Taxonomic Indian Buffet Process

May 2024

1 Introduction

Non-parametric Bayesian models are particularly used to deal with complex data structures and scenarios where the number of parameters is not fixed in advance. For characterizing infinite latent feature models, Indian buffet process (IBP) provides a flexible framework where each data point is associated with a potentially infinite set of features, allowing for a sparse representation and adaptive complexity as more data is observed (Griffiths and Ghahramani, 2011). As a popular prior, it is used to characterize binary matrices with finite number of exchangeable rows are independent columns.

Recently, significant extensions of the Indian Buffet Process (IBP) have been developed to characterize dependency structures across rows, making them non-exchangeable (Williamson et al., 2011, Miller et al., 2012, Gershman et al., 2014). Instead, we propose a random binary matrix with infinite columns for each taxonomic level such that each layer follows the marginal properties of a usual IBP while maintaining the appropriate dependence linkage across layers.

2 Background on the IBP

Firstly, we review the concepts of Indian Buffet Process (Griffiths and Ghahramani, 2011).

Denote $\boldsymbol{Y}$ as a binary matrix with N rows and infinite columns (features). $y_{ij} = 1$ means that sample i possesses the feature j.

It could be derived as Beta-Bernoulli model with infinite features. With parameter α and finite number of K features, $\boldsymbol{Y}$ could be regarded as a N x K matrix.

$$y_{ij}|\pi_j \sim \text{Bernoulli}(\pi_j), \quad \pi_j|\alpha \sim \text{Beta}\left(\frac{\alpha}{K}, 1\right) \tag{1}$$

The rows are exchangeable and the columns are independent. Therefore,

$$\sum_{j=1}^{K} y_{ij} \sim \text{Poisson}(\alpha), \quad \lim_{K\to\infty} \mathbb{E}\left[\sum_{j=1}^{K} y_{ij}\right] = \alpha \tag{2}$$

Let $K_+ = \sum_{h=1}^{2^N-1} K_h$ to be the number of non-empty columns. Therefore, for i = 1, ..., n, with $H_n = \sum_{i=1}^{n} \frac{1}{i}$:

$$K_+ \sim \text{Poisson}(\alpha H_n), \quad \mathbb{E}[K_+] = \alpha H_n \tag{3}$$

3 Taxonomic Beta-Bernoulli Matrix

3.1 For Taxonomic Layer l

We introduce a taxonomic IBP which allows a usual IBP marginally for each taxonomic layer with dependency linkage across taxonomic layers.

To make some definitions clear:

$P(j', l+1)$: the parent node for j' th mode at taxonomic layer (l + 1).

$C(j, l) = \{j : P(j', l+1) = j\}$: set of all children nodes from parent node j at taxonomic layer l.

We are given the hyperparameter, $\alpha^{(l)}$ and number of observed features $K^{(l)}$ at the taxonomic level l, with i = 1, ..., n and J = 1,..., $K^{(l)}$:

$$y_{ij}^{(l)} | \pi_j^{(l)} \sim \text{Bernoulli}(\pi_j^{(l)}), \quad \pi_j^{(l)} | \alpha^{(l)} \sim \text{Beta}\left(\frac{\alpha^{(l)}}{K^{(l)}}, 1\right) \tag{4}$$

Thus, we could sample a binary matrix $\boldsymbol{Y}^{(l)}$ from Beta-Bernoulli distribution.

3.2 For Taxonomic Layer (l + 1)

For the taxonomic layer (l + 1), we could infer the hyperparameter $\alpha^{(l+1)}$ given the number of observed features $K^{(l+1)}$.

We follow a dependency structure similar to Bayesian Pyramid (Gu et al., 2023).

For each $y_{ij}^{(l)}$ in the parent node, it has several children nodes $y_{ijj'}^{(l+1)}$ for every $j' \in C(j,l)$. (with j = 1, ..., $K^{(l)}$, we have j' = 1, ..., $|C(j,l)|$)

For each node in layer (l + 1):

$$y_{ijj'}^{(l+1)}|\pi_{jj'}^{(l+1)} \sim \text{Bernoulli}(\pi_{jj'}^{(l+1)}), \quad \pi_{jj'}^{(l+1)}|\alpha^{(l+1)} \sim \text{Beta}\left(\frac{\alpha^{(l+1)}}{K^{(l+1)}}, 1\right) \\ \text{with } j' \in C(j,l) \tag{5}$$

3.3 Dependency Linkage between Layer l and Layer (l + 1)

The relationship between level $y_{ij}^{(l)}$ and each $y_{ijj'}^{(l+1)}$ follows a Bayes Rule:

$$\begin{aligned} \Pr(y_{ijj'}^{(l+1)} = 1|\alpha^{(l)}) &= \Pr(y_{ijj'}^{(l+1)} = 1|y_{ij}^{(l)} = 1)\Pr(y_{ij}^{(l)} = 1|\alpha^{(l)}) \\ &\quad + \Pr(y_{ijj'}^{(l+1)} = 1|y_{ij}^{(l)} = 0)\Pr(y_{ij}^{(l)} = 0|\alpha^{(l)}) \\ \Pr(y_{ijj'}^{(l+1)} = 0|\alpha^{(l)}) &= \Pr(y_{ijj'}^{(l+1)} = 0|y_{ij}^{(l)} = 1)\Pr(y_{ij}^{(l)} = 1|\alpha^{(l)}) \\ &\quad + \Pr(y_{ijj'}^{(l+1)} = 0|y_{ij}^{(l)} = 0)\Pr(y_{ij}^{(l)} = 0|\alpha^{(l)}) \end{aligned} \tag{6}$$

For my proposal, set the dependency linkage for j = 1,..., $K^{(l)}$:

$$\begin{aligned} &P(y_{ijj'}^{(l+1)} = 1|y_{ij}^{(l)} = 0) = 0 \\ &P(y_{ijj'}^{(l+1)} = 0|y_{ij}^{(l)} = 0) = 1 \\ &P(y_{ijj'}^{(l+1)} = 1|y_{ij}^{(l)} = 1) = \pi_{jj'} \\ &P(y_{ijj'}^{(l+1)} = 0|y_{ij}^{(l)} = 1) = 1 - \pi_{jj'} \end{aligned} \tag{7}$$

Thus,

$$\begin{aligned} \Pr(y_{ijj'}^{(l+1)} = 1|\pi_j^{(l)}) &= \sum_{y_{ij}^{(l)} \in [0,1]} \Pr(y_{ijj'}^{(l+1)} = 1|y_{ij}^{(l)})\Pr(y_{ij}^{(l)}|\pi_j^{(l)}) \\ &= \Pr(y_{ijj'}^{(l+1)} = 1|y_{ij}^{(l)} = 1)\Pr(y_{ij}^{(l)} = 1|\pi_j^{(l)}) + \Pr(y_{ijj'}^{(l+1)} = 1|y_{ij}^{(l)} = 0)\Pr(y_{ij}^{(l)} = 0|\pi_j^{(l)}) \\ &= \pi_{jj'}\pi_j^{(l)} + 0(1 - \pi_j^{(l)}) \\ &= \pi_{jj'}\pi_j^{(l)} \end{aligned} \tag{8}$$

$$
\begin{aligned}
\Pr(y_{ijj'}^{(l+1)} = 0|\pi_j^{(l)}) &= \sum_{y_{ij}^{(l)} \in [0,1]} \Pr(y_{ijj'}^{(l+1)} = 0|y_{ij}^{(l)}) \Pr(y_{ij}^{(l)}|\pi_j^{(l)}) \\
&= \Pr(y_{ijj'}^{(l+1)} = 0|y_{ij}^{(l)} = 1) \Pr(y_{ij}^{(l)} = 1|\pi_j^{(l)}) + \Pr(y_{ijj'}^{(l+1)} = 0|y_{ij}^{(l)} = 0) \Pr(y_{ij}^{(l)} = 0|\pi_j^{(l)}) \\
&= (1 - \pi_{jj'})\pi_j^{(l)} + 1(1 - \pi_j^{(l)}) \\
&= 1 - \pi_{jj'}\pi_j^{(l)}
\end{aligned} \tag{9}
$$

