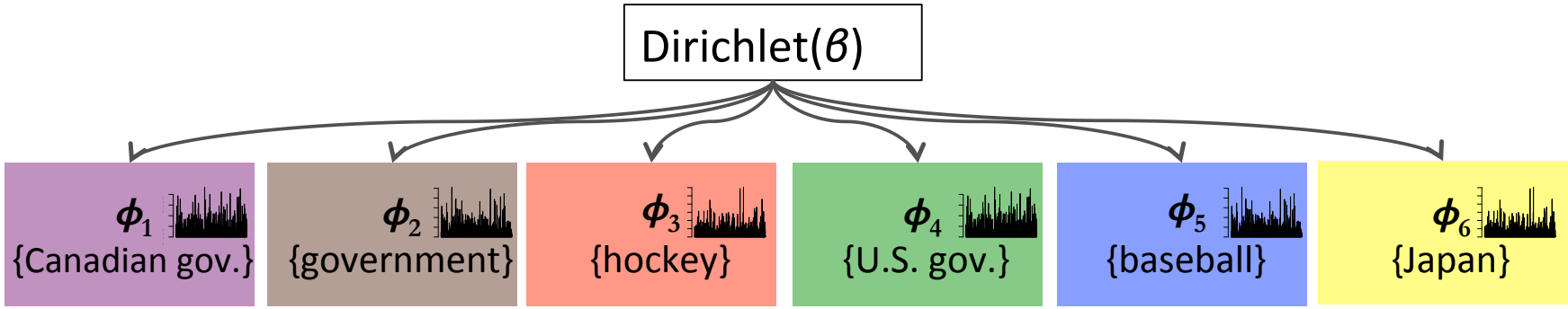
LDA for Topic Modeling

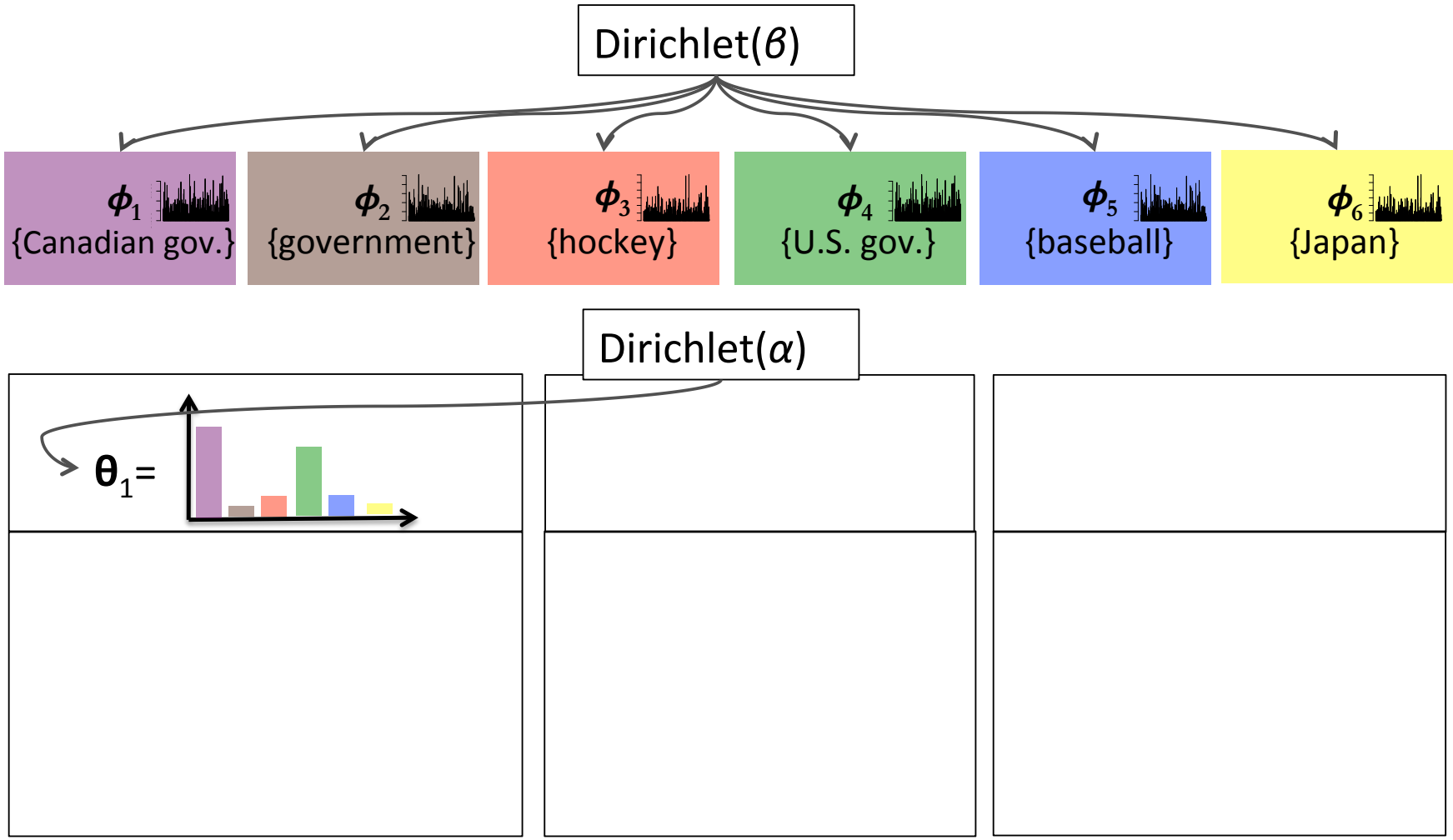


- A topic is visualized as its high probability words.
- A pedagogical **label** is used to identify the topic.

46

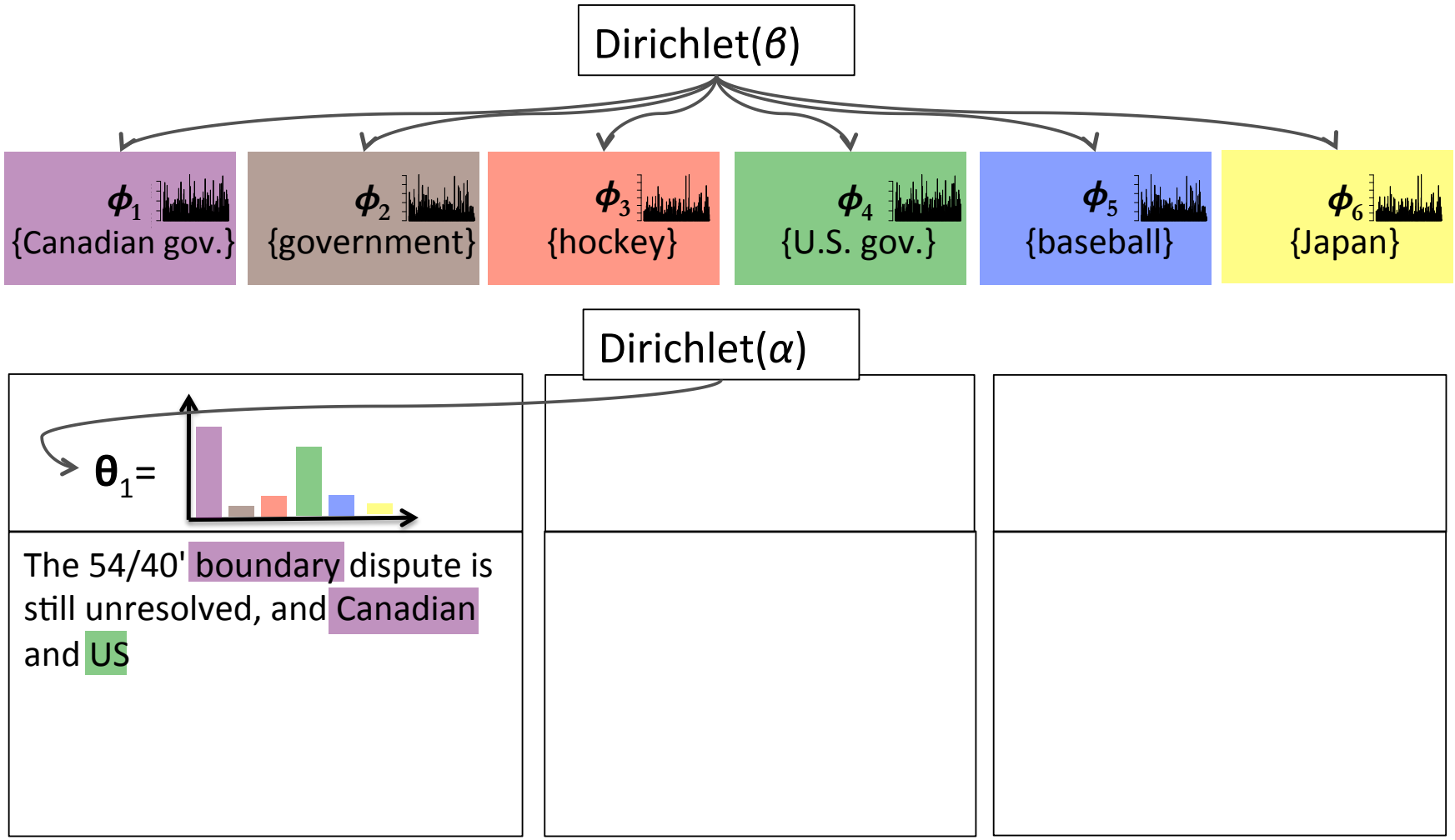
A slide from: Matt Gormley, Carnegie-Mellon University 10-710 (2016), Lecture 20: Topic Modelling

LDA for Topic Modeling



A slide from: Matt Gormley, Carnegie-Mellon University 10-710 (2016), Lecture 20: Topic Modelling

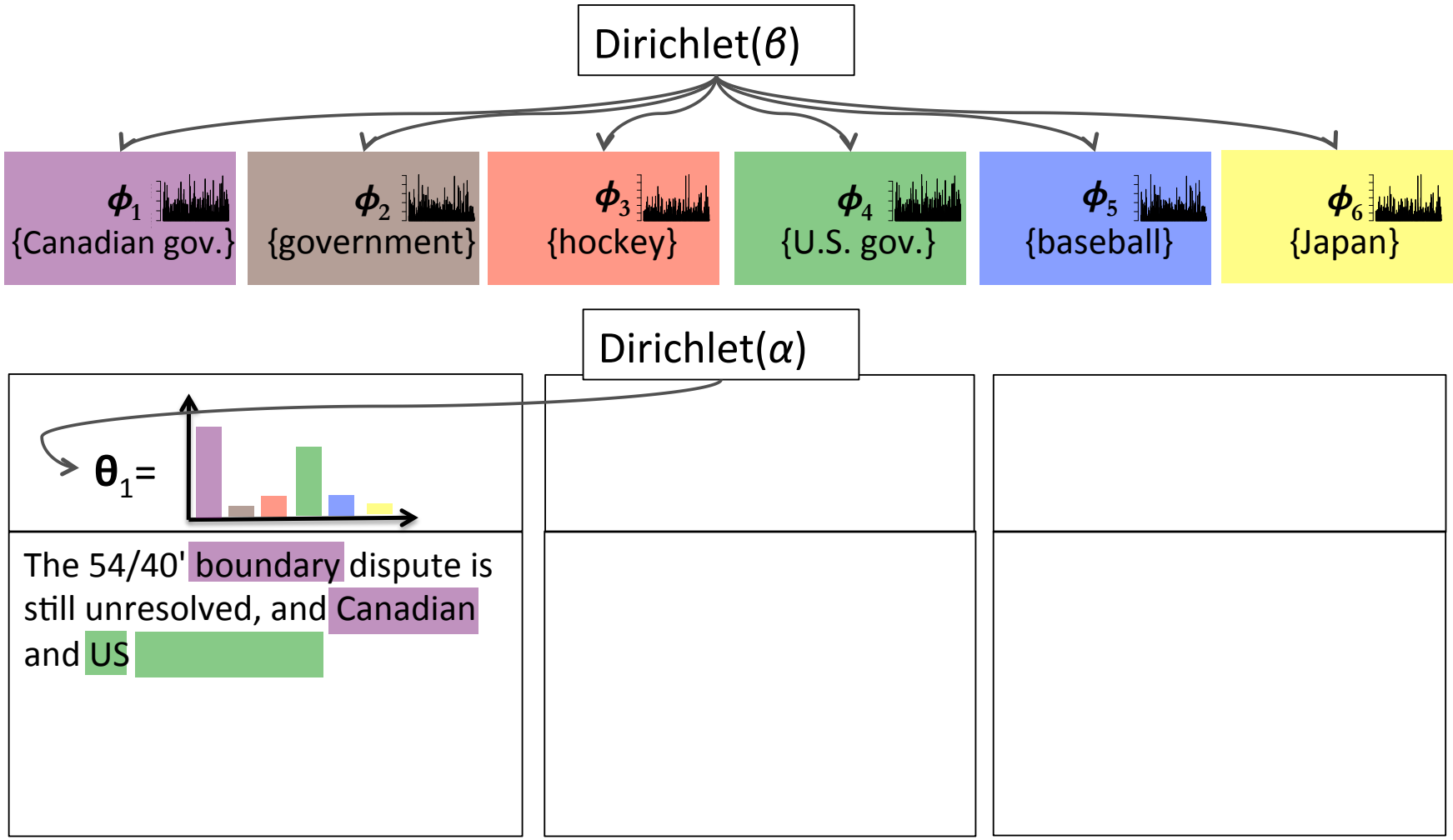
LDA for Topic Modeling



48

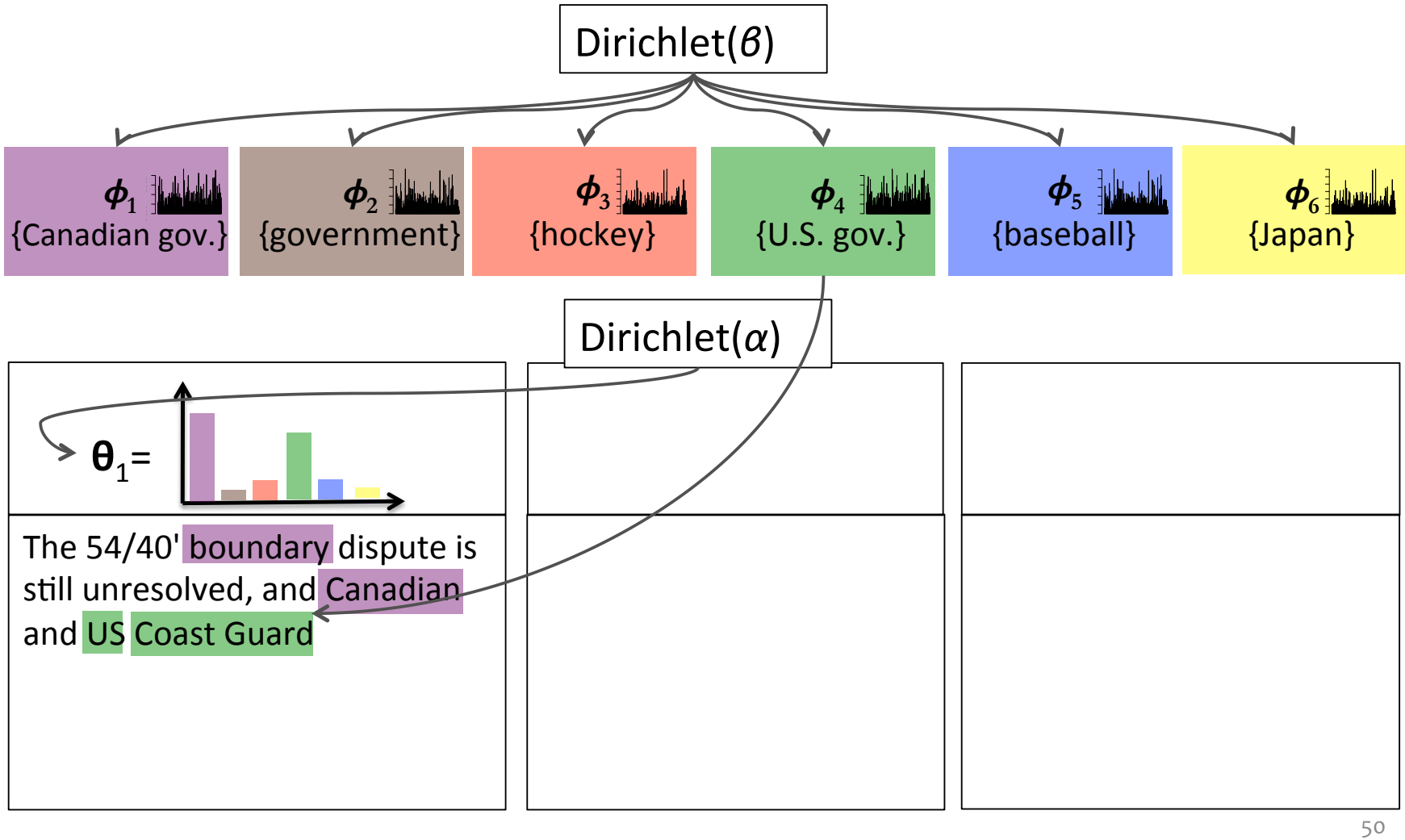
A slide from: Matt Gormley, Carnegie-Mellon University 10-710 (2016), Lecture 20: Topic Modelling

