

A Proof of Lemma 1

The following bounds are known [1]:

$$\left(y - \frac{1}{2}\right) \log y - y + \frac{1}{2} \log(2\pi) \leq \log \Gamma(y) \leq \left(y - \frac{1}{2}\right) \log y - y + \frac{1}{2} \log(2\pi) + \frac{1}{12y}, \quad (21)$$

$$\log y - \frac{1}{y} \leq \Psi(y) \leq \log y - \frac{1}{2y}. \quad (22)$$

For the Dirichlet distribution $p(\boldsymbol{\theta}|\check{\boldsymbol{\theta}}) \propto \prod_{h=1}^H a_h^{\check{\theta}_h-1}$, the mean and the variance are given as follows:

$$\hat{\theta}_h = \langle \theta_h \rangle_{p(\boldsymbol{\theta}|\check{\boldsymbol{\theta}})} = \frac{\check{\theta}_h}{\check{\theta}_0}, \quad \langle (\theta_h - \hat{\theta}_h)^2 \rangle_{p(\boldsymbol{\theta}|\check{\boldsymbol{\theta}})} = \frac{\check{\theta}_h(\check{\theta}_0 - \check{\theta}_h)}{\check{\theta}_0^2(\check{\theta}_0 + 1)},$$

where $\check{\theta}_0 = \sum_{h=1}^H \check{\theta}_h$.

For fixed N, R , defined by Eq.(15), diverges to $+\infty$ if $\check{\Theta}_{m,h} \rightarrow +0$ for any (m, h) or $\check{B}_{l,h} \rightarrow +0$ for any (l, h) . Therefore, the global minimizer of the free energy (14) is in the interior of the domain, where the free energy is differentiable. Consequently, the global minimizer is a stationary point. The stationary condition (12) implies that

$$\check{\Theta}_{m,h} \geq \alpha, \quad \check{B}_{l,h} \geq \eta, \quad (23)$$

$$\sum_{h=1}^H \check{\Theta}_{m,h} = \sum_{h=1}^H \alpha + N^{(m)}, \quad \sum_{l=1}^L \check{B}_{l,h} = \sum_{l=1}^L \eta + \sum_{m=1}^M (\check{\Theta}_{m,h} - \alpha). \quad (24)$$

Therefore, we have

$$\langle (\Theta_{m,h} - \hat{\Theta}_{m,h})^2 \rangle_{q(\boldsymbol{\Theta})} = O_p(N^{-2}) \quad \text{for all } (m, h), \quad (25)$$

$$\left(\max_m \hat{\Theta}_{m,h} \right)^2 \langle (B_{l,h} - \hat{B}_{l,h})^2 \rangle_{q(\boldsymbol{B})} = O_p(N^{-2}) \quad \text{for all } (l, h), \quad (26)$$

which leads to Eq.(17).

By using Eq.(22), Q is bounded as follows:

$$\underline{Q} \leq Q \leq \overline{Q},$$

where

$$\begin{aligned} \overline{Q} &= - \sum_{m=1}^M N^{(m)} \sum_{l=1}^L V_{l,m} \log \left(\sum_{h=1}^H \frac{\check{\Theta}_{m,h}}{\sum_{h'=1}^H \check{\Theta}_{m,h'}} \frac{\check{B}_{l,h}}{\sum_{l'=1}^L \check{B}_{l',h}} \frac{\exp\left(-\frac{1}{2\check{\Theta}_{m,h}}\right)}{\exp\left(-\frac{1}{2\sum_{h'=1}^H \check{\Theta}_{m,h'}}\right)} \frac{\exp\left(-\frac{1}{2\check{B}_{l,h}}\right)}{\exp\left(-\frac{1}{2\sum_{l'=1}^L \check{B}_{l',h}}\right)} \right), \\ \underline{Q} &= - \sum_{m=1}^M N^{(m)} \sum_{l=1}^L V_{l,m} \log \left(\sum_{h=1}^H \frac{\check{\Theta}_{m,h}}{\sum_{h'=1}^H \check{\Theta}_{m,h'}} \frac{\check{B}_{l,h}}{\sum_{l'=1}^L \check{B}_{l',h}} \frac{\exp\left(-\frac{1}{2\check{\Theta}_{m,h}}\right)}{\exp\left(-\frac{1}{2\sum_{h'=1}^H \check{\Theta}_{m,h'}}\right)} \frac{\exp\left(-\frac{1}{2\check{B}_{l,h}}\right)}{\exp\left(-\frac{1}{2\sum_{l'=1}^L \check{B}_{l',h}}\right)} \right). \end{aligned}$$

Using Eqs.(25) and (26), we have Eq.(18), which completes the proof of Lemma 1. \square

B Proof of Lemma 3

By using the bounds (21) and (22), R can be bounded as

$$\underline{R} \leq R \leq \overline{R}, \quad (27)$$

where

$$\begin{aligned} \underline{R} &= - \sum_{m=1}^M \log \left(\frac{\Gamma(H\alpha)}{\Gamma(\alpha)^H} \right) - \sum_{h=1}^H \log \left(\frac{\Gamma(L\eta)}{\Gamma(\eta)} \right) - \frac{M(H-1) + H(L-1)}{2} \log(2\pi) \\ &\quad + \sum_{m=1}^M \left\{ \left(H\alpha - \frac{1}{2} \right) \log \sum_{h=1}^H \check{\Theta}_{m,h} - \sum_{h=1}^H \left(\alpha - \frac{1}{2} \right) \log \check{\Theta}_{m,h} \right\} \end{aligned}$$

$$\begin{aligned}
& + \sum_{h=1}^H \left\{ \left(L\eta - \frac{1}{2} \right) \log \sum_{l=1}^L \check{B}_{l,h} - \sum_{l=1}^L \left(\eta - \frac{1}{2} \right) \log \check{B}_{l,h} \right\} \\
& + \sum_{m=1}^M \left\{ - \sum_{h=1}^H \frac{1}{12\check{\Theta}_{m,h}} - \sum_{h=1}^H \left(\check{\Theta}_{m,h} - \alpha \right) \left(\frac{1}{\check{\Theta}_{m,h}} - \frac{1}{2 \sum_{h'=1}^H \check{\Theta}_{m,h'}} \right) \right\} \\
& + \sum_{h=1}^H \left\{ - \sum_{l=1}^L \frac{1}{12\check{B}_{l,h}} - \sum_{l=1}^L \left(\check{B}_{l,h} - \eta \right) \left(\frac{1}{\check{B}_{l,h}} - \frac{1}{2 \sum_{l'=1}^L \check{B}_{l',h}} \right) \right\}, \tag{28} \\
\bar{R} = & - \sum_{m=1}^M \log \left(\frac{\Gamma(H\alpha)}{\Gamma(\alpha)^H} \right) - \sum_{h=1}^H \log \left(\frac{\Gamma(L\eta)}{\Gamma(\eta)^L} \right) - \frac{M(H-1) + H(L-1)}{2} \log(2\pi) \\
& + \sum_{m=1}^M \left\{ \left(H\alpha - \frac{1}{2} \right) \log \sum_{h=1}^H \check{\Theta}_{m,h} - \sum_{h=1}^H \left(\alpha - \frac{1}{2} \right) \log \check{\Theta}_{m,h} \right\} \\
& + \sum_{h=1}^H \left\{ \left(L\eta - \frac{1}{2} \right) \log \sum_{l=1}^L \check{B}_{l,h} - \sum_{l=1}^L \left(\eta - \frac{1}{2} \right) \log \check{B}_{l,h} \right\} \\
& + \sum_{m=1}^M \left\{ \frac{1}{12 \sum_{h=1}^H \check{\Theta}_{m,h}} - \sum_{h=1}^H \left(\check{\Theta}_{m,h} - \alpha \right) \left(\frac{1}{2\check{\Theta}_{m,h}} - \frac{1}{\sum_{h'=1}^H \check{\Theta}_{m,h'}} \right) \right\} \\
& + \sum_{h=1}^H \left\{ \frac{1}{12 \sum_{l=1}^L \check{B}_{l,h}} - \sum_{l=1}^L \left(\check{B}_{l,h} - \eta \right) \left(\frac{1}{2\check{B}_{l,h}} - \frac{1}{\sum_{l'=1}^L \check{B}_{l',h}} \right) \right\}. \tag{29}
\end{aligned}$$