Denote $m_{jj'}^{(l+1)} = \sum_{i=1}^n y_{ijj'}^{(l+1)}$. Set $u = \pi_{jj'}\pi_j^{(l)}$, $du = \pi_{jj'}d\pi_j^{(l)}$. Based on (8) and (9), we could calculate the probability of a column $\boldsymbol{Y}_{jj'}^{(l+1)}$ based on $\alpha^{(l)}$:

$$
\begin{aligned}
\Pr(\boldsymbol{Y}_{jj'}^{(l+1)}|\alpha^{(l)}) &= \int_0^1 \Pr(\boldsymbol{Y}_{jj'}^{(l+1)}|\pi_j^{(l)}) \Pr(\pi_j^{(l)}|\alpha^{(l)}) d\pi_j^{(l)} \\
&= \int_0^1 \prod_{i=1}^n \Pr(y_{ijj'}^{(l+1)}|\pi_j^{(l)}) \Pr(\pi_j^{(l)}|\alpha^{(l)}) d\pi_j^{(l)} \\
&= \int_0^1 \left[\prod_{i=1}^n (\pi_{jj'}\pi_j^{(l)})^{y_{ijj'}^{(l+1)}} (1 - \pi_{jj'}\pi_j^{(l)})^{1-y_{ijj'}^{(l+1)}} \right] \Pr(\pi_j^{(l)}|\alpha^{(l)}) d\pi_j^{(l)} \\
&= \int_0^1 \left[(\pi_{jj'}\pi_j^{(l)})^{\sum_{i=1}^n y_{ijj'}^{(l+1)}} (1 - \pi_{jj'}\pi_j^{(l)})^{n-\sum_{i=1}^n y_{ijj'}^{(l+1)}} \right] \alpha^{(l)} (\pi_j^{(l)})^{\alpha^{(l)}-1} d\pi_j^{(l)} \\
&= \int_0^{\pi_{jj'}} \left[u^{m_{jj'}^{(l+1)}} (1-u)^{n-m_{jj'}^{(l+1)}} \right] \frac{\alpha^{(l)}}{K^{(l)}} \left(\frac{u}{\pi_{jj'}} \right)^{\frac{\alpha^{(l)}}{K^{(l)}}-1} \frac{du}{\pi_{jj'}} \\
&= \frac{\alpha^{(l)}}{K^{(l)}} \left(\frac{1}{\pi_{jj'}} \right)^{\frac{\alpha^{(l)}}{K^{(l)}}} \int_0^{\pi_{jj'}} \left[u^{\frac{\alpha^{(l)}}{K^{(l)}} + m_{jj'}^{(l+1)} - 1} (1-u)^{n-m_{jj'}^{(l+1)}} \right] du \\
&= \frac{\alpha^{(l)}}{K^{(l)} (\pi_{jj'})^{\frac{\alpha^{(l)}}{K^{(l)}}}} B_{\pi_{jj'}} \left(\frac{\alpha^{(l)}}{K^{(l)}} + m_{jj'}^{(l+1)}, n - m_{jj'}^{(l+1)} + 1 \right) \text{ (incomplete beta function)} \\
&\leq \frac{\alpha^{(l)}}{K^{(l)} (\pi_{jj'})^{\frac{\alpha^{(l)}}{K^{(l)}}}} B \left(\frac{\alpha^{(l)}}{K^{(l)}} + m_{jj'}^{(l+1)}, n - m_{jj'}^{(l+1)} + 1 \right) \text{ (upper bound when } \pi_{jj'} = 1\text{)}
\end{aligned} \tag{10}
$$

By the concept of Beta-Binomial matrix for a finite feature model (Griffiths and Ghahramani, 2011), the marginal probability of a column $\boldsymbol{Y}_{jj'}^{(l+1)}$ would be:

$$
\begin{aligned}
\Pr(\boldsymbol{Y}_{jj'}^{(l+1)}|\alpha^{(l+1)}) &= \int \prod_{i=1}^n \Pr(y_{ijj'}^{(l+1)}|\pi_{jj'}^{(l+1)}) \Pr(\pi_{jj'}^{(l+1)}|\alpha^{(l+1)}) d\pi_{jj'}^{(l+1)} \\
&= \frac{\alpha^{(l+1)}}{K^{(l+1)}} B \left(\frac{\alpha^{(l+1)}}{K^{(l+1)}} + m_{jj'}^{(l+1)}, n - m_{jj'}^{(l+1)} + 1 \right)
\end{aligned} \tag{11}
$$

3.4 Find $\alpha^{(l+1)}$ for Taxonomic Truncated Beta-Binomial Matrix

Assume $\pi_{jj'} = \pi$ for j = 1, ..., $K^{(l)}$, j' = 1, ..., j' = 1, ..., $|C(j,l)|$). In other words, $\pi_{jj'}$ is constant for all linkages between level l and (l + 1).

By (10) and (11), with fixed value of $\alpha^{(l)}$, set the probability of a binary matrix $\boldsymbol{Y}^{(l+1)}$ in layer (l + 1) to be close to each other.

$$
\begin{aligned}
&\operatorname{argmin}_{\alpha^{(l+1)}} \left(\Pr(\boldsymbol{Y}^{(l+1)}|\alpha^{(l+1)}) - \Pr(\boldsymbol{Y}^{(l+1)}|\alpha^{(l)})\right) \\
&\operatorname{argmin}_{\alpha^{(l+1)}} \left(\prod_{jj'=1}^{K^{(l+1)}} \Pr(\boldsymbol{Y}_{jj'}^{(l+1)}|\alpha^{(l+1)}) - \prod_{jj'=1}^{K^{(l+1)}} \Pr(\boldsymbol{Y}_{jj'}^{(l+1)}|\alpha^{(l)})\right) \\
&\operatorname{argmin}_{\alpha^{(l+1)}} \left(\prod_{jj'=1}^{K^{(l+1)}} \frac{\alpha^{(l+1)}}{K^{(l+1)}} B\left(\frac{\alpha^{(l+1)}}{K^{(l+1)}} + m_{jj'}^{(l+1)}, n - m_{jj'}^{(l+1)} + 1\right)\right. \\
&\qquad \left. - \prod_{jj'=1}^{K^{(l+1)}} \frac{\alpha^{(l)}}{K^{(l)}(\pi_{jj'})^{\frac{\alpha^{(l)}}{K^{(l)}}}} B_{\pi_{jj'}}\left(\frac{\alpha^{(l)}}{K^{(l)}} + m_{jj'}^{(l+1)}, n - m_{jj'}^{(l+1)} + 1\right)\right)
\end{aligned}
$$