LDA for Topic Modeling



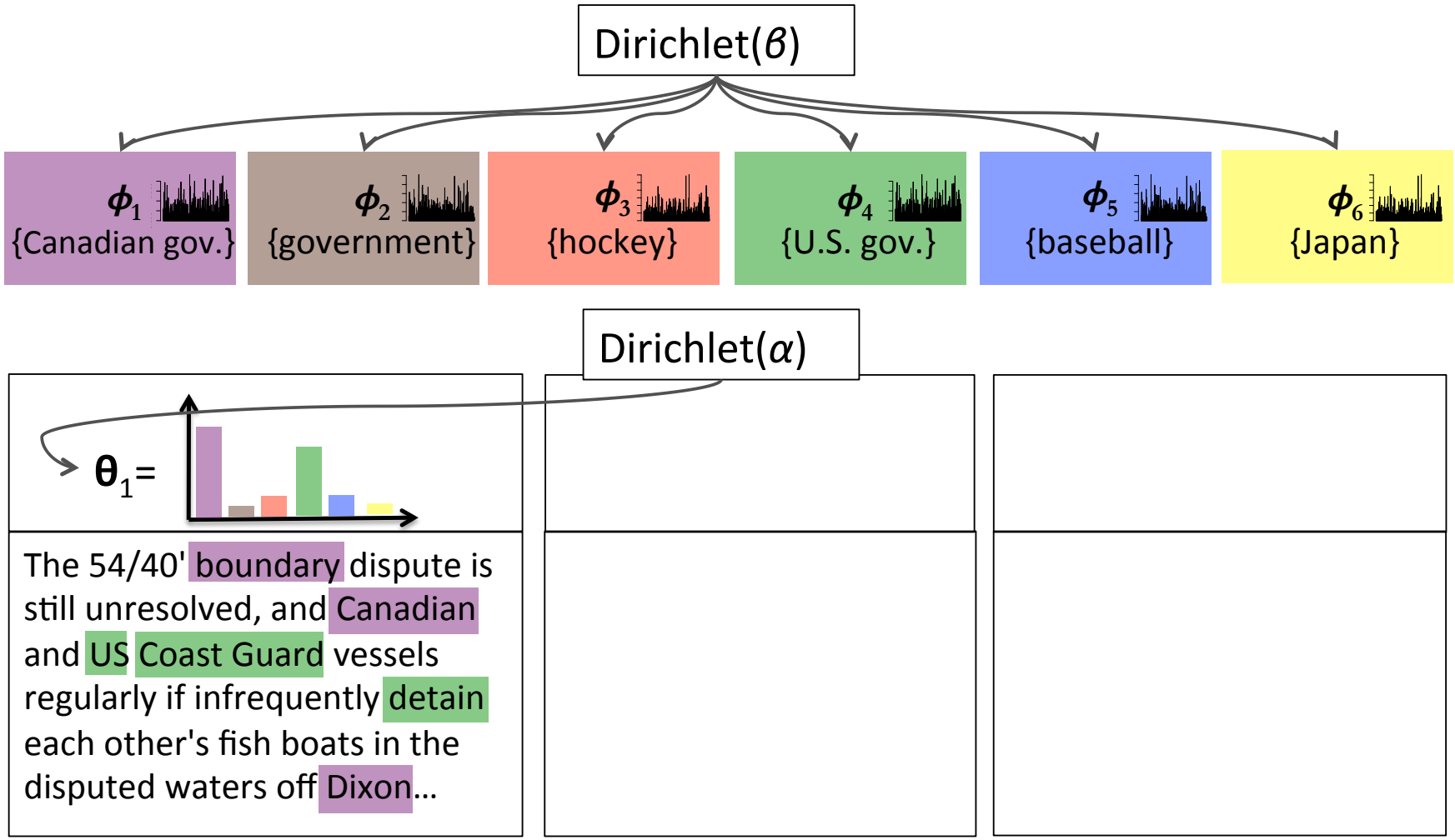
A slide from: Matt Gormley, Carnegie-Mellon University 10-710 (2016), Lecture 20: Topic Modelling

LDA for Topic Modeling



A slide from: Matt Gormley, Carnegie-Mellon University 10-710 (2016), Lecture 20: Topic Modelling

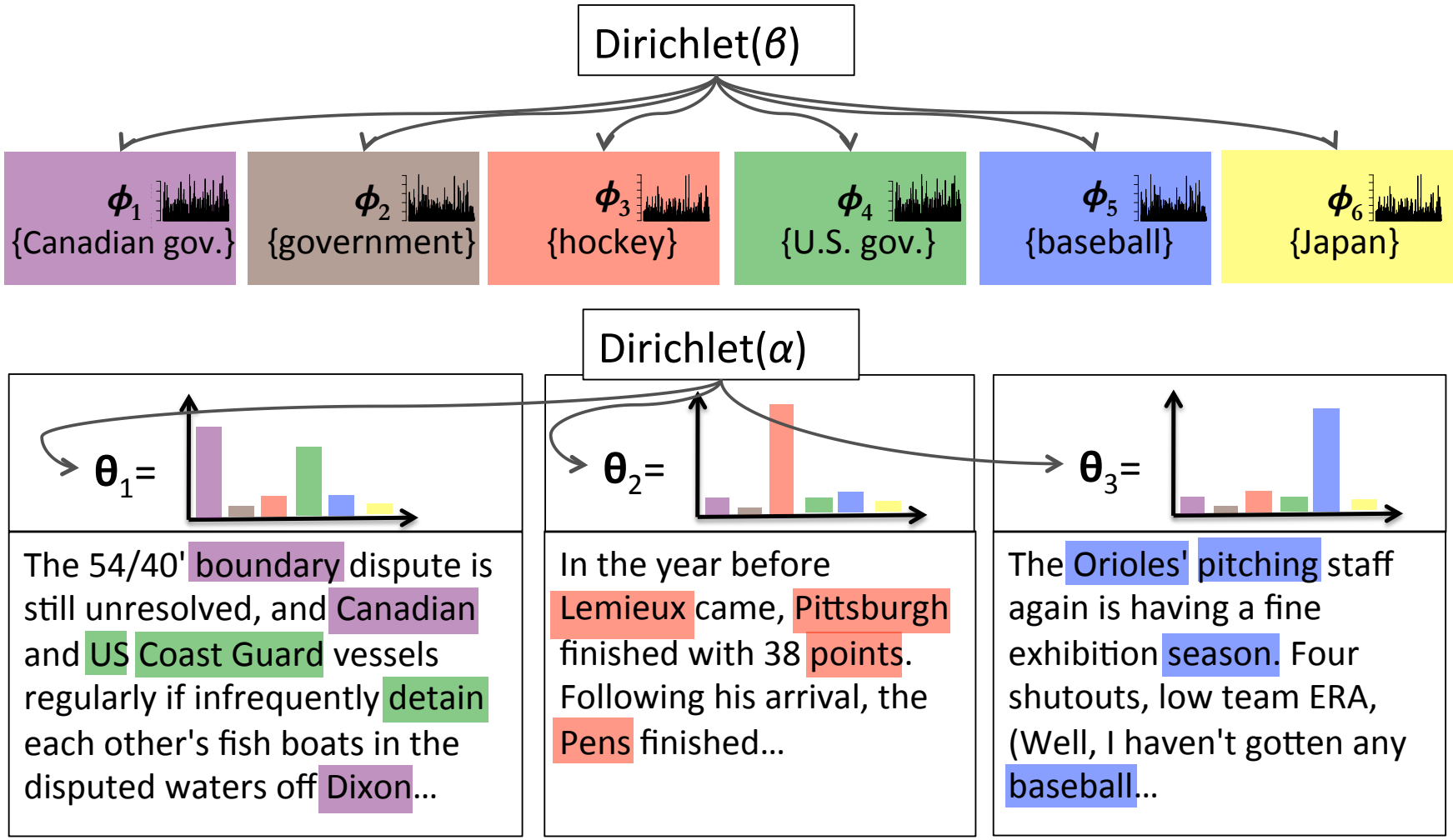
LDA for Topic Modeling



51

A slide from: Matt Gormley, Carnegie-Mellon University 10-710 (2016), Lecture 20: Topic Modelling

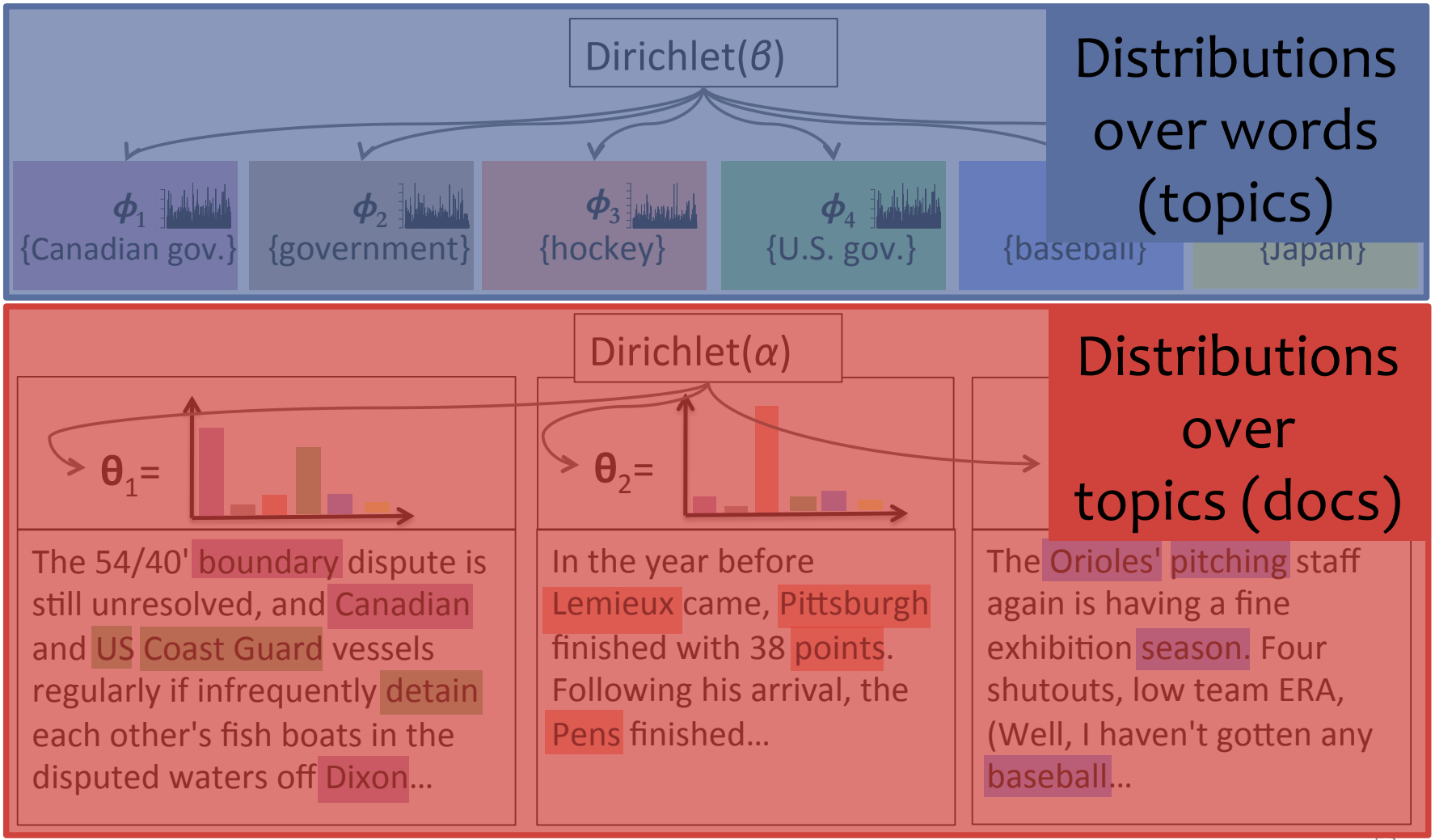
LDA for Topic Modeling



52

A slide from: Matt Gormley, Carnegie-Mellon University 10-710 (2016), Lecture 20: Topic Modelling

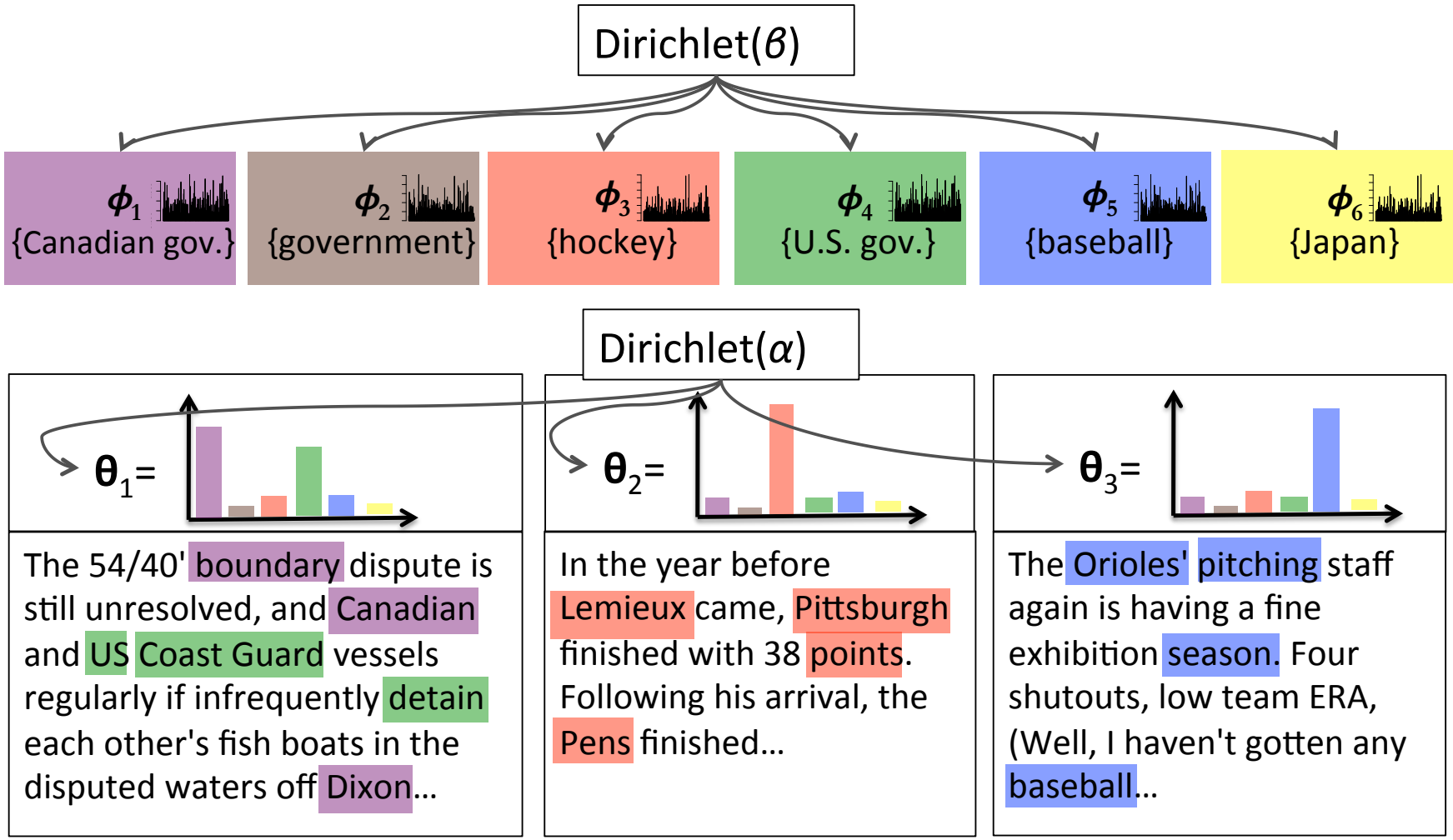
LDA for Topic Modeling



53

A slide from: Matt Gormley, Carnegie-Mellon University 10-710 (2016), Lecture 20: Topic Modelling

