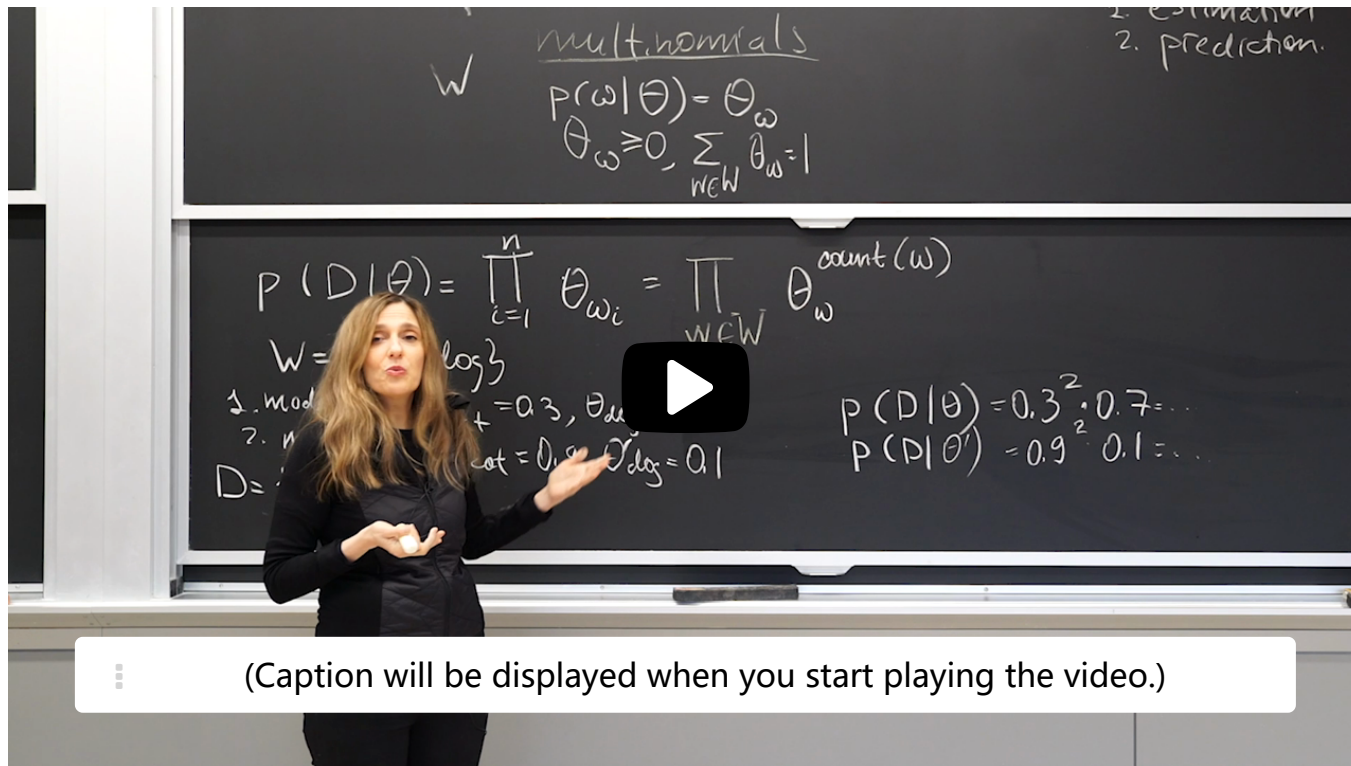


## 5. Maximum Likelihood Estimate

### Maximum Likelihood Estimate

[Start of transcript. Skip to the end.](#)



(Caption will be displayed when you start playing the video.)

So now we have our parameters.

We have ways to compute the likelihood of the documents

based on these parameters.

And the question is, how can we utilize

our training data to find the best parameters, correct?

So again, we are coming back to supervised scenario

that you've seen in the case of classifiers.

