Intro to Proba: continued

Motivation behind

While working with data you can make lots of mistakes due to absence of understanding probability concepts. After building the model which seems to show great results you might realize that none of the predictions came true. The problem is in the core. You are not allowed to build model without proving that you can do that!

Common question: “Why it is so important for linear regression to have independent variables?”

Answer: If variables are not independent you need to calculate conditional distribution among them and include it to your model. Too much theory, let’s dive inside.

Random Variables

Random Variable (RV) - variable whose value depend on some random phenomenon (wiki).

In fact it is every event that has a random outcome depending on some conditions. Body weight, building height, weather temperature or number of phone calls per minute are simple examples of RV.

There are 2 different types of RV:

1. Discrete RV
2. Continuous RV

Discrete RV

Sample space of discrete RV is finite and countable (0,1,2,...).

Example: gender.

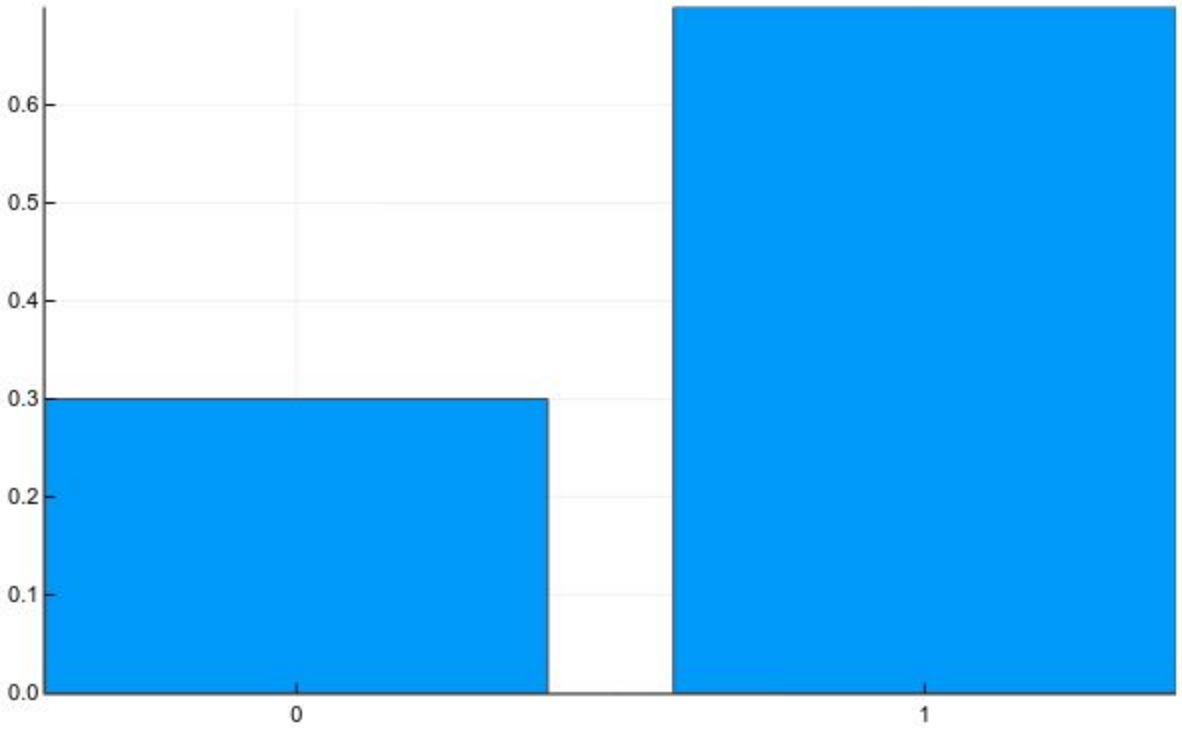
Main characteristics is:

Probability mass function (pmf) - function that defines probabilities for every value from sample space.

Example:

p(male) = 0.3;
p(female) = 0.7.

Example



Continuous RV

Sample space of continuous RV is infinite or uncountable.

Example: price of potatoes.

