

a random variable turns outcomes into numbers, and its distribution says how the probability is spread across them

Random Variables

- a random variable is a real-valued function on the sample space, written $X:S \rightarrow \mathbb{R}$
- a capital letter (X, Y, Z) names the random variable itself
- the matching small letter x names one particular value it takes
- the range of X is the set of values X actually assumes
- it can be far smaller than the sample space: ten coins give 1024 outcomes
- but only eleven possible head counts, 0 through 10
- the event $[X=x]$ collects every outcome mapped to that one value x
- $P[X=x]$ is really the probability of that whole collected event

Discrete vs Continuous Random Variable *

Discrete random variable

- possible values are finite or countably infinite
- values are obtained by COUNTING, and are usually non-negative integers
- e.g. the number of heads in ten tosses, or the number of defective items in a sample

Continuous random variable

- possible values fill a whole interval of real numbers, uncountably infinite
- values are obtained by MEASUREMENT, to whatever precision the instrument allows
- e.g. the height of a tree, or the waiting time for a bus

a discrete random variable is described with a probability mass function; a continuous one needs a probability density function instead, since no single value of a continuous X ever carries positive probability on its own

Probability Distribution: Sum on Two Dice

x	$P(x)$
2	$\frac{1}{36}$
3	$\frac{2}{36}$
4	$\frac{3}{36}$
5	$\frac{4}{36}$
6	$\frac{5}{36}$
7	$\frac{6}{36}$
8	$\frac{5}{36}$
9	$\frac{4}{36}$
10	$\frac{3}{36}$
11	$\frac{2}{36}$
12	$\frac{1}{36}$

check any probability distribution table by adding every entry in the $P(x)$ column: here the eleven fractions total $36/36=1$, exactly as they must. This description, values of X against their probabilities, is what is meant by the probability distribution of X

Probability Mass Function (p.m.f.)

p_i defined for $x_1, x_2, x_3, \dots$, the possible values of X

$$p_i = P[X=x_i]$$

every p.m.f. must satisfy both conditions at once

$$p_i \geq 0, \sum_i p_i = 1$$

the second condition is the fastest first check on a claimed distribution: add every probability in the $P(x)$ column, and if it does not total to 1 the table is wrong before anything else is computed from it