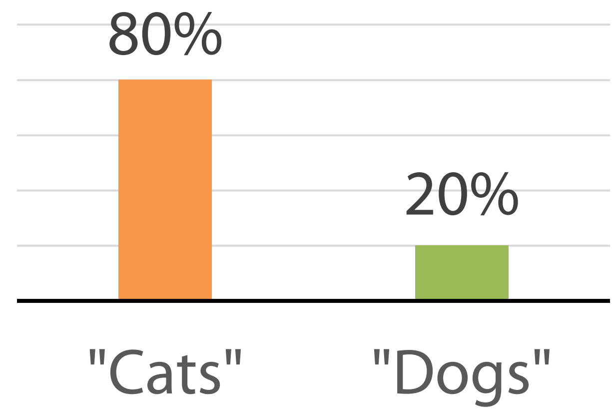
# MCMC for LDA

# Latent Dirichlet Allocation

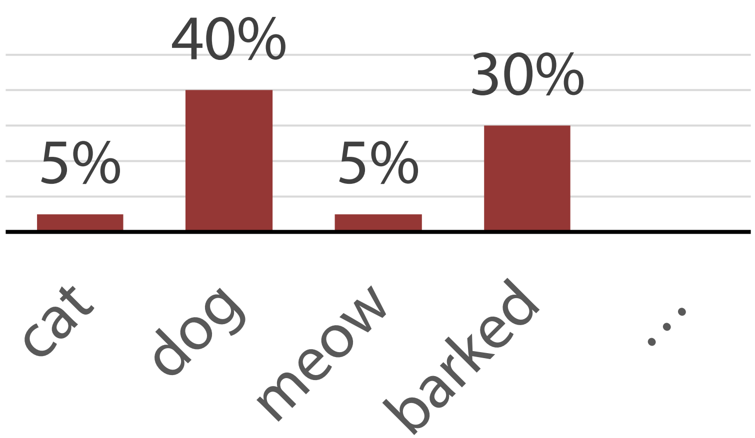
Document is a distribution over topics

