Random variables

Up to now we have studied probabilities of sets of outcomes. In practice, in many experiment we care about some numerical property of these outcomes. For example, if we sample a person from a particular population, we may want to measure her age, height, the time it takes her to solve a problem, etc. Here is where the concept of a random variable comes at hand. Intuitively, we can think of a random variable (rav) as a numerical measurement of outcomes. More precisely, a random variable is a rule (i.e., a function) that associates numbers to outcomes. In order to define the concept of random variable, we first need to see a few things about functions.

Functions: Intuitively a function is a rule that associates members of two sets.
The first set is called the **domain** and the second set is called the **target** or **codomain**. This rule has to be such that an element of the domain should not be associated to more than one element of the codomain. Functions are described using the following notation

$$f : A \longrightarrow B \qquad (2.1)$$

where f is the symbol identifying the function, A is the domain and B is the target. For example, $h : \quad \longrightarrow \quad$ tells us that h is a function whose inputs are real numbers and whose outputs are also real numbers.

The function $h(x) = (2)(x)+4$ would satisfy that description. Random variables are **functions** whose domain is the outcome space and whose codomain is the real numbers. In practice we can think of them as numerical measurements of outcomes. The input to a random variable is an elementary outcome and the output is a number.

Example: Consider the experiment of tossing a fair coin twice. In this case the outcome space is as follows:

$$\Omega = \{(H,H), (H, T), (T,H), (T, T)\} . \qquad (2.2)$$

One possible way to assign numbers to these outcomes is to count the number of heads in the outcome. I will name such a function with the symbol X, thus

$X : \Omega \longrightarrow \quad$ and

$$X(\omega) = \begin{cases} 0 & \text{if } \omega = (T,T) \\ 1 & \text{if } \omega = (T,H) \text{ or } \omega = (H,T) \\ 2 & \text{if } \omega = (H,H) \end{cases} \tag{2.3}$$

In many cases it is useful to define sets of Ω using the outcomes of the random variable X. For example the set $\{\omega : X(\omega) \leq 1\}$ is the set of outcomes for which X associates a number smaller or equal to 1. In other words

$$\{\omega : X(\omega) \leq 1\} = \{(T, T), (T,H), (H, T)\} \tag{2.4}$$

Another possible random variable for this experiment may measure whether the first element of an outcome is "heads". I will denote this random variable with the letter Y_1. Thus $Y_1 : \Omega \longrightarrow$ and

$$Y_1(\omega) = \begin{cases} 0 & \text{if } \omega = (T,T) \text{ or } \omega = (T,H) \\ 1 & \text{if } \omega = (H,H) \text{ or } \omega = (H,T) \end{cases} \tag{2.5}$$

Yet another random variable, which I will name Y_2 may tell us whether the second element of an outcome is heads.

$$Y_2(\omega) = \begin{cases} 0 & \text{if } \omega = (T,T) \text{ or } \omega = (H,T) \\ 1 & \text{if } \omega = (H,H) \text{ or } \omega = (T,H) \end{cases} \tag{2.6}$$

We can also describe relationships between random variables. For example, for all outcomes ω in Ω it is true that

$$X(\omega) = Y_1(\omega) + Y_2(\omega) \tag{2.7}$$

This relationship is represented succinctly as

$$X = Y_1 + Y_2 \tag{2.8}$$

Example: Consider an experiment in which we select a sample of 100 students from UCSD using simple random sampling (i.e., all the students have equal chance of being selected and the selection of each students does not constrain the selection of the rest of the students). In this case the sample space is the set of all possible samples of 100 students. In other words, each outcome is a sample that contains 100 students. A possible random variable for this experiment is the height of the first student in an outcome (remember each outcome is a sample with 100 students). We will refer to this random variable with the symbol H_1. Note given an outcome of the experiment, (i.e., a sample of 100 students) H_1 would assign a number to that outcome. Another random variable for this experiment is the height of the second student in an outcome. I will call this random variable H_2. More generally we may define the random variables $H_1, \ldots, H_{100}$ where $H_i : \Omega \longrightarrow$ such that $H_i(\omega)$ is the height of the subject number i in the sample ω . The average height of that sample would also be a random variable, which could be symbolized as $\ddot{H}$ and defined as follows

$$\bar{H}(\omega) = \frac{1}{100}(H_1(\omega) + \cdots + H_{100}(\omega)) \text{ for all } \omega \in \Omega \qquad (2.9)$$

or more succinctly

$$\bar{H} = \frac{1}{100}(H_1 + \cdots + H_{100}) \qquad (2.10)$$

I want you to remember that all these **random variables are not numbers**, they are functions (rules) that assign numbers to outcomes. The output of these functions may change with the outcome, thus the name random variable.

Definition A random variable X on a probability space (Ω, F, P) is a function $X : \Omega \longrightarrow$. The domain of the function is the outcome space and the target is the real numbers.

Notation: By convention random variables are represented with capital letters. For example $X : \Omega \longrightarrow$, tells us that X is a random variable. Specific values of a random variable are represented with small letters. For example, $X(\omega) = u$ tells us that the "measurement" assigned to the outcome ω by the random variable X is u. Also I will represents sets like

$$\{\omega : X(\omega) = u\} \tag{2.11}$$

with the simplified notation

$$\{X = u\} \tag{2.12}$$

I will also denote probabilities of such sets in a simplified, yet misleading, way. For example, the simplified notation

$$P(X = u) \tag{2.13}$$

or

$$P(\{X = u\}) \tag{2.14}$$

will stand for

$$P(\{\omega : \mathrm{X}(\omega) = u\}) \tag{2.15}$$

Note the simplified notation is a bit misleading since for example X cannot possibly equal u since the first is a function and the second is a number.

Definitions:

• A random variable X is **discrete** if there is a countable set or real numbers $\{x_1, x_2, . . .\}$ such that $P(X \in \{x_1, x_2, . . .\}) = 1$.

• A random variable X is **continuous** if for all real numbers u the probability that X takes that value is zero. More formally, for all $u \in$, $P(X = u) = 0$.

• A random variable X is **mixed** if it is not continuous and it is not discrete.