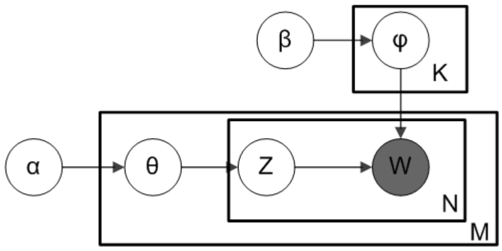
LDA Gibbs Sampling 详细推导过程， 2013.10.6， ys

作为基本的 topic model,熟悉 LDA 的推导过程还是很有必要的，当掌握了它的推导过程后，再遇到其他相关图模型的推断你就会觉得‘so easy!’。每次去推断别的模型，总在脑袋里回味一下这个基本的 topic model 的推断过程，趁现在跑程序总结下来，分享学习，做个记录，也欢迎各位不吝赐教。

下文假定各位读者已经熟悉了 LDA 的先验分布假设，生成过程等，只给出了主要的公式推导流程。

0. 概率图模型如下图：



1. 首先给出联合概率：$\mathrm{p}(\mathbf{w},\mathbf{z}|\alpha,\beta)=\mathrm{p}(\mathbf{w}|\mathbf{z},\alpha,\beta)\mathrm{p}(\mathbf{z}|\alpha,\beta)=\mathrm{p}(\mathbf{w}|\mathbf{z},\beta)\mathrm{p}(\mathbf{z}|\alpha)$

*注：当我们确定下来一个模型后，首先计算它的联合概率，这个是绝对不会错的，在后续的采样中是很有用的。

• 1.1 第一项$\mathrm{p}(\mathbf{w}|\mathbf{z},\beta)$:

$$\mathrm{p}(\mathbf{w}|\mathbf{z},\beta)=\int \mathrm{p}(\mathbf{w}|\mathbf{z},\Phi)\mathrm{p}(\Phi|\beta)\,d\Phi$$

$$\mathrm{p}(\mathbf{w}|\mathbf{z},\Phi)=\prod_{i=1}^{N*M}\varphi_{z_i,w_i}^{n(z_i,w_i)}=\prod_{k=1}^{K}\prod_{t=1}^{V}\varphi_{k,t}^{n(k,t)}$$

$$\mathrm{p}(\Phi|\beta)=\prod_{k=1}^{K}p(\varphi_k|\beta)=\prod_{k=1}^{K}\frac{\Gamma(\sum_{t=1}^{V}\beta_t)}{\prod_{t=1}^{V}\Gamma(\beta_t)}\prod_{t=1}^{V}\varphi_{k,t}^{\beta_t-1}$$

$$\Rightarrow \mathrm{p}(\mathbf{w}|\mathbf{z},\beta)=\int \mathrm{p}(\mathbf{w}|\mathbf{z},\Phi)\mathrm{p}(\Phi|\beta)\,d\Phi$$

$$=\int\prod_{k=1}^{K}\prod_{t=1}^{V}\varphi_{k,t}^{n(k,t)}\cdot\prod_{k=1}^{K}\frac{\Gamma(\sum_{t=1}^{V}\beta_t)}{\prod_{t=1}^{V}\Gamma(\beta_t)}\prod_{t=1}^{V}\varphi_{k,t}^{\beta_t-1}\,d\varphi_{k,t}$$

$$=\int\prod_{k=1}^{K}\frac{\Gamma(\sum_{t=1}^{V}\beta_t)}{\prod_{t=1}^{V}\Gamma(\beta_t)}\prod_{t=1}^{V}\varphi_{k,t}^{n(k,t)+\beta_t-1}\,d\varphi_{k,t}$$

$$=\int\prod_{k=1}^{K}\frac{\Delta(n_k+\beta)}{\Delta(\beta)}\frac{1}{\Delta(n_k+\beta)}\prod_{t=1}^{V}\varphi_{k,t}^{n(k,t)+\beta_t-1}\,\mathrm{d}\varphi_{k,t}\ (\text{其中}\frac{1}{\Delta(\beta)}=\frac{\Gamma(\sum_{t=1}^{V}\beta_t)}{\prod_{t=1}^{V}\Gamma(\beta_t)}(\text{以下 }\Delta(\alpha)\text{类似}))$$

$$=\prod_{k=1}^{K}\frac{\Delta(n_k+\beta)}{\Delta(\beta)}\quad(\text{利用}\int\frac{1}{\Delta(n_k+\beta)}\prod_{t=1}^{V}\varphi_{k,t}^{n(k,t)+\beta_t-1}\,d\varphi_{k,t}=1)\text{。}$$

