# Extending graphical models

ISTA 410 / INFO 510: Bayesian Modeling and Inference

U. of Arizona School of Information

November 16, 2020

# Beta-binomial model in the news

Before we go on: recent news!

- Pfizer COVID-19 vaccine
- Moderna COVID-19 vaccine

Pfizer's claim: "at least 90% effective" (I also saw some news outlets report "up to 90% effective")
Moderna's claim: "94.5% effective"

# Beta-binomial model

Defining parameters:

- $\pi_c$: probability that a control subject becomes ill
- $\pi_v$: probability that a vaccinated subject becomes ill
- Derived quantity: Vaccine efficacy:

$$VE = 1 - \frac{\pi_v}{\pi_c}$$

Parameter for the model:

$$\theta = \frac{1 - VE}{2 - VE} = \frac{\pi_v}{\pi_v + \pi_c}$$

# Pfizer's prior

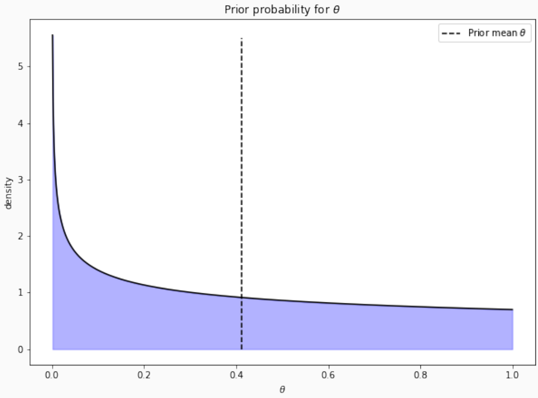
Let $y$ be the number of cases that come from the vaccinated group.

The model:

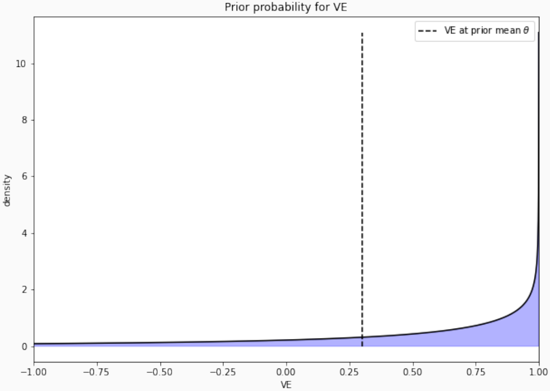
$$y \sim \mathrm{Binomial}(\theta, n)$$
$$\theta \sim \mathrm{Beta}(0.700102, 1)$$

Prior is stated in Pfizer's press release. Appears chosen so that the VE at prior mean $\theta$ is 30%. Prior 95% interval for VE is about (-26.2, 0.995).

# Pfizer’s prior



# Pfizer's prior



# What's the data?

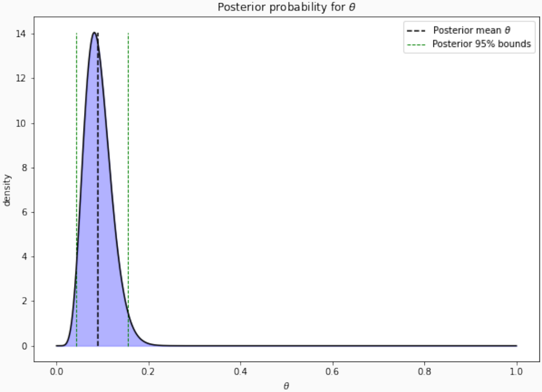
The Pfizer press release didn't state the number of cases from each group, but stated the overall number of cases as 94. So, we'd have to reverse-engineer the number of cases from the vaccine arm:

- If we interpret 90% as the sample efficacy, at most 8 cases came from the vaccine arm

