Complete proof of the solution of "Mutation Data Bridging" model in GDAMDB

Kaiyin Zhou[1#], Yuxing Wang[1#], Kevin Bretonnel Cohen[2], Jin-Dong Kim[3], Xiaohang Ma[1], Zhixue Shen[1], Xiangyu Meng[4,5], Jingbo Xia[1*]

[1]Hubei Key Lab of Agricultural Bioinformatics, College of Informatics, Huazhong Agricultural University, Wuhan, Hubei Province, P.R. China

[2]School of Medicine, University of Colorado Denver, Anschutz Medical Campus, Colorado, U.S

[3]Database Center for Life Science (DBCLS), Research Organization of Information and Systems (ROIS), Tokyo, Japan

[4]Department of Urology, Zhongnan Hospital of Wuhan University, Wuhan, Hubei Province, P.R. China

[5]Institut Curie, CNRS, Molecular Oncology Team, PSL Research University, Paris, France

1 Parameter setting

Disease: $d \in \{1, \cdots, D\}$;
Gene: $g \in \{1, \cdots, G\}$; *Mutation type* of gene g for disease d: $f_{dg} \in \{0,1\}^3$ refers to "LOF/GOF/NA";
The p-value of the mapped *mutation association* of gene g in disease d in GWAS: $p_{dg} \in (0,1)$.

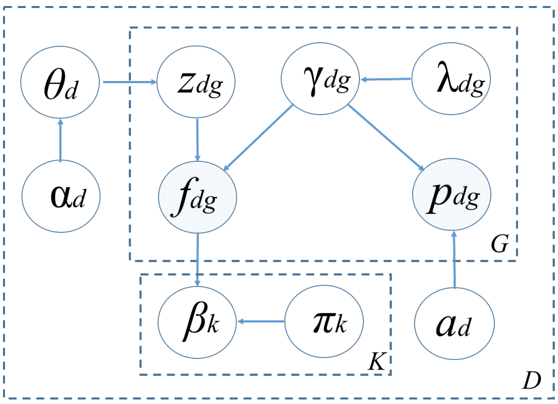

Figure 1: Graphical model of function change (f_{dg}) for the mentioned gene GM_{dg}.

As shown in Figure 1, all the random variables follow the following distributions.

$$
\begin{aligned}
&f_{dg} \sim \begin{cases} Multi(\beta_{z_{dg}}), & \text{if } \gamma_{dg} = 1 \\ \mathbf{0}, & \text{if } \gamma_{dg} = 0 \end{cases}, \\
&p_{dg} \sim \begin{cases} Beta(a_d, 1), & \text{if } \gamma_{dg} = 1 \\ U(0,1), & \text{if } \gamma_{dg} = 0 \end{cases}, \\
&\gamma_{dg} \sim Bernoulli(\lambda_{dg}), \text{ where } \lambda_{dg} \in (0,1), \\
&z_{dg} \sim Categorical(\theta_d), \;\; z_{dg} \in \{0,1\}^K, \text{ if there are } K \text{ latent factors,} \\
&\theta_d \sim Dir(\alpha_d), \text{ where } \alpha_d \in \mathbb{R}^K, \\
&\beta_k \sim Dir(\pi_k), \;\; k = 1, 2, \cdots, K, \text{ where } \pi \in \mathbb{R}^3.
\end{aligned} \tag{1}
$$

In the above notations, $z_{dg} \in \{0,1\}^K$ is a latent variable for factor index of β_k with respect to GM_{dg}, which is similar to topic index β_k in standard LDA [1]. Meanwhile, γ_{dg} is a "syncing variable" working for spike and slab prior so that f_{dg} and p_{dg} follow a consistent sparsity prior through value-syncing of γ_{dg}.

For a fixed disease d, when γ_{dg} equals to 1, p_{dg} follows a Beta distribution with parameter a_d being much greater than 1, which refers to a high significance with respect to the association between GM_{dg} and disease d. In the meantime, f_{dg} follows a Multinomial (Actually categorical) distribution and samples a $\{0,1\}^3$ vector which refers to function change, LOF/GOF/NA. In another case when γ_{dg} equals to zero, p_{dg} follows a Uniform distribution which leads to non-significance, while f_{dg} is valued as a zero vector.

For simplicity, all the parameters used in this work is shown as below:
Observation: $F = \{f_{dg}\}$, $P = \{p_{dg}\}$,
Latent variable: $\theta = \{\theta_d\}$, $\beta = \{\beta_k\}$, $\gamma = \{\gamma_{dg}\}$, $Z = \{z_{dg}\}$,
Model parameter: $\Theta = \{\alpha_d, \pi_k, \lambda_{dg}, a_d\}$.

2 Brief Flowchart of Model Computation Steps

0). Preliminaries to exponential family distributions.

Noted that all of the variables in (1) follow exponential family distribution [2], so each of them has Probability Density Function (pdf) with the form

$$p(x|\eta) = h(x) \cdot \exp(T(x)^T(\eta) - A(\eta)),$$

where $T(x)$ is a vector function of x, called as "sufficient statistics"; $A(\eta)$ is called as "log-normalizer; $h(x)$ is called as "base measure"; η is called as "natural parameter", which is the distribution parameter of the exponential family distribution.

1). The first step of the variational inference is to derive an ELBO (Evidence lower bound) by using Jensen's inequality to handle the logarithm of evidence $p(F, P)$.

In order to "detangle" f_{dg} in terms of binary value of γ_{dg}, a trick from Dai et al [3, 4] is used here. We assume $\tilde{f}_{dg} \sim Multi(\beta_{z_{dg}})$, then we know that $\tilde{f}_{dg}\gamma_{dg}$ has the same distribution of f_{dg}. So the consideration of logarithm of $P(F, P)$ is converted to that of $P(\tilde{F}, P)$.

$$
\begin{aligned}
\log p(\tilde{F}, P|\Theta) \quad &= \log \int_\theta \int_\beta \sum_\gamma \sum_z p(\tilde{F}, P, \theta, \beta, \gamma, z|\Theta) d\theta d\beta \\
&= \log \int_\theta \int_\beta \sum_\gamma \sum_z \frac{p(\tilde{F}, P, \theta, \beta, \gamma, z|\Theta)}{q(\theta, \beta, \gamma, z)} \cdot q(\theta, \beta, \gamma, z) d\theta d\beta \\
&\text{(Using Jensen's inequality)} \\
&\geq E_{q(\theta,\beta,\gamma,z)}[\log p(\tilde{F}, P, \theta, \beta, \gamma, z|\Theta)] - E_{q(\theta,\beta,\gamma,z)}[q(\theta, \beta, \gamma, z)] \\
&:= L(q) \quad \text{(also called as ELBO.)}
\end{aligned} \tag{2}
$$

$$\begin{aligned}\log p(\tilde{F}, P|\Theta) - ELBO \quad &= \log \int_\theta \int_\beta \sum_\gamma \sum_z p(\tilde{F}, P, \theta, \beta, \gamma, z|\Theta) d\theta d\beta \\ &= \log \int_\theta \int_\beta \sum_\gamma \sum_z \frac{p(\tilde{F}, P, \theta, \beta, \gamma, z|\Theta)}{q(\theta, \beta, \gamma, z)} \cdot q(\theta, \beta, \gamma, z) d\theta d\beta \\ &\text{(Using Jensen's inequality)} \\ &\geq E_{q(\theta,\beta,\gamma,z)}[\log p(\tilde{F}, P, \theta, \beta, \gamma, z|\Theta)] - E_{q(\theta,\beta,\gamma,z)}[q(\theta, \beta, \gamma, z)] \\ &:= L(q) \quad \text{(also called as ELBO.)}\end{aligned} \tag{3}$$

2). Second step: to maximize ELBO, which equals to minimize $\mathbb{KL}(p||q)$, the Kullback-Leibler divergence of $q(\theta, \beta, \gamma, z)$ and $p(\theta, \beta, \gamma, z|F, P; \Theta)$.