LDA for Topic Modeling

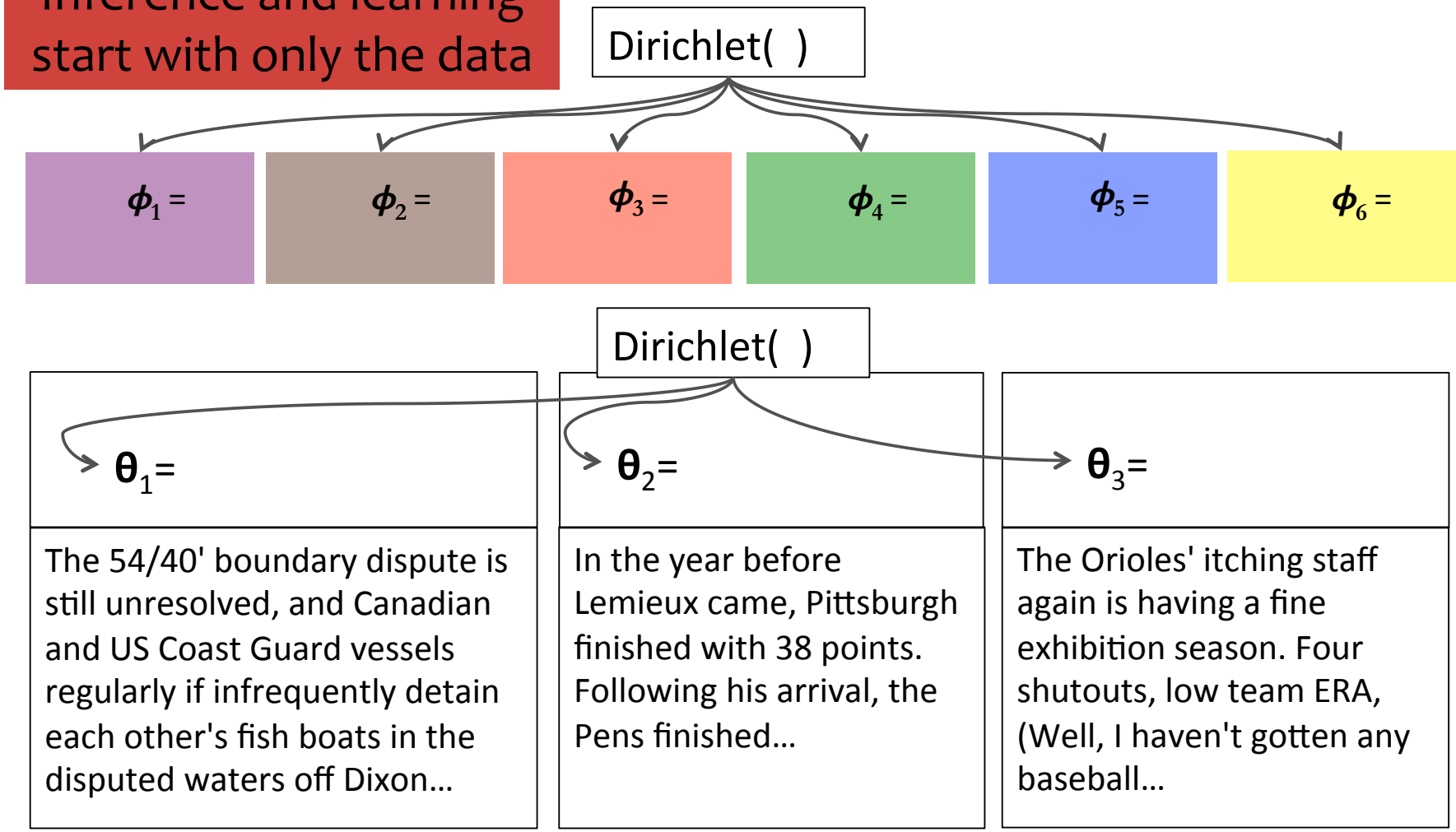


54

A slide from: Matt Gormley, Carnegie-Mellon University 10-710 (2016), Lecture 20: Topic Modelling

LDA for Topic Modeling

Inference and learning start with only the data



55

A slide from: Matt Gormley, Carnegie-Mellon University 10-710 (2016), Lecture 20: Topic Modelling

Latent Dirichlet Allocation

Questions:

- Is this a believable story for the generation of a corpus of documents?

- Why might it work well anyway?

A slide from: Matt Gormley, Carnegie-Mellon University 10-710 (2016), Lecture 20: Topic Modelling

Latent Dirichlet Allocation

Why does LDA "work"?

- LDA trades off two goals.
  1. For each document, allocate its words to as few topics as possible.
  2. For each topic, assign high probability to as few terms as possible.
- These goals are at odds.
  - Putting a document in a single topic makes #2 hard:
    All of its words must have probability under that topic.
  - Putting very few words in each topic makes #1 hard:
    To cover a document's words, it must assign many topics to it.
- Trading off these goals finds groups of tightly co-occurring words.

Slide from David Blei, MLSS 2012

A slide from: Matt Gormley, Carnegie-Mellon University 10-710 (2016), Lecture 20: Topic Modelling

Latent Dirichlet Allocation

How does this relate to my other favorite model for capturing low-dimensional representations of a corpus?

- Builds on latent semantic analysis (Deerwester et al., 1990; Hofmann, 1999)
- It is a mixed-membership model (Erosheva, 2004).
- It relates to PCA and matrix factorization (Jakulin and Buntine, 2002)
- Was independently invented for genetics (Pritchard et al., 2000)

Slide from David Blei, MLSS 2012

A slide from: Matt Gormley, Carnegie-Mellon University 10-710 (2016), Lecture 20: Topic Modelling

Content

Unsupervised Learning

Three learning paradigms:

1. Maximum likelihood

$$\arg\max_{\theta} p(X|\theta)$$

2. Maximum a posteriori (MAP)

$$\arg\max_{\theta} p(\theta|X) \propto p(X|\theta)p(\theta)$$

3. Bayesian approach

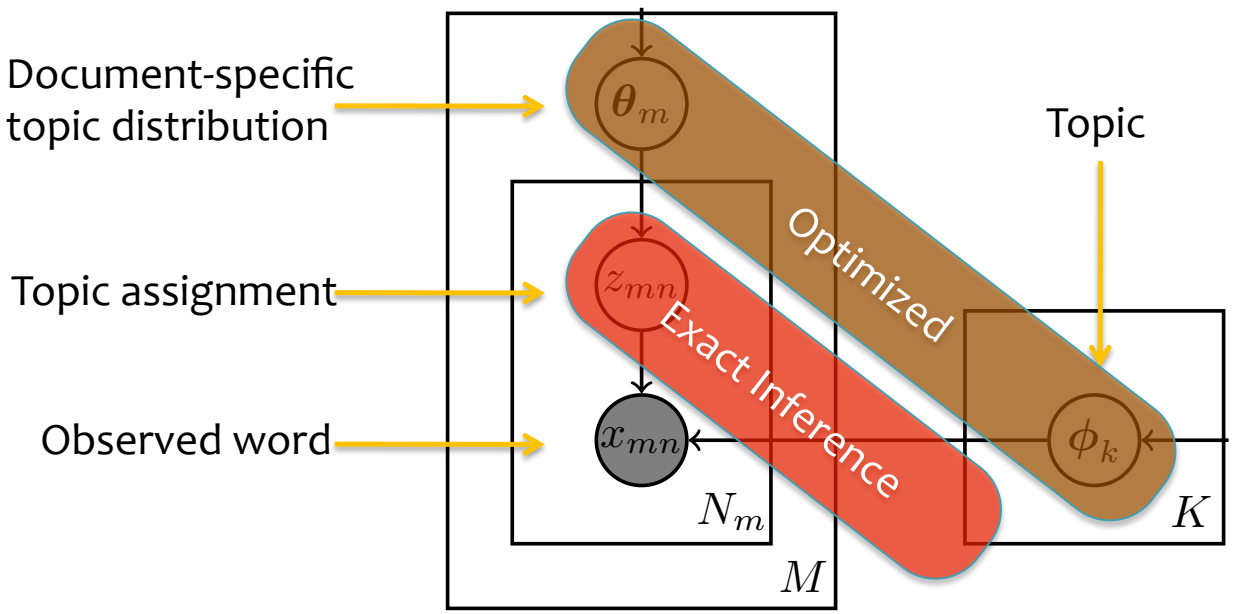
Estimate the posterior:

$$p(\theta|X) = \ldots$$

A slide from: Matt Gormley, Carnegie-Mellon University 10-710 (2016), Lecture 20: Topic Modelling

LDA Inference

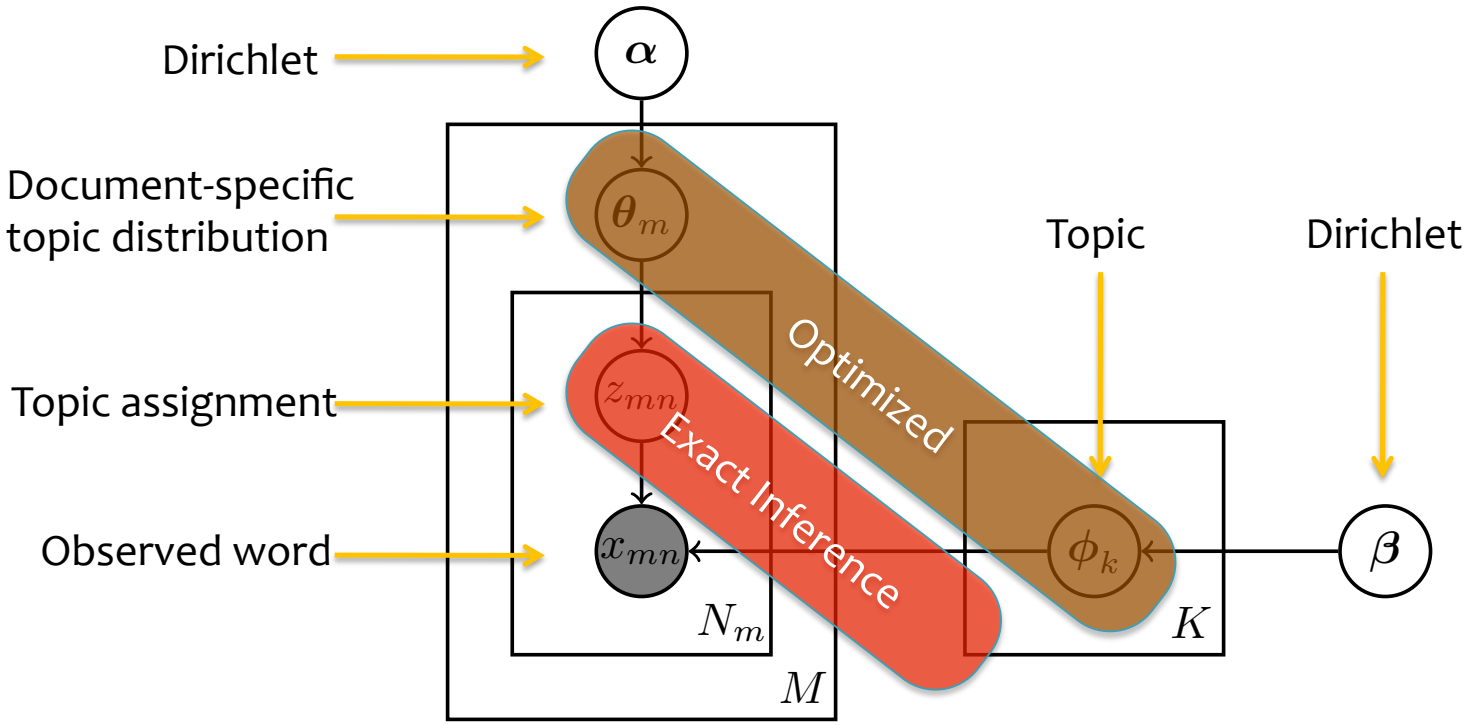
- Standard EM (Maximum Likelihood)



A slide from: Matt Gormley, Carnegie-Mellon University 10-710 (2016), Lecture 20: Topic Modelling

LDA Inference

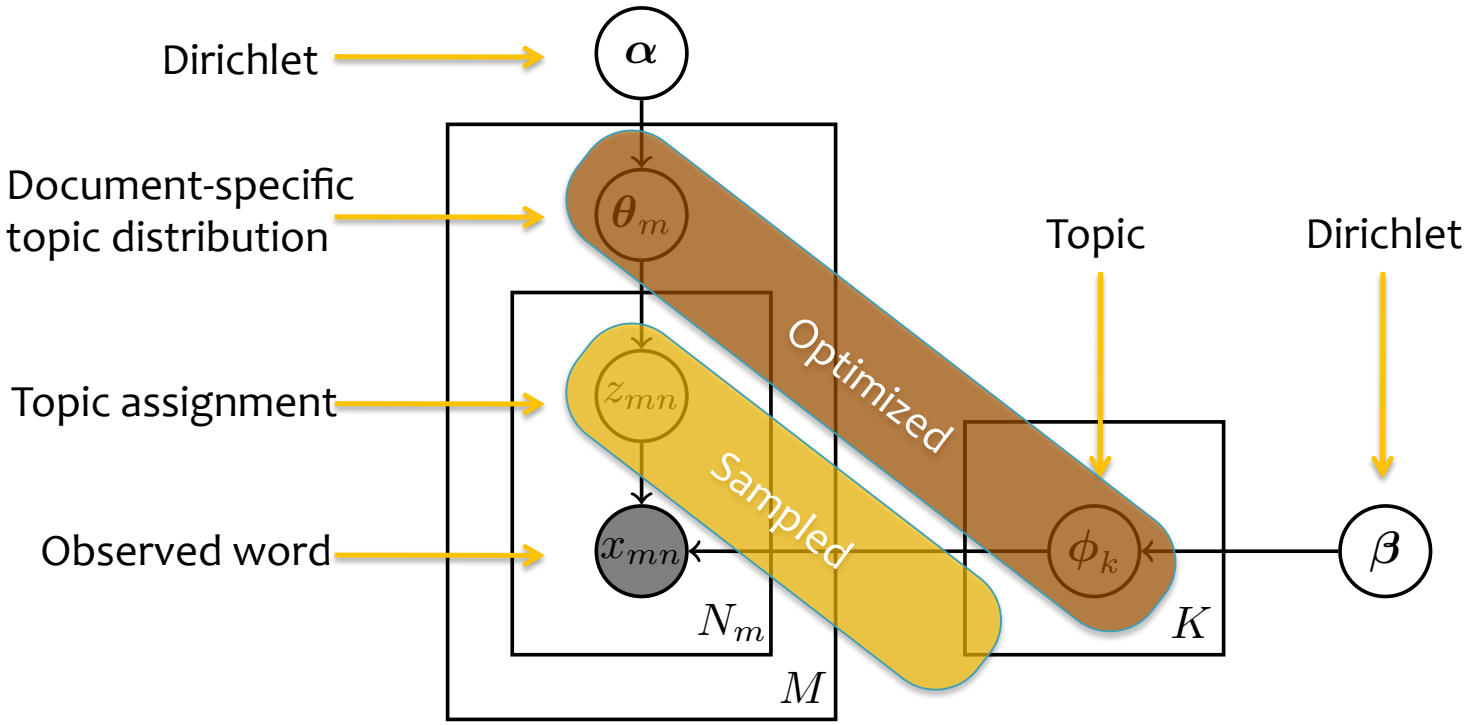
- Standard EM (MAP)



