Gibbs Sampling Derivation for LDA and TOT

Hewlett-Packard Laboratories, China
Han Xiao, Ping Luo
January 19, 2009

A. Notations

整个 corpus 中包含 D 个 document ，有 T 个 topics，且 vocabulary 的大小为 V；
粗体的 **w,z,t** 分别泛指整个 corpus 中的 word, topic 和 timestamp；
小写的 d 代表 document 的索引，表示某一篇 document, d=1 to D
N_d表示第 d 个 document 中 word 的数量
小写的 i 表示 word 的索引，i=1 to N_d，与 d 合用可以组成$w_{di}, z_{di}, d_{di}, t_{di}$：

- w_{di}表示第 d 个 document 中第 i 个 word,
- z_{di}表示第 d 个 document 中第 i 个 word 生成自哪个 topic,
- d_{di}表示第 d 个 document 中第 i 个 word 来自哪个 document,
- t_{di}表示第 d 个 document 中第 i 个 word 的 time stamp

注意对于不同的 i,j，w_{dj}, w_{di}可以指示相同的 word，也可以指示不同的 word，对于z_{di}, t_{di}, d_{di}也同理。对于整个 corpus，一共有 D*N_d个 word，也就对应有 D*N_d个 topic 和 D*N_d个 timestamp。但整个 corpus 中实际只有 V 个无重复的 word，T 个无重复的 topic，因此我们用：

- 小写的 z 表示无重复的 topic 索引,z=1 to T
- 小写的 v 表示无重复的 word 索引, v=1 to V

在后面的推导过程中，会用到

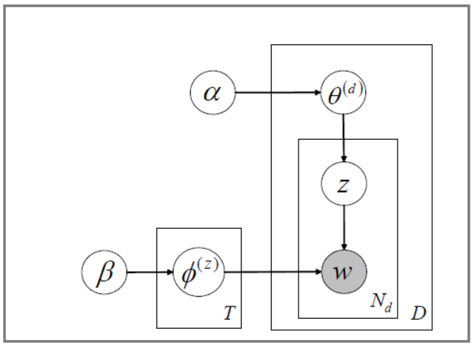
- $n_{z,v}$表示第 z 个 topic 和第 v 个 word 的共现次数（即 word v 多少次是由 topic z 生成的）。
- $n_{d,z}$表示第 d 个 document 和第 z 个 topic 的共现次数（即 topic z 在 document d 中出现了多少次）

在 LDA 中，每个 document 都有一个在所有 topics 上的分布，每个 topic 都有一个在整个 vocabulary 上的分布。对于 TOT 每个 topic 还有一个 beta 分布来表示时间。这里用大写的Ψ表示 beta 分布集，Φ, Θ表示多项分布集，其中每一个分布用如下小写符号表示：

- $\phi_{z_{di}}$表示z_{di}所对应的 topic-words 多项式分布
- ϕ_z表示第 z 个 topic 对应的 topic-words 多项式分布
- β为控制 topic-words 多项式分布的 dirichlet 先验分布参数，由dirichlet(β)抽出的都是 topic-words 多项式分布，即ϕ_z。在后面的推导中，为了不失一般化，使用的是非对称的 diricihlet 分布，参数用β_v表示，一共有 V 个β。
- θ_d表示第 d 个 document 的 document-topics 多项分布
- α为控制 document-topics 多项式分布的 dirichlet 先验分布参数，由dirichlet(α)抽出的都是 document-topics 多项式分布，即θ_d。同样的，对于非对称的 dirichlet 分布，参数用α_z表示，一共有 T 个α
- $\psi_{z_{di}}$表示z_{di}所对应的 beta 分布
- $\psi^1_{z_{di}}$和$\psi^2_{z_{di}}$表示z_{di}所对应的 beta 分布的两个参数

B. Generating Process

LDA(Latent Dirichlet Allocation, D.Blei etc., 2003)的 Graphical Model 如下：

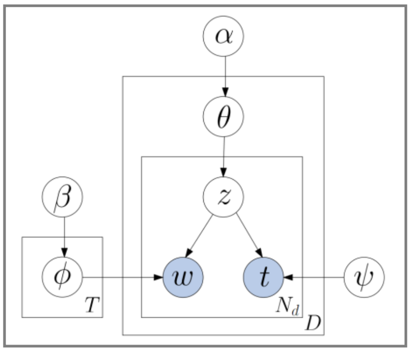


其生成过程为：

0. 设定 T，$\alpha_1,\dots,\alpha_T$，$\beta_1,\dots,\beta_V$
1. 对于每一个 topic，从参数为β的 Dirichlet 先验分布中抽样，并作为 1 个多项分布ϕ_z；重复 T 次
2. 对于每一个 document，从参数为α的 Dirichlet 先验分布中抽样，并作为 1 个多项分布θ_d，为了生成这个 document 中的每一个 word w_{di}：
    a) 从多项分布θ_d中抽取一个 topic z_{di}
    b) 从多项分布$\phi_{z_{di}}$中抽取一个 word w_{di}
    c) 重复(a,b) N_d次
3. 重复步骤 2 D 次，直到整个 corpus 生成完成。