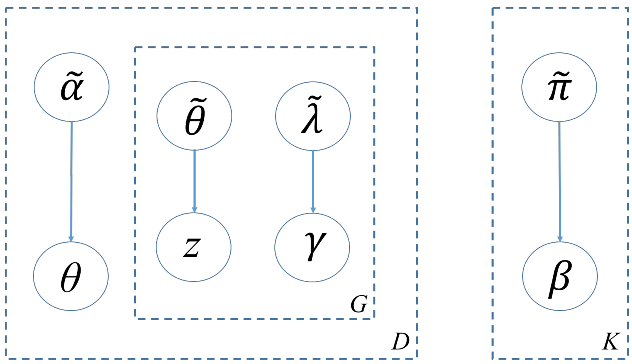

Figure 2: Graphical model of variational parameters

The graphical model of variational parameters setting is shown in Figure 2. According to "mean-field" algorithm [2] in variational inference, $q(\theta, \beta, \gamma, z) = q(\theta)q(\beta)q(\gamma)q(z)$ is represented by corresponding exponential family distributions:

$$\begin{cases} \text{Since } \theta_d \sim Dir(\alpha_d), \text{ we assume } q(\theta_d) = Dir(\tilde{\alpha}_d); \\ \text{Since } \beta_k \sim Dir(\pi_k), \text{ we assume } q(\beta_k) = Dir(\tilde{\pi}_k); \\ \text{Since } \gamma_{dg} \sim Bernoulli(\lambda_{dg}), \text{ we assume } q(\gamma_{dg}) = Bernoulli(\tilde{\lambda}_{dg}); \\ \text{Since } z_{dg} \sim Categorical(\theta_d), \text{ we assume } q(z_{dg}) = Categorical(\tilde{\theta}_{dg}). \end{cases} \tag{4}$$

Here, $\tilde{\alpha}$, $\tilde{\pi}$, $\tilde{\lambda}_{dg}$ and $\tilde{\theta}_{dg}$ are variational parameters under estimation. For simplicity, we denote $\tilde{\Theta} = \{\tilde{\alpha}, \tilde{\pi}, \tilde{\lambda}_{dg}, \tilde{\theta}_{dg}(d = 1, 2, \cdots, D; g = 1, 2, \cdots, G)\}$.

3). Variational parameters are estimated by differential computation of ELBO.

Here, the value of ELBO in formula (26) can be directly derived from Proposition 2 and Proposition 3. In addition, iteration formula of variational parameters and differential computation steps are shown in Theorem 1 and Theorem 2.

3 Model Computation Details

In order to perform computation in the "mean-field" variational inference algorithm, we assume $q(\theta_d) = Dir(\tilde{\alpha}_d)$, $q(\beta_k) = Dir(\tilde{\pi}_k)$, $q(\gamma_{dg}) = Bernoulli(\tilde{\lambda}_{dg})$, and $q(z_{dg}) = Categorical(\tilde{\theta}_{dg})$, where $\tilde{\alpha}_d$, $\tilde{\pi}_k$, $\tilde{\lambda}_{dg}$ and $\tilde{\theta}_{dg}$ are variational parameters.

The ELBO computation is given by Proposition 2 and Proposition 3 in Section 3.1. Subsequently, the variational parameters is computed via VI computations finally through Theorems 1 in Section 3.2.

3.1 ELBO Computation

Proposition 1. *For a fixed d,* $\tilde{F} = \{\tilde{f}_{dg}\}$*,* $P = \{p_{dg}\}$*,* $\theta = \theta_d$*,* $\beta = \{\beta_{z_{dg}}\}$*,* $\gamma = \{\gamma_{dg}\}$*,* $z = \{z_{dg}\}$*,* $g = 1, 2, \cdots, G$*. The joint probability* $p(\tilde{F}, P, \theta, \beta, \gamma, z|\Theta)$ *equals to*

$$\begin{aligned} p(\tilde{F}, P, \theta, \beta, \gamma, z|\Theta) &= \prod_{g=1}^{G}\prod_{k=1}^{K}\prod_{f=1}^{F}\beta_{k;f}^{\tilde{f}_{dg;f} z_{dg;k}} \cdot \prod_{g=1}^{G}((a_d p_{dg})^{a_d-1})^{\gamma_{dg}} \cdot \frac{\Gamma(\sum_{k=1}^{K}\alpha_{d;k})}{\prod_{k=1}^{K}\Gamma(\alpha_{d;k})}\prod_{k=1}^{K}\theta_{d;k}^{\alpha_{d;k}-1} \\ &\cdot \prod_{k=1}^{K}\frac{\Gamma(\sum_{f=1}^{F}\pi_{k;f})}{\prod_{f=1}^{F}\Gamma(\pi_{k;f})}\prod_{f=1}^{F}\beta_{k;f}^{\pi_{k;f}-1} \cdot \prod_{g=1}^{G}\lambda_{dg}^{\gamma_{dg}}(1-\lambda_{dg})^{1-\gamma_{dg}} \cdot \prod_{g=1}^{G}\prod_{k=1}^{K}\theta_{d;k}^{z_{dg;k}}. \end{aligned} \tag{5}$$

Proof. As shown in Figure 1, the decomposition of the joint probability is

$$p(\tilde{F}, P, \theta, \beta, \gamma, z|\Theta) = p(\tilde{F}|\beta, z; \Theta)p(P|\gamma; \Theta)p(\theta; \Theta)p(\beta; \Theta)p(\gamma; \Theta)p(z|\theta; \Theta). \tag{6}$$

Subsequently, each term in the joint probability, $p(\tilde{F}, P, \theta, \beta, \gamma, z|\Theta)$, is directly derived from the following results of Lemma 1 to Lemma 6. □

Lemma 1. *The pdf of* $p(\tilde{F}|\beta, z; \Theta)$ *is*

$$p(\tilde{F}|\beta, z; \Theta) = \prod_{g=1}^{G}\prod_{k=1}^{K}\prod_{f=1}^{F}\beta_{k;f}^{\tilde{f}_{dg;f}\cdot z_{dg;k}}. \tag{7}$$

Proof. It is clear that $\tilde{f}_{dg} \cdot z_{dg;k} \sim Multi(\beta_k)$, and it is known the probability for multinomial distribution is $p(x_1 = n_1, x_2 = n_2, \cdots, x_k = n_k) = (n_1 + n_2 + \cdots + n_k)! \prod_{k=1}^{K} \frac{p_k^{n_k}}{n_k!}$. So we have

$$p(\tilde{F}|\beta, z; \Theta) = \prod_{g=1}^{G} p(\tilde{f}_{dg}|z_{dg}, \beta) = \prod_{g=1}^{G}\prod_{k=1}^{K} p(\tilde{f}_{dg} \cdot z_{dg;k}, \beta_k) = \prod_{g=1}^{G}\prod_{k=1}^{K}\prod_{f=1}^{F}\beta_{k;f}^{\tilde{f}_{dg;f}\cdot z_{dg;k}}$$

□

Lemma 2. *The pdf of* $p(P|\gamma; \Theta)$ *is*

$$p(P|\gamma; \Theta) = \prod_{g=1}^{G}((a_d p_{dg})^{a_d-1})^{\gamma_{dg}} \tag{8}$$

.

