Why topic models?

- Suppose you have a huge number of documents
- Want to know what’s going on
- Can’t read them all (e.g. every New York Times article from the 90’s)
- Topic models offer a way to get a corpus-level view of major themes
Why topic models?

- Suppose you have a huge number of documents
- Want to know what’s going on
- Can’t read them all (e.g. every New York Times article from the 90’s)
- Topic models offer a way to get a corpus-level view of major themes
- Unsupervised
Roadmap

- What are topic models
- How to know if you have good topic model
- How to go from raw data to topics
What are Topic Models?

Conceptual Approach

From an **input corpus** and number of topics $K \rightarrow$ words to topics

![Diagram showing a corpus with topics extracted from various documents.]

- Forget the Bootleg, Just Download the Movie Legally
- Multiplex Heralded As Linchpin To Growth The Shape of Cinema, Transformed At the Click of a Mouse
- A Peaceful Crew Puts Muppets Where Its Mouth Is
- Stock Trades: A Better Deal
- The three big Internet
- Red Light, Green Light: A 2-Tone L.E.D. to Simplify Screens
What are Topic Models?

Conceptual Approach

From an input corpus and number of topics $K \rightarrow \text{words to topics}$

**TOPIC 1**
- computer, technology, system, service, site, phone, internet, machine

**TOPIC 2**
- sell, sale, store, product, business, advertising, market, consumer

**TOPIC 3**
- play, film, movie, theater, production, star, director, stage

Material adapted from David Mimno | UMD
What are Topic Models?

Conceptual Approach

- For each document, what topics are expressed by that document?

Material adapted from David Mimno | UMD
What are Topic Models?

Topics from Science

human  evolution  disease  computer
genome evolutionary host models
dna species bacteria information
genetic organisms bacteria data
genes life resistance computers
sequence origin bacterial system
gene biology new network
molecular groups strains model
sequencing phylogenetic control systems
map living infectious parallel
information diversity malaria methods
genetics group parasite networks
mapping new parasites software
project two united new
sequences common tuberculosis simulations
What are Topic Models?

Why should you care?

- Neat way to explore / understand corpus collections
  - E-discovery
  - Social media
  - Scientific data
- NLP Applications
  - Word Sense Disambiguation
  - Discourse Segmentation
  - Machine Translation
- Psychology: word meaning, polysemy
- Inference is (relatively) simple
What are Topic Models?

Matrix Factorization Approach

\[
\begin{bmatrix}
M 	imes K
\end{bmatrix}
\times
\begin{bmatrix}
K 	imes V
\end{bmatrix}
\approx
\begin{bmatrix}
M 	imes V
\end{bmatrix}
\]

- **M** Number of documents
- **K** Number of topics
- **V** Size of vocabulary

If you use singular value decomposition (SVD), this technique is called latent semantic analysis. Popular in information retrieval.

Material adapted from David Mimno | UMD
What are Topic Models?

**Matrix Factorization Approach**

\[
\begin{bmatrix}
M 	imes K
\end{bmatrix}
\times
\begin{bmatrix}
K 	imes V
\end{bmatrix}
\approx
\begin{bmatrix}
M 	imes V
\end{bmatrix}
\]

- Topic Assignment
- Topics
- Dataset

- If you use singular value decomposition (SVD), this technique is called latent semantic analysis.
- Popular in information retrieval.

**K** Number of topics

**M** Number of documents

**V** Size of vocabulary
Alternative: Generative Model

- How your data came to be
- Sequence of Probabilistic Steps
- Posterior Inference
Alternative: Generative Model

- How your data came to be
- Sequence of Probabilistic Steps
- Posterior Inference
What are Topic Models?

Multinomial Distribution

- Distribution over discrete outcomes
- Represented by non-negative vector that sums to one
- Picture representation

(1,0,0) (0,0,1) (0,1,0)
(1/2,1/2,0) (1/3,1/3,1/3) (1/4,1/4,1/2)
(0,1,0)
What are Topic Models?

**Multinomial Distribution**

- Distribution over discrete outcomes
- Represented by non-negative vector that sums to one
- Picture representation

- Come from a Dirichlet distribution

Material adapted from David Mimno | UMD

Topic Models | 10 / 51
What are Topic Models?

Dirichlet Distribution

\[
P(p | \alpha m) = \frac{\Gamma(\sum_k \alpha m_k)}{\prod_k \Gamma(\alpha m_k)} \prod_k p_k^{\alpha m_k - 1}
\]
What are Topic Models?

Dirichlet Distribution

\[ P(p | \alpha m) = \frac{\Gamma(\sum_k \alpha m_k)}{\prod_k \Gamma(\alpha m_k)} \prod_k p_k^{\alpha m_k - 1} \]

\[ \alpha = 3, \quad m = \left( \frac{1}{3}, \frac{1}{3}, \frac{1}{3} \right) \]

\[ \alpha = 6, \quad m = \left( \frac{1}{3}, \frac{1}{3}, \frac{1}{3} \right) \]

\[ \alpha = 30, \quad m = \left( \frac{1}{3}, \frac{1}{3}, \frac{1}{3} \right) \]
What are Topic Models?

