Infinite Latent Process Decomposition

Tomonari Masada
Department of Computer and Information Sciences
Nagasaki University
1-14 Bunkyo-machi, Nagasaki-shi, Nagasaki, Japan
masada@cis.nagasaki-u.ac.jp

October 7, 2010

Abstract

This paper proposes a new Bayesian probabilistic model targeting microarray data. We extend latent process decomposition (LPD) [3] so that we can assume that there are *infinite* latent processes. We call the proposed model *infinite latent process decomposition (iLPD)*. Further, we provide a collapsed variational Bayesian (CVB) inference for iLPD. Our CVB improves the CVB proposed for LPD in [8] with respect to the following two aspects. First, our CVB realizes a more efficient inference by treating full posterior distributions over the hyperparameters of Dirichlet variables based on the discussions in [6]. Second, we correct the weakness of CVB in [8], which makes the evaluation of the variational lower bound of the log evidence dependent on the ordering of genes. This dependency is removed by applying the second order approximation proposed in [6]. These two spects are independent of the assumption of infinite latent processes. Therefore, our CVB can also be applied to the original LPD. The experiment comparing iLPD with LPD by using the proposed CVB and also with LDA by using the CVB in [8] is under progress. This paper mainly includes the details of the model construction of iLPD.

1 Introduction

This paper proposes an extension of latent process decomposition (LPD) [3]. In this new version of LPD, we can assume that there are infinite latent processes. We denote the model as iLPD, which is an abbreviation of *infinite latent process decomposition*. Further, we provide a set of update formulas of collapsed variational Bayesian (CVB) inference for iLPD. This paper includes the full details of iLPD and its CVB inference.

The rest of the paper is organized as follows. Section 2 provides the previous works related to LPD and also to the assumtion of infinite topics in the field of text mining. In Section 3, iLPD is described in its full details. Section 4 provides all update formulas required for implementing CVB for iLPD. Section 5 shows how to obtain the lower bound of the log evidence, which is required when we evaluate the efficiency of iLPD and compare iLPD with LPD. Section 6 concludes the paper with planned future work.

2 Previous Works

By regarding samples as documents, genes as words, and latent processes as latent topics, we can grasp LPD [3] as a "bioinformatics variant" of latent Dirichlet allocation (LDA) [1]. Therefore, we can say that our iLPD extends LPD just as hierarchical Dirichlet process (HDP) [5] extends LDA by assuming that there are infinite latent processes. Further, the CVB for LDA originally proposed in [7] is greatly improved by the CVB proposed in [6], because we can treat full posterior distributions over the hyperparameters of Dirichlet variables based on the discussions in [6]. The improved CVB can be applied to both LDA and HDP. In a similar manner, we provide an improved CVB applicable to both LPD and iLPD. Our CVB inference is more efficient than the CVB for LPD proposed in [8] with respect to the following two aspects:

1. We introduce auxiliary variables by following the approach proposed in [6]. This approach treats full posterior distributions over the hyperparameters of Dirichlet variables and is independent of the assumption of infinite latent processes. Therefore, our CVB is also applicable to LPD after a small modification.

2. We use a more natural approximation technique in computing the variational lower bound of the log evidence than [8]. The approximation proposed in [8] is not technically natural, because the lower bound computation depends on the ordering of genes. Therefore, we use the second order approximation technique proposed in [6] and remove the dependence on the ordering of genes.

LPD has a completely different model construction with respect to the gene expression data when compared with the model construction of LDA related to the word frequencies. The expression data are continuous data and are modeled by Gaussian distributions, though the word frequencies are discrete data and are modeled by multinomial distributions in LDA and HDP. Therefore, while our proposal is heavily based on the discussions in [6], it is not a trivial task to obtain iLPD from LPD and further to obtain a CVB for iLPD from CVB for LPD.

3 Infinite Latent Process Decomposition (iLPD)

3.1 Generative description of iLPD

In this paper, we identify various types of entities appearing in our probabilistic model with their indices as below:

- $\{1, \ldots, D\}$: the set of samples,
- $\{1, \ldots, G\}$: the set of genes, and
- $\{1, \ldots, K\}$: the set of latent processes.

We give a generative description of iLPD below. Note that, by regarding samples as documents, genes as words, and latent processes as latent topics, we can grasp iLPD as a "bioinformatics variant" of HDP [5]. Therefore, the description below can be understand in parallel with the description of HDP.

- For each sample d, the parameter $\theta_d = (\theta_{d1}, \ldots, \theta_{dK})$ of the multinomial distribution $\texttt{Multi}(\theta_d)$, defined over latent processes $\{1, \ldots, K\}$, is drawn from the Dirichlet process $\texttt{DP}(\alpha, \pi)$.
  - We use the stick-breaking construction [4] for the center measure π of $\texttt{DP}(\alpha, \pi)$. We denote the parameter of the single parameter Beta distribution $\texttt{Beta}(1, \gamma)$ appearing in the stick-breaking construction as γ, which is in turn drawn from the Gamma distribution $\texttt{Gamma}(a_\gamma, b_\gamma)$.
  - The concentration parameter α of $\texttt{DP}(\alpha, \pi)$ is drawn from the Gamma distribution $\texttt{Gamma}(a_\alpha, b_\alpha)$.
- For each pair of gene g and latent process k, a mean parameter μ_{gk} and a precision parameter λ_{gk} of the Gaussian distribution $\texttt{Gauss}(\mu_{gk}, \lambda_{gk})$ are drawm from the Gaussian prior distribution $\texttt{Gauss}(\mu_0, \rho)$ and the Gamma prior distribution $\texttt{Gamma}(a_0, b_0)$, respectively.
  - We assume that the precision parameter ρ of the Gamma prior $\texttt{Gauss}(\mu_0, \rho)$ is in turn drawn from the Gamma distribution $\texttt{Gamma}(a_\rho, b_\rho)$.
- For each pair of sample d and gene g, a latent process is drawn from the multinomial distribution $\texttt{Multi}(\theta_d)$. Let z_{dg} be the latent variable whose value is this drawn process.
- Based on the process z_{dg} drawn from $\texttt{Multi}(\theta_d)$ for the pair of sample d and gene g, a real number is drawn from the Gaussian distribution $\texttt{Gauss}(\mu_{gz_{dg}}, \lambda_{gz_{dg}})$. Let x_{dg} be the observed variable whose value is this drawn real number. x_{dg} corresponds to the expression level in the microarray data.

By using the two Gamma distributions $\texttt{Gamma}(a_\alpha, b_\alpha)$ and $\texttt{Gamma}(a_\gamma, b_\gamma)$, we can treat full posterior distributions over the hyperparameters α and γ of the Dirichlet process $\texttt{DP}(\alpha, \pi)$. This is a remarkable achievement given in [6]. Therefore, we apply this technique to our CVB inference. This technique can also be applied to the latent topic Dirichlet prior of LDA and to the latent process Dirichlet prior of LPD. The details of this application can be deduced from the discussions on the word Dirichlet prior $\texttt{Dirichlet}(\beta, \tau)$ in [6], because the number of different words is assumed to be finite in [6] just like the number of latent topics (resp. latent processes) is assumed to be finite in LDA (resp. LPD).

3.2 Joint distribution

Based on the generative description in Section 3.1, we can give the full joint distribution of iLPD as follows:

$$
\begin{aligned}
&p(\mathbf{x}, \mathbf{z}, \theta, \mu, \lambda, \rho, \alpha, \pi, \gamma | \mu_0, a_0, b_0, a_\rho, b_\rho, a_\alpha, b_\alpha, a_\gamma, b_\gamma) \\
&= \prod_d p(\theta_d|\alpha,\pi) \cdot p(\alpha|a_\alpha, b_\alpha) \cdot p(\tilde{\pi}|\gamma) \cdot p(\gamma|a_\gamma, b_\gamma) \cdot \prod_{g,k} p(\mu_{gk}|\mu_0, \rho) \cdot p(\rho|a_\rho, b_\rho) \cdot \prod_{g,k} p(\lambda_{gk}|a_0, b_0) \\
&\quad \cdot \prod_d \prod_g p(z_{dg}|\theta_d) p(x_{dg}|\mu_{gz_{dg}}, \lambda_{gz_{dg}}) \\
&= \prod_d \frac{\Gamma(\alpha)}{\prod_k \Gamma(\alpha\pi_k)} \prod_k \theta_{dk}^{\alpha\pi_k - 1} \cdot \frac{a_\alpha^{b_\alpha}}{\Gamma(a_\alpha)} \alpha^{a_\alpha - 1} e^{-b_\alpha \alpha} \cdot \prod_{k=1}^{K} \frac{\Gamma(1+\gamma)}{\Gamma(1)\Gamma(\gamma)} \tilde{\pi}_k^{1-1} (1-\tilde{\pi}_k)^{\gamma-1} \cdot \frac{a_\gamma^{b_\gamma}}{\Gamma(a_\gamma)} \gamma^{a_\gamma - 1} e^{-b_\gamma \gamma} \\
&\quad \cdot \prod_g \prod_k \sqrt{\frac{\rho}{2\pi}} \exp\Big\{ -\frac{\rho}{2}(\mu_{gk} - \mu_0)^2 \Big\} \cdot \frac{a_\rho^{b_\rho}}{\Gamma(a_\rho)} \rho^{a_\rho - 1} e^{-b_\rho \rho} \cdot \prod_g \prod_k \frac{b_0^{a_0}}{\Gamma(a_0)} \lambda_{gk}^{a_0 - 1} e^{-b_0 \lambda_{gk}} \\
&\quad \cdot \prod_d \frac{n_d!}{\prod_k n_{dk}!} \prod_k \theta_{dk}^{n_{dk}} \cdot \prod_d \prod_g \prod_k \Bigg[\sqrt{\frac{\lambda_{gk}}{2\pi}} \exp\Big\{ -\frac{\lambda_{gk}}{2}(x_{dg} - \mu_{gk})^2 \Big\} \Bigg]^{n_{dgk}} , \qquad (1)
\end{aligned}
$$

