

DSC250: Advanced Data Mining

Topic Models

Zhiting Hu

Lecture 6, Jan 24, 2025

UC San Diego

HALICIOĞLU DATA SCIENCE INSTITUTE

In-class paper presentation

Adversarial Examples are not Bugs, they are Features



Andrew Ilyas*



Dimitris Tsipras*



Shibani Santurkar*



Logan Engstrom*



Brandon Tran



Aleksander Mądry

Topic Models

Outline

- Representations of Text and Topics
- Topic Model v1: Multinomial Mixture Model
- Topic Model v2: Probabilistic Latent Semantic Analysis (pLSA)
- Topic Model v3: Latent Dirichlet Allocation (LDA)

Slides adapted from:

- Y. Sun, CS 247: Advanced Data Mining
- M. Gormley, 10-701 Introduction to Machine Learning

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



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: Examples

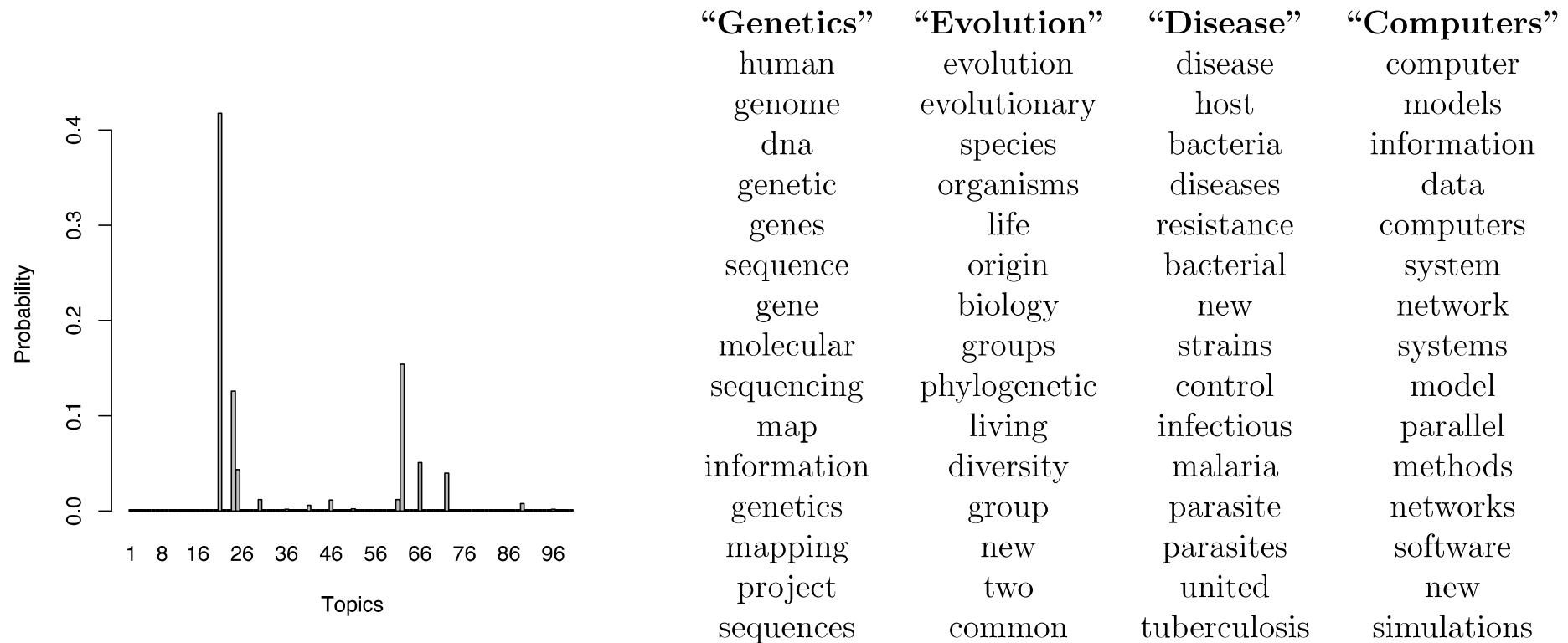


Figure from (Blei, 2011), shows topics and top words learned automatically from reading 17,000 Science articles

Topic Modeling: Examples

Dirichlet multinomial regression (DMR) topic model and CML (Mimno & McCallum, 2008)

Topic 0 [0.152]



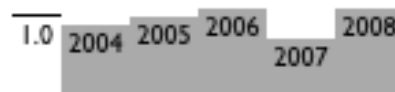
problem, optimization, problems, convex, convex optimization, linear, semidefinite programming, formulation, sets, constraints, proposed, margin, maximum margin, optimization problem, linear programming, programming, procedure, method, cutting plane, solutions

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]



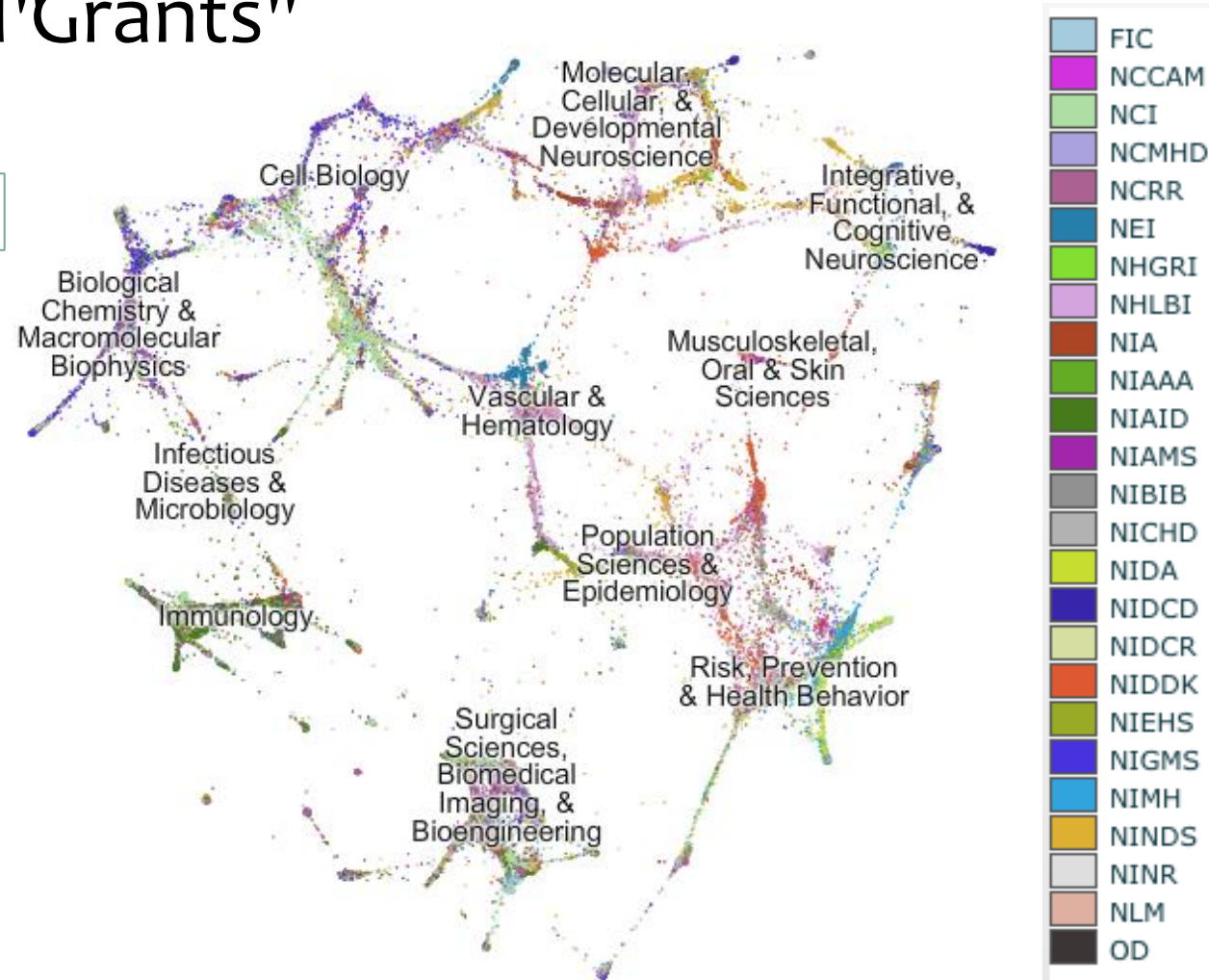
inference, approximate inference, exact inference, markov chain, models, approximate, gibbs sampling, variational, bayesian, variational inference, variational bayesian, approximation, sampling, methods, exact, bayesian inference, dynamic bayesian, process, mcmc, efficient

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

Topic Modeling: Examples

- Map'bf'NIH'Grants''

(Talley'et'al.,'2011)''



<https://app.nihmaps.org/>

Other Applications of Topic Models

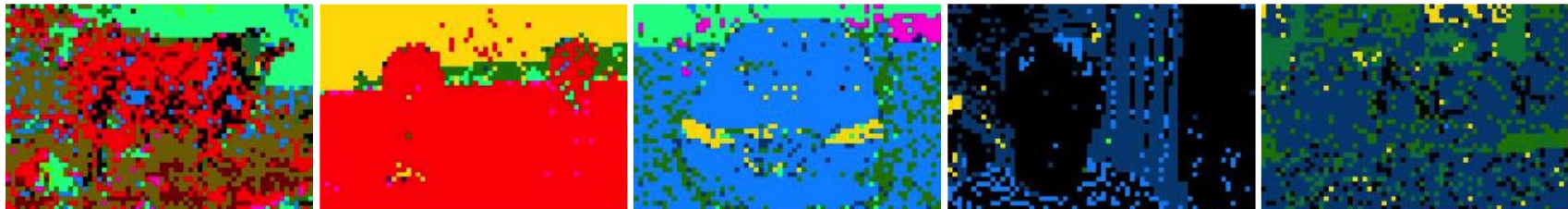
- Spatial LDA

(Wang & Grimson, 2007)

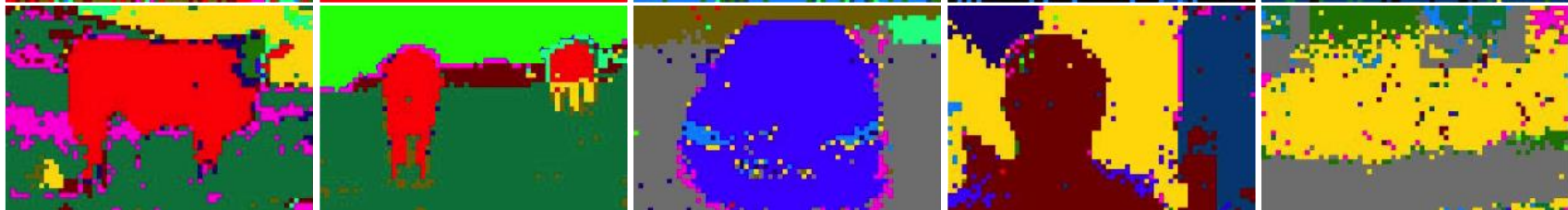
Manual



LDA



SLDA



Other Applications of Topic Models

- Word Sense Induction

(Brody & Lapata, 2009)

Senses of <i>drug</i> (WSJ)
1. U.S., administration, federal, against, war, dealer
2. patient, people, problem, doctor, company, abuse
3. company, million, sale, maker, stock, inc.
4. administration, food, company, approval, FDA

Senses of <i>drug</i> (BNC)
1. patient, treatment, effect, anti-inflammatory
2. alcohol, treatment, patient, therapy, addiction
3. patient, new, find, effect, choice, study
4. test, alcohol, patient, abuse, people, crime
5. trafficking, trafficker, charge, use, problem
6. abuse, against, problem, treatment, alcohol
7. people, wonder, find, prescription, drink, addict
8. company, dealer, police, enforcement, patient

- Selectional Preference

(Ritter et al., 2010)

Topic <i>t</i>	Arg1	Relations which assign highest probability to <i>t</i>	Arg2
18	The residue - The mixture - The reaction mixture - The solution - the mixture - the reaction mixture - the residue - The reaction - the solution - The filtrate - the reaction - The product - The crude product - The pellet - The organic layer - Thereto - This solution - The resulting solution - Next - The organic phase - The resulting mixture - C.)	was treated with, is treated with, was poured into, was extracted with, was purified by, was diluted with, was filtered through, is dissolved in, is washed with	EtOAc - CH ₂ Cl ₂ - H ₂ O - CH.sub.2Cl.sub.2 - H.sub.2O - water - MeOH - NaHCO ₃ - Et ₂ O - NHCl - CHCl.sub.3 - NHCl - drop-wise - CH ₂ Cl.sub.2 - Celite - Et.sub.2O - Cl.sub.2 - NaOH - AcOEt - CH ₂ Cl ₂ - the mixture - saturated NaHCO ₃ - SiO ₂ - H ₂ O - N hydrochloric acid - NHCl - preparative HPLC - to 0 C

