

Variational Inference

Material adapted from David Blei University of Maryland EXERCISE

Feedback

- Disagreement from you: flipped classroom (but most like), theory
- I agree with you: one size fits all, inconsistent notation, reading inconsistent, questions in random order
- Sorry: Slides w/ video, hard to hear questions, too much NLP
- Huh?: Online judge, ask questions on same interface, A cutoff, video speed
- Thinking about: second midterm

Announcements

Next class: unflipped VAE / GAN

Reading: Goodfellow

Next week: LSTM / RNN

Rest of course remains unchanged

Default project: more data, Kaggle site this week

Useful Reference: $\Psi(x)$

X	$\Psi(x)$
1	-0.577215664902
2	0.422784335098
3	0.922784335098
4	1.25611766843
5	1.50611766843
6	1.70611766843
7	1.8727843351
8	2.01564147796
9	2.14064147796

Also: from scipy.special import digamma

Example

Three topics, same documents as last time

$$\beta = \begin{bmatrix} \text{cat} & \text{dog} & \text{hamburger} & \text{iron} & \text{pig} \\ .26 & .185 & .185 & .185 & .185 \\ .185 & .185 & .26 & .185 & .185 \\ .185 & .185 & .185 & .26 & .185 \end{bmatrix}$$
(1)

- Assume uniform γ: (2.0, 2.0, 2.0)
- Compute update for ϕ

$$\phi_{ni} \propto \beta_{iv} \exp\left(\Psi(\gamma_i) - \Psi\left(\sum_j \gamma_j\right)\right)$$
 (2)

• For a the first word (dog) in the document: dog cat cat pig

Example

Three topics, same documents as last time

$$\beta = \begin{bmatrix} \text{cat} & \text{dog} & \text{hamburger} & \text{iron} & \text{pig} \\ .26 & .185 & .185 & .185 & .185 \\ .185 & .185 & .26 & .185 & .185 \\ .185 & .185 & .185 & .26 & .185 \end{bmatrix}$$
(1)

- Assume uniform γ: (2.0, 2.0, 2.0)
- Compute update for ϕ

$$\phi_{ni} \propto \beta_{iv} \exp\left(\Psi(\gamma_i) - \Psi\left(\sum_j \gamma_j\right)\right)$$
 (2)

For a the first word (dog) in the document: dog cat cat pig

$$\beta = \left[\begin{array}{ccccc} \text{cat} & \text{dog} & \text{hamburger} & \text{iron} & \text{pig} \\ .26 & .185 & .185 & .185 & .185 \\ .185 & .185 & .26 & .185 & .185 \\ .185 & .185 & .185 & .26 & .185 \end{array} \right] \qquad \qquad \begin{matrix} \phi_{\textit{ni}} \propto \\ \beta_{\textit{lv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \beta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \beta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \beta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \beta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \beta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right)$$

$$\gamma = (2.000, 2.000, 2.000)$$

$$\beta = \left[\begin{array}{ccccc} \text{cat} & \text{dog} & \text{hamburger} & \text{iron} & \text{pig} \\ .26 & .185 & .185 & .185 & .185 \\ .185 & .185 & .26 & .185 & .185 \\ .185 & .185 & .185 & .26 & .185 \end{array} \right] \qquad \qquad \begin{matrix} \phi_{\textit{ni}} \propto \\ \beta_{\textit{lv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \beta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \beta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \beta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \beta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \beta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right)$$

- $\gamma = (2.000, 2.000, 2.000)$
- $\phi(0) \propto 0.185 \times \exp(\Psi(2.000) \Psi(2.000 + 2.000 + 2.000)) = 0.051$

$$\beta = \left[\begin{array}{ccccc} \text{cat} & \text{dog} & \text{hamburger} & \text{iron} & \text{pig} \\ .26 & .185 & .185 & .185 & .185 \\ .185 & .185 & .26 & .185 & .185 \\ .185 & .185 & .185 & .26 & .185 \end{array} \right] \qquad \qquad \begin{matrix} \phi_{\textit{ni}} \propto \\ \beta_{\textit{lv}} \exp \left(\Psi(\gamma_{\textit{i}}) - \Psi\left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \end{matrix}$$

