Probability.

Consider an experiment (e.g., flipping a coin) whose outcome is determined by chance.

- Random variable = a variable whose value is the outcome of the experiment.
- The set of possible outcomes = sample space.

We can assign a probability of occurrence to each outcome.

E.g. 1) $\mathbb{P}(X = \text{Head}) = \frac{1}{2}$.

and $\mathbb{P}(X = \text{Tail}) = \frac{1}{2}$.

2) Let Y be the sum of two dice rolls

$\Rightarrow$ Possible values: $\{2, 3, 4, \dots, 12\}$.

Their probabilities:

$$\mathbb{P}(Y = 2) = 1/36.$$
$$\mathbb{P}(Y = 3) = 2/36$$
$$\mathbb{P}(Y = 4) = 3/36.$$

etc.

The probabilities assigned to the possible values of a random variable are its distribution.

Note: A r.v. is discrete if it has countably many possible values; otherwise, it is called continuous.

- Discrete distributions:

probability mass function. $\leftarrow$ $\mathbb{P}(X = x)$ = prob. that the value of X is x

E.g. If X is the outcome of the roll of a die,

$$\mathbb{P}(X=1) = \mathbb{P}(X=2) = \dots = \mathbb{P}(X=6) = 1/6.$$

(and $\mathbb{P}(X=x) = 0$ for all other values of x.)

- Continuous distributions: The distribution of a continuous random variable cannot be specified through a probability mass function because if X is continuous, then

$$\mathbb{P}(X = x) = 0 \quad \text{for all } x.$$

$\Rightarrow$ we must look at probabilities of ranges of values.

$$\mathbb{P}(a < X < b) = \int_a^b \underbrace{p(x)}_{\text{density function}} dx.$$

* Expectations of random variables:

Expected value (= Mean) of a r.v. X is denoted by $\mathbb{E}[X]$.

. For discrete r.v. X:

$$\mathbb{E}[X] = \sum_x x\, \mathbb{P}(X = x).$$

E.g. The expected value of the roll of a die:

$$\mathbb{E}[X] = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \dots + 6 \cdot \frac{1}{6} = \frac{21}{6}.$$

. For continuous r.v. X:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x\, p(x)\, dx$$

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

$N(\mu, \sigma^2)$

$X \sim N(0, 1)$ standard normal distribution.

$$\mathbb{E}[X] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-x^2/2}\, dx = 0$$

If g is a function,

$$\mathbb{E}[g(X)] = \sum_x g(x)\, \mathbb{P}(X = x).$$

E.g.

$$\mathrm{Var}[X] = \sigma_X^2 = \text{variance of a r.v. } X.$$
$$= \mathbb{E}\left[(X - \mathbb{E}[X])^2\right].$$

* Dependence and Independence:

In many situation, we may have several r.v.'s.
E.g. Think of the price of each stock in the New York exchange as a r.v. The movements of these variables are related.

X and Y are independent r.v.'s if

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\, \mathbb{P}(Y = y).$$

Event = a subset of the sample space.

- Union bound:

Let $A_1, A_2, \ldots, A_n$ be events.

$$\mathbb{P}(A_1 \cup A_2 \cup \ldots \cup A_n) \le \sum_{i=1}^{n} \mathbb{P}(A_i).$$

Properties of expected values:

- $\mathbb{E}[\alpha X] = \alpha \mathbb{E}[X]$ for any constant α.
- $\mathbb{E}_{X,Y}[X+Y] = \mathbb{E}_X[X] + \mathbb{E}_Y[Y]$.
- For any two <u>independent</u> r.v.'s X, Y:

$$\mathbb{E}_{X,Y}[XY] = \mathbb{E}_X[X]\,\mathbb{E}_Y[Y].$$

and

$$\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$$

* <u>The central limit Theorem:</u>

<u>E.g.</u> Consider n independent r.v.'s

$$X_i = \begin{cases} 0 & \text{with prob. } 1/2 \\ 1 & \text{with prob. } 1/2. \end{cases}$$

Let $S = X_1 + X_2 + \dots + X_n$.

$$\mathbb{E}[X_i] = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{1}{2}.$$

$$\sigma_i^2 = \left(\frac{1}{2} - 0\right)^2 \cdot \frac{1}{2} + \left(\frac{1}{2} - 1\right)^2 \cdot \frac{1}{2} = \frac{1}{4}.$$

Then, $\mathbb{E}[S] = \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \frac{n}{2}.$

$$\mathrm{Var}[S] = \frac{n}{4}$$

How concentrated S is around its mean depends on the standard deviation of S which is $\frac{\sqrt{n}}{2}$.

