Bayesian Analysis for Natural Language Processing
Lecture 1

Shay Cohen

January 28, 2013

Overview

- Bayesian analysis for NLP has been catching on since the last decade
- Before that: Bayesian analysis in NLP amounted to “MAP estimation”
- Bayesian Statistics, in general, is an approach to do Statistics
  - As opposed to classical Statistics

This class

- Discuss some of the recent advances in Bayesian NLP
- Some of the topics to touch on:
  - Bayesian analysis in general
  - Priors
  - Inference (sampling, variational, etc.)
  - Bayesian NLP models (generative models, nonparametric models, etc.)
  - Other things you request or want to read
- Prerequisites: probability, basic statistical principles and some general knowledge of NLP.

Administrativia

- Class schedule: Mondays, 10:10-12:00
- Things to do in the seminar:
  - Read papers / other material
  - Lead paper discussions
  - Participate in discussions
  - White paper (maybe? probably not.)
- Office hours: right after class (or email me)

Administrativia

- Please look around for three papers that you want to read here
- Send them by email to me
- You will lead a discussion on one or more of these papers
- You can use slides if you feel better supported this way

Homework for next class

- I will give a manuscript about Bayesian priors in NLP
- You should read it and send me in email by Saturday, 10pm:
  - At least two-three questions that you have about the material (more are welcome); or
  - Points that you noticed about the topic and you think others should be aware of.
- We will discuss these in class

