

A LPMRF Gibbs Sampling Derivation

As described in the main paper, we develop an LPMRF Gibbs sampler by considering the most common form of Multinomial sampling, namely by taking the sum of a sequence of L Categorical variables. Thus, if $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_L \sim \text{Categorical}(\boldsymbol{\theta})$ and $\mathbf{x} = \sum_{\ell=1}^L \mathbf{w}_\ell \sim \text{Multinomial}(\boldsymbol{\theta}|L)$, then the probability of any particular sequence is merely the Multinomial probability scaled by the inverse of the multinomial coefficient:

$$\Pr(\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_L | \boldsymbol{\theta}) = \left(\binom{L}{x_1, x_2, \dots, x_p} \right)^{-1} \Pr_{\text{Mult}}(\mathbf{x} = \sum_{\ell=1}^L \mathbf{w}_\ell | \boldsymbol{\theta}, L). \quad (4)$$

In a similar way, we can implicitly derive the probability for a particular sequence of words whose sum is distributed as an LPMRF:

$$\begin{aligned} \Pr(\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_L | \boldsymbol{\theta}, \Phi) &= \left(\binom{L}{x_1, x_2, \dots, x_p} \right)^{-1} \Pr_{\text{LPMRF}}(\mathbf{x} = \sum_{\ell=1}^L \mathbf{w}_\ell | \boldsymbol{\theta}, \Phi, L) \\ &= \exp(\boldsymbol{\theta}^T \mathbf{x} + \mathbf{x}^T \Phi \mathbf{x} - A_L(\boldsymbol{\theta}, \Phi) - \log(L!)). \end{aligned} \quad (5) \quad (6)$$

To develop a Gibbs sampler, we simply need to compute the conditional probability of one of these words given all the other words. Letting $\mathbf{x}_{-\ell} \equiv \sum_{m \neq \ell} \mathbf{w}_m$, then $\mathbf{x} = \mathbf{x}_{-\ell} + \mathbf{w}_\ell$. Thus, using the fact that the conditional distribution is proportional to the joint distribution, we can derive the form of the conditional distribution:

$$\Pr(\mathbf{w}_\ell = \mathbf{e}_s | \mathbf{w}_1, \dots, \mathbf{w}_{\ell-1}, \mathbf{w}_{\ell+1}, \dots, \mathbf{w}_L, \boldsymbol{\theta}, \Phi) \quad (7)$$

$$\propto \exp(\boldsymbol{\theta}^T (\mathbf{x}_{-\ell} + \mathbf{w}_\ell) + (\mathbf{x}_{-\ell} + \mathbf{w}_\ell)^T \Phi (\mathbf{x}_{-\ell} + \mathbf{w}_\ell)) \quad (8)$$

$$\propto \exp(\boldsymbol{\theta}_s + 2\Phi_s \mathbf{x}_{-\ell}). \quad (9)$$

Thus, each word can be sampled given the state of all the other words thus producing an LPMRF Gibbs sampler.

B Derivation of LPMRF Log Partition Upper Bound

$$A_L(\boldsymbol{\theta}, \Phi) \leq \log \left[\left(\sup_{\mathbf{x} \in \mathcal{X}_L} \exp(\mathbf{x}^T \Phi \mathbf{x}) \right) \sum_{\mathbf{x} \in \mathcal{X}_L} \exp(\boldsymbol{\theta}^T \mathbf{x} - \sum_s \log(x_s!)) \right] \quad (10) \quad \text{Hölder's Inequality}$$

$$= \log \left[\left(\sup_{\mathbf{x} \in \mathcal{X}_L} \exp(\mathbf{x}^T \Phi \mathbf{x}) \right) \exp(L \log(\sum_s \exp(\theta_s)) - \log(L!)) \right] \quad (11) \quad \text{Derived from Multinomial}$$

$$\leq \log \left[\exp(L^2 \lambda_{\Phi,1}) \exp(L \log(\sum_s \exp \theta_s) - \log(L!)) \right] \quad (12) \quad \text{Convex Relaxation of } \mathcal{X}_L$$

$$= L^2 \lambda_{\Phi,1} + L \log(\sum_s \exp \theta_s) - \log(L!), \quad (13)$$

where $\lambda_{\Phi,1}$ is the maximum eigenvalue of Φ . See next section for derivation of Eqn. 11.

B.1 Derivation of Multinomial Partition Function

Lemma 1. $\sum_{\mathbf{x} \in \mathcal{X}_L} \exp(\boldsymbol{\theta}^T \mathbf{x} - \sum_{s=1}^p \log(x_s!)) = \exp(L \log(\sum_s \exp \theta_s) - \log(L!))$

The derivation is based primarily on the fact that the above expression can be seen as the normalizing factor of a reparameterized Multinomial (or as a scaled version of a standard Multinomial parameterization).

Proof.