Document

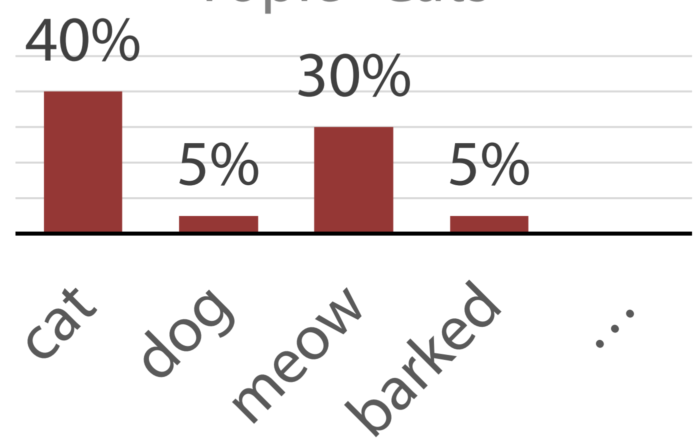



Topic is a distribution over words

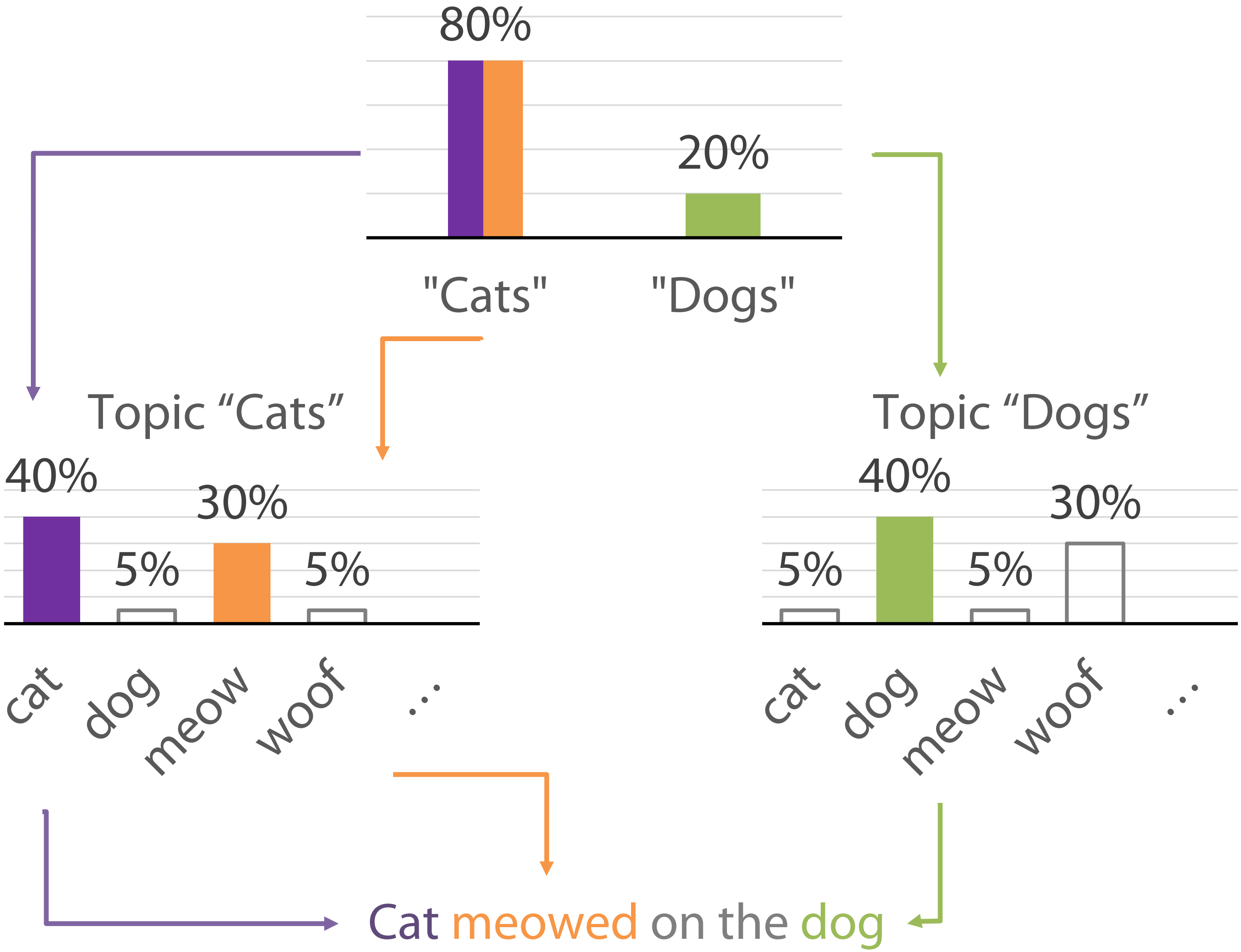
Topic “Dogs”