LDA 的目标：根据已有的 corpus(每个 word 是观察到的)重建θ_d和$\phi_{z_{di}}$，难点：每个 word w_{di} 对应的 topic z_{di}是未知的（z 是隐藏变量）

TOT(Topics over Time: A Non-Markov Continuous-Time Model of Topical Trends, Xuerui Wang etc., 2006)的 Graphical Model 如下:



其目的在于描述一个 corpus 中随时间变化 topic。其生成过程为：

0. 设定 T，$\alpha_1,\dots,\alpha_T$，$\beta_1,\dots,\beta_V$
1. 对于每一个 topic，从参数为β的 Dirichlet 先验分布中抽样，并作为 1 个多项分布ϕ_z；重复 T 次
2. 对于每一个 document，从参数为α的 Dirichlet 先验分布中抽样，并作为 1 个多项分布θ_d，为了生成这个 document 中的每一个 word w_{di}：
    a) 从多项分布θ_d中抽取一个 topic z_{di}
    b) 从多项分布$\phi_{z_{di}}$中抽取一个 word w_{di}
    c) 从 beta 分布$\psi_{z_{di}}$中抽取一个 timestamp t_{di}
    d) 重复(a,b,c) N_d次
3. 重复步骤 2 D 次，直到整个 corpus 生成完成。

TOT 的目标：根据已有的 corpus(每个 word 是观察到的，每个 word 上的 timestamp 用其所在的 document 的生成时间表示)重建$\theta_d\phi_{z_{di}}$和$\psi_{z_{di}}$，难点：每个 word w_{di} 对应的 topic z_{di}是未知的（z 是隐藏变量）

C. Derivation of Gibbs Sampling

Gibbs Sampling：为了对 X 进行估计，一般我们要从

$$P(\mathbf{X}) \equiv P(x_1, x_2, \dots, x_N)$$

中进行抽样。如果$P(\mathbf{X})$不易求得，我们可以通过对所有的$P(x_i|\mathbf{X}_{-i})$进行抽样来近似$\widehat{\mathbf{X}}$

其步骤如下：

1. 随机初始化$\mathbf{X}^{(0)} = \{x_1^{(0)}, x_2^{(0)}, \dots, x_N^{(0)}\}$
2. 重复进行 T 轮抽样，t=1,...,T

   在每轮抽样中，对于 i=1,...,N，每个$x_i^{(t)}$从$P\left(x_i^{(t)}|\mathbf{X}_1^{(t)}, \dots, \mathbf{X}_{i-1}^{(t)}, \mathbf{X}_{i+1}^{(t-1)}, \dots, \mathbf{X}_N^{(t)}\right)$抽样
3. 当 Burn-in 之后，可以通过几轮抽样计算$\mathbf{P}(\mathbf{X})$

为了不失一般性，下面对 TOT 的 Gibbs Sampling 过程进行推导

1. 在 TOT 的 Gibbs Sampling 中，我们要求

$$P(z_{di}|\mathbf{w}, \mathbf{t}, \mathbf{z}_{-\mathbf{di}}, \alpha, \beta, \Psi)$$

然后才能跟据它，对生成w_{di}, t_{di}的z_{di}进行抽样估计。因为z_{di}是隐藏变量，一旦抽样估计完成，对于每个w_{di}它生成自的 topic 就变成已知；对于每个 document，它包含的 topics 也变成已知。那么对于 document-topics 分布θ_d和 topic-words 上的分布ϕ_z也就可以非常容易的被拟合出来。

Step 1：根据

$$P(z_{di}, \mathbf{z}_{-\mathbf{di}}) = P(\mathbf{z})$$

和贝叶斯公式可以得到

$$P(z_{di}|\mathbf{w}, \mathbf{t}, \mathbf{z}_{-\mathbf{di}}, \alpha, \beta, \Psi) = \frac{P(\mathbf{w}, \mathbf{t}, \mathbf{z}, \alpha, \beta, \Psi)}{P(\mathbf{w}, \mathbf{t}, \mathbf{z}_{-\mathbf{di}}, \alpha, \beta, \Psi)} = \frac{P(\mathbf{w}, \mathbf{t}, \mathbf{z}|\alpha, \beta, \Psi)}{P(\mathbf{w}, \mathbf{t}, \mathbf{z}_{-\mathbf{di}}|\alpha, \beta, \Psi)}$$