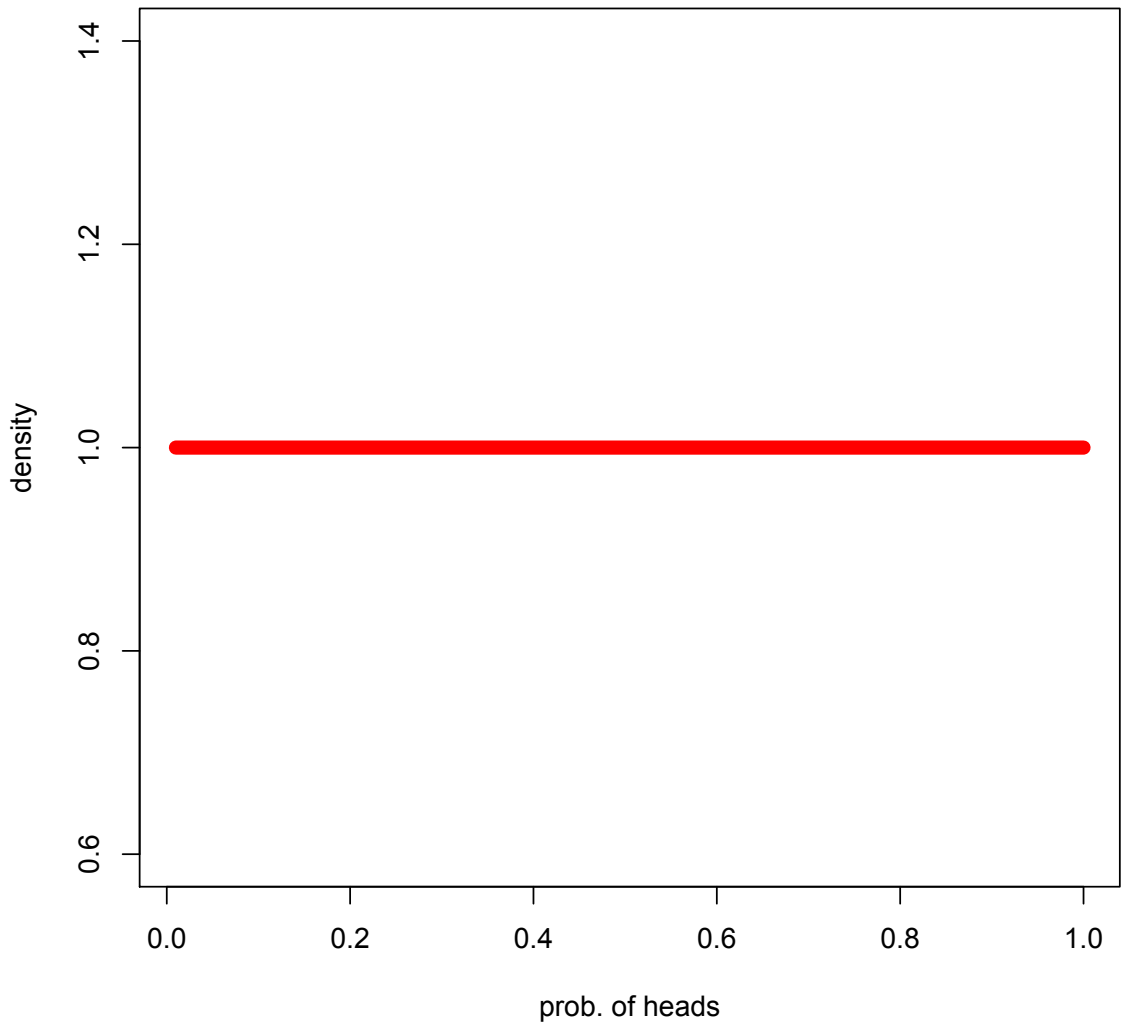
Topics for today

- What is the main idea about the Bayesian approach?
- Bayes' theorem and its use in Bayesian inference
- Bayesian updating
- Bayesian decision theory
- Hidden variables
- Maximum likelihood and maximum aposteriori estimation

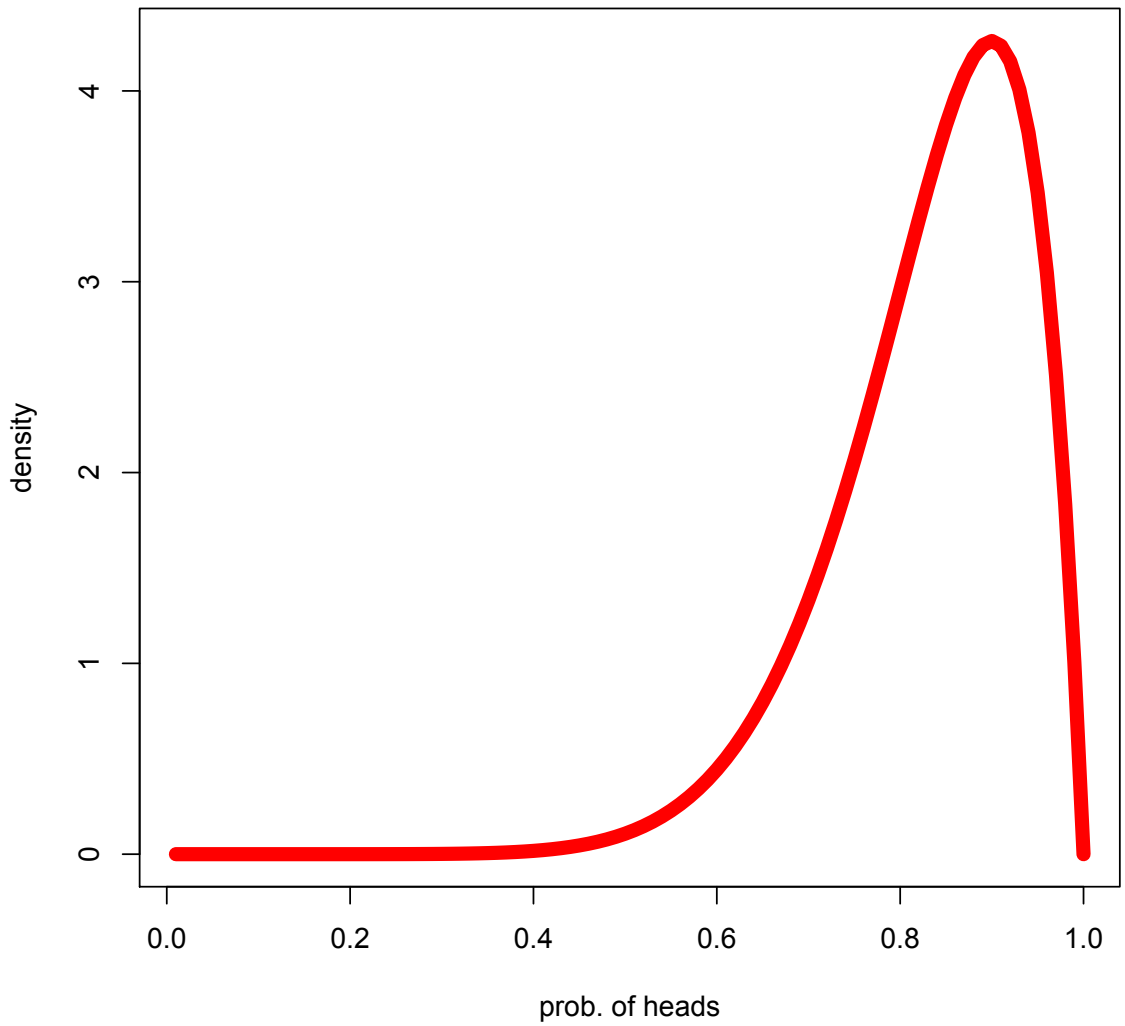
In general, today's goal is to play with Bayes' theorem in many ways!

