Machine Learning Department
School of Computer Science
Carnegie Mellon University

Bayesian Inference for Parameter Estimation + Topic Modeling

Matt Gormley
Lecture 14
Oct. 24, 2022

Reminders

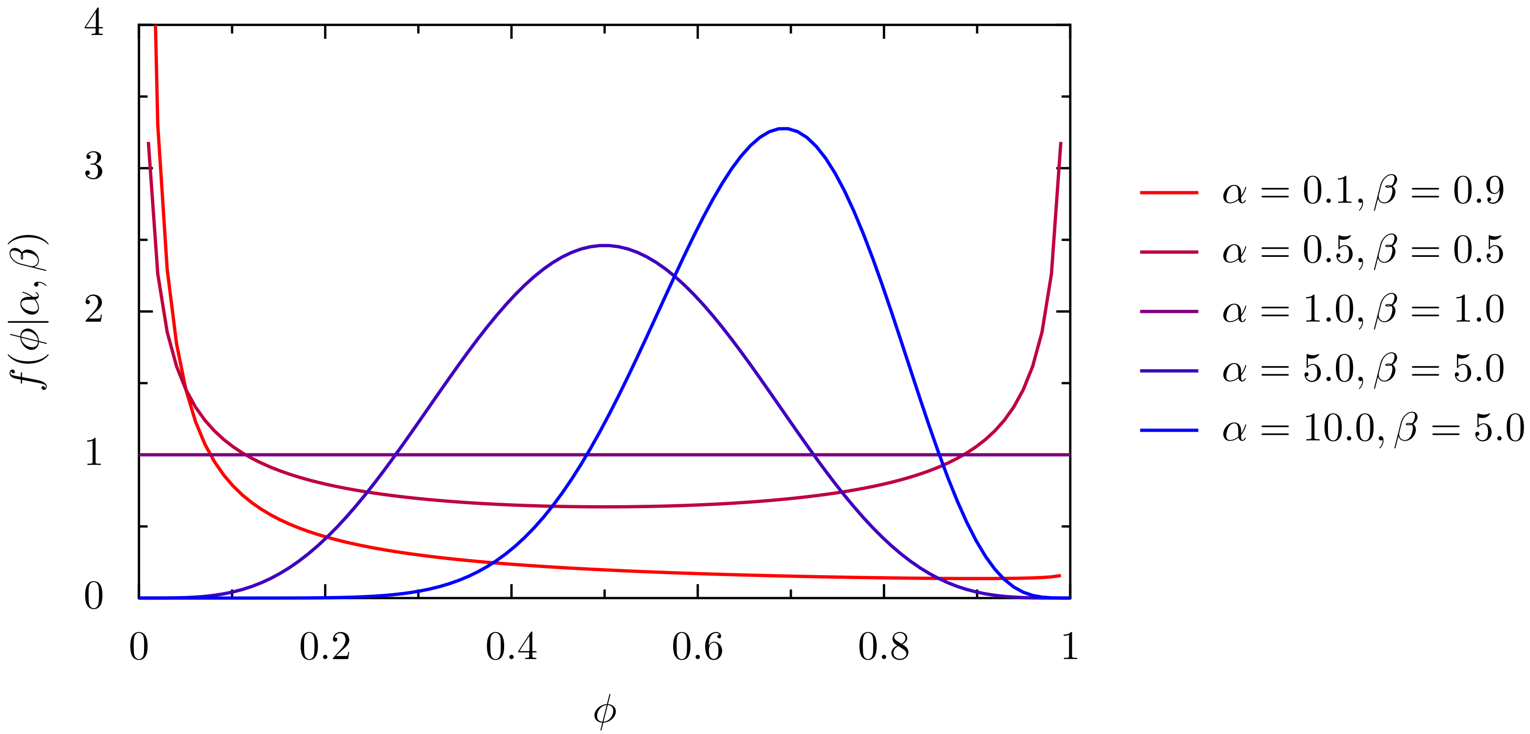
- Poll Questions 0a, 0b, 0c
- Grade Summary 1
- Homework 4: MCMC
  - Out: Mon, Oct 24
  - Due: Fri, Nov 3 at 11:59pm
- Recitation: Homework 4
  - today! 6pm, GHC 6121

BAYESIAN INFERENCE FOR NAÏVE BAYES

Beta-Bernoulli Model

- Beta Distribution

$$f(\phi|\alpha,\beta) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}$$



Beta-Bernoulli Model

- Generative Process

$\phi \sim \text{Beta}(\alpha, \beta)$	[*draw distribution over words*]
For each word $n \in \{1, \ldots, N\}$	
$\quad x_n \sim \text{Bernoulli}(\phi)$	[*draw word*]

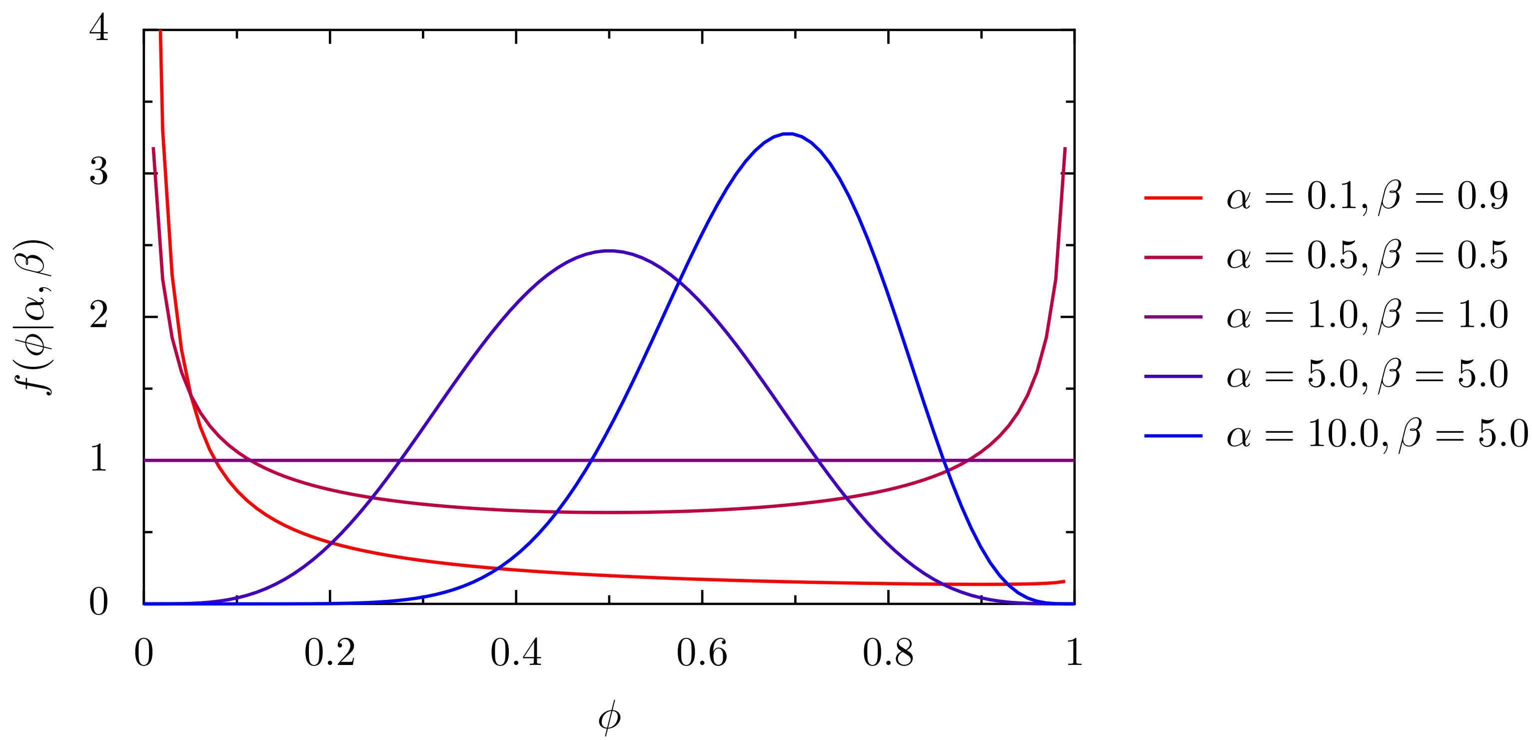
- Example corpus (heads/tails)

H	T	T	H	H	T	T	H	H	H
x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}

Dirichlet-Multinomial Model

- Dirichlet Distribution

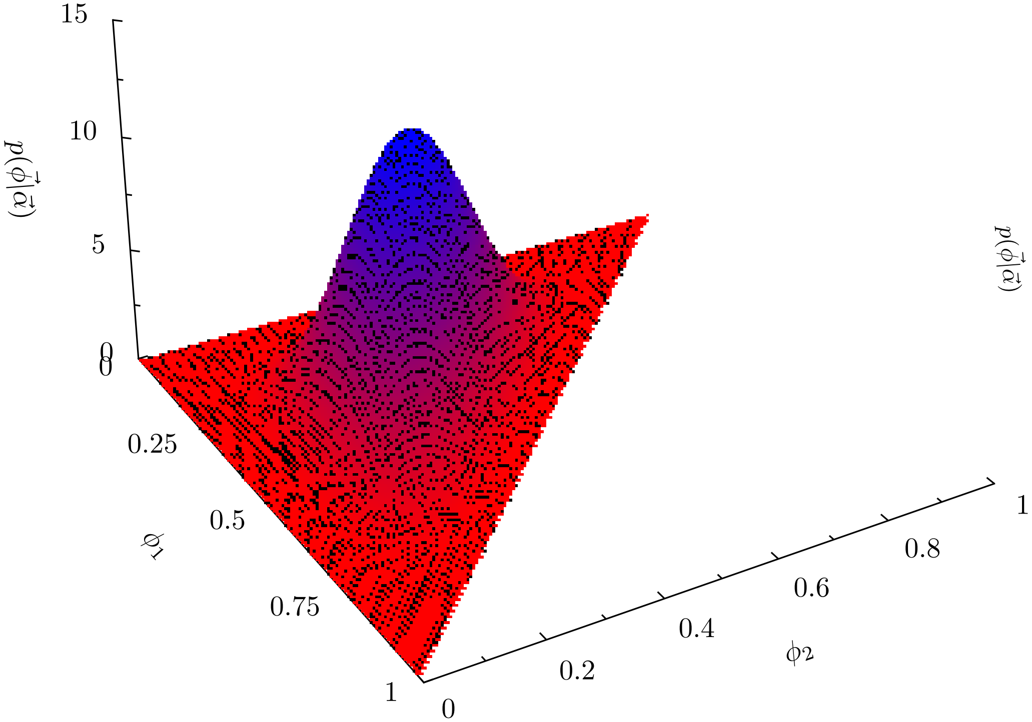
$$f(\phi|\alpha,\beta) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}$$

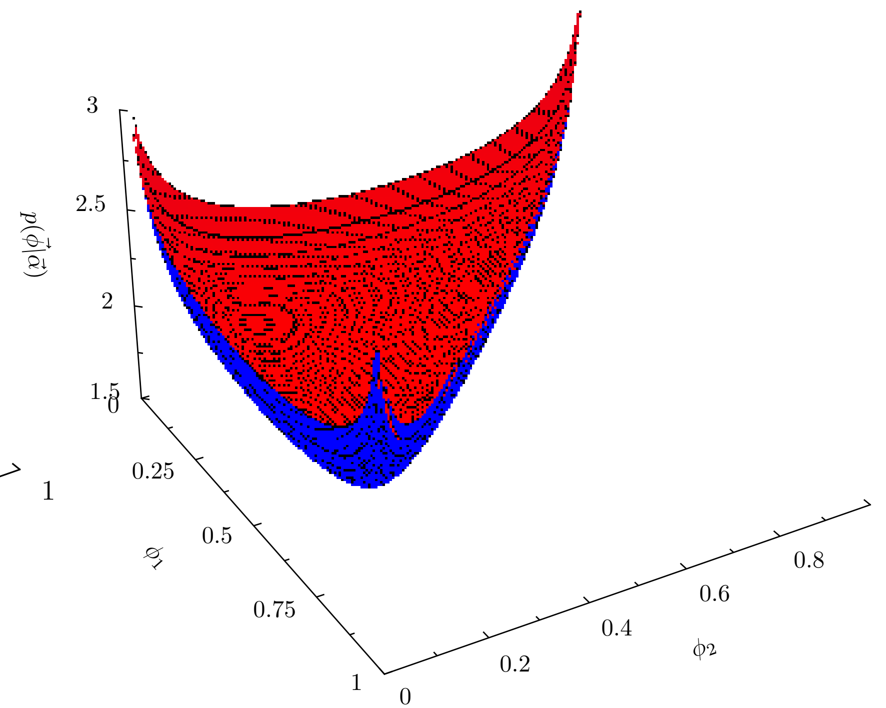


Dirichlet-Multinomial Model

- Dirichlet Distribution

$$p(\vec{\phi}|\boldsymbol{\alpha}) = \frac{1}{B(\boldsymbol{\alpha})} \prod_{k=1}^{K} \phi_k^{\alpha_k - 1} \quad \text{where } B(\boldsymbol{\alpha}) = \frac{\prod_{k=1}^{K} \Gamma(\alpha_k)}{\Gamma(\sum_{k=1}^{K} \alpha_k)}$$





Dirichlet-Multinomial Model

- Generative Process

$$\begin{aligned}
&\boldsymbol{\phi} \sim \mathrm{Dir}(\boldsymbol{\beta}) && \text{[\textit{draw distribution over words}]}\\
&\text{For each word } n \in \{1, \ldots, N\}\\
&\quad x_n \sim \mathrm{Mult}(1, \boldsymbol{\phi}) && \text{[\textit{draw word}]}
\end{aligned}$$

- Example corpus

the	he	is	the	and	the	she	she	is	is
x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}

Dirichlet-Multinomial Model

The Dirichlet is **conjugate** to the Multinomial

- 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: $p(\phi|X) \sim \text{Dir}(\boldsymbol{\beta} + \boldsymbol{n})$

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

$$p(\vec{\phi} \mid \vec{x}, \vec{\beta}) \propto p(\vec{x} \mid \vec{\phi})\, p(\vec{\phi})$$

$$= \left[\prod_{i=1}^{N} p(x^{(i)} \mid \vec{\phi})\right] p(\vec{\phi})$$

$$\propto \left[\prod_{k=1}^{K} \phi_k^{\beta_k - 1}\right] \left[\prod_{i=1}^{N} \prod_{k=1}^{K} \phi_k^{\mathbb{1}(x^{(i)} = k)}\right]$$