根据 Graphical Model，w_{di}, t_{di}都是由z_{di}生成的，如果不考虑z_{di}则无法考虑w_{di}, t_{di}。从而得到：

$$P(z_{di}|\mathbf{w}, \mathbf{t}, \mathbf{z}_{-\mathbf{di}}, \alpha, \beta, \Psi) \propto \frac{P(\mathbf{z}, \mathbf{w}, \mathbf{t}|\alpha, \beta, \Psi)}{P(\mathbf{z}_{-\mathbf{di}}, \mathbf{w}_{-\mathbf{di}}, \mathbf{t}_{-\mathbf{di}}|\alpha, \beta, \Psi)}$$

2. 由上式可知，在 Gibbs Sampling 中关键是要求出如下的联合概率

$$P(\mathbf{w}, \mathbf{t}, \mathbf{z}|\alpha, \beta, \Psi)$$

Step 1：根据 Graphical Model，略去Φ, Θ，可以将联合概率拆开

$$P(\mathbf{w}, \mathbf{t}, \mathbf{z}|\alpha, \beta, \Psi) = P(\mathbf{w}|\mathbf{z}, \beta)P(\mathbf{t}|\Psi, \mathbf{z})P(\mathbf{z}|\alpha)$$

Step 2：引入Φ, Θ，对Φ, Θ进行积分。再根据 Graphical Model，可以写出

$$P(\mathbf{w}, \mathbf{t}, \mathbf{z}|\alpha, \beta, \Psi) = P(\mathbf{t}|\Psi, \mathbf{z}) \int P(\mathbf{w}|\mathbf{z}, \Phi)P(\Phi|\beta)d\Phi \int P(\mathbf{z}|\Theta)P(\Theta|\alpha)d\Theta$$

Step 3：对于整个 corpus，拆开所有黑体和大写，条件概率中的条件$\Psi, \mathbf{z}$可以写做$\psi_{z_{di}}$；$\mathbf{z}, \Phi$写做$\phi_{z_{di}}$

$$= \prod_{d=1}^{D}\prod_{i=1}^{N_d} P(t_{di}|\psi_{z_{di}}) \int \prod_{d=1}^{D}\prod_{i=1}^{N_d} P(w_{di}|\phi_{z_{di}}) \prod_{z=1}^{T} P(\phi_z|\beta)\, d\phi_z \int \prod_{d=1}^{D}\prod_{i=1}^{N_d} P(z_{di}|\theta_d)P(\theta_d|\alpha)\, d\theta_d$$

Step 4：由于从第z_{di}个 topic 中抽去w_{di}是满足多项式分布$\phi_{z_{di}}$的，因此

$$\prod_{d=1}^{D}\prod_{i=1}^{N_d} P(w_{di}|\phi_{z_{di}}) = \prod_{z=1}^{T}\prod_{v=1}^{V} \phi_{z,v}^{n_{z,v}}$$

同理由于从第 d 个 document 中抽取z_{di}也是满足多项分布θ_d的，因此

$$\prod_{d=1}^{D}\prod_{i=1}^{N_d} P(z_{di}|\theta_d) = \prod_{d=1}^{D}\prod_{z=1}^{T} \theta_{d,z}^{n_{d,z}}$$

将两式带入(2.3)中可以得到，

$$= \prod_{d=1}^{D}\prod_{i=1}^{N_d} P(t_{di}|\psi_{z_{di}}) \int \prod_{z=1}^{T}\prod_{v=1}^{V} \phi_{z,v}^{n_{z,v}} \prod_{z=1}^{T} P(\phi_z|\beta)\, d\phi_z \int \prod_{d=1}^{D}\prod_{z=1}^{T} \theta_{d,z}^{n_{d,z}} \prod_{d=1}^{D} P(\theta_d|\alpha)\, d\theta_d$$

Step 5: 根据 dirichlet 的后验分布，可以将$P(\phi_z|\beta)$和$P(\theta_d|\alpha)$展开，得到

$$= \prod_{d=1}^{D}\prod_{i=1}^{N_d} P(t_{di}|\psi_{z_{di}}) \int \prod_{z=1}^{T}\prod_{v=1}^{V} \phi_{z,v}^{n_{z,v}} \prod_{z=1}^{T} \left(\frac{\Gamma(\sum_{v=1}^{V}\beta_v)}{\prod_{v=1}^{V}\Gamma(\beta_v)} \prod_{v=1}^{V} \phi_{z,v}^{\beta_v-1}\right) d\phi_z \int \prod_{d=1}^{D}\prod_{z=1}^{T} \theta_{d,z}^{n_{d,z}} \prod_{d=1}^{D} \left(\frac{\Gamma(\sum_{z=1}^{T}\alpha_z)}{\prod_{z=1}^{T}\Gamma(\alpha_z)} \prod_{z=1}^{T} \theta_{d,z}^{\alpha_z-1}\right) d\theta_d$$