We will provide our model with, let's say,

#### Video

[Download video file](#)

#### Transcripts

[Download SubRip \(.srt\) file](#)

[Download Text \(.txt\) file](#)



## Number of parameters

1/1 point (graded)

For the following set of questions, let us consider generating documents that are English letter sequences (assume no spaces or punctuation), i.e. the vocabulary  $W = \{a, b, c, \dots, z\}$  is made up of all the letters in the English alphabet.

We would like to generate documents using this vocabulary using a multinomial model  $M$ . As described in the lecture, what is the minimal number of parameters that the model  $M$  should have? Enter your answer below.

✓ Answer: 25

#### Solution:

Recall from the lecture that for multinomial generative models we have a parameter  $\theta_w$  for each word  $w \in W$ . However, since the parameters should sum up to one, we can express one of the parameters as 1 minus the sum of all others. Since the vocabulary size for this example is 26, our model  $M$  can have 25 parameters to express the probability of each letter.

Submit

You have used 2 of 2 attempts

❗ Answers are displayed within the problem

## Maximum Likelihood estimate

1/1 point (graded)

Let  $\theta^* = \theta_a^*, \theta_b^* \dots \theta_z^*$  be the parameters of the multinomial model  $M^*$  that maximize the likelihood of generating a document  $D$ .

Further, it is known that the letter 'e' is twice as more likely to occur than the letter 'z' in document  $D$ .

Which of the following options is a correct expression relating  $\theta_e^*$  and  $\theta_z^*$ ?

- ☐  $\theta_z^* = 2 * \theta_e^*$
- ☒  $\theta_e^* = 2 * \theta_z^*$  ✓
- ☐  $\theta_z^* = \theta_e^*$
- ☐  $\theta_z^* + \theta_e^* = 2$

Solution:

Recall from the lecture that for any  $w \in W$  we have,

$$\theta_w^* = \frac{count(w)}{\sum_{w' \in W} count(w')}$$

Since  $count(e) = 2 * count(z)$ , we can conclude that  $\theta_e^* = 2 * \theta_z^*$

Submit

You have used 1 of 1 attempt

📘 Answers are displayed within the problem

Maximum Likelihood Estimate for Poisson Distribution

2/2 points (graded)  
Maximum Likelihood Estimate (MLE) is a very general method that can be applied to both continuous and discrete distributions. In this problem, we assume we have a training data  $x_1, x_2, \dots x_n$  that are drawn from a Poisson distribution:

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

and we want to use MLE to fit the parameter  $\lambda$  with the training data. We can do so by first computing the log likelihood of our training data, which is:

- ☐  $\log \lambda \sum_i x_i + n\lambda + \sum_i \log(x_i!)$
- ☒  $\log \lambda \sum_i x_i - n\lambda - \sum_i \log(x_i!)$  ✓
- ☐  $\log \lambda \prod_i x_i - n\lambda - \prod_i \log(x_i!)$
- ☐  $\log \lambda \prod_i x_i + n\lambda + \prod_i \log(x_i!)$

In the next step, we maximize this log likelihood function by taking the derivative. What is the resulting estimator for  $\lambda$ ?

- ☒  $\frac{1}{n} \sum_i x_i$  ✓
- ☐  $\frac{1}{n} \prod_i x_i$

☐  $\sum_i x_i$

☐  $\prod_i x_i$

Is it in accordance with the definition of  $\lambda$  in Poisson distribution?

**Solution:**

The loglikelihood of the data is:

$$\begin{aligned}\log \prod_i P(X = x_i) &= \log \prod_i \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} \\ &= \sum_i \log(\lambda^{x_i}) + \log(e^{-\lambda}) - \log(x_i!) \\ &= \log \lambda \sum_i x_i - n\lambda - \sum_i \log(x_i!)\end{aligned}$$

Take the derivative with respect to  $\lambda$ , we have

$$\begin{aligned}\frac{1}{\lambda} \sum_i x_i - n &= 0 \\ \lambda &= \frac{1}{n} \sum_i x_i\end{aligned}$$

Submit

You have used 1 of 2 attempts

**i** Answers are displayed within the problem

## Discussion

Show Discussion

**Topic:** Unit 4 Unsupervised Learning (2 weeks) :Lecture 15. Generative Models / 5. Maximum Likelihood Estimate