

Random Variables

Notes

Random Variables

A **random variable** X is the numeric outcome of a random phenomenon. For instance, when we roll a die and X is the outcome, the possible values are $1, 2, \dots, 6$.

When we have a random variable X , we write $P(X = k)$ for the probability that X takes the value k . We will often also look at $P(X > k)$ and $P(X \geq k)$, and similarly for “less than”.

Random variables are separated in **discrete** and **continuous**.

Discrete Discrete random variables are those that take a finite set of possible values (for instance the result of a die roll, counting the number of heads in 5 coin flips and so on).

Continuous Continuous random variables are those that take values in a whole range. A variable that follows the normal distribution is a good example of that.

We will mostly study discrete distributions, with the exception of the normal distribution.

Random Variables and Sampling

We can think of any “fixed” variable with numerical values as a random variable. For instance imagine we have a group of 20 people, 4 of them with no siblings, 5 with 1 sibling, 10 of them with 2 siblings and 1 with 3 siblings.

The number of siblings in this case is one of the standard, fixed, variables we studied at the beginning of the class. In SPSS it would correspond to a column with 20 rows on it, 4 of which are 0, 5 of them are 1 and so on.

We can also think of a corresponding random variable: “Pick one person at random, and look at the number of siblings”. The possible outcomes are 0, 1, 2 and 3, and their probabilities are equal to the *relative frequencies* of those outcomes. For instance $P(1) = 5/20 = 1/4$.

$$P(X = k) \longleftrightarrow \text{relative frequency of } k$$

Probability Distribution Function

The *Probability Distribution Function* (often abbreviated as PDF or pdf) is the function that for a given k tells us what $P(X = k)$ is. For discrete distributions, we can express it as a table. Let us look at an example:

We roll a fair 6-sided die. If it is 3 or less, we don't win anything. If it is 4 or more we win \$1 and roll the die a second time. If that second time is a 5 or 6 then we win one more dollar.

X is the number of dollars we end up winning.

Make a tree/decision diagram showing how the play might go, and what values of X you get on each resulting path. You should end up with the following table:

X	0	1	2
$P(X)$	1/2	1/4	1/4

Let us consider one more example:

Alice and Bob play a game involving rolling a die. If the die is 4 or more, Bob wins \$1, if it is 5 or more then Alice wins \$2.

X denotes the difference between Bob's winnings and Alice's winnings.

There are two versions of the problem: One where Alice and Bob use the same, single, die roll, and one where they use two different die rolls.

Construct tree diagrams and corresponding probability tables for both cases.

And one more example for practice:

You play a game that involves a coin that has a 70% chance of coming up heads. You start by paying \$2 for the opportunity to play. You then get to flip the coin. Each time it lands heads you win a dollar, up to a total of 3 coin flips. BUT the first time you get tails the game stops.

X is the total net gain (accounting for the 2 dollars you had to pay at the beginning)

Create a tree diagram and a probability table for this problem.