Dirichlet Distribution

\[ P(p | \alpha m) = \frac{\Gamma(\sum_k \alpha m_k)}{\prod_k \Gamma(\alpha m_k)} \prod_k p_k^{\alpha m_k - 1} \]

\[ \alpha = 3, \ m = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3}) \]
\[ \alpha = 6, \ m = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3}) \]
\[ \alpha = 30, \ m = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3}) \]

\[ \alpha = 14, \ m = (\frac{1}{7}, \frac{5}{7}, \frac{1}{7}) \]
\[ \alpha = 14, \ m = (\frac{1}{7}, \frac{1}{7}, \frac{5}{7}) \]
\[ \alpha = 2.7, \ m = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3}) \]
What are Topic Models?

Dirichlet Distribution

\[
\alpha = (0.2, 0.1, 0.1)
\]
What are Topic Models?

**Dirichlet Distribution**

- If \( \vec{\phi} \sim \text{Dir}(\alpha) \), \( \vec{w} \sim \text{Mult}(\phi) \), and \( n_k = |\{w_i : w_i = k\}| \) then

\[
p(\phi | \alpha, \vec{w}) \propto p(\vec{w} | \phi) p(\phi | \alpha) \tag{1}
\]

\[
\propto \prod_k \phi^{n_k} \prod_k \phi^{\alpha_k-1} \tag{2}
\]

\[
\propto \prod_k \phi^{\alpha_k+n_k-1} \tag{3}
\]

- Conjugacy: this **posterior** has the same form as the **prior**
What are Topic Models?

**Dirichlet Distribution**

- If $\vec{\phi} \sim \text{Dir}(\alpha)$, $\vec{w} \sim \text{Mult}(\phi)$, and $n_k = |\{w_i : w_i = k\}|$ then

$$p(\phi | \alpha, \vec{w}) \propto p(\vec{w} | \phi) p(\phi | \alpha)$$  \hspace{1cm} (1)

$$\propto \prod_k \phi^{n_k} \prod_k \phi^{\alpha_k - 1}$$  \hspace{1cm} (2)

$$\propto \prod_k \phi^{\alpha_k + n_k - 1}$$  \hspace{1cm} (3)

- Conjugacy: this **posterior** has the same form as the **prior**
What are Topic Models?

Generative Model

**TOPIC 1**
- computer,
- technology,
- system,
- service, site,
- phone,
- internet,
- machine

**TOPIC 2**
- sell, sale,
- store, product,
- business,
- advertising,
- market,
- consumer

**TOPIC 3**
- play, film,
- movie, theater,
- production,
- star, director,
- stage

Material adapted from David Mimno | UMD
What are Topic Models?

Generative Model

The three big Internet portals begin to distinguish among themselves as shopping malls

Stock Trades: A Better Deal For Investors Isn't Simple

Red Light, Green Light: A 2-Tone L.E.D. to Simplify Screens

Forget the Bootleg, Just Download the Movie Legally

The Shape of Cinema, Transformed At the Click of a Mouse

Multiplex Heralded As Linchpin To Growth

A Peaceful Crew Puts Muppets Where Its Mouth Is

Material adapted from David Mimno | UMD
Hollywood studios are preparing to let people download and buy electronic copies of movies over the Internet, much as record labels now sell songs for 99 cents through Apple Computer's iTunes music store and other online services ...
Hollywood studios are preparing to let people download and buy electronic copies of movies over the Internet, much as record labels now sell songs for 99 cents through Apple Computer's iTunes music store and other online services ...
Hollywood studios are preparing to let people download and buy electronic copies of movies over the Internet, much as record labels now sell songs for 99 cents through Apple Computer's iTunes music store and other online services ...
Hollywood studios are preparing to let people download and buy electronic copies of movies over the Internet, much as record labels now sell songs for 99 cents through Apple Computer's iTunes music store and other online services ...
For each topic $k \in \{1, \ldots, K\}$, draw a multinomial distribution $\beta_k$ from a Dirichlet distribution with parameter $\lambda$. 
What are Topic Models?

Generative Model Approach

For each topic $k \in \{1, \ldots, K\}$, draw a multinomial distribution $\beta_k$ from a Dirichlet distribution with parameter $\lambda$.

For each document $d \in \{1, \ldots, M\}$, draw a multinomial distribution $\theta_d$ from a Dirichlet distribution with parameter $\alpha$.
Generative Model Approach

- For each topic $k \in \{1, \ldots, K\}$, draw a multinomial distribution $\beta_k$ from a Dirichlet distribution with parameter $\lambda$.
- For each document $d \in \{1, \ldots, M\}$, draw a multinomial distribution $\theta_d$ from a Dirichlet distribution with parameter $\alpha$.
- For each word position $n \in \{1, \ldots, N\}$, select a hidden topic $z_n$ from the multinomial distribution parameterized by $\theta$. 