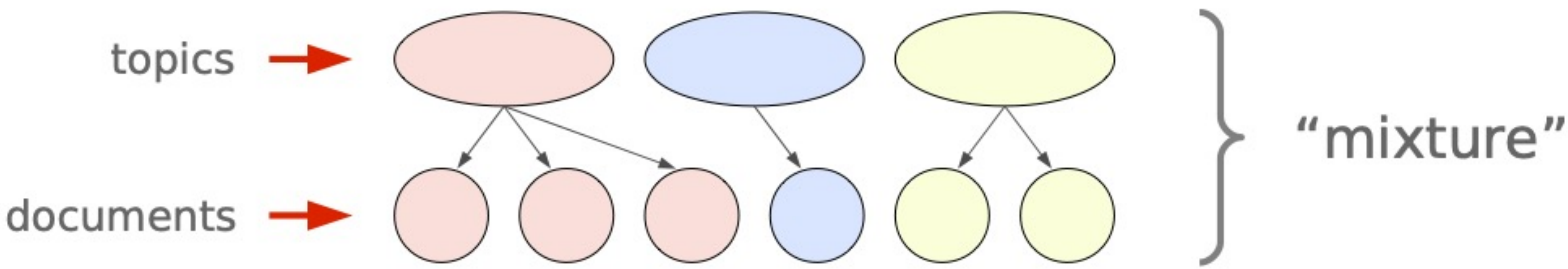
$$= \prod_{k=1}^{K} \phi_k^{\left[\beta_k - 1 + \sum_{i=1}^{N} \mathbb{1}(x^{(i)} = k)\right]}$$

$$\Rightarrow p(\vec{\phi} \mid \vec{x}, \vec{\beta}) \sim \text{Dirichlet}(\vec{\beta} + \vec{n})$$

where $n_k = \#$ times $x^{(i)} = k$

Dirichlet-Multinomial Mixture Model

- Generative Process



- Example corpus

the	he	is		the	and	the		she	she	is	is
x_{11}	x_{12}	x_{13}		x_{21}	x_{22}	x_{23}		x_{31}	x_{32}	x_{33}	x_{34}

Document 1 — Document 2 — Document 3

Figure from Wallach, JHU 2011, slides

Dirichlet-Multinomial Mixture Model

- Generative Process

For each topic $k \in \{1, \ldots, K\}$:
  $\boldsymbol{\phi}_k \sim \text{Dir}(\boldsymbol{\beta})$ [*draw distribution over words*]
$\boldsymbol{\theta} \sim \text{Dir}(\boldsymbol{\alpha})$ [*draw distribution over topics*]
For each document $m \in \{1, \ldots, M\}$
  $z_m \sim \text{Mult}(1, \boldsymbol{\theta})$ [*draw topic assignment*]
  For each word $n \in \{1, \ldots, N_m\}$
    $x_{mn} \sim \text{Mult}(1, \boldsymbol{\phi}_{z_m})$ [*draw word*]

- Example corpus

the	he	is
x_{11}	x_{12}	x_{13}

Document 1

the	and	the
x_{21}	x_{22}	x_{23}

Document 2

she	she	is	is
x_{31}	x_{32}	x_{33}	x_{34}

Document 3

Bayesian Inference

- The key idea behind **Bayesian Inference** is to treat your parameters as though they are like any other variables in your graphical model
- Bayesian Inference Summary:
  1. **Given:** data, D
  2. **Goal:** learn the posterior distribution over parameters θ given data D, i.e. $p(\theta \mid D)$
  3. **Store:** a distribution $p(\theta \mid D)$ as a probability mass function (pmf) or probability density function (pdf) or via some approximation
  4. **Afterwards:** marginalize over the parameters to work directly with other latent variables, e.g. $p(z \mid D) = \int_\theta p(z \mid \theta, D)\, p(\theta \mid D)\, d\theta$

Naïve Bayes Model

Data:

$$\mathcal{D} = \{(z^{(i)}, \mathbf{x}^{(i)}\}_{i=1}^{N}$$

where $z^{(i)} \in \{1, \ldots, L\}$,

$$\mathbf{x}^{(i)} \in \{1, \ldots, V\}^M$$

Generative Story:

$\boldsymbol{\theta} \sim \text{Dirichlet}(\boldsymbol{\alpha})$

$\boldsymbol{\phi}_k \sim \text{Dirichlet}(\boldsymbol{\beta}), \forall k$

$z^{(i)} \sim \text{Categorical}(\boldsymbol{\theta}), \forall i$

$x_m^{(i)} \sim \text{Categorical}(\boldsymbol{\phi}_{z^{(i)}}), \forall i, \forall m$

1) MLE:

$$\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\phi}} = \underset{\boldsymbol{\theta}, \boldsymbol{\phi}}{\operatorname{argmax}}\, p(\mathcal{D} \mid \boldsymbol{\theta}, \boldsymbol{\phi})$$

2) MAP Estimation:

$$\begin{aligned}\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\phi}} &= \underset{\boldsymbol{\theta}, \boldsymbol{\phi}}{\operatorname{argmax}}\, p(\mathcal{D} \mid \boldsymbol{\theta}, \boldsymbol{\phi}) p(\boldsymbol{\theta}, \boldsymbol{\phi}) \\ &= \underset{\boldsymbol{\theta}, \boldsymbol{\phi}}{\operatorname{argmax}}\, p(\boldsymbol{\theta}, \boldsymbol{\phi} \mid \mathcal{D})\end{aligned}$$

3) Bayesian Parameter Estimation:

$$p(\boldsymbol{\theta} \mid \mathcal{D}) = \text{Dirichlet}(\boldsymbol{\theta} \mid \boldsymbol{\alpha}')$$

$$p(\boldsymbol{\phi}_k \mid \mathcal{D}) = \text{Dirichlet}(\boldsymbol{\phi}_k \mid \boldsymbol{\beta}'_k)$$

Naïve Bayes Model

The standard presentation of Naïve Bayes (i.e. MLE and MAP Estimation) is not Bayesian!

1) MLE:

$$\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\phi}} = \operatorname*{argmax}_{\boldsymbol{\theta}, \boldsymbol{\phi}} p(\mathcal{D} \mid \boldsymbol{\theta}, \boldsymbol{\phi})$$

2) MAP Estimation:

$$\begin{aligned} \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\phi}} &= \operatorname*{argmax}_{\boldsymbol{\theta}, \boldsymbol{\phi}} p(\mathcal{D} \mid \boldsymbol{\theta}, \boldsymbol{\phi}) p(\boldsymbol{\theta}, \boldsymbol{\phi}) \\ &= \operatorname*{argmax}_{\boldsymbol{\theta}, \boldsymbol{\phi}} p(\boldsymbol{\theta}, \boldsymbol{\phi} \mid \mathcal{D}) \end{aligned}$$

3) Bayesian Parameter Estimation:

$$p(\boldsymbol{\theta} \mid \mathcal{D}) = \text{Dirichlet}(\boldsymbol{\theta} \mid \boldsymbol{\alpha}')$$

$$p(\boldsymbol{\phi}_k \mid \mathcal{D}) = \text{Dirichlet}(\boldsymbol{\phi}_k \mid \boldsymbol{\beta}'_k)$$

Naïve Bayes Model

Generative Story:

$$\boldsymbol{\theta} \sim \text{Dirichlet}(\boldsymbol{\alpha})$$

$$\boldsymbol{\phi}_k \sim \text{Dirichlet}(\boldsymbol{\beta}), \forall k$$

$$z^{(i)} \sim \text{Categorical}(\boldsymbol{\theta}), \forall i$$

$$x_m^{(i)} \sim \text{Categorical}(\boldsymbol{\phi}_{z^{(i)}}), \forall i, \forall m$$

1) MLE:

$$\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\phi}} = \operatorname*{argmax}_{\boldsymbol{\theta}, \boldsymbol{\phi}} p(\mathcal{D} \mid \boldsymbol{\theta}, \boldsymbol{\phi})$$

- pro: solved by counting

$$\theta_k \propto \sum_{i=1}^{N} \mathbb{1}(z^{(i)} = k)$$

$$\phi_{kv} \propto \sum_{i=1}^{N} \sum_{m=1}^{M} \mathbb{1}(x_m^{(i)} = v)\mathbb{1}(z^{(i)} = k)$$

- con: single point estimate
- con: ignores the priors over parameters

2) MAP Estimation:

$$\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\phi}} = \operatorname*{argmax}_{\boldsymbol{\theta}, \boldsymbol{\phi}} p(\mathcal{D} \mid \boldsymbol{\theta}, \boldsymbol{\phi}) p(\boldsymbol{\theta}, \boldsymbol{\phi})$$

$$= \operatorname*{argmax}_{\boldsymbol{\theta}, \boldsymbol{\phi}} p(\boldsymbol{\theta}, \boldsymbol{\phi} \mid \mathcal{D})$$

- pro: takes prior into account
- pro: solved by counting

$$\theta_k \propto (\alpha - 1) + \sum_{i=1}^{N} \mathbb{1}(z^{(i)} = k)$$

$$\phi_{kv} \propto (\beta - 1) + \sum_{i=1}^{N} \sum_{m=1}^{M} \mathbb{1}(x_m^{(i)} = v)\mathbb{1}(z^{(i)} = k)$$

- con: single point estimate

Naïve Bayes Model

1) MLE:

$$\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\phi}} = \operatorname*{argmax}_{\boldsymbol{\theta}, \boldsymbol{\phi}} p(\mathcal{D} \mid \boldsymbol{\theta}, \boldsymbol{\phi})$$

2) MAP Estimation:

$$\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\phi}} = \operatorname*{argmax}_{\boldsymbol{\theta}, \boldsymbol{\phi}} p(\mathcal{D} \mid \boldsymbol{\theta}, \boldsymbol{\phi}) p(\boldsymbol{\theta}, \boldsymbol{\phi})$$

$$= \operatorname*{argmax}_{\boldsymbol{\theta}, \boldsymbol{\phi}} p(\boldsymbol{\theta}, \boldsymbol{\phi} \mid \mathcal{D})$$

- pro: takes prior into account
- pro: solved by counting

$$\theta_k \propto (\alpha - 1) + \sum_{i=1}^{N} \mathbb{1}(z^{(i)} = k)$$

$$\phi_{kv} \propto (\beta - 1) + \sum_{i=1}^{N} \sum_{m=1}^{M} \mathbb{1}(x_m^{(i)} = v) \mathbb{1}(z^{(i)} = k)$$

- con: single point estimate

3) Bayesian Parameter Estimation:

$$p(\boldsymbol{\theta} \mid \mathcal{D}) = \text{Dirichlet}(\boldsymbol{\theta} \mid \boldsymbol{\alpha}')$$

$$p(\boldsymbol{\phi}_k \mid \mathcal{D}) = \text{Dirichlet}(\boldsymbol{\phi}_k \mid \boldsymbol{\beta}'_k)$$

- pro: takes uncertainty over parameters into account
- pro: compactly represented b/c of Dirichlet-Multinomial conjugacy

$$\alpha'_k = \alpha_k + \sum_{i=1}^{N} \mathbb{1}(z^{(i)} = k)$$

$$\beta'_{kv} = \beta_k + \sum_{i=1}^{N} \sum_{m=1}^{M} \mathbb{1}(x_m^{(i)} = v) \mathbb{1}(z^{(i)} = k)$$

Naïve Bayes Model

2) MAP Estimation:

$$\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\phi}} = \underset{\boldsymbol{\theta},\boldsymbol{\phi}}{\operatorname{argmax}}\, p(\mathcal{D} \mid \boldsymbol{\theta}, \boldsymbol{\phi}) p(\boldsymbol{\theta}, \boldsymbol{\phi})$$
$$= \underset{\boldsymbol{\theta},\boldsymbol{\phi}}{\operatorname{argmax}}\, p(\boldsymbol{\theta}, \boldsymbol{\phi} \mid \mathcal{D})$$

3) Bayesian Parameter Estimation:

$$p(\boldsymbol{\theta} \mid \mathcal{D}) = \text{Dirichlet}(\boldsymbol{\theta} \mid \boldsymbol{\alpha}')$$
$$p(\boldsymbol{\phi}_k \mid \mathcal{D}) = \text{Dirichlet}(\boldsymbol{\phi}_k \mid \boldsymbol{\beta}'_k)$$

Question:

Given a new point **x** and point estimates of θ and ϕ how do we do inference over z?

Question:

Given a new point **x** and distributions p(θ | D) and p(ϕ | D) how do we do inference over z?

Answer:

Answer:

Plate Diagrams

***Whiteboard*:**

- Example: Dirichet-Multinomial as a directed graphical model
- Example: Plate diagram for Dirichlet-Multinomial model

TOPIC MODELING

Topic Modeling

Motivation:

Suppose you're given a massive corpora and asked to carry out the following tasks

- **Organize** the documents into **thematic categories**
- **Describe** the evolution of those categories **over time**
- Enable a domain expert to **analyze and understand** the content
- Find **relationships** between the categories
- Understand how **authorship** influences the content

Topic Modeling

Motivation:

Suppose you're given a massive corpora and asked to carry out the following tasks

- **Organize** the documents into **thematic categories**
- **Describe** the evolution of those categories **over time**
- Enable a domain expert to **analyze and understand** the content
- Find **relationships** between the categories
- Understand how **authorship** influences the content

Topic Modeling:

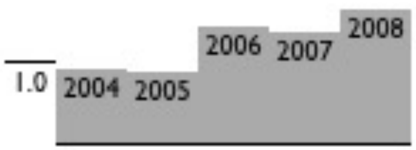
A method of (usually unsupervised) discovery of latent or hidden structure in a corpus

- Applied primarily to text corpora, but **techniques are more general**
- Provides a **modeling toolbox**
- Has prompted the exploration of a variety of new **inference methods** to accommodate **large-scale datasets**

Topic Modeling

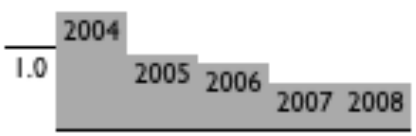
Dirichlet-multinomial regression (DMR) topic model on ICML
(Mimno & McCallum, 2008)

Topic 0 [0.152]



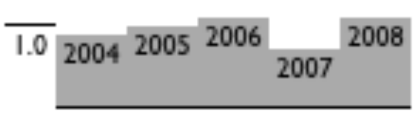
problem, optimization, problems, convex, convex optimization, linear, semidefinite programming, formulation, sets, constraints, proposed, margin, maximum margin, optimization problem, linear programming, programming, procedure, method, cutting plane, solutions

Topic 54 [0.051]



decision trees, trees, tree, decision tree, decision, tree ensemble, junction tree, decision tree learners, leaf nodes, arithmetic circuits, ensembles modts, skewing, ensembles, anytime induction decision trees, trees trees, random forests, objective decision trees, tree learners, trees grove, candidate split

Topic 99 [0.066]



