s): gives the probability for each outcome $\omega \in S$



Recap of Common Discrete RV

1.Bernoulli distribution

- Take value 1 with probability p and value 0 with probability 1 − p.

- $X \sim Bernoulli(p)$

- Mean=p
- Variance=p(1-p)

2.Binomial distribution

- N independent experiments
- Each experiment: success (with probability p) or failure (with probability 1 − p).
- X: the number of success (failure).

- $X \sim Binomial(N, p)$

- Mean=Np
- Variance=Np(1-p)

3.Geometric distribution

- Continuously draw a Bernoulli R.V.
- The X-th sample is the first success.
- X follows a geometric distribution.

- $X\sim Geometric(p)$ or $X\sim Geo(p)$

- Mean=1/p
- Variance=(1-p)/p^2

Continuous Random Variable

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

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

Motivation Example

True or not?

- If $P(E) = 0$, then it is impossible to observe E.

What’s the meaning of impossibility?

- **Random Experiment**: a repeatable procedure
- **Sample space**: set of all possible outcomes S.
- An event is a subset of possible outcomes.

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

Is a zero-probability event impossible?

Example

- We have a probability model.
  - X represents the time a machine needs to complete a task.
  - The values X can take are within [0,1].
  - For any $x, y \in [0,1]$, we have $P(X = x)$= $P(X = y)$
  - E = {0.5}
  - Then what's the probability of $P(E)$?

Example

- We have a probability model.
  - X represents the time a machine needs to complete a task.
  - The values X can take are within [0,1].
  - For any $x, y \in [0,1]$, we have $P(X = x) = P(X = y)$
  - E = {0.5}
  - Then what's the probability of $P(E)$?

Although the impossible event has zero probability,
not all zero-probability events are impossible.

Example

- We have a probability model.
  - X represents the time a machine needs to complete a task.
  - The values X can take are within [0,1].
  - For any $x, y \in [0,1]$, we have $P(X = x)= P(X = y)$
  - E = [0, 0.5]
  - Then what’s the probability of $P(E)$?

General Case

- X is a random variable.
- Define its range as S.
- For any $x, y \in S$, we have $\frac{P(X=x)}{P(X=y)} = \frac{f(x)}{f(y)}$.
    - $\boldsymbol{f(x) > 0}$ **for any** $x \in S$**.**
    - $\boldsymbol{f(x) = 0}$ **for any** $x \notin S$**.**
- E = [a, b] $\subseteq S$
- Then what's the probability of $P(E)$?

General Case

An intuitive interpretation, but not rigorous

- X is a random variable.
- Define its range as S.
- For any $x, y \in S$, we have $\frac{P(X=x)}{P(X=y)} = \frac{f(x)}{f(y)}$.
  - $\boldsymbol{f(x) > 0}$ **for any** $x \in S$.
  - $\boldsymbol{f(x) = 0}$ **for any** $x \notin S$.
- E = [a, b] $\subseteq S$
- Then what's the probability of $P(E)$?

$$\frac{\int_a^b f(x)dx}{\int_S f(x)dx}$$

Continuous RVs

1. How to describe?

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.

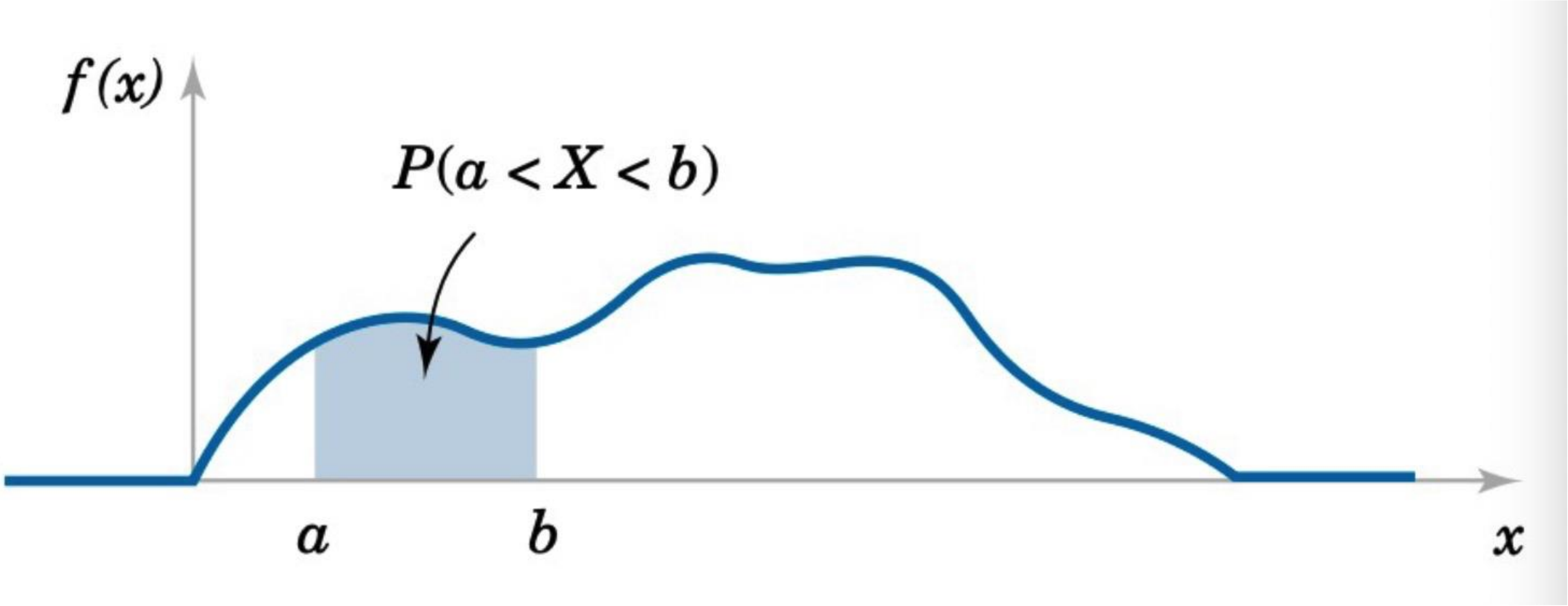
How to describe the probability?

