

Summary

- **Language Model** = probability distribution over text = generative model for text data
- **Unigram Language Model** = **word distribution**
- **Likelihood function: $p(\mathbf{X}|\theta)$**
 - **Given $\theta \rightarrow$** which \mathbf{X} has a higher likelihood?
 - **Given $\mathbf{X} \rightarrow$** which θ maximizes $p(\mathbf{X}|\theta)$? [**ML estimate**]
- **Bayesian estimation/inference**
 - Must define a **prior: $p(\theta)$**
 - **Posterior distribution: $p(\theta|\mathbf{X}) \propto p(\mathbf{X}|\theta)p(\theta)$**
 - \rightarrow Allows for inferring any “derived value” from θ !**