On Random Variables

François Fleuret

February 12, 2024

Random variables (RVs) are central to any model of a random phenomenon, but their mathematical definition is unclear to most. This is an attempt at giving an intuitive understanding of their definition and utility.

1 Modeling randomness

To formalize something "random", the natural strategy is to define a distribution, that is, in the finite case, a list of values / probabilities. For instance, the head / tail result of a coin flipping would be

$$\{(H, 0.5), (T, 0.5)\}.$$

This is perfectly fine, until you have several such objects. To model two coins A and B, it seems intuitively okay: they have nothing to do with each other, they are "independent", so defining how they behave individually is sufficient.

2 Non-independent variables

The process to generate two random values can be such that they are related. Consider for instance that A is the result of flipping a coin, and B as *the inverse value of A*.

Both A and B are legitimate RVs, a both have the same distribution (H, 0.5) (T, 0.5). So where is the information that they have a relation?

With models of the respective distributions of A and B, this is nowhere. This can be fixed in some way by specifying the distribution of the pair (A, B). That would be here

$$\{(H/H, 0.0), (H/T, 0.5), (T/H, 0.5), (T/T, 0.0)\}.$$

The distribution of A and B individually are called the marginal distributions, and this is the joint distribution.

Note that the joint is a far richer object than the two marginals, and in general many different joints are consistent with given marginals. Here for instance, the marginals are the same as if A and B where two independent coins, even though they are not.

Even though this could somehow work, the notion of a RV here is very unclear: it is not simply a

distribution, and every time a new one is defined, it require the specification of the joint with all the variables already defined.

3 Random Variables

The actual definition of a RV is a bit technical. Intuitively, in some way, it consists of defining first "the source of all randomness", and then every RV is a deterministic function of it.

Formally, it relies first on the definition of a set Ω such that its subsets can be measured, with all the desirable properties, such as $\mu(\Omega) = 1, \mu(\emptyset) = 0$ and $A \cap B = \emptyset \Rightarrow \mu(A \cup B) = \mu(A) + \mu(B)$.

There is a technical point: for some Ω it may be impossible to define such a measure on all its subsets due to tricky infinity-related pathologies. So the set Σ of measurable subsets is explicitly specified and called a σ-algebra. In any practical situation this technicality does not matter, since Σ contains anything needed.

The triplet (Ω, Σ, μ) is a measured set.

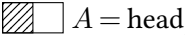
Given such a measured set, an random variable X is a mapping from Ω into another set, and the probability that X takes the value x is the mea-

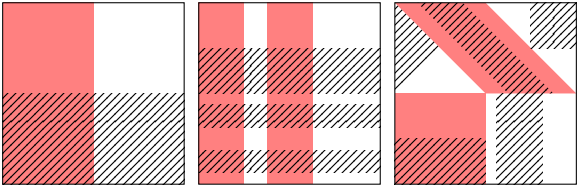
sure of the subset of Ω where X takes the value x:

$$P(X = x) = \mu(X^{-1}(x))$$

You can imagine Ω as the square $[0,1]^2$ in $\mathbb{R}^2$ with the usual geometrical area for μ.

For instance if the two coins A and B are flipped independently, we could picture possible random variables with the proper distribution as follows:





And if A is flipped and B is the inverse of A, possible RV would be