Eqs.(23) and (24) imply that

$$\begin{aligned}
R = & \sum_{m=1}^M \left\{ \left(H\alpha - \frac{1}{2} \right) \log \sum_{h=1}^H \check{\Theta}_{m,h} - \sum_{h=1}^H \left(\alpha - \frac{1}{2} \right) \log \check{\Theta}_{m,h} \right\} \\
& + \sum_{h=1}^H \left\{ \left(L\eta - \frac{1}{2} \right) \log \sum_{l=1}^L \check{B}_{l,h} - \sum_{l=1}^L \left(\eta - \frac{1}{2} \right) \log \check{B}_{l,h} \right\} + O_p(H(M+L)),
\end{aligned}$$

which leads to Eq.(20). This completes the proof of Lemma 3. \square

C Proof of Theorem 1

Lemma 2 and Lemma 3 imply that the free energy can be written as follows:

$$\begin{aligned}
F - S = & \left\{ M \left(H\alpha - \frac{1}{2} \right) + \hat{H} \left(L\eta - \frac{1}{2} \right) - \sum_{h=1}^{\hat{H}} \left(\widehat{M}^{(h)} \left(\alpha - \frac{1}{2} \right) + \widehat{L}^{(h)} \left(\eta - \frac{1}{2} \right) \right) \right\} \log N \\
& + (H - \hat{K}) \left(L\eta - \frac{1}{2} \right) \log L + O_p(\hat{J}N + LM). \tag{30}
\end{aligned}$$

Below, we investigate the leading term of the free energy (30) in different asymptotic limits.

In the limit when $N \rightarrow \infty$ with $L, M \sim O(1)$

In this case, the minimizer should satisfy

$$\widehat{\mathbf{B}} \widehat{\boldsymbol{\Theta}}^\top = \mathbf{B}^* \boldsymbol{\Theta}^{*\top} + O_p(N^{-1}), \tag{31}$$

making $\hat{J} = 0$ with probability 1, and the leading term of the order of $O_p(\log N)$:

$$F - S = \left\{ M \left(H\alpha - \frac{1}{2} \right) + \hat{H} \left(L\eta - \frac{1}{2} \right) - \sum_{h=1}^{\hat{H}} \left(\widehat{M}^{(h)} \left(\alpha - \frac{1}{2} \right) + \widehat{L}^{(h)} \left(\eta - \frac{1}{2} \right) \right) \right\} \log N + O_p(1).$$

Note that Eq.(31) implies the consistency of the predictive distribution.

In the limit when $N, M \rightarrow \infty$ with $\frac{M}{N}, L \sim O(1)$

In this case,

$$\hat{J} = o_p(\log N), \tag{32}$$

making the leading term of the order of $O_p(N \log N)$:

$$F - S = \left\{ M \left(H\alpha - \frac{1}{2} \right) - \sum_{h=1}^{\hat{H}} \widehat{M}^{(h)} \left(\alpha - \frac{1}{2} \right) \right\} \log N + o_p(N \log N).$$

Eq.(32) implies that the predictive distribution is not necessarily consistent— $o_p(\log N)$ components can deviate from the true matrix $\mathbf{B}^* \boldsymbol{\Theta}^{*\top}$ by the order of $O_p(1)$.

In the limit when $N, L \rightarrow \infty$ with $\frac{L}{N}, M \sim O(1)$

In this case, Eq.(32) holds, and the leading term of the free energy is of the order of $O_p(N \log N)$:

$$F - S = HL\eta \log N + o_p(N \log N).$$

In the limit when $N, L, M \rightarrow \infty$ with $\frac{L}{N}, \frac{M}{N} \sim O(1)$

In this case,

$$\hat{J} = o_p(N \log N), \quad (33)$$

and the leading term of the free energy is of the order of $O_p(N^2 \log N)$:

$$F - S = H(M\alpha + L\eta) \log N + o_p(N^2 \log N).$$

This completes the proof of Theorem 1. \square

D Proof of Corollary 1

From the compact representation when $\hat{H} = H^*, \hat{M}^{(h)} = M^{*(h)}$, and $\hat{L}^{(h)} = L^{*(h)}$, we can decompose a singular component into two, keeping $\hat{B}\hat{\Theta}^\top$ unchanged, so that

$$\hat{H} \rightarrow \hat{H} + 1, \quad (34)$$

$$\sum_{h=1}^H \hat{M}^{(h)} \rightarrow \sum_{h=1}^{H^*} \hat{M}^{(h)} + \Delta M \quad \text{for} \quad \min_h M^{*(h)} \leq \Delta M \leq \max_h M^{*(h)}, \quad (35)$$

$$\sum_{h=1}^H \hat{L}^{(h)} \rightarrow \sum_{h=1}^{H^*} \hat{L}^{(h)} + \Delta L \quad \text{for} \quad 0 \leq \Delta L \leq \max_h L^{*(h)}. \quad (36)$$

Here the lower-bound for ΔM in Eq.(35) corresponds to the case that the least frequent topic is chosen to be decomposed, while the upper-bound to the case that the most frequent topic is chosen. The lower-bound for ΔL in Eq.(36) corresponds to the decomposition such that some of the word-occurrences are moved to a new topic, while the upper-bound to the decomposition such that the topic with the widest vocabulary is copied to a new topic. Note that the bounds both for ΔM and ΔL are not always achievable simultaneously, when we choose one topic to decompose.

Below, we investigate the relation between the sparsity of the solution and the hyperparameter setting in different asymptotic limits.

In the limit when $N \rightarrow \infty$ with $L, M \sim O(1)$

The coefficient of the leading term of the free energy is

$$\lambda = M \left(H\alpha - \frac{1}{2} \right) + \sum_{h=1}^{\hat{H}} \left(L\eta - \frac{1}{2} - \hat{M}^{(h)} \left(\alpha - \frac{1}{2} \right) - \hat{L}^{(h)} \left(\eta - \frac{1}{2} \right) \right). \quad (37)$$

Note that the solution is sparse if Eq.(37) is increasing of \hat{H} , and dense if it is decreasing. Eqs.(34)–(36) imply the following:

1. When $0 < \eta \leq \frac{1}{2L}$ and $\alpha \leq \frac{1}{2}$, the solution is sparse if

$$L\eta - \frac{1}{2} - \min_h M^{*(h)} \left(\alpha - \frac{1}{2} \right) > 0, \text{ or equivalently, } \alpha < \frac{1}{2} - \frac{1}{\min_h M^{*(h)}} \left(\frac{1}{2} - L\eta \right),$$

and dense if

$$\alpha > \frac{1}{2} - \frac{1}{\min_h M^{*(h)}} \left(\frac{1}{2} - L\eta \right).$$

2. When $0 < \eta \leq \frac{1}{2L}$ and $\alpha > \frac{1}{2}$, the solution is sparse if

$$L\eta - \frac{1}{2} - \max_h M^{*(h)} \left(\alpha - \frac{1}{2} \right) > 0, \text{ or equivalently, } \alpha < \frac{1}{2} - \frac{1}{\max_h M^{*(h)}} \left(\frac{1}{2} - L\eta \right),$$

and dense if

$$\alpha > \frac{1}{2} - \frac{1}{\max_h M^{*(h)}} \left(\frac{1}{2} - L\eta \right).$$

Therefore, the solution is always dense in this case.