Text Data

- Word/term
- Document
 - A sequence of words
- Corpus
 - A collection of documents



Represent a Document

- Most common way: Bag-of-Words
 - Ignore the order of words
 - keep the count

	c1	c2	c3	c4	c5	m1	m2	m3	m4
<i>human</i>	1	0	0	1	0	0	0	0	0
<i>interface</i>	1	0	1	0	0	0	0	0	0
<i>computer</i>	1	1	0	0	0	0	0	0	0
<i>user</i>	0	1	1	0	1	0	0	0	0
<i>system</i>	0	1	1	2	0	0	0	0	0
<i>response</i>	0	1	0	0	1	0	0	0	0
<i>time</i>	0	1	0	0	1	0	0	0	0
<i>EPS</i>	0	0	1	1	0	0	0	0	0
<i>survey</i>	0	1	0	0	0	0	0	0	1
<i>trees</i>	0	0	0	0	0	1	1	1	0
<i>graph</i>	0	0	0	0	0	0	1	1	1
<i>minors</i>	0	0	0	0	0	0	0	1	1

Vector space model

Represent a Document

- Represent the doc as a vector where each entry corresponds to a different word and the number at that entry corresponds to how many times that word was present in the document (or some function of it)
 - Number of words is huge
 - Select and use a smaller set of words that are of interest
 - E.g. uninteresting words: 'and', 'the' 'at', 'is', etc. These are called stop-words
 - Stemming: remove endings. E.g. 'learn', 'learning', 'learnable', 'learned' could be substituted by the single stem 'learn'
 - Other simplifications can also be invented and used
 - The set of different remaining words is called dictionary or vocabulary. Fix an ordering of the terms in the dictionary so that you can operate them by their index.
 - Can be extended to bi-gram, tri-gram, or so

Limitations of Bag-of-Words

- Dimensionality
 - High dimensionality
- Sparseness
 - Most of the entries are zero
- Shallow representation
 - The vector representation does not capture semantic relations between words

Ex: "Tom loves Kate."

Represent a Topic

- A topic is represented by a word distribution
- Relate to an issue

universe	0.0439	drug	0.0672	cells	0.0675	sequence	0.0818	years	0.156
galaxies	0.0375	patients	0.0493	stem	0.0478	sequences	0.0493	million	0.0556
clusters	0.0279	drugs	0.0444	human	0.0421	genome	0.033	ago	0.045
matter	0.0233	clinical	0.0346	cell	0.0309	dna	0.0257	time	0.0317
galaxy	0.0232	treatment	0.028	gene	0.025	sequencing	0.0172	age	0.0243
cluster	0.0214	trials	0.0277	tissue	0.0185	map	0.0123	year	0.024
cosmic	0.0137	therapy	0.0213	cloning	0.0169	genes	0.0122	record	0.0238
dark	0.0131	trial	0.0164	transfer	0.0155	chromosome	0.0119	early	0.0233
light	0.0109	disease	0.0157	blood	0.0113	regions	0.0119	billion	0.0177
density	0.01	medical	0.00997	embryos	0.0111	human	0.0111	history	0.0148

bacteria	0.0983	male	0.0558	theory	0.0811	immune	0.0909	stars	0.0524
bacterial	0.0561	females	0.0541	physics	0.0782	response	0.0375	star	0.0458
resistance	0.0431	female	0.0529	physicists	0.0146	system	0.0358	astrophys	0.0237
coli	0.0381	males	0.0477	einstein	0.0142	responses	0.0322	mass	0.021
strains	0.025	sex	0.0339	university	0.013	antigen	0.0263	disk	0.0173
microbiol	0.0214	reproductive	0.0172	gravity	0.013	antigens	0.0184	black	0.0161
microbial	0.0196	offspring	0.0168	black	0.0127	immunity	0.0176	gas	0.0149
strain	0.0165	sexual	0.0166	theories	0.01	immunology	0.0145	stellar	0.0127
salmonella	0.0163	reproduction	0.0143	aps	0.00987	antibody	0.014	astron	0.0125
resistant	0.0145	eggs	0.0138	matter	0.00954	autoimmune	0.0128	hole	0.00824



Topic Models

- Topic modeling
 - Get topics automatically from a corpus
 - Assign documents to topics automatically
- Most frequently used topic models
 - pLSA
 - LDA

“Arts”	“Budgets”	“Children”	“Education”
NEW	MILLION	CHILDREN	SCHOOL
FILM	TAX	WOMEN	STUDENTS
SHOW	PROGRAM	PEOPLE	SCHOOLS
MUSIC	BUDGET	CHILD	EDUCATION
MOVIE	BILLION	YEARS	TEACHERS
PLAY	FEDERAL	FAMILIES	HIGH
MUSICAL	YEAR	WORK	PUBLIC
BEST	SPENDING	PARENTS	TEACHER
ACTOR	NEW	SAYS	BENNETT
FIRST	STATE	FAMILY	MANIGAT
YORK	PLAN	WELFARE	NAMPHY
OPERA	MONEY	MEN	STATE
THEATER	PROGRAMS	PERCENT	PRESIDENT
ACTRESS	GOVERNMENT	CARE	ELEMENTARY
LOVE	CONGRESS	LIFE	HAITI

The William Randolph Hearst Foundation will give \$1.25 million to Lincoln Center, Metropolitan Opera Co., New York Philharmonic and Juilliard School. “Our board felt that we had a real opportunity to make a mark on the future of the performing arts with these grants an act every bit as important as our traditional areas of support in health, medical research, education and the social services,” Hearst Foundation President Randolph A. Hearst said Monday in announcing the grants. Lincoln Center’s share will be \$200,000 for its new building, which will house young artists and provide new public facilities. The Metropolitan Opera Co. and New York Philharmonic will receive \$400,000 each. The Juilliard School, where music and the performing arts are taught, will get \$250,000. The Hearst Foundation, a leading supporter of the Lincoln Center Consolidated Corporate Fund, will make its usual annual \$100,000 donation, too.

Notations

- Word, document, topic
 - w, d, z
- Word count in document:
 - $c(w, d)$: number of times word w occurs in document d
 - or x_{dn} : number of times the n th word in the vocabulary occurs in document d
- Word distribution for each topic (β_z)
 - β_{zw} : $p(w|z)$



Recap: Multinomial distribution

- Multinomial distribution

- Discrete random variable x that takes one of M values $\{1, \dots, M\}$
- $p(x = i) = \pi_i, \quad \sum_i \pi_i = 1$
- Out of n independent trials, let k_i be the number of times $x = i$ was observed
- The probability of observing a vector of occurrences $\mathbf{k} = [k_1, \dots, k_M]$ is given by the *multinomial distribution* parametrized by $\boldsymbol{\pi}$

$$p(\mathbf{k}|\boldsymbol{\pi}, n) = p(k_1, \dots, k_M | \pi_1, \dots, \pi_M, n) = \frac{n!}{k_1! k_2! \dots k_M!} \prod_{i=1}^M \pi_i^{k_i}$$

- E.g., describing a text document by the frequency of occurrence of every distinct word
- For $n = 1$, a.k.a. *categorical distribution*
 - $p(x = i | \boldsymbol{\pi}) = \pi_i$
 - In $\mathbf{k} = [k_1, \dots, k_M]$: $k_i = 1$, and $k_j = 0$ for all $j \neq i \rightarrow$ a.k.a., *one-hot representation* of i

Topic Model v1: Multinomial Mixture Model

- For documents with bag-of-words representation
 - $\mathbf{x}_d = (x_{d1}, x_{d2}, \dots, x_{dN})$, x_{dn} is the number of words for nth word in the vocabulary
- Generative model

Topic Model v1: Multinomial Mixture Model

- For documents with bag-of-words representation
 - $\mathbf{x}_d = (x_{d1}, x_{d2}, \dots, x_{dN})$, x_{dn} is the number of words for n th word in the vocabulary
- Generative model

Formulating the statistical relationship between words, documents and latent topics as a generative process describing how documents are created

Topic Model v1: Multinomial Mixture Model

- For documents with bag-of-words representation
 - $\mathbf{x}_d = (x_{d1}, x_{d2}, \dots, x_{dN})$, x_{dn} is the number of words for nth word in the vocabulary
- Generative model
 - For each document
 - Sample its cluster label $z \sim \text{Categorical}(\boldsymbol{\pi})$
 - $\boldsymbol{\pi} = (\pi_1, \pi_2, \dots, \pi_K)$, π_k is the proportion of jth cluster
 - $p(z = k) = \pi_k$
 - Sample its word vector $\mathbf{x}_d \sim \text{multinomial}(\boldsymbol{\beta}_z)$
 - $\boldsymbol{\beta}_z = (\beta_{z1}, \beta_{z2}, \dots, \beta_{zN})$, β_{zn} is the parameter associate with nth word in the vocabulary
 - $p(\mathbf{x}_d | z = k) = \frac{(\sum_n x_{dn})!}{\prod_n x_{dn}!} \prod_n \beta_{kn}^{x_{dn}} \propto \prod_n \beta_{kn}^{x_{dn}}$

Topic Model v1: Multinomial Mixture Model

Graphical
Model

- *Plates indicate replicated variables.*
- *Shaded nodes are observed; unshaded nodes are hidden.*

- Generative model

- For each document

- Sample its cluster label $z \sim \text{Categorical}(\boldsymbol{\pi})$

- $\boldsymbol{\pi} = (\pi_1, \pi_2, \dots, \pi_K)$, π_k is the proportion of j th cluster

- $p(z = k) = \pi_k$

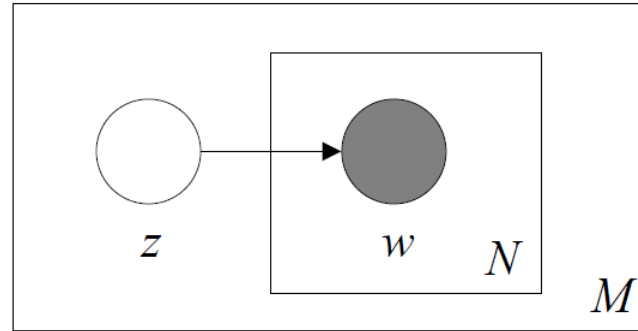
- Sample its word vector $\mathbf{x}_d \sim \text{multinomial}(\boldsymbol{\beta}_z)$

- $\boldsymbol{\beta}_z = (\beta_{z1}, \beta_{z2}, \dots, \beta_{zN})$, β_{zn} is the parameter associate with n th word in the vocabulary

- $p(\mathbf{x}_d | z = k) = \frac{(\sum_n x_{dn})!}{\prod_n x_{dn}!} \prod_n \beta_{kn}^{x_{dn}} \propto \prod_n \beta_{kn}^{x_{dn}}$

Topic Model v1: Multinomial Mixture Model

Graphical
Model



- *Plates indicate replicated variables.*
- *Shaded nodes are observed; unshaded nodes are hidden.*

- Generative model

- For each document

- Sample its cluster label $z \sim \text{Categorical}(\boldsymbol{\pi})$

- $\boldsymbol{\pi} = (\pi_1, \pi_2, \dots, \pi_K)$, π_k is the proportion of j th cluster

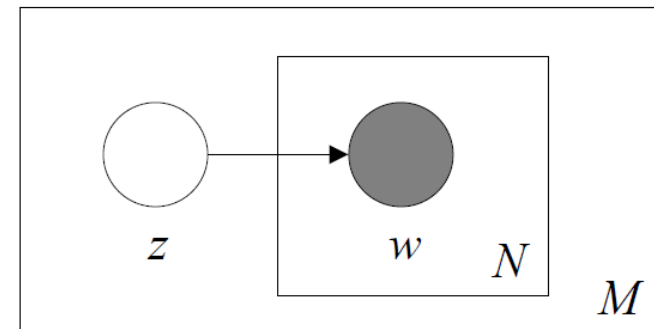
- $p(z = k) = \pi_k$

- Sample its word vector $\mathbf{x}_d \sim \text{multinomial}(\boldsymbol{\beta}_z)$

- $\boldsymbol{\beta}_z = (\beta_{z1}, \beta_{z2}, \dots, \beta_{zN})$, β_{zn} is the parameter associate with n th word in the vocabulary

- $p(\mathbf{x}_d | z = k) = \frac{(\sum_n x_{dn})!}{\prod_n x_{dn}!} \prod_n \beta_{kn}^{x_{dn}} \propto \prod_n \beta_{kn}^{x_{dn}}$

