Johns Hopkins
Whiting School of Engineering

Algorithms for Data Science

Probabilities and Distributions

Module Learning Objectives

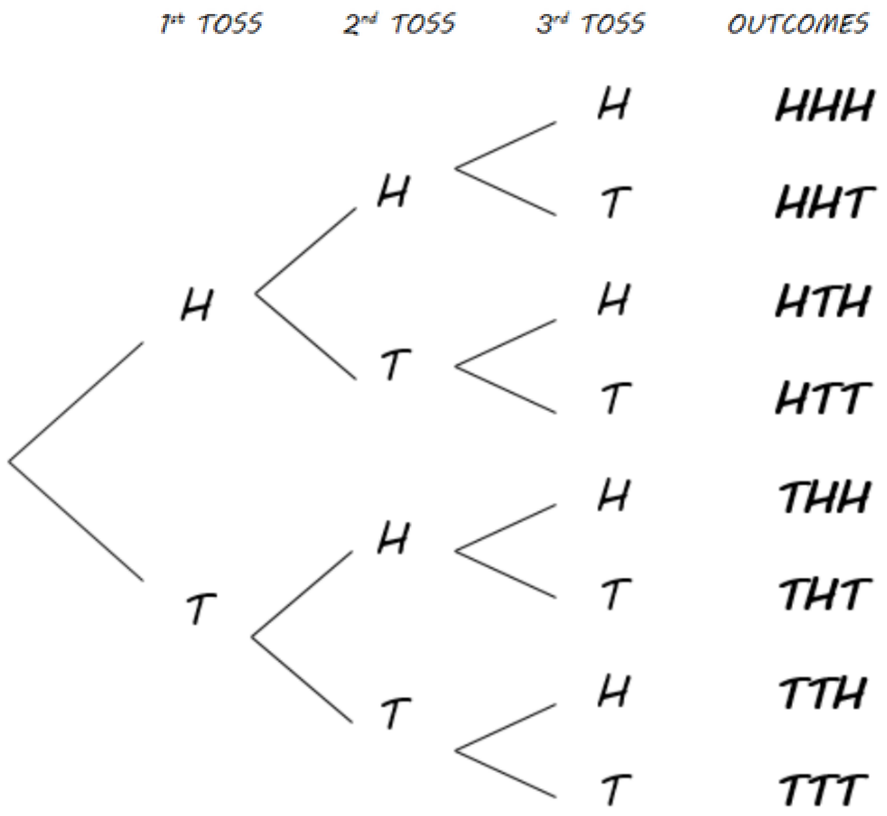
- Recognize the probability types, axioms, and laws such as joint probability, marginal probability, conditional probability, non-negativity axiom, normalization axiom, and additivity axiom, CLT, LLN, including their uses and examples.
- Describe multiple probability distributions such as hypergeometric, geometric, negative binomial, binomial, Bernoulli, Poisson, uniform, normal, exponential, gamma, and beta.
- Compute the moments of distributions like mean, standard deviation, variance, degrees of freedom, skewness, and kurtosis
- Assess the appropriateness of randomization in algorithm design.
- Utilize randomization in the design of an algorithm.
- Apply the Markov principle to analyze a randomized algorithm.
- Assess the empirical performance of an algorithm with and without randomization incorporated.

JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

Probability Review / Definitions

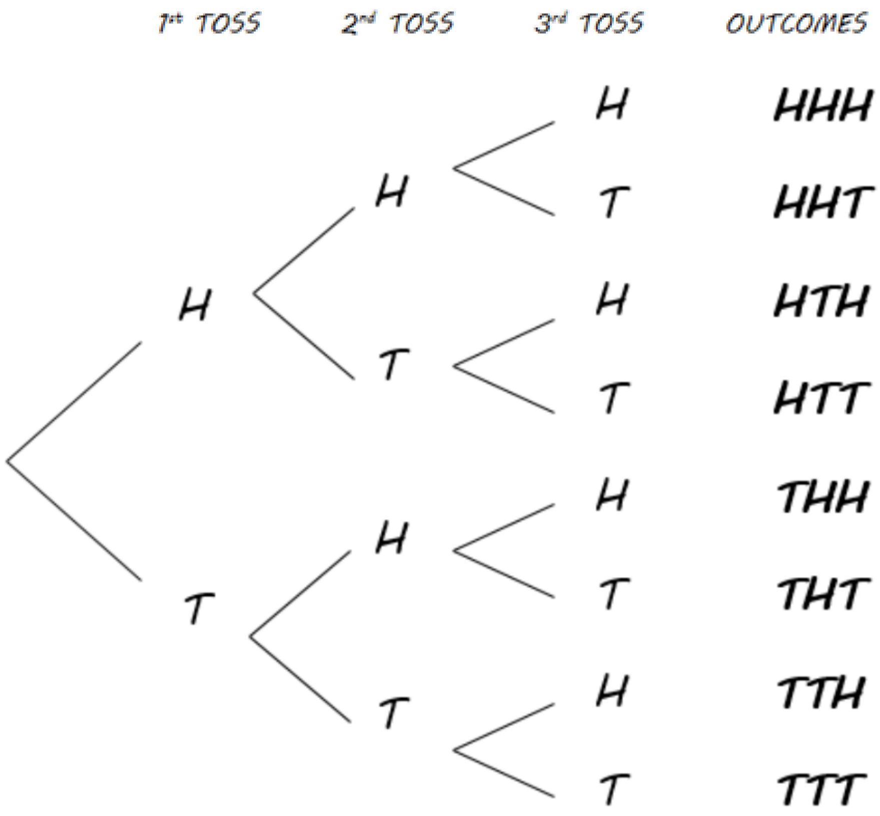
Sample Space

- A **sample space** is a set **S** whose elements are called elementary events. The sample space **S** is the set of all possible outcomes of a random experiment.
- Sometimes a tree diagram is helpful in determining the sample space.
- Three fair coins are tossed. The sample space is shown here.



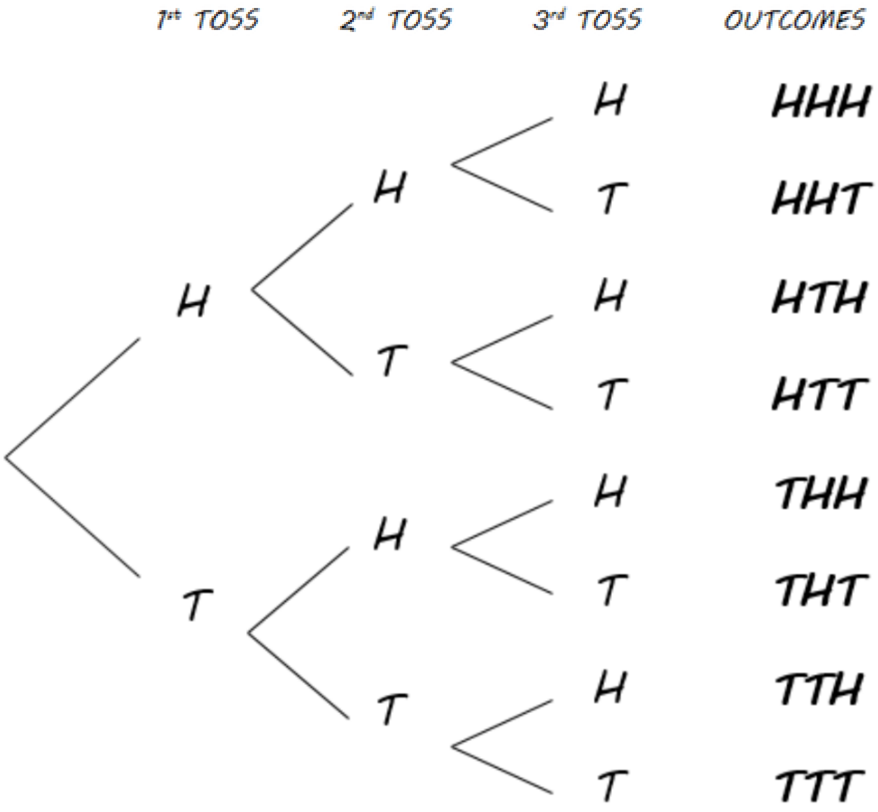
Outcome

- An **outcome** is a single result from the sample space of a random experiment.
- In this case HHH is an outcome.



Event

- An **event** is a subset of the sample space.
- It is a set of outcomes to which a probability is assigned.
- An event could be that the first toss is a **H**. The event **A** of having a **H** on the first toss could be expressed as **A** = {HHH, HHT, HTH, HTT}
- Two events **A** and **B**, are mutually exclusive if and only if **A** intersect **B** is an empty set.