3. When $\frac{1}{2L} < \eta \leq \frac{1}{2}$ and $\alpha < \frac{1}{2}$, the solution is sparse if

$$L\eta - \frac{1}{2} - \min_h M^{*(h)} \left(\alpha - \frac{1}{2} \right) > 0, \text{ or equivalently, } \alpha < \frac{1}{2} + \frac{1}{\min_h M^{*(h)}} \left(L\eta - \frac{1}{2} \right),$$

and dense if

$$\alpha > \frac{1}{2} + \frac{1}{\min_h M^{*(h)}} \left(L\eta - \frac{1}{2} \right).$$

Therefore, the solution is always sparse in this case.

4. When $\frac{1}{2L} < \eta \leq \frac{1}{2}$ and $\alpha \geq \frac{1}{2}$, the solution is sparse if

$$L\eta - \frac{1}{2} - \max_h M^{*(h)} \left(\alpha - \frac{1}{2} \right) > 0, \text{ or equivalently, } \alpha < \frac{1}{2} + \frac{1}{\max_h M^{*(h)}} \left(L\eta - \frac{1}{2} \right),$$

and dense if

$$\alpha > \frac{1}{2} + \frac{1}{\max_h M^{*(h)}} \left(L\eta - \frac{1}{2} \right).$$

5. When $\eta > \frac{1}{2}$ and $\alpha < \frac{1}{2}$, the solution is sparse if

$$L\eta - \frac{1}{2} - \max_h \left(M^{*(h)} \left(\alpha - \frac{1}{2} \right) + L^{*(h)} \left(\eta - \frac{1}{2} \right) \right) > 0, \quad (38)$$

and dense if

$$L\eta - \frac{1}{2} - \max_h \left(M^{*(h)} \left(\alpha - \frac{1}{2} \right) + L^{*(h)} \left(\eta - \frac{1}{2} \right) \right) < 0. \quad (39)$$

Therefore, the solution is sparse if

$$L\eta - \frac{1}{2} - \min_h M^{*(h)} \left(\alpha - \frac{1}{2} \right) - \max_h L^{*(h)} \left(\eta - \frac{1}{2} \right) > 0,$$

$$\text{or equivalently, } \alpha < \frac{1}{2} + \frac{1}{\min_h M^{*(h)}} \left(L\eta - \frac{1}{2} - \max_h L^{*(h)} \left(\eta - \frac{1}{2} \right) \right),$$

and dense if

$$L\eta - \frac{1}{2} - \max_h M^{*(h)} \left(\alpha - \frac{1}{2} \right) - \max_h L^{*(h)} \left(\eta - \frac{1}{2} \right) < 0,$$

$$\text{or equivalently, } \alpha > \frac{1}{2} + \frac{1}{\max_h M^{*(h)}} \left(L\eta - \frac{1}{2} - \max_h L^{*(h)} \left(\eta - \frac{1}{2} \right) \right).$$

Therefore, the solution is always sparse in this case.

6. When $\eta > \frac{1}{2}$ and $\alpha \geq \frac{1}{2}$, the solution is sparse if Eq.(38) holds, and dense if Eq.(39) holds. Therefore, the solution is sparse if

$$L\eta - \frac{1}{2} - \max_h M^{*(h)} \left(\alpha - \frac{1}{2} \right) - \max_h L^{*(h)} \left(\eta - \frac{1}{2} \right) > 0,$$

$$\text{or equivalently, } \alpha < \frac{1}{2} + \frac{1}{\max_h M^{*(h)}} \left(L\eta - \frac{1}{2} - \max_h L^{*(h)} \left(\eta - \frac{1}{2} \right) \right),$$

and dense if

$$L\eta - \frac{1}{2} - \min_h M^{*(h)} \left(\alpha - \frac{1}{2} \right) - \max_h L^{*(h)} \left(\eta - \frac{1}{2} \right) < 0,$$

$$\text{or equivalently, } \alpha > \frac{1}{2} + \frac{1}{\min_h M^{*(h)}} \left(L\eta - \frac{1}{2} - \max_h L^{*(h)} \left(\eta - \frac{1}{2} \right) \right).$$

Thus, we can conclude that, in this case, the solution is sparse if

$$\alpha < \frac{1}{2} + \frac{L-1}{2 \max_h M^{*(h)}},$$

and dense if

$$\alpha > \frac{1}{2} + \frac{L\eta - \frac{1}{2}}{\min_h M^{*(h)}}.$$

Summarizing the above, we have the following lemma:

Lemma 4 When $0 < \eta \leq \frac{1}{2L}$, the solution is sparse if $\alpha < \frac{1}{2} - \frac{\frac{1}{2} - L\eta}{\min_h M^{*(h)}}$, and dense if $\alpha > \frac{1}{2} - \frac{\frac{1}{2} - L\eta}{\min_h M^{*(h)}}$. When $\frac{1}{2L} < \eta \leq \frac{1}{2}$, the solution is sparse if $\alpha < \frac{1}{2} + \frac{L\eta - \frac{1}{2}}{\max_h M^{*(h)}}$, and dense if $\alpha > \frac{1}{2} + \frac{L\eta - \frac{1}{2}}{\max_h M^{*(h)}}$. When $\eta > \frac{1}{2}$, the solution is sparse if $\alpha < \frac{1}{2} + \frac{L-1}{2 \max_h M^{*(h)}}$, and dense if $\alpha > \frac{1}{2} + \frac{L\eta - \frac{1}{2}}{\min_h M^{*(h)}}$.

In the limit when $N, M \rightarrow \infty$ with $\frac{M}{N}, L \sim O(1)$

The coefficient of the leading term of the free energy is given by

$$\lambda = M \left(H\alpha - \frac{1}{2} \right) - \sum_{h=1}^{\hat{H}} \widehat{M}^{(h)} \left(\alpha - \frac{1}{2} \right). \quad (40)$$

Although the predictive distribution does not necessarily converges to the true distribution, we can investigate the sparsity of the solution by considering the duplication rules (34)–(36) that keep $\widehat{\mathbf{B}}\widehat{\boldsymbol{\Theta}}^\top$ unchanged. It is clear that Eq.(40) is increasing of \widehat{H} if $\alpha < \frac{1}{2}$, and decreasing if $\alpha > \frac{1}{2}$. Combing this result with Lemma 4 completes the proof of Corollary 1. \square

E Proof of Theorem 2

We analyze PBA learning, PBB learning, and MAP estimation, and then summarize the results.