Step 6: 由于$\frac{\Gamma(\sum_{v=1}^{V}\beta_v)}{\prod_{v=1}^{V}\Gamma(\beta_v)}$与$\phi$无关，$\frac{\Gamma(\sum_{z=1}^{T}\alpha_z)}{\prod_{z=1}^{T}\Gamma(\alpha_z)}$与$\theta$无关，可以将它们提出，得：

$$= \prod_{d=1}^{D}\prod_{i=1}^{N_d} P(t_{di}|\psi_{z_{di}}) \left(\frac{\Gamma(\sum_{v=1}^{V}\beta_v)}{\prod_{v=1}^{V}\Gamma(\beta_v)}\right)^T \left(\frac{\Gamma(\sum_{z=1}^{T}\alpha_z)}{\prod_{z=1}^{T}\Gamma(\alpha_z)}\right)^D \int \prod_{z=1}^{T}\prod_{v=1}^{V} \phi_{z,v}^{n_{z,v}} \prod_{z=1}^{T} \left(\prod_{v=1}^{V} \phi_{z,v}^{\beta_v-1}\right) d\phi_z \int \prod_{d=1}^{D}\prod_{z=1}^{T} \theta_{d,z}^{n_{d,z}} \prod_{d=1}^{D} \left(\prod_{z=1}^{T} \theta_{d,z}^{\alpha_z-1}\right) d\theta_d$$

Step 7: 由于不同的 topic 的 topic-words 分布是独立的（比如ϕ_1与ϕ_2是独立的，可以通过 d-separation 判定），因此连乘的积分可以写作积分的连乘；同理，不同 document 的 document-topics 分布也是独立的（θ_1与θ_2是独立的），因此可以上式可以写为：

$$= \prod_{d=1}^{D}\prod_{i=1}^{N_d} P(t_{di}|\psi_{z_{di}}) \left(\frac{\Gamma(\sum_{v=1}^{V}\beta_v)}{\prod_{v=1}^{V}\Gamma(\beta_v)}\right)^T \left(\frac{\Gamma(\sum_{z=1}^{T}\alpha_z)}{\prod_{z=1}^{T}\Gamma(\alpha_z)}\right)^D \prod_{z=1}^{T}\left(\int \prod_{v=1}^{V} \phi_{z,v}^{n_{z,v}+\beta_v-1}\, d\phi_z\right) \prod_{d=1}^{D}\left(\int \prod_{z=1}^{T} \theta_{d,z}^{n_{d,z}+\alpha_z-1}\, d\theta_d\right)$$

Step 8: 根据欧拉积分

$$\int_{\sum x_i=1} \prod_{i=1}^{N} x_i^{a_i-1}\, d^N\vec{x} = \frac{\prod_{i=1}^{N}\Gamma(a_i)}{\Gamma(\sum_{i=1}^{N} a_i)}$$

对(2.7)式中后面两项使用欧拉积分可得

$$P(\mathbf{w},\mathbf{t},\mathbf{z}|\alpha,\beta,\Psi) = \prod_{d=1}^{D}\prod_{i=1}^{N_d} P(t_{di}|\psi_{z_{di}}) \left(\frac{\Gamma(\sum_{v=1}^{V}\beta_v)}{\prod_{v=1}^{V}\Gamma(\beta_v)}\right)^T \left(\frac{\Gamma(\sum_{z=1}^{T}\alpha_z)}{\prod_{z=1}^{T}\Gamma(\alpha_z)}\right)^D \prod_{z=1}^{T} \frac{\prod_{v=1}^{V}\Gamma(n_{z,v}+\beta_v)}{\Gamma(\sum_{v=1}^{V} n_{z,v}+\beta_v)} \prod_{d=1}^{D} \frac{\prod_{z=1}^{T}\Gamma(n_{d,z}+\alpha_z)}{\Gamma(\sum_{z=1}^{T}(n_{d,z}+\alpha_z))}$$