Material adapted from David Mimno | UMD
Generative Model Approach

- For each topic $k \in \{1, \ldots, K\}$, draw a multinomial distribution $\beta_k$ from a Dirichlet distribution with parameter $\lambda$.
- For each document $d \in \{1, \ldots, M\}$, draw a multinomial distribution $\theta_d$ from a Dirichlet distribution with parameter $\alpha$.
- For each word position $n \in \{1, \ldots, N\}$, select a hidden topic $z_n$ from the multinomial distribution parameterized by $\theta$.
- Choose the observed word $w_n$ from the distribution $\beta_{z_n}$.
What are Topic Models?

Generative Model Approach

For each topic \( k \in \{1, \ldots, K\} \), draw a multinomial distribution \( \beta_k \) from a Dirichlet distribution with parameter \( \lambda \).

For each document \( d \in \{1, \ldots, M\} \), draw a multinomial distribution \( \theta_d \) from a Dirichlet distribution with parameter \( \alpha \).

For each word position \( n \in \{1, \ldots, N\} \), select a hidden topic \( z_n \) from the multinomial distribution parameterized by \( \theta \).

Choose the observed word \( w_n \) from the distribution \( \beta_{z_n} \).

We use statistical inference to uncover the most likely underlying topics.
What are Topic Models?

Topic Models: What’s Important

- Topic models
  - Topics to word types
  - Documents to topics
  - Topics to word types—multinomial distribution
  - Documents to topics—multinomial distribution

- Focus in this talk: statistical methods
  - Model: story of how your data came to be
  - Latent variables: missing pieces of your story
  - Statistical inference: filling in those missing pieces

- We use latent Dirichlet allocation (LDA), a fully Bayesian version of pLSI, probabilistic version of LSA
What are Topic Models?

**Topic Models: What’s Important**

- **Topic models (latent variables)**
  - Topics to word types
  - Documents to topics
  - Topics to word types—multinomial distribution
  - Documents to topics—multinomial distribution

- **Focus in this talk: statistical methods**
  - Model: story of how your data came to be
  - Latent variables: missing pieces of your story
  - Statistical inference: filling in those missing pieces

- **We use latent Dirichlet allocation (LDA), a fully Bayesian version of pLSI, probabilistic version of LSA**

Material adapted from David Mimno | UMD
How you compute it is important too (Wallach et al. 2009)
Measures predictive power, not what the topics are

$$P(w \mid w', z', \alpha m, \beta u) = \sum_z P(w, z \mid w', z', \alpha m, \beta u)$$

How you compute it is important too (Wallach et al. 2009)
Word Intrusion

**TOPIC 1**
- computer,
- technology,
- system,
- service, site,
- phone,
- internet,
- machine

**TOPIC 2**
- sell, sale,
- store, product,
- business,
- advertising,
- market,
- consumer

**TOPIC 3**
- play, film,
- movie, theater,
- production,
- star, director,
- stage

Material adapted from David Mimno | UMD
Word Intrusion

1. Take the highest probability words from a topic

**Original Topic**

dog, cat, horse, pig, cow
**Word Intrusion**

1. Take the highest probability words from a topic

   **Original Topic**
   
   dog, cat, horse, pig, cow

2. Take a high-probability word from another topic and add it

   **Topic with Intruder**
   
   dog, cat, apple, horse, pig, cow
# Word Intrusion

1. Take the highest probability words from a topic

**Original Topic**

- dog, cat, horse, pig, cow

2. Take a high-probability word from another topic and add it

**Topic with Intruder**

- dog, cat, apple, horse, pig, cow

3. We ask users to find the word that doesn’t belong

**Hypothesis**

If the topics are interpretable, users will consistently choose true intruder
Word Intrusion

1 / 10
- crash
- accident
- board
- agency
- tibetan
- safety

2 / 10
- commercial
- network
- television
- advertising
- viewer
- layoff

3 / 10
- arrest
- crime
- inmate
- pitcher
- prison
- death

4 / 10
- hospital
- doctor
- health
- care
- medical
- tradition
Word Intrusion

- **Order of words was shuffled**
- **Which intruder was selected varied**
- **Model precision:** percentage of users who clicked on intruder

<table>
<thead>
<tr>
<th>1 / 10</th>
<th>Reveal additional response</th>
</tr>
</thead>
<tbody>
<tr>
<td>crash</td>
<td>accident</td>
</tr>
<tr>
<td>board</td>
<td>agency</td>
</tr>
<tr>
<td></td>
<td>tibetan</td>
</tr>
<tr>
<td></td>
<td>safety</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2 / 10</th>
<th>commercial</th>
<th>network</th>
<th>television</th>
<th>advertising</th>
<th>viewer</th>
<th>layoff</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>3 / 10</th>
<th>arrest</th>
<th>crime</th>
<th>inmate</th>
<th>pitcher</th>
<th>prison</th>
<th>death</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>4 / 10</th>
<th>hospital</th>
<th>doctor</th>
<th>health</th>
<th>care</th>
<th>medical</th>
<th>tradition</th>
</tr>
</thead>
</table>

Material adapted from David Mimno | UMD
Word Intrusion: Which Topics are Interpretable?