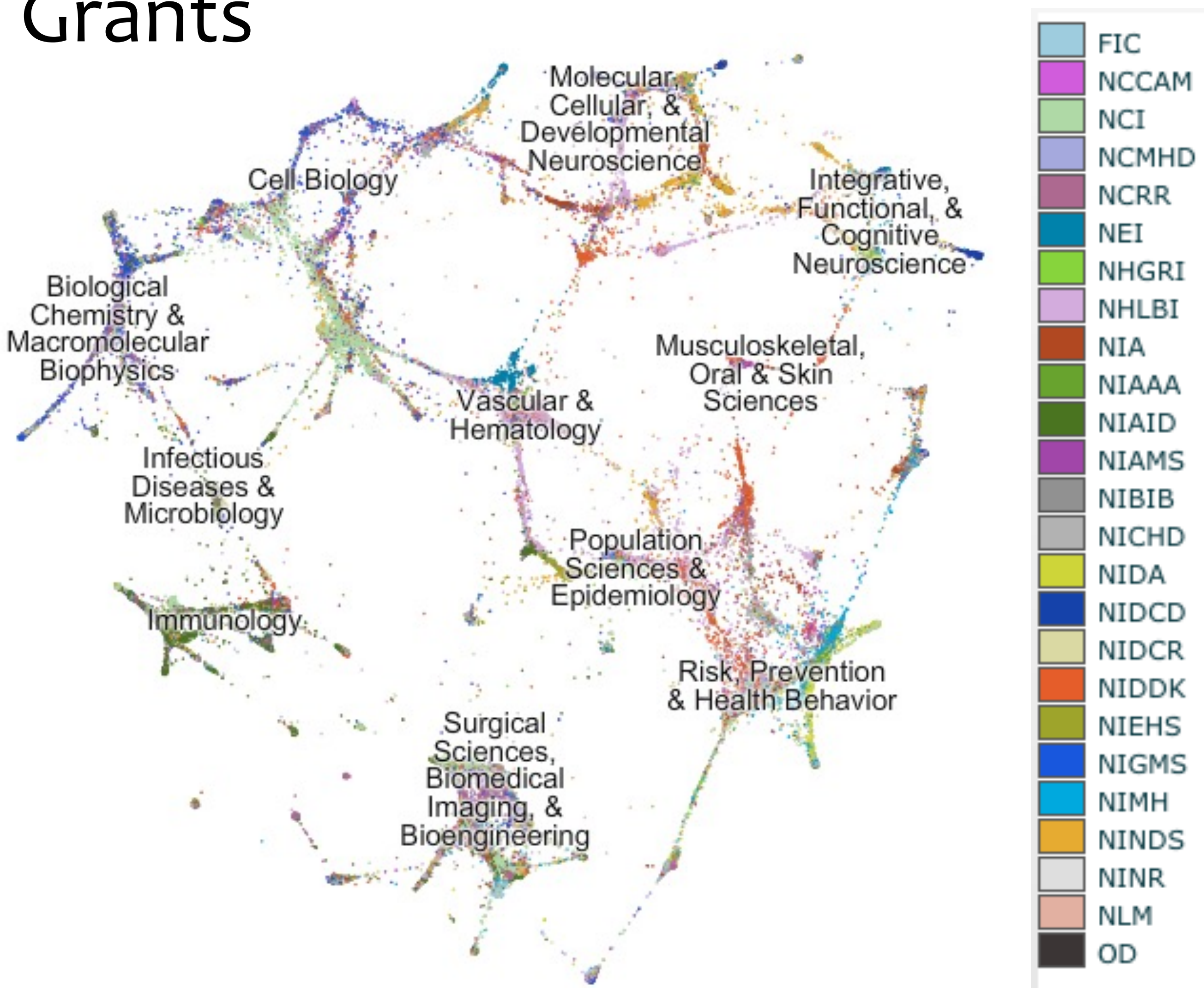
inference, approximate inference, exact inference, markov chain, models, approximate, gibbs sampling, variational, bayesian, variational inference, variational bayesian, approximation, sampling, methods, exact, bayesian inference, dynamic bayesian, process, mcmc, efficient

http:// www.cs.umass.edu/~mimno/icml100.html

Topic Modeling

- Map of NIH Grants

(Talley et al., 2011)

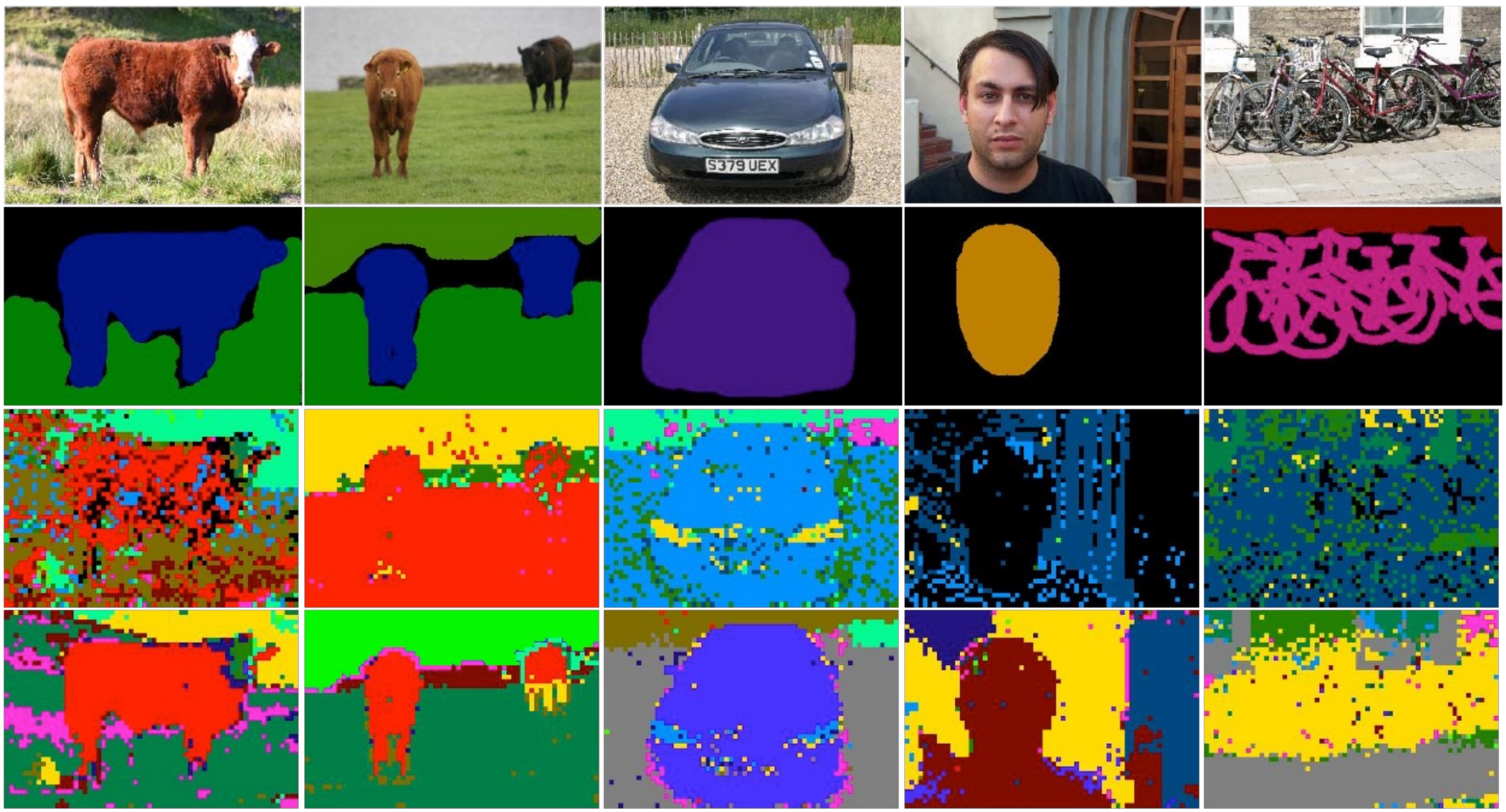


https://app.nihmaps.org/

Other Applications of Topic Models

- Spacial LDA

(Wang & Grimson, 2007)



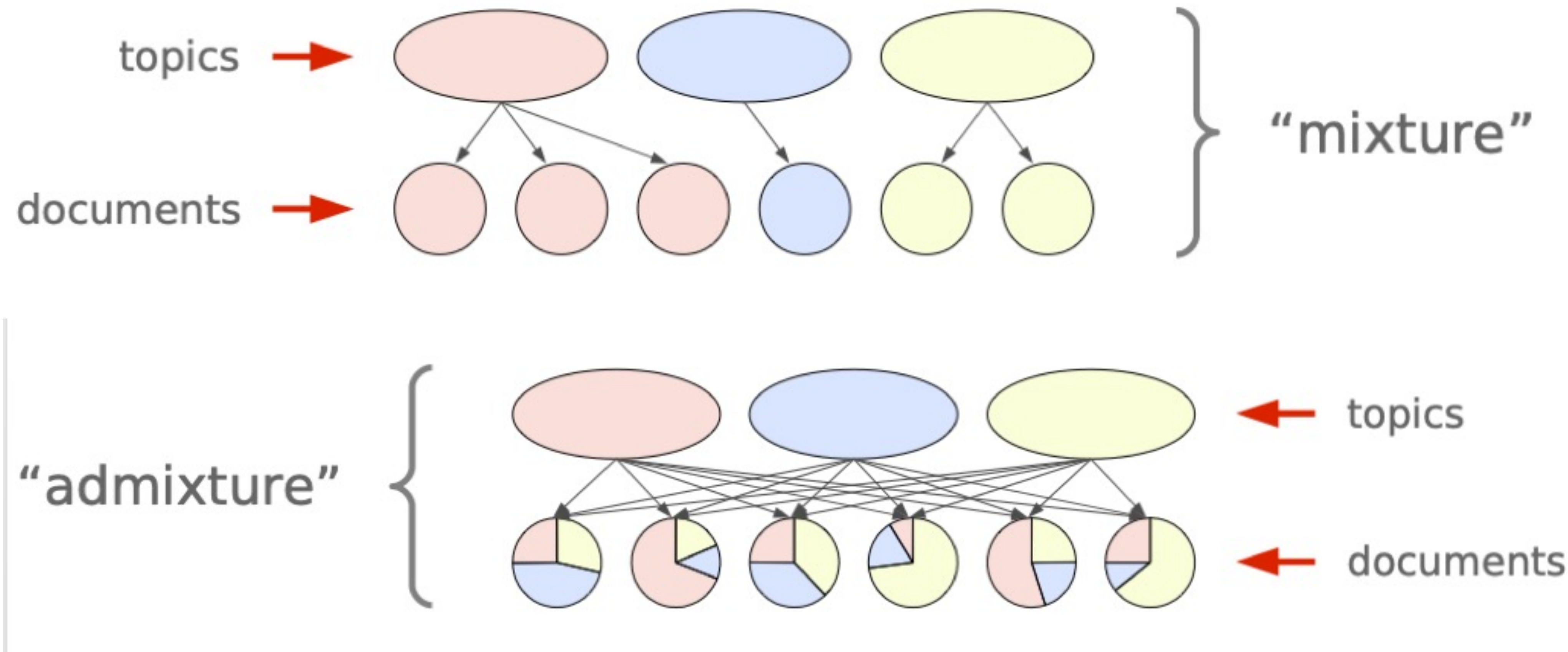
Manual

LDA

SLDA

LATENT DIRICHLET ALLOCATION (LDA)

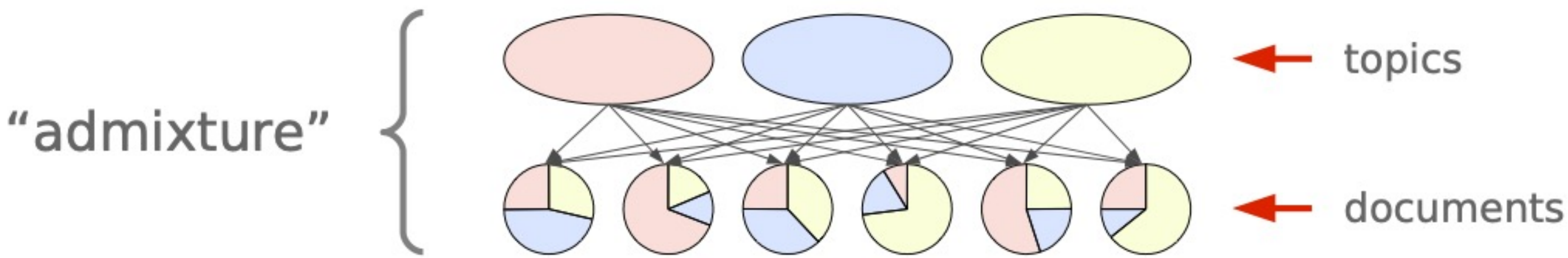
Mixture vs. Admixture (LDA)



Figures from Wallach, JHU 2011, slides

Latent Dirichlet Allocation

- Generative Process



- Example corpus

the	he	is
x_{11}	x_{12}	x_{13}

Document 1

the	and	the
x_{21}	x_{22}	x_{23}

Document 2

she	she	is	is
x_{31}	x_{32}	x_{33}	x_{34}

Document 3

Figure from Wallach, JHU 2011, slides

Latent Dirichlet Allocation

- Generative Process

For each topic $k \in \{1, \ldots, K\}$:
$\quad \boldsymbol{\phi}_k \sim \text{Dir}(\boldsymbol{\beta})$ [*draw distribution over words*]
For each document $m \in \{1, \ldots, M\}$
$\quad \boldsymbol{\theta}_m \sim \text{Dir}(\boldsymbol{\alpha})$ [*draw distribution over topics*]
$\quad$ For each word $n \in \{1, \ldots, N_m\}$
$\quad\quad z_{mn} \sim \text{Mult}(1, \boldsymbol{\theta}_m)$ [*draw topic assignment*]
$\quad\quad x_{mn} \sim \boldsymbol{\phi}_{z_{mi}}$ [*draw word*]