We could take the argmin of difference between their log-likelihoods:

$$
\begin{aligned}
&\operatorname{argmin}_{\alpha^{(l+1)}} \left(\sum_{jj'=1}^{K^{(l+1)}} \left[\log \alpha^{(l+1)} - \log K^{(l+1)} + \log B\left(\frac{\alpha^{(l+1)}}{K^{(l+1)}} + m_{jj'}^{(l+1)}, n - m_{jj'}^{(l+1)} + 1\right)\right]\right. \\
&\qquad \left. - \sum_{jj'=1}^{K^{(l+1)}} \left[\log \alpha^{(l)} - \log K^{(l)} - \frac{\alpha^{(l)}}{K^{(l)}} \log \pi_{jj'} + \log B_{\pi_{jj'}}\left(\frac{\alpha^{(l)}}{K^{(l)}} + m_{jj'}^{(l+1)}, n - m_{jj'}^{(l+1)} + 1\right)\right]\right)
\end{aligned}
\tag{12}
$$

Since $\pi_{jj'}$ changes, it becomes complicated to infer the value of $\alpha^{(l+1)}$ in (12). It is hard to find the exact analytical solution for $\alpha^{(l+1)}$ since the $\Pr(\boldsymbol{Y}^{(l+1)}|\alpha^{(l+1)})$ and $\Pr(\boldsymbol{Y}^{(l+1)}|\alpha^{(l)})$ may not be equal. In the next section, we could infer the empirical value of $\hat{\alpha}^{(l+1)}$ by statistical inference.

4 Posterior Computation and Inference for Truncated Beta-Binomial Matrix

4.1 Gibbs Sampling

To get the posterior inference, the most direct way is to use Gibbs Sampling:

With given $\alpha^{(l)}$, we could sample a number of n x $K^{(l)}$ binary matrix $\boldsymbol{Y}^{(l)}$. By dependency data for each column of $\boldsymbol{Y}^{(l)}$, generate an n x $K^{(l+1)}$ binary matrix $\boldsymbol{Y}^{(l+1)}$.

With $\boldsymbol{Y}^{(l+1)}$, set $\alpha^{(l+1)} \sim Gamma(a, b)$ as prior:

$$\Pr(\alpha^{(l+1)}) \propto (\alpha^{(l+1)})^{a-1} \exp\{-b\alpha^{(l+1)}\} \tag{13}$$

Based on the concept of Beta-Binomial matrix in (13), the likelihood becomes:

$$\begin{aligned}
\Pr(\boldsymbol{Y}^{(l+1)}|\alpha^{(l+1)}) &= \prod_{jj'=1}^{K^{(l+1)}} \Pr(\boldsymbol{Y}_{jj'}^{(l+1)}|\alpha^{(l+1)}) \\
&= \prod_{jj'=1}^{K^{(l+1)}} \frac{\alpha^{(l+1)}}{K^{(l+1)}} B\left(\frac{\alpha^{(l+1)}}{K^{(l+1)}} + m_{jj'}^{(l+1)}, n - m_{jj'}^{(l+1)} + 1\right) \\
&= \exp\left\{\sum_{jj'=1}^{K^{(l+1)}} \log\left[\frac{\alpha^{(l+1)}}{K^{(l+1)}} B\left(\frac{\alpha^{(l+1)}}{K^{(l+1)}} + m_{jj'}^{(l+1)}, n - m_{jj'}^{(l+1)} + 1\right)\right]\right\}
\end{aligned} \tag{14}$$

Based on (13) and (14), the posterior probability of $\alpha^{(l+1)}$ would be:

$$\begin{aligned}
\Pr(\alpha^{(l+1)}|\boldsymbol{Y}^{(l+1)}) &\propto \Pr(\boldsymbol{Y}^{(l+1)}|\alpha^{(l+1)}) \Pr(\alpha^{(l+1)}) \\
&\propto (\alpha^{(l+1)})^{a-1} \exp\left\{-b\alpha^{(l+1)} + \sum_{jj'=1}^{K^{(l+1)}} \log\left[\frac{\alpha^{(l+1)}}{K^{(l+1)}} B\left(\frac{\alpha^{(l+1)}}{K^{(l+1)}} + m_{jj'}^{(l+1)}, n - m_{jj'}^{(l+1)} + 1\right)\right]\right\} \\
&= (\alpha^{(l+1)})^{a-1} \exp\left\{-\alpha^{(l+1)}\left[b - \sum_{jj'=1}^{K^{(l+1)}} \log\left(\frac{B\left(\frac{\alpha^{(l+1)}}{K^{(l+1)}} + m_{jj'}^{(l+1)}, n - m_{jj'}^{(l+1)} + 1\right)}{K^{(l+1)}}\right)\right]\right\} \\
&\sim Gamma\left(a, b - \sum_{jj'=1}^{K^{(l+1)}} \log\left(\frac{B\left(\frac{\alpha^{(l+1)}}{K^{(l+1)}} + m_{jj'}^{(l+1)}, n - m_{jj'}^{(l+1)} + 1\right)}{K^{(l+1)}}\right)\right) \\
&\sim Gamma\left(a, b - \log \Pr(\boldsymbol{Y}^{(l+1)}|\alpha^{(l+1)})\right)
\end{aligned} \tag{15}$$

For iteration t:

1. Draw $(\log \Pr(\boldsymbol{Y}^{(l+1)}|\alpha^{(l+1)}))^{(t)}$ with $(\alpha^{(l+1)})^{(t-1)}$.
2. Draw $(\alpha^{(l+1)})^{(t)} \sim \text{Gamma}\left(a, b - (\log \Pr(\boldsymbol{Y}^{(l+1)}|\alpha^{(l+1)}))^{(t)}\right)$.

4.2 Maximum Likelihood Estimation

To sample α from $Gamma(a, b)$, we could set up proper initial value of α based on Maximum Likelihood Estimation.

Given a n x K binary matrix $\boldsymbol{Y}$, we could calculate estimated $\hat{\pi}_j$ for j = 1, ..., K for each column by MLE:

$$\hat{\pi}_j = \frac{\sum_{i=1}^{n} y_{ij}}{n}$$

With estimated $\hat{\pi}_1$, ..., $\hat{\pi}_K$, we could calculate $\hat{\alpha}_{MLE}$ by MLE

$$\frac{\hat{\alpha}_{MLE}}{K} = \frac{K}{\sum_{j=1}^{K} \log(\hat{\pi}_j)}$$

$$\hat{\alpha}_{MLE} = \frac{K^2}{\sum_{j=1}^{K} \log(\hat{\pi}_j)}$$

To prevent $\log(\hat{\pi}_j)$ to be infinity when $\hat{\pi}_j = 0$, add $\epsilon = 5 \times 10^{-5}$ to the index where $\hat{\pi}_j = 0$. With Maximum likelihood estimator $\hat{\alpha}_{MLE}$, we could set up proper initial values of a and b such that $\hat{\alpha}_{MLE} = \frac{a}{b}$ for Gibbs Sampling.

After getting the value of $\alpha^{(l+1)}$, when $K^{(l+1)} \to \infty$, convert this truncated Beta-Binomial matrix into an IBP with parameter $\alpha^{(l+1)}$ and $K^{(l+1)}$.

4.3 Limitation

From the code and grid plot, we observe the difference between $\log \Pr(\boldsymbol{Y}^{(l+1)} \mid \alpha^{(l+1)})$ and $\log \Pr(\boldsymbol{Y}^{(l+1)} \mid \alpha^{(l)})$. However, the choice of $\alpha^{(l+1)}$ that minimizes this difference is empirically close to $\alpha^{(l+1)}_{\text{MLE}}$.