E.1 PBA Learning

The free energy for PBA learning is given as follows:

$$F^{\text{PBA}} = \chi_B + R^{\text{PBA}} + Q^{\text{PBA}}, \quad (41)$$

where χ_B is a large constant corresponding to the negative entropy of the delta functions, and

$$\begin{aligned} R^{\text{PBA}} &= \left\langle \log \frac{q(\boldsymbol{\Theta})q(\mathbf{B})}{p(\boldsymbol{\Theta}|\alpha)p(\mathbf{B}|\eta)} \right\rangle_{q^{\text{PBA}}(\boldsymbol{\Theta}, \mathbf{B})} \\ &= \sum_{m=1}^M \left(\log \frac{\Gamma(\sum_{h=1}^H \check{\Theta}_{m,h}^{\text{PBA}})}{\prod_{h=1}^H \Gamma(\check{\Theta}_{m,h}^{\text{PBA}})} \frac{\Gamma(\alpha)^H}{\Gamma(H\alpha)} + \sum_{h=1}^H \left(\check{\Theta}_{m,h}^{\text{PBA}} - \alpha \right) \left(\Psi(\check{\Theta}_{m,h}^{\text{PBA}}) - \Psi(\sum_{h'=1}^H \check{\Theta}_{m,h'}^{\text{PBA}}) \right) \right) \\ &\quad + \sum_{h=1}^H \left(\log \frac{\Gamma(\eta)^L}{\Gamma(L\eta)} + \sum_{l=1}^L (1 - \eta) \left(\log(\check{B}_{l,h}^{\text{PBA}}) - \log(\sum_{l'=1}^L \check{B}_{l',h}^{\text{PBA}}) \right) \right), \end{aligned} \quad (42)$$

$$\begin{aligned} Q^{\text{PBA}} &= \left\langle \log \frac{q(\{\mathbf{z}^{(n,m)}\})}{p(\{\mathbf{w}^{(n,m)}\}, \{\mathbf{z}^{(n,m)}\}|\boldsymbol{\Theta}, \mathbf{B})} \right\rangle_{q^{\text{PBA}}(\boldsymbol{\Theta}, \mathbf{B}, \{\mathbf{z}^{(n,m)}\})} \\ &= - \sum_{m=1}^M N^{(m)} \sum_{l=1}^L V_{l,m} \log \left(\sum_{h=1}^H \frac{\exp(\Psi(\check{\Theta}_{m,h}^{\text{PBA}}))}{\exp(\Psi(\sum_{h'=1}^H \check{\Theta}_{m,h'}^{\text{PBA}}))} \frac{\check{B}_{l,h}^{\text{PBA}}}{\sum_{l'=1}^L \check{B}_{l',h}^{\text{PBA}}} \right). \end{aligned} \quad (43)$$

Let us first consider the case when $\eta < 1$. In this case, F diverges to $F \rightarrow -\infty$ with fixed N , when $\check{B}_{l,h} = O(1)$ for any (l, h) and $\check{B}_{l',h} \rightarrow +0$ for all other $l' \neq l$. Therefore, the solution is useless.

When $\eta \geq 1$, the solution satisfies the following stationary condition:

$$\check{\Theta}_{m,h}^{\text{PBA}} = \alpha + \sum_{n=1}^{N^{(m)}} \hat{z}_h^{\text{PBA}(n,m)}, \quad \check{B}_{l,h}^{\text{PBA}} = \eta - 1 + \sum_{m=1}^M \sum_{n=1}^{N^{(m)}} w_l^{(n,m)} \hat{z}_h^{\text{PBA}(n,m)}, \quad (44)$$

$$\hat{z}_h^{\text{PBA}(n,m)} = \frac{\exp(\Psi(\check{\Theta}_{m,h}^{\text{PBA}})) \prod_{l=1}^L (\check{B}_{l,h}^{\text{PBA}})^{w_l^{(n,m)}}}{\sum_{h'=1}^H \exp(\Psi(\check{\Theta}_{m,h'}^{\text{PBA}})) \prod_{l=1}^L (\check{B}_{l,h'}^{\text{PBA}})^{w_l^{(n,m)}}}. \quad (45)$$

In the same way as for VB learning, we can obtain the following lemma:

Lemma 5 Let $\widehat{\mathbf{B}}^{\text{PBA}} \widehat{\boldsymbol{\Theta}}^{\text{PBA}\top} = \langle \mathbf{B} \boldsymbol{\Theta}^\top \rangle_{q^{\text{PBA}}(\boldsymbol{\Theta}, \mathbf{B})}$. Then, it holds that

$$\langle (\mathbf{B} \boldsymbol{\Theta}^\top - \widehat{\mathbf{B}}^{\text{PBA}} \widehat{\boldsymbol{\Theta}}^{\text{PBA}\top})_{l,m}^2 \rangle_{q^{\text{PBA}}(\boldsymbol{\Theta}, \mathbf{B})} = O_p(N^{-2}), \quad (46)$$

$$Q^{\text{PBA}} = - \sum_{m=1}^M N^{(m)} \sum_{l=1}^L V_{l,m} \log(\widehat{\mathbf{B}}^{\text{PBA}} \widehat{\boldsymbol{\Theta}}^{\text{PBA}\top})_{l,m} + O_p(N^{-1}). \quad (47)$$

Q^{PBA} is minimized when $\widehat{\mathbf{B}}^{\text{PBA}} \widehat{\boldsymbol{\Theta}}^{\text{PBA}\top} = \mathbf{B}^* \boldsymbol{\Theta}^{*\top} + O_p(N^{-1})$, and it holds that

$$Q^{\text{PBA}} = S + O_p(\widehat{J}N + LM).$$

R^{PBA} is written as follows:

$$\begin{aligned} R^{\text{PBA}} &= \left\{ M \left(H\alpha - \frac{1}{2} \right) + \widehat{H}L(\eta - 1) - \sum_{h=1}^{\widehat{H}} \left(\widehat{M}^{(h)} \left(\alpha - \frac{1}{2} \right) + \widehat{L}^{(h)} (\eta - 1) \right) \right\} \log N \\ &\quad + (H - \widehat{H})L(\eta - 1) \log L + O_p(H(M + L)). \end{aligned} \quad (48)$$

Taking the different asymptotic limits, we obtain the following theorem:

Theorem 3 When $\eta < 1$, each column vector of $\hat{\mathbf{B}}^{\text{PBA}}$ has only one non-zero entry. Assume below that $\eta \geq 1$. In the limit when $N \rightarrow \infty$ with $L, M \sim O(1)$, it holds that $\hat{\mathcal{J}} = 0$ with probability 1, and

$$F^{\text{PBA}} = S + \left\{ M \left(H\alpha - \frac{1}{2} \right) + \hat{H}L(\eta - 1) - \sum_{h=1}^{\hat{H}} \left(\hat{M}^{(h)} \left(\alpha - \frac{1}{2} \right) + \hat{L}^{(h)} (\eta - 1) \right) \right\} \log N + O_p(1).$$

In the limit when $N, M \rightarrow \infty$ with $\frac{M}{N}, L \sim O(1)$, it holds that $\hat{\mathcal{J}} = o_p(\log N)$, and

$$F^{\text{PBA}} = S + \left\{ M \left(H\alpha - \frac{1}{2} \right) - \sum_{h=1}^{\hat{H}} \hat{M}^{(h)} \left(\alpha - \frac{1}{2} \right) \right\} \log N + o_p(N \log N).$$

In the limit when $N, L \rightarrow \infty$ with $\frac{L}{N}, M \sim O(1)$, it holds that $\hat{\mathcal{J}} = o_p(\log N)$, and

$$F^{\text{PBA}} = S + HL(\eta - 1) \log N + o_p(N \log N).$$

In the limit when $N, L, M \rightarrow \infty$ with $\frac{L}{N}, \frac{M}{N} \sim O(1)$, it holds that $\hat{\mathcal{J}} = o_p(N \log N)$, and

$$F^{\text{PBA}} = S + H(M\alpha + L(\eta - 1)) \log N + o_p(N^2 \log N).$$

Note that Theorem 3 provides no information on the sparsity of the PBA solution for $\eta < 1$. Below, we investigate the sparsity of the solution for $\eta \geq 1$.