- $\gamma = (2.000, 2.000, 2.000)$
- $\phi(0) \propto 0.185 \times \exp(\Psi(2.000) \Psi(2.000 + 2.000 + 2.000)) = 0.051$
- $\phi(1) \propto 0.185 \times \exp(\Psi(2.000) \Psi(2.000 + 2.000 + 2.000)) = 0.051$

$$\beta = \left[\begin{array}{ccccc} \text{cat} & \text{dog} & \text{hamburger} & \text{iron} & \text{pig} \\ .26 & .185 & .185 & .185 & .185 \\ .185 & .185 & .26 & .185 & .185 \\ .185 & .185 & .185 & .26 & .185 \end{array} \right] \qquad \qquad \phi_{ni} \propto \\ \beta_{ni} \exp \left(\Psi(\gamma_i) - \Psi\left(\sum_j \gamma_j \right) \right)$$

- $\gamma = (2.000, 2.000, 2.000)$
- $\phi(0) \propto 0.185 \times \exp(\Psi(2.000) \Psi(2.000 + 2.000 + 2.000)) = 0.051$
- $\phi(1) \propto 0.185 \times \exp(\Psi(2.000) \Psi(2.000 + 2.000 + 2.000)) = 0.051$
- $\phi(2) \propto 0.185 \times \exp(\Psi(2.000) \Psi(2.000 + 2.000 + 2.000)) = 0.051$

$$\beta = \left[\begin{array}{ccccc} \text{cat} & \text{dog} & \text{hamburger} & \text{iron} & \text{pig} \\ .26 & .185 & .185 & .185 & .185 \\ .185 & .185 & .26 & .185 & .185 \\ .185 & .185 & .185 & .26 & .185 \end{array} \right] \qquad \qquad \phi_{ni} \propto \\ \beta_{nv} \exp \left(\Psi(\gamma_i) - \Psi\left(\sum_j \gamma_j \right) \right)$$

- $\gamma = (2.000, 2.000, 2.000)$
- $\phi(0) \propto 0.185 \times \exp(\Psi(2.000) \Psi(2.000 + 2.000 + 2.000)) = 0.051$
- $\phi(1) \propto 0.185 \times \exp(\Psi(2.000) \Psi(2.000 + 2.000 + 2.000)) = 0.051$
- $\phi(2) \propto 0.185 \times \exp(\Psi(2.000) \Psi(2.000 + 2.000 + 2.000)) = 0.051$
- After normalization: {0.333, 0.333, 0.333}

$$\beta = \left[\begin{array}{ccccc} \text{cat} & \text{dog} & \text{hamburger} & \text{iron} & \text{pig} \\ .26 & .185 & .185 & .185 & .185 \\ .185 & .185 & .26 & .185 & .185 \\ .185 & .185 & .185 & .26 & .185 \end{array} \right] \qquad \qquad \begin{matrix} \phi_{\textit{ni}} \propto \\ \beta_{\textit{lv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \beta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \beta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \beta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \beta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \beta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right)$$

 $\gamma = (2.000, 2.000, 2.000)$

$$\beta = \left[\begin{array}{ccccc} \text{cat} & \text{dog} & \text{hamburger} & \text{iron} & \text{pig} \\ .26 & .185 & .185 & .185 & .185 \\ .185 & .185 & .26 & .185 & .185 \\ .185 & .185 & .185 & .26 & .185 \end{array} \right] \qquad \qquad \begin{matrix} \phi_{\textit{ni}} \propto \\ \beta_{\textit{lv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \beta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \beta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \beta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \beta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \beta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right)$$

- $\gamma = (2.000, 2.000, 2.000)$
- $\phi(0) \propto 0.185 \times \exp(\Psi(2.000) \Psi(2.000 + 2.000 + 2.000)) = 0.051$

Material adapted from David Blei UMD

$$\beta = \left[\begin{array}{ccccc} \text{cat} & \text{dog} & \text{hamburger} & \text{iron} & \text{pig} \\ .26 & .185 & .185 & .185 & .185 \\ .185 & .185 & .26 & .185 & .185 \\ .185 & .185 & .185 & .26 & .185 \end{array} \right] \qquad \qquad \begin{matrix} \phi_{\textit{ni}} \propto \\ \beta_{\textit{lv}} \exp \left(\Psi(\gamma_{\textit{i}}) - \Psi\left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \end{matrix}$$