• 1.2 第二项$\mathrm{p}(\mathbf{z}|\alpha)$：

$$\mathrm{p}(\mathbf{z}|\alpha)=\int p(\mathbf{z}|\Theta)p(\Theta|\alpha)d\Theta$$

$$\mathrm{p}(\mathbf{z}|\Theta,\alpha)=\prod_{i=1}^{N*M}\theta_{di,zi}=\prod_{d=1}^{M}\prod_{k=1}^{K}\theta_{d,k}^{n(d,k)}$$

$$\mathrm{p}(\Theta|\alpha)=\prod_{d=1}^{M}p(\theta_d|\alpha)=\prod_{d=1}^{M}\frac{\Gamma(\sum_{k=1}^{K}\alpha_k)}{\prod_{k=1}^{K}\Gamma(\alpha_k)}\prod_{k=1}^{K}\theta_{d,k}^{\alpha_k-1}$$

$$\Rightarrow\mathrm{p}(\mathbf{z}|\alpha)=\int p(\mathbf{z}|\Theta)p(\Theta|\alpha)d\Theta$$

$$=\int\prod_{d=1}^{M}\prod_{k=1}^{K}\theta_{d,k}^{n(d,k)}\cdot\prod_{d=1}^{M}\frac{\Gamma(\sum_{k=1}^{K}\alpha_k)}{\prod_{k=1}^{K}\Gamma(\alpha_k)}\prod_{k=1}^{K}\theta_{d,k}^{\alpha_k-1}\,d\theta_{d,k}$$

$$=\int\prod_{d=1}^{M}\frac{\Gamma(\sum_{k=1}^{K}\alpha_k)}{\prod_{k=1}^{K}\Gamma(\alpha_k)}\prod_{k=1}^{K}\theta_{d,k}^{n(d,k)+\alpha_k-1}\,d\theta_{d,k}$$

$$=\prod_{d=1}^{M}\frac{\Delta(n_d+\alpha)}{\Delta(\alpha)}(\text{同上述 }\mathrm{p}(\mathbf{w}|\mathbf{z},\beta)\text{的最后一步推导})$$

• 1.3 联合 1.1,1.2 有联合概率：

$$\mathrm{p}(\mathbf{w},\mathbf{z}|\alpha,\beta)=\mathrm{p}(\mathbf{w}|\mathbf{z},\alpha,\beta)\mathrm{p}(\mathbf{z}|\alpha,\beta)=\mathrm{p}(\mathbf{w}|\mathbf{z},\beta)\mathrm{p}(\mathbf{z}|\alpha)=\prod_{k=1}^{K}\frac{\Delta(n_k+\beta)}{\Delta(\beta)}\cdot\prod_{d=1}^{M}\frac{\Delta(n_d+\alpha)}{\Delta(\alpha)}$$

2. 计算采样的条件概率:第 i 个 word 的 topic 为 k 的概率 $p(z_i = k|z_{-i}, \mathbf{w};\alpha,\beta)$

$$p(z_i = k|\boldsymbol{z}_{-i}, \mathbf{w};\alpha,\beta) = \frac{p(z_i = k, \boldsymbol{z}_{-i}, \boldsymbol{w};\alpha,\beta)}{p(\boldsymbol{z}_{-i}, \boldsymbol{w};\alpha,\beta)} \propto p(z_i = k, \boldsymbol{z}_{-i}, \boldsymbol{w};\alpha,\beta) = p(\boldsymbol{w}|z_i = k, \boldsymbol{z}_{-i};\alpha,\beta) \cdot p(z_i = k, \boldsymbol{z}_{-i};\alpha,\beta)$$

- 2.1 第一项$p(\boldsymbol{w}|z_i = k, \boldsymbol{z}_{-i};\alpha,\beta)$分为两步走:

  (1).不包含第 i 个 word 及其所在 z_i=k 的项:

$$\prod_{\mathrm{z=1,z\neq zi}}^{K} \frac{\Delta(n_{\mathrm{z,-i}} + \beta)}{\Delta(\beta)}$$