3. 求 full conditional probability

Step 1: 将(2.8)中的式子代入，可以得到

$$\frac{P(\mathbf{z},\mathbf{w},\mathbf{t}|\alpha,\beta,\Psi)}{P(\mathbf{z}_{-di},\mathbf{w}_{-di},\mathbf{t}_{-di}|\alpha,\beta,\Psi)}$$

$$= \frac{\prod_{d=1}^{D}\prod_{i=1}^{N_d} P(t_{di}|\psi_{z_{di}}) \left(\frac{\Gamma(\sum_{v=1}^{V}\beta_v)}{\prod_{v=1}^{V}\Gamma(\beta_v)}\right)^T \left(\frac{\Gamma(\sum_{z=1}^{T}\alpha_z)}{\prod_{z=1}^{T}\Gamma(\alpha_z)}\right)^D \prod_{z=1}^{T} \frac{\prod_{v=1}^{V}\Gamma(n_{z,v}+\beta_v)}{\Gamma(\sum_{v=1}^{V} n_{z,v}+\beta_v)} \prod_{d=1}^{D} \frac{\prod_{z=1}^{T}\Gamma(n_{d,z}+\alpha_z)}{\Gamma(\sum_{z=1}^{T}(n_{d,z}+\alpha_z))}}{\left[\prod_{d=1}^{D}\prod_{i=1}^{N_d} P(t_{di}|\psi_{z_{di}}) \left(\frac{\Gamma(\sum_{v=1}^{V}\beta_v)}{\prod_{v=1}^{V}\Gamma(\beta_v)}\right)^T \left(\frac{\Gamma(\sum_{z=1}^{T}\alpha_z)}{\prod_{z=1}^{T}\Gamma(\alpha_z)}\right)^D \prod_{z=1}^{T} \frac{\prod_{v=1}^{V}\Gamma(n_{z,v}+\beta_v)}{\Gamma(\sum_{v=1}^{V} n_{z,v}+\beta_v)} \prod_{d=1}^{D} \frac{\prod_{z=1}^{T}\Gamma(n_{d,z}+\alpha_z)}{\Gamma(\sum_{z=1}^{T}(n_{d,z}+\alpha_z))}\right]_{-di}}$$

Step 2: 此时，要留意所有角标与 di 有关的$n_{d,z}$和$n_{z,v}$，由于不考虑z_{di}，因此w_{di} t_{di}不用考虑（因为它们没有被生成），也就是说在考虑 topic z 和 word v 的所有共现次数时(即$n_{z,v}$)，我们忽略了这一次z_{di}与w_{di}的共现，这仅仅会让$n_{z_{di},w_{di}}$减 1，而对于其他的$n_{z,w_{di}}$并无影响；同样的，这也会使$n_{d_{di},z_{di}}$减 1，而并不会使其他的$n_{d,z_{di}}$发生改变。注意：V,T,D 的大小并没有发生改变

将上式拆成 3 部分来看，可以得到如下三式：

$$1)\quad \frac{\prod_{d=1}^{D}\prod_{i=1}^{N_d} P(t_{di}|\psi_{z_{di}})}{\left[\prod_{d=1}^{D}\prod_{i=1}^{N_d} P(t_{di}|\psi_{z_{di}})\right]_{-di}} = P(t_{di}|\psi_{z_{di}})$$

$$2)\quad \frac{\left(\frac{\Gamma\left(\sum_{v=1}^{V}\beta_v\right)}{\prod_{v=1}^{V}\Gamma(\beta_v)}\right)^T\left(\frac{\Gamma\left(\sum_{z=1}^{T}\alpha_z\right)}{\prod_{z=1}^{T}\Gamma(\alpha_z)}\right)^D}{\left[\left(\frac{\Gamma\left(\sum_{v=1}^{V}\beta_v\right)}{\prod_{v=1}^{V}\Gamma(\beta_v)}\right)^T\left(\frac{\Gamma\left(\sum_{z=1}^{T}\alpha_z\right)}{\prod_{z=1}^{T}\Gamma(\alpha_z)}\right)^D\right]_{-di}} = 1$$

$$3)\quad \frac{\prod_{z=1}^{T}\frac{\prod_{v=1}^{V}\Gamma(n_{z,v}+\beta_v)}{\Gamma\left(\sum_{v=1}^{V}(n_{z,v}+\beta_v)\right)}\prod_{d=1}^{D}\frac{\prod_{z=1}^{T}\Gamma(n_{d,z}+\alpha_z)}{\Gamma\left(\sum_{z=1}^{T}(n_{d,z}+\alpha_z)\right)}}{\left[\prod_{z=1}^{T}\frac{\prod_{v=1}^{V}\Gamma(n_{z,v}+\beta_v)}{\Gamma\left(\sum_{v=1}^{V}(n_{z,v}+\beta_v)\right)}\prod_{d=1}^{D}\frac{\prod_{z=1}^{T}\Gamma(n_{d,z}+\alpha_z)}{\Gamma\left(\sum_{z=1}^{T}(n_{d,z}+\alpha_z)\right)}\right]_{-di}}$$