- $\gamma = (2.000, 2.000, 2.000)$
- $\phi(0) \propto 0.185 \times \exp(\Psi(2.000) \Psi(2.000 + 2.000 + 2.000)) = 0.051$
- $\phi(1) \propto 0.185 \times \exp(\Psi(2.000) \Psi(2.000 + 2.000 + 2.000)) = 0.051$

Material adapted from David Blei | UMD

$$\beta = \begin{bmatrix} \text{cat} & \text{dog} & \text{hamburger} & \text{iron} & \text{pig} \\ .26 & .185 & .185 & .185 & .185 \\ .185 & .185 & .26 & .185 & .185 \\ .185 & .185 & .185 & .26 & .185 \end{bmatrix}^{\phi_{ni} \propto} \beta_{n_r} \exp\left(\Psi(\gamma_i) - \Psi\left(\sum_j \gamma_j\right)\right)$$

- $\gamma = (2.000, 2.000, 2.000)$
- $\phi(0) \propto 0.185 \times \exp(\Psi(2.000) \Psi(2.000 + 2.000 + 2.000)) = 0.051$
- $\phi(1) \propto 0.185 \times \exp(\Psi(2.000) \Psi(2.000 + 2.000 + 2.000)) = 0.051$
- $\phi(2) \propto 0.185 \times \exp(\Psi(2.000) \Psi(2.000 + 2.000 + 2.000)) = 0.051$

$$\beta = \left[\begin{array}{ccccc} \text{cat} & \text{dog} & \text{hamburger} & \text{iron} & \text{pig} \\ .26 & .185 & .185 & .185 & .185 \\ .185 & .185 & .26 & .185 & .185 \\ .185 & .185 & .185 & .26 & .185 \end{array} \right] \qquad \qquad \phi_{ni} \propto \\ \beta_{nv} \exp \left(\Psi(\gamma_i) - \Psi\left(\sum_j \gamma_j \right) \right)$$

- $\gamma = (2.000, 2.000, 2.000)$
- $\phi(0) \propto 0.185 \times \exp(\Psi(2.000) \Psi(2.000 + 2.000 + 2.000)) = 0.051$
- $\phi(1) \propto 0.185 \times \exp(\Psi(2.000) \Psi(2.000 + 2.000 + 2.000)) = 0.051$
- $\phi(2) \propto 0.185 \times \exp(\Psi(2.000) \Psi(2.000 + 2.000 + 2.000)) = 0.051$
- After normalization: {0.333, 0.333, 0.333}

$$\beta = \left[\begin{array}{ccccc} \text{cat} & \text{dog} & \text{hamburger} & \text{iron} & \text{pig} \\ .26 & .185 & .185 & .185 & .185 \\ .185 & .185 & .26 & .185 & .185 \\ .185 & .185 & .185 & .26 & .185 \end{array} \right] \qquad \qquad \begin{matrix} \phi_{\textit{ni}} \propto \\ \beta_{\textit{lv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \beta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \beta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \beta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \beta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \beta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right)$$

$$\gamma = (2.000, 2.000, 2.000)$$

$$\beta = \left[\begin{array}{ccccc} \text{cat} & \text{dog} & \text{hamburger} & \text{iron} & \text{pig} \\ .26 & .185 & .185 & .185 & .185 \\ .185 & .185 & .26 & .185 & .185 \\ .185 & .185 & .185 & .26 & .185 \end{array} \right] \qquad \qquad \begin{matrix} \phi_{\textit{ni}} \propto \\ \beta_{\textit{lv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \beta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \beta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \beta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \beta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \beta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \\ \theta_{\textit{nv}} \exp \left(\Psi(\gamma_{\textit{l}}) - \Psi \left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right)$$

- $\gamma = (2.000, 2.000, 2.000)$
- $\phi(0) \propto 0.260 \times \exp(\Psi(2.000) \Psi(2.000 + 2.000 + 2.000)) = 0.072$

$$\beta = \left[\begin{array}{ccccc} \text{cat} & \text{dog} & \text{hamburger} & \text{iron} & \text{pig} \\ .26 & .185 & .185 & .185 & .185 \\ .185 & .185 & .26 & .185 & .185 \\ .185 & .185 & .185 & .26 & .185 \end{array} \right] \qquad \qquad \begin{matrix} \phi_{\textit{ni}} \propto \\ \beta_{\textit{lv}} \exp \left(\Psi(\gamma_{\textit{i}}) - \Psi\left(\sum_{\textit{j}} \gamma_{\textit{j}} \right) \right) \end{matrix}$$