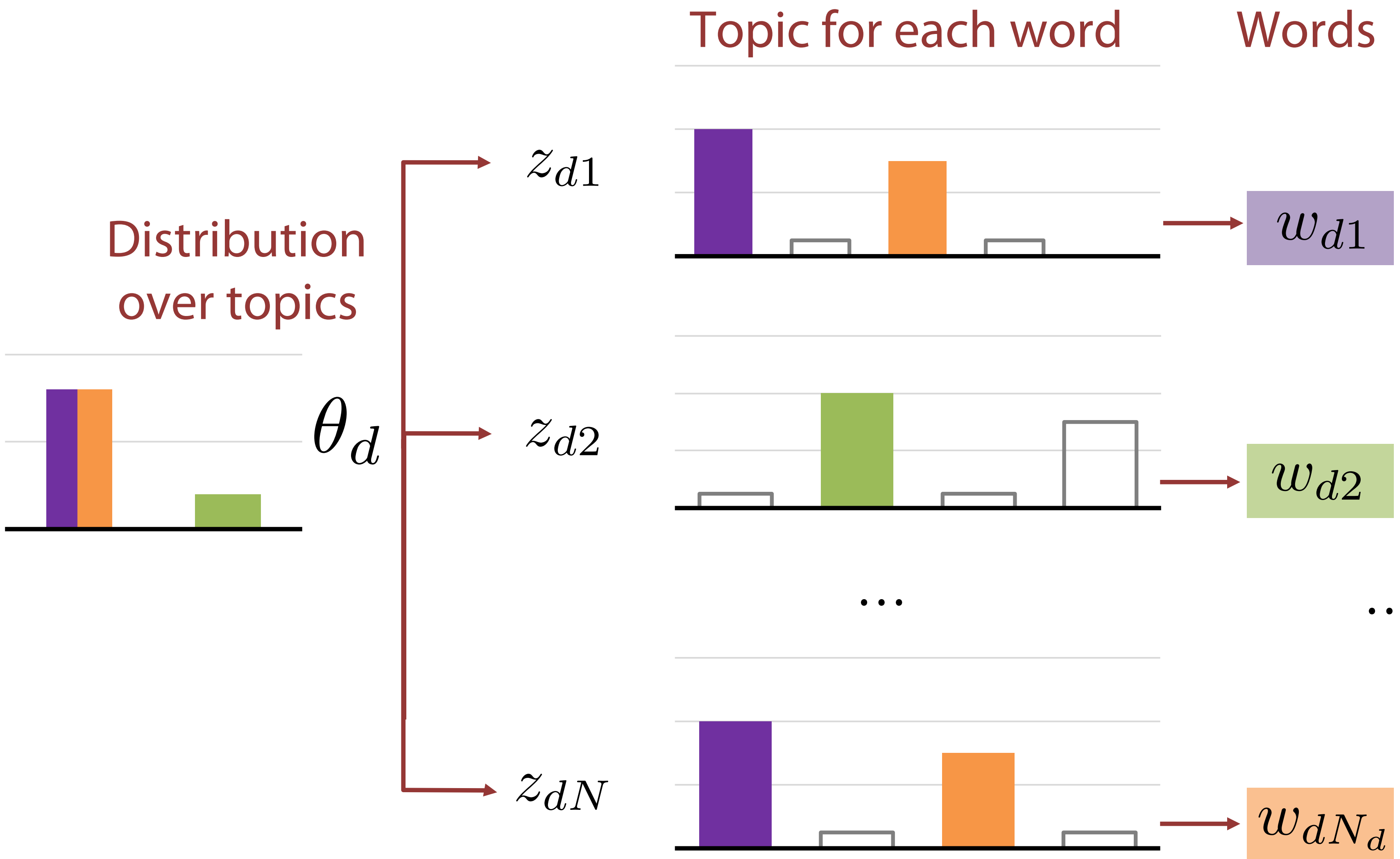

Topic “Cats”



# Text generation



# Model




$$z_{dn} \in \{1 \,..\, T\} \quad w_{dn} \in \{1..V\}$$

# Mean Field for LDA (week 3)

$$\max_{\Phi} p(W \mid \Phi)$$

**E-step** $$p(Z, \Theta \mid W, \Phi) \approx q(Z, \Theta)$$

**M-step** $$\max_{\Phi} \mathbb{E}_q \log p(Z, \Theta, W \mid \Phi)$$

# Mean Field for LDA (week 3)

$$\log q(Z) = \mathbb{E}_{q(\Theta)} \log p(\Theta, Z, W) + \text{const}$$

$$= \mathbb{E}_{q(\Theta)} \sum_{d=1}^{D} \sum_{t=1}^{T} (\alpha - 1) \log \theta_{dt} + \text{const}$$

$$+ \mathbb{E}_{q(\Theta)} \sum_{d=1}^{D} \sum_{n=1}^{N_d} \sum_{t=1}^{T} [z_{d_n} = 1](\log \theta_{dt} + \log \phi_{tw_{dn}})$$

$$= \sum_{d=1}^{D} \sum_{n=1}^{N_d} \sum_{t=1}^{T} [z_{d_n} = 1] \left(\mathbb{E}_{q(\Theta)} \log \theta_{dt} + \log \phi_{tw_{dn}}\right)$$