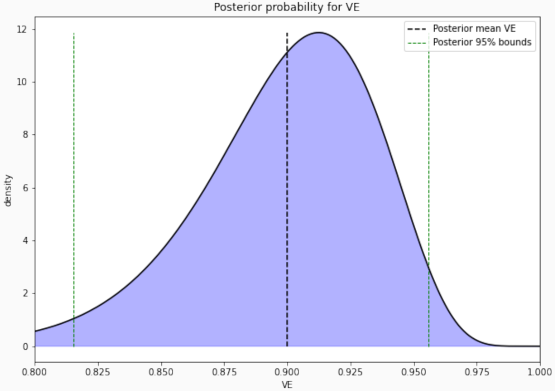
Then the posterior for $\theta$ is

$$p(\theta|y) = \mathrm{Beta}(0.700102 + 8, 1 + 86)$$

# Posterior distribution



# Posterior distribution



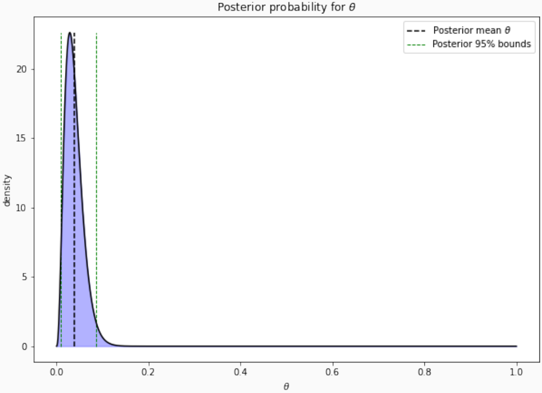
# Another reverse-engineering

Another interpretation of the 90% figure:

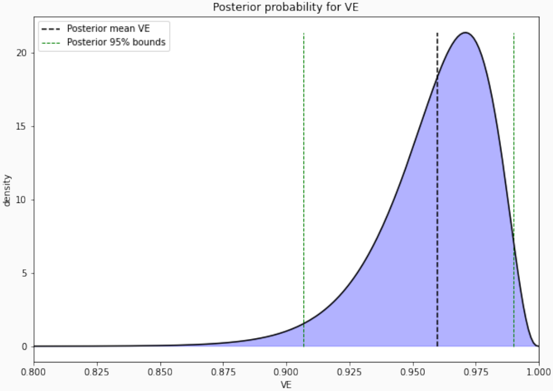
- The 95% credible interval for vaccine effectiveness lies entirely above 90%

This is a more optimistic interpretation – for this, we need at most 3 cases in the vaccine arm.

# Posterior distribution



# Posterior distribution



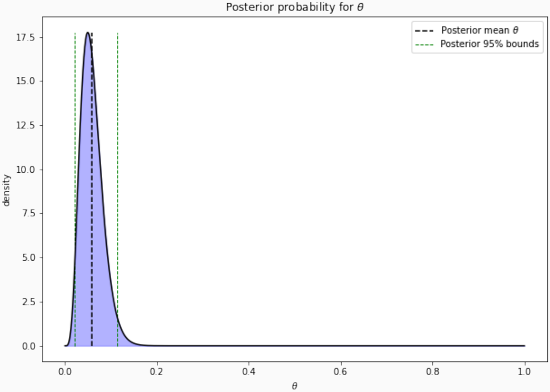
## Moderna's vaccine

We don't have to do as much reverse engineering for Moderna's data, because they published case counts:

- 90 cases in the placebo arm
- 5 cases in the vaccine arm

However, the press release doesn't appear to indicate the type of analysis that resulted in the 94.5% figure.

# Posterior distribution



# Posterior distribution

## Further reading

A few sources:

- "A look at Biontech/Pfizer's Bayesian analysis of their Covid-19 vaccine trial" skranz.github.io//r/2020/11/11/CovidVaccineBayesian.html
- "The Pfizer-Biontech Vaccine May Be A Lot More Effective Than You Think" http://blog.fellstat.com/?p=440
- "How to describe Pfizer's beta(0.7, 1) prior on vaccine effect?" https://statmodeling.stat.columbia.edu/2020/11/13/pfizer-beta-prior-vaccine-effect/

# Graphical models, mixtures, plate notation

# Bayesian networks

A *Bayesian network* is a probabilistic model based on a DAG.

- DAG represents relationships between variables
- Each variable equipped with a probability distribution conditional on its parents in the graph
- DAG implies certain factorization / conditional independence properties