- $\gamma = (2.000, 2.000, 2.000)$
- $\phi(0) \propto 0.260 \times \exp(\Psi(2.000) \Psi(2.000 + 2.000 + 2.000)) = 0.072$
- $\phi(1) \propto 0.185 \times \exp(\Psi(2.000) \Psi(2.000 + 2.000 + 2.000)) = 0.051$

$$\beta = \left[\begin{array}{ccccc} \text{cat} & \text{dog} & \text{hamburger} & \text{iron} & \text{pig} \\ .26 & .185 & .185 & .185 & .185 \\ .185 & .185 & .26 & .185 & .185 \\ .185 & .185 & .185 & .26 & .185 \end{array} \right] \qquad \qquad \phi_{ni} \propto \\ \beta_{nv} \exp \left(\Psi(\gamma_i) - \Psi\left(\sum_j \gamma_j \right) \right)$$

- $\gamma = (2.000, 2.000, 2.000)$
- $\phi(0) \propto 0.260 \times \exp(\Psi(2.000) \Psi(2.000 + 2.000 + 2.000)) = 0.072$
- $\phi(1) \propto 0.185 \times \exp(\Psi(2.000) \Psi(2.000 + 2.000 + 2.000)) = 0.051$
- $\phi(2) \propto 0.185 \times \exp(\Psi(2.000) \Psi(2.000 + 2.000 + 2.000)) = 0.051$

Material adapted from David Blei

$$\beta = \left[\begin{array}{ccccc} \text{cat} & \text{dog} & \text{hamburger} & \text{iron} & \text{pig} \\ .26 & .185 & .185 & .185 & .185 \\ .185 & .185 & .26 & .185 & .185 \\ .185 & .185 & .185 & .26 & .185 \end{array} \right] \qquad \qquad \phi_{ni} \propto \\ \beta_{nv} \exp \left(\Psi(\gamma_i) - \Psi\left(\sum_j \gamma_j \right) \right)$$

- $\gamma = (2.000, 2.000, 2.000)$
- $\phi(0) \propto 0.260 \times \exp(\Psi(2.000) \Psi(2.000 + 2.000 + 2.000)) = 0.072$
- $\phi(1) \propto 0.185 \times \exp(\Psi(2.000) \Psi(2.000 + 2.000 + 2.000)) = 0.051$
- $\phi(2) \propto 0.185 \times \exp(\Psi(2.000) \Psi(2.000 + 2.000 + 2.000)) = 0.051$
- After normalization: {0.413, 0.294, 0.294}

Update γ

- Document: dog cat cat pig
- Update equation

$$\gamma_i = \alpha_i + \sum_n \phi_{ni} \tag{3}$$

• Assume $\alpha = (.1, .1, .1)$

Update γ

- Document: dog cat cat pig
- Update equation

$$\gamma_i = \alpha_i + \sum_n \phi_{ni} \tag{3}$$

• Assume $\alpha = (.1, .1, .1)$

	ϕ_{0}	$\phi_{ extsf{1}}$	$\phi_{ t 2}$
dog	.333	.333	.333
cat	.413	.294	.294
pig	.333	.333	.333
α	0.1	0.1	0.1
sum	1.592	1.354	1.354

Note: do not normalize!

Update γ

- Document: dog cat cat pig
- Update equation

$$\gamma_i = \alpha_i + \sum_n \phi_{ni} \tag{3}$$

• Assume $\alpha = (.1, .1, .1)$

	$oldsymbol{\phi}_{0}$	$oldsymbol{\phi}_{1}$	$\phi_{ exttt{2}}$	
dog	.333	.333	.333	
cat	.413	.294	.294	x2
pig	.333	.333	.333	
α	0.1	0.1	0.1	
sum	1.592	1.354	1.354	

Note: do not normalize!