New York Times, 50 LDA Topics

Model Precision: percentage of correct intruders found

Number of Topics: committee legislation proposal republican taxis

Model Precision: fireplace garage house kitchen list

Americans, Japanese, Jewish, states, terrorist

Artist, exhibition, gallery, museum, painting

Material adapted from David Mimno | UMD
Interpretability and Likelihood

Model Precision on New York Times

within a model, higher likelihood $\neq$ higher interpretability

Material adapted from David Mimno | UMD
Interpretability and Likelihood

Topic Log Odds on Wikipedia

across models, higher likelihood \neq higher interpretability
Evaluation Takeaway

- Measure what you care about
- If you care about prediction, likelihood is good
- If you care about a particular task, measure that
Inference

- We are interested in posterior distribution

\[ p(Z|X, \Theta) \] (4)
Inference

- We are interested in posterior distribution

\[ p(Z|X, \Theta) \]  

(4)

- Here, latent variables are topic assignments \( z \) and topics \( \theta \). \( X \) is the words (divided into documents), and \( \Theta \) are hyperparameters to Dirichlet distributions: \( \alpha \) for topic proportion, \( \lambda \) for topics.

\[ p(\vec{Z}, \vec{\beta}, \vec{\theta} | \vec{w}, \alpha, \lambda) \]  

(5)
Inference

- We are interested in posterior distribution

\[ p(Z|X, \Theta) \]  (4)

- Here, latent variables are topic assignments \( z \) and topics \( \theta \). \( X \) is the words (divided into documents), and \( \Theta \) are hyperparameters to Dirichlet distributions: \( \alpha \) for topic proportion, \( \lambda \) for topics.

\[ p(\tilde{Z}, \tilde{\beta}, \tilde{\theta} | \tilde{w}, \alpha, \lambda) \]  (5)

\[
p(\tilde{w}, \tilde{z}, \tilde{\theta}, \tilde{\beta} | \alpha, \lambda) = \prod_k p(\beta_k | \lambda) \prod_d p(\theta_d | \alpha) \prod_n p(z_{d,n} | \theta_d) p(w_{d,n} | \beta_{z_{d,n}})
\]
Gibbs Sampling

- A form of Markov Chain Monte Carlo
- Chain is a sequence of random variable states
- Given a state \{z_1, \ldots, z_N\} given certain technical conditions, drawing \(z_k \sim p(z_1, \ldots, z_{k-1}, z_{k+1}, \ldots, z_N | \mathbf{X}, \Theta)\) for all \(k\) (repeatedly) results in a Markov Chain whose stationary distribution is the posterior.
- For notational convenience, call \(\tilde{\mathbf{z}}\) with \(z_d, n\) removed \(\tilde{\mathbf{z}}_{-d, n}\)
Hollywood studios are preparing to let people download and buy electronic copies of movies over the Internet, much as record labels now sell songs for 99 cents through Apple Computer's iTunes music store and other online services ...
Hollywood studios are preparing to let people download and buy electronic copies of movies over the Internet, much as record labels now sell songs for 99 cents through Apple Computer's iTunes music store and other online services ...
Hollywood studios are preparing to let people download and buy electronic copies of movies over the Internet, much as record labels now sell songs for 99 cents through Apple Computer's iTunes music store and other online services."
Hollywood studios are preparing to let people download and buy electronic copies of movies over the Internet, much as record labels now sell songs for 99 cents through Apple Computer's iTunes music store and other online services ...
Hollywood studios are preparing to let people download and buy electronic copies of movies over the Internet, much as record labels now sell songs for 99 cents through Apple Computer's iTunes music store and other online services ...
Hollywood studios are preparing to let people download and buy electronic copies of movies over the Internet, much as record labels now sell songs for 99 cents through Apple Computer's iTunes music store and other online services ...
Hollywood studios are preparing to let people download and buy electronic copies of movies over the Internet, much as record labels now sell songs for 99 cents through Apple Computer's iTunes music store and other online services ...
Gibbs Sampling

- For LDA, we will sample the topic assignments
- Thus, we want:

\[
p(z_{d,n} = k | \tilde{z}_{d,n}, \tilde{w}, \alpha, \lambda) = \frac{p(z_{d,n} = k, \tilde{z}_{d,n} | \tilde{w}, \alpha, \lambda)}{p(\tilde{z}_{d,n} | \tilde{w}, \alpha, \lambda)}
\]
Gibbs Sampling

- For LDA, we will sample the topic assignments
- Thus, we want:

\[
p(z_{d,n} = k | \tilde{z}_{-d,n}, \tilde{w}, \alpha, \lambda) = \frac{p(z_{d,n} = k, \tilde{z}_{-d,n} | \tilde{w}, \alpha, \lambda)}{p(\tilde{z}_{-d,n} | \tilde{w}, \alpha, \lambda)}
\]

- The topics and per-document topic proportions are integrated out / marginalized
- Let \( n_{d,i} \) be the number of words taking topic \( i \) in document \( d \). Let \( v_{k,w} \) be the number of times word \( w \) is used in topic \( k \).

