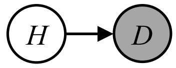
Math Modeling, Week 5

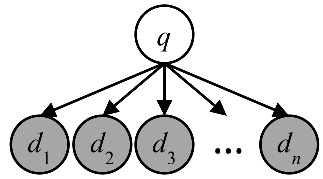
Bayes nets

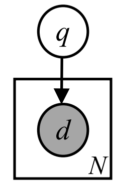
Simple Bayesian inference



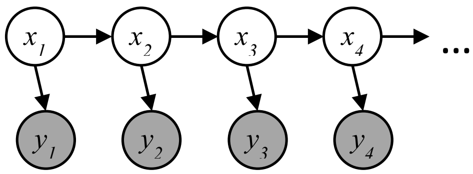
IID sampling (e.g., Bernoulli/multinomial)

plates





Kalman filter



Latent Dirichlet Allocation

Hierarchical Bayesian model with latent structure

Used for modeling corpora (topic models)

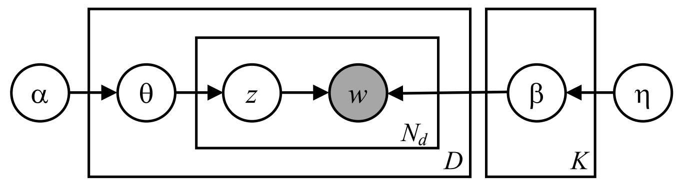
Observed data:

Documents $d \in \{1,\ldots,D\}$, words $w_{d,n}$ for $n \in \{1,\ldots,N_d\}$

Latent structure:

Topics $k \in \{1,\ldots,K\}$, topic–word-type distributions β_k

Document-topic distributions θ_d, word-token–topic assignments $z_{d,n} \in \{1,\ldots,K\}$ for $n \in \{1,\ldots,N_d\}$



Joint distribution:

$$p(\beta,\theta,z,w|\alpha,\eta) = p(\beta|\eta)\cdot p(\theta|\alpha)\cdot p(z|\theta)\cdot p(w|z,\beta)$$

$$= \prod_k \mathrm{Dir}(\beta_k;\eta)\cdot\prod_{d=1}^{D}\left[\mathrm{Dir}(\theta_d;\alpha)\cdot\prod_{n=1}^{N_d}\left[\theta_d(z_{d,n})\cdot\beta_{z_{d,n}}(w_{d,n})\right]\right]$$

Marginal likelihood (normalization for posterior on β,θ,z) is intractable:

$$p(w_{d,n}|\alpha,\eta) = \iint_{\beta,\theta}\left[\prod_{\mathrm{k}}\mathrm{Dir}(\beta_k;\eta)\prod_{d=1}^{D}\left[\mathrm{Dir}(\theta;\alpha)\prod_{n=1}^{N_d}\sum_{z_{d,n}}\left[\theta_d(z_{d,n})\cdot\beta_{z_{d,n}}(w_{d,n})\right]\right]\right]d\theta d\beta$$

Gibbs sampling in Bayes nets

$p(x_i \mid \mathbf{x}_{-i}) = p(x_i \mid \mathbf{x}_{\mathrm{An}(i)}, \mathbf{x}_{\mathrm{Pa}(i)}, \mathbf{x}_{\mathrm{Ch}(i)}, \mathbf{x}_{\mathrm{De}(i)})$ [Ancestors, Parents, Children, Descendants]

$\propto p(x_i \mid \mathbf{x}_{\mathrm{An}(i)}, \mathbf{x}_{\mathrm{Pa}(i)}) \cdot p(\mathbf{x}_{\mathrm{Ch}(i)}, \mathbf{x}_{\mathrm{De}(i)} \mid x_i, \mathbf{x}_{\mathrm{An}(i)}, \mathbf{x}_{\mathrm{Pa}(i)})$

$= p(x_i \mid \mathbf{x}_{\mathrm{Pa}(i)}) \cdot \prod_{j\in\mathrm{Ch}(i)\cup\mathrm{De}(i)} p(x_j \mid \mathbf{x}_{\mathrm{Pa}(j)})$

$\propto p(x_i \mid \mathbf{x}_{\mathrm{Pa}(i)}) \cdot \prod_{j\in\mathrm{Ch}(i)} p(x_j \mid \mathbf{x}_{\mathrm{Pa}(j)})$

Alternatively, $p(x_i|\mathbf{x}_{-i}) = p(x_i,\mathbf{x}_{-i})/\Sigma p(x_i',\mathbf{x}_{-i})$

Joint distribution $\prod_j p(x_j|\mathbf{x}_{\mathrm{Pa}(j)})$

All terms cancel (same for all x_i') except $j=i$ and $i\in\mathrm{Pa}(j)$

LDA

$$p(z|w,\beta,\theta,\alpha,\eta) \propto p(z|\theta)\cdot p(w|z,\beta)$$

$$= \prod_{d=1}^{D}\prod_{n=1}^{N_d}\left[\theta_d(z_{d,n})\cdot\beta_{z_{d,n}}(w_{d,n})\right]$$

$$\text{normalization: } \prod_{d=1}^{D}\prod_{n=1}^{N_d}\sum_{z}\left[\theta_d(z)\cdot\beta_z(w_{d,n})\right]$$

$$p(\theta|w,z,\beta,\alpha,\eta) \propto p(\theta|\alpha)\cdot p(z|\theta)$$

$$\propto \prod_{d=1}^{D}\mathrm{Dir}(\theta_d;\alpha+|z_d|)$$

$|\cdot|$ = vector of counts

$$p(\beta|w,z,\theta,\alpha,\eta) \propto p(\beta|\eta)\cdot p(w|z,\beta)$$

$$\propto \prod_{k=1}^{K}\mathrm{Dir}(\beta_k;\eta+|w_{z=k}|)$$