For $n = 100 \Rightarrow \mathbb{E}[S] = 50$ and $\sigma_S = 5 = 0.1\,\mathbb{E}[S]$.

$n = 10000 \Rightarrow \mathbb{E}[S] = 5000$ and $\sigma_S = 50 = 0.01\,\mathbb{E}[S]$.

$\Rightarrow$ as n increases, σ_S increases

but $\frac{\sigma_S}{\mathbb{E}[S]} \to 0$.

Thm: (CLT) Suppose $X_1, X_2, \ldots, X_n$ is a sequence of identically distributed independent (i.i.d.) variables, each with mean μ and variance σ^2. The distribution of the random variable

$$\frac{1}{\sqrt{n}}\left(X_1 + X_2 + \ldots + X_n - n\mu\right)$$

converges to the distribution of the Gaussian with mean 0 and variance σ^2.

- Gaussian or normal distribution:

$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-1/2 \frac{(x-\mu)^2}{2\sigma^2}}.$$

* Important inequalities:

- Used for bounding the tail distribution, the probability that a r.v. assumes values that are far from its expectation
- In the context of analysis of algorithms; these inequalities are the major tool for estimating the failure probability of algorithms and for establishing high prob. bounds on their run-time.

Thm: (Markov's inequality) Let X be a r.v. and $X \geq 0$. Then, for all $a > 0$;

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}.$$

pf: For $a > 0$, let

$$I = \begin{cases} 1 & \text{if } X \geq a \\ 0 & \text{otherwise} \end{cases} \quad (\text{indicator r.v.})$$

since $X \geq 0$,

$$I \leq \frac{X}{a}.$$

And

$$\mathbb{E}[I] = \mathbb{P}(I=1) = \mathbb{P}(X \geq a)$$

$$\Rightarrow \quad \mathbb{P}(X \geq a) = \mathbb{E}[I] \leq \mathbb{E}\left[\frac{X}{a}\right] = \frac{\mathbb{E}[X]}{a}. \quad \blacksquare$$

E.g. Suppose we want to bound the prob. of obtaining more than $3n/4$ heads in a sequence of n coin flips.

Let $X_i = \begin{cases} 1 & \text{if the ith coin flip is head,} \\ 0 & \text{otherwise} \end{cases}$

and let $X = \sum_{i=1}^{n} X_i$ = the number of heads in the n coin flips.

Since $\mathbb{E}[X_i] = \mathbb{P}(X_i = 1) = \frac{1}{2}$,

$$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \frac{n}{2}.$$

$\Rightarrow$ Applying Markov's inequality,

$$\mathbb{P}\left(X \geq \frac{3n}{4}\right) \leq \frac{\mathbb{E}[X]}{\frac{3n}{4}} = \frac{n/2}{3n/4} = \frac{2}{3}.$$

Note: Even though Markov's inequality is simple, it's "weak".

Thm: (Chebyshev's inequality). For any $a > 0$,

$$\mathbb{P}[\,|X - \mathbb{E}[X]| \geq a] \leq \frac{Var[X]}{a^2}$$

Pf: $\mathbb{P}(\,|X - \mathbb{E}[X]| \geq a) = \mathbb{P}((X - \mathbb{E}[X])^2 \geq a^2)$

$$\underset{\text{Markov}}{\leq} \frac{\mathbb{E}[(X - \mathbb{E}[X])^2]}{a^2} = \frac{Var[X]}{a^2}.$$

(# of heads in n tosses of a fair coin)

E.g. The probability that X deviates from $\mu = \mathbb{E}[X] = \frac{n}{2}$ by more than $\sqrt{n}$ is at most $\frac{1}{4}$.

The probability that it deviates by more than $5\sqrt{n}$ is at most $\frac{1}{100}$.