\[
\begin{align*}
= & \int_{\theta_d} \left( \prod_{i \neq k} \theta_d^{\alpha_i + n_{d,i} - 1} \right) \theta_d^{\alpha_k + n_{d,i}} d\theta_d \int_{\beta_k} \left( \prod_{i \neq w_{d,n}} \beta_{k,i}^{\lambda_i + v_{k,i} - 1} \right) \beta_{k,w_{d,n}}^{\lambda_i + v_{k,i}} d\beta_k \\
= & \int_{\theta_d} \left( \prod_i \theta_d^{\alpha_i + n_{d,i} - 1} \right) d\theta_d \int_{\beta_k} \left( \prod_i \beta_{k,i}^{\lambda_i + v_{k,i} - 1} \right) d\beta_k
\end{align*}
\]
Gibbs Sampling

- For LDA, we will sample the topic assignments
- The topics and per-document topic proportions are integrated out / marginalized / Rao-Blackwellized
- Thus, we want:

\[
p(z_{d,n} = k | \vec{z}_{-d,n}, \vec{w}, \alpha, \lambda) = \frac{n_{d,k} + \alpha_k}{\sum_i n_{d,i} + \alpha_i} \frac{v_{k,w_{d,n}} + \lambda w_{d,n}}{\sum_i v_{k,i} + \lambda_i}
\]
Gibbs Sampling

- Integral is normalizer of Dirichlet distribution

\[
\int_{\beta_k} \left( \prod_i \beta_{k,i}^{\lambda_i + v_{k,i} - 1} \right) d\beta_k = \frac{\prod_i^V \Gamma(\beta_i + v_{k,i})}{\Gamma(\sum_i^V \beta_i + v_{k,i})}
\]
Gibbs Sampling

- Integral is normalizer of Dirichlet distribution

\[
\int_{\beta_k} \left( \prod_i \beta_{k,i}^{\lambda_i + v_{k,i} - 1} \right) d\beta_k = \frac{\prod_i^V \Gamma(\beta_i + v_{k,i})}{\Gamma(\sum_i^V \beta_i + v_{k,i})}
\]

- So we can simplify

\[
\begin{align*}
\int_{\theta_d} \left( \prod_{i \neq k} \theta_d^{\alpha_i + n_{d,i} - 1} \right) \theta_d^{\alpha_k + n_{d,k}} d\theta_d \int_{\beta_k} \left( \prod_{i \neq w_{d,n}} \beta_{k,i}^{\lambda_i + v_{k,i} - 1} \right) \beta_{k,w_{d,n}}^{\lambda_i + v_{k,i}} d\beta_k = \\
\frac{\Gamma(\alpha_k + n_{d,k} + 1)}{\Gamma(\sum_i^K \alpha_i + n_{d,i} + 1)} \prod_{i \neq k}^K \Gamma(\alpha_k + n_{d,k}) \frac{\Gamma(\lambda_{w_{d,n}} + v_{k,w_{d,n}} + 1)}{\Gamma(\sum_i^V \lambda_i + v_{k,i} + 1)} \prod_{i \neq w_{d,n}}^V \Gamma(\lambda_k + v_{k,w_{d,n}}) \\
\frac{\prod_i^K \Gamma(\alpha_i + n_{d,i})}{\Gamma(\sum_i^K \alpha_i + n_{d,i})} \frac{\prod_i^V \Gamma(\lambda_i + v_{k,i})}{\Gamma(\sum_i^V \lambda_i + v_{k,i})}
\end{align*}
\]
Gamma Function Identity

\[ z = \frac{\Gamma(z + 1)}{\Gamma(z)} \]  

\[
\frac{\Gamma(\alpha_k + n_{d,k} + 1)}{\Gamma(\sum_i^{K} \alpha_i + n_{d,i} + 1)} \prod_{i \neq k}^{K} \Gamma(\alpha_k + n_{d,k}) = \frac{n_{d,k} + \alpha_k}{\sum_i^{K} n_{d,i} + \alpha_i} \frac{v_{k,wd,n} + \lambda_{wd,n}}{\sum_i v_{k,i} + \lambda_i}
\]

\[
\frac{\Gamma(\lambda_{wd,n} + v_{k,wd,n} + 1)}{\Gamma(\sum_i^{V} \lambda_i + v_{k,i} + 1)} \prod_{i \neq wd,n}^{V} \Gamma(\lambda_k + v_{k,d,n}) = \frac{\prod_i^{V} \Gamma(\lambda_i + v_{k,i})}{\Gamma(\sum_i^{V} \lambda_i + v_{k,i})}
\]
Gibbs Sampling Equation

\[
\frac{n_{d,k} + \alpha_k}{\sum_i^K n_{d,i} + \alpha_i} \quad \frac{v_{k,w_{d,n}} + \lambda_{w_{d,n}}}{\sum_i v_{k,i} + \lambda_i}
\]  

- Number of times document \( d \) uses topic \( k \)
- Number of times topic \( k \) uses word type \( w_{d,n} \)
- Dirichlet parameter for document to topic distribution
- Dirichlet parameter for topic to word distribution
- How much this document likes topic \( k \)
- How much this topic likes word \( w_{d,n} \)
Gibbs Sampling Equation

\[ \frac{n_{d,k} + \alpha_k}{\sum_i^K n_{d,i} + \alpha_i} \frac{v_{k,w_{d,n}} + \lambda_{w_{d,n}}}{\sum_i v_{k,i} + \lambda_i} \]  