It means that with the dependency structure, the marginal probabilities $\pi^{(l+1)}$ in layer (l + 1) is sometimes not strictly follow Beta(a, 1) distribution with any choice of parameter a. We could only choose the parameter a to make it close to a Beta distribution. (Note that $a = \frac{\alpha^{(l+1)}}{K^{(l+1)}}$)

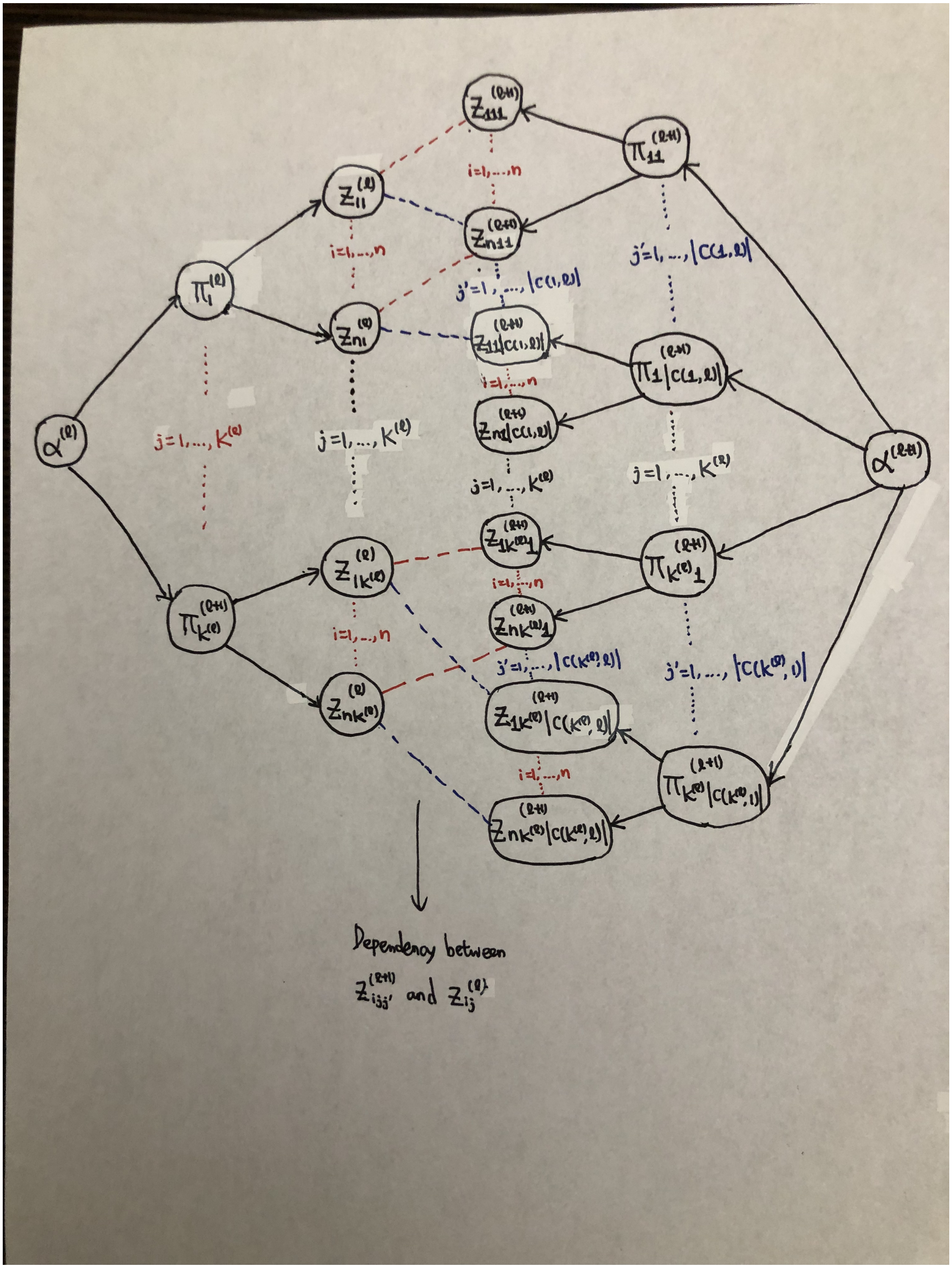

Figure 1: Hierarchical Model for Taxonomic IBP

5 Taxonomic Indian Buffet Process

From Truncated-Beta Binomial Matrix, it is straightforward to find empirical $\alpha^{(l+1)}$ value from $\alpha^{(l)}$ by MLE and Gibbs Sampling.

Extending this idea to Indian Buffet Process with given $\alpha^{(l)}$, $K^{(l)}$, and $K^{(l+1)}$.

Firstly, we convert the regularized incomplete beta function into the product of Beta function and unregularized beta function:

$$
\begin{aligned}
&B_{\pi_{jj'}}\left(\frac{\alpha^{(l)}}{K^{(l)}}+m_{jj'}^{(l+1)}, n-m_{jj'}^{(l+1)}+1\right)=B\left(\frac{\alpha^{(l)}}{K^{(l)}}+m_{jj'}^{(l+1)}, n-m_{jj'}^{(l+1)}+1\right)\cdot \\
&I_{\pi_{jj'}}\left(\frac{\alpha^{(l)}}{K^{(l)}}+m_{jj'}^{(l+1)}, n-m_{jj'}^{(l+1)}+1\right)
\end{aligned}
$$

Based on (10), for a specific column $\boldsymbol{Y}_{jj'}^{(l+1)}$ in binary matrix $\boldsymbol{Y}^{(l+1)}$, the upper bound would be:

$$
\begin{aligned}
\Pr(\boldsymbol{Y}_{jj'}^{(l+1)}|\alpha^{(l)}) &= \frac{\alpha^{(l)}}{K^{(l)}(\pi_{jj'})^{\frac{\alpha^{(l)}}{K^{(l)}}}} B_{\pi_{jj'}}\left(\frac{\alpha^{(l)}}{K^{(l)}}+m_{jj'}^{(l+1)}, n-m_{jj'}^{(l+1)}+1\right) \\
&= \frac{\alpha^{(l)}}{K^{(l)}(\pi_{jj'})^{\frac{\alpha^{(l)}}{K^{(l)}}}} B\left(\frac{\alpha^{(l)}}{K^{(l)}}+m_{jj'}^{(l+1)}, n-m_{jj'}^{(l+1)}+1\right) I_{\pi_{jj'}}\left(\frac{\alpha^{(l)}}{K^{(l)}}+m_{jj'}^{(l+1)}, n-m_{jj'}^{(l+1)}+1\right) \\
&= (\pi_{jj'})^{-\frac{\alpha^{(l)}}{K^{(l)}}} \frac{\alpha^{(l)}}{K^{(l)}} \frac{\Gamma(\frac{\alpha^{(l)}}{K^{(l)}}+m_{jj'}^{(l+1)})\Gamma(n-m_{jj'}^{(l+1)}+1)}{\Gamma(n+1+\frac{\alpha^{(l)}}{K^{(l)}})} I_{\pi_{jj'}}\left(\frac{\alpha^{(l)}}{K^{(l)}}+m_{jj'}^{(l+1)}, n-m_{jj'}^{(l+1)}+1\right)
\end{aligned}
$$