In the limit when $N \rightarrow \infty$ with $L, M \sim O(1)$

The coefficient of the leading term of the free energy is

$$\lambda^{\text{PBA}} = M \left(H\alpha - \frac{1}{2} \right) + \sum_{h=1}^{\hat{H}} \left(L(\eta - 1) - \hat{M}^{(h)} \left(\alpha - \frac{1}{2} \right) - \hat{L}^{(h)} (\eta - 1) \right).$$

The solution is sparse if λ^{PBA} is increasing of \hat{H} , and dense if it is decreasing. We focus on the case when $\eta \geq 1$. Eqs.(34)–(36) imply the following:

1. When $\alpha < \frac{1}{2}$, the solution is sparse if

$$L(\eta - 1) - \max_h \left(M^{*(h)} \left(\alpha - \frac{1}{2} \right) + L^{*(h)} (\eta - 1) \right) > 0, \quad (49)$$

and dense if

$$L(\eta - 1) - \max_h \left(M^{*(h)} \left(\alpha - \frac{1}{2} \right) + L^{*(h)} (\eta - 1) \right) < 0. \quad (50)$$

Therefore, the solution is sparse if

$$L(\eta - 1) - \min_h M^{*(h)} \left(\alpha - \frac{1}{2} \right) - \max_h L^{*(h)} (\eta - 1) > 0,$$

or equivalently, $\alpha < \frac{1}{2} + \frac{(L - \max_h L^{*(h)}) (\eta - 1)}{\min_h M^{*(h)}}$,

and dense if

$$L(\eta - 1) - \max_h M^{*(h)} \left(\alpha - \frac{1}{2} \right) - \max_h L^{*(h)} (\eta - 1) < 0,$$

or equivalently, $\alpha > \frac{1}{2} + \frac{(L - \max_h L^{*(h)}) (\eta - 1)}{\max_h M^{*(h)}}$.

Therefore, the solution is always sparse in this case.

2. When $\alpha \geq \frac{1}{2}$, the solution is sparse if Eq.(49) holds, and dense if Eq.(50) holds. Therefore, the solution is sparse if

$$L(\eta - 1) - \max_h M^{*(h)} \left(\alpha - \frac{1}{2} \right) - \max_h L^{*(h)} (\eta - 1) > 0,$$

$$\text{or equivalently, } \alpha < \frac{1}{2} + \frac{(L - \max_h L^{*(h)}) (\eta - 1)}{\max_h M^{*(h)}},$$

and dense if

$$L(\eta - 1) - \min_h M^{*(h)} \left(\alpha - \frac{1}{2} \right) - \max_h L^{*(h)} (\eta - 1) < 0,$$

$$\text{or equivalently, } \alpha > \frac{1}{2} + \frac{(L - \max_h L^{*(h)}) (\eta - 1)}{\min_h M^{*(h)}}.$$

Thus, we can conclude that, in this case, the solution is sparse if

$$\alpha < \frac{1}{2},$$

and dense if

$$\alpha > \frac{1}{2} + \frac{L(\eta - 1)}{\min_h M^{*(h)}}.$$

Summarizing the above, we have the following lemma:

Lemma 6 Assume that $\eta \geq 1$. The solution is sparse if $\alpha < \frac{1}{2}$, and dense if $\alpha > \frac{1}{2} + \frac{L(\eta-1)}{\min_h M^{*(h)}}$.

In the limit when $N, M \rightarrow \infty$ with $\frac{M}{N}, L \sim O(1)$

The coefficient of the leading term of the free energy is given by

$$\lambda = M \left(H\alpha - \frac{1}{2} \right) - \sum_{h=1}^{\hat{H}} \widehat{M}^{(h)} \left(\alpha - \frac{1}{2} \right). \quad (51)$$

Although the predictive distribution does not necessarily converges to the true distribution, we can investigate the sparsity of the solution by considering the duplication rules (34)–(36) that keep $\widehat{\mathbf{B}}\widehat{\boldsymbol{\Theta}}^\top$ unchanged. It is clear that Eq.(51) is increasing of \hat{H} if $\alpha < \frac{1}{2}$, and decreasing if $\alpha > \frac{1}{2}$. Combing this result with Lemma 6, we obtain the following corollary:

Corollary 2 Assume that $\eta \geq 1$. In the limit when $N \rightarrow \infty$ with $L, M \sim O(1)$, the PBA solution is sparse if $\alpha < \frac{1}{2}$, and dense if $\alpha > \frac{1}{2} + \frac{L(\eta-1)}{\min_h M^{*(h)}}$. In the limit when $N, M \rightarrow \infty$ with $\frac{M}{N}, L \sim O(1)$, the PBA solution is sparse if $\alpha < \frac{1}{2}$, and dense if $\alpha > \frac{1}{2}$.

E.2 PBB Learning

The free energy for PBB learning is given as follows:

$$F^{\text{PBB}} = \chi_\Theta + R^{\text{PBB}} + Q^{\text{PBB}}, \quad (52)$$

where χ_Θ is a large constant corresponding to the negative entropy of the delta functions, and

$$\begin{aligned} R^{\text{PBB}} &= \left\langle \log \frac{q(\boldsymbol{\Theta})q(\mathbf{B})}{p(\boldsymbol{\Theta}|\alpha)p(\mathbf{B}|\eta)} \right\rangle_{q^{\text{PBB}}(\boldsymbol{\Theta}, \mathbf{B})} \\ &= \sum_{m=1}^M \left(\log \frac{\Gamma(\alpha)^H}{\Gamma(H\alpha)} + \sum_{h=1}^H (1 - \alpha) \left(\log(\check{\Theta}_{m,h}^{\text{PBB}}) - \log(\sum_{h'=1}^H \check{\Theta}_{m,h'}^{\text{PBB}}) \right) \right) \\ &\quad + \sum_{h=1}^H \left(\log \frac{\Gamma(\sum_{l=1}^L \check{B}_{l,h}^{\text{PBB}})}{\prod_{l=1}^L \Gamma(\check{B}_{l,h}^{\text{PBB}})} \frac{\Gamma(\eta)^L}{\Gamma(L\eta)} + \sum_{l=1}^L \left(\check{B}_{l,h}^{\text{PBB}} - \eta \right) \left(\Psi(\check{B}_{l,h}^{\text{PBB}}) - \Psi(\sum_{l'=1}^L \check{B}_{l',h}^{\text{PBB}}) \right) \right), \end{aligned} \quad (53)$$

$$\begin{aligned}
Q^{\text{PBB}} &= \left\langle \log \frac{q(\{\mathbf{z}^{(n,m)}\})}{p(\{\mathbf{w}^{(n,m)}\}, \{\mathbf{z}^{(n,m)}\} | \boldsymbol{\Theta}, \mathbf{B})} \right\rangle_{q^{\text{PBB}}(\boldsymbol{\Theta}, \mathbf{B}, \{\mathbf{z}^{(n,m)}\})} \\
&= - \sum_{m=1}^M N^{(m)} \sum_{l=1}^L V_{l,m} \log \left(\sum_{h=1}^H \frac{\check{\Theta}_{m,h}^{\text{PBB}}}{\sum_{h'=1}^H \check{\Theta}_{m,h'}^{\text{PBB}}} \frac{\exp(\Psi(\check{B}_{l,h}^{\text{PBB}}))}{\exp(\Psi(\sum_{l'=1}^L \check{B}_{l',h}^{\text{PBB}}))} \right). \tag{54}
\end{aligned}$$

Let us first consider the case when $\alpha < 1$. In this case, F diverges to $F \rightarrow -\infty$ with fixed N , when $\check{\Theta}_{m,h} = O(1)$ for any (m, h) and $\check{\Theta}_{m,h'} \rightarrow +0$ for all other $h' \neq h$. Therefore, the solution is sparse (so sparse that the estimator is useless).