where n_{dgk} is equal to one when gene g in sample d is assigned to latent process k and is equal to zero otherwise. Further, we define $n_{dk} \equiv \sum_g n_{dgk}$.

In Eq. (1), $p(\tilde{\pi}|\gamma)$ denotes the density function of the the single parameter Beta distribution $\texttt{Beta}(1,\gamma)$ appearing in the stick breaking construction for π. Between the values $\tilde{\pi}_k$ drawn from $\texttt{Beta}(1,\gamma)$ and the parameters π_k of the center measure of the Dirichlet process $\texttt{DP}(\alpha,\pi)$, the following equation holds:

$$
\pi_k = \tilde{\pi}_k \prod_{l=1}^{k-1} (1 - \tilde{\pi}_l) \ . \qquad (2)
$$

3.3 Introducing augxiliary variables

By marginalizing out the latent process multinomial parameters θ_d for each sample d, we obtain the following:

$$
\begin{aligned}
&p(\mathbf{x}, \mathbf{z}, \mu, \lambda, \alpha, \rho, \pi, \gamma | \mu_0, a_0, b_0, a_\rho, b_\rho, a_\alpha, b_\alpha, a_\gamma, b_\gamma) = \int p(\mathbf{x}, \mathbf{z}, \theta, \mu, \lambda, \alpha, \rho, \pi, \gamma | \mu_0, a_0, b_0, a_\rho, b_\rho, a_\alpha, b_\alpha, a_\gamma, b_\gamma) d\theta \\
&= \prod_d \frac{\Gamma(\alpha)}{\Gamma(n_d + \alpha)} \prod_k \frac{\Gamma(n_{dk} + \alpha\pi_k)}{\Gamma(\alpha\pi_k)} \cdot \frac{a_\alpha^{b_\alpha}}{\Gamma(a_\alpha)} \alpha^{a_\alpha - 1} e^{-b_\alpha \alpha} \cdot \prod_{k=1}^{K} \frac{\Gamma(1+\gamma)}{\Gamma(1)\Gamma(\gamma)} \tilde{\pi}_k^{1-1} (1-\tilde{\pi}_k)^{\gamma-1} \cdot \frac{a_\gamma^{b_\gamma}}{\Gamma(a_\gamma)} \gamma^{a_\gamma - 1} e^{-b_\gamma \gamma} \\
&\quad \cdot \prod_g \prod_k \sqrt{\frac{\rho}{2\pi}} \exp\Big\{ -\frac{\rho}{2}(\mu_{gk} - \mu_0)^2 \Big\} \cdot \frac{a_\rho^{b_\rho}}{\Gamma(a_\rho)} \rho^{a_\rho - 1} e^{-b_\rho \rho} \cdot \prod_g \prod_k \frac{b_0^{a_0}}{\Gamma(a_0)} \lambda_{gk}^{a_0 - 1} e^{-b_0 \lambda_{gk}} \\
&\quad \cdot \prod_d \prod_g \prod_k \Bigg[\sqrt{\frac{\lambda_{gk}}{2\pi}} \exp\Big\{ -\frac{\lambda_{gk}}{2}(x_{dg} - \mu_{gk})^2 \Big\} \Bigg]^{n_{dgk}} . \qquad (3)
\end{aligned}
$$

Now we introduce the auxiliary variables $\boldsymbol{\eta}$ and $\mathbf{s}$ to obtain efficient variational updates [6] as follows:

$$
p(\mathbf{z}, \boldsymbol{\eta}, \mathbf{s}|\alpha, \pi) = p(\boldsymbol{\eta}|\alpha) p(\mathbf{s}, \mathbf{z}|\alpha, \pi) = \prod_d \frac{\eta_d^{\alpha-1}(1-\eta_d)^{n_d - 1} \prod_k \begin{bmatrix} n_{dk} \\ s_{dk} \end{bmatrix} (\alpha\pi_k)^{s_{dk}}}{\Gamma(n_d)} \ . \qquad (4)
$$

Then, the following equation holds by marginalizing out these auxiliary variables:

$$
p(\mathbf{z}|\alpha, \pi) = \int \sum_{\mathbf{s}} p(\mathbf{z}, \boldsymbol{\eta}, \mathbf{s}|\alpha, \pi) d\boldsymbol{s} d\eta = \prod_d \frac{\Gamma(\alpha)}{\Gamma(n_d + \alpha)} \prod_k \frac{\Gamma(n_{dk} + \alpha\pi_k)}{\Gamma(\alpha\pi_k)} \ . \qquad (5)
$$

After introducing the auxiliary variables, the distribution in Eq. (3) can be rewritten as follows:

$$\begin{aligned}
&p(\mathbf{x}, \mathbf{z}, \mu, \lambda, \rho, \alpha, \pi, \gamma | \mu_0, a_0, b_0, a_\rho, b_\rho, a_\alpha, b_\alpha, a_\gamma, b_\gamma) \\
&= p(\boldsymbol{\eta}|\alpha) p(\mathbf{s}, \mathbf{z}|\alpha, \pi) p(\alpha|a_\alpha, b_\alpha) p(\tau|a_\pi) p(\tilde{\pi}|\gamma) p(\gamma|a_\gamma, b_\gamma) p(\mathbf{x}|\mathbf{z}, \mu, \lambda) p(\lambda|a_0, b_0) p(\mu|\mu_0, \rho) p(\rho|a_\rho, b_\rho) p(\alpha|a_\alpha, b_\alpha) \\
&= \prod_d \frac{\eta_d^{\alpha-1}(1-\eta_d)^{n_d-1} \prod_k \begin{bmatrix} n_{dk} \\ s_{dk} \end{bmatrix} (\alpha \pi_k)^{s_{dk}}}{\Gamma(n_d)} \cdot \frac{a_\alpha^{b_\alpha}}{\Gamma(a_\alpha)} \alpha^{a_\alpha - 1} e^{-b_\alpha \alpha} \cdot \prod_{k=1}^K \gamma (1-\tilde{\pi}_k)^{\gamma-1} \cdot \frac{a_\gamma^{b_\gamma}}{\Gamma(a_\gamma)} \gamma^{a_\gamma - 1} e^{-b_\gamma \gamma} \\
&\quad \cdot \prod_g \prod_k \sqrt{\frac{\rho}{2\pi}} \exp\left\{ -\frac{\rho}{2}(\mu_{gk} - \mu_0)^2 \right\} \cdot \frac{a_\rho^{b_\rho}}{\Gamma(a_\rho)} \rho^{a_\rho - 1} e^{-b_\rho \rho} \cdot \prod_g \prod_k \frac{b_0^{a_0}}{\Gamma(a_0)} \lambda_{gk}^{a_0 - 1} e^{-b_0 \lambda_{gk}} \\
&\quad \cdot \prod_d \prod_g \prod_k \left[\sqrt{\frac{\lambda_{gk}}{2\pi}} \exp\left\{ -\frac{\lambda_{gk}}{2} (x_{dg} - \mu_{gk})^2 \right\} \right]^{n_{dgk}} . \qquad (6)
\end{aligned}$$