Then, for the binary matrix $\boldsymbol{Y}^{(l+1)}$:

$$
\begin{aligned}
\Pr(\boldsymbol{Y}^{(l+1)}|\alpha^{(l)}) &= \prod_{jj'=1}^{K^{(l+1)}} \Pr(\boldsymbol{Y}_{jj'}^{(l+1)}|\alpha^{(l)}) \\
&= \prod_{jj'=1}^{K^{(l+1)}} (\pi_{jj'})^{-\frac{\alpha^{(l)}}{K^{(l)}}} \frac{\alpha^{(l)}}{K^{(l)}} \frac{\Gamma(\frac{\alpha^{(l)}}{K^{(l)}} + m_{jj'}^{(l+1)})\Gamma(n - m_{jj'}^{(l+1)} + 1)}{\Gamma(n+1+\frac{\alpha^{(l)}}{K^{(l)}})} \\
&\times I_{\pi_{jj'}}\left(\frac{\alpha^{(l)}}{K^{(l)}} + m_{jj'}^{(l+1)}, n - m_{jj'}^{(l+1)} + 1\right) \\
&= \left(\prod_{jj'=1}^{K^{(l+1)}} \pi_{jj'}\right)^{-\frac{\alpha^{(l)}}{K^{(l)}}} \prod_{jj'=1}^{K^{(l+1)}} \frac{\frac{\alpha^{(l)}}{K^{(l)}}\Gamma(\frac{\alpha^{(l)}}{K^{(l)}} + m_{jj'}^{(l+1)})\Gamma(n - m_{jj'}^{(l+1)} + 1)}{\Gamma(n+1+\frac{\alpha^{(l)}}{K^{(l)}})} \\
&\times \prod_{k=1}^{K^{(l+1)}} I_{\pi_{jj'}}\left(\frac{\alpha^{(l)}}{K^{(l)}} + m_{jj'}^{(l+1)}, n - m_{jj'}^{(l+1)} + 1\right) \\
&= \left(\prod_{jj'=1}^{K^{(l+1)}} \pi_{jj'}\right)^{-\frac{\alpha^{(l)}}{K^{(l)}}} \left(\frac{\frac{\alpha^{(l)}}{K^{(l)}}\Gamma(\frac{\alpha^{(l)}}{K^{(l)}})\Gamma(n+1)}{\Gamma(n+1+\frac{\alpha^{(l)}}{K^{(l)}})}\right)^{K_0^{(l+1)}} \\
&\times \prod_{k=1}^{K_+^{(l+1)}} \frac{\frac{\alpha^{(l)}}{K^{(l)}}\Gamma(\frac{\alpha^{(l)}}{K^{(l)}} + m_k^{(l+1)})\Gamma(n - m_k^{(l+1)} + 1)}{\Gamma(n+1+\frac{\alpha^{(l)}}{K^{(l)}})} \prod_{k=1}^{K^{(l+1)}} I_{\pi_{jj'}}\left(\frac{\alpha^{(l)}}{K^{(l)}} + m_{jj'}^{(l+1)}, n - m_{jj'}^{(l+1)} + 1\right)
\end{aligned}
$$

(continued in next page)

$$
\begin{aligned}
&= \left(\prod_{jj'=1}^{K^{(l+1)}} \pi_{jj'}\right)^{-\frac{\alpha^{(l)}}{K^{(l)}}} \left(\frac{\frac{\alpha^{(l)}}{K^{(l)}}\Gamma(\frac{\alpha^{(l)}}{K^{(l)}})\Gamma(n+1)}{\Gamma(n+1+\frac{\alpha^{(l)}}{K^{(l)}})}\right)^{K^{(l+1)}-K_+^{(l+1)}} \\
&\times \prod_{k=1}^{K_+^{(l+1)}} \frac{\frac{\alpha^{(l)}}{K^{(l)}}\Gamma(\frac{\alpha^{(l)}}{K^{(l)}}+m_k^{(l+1)})\Gamma(n-m_k^{(l+1)}+1)}{\Gamma(n+1+\frac{\alpha^{(l)}}{K^{(l)}})} \prod_{k=1}^{K^{(l+1)}} I_{\pi_{jj'}}\left(\frac{\alpha^{(l)}}{K^{(l)}}+m_{jj'}^{(l+1)}, n-m_{jj'}^{(l+1)}+1\right) \\
&= \left(\prod_{jj'=1}^{K^{(l+1)}} \pi_{jj'}\right)^{-\frac{\alpha^{(l)}}{K^{(l)}}} \left(\frac{\frac{\alpha^{(l)}}{K^{(l)}}\Gamma(\frac{\alpha^{(l)}}{K^{(l)}})\Gamma(n+1)}{\Gamma(n+1+\frac{\alpha^{(l)}}{K^{(l)}})}\right)^{K^{(l+1)}} \\
&\times \prod_{k=1}^{K_+^{(l+1)}} \frac{\Gamma(\frac{\alpha^{(l)}}{K^{(l)}}+m_k^{(l+1)})\Gamma(n-m_k^{(l+1)}+1)}{\Gamma(\frac{\alpha^{(l)}}{K^{(l)}})\Gamma(n+1)} \prod_{k=1}^{K^{(l+1)}} I_{\pi_{jj'}}\left(\frac{\alpha^{(l)}}{K^{(l)}}+m_{jj'}^{(l+1)}, n-m_{jj'}^{(l+1)}+1\right) \\
&= \left(\prod_{jj'=1}^{K^{(l+1)}} \pi_{jj'}\right)^{-\frac{\alpha^{(l)}}{K^{(l)}}} \left(\frac{\Gamma(\frac{\alpha^{(l)}}{K^{(l)}}+1)\cdot n!}{(n+\frac{\alpha^{(l)}}{K^{(l)}})\cdots(\frac{\alpha^{(l)}}{K^{(l)}}+1)\Gamma(\frac{\alpha^{(l)}}{K^{(l)}}+1)}\right)^{K^{(l+1)}} \\
&\times \prod_{k=1}^{K_+^{(l+1)}} \frac{(\frac{\alpha^{(l)}}{K^{(l)}}+m_k^{(l+1)}-1)\cdots(\frac{\alpha^{(l)}}{K^{(l)}}+1)\cdot(\frac{\alpha^{(l)}}{K^{(l)}})\cdot\Gamma(\frac{\alpha^{(l)}}{K^{(l)}})\cdot(n-m_k^{(l+1)})!}{\Gamma(\frac{\alpha^{(l)}}{K^{(l)}})\cdot n!} \\
&\times \prod_{k=1}^{K^{(l+1)}} I_{\pi_{jj'}}\left(\frac{\alpha^{(l)}}{K^{(l)}}+m_{jj'}^{(l+1)}, n-m_{jj'}^{(l+1)}+1\right) \\
&= \left(\prod_{jj'=1}^{K^{(l+1)}} \pi_{jj'}\right)^{-\frac{\alpha^{(l)}}{K^{(l)}}} \left(\frac{n!}{\prod_{j=1}^{n}(j+\frac{\alpha^{(l)}}{K^{(l)}})}\right)^{K^{(l+1)}} \prod_{k=1}^{K^{(l+1)}} I_{\pi_{jj'}}\left(\frac{\alpha^{(l)}}{K^{(l)}}+m_{jj'}^{(l+1)}, n-m_{jj'}^{(l+1)}+1\right) \\
&\times \prod_{k=1}^{K_+^{(l+1)}} \frac{\alpha^{(l)}}{K^{(l)}} \frac{\prod_{j=1}^{m_k^{(l+1)}-1}(\frac{\alpha^{(l)}}{K^{(l)}}+j)\cdot(n-m_k^{(l+1)})!}{n!}
\end{aligned}
\tag{16}
$$