Proof. It is noted that p_{dg} follows the distribution $p_{dg} \sim \begin{cases} Beta(a_d, 1), & \text{if } \gamma_{dg} = 1 \\ U(0,1), & \text{if } \gamma_{dg} = 0 \end{cases}$, and the pdf of Beta$(a_d, 1)$ is $f(x, a_d, 1) = \frac{\Gamma(\alpha+1)}{\Gamma(a_d)\Gamma(1)x^{a_d-1}(1-x)^0} = a_d x^{a_d-1}$. Taking a trick for fast computation [3], the pdf of $p(P|\gamma; \Theta)$ is

$$p(P|\gamma, a_d; \Theta) = \prod_{g=1}^{G} p(p_{dg}|\gamma_{dg}, a_d; \Theta) = \prod_{g=1}^{G}((a_d p_{dg})^{a_d-1})^{\gamma_{dg}} \quad .$$

□

Lemma 3. *The pdf of $p(\theta;\Theta)$ is*

$$p(\theta;\Theta) = \frac{\Gamma(\sum_{k=1}^{K} \alpha_{d;k})}{\prod_{k=1}^{K} \Gamma(\alpha_{d;k})} \prod_{k=1}^{K} \theta_{d;k}^{\alpha_{d;k}-1}. \tag{9}$$

Proof. Since $\theta_d \sim Dir(\alpha_d)$, the result is obtained in a straightforward way. □

Lemma 4. *The pdf of $p(\beta;\Theta)$ is*

$$p(\beta;\Theta) = \prod_{k=1}^{K} \frac{\Gamma(\sum_{f=1}^{F} \pi_{k;f})}{\prod_{f=1}^{F} \Gamma(\pi_{k;f})} \prod_{f=1}^{F} \beta_{k;f}^{\pi_{k;f}-1}. \tag{10}$$

Proof. Since $\beta_k \sim Dir(\pi_k)$, it is straightforward to have

$$p(\beta;\Theta) = \prod_{k=1}^{K} p(\beta_k;\Theta) = \prod_{k=1}^{K} \frac{\Gamma(\sum_{f=1}^{F} \pi_{k;f})}{\prod_{f=1}^{F} \Gamma(\pi_{k;f})} \prod_{f=1}^{F} \beta_{k;f}^{\pi_{k;f}-1}.$$

Thus lemma follows. □

Lemma 5. *The pdf of $p(\gamma;\Theta)$ is*

$$p(\gamma;\Theta) = \prod_{g=1}^{G} \lambda^{\gamma_{dg}} (1-\lambda_{dg})^{1-\gamma_{dg}}. \tag{11}$$

Proof. Since $\gamma_{dg} \sim Bernoulli(\lambda_{dg})$, while probability of Bernoulli distribution is $p(x|p) = p^x(1-p)^{1-x}$, so we have

$$p(\gamma;\Theta) = \prod_{g=1}^{G} p(\gamma_{dg}) = \prod_{g=1}^{G} \lambda^{\gamma_{dg}} (1-\lambda_{dg})^{1-\gamma_{dg}}.$$

□

Lemma 6. *The pdf of $p(z|\theta;\Theta)$ is*

$$p(z|\theta;\Theta) = \prod_{g=1}^{G} \prod_{k=1}^{K} \theta_{d;k}^{z_{dg;k}}. \tag{12}$$

Proof. Similar to the proof of Lemma 1, we know that

$$p(z;\Theta) = \prod_{g=1}^{G} p(z_{dg}|\theta_d;\Theta) = \prod_{g=1}^{G} \prod_{k=1}^{K} \theta_{d;k}^{z_{dg;k}},$$

when $z_{dg} \sim Multi(\theta_d)$. □

Subsequently, Lemma 7 to 12 are to show Proposition 2, which presented the expectation of $E_{q(\theta,\beta,\gamma,z)}[\log p(\tilde{F}, P, \theta, \beta, \gamma, z|\Theta)]$.

Proposition 2. *Let $\tilde{F}, P, \theta, \beta, \gamma, z$ being defined as in Proposition 1, and without loss of generality assume $q(\theta, \beta, \gamma, z) = q(\theta)q(\beta)q(\gamma)q(z)$. The expectation of logarithm of joint probability, $\log p(\tilde{F}, P, \theta, \beta, \gamma, z|\Theta)$, is*

$$
\begin{aligned}
& E_{q(\theta,\beta,\gamma,z)}[\log p(\tilde{F}, P, \theta, \beta, \gamma, z|\Theta)] \\
& = \sum_{g=1}^{G} \sum_{k=1}^{K} \sum_{f=1}^{F} [\tilde{f}_{dg;f} \cdot \tilde{\theta}_{dg;k} \cdot (\psi(\tilde{\pi}_{k;f}) - \psi(\sum_{f=1}^{F} \tilde{\pi}_{k;f}))] \\
& \quad + \sum_{g=1}^{G} \tilde{\lambda}_{dg} \log[a_d p_{dg}^{a_d-1}] \\
& \quad + \log(\Gamma(\sum_{k=1}^{K} \alpha_{d;k})) - \sum_{k=1}^{K} \log(\Gamma(\alpha_{d;k})) + \sum_{k=1}^{K} (\alpha_{d;k} - 1)(\psi(\tilde{\alpha}_{d;k}) - \psi(\sum_{k=1}^{K} \tilde{\alpha}_{d;k})) \\
& \quad + \sum_{k=1}^{K} \log(\Gamma(\sum_{f=1}^{F} \pi_{k;f})) - \sum_{k=1}^{K} \sum_{f=1}^{F} \log \Gamma(\pi_{k;f}) + \sum_{k=1}^{K} \sum_{f=1}^{F} (\pi_{k;f} - 1)(\psi(\tilde{\pi}_{k;f}) - \psi(\sum_{f=1}^{F} \tilde{\pi}_{k;f})) \\
& \quad + \sum_{g=1}^{G} [\tilde{\lambda}_{dg} \log \lambda_{dg} + (1 - \tilde{\lambda}_{dg}) \log(1 - \lambda_{dg})] \\
& \quad + \sum_{g=1}^{G} \sum_{k=1}^{K} \tilde{\theta}_{dg;k}(\psi(\tilde{\alpha}_{d;k}) - \psi(\sum_{k=1}^{K} \tilde{\alpha}_{d;k}))
\end{aligned} \tag{13}
$$

Proof. Using the joint probability in Proposition 1, we have $E_{q(\theta,\beta,\gamma,z)}[\log p(\tilde{F}, P, \theta, \beta, \gamma, z|\Theta)] = E_{q(\theta,\beta,\gamma,z)}[\log[p(\tilde{F}|\beta, z; \Theta)] + \log[p(p|\gamma; \Theta)] + \log[p(\theta; \Theta)] + \log[p(\gamma; \Theta)] + \log[p(z|\theta; \Theta)] + \log[p(\beta; \Theta)]]$. In addition, Lemma 7 to 12 suffice to show the rest part of this proposition. □

Lemma 7. *Expectation of $\log[p(\tilde{F}|\beta, z; \Theta)]$ equals to*

$$E_{q(\theta,\beta,\gamma,z)}[\log[p(\tilde{F}|\beta, z; \Theta)]] = \sum_{g=1}^{G} \sum_{k=1}^{K} \sum_{f=1}^{F} [\tilde{f}_{dg;f} \tilde{\theta}_{dg;k}(\psi(\tilde{\pi}_{k;f}) - \psi(\sum_{f=1}^{F} \tilde{\pi}_{k;f}))]. \tag{14}$$

Proof. First, we need to show that $E_{q(z)}[z_{dg;k}] = \tilde{\theta}_{dg;k}$. Here, $E_{q(z)}[z_{dg;k}] = E_{q(z_{dg})}[z_{dg;k}] = \sum_{z_{dg}} q(z_{dg}) z_{dg;k} = \sum_{k=1}^{K} \sum_{z_{dg;k}=0}^{1} q(z_{dg}) z_{dg;k} = \sum_{k=1}^{K} \sum_{z_{dg;k}=0}^{1} \prod_{m=1}^{K} \tilde{\theta}_{d;m}^{z_{dg;m}} z_{dg;k} = \sum_{k=1}^{K} \sum_{z_{dg;k}=0}^{1} \tilde{\theta}_{dg;k} z_{dg;k} = \tilde{\theta}_{dg;k}$.

Second, from [1], it is a known result that if $\theta \sim Dir(\alpha_d)$, the expectation $E_{p(\theta|\alpha_d)}[\log \theta_k] = \psi(\alpha_{d;k}) - \psi(\sum_{k=1}^{K} \alpha_{d;k})$, where digamma function, $\psi(x) = \frac{\Gamma'(x)}{\Gamma(x)}$, is the logarithmic derivative of the gamma function. As in (4), $q(\beta_k) = Dir(\tilde{\pi}_k)$, and we have $E_q(\beta)[\log \beta_{k;f}] = E_{q(\beta_k)}[\log \beta_{k;f}]$ $= \psi(\tilde{\pi}_{k;f}) - \psi(\sum_{f=1}^{F} \tilde{\pi}_{k;f})$.