- Number of times document \( d \) uses topic \( k \)
- Number of times topic \( k \) uses word type \( w_{d,n} \)
- Dirichlet parameter for document to topic distribution
- Dirichlet parameter for topic to word distribution
- How much this document likes topic \( k \)
- How much this topic likes word \( w_{d,n} \)
Inference

Gibbs Sampling Equation

\[
\frac{n_{d,k} + \alpha_k}{\sum_i^K n_{d,i} + \alpha_i} \quad \frac{v_{k,w_{d,n}} + \lambda_{w_{d,n}}}{\sum_i v_{k,i} + \lambda_i}
\]  

- Number of times document \( d \) uses topic \( k \)
- Number of times topic \( k \) uses word type \( w_{d,n} \)
- Dirichlet parameter for document to topic distribution
- Dirichlet parameter for topic to word distribution
- How much this document likes topic \( k \)
- How much this topic likes word \( w_{d,n} \)
Gibbs Sampling Equation

\[
\frac{n_{d,k} + \alpha_k}{\sum_i^K n_{d,i} + \alpha} \cdot \frac{v_{k,w_{d,n}} + \lambda_{w_{d,n}}}{\sum_i v_{k,i} + \lambda_i}
\]

- Number of times document \(d\) uses topic \(k\)
- Number of times topic \(k\) uses word type \(w_{d,n}\)
- Dirichlet parameter for document to topic distribution
- Dirichlet parameter for topic to word distribution
- How much this document likes topic \(k\)
- How much this topic likes word \(w_{d,n}\)
Inference

Gibbs Sampling Equation

\[
\frac{n_{d,k} + \alpha_k}{\sum_i^K n_{d,i} + \alpha_i} \frac{v_{k,w_{d,n}} + \lambda_{w_{d,n}}}{\sum_i v_{k,i} + \lambda_i}
\]

- Number of times document \(d\) uses topic \(k\)
- Number of times topic \(k\) uses word type \(w_{d,n}\)
- Dirichlet parameter for document to topic distribution
- Dirichlet parameter for topic to word distribution
- How much this document likes topic \(k\)
- How much this topic likes word \(w_{d,n}\)
Gibbs Sampling Equation

$$\frac{n_{d,k} + \alpha_k}{\sum_i^K n_{d,i} + \alpha_i} \frac{v_{k,w_{d,n}} + \lambda_{w_{d,n}}}{\sum_i v_{k,i} + \lambda_i}$$

- Number of times document $d$ uses topic $k$
- Number of times topic $k$ uses word type $w_{d,n}$
- Dirichlet parameter for document to topic distribution
- Dirichlet parameter for topic to word distribution
- How much this document likes topic $k$
- How much this topic likes word $w_{d,n}$
### Sample Document

<table>
<thead>
<tr>
<th>Etruscan</th>
<th>trade</th>
<th>price</th>
<th>temple</th>
<th>market</th>
</tr>
</thead>
</table>
Sample Document

<table>
<thead>
<tr>
<th>Etruscan</th>
<th>trade</th>
<th>price</th>
<th>temple</th>
<th>market</th>
</tr>
</thead>
</table>
Randomly Assign Topics

<table>
<thead>
<tr>
<th></th>
<th>3</th>
<th>2</th>
<th>1</th>
<th>3</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Etruscan</td>
<td>trade</td>
<td>price</td>
<td>temple</td>
<td>market</td>
</tr>
</tbody>
</table>
Randomly Assign Topics

<table>
<thead>
<tr>
<th>z</th>
<th>3</th>
<th>2</th>
<th>1</th>
<th>3</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Etruscan</td>
<td>trade</td>
<td>price</td>
<td>temple</td>
<td>market</td>
</tr>
</tbody>
</table>

Material adapted from David Mimno | UMD
### Total Topic Counts

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Etruscan</td>
<td>1</td>
<td>0</td>
<td>35</td>
</tr>
<tr>
<td>market</td>
<td>50</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>price</td>
<td>42</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>temple</td>
<td>0</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td>trade</td>
<td>10</td>
<td>8</td>
<td>1</td>
</tr>
</tbody>
</table>

*Material adapted from David Mimno*
Total Topic Counts

<table>
<thead>
<tr>
<th></th>
<th>3</th>
<th>2</th>
<th>1</th>
<th>3</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Etruscan</td>
<td>trade</td>
<td>price</td>
<td>temple</td>
<td>market</td>
<td></td>
</tr>
</tbody>
</table>

Sampling Equation

\[
\frac{n_{d,k} + \alpha_k}{\sum_i n_{d,i} + \alpha_i} \quad \frac{v_{k,w_{d,n}} + \lambda_{w_{d,n}}}{\sum_i v_{k,i} + \lambda_i}
\]
Total Topic Counts