Likelihood Function

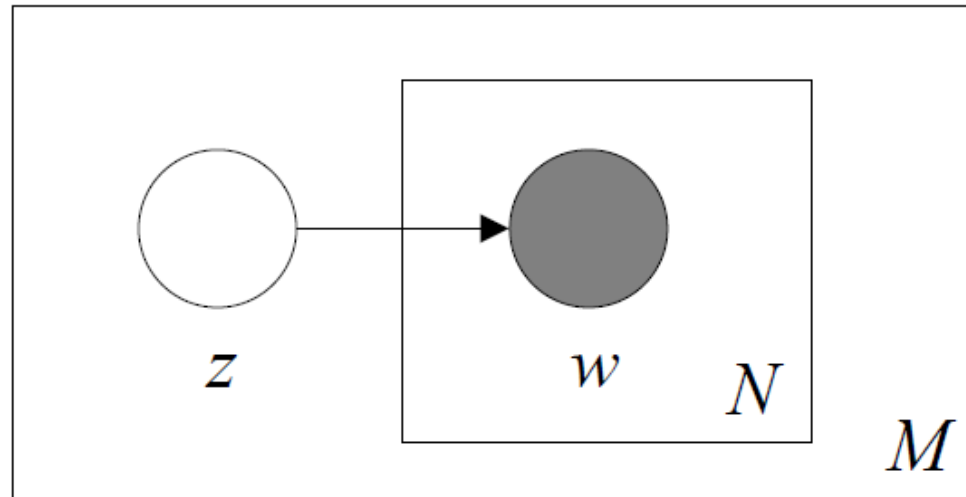


Likelihood Function

$$\begin{aligned} L &= \prod_d p(\mathbf{x}_d) = \prod_d \sum_k p(\mathbf{x}_d, z = k) \\ &= \prod_d \sum_k p(\mathbf{x}_d | z = k) p(z = k) \\ &= \prod_d \frac{(\sum_n x_{dn})!}{\prod_n x_{dn}!} \sum_k p(z = k) \prod_n \beta_{kn}^{x_{dn}} \end{aligned}$$

Limitations of Multinomial Mixture Model

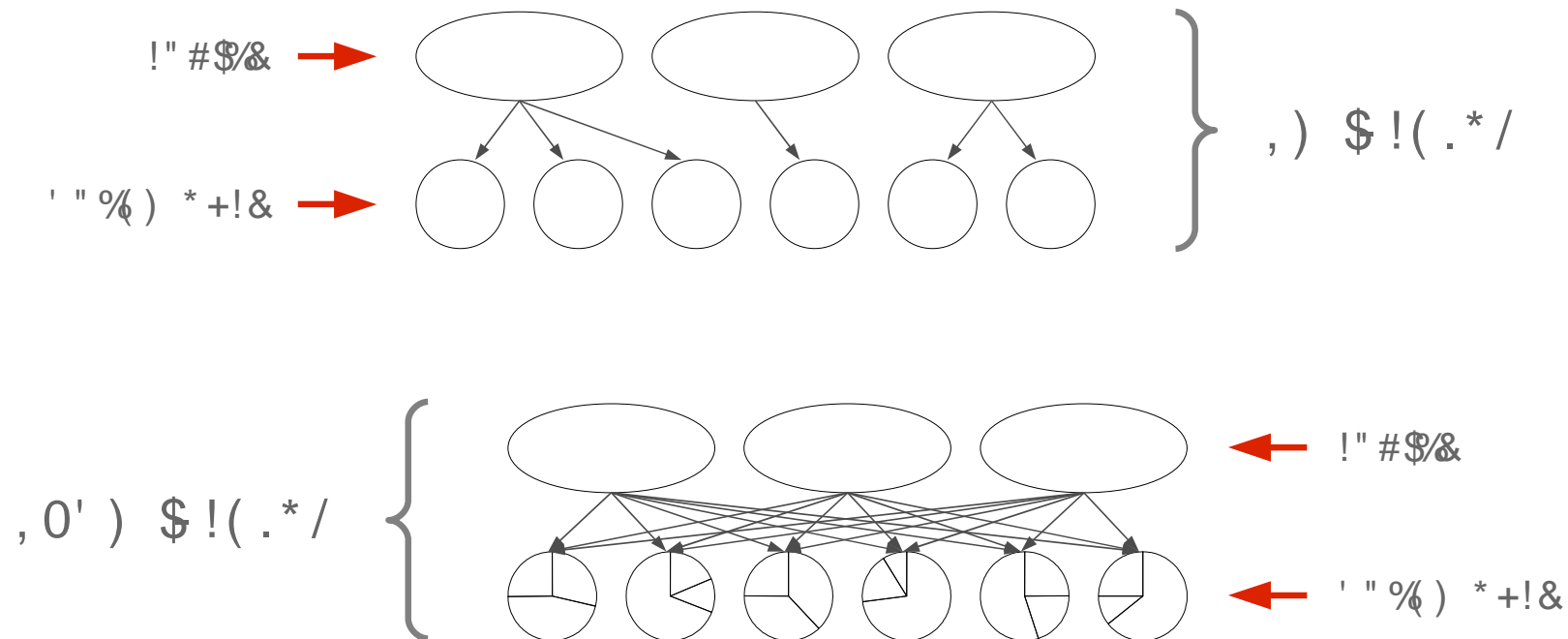
- All the words in the same documents are sampled from the same topic



- In practice, people switch topics during their writing

Limitations of Multinomial Mixture Model

Mixture vs. 'Admixture'



Diagrams from Wallach, JHU 2011, slides

Topic Model v2: Probabilistic Latent Semantic Analysis (pLSA)

“Arts” “Budgets” “Children” “Education”

NEW	MILLION	CHILDREN	SCHOOL
FILM	TAX	WOMEN	STUDENTS
SHOW	PROGRAM	PEOPLE	SCHOOLS
MUSIC	BUDGET	CHILD	EDUCATION
MOVIE	BILLION	YEARS	TEACHERS
PLAY	FEDERAL	FAMILIES	HIGH
MUSICAL	YEAR	WORK	PUBLIC
BEST	SPENDING	PARENTS	TEACHER
ACTOR	NEW	SAYS	BENNETT
FIRST	STATE	FAMILY	MANIGAT
YORK	PLAN	WELFARE	NAMPHY
OPERA	MONEY	MEN	STATE
THEATER	PROGRAMS	PERCENT	PRESIDENT
ACTRESS	GOVERNMENT	CARE	ELEMENTARY
LOVE	CONGRESS	LIFE	HAITI

The William Randolph Hearst Foundation will give \$1.25 million to Lincoln Center, Metropolitan Opera Co., New York Philharmonic and Juilliard School. “Our board felt that we had a real opportunity to make a mark on the future of the performing arts with these grants an act every bit as important as our traditional areas of support in health, medical research, education and the social services,” Hearst Foundation President Randolph A. Hearst said Monday in announcing the grants. Lincoln Center’s share will be \$200,000 for its new building, which will house young artists and provide new public facilities. The Metropolitan Opera Co. and New York Philharmonic will receive \$400,000 each. The Juilliard School, where music and the performing arts are taught, will get \$250,000. The Hearst Foundation, a leading supporter of the Lincoln Center Consolidated Corporate Fund, will make its usual annual \$100,000 donation, too.

Generative Model for pLSA

- For each position in d , $n = 1, \dots, N_d$

- Generate the topic for the position as

$$z_n | d \sim \text{Categorical}(\boldsymbol{\theta}_d), \text{ i.e., } p(z_n = k | d) = \theta_{dk}$$

(Note, 1 trial multinomial)

- Generate the word for the position as

$$w_n | z_n \sim \text{Categorical}(\boldsymbol{\beta}_{z_n}), \text{ i.e., } p(w_n = w | z_n) = \beta_{z_n w}$$

Generative Model for pLSA

- For each position in d , $n = 1, \dots, N_d$

- Generate the topic for the position as

$$z_n | d \sim \text{Categorical}(\boldsymbol{\theta}_d), \text{ i.e., } p(z_n = k | d) = \theta_{dk}$$

(Note, 1 trial multinomial)

- Generate the word for the position as

$$w_n | z_n \sim \text{Categorical}(\boldsymbol{\beta}_{z_n}), \text{ i.e., } p(w_n = w | z_n) = \beta_{z_n w}$$

Graphical
Model

Generative Model for pLSA

- For each position in d , $n = 1, \dots, N_d$

- Generate the topic for the position as

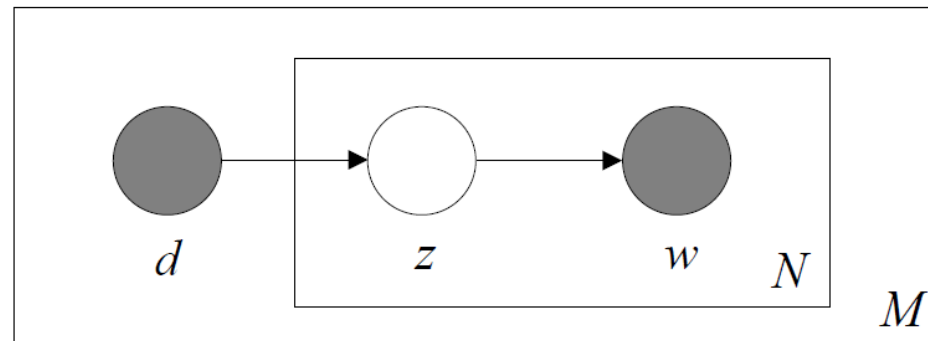
$$z_n | d \sim \text{Categorical}(\boldsymbol{\theta}_d), \text{ i.e., } p(z_n = k | d) = \theta_{dk}$$

(Note, 1 trial multinomial)

- Generate the word for the position as

$$w_n | z_n \sim \text{Categorical}(\boldsymbol{\beta}_{z_n}), \text{ i.e., } p(w_n = w | z_n) = \beta_{z_n w}$$

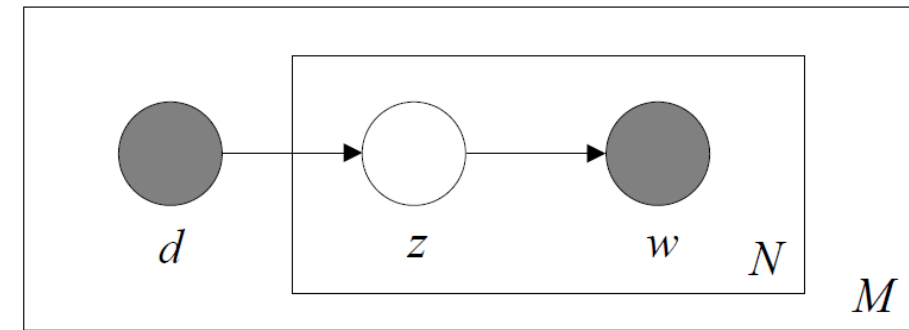
Graphical
Model



Likelihood Function

- Probability of a word w

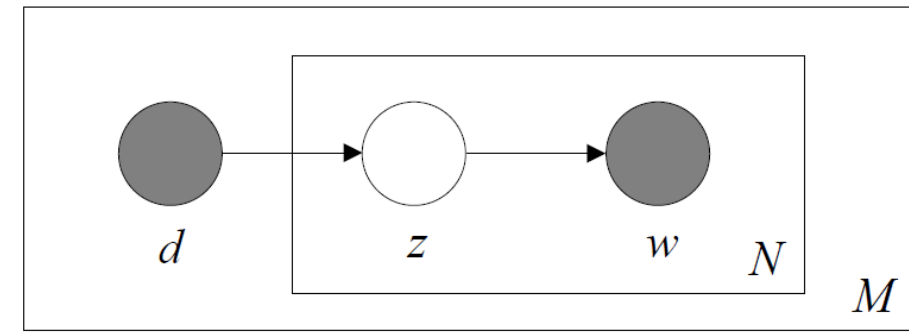
$$p(w|d, \theta, \beta)$$



Likelihood Function

- Probability of a word w

$$\begin{aligned} p(w|d, \theta, \beta) &= \sum_k p(w, z = k|d, \theta, \beta) \\ &= \sum_k p(w|z = k, d, \theta, \beta) p(z = k|d, \theta, \beta) = \sum_k \beta_{kw} \theta_{dk} \end{aligned}$$



Likelihood Function

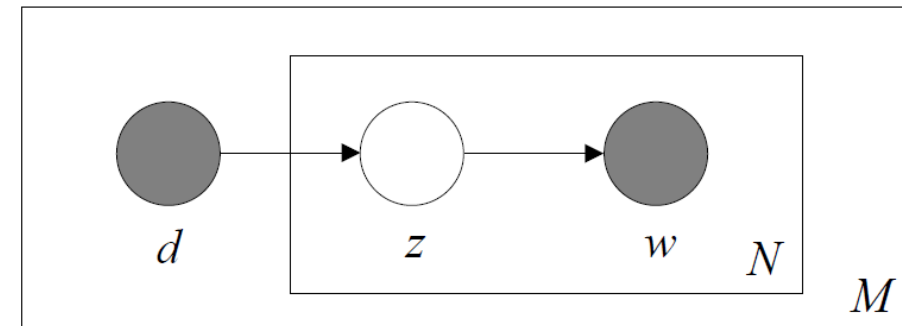
- Probability of a word w

$$\begin{aligned} p(w|d, \theta, \beta) &= \sum_k p(w, z = k|d, \theta, \beta) \\ &= \sum_k p(w|z = k, d, \theta, \beta) p(z = k|d, \theta, \beta) = \sum_k \beta_{kw} \theta_{dk} \end{aligned}$$

- Likelihood of a corpus

$$\begin{aligned} &\prod_{d=1} P(w_1, \dots, w_{N_d}, d|\theta, \beta, \pi) \\ &= \prod_{d=1} P(d) \left\{ \prod_{n=1}^{N_d} \left(\sum_k P(z_n = k|d, \theta_d) P(w_n|\beta_k) \right) \right\} \\ &= \prod_{d=1} \pi_d \left\{ \prod_{n=1}^{N_d} \left(\sum_k \theta_{dk} \beta_{kw_n} \right) \right\} \end{aligned}$$

π_d is usually considered as uniform, i.e., $1/M$



Re-arrange the Likelihood Function

- Group the same word from different positions together

$$\max \log L = \sum_{dw} c(w, d) \log \sum_z \theta_{dz} \beta_{zw}$$

$$s. t. \sum_z \theta_{dz} = 1 \text{ and } \sum_w \beta_{zw} = 1$$

Limitations of pLSA

- Not a proper generative model
 - θ_d is treated as a parameter
 - Cannot model new documents
- Solution:
 - Make it a proper generative model by adding priors to θ and β

Limitations of pLSA

- Not a proper generative model
 - θ_d is treated as a parameter
 - Cannot model new documents
- Solution:
 - Make it a proper generative model by adding priors to θ and β



Topic Model v3: Latent Dirichlet Allocation (LDA)

Review: Dirichlet Distribution

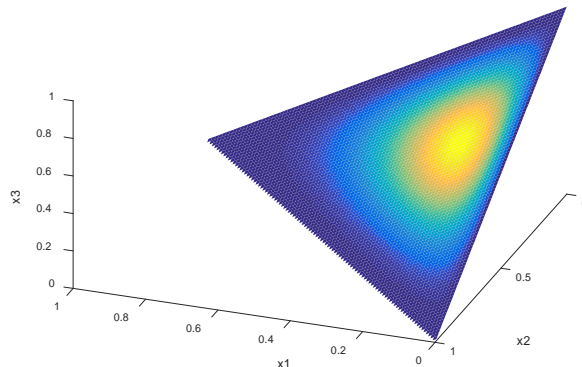
- Dirichlet distribution: $\boldsymbol{\theta} \sim \text{Dirichlet}(\boldsymbol{\alpha})$
 - i. e., $p(\boldsymbol{\theta}|\boldsymbol{\alpha}) = \frac{\Gamma(\sum_k \alpha_k)}{\prod_k \Gamma(\alpha_k)} \prod_k \theta_k^{\alpha_k-1}$, where $\alpha_k > 0$
 - $\Gamma(\cdot)$ is *gamma function*: $\Gamma(z) = \int_0^\infty e^{-t} t^{z-1} dt$
 - $\Gamma(z+1) = z\Gamma(z)$

Review: Dirichlet Distribution

- Dirichlet distribution: $\theta \sim \text{Dirichlet}(\alpha)$
 - i. e., $p(\theta|\alpha) = \frac{\Gamma(\sum_k \alpha_k)}{\prod_k \Gamma(\alpha_k)} \prod_k \theta_k^{\alpha_k - 1}$, where $\alpha_k > 0$
 - $\Gamma(\cdot)$ is gamma function: $\Gamma(z) = \int_0^\infty e^{-t} t^{z-1} dt$
 - $\Gamma(z+1) = z\Gamma(z)$

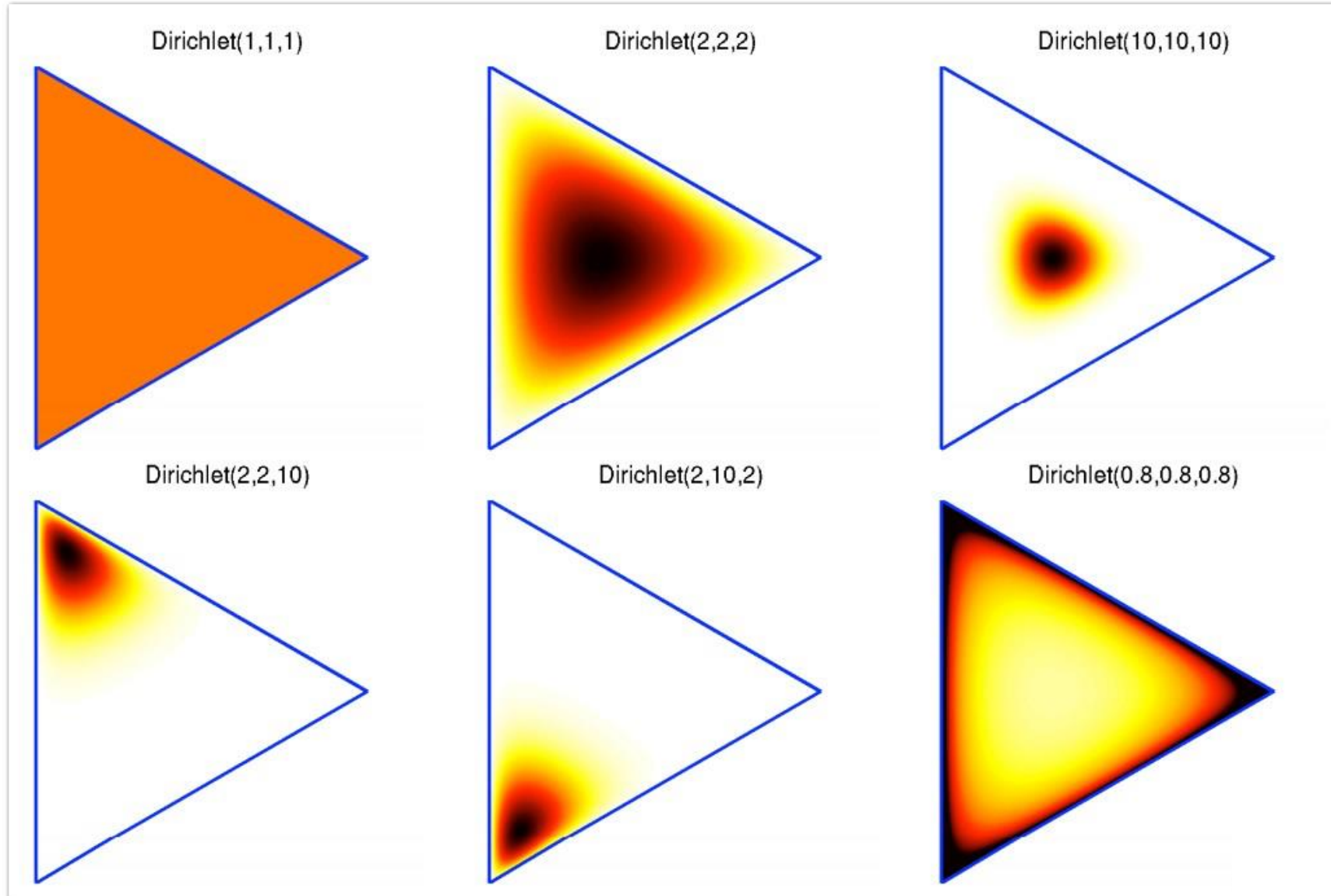
Simplex view:

- $x = x_1(1,0,0) + x_2(0,1,0) + x_3(0,0,1)$
- Where $0 \leq x_1, x_2, x_3 \leq 1$ and $x_1 + x_2 + x_3 = 1$

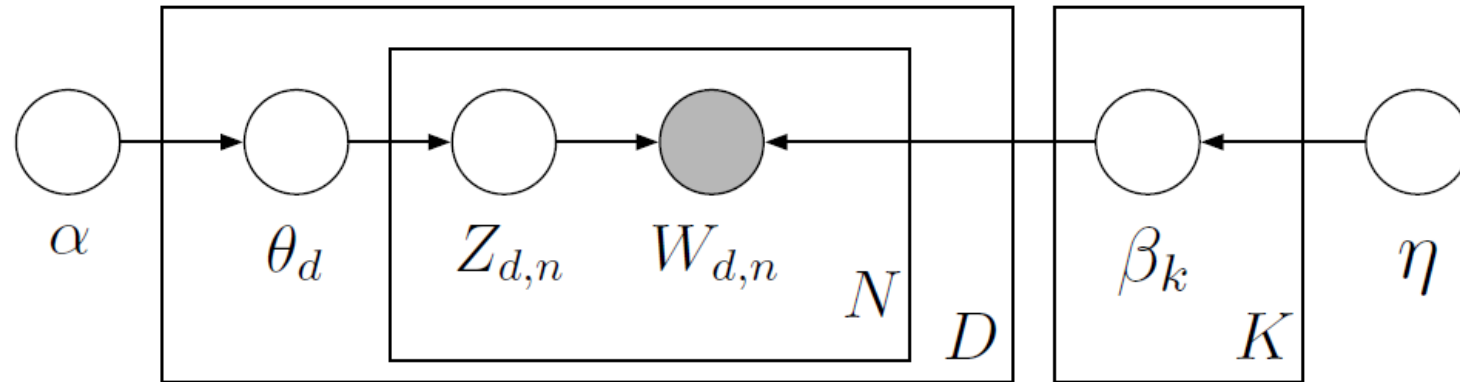


$x|\alpha \sim \text{Dir}(\alpha), \alpha = (2,3,4)$

More Examples in the Simplex View



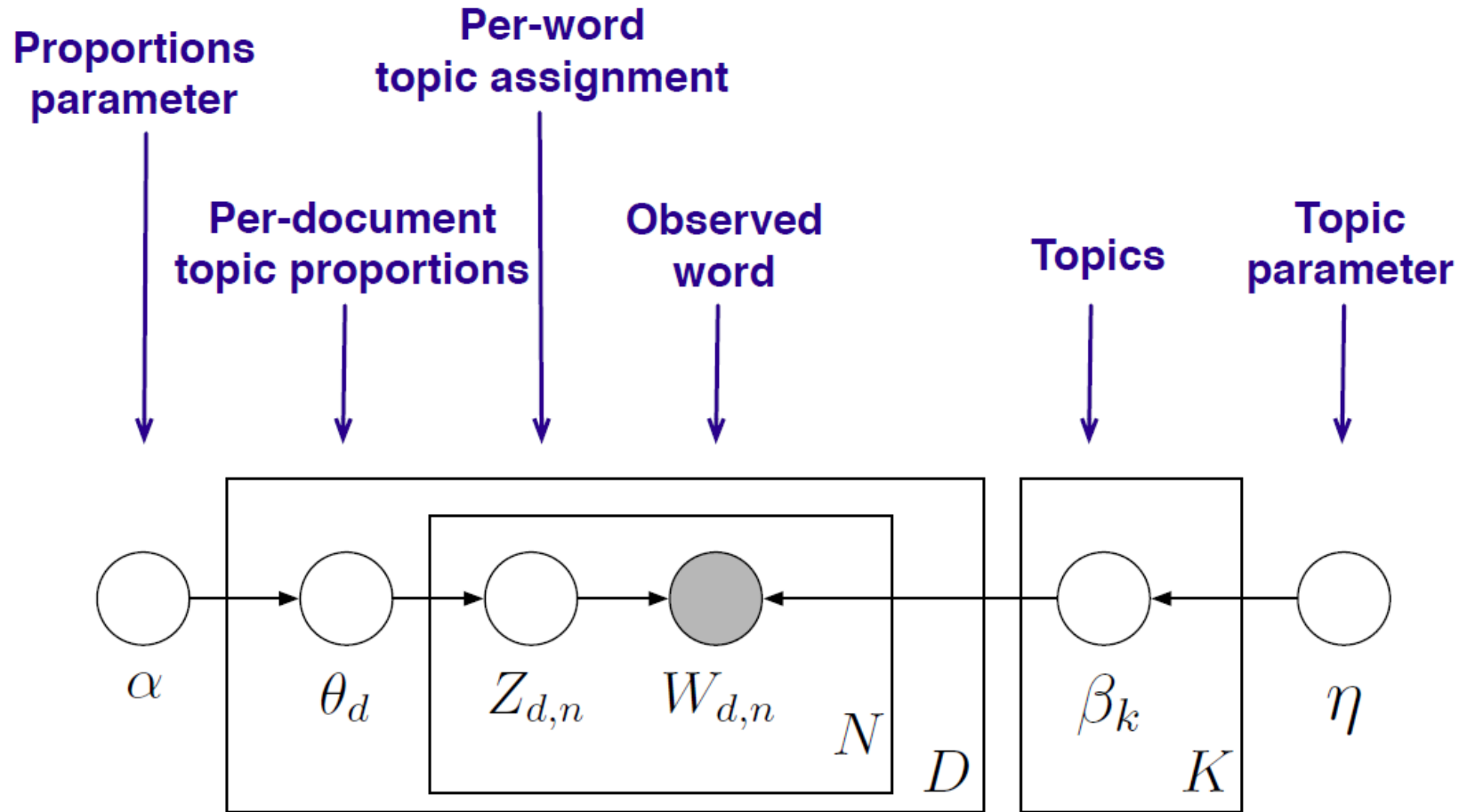
Topic Model v3: Latent Dirichlet Allocation (LDA)



$\theta_d \sim \text{Dirichlet}(\alpha)$: address topic distribution for unseen documents

$\beta_k \sim \text{Dirichlet}(\eta)$: smoothing over words

Topic Model v3: Latent Dirichlet Allocation (LDA)



$\theta_d \sim \text{Dirichlet}(\alpha)$: address topic distribution for unseen documents

$\beta_k \sim \text{Dirichlet}(\eta)$: smoothing over words

Generative Model for LDA

For each topic $k \in \{1, \dots, K\}$:

$\beta_k \sim \text{Dir}(\eta)$ *[draw distribution over words]*

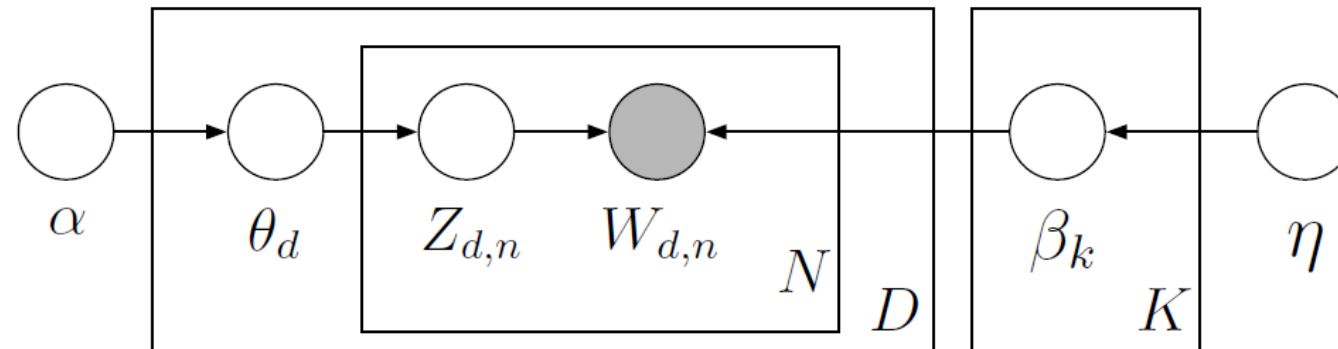
For each document $d \in \{1, \dots, D\}$

$\theta_d \sim \text{Dir}(\alpha)$ *[draw distribution over topics]*

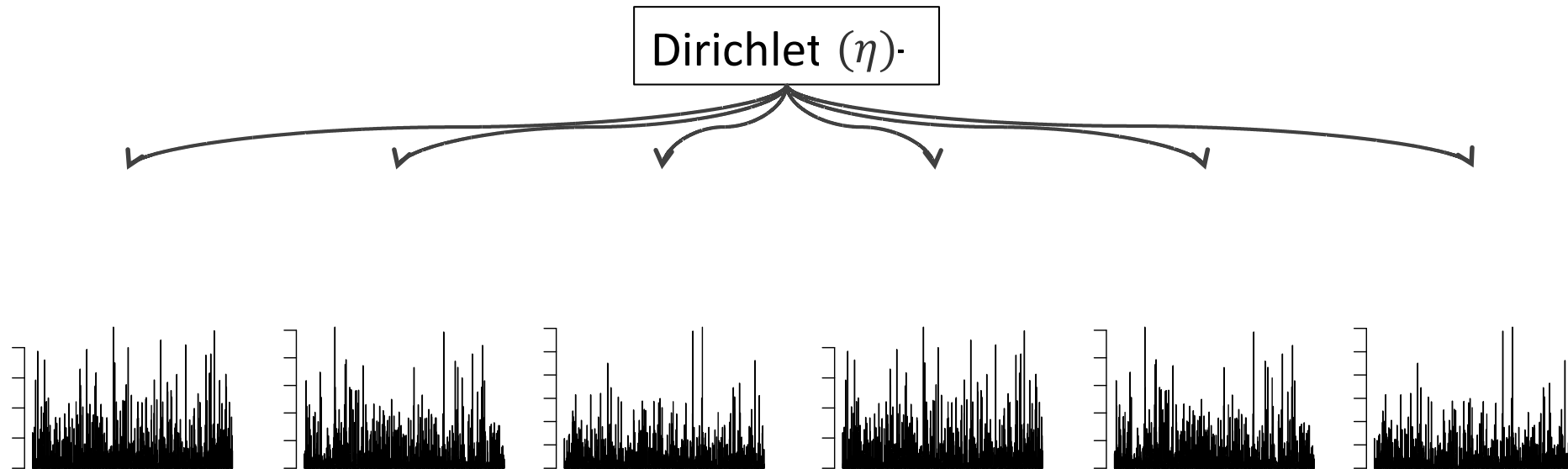
For each word $n \in \{1, \dots, N_d\}$

$z_{d,n} \sim \text{Mult}(1, \theta_d)$ *[draw topic assignment]*

$w_{d,n} \sim \theta_{z_{d,n}}$ *[draw word]*

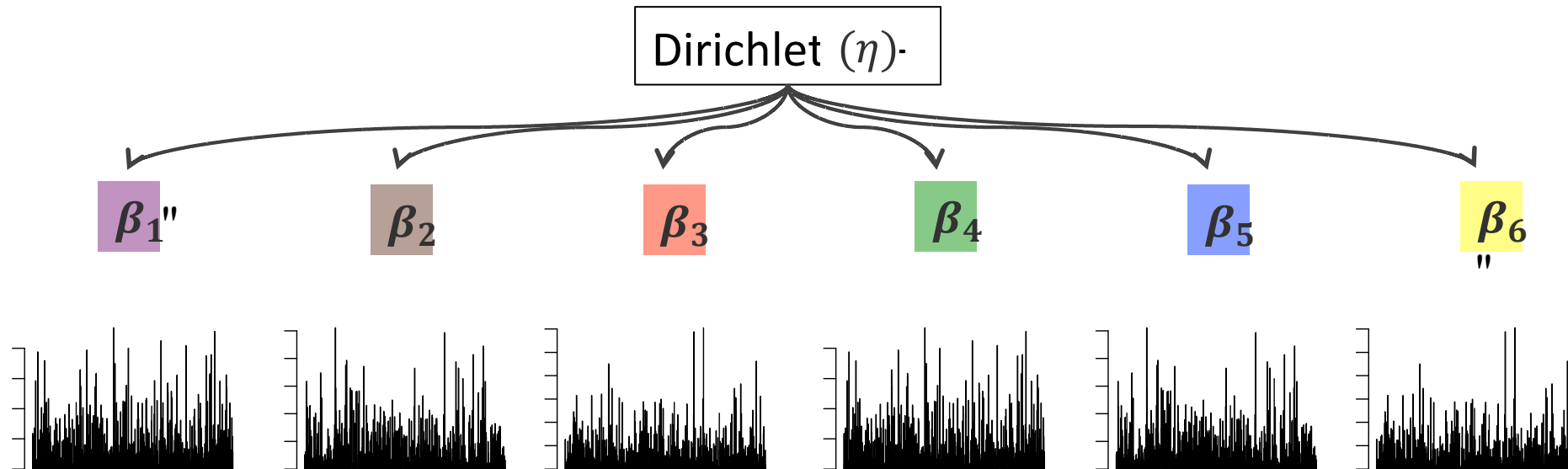


LDA 'for' 'Topic' Modeling''



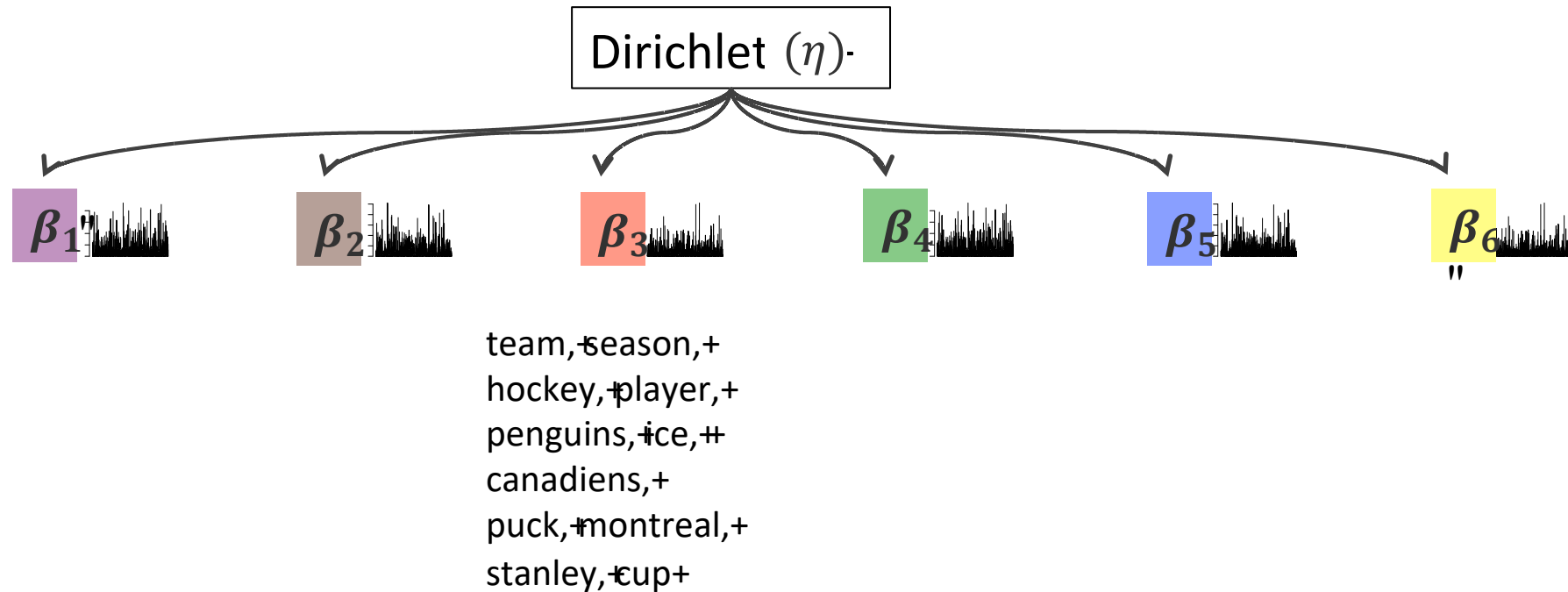
- The 'generative story' begins with only a **Dirichlet prior** over the topics."
- Each **topic** defines a **Multinomial distribution** over the vocabulary, parameterized by ' β_k '

LDA 'for' 'Topic' Modeling''



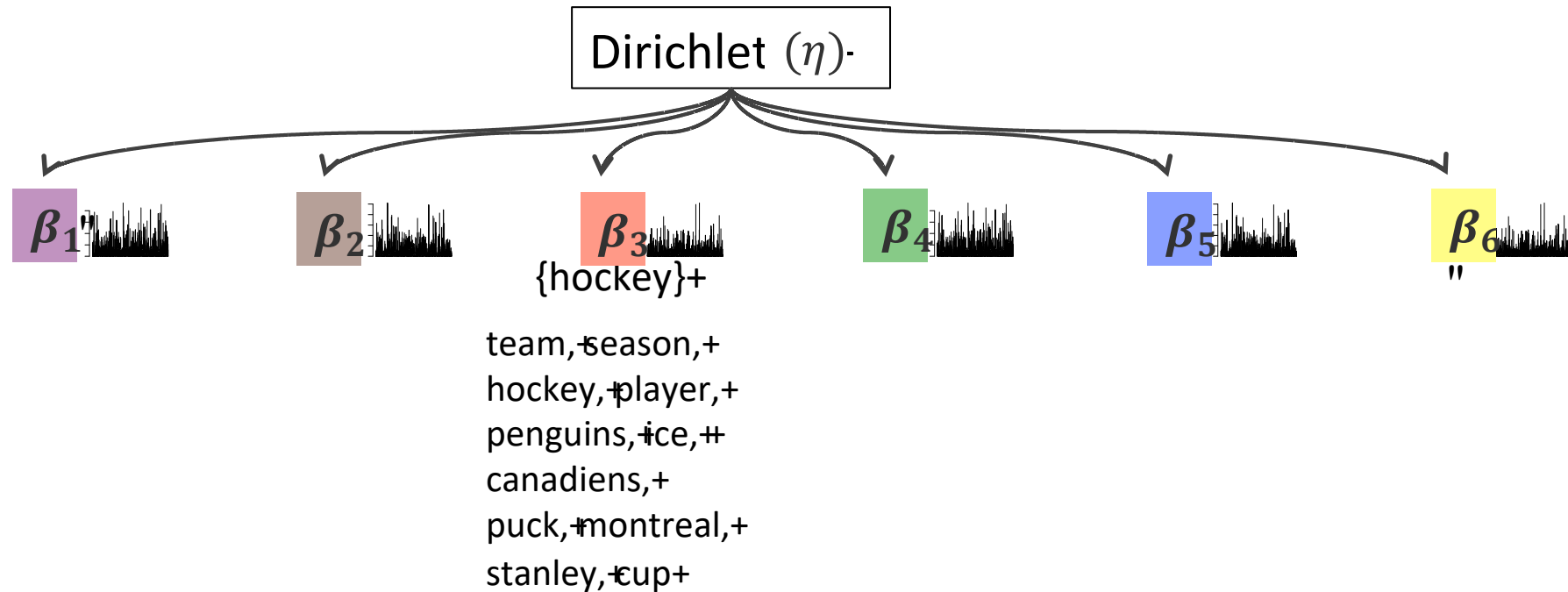
- The 'generative story' begins with only a Dirichlet prior over the topics.
- Each topic defines a 'Multinomial distribution' over the vocabulary, parameterized by ' β_k '

LDA 'for' 'Topic' Modeling''



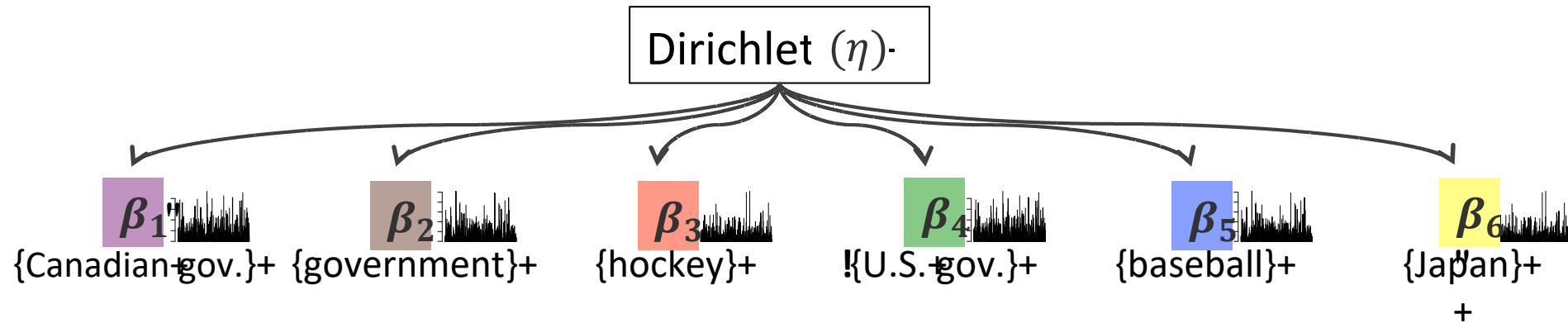
- A 'topic' is visualized as its high probability words."

LDA 'for' 'Topic' Modeling''



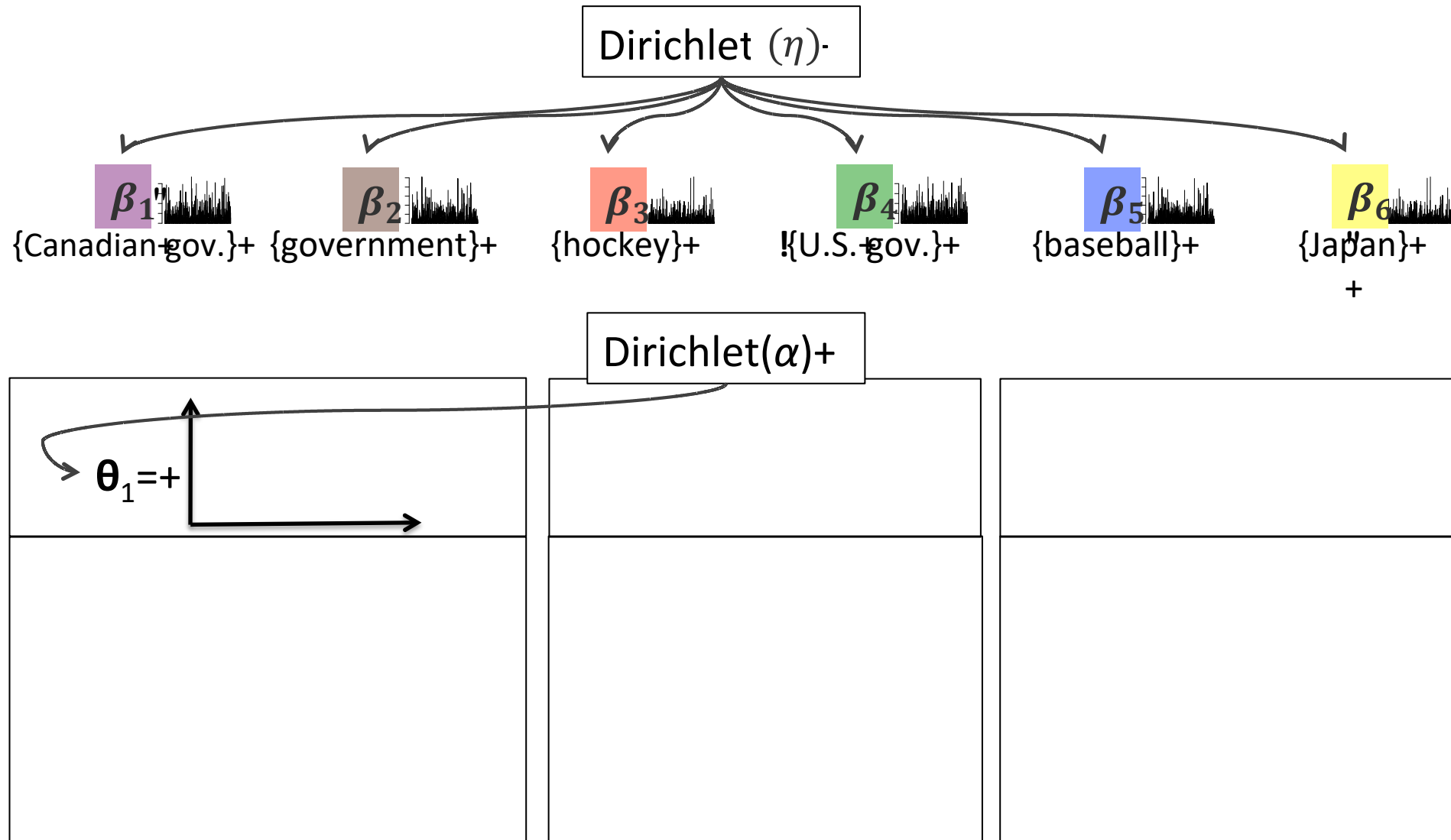
- A 'topic' is visualized as its high probability words."
- A pedagogical label is used to identify the topic."

LDA for Topic Modeling

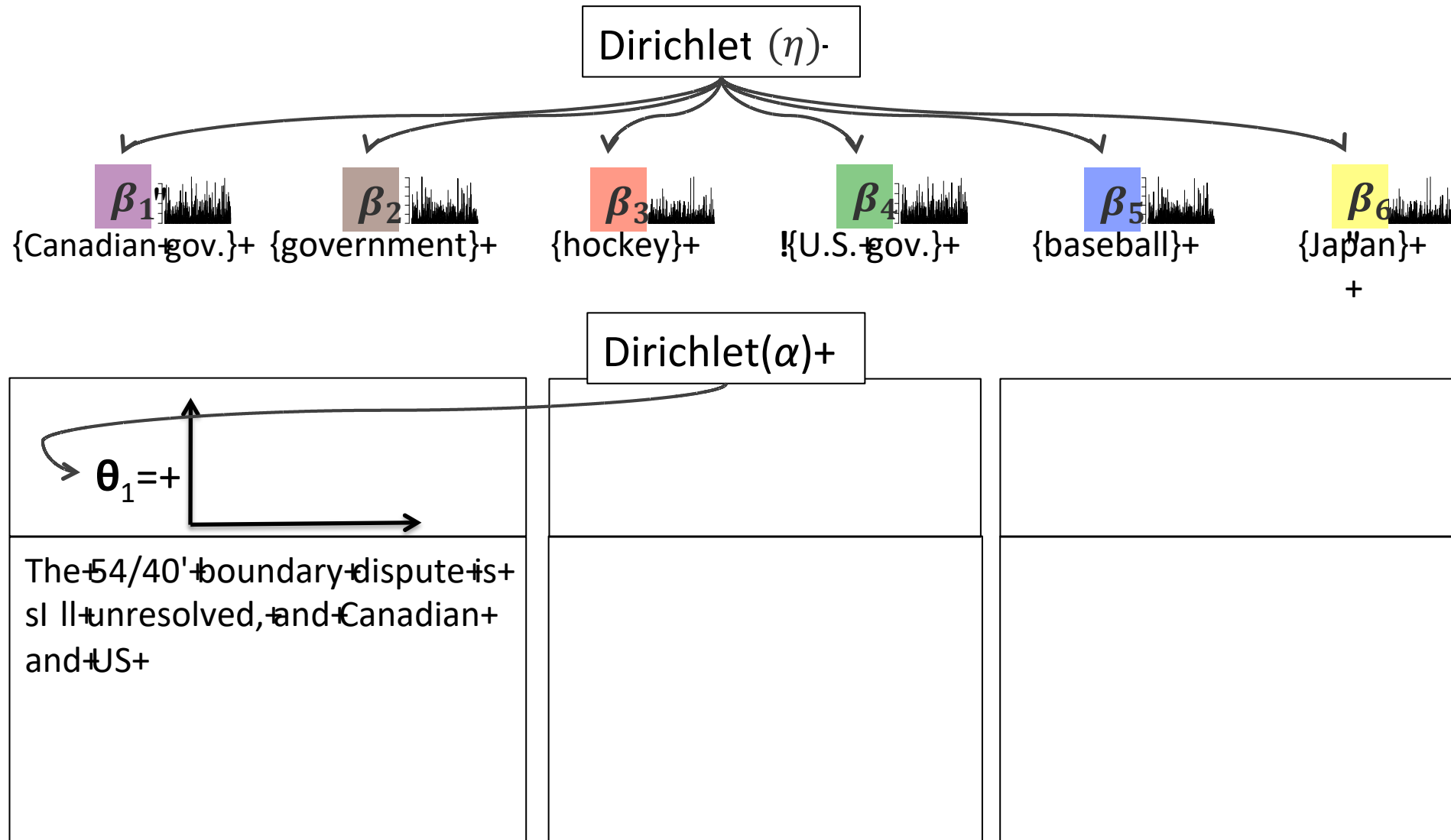


- A topic is visualized as its high probability words."
- A pedagogical **label** is used to identify the topic."

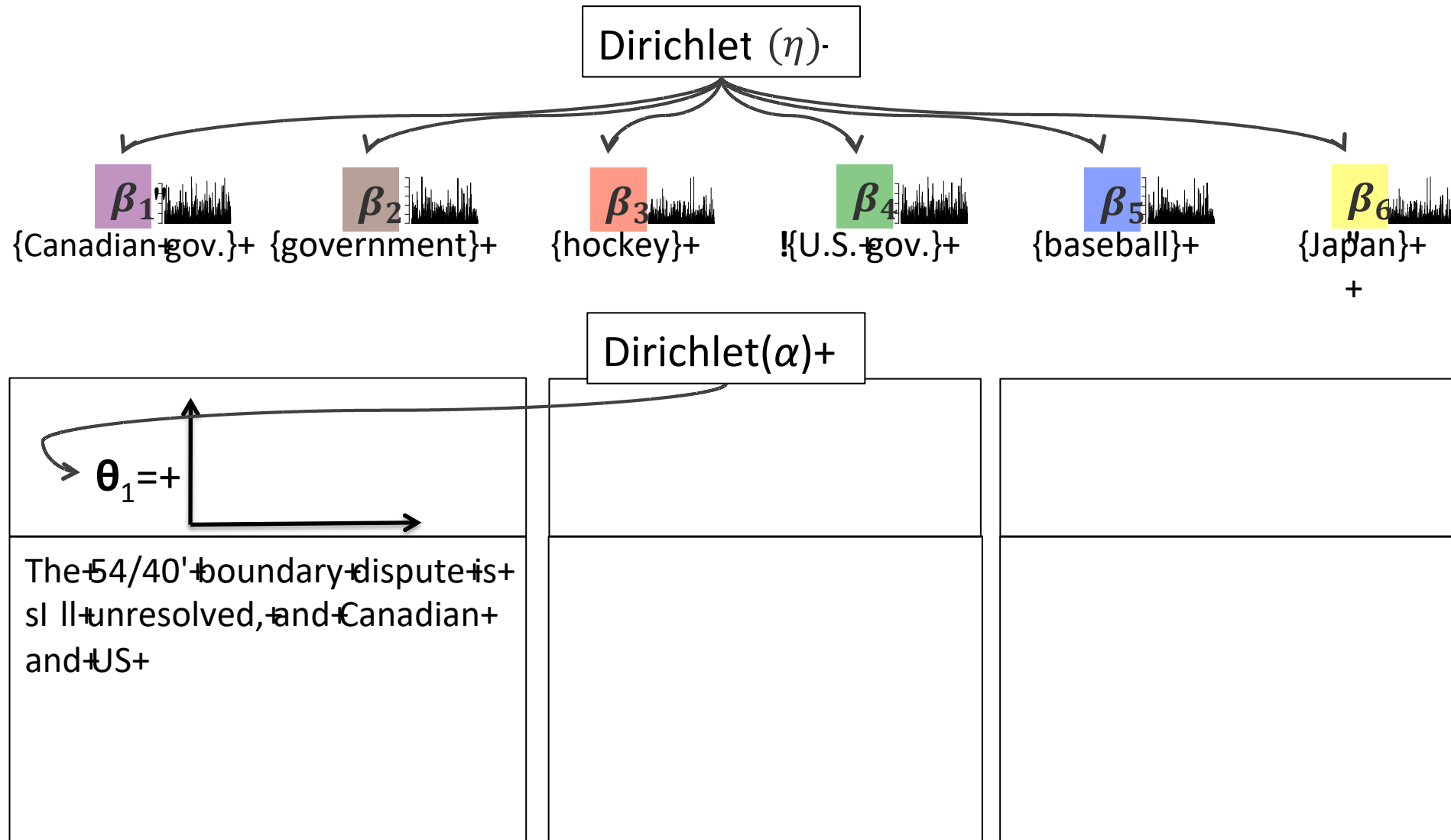
LDA'for'Topic'Modeling''



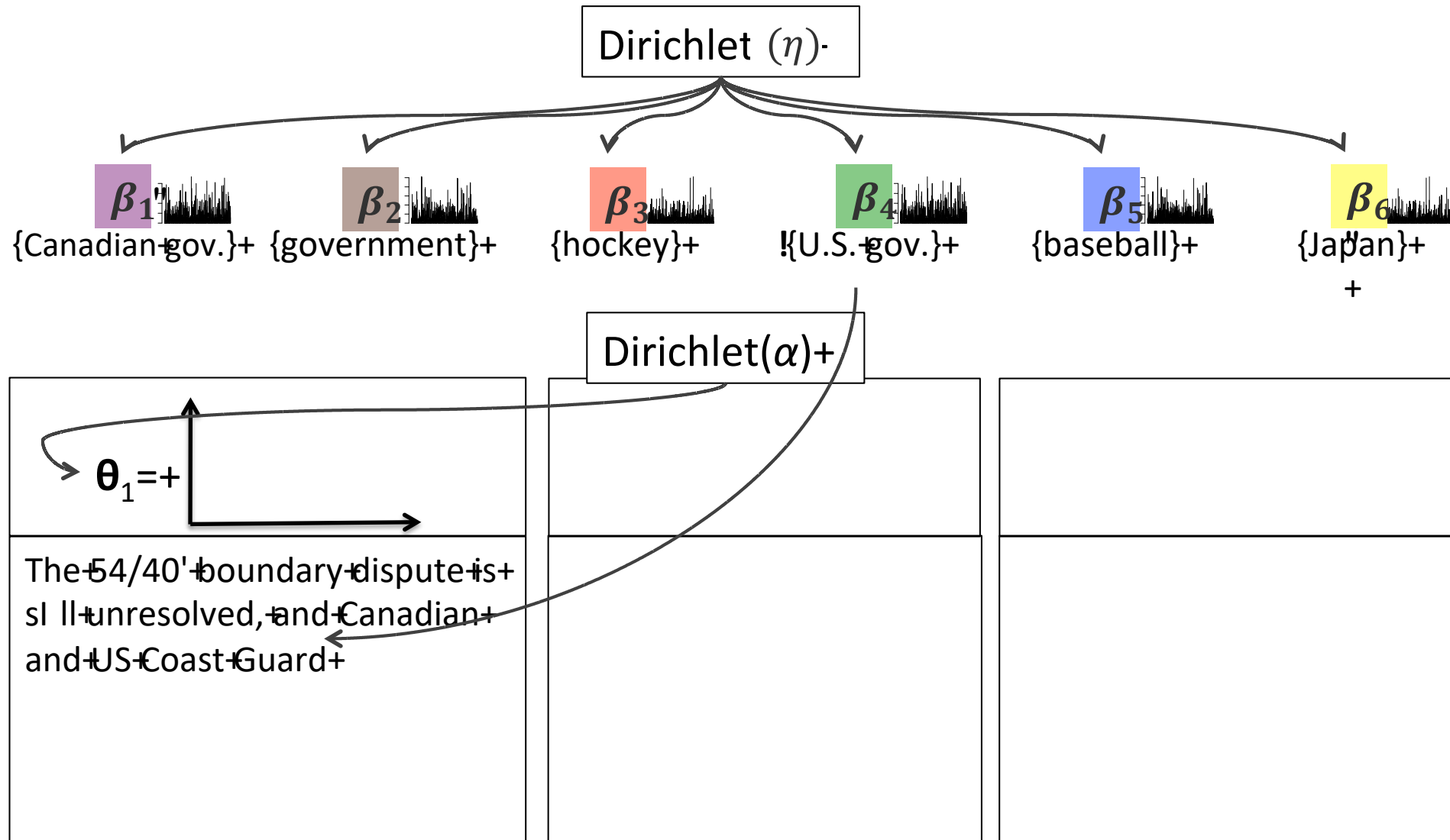
LDA'for'Topic'Modeling''



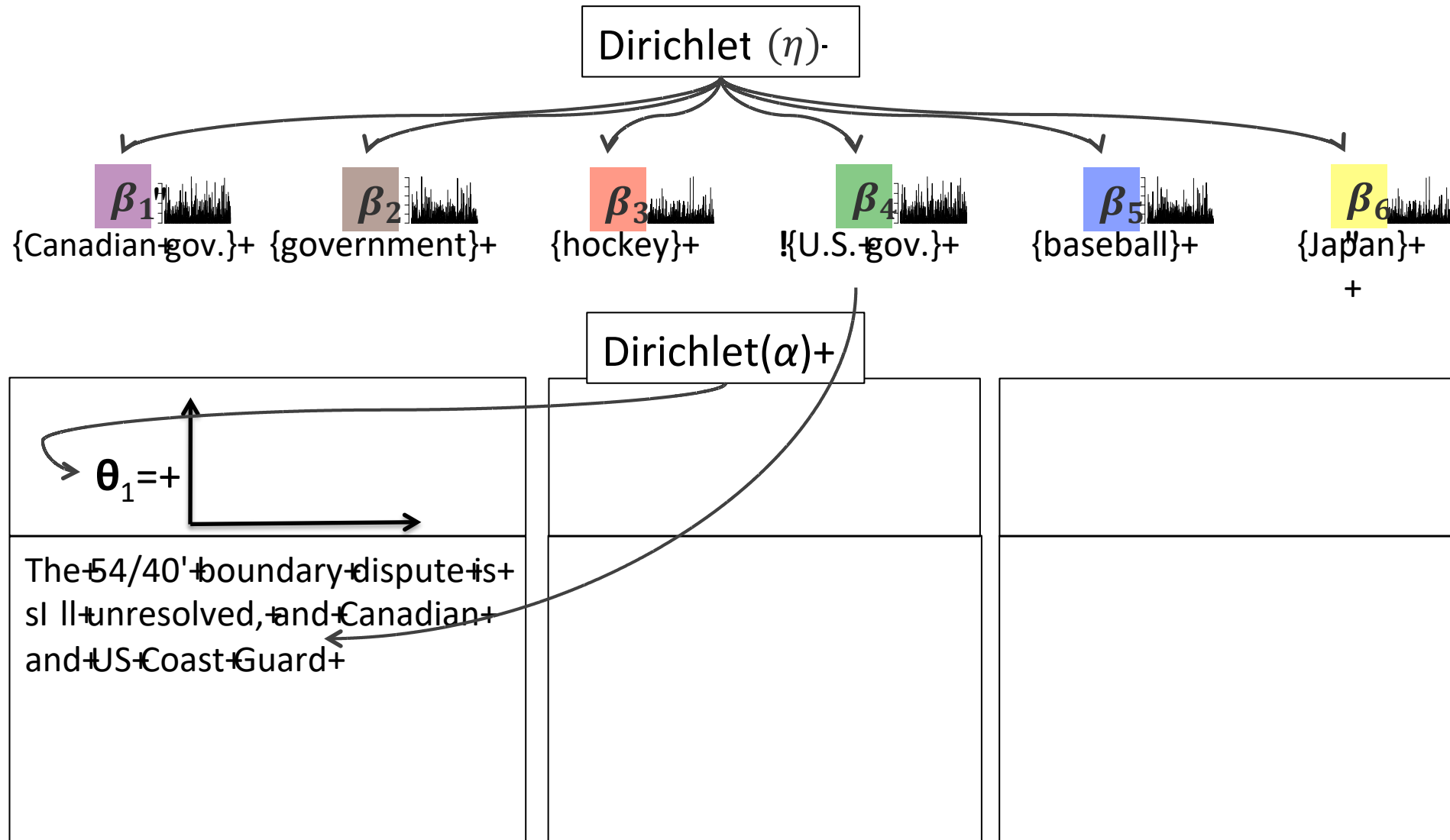
LDA'for'Topic'Modeling''



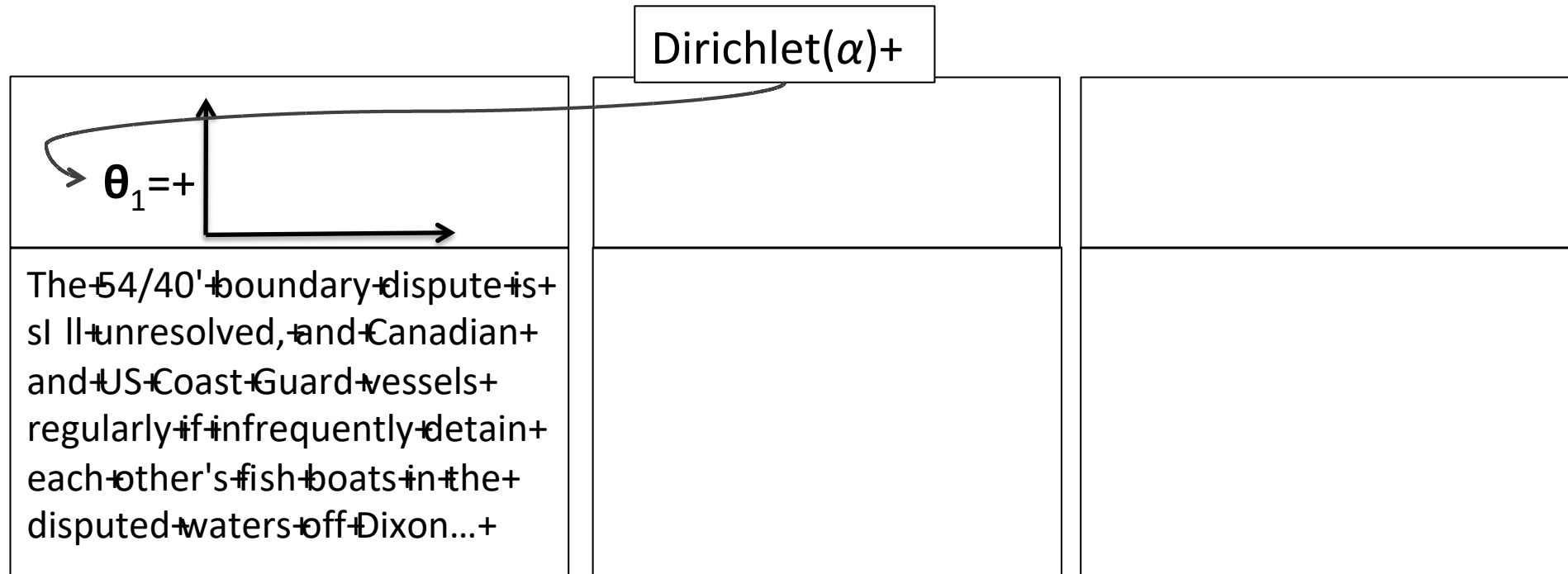
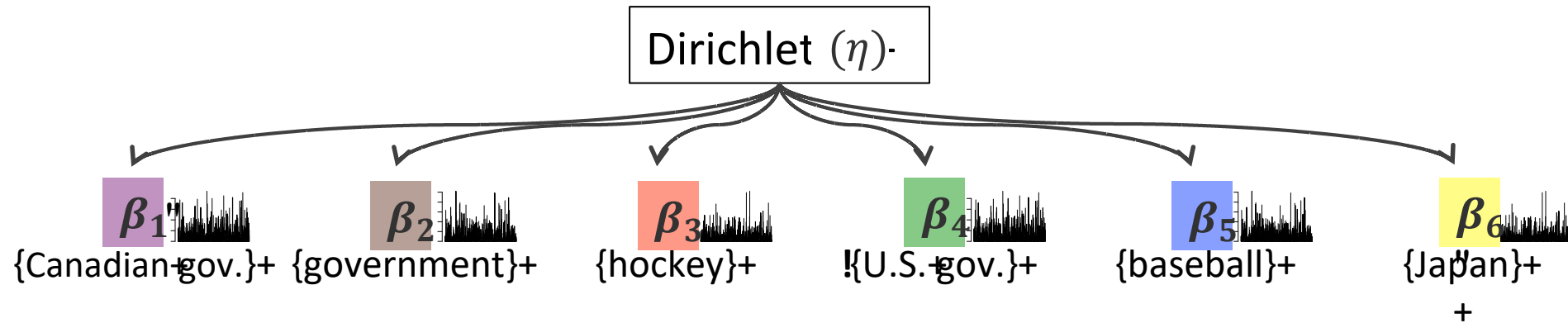
LDA'for'Topic'Modeling''



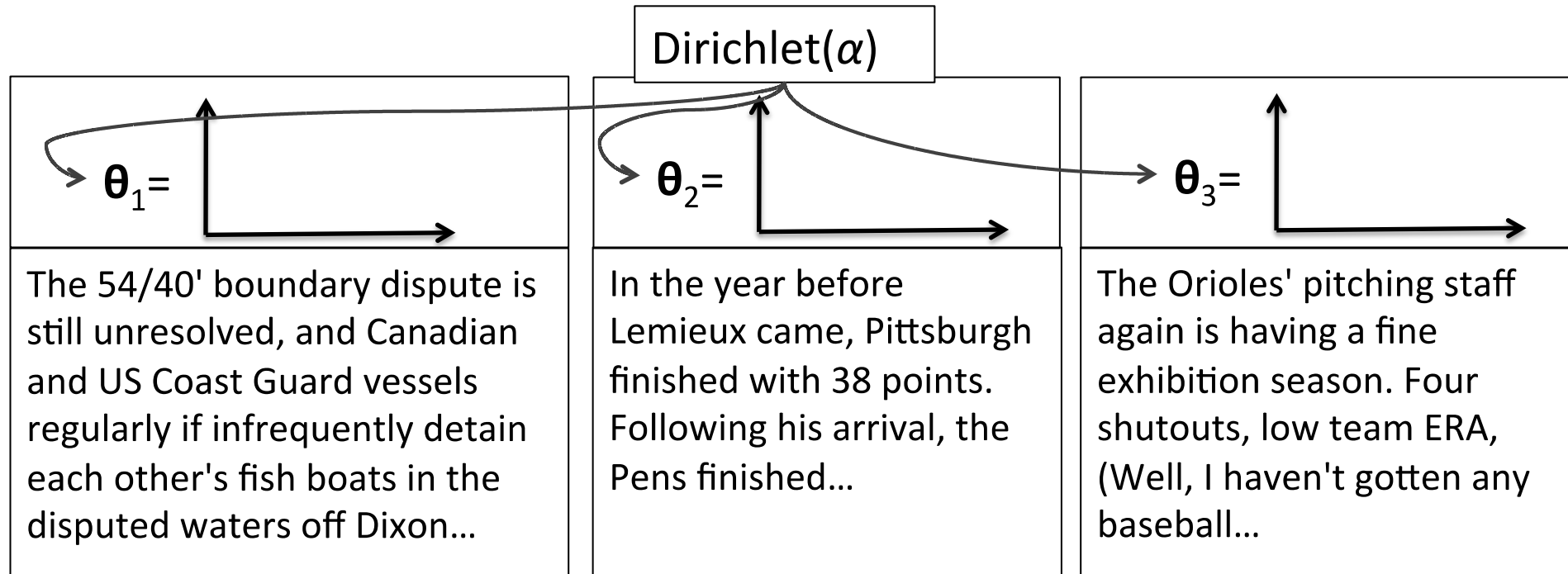
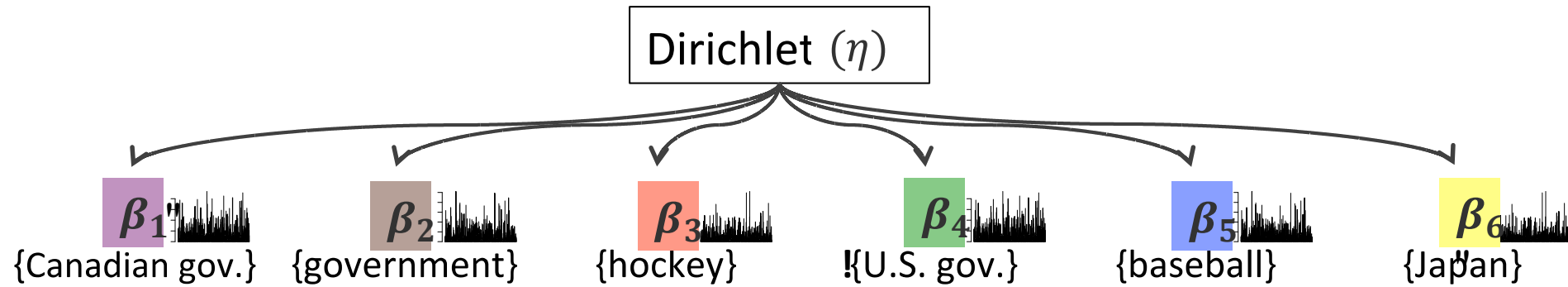
LDA'for'Topic'Modeling''



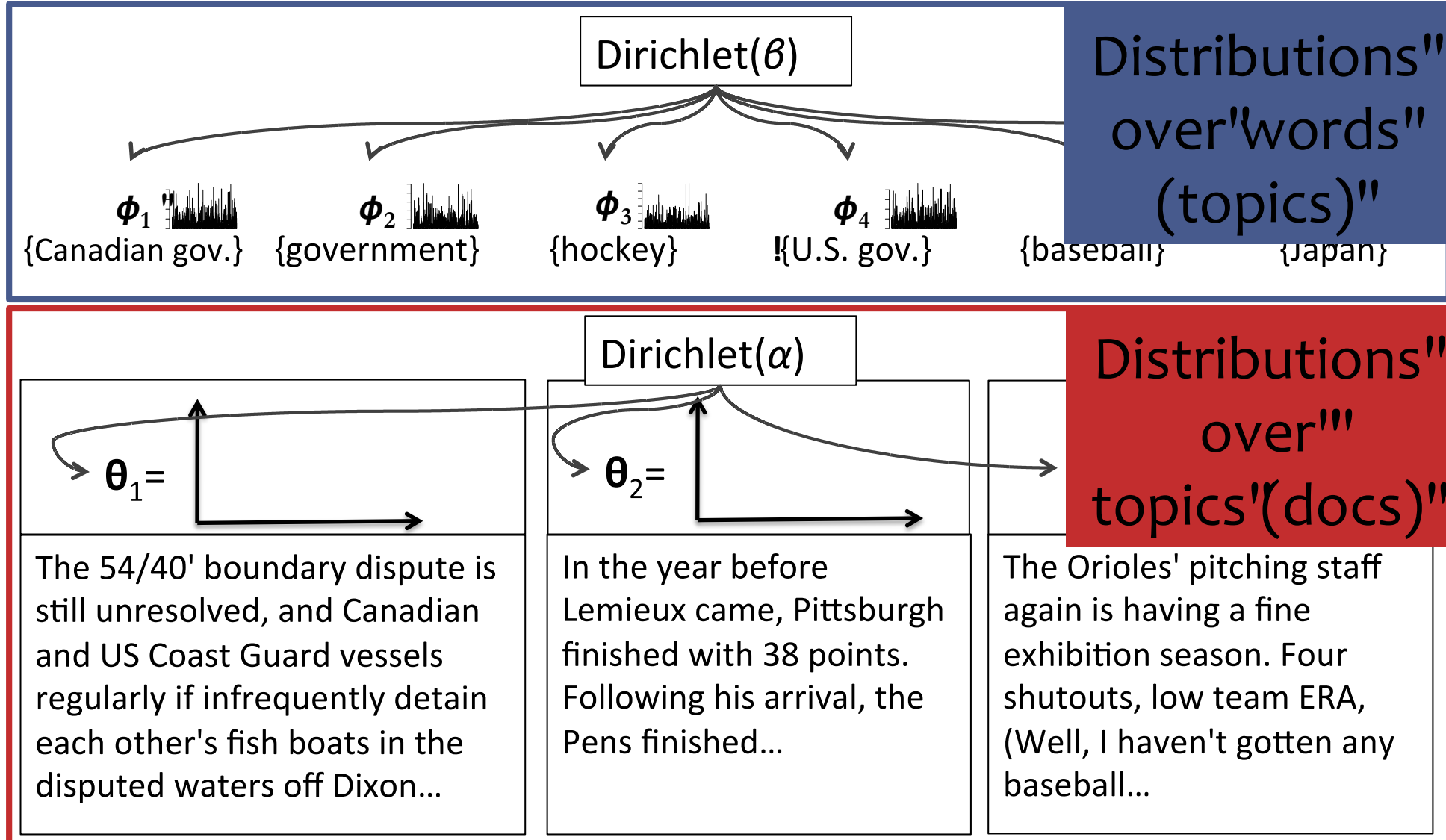
LDA 'for' 'Topic' Modeling''



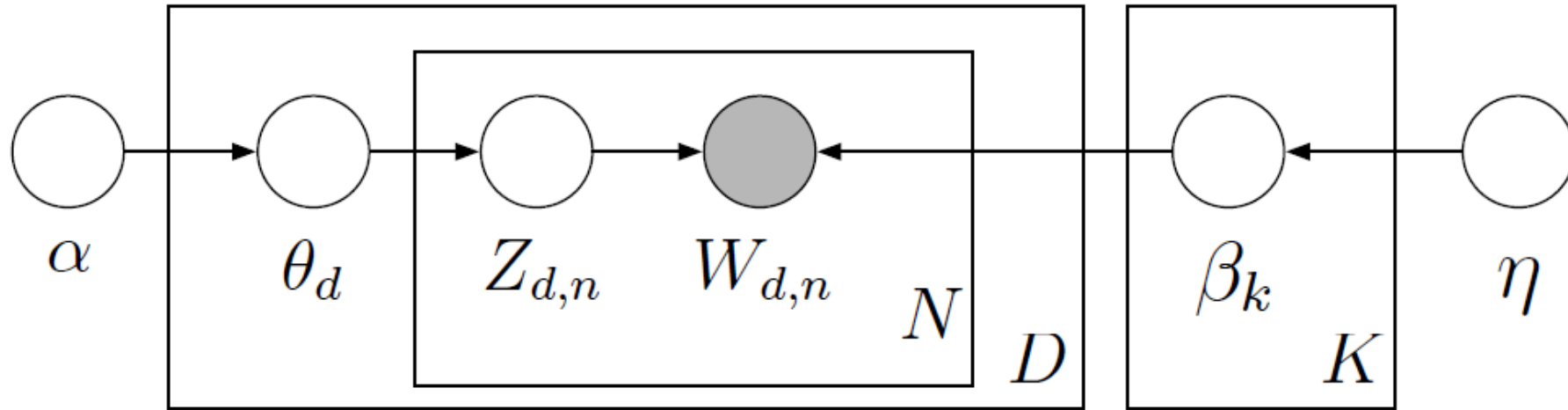
LDA for Topic Modeling



LDA'for'Topic'Modeling''



Joint Distribution for LDA



- Joint distribution of latent variables and documents is:

$$p(\boldsymbol{\beta}_{1:K}, \mathbf{z}_{1:D}, \boldsymbol{\theta}_{1:D}, \mathbf{w}_{1:D} | \alpha, \eta) = \prod_{i=1}^K p(\beta_i | \eta) \prod_{d=1}^D p(\theta_d | \alpha) \left(\prod_{n=1}^N p(z_{d,n} | \theta_d) p(w_{d,n} | \beta_{1:K}, z_{d,n}) \right)$$

Questions?