Feel free to interrupt to ask questions!

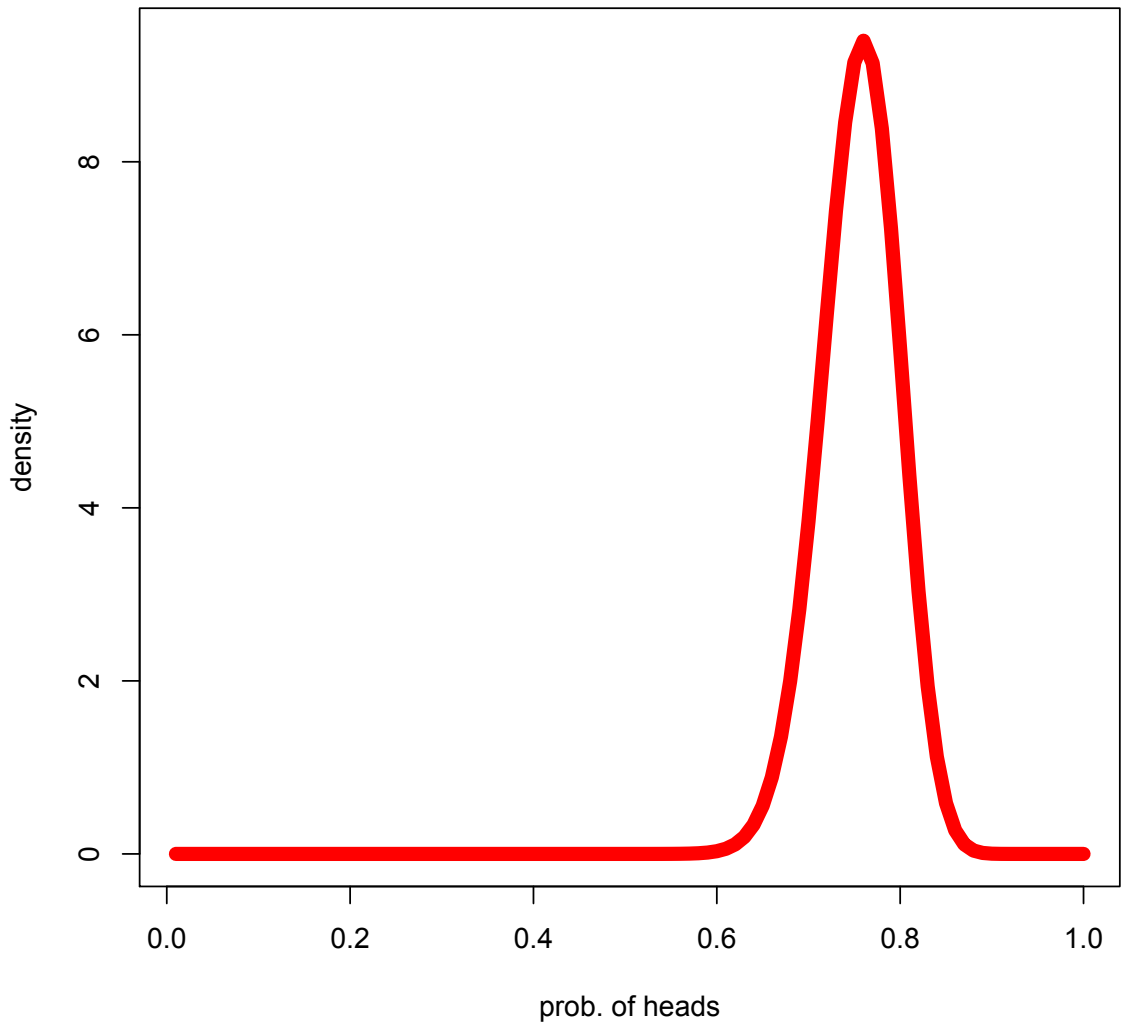
Uniform prior, coin with 0.7 prob. for heads



