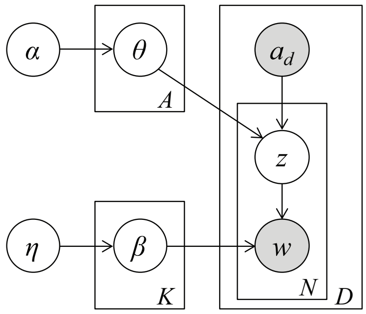

Figure 1: TATM2 in plate notation

1 Tweet Author-Topic Model (with per-word topics)

The model is shown in Figure 1. The generative process is as follows:

1. For $k \in 1, \ldots, K$:
   - Draw topic $\beta_k \sim \mathrm{Dir}_V(\eta)$.
2. For $a \in 1, \ldots, A$:
   - Draw author's topic proportions $\theta_a \sim \mathrm{Dir}_K(\alpha)$.
3. For each document (tweet) $d \in 1, \ldots, D$:
   - For each word $w \in 1, \ldots, N$:
     - Draw topic assignment $z_{d,n} \sim \mathrm{Mult}_K(\theta_a)$.
     - Draw word $w_{d,n} \sim \mathrm{Mult}_V(\beta_{z_{d,n}})$.

Given the hyperparameters α and η, the joint distribution of topics β, author-topic mixture θ, topic assignments $\mathbf{z}$ and words $\mathbf{w}$ is given by:

$$p(\theta, \beta, \mathbf{z}, \mathbf{w}|\alpha, \eta) = p(\theta|\alpha)p(\beta|\eta)\prod_{n=1}^{N} p(z_n|\theta)p(w_n|z_n, \beta_{z_{d,n}}) \tag{1}$$

1.1 Complete conditionals

First, we denote θ_a to be the topic proportions of the author a of document d, i.e. $\theta_a = \theta_{a_d}$.

Local hidden variables. The complete conditional of the topic assignment $z_{d,n}$ is a multinomial,

$$p(z_{d,n} = k|\theta_a, \beta, w_{d,n}) \propto \exp\{\log \theta_{a,k} + \log \beta_{k,w_{d,n}}\} \tag{2}$$

Global hidden variables. In contrast with standard LDA, the topic proportions are now a global hidden variable. The complete conditional of the author-topic proportions is a posterior Dirichlet,

$$p(\theta_a|\beta, z_d) \propto \mathrm{Dir}(\alpha + \sum_{d \in D_a} \sum_{n=1}^{N} z_{d,n}^k), \tag{3}$$

where D_a is the set of documents whose author is a.

Finally, the complete conditional for the topic β_k is also a posterior Dirichlet,

$$p(\beta_k|\mathbf{z}, \mathbf{w}) \propto \mathrm{Dir}(\eta + \sum_{d=1}^{D} \sum_{n=1}^{N} z_{d,n}^k w_{d,n}) \tag{4}$$

1.2 Variational parameters for batch variational inference

The variational parameters are:

- Global per-topic Dirichlets $\lambda_{1:K}$
- Global per-author Dirichlets $\gamma_{1:A}$
- Local per-word multinomials $\phi_{1:D,1:N}$

Each update of the local variables is defined as

$$\phi_{d,n} \propto \exp\{\Psi(\gamma_a) + \Psi(\lambda_{\cdot,w_{d,n}}) - \Psi(\sum_v \lambda_{\cdot,w_{d,n}})\} \tag{5}$$

Each update of the global variational Dirichlets is defined as

$$\gamma_a = \alpha + \sum_{d \in D_a} \sum_{n=1}^{N} \phi_{d,n} \tag{6}$$

$$\lambda_k = \eta + \sum_{d=1}^{D_a} \sum_{n=1}^{N} \phi_{d,n}^k w_{d,n} \tag{7}$$

1.3 Stochastic variational inference

Each update of the local variables is defined as

$$\phi_{d,n}^k \propto \exp\{\mathbb{E}[\log \theta_{a,k}] + \mathbb{E}[\log \beta_{k,w_{d,n}}]\} \tag{8}$$

During the updates of local variables, we utilize an intermediate local variable $\hat{\gamma}_d$, which serves as an indicator of convergence of ϕ_d during the inner loop of the E-step. Specifically,

$$\hat{\gamma}_d = \sum_{n=1}^{N} \phi_{d,n} \tag{9}$$

After fitting the local variables, we set the intermediate topics as

$$\hat{\lambda}_k = \eta + D \sum_{d=1}^{D_a} \sum_{n=1}^{N} \phi_{d,n}^k w_{d,n} \tag{10}$$

Finally, the global topics and the global per-author distributions are updated as

$$\lambda_k^{t+1} = (1 - \rho_t)\lambda^t + \rho_t \hat{\lambda}_k \tag{11}$$

$$\gamma_a^{t+1} = \gamma_a^t + \hat{\gamma}_d \tag{12}$$