3.4 A lower bound of the log evidence

The marginalized likelihood $p(\mathbf{x})$ of the observed data $\mathbf{x}$ is often called *evidence*. By following a regular habit of variational inferences, we introduce a variational posterior distribution $q(\mathbf{z}, \boldsymbol{\eta}, \mathbf{s}, \mu, \lambda, \rho, \alpha, \pi)$ and apply Jensen's inequality to obtain a lower bound of the log of the evidence as follows:

$$\begin{aligned}
&\log p(\mathbf{x}|\mu_0, a_0, b_0, a_\rho, b_\rho, a_\alpha, b_\alpha, a_\gamma, b_\gamma) \\
&= \log \int \sum_{\mathbf{z}} \sum_{\mathbf{s}} p(\mathbf{x}, \mathbf{z}, \boldsymbol{\eta}, \mathbf{s}, \mu, \lambda, \rho, \alpha, \pi, \gamma | \mu_0, a_0, b_0, a_\rho, b_\rho, a_\alpha, b_\alpha, a_\gamma, b_\gamma) d\boldsymbol{\eta} d\mu d\lambda d\rho d\alpha d\pi d\gamma \\
&= \log \int \sum_{\mathbf{z}} \sum_{\mathbf{s}} q(\mathbf{z}, \boldsymbol{\eta}, \mathbf{s}, \mu, \lambda, \rho, \alpha, \pi, \gamma) \frac{p(\mathbf{x}, \mathbf{z}, \boldsymbol{\eta}, \mathbf{s}, \mu, \lambda, \rho, \alpha, \pi, \gamma | \mu_0, a_0, b_0, a_\rho, b_\rho, a_\alpha, b_\alpha, a_\gamma, b_\gamma)}{q(\mathbf{z}, \boldsymbol{\eta}, \mathbf{s}, \mu, \lambda, \rho, \alpha, \pi)} d\boldsymbol{\eta} d\mu d\lambda d\rho d\alpha d\pi d\gamma \\
&\geq \int \sum_{\mathbf{z}} \sum_{\mathbf{s}} q(\mathbf{z}, \boldsymbol{\eta}, \mathbf{s}, \mu, \lambda, \rho, \alpha, \pi, \gamma) \log \frac{p(\mathbf{x}, \mathbf{z}, \boldsymbol{\eta}, \mathbf{s}, \mu, \lambda, \rho, \alpha, \pi, \gamma | \mu_0, a_0, b_0, a_\rho, b_\rho, a_\alpha, b_\alpha, a_\gamma, b_\gamma)}{q(\mathbf{z}, \boldsymbol{\eta}, \mathbf{s}, \mu, \lambda, \rho, \alpha, \pi)} d\boldsymbol{\eta} d\mu d\lambda d\rho d\alpha d\pi d\gamma
\end{aligned} \qquad (7)$$

Let the right hand side, i.e., the lower bound of the log evidence, be referred to by $\mathcal{L}$ for the rest of the paper.

3.5 Posterior factorization assumption

We assume that $q(\mathbf{z}, \boldsymbol{\eta}, \mathbf{s}, \mu, \lambda, \rho, \alpha, \pi)$ can be factorized as$q(\boldsymbol{\eta}, \mathbf{s}|\mathbf{z}) q(\mathbf{z}) q(\mu) q(\lambda) q(\rho) q(\alpha) q(\pi) q(\gamma)$. Then, $\mathcal{L}$ can be written as follows:

$$\begin{aligned}
&\mathcal{L} \\
&= \int \sum_{\mathbf{z}} \sum_{\mathbf{s}} q(\boldsymbol{\eta}, \mathbf{s}|\mathbf{z}) q(\mathbf{z}) q(\mu) q(\lambda) q(\rho) q(\alpha) q(\pi) q(\gamma) \\
&\qquad \log \frac{p(\boldsymbol{\eta}|\alpha) p(\mathbf{s}, \mathbf{z}|\alpha, \tau) p(\mathbf{x}|\mathbf{z}, \mu, \lambda) p(\alpha|a_\alpha, b_\alpha) p(\tilde{\pi}|\gamma)(\mu|\mu_0, \rho) p(\lambda|a_0, b_0) p(\rho|a_\rho, b_\rho) p(\gamma|a_\gamma, b_\gamma)}{q(\boldsymbol{\eta}, \mathbf{s}|\mathbf{z}) q(\mathbf{z}) q(\alpha) q(\pi) q(\mu) q(\lambda) q(\rho) q(\gamma)} d\boldsymbol{\eta} d\alpha d\pi d\mu d\lambda d\rho d\gamma
\end{aligned} \qquad (8)$$

By taking a functional derivative of $\mathcal{L}$ in Eq. (8) with respect to $q(\boldsymbol{\eta}, \mathbf{s}|\mathbf{z})$, it can be shown that $\mathcal{L}$ is maximized when $q(\boldsymbol{\eta}, \mathbf{s}|\mathbf{z})$ is equal to $p(\boldsymbol{\eta}, \mathbf{s}|\mathbf{x}, \mathbf{z}, \alpha, \pi, \mu, \lambda)$. By replacing $q(\boldsymbol{\eta}, \mathbf{s}|\mathbf{z})$ with $p(\boldsymbol{\eta}, \mathbf{s}|\mathbf{x}, \mathbf{z}, \alpha, \pi, \mu, \lambda)$ in Eq. (8), we obtain the following simplified form of $\mathcal{L}$:

$$\begin{aligned}
\mathcal{L} = &\int \sum_{\mathbf{z}} q(\mathbf{z}) q(\mu) q(\lambda) q(\rho) q(\alpha) q(\pi) q(\gamma) \\
&\qquad \log \frac{p(\mathbf{x}, \mathbf{z}|\alpha, \pi, \mu, \lambda) p(\alpha|a_\alpha, b_\alpha) p(\tilde{\pi}|\gamma) p(\mu|\mu_0, \rho) p(\lambda|a_0, b_0) p(\rho|a_\rho, b_\rho) p(\gamma|a_\gamma, b_\gamma)}{q(\mathbf{z}) q(\mu) q(\lambda) q(\rho) q(\alpha) q(\pi) q(\gamma)} d\mu d\lambda d\rho d\alpha d\pi d\gamma .
\end{aligned} \qquad (9)$$

The lower bound in Eq. (9) will be used to derive the update formula for $q(\mathbf{z})$ in Section 4.5.

In Eq. (9), we set $q(\boldsymbol{\eta}, \mathbf{s}|\mathbf{z})$ to be equal to $p(\boldsymbol{\eta}, \mathbf{s}|\mathbf{x}, \mathbf{z}, \alpha, \tau, \mu, \lambda)$ and maximize $\mathcal{L}$. On the other hand, $\boldsymbol{\eta}$ and $\mathbf{s}$ are decoupled in Eq. (4). Therefore, we further assume that $q(\boldsymbol{\eta}, \mathbf{s}|\mathbf{z})$ are factorized as $q(\mathbf{s}|\mathbf{z})q(\mathbf{s}|\mathbf{z})$. Then, by rewriting $\mathcal{L}$ in Eq. (8), we obtain the following result:

$$
\begin{aligned}
\mathcal{L} = & \int \sum_{\mathbf{z}} q(\boldsymbol{\eta}|\mathbf{z})q(\mathbf{z})q(\alpha) \log p(\boldsymbol{\eta}|\alpha) d\boldsymbol{\eta} d\alpha + \int \sum_{\mathbf{z}} \sum_{\mathbf{s}} q(\mathbf{s}|\mathbf{z})q(\mathbf{z})q(\alpha)q(\pi) \log p(\mathbf{s}|\mathbf{z}, \alpha, \pi) d\alpha d\tau \\
& + \int q(\alpha) \log p(\alpha|a_\alpha, b_\alpha) d\alpha + \int q(\pi) \log p(\tilde{\pi}|\gamma) d\pi + \int q(\gamma) \log p(\gamma|a_\gamma, b_\gamma) d\gamma \log p(\mathbf{x}|\mathbf{z}, \mu, \lambda) d\mu d\lambda \\
& + \int \sum_{\mathbf{z}} q(\mathbf{z})q(\mu)q(\lambda) + \int q(\mu)q(\rho) \log p(\mu|\mu_0, \rho) d\mu + \int q(\lambda) \log p(\lambda|a_0, b_0) d\lambda + \int q(\rho) \log p(\rho|a_\rho, b_\rho) d\rho \\
& - \int \sum_{\mathbf{z}} q(\boldsymbol{\eta}|\mathbf{z})q(\mathbf{z}) \log q(\boldsymbol{\eta}|\mathbf{z}) d\boldsymbol{\eta} - \sum_{\mathbf{z}} \sum_{\mathbf{s}} q(\mathbf{s}|\mathbf{z})q(\mathbf{z}) \log q(\mathbf{s}|\mathbf{z}) - \sum_{\mathbf{z}} q(\mathbf{z}) \log q(\mathbf{z}) \\
& - \int q(\alpha) \log q(\alpha) d\alpha - \int q(\pi) \log q(\pi) d\pi - \int q(\gamma) \log q(\gamma) d\gamma \\
& - \int q(\mu) \log q(\mu) d\mu - \int q(\lambda) \log q(\lambda) d\lambda - \int q(\rho) \log q(\rho) d\rho \ . \qquad (10)
\end{aligned}
$$

The lower bound in Eq. (10) will be used to derive the update formulas in Section 4.

