Cluster LDA

March 7, 2018

1 Standard Generative Process

For this reduced case of IPTM, we assume that the document-cluster assignments $c_d \in \{1, \ldots, C\}$ for $d \in [D]$ are known and fixed. Following the cluster-based topic model, document d has the document-topic distribution

$$\boldsymbol{\theta}_d \sim \text{Dirichlet}(\alpha, \boldsymbol{\xi}_{c_d}), \tag{1}$$

where α are the concentration parameter and $\boldsymbol{\xi}_{c_d}$ is the base measure corresponding to the cluster of document d. In order to capture the overall prevalence of each topic in the corpus, we assume that each $\boldsymbol{\xi}_c$ is given Dirichlet priors with a single corpus-level base measure $\boldsymbol{m}$

$$\boldsymbol{\xi}_c \sim \text{Dirichlet}\Big(\alpha_1, \boldsymbol{m}\Big), \tag{2}$$

where α_1 is the concentration parameter determining the extent to which the group-specific base measures are affected by the corpus-level base measure. Finally, the corpus-level base measure is assumed to have Dirichlet prior with uniform base measure

$$\boldsymbol{m} \sim \text{Dirichlet}\Big(\alpha_0, (\frac{1}{K}, \ldots, \frac{1}{K})\Big). \tag{3}$$

Given that $\bar{N}_d = \max(1, N_d)$ where N_d is known, a topic z_{dn} is drawn from the document-topic distribution for each $n \in [\bar{N}_d]$—i.e.,

$$z_{dn} \sim \text{Multinomial}(\boldsymbol{\theta}_d). \tag{4}$$

2   Current Derivatoin

The predictive probability

$$\Pr(z_{N_d+1} = k | \boldsymbol{z}, \alpha, \boldsymbol{\xi}_{c_d}) = \int_{\boldsymbol{\theta}_d} \Pr(k|\boldsymbol{\theta}_d) \times \Pr(\boldsymbol{\theta}_d | \boldsymbol{z}, \alpha, \boldsymbol{\xi}_{c_d}) d\boldsymbol{\theta}_d$$

$$= \int_{\boldsymbol{\theta}_d} \prod_{k=1}^{K} (\theta_{dk}) \times \Big(\prod_{n=1}^{N_d} \Pr(z_{dn}|\boldsymbol{\theta}_d) \times \Pr(\boldsymbol{\theta}_d | \alpha, \boldsymbol{\xi}_{c_d}) \Big) d\boldsymbol{\theta}_d$$

$$= \int_{\boldsymbol{\theta}_d} \prod_{k=1}^{K} (\theta_{dk}) \times \prod_{k=1}^{K} (\theta_{dk})^{N_{dk}} \times \frac{\Gamma(\sum_{k=1}^{K} \alpha\xi_{c_dk})}{\prod_{k=1}^{K} \Gamma(\alpha\xi_{c_dk})} \prod_{k=1}^{K} (\theta_{dk})^{(\alpha\xi_{c_dk})-1} d\boldsymbol{\theta}_d$$

$$\propto \int_{\boldsymbol{\theta}_d} \prod_{k=1}^{K} (\theta_{dk}) \times \mathrm{Dir}(N_{dk} + \alpha\xi_{c_dk}) d\boldsymbol{\theta}_d \qquad \text{(Expectation!)}$$

$$= \frac{N_{dk} + \alpha\xi_{c_dk}}{N_d + \alpha}.$$

$$(5)$$

Then, we move to the next cluster-level hierarchy:

$$\Pr(z_{N_d+1} = k | \boldsymbol{z}, \alpha, \alpha_1, \boldsymbol{m})$$

$$= \int_{\boldsymbol{\xi}_{c_d}} \Pr(z_{N_d+1} = k | \boldsymbol{z}, \alpha, \boldsymbol{\xi}_{c_d}) \times \Pr(\boldsymbol{\xi}_{c_d} | \boldsymbol{z}, \alpha_1, \boldsymbol{m}) d\boldsymbol{\xi}_{c_d}$$