Update β

- lacksquare Count up all of the ϕ across all documents
- For each topic, divide by total
- Corresponds to maximum likelihood of expected counts

Update β

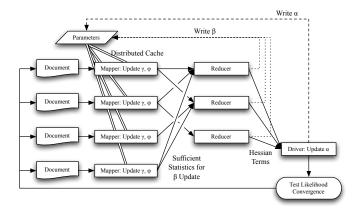
- Count up all of the ϕ across all documents
- For each topic, divide by total
- Corresponds to maximum likelihood of expected counts
- Unlike Gibbs sampling, no Dirichlet prior

Automatic Inference



Parallel LDA

Zhai et al, 2012



Online LDA

Hoffman and Blei, 2010

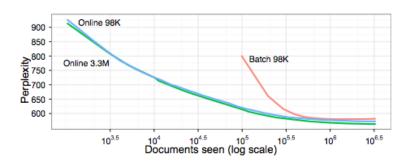
Algorithm 2 Online variational Bayes for LDA

```
Define \rho_t \triangleq (\tau_0 + t)^{-\kappa}
Initialize \lambda randomly.
for t = 0 to \infty do
    E step:
    Initialize \gamma_{tk} = 1. (The constant 1 is arbitrary.)
    repeat
        Set \phi_{twk} \propto \exp\{\mathbb{E}_{\sigma}[\log \theta_{tk}] + \mathbb{E}_{\sigma}[\log \beta_{kw}]\}
        Set \gamma_{tk} = \alpha + \sum_{w} \phi_{twk} n_{tw}
    until \frac{1}{k'}\sum_{k} |\text{change in}\gamma_{tk}| < 0.00001
    M step:
    Compute \bar{\lambda}_{kw} = \eta + D n_{tw} \phi_{twk}
    Set \lambda = (1 - \rho_t)\lambda + \rho_t \tilde{\lambda}.
end for
```

Material adapted from David Blei | UMD

Online LDA

Hoffman and Blei, 2010



Online I DA

Hoffman and Blei, 2010

- Initialize λ⁽⁰⁾ randomly.
- 2: Set the step-size schedule ρ_t appropriately.
- 3: repeat
- Sample a data point x_i uniformly from the data set. 4:
- Compute its local variational parameter, 5:

$$\phi = \mathbb{E}_{\lambda^{(t-1)}}[\eta_g(x_i^{(N)}, z_i^{(N)})].$$

Compute intermediate global parameters as though x_i is replicated N times, 6:

$$\hat{\lambda} = \mathbb{E}_{\phi}[\eta_g(x_i^{(N)}, z_i^{(N)})].$$

7: Update the current estimate of the global variational parameters,

$$\lambda^{(t)} = (1 - \rho_t)\lambda^{(t-1)} + \rho_t \hat{\lambda}.$$

8: until forever

Material adapted from David Blei | UMD

Best of Both Worlds

Algorithm 1 Algorithm for hybrid stochastic variational-Gibbs inference.

```
for t \in 1, ..., \infty do
   \rho_t \leftarrow \left(\frac{1}{t_0 + t}\right)^{\kappa}
   sample minibatch \mathcal{B}
   for d \in \mathcal{B} do
        initialize z_d^0
        discard B burn-in sweeps
        for sample s \in 1, ..., S do
            for token i \in 1, ..., N_d do
                sample z_{di}^s \propto (\alpha + N_{dk}) e^{\mathbb{E}_q[\log \beta_{kw}]}
            end for
        end for
   end for
   \lambda_{kw}^{t} \leftarrow (1 - \rho_t)\lambda_{kw}^{t-1} + \rho_t \left(\eta + \frac{D}{|\mathcal{B}|}\hat{N}_{kw}\right)
end for
```

Zhai and Boyd-Graber, 2013



Zhai and Boyd-Graber, 2013

minibatch-5	minibatch-8	minibatch-10			minibatch-120
102-club	118-club	132-rock			0-appear
115-issuee	128-copy	194-issue			1-hulk
127-cover	137-cover	215-series			2-wolverin
130-copy	138-issue	217-copy			3-annual
197-appear	180-appear	226-cover			4-copy 5-rider
	319-rock				6-comicstrip
289-rock	493-annual	261-appear			7-cover
450-annual		588-annual			8-force
584-series	639-series	949-force			9-captain
811-forcee	877-force	1074-rider			11-issue
1090-rider	1003-rider	6038-comicstrip			12-series
	7075-captain	6520-mutant			16-mutant
F		9569-captain			
		oooo oaptam			41-rock
5	8	10			
captain	comicstrip	hulk			
seqitur	mutant	mazelyah			
	patlafontain				