When $\alpha \geq 1$, the solution satisfies the following stationary condition:

$$\check{\Theta}_{m,h}^{\text{PBB}} = \alpha - 1 + \sum_{n=1}^{N^{(m)}} \hat{z}_h^{\text{PBB}(n,m)}, \quad \check{B}_{l,h}^{\text{PBB}} = \eta + \sum_{m=1}^M \sum_{n=1}^{N^{(m)}} w_l^{(n,m)} \hat{z}_h^{\text{PBB}(n,m)}, \tag{55}$$

$$\hat{z}_h^{\text{PBB}(n,m)} = \frac{\check{\Theta}_{m,h}^{\text{PBB}} \exp\left\{\sum_{l=1}^L w_l^{(n,m)} (\Psi(\check{B}_{l,h}^{\text{PBB}}) - \Psi(\sum_{l'=1}^L \check{B}_{l',h}^{\text{PBB}}))\right\}}{\sum_{h'=1}^H \check{\Theta}_{m,h'}^{\text{PBB}} \exp\left\{\sum_{l=1}^L w_l^{(n,m)} (\Psi(\check{B}_{l,h'}^{\text{PBB}}) - \Psi(\sum_{l'=1}^L \check{B}_{l',h'}^{\text{PBB}}))\right\}}. \tag{56}$$

In the same way as for VB and PBA learning, we can obtain the following lemma:

Lemma 7 Let $\hat{\mathbf{B}}^{\text{PBB}} \hat{\boldsymbol{\Theta}}^{\text{PBB}\top} = \langle \mathbf{B} \boldsymbol{\Theta}^\top \rangle_{q^{\text{PBB}}(\boldsymbol{\Theta}, \mathbf{B})}$. Then, it holds that

$$\langle (\mathbf{B} \boldsymbol{\Theta}^\top - \hat{\mathbf{B}}^{\text{PBB}} \hat{\boldsymbol{\Theta}}^{\text{PBB}\top})_{l,m}^2 \rangle_{q^{\text{PBB}}(\boldsymbol{\Theta}, \mathbf{B})} = O_p(N^{-2}), \tag{57}$$

$$Q^{\text{PBB}} = - \sum_{m=1}^M N^{(m)} \sum_{l=1}^L V_{l,m} \log(\hat{\mathbf{B}}^{\text{PBB}} \hat{\boldsymbol{\Theta}}^{\text{PBB}\top})_{l,m} + O_p(N^{-1}). \tag{58}$$

Q^{PBB} is minimized when $\hat{\mathbf{B}}^{\text{PBB}} \hat{\boldsymbol{\Theta}}^{\text{PBB}\top} = \mathbf{B}^* \boldsymbol{\Theta}^{*\top} + O_p(N^{-1})$, and it holds that

$$Q^{\text{PBB}} = S + O_p(\hat{J}N + LM).$$

R^{PBB} is written as follows:

$$\begin{aligned}
R^{\text{PBB}} &= \left\{ MH(\alpha - 1) + \hat{H} \left(L\eta - \frac{1}{2} \right) - \sum_{h=1}^{\hat{H}} \left(\widehat{M}^{(h)} (\alpha - 1) + \widehat{L}^{(h)} \left(\eta - \frac{1}{2} \right) \right) \right\} \log N \\
&\quad + (H - \hat{H}) \left(L\eta - \frac{1}{2} \right) \log L + O_p(H(M + L)). \tag{59}
\end{aligned}$$

Taking the different asymptotic limits, we obtain the following theorem:

Theorem 4 When $\alpha < 1$, each row vector of $\hat{\boldsymbol{\Theta}}^{\text{PBB}}$ has only one non-zero entry, and the PBB solution is sparse. Assume below that $\alpha \geq 1$. In the limit when $N \rightarrow \infty$ with $L, M \sim O(1)$, it holds that $\hat{J} = 0$ with probability 1, and

$$\begin{aligned}
F^{\text{PBB}} &= S + \left\{ MH(\alpha - 1) + \hat{H} \left(L\eta - \frac{1}{2} \right) - \sum_{h=1}^{\hat{H}} \left(\widehat{M}^{(h)} (\alpha - 1) + \widehat{L}^{(h)} \left(\eta - \frac{1}{2} \right) \right) \right\} \log N \\
&\quad + O_p(1).
\end{aligned}$$

In the limit when $N, M \rightarrow \infty$ with $\frac{M}{N}, L \sim O(1)$, it holds that $\hat{J} = o_p(\log N)$, and

$$F^{\text{PBB}} = S + \left\{ MH(\alpha - 1) - \sum_{h=1}^{\hat{H}} \widehat{M}^{(h)} (\alpha - 1) \right\} \log N + o_p(N \log N).$$

In the limit when $N, L \rightarrow \infty$ with $\frac{L}{N}, M \sim O(1)$, it holds that $\hat{J} = o_p(\log N)$, and

$$F^{\text{PBB}} = S + HL\eta \log N + o_p(N \log N).$$

In the limit when $N, L, M \rightarrow \infty$ with $\frac{L}{N}, \frac{M}{N} \sim O(1)$, it holds that $\hat{J} = o_p(N \log N)$, and

$$F^{\text{PBB}} = S + H(M(\alpha - 1) + L\eta) \log N + o_p(N^2 \log N).$$

Theorem 4 states that the PBB solution is sparse when $\alpha < 1$. Below, we investigate the sparsity of the solution for $\alpha \geq 1$.

In the limit when $N \rightarrow \infty$ with $L, M \sim O(1)$

The coefficient of the leading term of the free energy is

$$\lambda^{\text{PBB}} = MH(\alpha - 1) + \sum_{h=1}^{\widehat{H}} \left(L\eta - \frac{1}{2} - \widehat{M}^{(h)}(\alpha - 1) - \widehat{L}^{(h)}\left(\eta - \frac{1}{2}\right) \right).$$

The solution is sparse if λ^{PBB} is increasing of \widehat{H} , and dense if it is decreasing. We focus on the case when $\alpha \geq 1$. Eqs.(34)–(36) imply the following:

1. When $0 < \eta \leq \frac{1}{2L}$, the solution is sparse if

$$L\eta - \frac{1}{2} - \max_h M^{*(h)}(\alpha - 1) > 0, \text{ or equivalently, } \alpha < 1 - \frac{1}{\max_h M^{*(h)}} \left(\frac{1}{2} - L\eta \right),$$

and dense if

$$\alpha > 1 - \frac{1}{\max_h M^{*(h)}} \left(\frac{1}{2} - L\eta \right).$$

Therefore, the solution is always dense in this case.