A slide from: Matt Gormley, Carnegie-Mellon University 10-710 (2016), Lecture 20: Topic Modelling

LDA Inference

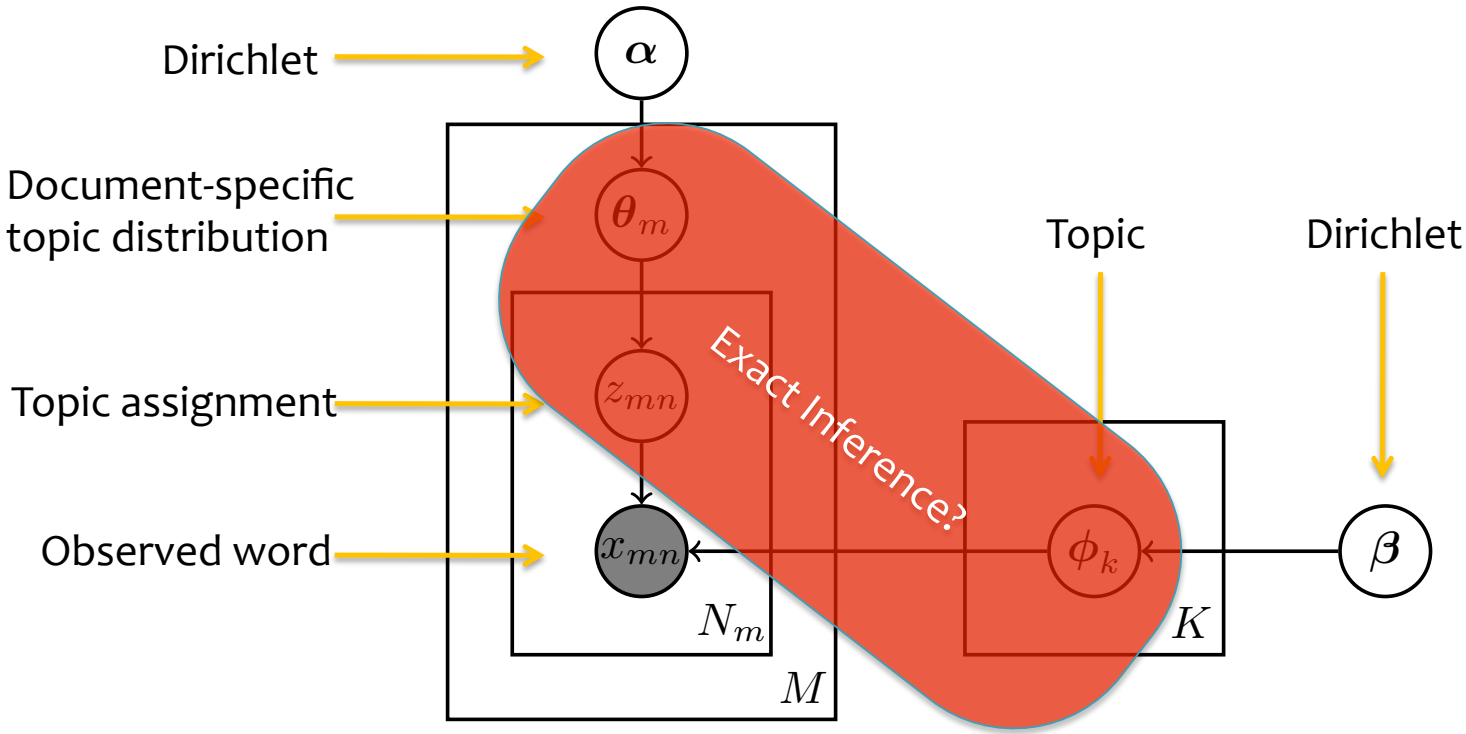
- Monte Carlo EM



A slide from: Matt Gormley, Carnegie-Mellon University 10-710 (2016), Lecture 20: Topic Modelling

LDA Inference

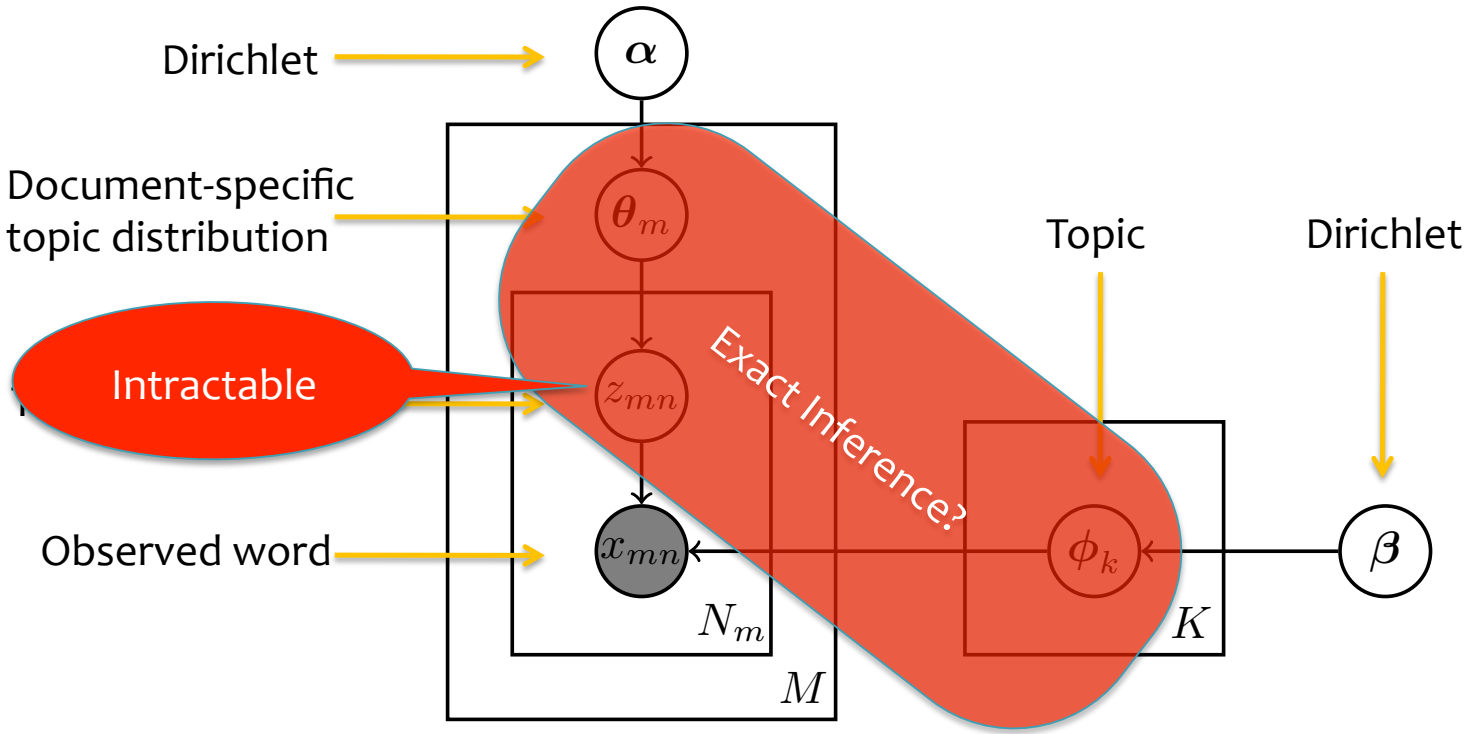
- Bayesian Approach



A slide from: Matt Gormley, Carnegie-Mellon University 10-710 (2016), Lecture 20: Topic Modelling

LDA Inference

- Bayesian Approach



A slide from: Matt Gormley, Carnegie-Mellon University 10-710 (2016), Lecture 20: Topic Modelling

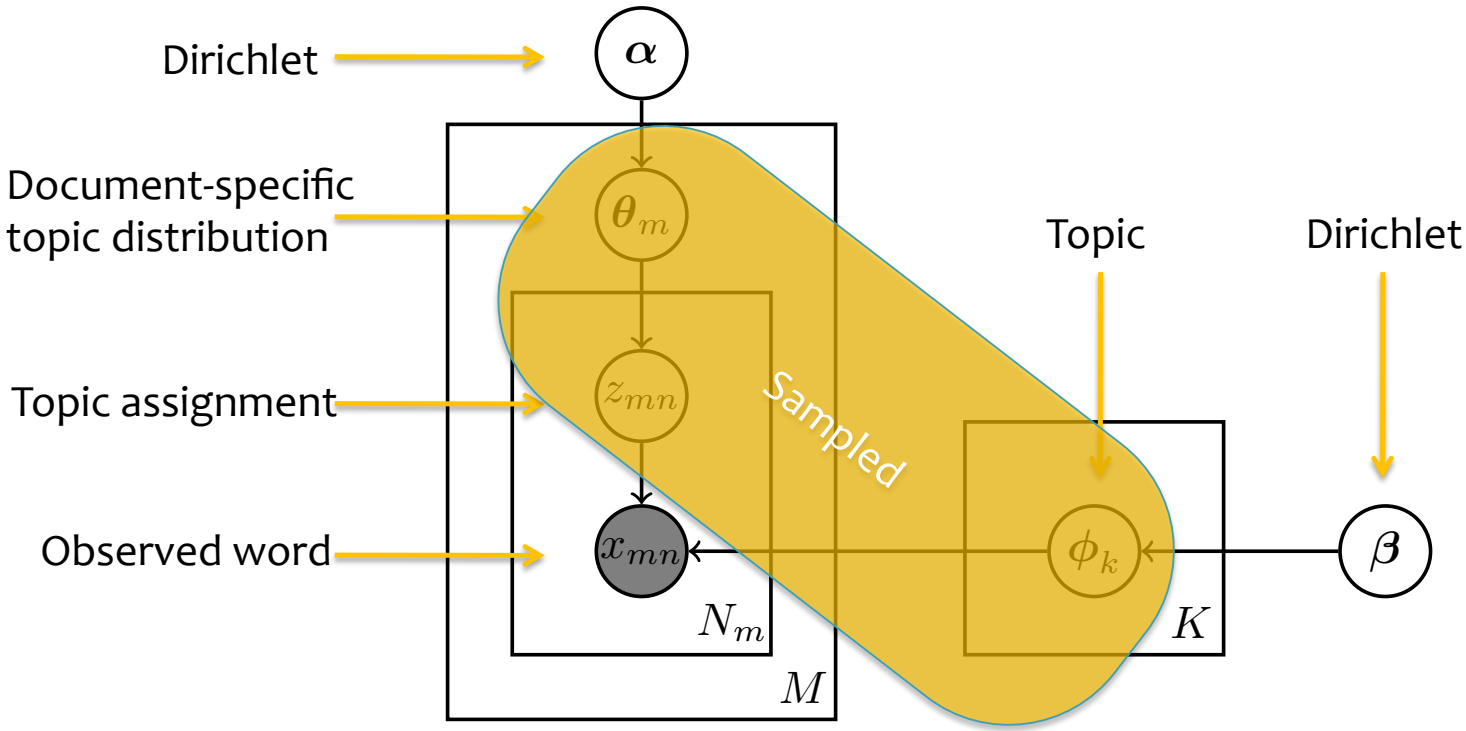
Exact Inference in LDA

- Exactly computing the posterior is intractable in LDA
  - Junction tree algorithm: exact inference in general graphical models
    1. "moralization" converts directed to undirected
    2. "triangulation" breaks 4-cycles by adding edges
    3. Cliques arranged into a junction tree
  - Time complexity is exponential in size of cliques
  - LDA cliques will be large (at least O(# topics)), so complexity is $O(2^{\#\text{ topics}})$
- Exact MAP inference in LDA is NP-hard for a large number of topics (Sontag & Roy, 2011)

A slide from: Matt Gormley, Carnegie-Mellon University 10-710 (2016), Lecture 20: Topic Modelling

LDA Inference

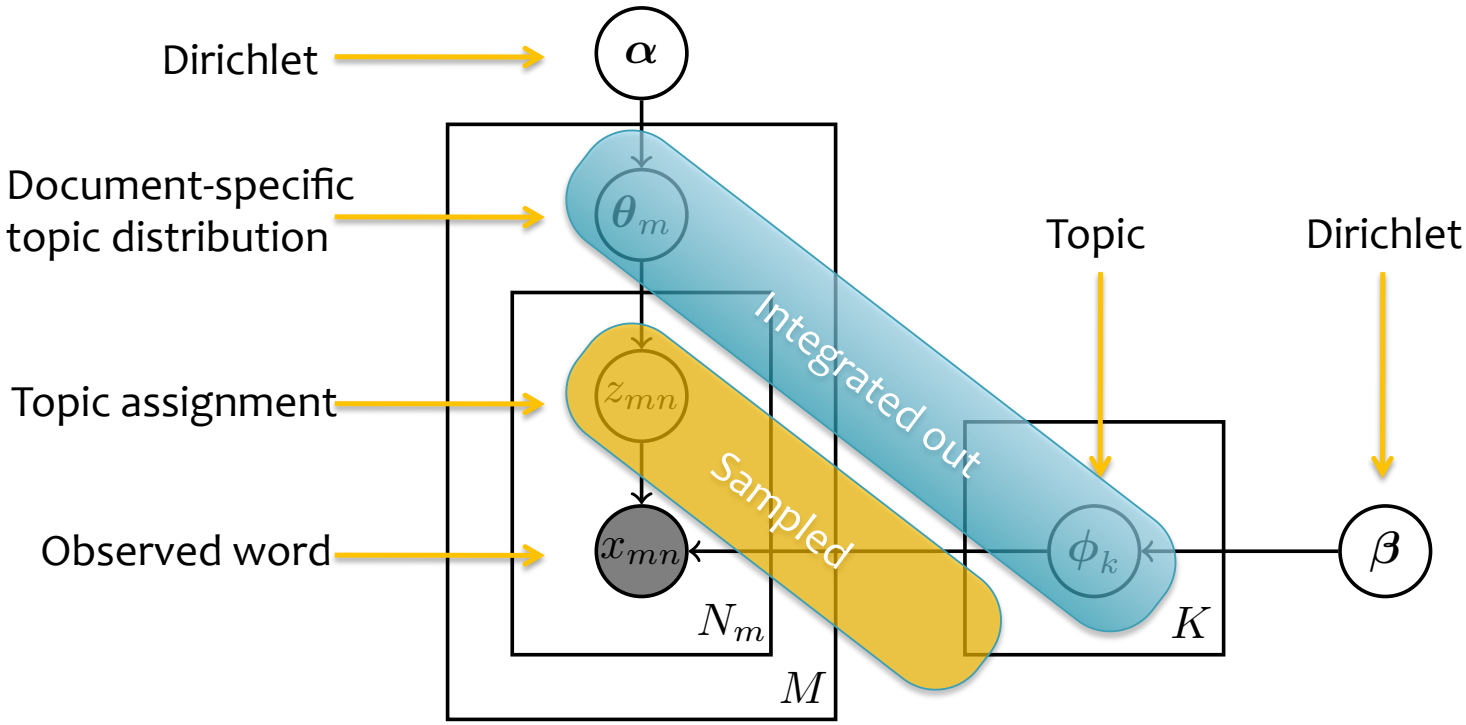
- Explicit Gibbs Sampler



A slide from: Matt Gormley, Carnegie-Mellon University 10-710 (2016), Lecture 20: Topic Modelling