Step 3: 关键是对(3.2.3 式的化简)，首先看其前一部分的分母

$$\left[\prod_{z=1}^{T}\frac{\prod_{v=1}^{V}\Gamma(n_{z,v}+\beta_v)}{\Gamma\left(\sum_{v=1}^{V}(n_{z,v}+\beta_v)\right)}\right]_{-di} = \frac{\Gamma\left(n_{z_{di},w_{di}}+\beta_{w_{di}}-1\right)\times\left[\prod_{z=1}^{T}\prod_{v=1}^{V}\Gamma(n_{z,v}+\beta_v)\right]_{-z_{di},w_{di}}}{\Gamma\left(\sum_{v=1}^{V}(n_{z_{di},v}+\beta_v)-1\right)\times\prod_{z\neq z_{di}}^{T}\Gamma\left(\sum_{v=1}^{V}(n_{z,v}+\beta_v)\right)}$$

Step 4: 看前一部分的分子

$$\prod_{z=1}^{T}\frac{\prod_{v=1}^{V}\Gamma(n_{z,v}+\beta_v)}{\Gamma\left(\sum_{v=1}^{V}(n_{z,v}+\beta_v)\right)} = \frac{\Gamma\left(n_{z_{di},w_{di}}+\beta_{w_{di}}\right)\times\left[\prod_{z=1}^{T}\prod_{v=1}^{V}\Gamma(n_{z,v}+\beta_v)\right]_{-z_{di},w_{di}}}{\Gamma\left(\sum_{v=1}^{V}(n_{z_{di},v}+\beta_v)\right)\times\prod_{z\neq z_{di}}^{T}\Gamma\left(\sum_{v=1}^{V}(n_{z,v}+\beta_v)\right)}$$

Step 5: (3.4)/(3.3)，借助$\Gamma(x+1)=x\Gamma(x)$可以得到

$$\frac{\prod_{z=1}^{T}\frac{\prod_{v=1}^{V}\Gamma(n_{z,v}+\beta_v)}{\Gamma\left(\sum_{v=1}^{V}(n_{z,v}+\beta_v)\right)}}{\left[\prod_{z=1}^{T}\frac{\prod_{v=1}^{V}\Gamma(n_{z,v}+\beta_v)}{\Gamma\left(\sum_{v=1}^{V}(n_{z,v}+\beta_v)\right)}\right]_{-di}} = \frac{n_{z_{di},w_{di}}+\beta_{w_{di}}-1}{\sum_{v=1}^{V}(n_{z_{di},v}+\beta_v)-1}$$

Step 6: 看后一部分的分母

$$\left[\prod_{d=1}^{D}\frac{\prod_{z=1}^{T}\Gamma(n_{d,z}+\alpha_z)}{\Gamma\left(\sum_{z=1}^{T}(n_{d,z}+\alpha_z)\right)}\right]_{-di} = \frac{\Gamma\left(n_{d_{di},z_{di}}+\alpha_{z_{di}}-1\right)\times\left[\prod_{d=1}^{D}\prod_{z=1}^{T}\Gamma(n_{d,z}+\alpha_z)\right]_{-d_{di},z_{di}}}{\Gamma\left(\sum_{z=1}^{T}(n_{d_{di},z}+\alpha_z)-1\right)\times\prod_{d\neq d_{di}}^{D}\Gamma\left(\sum_{z=1}^{T}(n_{d,z}+\alpha_z)\right)}$$

Step 7: 看后一部分的分子：

$$\prod_{d=1}^{D}\frac{\prod_{z=1}^{T}\Gamma(n_{d,z}+\alpha_z)}{\Gamma\left(\sum_{z=1}^{T}(n_{d,z}+\alpha_z)\right)} = \frac{\Gamma\left(n_{d_{di},z_{di}}+\alpha_{z_{di}}\right)\times\left[\prod_{d=1}^{D}\prod_{z=1}^{T}\Gamma(n_{d,z}+\alpha_z)\right]_{-d_{di},z_{di}}}{\Gamma\left(\sum_{z=1}^{T}(n_{d_{di},z}+\alpha_z)\right)\times\prod_{d\neq d_{di}}^{D}\Gamma\left(\sum_{z=1}^{T}(n_{d,z}+\alpha_z)\right)}$$