$$+ \text{const}$$

$$q(Z) = \prod_{d=1}^{D} q(z_d) = \ldots$$

# MCMC for LDA

**Known:** $W$ data

**Unknown:** $\Phi$ parameters, distribution over words for each topic

**Unknown:** $Z$ latent variables, topic of each word

**Unknown:** $\Theta$ latent variables, distribution over topics for each document

# MCMC for LDA

Lets go full Bayesian

**Known:** $W$ data

**Unknown:** $\Phi$ latent variables, distribution over words for each topic

**Unknown:** $Z$ latent variables, topic of each word

**Unknown:** $\Theta$ latent variables, distribution over topics for each document

# MCMC for LDA

Lets go full Bayesian

**Known:** $W$ data

**Unknown:** $\Phi$ latent variables, distribution over words for each topic

**Unknown:** $Z$ latent variables, topic of each word

**Unknown:** $\Theta$ latent variables, distribution over topics for each document

$$p(\Phi, \Theta, Z|W) \sim \{Gibbs\ Sampling\}$$

# MCMC for LDA

$$p(\Phi, \Theta, Z|W) \sim \{Gibbs\ Sampling\}$$

$$\text{Init: } \Phi^0, \Theta^0, Z^0$$

# MCMC for LDA

$$p(\Phi, \Theta, Z | W) \sim \{Gibbs\ Sampling\}$$

Init: $\Phi^0, \Theta^0, Z^0$

$$\phi_1^1 \sim p(\phi_1 | \phi_2^0, \phi_3^0, \ldots, \Theta^0, Z^0, W)$$

# MCMC for LDA

$$p(\Phi, \Theta, Z|W) \sim \{Gibbs\ Sampling\}$$

$$\text{Init: } \Phi^0, \Theta^0, Z^0$$

$$\phi_1^1 \sim p(\phi_1|\phi_2^0, \phi_3^0, \ldots, \Theta^0, Z^0, W)$$

$$\phi_2^1 \sim p(\phi_2|\phi_1^1, \phi_3^0, \ldots, \Theta^0, Z^0, W)$$

# MCMC for LDA

$$p(\Phi, \Theta, Z|W) \sim \{Gibbs\ Sampling\}$$

$$\text{Init: } \Phi^0, \Theta^0, Z^0$$

$$\phi_i^1 \sim p(\phi_i|\phi_1^1, \ldots, \phi_{i-1}^1, \phi_{i+1}^0, \ldots, \Theta^0, Z^0, W)$$

# MCMC for LDA

$$p(\Phi, \Theta, Z|W) \sim \{Gibbs\ Sampling\}$$

Init: $\Phi^0, \Theta^0, Z^0$

$$\phi_i^1 \sim p(\phi_i|\phi_1^1, \ldots, \phi_{i-1}^1, \phi_{i+1}^0, \ldots, \Theta^0, Z^0, W)$$

$$\theta_i^1 \sim p(\theta_i|\Phi^1, \theta_1^1, \ldots, \theta_{i-1}^1, \theta_{i+1}^0, \ldots, Z^0, W)$$

# MCMC for LDA

$$p(\Phi, \Theta, Z|W) \sim \{Gibbs\ Sampling\}$$

$$\text{Init: } \Phi^0, \Theta^0, Z^0$$

$$\phi_i^1 \sim p(\phi_i|\phi_1^1, \ldots, \phi_{i-1}^1, \phi_{i+1}^0, \ldots, \Theta^0, Z^0, W)$$

$$\theta_i^1 \sim p(\theta_i|\Phi^1, \theta_1^1, \ldots, \theta_{i-1}^1, \theta_{i+1}^0, \ldots, Z^0, W)$$

$$z_i^1 \sim p(z_i|\Phi^1, \Theta^1, z_1^1, \ldots, z_{i-1}^1, z_{i+1}^0, \ldots, W)$$

# MCMC for LDA

$$p(\Phi, \Theta, Z|W) \sim \{\mathit{Gibbs\ Sampling}\}$$

Init: $\Phi^0, \Theta^0, Z^0$

For $\mathbf{k} = 1, 2, \ldots$

$$\phi_i^{\mathbf{k+1}} \sim p(\phi_i|\phi_1^{\mathbf{k+1}}, \ldots, \phi_{i-1}^{\mathbf{k+1}}, \phi_{i+1}^{\mathbf{k}}, \ldots, \Theta^{\mathbf{k}}, Z^{\mathbf{k}}, W)$$

$$\theta_i^{\mathbf{k+1}} \sim p(\theta_i|\Phi^{\mathbf{k+1}}, \theta_1^{\mathbf{k+1}}, \ldots, \theta_{i-1}^{\mathbf{k+1}}, \theta_{i+1}^{\mathbf{k}}, \ldots, Z^{\mathbf{k}}, W)$$

$$z_i^{\mathbf{k+1}} \sim p(z_i|\Phi^{\mathbf{k+1}}, \Theta^{\mathbf{k+1}}, z_1^{\mathbf{k+1}}, \ldots, z_{i-1}^{\mathbf{k+1}}, z_{i+1}^{\mathbf{k}}, \ldots, W)$$

