CIS 678 - Machine Learning

Basics of Probability

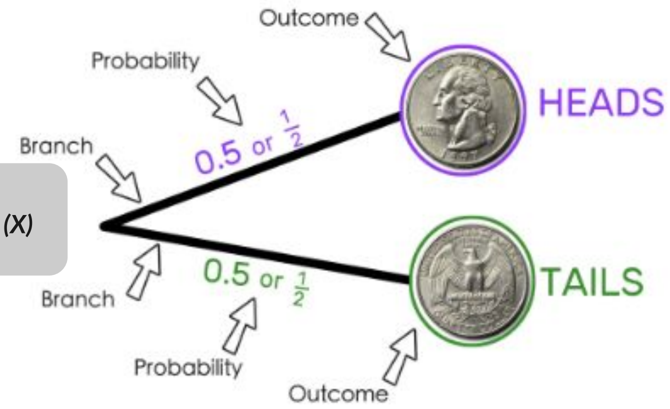
Probability distributions

Distribution: Generally, **a function** showing **all possible values (or intervals)** of the data (**variable**) and how often they occur.

Probability distributions

Distribution: Generally, **a function** showing **all possible values (or intervals)** of the data (**variable**) and how often they occur.

Tossing a fair coin *(x)*



Probability distributions

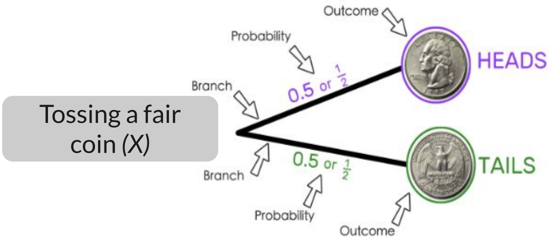
Distribution: Generally, **a function** showing **all possible values (or intervals)** of the data (**variable**) and how often they occur.

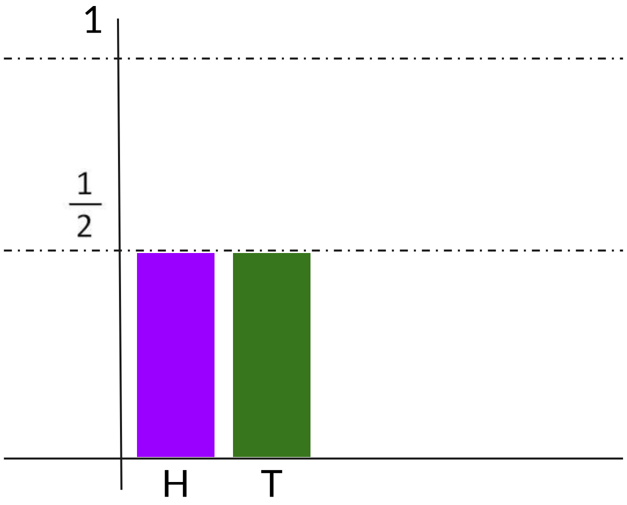
Popular probability distributions

- *Uniform distribution (discrete and continuous)*
- *Binomial distribution (discrete, binary)*
- *Multinomial distribution (discrete, general)*
- *Normal/Gaussian distribution (continuous)*

Discrete Uniform Distribution

Distribution: Generally, **a function** showing **all possible values (or intervals)** of the data and how often they occur.



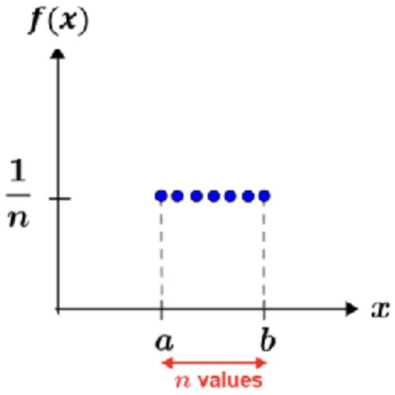


Discrete Uniform Distribution

- Uniform/Equal probability
- ***X*** is a random variable
- ***n*** is the number of different choices ***X*** has

$$f(x) = \frac{1}{n}\ ,\ \mathrm{x} = 1, 2, 3, \ldots, \mathrm{n}$$

Tossing a (**fair**) coin, throwing a (**fair**) dice

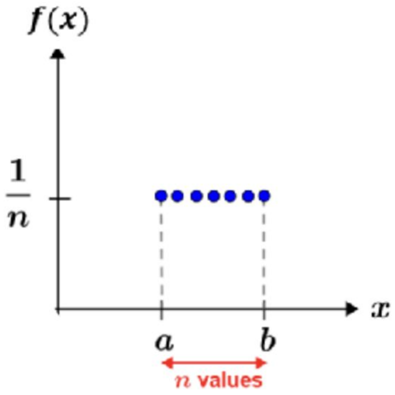


Discrete Uniform Distribution

- Uniform/Equal probability
- ***X*** is a random variable
- ***n*** is the number of different choices ***X*** has

$$f(x) = \frac{1}{n}\ ,\ \mathrm{x} = 1, 2, 3, \ldots, \mathrm{n}$$

Probability mass function

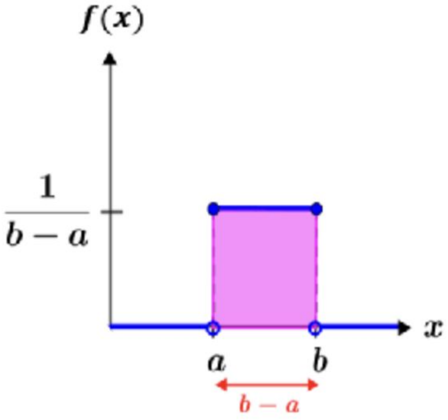


Continuous Uniform Distribution

- Uniform/Equal probability

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b \\ 0, & otherwise \end{cases}$$