Now we have

$$\begin{aligned}
E_{q(\theta,\beta,\gamma,z)}[\log[p(\tilde{F}|\beta,z;\Theta)]] &= E_{q(\theta,\beta,\gamma,z)}[\sum_{g=1}^{G}\sum_{k=1}^{K}\sum_{f=1}^{F}\log(\beta_{k;f}^{\tilde{f}_{dg;f}z_{dg;k}})] \\
&= E_{q(\theta,\beta,\gamma,z)}[\sum_{g=1}^{G}\sum_{k=1}^{K}\sum_{f=1}^{F}(\tilde{f}_{dg;f}z_{dg;k}\cdot\log\beta_{k;f})] \\
&= \sum_{g=1}^{G}\sum_{k=1}^{K}\sum_{f=1}^{F}(\tilde{f}_{dg;f}\cdot E_{q(z)}[z_{dg;k}]\cdot E_{q(\beta)}[\log\beta_{k;f}]) \\
&= \sum_{g=1}^{G}\sum_{k=1}^{K}\sum_{f=1}^{F}(\tilde{f}_{dg;f}\tilde{\theta}_{dg;k}(\psi(\tilde{\pi}_{k;f})-\psi(\sum_{f=1}^{F}\tilde{\pi}_{k;f}))).
\end{aligned}$$

Thus the lemma follows. □

Lemma 8. *Expectation of* $\log p(P|\gamma;\Theta)$ *is*

$$E_{q(\theta,\beta,\gamma,z)}[\log(p(P|\gamma;\Theta))] = \sum_{g=1}^{G}\tilde{\lambda}_{dg}\log(a_d p_{dg}^{a_d-1}) \tag{15}$$

Proof. For a fixed d, $\log p(P) = \sum_{g=1}^{G}\log p(p_{dg})$. So, combined with the result in formula (8), we have $E_{q(\theta,\beta,\gamma,z)}[\log[p(P|\gamma;\Theta)]] = \sum_{g=1}^{G}E_{q(\gamma)}[\gamma_{dg}]\log(a_d p_{dg}^{a_d-1}) = \sum_{g=1}^{G}\tilde{\lambda}_{dg}\log(a_d p_{dg}^{a_d-1})$.

□

Lemma 9. *The expectation of* $\log p(\theta;\Theta)$ *equals to*

$$E_{q(\theta,\beta,\gamma,z)}[\log p(\theta;\Theta)] = \log(\Gamma(\sum_{k=1}^{K}\alpha_{d;k})) - \sum_{k=1}^{K}\log(\Gamma(\alpha_{d;k})) + \sum_{k=1}^{K}(\alpha_{d;k}-1)(\psi(\tilde{\alpha}_{d;k})-\psi(\sum_{k=1}^{K}\tilde{\alpha}_{d;k})). \tag{16}$$

Proof. As in formula (4), $q(\theta_d)\sim Dir(\tilde{\alpha}_d)$. From (9), we know that $E_{q(\theta,\beta,\gamma,z)}[\log p(\theta;\Theta)] = E_{q(\theta)}[\log(\frac{\Gamma(\sum_{k=1}^{K}\alpha_{d;k})}{\prod_{k=1}^{K}\Gamma(\alpha_{d;k})}\prod_{k=1}^{K}\theta_{d;k}^{\alpha_{d;k}-1})] = \log(\Gamma(\sum_{k=1}^{K}\alpha_{d;k})) - \sum_{k=1}^{K}\log(\Gamma(\alpha_{d;k})) + \sum_{k=1}^{K}(\alpha_{d;k}-1)E_{q(\theta)}[\log\theta_{d;k}] = \log(\Gamma(\sum_{k=1}^{K}\alpha_{d;k})) - \sum_{k=1}^{K}\log(\Gamma(\alpha_{d;k})) + \sum_{k=1}^{K}(\alpha_{d;k}-1)(\psi(\tilde{\alpha}_{d;k})-\psi(\sum_{k=1}^{K}\tilde{\alpha}_{d;k}))$. □

Lemma 10. *Expectation of* $\log p(\beta;\Theta)$ *is*

$$\begin{aligned}
&E_{q(\theta,\beta,\gamma,z)}[\log p(\beta;\Theta)] = \\
&\sum_{k=1}^{K}\log(\Gamma(\sum_{f=1}^{F}\pi_{k;f})) - \sum_{k=1}^{K}\sum_{f=1}^{F}\log\Gamma(\pi_{k;f}) + \sum_{k=1}^{K}\sum_{f=1}^{F}(\pi_{k;f}-1)(\psi(\tilde{\pi}_{k;f})-\psi(\sum_{f=1}^{F}\tilde{\pi}_{k;f}))
\end{aligned} \tag{17}$$

Proof. As in (4), $q(\beta_k) = Dir(\tilde{\pi}_k)$, and combined with the result of formula (10), we have $E_{q(\theta,\beta,\gamma,z)}[\log p(\beta;\Theta)] = E_{q(\beta)}[\sum_{k=1}^{K}\log(\Gamma(\sum_{f=1}^{F}\pi_{k;f})) - \sum_{k=1}^{K}\sum_{f=1}^{F}\log\Gamma(\pi_{k;f}) + \sum_{f=1}^{F}\sum_{k=1}^{K}(\pi_{k;f}-1)\log\beta_{k;f}] = \sum_{k=1}^{K}\log(\Gamma(\sum_{f=1}^{F}\pi_{k;f})) - \sum_{k=1}^{K}\sum_{f=1}^{F}\log\Gamma(\pi_{k;f}) + \sum_{k=1}^{K}\sum_{f=1}^{F}(\pi_{k;f}-1)(\psi(\tilde{\pi}_{k;f})-\psi(\sum_{f=1}^{F}\tilde{\pi}_{k;f}))$. □

Lemma 11. *The expectation of* $\log p(\gamma;\Theta)$ *equals to*

$$E_{q(\theta,\beta,\gamma,z)}[\log p(\gamma;\Theta)] = \sum_{g=1}^{G}[\tilde{\lambda}_{dg}\log\lambda_{dg} + (1-\tilde{\lambda}_{dg})\log(1-\lambda_{dg})] \tag{18}$$

Proof. As in formula (4), $q(\gamma_{dg}) = Bernoulli(\tilde{\lambda}_{dg})$. Then combined with the result in formula 11, we have $E_{q(\theta,\beta,\gamma,z)}[\log p(\gamma;\Theta)] = E_{q(\theta,\beta,\gamma,z)}[\sum_{g=1}^{G}(\gamma_{dg}\log\lambda_{dg} + (1-\gamma_{dg})\log(1-\lambda_{dg}))] = \sum_{g=1}^{G}\log\lambda_{dg}E_{q(\gamma)}[\gamma_{dg}] + \sum_{g=1}^{G}\log(1-\lambda_{dg})E_{q(\gamma)}[1-\gamma_{dg}] = \sum_{g=1}^{G}[\tilde{\lambda}_{dg}\log\lambda_{dg}+(1-\tilde{\lambda}_{dg})\log(1-\lambda_{dg})]$. □

Lemma 12. *The expectation of* $\log p(z|\theta;\Theta)$ *equals to*

$$E_{q(\theta,\beta,\gamma,z)}[\log p(z|\theta;\Theta)] = \sum_{g=1}^{G}\sum_{k=1}^{K}\tilde{\theta}_{dg;k}(\psi(\tilde{\alpha}_{d;k}) - \psi(\sum_{k=1}^{K}\tilde{\alpha}_{d;k})) \tag{19}$$

Proof. As in formula (4), $q(\theta_d) = Dir(\tilde{\alpha}_d)$ and $q(z_{dg}) = Categorical(\tilde{\theta}_d)$. From (12), we directly have $E_{q(\theta,\beta,\gamma,z)}[\log p(z|\theta;\Theta)] = E_{q(\theta,\beta,\gamma,z)}[\sum_{g=1}^{G}\sum_{k=1}^{K}\log\theta_{d;k}^{z_{dg;k}}] = \sum_{g=1}^{G}\sum_{k=1}^{K}E_{q(z)}[z_{dg;k}]E_{q(\theta)}[\log\theta_{d;k}] = \sum_{g=1}^{G}\sum_{k=1}^{K}\tilde{\theta}_{dg;k}(\psi(\tilde{\alpha}_{d;k}) - \psi(\sum_{k=1}^{K}\tilde{\alpha}_{d;k}))$. □

Moreover, Lemma 13 to 16 are to show proposition 3, which compute the expectation of $q(\theta,\beta,\gamma,z)$.