<table>
<thead>
<tr>
<th></th>
<th>3</th>
<th>2</th>
<th>1</th>
<th>3</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Etruscan</td>
<td>trade</td>
<td>price</td>
<td>temple</td>
<td>market</td>
<td></td>
</tr>
</tbody>
</table>

Material adapted from David Mimno | UMD
We want to sample this word …

<table>
<thead>
<tr>
<th>3</th>
<th>2</th>
<th>1</th>
<th>3</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Etruscan</td>
<td>trade</td>
<td>price</td>
<td>temple</td>
<td>market</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Etruscan</td>
<td>1</td>
<td>0</td>
<td>35</td>
</tr>
<tr>
<td>market</td>
<td>50</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>price</td>
<td>42</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>temple</td>
<td>0</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td>trade</td>
<td>10</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Inference

We want to sample this word ...

<table>
<thead>
<tr>
<th></th>
<th>3</th>
<th>2</th>
<th>1</th>
<th>3</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Etruscan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>trade</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>price</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>temple</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>market</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Etruscan</td>
<td>1</td>
<td>0</td>
<td>35</td>
</tr>
<tr>
<td>market</td>
<td>50</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>price</td>
<td>42</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>temple</td>
<td>0</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td>trade</td>
<td>10</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Decrement its count

<table>
<thead>
<tr>
<th></th>
<th>3</th>
<th>?</th>
<th>1</th>
<th>3</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Etruscan</td>
<td>trade</td>
<td>price</td>
<td>temple</td>
<td>market</td>
<td></td>
</tr>
</tbody>
</table>

```

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Etruscan</td>
<td>1</td>
<td>0</td>
<td>35</td>
</tr>
<tr>
<td>market</td>
<td>50</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>price</td>
<td>42</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>temple</td>
<td>0</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td>trade</td>
<td>10</td>
<td>7</td>
<td>1</td>
</tr>
</tbody>
</table>

...  
```
What is the conditional distribution for this topic?

<table>
<thead>
<tr>
<th></th>
<th>?</th>
<th>1</th>
<th>3</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Etruscan</td>
<td>trade</td>
<td>price</td>
<td>temple</td>
<td>market</td>
</tr>
</tbody>
</table>
Part 1: How much does this document like each topic?

<table>
<thead>
<tr>
<th></th>
<th>?</th>
<th>1</th>
<th>3</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Etruscan</td>
<td>trade</td>
<td>price</td>
<td>temple</td>
<td>market</td>
</tr>
</tbody>
</table>
Part 1: How much does this document like each topic?

<table>
<thead>
<tr>
<th>Topic 1</th>
<th>Topic 2</th>
<th>Topic 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>?</td>
<td>1</td>
</tr>
<tr>
<td>Etruscan</td>
<td>trade</td>
<td>price</td>
</tr>
<tr>
<td>3</td>
<td>temple</td>
<td>market</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Part 1: How much does this document like each topic?

<table>
<thead>
<tr>
<th>3</th>
<th>?</th>
<th>1</th>
<th>3</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Etruscan</td>
<td>trade</td>
<td>price</td>
<td>temple</td>
<td>market</td>
</tr>
</tbody>
</table>

**Sampling Equation**

\[
\frac{n_{d,k} + \alpha_k}{\sum_i n_{d,i} + \alpha_i} \quad \frac{v_{k,w_{d,n}} + \lambda_{w_{d,n}}}{\sum_i v_{k,i} + \lambda_i}
\]
Part 1: How much does this document like each topic?

<table>
<thead>
<tr>
<th></th>
<th>?</th>
<th>1</th>
<th>3</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Etruscan</td>
<td>trade</td>
<td>price</td>
<td>temple</td>
<td>market</td>
</tr>
</tbody>
</table>

Sampling Equation

\[
\frac{n_{d,k} + \alpha_k}{\sum_i n_{d,i} + \alpha_i} \quad \frac{v_{k,w_{d,n}} + \lambda_{w_{d,n}}}{\sum_i v_{k,i} + \lambda_i}
\]
Part 2: How much does each topic like the word?

<table>
<thead>
<tr>
<th></th>
<th>3</th>
<th>?</th>
<th>1</th>
<th>3</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Etruscan</td>
<td>trade</td>
<td>price</td>
<td>temple</td>
<td>market</td>
<td></td>
</tr>
</tbody>
</table>

**Topic 1**

**Topic 2**

**Topic 3**

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>trade</td>
<td>10</td>
<td>7</td>
<td>1</td>
</tr>
</tbody>
</table>
Part 2: How much does each topic like the word?

<table>
<thead>
<tr>
<th>3</th>
<th>?</th>
<th>1</th>
<th>3</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Etruscan</td>
<td>trade</td>
<td>price</td>
<td>temple</td>
<td>market</td>
</tr>
</tbody>
</table>

**Sampling Equation**

\[
\frac{n_{d,k} + \alpha_k}{\sum_i n_{d,i} + \alpha_i} \quad \frac{v_{k,w_{d,n}} + \lambda_{w_{d,n}}}{\sum_i v_{k,i} + \lambda_i}
\]
Part 2: How much does each topic like the word?