Finally, we assume that $q(\mathbf{z})$ can be factorized as $q(\mathbf{z}) = \prod_d \prod_g q(z_{dg})$. Note that $\sum_{k=1}^{K} q(z_{dg} = k) = 1$ is satisfied for every pair of sample d and gene g.

4 Posterior Updates

In this section, by taking the functional derivative of the lower bound $\mathcal{L}$ with respect to each factor of the variational posterior $q(\mathbf{z}, \boldsymbol{\eta}, \mathbf{s}, \mu, \lambda, \rho, \alpha, \pi) = q(\boldsymbol{\eta}|\mathbf{z})q(\mathbf{s}|\mathbf{z})q(\mathbf{z})q(\mu)q(\lambda)q(\alpha)q(\pi)$, we obtain the function form of each factor.

4.1 Posteriors inheritable from CVB for HDP

For the variational posteriors $q(\alpha)$, $q(\pi)$, $q(\gamma)$, $q(\eta_d|\mathbf{z}_d)$, and $q(s_{dk}|z_{dk})$, we can use the results of CVB for HDP [6] as is. Therefore, we only show the resulting function forms below.

$$
\begin{aligned}
q(\alpha) &\propto e^{\alpha(-b_\alpha + \sum_d \mathbb{E}[\log \eta_d])} \alpha^{a_\alpha + \mathbb{E}[s_{\cdot\cdot}] - 1} && (11) \\
q(\tilde{\pi}_k) &\propto \tilde{\pi}_k^{\mathbb{E}[s_{\cdot k}]} (1 - \tilde{\pi}_k)^{\mathbb{E}[s_{\cdot > k}] + \mathbb{E}[\gamma] - 1} && (12) \\
q(\gamma) &\propto e^{-\gamma(b_\gamma - \sum_k \mathbb{E}[\log(1-\tilde{\pi}_k)])} \gamma^{a_\gamma + K - 1} && (13) \\
q(\eta_d) &\propto \eta_d^{\mathbb{E}[\alpha] - 1} (1 - \eta_d)^{n_d - 1} && (14) \\
q(s_{dk}|z_{dk}) &\propto \begin{bmatrix} n_{dk} \\ s_{dk} \end{bmatrix} e^{s_{dk} \mathbb{E}[\log \alpha]} e^{s_{dk} \mathbb{E}[\log \pi_k]} \ , && (15)
\end{aligned}
$$

where we define $s_{\cdot k} \equiv \sum_d s_{dk}$, $s_{\cdot\cdot} \equiv \sum_d \sum_k s_{dk}$, and $s_{\cdot > k} \equiv \sum_d \sum_{l > k} s_{dk}$. The formulas above include many expectations $\mathbb{E}[\cdot]$ taken with respect to the variational posteriors. Also for these expectations, we can use the results presented in [6] as is. For completeness, we include the evaluation formulas of these expectations below.

$$
\begin{aligned}
\mathbb{E}[\log \eta_d] &= \Psi(\mathbb{E}[\alpha]) - \Psi(n_d + \mathbb{E}[\alpha]) && (16) \\
\mathbb{E}[s_{dk}] &\approx \mathbb{G}[\alpha]\mathbb{G}[\pi_k] \Big\{ 1 - \prod_i q(z_{id} \neq k) \Big\} \\
& \quad \cdot \Big\{ \Psi(\mathbb{E}_+[n_{dk}] + \mathbb{G}[\alpha]\mathbb{G}[\pi_k]) - \Psi(\mathbb{G}[\alpha]\mathbb{G}[\pi_k]) + \frac{1}{2} \mathbb{V}_+[n_{dk}] \Psi''(\mathbb{E}_+[n_{dk}] + \mathbb{G}[\alpha]\mathbb{G}[\pi_k]) \Big\} && (17) \\
\mathbb{E}[\log \alpha] &= \Psi(a_\alpha + \mathbb{E}[s_{\cdot\cdot}]) - \log \Big(b_\alpha - \sum_d \mathbb{E}[\log \eta_d] \Big) && (18) \\
\mathbb{E}[\log \pi_k] &= \Psi(\mathbb{E}[s_{\cdot k}] + 1) + \sum_{l=1}^{k-1} \Psi(\mathbb{E}[s_{\cdot > l}] + \mathbb{E}[\gamma]) - \sum_{l=1}^{k} \Psi(\mathbb{E}[s_{\cdot \geq l}] + \mathbb{E}[\gamma] + 1) && (19) \\
\mathbb{E}[\log(1 - \tilde{\pi}_k)] &= \Psi(\mathbb{E}[s_{\cdot > k}] + \mathbb{E}[\gamma]) - \Psi(\mathbb{E}[s_{\cdot \geq k}] + \mathbb{E}[\gamma] + 1) \ , && (20)
\end{aligned}
$$

where

$$\mathbb{E}_+[n_{dk}] = \frac{\mathbb{E}[n_{dk}]}{1-\prod_g q(z_{dg} \neq k)} \ , \tag{21}$$

$$\mathbb{V}_+[n_{dk}] = \frac{\mathbb{V}[n_{dk}]}{1-\prod_g q(z_{dg} \neq k)} - \mathbb{E}_+[n_{dk}]^2 \prod_g q(z_{dg} \neq k) \ . \tag{22}$$

Our CVB inference uses the special mean $\mathbb{E}_+[n_{dk}]$ and the special variance $\mathbb{V}_+[n_{dk}]$ for n_{dk}, which are proposed in [6] for treating the case $n_{dk} = 0$ exactly. This technique makes our CVB more efficient than the CVB in [8].

4.2 Mean posteriors $q(\mu)$

By taking a functional derivative of $\mathcal{L}$ with respect to $q(\mu)$, we obtain

$$\frac{\delta \mathcal{L}}{\delta q(\mu)} = \int \sum_{\mathbf{z}} q(\mathbf{z}) q(\lambda) \log p(\mathbf{x}|\mathbf{z}, \mu, \lambda) d\lambda + \int q(\rho) \log p(\mu|\mu_0, \rho) d\rho - \log q(\mu) + const. \tag{23}$$

Therefore, $q(\mu)$ can be written as follows:

$$q(\mu) \propto \exp\Big\{ \int q(\rho) \log p(\mu|\mu_0, \rho) d\rho \Big\} \exp\Big\{ \int \sum_{\mathbf{z}} q(\mathbf{z}) q(\lambda) \log p(\mathbf{x}|\mathbf{z}, \mu, \lambda) d\lambda \Big\} \ . \tag{24}$$

The integral appearing in the first exponential function can be evaluated as follows:

$$\int q(\rho) \log p(\mu|\mu_0, \rho) d\rho = -\frac{1}{2} \sum_{g,k} (\mu_{gk} - \mu_0)^2 \int q(\rho) \rho d\rho + const. = -\frac{\mathbb{E}[\rho]}{2} \sum_{g,k} (\mu_{gk} - \mu_0)^2 + const. \tag{25}$$

where we regard every term not related to μ as a constant. The integral inside the second exponential function in Eq. (24) can be evaluated as follows:

$$\begin{aligned}
&\int \sum_{\mathbf{z}} q(\mathbf{z}) q(\lambda) \log p(\mathbf{x}|\mathbf{z}, \mu, \lambda) d\lambda = \int \sum_{\mathbf{z}} q(\mathbf{z}) q(\lambda) \log \Bigg[\prod_g \prod_k \prod_d \Bigg[\sqrt{\frac{\lambda_{gk}}{2\pi}} \exp\Big\{ -\frac{\lambda_{gk}}{2}(x_{dg} - \mu_{gk})^2 \Big\} \Bigg]^{n_{dgk}} \Bigg] d\lambda_{gk} \\
&= \sum_{d,g} \sum_{k=1}^{K} q(z_{dg} = k) \int q(\lambda_{gk}) \log \Bigg[\sqrt{\frac{\lambda_{gk}}{2\pi}} \exp\Big\{ -\frac{\lambda_{gk}}{2}(x_{dg} - \mu_{gk})^2 \Big\} \Bigg] d\lambda_{gk} \\
&= -\sum_{g,k} \mathbb{E}[\lambda_{gk}] \sum_d \frac{q(z_{dg} = k)(x_{dg} - \mu_{gk})^2}{2} + const. = -\sum_{g,k} \mathbb{E}[\lambda_{gk}] \frac{\mathbb{E}[n_{gk}]\mu_{gk}^2 - 2\mu_{gk}\overline{x}_{gk}}{2} + const. \ ,
\end{aligned} \tag{26}$$

where we regard every term not related to μ as a constant. In Eq. (26), we refer to $\sum_d q(z_{dg} = k)$ by $\mathbb{E}[n_{gk}]$, i.e., the expected frequency of the assignment of gene g to latent process k. Further, we define $\overline{x}_{gk} \equiv \sum_d q(z_{dg} = k) x_{dg}$.