# Collapsed Gibbs for LDA

**Model**

$$p(\theta_d) = \mathrm{Dir}(\beta) \qquad p(\phi_t) = \mathrm{Dir}(\alpha)$$

$$p(z_{dn}|\theta_d) = \Theta_{dz_{dn}} \qquad p(w_{dn}|z_{dn}, \Phi) = \Phi_{z_{dn}w_{dn}}$$

# Collapsed Gibbs for LDA

**Model**

$$p(\theta_d) = \mathrm{Dir}(\beta) \qquad p(\phi_t) = \mathrm{Dir}(\alpha)$$

$$p(z_{dn} | \theta_d) = \Theta_{d z_{dn}} \qquad p(w_{dn} | z_{dn}, \Phi) = \Phi_{z_{dn} w_{dn}}$$

**Conjugate**

# Collapsed Gibbs for LDA

## Model

$$p(\theta_d) = \mathrm{Dir}(\beta) \qquad p(\phi_t) = \mathrm{Dir}(\alpha)$$

$$p(z_{dn}|\theta_d) = \Theta_{dz_{dn}} \qquad p(w_{dn}|z_{dn}, \Phi) = \Phi_{z_{dn}w_{dn}}$$

## Can compute analytically

$$p(\Theta \mid Z)$$

# Collapsed Gibbs for LDA

## Model

$$p(\theta_d) = \mathrm{Dir}(\beta) \qquad p(\phi_t) = \mathrm{Dir}(\alpha)$$

$$p(z_{dn}|\theta_d) = \Theta_{dz_{dn}} \qquad p(w_{dn}|z_{dn}, \Phi) = \Phi_{z_{dn}w_{dn}}$$

## Can compute analytically

$$p(\Theta \mid Z)$$

$$p(Z) = \int p(Z \mid \Theta)p(\Theta)d\Theta$$

# Collapsed Gibbs for LDA

## Model

$$p(\theta_d) = \mathrm{Dir}(\beta) \qquad p(\phi_t) = \mathrm{Dir}(\alpha)$$

$$p(z_{dn}|\theta_d) = \Theta_{dz_{dn}} \qquad p(w_{dn}|z_{dn}, \Phi) = \Phi_{z_{dn}w_{dn}}$$

## Can compute analytically

$$p(\Theta \mid Z)$$

$$p(Z) = \int p(Z \mid \Theta)p(\Theta)d\Theta$$

$$= \frac{p(Z \mid \Theta)p(\Theta)}{p(\Theta \mid Z)}$$

# Collapsed Gibbs for LDA

## Model

$$p(\theta_d) = \mathrm{Dir}(\beta) \qquad p(\phi_t) = \mathrm{Dir}(\alpha)$$

$$p(z_{dn}|\theta_d) = \Theta_{dz_{dn}} \qquad p(w_{dn}|z_{dn}, \Phi) = \Phi_{z_{dn}w_{dn}}$$

## Can compute analytically

$$p(\Theta \mid Z)$$

$$p(Z)$$

# Collapsed Gibbs for LDA

**Model**

$$p(\theta_d) = \mathrm{Dir}(\beta) \qquad p(\phi_t) = \mathrm{Dir}(\alpha)$$

$$p(z_{dn}|\theta_d) = \Theta_{dz_{dn}} \qquad p(w_{dn}|z_{dn}, \Phi) = \Phi_{z_{dn}w_{dn}}$$

**Can compute analytically**

**Conjugate**

$$p(\Theta \mid Z)$$

$$p(Z)$$

# Collapsed Gibbs for LDA

**Model**

$$p(\theta_d) = \mathrm{Dir}(\beta) \qquad p(\phi_t) = \mathrm{Dir}(\alpha)$$

$$p(z_{dn} | \theta_d) = \Theta_{d z_{dn}} \qquad p(w_{dn} | z_{dn}, \Phi) = \Phi_{z_{dn} w_{dn}}$$

**Can compute analytically**

$$p(\Theta \mid Z) \qquad p(\Phi \mid Z, W)$$

$$p(Z)$$

# Collapsed Gibbs for LDA

**Model**

$$p(\theta_d) = \mathrm{Dir}(\beta) \qquad p(\phi_t) = \mathrm{Dir}(\alpha)$$

$$p(z_{dn}|\theta_d) = \Theta_{dz_{dn}} \qquad p(w_{dn}|z_{dn}, \Phi) = \Phi_{z_{dn}w_{dn}}$$