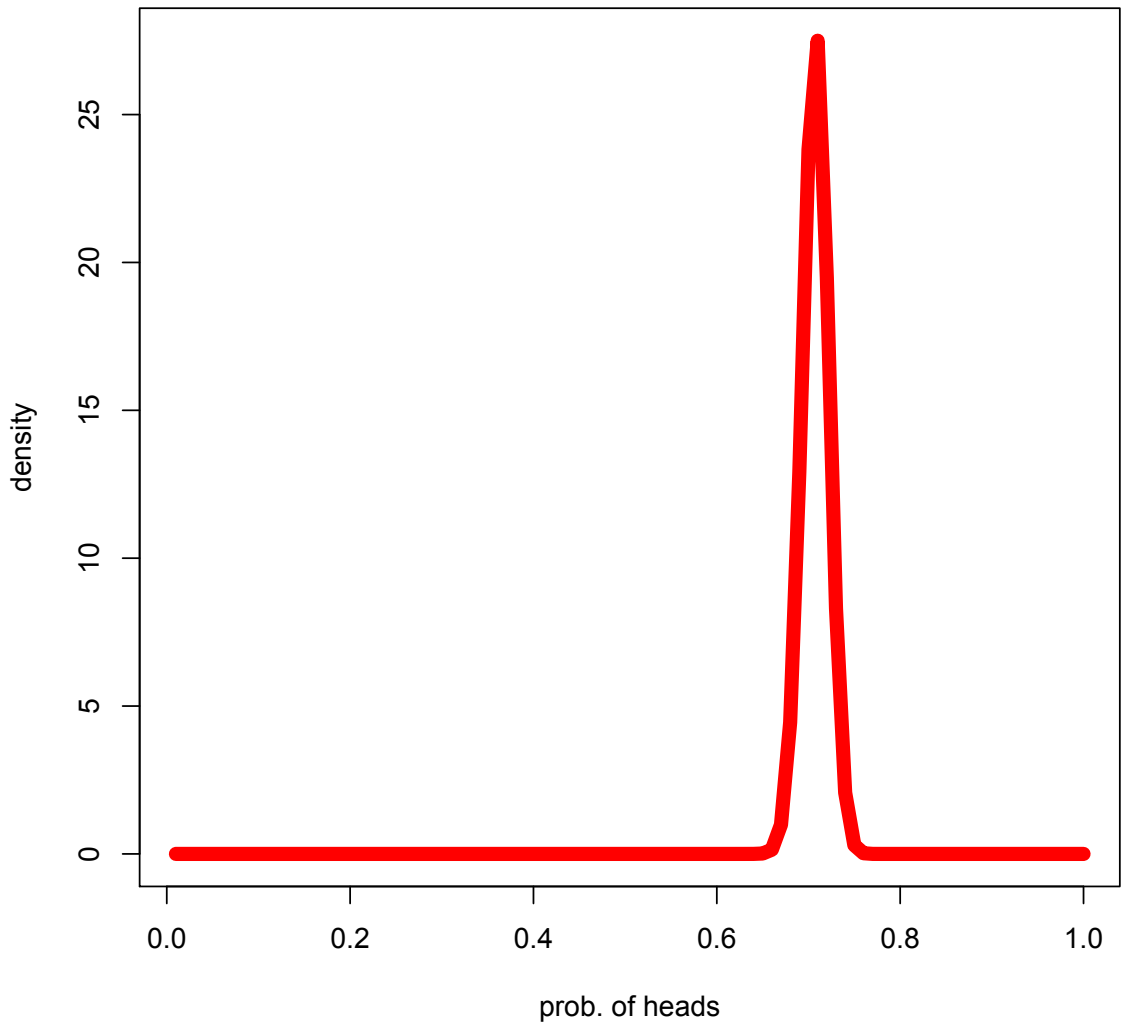
Posterior after 10 tosses, coin with 0.7 prob. for heads