Denote lof-equivalence class of binary matrix $\boldsymbol{Y}^{(l+1)}$ as $[\boldsymbol{Y}^{(l+1)}]$ (Griffiths and Ghahramani, 2011). The cardinality of $[\boldsymbol{Y}^{(l+1)}]$:

$$
|[\boldsymbol{Y}^{(l+1)}]| = \frac{K^{(l+1)}!}{\prod_{h=0}^{2^n-1} K_h^{(l+1)}!} = \frac{K^{(l+1)}!}{K_0^{(l+1)}!\prod_{h=1}^{2^n-1} K_h^{(l+1)}!}
$$

Then,

$$
\begin{aligned}
\Pr\left([\boldsymbol{Y}^{(l+1)}]|\alpha^{(l)}\right) &= \frac{K^{(l+1)}!}{K_0^{(l+1)}!\prod_{h=1}^{2^n-1}K_h^{(l+1)}!}\left(\prod_{jj'=1}^{K^{(l+1)}}\pi_{jj'}\right)^{-\frac{\alpha^{(l)}}{K^{(l)}}}\left(\frac{n!}{\prod_{j=1}^{n}(j+\frac{\alpha^{(l)}}{K^{(l)}})}\right)^{K^{(l+1)}} \\
&\times\prod_{k=1}^{K_+^{(l+1)}}\frac{\alpha^{(l)}}{K^{(l)}}\frac{\prod_{j=1}^{m_k^{(l+1)}-1}(\frac{\alpha^{(l)}}{K^{(l)}}+j)\cdot(n-m_k^{(l+1)})!}{n!} \\
&\times\prod_{k=1}^{K^{(l+1)}}I_{\pi_{jj'}}\left(\frac{\alpha^{(l)}}{K^{(l)}}+m_{jj'}^{(l+1)},n-m_{jj'}^{(l+1)}+1\right) \\
&= \frac{K^{(l+1)}!}{K_0^{(l+1)}!}\frac{1}{\prod_{h=1}^{2^n-1}K_h^{(l+1)}!}\left(\prod_{jj'=1}^{K^{(l+1)}}\pi_{jj'}\right)^{-\frac{\alpha^{(l)}}{K^{(l)}}}\left(\frac{n!}{\prod_{j=1}^{n}(j+\frac{\alpha^{(l)}}{K^{(l)}})}\right)^{K^{(l+1)}} \\
&\times\left(\frac{\alpha^{(l)}}{K^{(l)}}\right)^{K_+^{(l+1)}}\prod_{k=1}^{K_+^{(l+1)}}\frac{\prod_{j=1}^{m_k^{(l+1)}-1}(\frac{\alpha^{(l)}}{K^{(l)}}+j)\cdot(n-m_k^{(l+1)})!}{n!} \\
&\times\prod_{k=1}^{K^{(l+1)}}I_{\pi_{jj'}}\left(\frac{\alpha^{(l)}}{K^{(l)}}+m_{jj'}^{(l+1)},n-m_{jj'}^{(l+1)}+1\right) \\
&= \frac{K^{(l+1)}!}{K_0^{(l+1)}!K^{(l+1)K_+^{(l+1)}}}\frac{\left(\frac{\alpha^{(l)}K^{(l+1)}}{K^{(l)}}\right)^{K_+^{(l+1)}}}{\prod_{h=1}^{2^n-1}K_h^{(l+1)}!}\left(\prod_{jj'=1}^{K^{(l+1)}}\pi_{jj'}\right)^{-\frac{\alpha^{(l)}}{K^{(l)}}}\left(\frac{n!}{\prod_{j=1}^{n}(j+\frac{\alpha^{(l)}}{K^{(l)}})}\right)^{K^{(l+1)}} \\
&\times\prod_{k=1}^{K_+^{(l+1)}}\frac{\prod_{j=1}^{m_k^{(l+1)}-1}(\frac{\alpha^{(l)}}{K^{(l)}}+j)\cdot(n-m_k^{(l+1)})!}{n!}\prod_{k=1}^{K^{(l+1)}}I_{\pi_{jj'}}\left(\frac{\alpha^{(l)}}{K^{(l)}}+m_{jj'}^{(l+1)},n-m_{jj'}^{(l+1)}+1\right)
\end{aligned}
\tag{17}
$$

Before taking the limit as $K^{(l+1)}\to\infty$, there are 2 assumptions:

(1). Assume $0<\pi_{jj'}\leq 1$. Then

$$
\begin{aligned}
0<\prod_{jj'=1}^{K^{(l+1)}}\pi_{jj'}\leq 1 \\
0<\lim_{K^{(l+1)}\to\infty}\prod_{jj'=1}^{K^{(l+1)}}\pi_{jj'}\leq 1
\end{aligned}
\tag{18}
$$

(2). Both layer l and (l + 1) have infinite features. The observed number of features in layer l is $K^{(l)}$ and the observed number of features in layer (l + 1) is $K^{(l+1)}$. Assume the increasing rates of $K^{(l+1)}$ and $K^{(l)}$ have the same ratio:

$$
\begin{aligned}
K^{(l+1)} &= sK^{(l)} \\
\lim_{(K^{(l)},K^{(l+1)})\to(\infty,\infty)} \frac{K^{(l+1)}}{K^{(l)}} &= s
\end{aligned} \tag{19}
$$

Therefore,

$$
\begin{aligned}
\lim_{K^{(l+1)}\to\infty} \Pr\left([\boldsymbol{Y}^{(l+1)}]|\alpha^{(l)}\right) = &\lim_{K^{(l+1)}\to\infty} \frac{K^{(l+1)}!}{K_0^{(l+1)}!K^{(l+1)K_+^{(l+1)}}} \frac{\left(\frac{\alpha^{(l)}K^{(l+1)}}{K^{(l)}}\right)^{K_+^{(l+1)}}}{\prod_{h=1}^{2^n-1} K_h^{(l+1)}!} \left(\prod_{jj'=1}^{K^{(l+1)}} \pi_{jj'}\right)^{-\frac{\alpha^{(l)}}{K^{(l)}}} \\
&\times \left(\frac{n!}{\prod_{j=1}^{n}(j+\frac{\alpha^{(l)}}{K^{(l)}})}\right)^{K^{(l+1)}} \prod_{k=1}^{K_+^{(l+1)}} \frac{\prod_{j=1}^{m_k^{(l+1)}-1}(\frac{\alpha^{(l)}}{K^{(l)}}+j)\cdot(n-m_k^{(l+1)})!}{n!} \\
&\times \prod_{k=1}^{K^{(l+1)}} I_{\pi_{jj'}}\left(\frac{\alpha^{(l)}}{K^{(l)}}+m_{jj'}^{(l+1)}, n-m_{jj'}^{(l+1)}+1\right)
\end{aligned}
$$

The upper bound is achieved when all $\pi_{jj'}=1$ for all jj', with

$$
\prod_{k=1}^{K^{(l+1)}} I_{\pi_{jj'}}\left(\frac{\alpha^{(l)}}{K^{(l)}}+m_{jj'}^{(l+1)}, n-m_{jj'}^{(l+1)}+1\right) = 1
$$

The upper bound would be:

$$
\begin{aligned}
\lim_{K^{(l+1)}\to\infty} \Pr\left([\boldsymbol{Y}^{(l+1)}]|\alpha^{(l)}\right) &= 1\cdot\frac{\left(\alpha^{(l)}s\right)^{K_+^{(l+1)}}}{\prod_{h=1}^{2^n-1} K_h^{(l+1)}!}\cdot 1\cdot\exp\{-\alpha^{(l)}sH_n\}\cdot\prod_{k=1}^{K_+^{(l+1)}} \frac{(m_k^{(l+1)}-1)!(n-m_k^{(l+1)})!}{n!} \\
&= \frac{\left(\alpha^{(l)}s\right)^{K_+^{(l+1)}}}{\prod_{h=1}^{2^n-1} K_h^{(l+1)}!} \exp\{-\alpha^{(l)}sH_n\} \prod_{k=1}^{K_+^{(l+1)}} \frac{(m_k^{(l+1)}-1)!(n-m_k^{(l+1)})!}{n!}
\end{aligned} \tag{20}
$$

The specific steps to take the limit for each element in $\Pr\left([\boldsymbol{Y}^{(l+1)}]|\alpha^{(l)}\right)$ is in appendix.