$$= \int_{\boldsymbol{\xi}_{c_d}} \prod_{k=1}^{K} (\frac{N_{dk} + \alpha\xi_{c_dk}}{N_d + \alpha}) \times \Big(\prod_{d':c_d'=c_d} \prod_{n=1}^{N_d} \Pr(z_{dn}|\boldsymbol{\xi}_{c_d}) \times \Pr(\boldsymbol{\xi}_{c_d} | \alpha_1, \boldsymbol{m}) \Big) d\boldsymbol{\xi}_{c_d}$$

$$= \frac{N_{dk}}{N_d + \alpha} + \frac{\alpha}{N_d + \alpha} \int_{\boldsymbol{\xi}_{c_d}} \prod_{k=1}^{K} (\xi_{c_dk}) \times \prod_{k=1}^{K} (\xi_{c_dk})^{N_{c_dk}} \times \frac{\Gamma(\sum_{k=1}^{K} \alpha_1 m_k)}{\prod_{k=1}^{K} \Gamma(\alpha_1 m_k)} \prod_{k=1}^{K} (\xi_{c_dk})^{(\alpha_1 m_k)-1} d\boldsymbol{\xi}_{c_d}$$

$$\propto \frac{N_{dk}}{N_d + \alpha} + \frac{\alpha}{N_d + \alpha} \int_{\boldsymbol{\xi}_{c_d}} \prod_{k=1}^{K} (\xi_{c_dk}) \times \mathrm{Dir}(N_{c_dk} + \alpha_1 m_k) d\boldsymbol{\xi}_{c_d} \qquad \text{(Expectation!)}$$

$$= \frac{N_{dk} + \alpha\frac{N_{c_dk}+\alpha_1 m_k}{N_{c_d}+\alpha_1}}{N_d + \alpha}.$$

$$(6)$$

NOTE: I am not sure if we can directly use $\Pr(z_{dn}|\boldsymbol{\xi}_{c_d})$ as Multinomial when we move from the first line to the second one.

Finally, we move to the next corpus-level hierarchy:

$$
\Pr(z_{N_d+1} = k | \boldsymbol{z}, \alpha, \alpha_1, \alpha_0)
$$

$$
= \int_{\boldsymbol{m}} \Pr(z_{N_d+1} = k | \boldsymbol{z}, \alpha, \alpha_1, \boldsymbol{m}) \times \Pr(\boldsymbol{m} | \boldsymbol{z}, \alpha_0, \boldsymbol{u}) d\boldsymbol{m}
$$

$$
= \int_{\boldsymbol{m}} \prod_{k=1}^{K} \left(\frac{N_{dk} + \alpha \frac{N_{c_d k} + \alpha_1 m_k}{N_{c_d} + \alpha_1}}{N_d + \alpha} \right) \times \left(\prod_{d=1}^{D} \prod_{n=1}^{N_d} \Pr(z_{dn} | \boldsymbol{m}) \times \Pr(\boldsymbol{m} | \alpha_0, \boldsymbol{u}) \right) d\boldsymbol{m}
$$

$$
= \frac{N_{dk}}{N_d + \alpha} + \frac{\alpha}{N_d + \alpha} \times
$$

$$
\left(\frac{N_{c_d k}}{N_{c_d} + \alpha_1} + \frac{\alpha_1}{N_{c_d} + \alpha_1} \int_{\boldsymbol{m}} \prod_{k=1}^{K} (m_k) \times \prod_{k=1}^{K} (m_k)^{N_k} \times \frac{\Gamma(\sum_{k=1}^{K} \alpha_0/K)}{\prod_{k=1}^{K} \Gamma(\alpha_0/K)} \prod_{k=1}^{K} (m_k)^{(\alpha_0/K)-1} d\boldsymbol{m} \right)
$$

$$
\propto \frac{N_{dk}}{N_d + \alpha} + \frac{\alpha}{N_d + \alpha} \times
$$

$$
\left(\frac{N_{c_d k}}{N_{c_d} + \alpha_1} + \frac{\alpha_1}{N_{c_d} + \alpha_1} \int_{\boldsymbol{m}} \prod_{k=1}^{K} (m_k) \times \mathrm{Dir}(N_k + \alpha_0/K) d\boldsymbol{m} \right) \qquad \text{(Expectation!)}
$$