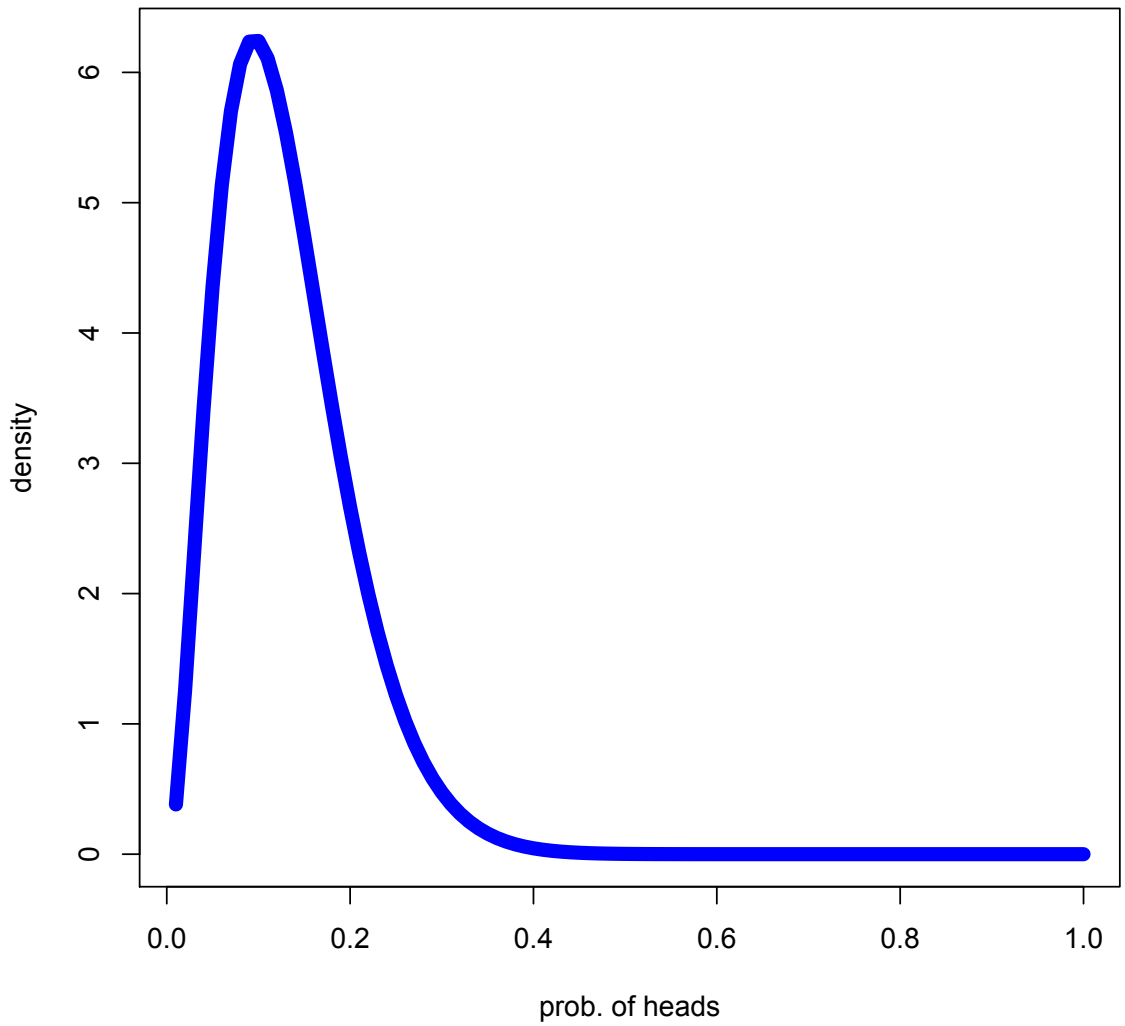
Posterior after 100 tosses, coin with 0.7 prob. for heads