- 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 $X \in [a, b]$

In our course, for continuous RV, we mainly focus on the case where the range is only one interval, not a union of multiple intervals.

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



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.
  - $f(\omega) = 2$, if $\omega \in [0, 0.5]$
  - $f(\omega) = 0$, if $\omega \notin [0, 0.5]$

Continuous RVs

2. CDF, Mean, and Variance

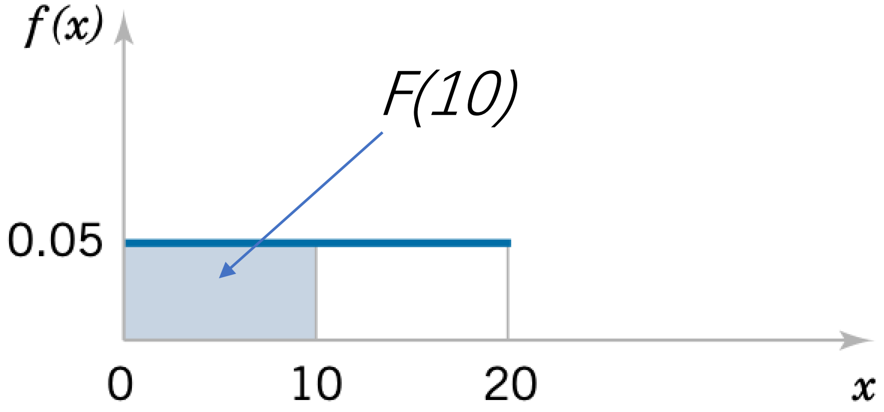
CDF

- 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.

Summation ↔ 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

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

Expectation of g(X)

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

Continuous RVs

3. Common Distributions

- 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?

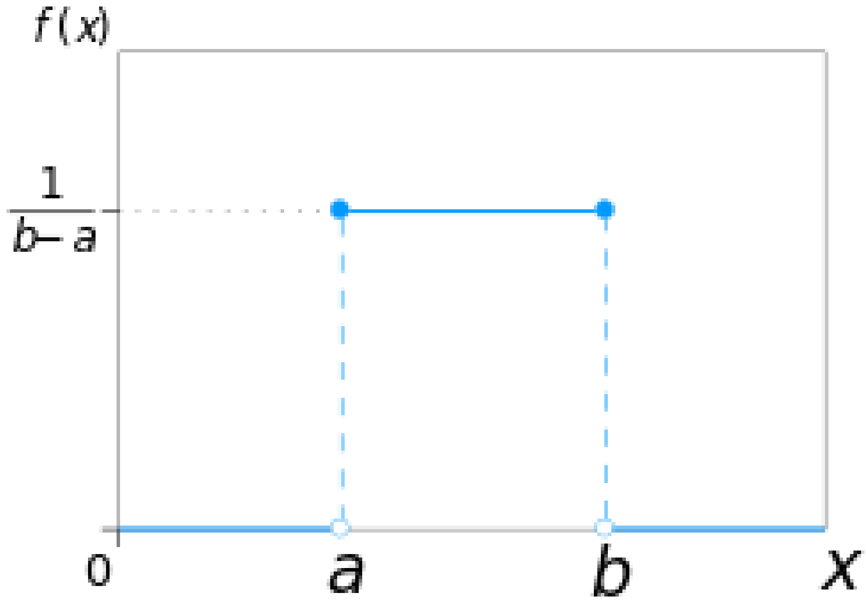
- With the same probability, X takes a value within [a, b], where b>a.

- 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

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

- Draw N samples of X: $X_1, X_2, X_3, \ldots, X_N$
- Calculate $\frac{\Sigma_i g(X_i)}{N}$

General Case

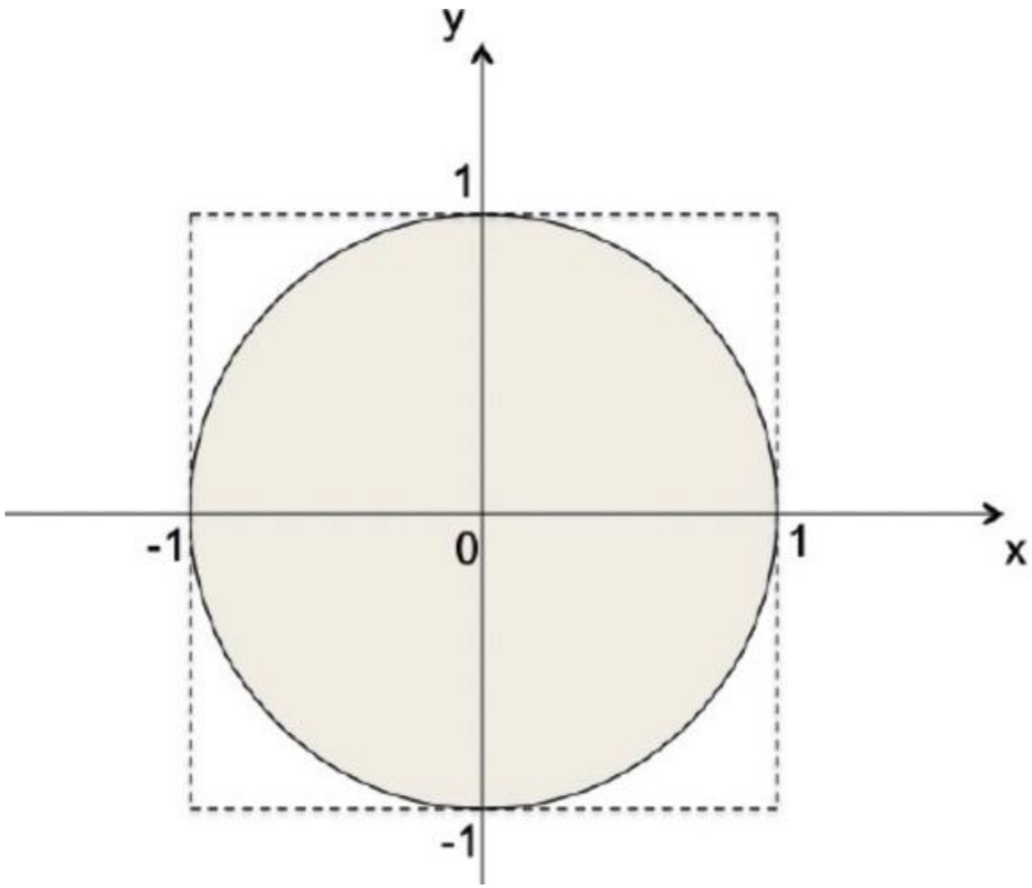
- How to calculate $\int_a^b h(x)dx$?