$$
= \frac{N_{dk}}{N_d + \alpha} + \frac{\alpha}{N_d + \alpha} \left(\frac{N_{c_d k}}{N_{c_d} + \alpha_1} + \frac{\alpha_1}{N_{c_d} + \alpha_1} \frac{N_k + \alpha_0/K}{N + \alpha_0} \right)
$$

$$
= \frac{N_{dk} + \alpha \frac{N_{c_d k} + \alpha_1 \frac{N_k + \alpha_0/K}{N + \alpha_0}}{N_{c_d} + \alpha_1}}{N_d + \alpha}.
$$

$$
(7)
$$

NOTE: Again, I am not sure if we can directly use $\Pr(z_{dn}|\boldsymbol{m})$ as Multinomial when we move from the first line to the second one.

3 Integrating out $\Theta, \Xi, \boldsymbol{m}$

The big joint distribution is:

$$
\prod_{d=1}^{D} \prod_{n=1}^{N_d} \Pr(z_{dn}|\boldsymbol{\theta}_d) \times \prod_{d=1}^{D} \Pr(\boldsymbol{\theta}_d | \alpha, \boldsymbol{\xi}_{c_d}) \times \prod_{c=1}^{C} \Pr(\boldsymbol{\xi}_c | \alpha_1, \boldsymbol{m}) \times \Pr(\boldsymbol{m} | \alpha_0)
$$

$$
= \prod_{d=1}^{D} \prod_{n=1}^{N_d} \mathrm{Multinom}(z_{dn}|\boldsymbol{\theta}_d) \times \prod_{d=1}^{D} \mathrm{Dir}(\boldsymbol{\theta}_d | \alpha, \boldsymbol{\xi}_{c_d}) \times \prod_{c=1}^{C} \mathrm{Dir}(\boldsymbol{\xi}_c | \alpha_1, \boldsymbol{m}) \times \mathrm{Dir}(\boldsymbol{m} | \alpha_0).
$$

$$
(8)
$$

We want to integrate out $\boldsymbol{\theta}$, $\boldsymbol{\xi}$, and $\boldsymbol{m}$:

$$\int_{\boldsymbol{m}} \int_{\Xi} \int_{\Theta} \prod_{d=1}^{D} \prod_{n=1}^{N_d} \Pr(z_{dn}|\boldsymbol{\theta}_d) \times \prod_{d=1}^{D} \Pr(\boldsymbol{\theta}_d|\alpha, \boldsymbol{\xi}_{c_d}) \times \prod_{c=1}^{C} \Pr(\boldsymbol{\xi}_c|\alpha_1, \boldsymbol{m}) \times \Pr(\boldsymbol{m}|\alpha_0) d\Theta d\Xi dM$$

$$= \int_{\boldsymbol{m}} \int_{\Xi} \prod_{c=1}^{C} \Pr(\boldsymbol{\xi}_c|\alpha_1, \boldsymbol{m}) \times \Pr(\boldsymbol{m}|\alpha_0) \times \Big(\int_{\Theta} \prod_{d=1}^{D} \prod_{n=1}^{N_d} \Pr(z_{dn}|\boldsymbol{\theta}_d) \times \prod_{d=1}^{D} \Pr(\boldsymbol{\theta}_d|\alpha, \boldsymbol{\xi}_{c_d}) d\Theta \Big) d\Xi dM.$$

$$(9)$$

First, we can work on the integration of $\boldsymbol{\theta}$ as below:

$$\int_{\Theta} \prod_{d=1}^{D} \prod_{n=1}^{N_d} \Pr(z_{dn}|\boldsymbol{\theta}_d) \times \prod_{d=1}^{D} \Pr(\boldsymbol{\theta}_d|\alpha, \boldsymbol{\xi}_{c_d}) d\Theta$$

$$= \prod_{d=1}^{D} \int_{\boldsymbol{\theta}_d} \prod_{n=1}^{N_d} \mathrm{Multinom}(z_{dn}|\boldsymbol{\theta}_d) \times \mathrm{Dir}(\boldsymbol{\theta}_d|\alpha, \boldsymbol{\xi}_{c_d}) d\boldsymbol{\theta}_d$$