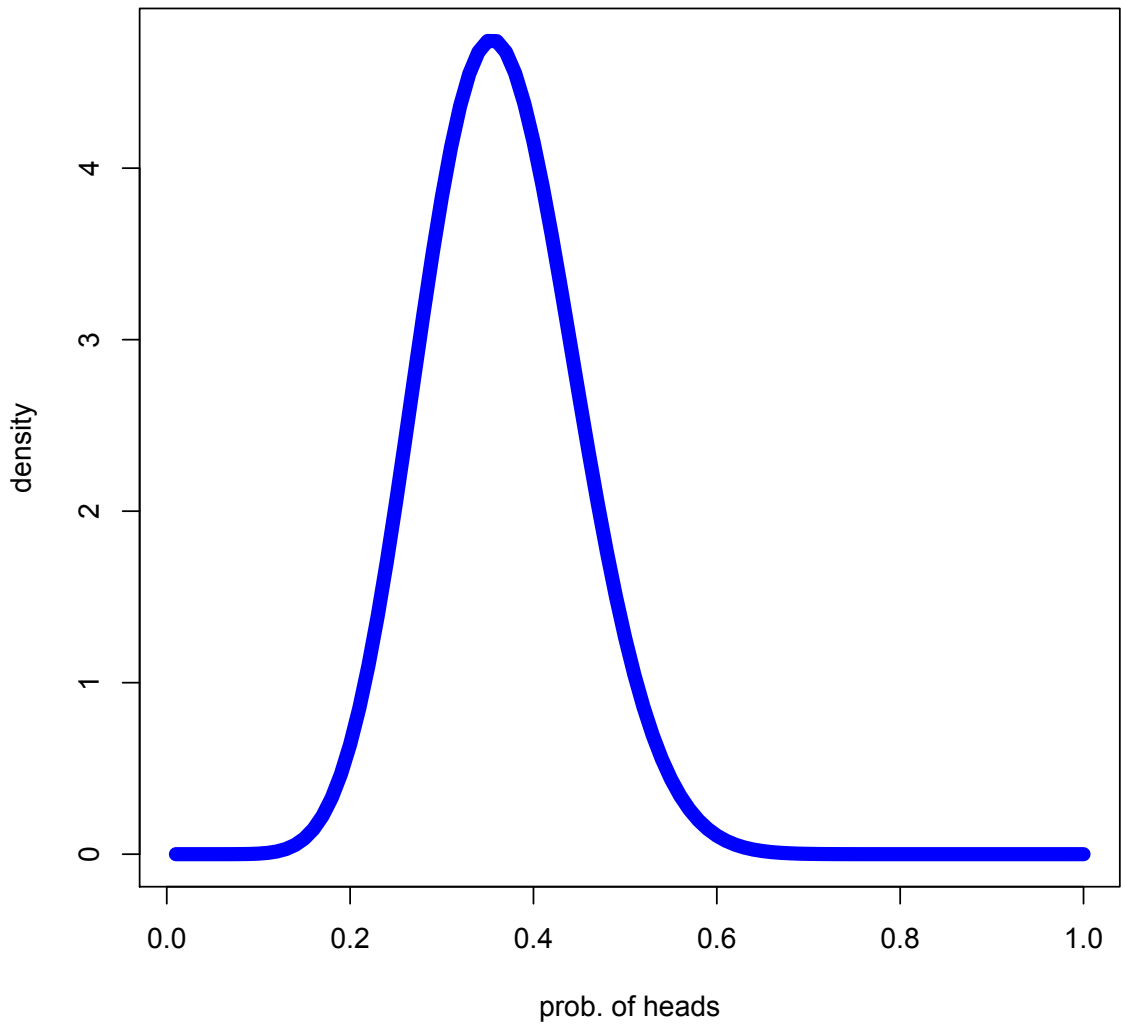
Posterior after 1000 tosses, coin with 0.7 prob. for heads