**Can compute analytically**

$$p(\Theta \mid Z) \qquad p(\Phi \mid Z, W)$$

$$p(Z) \qquad p(W \mid Z) = \frac{p(W \mid Z, \Phi)\, p(\Phi)}{p(\Phi \mid Z, W)}$$

# Collapsed Gibbs for LDA

**Model**

$$p(\theta_d) = \mathrm{Dir}(\beta) \qquad p(\phi_t) = \mathrm{Dir}(\alpha)$$

$$p(z_{dn}|\theta_d) = \Theta_{d z_{dn}} \qquad p(w_{dn}|z_{dn}, \Phi) = \Phi_{z_{dn} w_{dn}}$$

**Can compute analytically**

$$p(\Theta \mid Z) \qquad p(\Phi \mid Z, W)$$

$$p(Z) \qquad p(W \mid Z)$$

# Collapsed Gibbs for LDA

**Model**

$$p(\theta_d) = \mathrm{Dir}(\beta) \qquad p(\phi_t) = \mathrm{Dir}(\alpha)$$

$$p(z_{dn}|\theta_d) = \Theta_{dz_{dn}} \qquad p(w_{dn}|z_{dn}, \Phi) = \Phi_{z_{dn}w_{dn}}$$

**Can compute analytically**

$$p(\Theta \mid Z) \qquad p(\Phi \mid Z, W) \qquad p(Z \mid W) = \frac{p(W \mid Z)p(Z)}{C}$$

$$p(Z) \qquad p(W \mid Z)$$

# Collapsed Gibbs for LDA

**Model**

$$p(\theta_d) = \mathrm{Dir}(\beta) \qquad p(\phi_t) = \mathrm{Dir}(\alpha)$$

$$p(z_{dn}|\theta_d) = \Theta_{dz_{dn}} \qquad p(w_{dn}|z_{dn}, \Phi) = \Phi_{z_{dn}w_{dn}}$$

**Can compute analytically**

$$p(\Theta \mid Z) \qquad p(\Phi \mid Z, W) \qquad p(Z \mid W) = \frac{p(W \mid Z)p(Z)}{C}$$

$$p(Z) \qquad p(W \mid Z)$$

$$p(Z \mid W) \sim \{Gibbs\ Sampling\}$$

# Collapsed Gibbs for LDA

**Model**

$$p(\theta_d) = \mathrm{Dir}(\beta) \qquad p(\phi_t) = \mathrm{Dir}(\alpha)$$

$$p(z_{dn} | \theta_d) = \Theta_{d z_{dn}} \qquad p(w_{dn} | z_{dn}, \Phi) = \Phi_{z_{dn} w_{dn}}$$

$$p(Z \mid W) \sim \{Gibbs\ Sampling\}$$

# Collapsed Gibbs for LDA

**Model**

$$p(\theta_d) = \mathrm{Dir}(\beta) \qquad p(\phi_t) = \mathrm{Dir}(\alpha)$$

$$p(z_{dn}|\theta_d) = \Theta_{dz_{dn}} \qquad p(w_{dn}|z_{dn}, \Phi) = \Phi_{z_{dn}w_{dn}}$$

$$p(Z \mid W) \sim \{\textit{Gibbs Sampling}\}$$

$$\begin{aligned} p(\Phi|W) &= \int p(\Phi|W, Z)p(Z|W)dZ \\ &= \mathbb{E}_{p(Z|W)}p(\Phi|W, Z) \end{aligned}$$