- Example corpus

the	he	is
x_{11}	x_{12}	x_{13}

Document 1

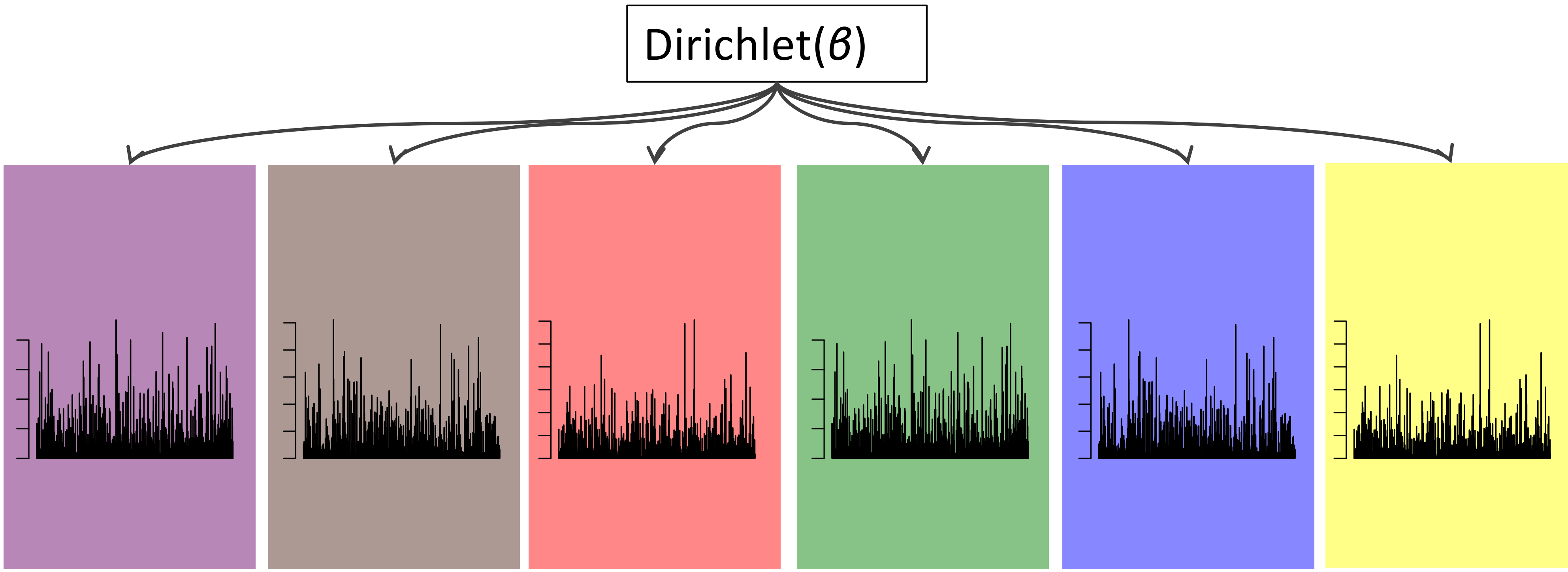
the	and	the
x_{21}	x_{22}	x_{23}

Document 2

she	she	is	is
x_{31}	x_{32}	x_{33}	x_{34}

Document 3

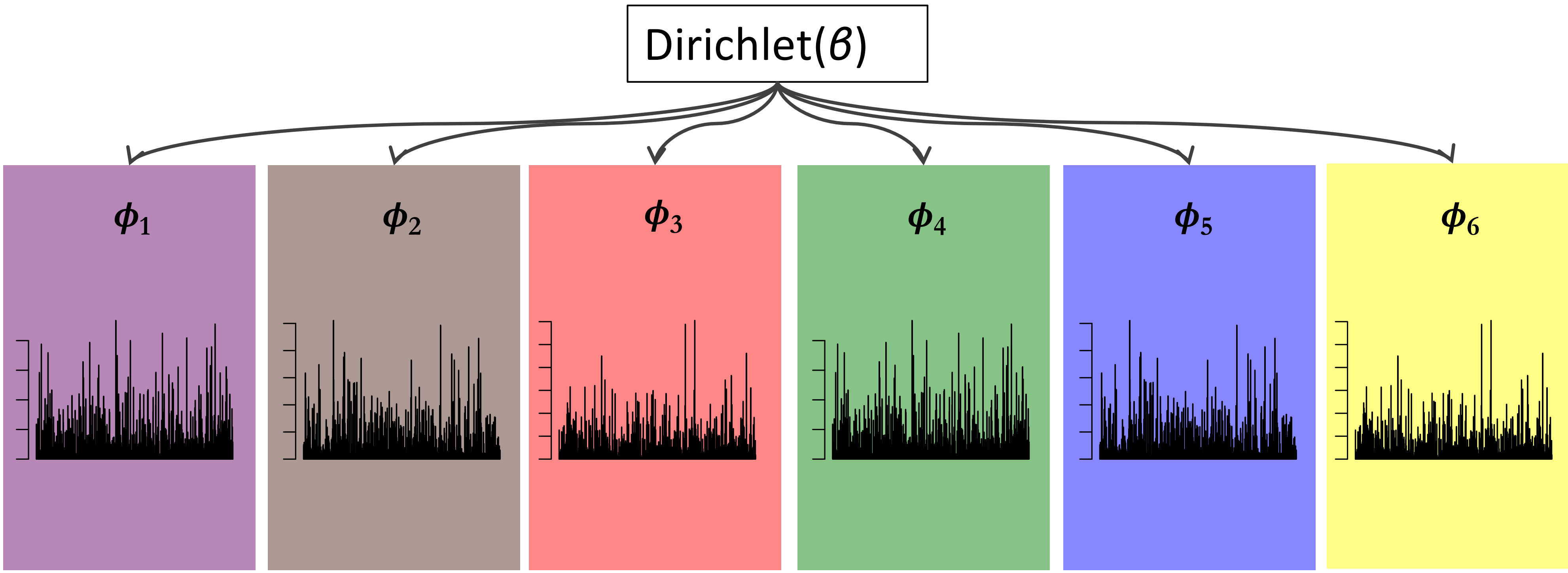
LDA for Topic Modeling



- The **generative story** begins with only a **Dirichlet prior** over the topics.
- Each **topic** is defined as a **Multinomial distribution** over the vocabulary, parameterized by $\boldsymbol{\phi}_k$

(Blei, Ng, & Jordan, 2003)

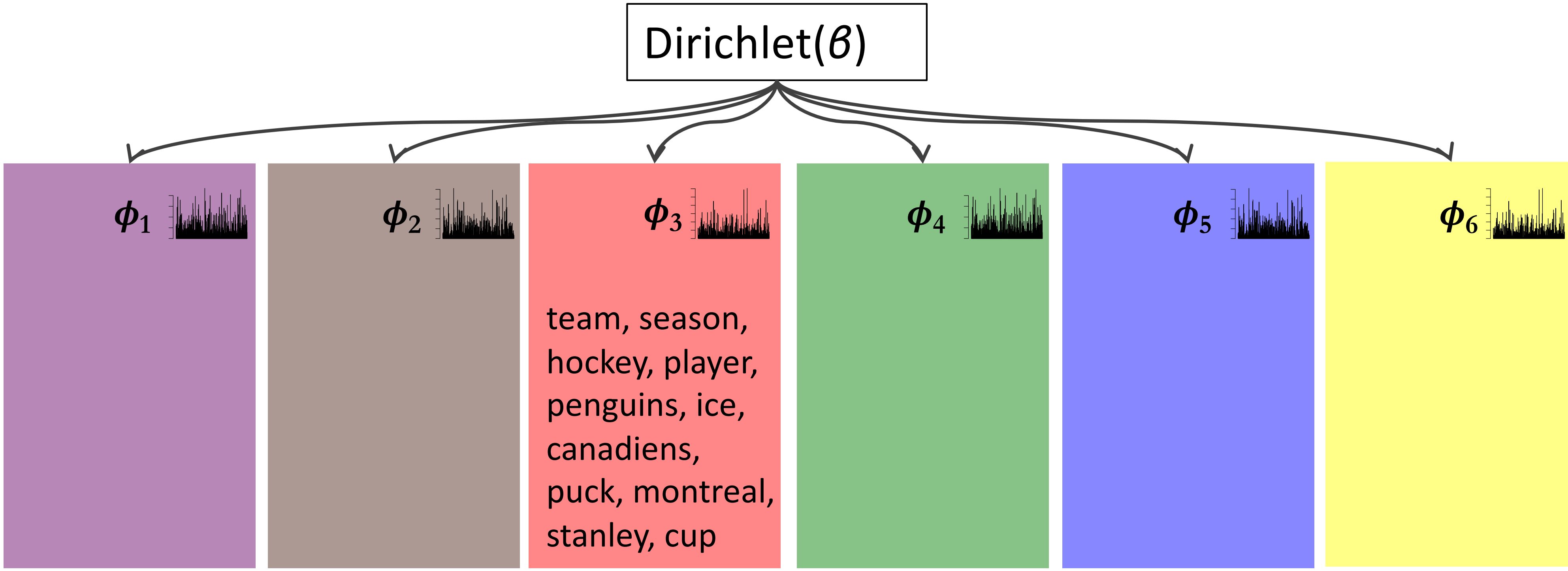
LDA for Topic Modeling



- The **generative story** begins with only a **Dirichlet prior** over the topics.
- Each **topic** is defined as a **Multinomial distribution** over the vocabulary, parameterized by $\boldsymbol{\phi}_k$

(Blei, Ng, & Jordan, 2003)

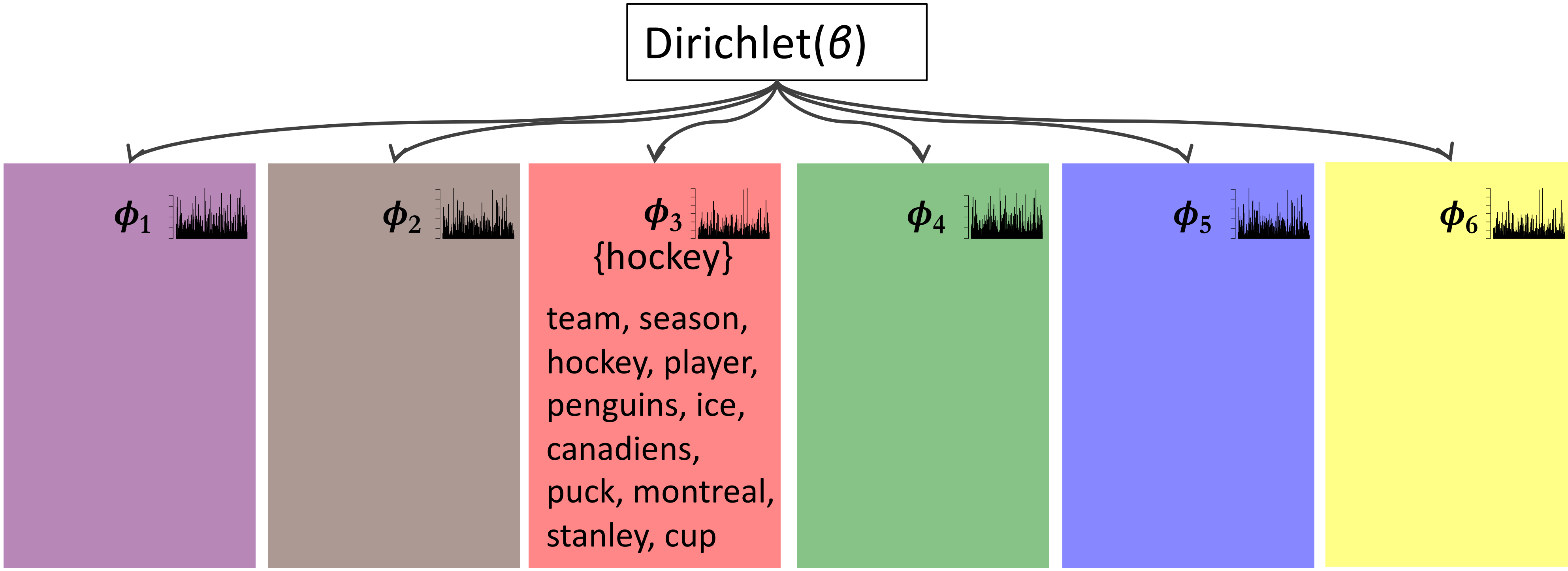
LDA for Topic Modeling



- A topic is visualized as its **high probability words.**

(Blei, Ng, & Jordan, 2003)

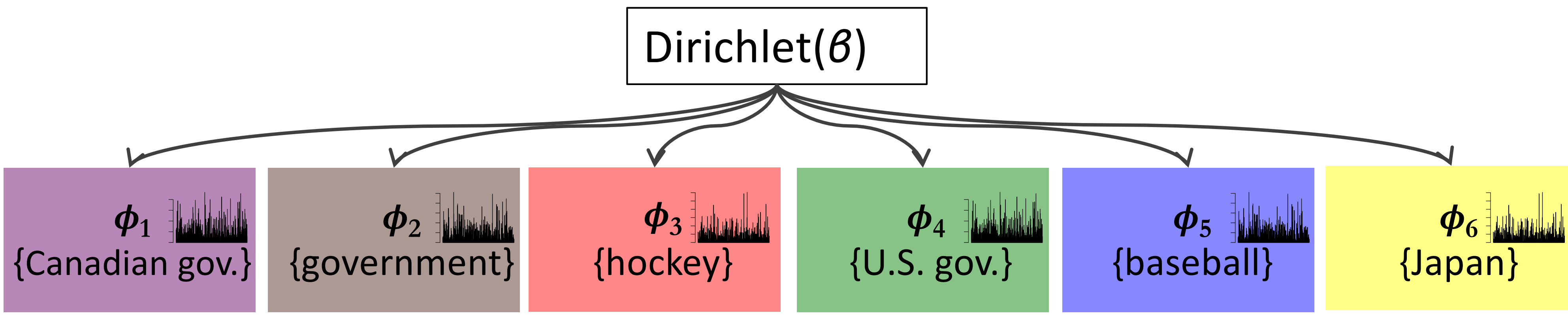
LDA for Topic Modeling



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

(Blei, Ng, & Jordan, 2003)

