Introduction to Bayesian Data Analysis

Xuweiyi Chen

UNIVERSITY of WASHINGTON

What is Bayesian Statistics?

Suppose we have y as a data vector, ϴ is the vector for parameters of model, then we have

- $L(y \mid \Theta)$ Likelihood
- $\pi(\Theta)$ Prior
- $\pi(\Theta \mid y) \propto L(y \mid \Theta)\, \pi(\Theta)$ Posterior

Difference between frequentist statistics and Bayesian statistics

Frequentist:

- confidence interval
- point estimation
- p-value, power,
- significance

Bayesian:

- credible interval
- Bayes Factor
- prior
- posterior

Example

Suppose that during a recent doctor's visit, you tested positive for a very rare disease. If you only get to ask the doctor one question, which would it be?

a. What's the chance that I actually have the disease?

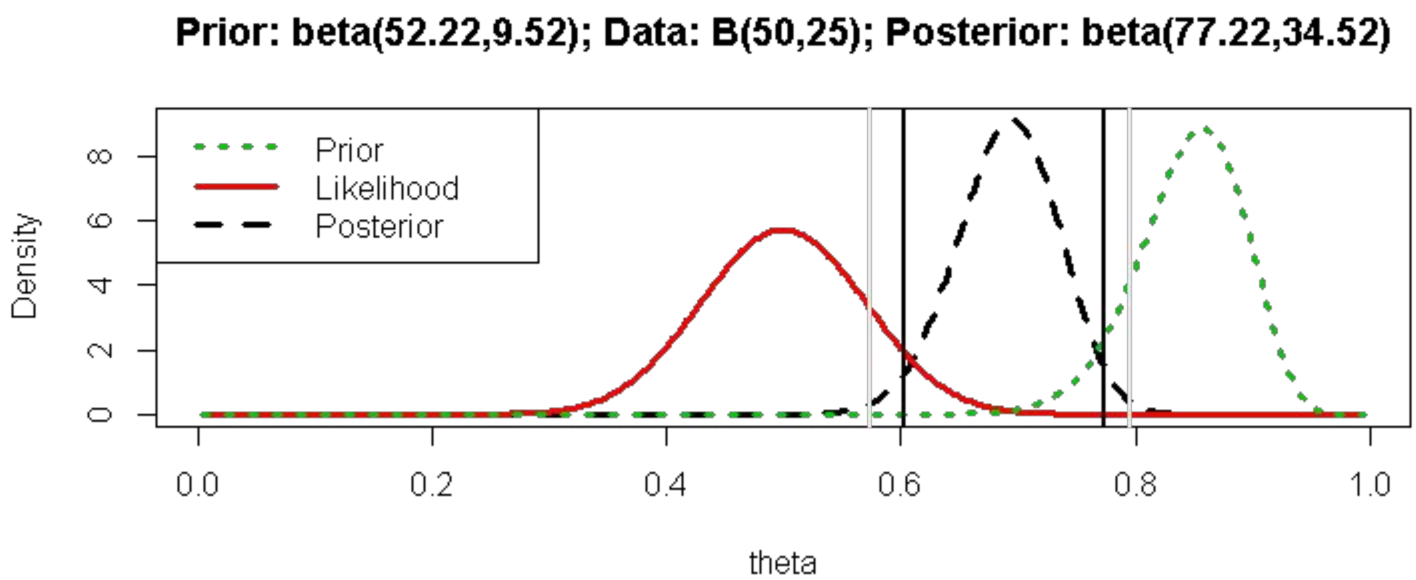
b. If in fact I don't have the disease, what's the chance that I would've gotten this positive test result?

TABLE 1.1: Disease status and test outcomes for 100 people.

	test positive	test negative	total
disease	3	1	4
no disease	9	87	96
total	12	88	100

Borrowed from
Bayes rules book

Example



Borrowed from Wesley

MCMC sampling

In order to sample from the posterior :

$$\pi(\Theta \mid y) \propto L(y \mid \Theta)\, \pi(\Theta)$$

- Metropolis-Hastings
- Gibbs Resampling

Data

We used data from a kaggle challenge: Twitter tweets data to do sentiment analysis

34	0	it was a hard monday due to cloudy weather. disabling oxygen production for today. #goodnight #badmo...
35	1	it's unbelievable that in the 21st century we'd need something like this. again. #neverump #xenopho...
36	0	#taylorswift1989 bull up: you will dominate your bull and you will direct it whatever you want it ...
37	0	morning~~ #travelingram #dalat #ripinkylife

Brief Introduction on Topic Model

In machine learning and natural language processing, a topic model is a type of statistical model for discovering the abstract "topics" that occur in a collection of documents.