Three Probability Axioms

- The probability of any event **A** is a non-negative number

$$P(A) \geq 0$$

Non-Negativity

- The probability of the sample space **S** is equal to 1.

$$P(S) = 1$$

Normalization

- For any two mutually exclusive events **A** and **B**, the probability of their union is the sum of their individual probabilities

$$P(A \cup B) = P(A) + P(B)$$

$$\text{if } A \cap B = \emptyset$$

Additivity

Expected Value

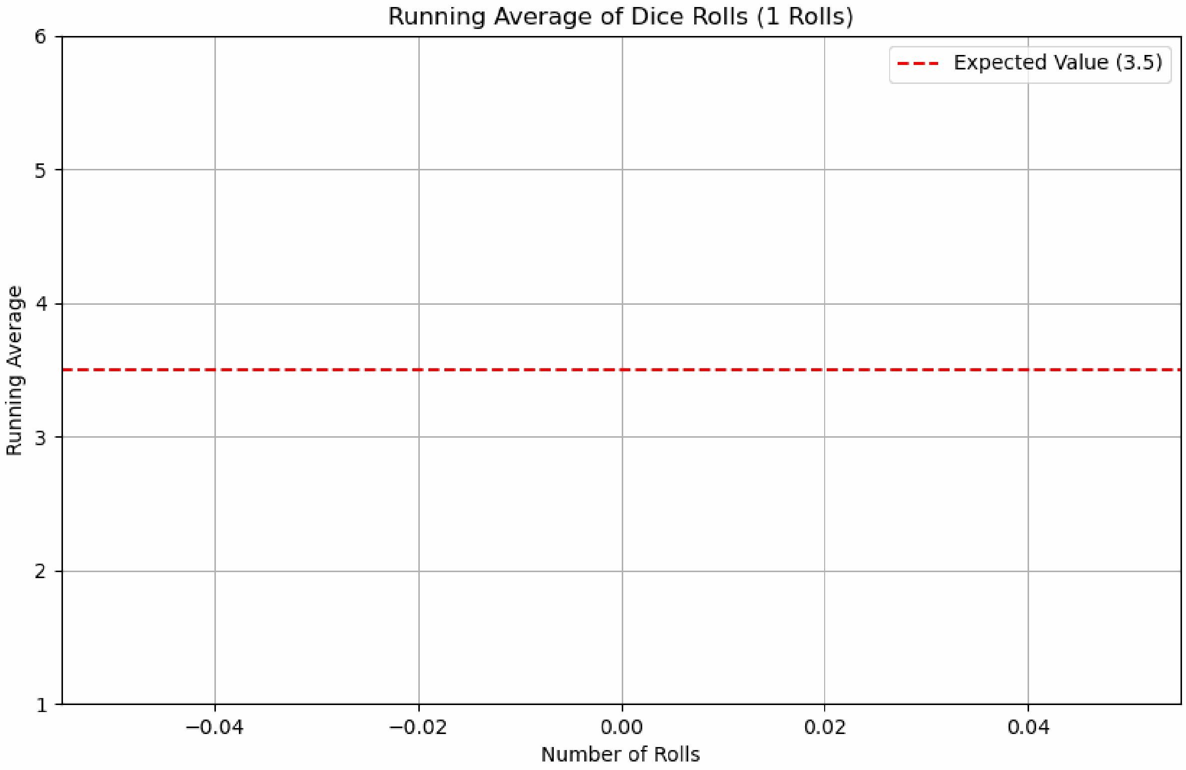
- **Expected Value,** often referred to as the mean, is a fundamental concept in probability and statistics that provides a measure of the central tendency of a random variable.

- **Random Variable**, mapping outcomes of a random process to numbers.
- **Discrete Random Variable**, the possible outcomes are countable or finite.
- **Continuous Random Variable**, an uncountable number of outcomes represented by an interval on a number line.

