Esssentials of Applied Data Analysis
IPSA-USP Summer School 2018

Introduction to Random Variables

Leonardo Sangali Barone
leonardo.barone@usp.br

jan/18

Random Variables

Theory, Concepts and Variables

- A theory is a set of statements that involve concepts.

- Concepts are inventions that help us understand the world.

- Theories (and the hypotheses they imply) concern relationships among abstract concepts.

What is a variable? A variable is simply an indicator we develop to measure our concepts.

Variables *versus* constants

A constant is a variable that can assume only one value. For example: if we want to use survey data of women teenagers to compute abortion rates, sex of the subjects is not a variable, is a constant

Golden rule: a variable must vary!

Types of Variables

A) Discrete - values are countable, even if infinite.

- Nominal - can be counted, but not ordered

- Ordinal - can be counted and ordered

- Integer - can be counted, ordered and the distance between categories are mathematically meaningful (yes, numbers with no decimals).

B) Continuous - Infinite set of values. Not countable (yes, whatever number there is, including $\sqrt{2}$ and π).

Can you think of some examples of each type of variable?

Whatch out! Nominal and Ordinal variable's categories can be coded with numbers, but mathematical operations with those numbers are not possible (if it were, they would be Integer variables).

Random variables, distributions and probability functions

Probability function is a function that describes the likelihood of the values a concept (variable) might take given a description of the process that generates it.

What?!?!?!?

We have already studied the rules of probability. Now, we are going to use our knowledge of probability to learn about random variables.

A *probability function* is just a function that indicates the probability of each value (events!!!) of a random variable (measurements of concepts!!!). We usually represent it as any other function, $f(x)$, and it is defined as:

$$f(x) = P(X = x)$$

One way to think about it is that it is a mapping function, since it "maps" the probability associated with each value of a random variable (X).
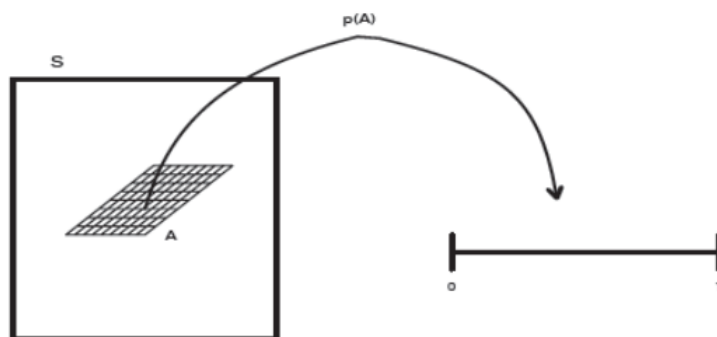
Fig. 7.2. THE MAPPING OF A PROBABILITY FUNCTION

Figure 1: Mapping function

Random variable doesn't mean we are picking any variable randomly, but a variable that can take some array of values, with the probability that it takes any particular value defined according to some random process.

The set of values the variable might take is the *distribution* of the random variable.

The probability function gives us the probability distribution of a random variable. Conventionally, we call it *probability mass function* if the variable is discrete and *probability density function* if the variable is continuous.

Some random variables in political science:

- Voter's choice
- Government regimes; electoral systems
- Choices of policy and policy paths
- % of votes in an election
- Number of parties in a cabinet

- Time past since the last war/disaster

- Public health care system expenditure

- Number of families/individuals in a program

Can you think of some more?