LDA Inference

- Collapsed Gibbs Sampler



A slide from: Matt Gormley, Carnegie-Mellon University 10-710 (2016), Lecture 20: Topic Modelling

Sampling

Goal:

- Draw samples from the posterior $p(Z|X, \alpha, \beta)$
- Integrate out topics ϕ and document-specific distribution over topics θ

Algorithm:

- While not done…
  - For each document, m:
    - For each word, n:
      - » Resample a single topic assignment using the full conditionals for z_{mn}

A slide from: Matt Gormley, Carnegie-Mellon University 10-710 (2016), Lecture 20: Topic Modelling

Sampling

- What can we do with samples of z_{mn}?
  - Mean of z_{mn}
  - Mode of z_{mn}
  - Estimate posterior over z_{mn}
  - Estimate of topics ϕ and document-specific distribution over topics θ

A slide from: Matt Gormley, Carnegie-Mellon University 10-710 (2016), Lecture 20: Topic Modelling

Gibbs Sampling for LDA

- Full conditionals

$$p(z_i = k|Z^{-i}, X, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{n_{kt}^{-i} + \beta_t}{\sum_{v=1}^{T} n_{kv}^{-i} + \beta_v} \cdot \frac{n_{mk}^{-i} + \alpha_k}{\sum_{j=1}^{K} n_{mj}^{-i} + \alpha_j}$$

where t, m are given by i

n_{kt} = # times topic k appears with type t

n_{mk} = # times topic k appears in document m

A slide from: Matt Gormley, Carnegie-Mellon University 10-710 (2016), Lecture 20: Topic Modelling

Gibbs Sampling for LDA

- Sketch of the derivation of the full conditionals

$$
\begin{aligned}
p(z_i = k|Z^{-i}, X, \boldsymbol{\alpha}, \boldsymbol{\beta}) &= \frac{p(X, Z|\boldsymbol{\alpha}, \boldsymbol{\beta})}{p(X, Z^{-i}|\boldsymbol{\alpha}, \boldsymbol{\beta})} \\
&\propto p(X, Z|\boldsymbol{\alpha}, \boldsymbol{\beta}) \\
&= p(X|Z, \boldsymbol{\beta})p(Z|\boldsymbol{\alpha}) \\
&= \int_{\Phi} p(X|Z, \Phi)p(\Phi|\boldsymbol{\beta})\, d\Phi \int_{\Theta} p(Z|\Theta)p(\Theta|\boldsymbol{\alpha})\, d\Theta \\
&= \left(\prod_{k=1}^{K} \frac{B(\vec{n}_k + \boldsymbol{\beta})}{B(\boldsymbol{\beta})}\right)\left(\prod_{m=1}^{M} \frac{B(\vec{n}_m + \boldsymbol{\alpha})}{B(\boldsymbol{\alpha})}\right) \\
&= \frac{n_{kt}^{-i} + \beta_t}{\sum_{v=1}^{T} n_{kv}^{-i} + \beta_v} \cdot \frac{n_{mk}^{-i} + \alpha_k}{\sum_{j=1}^{K} n_{mj}^{-i} + \alpha_j} \\
&\quad \text{where } t, m \text{ are given by } i
\end{aligned}
$$

A slide from: Matt Gormley, Carnegie-Mellon University 10-710 (2016), Lecture 20: Topic Modelling

Dirichlet-Multinomial Model

- The Dirichlet is conjugate to the Multinomial

$\boldsymbol{\phi} \sim \text{Dir}(\boldsymbol{\beta})$ [*draw distribution over words*]
For each word $n \in \{1, \ldots, N\}$
$x_n \sim \text{Mult}(1, \boldsymbol{\phi})$ [*draw word*]

- The posterior of ϕ is $p(\phi|X) = \frac{p(X|\phi)p(\phi)}{P(X)}$
- Define the count vector $\boldsymbol{n}$ such that n_t denotes the number of times word t appeared
- Then the posterior is also a Dirichlet distribution: $p(\phi|X) \sim \text{Dir}(\boldsymbol{\beta} + \boldsymbol{n})$

A slide from: Matt Gormley, Carnegie-Mellon University 10-710 (2016), Lecture 20: Topic Modelling

Dirichlet-Multinomial Model

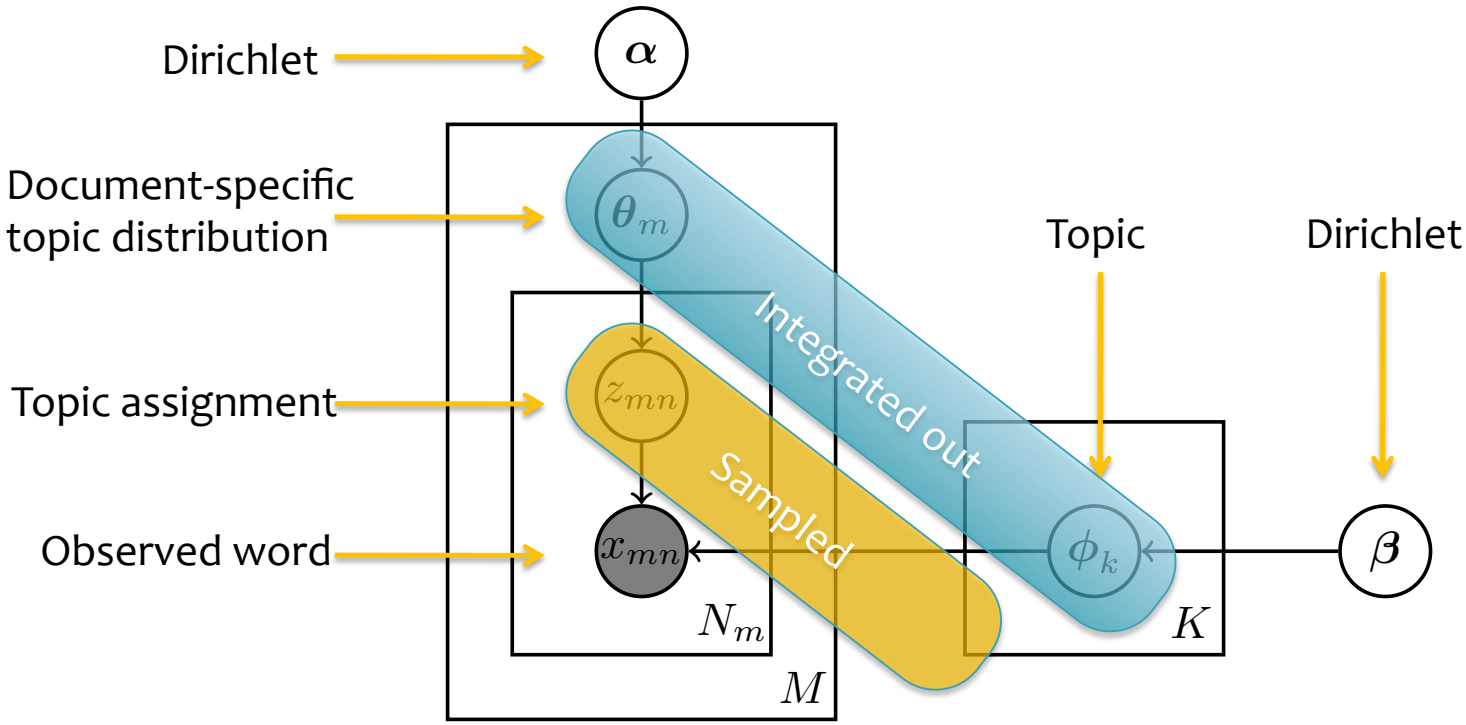
- Why conjugacy is so useful

$$
\begin{aligned}
p(X|\boldsymbol{\alpha}) &= \int_{\phi} p(X|\vec{\phi})p(\vec{\phi}|\boldsymbol{\alpha})\, d\phi \\
&= \int_{\phi} \left(\prod_{v=1}^{V} \phi_v^{n_v} \right) \left(\frac{1}{B(\boldsymbol{\alpha})} \prod_{v=1}^{V} \phi_v^{\alpha_v - 1} \right) d\phi \\
&= \frac{1}{B(\boldsymbol{\alpha})} \int_{\phi} \prod_{v=1}^{V} \phi_v^{n_v + \alpha_v - 1}\, d\phi \\
&= \frac{1}{B(\boldsymbol{\alpha})} \int_{\phi} \frac{B(\vec{n} + \boldsymbol{\alpha})}{B(\vec{n} + \boldsymbol{\alpha})} \prod_{v=1}^{V} \phi_v^{n_v + \alpha_v - 1}\, d\phi \\
&= \frac{B(\vec{n} + \boldsymbol{\alpha})}{B(\boldsymbol{\alpha})} \int_{\phi} \underbrace{\frac{1}{B(\vec{n} + \boldsymbol{\alpha})} \prod_{v=1}^{V} \phi_v^{n_v + \alpha_v - 1}}_{Dir(\vec{n} + \boldsymbol{\alpha})}\, d\phi \\
&= \frac{B(\vec{n} + \boldsymbol{\alpha})}{B(\boldsymbol{\alpha})}
\end{aligned}
$$

A slide from: Matt Gormley, Carnegie-Mellon University 10-710 (2016), Lecture 20: Topic Modelling

LDA Inference

- Collapsed Gibbs Sampler



A slide from: Matt Gormley, Carnegie-Mellon University 10-710 (2016), Lecture 20: Topic Modelling

Gibbs Sampling for LDA

Algorithm

`// initialisation`

zero all count variables, $n_m^{(k)}, n_m, n_k^{(t)}, n_k$

for all documents $m \in [1, M]$ **do**

  for all words $n \in [1, N_m]$ in document m **do**

    sample topic index $z_{m,n}=k \sim \text{Mult}(1/K)$

    increment document–topic count: $n_m^{(k)}$ += 1

    increment document–topic sum: n_m += 1

    increment topic–term count: $n_k^{(t)}$ += 1

    increment topic–term sum: n_k += 1

Figure from Heinrich (2009)

A slide from: Matt Gormley, Carnegie-Mellon University 10-710 (2016), Lecture 20: Topic Modelling

Gibbs Sampling for LDA

Algorithm

// Gibbs sampling over burn-in period and sampling period

while not finished **do**

  for all documents $m \in [1, M]$ **do**

    for all words $n \in [1, N_m]$ in document m **do**

      // for the current assignment of k to a term t for word $w_{m,n}$:

      decrement counts and sums: $n_m^{(k)} -= 1; n_m -= 1; n_k^{(t)} -= 1; n_k -= 1$

      // multinomial sampling acc. to Eq. 78 (decrements from previous step):

      sample topic index $\tilde{k} \sim p(z_i | \vec{z}_{\neg i}, \vec{w})$

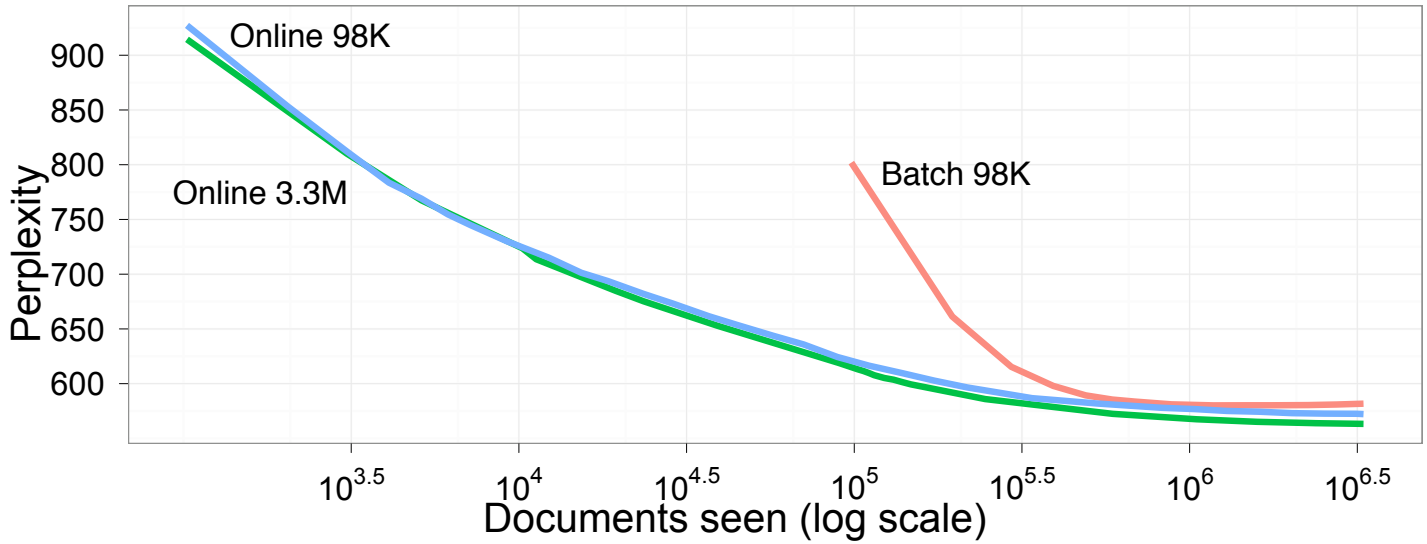
      // for the new assignment of $z_{m,n}$ to the term t for word $w_{m,n}$:

      increment counts and sums: $n_m^{(\tilde{k})} += 1; n_m += 1; n_{\tilde{k}}^{(t)} += 1; n_{\tilde{k}} += 1$

Figure from Heinrich (2009)

A slide from: Matt Gormley, Carnegie-Mellon University 10-710 (2016), Lecture 20: Topic Modelling

Online Variational Inference for LDA



Documents analyzed	**2048**	**4096**	**8192**	**12288**	**16384**	**32768**	**49152**	**65536**
Top eight words	systems road made service announced national west language	systems health communication service billion language care road	service systems health companies market communication company billion	service systems companies business company billion health industry	service companies systems business company industry market billion	business service companies industry company management systems services	business service companies industry services company management public	business industry service companies services company management public

78

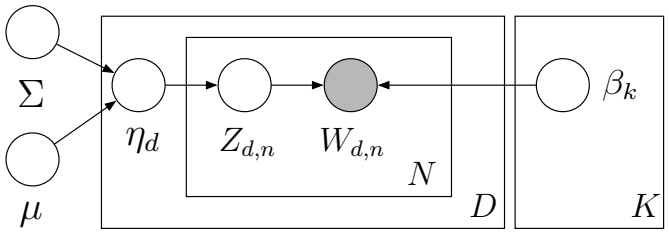
A slide from: Matt Gormley, Carnegie-Mellon University 10-710 (2016), Lecture 20: Topic Modelling

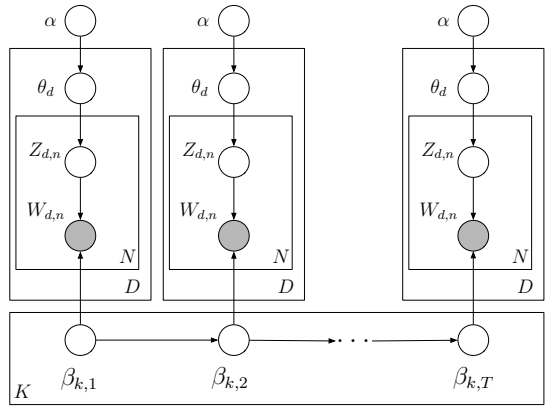
Content

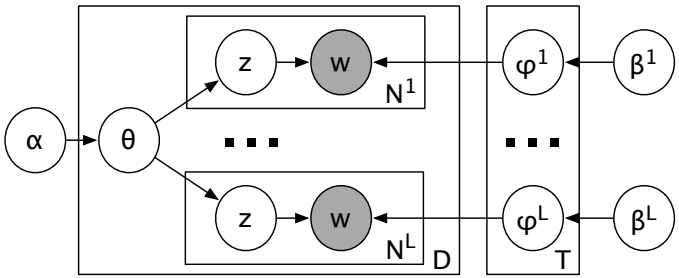
Extensions to the LDA Model

- Correlated topic models
  - Logistic normal prior over topic assignments
- Dynamic topic models
  - Learns topic changes over time
- Polylingual topic models
  - Learns topics aligned across multiple languages

...