By combining Eq. (25) and Eq. (26), we obtain

$$\begin{aligned}
q(\mu_{gk}) &\propto \exp\Big\{ -\frac{\mathbb{E}[\rho]}{2}(\mu_{gk} - \mu_0)^2 \Big\} \exp\Big(-\mathbb{E}[\lambda_{gk}] \frac{\mathbb{E}[n_{gk}]\mu_{gk}^2 - 2\mu_{gk}\overline{x}_{gk}}{2} \Big) \\
&\propto \exp\Big\{ -\frac{1}{2}\Big(\mathbb{E}[\rho]\mu_{gk}^2 - 2\mathbb{E}[\rho]\mu_0\mu_{gk} + \mathbb{E}[n_{gk}]\mathbb{E}[\lambda_{gk}]\mu_{gk}^2 - 2\mathbb{E}[\lambda_{gk}]\overline{x}_{gk}\mu_{gk} \Big) \Big\} \\
&\propto \exp\Big\{ -\frac{\mathbb{E}[\rho] + \mathbb{E}[n_{gk}]\mathbb{E}[\lambda_{gk}]}{2} \Big(\mu_{gk} - \frac{\mu_0\mathbb{E}[\rho] + \overline{x}_{gk}\mathbb{E}[\lambda_{gk}]}{\mathbb{E}[\rho] + \mathbb{E}[n_{gk}]\mathbb{E}[\lambda_{gk}]} \Big)^2 \Big\} \ .
\end{aligned} \tag{27}$$

Eq. (27) tells that the mean parameter m_{gk} and the precision parameter r_{gk} of the variational Gaussian posterior $q(\mu_{gk})$ can be written as follows:

$$m_{gk} = \frac{\mu_0\mathbb{E}[\rho] + \overline{x}_{gk}\mathbb{E}[\lambda_{gk}]}{r_{gk}} \ , \ \ r_{gk} = \mathbb{E}[\rho] + \mathbb{E}[n_{gk}]\mathbb{E}[\lambda_{gk}] \ . \tag{28}$$

4.3 Precision posteriors $q(\lambda)$

By taking a functional derivative of $\mathcal{L}$ with respect to $q(\lambda)$, we obtain

$$\frac{\delta \mathcal{L}}{\delta q(\lambda)} = \int \sum_{\mathbf{z}} q(\mathbf{z}) q(\mu) \log p(\mathbf{x}|\mathbf{z}, \mu, \lambda) d\mu + \log p(\lambda | a_0, b_0) - q(\lambda) + const. \tag{29}$$

Therefore, $q(\lambda)$ can be written as follows:

$$q(\lambda) \propto p(\lambda|a_0, b_0) \exp \left\{ \int \sum_{\mathbf{z}} q(\mathbf{z}) q(\mu) \log p(\mathbf{x}|\mathbf{z}, \mu, \lambda) d\mu \right\} . \tag{30}$$

The integral inside the exponential function can be evaluated as follows:

$$\begin{aligned}
\int \sum_{\mathbf{z}} q(\mathbf{z}) q(\mu) \log p(\mathbf{x}|\mathbf{z}, \mu, \lambda) d\mu &= \sum_{d,g} \sum_{k=1}^{K} q(z_{dg} = k) \int q(\mu_{gk}) \log \left[\sqrt{\frac{\lambda_{gk}}{2\pi}} \exp \left\{ - \frac{\lambda_{gk}}{2} (x_{dg} - \mu_{gk})^2 \right\} \right] d\mu_{gk} \\
&= \sum_{g,k} \log \sqrt{\frac{\lambda_{gk}}{2\pi}} \sum_d q(z_{dg} = k) - \sum_{g,k} \frac{\lambda_{gk}}{2} \sum_d q(z_{dg} = k) \int q(\mu_{gk}) (\mu_{gk}^2 - 2 x_{dg} \mu_{gk} + x_{dg}^2) d\mu_{gk} \\
&= \sum_{g,k} \mathbb{E}[n_{gk}] \log \sqrt{\frac{\lambda_{gk}}{2\pi}} - \sum_{g,k} \frac{\lambda_{gk}}{2} \Big(\mathbb{E}[n_{gk}] \mathbb{E}[\mu_{gk}^2] - 2 \overline{x}_{gk} \mathbb{E}[\mu_{gk}] + \overline{v}_{gk} \Big) \\
&= \sum_{g,k} \mathbb{E}[n_{gk}] \log \sqrt{\frac{\lambda_{gk}}{2\pi}} - \sum_{g,k} \lambda_{gk} \left\{ \frac{\mathbb{E}[n_{gk}]/r_{gk} + \sum_d q(z_{dg} = k)(x_{dg} - m_{gk})^2}{2} \right\}
\end{aligned} \tag{31}$$

where we define $\overline{v}_{gk} \equiv \sum_d q(z_{dg} = k) x_{dg}^2$ and replace the variance $\mathbb{E}[\mu_{gk}^2] - \mathbb{E}[\mu_{gk}]^2$ of μ_{gk} with the inversion of the precision r_{gk}, which is introduced in Eq. (28). Consequently, we can obtain $q(\lambda_{gk})$ as follows:

$$\begin{aligned}
q(\lambda_{gk}) &\propto \lambda_{gk}^{a_0 - 1} e^{-b_0 \lambda_{gk}} \cdot \exp \left[\mathbb{E}[n_{gk}] \log \sqrt{\frac{\lambda_{gk}}{2\pi}} - \lambda_{gk} \left\{ \frac{\mathbb{E}[n_{gk}]/r_{gk} + \sum_d q(z_{dg} = k)(x_{dg} - m_{gk})^2}{2} \right\} \right] \\
&\propto \lambda_{gk}^{\mathbb{E}[n_{gk}]/2 + a_0 - 1} \exp \left[- \lambda_{gk} \left\{ \frac{\mathbb{E}[n_{gk}]/r_{gk} + \sum_d q(z_{dg} = k)(x_{dg} - m_{gk})^2}{2} + b_0 \right\} \right] .
\end{aligned} \tag{32}$$

Eq. (32) tells that the shape parameter a_{gk} and the rate parameter b_{gk} of the variational Gamma posterior $q(\lambda_{gk})$ can be written as follows:

$$a_{gk} = \frac{\mathbb{E}[n_{gk}]}{2} + a_0 \ , \quad b_{gk} = \frac{\mathbb{E}[n_{gk}]/r_{gk} + \sum_d q(z_{dg} = k)(x_{dg} - m_{gk})^2}{2} + b_0 \ . \tag{33}$$

By using Eq. (33), the expectation $\mathbb{E}[\lambda_{gk}]$ appearing in Eq. (28) can be evaluated as a_{gk}/b_{gk}.

4.4 Precision hyperparameter posterior $q(\rho)$

We take a functional derivative of $\mathcal{L}$ with respect to $q(\rho)$ as follows:

$$\frac{\delta \mathcal{L}}{\delta q(\lambda)} = \int q(\mu) \log p(\mu | \mu_0, \rho) d\mu + \log p(\rho | a_\rho, b_\rho) - \log q(\rho) + const. \tag{34}$$

Then, $q(\rho)$ can be written as

$$q(\rho) \propto p(\rho | a_\rho, b_\rho) \cdot \exp \left\{ \int q(\mu) \log p(\mu | \mu_0, \rho) d\mu \right\} . \tag{35}$$

The integral inside the exponential function can be evaluated as follows:

$$
\begin{aligned}
&\int q(\mu)\log p(\mu|\mu_0,\rho)d\mu = \sum_{g,k}\int q(\mu_{gk})\log\left[\sqrt{\frac{\rho}{2\pi}}\exp\Big\{-\frac{\rho}{2}(\mu_{gk}-\mu_0)^2\Big\}\right]d\mu_{gk} \\
&= \frac{GK}{2}\log\Big(\frac{\rho}{2\pi}\Big) - \frac{\rho}{2}\sum_{g,k}\int q(\mu_{gk})(\mu_{gk}-\mu_0)^2 d\mu_{gk} = \frac{GK}{2}\log\Big(\frac{\rho}{2\pi}\Big) - \frac{\rho}{2}\sum_{g,k}\big(\mathbb{E}[\mu_{gk}^2] - 2\mu_0\mathbb{E}[\mu_{gk}] + \mu_0^2\big) \\
&= \frac{GK}{2}\log\Big(\frac{\rho}{2\pi}\Big) - \frac{\rho}{2}\sum_{g,k}\Big\{\big(\mathbb{E}[\mu_{gk}^2] - \mathbb{E}[\mu_{gk}]^2\big) + \mathbb{E}[\mu_{gk}]^2 - 2\mu_0\mathbb{E}[\mu_{gk}] + \mu_0^2\Big\} \\
&= \frac{GK}{2}\log\Big(\frac{\rho}{2\pi}\Big) - \frac{\rho}{2}\sum_{g,k}\Big\{\frac{1}{r_{gk}} + \big(\mathbb{E}[\mu_{gk}] - \mu_0\big)^2\Big\} \qquad (36)
\end{aligned}
$$