Proposition 3.

$$\begin{aligned}E_{q(\theta,\beta,\gamma,z)}[\log q(\theta,\beta,\gamma,z)] &= \log\Gamma(\sum_{k=1}^{K}\tilde{\alpha}_{d;k}) - \sum_{k=1}^{K}\log\Gamma(\tilde{\alpha}_{d;k}) + \sum_{k=1}^{K}(\tilde{\alpha}_{d;k}-1)(\psi(\tilde{\alpha}_{d;k}) - \psi(\sum_{k=1}^{K}\tilde{\alpha}_{d;k}))\\ &+ \sum_{k=1}^{K}\log\Gamma(\sum_{f=1}^{F}\tilde{\pi}_{k;f}) - \sum_{k=1}^{K}\sum_{f=1}^{F}\log\Gamma(\tilde{\pi}_{k;f}) + \sum_{k=1}^{K}\sum_{f=1}^{F}(\tilde{\pi}_{k;f}-1)(\psi(\tilde{\pi}_{k;f}) - \psi(\sum_{f=1}^{F}\tilde{\pi}_{k;f}))\\ &+ \sum_{g=1}^{G}[\tilde{\lambda}_{dg}\log\tilde{\lambda}_{dg} + (1-\tilde{\lambda}_{dg})\log(1-\tilde{\lambda}_{dg})]\\ &+ \sum_{g=1}^{G}\sum_{k=1}^{K}\tilde{\theta}_{dg;k}\log\tilde{\theta}_{dg;k}\end{aligned} \tag{20}$$

Proof. Since $q(\theta,\beta,\gamma,z) = q(\theta|\tilde{\alpha})q(\beta|\tilde{\pi})q(\gamma|\tilde{\lambda})q(z|\tilde{\theta})$, the expectation computation directly yields $E_{q(\theta,\beta,\gamma,z)}[\log q(\theta,\beta,\gamma,z)] = E_{q(\theta)}[\log q(\theta|\tilde{\alpha})]+E_{q(\beta)}[\log q(\beta|\tilde{\pi})]+E_{q(\gamma)}[\log q(\gamma|\tilde{\lambda})]+E_{q(z)}[\log q(z|\tilde{\theta}_{dg})]$. Furthermore, the results in Lemma (13) to (16) suffice to show the rest part of this proposition. □

Lemma 13. *The expectation of* $\log q(\theta|\tilde{\alpha})$ *equals to*

$$E_{q(\theta,\beta,\gamma,z)}[\log q(\theta|\tilde{\alpha})] = \log\Gamma(\sum_{k=1}^{K}\tilde{\alpha}_{d;k}) - \sum_{k=1}^{K}\log\Gamma(\tilde{\alpha}_{d;k}) + \sum_{k=1}^{K}(\tilde{\alpha}_{d;k}-1)(\psi(\tilde{\alpha}_{d;k}) - \psi(\sum_{k=1}^{K}\tilde{\alpha}_{d;k})). \tag{21}$$

Proof. As in formula (4), $q(\theta_d) = Dir(\tilde{\alpha}_d)$, so we have $q(\theta_d|\tilde{\alpha}_d) = \frac{\Gamma(\sum_{k=1}^{K}\tilde{\alpha}_{d;k})}{\prod_{k=1}^{K}\Gamma(\tilde{\alpha}_{d;k})}\prod_{k=1}^{K}\theta_{d;k}^{\tilde{\alpha}_{d;k}-1}$. Hence, we obtain

$$\begin{aligned} E_{q(\theta,\beta,\gamma,z)}[\log q(\theta|\tilde{\alpha})] &= \log\Gamma(\sum_{k=1}^{K}\tilde{\alpha}_{d;k}) - \sum_{k=1}^{K}\log\Gamma(\tilde{\alpha}_{d;k}) + \sum_{k=1}^{K}(\tilde{\alpha}_{d;k}-1)E_{q(\theta)}[\log\theta_{d;k}] \\ &= \log\Gamma(\sum_{k=1}^{K}\tilde{\alpha}_{d;k}) - \sum_{k=1}^{K}\log\Gamma(\tilde{\alpha}_{d;k}) + \sum_{k=1}^{K}(\tilde{\alpha}_{d;k}-1)(\psi(\tilde{\alpha}_{d;k}) - \psi(\sum_{k=1}^{K}\tilde{\alpha}_{d;k})). \end{aligned}$$

Thus the lemma follows. □

Lemma 14. *Expectation of* $\log q(\beta|\tilde{\pi})$ *equals to*

$$E_{q(\theta,\beta,\gamma,z)}[\log q(\beta|\tilde{\pi})] = \sum_{k=1}^{K}\log\Gamma(\sum_{f=1}^{F}\tilde{\pi}_{k;f}) - \sum_{k=1}^{K}\sum_{f=1}^{F}\log\Gamma(\tilde{\pi}_{k;f}) + \sum_{k=1}^{K}\sum_{f=1}^{F}(\tilde{\pi}_{k;f}-1)(\psi(\tilde{\pi}_{k;f}) - \psi(\sum_{f=1}^{F}\tilde{\pi}_{k;f})). \tag{22}$$

Proof. As in formula (4), $q(\beta_k) = Dir(\tilde{\pi}_k)$, it is straightforward to have $q(\beta|\tilde{\pi}) = \prod_{k=1}^{K} q(\beta_k|\tilde{\pi}_k) = \prod_{k=1}^{K}\frac{\Gamma(\sum_{f=1}^{F}\tilde{\pi}_{k;f})}{\prod_{f=1}^{F}\Gamma(\tilde{\pi}_{k;f})}\prod_{f=1}^{F}\beta_{k;f}^{\tilde{\pi}_{k;f}-1}$. Hence, we have

$$\begin{aligned} E_{q(\theta,\beta,\gamma,z)}[\log q(\beta|\tilde{\pi})] &= E_{q(\beta)}[\sum_{k=1}^{K}[\log\Gamma(\sum_{f=1}^{F}\tilde{\pi}_{k;f}) - \sum_{f=1}^{F}\log\Gamma(\tilde{\pi}_{k;f}) + \sum_{f=1}^{F}(\tilde{\pi}_{k;f}-1)\log\beta_{k;f}^{\tilde{\pi}_{k;f}-1}]] \\ &= \sum_{k=1}^{K}\log\Gamma(\sum_{f=1}^{F}\tilde{\pi}_{k;f}) - \sum_{k=1}^{K}\sum_{f=1}^{F}\log\Gamma(\tilde{\pi}_{k;f}) + \sum_{k=1}^{K}\sum_{f=1}^{F}(\tilde{\pi}_{k;f}-1)E_{q(\beta)}[\log\beta_{k;f}^{\tilde{\pi}_{k;f}-1}] \\ &= \sum_{k=1}^{K}\log\Gamma(\sum_{f=1}^{F}\tilde{\pi}_{k;f}) - \sum_{k=1}^{K}\sum_{f=1}^{F}\log\Gamma(\tilde{\pi}_{k;f}) + \sum_{k=1}^{K}\sum_{f=1}^{F}(\tilde{\pi}_{k;f}-1)(\psi(\tilde{\pi}_{k;f}) - \psi(\sum_{f=1}^{F}\tilde{\pi}_{k;f})). \end{aligned}$$

Thus the lemma follows. □

Lemma 15. *Expectation of* $\log q(\gamma|\tilde{\lambda})$ *equals to*

$$E_{q(\theta,\beta,\gamma,z)}[\log q(\gamma|\tilde{\lambda})] = \sum_{g=1}^{G}[\tilde{\lambda}_{dg}\log\tilde{\lambda}_{dg} + (1-\tilde{\lambda}_{dg})\log(1-\tilde{\lambda}_{dg})]. \tag{23}$$