Zhai and Boyd-Graber, 2013

minibatch-5	minibatch-8	minibatch-10	minibatch-16		
102-club	118-club	132-rock	87-series		
115-issuee	128-copy	194-issue	161-issue		
127-cover	137-cover	215-series	283-сору		
130-copy	138-issue	217-copy	306-appear		
197-appear	180-appear	226-cover	307-cover		
289-rock	319-rock	261-appear	502-annual		
450-annual	493-annual	588-annual	814-force		7-cover
584-series	639-series	949-force	1194-rider		8-force
811-forcee	877-force	1074-rider	8944-hulk		9-captain
1090-rider	1003-rider		10819-comicstrip		11-issue
1090-rider		6038-comicstrip	11301-mutant		12-series
	7075-captain	6520-mutant			16-mutant
		9569-captain	14335-captain		41-rock
5	8	10	16		
captain	comicstrip	hulk	wolverin		
seqitur	mutant	mazelyah	albion		
	patlafontain				

Zhai and Boyd-Graber, 2013

minibatch-5	minibatch-8	minibatch-10	minibatch-16	minibatch-17		minibatch-120
102-club	118-club	132-rock	87-series	82-series		0-appear
115-issuee	128-copy	194-issue	161-issue	162-issue		1-hulk
127-cover	137-cover	215-series	283-copy	288-copy		2-wolverin 3-annual
130-copy	138-issue	217-copy	306-appear	294-appear		4-copy
197-appear	180-appear	226-cover	307-cover	311-cover		5-rider
289-rock	319-rock	261-appear	502-annual	512-annual		6-comicstrip
450-annual	493-annual	588-annual	814-force	830-force		7-cover
584-series	639-series	949-force	1194-rider	4782-wolverin		8-force
811-forcee	877-force	1074-rider	8944-hulk	9659-hulk		9-captain
1090-rider	1003-rider	6038-comicstrip	10819-comicstrip	11527-comicstrip		11-issue
1090-11dei	7075-captain		11301-mutant	12009-mutant		12-series
	7075-Capiain	6520-mutant	14335-captain			16-mutant
		9569-captain	14335-captain	15040-captain		41-rock
5	8	10	16	17		
captain	comicstrip	hulk	wolverin	lacy		
seqitur	mutant	mazelyah	albion			
	patlafontain					

Zhai and Boyd-Graber, 2013

minibatch-5	minibatch-8	minibatch-10	minibatch-16	minibatch-17	minibatch-39	minibatch-83	minibatch-120
102-club	118-club	132-rock	87-series	82-series	1-annual		
115-issuee	128-copy	194-issue	161-issue	162-issue	2-rock		
127-cover	137-cover	215-series	283-copy	288-copy	3-wolverin		
130-copy	138-issue	217-copy	306-appear	294-appear	4-appear		
197-appear	180-appear	226-cover	307-cover	311-cover	5-comicstrip 6-series		
289-rock	319-rock	261-appear	502-annual	512-annual	7-mutant		
450-annual	493-annual	588-annual	814-force	830-force	8-cover		
584-series	639-series	949-force	1194-rider	4782-wolverin	12-issue		
	877-force		8944-hulk	9659-hulk	14-hulk		
811-forcee	1003-rider	1074-rider	10819-comicstrip		16-copy		
1090-rider		6038-comicstrip		11527-comicstrip	53-force		
	7075-captain	6520-mutant	11301-mutant	12009-mutant	57-rider		
		9569-captain	14335-captain	15040-captain			
					86-captain		
5	8	10	16	17	39		
captain	comicstrip	hulk	wolverin	lacy	izzo		
seqitur	mutant	mazelyah	albion				
	patlafontain						

Zhai and Boyd-Graber, 2013

minibatch-5	minibatch-8	minibatch-10	minibatch-16	minibatch-17	minibatch-39	minibatch-83	minibatch-120
102-club	118-club	132-rock	87-series	82-series	1-annual	0-captain	0-appear
115-issuee	128-copy	194-issue	161-issue	162-issue	2-rock	1-appear	1-hulk
127-cover	137-cover	215-series	283-copy	288-copy	3-wolverin	3-hulk	2-wolverin
130-copy	138-issue	217-copy	306-appear	294-appear	4-appear	5-rock	3-annual 4-copy
197-appear	180-appear	226-cover	307-cover	311-cover	5-comicstrip 6-series	6-wolverin	5-rider
289-rock	319-rock	261-appear	502-annual	512-annual	7-mutant	9-comicstrip	6-comicstrip
	493-annual		814-force	830-force	8-cover	12-annal	7-cover
450-annual	639-series	588-annual	1194-rider		12-issue	13-mutant	8-force
584-series		949-force		4782-wolverin	14-hulk	15-series	9-captain
811-forcee	877-force	1074-rider	8944-hulk	9659-hulk		16-cover	11-issue
1090-rider	1003-rider	6038-comicstrip	10819-comicstrip	11527-comicstrip	16-copy	19-сору	12-series
	7075-captain	6520-mutant	11301-mutant	12009-mutant	53-force	23-issue	16-mutant
		9569-captain	14335-captain	15040-captain	57-rider	280-rider	41-rock
					86-captain		41-10CK
5	8	10	16	17	39	83	
captain	comicstrip	hulk	wolverin	lacy	izzo	gown	
seqitur	mutant	mazelyah	albion				
	patlafontain						