  (2). 第 i 个 word 所在的 zi=k 项:

$$\frac{\Delta(n_k + \beta)}{\Delta(\beta)}$$

  (这里的 n_k是包含了第 k 个 topic 下的 word 数目，已经包含了第 i 个 word)。

  由此有:

$$
\begin{aligned}
&p(z_i = k|\boldsymbol{z}_{-i}, \mathbf{w};\alpha,\beta) \propto p(\boldsymbol{w}|z_i = k, \boldsymbol{z}_{-i};\alpha,\beta)\\
&\quad = \left(\prod_{\mathrm{z=1,z\neq k}}^{K} \frac{\Delta\left(n_{\mathrm{z,-i}} + \beta\right)}{\Delta(\beta)}\right) \cdot \frac{\Delta(n_k + \beta)}{\Delta(\beta)}\\
&\quad = \prod_{\mathrm{z=1}}^{K} \frac{\Delta(n_{\mathrm{z,-i}} + \beta)}{\Delta(\beta)} \cdot \frac{\Delta(\beta)}{\Delta\left(n_{z=k,-\mathrm{i}} + \beta\right)} \cdot \frac{\Delta(n_k + \beta)}{\Delta(\beta)} \propto \frac{\Delta(n_k + \beta)}{\Delta(n_{z=k,-\mathrm{i}} + \beta)}\\
&\quad = \frac{\Gamma(\sum_{t=1}^{V}(n_{k,-i}^t + \beta_t))}{\prod_{t=1}^{V}\Gamma\left(n_{k,-i}^t + \beta_t\right)} \cdot \frac{\prod_{t=1}^{V}\Gamma\left(n_k^t + \beta_t\right)}{\Gamma(\sum_{t=1}^{V}(n_k^t + \beta_t))} \ \text{(利用 } \Gamma(a+1) = a\Gamma(a), \text{及} n_k^t = n_{k,-i}^t + 1)\\
&\quad = \frac{n_{k,-i}^t + \beta_t}{\sum_{t=1}^{V}(n_{k,-i}^t + \beta_t)}
\end{aligned}
$$

- 2.2 关于第二项$p(z_i = k, \boldsymbol{z}_{-i};\alpha,\beta) = p(z_i = k, \boldsymbol{z}_{-i};\alpha)$同样分为两步走:

  (1). 不包含第 i 个 word 所在的 doc di 项:

$$\prod_{d=1,\mathrm{d}\neq d_i}^{M} \frac{\Delta(n_{d,-i} + \alpha)}{\Delta(\alpha)}$$

  (2). 第 i 个 word 所在的项:

$$\frac{\Delta(n_{di} + \alpha)}{\Delta(\alpha)}$$

  从而有

$$
\begin{aligned}
&p(z_i = k, \boldsymbol{z}_{-i};\alpha,\beta) = p(z_i = k, \boldsymbol{z}_{-i};\alpha)\\
&\quad = \prod_{d=1,\mathrm{d}\neq d_i}^{M} \frac{\Delta\left(n_{d,-i} + \alpha\right)}{\Delta(\alpha)} \cdot \frac{\Delta\left(n_{di} + \alpha\right)}{\Delta(\alpha)}\\
&\quad = \prod_{d=1}^{M} \frac{\Delta(n_{d,-i} + \alpha)}{\Delta(\alpha)} \cdot \frac{\Delta(\alpha)}{\Delta\left(n_{\mathrm{di},-i} + \alpha\right)} \cdot \frac{\Delta(n_{di} + \alpha)}{\Delta(\alpha)} \ \text{(注意此时在第一个分式中加入不含第 i 个 word 的 doc di)}\\
&\quad = \frac{\Delta(n_{di} + \alpha)}{\Delta\left(n_{\mathrm{di},-i} + \alpha\right)}\\
&\quad = \frac{\Gamma(\sum_{k=1}^{K}(n_{di,-i}^k + \alpha_k))}{\prod_{k=1}^{K}\Gamma\left(n_{di,-i}^k + \alpha_k\right)} \cdot \frac{\prod_{k=1}^{K}\Gamma\left(n_{di}^k + \alpha_k\right)}{\Gamma(\sum_{k=1}^{K}(n_{di}^k + \alpha_k))} \ \text{(利用 } \Gamma(a+1) = a\Gamma(a), \text{及} n_{di}^k = n_{di,-i}^k + 1)\\
&\quad = \frac{n_{di,-i}^k + \alpha_k}{\sum_{k=1}^{K}(n_{di,-i}^k + \alpha_k)}
\end{aligned}
$$

- 2.3 联合 2.1，2.2 从而有

$$p(z_i = k|\boldsymbol{z}_{-i}, \mathbf{w};\alpha,\beta) \propto p(\boldsymbol{w}|z_i = k, \boldsymbol{z}_{-i};\alpha,\beta) \cdot p(z_i = k, \boldsymbol{z}_{-i};\alpha,\beta) = \frac{n_{k,-i}^t + \beta_t}{\sum_{t=1}^{V}(n_{k,-i}^t + \beta_t)} \cdot \frac{n_{di,-i}^k + \alpha_k}{\sum_{k=1}^{K}(n_{di,-i}^k + \alpha_k)}$$