$$\begin{aligned}
& \sum_{\mathbf{x} \in \mathcal{X}_L} \exp(\boldsymbol{\theta}^T \mathbf{x} - \sum_{s=1}^p \log(x_s!)) \\
&= \sum_{\mathbf{x} \in \mathcal{X}_L} \exp(\boldsymbol{\theta}^T \mathbf{x} - \sum_{s=1}^p \log(x_s!) + c\mathbf{e}^T \mathbf{x} - c\mathbf{e}^T \mathbf{x} + \log(L!) - \log(L!)) \\
&\quad (\text{where } \mathbf{e} = [1, 1, \dots, 1]^T \text{ and } c \text{ is a constant}) \\
&= \frac{1}{L!} \sum_{\mathbf{x} \in \mathcal{X}_L} \exp((\boldsymbol{\theta} - c)^T \mathbf{x} - \sum_{s=1}^p \log(x_s!) + c\mathbf{e}^T \mathbf{x} + \log(L!)) \\
&= \frac{1}{L!} \sum_{\mathbf{x} \in \mathcal{X}_L} \exp((\boldsymbol{\theta} - c)^T \mathbf{x} - \sum_{s=1}^p \log(x_s!) + Lc + \log(L!)) \\
&= \frac{1}{L!} \exp(cL) \sum_{\mathbf{x} \in \mathcal{X}_L} \exp((\boldsymbol{\theta} - c)^T \mathbf{x} - \sum_{s=1}^p \log(x_s!) + \log(L!)) \\
&= \exp(Lc - \log(L!)) \sum_{\mathbf{x} \in \mathcal{X}_L} \exp((\boldsymbol{\theta} - c)^T \mathbf{x} - \sum_{s=1}^p \log(x_s!) + \log(L!)) \\
&\quad (\text{Letting } c = \log(\sum_s \exp(\theta_s))) \\
&= \exp(L \log(\sum_s \exp(\theta_s)) - \log(L!)) \sum_{\mathbf{x} \in \mathcal{X}_L} \exp((\boldsymbol{\theta} - \log(\sum_s \exp(\theta_s)))^T \mathbf{x} - \sum_{s=1}^p \log(x_s!) + \log(L!)) \\
&= \exp(L \log(\sum_s \exp(\theta_s)) - \log(L!)) \underbrace{\sum_{\mathbf{x} \in \mathcal{X}_L} \exp(\log(\boldsymbol{\rho})^T \mathbf{x} - \sum_{s=1}^p \log(x_s!) + \log(L!))}_{\text{Partition function of Multinomial} = 1} \tag{14} \\
&= \exp(L \log(\sum_s \exp(\theta_s)) - \log(L!)) \underbrace{\sum_{\mathbf{x} \in \mathcal{X}_L} \frac{L!}{\prod_{s=1}^p x_s!} \prod_{s=1}^p \rho_s^{x_s}}_{\text{Partition function of Multinomial} = 1} \\
&= \exp(L \log(\sum_s \exp(\theta_s)) - \log(L!))
\end{aligned}$$

where 14 is derived by showing that $\boldsymbol{\rho} = \exp(\boldsymbol{\theta} - \log(\sum_s \exp(\theta_s)))$ is a valid standard Multinomial parameter vector because the vector is positive and sums to 1:

$$\sum_{t=1}^p \exp\left(\theta_t - \log\left(\sum_s \exp(\theta_s)\right)\right) = \frac{1}{\sum_{s=1}^p \exp(\theta_s)} \sum_{t=1}^p \exp(\theta_t) = \frac{\sum_{t=1}^p \exp(\theta_t)}{\sum_{s=1}^p \exp(\theta_s)} = 1.$$

□

C Example of Log Partition Estimation

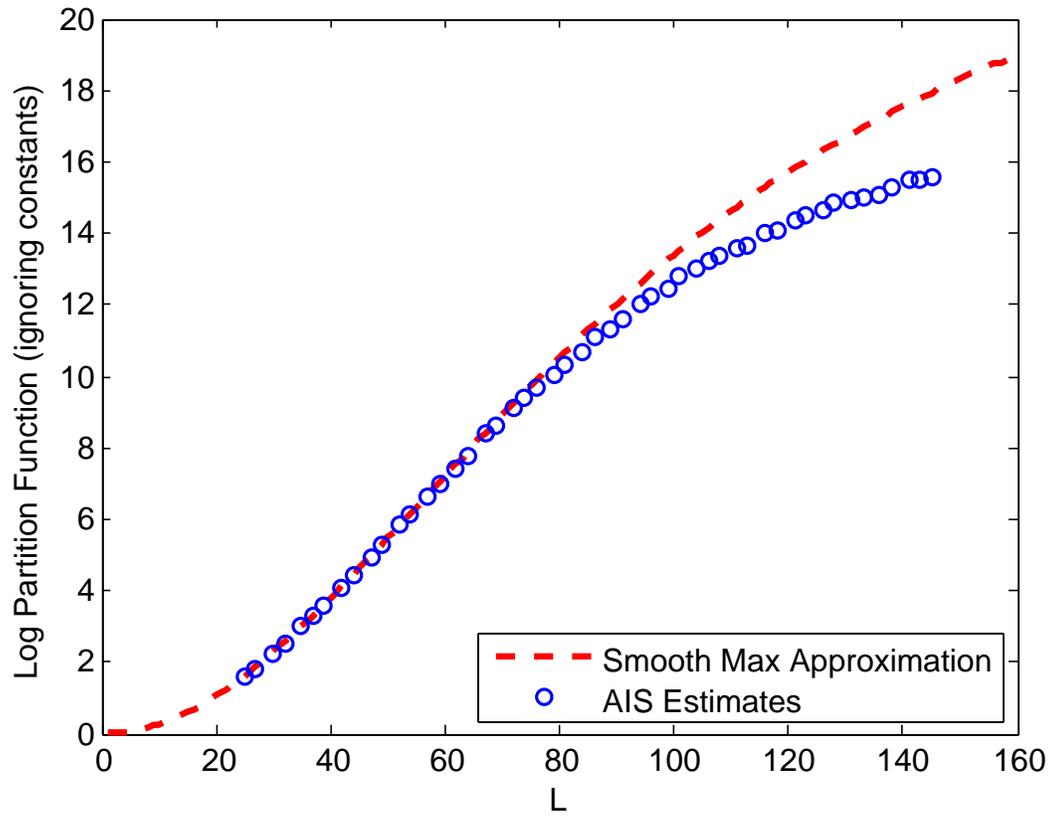


Figure 5: Example of log partition estimation for all values of L .