Therefore, $q(\rho)$ is obtained as

$$
\begin{aligned}
q(\rho) &\propto \rho^{a_\rho - 1}e^{-b_\rho\rho}\cdot\exp\left[\frac{GK}{2}\log\Big(\frac{\rho}{2\pi}\Big) - \frac{\rho}{2}\sum_{g,k}\Big\{\frac{1}{r_{gk}} + \big(\mathbb{E}[\mu_{gk}] - \mu_0\big)^2\Big\}\right] \\
&\propto \rho^{GK/2 + a_\rho - 1}\cdot\exp\left[-\rho\Big\{b_\rho + \sum_{g,k}\frac{r_{gk}^{-1} + \big(m_{gk}-\mu_0\big)^2}{2}\Big\}\right]. \qquad (37)
\end{aligned}
$$

Eq. (37) tells that the shape parameter a and the rate parameter b of the variational Gamma posterior $q(\rho)$ can be written as follows:

$$
a = a_\rho + \frac{GK}{2}\ ,\ \ b = b_\rho + \sum_{g,k}\frac{r_{gk}^{-1} + \big(m_{gk}-\mu_0\big)^2}{2}\ . \qquad (38)
$$

4.5 Latent process assignment posteriors $q(\mathbf{z})$

Recall that $q(\mathbf{z})$ is factorized as $\prod_d\prod_g q(z_{dg})$. For each $q(z_{dg})$, we take a functional derivative of $\mathcal{L}$ in Eq. (9) as follows:

$$
\frac{\delta\mathcal{L}}{\delta q(z'_{dg})} = \int\sum_{\mathbf{z}^{\neg dg}} q(\mathbf{z}^{\neg dg})q(\alpha)q(\pi)q(\mu)q(\lambda)\log p(\mathbf{x},\mathbf{z}^{\neg dg}, z'_{dg}|\alpha,\pi,\mu,\lambda)d\alpha d\tau d\mu d\lambda - \log q(z'_{dg}) + const. \qquad (39)
$$

Therefore, we obtain a function form of $q(z_{dg}=k)$ as follows:

$$
q(z_{dg}=k) \propto \exp\Big\{\int\sum_{\mathbf{z}^{\neg dg}} q(\mathbf{z}^{\neg dg})q(\alpha)q(\pi)q(\mu)q(\lambda)\log p(\mathbf{x},\mathbf{z}^{\neg dg}, z_{dg}=k|\alpha,\pi,\mu,\lambda)d\alpha d\pi d\mu d\lambda\Big\}\ . \qquad (40)
$$

The integral inside the exponential function can be evaluated as below. First, we rewrite $p(\mathbf{x},\mathbf{z}|\alpha,\pi,\mu,\lambda)$ as:

$$
\begin{aligned}
&p(\mathbf{x},\mathbf{z}|\alpha,\pi,\mu,\lambda) = p(\mathbf{z}|\alpha,\pi)p(\mathbf{x}|\mathbf{z},\mu,\lambda) \\
&= \prod_d\frac{\Gamma(\alpha)}{\Gamma(n_d+\alpha)}\prod_k\frac{\Gamma(n_{dk}+\alpha\pi_k)}{\Gamma(\alpha\pi_k)}\cdot\prod_d\prod_g\prod_k\left[\sqrt{\frac{\lambda_{gk}}{2\pi}}\exp\Big\{-\frac{\lambda_{gk}}{2}(x_{dg}-\mu_{gk})^2\Big\}\right]^{n_{dgk}}\ . \qquad (41)
\end{aligned}
$$

By removing gene g from sample d, we obtain the following distribution:

$$
\begin{aligned}
&p(\mathbf{x}^{\neg dg},\mathbf{z}^{\neg dg}|\alpha,\pi,\mu,\lambda) = p(\mathbf{z}^{\neg dg}|\alpha,\pi)p(\mathbf{x}^{\neg dg}|\mathbf{z}^{\neg dg},\mu,\lambda) \\
&= \prod_d\frac{\Gamma(\alpha)}{\Gamma(n_d^{\neg dg}+\alpha)}\prod_k\frac{\Gamma(n_{dk}^{\neg dg}+\alpha\pi_k)}{\Gamma(\alpha\pi_k)}\cdot\prod_d\prod_{g'\neq g}\prod_k\left[\sqrt{\frac{\lambda_{g'k}}{2\pi}}\exp\Big\{-\frac{\lambda_{g'k}}{2}(x_{dg'}-\mu_{g'k})^2\Big\}\right]^{n_{dg'k}}\ . \qquad (42)
\end{aligned}
$$

We divide $p(\mathbf{x},\mathbf{z}|\alpha,\pi,\mu,\lambda)$ by $p(\mathbf{x}^{\neg dg},\mathbf{z}^{\neg dg}|\alpha,\pi,\mu,\lambda)$ and obtain the following result:

$$
\begin{aligned}
p(x_{dg}, z_{dg}=k|\mathbf{x}^{\neg dg},\mathbf{z}^{\neg dg},\alpha,\pi,\mu,\lambda) &= \frac{p(\mathbf{x},\mathbf{z}^{\neg dg},z_{dg}=k|\alpha,\pi,\mu,\lambda)}{p(\mathbf{x}^{\neg dg},\mathbf{z}^{\neg dg}|\alpha,\pi,\mu,\lambda)} \\
&= \frac{n_{dk}^{\neg dg}+\alpha\pi_k}{n_d^{\neg dg}+\alpha}\cdot\sqrt{\frac{\lambda_{gk}}{2\pi}}\exp\Big\{-\frac{\lambda_{gk}}{2}(x_{dg}-\mu_{gk})^2\Big\}\ . \qquad (43)
\end{aligned}
$$

Therefore, the integral inside the exponential function in Eq. (40) can be evaluated as follows:

$$
\begin{aligned}
&\int \sum_{\mathbf{z}^{\neg dg}} q(\mathbf{z}^{\neg dg}) q(\alpha) q(\pi) q(\mu) q(\lambda) \log p(\mathbf{x}, \mathbf{z}^{\neg dg}, z_{dg} = k | \alpha, \pi, \mu, \lambda) d\alpha d\pi d\mu d\lambda \\
&= \int \sum_{\mathbf{z}^{\neg dg}} q(\mathbf{z}^{\neg dg}) q(\alpha) q(\pi) q(\mu) q(\lambda) \log \Big\{ p(x_{dg}, z_{dg} = k | \mathbf{x}^{\neg dg}, \mathbf{z}^{\neg dg}, \alpha, \pi, \mu, \lambda) p(\mathbf{x}^{\neg dg}, \mathbf{z}^{\neg dg} | \alpha, \pi, \mu, \lambda) \Big\} d\alpha d\pi d\mu d\lambda \\
&= \int \sum_{\mathbf{z}^{\neg dg}} q(\mathbf{z}^{\neg dg}) q(\alpha) q(\pi) \log \frac{n_{dk}^{\neg dg} + \alpha \pi_k}{n_d^{\neg dg} + \alpha} d\alpha d\pi \\
&\quad + \int q(\mu_{gk}) q(\lambda_{gk}) \log \bigg[\sqrt{\frac{\lambda_{gk}}{2\pi}} \exp\Big\{ -\frac{\lambda_{gk}}{2} (x_{dg} - \mu_{gk})^2 \Big\} \bigg] d\mu_{gk} d\lambda_{gk} + const.
\end{aligned} \tag{44}
$$

The first term in Eq. (44) can be approximated as is discussed in [6]:

$$
\int \sum_{\mathbf{z}^{\neg dg}} q(\mathbf{z}^{\neg dg}) q(\alpha) q(\pi) \log \frac{n_{dk}^{\neg dg} + \alpha \pi_k}{n_d^{\neg dg} + \alpha} d\alpha d\pi \approx \log(\mathbb{G}[\alpha\pi_k] + \mathbb{E}[n_{dk}^{\neg dg}]) - \frac{\mathbb{V}[n_{dk}^{\neg dg}]}{2(\mathbb{G}[\alpha\pi_k] + \mathbb{E}[n_{dk}^{\neg dg}])^2} , \tag{45}
$$

where $\mathbb{G}[\cdot]$ means the geometric expectation $\mathbb{G}[\cdot] \equiv e^{\mathbb{E}[\log \cdot]}$. The second term in Eq. (44) can be evaluated as follows:

$$
\begin{aligned}
&\int q(\mu_{gk}) q(\lambda_{gk}) \log \bigg[\sqrt{\frac{\lambda_{gk}}{2\pi}} \exp\Big\{ -\frac{\lambda_{gk}}{2} (x_{dg} - \mu_{gk})^2 \Big\} \bigg] d\mu_{gk} d\lambda_{gk} \\
&= \frac{1}{2}\mathbb{E}[\log \lambda_{gk}] - \frac{1}{2} x_{dg}^2 \mathbb{E}[\lambda_{gk}] + 2 x_{dg} \mathbb{E}[\lambda_{gk}] \mathbb{E}[\mu_{gk}] - \frac{1}{2}\mathbb{E}[\lambda_{gk}] \mathbb{E}[\mu_{gk}]^2 - \frac{1}{2}\mathbb{E}[\lambda_{gk}] (\mathbb{E}[\mu_{gk}^2] - \mathbb{E}[\mu_{gk}]^2) + const. \\
&= \frac{1}{2}\mathbb{E}[\log \lambda_{gk}] - \frac{a_{gk}}{b_{gk}} \cdot \frac{(x_{dg} - m_{gk})^2 + r_{gk}^{-1}}{2} + const.
\end{aligned} \tag{46}
$$