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

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

Exercise

- Estimation of π



Exercise

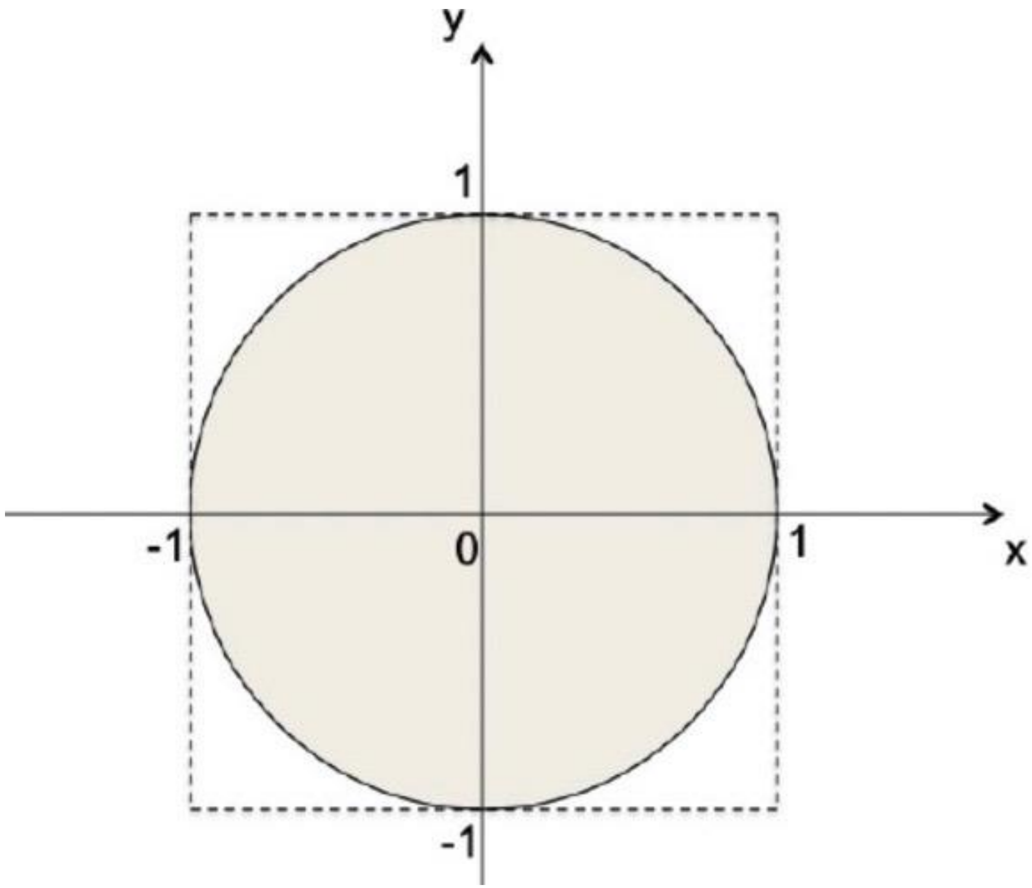
- Estimation of π

Draw a two-dimensional point from the square

$X \sim Uniform(-1,1)$

$Y \sim Uniform(-1,1)$

The sample proportion that $X^2 + Y^2 \leq 1$ is approximately $\pi/4$

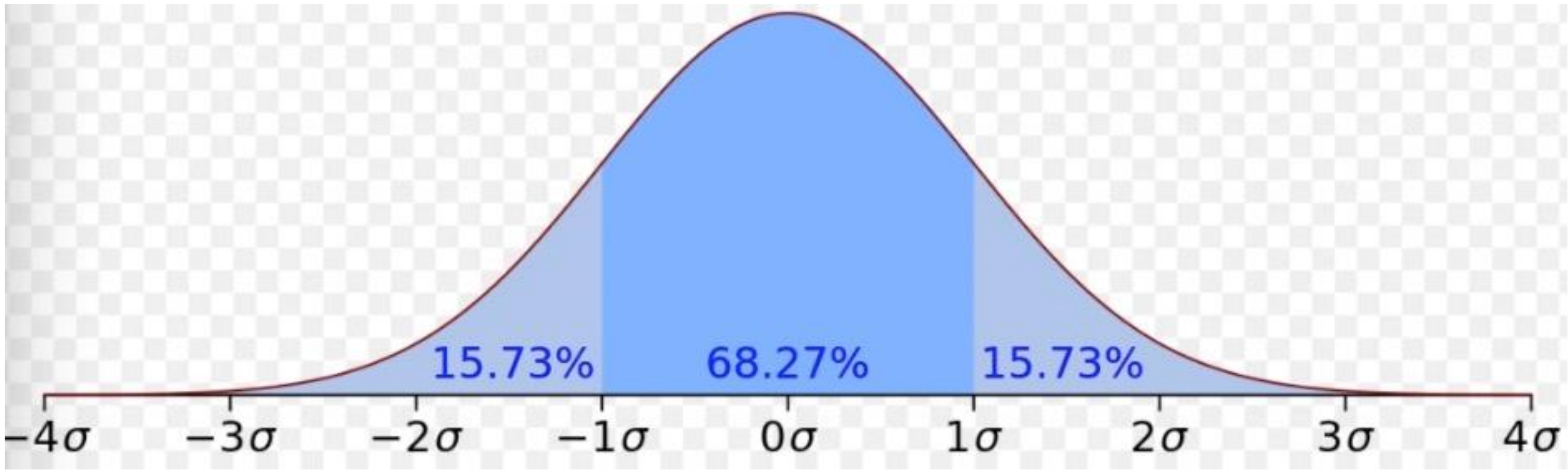


Most important:
Normal distribution