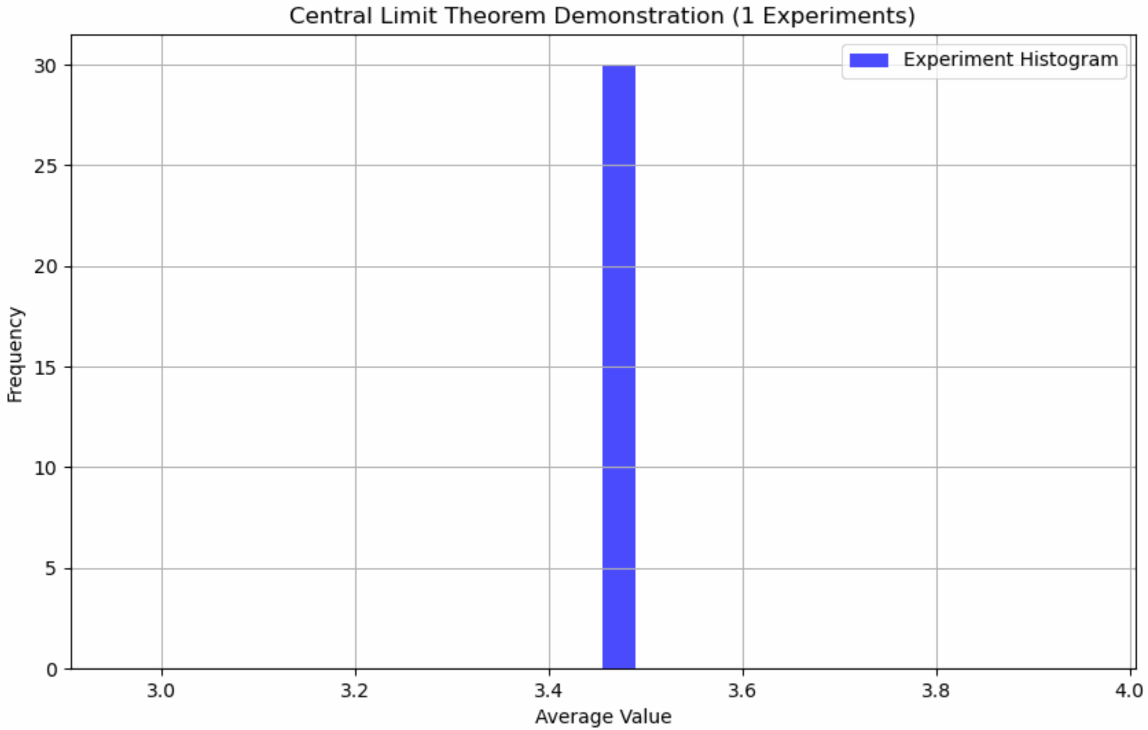
$$E(X) = \sum_{i=1}^{n} x_i P(X = x_i)$$

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

Law of Large Numbers



Central Limit Theorem



Variance

- **Variance**, measures the dispersion or spread of a random variable around its mean (expected value).

- Just like with Expected Value there are different formulations for discrete vs continuous calculations for variance

Discrete Random Variable Variance

$$\mathrm{Var}(X) = E[(X-\mu)^2] = \sum_{i=1}^{n} (x_i - \mu)^2 P(X = x_i)$$

Continuous Random Variable Variance

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

Standard Deviation

- **Standard Deviation**, quantifies the amount of variation or dispersion in a set of data values.

- Just like with Variance there are different formulations for discrete vs continuous calculations for standard deviation

Discrete Random Variable Standard Deviation

$$\mathrm{Var}(X) = E[(X-\mu)^2] = \sum_{i=1}^{n}(x_i-\mu)^2 P(X=x_i)$$

Continuous Random Variable Standard Deviation

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

Skewness

- **Skewness**, describes the asymmetry of the probability distribution

$$\frac{\frac{1}{N}\sum_{n=1}^{N}(\mathbf{x}_n-\mu)^3}{\sigma^3}$$

- The distribution of X is said to be positively skewed, negatively skewed or unskewed depending on whether skew(X) is positive, negative, or 0.
- This is the 3rd Moment

$$\text{skew}(X)=\frac{\mathbb{E}\left(X^3\right)-3\mu\mathbb{E}\left(X^2\right)+2\mu^3}{\sigma^3}=\frac{\mathbb{E}\left(X^3\right)-3\mu\sigma^2-\mu^3}{\sigma^3}$$

Kurtosis

- **Kurtosis**, a measure of the "tailedness" of a distribution, which shows how often outliers occur.

$$= \frac{\frac{1}{N}\sum_{n=1}^{N}(\mathbf{x}_n - \mu)^4}{\sigma^4}$$

- This is the 4th moment

$$\operatorname{kurt}(X) = \frac{\mathbb{E}\left(X^4\right) - 4\mu\mathbb{E}\left(X^3\right) + 6\mu^2\mathbb{E}\left(X^2\right) - 3\mu^4}{\sigma^4} = \frac{\mathbb{E}\left(X^4\right) - 4\mu\mathbb{E}\left(X^3\right) + 6\mu^2\sigma^2 + 3\mu^4}{\sigma^4}$$

JOHNS HOPKINS

WHITING SCHOOL of ENGINEERING