Zhai and Boyd-Graber, 2013

minibatch-5	minibatch-8	minibatch-10	minibatch-16	minibatch-17	minibatch-39	minibatch-83	minibatch-120
102-club	118-club	132-rock	87-series	82-series	1-annual	0-captain	0-appear
115-issuee	128-copy	194-issue	161-issue	162-issue	2-rock	1-appear	1-hulk
127-cover	137-cover	215-series	283-copy	288-copy	3-wolverin	3-hulk	2-wolverin 3-annual
130-copy	138-issue	217-copy	306-appear	294-appear	4-appear 5-comicstrip	5-rock	4-copy
197-appear	180-appear	226-cover	307-cover	311-cover	6-series	6-wolverin	5-rider
289-rock	319-rock	261-appear	502-annual	512-annual	7-mutant	9-comicstrip	6-comicstrip
450-annual	493-annual	588-annual	814-force	830-force	8-cover	12-annal	7-cover
584-series	639-series	949-force	1194-rider	4782-wolverin	12-issue	13-mutant	8-force
811-forcee	877-force	1074-rider	8944-hulk	9659-hulk	14-hulk	15-series	9-captain
1090-rider	1003-rider	6038-comicstrip	10819-comicstrip	11527-comicstrip	16-copy	16-cover	11-issue
1090-Huel	7075-captain		11301-mutant	12009-mutant	53-force	19-copy	12-series
	7075-captairi	6520-mutant		15040-captain	57-rider	23-issue	16-mutant
		9569-captain	14335-captain	15040-captain	86-captain	280-rider	41-rock
					00-captairi		
5	8	10	16	17	39	83	
captain	comicstrip	hulk	wolverin	lacy	izzo	gown	
seqitur	mutant	mazelyah	albion				
	patlafontain						

Zhai and Boyd-Graber, 2013

minibatch-5	minibatch-8	minibatch-10	minibatch-16	minibatch-17	minibatch-39	minibatch-83	minibatch-120
102-club	118-club	132-rock	87-series	82-series	1-annual	0-captain	0-appear
115-issuee	128-copy	194-issue	161-issue	162-issue	2-rock	1-appear	1-hulk
127-cover	137-cover	215-series	283-copy	288-copy	3-wolverin	3-hulk	2-wolverin
130-copy	138-issue	217-copy	306-appear	294-appear	4-appear	5-rock	3-annual
197-appear	180-appear	226-cover	307-cover	311-cover	5-comicstrip	6-wolverin	4-copy 5-rider
	319-rock		502-annual		6-series	9-comicstrip	6-comicstrip
289-rock		261-appear	A	512-annual	7-mutant	12-annal	
450-annual	493-annual	588-annual	814-force	830-force	8-cover	13-mutant	7-cover
584-series	639-series	949-force	1194-rider	4782-wolverin	12-issue		8-force
811-forcee	877-force	1074-rider	8944-hulk	9659-hulk	14-hulk	15-series	9-captain
	1003-rider		10819-comicstrip		16-copy	16-cover	11-issue
1090-rider	1003-Huei	6038-comicstrip		11527-comicsurp		19-copy	12-series
	7075-captain	6520-mutant	11301-mutant	12009-mutant	53-force	23-issue	16-mutant
		9569-captain	14335-captain	15040-captain	57-rider	280-rider	41-rock
					86-captain		41-10CK
5	8	10	16	17	39	83	1
-	_						
captain	comicstrip	hulk	wolverin	lacy	izzo	gown	
seqitur	mutant	mazelyah	albion				I
	patlafontain						l