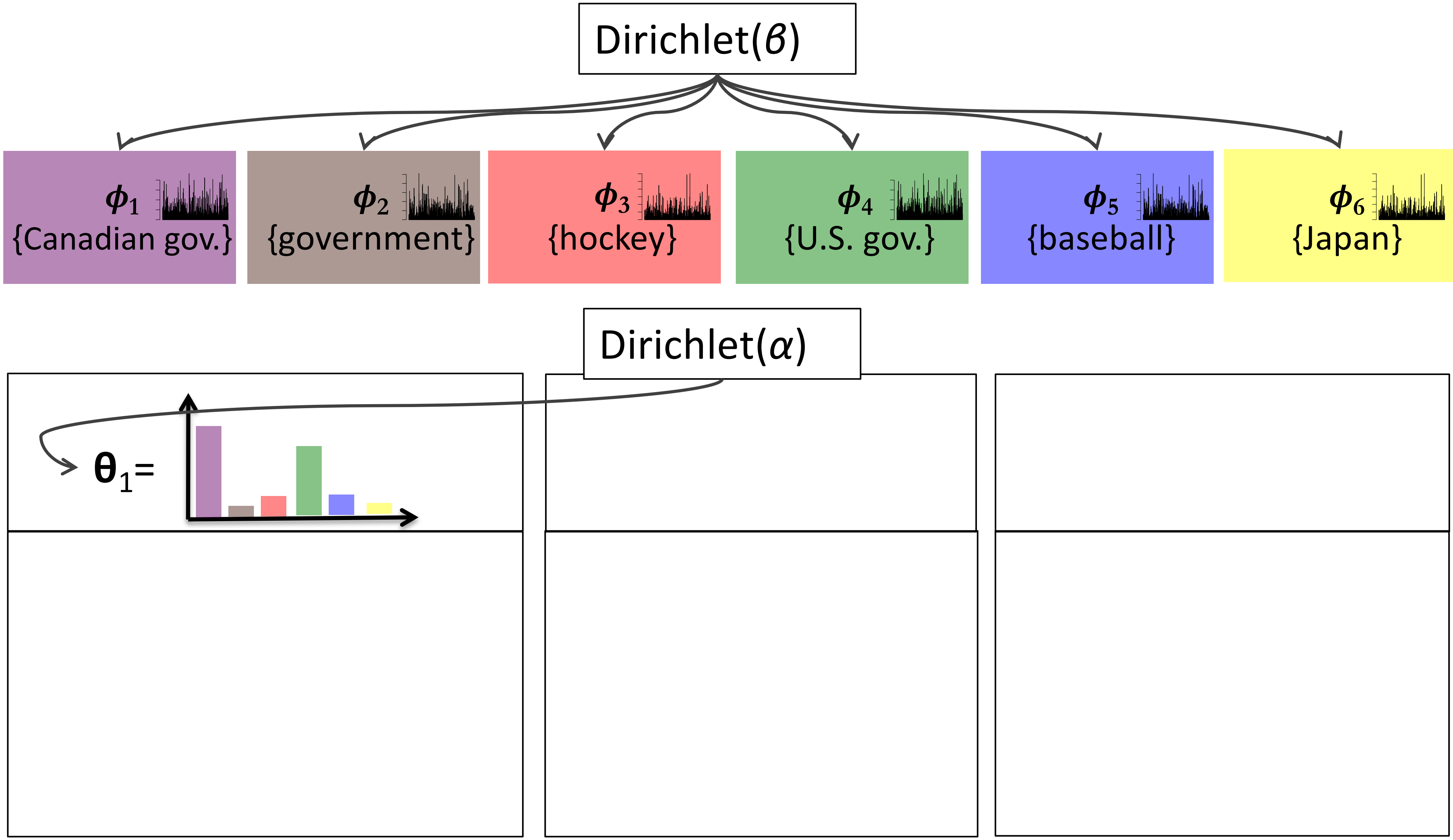
LDA for Topic Modeling



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

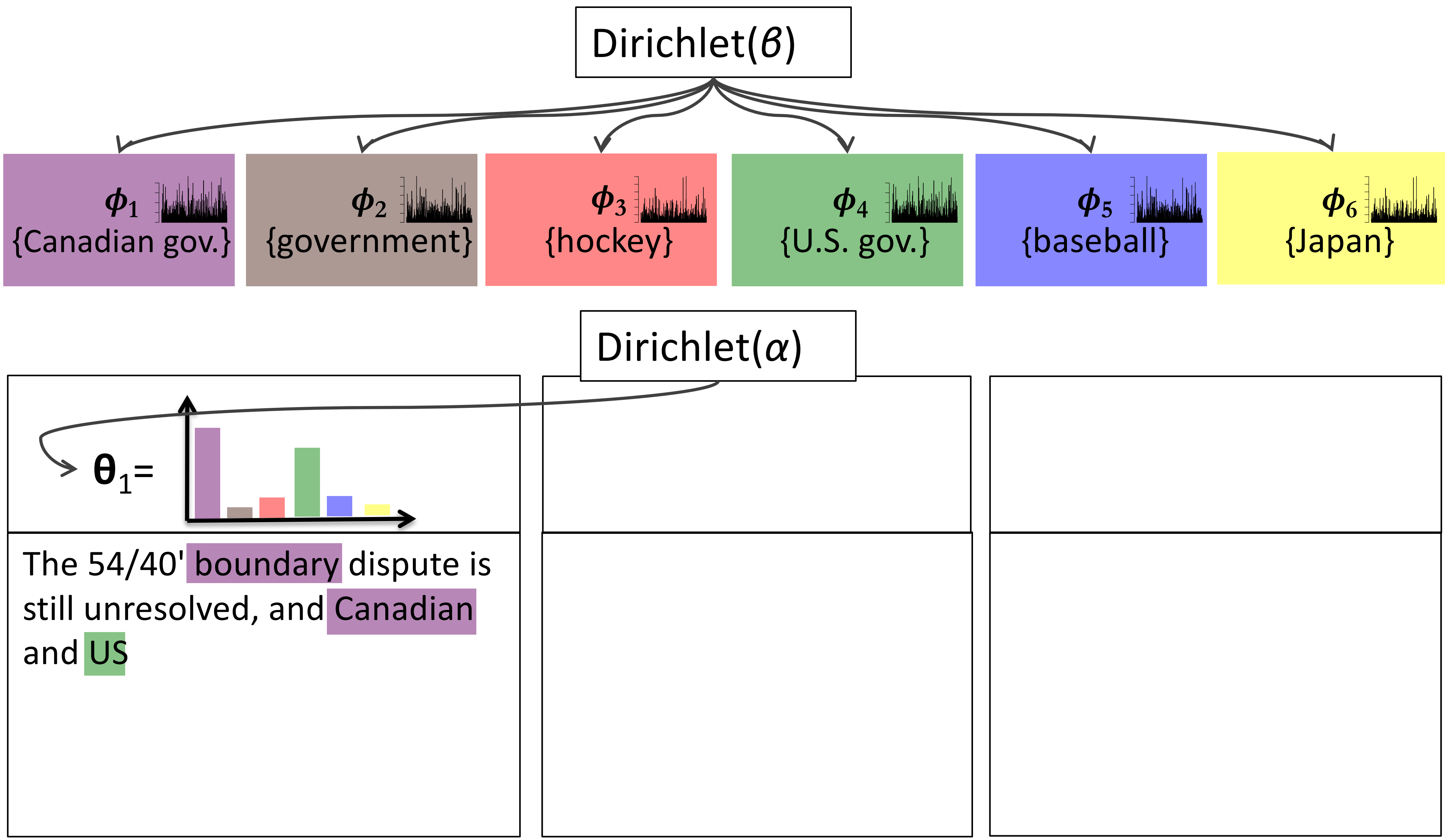
(Blei, Ng, & Jordan, 2003)

LDA for Topic Modeling



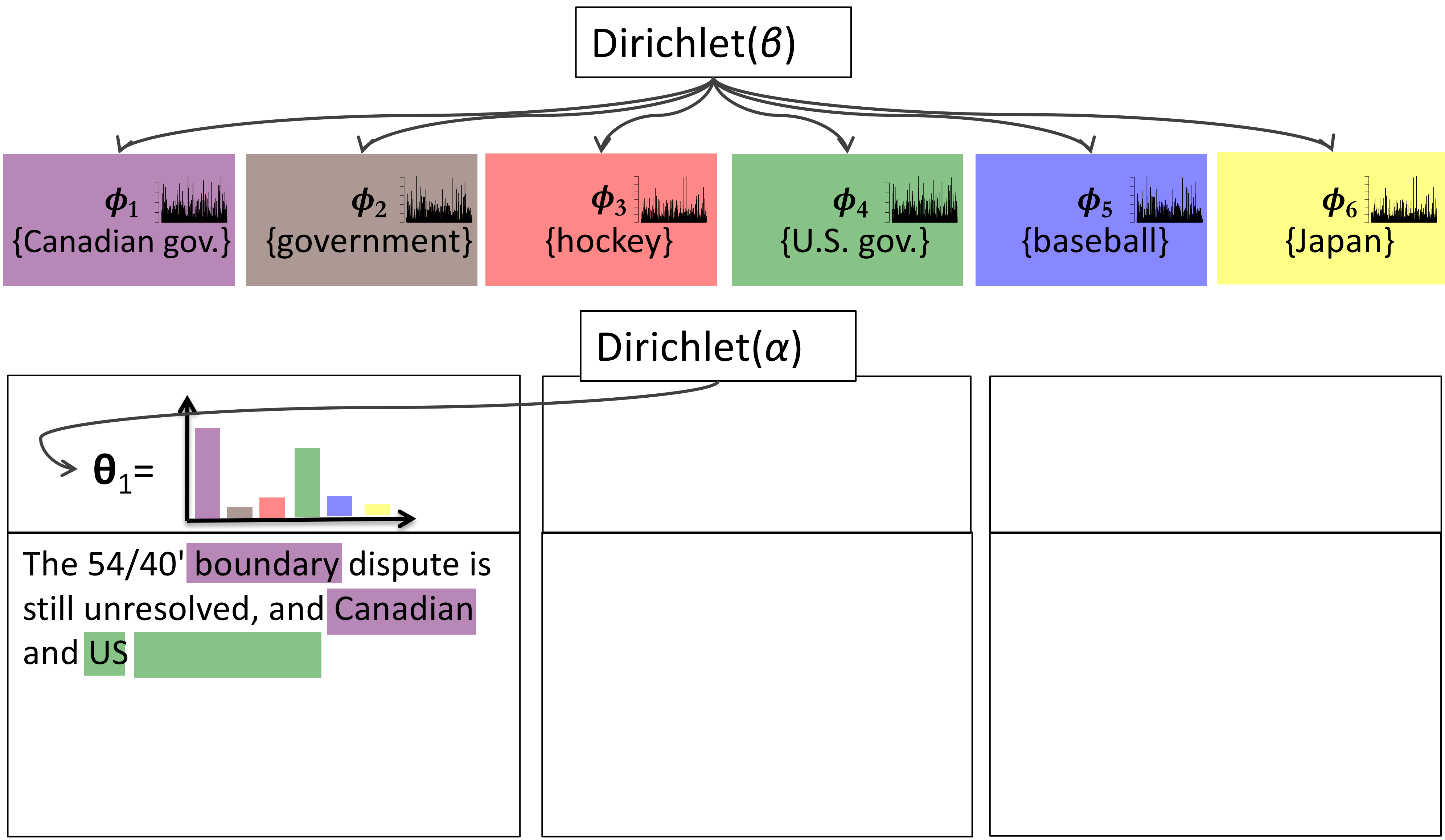
(Blei, Ng, & Jordan, 2003)

LDA for Topic Modeling



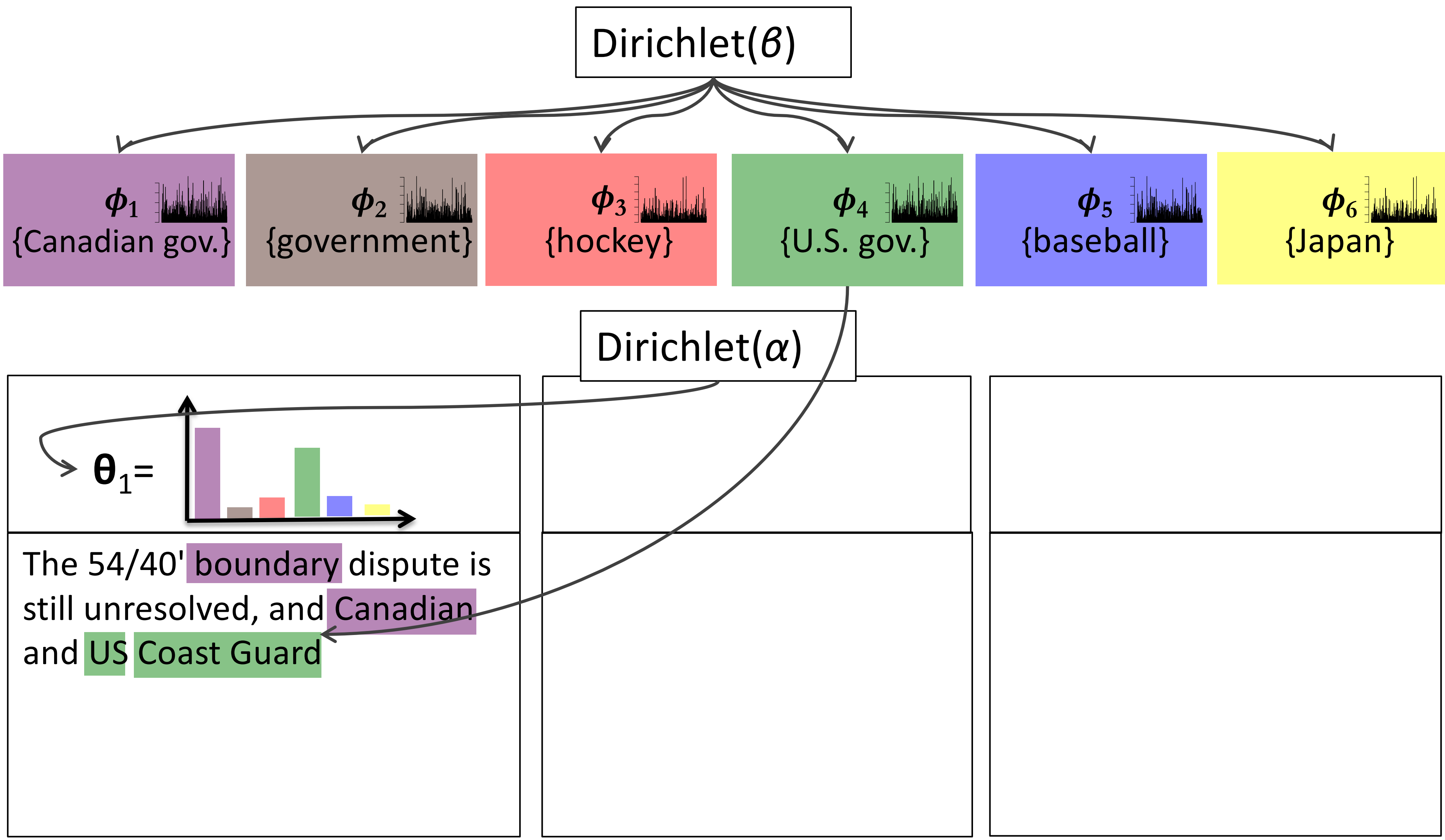
(Blei, Ng, & Jordan, 2003)

