



**Data Science  
Academy**

[www.datascienceacademy.com.br](http://www.datascienceacademy.com.br)

Processamento de Linguagem Natural

Maximum Likelihood Estimation (MLE)

## Maximum Likelihood Estimation (MLE)

Esta estimativa é denominada “estimativa da probabilidade máxima” (MLE) porque corresponde à escolha dos valores dos parâmetros que geram a mais alta probabilidade de um evento, como um n-grama, no corpus de treinamento. A MLE de um n-grama  $w_1, \dots, w_n$  é:

$$P_{MLE}(w_1, \dots, w_n) = \frac{C(w_1, \dots, w_n)}{N}$$

onde  $C(w_1, \dots, w_n)$  é a frequência do n-grama  $w_1, \dots, w_n$ , e  $N$  é o número de instâncias da classe de equivalência a que o n-grama pertence; ou

$$P_{MLE}(w_n | w_1, \dots, w_{n-1}) = \frac{C(w_1, \dots, w_n)}{C(w_1, \dots, w_{n-1})}$$

A MLE é geralmente inadequada ao processamento da linguagem natural, devido a os dados serem esparsos - enquanto algumas palavras são bastante comuns, a vasta maioria é pouco frequente (e os n-gramas que as contêm são ainda mais incomuns). Por consequência, alguns eventos do corpus, sequências de palavras, podem não estar contidos nos dados de treinamento. Para estimar a probabilidade de ocorrência destes eventos, podem ser utilizadas técnicas denominadas de smoothing.