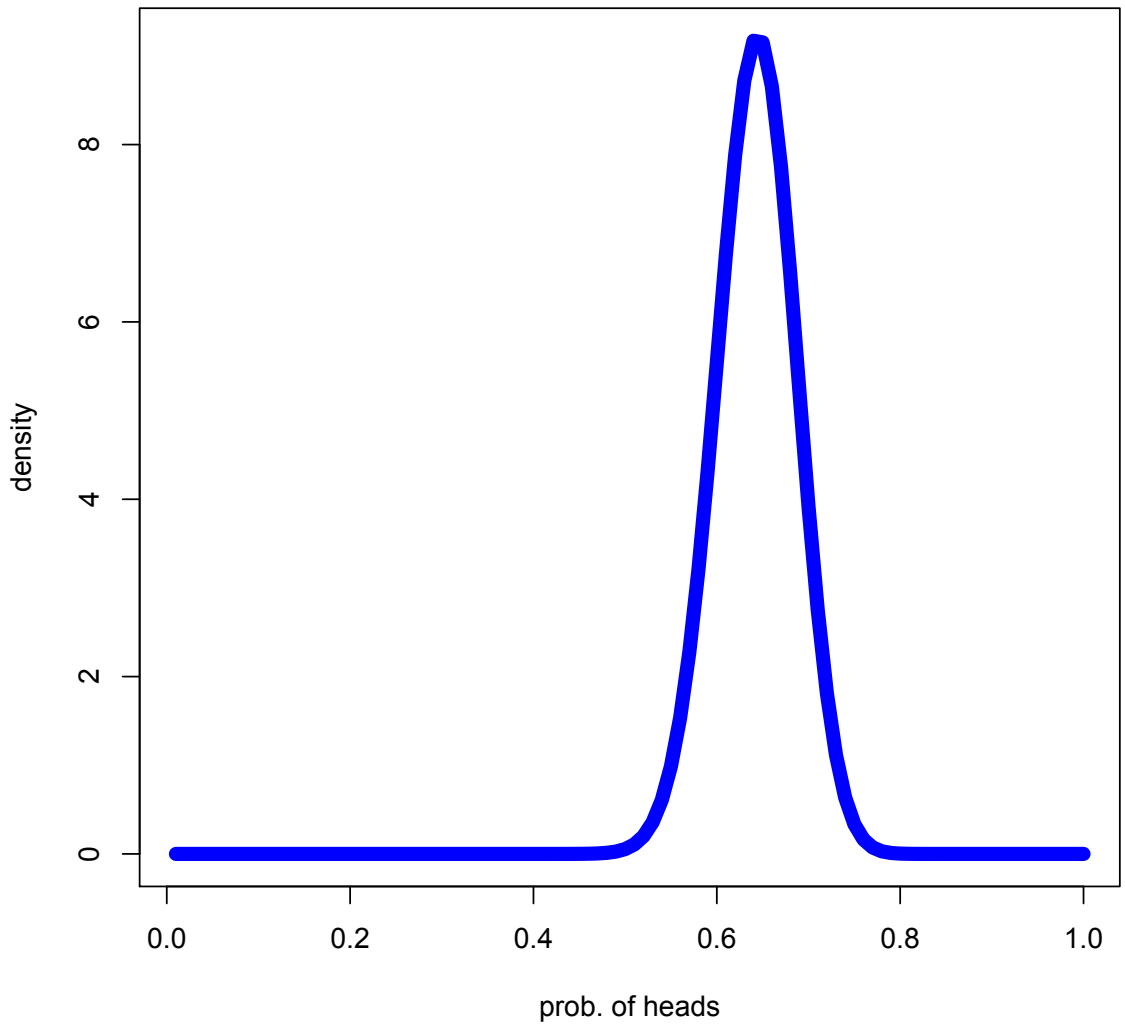
Non-uniform prior, coin with 0.7 prob. for heads