LDA for Topic Modeling



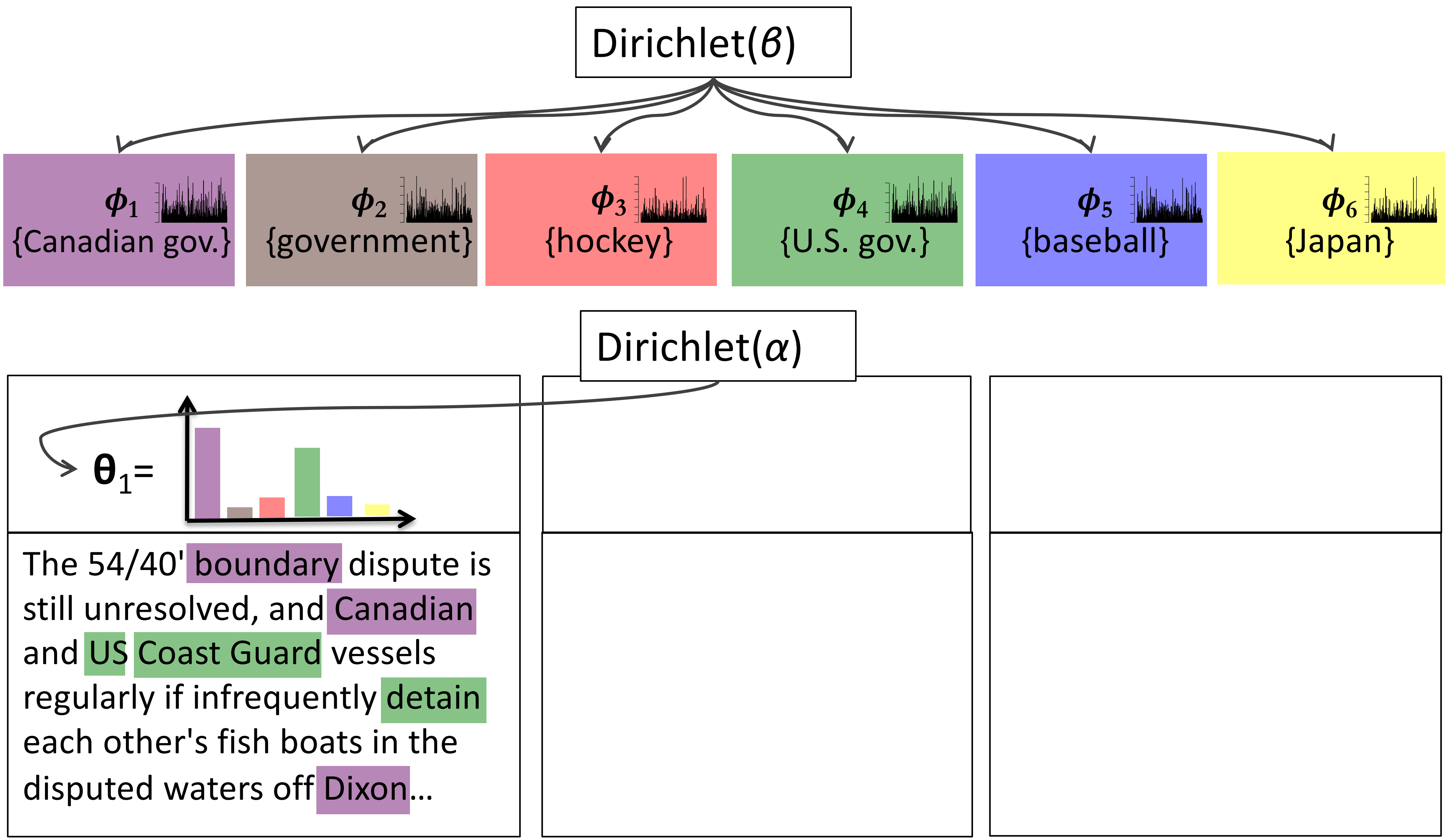
(Blei, Ng, & Jordan, 2003)

LDA for Topic Modeling



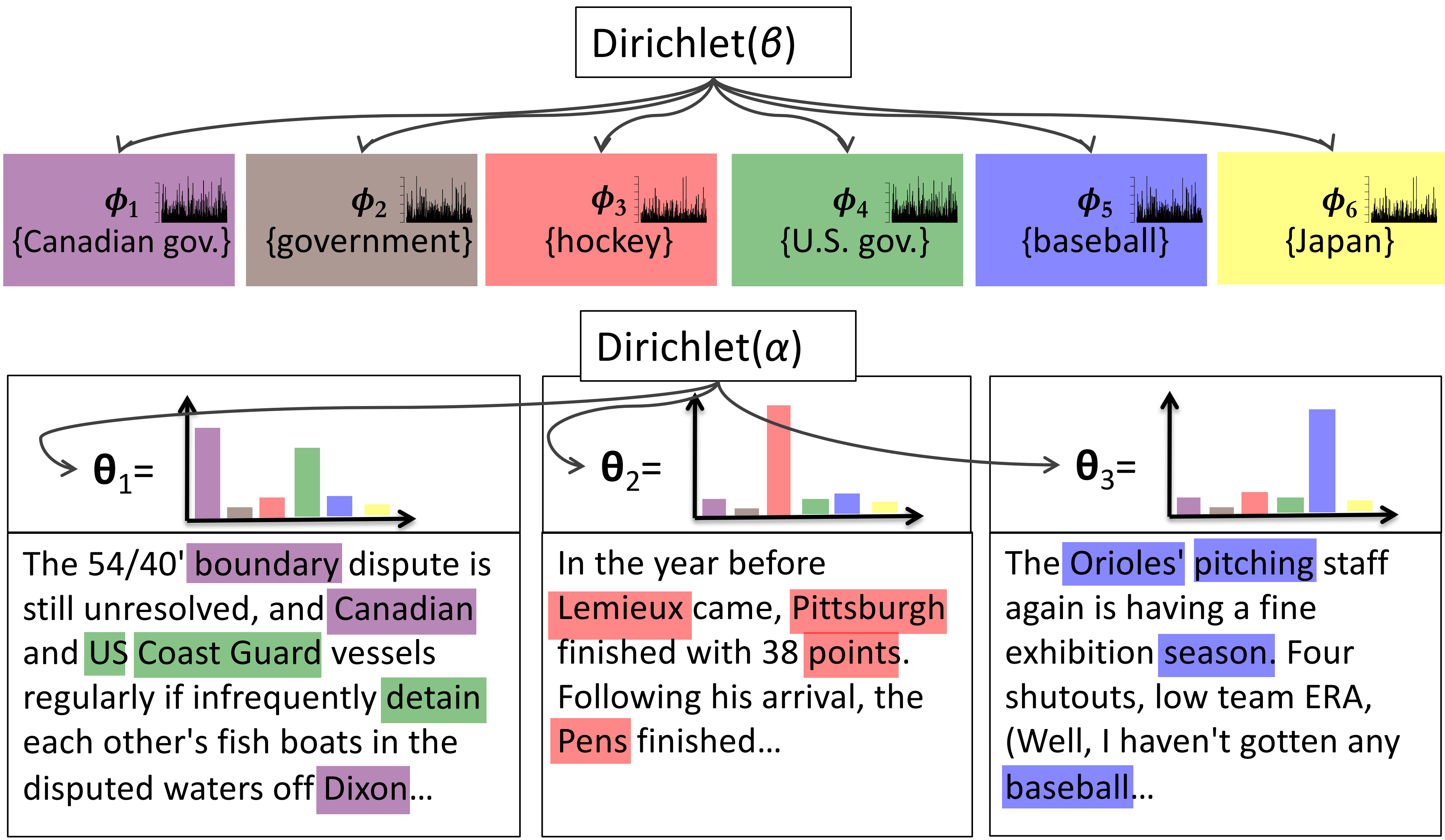
(Blei, Ng, & Jordan, 2003)

LDA for Topic Modeling



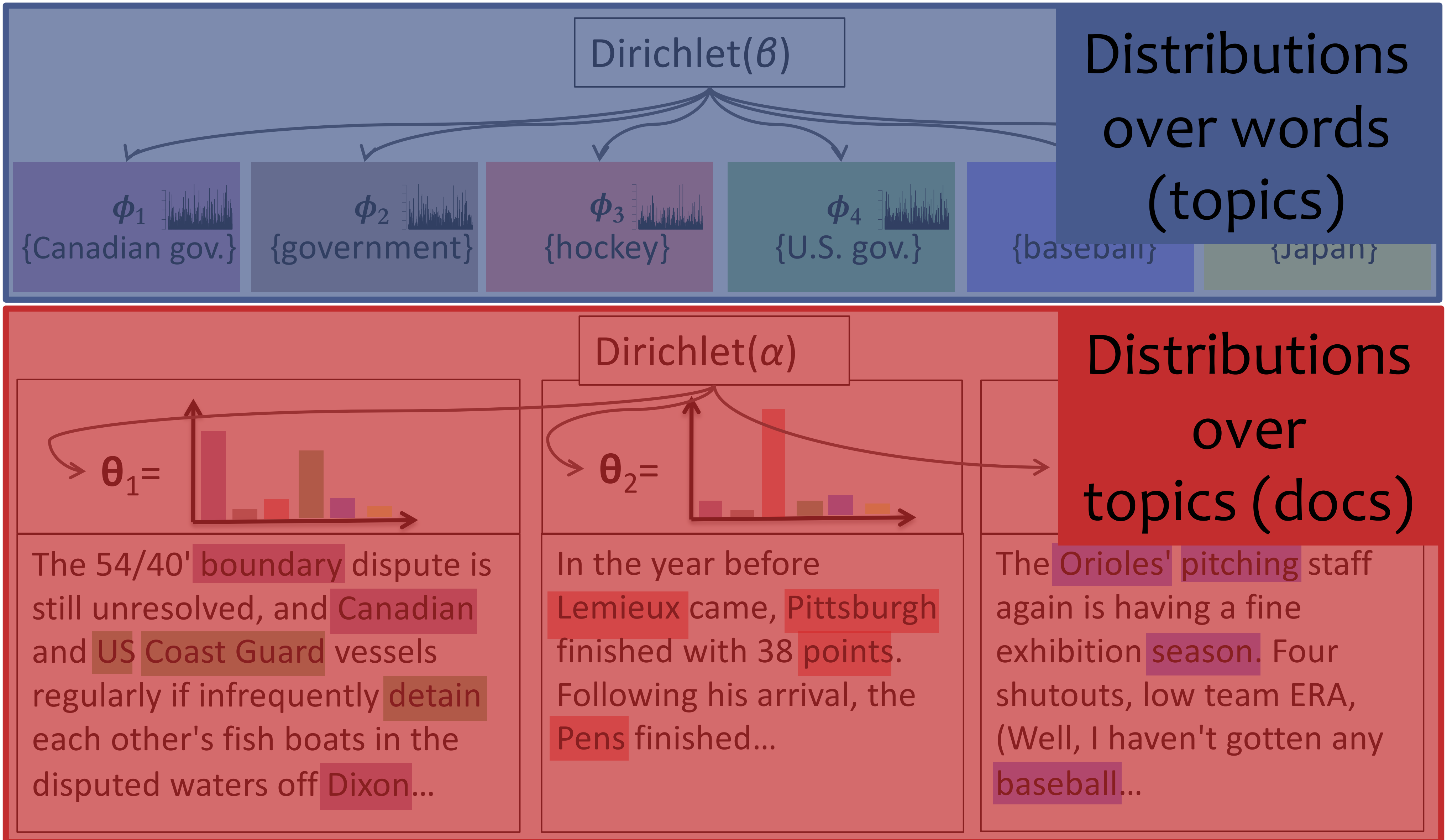
(Blei, Ng, & Jordan, 2003)

LDA for Topic Modeling



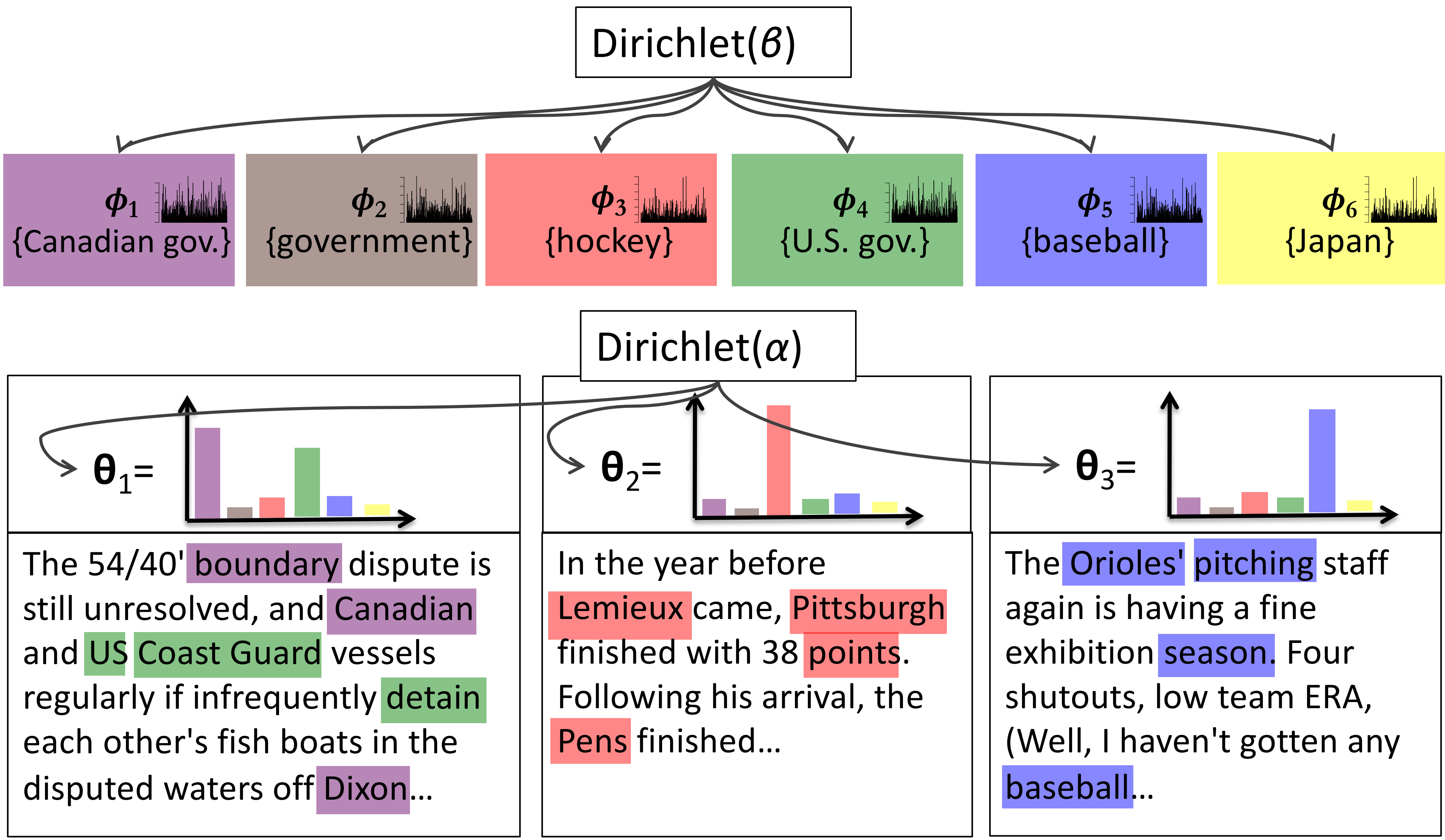
(Blei, Ng, & Jordan, 2003)

LDA for Topic Modeling



(Blei, Ng, & Jordan, 2003)

LDA for Topic Modeling



(Blei, Ng, & Jordan, 2003)

LDA for Topic Modeling

Inference and learning start with only the data

Dirichlet( )