Latent Dirichlet Allocation

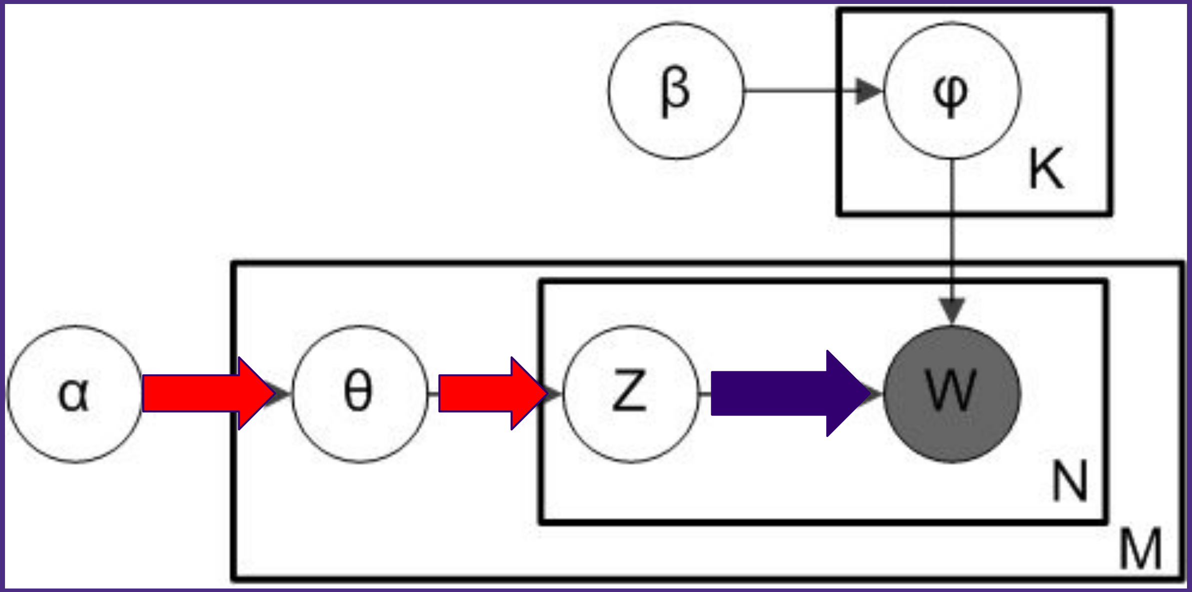
Latent Dirichlet allocation (LDA) is a generative probabilistic model of a corpus. The basic idea is that documents are represented as random mixtures over latent topics, where each topic is characterized by a distribution over words.[1]

LDA assumes the following generative process for each document $\mathbf{w}$ in a corpus D:

1. Choose $N \sim \text{Poisson}(\xi)$.
2. Choose $\theta \sim \text{Dir}(\alpha)$.
3. For each of the N words w_n:
   (a) Choose a topic $z_n \sim \text{Multinomial}(\theta)$.
   (b) Choose a word w_n from $p(w_n \mid z_n, \beta)$, a multinomial probability conditioned on the topic z_n.

Borrowed from original Paper

LDA Diagram



α is the per-document topic distributions,

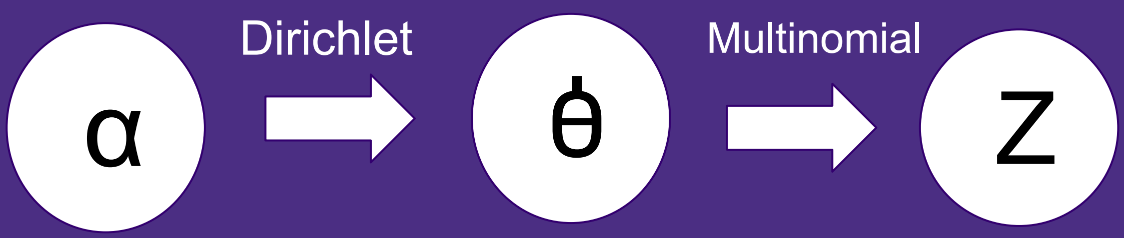
θ is the topic distribution for document m,