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

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

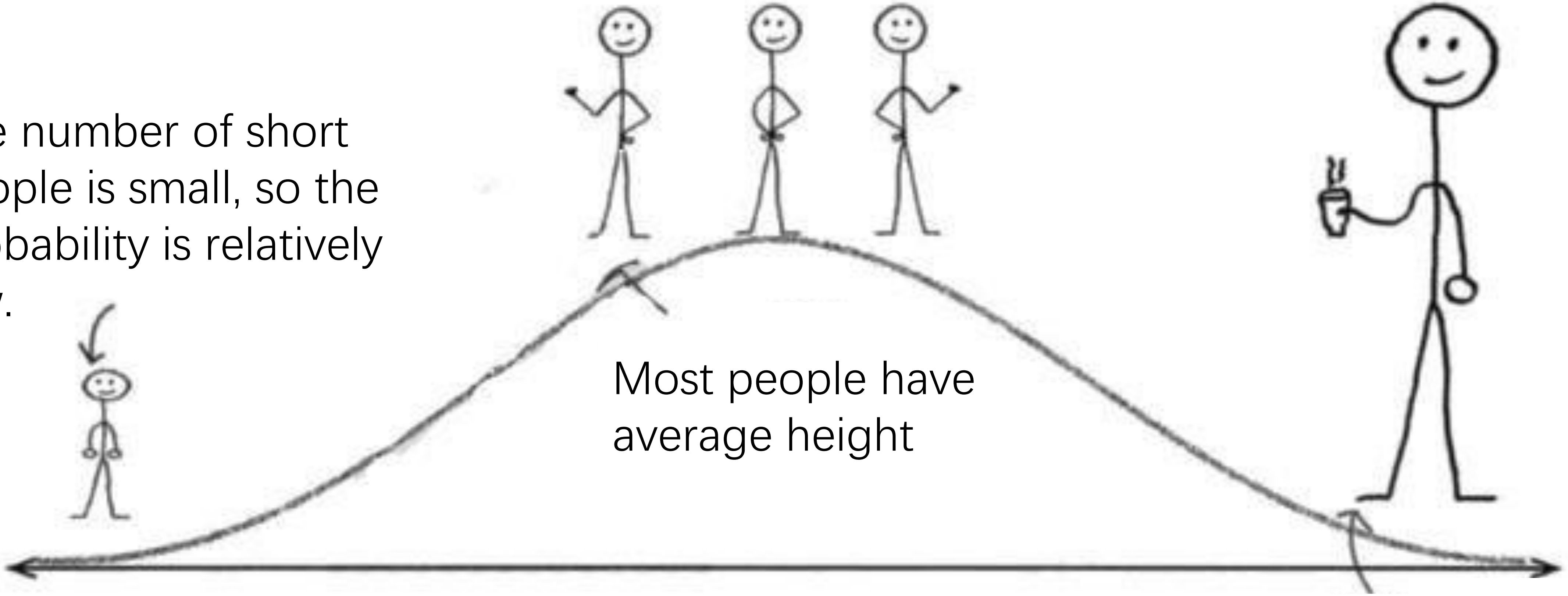
- $X \sim Normal(\mu, \sigma)$



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

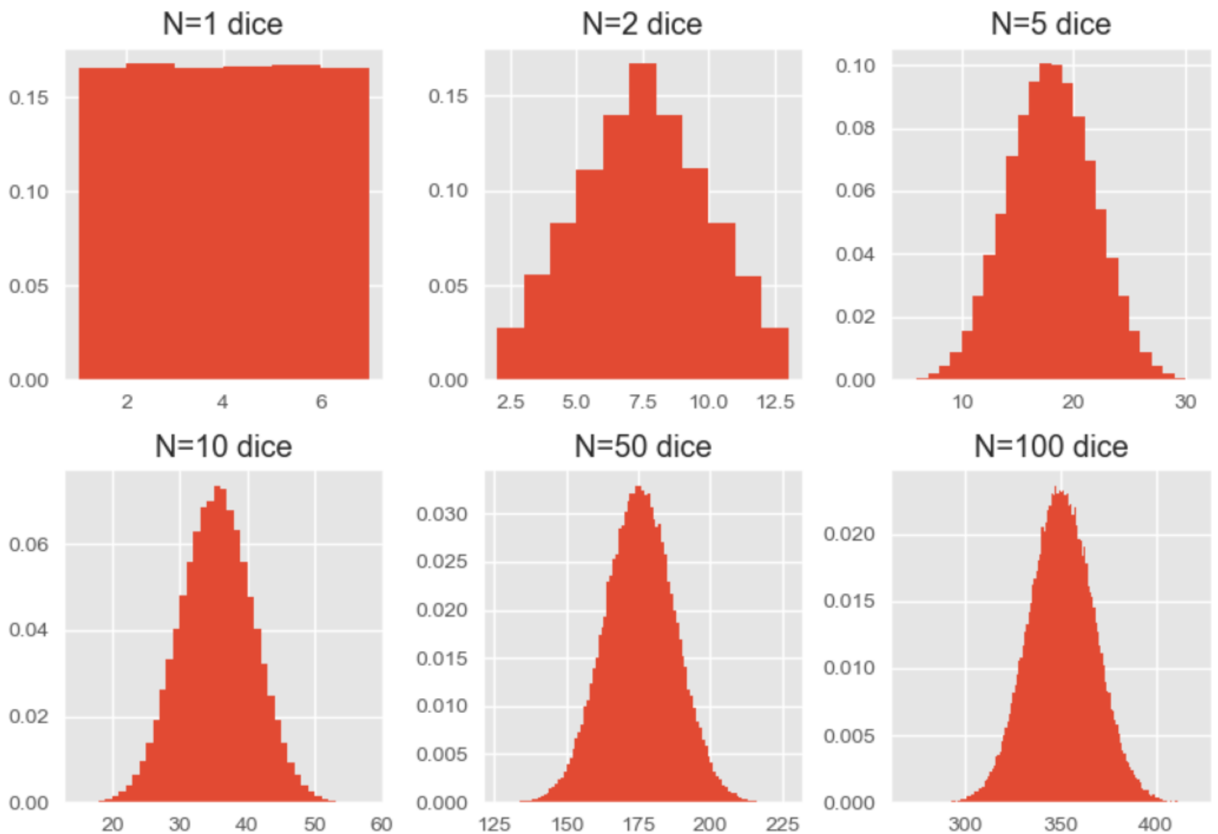
Human Being's Height

Normal Distribution: examples

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

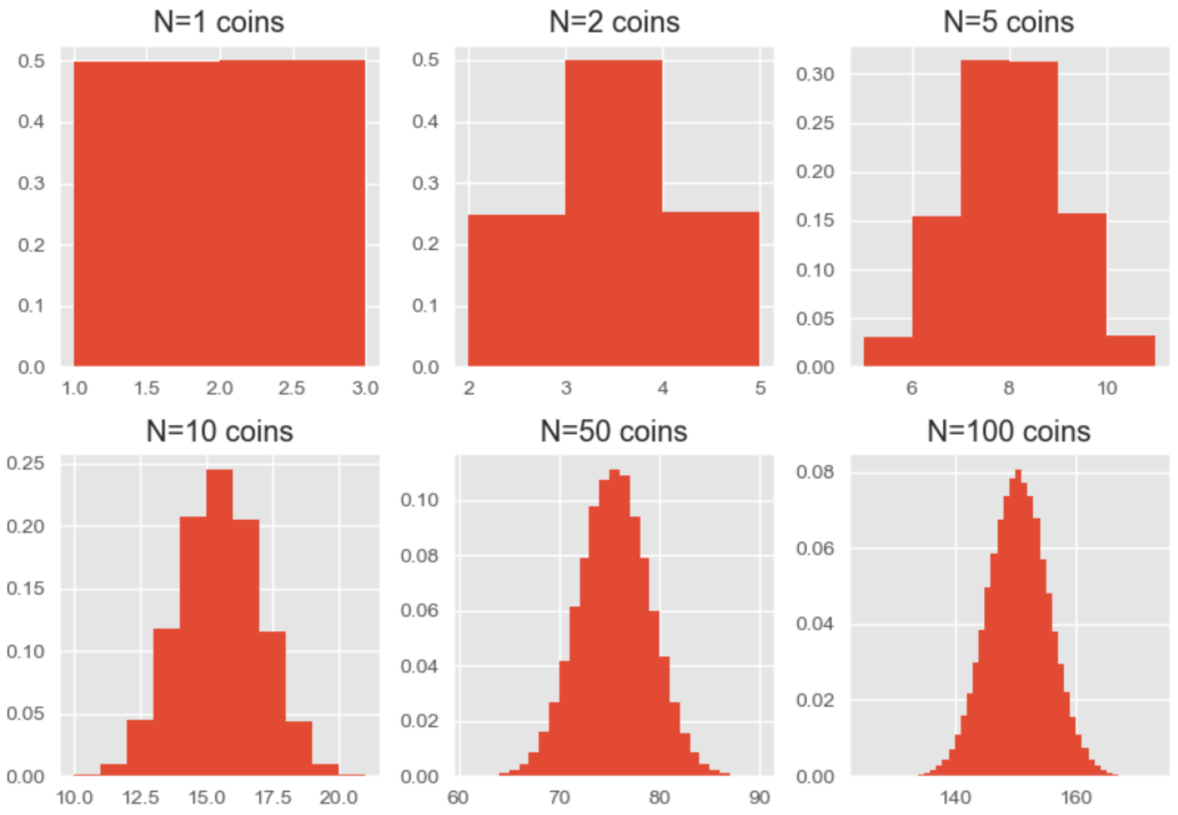
An example

- 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!!!