Therefore, we have obtained $q(z_{dg} = k)$ as below:

$$
\begin{aligned}
q(z_{dg} = k) \propto\approx & (\mathbb{G}[\alpha\pi_k] + \mathbb{E}[n_{dk}^{\neg dg}]) \exp\bigg\{ - \frac{\mathbb{V}[n_{dk}^{\neg dg}]}{2(\mathbb{G}[\alpha\pi_k] + \mathbb{E}[n_{dk}^{\neg dg}])^2} \bigg\} \\
& \cdot \sqrt{\mathbb{G}[\lambda_{gk}]} \exp\bigg\{ - \frac{a_{gk}}{b_{gk}} \cdot \frac{(x_{dg} - m_{gk})^2 + r_{gk}^{-1}}{2} \bigg\} .
\end{aligned} \tag{47}
$$

5 Lower Bound

When we implement the inference for Bayesian probabilistic models, we often monitor the progress of the inference by evaluating the lower bound of the log evidence per several iterations. The lower bound is expected to be increased as the inference proceeds. Therefore, we can use the lower bound evaluation for checking the correctness of the implementation. Further, we can also use the lower bound achieved at the final iteration of the inference to compare e.g. the convergence efficiency of different inference approaches over the same training data.

In this section, we try to rewrite $\mathcal{L}$ only by using the parameters and their expectations so as to evaluate $\mathcal{L}$ based on the results given in the preceding sections. First, we rewrite $\mathcal{L}$ in Eq. (9) as a sum of various terms depending on different sets of parameters.

$$
\begin{aligned}
\mathcal{L} = & \int \sum_{\mathbf{z}} q(\mathbf{z}) q(\alpha) q(\pi) \log p(\mathbf{z}|\alpha, \pi) d\alpha d\pi + \int \sum_{\mathbf{z}} q(\mathbf{z}) q(\mu) q(\lambda) \log p(\mathbf{x}|\mathbf{z}, \mu, \lambda) d\mu d\lambda \\
& + \int q(\alpha) \log p(\alpha | a_\alpha, b_\alpha) d\alpha + \int q(\pi) \log p(\tilde{\pi}|\gamma) d\pi + \int q(\gamma) \log p(\gamma | a_\gamma, b_\gamma) d\gamma \\
& + \int q(\mu) q(\rho) \log p(\mu | \mu_0, \rho) d\mu d\rho + \int q(\lambda) \log p(\lambda | a_0, b_0) d\lambda + \int q(\rho) \log p(\rho | a_\rho, b_\rho) d\rho \\
& - \sum_{\mathbf{z}} q(\mathbf{z}) \log q(\mathbf{z}) - \int q(\alpha) \log q(\alpha) d\alpha - \int q(\pi) \log q(\pi) d\pi - \int q(\gamma) \log q(\gamma) d\gamma \\
& - \int q(\mu) \log q(\mu) d\mu - \int q(\lambda) \log q(\lambda) d\lambda - \int q(\rho) \log q(\rho) d\rho .
\end{aligned} \tag{48}
$$

From now on, we explain how to evaluate the terms in the right hand side of Eq.(48) one by one.

1. The first term in Eq. (48) is related to the posterior distribution of latent process assignments. Here we use the approximation technique proposed in [6] as is and obtain the following result:

$$\begin{aligned}
&\int \sum_{\mathbf{z}} q(\mathbf{z})q(\alpha)q(\pi)\log p(\mathbf{z}|\alpha,\pi)d\alpha d\pi \\
&= D\int q(\alpha)\log\Gamma(\alpha)d\alpha - \sum_d \int q(\alpha)\log\Gamma(\alpha+n_d)d\alpha \\
&\quad + \sum_d\sum_k\sum_{\mathbf{z}} q(\mathbf{z})\int q(\alpha)q(\pi)\log\Gamma(\alpha\pi_k+n_{dk})d\pi d\alpha - D\sum_k\int q(\alpha)q(\pi)\log\Gamma(\alpha\pi_k)d\pi d\alpha \\
&\approx D\log\Gamma(\mathbb{E}[\alpha]) - \sum_d \log\Gamma(\mathbb{E}[\alpha]+n_d) \\
&\quad + \sum_d\sum_k\Big\{1-\prod_g q(z_{dg}\neq k)\Big\}\bigg\{\log\Gamma(\mathbb{G}[\alpha\pi_k]+\mathbb{E}_+[n_{dk}]) + \frac{\mathbb{V}_+[n_{dk}]\Psi'(\mathbb{G}[\alpha\pi_k]+\mathbb{E}_+[n_{dk}])}{2}\bigg\} \\
&\quad .
\end{aligned} \tag{49}$$

The CVB for LPD [8] adopts an approximation method where we do not need to evaluate the trigamma function $\Psi'(\cdot)$, which appears in Eq. (49). However, this approximation has a serious drawback. The evaluation of this term, i.e., $\int\sum_{\mathbf{z}} q(\mathbf{z})q(\alpha)q(\pi)\log p(\mathbf{z}|\alpha,\pi)d\alpha d\pi$, depends on the ordering of genes $\{1,\ldots,G\}$. Therefore, we use an approximation proposed in [6] and remove this dependence on the ordering of genes. Further, we can treat the case $n_{dk\cdot}=0$ exactly. The approximation method used in Eq. (49) is independent of the assumption of infinite latent processes.

2. We focus on the second term related to the posterior of the observed data and rewrite it as follows:

$$\begin{aligned}
&\int\sum_{\mathbf{z}} q(\mathbf{z})q(\lambda)q(\mu)\log p(\mathbf{x}|\mathbf{z},\mu,\lambda)d\lambda d\mu \\
&= \sum_{d,g}\sum_{k=1}^{K} q(z_{dg}=k)\int q(\lambda_{gk})q(\mu_{gk})\log\bigg[\sqrt{\frac{\lambda_{gk}}{2\pi}}\exp\Big\{-\frac{\lambda_{gk}}{2}(x_{dg}-\mu_{gk})^2\Big\}\bigg]d\lambda_{gk}d\mu_{gk} \\
&= \sum_{g,k}\frac{\mathbb{E}[n_{gk}]\big(\mathbb{E}[\log\lambda_{gk}]-\log 2\pi\big)}{2} - \sum_{g,k}\mathbb{E}[\lambda_{gk}]\Big\{\frac{\mathbb{E}[n_{gk}]/r_{gk}+\sum_d q(z_{dg}=k)(x_{dg}-m_{gk})^2}{2}\Big\}\ .
\end{aligned} \tag{50}$$

3. The four terms $\int q(\pi)\log p(\tilde{\pi}|\gamma)d\pi$, $\int q(\gamma)\log p(\gamma|a_\gamma,b_\gamma)d\gamma$, $-\int q(\pi)\log q(\pi)d\pi$, and $-\int q(\gamma)\log q(\gamma)d\gamma$ in the right hand side of Eq. (48) can be combined as $\int q(\gamma)q(\pi)\log\frac{p(\pi,\gamma|a_\gamma,b_\gamma)}{q(\pi)q(\gamma)}d\pi d\gamma$. This is the negative of the Kullback-Leibler divergence of $q(\gamma)q(\pi)$ from $p(\pi,\gamma|a_\gamma,b_\gamma)$ and can be evaluated as follows:

$$\begin{aligned}
&\int q(\gamma)q(\pi)\log\frac{p(\pi,\gamma|a_\gamma,b_\gamma)}{q(\pi)q(\gamma)}d\pi d\gamma \\
&= a_\gamma\log b_\gamma - \log\Gamma(a_\gamma) - (a_\gamma+K)\log\Big(b_\gamma-\sum_k\mathbb{E}[\log(1-\tilde{\pi}_k)]\Big) + \log\Gamma(a_\gamma+K) \\
&\quad -\sum_k\log\Gamma(\mathbb{E}[s_{\cdot\geq k}]+\mathbb{E}[\gamma]+1) + \sum_k\log\Gamma(\mathbb{E}[s_{\cdot k}]+1) + \sum_k\log\Gamma(\mathbb{E}[s_{\cdot>k}]+\mathbb{E}[\gamma]) \\
&\quad -\sum_k\mathbb{E}[s_{\cdot k}]\mathbb{E}[\log\tilde{\pi}_k] - \sum_k(\mathbb{E}[s_{\cdot>k}]+\mathbb{E}[\gamma])\mathbb{E}[\log(1-\tilde{\pi}_k)]\ ,
\end{aligned} \tag{51}$$

where $\mathbb{E}[\log\tilde{\pi}_k]$ can be evaluated as $\Psi(\mathbb{E}[s_{\cdot k}]+1)-\Psi(\mathbb{E}[s_{\cdot\geq k}]+\mathbb{E}[\gamma]+1)$ based on Eq. (12).

