Statistics and probability notebook

Mateus Melo

2022-02-08

Contents

ET586 - Estatistica e Probabilidade

Discrete Probability Distributions

An introduction to discrete random variables and discrete probability distributions

An informal definition:

A random variable is a quantitative variable whose value depends on chance in some way.

Suppose we are to toss a coin 3 times.

Let X represent the number of heads in the 3 tosses.

X is a random variable that will take on of the values:

$$0, 1, 2, 3$$

Here, X is a **Discrete Random Variable**

Discrete random variables can take on a ***countable*** *number of possible values.*

Continuous random variables can take on any value in an interval

Examples of discrete random variables:

- The number of free throws an NBA player makes in his next 20 attempts. Possible values: $0, 1, 2, ..., 20$
- The number of rolls of a die needed to roll a 3 for the first time. Possible values: $1, 2, 3, ...$ - It's a countably infinite number of possible values.
- The profit on a \$1.50 bet on black in roulette. Possible values: $-1.50, 1.50$

Examples of continuous random variables:

- The velocity of the next pitch in Major League Baseball.
- The time between lightning strikes in a thunderstorm.

Because of the difference between discrete and continuous variables, we must treat them diferently.

The probability distribution of a discrete random variable X is a listing of all possible values of X and their probabilities of occurring.

Approximately 60 of full-term newborn babies develop jaundice.

Suppose we randomly sample 2 full-term newborn babies, and let X represent the number that develop jaundice.

What is the probability distribution of X?

$$0, 1, 2$$

Possible outcomes:

-	JJ	JN	NJ	NN
Value of X	2	1	1	0
Probability	$0.6 \cdot 0.6$	$0.6 \cdot 0.4$	$0.4 \cdot 0.6$	$0.4 \cdot 0.4$

Value of X	0	1	2
Probability	0.16	0.48	0.36

Now we can read probabilities out of the table, like so:

$$P(X = 0) = 0.16$$

We read it like this:

$$P(X = x)$$

X is the random variable, and x is a value that the random variable can assume.

We usually use a simpler equivalent notation:

$$p(x)$$

Where $p(x)$ denotes the probability of X when it assumes the x value.

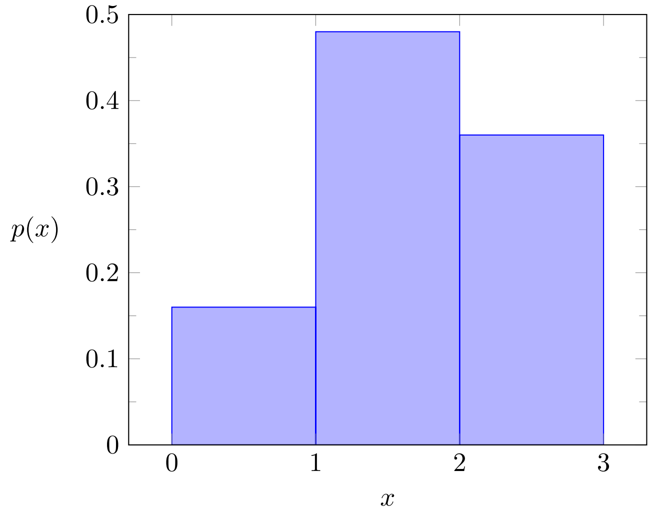
We then can replace the previous table with:

x	0	1	2
$p(x)$	0.16	0.48	0.36

All discrete probability distributions must satisfy these 2 conditions:

1. $0 \leq p(x) \leq 1$ for all x
2. $\sum_{all x} p(x) = 1$

We may plot the table above in a probability histogram:



Insert here continuous probability distribution

In continuous probability distribution, the y axis doesn't displays probability, but *probability density.*

Since tables may get hard to keep track for many probabilities, we usualy make a formula to track a probability distribution based on a that formula. The example above can be described by the following formula:

$$p(x) = \binom{2}{x} 0.6^x (1 - 0.6)^{2-x}$$
$$for x = 0, 1, 2$$

description: A probability mass function

There are many common types of discrete probability distributions:

- Geometric
- Hypergeometric
- Binomial
- Poisson