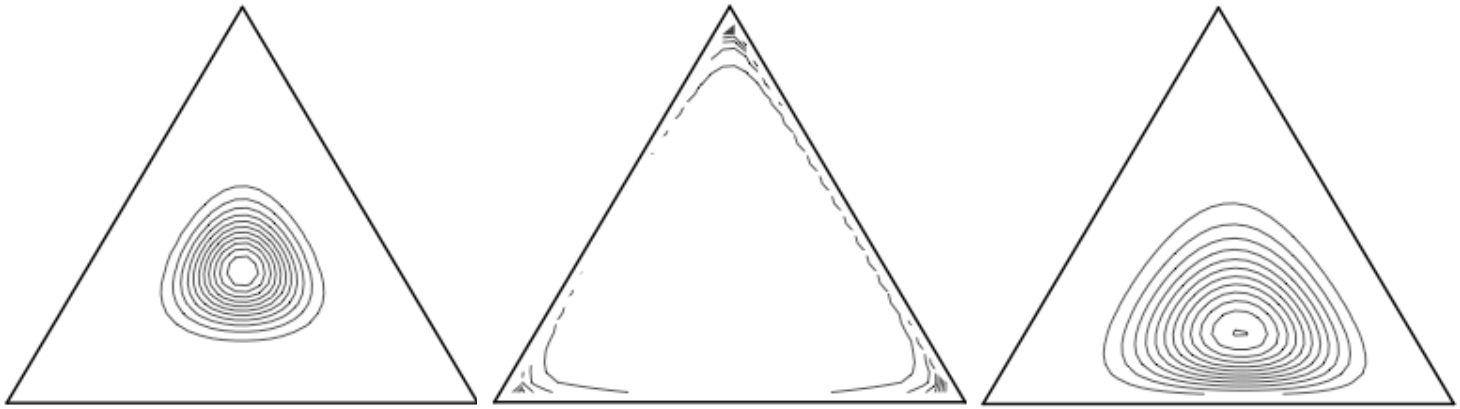






A slide from: Matt Gormley, Carnegie-Mellon University 10-710 (2016), Lecture 20: Topic Modelling

Correlated Topic Models



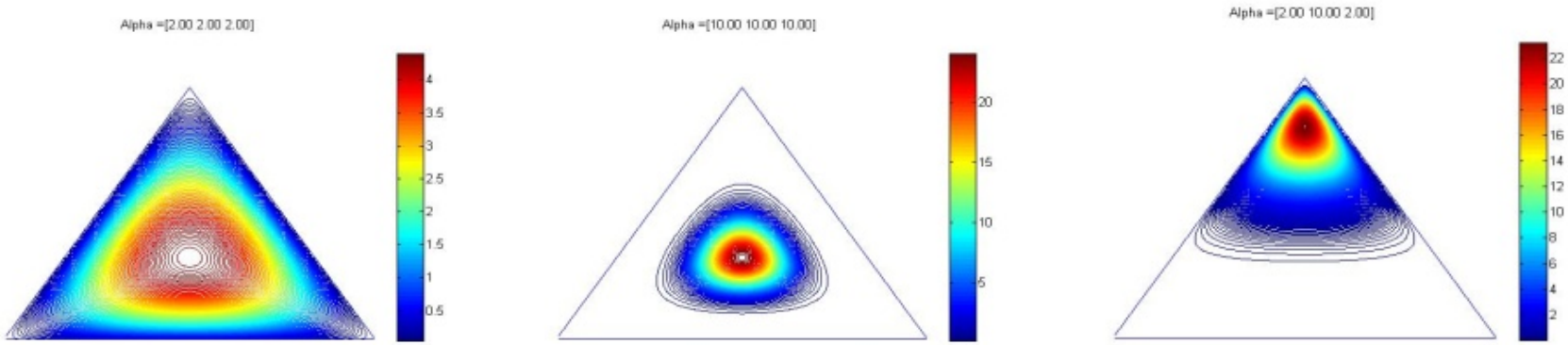
- The Dirichlet is a distribution on the simplex, positive vectors that sum to 1.
- It assumes that components are nearly independent.
- In real data, an article about *fossil fuels* is more likely to also be about *geology* than about *genetics*.

Slide from David Blei, MLSS 2012

81

A slide from: Matt Gormley, Carnegie-Mellon University 10-710 (2016), Lecture 20: Topic Modelling

Correlated Topic Models



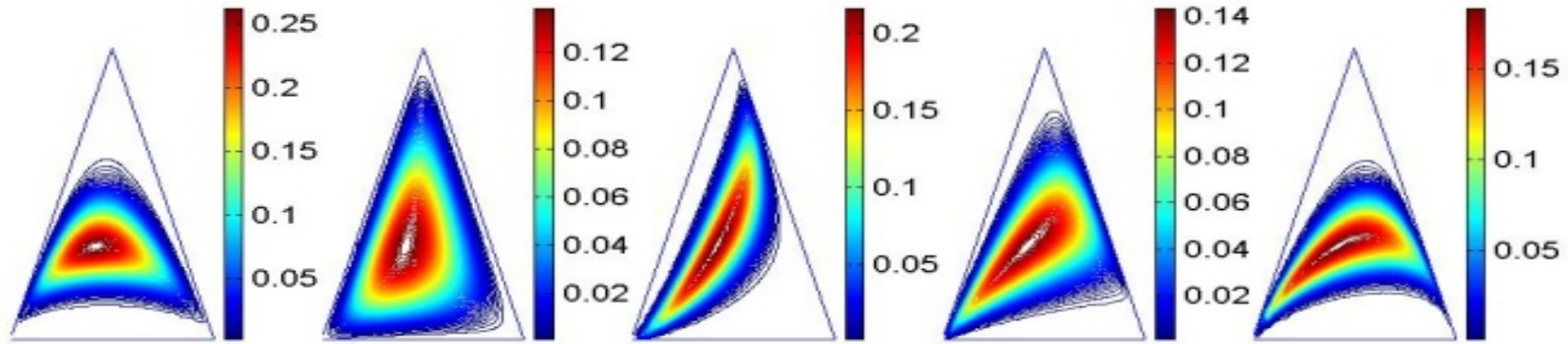
- The Dirichlet is a distribution on the simplex, positive vectors that sum to 1.
- It assumes that components are nearly independent.
- In real data, an article about *fossil fuels* is more likely to also be about *geology* than about *genetics*.

Slide from David Blei, MLSS 2012

82

A slide from: Matt Gormley, Carnegie-Mellon University 10-710 (2016), Lecture 20: Topic Modelling

Correlated Topic Models



- The **logistic normal** is a distribution on the simplex that can model dependence between components (Aitchison, 1980).
- The log of the parameters of the multinomial are drawn from a multivariate Gaussian distribution,

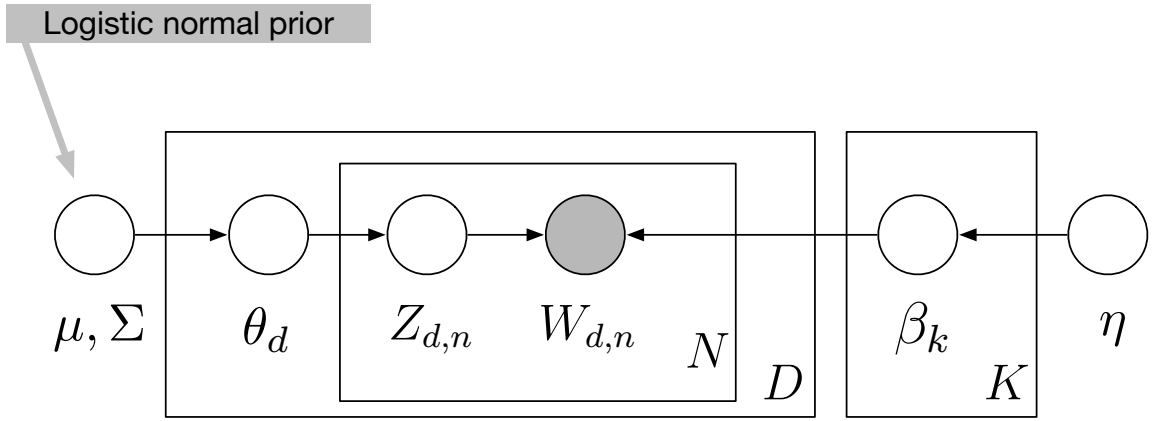
$$\begin{aligned} X &\sim \mathcal{N}_K(\mu, \Sigma) \\ \theta_i &\propto \exp\{x_i\}. \end{aligned}$$

Slide from David Blei, MLSS 2012

83

A slide from: Matt Gormley, Carnegie-Mellon University 10-710 (2016), Lecture 20: Topic Modelling

Correlated Topic Models



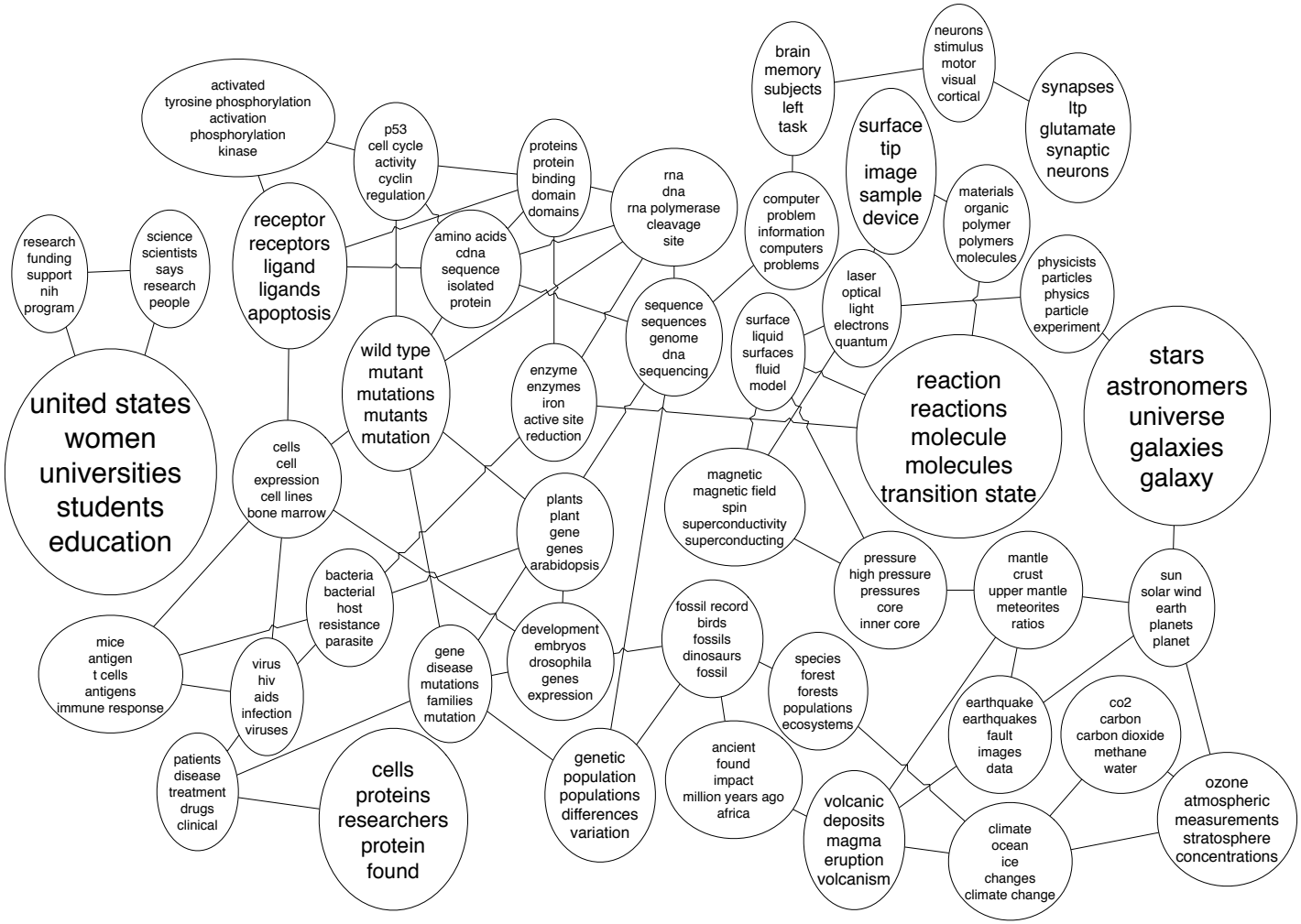
- Draw topic proportions from a logistic normal
- This allows topic occurrences to exhibit correlation.
- Provides a "map" of topics and how they are related
- Provides a better fit to text data, but computation is more complex

Slide from David Blei, MLSS 2012

84

A slide from: Matt Gormley, Carnegie-Mellon University 10-710 (2016), Lecture 20: Topic Modelling

Correlated Topic Models



Slide from David Blei, MLSS 2012

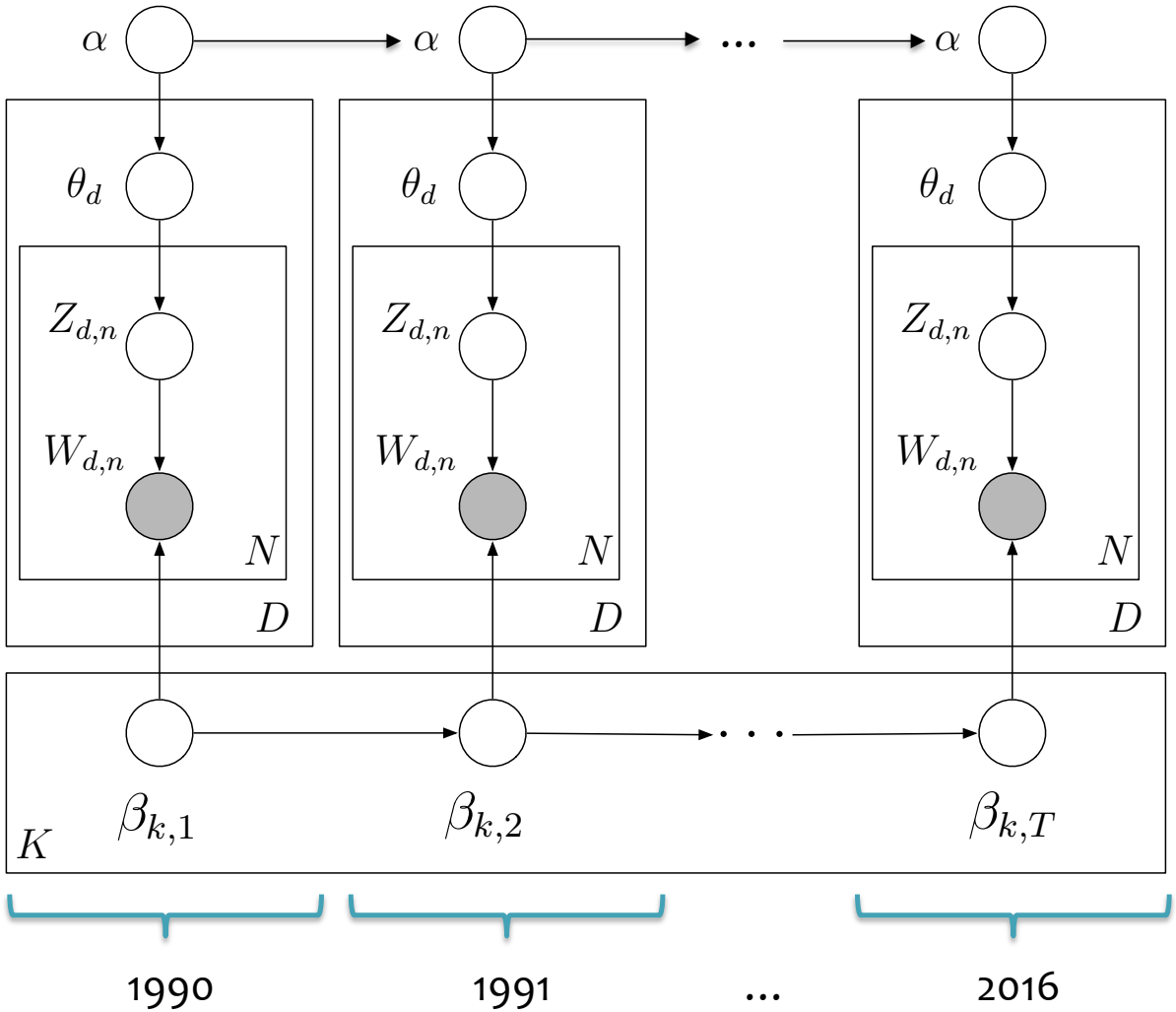
85

A slide from: Matt Gormley, Carnegie-Mellon University 10-710 (2016), Lecture 20: Topic Modelling

