An Infinite Factor Model Hierarchy Via a Noisy-Or Mechanism: Supplemental Material

Aaron C. Courville, Douglas Eck and Yoshua Bengio
Department of Computer Science and Operations Research
University of Montréal
Montréal, Québec, Canada
{courvila,eckdoug,bengioy}@iro.umontreal.ca

In sampling from the infinite factor model hierarchy we require an expression for the posterior $p(\nu_k^o \mid \nu_{k-1}^o, Y^+, z_{:,\geq k}^o = 0)$. We first note that we will use the ARS sampling scheme and therefore only require the posterior density up to a normalizing constant. We decompose the posterior as follows:

$$p(\nu_k^o \mid \nu_{k-1}^o, Y^+, z_{:,\geq k}^o = 0) \propto P(z_{:,>k}^o = 0 \mid \nu_k^o, Y^+) P(z_{:,k}^o = 0 \mid \nu_k^o, Y^+) p(\nu_k^o \mid \nu_{k-1}^o) \quad (1)$$

The prior term is readily available from the stick breaking construction of the IBP [1]: $p(\nu_k^o \mid \nu_{k-1}^o) = \alpha_\nu \nu_{k-1}^{-\alpha_\nu} \nu_k^{\alpha_\nu - 1} \mathbb{I}(0 \leq \nu_k \leq \nu_{k-1})$. According to the noisy-or mechanism defined in Eq. 1 in the main document, $z_{n,k}^o = 0$ implies that the binary trials associated with each (active) $y_{n,k}^+ = 1$ failed. Thus we have the following expression for $p(z_{:,k}^o = \mathbf{0} \mid \nu_k^o, Y^+)$:

$$
\begin{aligned}
P(z_{:,k}^o = 0 \mid \nu_k^o, Y^+) &= \prod_{j=1}^{J^+} \int P(z_{:,k}^o \mid V_{jk}, Y^+) p(V_{jk} \mid \nu_k^o) \, dV_{jk} \\
&= \prod_{j=1}^{J^+} \int_0^1 V_{jk}^{c\nu_k^o - 1} (1 - V_{jk})^{N_j} \frac{\Gamma(c+1)}{\Gamma(c\nu_k^o)\Gamma(c(1-\nu_k^o)+1)} \, dV_{jk} \\
&= \prod_{j=1}^{J^+} \frac{\Gamma(c+1)}{\Gamma(c+1+N_j)} \frac{\Gamma(c(1-\nu_k^o)+N_j+1)}{\Gamma(c(1-\nu_k^o)+1)}
\end{aligned}
$$

Here J^+ is the number of binary variables $y_{:,j}$ active on any data example (those with at least one $y_{ij} = 1$, for at least one element i). It remains to derive the expression for $P(z_{:,>k}^o = 0 \mid \nu_k^o, Y^+)$. From the standard stick-breaking construction of the IBP [1], given ν_k^o, ν_l^o and $z_{n,l}$ (with $l > k$) are conditionally i.i.d. across l. Therefore,

$$P(z_{:,>k}^+ = 0 \mid \nu_k^o, Y^+) = \lim_{K \to \infty} \left(\int_0^{\nu_k^o} \frac{\alpha}{K} (\nu_k^o)^{-\alpha/K} \nu^{\alpha/K - 1} \prod_{j=1}^{J^+} \frac{\Gamma(c+1)}{\Gamma(c+1+N_j)} \frac{\Gamma(c(1-\nu)+N_j+1)}{(c(1-\nu))\Gamma(c(1-\nu))} \right)^{K-k} d\nu$$

