5 Discrete Probability Distributions

5.1 Random Variables

A **variable** may be **categorical** (qualitative) or **numerical** (quantitative).

Definition: A **random variable** assigns a **number** to each outcome in a population.

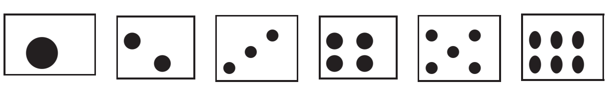
Two types of **random variables** are **discrete** and **continuous**.

What are some **discrete** random variables?

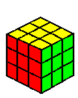
What are some **continuous** random variables?

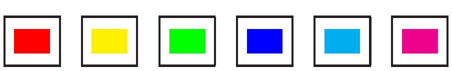
Discrete distributions

Example: *(Discrete case)* Roll a (not necessarily fair) six-sided die once. The possible outcomes are the **faces** (i.e., dots) on the die.



A different die might have six colors for the six sides (like a Rubik's cube).





□

Example: *(Discrete case)* Toss a (not necessarily fair) coin once.

□

Definition: The **probability distribution** of a **discrete** random variable X consists of the possible values of X along with their associated probabilities.

We sometimes use the terminology **population distribution**.

Example: Fair (six-sided) die. Let X be the numerical outcome. Determine the probability distribution of X, graph this probability distribution, and compute the mean of the probability distribution.

Example: Toss a fair coin 3 times. Let X = number of heads.

Let Y = number of matching coins among the three tosses.

□

Terminology: The **expected value** of X is the **mean** of a random variable X.

Example: (hypothetical) Suppose an airline often overbooks flights, because past experience shows that some passengers fail to show.

Let the random variable X be the number of passengers who cannot be boarded because there are more passengers than seats.

x	$P(x)$
0	0.6
1	0.2
2	0.1
3	0.1
sum	1

(a) Compute the expected value of X.

(b) Suppose each unboarded passenger costs the airline \$100 in a ticket voucher. Compute the average cost to the airline per flight in ticket vouchers.

□

Law of Large Numbers

$\bar{X}$ "converges" to μ as n gets large.

Likewise, s^2 "converges" to the population variance, σ^2, as n gets large.

Similarly, s "converges" to the population standard deviation, σ, as n gets large.

5.2 The Binomial Distribution

Finding Probabilities When Each Observation Has Two Possible Outcomes

The binomial distribution

Example: Toss an unfair coin 5 times, where $p = P(\text{heads}) = 0.4$.
Let X be the number of heads.
Suppose we want to determine $P(X = 2)$.

$\square$

A **Bernoulli** trial can have two possible outcomes, *success* or *failure*.

Definition of a **binomial** random variable X.

1. Let n, the number of Bernoulli trials, be **fixed** in advance.
2. The Bernoulli trials are **independent**.
3. The probability of success of a Bernoulli trial is p, which is the same for all observations.

Let X be the number of *successes*. Then, X is a binomial(n, p) random variable.

$$P(X = x) = \frac{n!}{x!\,(n-x)!}\, p^x\, (1-p)^{n-x}, \quad \text{for } x = 0, 1, 2, \ldots, n$$

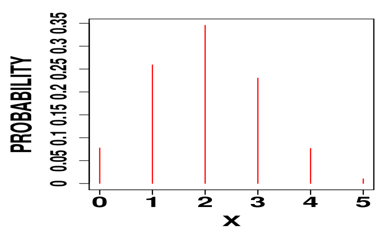
Example: $n = 5$ tosses of an unfair coin.

Assume $p = P(heads) = 0.4$ (OR 40% Democrats from a huge population)

Let X be the number of heads.

Determine the probability distribution of X and construct the line graph.

x	$P(x)$
0	0.07776
1	0.2592
2	0.3456
3	0.2304
4	0.0768
5	0.01024



Determine the probability of obtaining *at least one* heads among the five coin tosses.

In 100 tosses of this coin, on average, how many heads do you expect?

□

Mean and standard deviation of a binomial random variable

$$\mu = np, \qquad \sigma^2 = np(1-p), \qquad \sigma = \sqrt{np(1-p)}$$

Example: *Revisit.* Let $X \sim \text{Binomial}(n = 5, p = 0.4)$. Compute the *mean* and *standard deviation* of X.

□

Example: Consider a *huge* population where 30% of the people are Democrats. Let X be the number of Democrats in a sample of size 1000. Compute the mean and standard deviation of X.

Remark: For large sample sizes, a Binomial(n, p) random variable is approximately Normal(μ, σ), in which case the Empirical Rule may be used; details are in section 6.3.