Random Variables

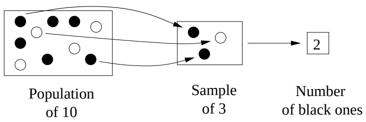
We are often interested in numbers that are obtained randomly:

- The number who support Joe Blowhard in a random sample of 100 eligible voters.
- The average of 10 measurements of the charge of the electron.
- The maximum age of a person in a random sample of 1000 Canadians.

These are called *random variables*.

Mathematically, a r. v. is a mapping from an outcome in the sample space to a number.

For example:



Distribution of a Discrete Random Variable

Sometimes, the value of some random variable is the *only* thing about an outcome that we are interested in.

Then, rather than look at probabilities of outcomes, we look at just the probability distribution of the random variable.

If a random variable, say X, has a finite range, say $\{0, 1, 2\}$, we can make a table of:

$$\begin{aligned} P(X=0) &= P(\text{all outcomes for which } X=0) \\ P(X=1) &= P(\text{all outcomes for which } X=1) \\ P(X=2) &= P(\text{all outcomes for which } X=2) \end{aligned}$$

Once we have this table, we can figure out the probabilities of other events involving X, such as $P(X<2)$, without referring to the detailed outcomes.

Example: Number of Heads in Four Tosses of a Fair Coin

Here is the sample space for tossing a coin four times, arranged according to the value of the random variable $X =$ number of heads:

		HTTH		
		HTHT		
	HTTT	THTH	HHHT	
	THTT	HHTT	HHTH	
	TTHT	THHT	HTHH	
TTTT	TTTH	TTHH	THHH	HHHH
$X=0$	$X=1$	$X=2$	$X=3$	$X=4$

If the coin is fair, and tosses are independent, the 16 outcomes will be equally likely. We can then compute the table

$$\begin{aligned} P(X=0) &= 1/16 \\ P(X=1) &= 4/16 \\ P(X=2) &= 6/16 \\ P(X=3) &= 4/16 \\ P(X=4) &= 1/16 \end{aligned}$$

Joint Probabilities and Independence of Random Variables

We can define two random variables, X and Y, on the same sample space. We then talk about *joint probabilities*, such as

$$P(X=5 \text{ and } Y=2)$$

Random variables X and Y are *independent* if we can get these joint probabilities by multiplying "marginal" probabilities, that is

$$P(X=i \text{ and } Y=j) \;=\; P(X=i)\,P(Y=j)$$

for all i and j in the range of X and Y.

Assigning probabilities in the model is much easier if we can assume independence of the random variables: We need only specify the marginal distributions for each one; the joint probabilities then follow.

Example: Rolling Two Dice

Suppose we roll a red die and a green die. Let R and G be the random variables whose values are the numbers they show.

Dice behave so that R and G are independent. So, for example:

$$P(R{=}3 \text{ and } G{=}6) = P(R{=}3)\,P(G{=}6) = (1/6)(1/6)$$

We can define some other random variables:

$$S = R+G$$

$$E_S = \begin{cases} 0 & \text{if } S \text{ is even} \\ 1 & \text{if } S \text{ is odd} \end{cases}$$

$$E_R = \begin{cases} 0 & \text{if } R \text{ is even} \\ 1 & \text{if } R \text{ is odd} \end{cases}$$

Are S and R independent?
Are E_S and E_R independent?

When Can We Assume Independence?

When can we assume independence? The answer may depend not only on the meaning of the variables, but also on the sample space.

Suppose you pick a person at random, and ask them two questions:

- How many Maple Leaf hockey games did you attend last winter?
- How many times did you swim in Lake Ontario last summer?

The answers are the values of two random variables. Are these variables independent? Consider two sample spaces, that result from the following sampling methods:

- The person is picked uniformly from among all residents of Toronto.
- The person is picked uniformly from among all residents of Canada.