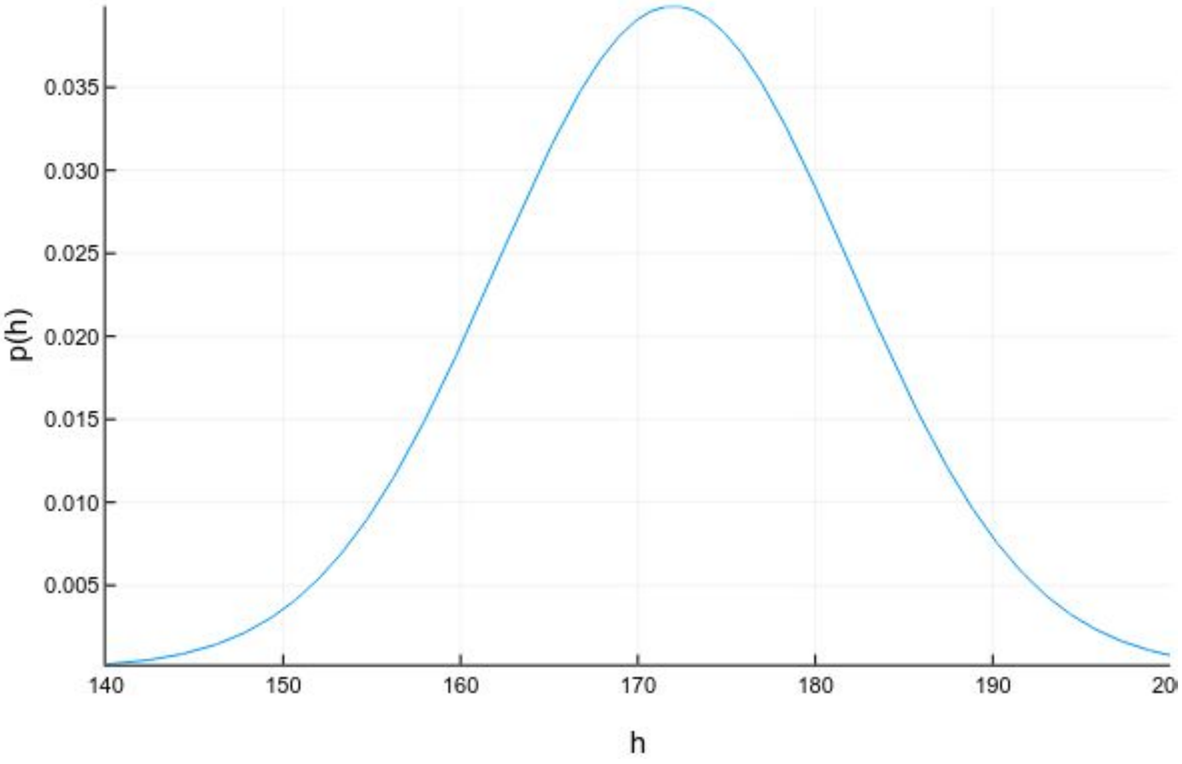
Main characteristics is:

Probability density function (pdf) - function that defines probability densities for every value from sample space. Difference from pmf - every value of pdf is not a probability of having a certain value, but the relative share of probability at this value. Probability of every single number in this case is equal to 0 (total probability is equal to 1, number of datapoints is infinite, so 1/infinite is 0), so we use it to calculate probability of a range.

Example:

$$\begin{cases} \frac{1}{b-a} & \text{for } x \in [a, b] \\ 0 & \text{otherwise} \end{cases}$$

Example



IMPORTANT INFO

The most important distributions to keep an eye on are:

1. Bernoulli
2. Binomial
3. Uniform
4. Poisson
5. Categorical
6. Normal
7. Exponential
8. Log-Normal

Marginal, conditional and joint distributions

Similarly with probabilities, we can define marginal, conditional and joint probability distributions.

Marginal probability distribution - distribution of marginal probability

Conditional probability distribution - distribution of conditional probability

Joint probability distribution - distribution of joint probability

Example

In the table below we can see the distribution of probabilities to meet a person of certain gender with certain number of children in the building.

In grey we can find joint probabilities of 2 parameters (share of people of certain type in the total number of people in the building), while in blue we see the marginal probabilities of 2 parameters.

Gender	Number of children			Marginal Probability
	0	1	2	
Male	0,1	0,1	0,1	0,3
Female	0,2	0,4	0,1	0,7
Marginal Probability	0,3	0,5	0,2	1

Example cont'd

Gender	Number of children		
	0	1	2
Male	0,1/0,3	0,1/0,5	0,1/0,2
Female	0,2/0,3	0,4/0,5	0,1/0,2

In the table above we can find the conditional probability of gender given the number of children (percentage per column).

Gender	Number of children		
	0	1	2
Male	0,1/0,3	0,1/0,3	0,1/0,3
Female	0,2/0,7	0,4/0,7	0,1/0,7

In the table below we can find the conditional probability of number of children given the gender (percentage per row).

TL;DR: Discussion

If you have a table with joint distribution defined you can easily define both marginal and conditional distribution. Unfortunately, you can define such a table only for discrete rv.

Another problem of such tables - the more values your random variable can take, the more memory you need to define it.

Law of Total Probability

Joint probability density function is equal to the product of marginal and conditional probabilities.

$$p(x,y) = p(y|x) * p(x) = p(x|y) * p(x)$$

Which can be explained on example:

Probability of a person getting into accident while crossing the street on a red light of traffic lights is equal to the product of probability to cross the street on a red light and probability of getting into accident given that you are crossing the street on a red light. Don’t cross the street on a red light. Don’t be dumb.

Chain Rule

What if we have more than 2 events happening at a time? Apply the chain rule:

$$p(w_1, w_2, \ldots, w_{1000}) = p(w_1) * p(w_2|w_1) * p(w_3|w_1w_2) * \ldots * p(w_{1000}|w_1w_2\ldots)$$

Example:

Text analysis. Clusterization of communication patterns - we need to calculate the probability that text consisting of 1000 words has the same meaning as another text of 1000 words. The meaning of the text is explained by joint distribution of words in the text. In order to build the table representation of probabilities one will have to build the table containing of more than 10^{301} cells which is more than number of atoms in the universe.

Exercises to practise

Let human height be a normally distributed random variable with the following parameters: μ=172 and σ^2=10. What is the probability of meeting a person exactly 178cm tall?

Answer

Since the height is continuous random variable, the probability of meeting a person of a certain height is equal to 0.

Problems of Probability Theory

1. Irrationality in decision making.
2. Information limitations.
3. Billions of random processes surrounding us.
4. Lazyness

Examples:

1. Airplane vs Train vs Automobile accident death rate
2. Trump elections
3. ‘Fortunately’ by Remi Charlip or ‘The Drunkards’ Walk’ by Leonard Mlodinow
4. Instinct vs Analysis