2. When $\frac{1}{2L} < \eta \leq \frac{1}{2}$, the solution is sparse if

$$L\eta - \frac{1}{2} - \max_h M^{*(h)}(\alpha - 1) > 0, \text{ or equivalently, } \alpha < 1 + \frac{L\eta - \frac{1}{2}}{\max_h M^{*(h)}},$$

and dense if

$$\alpha > 1 + \frac{L\eta - \frac{1}{2}}{\max_h M^{*(h)}}.$$

3. When $\eta > \frac{1}{2}$, the solution is sparse if

$$L\eta - \frac{1}{2} - \max_h \left(M^{*(h)}(\alpha - 1) + L^{*(h)}\left(\eta - \frac{1}{2}\right) \right) > 0, \quad (60)$$

and dense if

$$L\eta - \frac{1}{2} - \max_h \left(M^{*(h)}(\alpha - 1) + L^{*(h)}\left(\eta - \frac{1}{2}\right) \right) < 0. \quad (61)$$

Therefore, the solution is sparse if

$$L\eta - \frac{1}{2} - \max_h M^{*(h)}(\alpha - 1) - \max_h L^{*(h)}\left(\eta - \frac{1}{2}\right) > 0,$$

$$\text{or equivalently, } \alpha < 1 + \frac{1}{\max_h M^{*(h)}} \left(L\eta - \frac{1}{2} - \max_h L^{*(h)}\left(\eta - \frac{1}{2}\right) \right),$$

and dense if

$$L\eta - \frac{1}{2} - \min_h M^{*(h)}(\alpha - 1) - \max_h L^{*(h)}\left(\eta - \frac{1}{2}\right) < 0,$$

$$\text{or equivalently, } \alpha > 1 + \frac{1}{\min_h M^{*(h)}} \left(L\eta - \frac{1}{2} - \max_h L^{*(h)}\left(\eta - \frac{1}{2}\right) \right).$$

Thus, we can conclude that, in this case, the solution is sparse if

$$\alpha < 1 + \frac{L - 1}{2 \max_h M^{*(h)}},$$

and dense if

$$\alpha > 1 + \frac{L\eta - \frac{1}{2}}{\min_h M^{*(h)}}.$$

Summarizing the above, we have the following lemma:

Lemma 8 Assume that $\alpha \geq 1$. When $0 < \eta \leq \frac{1}{2L}$, the solution is always dense. When $\frac{1}{2L} < \eta \leq \frac{1}{2}$, the solution is sparse if $\alpha < 1 + \frac{L\eta - \frac{1}{2}}{\max_h M^{*(h)}}$, and dense if $\alpha > 1 + \frac{L\eta - \frac{1}{2}}{\max_h M^{*(h)}}$. When $\eta > \frac{1}{2}$, the solution is sparse if $\alpha < 1 + \frac{L-1}{2 \max_h M^{*(h)}}$, and dense if $\alpha > 1 + \frac{L\eta - \frac{1}{2}}{\min_h M^{*(h)}}$.

In the limit when $N, M \rightarrow \infty$ with $\frac{M}{N}, L \sim O(1)$

The coefficient of the leading term of the free energy is given by

$$\lambda = M(H\alpha - 1) - \sum_{h=1}^{\hat{H}} \widehat{M}^{(h)}(\alpha - 1). \quad (62)$$

Although the predictive distribution does not necessarily converges to the true distribution, we can investigate the sparsity of the solution by considering the duplication rules (34)–(36) that keep $\widehat{\mathbf{B}}\widehat{\boldsymbol{\Theta}}^\top$ unchanged. It is clear that Eq.(62) is decreasing of \hat{H} if $\alpha > 1$. Combing this result with Theorem 4, which states that the PBB solution is sparse when $\alpha < 1$, and Lemma 8, we obtain the following corollary:

Corollary 3 *Consider the limit when $N \rightarrow \infty$ with $L, M \sim O(1)$. When $0 < \eta \leq \frac{1}{2L}$, the PBB solution is sparse if $\alpha < 1$, and dense if $\alpha > 1$. When $\frac{1}{2L} < \eta \leq \frac{1}{2}$, the PBB solution is sparse if $\alpha < 1 + \frac{L\eta - \frac{1}{2}}{\max_h M^{*(h)}}$, and dense if $\alpha > 1 + \frac{L\eta - \frac{1}{2}}{\max_h M^{*(h)}}$. When $\eta > \frac{1}{2}$, the PBB solution is sparse if $\alpha < 1 + \frac{L-1}{2\max_h M^{*(h)}}$, and dense if $\alpha > 1 + \frac{L\eta - \frac{1}{2}}{\min_h M^{*(h)}}$. In the limit when $N, M \rightarrow \infty$ with $\frac{M}{N}, L \sim O(1)$, the PBB solution is sparse if $\alpha < 1$, and dense if $\alpha > 1$.*

E.3 MAP Learning

The free energy for MAP learning is given as follows:

$$F^{\text{MAP}} = \chi_{\boldsymbol{\Theta}} + \chi_{\mathbf{B}} + R^{\text{MAP}} + Q^{\text{MAP}}, \quad (63)$$

where $\chi_{\boldsymbol{\Theta}}$ and $\chi_{\mathbf{B}}$ are large constants corresponding to the negative entropies of the delta functions, and

$$\begin{aligned} R^{\text{MAP}} &= \left\langle \log \frac{q(\boldsymbol{\Theta})q(\mathbf{B})}{p(\boldsymbol{\Theta}|\alpha)p(\mathbf{B}|\eta)} \right\rangle_{q^{\text{MAP}}(\boldsymbol{\Theta}, \mathbf{B})} \\ &= \sum_{m=1}^M \left(\log \frac{\Gamma(\alpha)^H}{\Gamma(H\alpha)} + \sum_{h=1}^H (1 - \alpha) \left(\log(\check{\Theta}_{m,h}^{\text{MAP}}) - \log(\sum_{h'=1}^H \check{\Theta}_{m,h'}^{\text{MAP}}) \right) \right) \\ &\quad + \sum_{h=1}^H \left(\log \frac{\Gamma(\eta)^L}{\Gamma(L\eta)} + \sum_{l=1}^L (1 - \eta) \left(\log(\check{B}_{l,h}^{\text{MAP}}) - \log(\sum_{l'=1}^L \check{B}_{l',h}^{\text{MAP}}) \right) \right), \end{aligned} \quad (64)$$

$$\begin{aligned} Q^{\text{MAP}} &= \left\langle \log \frac{q(\{\mathbf{z}^{(n,m)}\})}{p(\{\mathbf{w}^{(n,m)}\}, \{\mathbf{z}^{(n,m)}\}|\boldsymbol{\Theta}, \mathbf{B})} \right\rangle_{q^{\text{MAP}}(\boldsymbol{\Theta}, \mathbf{B}, \{\mathbf{z}^{(n,m)}\})} \\ &= - \sum_{m=1}^M N^{(m)} \sum_{l=1}^L V_{l,m} \log \left(\sum_{h=1}^H \frac{\check{\Theta}_{m,h}^{\text{MAP}}}{\sum_{h'=1}^H \check{\Theta}_{m,h'}^{\text{MAP}}} \frac{\check{B}_{l,h}^{\text{MAP}}}{\sum_{l'=1}^L \check{B}_{l',h}^{\text{MAP}}} \right). \end{aligned} \quad (65)$$

Let us first consider the case when $\alpha < 1$. In this case, F diverges to $F \rightarrow -\infty$ with fixed N , when $\check{\Theta}_{m,h} = O(1)$ for any (h, m) and $\check{\Theta}_{m,h'} \rightarrow +0$ for all other $h' \neq h$. Therefore, the solution is sparse (so sparse that the estimator is useless). Similarly, assume that $\eta < 1$. Then, F diverges to $F \rightarrow -\infty$ with fixed N , when $\check{B}_{l,h} = O(1)$ for any (l, h) and $\check{B}_{l',h} \rightarrow +0$ for all other $l' \neq l$. Therefore, the solution is useless.