Step 8: (3.7)/(3.6)，借助$\Gamma(x+1)=x\Gamma(x)$可得：

$$\frac{\prod_{d=1}^{D}\frac{\prod_{z=1}^{T}\Gamma(n_{d,z}+\alpha_z)}{\Gamma\left(\sum_{z=1}^{T}(n_{d,z}+\alpha_z)\right)}}{\left[\prod_{d=1}^{D}\frac{\prod_{z=1}^{T}\Gamma(n_{d,z}+\alpha_z)}{\Gamma\left(\sum_{z=1}^{T}(n_{d,z}+\alpha_z)\right)}\right]_{-di}} = \frac{n_{d_{di},z_{di}}+\alpha_{z_{di}}-1}{\sum_{z=1}^{T}(n_{d_{di},z}+\alpha_z)-1}$$

Step 9: 将(3.8)(3.5)(3.2)代入到(3.1)可得

$$P(z_{di}|\mathbf{w},\mathbf{t},\mathbf{z_{-di}},\alpha,\beta,\Psi) \propto \frac{P(\mathbf{z},\mathbf{w},\mathbf{t}|\alpha,\beta,\Psi)}{P(\mathbf{z_{-di}},\mathbf{w_{-di}},\mathbf{t_{-di}}|\alpha,\beta,\Psi)} = P(t_{di}|\psi_{z_{di}})\times\frac{n_{z_{di},w_{di}}+\beta_{w_{di}}-1}{\sum_{v=1}^{V}(n_{z_{di},v}+\beta_v)-1}\times\frac{n_{d_{di},z_{di}}+\alpha_{z_{di}}-1}{\sum_{z=1}^{T}(n_{d_{di},z}+\alpha_z)-1}$$

$$=\frac{(1-t_{di})^{\psi_{z_{di}}^{1}-1}t_{di}^{\psi_{z_{di}}^{2}-1}}{B(\psi_{z_{di}}^{1}-1,\psi_{z_{di}}^{2}-1)}\times\frac{n_{z_{di},w_{di}}+\beta_{w_{di}}-1}{\sum_{v=1}^{V}(n_{z_{di},v}+\beta_v)-1}\times\frac{n_{d_{di},z_{di}}+\alpha_{z_{di}}-1}{\sum_{z=1}^{T}(n_{d_{di},z}+\alpha_z)-1}$$

其中$B(\alpha,\beta)=\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$。我们可以利用上式对每个$z_{di}$进行抽样，当迭代次数足够大时，抽样结果趋于稳定。

LDA 中 Gibbs Sampling 中$P(z_{di}|\mathbf{w},\mathbf{z_{-di}},\alpha,\beta)\propto\frac{n_{z_{di},w_{di}}+\beta_{w_{di}}-1}{\sum_{v=1}^{V}(n_{z_{di},v}+\beta_v)-1}\times\frac{n_{d_{di},z_{di}}+\alpha_{z_{di}}-1}{\sum_{z=1}^{T}(n_{d_{di},z}+\alpha_z)-1}$

对比 LDA Gibbs Sampling 发现，模型中加入时间信息后，对z_{di}抽样时依据的概率分布做了一些改变，相当于在 LDA 的基础上增加了一个因子。

当对z_{di}的抽样完成后，我们可以方便的根据$n_{z,v}$和$n_{d,z}$对ϕ_z与θ_d进行估计

例如，对于第 z 个 topic 的 topic-words 分布，可以用：

$$\phi_z^v = \frac{n_{z,v} + \beta_v - 1}{\sum_{v=1}^{V}\left(n_{z,v} + \beta_v\right) - 1}$$

求得$\phi_z = (\phi_z^1, \dots, \phi_z^V)$

D. Quick Derivation

对于 TOT 的 Gibbs Sampling，我们要求的是

$$P(z_{di}|\mathbf{w}, \mathbf{t}, \mathbf{z}_{-\mathbf{di}}, \alpha, \beta, \Psi)$$

根据 Bayes Rule，上式可以写作：

$$P(z_{di}|\mathbf{w}, \mathbf{t}, \mathbf{z}_{-\mathbf{di}}, \alpha, \beta, \Psi) = \frac{P(\mathbf{w}, \mathbf{t}, \mathbf{z}, \alpha, \beta, \Psi)}{P(\mathbf{w}, \mathbf{t}, \mathbf{z}_{-\mathbf{di}}, \alpha, \beta, \Psi)}$$

根据 Graphical Model，w_{di}, t_{di}都是由z_{di}生成的，如果不考虑z_{di}则无法考虑w_{di}, t_{di}，因此：

$$P(z_{di}|\mathbf{w}, \mathbf{t}, \mathbf{z}_{-\mathbf{di}}, \alpha, \beta, \Psi) \propto \frac{P(\mathbf{w}, \mathbf{t}, \mathbf{z}, \alpha, \beta, \Psi)}{P(\mathbf{z}_{-\mathbf{di}}, \mathbf{w}_{-\mathbf{di}}, \mathbf{t}_{-\mathbf{di}}, \alpha, \beta, \Psi)} = \frac{P(\mathbf{w}, \mathbf{t}, \mathbf{z}|\alpha, \beta, \Psi)}{P(\mathbf{z}_{-\mathbf{di}}, \mathbf{w}_{-\mathbf{di}}, \mathbf{t}_{-\mathbf{di}}|\alpha, \beta, \Psi)}$$