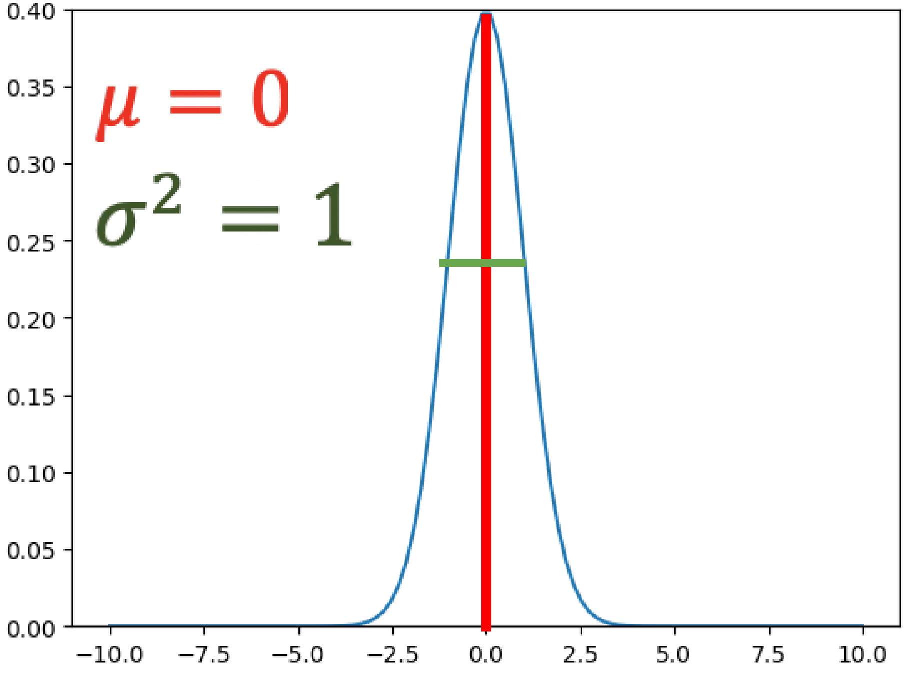
What is the meaning of the parameters?