# Recap: directed graphs

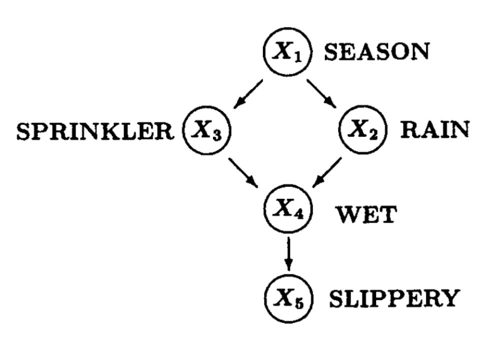



Figure from *Causality*

# Recap: directed graphs

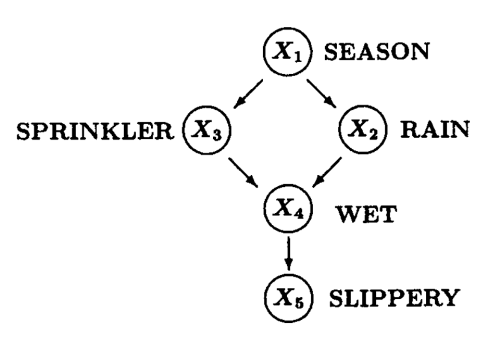



Figure from *Causality*

Joint probability distribution

$$P(x_1, \ldots, x_5) = P(x_1)P(x_2|x_1)P(x_3|x_1)P(x_4|x_2, x_3)P(x_5|x_4)$$

# Plate models

Plate models: a formalism for specifying graphical / hierarchical models with multiple observations / groups

- Similar to a directed graph, but groups repeated structures into "plates"
- Plates indicated by rectangles surrounding variables
- Can indicate observable/latent variables by shading

## Example: election forecast model

As an example, we can write our hierarchical election forecasting model in plate notation:

# Example: election forecast model

As an example, we can write our hierarchical election forecasting model in plate notation:

Joint distribution:

$$p(y, x, \delta, \tau) = p(\tau) \prod_{i=1}^{11} p(\delta_i|\tau) \prod_{i=1}^{5} 0p(y_i|x, \delta)$$

# Expanding to the full model graph

# Example: latent Dirichlet allocation

Example in practice: latent Dirichlet allocation (Blei et al, 2003)

- Model for topics and words in documents
- A document is modeled as a sequence of words; each word is associated with a topic
- Idea: each document has an underlying distribution of topics, and each word is chosen from a randomly selected topic

# Data generating process as an algorithm

A document is represented as a vector of words $\mathrm{w} = w_i$.

Steps for generating a document:

1. Choose $\theta \sim \mathrm{Dirichlet}(\alpha)$.
2. For each word:
   2.1 Choose a topic $z \sim \mathrm{Multinomial}(\theta)$
   2.2 Choose a word $w \sim \mathrm{Multinomial}(\beta|\mathrm{z})$

# Writing down the model

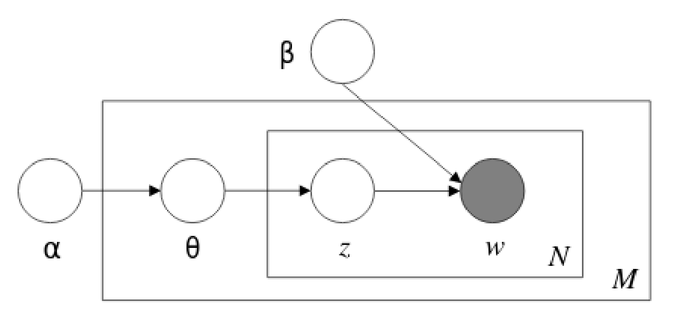
A document is represented as a vector of words w $= w_i$.

$$
\begin{aligned}
w_i &\sim \mathrm{Multinomial}(\beta_z) \\
z &\sim \mathrm{Multinomial}(\theta) \\
\theta &\sim \mathrm{Dirichlet}(\alpha)
\end{aligned}
$$

(priors omitted for now)

# Example: latent Dirichlet allocation

In plate notation:

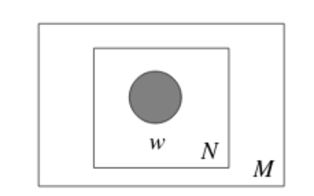


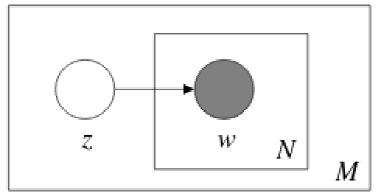
## Comparison to other models

Two simpler models:

- Pure unigram model
    - Each word drawn independently from underlying distribution of words
    - No topic variable
- Mixture of unigrams
    - Documentwise topic variable
    - Each word drawn from a distribution of words according to topic

# Comparison to other models



# Summary

Today:

- Some current events
- Plate notation, LDA example

Going forward:

- More on the LDA example
- Temporal models