根据 d-sperate 可以判定对于不同的 i,j: z_{di}, w_{di}, t_{di}与z_{dj}, w_{dj}, t_{dj} 在条件α, β, Ψ下是独立的。因此，上式右边可以拆为各项连乘的形式，分子分母消去公共项，只剩$P(z_{di}, w_{di}, t_{di}|\alpha, \beta, \Psi)$，即

$$P(z_{di}|\mathbf{w}, \mathbf{t}, \mathbf{z}_{-\mathbf{di}}, \alpha, \beta, \Psi) \propto P(z_{di}, w_{di}, t_{di}|\alpha, \beta, \Psi)$$

又由于z_{di}, w_{di}, t_{di}与z_{dj}, w_{dj}, t_{dj}实际上在条件$\Phi, \Theta, \alpha, \beta, \Psi$下也是独立的，同理可得

$$P(z_{di}|\mathbf{w}, \mathbf{t}, \mathbf{z}_{-\mathbf{di}}, \Phi, \Theta, \alpha, \beta, \Psi) \propto P(z_{di}, w_{di}, t_{di}|\Phi, \Theta, \alpha, \beta, \Psi)$$

根据 Graphical Model $P\left(z_{dj}, w_{dj}, t_{dj}\middle|\Phi, \Theta, \alpha, \beta, \Psi\right)$可以写作：

$$P(z_{di}, w_{di}, t_{di}|\Phi, \Theta, \alpha, \beta, \Psi) = P(w_{di}|z_{di}, \Phi, \beta)P(t_{di}|z_{di}, \Psi)P(z_{di}|\Theta, \alpha)$$

由于抽w_{di}和抽z_{di}都满足多项式分布，抽t_{di}满足 beta 分布，于是有：

$$P(w_{di}|z_{di}, \Phi, \beta) = \frac{n_{z_{di}, w_{di}} + \beta_{w_{di}}}{\sum_{v=1}^{V}\left(n_{z_{di}, v} + \beta_v\right)}$$

$$P(z_{di}|\Theta, \alpha) = \frac{n_{d_{di}, z_{di}} + \alpha_{z_{di}}}{\sum_{z=1}^{T}\left(n_{d_{di}, z} + \alpha_z\right)}$$

$$P(t_{di}|z_{di}, \Psi) = \frac{(1 - t_{di})^{\psi_{z_{di}}^1 - 1} t_{di}^{\psi_{z_{di}}^2 - 1}}{B(\psi_{z_{di}}^1 - 1, \psi_{z_{di}}^2 - 1)}$$

于是得到$P(z_{di}|\mathbf{w}, \mathbf{t}, \mathbf{z}_{-\mathbf{di}}, \alpha, \beta, \Psi) \propto \frac{(1-t_{di})^{\psi_{z_{di}}^1 - 1} t_{di}^{\psi_{z_{di}}^2 - 1}}{B(\psi_{z_{di}}^1 - 1, \psi_{z_{di}}^2 - 1)} \times \frac{n_{z_{di}, w_{di}} + \beta_{w_{di}}}{\sum_{v=1}^{V}\left(n_{z_{di}, v} + \beta_v\right)} \times \frac{n_{d_{di}, z_{di}} + \alpha_{z_{di}}}{\sum_{z=1}^{T}\left(n_{d_{di}, z} + \alpha_z\right)}$

对于 LDA，推导过程同上，只是我们无需考虑t_{di}，于是$P(z_{di}|\mathbf{w}, \mathbf{z}_{-\mathbf{di}}, \alpha, \beta, \Psi) = \frac{n_{z_{di}, w_{di}} + \beta_{w_{di}}}{\sum_{v=1}^{V}\left(n_{z_{di}, v} + \beta_v\right)} \times \frac{n_{d_{di}, z_{di}} + \alpha_{z_{di}}}{\sum_{z=1}^{T}\left(n_{d_{di}, z} + \alpha_z\right)}$

E. References

[1] D. Blei, A. Ng, and M. Jordan. Latent Dirichlet allocation. Journal of Machine Learning Research, 3:993–1022, 2003.

[2] Xuerui Wang, Andrew McCallum. Topics over Time: A Non-Markov Continuous-Time Model of Topical Trends. KDD'06 August 20–23, 2006, Philadelphia, Pennsylvania, USA