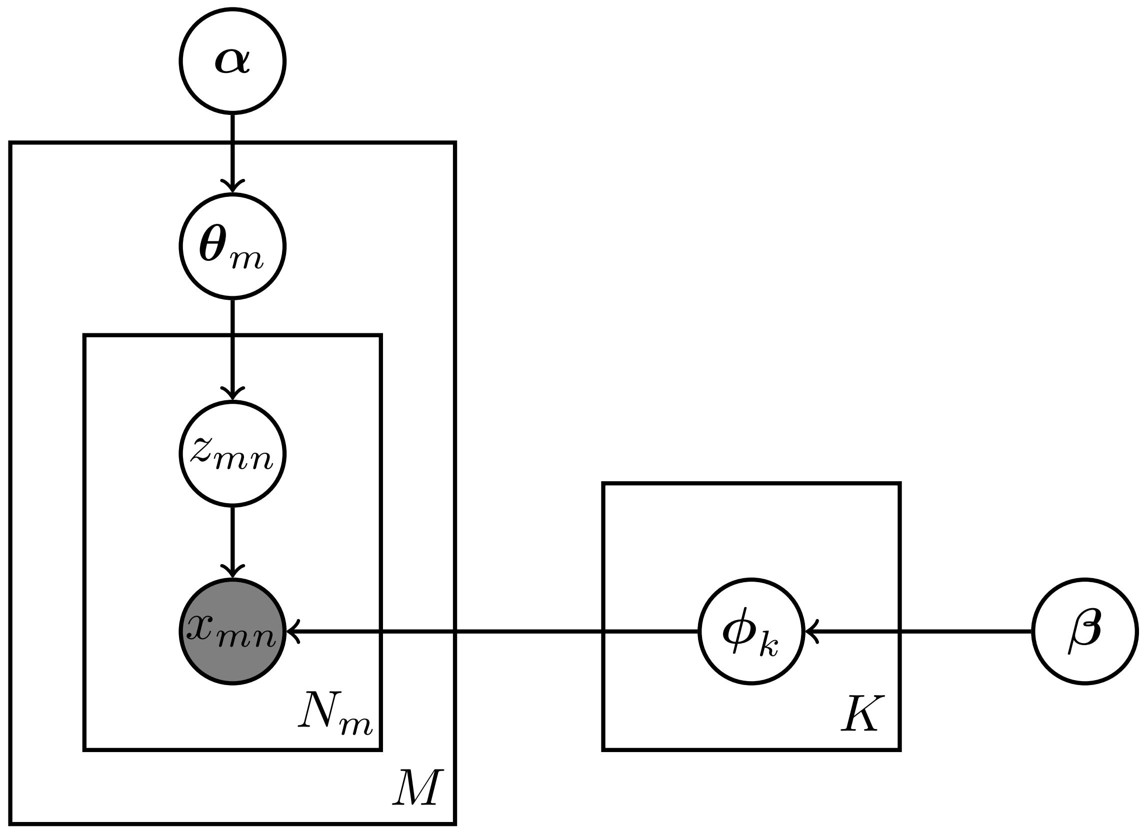
ϕ_1 =	ϕ_2 =	ϕ_3 =	ϕ_4 =	ϕ_5 =	ϕ_6 =

Dirichlet( )

θ_1=	θ_2=	θ_3=
The 54/40' boundary dispute is still unresolved, and Canadian and US Coast Guard vessels regularly if infrequently detain each other's fish boats in the disputed waters off Dixon…	In the year before Lemieux came, Pittsburgh finished with 38 points. Following his arrival, the Pens finished…	The Orioles' itching staff again is having a fine exhibition season. Four shutouts, low team ERA, (Well, I haven't gotten any baseball…

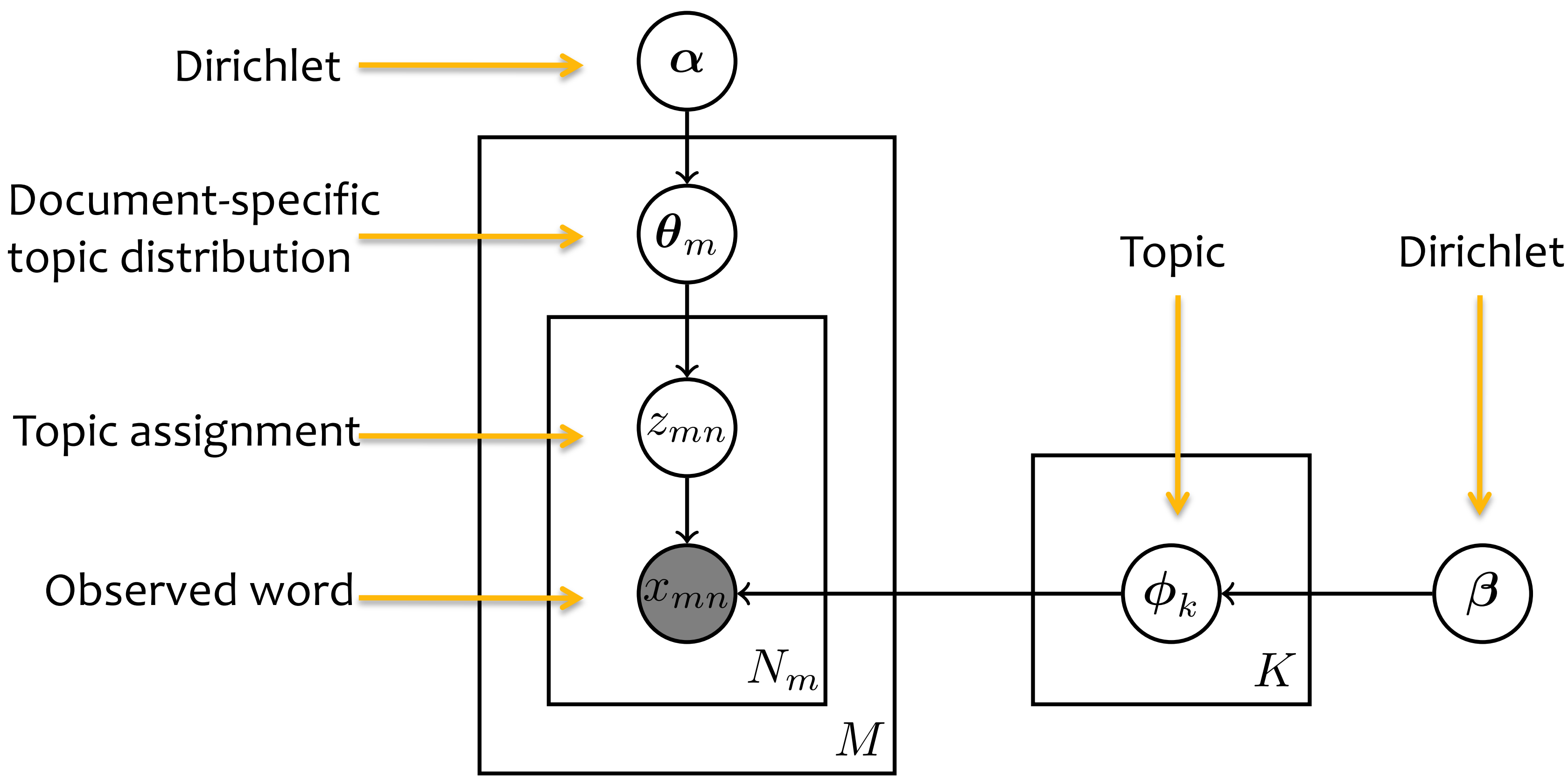
Latent Dirichlet Allocation

- Plate Diagram



Latent Dirichlet Allocation

- Plate Diagram



Latent Dirichlet Allocation

Question:

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

Answer:

Question:

Why might it work well anyway?

Answer:

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 non-negative matrix factorization (Jakulin and Buntine, 2002)
- Was independently invented for genetics (Pritchard et al., 2000)

Slide from David Blei, MLSS 2012

Outline

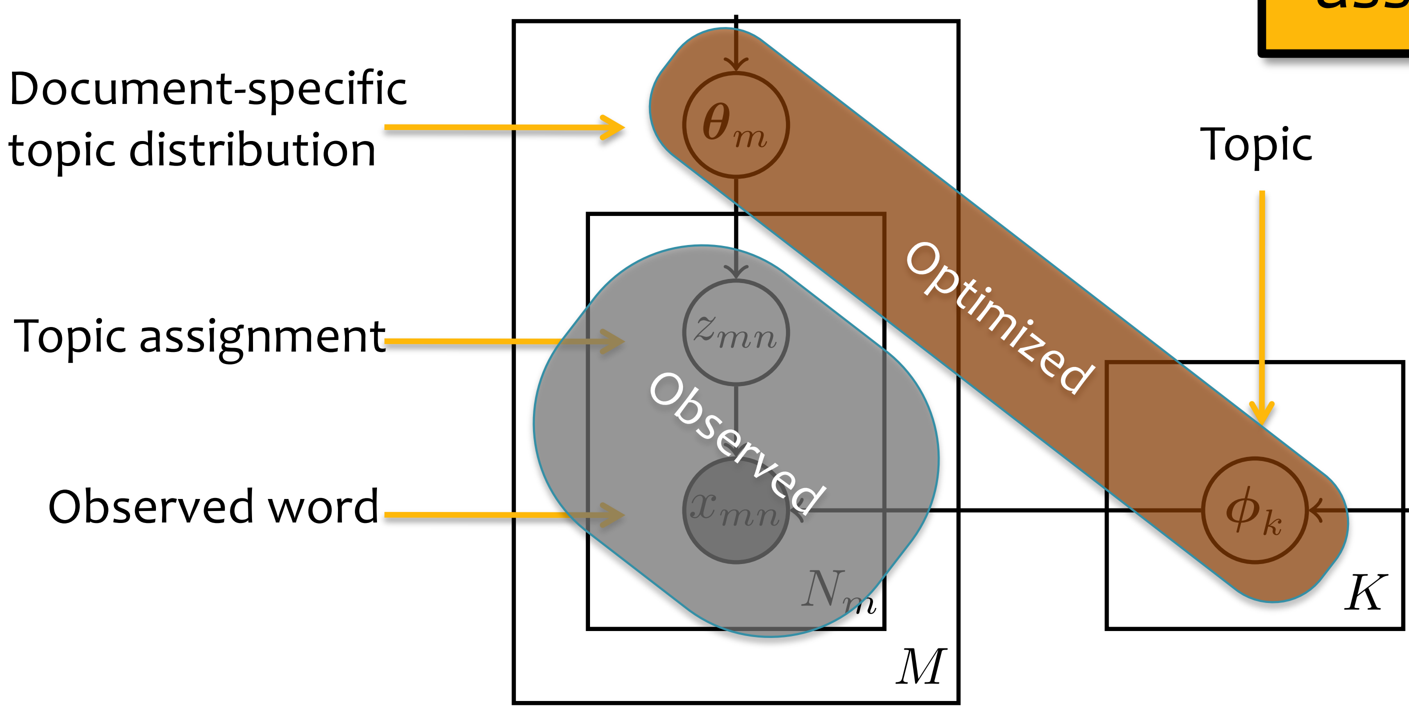
- **Applications of Topic Modeling**
- **Latent Dirichlet Allocation (LDA)**
  1. Beta-Bernoulli
  2. Dirichlet-Multinomial
  3. Dirichlet-Multinomial Mixture Model
  4. LDA
- **Bayesian Inference for Parameter Estimation**
  – Exact inference
  – EM
  – Monte Carlo EM
  – Gibbs sampler
  – Collapsed Gibbs sampler
- **Extensions of LDA**
  – Correlated topic models
  – Dynamic topic models
  – Polylingual topic models
  – Supervised LDA

BAYESIAN INFERENCE FOR PARAMETER ESTIMATION

LDA Inference

- Fully Observed MLE

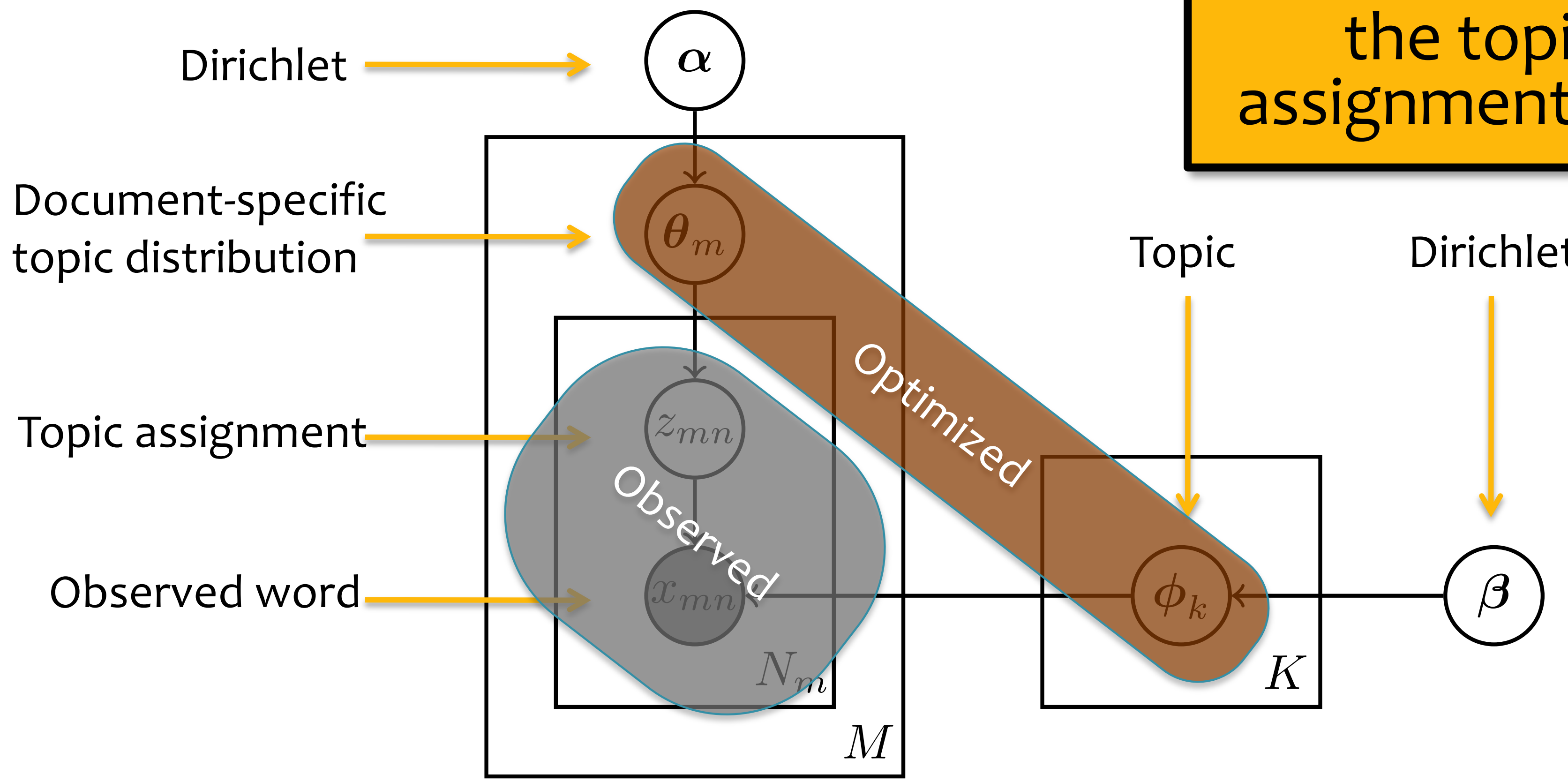
Learning like this would be easy, but in practice we do not observe the topic assignments z_{mn}



LDA Inference

- Full Observed MAP Estimation

Learning like this would be easy, but in practice we do not observe the topic assignments z_{mn}



Unsupervised Learning

Three learning paradigms:

