3 Discrete Random Variables

We need good data and good statistical thinking in order to make good predictions.

3.1 Random variables

It is better to define a random variable as a *function* mapping the sample space to the real line.

Definition of Random Variable:

Given an experiment with sample space S, a *random variable* is a function from the sample space S to the real numbers $\mathbb{R}$.

A random variable X assigns a numerical value $X(s)$ to each possible outcome s of the experiment. The randomness comes from the fact that we have **a random experiment** the mapping itself is **deterministic**.

Example: coin tosses

We toss a fair coin twice. The sample space $S = HH, HT, TH, TT$.

Some random variables on this space:

- X is the number of Heads. So $X(HH) = 2, X(TH) = X(HT) = 1, X(TT) = 0$;
- Y is the number of Tails. $Y = 2 - X$ or $Y(s) = 2 - X(s) \forall s$;
- I is 1 is the first toss lands Heads and 0 otherwise. This is an example of *indicator random variable*.

We also can encode sample space as $(1, 1), (1, 0), (0, 1), (0, 0)$. Then we can give explicit formulas for X, Y, I :

$$X(s_1, s_2) = s_1 + s_2, Y(s_1, s_2) = 2 - s_1 - s_2, I(s_1, s_2) = s_1$$

After we perform the experiment and the outcome s has been realized, the random variable *crystallizes* into the numerical value $X(s)$.

3.2 Distributions and probability mass functions