As $K^{(l+1)}\to\infty$, the upper bound of occurrence for each row would be $\alpha^{(l)}s$ in layer $(l+1)$, given that the expected occurrence for each row in layer l is $\alpha^{(l)}$. From the definition in (3):

$$
K_+^{(l+1)} \sim \text{Poisson}(\alpha^{(l)}sH_n), \quad \mathbb{E}\left[K_+^{(l+1)}\right] = \alpha^{(l)}sH_n \tag{21}
$$

We could see that the dependency structure between layer l and layer (l + 1) doesn't affect the upper bound of expected occurrence in layer (l + 1). Only the ratio of observed number of features, s, affects the upper bound of expected occurrence in layer (l + 1).

6 Posterior Computation and Inference for Taxonomic Indian Buffet Process

6.1 Gibbs Sampling

For layer (l + 1), we could use Gibbs sampling to compute the predictive posterior distribution of $y_{ik}^{(l+1)}|\boldsymbol{y}_{-i,k}^{(l+1)}$ (Griffiths and Ghahramani, 2011):

$$\begin{aligned}\Pr(\pi_k^{(l+1)}|\boldsymbol{y}_{-i,k}^{(l+1)}) &\propto \Pr(\boldsymbol{y}_{-i,k}^{(l+1)}|\pi_k^{(l+1)})\Pr(\pi_k^{(l+1)}) \\ &\propto (\pi_k^{(l+1)})^{m_{-i,k}}(1-\pi_k^{(l+1)})^{n-1-m_{-i,k}}(\pi_k^{(l+1)})^{\frac{\alpha^{(l+1)}}{K^{(l+1)}}-1} \\ &\sim \mathrm{Beta}\left(\frac{\alpha^{(l+1)}}{K^{(l+1)}}+m_{-i,k}, n-m_{-i,k}\right)\end{aligned} \tag{22}$$

From equation (22):

$$\begin{aligned}\Pr(y_{ik}^{(l+1)}|\boldsymbol{y}_{-i,k}^{(l+1)}) &= \int_0^1 \Pr(y_{ik}^{(l+1)}|\pi_k^{(l+1)})\Pr(\pi_k^{(l+1)}|\boldsymbol{y}_{-i,k}^{(l+1)})d\pi_k^{(l+1)} \\ &= \int_0^1 (\pi_k^{(l+1)})^{y_{ik}^{(l+1)}}(1-\pi_k^{(l+1)})^{1-y_{ik}^{(l+1)}}\frac{(\pi_k^{(l+1)})^{\frac{\alpha^{(l+1)}}{K^{(l+1)}}+m_{-i,k}-1}(1-\pi_k^{(l+1)})^{n-m_{-i,k}-1}}{B\left(\frac{\alpha^{(l+1)}}{K^{(l+1)}}+m_{-i,k}, n-m_{-i,k}\right)}d\pi_k^{(l+1)} \\ &= \frac{B\left(\frac{\alpha^{(l+1)}}{K^{(l+1)}}+m_{-i,k}+y_{ik}^{(l+1)}, n-m_{-i,k}+1-y_{ik}^{(l+1)}\right)}{B\left(\frac{\alpha^{(l+1)}}{K^{(l+1)}}+m_{-i,k}, n-m_{-i,k}\right)}\end{aligned} \tag{23}$$

For the Indian Buffet Process, we could set $\alpha^{(l+1)} = \alpha^{(l)}s$

$$\Pr(y_{ik}^{(l+1)} = 1 | \boldsymbol{y}_{-i,k}^{(l+1)}) = \frac{B\left(\frac{\alpha^{(l+1)}}{K^{(l+1)}} + m_{-i,k} + 1, n - m_{-i,k} + 1 - 1\right)}{B\left(\frac{\alpha^{(l+1)}}{K^{(l+1)}} + m_{-i,k}, n - m_{-i,k}\right)}$$
$$= \frac{m_{-i,k} + \frac{\alpha^{(l+1)}}{K^{(l+1)}}}{n + \frac{\alpha^{(l+1)}}{K^{(l+1)}}}$$
$$= \frac{m_{-i,k} + \frac{\alpha^{(l)}s}{K^{(l+1)}}}{n + \frac{\alpha^{(l)}s}{K^{(l+1)}}}$$

In the infinity case, as $K^{(l+1)} \to \infty$,

$$\Pr(y_{ik}^{(l+1)} = 1 | \boldsymbol{y}_{-i,k}^{(l+1)}) = \frac{m_{-i,k}}{n}$$

7 Appendix A.

Note that we use assumption (1), equation (18) and assumption (2), equation (19) all the time.

7.1

$$\frac{K^{(l+1)}!}{K_0^{(l+1)}! K^{(l+1)K_+^{(l+1)}}} = \frac{\prod_{k=1}^{K_+^{(l+1)}} (K^{(l+1)} - k + 1)}{K^{(l+1)K_+^{(l+1)}}}$$
$$= \frac{K^{(l+1)K_+^{(l+1)}} - \frac{(K_+^{(l+1)}-1)K_+^{(l+1)}}{2} K^{(l+1)K_+^{(l+1)}-1} + \cdots + (-1)^{K_+^{(l+1)}-1}(K_+^{(l+1)} - 1)K^{(l+1)}}{K^{(l+1)K_+^{(l+1)}}}$$
$$= 1 - \frac{(K_+^{(l+1)} - 1)K_+^{(l+1)}}{2K^{(l+1)}} + \cdots + \frac{(-1)^{K_+^{(l+1)}-1}(K_+^{(l+1)} - 1)}{K^{(l+1)K_+^{(l+1)}-1}}$$

Then

$$\lim_{K^{(l+1)} \to \infty} \frac{K^{(l+1)}!}{K_0^{(l+1)}! K^{(l+1)K_+^{(l+1)}}} = \lim_{K^{(l+1)} \to \infty} \left[1 - \frac{(K_+^{(l+1)} - 1)K_+^{(l+1)}}{2K^{(l+1)}} + \cdots + \frac{(-1)^{K_+^{(l+1)}-1}(K_+^{(l+1)} - 1)}{K^{(l+1)K_+^{(l+1)}-1}}\right]$$
$$= 1 - 0 + \cdots + 0 = 1$$

(See the appendix A. in Griffiths and Ghahramani, 2011)