<table>
<thead>
<tr>
<th>3</th>
<th>?</th>
<th>1</th>
<th>3</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Etruscan</td>
<td>trade</td>
<td>price</td>
<td>temple</td>
<td>market</td>
</tr>
</tbody>
</table>

**Sampling Equation**

\[
\frac{n_{d,k} + \alpha_k}{\sum_i n_{d,i} + \alpha_i} \quad \frac{v_{k,w_{d,n}} + \lambda_{w_{d,n}}}{\sum_i v_{k,i} + \lambda_i}
\]
Geometric interpretation

<table>
<thead>
<tr>
<th>3</th>
<th>?</th>
<th>1</th>
<th>3</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Etruscan</td>
<td>trade</td>
<td>price</td>
<td>temple</td>
<td>market</td>
</tr>
</tbody>
</table>

Topic 1

Topic 2

Topic 3
**Geometric interpretation**

<table>
<thead>
<tr>
<th>3</th>
<th>?</th>
<th>1</th>
<th>3</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Etruscan</td>
<td>trade</td>
<td>price</td>
<td>temple</td>
<td>market</td>
</tr>
</tbody>
</table>

**Diagram**

- **Topic 1**
- **Topic 2**
- **Topic 3**
Geometric interpretation

<table>
<thead>
<tr>
<th></th>
<th>?</th>
<th>1</th>
<th>3</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Etruscan</td>
<td>trade</td>
<td>price</td>
<td>temple</td>
<td>market</td>
</tr>
</tbody>
</table>

Topic 1

Topic 2

Topic 3
Update counts

<table>
<thead>
<tr>
<th></th>
<th>Etruscan</th>
<th>trade</th>
<th>price</th>
<th>temple</th>
<th>market</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>?</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Etruscan</td>
<td>1</td>
<td>0</td>
<td>35</td>
</tr>
<tr>
<td>market</td>
<td>50</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>price</td>
<td>42</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>temple</td>
<td>0</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td>trade</td>
<td>10</td>
<td>7</td>
<td>1</td>
</tr>
</tbody>
</table>

...
Inference

Update counts

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>1</th>
<th>3</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Etruscan</td>
<td>trade</td>
<td>price</td>
<td>temple</td>
<td>market</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Etruscan</td>
<td>1</td>
<td>0</td>
<td>35</td>
</tr>
<tr>
<td>market</td>
<td>50</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>price</td>
<td>42</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>temple</td>
<td>0</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td>trade</td>
<td>11</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Update counts

<table>
<thead>
<tr>
<th></th>
<th>3</th>
<th>1</th>
<th>1</th>
<th>3</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Etruscan</td>
<td>trade</td>
<td>price</td>
<td>temple</td>
<td>market</td>
<td></td>
</tr>
</tbody>
</table>

Theme 1                        | Theme 2                          | Theme 3

Material adapted from David Mimno | UMD
Details: how to sample from a distribution

\[
\frac{n_{d,k} + \alpha_k}{\sum_i^K n_{d,i} + \alpha_i} \frac{v_{k,w_{d,n}} + \lambda_{w_{d,n}}}{\sum_i v_{k,i} + \lambda_i}
\]
Algorithm

1. For each iteration $i$:
   1.1 For each document $d$ and word $n$ currently assigned to $z_{old}$:
      1.1.1 Decrement $n_{d,z_{old}}$ and $v_{z_{old},w_{d,n}}$
      1.1.2 Sample $z_{new} = k$ with probability proportional to
      
      $\frac{n_{d,k} + \alpha_k}{\sum_{i} n_{d,i} + \alpha_i} \frac{v_{k,w_{d,n}} + \lambda_{w_{d,n}}}{\sum_{i} v_{k,i} + \lambda_i}$
      
      1.1.3 Increment $n_{d,z_{new}}$ and $v_{z_{new},w_{d,n}}$
Inference

Implementation

Algorithm

1. For each iteration $i$:
   1.1 For each document $d$ and word $n$ currently assigned to $z_{old}$:
      1.1.1 Decrement $n_{d,z_{old}}$ and $v_{z_{old},w_{d,n}}$
      1.1.2 Sample $z_{new} = k$ with probability proportional to $\frac{n_{d,k} + \alpha_k}{\sum_i n_{d,i} + \alpha_i}$ $\frac{v_{k,w_{d,n}} + \lambda_{w_{d,n}}}{\sum_i v_{k,i} + \lambda_i}$
      1.1.3 Increment $n_{d,z_{new}}$ and $v_{z_{new},w_{d,n}}$
Desiderata

- Hyperparameters: Sample them too (slice sampling)
- Initialization: Random
- Sampling: Until likelihood converges
- Lag / burn-in: Difference of opinion on this
- Number of chains: Should do more than one
Available implementations

- Mallet (http://mallet.cs.umass.edu)
- LDAC (http://www.cs.princeton.edu/blei/lda-c)
- Topicmod (http://code.google.com/p/topicmod)
Wrapup

- Topic Models: Tools to uncover themes in large document collections
- Another example of Gibbs Sampling
- In class: Gibbs sampling example