

# Lecture 15 - 4: Likelihood Function

Tabaré Pérez

May 6, 2020

So now, given that:

1.  $\theta_w \geq 0$
2.  $\sum_{w \in \mathcal{W}} \theta_w = 1$

How can I compute the likelihood of generating a second set of documents?

So let's say somebody gave me all this  $\theta_w$ 's, I have them, how do I compute the likelihood of generating a document?

$$\mathbb{P}(D|\theta) = \prod_{i=1}^n \theta_{w_i} \quad (1)$$

So we would assume that we have our document, and this is the capital  $D$ , and I will use it throughout this lecture, given particular  $\theta$ s.

So if we assume that our documents has  $n$  words, again as I said earlier every single word is generated independently, so we take probability of every word in our document and just multiply them all together.

So now, we will take this same formula and write it in slightly different way. The reason we want to do it is because we have the same word which appears multiple times in the document.

So instead, we can actually refactorize it, and just say we will take the probability of this word, let's say that it appear five times, and just make this probability to the fifth degree.

So then, instead of making a product over all the words in the document, we will actually go over all the words in our vocabulary  $\mathcal{W}$  and now we will take the probability  $\theta_w$  for each word  $w$  to the  $\text{count}(w)$  in our document.

$$\mathbb{P}(D|\theta) = \prod_{i=1}^n \theta_{w_i} = \prod_{w \in \mathcal{W}} \theta_w^{\text{count}(w)} \quad (2)$$

So those are the two forms to write exactly the same things, but I will primarily use this representation.

So now we are ready to start addressing the first questions that I wrote over there, which is a question of estimation.

But before we go there, I just want to give you an example of how this model, very simple model, can be used to evaluate the likelihoods of the document.

So let's take the primitive example, and I will just look at the case whenever my vocabulary just have two words, cat and dog:

$$\mathcal{W} = \{cat, dog\} \quad (3)$$

And now I am going to write two different models that operate over this vocabulary.

So this would be the first model which takes parameters  $\theta$ . And I will decide that in this particular model, the likelihood of generating the word cat would be 0.3 and the likelihood of the word dog, in this case obvious correct, is going to be 0.7:

$$\theta; \theta_{cat} = 0.3, \theta_{dog} = 0.7 \quad (4)$$

As you can see, they sum to 1 and they both are not negative.

Now, I am going to look at model number two. Model number two, and to distinguish the parameter, I just wrote  $\theta'$ , and here, let's say we want the  $\theta'$  to generate the word cat be 0.9:

$$\theta'; \theta'_{cat} = 0.9, \theta'_{dog} = 0.1 \quad (5)$$

So you can see this model really strongly prefers cats. And then for the dog, again, because they have to sum to 1, it's going to be 0.1.

So we have two models. And as you can expect, they will be generating different types of documents:

- $\theta$  would love dog style of documents.
- $\theta'$  would love cat style of documents.

So now assume that I give you a document  $D$ :

$$D = \{cat, cat, dog\} \quad (6)$$

This particular document will have 2 cats, and 1 dog.

And what we can do, we can now compute the likelihood of these document being generated by the first model or by the second model.

And the notations that we are going to be using here, we're going to say, what is the probability of this document  $D$ , given parameters  $\theta$ .

And again, this is the values of these parameters. So in this case, if I just straightforwardly apply this formula, we are going to get:

$$\mathbb{P}(D|\theta) = 0.3^2 \times 0.7 = 0.063 \quad (7)$$

Similarly, for the second model:

$$\mathbb{P}(D|\theta') = 0.9^2 \times 0.1 = 0.081 \quad (8)$$

Again, we have the document,  $D$ , but now our parameters,  $\theta'$ , the cats, are going to have 0.9 squared multiplied by 0.1.

It's quite easy to see that this document would have higher probability when generated by the second model.