7.2

By assumption (2), equation (19),

$$K^{(l+1)} = sK^{(l)}$$

Then,

$$\frac{\left(\frac{\alpha^{(l)}K^{(l+1)}}{K^{(l)}}\right)^{K_+^{(l+1)}}}{\prod_{h=1}^{2^n-1} K_h^{(l+1)}!} = \frac{\left(\alpha^{(l)}s\right)^{K_+^{(l+1)}}}{\prod_{h=1}^{2^n-1} K_h^{(l+1)}!}$$

$$\lim_{K^{(l+1)}\to\infty} \frac{\left(\frac{\alpha^{(l)}K^{(l+1)}}{K^{(l)}}\right)^{K_+^{(l+1)}}}{\prod_{h=1}^{2^n-1} K_h^{(l+1)}!} = \frac{\left(\alpha^{(l)}s\right)^{K_+^{(l+1)}}}{\prod_{h=1}^{2^n-1} K_h^{(l+1)}!}$$

7.3

By asssumption (1), equation (18),

$$0 < \lim_{K^{(l+1)}\to\infty} \prod_{jj'=1}^{K^{(l+1)}} \pi_{jj'} \leq 1$$

Then,

$$\begin{aligned} \lim_{K^{(l+1)}\to\infty} \left(\prod_{jj'=1}^{K^{(l+1)}} \pi_{jj'}\right)^{-\frac{\alpha^{(l)}}{K^{(l)}}} &= \lim_{K^{(l+1)}\to\infty} \left(\prod_{jj'=1}^{K^{(l+1)}} \pi_{jj'}\right)^{-\frac{\alpha^{(l)}s}{K^{(l+1)}}} \\ &= 1 \end{aligned}$$

7.4

$$
\begin{aligned}
\left(\frac{n!}{\prod_{j=1}^{n}(j+\frac{\alpha^{(l)}}{K^{(l)}})}\right)^{K^{(l+1)}} &= \left(\frac{\prod_{j=1}^{n} j}{\prod_{j=1}^{n}(j+\frac{\alpha^{(l)}}{K^{(l)}})}\right)^{K^{(l+1)}} \\
&= \prod_{j=1}^{n}\left(\frac{j}{j+\frac{\alpha^{(l)}}{K^{(l)}}}\right)^{K^{(l+1)}} \\
&= \prod_{j=1}^{n}\left(\frac{1}{1+\frac{\alpha^{(l)}\frac{1}{j}}{K^{(l)}}}\right)^{K^{(l+1)}} \\
&= \prod_{j=1}^{n}\left(\frac{1}{1+\frac{\alpha^{(l)}s\frac{1}{j}}{K^{(l+1)}}}\right)^{K^{(l+1)}}
\end{aligned}
$$

Use the fact that

$$
\lim_{K^{(l+1)}\to\infty}\left(\frac{1}{1+\frac{x}{K^{(l+1)}}}\right)^{K^{(l+1)}} = \exp\{-x\}
$$

$$
\begin{aligned}
\lim_{K^{(l+1)}\to\infty}\left(\frac{n!}{\prod_{j=1}^{n}(j+\frac{\alpha^{(l)}}{K^{(l)}})}\right)^{K^{(l+1)}} &= \lim_{K^{(l+1)}\to\infty}\prod_{j=1}^{n}\left(\frac{1}{1+\frac{\alpha^{(l)}s\frac{1}{j}}{K^{(l+1)}}}\right)^{K^{(l+1)}} \\
&= \prod_{j=1}^{n}\exp\left\{-\alpha^{(l)}s\frac{1}{j}\right\} \\
&= \exp\left\{-\alpha^{(l)}s\sum_{j=1}^{n}\frac{1}{j}\right\} \\
&= \exp\left\{-\alpha^{(l)}sH_n\right\}
\end{aligned}
$$

(See the appendix A. in Griffiths and Ghahramani, 2011)

7.5

$$
\prod_{k=1}^{K_+^{(l+1)}}\frac{\prod_{j=1}^{m_k^{(l+1)}-1}(\frac{\alpha^{(l)}}{K^{(l)}}+j)\cdot(n-m_k^{(l+1)})!}{n!} = \prod_{k=1}^{K_+^{(l+1)}}\frac{\prod_{j=1}^{m_k^{(l+1)}-1}(\frac{\alpha^{(l)}s}{K^{(l+1)}}+j)\cdot(n-m_k^{(l+1)})!}{n!}
$$

$$\begin{aligned}\lim_{K^{(l+1)}\to\infty}\prod_{k=1}^{K_+^{(l+1)}}\frac{\prod_{j=1}^{m_k^{(l+1)}-1}(\frac{\alpha^{(l)}}{K^{(l)}}+j)\cdot(n-m_k^{(l+1)})!}{n!} &= \lim_{K^{(l+1)}\to\infty}\prod_{k=1}^{K_+^{(l+1)}}\frac{\prod_{j=1}^{m_k^{(l+1)}-1}(\frac{\alpha^{(l)}s}{K^{(l+1)}}+j)\cdot(n-m_k^{(l+1)})!}{n!} \\ &= \prod_{k=1}^{K_+^{(l+1)}}\frac{\prod_{j=1}^{m_k^{(l+1)}-1}(0+j)\cdot(n-m_k^{(l+1)})!}{n!} \\ &= \prod_{k=1}^{K_+^{(l+1)}}\frac{(m_k^{(l+1)}-1)!\cdot(n-m_k^{(l+1)})!}{n!}\end{aligned}$$

8 Key Questions

1. How to set the Dependency Structure for the root (1st layer) and top layers when there are only a few variables there (i.e. the number of features is finite)?

2. Are there any additional constraints ignored in this dependency structure?

3. How to test the validity of this model?

4. How to guarantee that at least one of the $\Pr(y_{ijj'} = 1)$ when $\Pr(y_{ij} = 1)$? When we sample a binary matrix $\boldsymbol{Y}^{(l+1)}$ from $\boldsymbol{Y}^{(l)}$, should I accept only the binary matrix with at least one species present for that index when the corresponding genus present in that index?

5. We could find the upper bound of taxonomic IBP for layer (l + 1) given the layer l. For specific derivation of α^{l+1}, we could use MLE / Gibbs sampling for truncated Beta-Bernoulli matrix of $\alpha^{(l+1)}|\boldsymbol{Y}^{(l+1)}$. To derive the $\alpha^{(l+1)}$ in a closed form for taxonomic IBP, we need to solve the regularized incomplete beta function.

9 References

- Griffiths, Thomas L., and Zoubin Ghahramani. "The Indian Buffet Process: An Introduction and Review." Journal of Machine Learning Research 12.4 (2011).
- Williamson, Sinead, Peter Orbanz, and Zoubin Ghahramani. "Dependent Indian buffet processes." Proceedings of the thirteenth international conference on artificial intelligence and statistics. JMLR Workshop and Conference Proceedings, 2010.
- Miller, Kurt T., Thomas Griffiths, and Michael I. Jordan. "The phylogenetic Indian buffet process: A non-exchangeable nonparametric prior for latent features." arXiv preprint arXiv:1206.3279 (2012).

- Gershman, Samuel J., Peter I. Frazier, and David M. Blei. "Distance dependent infinite latent feature models." IEEE transactions on pattern analysis and machine intelligence 37.2 (2014): 334-345.
- Gu, Yuqi, and David B. Dunson. "Bayesian pyramids: Identifiable multilayer discrete latent structure models for discrete data." Journal of the Royal Statistical Society Series B: Statistical Methodology 85.2 (2023): 399-426.