Example 1. You toss a coin infinitely often. The outcomes of the coin tosses are independent, and $\mathbb{P}(H) = p$ and $\mathbb{P}(T) = 1 - p$.

(a) What is the probability of at least 1 heads in the first n tosses?

(b) What is the probability of exactly k heads in the first n tosses.

A. **Discrete Random Variables.** Recall the sample space Ω of infinite sequences of independent coin tosses, with $\mathbb{P}(H) = p$ and $\mathbb{P}(T) = 1 - p$.

Consider the function $X : \Omega \to \mathbb{N}$ that assigns, to a sequence, the number of heads that occur in the first n tosses. We define:

$$p(k) = \mathbb{P}(X = k) = \mathbb{P}(\text{exactly } k \text{ heads}) = \binom{n}{k} p^k (1-p)^{n-k}$$

Discrete Random Variable. Let Ω be the sample space of a probability space.

A **discrete random variable** is a function: $X : \Omega \to \mathbb{R}$ having **countably** many values. Its **probability mass function** is:

$p(x) = \mathbb{P}(X = x) = \mathbb{P}(\text{set of outcomes with value } x)$

This tells us the **probability distribution** of the random variable, i.e. how probability is distributed over values of the random variable.

If two discrete random variables X and Y have the same probability distribution, then we say X and Y are **identically distributed** and write $X \sim Y$.

By countably many values, we mean countably many outputs, or that its range is countable.

In the example of the number X of heads in n tosses, we consider:

$$\sum_{k=0}^{n} p(k) = \sum_{k=0}^{n} \mathbb{P}(X = k) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} =$$

Total Mass. If X is a discrete random variable, with probability mass function:

$p(x) = \mathbb{P}(X = x)$

then:

$$\sum_{x} p(x) =$$

where the sum is over all countably many values x of the random variable X.

B. **Binomial Random Variables.** The random variable equal to the number X of heads after n independent tosses, where heads occurs with probability p, is an example of a **binomial random variable**.

Binomial Random Variable. Let Ω be the sample space of sequences of n independent trials, each with success rate p, and failure rate $1-p$.

Then $X : \Omega \to \mathbb{N}$, defined by:

$X(\text{sequence}) = \text{number of successes in those } n \text{ trials}$

is the **binomial random variable** with parameter n and p:

$X \sim \mathbf{Binom}(n, p)$

Its probability mass function is:

$p(k) = \mathbb{P}(X = k) = \mathbb{P}(k \text{ successes in } n \text{ trials}) =$

C. **Bernoulli Random Variables.** A binomial random variable with parameter $n = 1$ is called a Bernoulli random variable.

Bernoulli Random Variable. Let Ω be the sample space of a **single** trial of an experiment, with success rate p, and failure rate $1 - p$.

Then $X : \Omega \to \{0, 1\}$, defined by:

$X(\text{success}) = 1$ and $X(\text{failure}) = 0$

is the **Bernoulli random variable** with parameter p:

$X \sim \mathbf{Bernoulli}(p) \sim$

Its probability mass function is:

$$p(k) = \mathbb{P}(X = k) = \mathbb{P}(k \text{ successes in 1 trial}) = \left\{ \right.$$

A Bernoulli variable is also called an **indicator function** in that it indicates whether an event occurs, specifically the event of a success.

Indicator Function. Let E be an event in probability space with sample space Ω. Then the **indicator function** of E is the random variable:

$\mathbb{1}_E : \Omega \to \{0, 1\}$

defined by:

$$\mathbb{1}_E(\text{outcome}) = \left\{ \right.$$

If the event E occurs with probability p, then:

$\mathbb{1}_E \sim$

Example 2. Recall the matching hats problem: Suppose that n people throw their hats into the center of a room, then the hats are mixed and returned to people uniformly at random.

Consider the random variable $X : \Omega \to \mathbb{N}$ that assigns, to an outcome, the number of people that get their hat back.

An outcome is an assignment of hats back to people. The sample space Ω is made up of these outcomes.

We had calculated:

$$p(k) = \mathbb{P}(X = k) = \mathbb{P}(\text{exactly } k \text{ people get their hat back}) = \frac{1}{k!}\sum_{r=0}^{n-k}\frac{(-1)^r}{r!}$$

We can also write X in terms of indicator functions.

Recall, an indicator function is equivalent to a Bernoulli random variable with probability of success equal to the probability of the event being indicated.

$E_i = [\text{event the } i\text{th person gets their hat back}]$

$\mathbb{1}_{E_i} \sim$

$X =$

Caution. Probability mass functions do **not** add.

One easy bit of evidence they do not add is that the total mass of a random variable should be 1. But if you add two random variables, the total mass will be 2! And no, it is not as simple as dividing by 2 to correct the error. We will explore sums of random variables more in Unit B.

D. **Independent Discrete Random Variables.** If X and Y are random variables on the same probability space, then when we say they are **independent**, we mean that, that the event that X takes on any particular value is independent from the event that Y takes on any particular value.

Independent Discrete Random Variables. Two discrete random variables X and Y, on the same probability space are **independent** if and only if:

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

for all values x and y of X and Y.

Intuitively, if a Bernoulli random variable counts the success of one trial of an experiment, then if we add n **independent and identically distributed (iid)** Bernoulli variables, then we count the number of successes of n independent trials of that experiment. Is this not a binomial random variable?

Remember that identically distributed meant they had the same probability mass functions. In this case, this means the Bernoulli variables all have the same parameter p.

Sum of iid Bernoullis is Binomial. Let $X_1, X_2, \ldots, X_n$ be **iid** Bernoulli random variables, with p as their common parameter. Then:

$X_1 + X_2 + \cdots + X_n \sim \mathrm{Binom}(n, p)$

In the matching hats problem, we expressed the number of people who get their hat back as a sum of indicator functions of the event that exactly k get their hat back, i.e. as a sum of Bernoulli random variables. However, the result the the left does not apply? Why? The events are **not** independent: if one person gets their that back, that changes the probability that another person gets their hat back.

Let's confirm using computations that the sum has the binomial distribution:

$\mathbb{P}(X_1 + X_2 + \cdots + X_n = k) =$