Mean and Variance

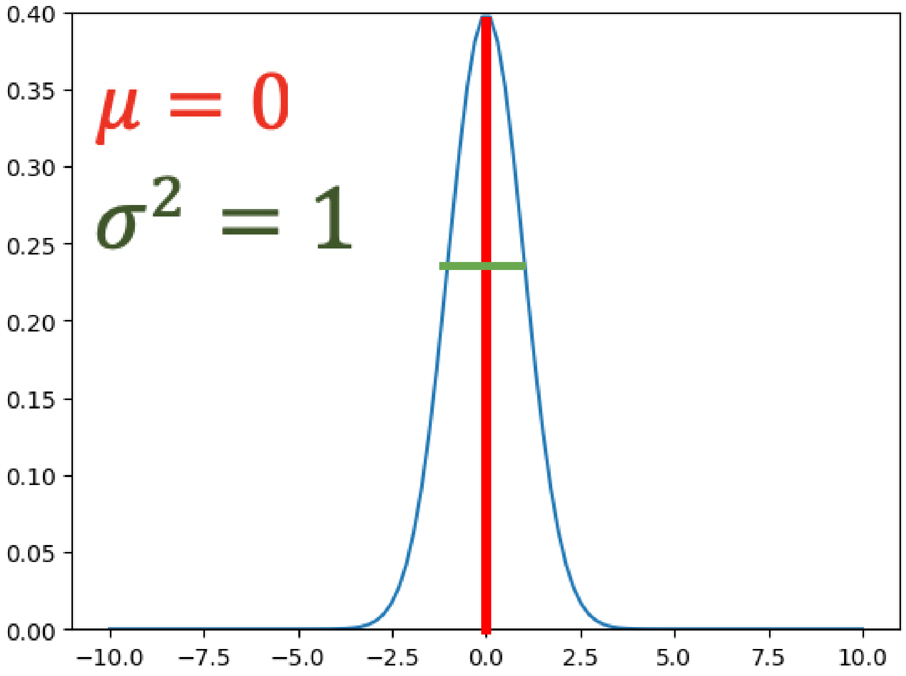
- Mean: μ

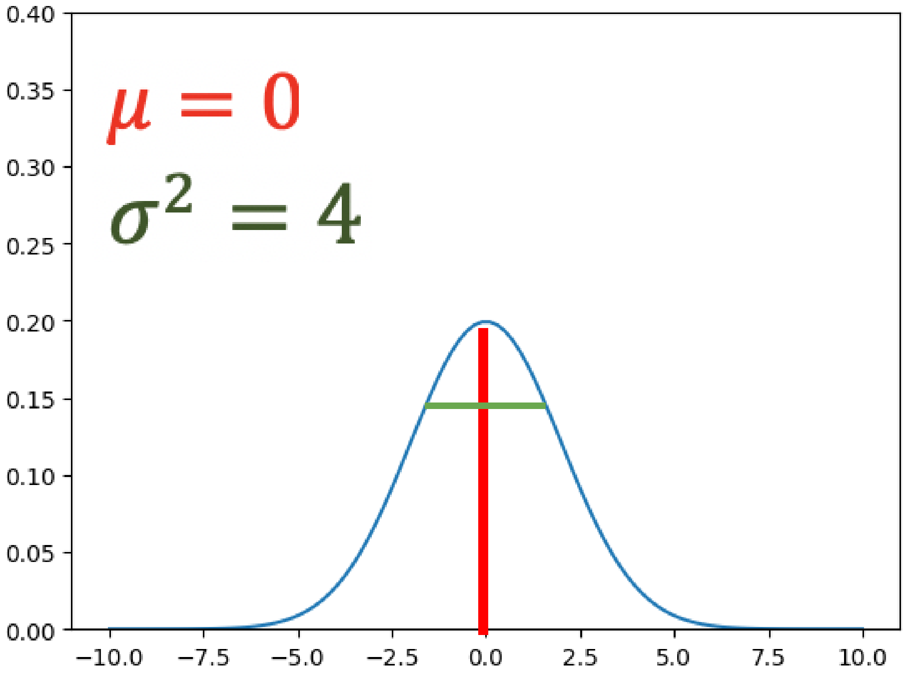
- Variance: σ^2

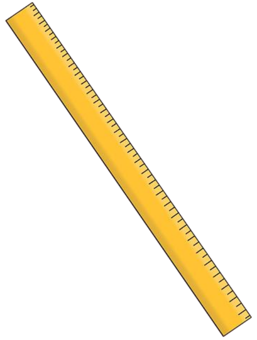
The Normal PDF



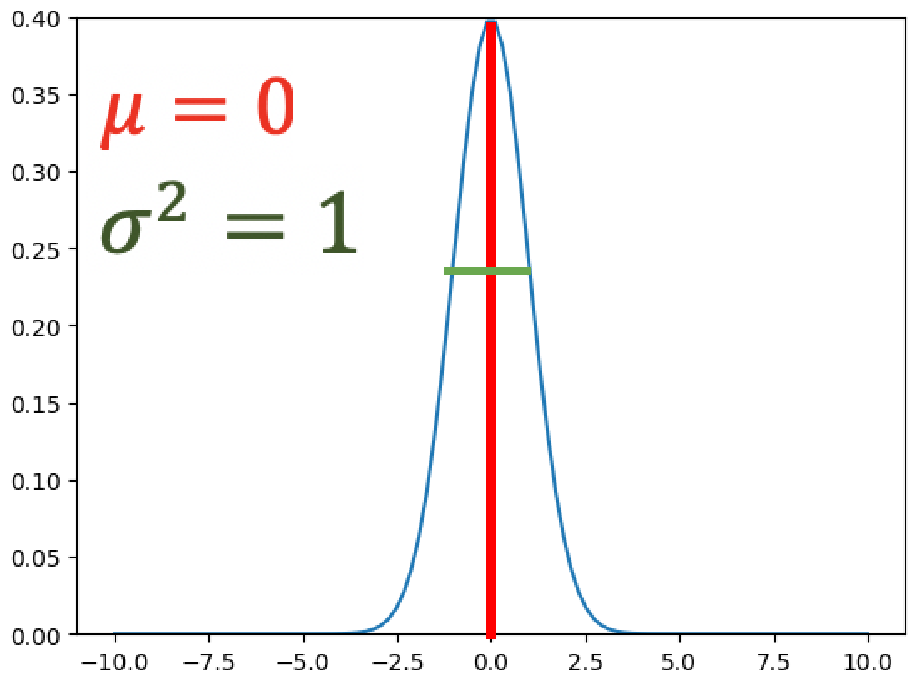
The Normal PDF