$$= \prod_{d=1}^{D} \int_{\boldsymbol{\theta}_d} \prod_{k=1}^{K} (\theta_{dk})^{N_{dk}} \times \frac{\Gamma(\sum_{k=1}^{K} \alpha\xi_{c_dk})}{\prod_{k=1}^{K} \Gamma(\alpha\xi_{c_dk})} \prod_{k=1}^{K} (\theta_{dk})^{(\alpha\xi_{c_dk})-1} d\boldsymbol{\theta}_d$$

$$= \prod_{d=1}^{D} \frac{\Gamma(\sum_{k=1}^{K} \alpha\xi_{c_dk})}{\prod_{k=1}^{K} \Gamma(\alpha\xi_{c_dk})} \int_{\boldsymbol{\theta}_d} \prod_{k=1}^{K} (\theta_{dk})^{(\alpha\xi_{c_dk})+N_{dk}-1} d\boldsymbol{\theta}_d$$

$$= \prod_{d=1}^{D} \frac{\Gamma(\sum_{k=1}^{K} \alpha\xi_{c_dk})}{\prod_{k=1}^{K} \Gamma(\alpha\xi_{c_dk})} \frac{\prod_{k=1}^{K} \Gamma(\alpha\xi_{c_dk} + N_{dk})}{\Gamma(\sum_{k=1}^{K} \alpha\xi_{c_dk} + N_{dk})} \int_{\boldsymbol{\theta}_d} \frac{\Gamma(\sum_{k=1}^{K} \alpha\xi_{c_dk} + N_{dk})}{\prod_{k=1}^{K} \Gamma(\alpha\xi_{c_dk} + N_{dk})} \prod_{k=1}^{K} (\theta_{dk})^{(\alpha\xi_{c_dk})+N_{dk}-1} d\boldsymbol{\theta}_d$$

$$= \prod_{d=1}^{D} \frac{\Gamma(\sum_{k=1}^{K} \alpha\xi_{c_dk})}{\Gamma(\sum_{k=1}^{K} \alpha\xi_{c_dk} + N_{dk})} \frac{\prod_{k=1}^{K} \Gamma(\alpha\xi_{c_dk} + N_{dk})}{\prod_{k=1}^{K} \Gamma(\alpha\xi_{c_dk})}$$

$$= \prod_{d=1}^{D} \frac{\Gamma(\alpha)}{\Gamma(\alpha + N_d)} \frac{\prod_{k=1}^{K} \Gamma(\alpha\xi_{c_dk} + N_{dk})}{\prod_{k=1}^{K} \Gamma(\alpha\xi_{c_dk})},$$

$$(10)$$

where N_{dk} is the number of times topic k is assigned in document d. Since $\frac{\Gamma(\alpha)}{\Gamma(\alpha+N_d)}$ is only a function of hyperparameters, we drop the term and then

Equation (6) can be re-written as:

$$\int_{\boldsymbol{m}} \int_{\Xi} \prod_{c=1}^{C} \Pr(\boldsymbol{\xi}_c | \alpha_1, \boldsymbol{m}) \times \Pr(\boldsymbol{m} | \alpha_0) \times \Big(\int_{\Theta} \prod_{d=1}^{D} \prod_{n=1}^{N_d} \Pr(z_{dn} | \boldsymbol{\theta}_d) \times \prod_{d=1}^{D} \Pr(\boldsymbol{\theta}_d | \alpha, \boldsymbol{\xi}_{c_d}) d\Theta \Big) d\Xi dM$$

$$\propto \int_{\boldsymbol{m}} \int_{\Xi} \prod_{c=1}^{C} \Pr(\boldsymbol{\xi}_c | \alpha_1, \boldsymbol{m}) \times \Pr(\boldsymbol{m} | \alpha_0) \times \prod_{d=1}^{D} \prod_{k=1}^{K} \frac{\Gamma(\alpha \xi_{c_d k} + N_{dk})}{\Gamma(\alpha \xi_{c_d k})} d\Xi dM$$

$$= \int_{\boldsymbol{m}} \Pr(\boldsymbol{m} | \alpha_0) \Big(\int_{\Xi} \prod_{c=1}^{C} \Pr(\boldsymbol{\xi}_c | \alpha_1, \boldsymbol{m}) \times \prod_{d=1}^{D} \prod_{k=1}^{K} \frac{\Gamma(\alpha \xi_{c_d k} + N_{dk})}{\Gamma(\alpha \xi_{c_d k})} d\Xi \Big) dM.$$