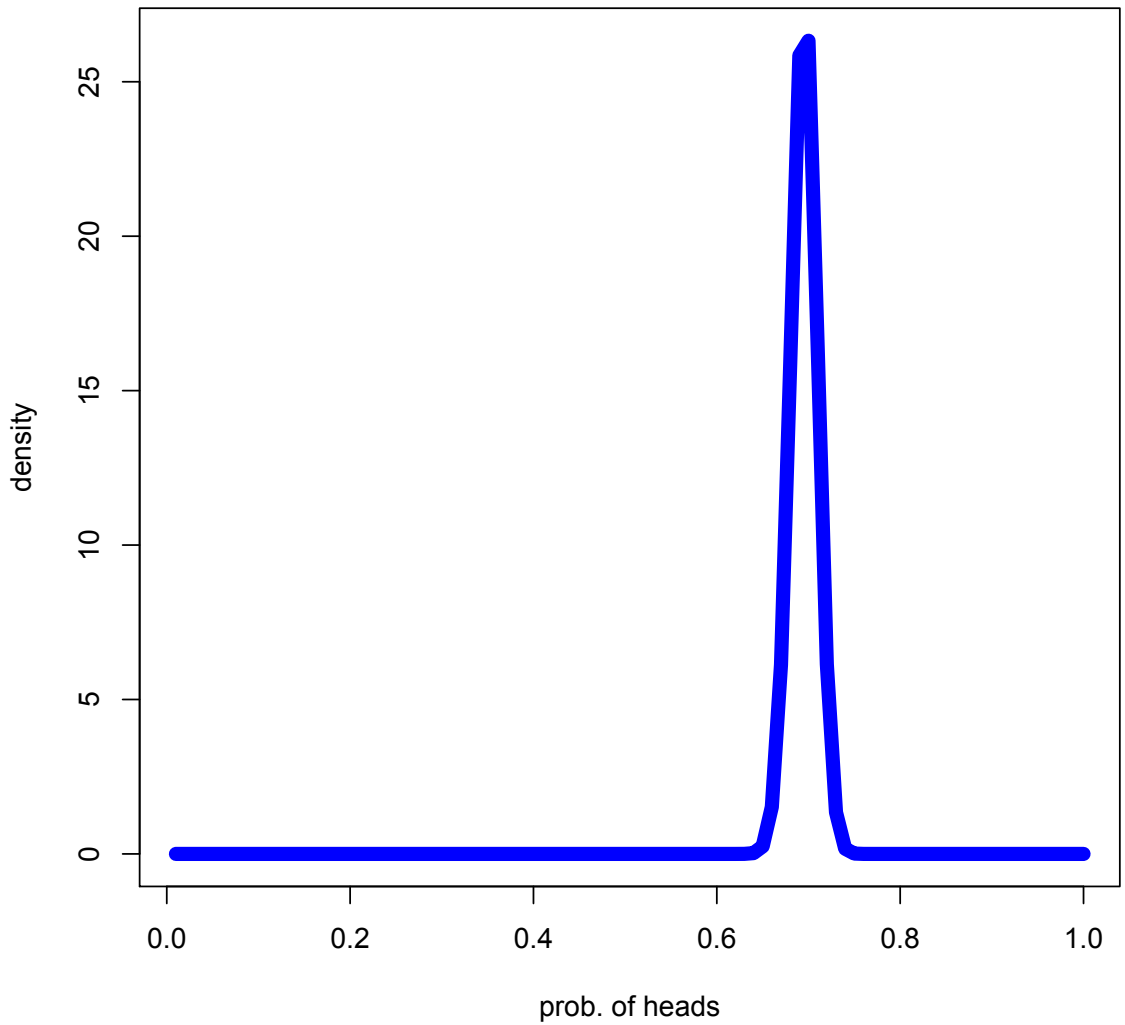
Posterior after 10 tosses, coin with 0.7 prob. for heads



Posterior after 100 tosses, coin with 0.7 prob. for heads



Posterior after 1000 tosses, coin with 0.7 prob. for heads



Modeling with latent variables

Why is Bayesian statistics now often used with incomplete data in NLP?

- Discriminative models do best in the supervised case
- Priors play much more important role in the unsupervised case
- Parameters are latent variables, easy to add more latent variables

Summary

Advantages of the Bayesian approach:

- Mananging uncertainty over the parameters as a distribution - diversity
- Simple theory, elegant inference (in theory!)
- Incorporate prior knowledge through the prior distribution

“Disadvantage”: *always* need to pick a prior

- Almost...