# Collapsed Gibbs for LDA

**Model**

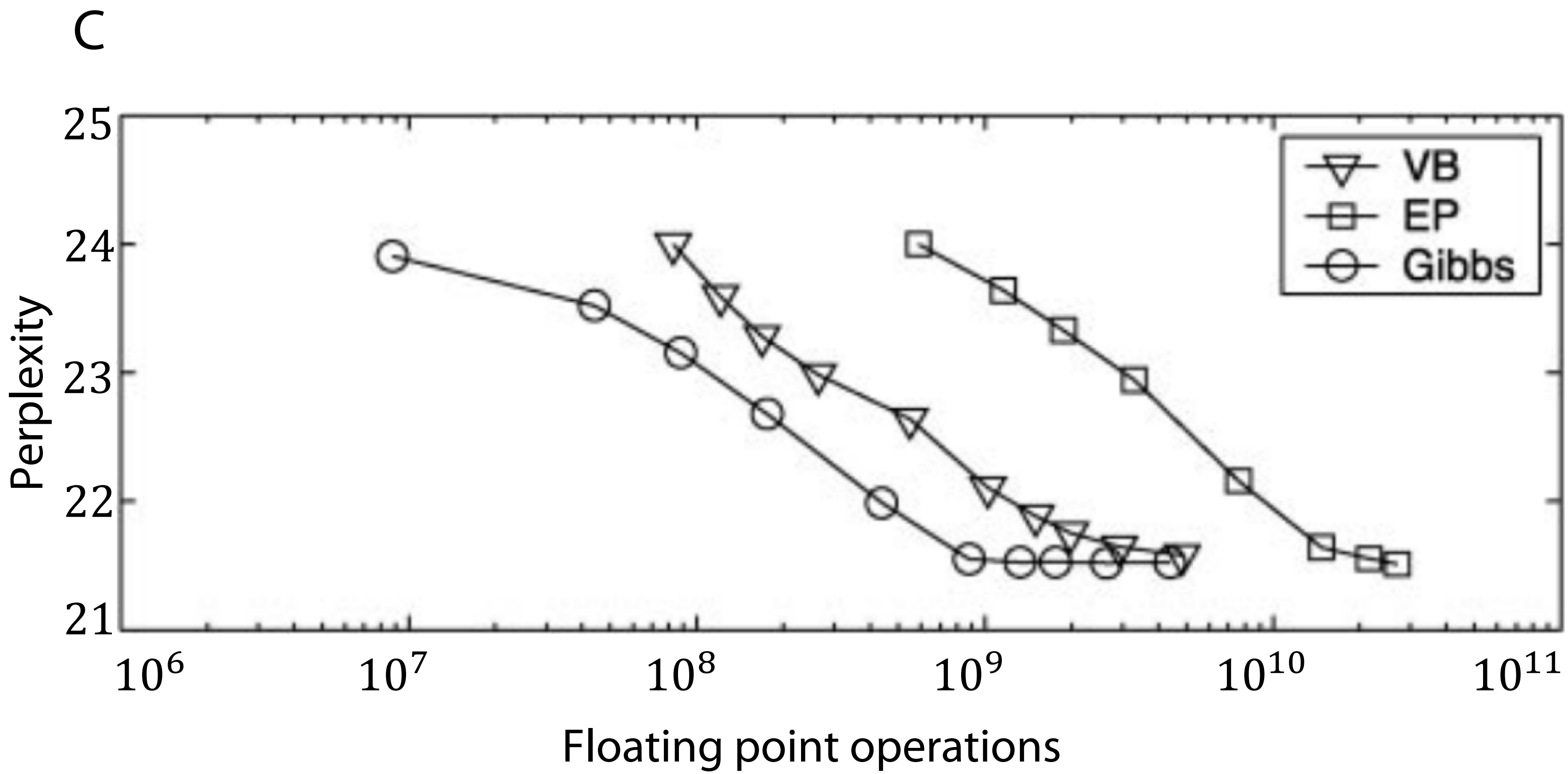
$$p(\theta_d) = \mathrm{Dir}(\beta) \qquad p(\phi_t) = \mathrm{Dir}(\alpha)$$

$$p(z_{dn}|\theta_d) = \Theta_{d z_{dn}} \qquad p(w_{dn}|z_{dn}, \Phi) = \Phi_{z_{dn} w_{dn}}$$

$$p(Z \mid W) \sim \{\textit{Gibbs Sampling}\}$$

$$\begin{aligned} p(\Phi|W) &= \int p(\Phi|W, Z)p(Z|W)dZ \\ &= \mathbb{E}_{p(Z|W)} p(\Phi|W, Z) \\ &\approx p(\Phi|W, \widehat{Z}) \end{aligned}$$

# Collapsed Gibbs for LDA




[Source: Griffiths, Thomas L., and Mark Steyvers. "Finding scientific topics.
" *Proceedings of the National academy of Sciences* 101.suppl 1 (2004): 5228-5235.]