Proof. As in formula (4), $q(\gamma_{dg}) = Bernoulli(\tilde{\lambda}_{dg})$, so we have $q(\gamma|\tilde{\lambda}) = \prod_{g=1}^{G} q(\gamma_{dg}|\tilde{\lambda}_{dg}) = \prod_{g=1}^{G}\tilde{\lambda}_{dg}^{\gamma_{dg}}(1-\tilde{\lambda}_{dg})^{(1-\gamma_{dg})}$. Hence, it is straightforward to have that $E_{q(\theta,\beta,\gamma,z)}[\log q(\gamma|\tilde{\lambda})] = \sum_{g=1}^{G}(E_{q(\gamma)}[\gamma_{dg}]\log\tilde{\lambda}_{dg} + E_{q(\gamma)}[(1-\gamma_{dg})]\log(1-\tilde{\lambda}_{dg})) = \sum_{g=1}^{G}(\tilde{\lambda}_{dg}\log\tilde{\lambda}_{dg} + (1-\tilde{\lambda}_{dg})\log(1-\tilde{\lambda}_{dg}))$. □

Lemma 16. *Expectation of* $\log q(z|\tilde{\theta})$ *equals to*

$$E_{q(\theta,\beta,\gamma,z)}[\log q(z|\tilde{\theta})] = \sum_{g=1}^{G}\sum_{k=1}^{K}\tilde{\theta}_{dg;k}\log\tilde{\theta}_{dg;k}. \tag{24}$$

Proof. As in formula (4), $q(z) = Multi(\tilde{\theta})$. Similar to equation (12), it is straightforward to have $q(z|\tilde{\theta}) = \prod_{g=1}^{G} q(z_{dg}|\tilde{\theta}_{dg}) = \prod_{g=1}^{G}\prod_{k=1}^{K} \tilde{\theta}_{dg;k}^{z_{dg;k}}$. Substituting $q(z_{dg})$, we directly have $E_{q(\theta,\beta,\gamma,z)}[\log q(z|\tilde{\theta})]$ $= \sum_{g=1}^{G}\sum_{k=1}^{K} E_{q(z)}[z_{dg;k}] \log(\tilde{\theta}_{dg;k}) = \sum_{g=1}^{G}\sum_{k=1}^{K} \tilde{\theta}_{dg;k} \log \tilde{\theta}_{dg;k}$. □

3.2 Iteration Computation for Variational Parameters

Theorem 1. *Iteration formulas for variational parameters, $\tilde{\alpha}$, $\tilde{\pi}$, $\tilde{\lambda}$ and $\tilde{\theta}$, are:*

$$\begin{cases} \tilde{\alpha}_d^{(t+1)} = \alpha^{(t)} + \sum_{g=1}^{G} \tilde{\theta}_{dg}^{(t)} \\ \tilde{\pi}_k^{(t+1)} = \pi_k^{(t)} + \sum_{g=1}^{G} \cdot\tilde{\theta}_{dg;k}^{(t)} \tilde{f}_{dg}^{(t)} \\ \tilde{\lambda}_{dg}^{(t+1)} = sigmoid(\log \frac{\lambda_{dg}^{(t)}}{1-\lambda_{dg}^{(t)}} a_d^{(t)} p_{dg}^{(a_d^{(t)}-1)}) \\ \tilde{\theta}_{dg;k}^{(t+1)} \propto exp(\sum_{f=1}^{F} \tilde{f}_{dg;f}^{(t)}(\psi(\tilde{\pi}_{k;f}^{(t)}) - \psi(\sum_{f=1}^{F} \tilde{\pi}_{k;f}^{(t)})) + \psi(\tilde{\alpha}_{d;k}^{(t)}) - \psi(\sum_{f=1}^{F} \tilde{\alpha}_{d;k}^{(t)})) \end{cases} \quad (25)$$

Proof. From Proposition 2 and Proposition 3, we know that ELBO equals to

$$\begin{aligned} L(q) = & \sum_{g=1}^{G}\sum_{k=1}^{K}\sum_{f=1}^{F}[\tilde{f}_{dg;f} \cdot \tilde{\theta}_{dg;k} \cdot (\psi(\tilde{\pi}_{k;f}) - \psi(\sum_{f=1}^{F} \tilde{\pi}_{k;f}))] \\ & + \sum_{g=1}^{G} \tilde{\lambda}_{dg} \log[a_d p_{dg}^{a_d-1}] \\ & + \log(\Gamma(\sum_{k=1}^{K} \alpha_{d;k})) - \sum_{k=1}^{K} \log(\Gamma(\alpha_{d;k})) + \sum_{k=1}^{K} (\alpha_{d;k} - 1)(\psi(\tilde{\alpha}_{d;k}) - \psi(\sum_{k=1}^{K} \tilde{\alpha}_{d;k})) \\ & + \sum_{k=1}^{K} \log(\Gamma(\sum_{f=1}^{F} \pi_{k;f})) - \sum_{k=1}^{K}\sum_{f=1}^{F} \log\Gamma(\pi_{k;f}) + \sum_{k=1}^{K}\sum_{f=1}^{F} (\pi_{k;f} - 1)(\psi(\tilde{\pi}_{k;f}) - \psi(\sum_{f=1}^{F} \tilde{\pi}_{k;f})) \\ & + \sum_{g=1}^{G} [\tilde{\lambda}_{dg} \log \lambda_{dg} + (1 - \tilde{\lambda}_{dg}) \log(1 - \lambda_{dg})] \\ & + \sum_{g=1}^{G}\sum_{k=1}^{K} \tilde{\theta}_{dg;k}(\psi(\tilde{\alpha}_{d;k}) - \psi(\sum_{k=1}^{K} \tilde{\alpha}_{d;k})) \\ & - \log\Gamma(\sum_{k=1}^{K} \tilde{\alpha}_{d;k}) + \sum_{k=1}^{K} \log\Gamma(\tilde{\alpha}_{d;k}) - \sum_{k=1}^{K} (\tilde{\alpha}_{d;k} - 1)(\psi(\tilde{\alpha}_{d;k}) - \psi(\sum_{k=1}^{K} \tilde{\alpha}_{d;k})) \\ & - \sum_{k=1}^{K} \log\Gamma(\sum_{f=1}^{F} \tilde{\pi}_{k;f}) + \sum_{k=1}^{K}\sum_{f=1}^{F} \log\Gamma(\tilde{\pi}_{k;f}) - \sum_{k=1}^{K}\sum_{f=1}^{F} (\tilde{\pi}_{k;f} - 1)(\psi(\tilde{\pi}_{k;f}) - \psi(\sum_{f=1}^{F} \tilde{\pi}_{k;f})) \\ & - \sum_{g=1}^{G} [\tilde{\lambda}_{dg} \log \tilde{\lambda}_{dg} + (1 - \tilde{\lambda}_{dg}) \log(1 - \tilde{\lambda}_{dg})] \\ & - \sum_{g=1}^{G}\sum_{k=1}^{K} \tilde{\theta}_{dg;k} \log \tilde{\theta}_{dg;k}. \end{aligned} \quad (26)$$

Based on the variational inference algorithm, the variational parameters $\tilde{\Theta}$ are updated by solving $\frac{\partial L(q)}{\partial \tilde{\Theta}} = 0$. Denote $L(\tilde{\alpha})$ as the sum of all terms that contain $\tilde{\alpha}$ in formula (26), we have $L(\tilde{\alpha}) = \sum_{k=1}^{K} (\alpha_{d;k} - 1)(\psi(\tilde{\alpha}_{d;k}) - \psi(\sum_{k=1}^{K} \tilde{\alpha}_{d;k})) + \sum_{g=1}^{G}\sum_{k=1}^{K} \tilde{\theta}_{dg;k}(\psi(\tilde{\alpha}_{d;k}) - \psi(\sum_{k=1}^{K} \tilde{\alpha}_{d;k})) -$