4. By combining the terms related to the concentration parameter α, we obtain the negative of the Kullback-

Leibler divergence of $q(\alpha)$ from $p(\alpha|a_\alpha, b_\alpha)$ as follows:

$$\begin{aligned}
&\int q(\alpha)\frac{\log p(\alpha|a_\alpha, b_\alpha)}{q(\alpha)}d\alpha \\
&= \Big\{a_\alpha \log b_\alpha - \log\Gamma(a_\alpha) + (a_\alpha - 1)\Psi(a_\alpha + \mathbb{E}[s_{\cdot\cdot}]) - (a_\alpha - 1)\log\Big(b_\alpha - \sum_d \mathbb{E}[\log \eta_d]\Big) - b_\alpha \mathbb{E}[\alpha]\Big\} \\
&\quad - \Big\{\log\Big(b_\alpha - \sum_d \mathbb{E}[\log\eta_d]\Big) - \log\Gamma(a_\alpha + \mathbb{E}[s_{\cdot\cdot}]) + (a_\alpha + \mathbb{E}[s_{\cdot\cdot}] - 1)\Psi(a_\alpha + \mathbb{E}[s_{\cdot\cdot}]) - (a_\alpha + \mathbb{E}[s_{\cdot\cdot}])\Big\} . \\
&= a_\alpha \log\frac{b_\alpha}{b_\alpha - \sum_d \mathbb{E}[\log\eta_d]} + \log\frac{\Gamma(a_\alpha + \mathbb{E}[s_{\cdot\cdot}])}{\Gamma(a_\alpha)} - \mathbb{E}[s_{\cdot\cdot}]\Psi(a_\alpha + \mathbb{E}[s_{\cdot\cdot}]) - \mathbb{E}[\alpha]\sum_d \mathbb{E}[\log\eta_d] .
\end{aligned} \tag{52}$$

5. We can rewrite the term $\int q(\mu)q(\rho)\log p(\mu|\mu_0,\rho)d\mu$ as follows:

$$\begin{aligned}
\int q(\mu)q(\rho)\log p(\mu|\mu_0,\rho)d\mu_{gk}d\rho &= \sum_{g,k}\int q(\rho)\log\sqrt{\frac{\rho}{2\pi}}d\rho - \sum_{g,k}\int q(\mu_{gk})q(\rho_{gk})\frac{\rho}{2}(\mu_{gk} - \mu_0)^2 d\mu_{gk}d\rho \\
&= \frac{GK}{2}\{\mathbb{E}[\log\rho] - \log(2\pi)\} - \frac{\mathbb{E}[\rho]}{2}\sum_{g,k}\Big\{\frac{1}{r_{gk}} + \big(\mathbb{E}[\mu_{gk}] - \mu_0\big)^2\Big\} .
\end{aligned} \tag{53}$$

6. The terms $\int q(\lambda)\log p(\lambda|a_0,b_0)d\lambda$ and $\int q(\rho)\log p(\rho|a_\rho,b_\rho)d\rho$ can be evaluated as follows:

$$\begin{aligned}
\int q(\lambda)\log p(\lambda|a_0,b_0)d\lambda &= \sum_{g,k}\int q(\lambda_{gk})\log\frac{b_0^{a_0}}{\Gamma(a_0)}\lambda_{gk}^{a_0-1}e^{-b_0\lambda_{gk}}d\lambda_{gk} \\
&= GKa_0\log b_0 - GK\log\Gamma(a_0) + (a_0 - 1)\sum_{g,k}\mathbb{E}[\log\lambda_{gk}] - b_0\sum_{g,k}\mathbb{E}[\lambda_{gk}] ,
\end{aligned} \tag{54}$$

$$\int q(\rho)\log p(\rho|a_\rho,b_\rho)d\rho = a_\rho\log b_\rho - \log\Gamma(a_\rho) + (a_\rho - 1)\mathbb{E}[\log\rho] - b_\rho\mathbb{E}[\rho] . \tag{55}$$

7. The term $\int q(\mu)\log q(\mu)d\mu$ is evaluated by using the parameters m_{gk} and r_{gk} obtained in Eq. (28) as follows:

$$\begin{aligned}
&\int q(\mu)\log q(\mu)d\mu = \sum_{g,k}\int q(\mu_{gk})\log q(\mu_{gk})d\mu_{gk} = \sum_{g,k}\int q(\mu_{gk})\log\sqrt{\frac{r_{gk}}{2\pi}}\exp\Big\{-\frac{r_{gk}}{2}(\mu_{gk} - m_{gk})^2\Big\}d\mu_{gk} \\
&= \sum_{g,k}\log\sqrt{\frac{r_{gk}}{2\pi}} - \sum_{g,k}\frac{r_{gk}}{2}\int q(\mu_{gk})(\mu_{gk} - m_{gk})^2 d\mu_{gk} = \sum_{g,k}\frac{\log r_{gk}}{2} - \frac{GK\log(2\pi)}{2} - \frac{GK}{2}
\end{aligned} \tag{56}$$

8. The terms $\int q(\lambda)\log q(\lambda)d\lambda$ and $\int q(\rho)\log q(\rho)d\rho$ are evaluated based on Eq. (33) and Eq. (37), respectively:

$$\begin{aligned}
\int q(\lambda)\log q(\lambda)d\lambda &= \sum_{g,k}a_{gk}\log b_{gk} - \sum_{g,k}\log\Gamma(a_{gk}) + \sum_{g,k}(a_{gk} - 1)\mathbb{E}[\log\lambda_{gk}] - \sum_{g,k}b_{gk}\mathbb{E}[\lambda_{gk}] \\
&= \sum_{g,k}\log b_{gk} - \sum_{g,k}\log\Gamma(a_{gk}) + \sum_{g,k}(a_{gk} - 1)\Psi(a_{gk}) - \sum_{g,k}a_{gk}
\end{aligned} \tag{57}$$

$$\int q(\rho)\log q(\rho)d\rho = \log b - \log\Gamma(a) + (a - 1)\Psi(a) - a . \tag{58}$$

9. Finally, the term $\sum_{\mathbf{z}} q(\mathbf{z})\log q(\mathbf{z}) = \sum_{d,g,k} q(z_{dg} = k)\log q(z_{dg} = k)$ can be evaluated by using the probabilities obtained in Eq. (47).

6 Conclusion

We have implemented the proposed CVB based on the mathematical descriptions given in this paper. To achieve the efficiency in computational cost, we have parallelized the inference with OpenMP library, because we have already confirmed the efficiency of OpenMP parallelization in text mining using LDA-like topic models [2]. Further, the experiment comparing iLPD with LPD is now being conducted on the microarray data available at `http://www.gems-system.org/`.

References

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

[2] T. Masada, D. Fukagawa, A. Takasu, Y. Shibata, and K. Oguri. Modeling topical trends over continuous time with priors. in *Proceedings of ISNN'10*, pages 302–311, 2010.

[3] S. Rogers, M. Girolami, C. Campbell, and R. Breitling. The latent process decomposition of cDNA microarray data sets. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2(2):143–156, 2005.

[4] J. Sethuraman. A constructive definition of Dirichlet priors. *Statistica Sinica*, 4:639–650, 1994.

[5] Y.-W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei. Hierarchical Dirichlet processes. *Journal of the American Statistical Association*, 101(476):1566–1581, 2006.

[6] Y.-W. Teh, K. Kurihara, and M. Welling. Collapsed variational inference for HDP. in *Advances in Neural Information Processing Systems 20*, pages 1481–1488, 2008.

[7] Y.-W. Teh, D. Newman, and M. Welling. A collapsed variational Bayesian inference algorithm for latent Dirichlet allocation. in *Advances in Neural Information Processing Systems 19*, pages 1353–1360, 2007.

[8] Y.-M. Ying, P. Li, and C. Campbell. A marginalized variational Bayesian approach to the analysis of array data. *BMC Proceedings*, 2(Suppl 4):S7, 2008.