$$(11)$$

Then, we can work on the integration of Ξ as below:

$$\int_{\Xi} \prod_{c=1}^{C} \Pr(\boldsymbol{\xi}_c | \alpha_1, \boldsymbol{m}) \times \prod_{d=1}^{D} \prod_{k=1}^{K} \frac{\Gamma(\alpha \xi_{c_d k} + N_{dk})}{\Gamma(\alpha \xi_{c_d k})} d\Xi$$

$$= \prod_{c=1}^{C} \int_{\xi_c} \mathrm{Dir}(\boldsymbol{\xi}_c | \alpha_1, \boldsymbol{m}) \times \prod_{d:c_d=c} \prod_{k=1}^{K} \frac{\Gamma(\alpha \xi_{ck} + N_{dk})}{\Gamma(\alpha \xi_{ck})} d\xi_c$$

$$= \prod_{c=1}^{C} \int_{\xi_c} \frac{\Gamma(\sum_{k=1}^{K} \alpha_1 m_k)}{\prod_{k=1}^{K} \Gamma(\alpha_1 m_k)} \prod_{k=1}^{K} \Big((\xi_{ck})^{(\alpha_1 m_k)-1} \times \prod_{d:c_d=c} \frac{\Gamma(\alpha \xi_{ck} + N_{dk})}{\Gamma(\alpha \xi_{ck})} \Big) d\xi_c$$

$$= \prod_{c=1}^{C} \frac{\Gamma(\sum_{k=1}^{K} \alpha_1 m_k)}{\prod_{k=1}^{K} \Gamma(\alpha_1 m_k)} \int_{\xi_c} \prod_{k=1}^{K} \Big((\xi_{ck})^{(\alpha_1 m_k)-1} \times \prod_{d:c_d=c} \frac{\Gamma(\alpha \xi_{ck} + N_{dk})}{\Gamma(\alpha \xi_{ck})} \Big) d\xi_c.$$

$$(12)$$

Here, we need to know from Equation (36) of Hierarchical Dirichlet Process (http://qwone.com/~jason/trg/papers/teh-hierarchical-04.pdf) that

$$\frac{\Gamma(\alpha \xi_{ck} + N_{dk})}{\Gamma(\alpha \xi_{ck})} = \prod_{i_{dk}=1}^{N_{dk}} (i_{dk} - 1 + \alpha \xi_{ck})$$

$$= \sum_{i_{dk}=0}^{N_{dk}} s(N_{dk}, i_{dk})(\alpha \xi_{ck})^{i_{dk}},$$

$$(13)$$

5

where $s(i_{dk}, N_{dk})$ is the coefficient of $(\alpha \xi_{ck})^{i_{dk}}$, which are unsigned Stirling numbers of the first kind. Plugging this into Equation (9), we have

$$\prod_{c=1}^{C} \frac{\Gamma(\sum_{k=1}^{K} \alpha_1 m_k)}{\prod_{k=1}^{K} \Gamma(\alpha_1 m_k)} \int_{\xi_c} \prod_{k=1}^{K} \left((\xi_{ck})^{(\alpha_1 m_k)-1} \times \prod_{d:c_d=c} \sum_{i_{dk}=0}^{N_{dk}} s(N_{dk}, i_{dk})(\alpha \xi_{ck})^{i_{dk}} \right) d\xi_c.$$
(14)

Maybe we need to introduce auxiliary variable $\boldsymbol{i} = (i_{dk} : \forall d, k)$ and move forward, but I have no idea....

$$\propto \prod_{c=1}^{C} \frac{\Gamma(\sum_{k=1}^{K} \alpha_1 m_k)}{\prod_{k=1}^{K} \Gamma(\alpha_1 m_k)} \int_{\xi_c} \prod_{k=1}^{K} \left((\xi_{ck})^{(\alpha_1 m_k)-1} \times \prod_{d:c_d=c} \sum_{i_{dk}=0}^{N_{dk}} s(N_{dk}, i_{dk})(\alpha \xi_{ck})^{i_{dk}} \right) d\xi_c.$$
(15)