$\log\Gamma(\sum_{k=1}^{K}\tilde{\alpha}_{d;k})+\sum_{k=1}^{K}\log\Gamma(\tilde{\alpha}_{d;k})-\sum_{k=1}^{K}(\tilde{\alpha}_{d;k}-1)(\psi(\tilde{\alpha}_{d;k})-\psi(\sum_{k=1}^{K}\tilde{\alpha}_{d;k}))=\sum_{k=1}^{K}(\psi(\tilde{\alpha}_{d;k})-\psi(\sum_{k=1}^{K}\tilde{\alpha}_{d;k}))$ $(\sum_{g=1}^{G}\tilde{\theta}_{dg;k}+\alpha_{d;k}-\tilde{\alpha}_{d;k})-\log\Gamma(\sum_{k=1}^{K}\tilde{\alpha}_{d;k})+\sum_{k=1}^{K}\log\Gamma(\tilde{\alpha}_{d;k}$. For simpler illustration, the partial computation of $\frac{\partial L(q)}{\partial\tilde{\alpha}_d}$ is performed in terms of each component of the vector $\tilde{\alpha}_d$. Henthforce, we immediately have $\frac{\partial L(q)}{\partial\tilde{\alpha}_{d;k}}=\frac{\partial L(\tilde{\alpha})}{\partial\tilde{\alpha}_{d;k}}=(\psi'(\tilde{\alpha}_{d;k})-\psi'(\sum_{k=1}^{K}\tilde{\alpha}_{d;k}))(\sum_{g=1}^{G}\tilde{\theta}_{dg;k}+\alpha_{d;k}-\tilde{\alpha}_{d;k})$. Let it equals to zero, and we obtain the iteration formula $\tilde{\alpha}_{d;k}^{(t+1)}=\alpha_{d;k}^{(t)}+\sum_{g=1}^{G}\tilde{\theta}_{dg;k}^{(t)}$. Thus the vectorised form of the iteration is $\tilde{\alpha}_d^{(t+1)}=\alpha^{(t)}+\sum_{g=1}^{G}\tilde{\theta}_{dg}^{(t)}$.

Similarly, we denote $L(\tilde{\pi})$ as the sum of all terms with $\tilde{\pi}$ in formula (26). One immediately have $L(\tilde{\pi})=\sum_{g=1}^{G}\sum_{k=1}^{K}\sum_{f=1}^{F}[\tilde{f}_{dg;f}\cdot\tilde{\theta}_{dg;k}\cdot(\psi(\tilde{\pi}_{k;f})-\psi(\sum_{f=1}^{F}\tilde{\pi}_{k;f}))]+\sum_{k=1}^{K}\sum_{f=1}^{F}(\pi_{k;f}-1)(\psi(\tilde{\pi}_{k;f})-\psi(\sum_{f=1}^{F}\tilde{\pi}_{k;f}))-\sum_{k=1}^{K}\log\Gamma(\sum_{f=1}^{F}\tilde{\pi}_{k;f})+\sum_{k=1}^{K}\sum_{f=1}^{F}\log\Gamma(\tilde{\pi}_{k;f})-\sum_{k=1}^{K}\sum_{f=1}^{F}(\tilde{\pi}_{k;f}-1)(\psi(\tilde{\pi}_{k;f})-\psi(\sum_{f=1}^{F}\tilde{\pi}_{k;f}))=\sum_{k=1}^{K}\sum_{f=1}^{F}(\psi(\tilde{\pi}_{k;f})-\psi(\sum_{f=1}^{F}\tilde{\pi}_{k;f}))(\sum_{g=1}^{G}\tilde{f}_{dg;f}\cdot\tilde{\theta}_{dg;k}+\pi_{k;f}-\tilde{\pi}_{k;f})-\sum_{k=1}^{K}\log\Gamma(\sum_{f=1}^{F}\tilde{\pi}_{k;f})+\sum_{k=1}^{K}\sum_{f=1}^{F}\log\Gamma(\tilde{\pi}_{k;f})$.

Straightforward computation shows that $\frac{\partial L(q)}{\partial\tilde{\pi}_{k;f}}=\frac{\partial L(\tilde{\pi})}{\partial\tilde{\pi}_{k;f}}=(\psi'(\tilde{\pi}_{k;f})-\psi'(\sum_{f=1}^{F}\tilde{\pi}_{k;f}))(\pi_{k;f}-\tilde{\pi}_{k;f}+\sum_{g=1}^{G}\tilde{f}_{dg;f}\cdot\tilde{\theta}_{dg;k})$. Subsequently, $\frac{\partial L(\tilde{\pi})}{\partial\tilde{\pi}_{k;f}}=0$ leads to the iteration formula, $\tilde{\pi}_{k;f}^{(t+1)}=\pi_{k;f}^{(t)}+\sum_{g=1}^{G}\tilde{f}_{dg;f}\cdot\tilde{\theta}_{dg;k}^{(t)}$. Meanwhile, its vectorised form is $\tilde{\pi}_k^{(t+1)}=\pi_k^{(t)}+\sum_{g=1}^{G}\tilde{\theta}_{dg;k}^{(t)}\cdot\tilde{f}_{dg}$.

We denote $L(\tilde{\lambda})$ as the sum of all terms with $\tilde{\lambda}_{dg}$ in formula (26). Then we have $L(\tilde{\lambda})=\sum_{g=1}^{G}(\tilde{\lambda}_{dg}\log\lambda_{dg}+(1-\tilde{\lambda}_{dg})\log(1-\lambda_{dg}))-\sum_{g=1}^{G}(\tilde{\lambda}_{dg}\log\tilde{\lambda}_{dg}+(1-\tilde{\lambda}_{dg})\log(1-\tilde{\lambda}_{dg}))+\sum_{g=1}^{G}\tilde{\lambda}_{dg}\log(a_d p_{dg}^{a_d-1})$. The partial computation of $\frac{\partial L(\tilde{\lambda})}{\partial\tilde{\lambda}_{dg}}$ is $\frac{\partial L(\tilde{\lambda})}{\partial\tilde{\lambda}_{dg}}=\log\lambda_{dg}-\log(1-\lambda_{dg})-\log\tilde{\lambda}_{dg}+\log(1-\tilde{\lambda}_{dg})+\log(a_d p_{dg}^{a_d-1})$. Let $\frac{\partial L(\tilde{\lambda})}{\partial\tilde{\lambda}_{dg}}=0$ we get the iteration formula, $\tilde{\lambda}_{dg}=\frac{1}{1+e^{-w_j}}$ where $w_j=\log\frac{\lambda}{1-\lambda}+\log(a_d p_{dg}^{(a_d-1)})$. By using the notation of Sigmoid function, the iterative formula is simplified as $\tilde{\lambda}_{dg}^{(t+1)}=sigmoid(\log\frac{\lambda_{dg}^{(t)}}{1-\lambda_{dg}^{(t)}}a_d^{(t)}p_{dg}^{(a_d^{(t)}-1)})$.

Besides, we also need calculate the iteration formula of $\tilde{\theta}_{dg;k}$. Taking into account the constraint of $\sum_{k=1}^{K}\tilde{\theta}_{dg;k}=1$ for each g, we denote $L(\tilde{\theta})$ as the sum of all terms with $\tilde{\theta}_{dg;k}$ in formula (26), plus the constraints. One immediately have $L(\tilde{\theta})=\sum_{g=1}^{G}\sum_{k=1}^{K}\sum_{f=1}^{F}[\tilde{f}_{dg;f}\cdot\tilde{\theta}_{dg;k}\cdot(\psi(\tilde{\pi}_{k;f})-\psi(\sum_{f=1}^{F}\tilde{\pi}_{k;f}))]+\sum_{g=1}^{G}\sum_{k=1}^{K}\tilde{\theta}_{dg;k}(\psi(\tilde{\alpha}_{d;k})-\psi(\sum_{k=1}^{K}\tilde{\alpha}_{d;k}))-\sum_{g=1}^{G}\sum_{k=1}^{K}\tilde{\theta}_{dg;k}\log\tilde{\theta}_{dg;k}+\sum_{g=1}^{G}\eta_g(\sum_{k=1}^{K}\tilde{\theta}_{dg;k}-1)$, where η_g is Lagrangian multiplier. Therefore, the partial computation of $L(\tilde{\theta})$ can be calculate as follow, $\frac{\partial L(\tilde{\theta})}{\partial\tilde{\theta}_{dg;k}}=\sum_{f=1}^{F}[\tilde{f}_{dg;f}\cdot(\psi(\tilde{\pi}_{k;f})-\psi(\sum_{f=1}^{F}\tilde{\pi}_{k;f}))+\psi(\tilde{\alpha}_{d;k})-\psi(\sum_{k=1}^{K}\tilde{\alpha}_{d;k})-\log\tilde{\theta}_{dg;k}-1+\eta_g$.