Throwing a (**fair**) dirt

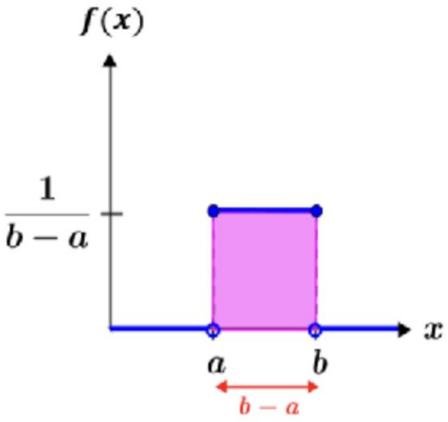


Continuous Uniform Distribution

- Uniform/Equal probability

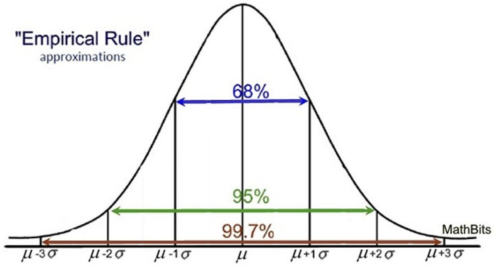
$$f(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b \\ 0, & otherwise \end{cases}$$

Probability density function



Normal (Gaussian) Distribution

- **Definition:** A continuous, symmetric, bell-shaped probability distribution.
- **Applications:** Test scores, heights, errors, finance, etc.



Carl Friedrich Gauss

Portrait by Christian Albrecht Jensen, 1840 (copy from Gottlieb Biermann, 1887)[1]

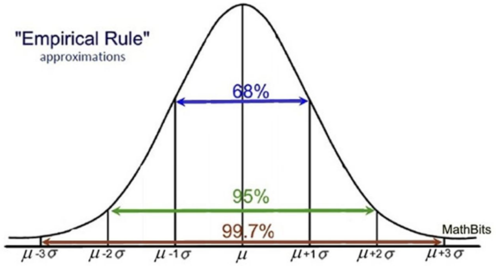
Probability density function	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$
parameters	$\mu =$ mean of x $\sigma =$ standard deviation of x

$\pi \approx 3.14159 \ldots$

$e \approx 2.71828 \ldots$

Normal (Gaussian) Distribution

- **Definition:** A continuous, symmetric, bell-shaped probability distribution.
- **Applications:** Test scores, heights, errors, finance, etc.
- **Parameters:**
  - Mean (μ): center of the distribution
  - Standard deviation (σ): spread of the data
- **Empirical Rule:**
  - 68% within $\mu \pm 1\sigma$
  - 95% within $\mu \pm 2\sigma$
  - 99.7% within $\mu \pm 3\sigma$



Probability density function	$f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$
parameters	$\mu =$ mean of x $\sigma =$ standard deviation of x

$\pi \approx$ 3.14159 ...
$e \approx$ 2.71828 ...

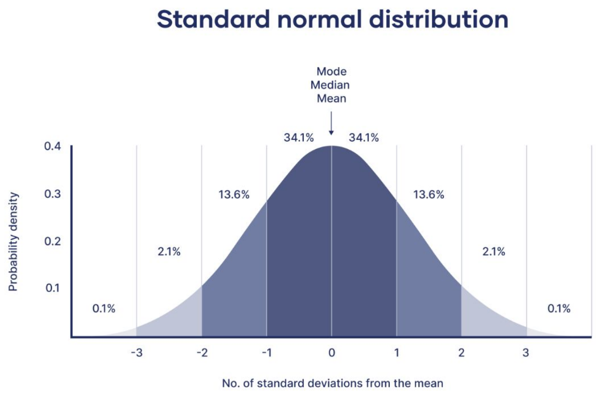
Standard-normal/Z-distribution

If a data set has a normal distribution, and you standardize all the data to obtain standard scores, those standard scores are called *z*-values. All *z*-values have what is known as a standard normal distribution (or *Z*-distribution). The *standard normal distribution* is a special normal distribution with a mean equal to 0 and a standard deviation equal to 1.

$$z = \frac{x - \mu}{\sigma}$$

$\mu =$ Mean

$\sigma =$ Standard Deviation

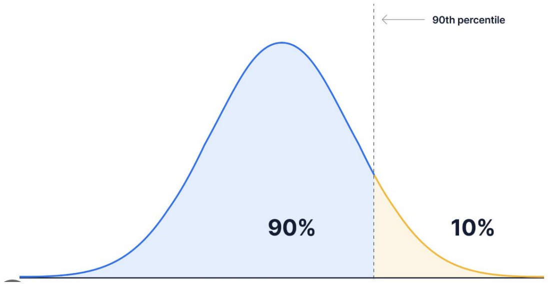


ref link

Percentile

Percentile: The **percentile** reported for a given score is the percentage of values in the data set that fall below that certain score. For example,

- If **your score** was reported to be at the **90th percentile**, that means 90% of the other people who took the test scored lower than you did.



QA