Dynamic Topic Models

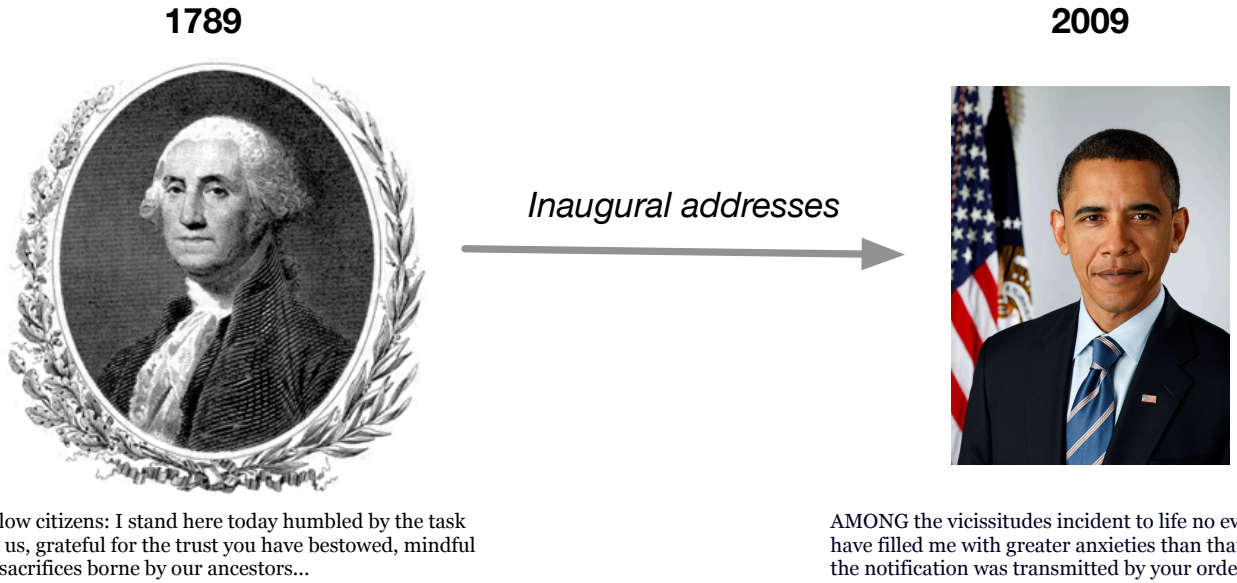
High-level idea:

- Divide the documents up by year
- Start with a separate topic model for each year
- Then add a dependence of each year on the previous one



A slide from: Matt Gormley, Carnegie-Mellon University 10-710 (2016), Lecture 20: Topic Modelling

Dynamic Topic Models



- LDA assumes that the order of documents does not matter.
- Not appropriate for sequential corpora (e.g., that span hundreds of years)
- Further, we may want to track how language changes over time.
- Dynamic topic models let the topics *drift* in a sequence.

Slide from David Blei, MLSS 2012

87

A slide from: Matt Gormley, Carnegie-Mellon University 10-710 (2016), Lecture 20: Topic Modelling

Dynamic Topic Models

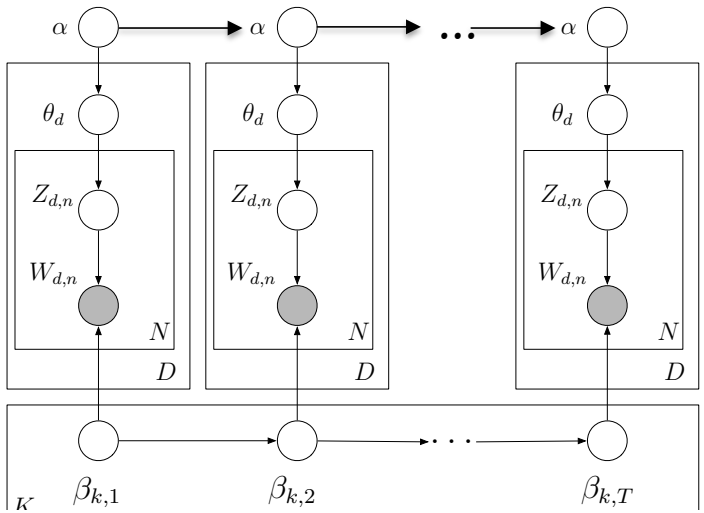
Generative Story

1. Draw topics $\beta_t \mid \beta_{t-1} \sim \mathcal{N}(\beta_{t-1}, \sigma^2 I)$.
2. Draw $\alpha_t \mid \alpha_{t-1} \sim \mathcal{N}(\alpha_{t-1}, \delta^2 I)$.
3. For each document:
   (a) Draw $\eta \sim \mathcal{N}(\alpha_t, a^2 I)$
   (b) For each word:
      i. Draw $Z \sim \mathit{Mult}(\pi(\eta))$.
      ii. Draw $W_{t,d,n} \sim \mathit{Mult}(\pi(\beta_{t,z}))$.

Logistic-normal priors

The pi function maps from the natural parameters to the mean parameters:

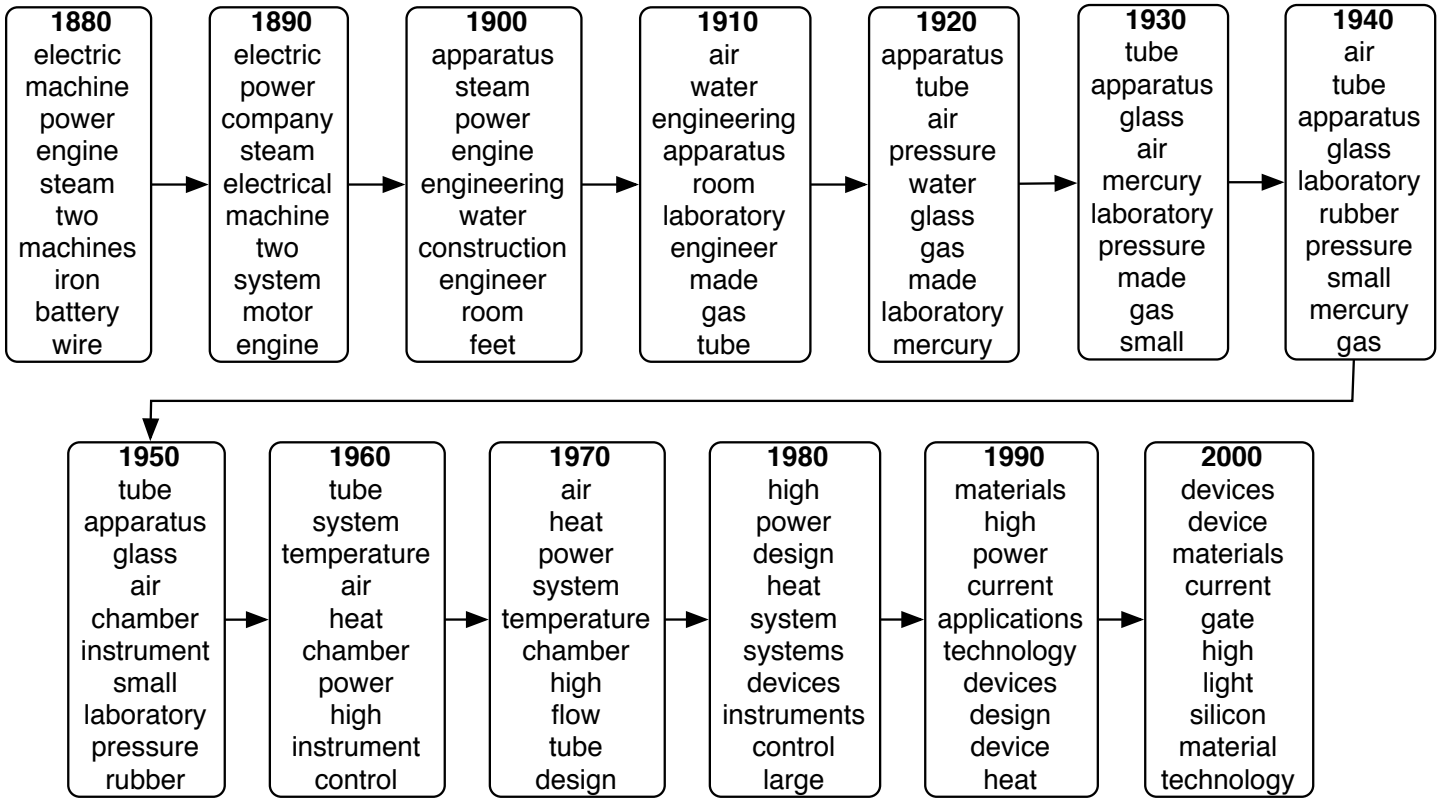
$$\pi(\beta_{k,t})_w = \frac{\exp(\beta_{k,t,w})}{\sum_w \exp(\beta_{k,t,w})}.$$



A slide from: Matt Gormley, Carnegie-Mellon University 10-710 (2016), Lecture 20: Topic Modelling

Dynamic Topic Models

Top ten most likely words in a "drifting" topic shown at 10-year increments

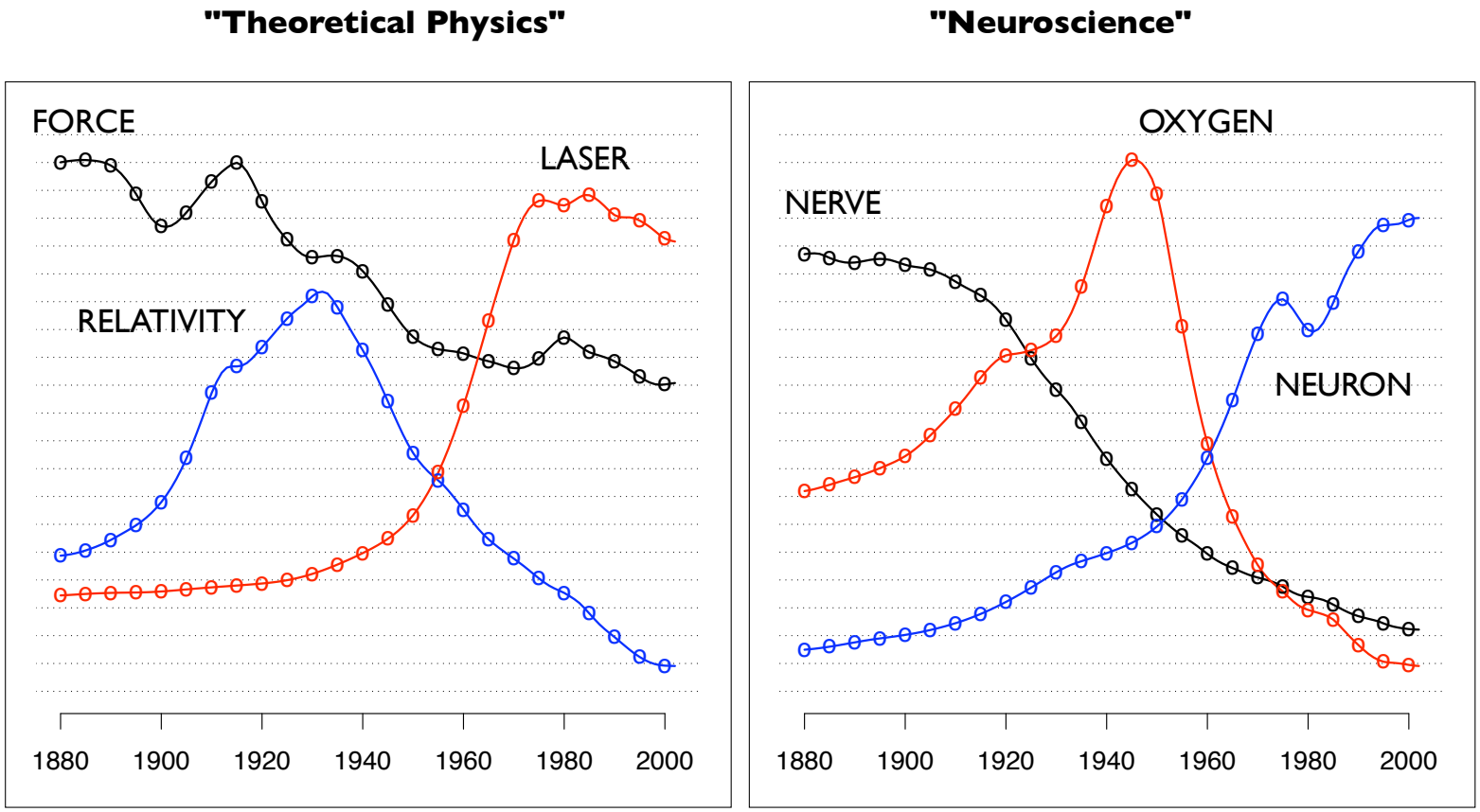


89

A slide from: Matt Gormley, Carnegie-Mellon University 10-710 (2016), Lecture 20: Topic Modelling

Dynamic Topic Models

Posterior estimate of **word frequency as a function of year** for three words each in two separate topics:

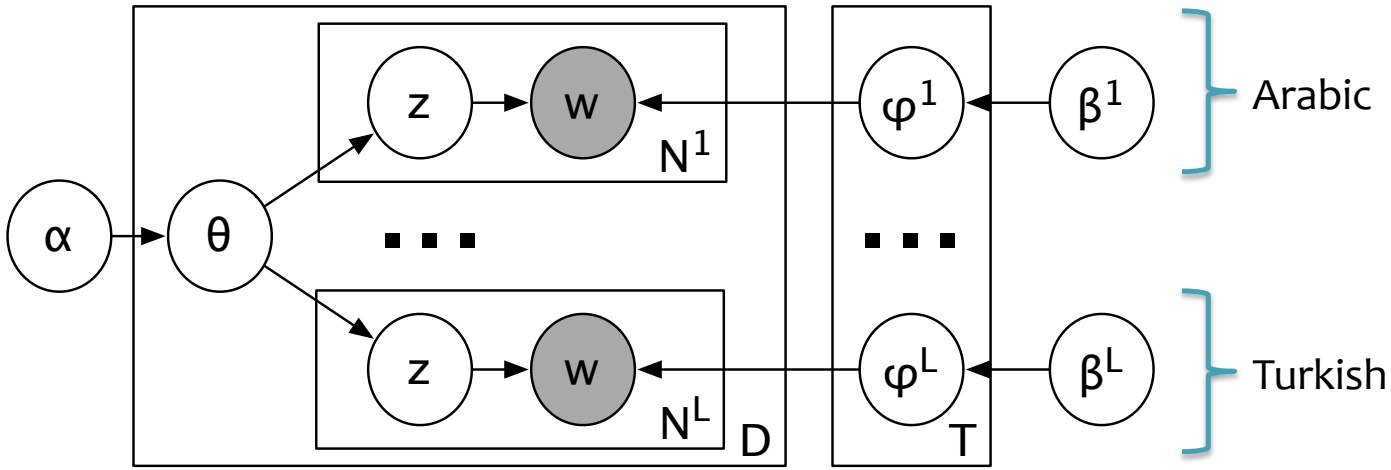


90

A slide from: Matt Gormley, Carnegie-Mellon University 10-710 (2016), Lecture 20: Topic Modelling

Polylingual Topic Models

- **Data Setting:** Comparable versions of each document exist in multiple languages (e.g. the Wikipedia article for "Barak Obama" in twelve languages)
- **Model:** Very similar to LDA, except that the topic assignments, z, and words, w, are sampled separately for each language.



91

A slide from: Matt Gormley, Carnegie-Mellon University 10-710 (2016), Lecture 20: Topic Modelling

Polylingual Topic Models

Topic 1 (twelve languages)

CY	sadwrn blaned gallair at lloeren mytholeg
DE	space nasa sojus flug mission
EL	διαστημικό sts nasa αγγλ small
EN	**space mission launch satellite nasa spacecraft**
FA	فضایی ماموریت ناسا مدار فضانورد ماهواره
FI	sojuz nasa apollo ensimmäinen space lento
FR	spatiale mission orbite mars satellite spatial
HE	החלל הארץ חלל כדור א תוכנית
IT	spaziale missione programma space sojuz stazione
PL	misja kosmicznej stacji misji space nasa
RU	космический союз космического спутник станции
TR	uzay soyuz ay uzaya salyut sovyetler

92

A slide from: Matt Gormley, Carnegie-Mellon University 10-710 (2016), Lecture 20: Topic Modelling

Polylingual Topic Models

Topic 2 (twelve languages)

CY	sbaen madrid el la josé sbaeneg
DE	de spanischer spanischen spanien madrid la
EL	ισπανίας ισπανία de ισπανός ντε μαδρίτη
EN	**de spanish spain la madrid y**
FA	اسپانیا اسپانیایی کوبا مادرید de ترین
FI	espanja de espanjan madrid la real
FR	espagnol espagne madrid espagnole juan y
HE	ספרד ספרדית דה מדריד הספרדית קובה
IT	de spagna spagnolo spagnola madrid el
PL	de hiszpański hiszpanii la juan y
RU	де мадрид испании испания испанский de
TR	ispanya ispanyol madrid la küba real

93

A slide from: Matt Gormley, Carnegie-Mellon University 10-710 (2016), Lecture 20: Topic Modelling

Polylingual Topic Models

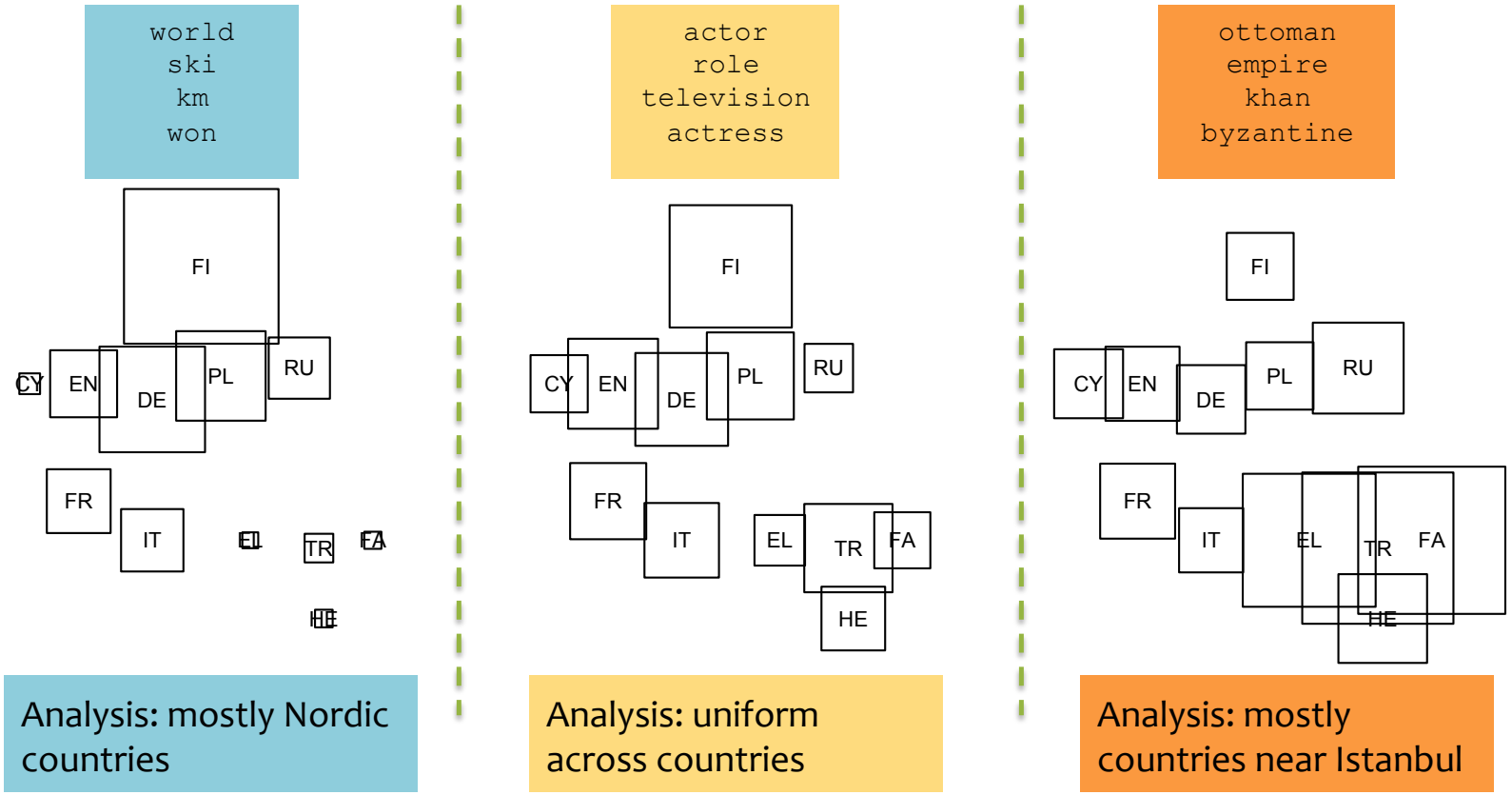
Topic 3 (twelve languages)

CY	bardd gerddi iaith beirdd fardd gymraeg
DE	dichter schriftsteller literatur gedichte gedicht werk
EL	ποιητής ποίηση ποιητή έργο ποιητές ποιήματα
EN	**poet poetry literature literary poems poem**
FA	شاعر شعر ادبیات فارسی ادبی آثار
FI	runoilija kirjailija kirjallisuuden kirjoitti runo julkaisi
FR	poète écrivain littérature poésie littéraire ses
HE	משורר ספרות שירה סופר שירים המשורר
IT	poeta letteratura poesia opere versi poema
PL	poeta literatury poezji pisarz in jego
RU	поэт его писатель литературы поэзии драматург
TR	şair edebiyat şiir yazar edebiyatı adlı

A slide from: Matt Gormley, Carnegie-Mellon University 10-710 (2016), Lecture 20: Topic Modelling

Polylingual Topic Models

Size of each square represents proportion of tokens assigned to the specified topic.



Analysis: mostly Nordic countries

Analysis: uniform across countries

Analysis: mostly countries near Istanbul

95

A slide from: Matt Gormley, Carnegie-Mellon University 10-710 (2016), Lecture 20: Topic Modelling

Supervised LDA

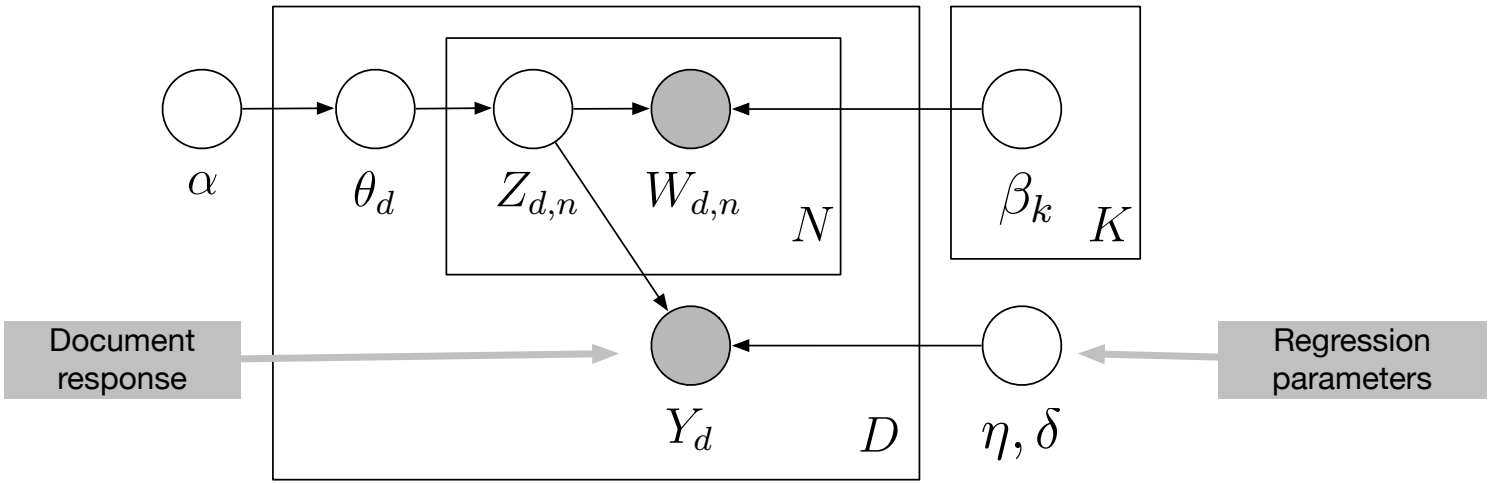
- LDA is an unsupervised model. How can we build a topic model that is good at the task we care about?
- Many data are paired with **response variables**.
  - User reviews paired with a number of stars
  - Web pages paired with a number of “likes”
  - Documents paired with links to other documents
  - Images paired with a category
- **Supervised LDA** are topic models of documents and responses. They are fit to find topics predictive of the response.

Slide from David Blei, MLSS 2012

96

A slide from: Matt Gormley, Carnegie-Mellon University 10-710 (2016), Lecture 20: Topic Modelling

Supervised LDA



1. Draw topic proportions $\theta \mid \alpha \sim \mathrm{Dir}(\alpha)$.
2. For each word
   - Draw topic assignment $z_n \mid \theta \sim \mathrm{Mult}(\theta)$.
   - Draw word $w_n \mid z_n, \beta_{1:K} \sim \mathrm{Mult}(\beta_{z_n})$.
3. Draw response variable $y \mid z_{1:N}, \eta, \sigma^2 \sim \mathrm{N}\left(\eta^\top \bar{z}, \sigma^2\right)$, where

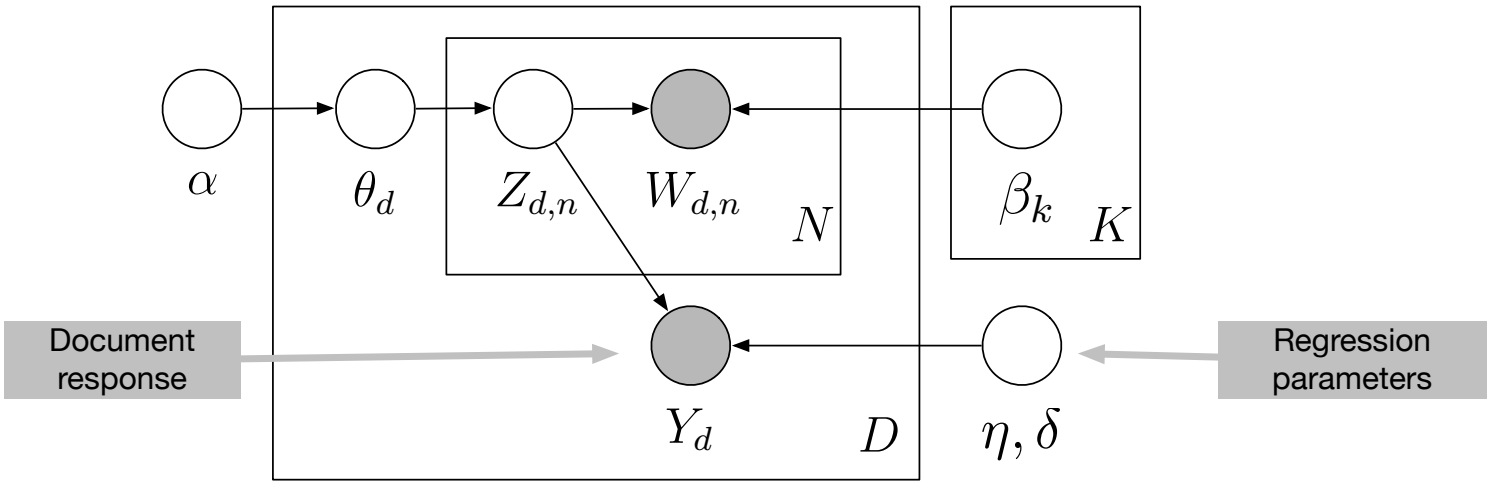
$$\bar{z} = (1/N) \sum_{n=1}^{N} z_n.$$

Slide from David Blei, MLSS 2012

97

A slide from: Matt Gormley, Carnegie-Mellon University 10-710 (2016), Lecture 20: Topic Modelling

Supervised LDA



- Fit sLDA parameters to documents and responses.
  This gives: topics $\beta_{1:K}$ and coefficients $\eta_{1:K}$.
- Given a new document, predict its response using the expected value:

$$\mathrm{E}\left[Y \mid w_{1:N}, \alpha, \beta_{1:K}, \eta, \sigma^2\right] = \eta^\top \mathrm{E}\left[\bar{Z} \mid w_{1:N}\right]$$

- This blends generative and discriminative modeling.

Slide from David Blei, MLSS 2012

A slide from: Matt Gormley, Carnegie-Mellon University 10-710 (2016), Lecture 20: Topic Modelling

What if we don't know the number of topics, K, ahead of time?

Take 10-708 to learn **Bayesian Nonparametrics**

- New modeling constructs:
  - Chinese Restaurant Process (Dirichlet Process)
  - Indian Buffet Process
- e.g. an **infinite number of topics** in a finite amount of space

99

A slide from: Matt Gormley, Carnegie-Mellon University 10-710 (2016), Lecture 20: Topic Modelling

Summary: Topic Modeling

- **The Task of Topic Modeling**
  - Topic modeling enables the **analysis of large** (possibly unannotated) **corpora**
  - Applicable to more than just bags of words
  - Extrinsic evaluations are often appropriate for these unsupervised methods
- **Constructing Models**
  - LDA is comprised of **simple building blocks** (Dirichlet, Multinomial)
  - LDA itself can act as a building block **for other models**
- **Approximate Inference**
  - Many different approaches to inference (and learning) can be applied to the same model

100

A slide from: Matt Gormley, Carnegie-Mellon University 10-710 (2016), Lecture 20: Topic Modelling

Content