1. Maximum likelihood estimation (MLE)

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

2. Maximum a posteriori (MAP) estimation

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

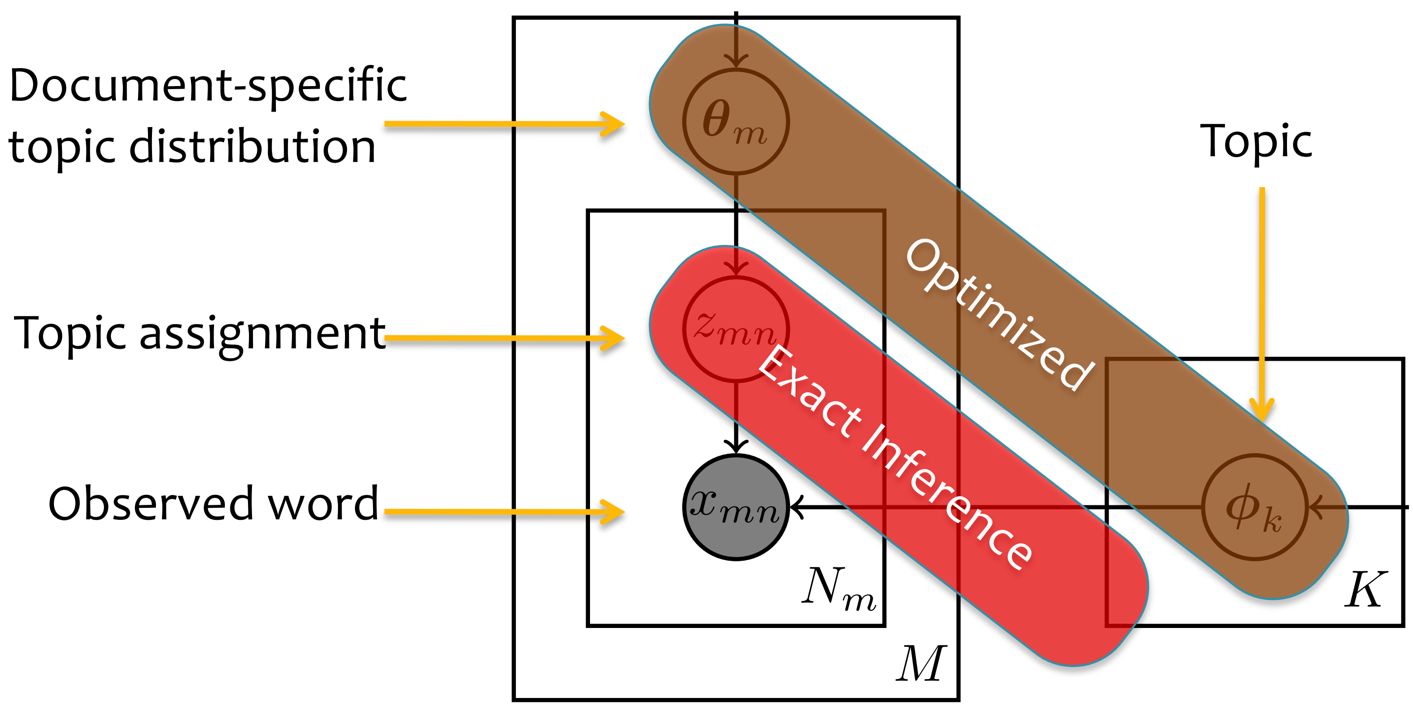
3. Bayesian approach

Estimate the posterior:

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

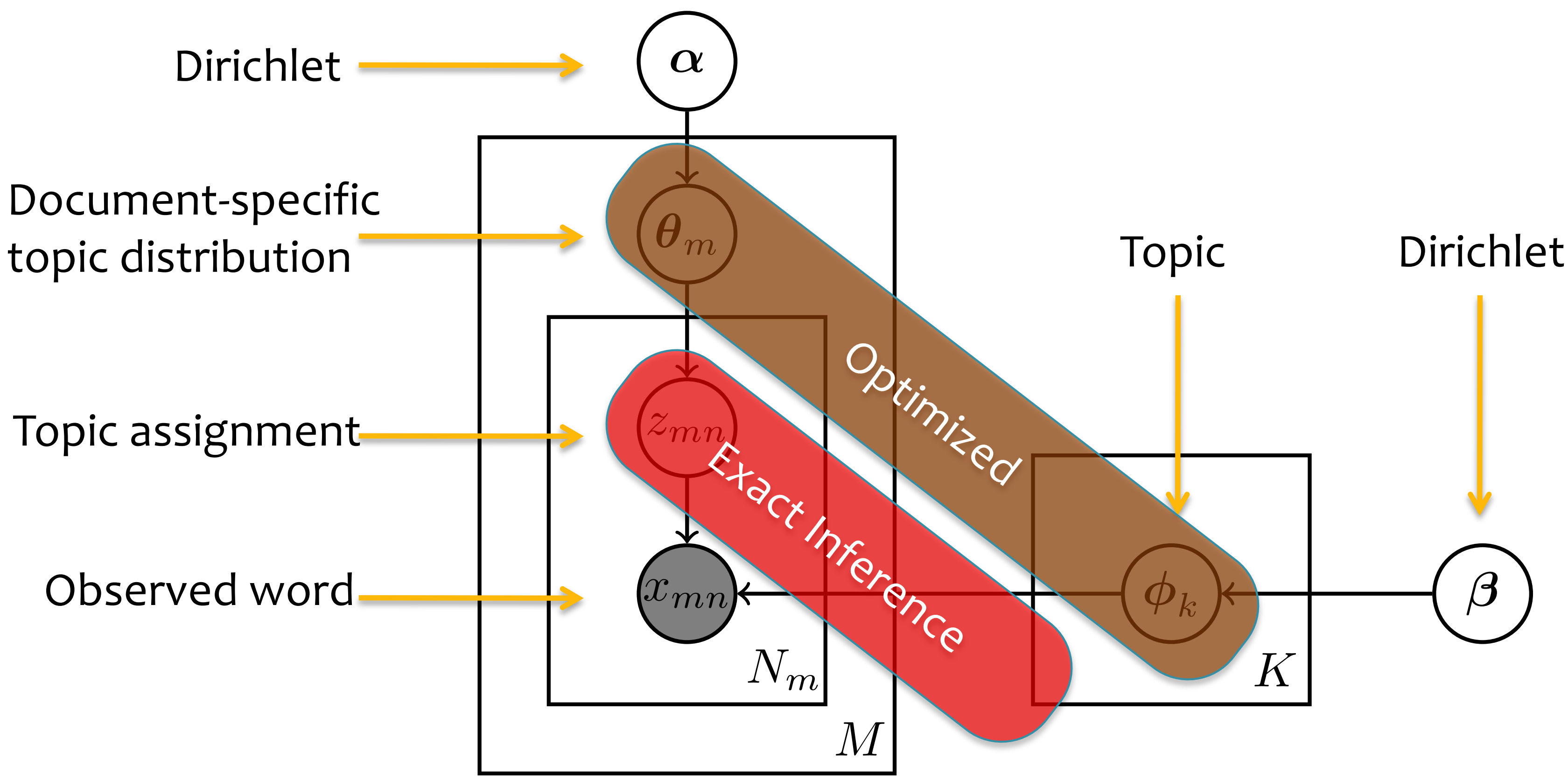
LDA Inference

- Standard EM (MLE)



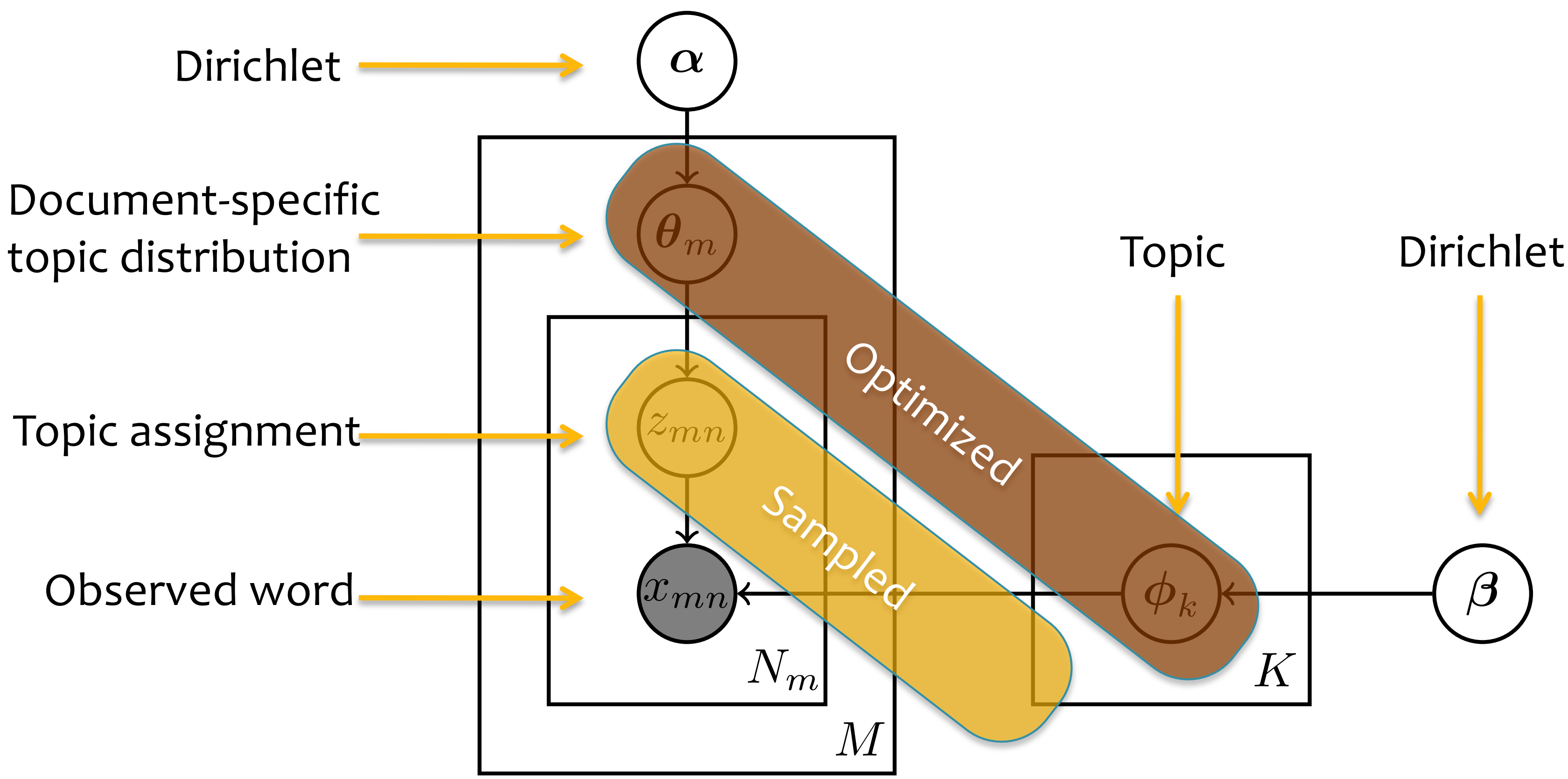
LDA Inference

- Standard EM (MAP Estimation)



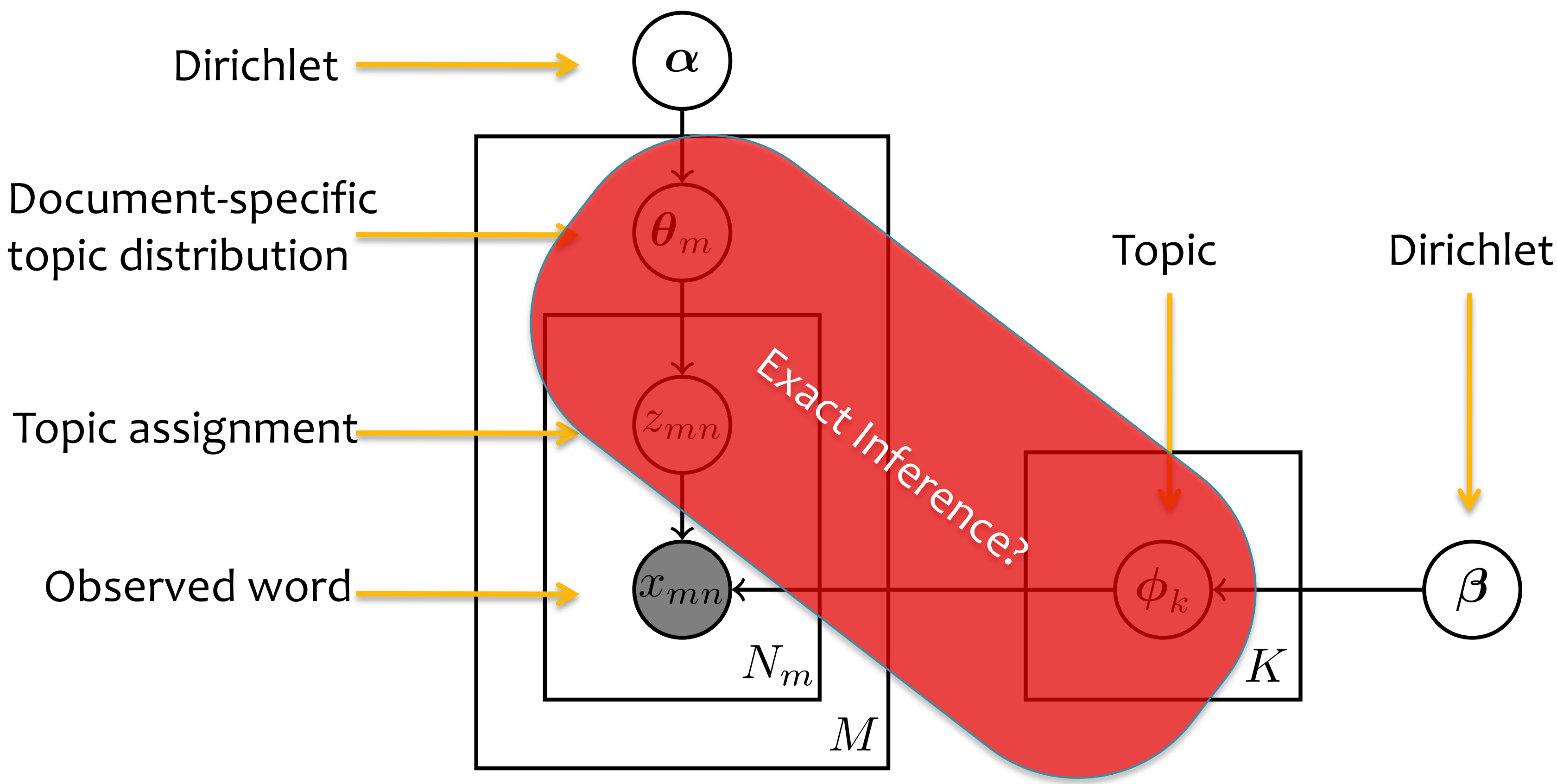
LDA Inference

- Monte Carlo EM (MAP Estimation)



LDA Inference

- Bayesian Approach



LDA Inference

- Bayesian Approach