z is the topic for the n-th word in document m

w is the specific word

Source from
Tyler Doll

Document to Topic for each word



α is the per-document topic distributions,

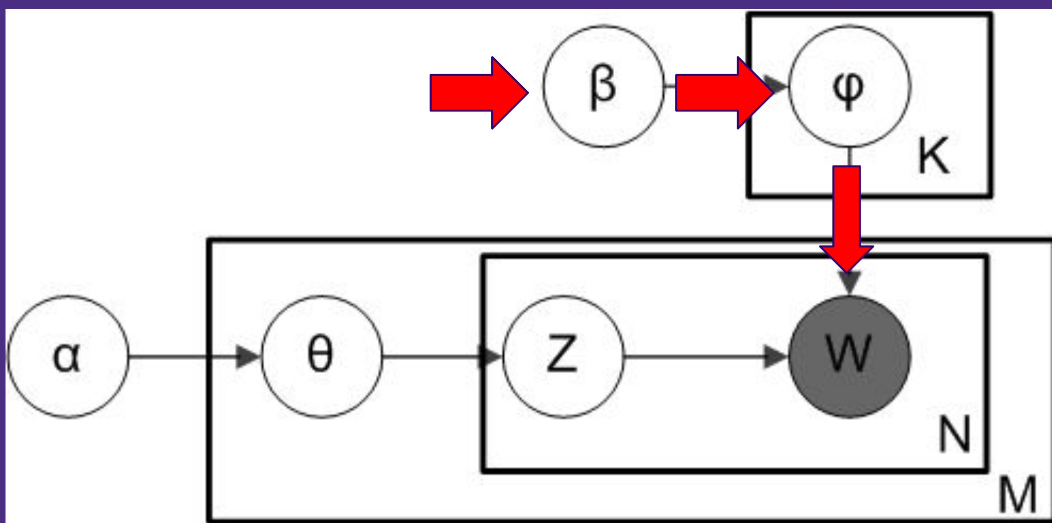
θ is the topic distribution for document m,

z is the topic for the n-th word in document m

w is the specific word

topic for word z: k

LDA Diagram



β is the per-topic word distribution,

φ is the word distribution for topic k,

w is the specific word

Source from Tyler Doll

Topic generate each word based on k

$$\beta \xrightarrow{\text{Dirichlet}} \varphi \xrightarrow{\text{Multinomial}} w$$

β is the per-topic word distribution,

φ is the word distribution for topic k,

w is the specific word

Simulation with Rstan

Marginal Posterior:

α is the per-document topic distributions,

β is the per-topic word distribution,

θ is the topic distribution for document m,

φ is the word distribution for topic k,

z is the topic for the n-th word in document m

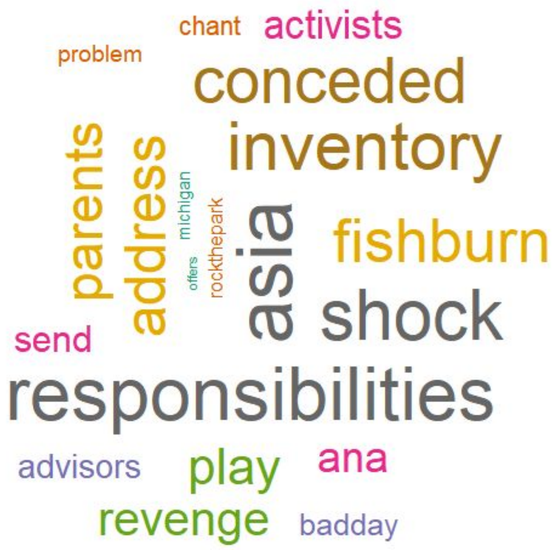
w is the specific word

$$
\begin{aligned}
p(\theta, \phi | w, \alpha, \beta) &\propto p(\theta|\alpha)p(\phi|\beta)p(w|\theta, \phi) \\
&= \prod_{m=1}^{M} p(\theta_m|\alpha) * \prod_{k=1}^{K} p(\phi_k|\beta) * \prod_{m=1}^{M} \prod_{n=1}^{M[n]} p(w_{m,n}|\theta_m, \phi).
\end{aligned}
$$

Source from Stan

Result

Label 0:



Result

label 1: