Spectral Methods for Correlated Topic Models

Forough Arabshahi* Animashree Anandkumar*

*University of California, Irvine

Topic Modeling

- Exchangeable topic models are probabilistic admixture models
  - Observations: Words in a document corpus
  - Assumption: Words are conditionally i.i.d given hidden topics
  - Goal: Recover a distribution over the words and topics

- Latent Dirichlet Allocation (LDA)
  - Topic prior distribution: Dirichlet
  - Limitations:
    - Only capable of modeling positive correlations
    - Elements with similar mean need to have similar variance

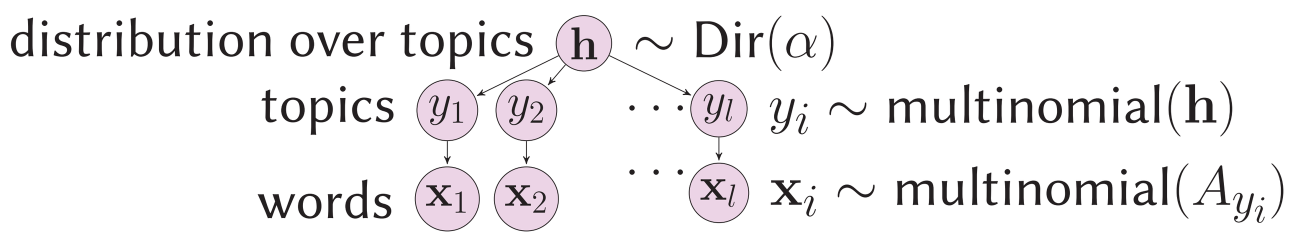


Figure: The exchangeable topic model.

- Overcoming the limitations of LDA
  - Latent Normalized Infinitely Divisible (NID) topic models
    - A generalization of the Dirichlet distribution
    - Capable of modeling arbitrary correlations
    - Do not require fixing a distribution over the topic space

Normalized Infinitely Divisible (NID) Distributions

- ID distributions
  - A class of positive or negative distributions
  - Representable as the sum of an arbitrary number of i.i.d rv's
  - No closed form pdf in general
  - Uniquely identified by their underlying Lévy measure

- NID distributions
  - A class of distributions on the simplex
  - Representable as normalized independent ID rv's
  - No closed form pdf in general
  - Uniquely identified by their underlying Lévy measure

- 3 examples of NID distributions on the 2-d simplex

  - Dirichlet:

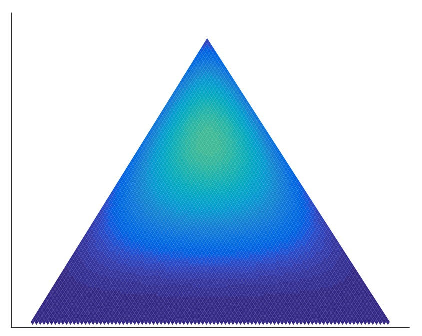


Figure: $\beta = 1$

  - γ-stable:

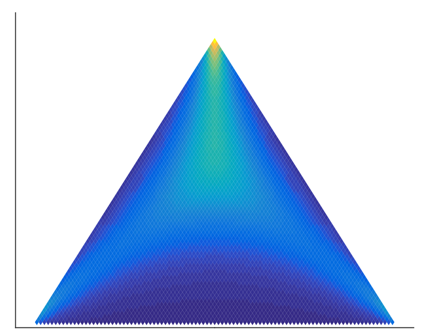


Figure: $\gamma = 0.75$

  - Inverse Gaussian:

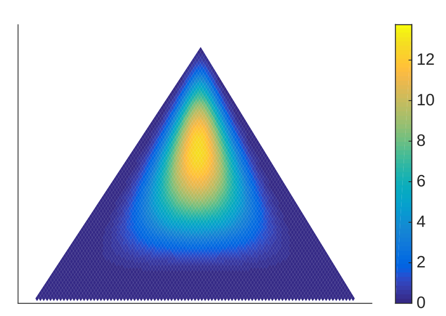
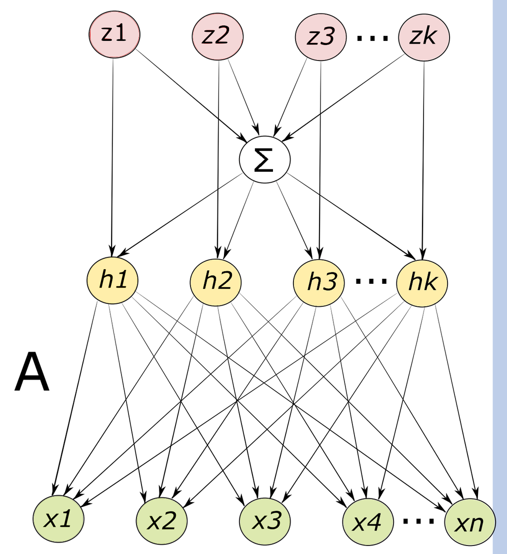


Figure: $\lambda = 0.01$

Latent NID Topic Models

- Generative model:

  For each document d of length n:
  - Draw z_i's independently from an ID distribution
  - Generate h_i's by normalization
  - For each word $i \in [0, n]$ in d:
    - Draw topic $y_i \sim \text{multinomial}(\mathbf{h})$
    - Draw word $x_i \sim \text{multinomial}(A_{y_i})$



- Main idea: sample generation from a k-d Dirichlet distribution
  - Generate k independent Gamma random variables
  - Normalize by their sum

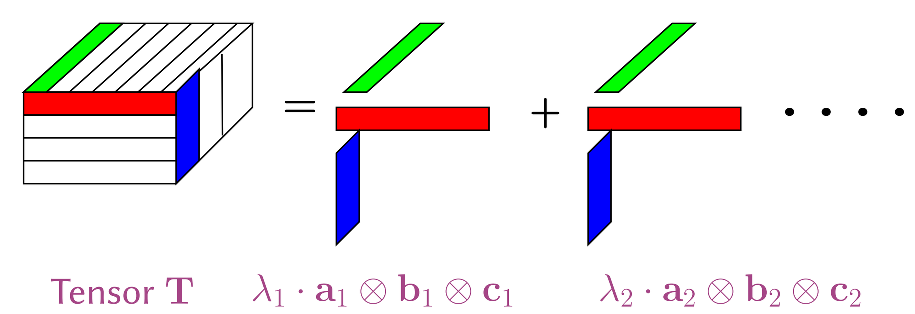
- Learning problem
  - Observation: Words
  - Goals:
    - Recover topic-word matrix $\mathbf{A}$
    - Recover the parameters of the topic distribution

The Learning Problem and Spectral Methods

- Unsupervised learning methods
  - Maximum likelihood estimation
    - Expectation Maximization (EM)
    - MCMC sampling
    - Variational methods
  - Matrix factorization
  - Spectral methods

- Spectral methods for latent variable models
  - Combination of low order moments that yield a CP decomposition

$$\mathbf{T} = \sum_{i \in [k]} \lambda_i \mathbf{a}_i \otimes \mathbf{b}_i \otimes \mathbf{c}_i,$$



Lemma: Decomposibility of Moments

- A linear combination of the low order moments has a CP decomposition
- The combination weights depend on the underlying NID distribution

$$\mathbf{T} = \mathbb{E}[\mathbf{x}_1 \otimes \mathbf{x}_2 \otimes \mathbf{x}_3]$$
$$+ v_1 (\mathbb{E}[\mathbf{x}_1 \otimes \mathbf{x}_2 \otimes \mathbb{E}[\mathbf{x}_3]] + \mathbb{E}[\mathbf{x}_1 \otimes \mathbb{E}[\mathbf{x}_2] \otimes \mathbf{x}_3] + \mathbb{E}[\mathbb{E}[\mathbf{x}_1] \otimes \mathbf{x}_2 \otimes \mathbf{x}_3])$$
$$+ v_2 \mathbb{E}[\mathbf{x}_1] \otimes \mathbb{E}[\mathbf{x}_2] \otimes \mathbb{E}[\mathbf{x}_3]$$

Learning Latent NID topic Models

- weights v_1 and v_2 from the Lemma can be computed efficiently
  - Numerical univariate integration
- Third order moments suffice for efficient learning
- T can be estimated in Polynomial time

Theorem: Learning Result

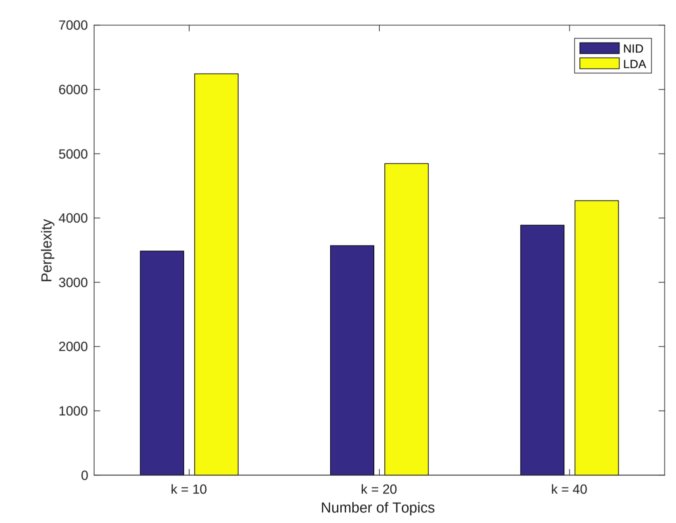
- Given linear independence of the columns of A, the topic-word matrix can be learned with polynomial sample complexity by decomposing the estimated tensor T of the Lemma into its rank-1 components.

Real Data experiments

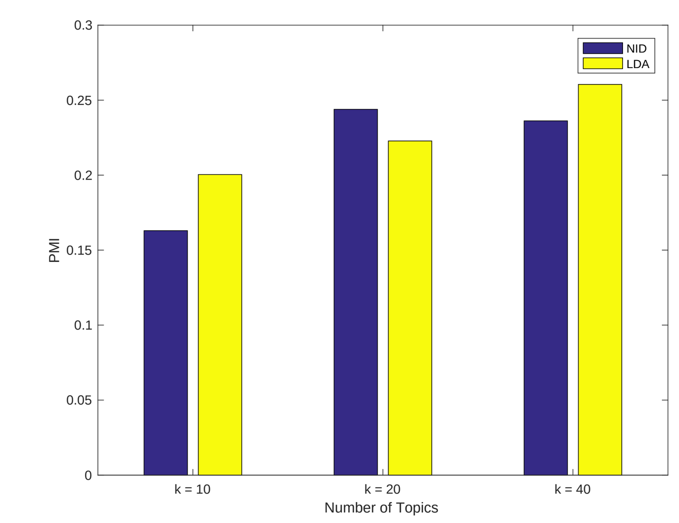
- Data sets:
  - New York Times:
    - $300,000$ documents
    - $102,660$ vocabulary size
  - Pubmed
    - 8.2 M documents
    - $141,044$ vocabulary size

- Evaluation Measures:
  - Likelihood Perplexity:
    - Evaluates generalization
    - The less the better
  - PMI:
    - Evaluates topic coherence
    - The more the better

- Perplexity and PMI for New York Times
  - Likelihood Perplexity



  - PMI



- Likelihood and PMI comparison across New York Times and Pubmed
  - Likelihood Perplexity

Dataset	NYtimes	Pubmed
NID	$\mathbf{3.5702e+03}$	$\mathbf{4.0771e+03}$
LDA	$4.8464e+03$	$4.3702e+03$

  - PMI

Dataset	NYtimes	Pubmed
NID	$\mathbf{0.2439}$	0.3080
LDA	0.2362	$\mathbf{0.4487}$

- Conclusions: Latent NID topic models
  - Can capture arbitrary correlations in the data
  - Can be learned in polynomial time with guarantees
  - Achieve better generalization and topic coherence

MEGA DatA Lab, EECS department, University of California, Irvine, USA   Email: farabsha@uci.edu   WWW: http://newport.eecs.uci.edu/anandkumar/#lab