When $\alpha \geq 1$ and $\eta \geq 1$, the solution satisfies the following stationary condition:

$$\check{\Theta}_{m,h}^{\text{MAP}} = \alpha - 1 + \sum_{n=1}^{N^{(m)}} \widehat{z}_h^{\text{MAP}(n,m)}, \quad \check{B}_{l,h}^{\text{MAP}} = \eta - 1 + \sum_{m=1}^M \sum_{n=1}^{N^{(m)}} w_l^{(n,m)} \widehat{z}_h^{\text{MAP}(n,m)}, \quad (66)$$

$$\widehat{z}_h^{\text{MAP}(n,m)} = \frac{\check{\Theta}_{m,h}^{\text{MAP}} \prod_{l=1}^L (\check{B}_{l,h}^{\text{MAP}})^{w_l^{(n,m)}}}{\sum_{h'=1}^H \left(\check{\Theta}_{m,h'}^{\text{MAP}} \prod_{l=1}^L (\check{B}_{l,h'}^{\text{MAP}})^{w_l^{(n,m)}} \right)}. \quad (67)$$

In the same way as for VB, PBA, and PBB learning, we can obtain the following lemma:

Lemma 9 Let $\hat{\mathbf{B}}^{\text{MAP}} \hat{\boldsymbol{\Theta}}^{\text{MAP}\top} = \langle \mathbf{B} \boldsymbol{\Theta}^\top \rangle_{q^{\text{MAP}}(\boldsymbol{\Theta}, \mathbf{B})}$. Then, Q^{MAP} is minimized when $\hat{\mathbf{B}}^{\text{MAP}} \hat{\boldsymbol{\Theta}}^{\text{MAP}\top} = \mathbf{B}^* \boldsymbol{\Theta}^{*\top} + O_p(N^{-1})$, and it holds that

$$Q^{\text{MAP}} = S + O_p(\hat{J}N + LM).$$

R^{MAP} is written as follows:

$$\begin{aligned} R^{\text{MAP}} = & \left\{ MH(\alpha - 1) + \hat{H}L(\eta - 1) - \sum_{h=1}^{\hat{H}} \left(\widehat{M}^{(h)}(\alpha - 1) + \widehat{L}^{(h)}(\eta - 1) \right) \right\} \log N \\ & + (H - \hat{H})L(\eta - 1) \log L + O_p(H(M + L)). \end{aligned} \quad (68)$$

Taking the different asymptotic limits, we obtain the following theorem:

Theorem 5 When $\alpha < 1$, each row vector of $\hat{\boldsymbol{\Theta}}^{\text{MAP}}$ has only one non-zero entry, and the MAP solution is sparse. When $\eta < 1$, each column vector of $\hat{\mathbf{B}}^{\text{MAP}}$ has only one non-zero entry. Assume below that $\alpha, \eta \geq 1$. In the limit when $N \rightarrow \infty$ with $L, M \sim O(1)$, it holds that $\hat{J} = 0$ with probability 1, and

$$\begin{aligned} F^{\text{MAP}} = & S + \left\{ MH(\alpha - 1) + \hat{H}L(\eta - 1) - \sum_{h=1}^{\hat{H}} \left(\widehat{M}^{(h)}(\alpha - 1) + \widehat{L}^{(h)}(\eta - 1) \right) \right\} \log N \\ & + O_p(1). \end{aligned}$$

In the limit when $N, M \rightarrow \infty$ with $\frac{M}{N}, L \sim O(1)$, it holds that $\hat{J} = o_p(\log N)$, and

$$F^{\text{MAP}} = S + \left\{ MH(\alpha - 1) - \sum_{h=1}^{\hat{H}} \widehat{M}^{(h)}(\alpha - 1) \right\} \log N + o_p(N \log N).$$

In the limit when $N, L \rightarrow \infty$ with $\frac{L}{N}, M \sim O(1)$, it holds that $\hat{J} = o_p(\log N)$, and

$$F^{\text{MAP}} = S + HL(\eta - 1) \log N + o_p(N \log N).$$

In the limit when $N, L, M \rightarrow \infty$ with $\frac{L}{N}, \frac{M}{N} \sim O(1)$, it holds that $\hat{J} = o_p(N \log N)$, and

$$F^{\text{MAP}} = S + H(M(\alpha - 1) + L(\eta - 1)) \log N + o_p(N^2 \log N).$$

Theorem 5 states that the MAP solution is sparse when $\alpha < 1$. However it provides no information on the sparsity of the MAP solution for $\eta < 1$. Below, we investigate the sparsity of the solution for $\alpha, \eta \geq 1$.

In the limit when $N \rightarrow \infty$ with $L, M \sim O(1)$

The coefficient of the leading term of the free energy is

$$\lambda^{\text{MAP}} = MH(\alpha - 1) + \sum_{h=1}^{\hat{H}} \left(L(\eta - 1) - \widehat{M}^{(h)}(\alpha - 1) - \widehat{L}^{(h)}(\eta - 1) \right).$$

The solution is sparse if λ^{MAP} is increasing of \hat{H} , and dense if it is decreasing. We focus on the case when $\alpha, \eta \geq 1$. Eqs.(34)–(36) imply the following:

The solution is sparse if

$$L(\eta - 1) - \max_h \left(M^{*(h)}(\alpha - 1) + L^{*(h)}(\eta - 1) \right) > 0, \quad (69)$$

and dense if

$$L(\eta - 1) - \max_h \left(M^{*(h)}(\alpha - 1) + L^{*(h)}(\eta - 1) \right) < 0. \quad (70)$$

Therefore, the solution is sparse if

$$L(\eta - 1) - \max_h M^{*(h)}(\alpha - 1) - \max_h L^{*(h)}(\eta - 1) > 0,$$

$$\text{or equivalently, } \alpha < 1 + \frac{(L - \max_h L^{*(h)})(\eta - 1)}{\max_h M^{*(h)}},$$

ad dense if

$$L(\eta - 1) - \min_h M^{*(h)} (\alpha - 1) - \max_h L^{*(h)} (\eta - 1) < 0,$$

or equivalently, $\alpha > 1 + \frac{(L - \max_h L^{*(h)})(\eta - 1)}{\min_h M^{*(h)}}.$

Thus, we can conclude that the solution is sparse if

$$\alpha < 1,$$

and dense if

$$\alpha > 1 + \frac{L(\eta - 1)}{\min_h M^{*(h)}}.$$

Summarizing the above, we have the following lemma:

Lemma 10 *Assume that $\eta \geq 1$. The solution is sparse if $\alpha < 1$, and dense if $\alpha > 1 + \frac{L(\eta-1)}{\min_h M^{*(h)}}.$*

In the limit when $N, M \rightarrow \infty$ with $\frac{M}{N}, L \sim O(1)$

The coefficient of the leading term of the free energy is given by

$$\lambda^{\text{MAP}} = MH (\alpha - 1) - \sum_{h=1}^{\hat{H}} \widehat{M}^{(h)} (\alpha - 1). \quad (71)$$

Although the predictive distribution does not necessarily converges to the true distribution, we can investigate the sparsity of the solution by considering the duplication rules (34)–(36) that keep $\widehat{\mathbf{B}}\widehat{\boldsymbol{\Theta}}^\top$ unchanged. It is clear that Eq.(71) is decreasing of \widehat{H} if $\alpha > 1$. Combing this result with Theorem 5, which states that the MAP solution is sparse if $\alpha < 1$, and Lemma 10, we obtain the following corollary:

Corollary 4 *Assume that $\eta \geq 1$. In the limit when $N \rightarrow \infty$ with $L, M \sim O(1)$, the MAP solution is sparse if $\alpha < 1$, and dense if $\alpha > 1 + \frac{L(\eta-1)}{\min_h M^{*(h)}}.$ In the limit when $N, M \rightarrow \infty$ with $\frac{M}{N}, L \sim O(1)$, the MAP solution is sparse if $\alpha < 1$, and dense if $\alpha > 1$.*

E.4 Summary of Results

Summarizing Corollary 1, Corollary 2, Corollary 3, and Corollary 4 completes the proof of Theorem 2. \square