Finally, we get the iteration formula, $\tilde{\theta}_{dg;k}^{(t+1)} \propto exp(\sum_{f=1}^{F} \tilde{f}_{dg;f}(\psi(\tilde{\pi}_{k;f}^{(t)}) - \psi(\sum_{f=1}^{F} \tilde{\pi}_{k;f}^{(t)})) + \psi(\tilde{\alpha}_{d;k}^{(t)}) - \psi(\sum_{f=1}^{F} \tilde{\alpha}_{d;k}^{(t)}))$.

From the above statements, the theorem is proven. □

Theorem 2. *Iteration for computing model parameters, α_d and π_k, is based on Newton's method, $\Theta^{(n+1)} = \Theta^{(n)} - (\mathbf{H}f(\Theta))^{-1} \cdot \nabla f(\Theta)$, where $\mathbf{H}f(\Theta)$ is the Hessian matrix and $\nabla f(\Theta)$ is the gradient of $f(\Theta)$. Then the Newton method iteration of α_d and π_k is based on:*

$$\begin{cases} \mathbf{H}L(\alpha_d)_{kj} = \frac{\partial L(\alpha_d)}{\partial \alpha_{d;k} \partial \alpha_j} = D(\psi'(\sum_{k=1}^{K} \alpha_{d;k}) - \delta(k,j)\psi'(\alpha_{d;k})), \\ \nabla L(\alpha_d) = \frac{\partial L(\alpha_d)}{\partial \alpha_{d;k}} = D(\psi(\sum_{k=1}^{K} \alpha_{d;k}) - \psi(\alpha_{d;k})) + \sum_{d=1}^{D}(\psi(\tilde{\alpha}_{d;k}) - \psi(\sum_{k=1}^{K} \tilde{\alpha}_{d;k})), \\ \mathbf{H}L(\pi_k)_{fj} = \frac{\partial L(\pi_k)}{\partial \pi_{k;f} \partial \pi_{k;j}} = \psi'(\sum_{f=1}^{F} \pi_{k;f}) - \delta(f,j)\psi'(\pi_{k;f}), \\ \nabla L(\pi_k) = \frac{\partial L(\pi_k)}{\partial \pi_{k;f}} = \psi(\sum_{f=1}^{F} \pi_{k;f}) - \psi(\pi_{k;f}) + \psi'(\tilde{\pi}_{k;f}) - \psi'(\sum_{f=1}^{F} \tilde{\pi}_{k;f}), \end{cases} \tag{27}$$

where $\delta(i,j) = \begin{cases} 1, & if\ i = j \\ 0, & if\ i \neq j \end{cases}$.

In the meantime, the iteration of model parameters, λ_{dg} and a_d, is

$$\begin{cases} \lambda_{dg}^{(t+1)} = \tilde{\lambda}_{dg}^{(t)}, \\ a_d^{(t+1)} = -\frac{\sum_{g=1}^{G} \tilde{\lambda}_{dg}^{(t)}}{\sum_{g=1}^{G} (\tilde{\lambda}_{dg}^{(t)} \log p_{dg})}. \end{cases} \tag{28}$$

Proof. Based on the EM algorithm, we need to maximize formula (26) with fixed variational parameters. For model parameters α_d, we denote $L(\alpha_d)$ as the sum of all terms with α_d in formula (26). $L(\alpha_d) = \sum_{d=1}^{D} \log(\Gamma(\sum_{k=1}^{K})) - \sum_{k=1}^{K} \log(\Gamma(\alpha_{d;k})) + \sum_{d=1}^{D} \sum_{k=1}^{K} (\alpha_{d;k} - 1)(\psi(\tilde{\alpha}_{d;k}) - \psi(\sum_{k=1}^{K} \tilde{\alpha}_{d;k})))$

Then we use Newton method to get the optimal solution, and we obtain $\frac{\partial L(\alpha_d)}{\partial \alpha_{d;k}} = D(\psi(\sum_{k=1}^{K} \alpha_{d;k}) - \psi(\alpha_{d;k})) + \sum_{d=1}^{D}(\psi(\tilde{\alpha}_{d;k}) - \psi(\sum_{k=1}^{K} \tilde{\alpha}_{d;k}))$ and $\frac{\partial L(\alpha_d)}{\partial \alpha_d \partial \alpha_j} = D(\psi'(\sum_{k=1}^{K} \alpha_{d;k}) - \delta(k,j)\psi'(\alpha_{d;k}))$.

Similarly, we denote $L(\pi_k)$ as the sum of all terms with π_k in formula (26). $L(\pi_k) = \log(\Gamma(\sum_{f=1}^{F} \pi_{k;f})) - \sum_{f=1}^{F} \log\Gamma(\pi_{k;f}) + \sum_{f=1}^{F} (\pi_{k;f} - 1)(\psi(\tilde{\pi}_{k;f}) - \psi(\sum_{f=1}^{F} \tilde{\pi}_{k;f}))$. Then we calculate gradient $\frac{\partial L(\pi_k)}{\partial \pi_{k;f}} = \psi(\sum_{f=1}^{F} \pi_{k;f}) - \psi(\pi_{k;f}) + \psi'(\tilde{\pi}_{k;f}) - \psi'(\sum_{f=1}^{F} \tilde{\pi}_{k;f})$, and Hessian matrix $\frac{\partial L(\pi_k)}{\partial \pi_{k;f} \partial \pi_{k;j}} = \psi'(\sum_{f=1}^{F} \pi_{k;f}) - \delta(f,j)\psi'(\pi_{k;f})$.

For model parameters λ_{dg}, we denote $L(\lambda_{dg})$ as the sum of all terms with λ_{dg} in formula (26). $L(\lambda_{dg}) = \tilde{\lambda}_{dg} \log \lambda_{dg} + (1 - \tilde{\lambda}_{dg}) \log(1 - \lambda_{dg})$. Through partial computation, $\frac{\partial L(\lambda_{dg})}{\partial \lambda_{dg}} = \frac{\tilde{\lambda}_{dg}}{\lambda_{dg}} - \frac{1 - \tilde{\lambda}_{dg}}{1 - \lambda_{dg}}$, the iteration formula is obtained as $\lambda_{dg}^{(t+1)} = \tilde{\lambda}_{dg}^{(t)}$.

Finally, we calculate $L(a)$ and denote it as the sum of all terms with a_d in formula (26). $L(a) = \sum_{d=1}^{D} \sum_{g=1}^{G} \tilde{\lambda}_{dg} \log(a_d p_{dg}^{(a_d - 1)})$. By partial computation, $\frac{\partial L(a)}{\partial a_d} = \sum_{d=1}^{D} \sum_{g=1}^{G} \tilde{\lambda}_{dg} [\frac{1}{a_d} + \log p_{dg}]$, we obtain the iteration formula, $a_d^{(t+1)} = -\frac{\sum_{g=1}^{G} \tilde{\lambda}_{dg}^{(t)}}{\sum_{g=1}^{G} (\tilde{\lambda}_{dg}^{(t)} \log p_{dg})}$. By concluding all the above statements, the theorem is proven. □

References

[1] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.

[2] Christopher M Bishop. *Pattern recognition and machine learning.* springer, 2006.

[3] Mingwei Dai, Jingsi Ming, Mingxuan Cai, Jin Liu, Can Yang, Xiang Wan, and Zongben Xu. Igess: a statistical approach to integrating individual-level genotype data and summary statistics in genome-wide association studies. *Bioinformatics*, 33(18):2882–2889, 2017.

[4] Mingwei Dai, Xiang Wan, Hao Peng, Yao Wang, Yue Liu, Jin Liu, Zongben Xu, and Can Yang. Joint analysis of individual-level and summary-level GWAS data by leveraging pleiotropy. *arXiv preprint arXiv:1804.11011*, 2018.