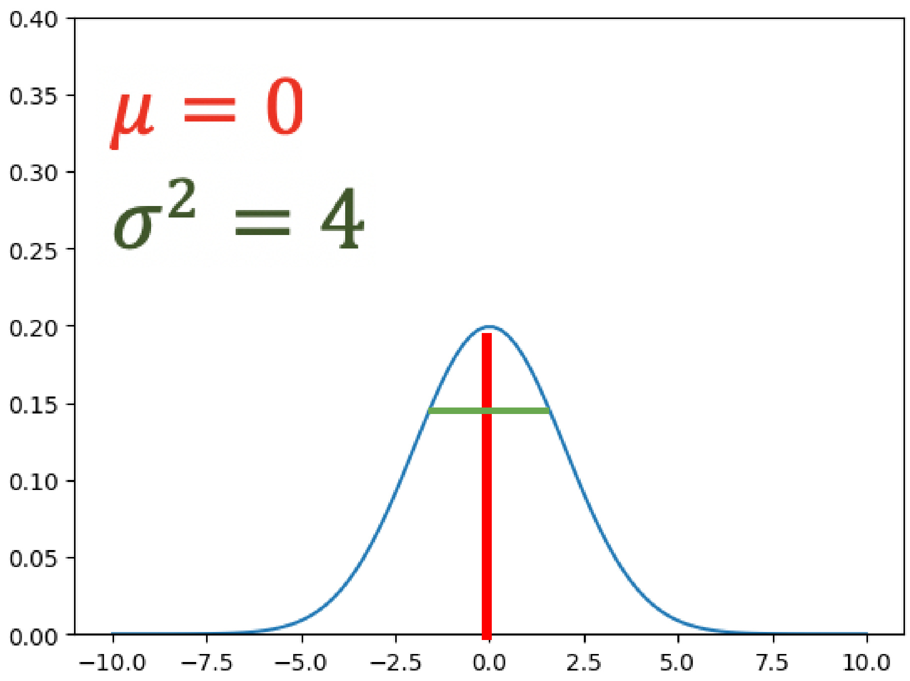


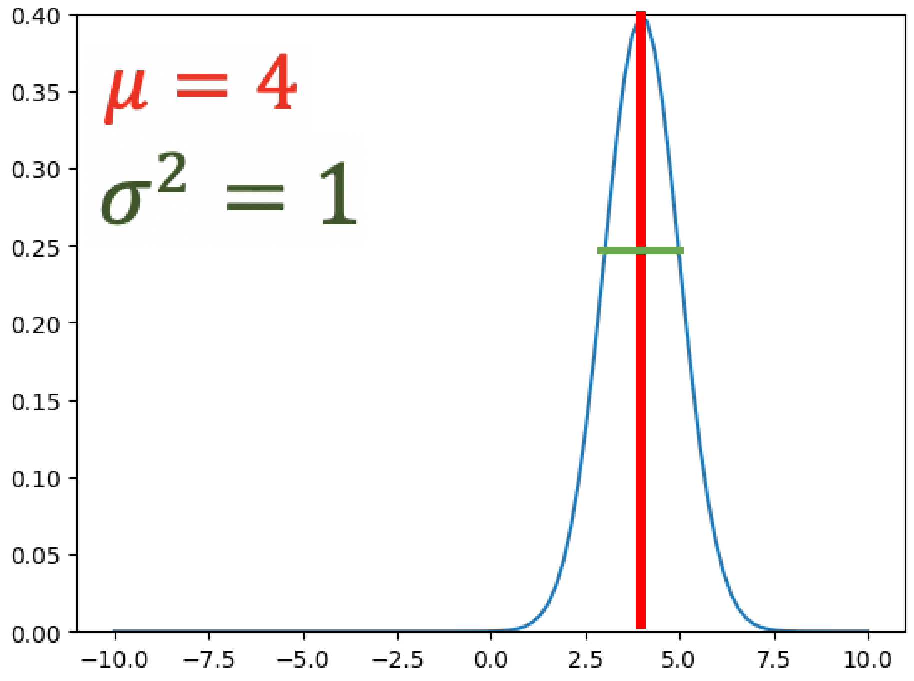


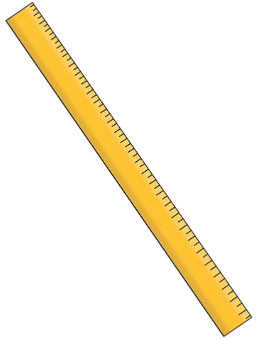


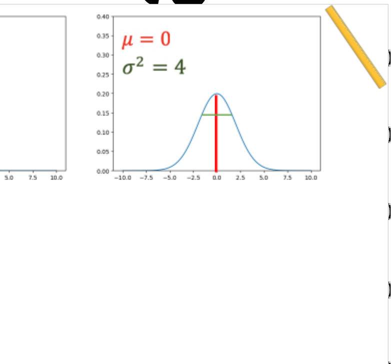
The Normal PDF

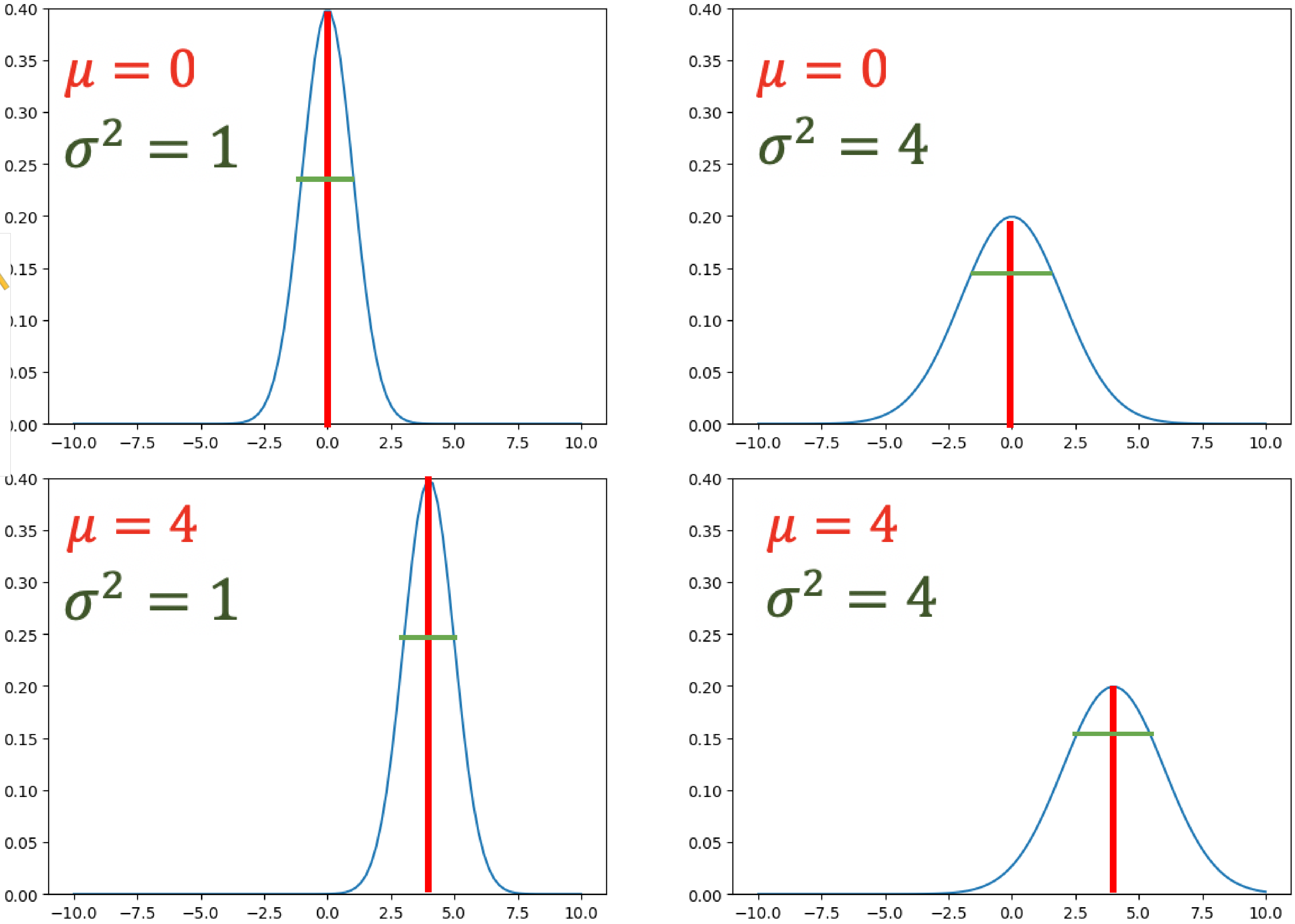









$\mu = 0$
$\sigma^2 = 4$


$\mu = 0$
$\sigma^2 = 1$
$\mu = 0$
$\sigma^2 = 4$
$\mu = 4$
$\sigma^2 = 1$
$\mu = 4$
$\sigma^2 = 4$
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
−10.0
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5
10.0