For convenience, we define: $F = \left[\prod_{j=1}^{J^+} \frac{\Gamma(c+1)}{\Gamma(c+1+N_j)} \right]$. We now execute a change of variables. Letting $\eta = \nu/\nu_k^o$ and the corresponding differential $d\eta = 1/\nu_k^o \, d\nu$, then:

$$P(z^o_{:,>k} = 0 \mid \nu^o_k, Y^+)$$

$$= \lim_{K \to \infty} \left(\int_0^1 \frac{\alpha}{K} (\nu^o_k)^{-\alpha/K} (\eta \nu^o_k)^{\alpha/K-1} \nu^o_k \prod_{j=1}^{J^+} \frac{\Gamma(c+1)}{\Gamma(c+1+N_j)} \frac{\Gamma(c(1-\eta\nu^o_k)+N_j+1)}{(c(1-\eta\nu^o_k))\Gamma(c(1-\eta\nu^o_k))} \, d\eta \right)^{K-k}$$

$$= \lim_{K \to \infty} \left(\frac{\alpha}{K} F \int_0^1 \eta^{\alpha/K-1} \prod_{j=1}^{J^+} \frac{\Gamma(c(1-\eta\nu^o_k)+N_j+1)}{(c(1-\eta\nu^o_k))\Gamma(c(1-\eta\nu^o_k))} \, d\eta \right)^{K-k}$$

$$= \lim_{K \to \infty} \left(\frac{\alpha}{K} F \int_0^1 \eta^{\alpha/K-1} \prod_{j=1}^{J^+} \frac{1}{(c(1-\eta\nu^o_k))} \sum_{i'=0}^{N_j+1} (-1)^{N_j+1+i'} S^{(i')}_{N_j+1} (c(1-\eta\nu^o_k))^{i'} \, d\eta \right)^{K-k} \quad \text{with Stirling num. } S^{(i')}_{N_j+1}$$

$$= \lim_{K \to \infty} \left(\frac{\alpha}{K} F \int_0^1 \eta^{\alpha/K-1} \prod_{j=1}^{J^+} \sum_{i=0}^{N_j} (-1)^{N_j+i} S^{(i+1)}_{N_j+1} (c(1-\eta\nu^o_k))^i \, d\eta \right)^{K-k} \quad \text{where } i = i'-1 \text{ and } S^{(0)}_{N_j} = 0$$

$$= \lim_{K \to \infty} \left(\frac{\alpha}{K} F \int_0^1 \eta^{\alpha/K-1} \sum_{i=0}^{N_1+\cdots+N_{J^+}} w_i c^i (1-\eta\nu^o_k)^i \, d\eta \right)^{K-k} \quad w_i \text{ is def. via the convolution of the Stirlings}$$

$$= \lim_{K \to \infty} \left(\frac{\alpha}{K} F \sum_{i=0}^{N_1+\cdots+N_{J^+}} w_i c^i \sum_{l=0}^{i} \binom{i}{l} (1-\nu^o_k)^l \frac{\Gamma(i-l+1)\Gamma(l+\alpha/K)}{\Gamma(i+\alpha/K+1)} \right)^{K-k}$$

$$= \lim_{K \to \infty} \left(\frac{\alpha}{K} F \sum_{i=0}^{N_1+\cdots+N_{J^+}} w_i c^i \left[\frac{i!}{\prod_{n=0}^{i}(n+\alpha/K)} + \sum_{l=1}^{i} \frac{i! \prod_{m=0}^{l-1}(m+\alpha/K)}{l! \prod_{q=0}^{i}(q+\alpha/K)} (1-\nu^o_k)^l \right] \right)^{K-k}$$

$$= \lim_{K \to \infty} \left(F \sum_{i=0}^{N_1+\cdots+N_{J^+}} w_i c^i F \left[1 + \sum_{l=1}^{i} \frac{\alpha}{K} \frac{(1-\nu^o_k)^l}{l} \prod_{m=1}^{l-1} \left(1 + \frac{\alpha}{mK} \right) \right] \right)^{K-k}$$

Applying l'Hôpital's rule and simplifying, we get:

$$P(z^o_{:,>k} = 0 \mid \nu^o_k, Y^+)$$

$$= \exp \left(\left[\alpha \prod_{j=1}^{J^+} \frac{\Gamma(c+1)}{\Gamma(c+1+N_j)} \right] \sum_{i=0}^{N_1+\cdots+N_{J^+}} w_i c^i \left(-H_i + \left[\sum_{l=1}^{i} \frac{(1-\nu^o_k)^l}{l} \right] \right) \right).$$

References

[1] Yee Whye Teh, Dilan Görür, and Zoubin Ghahramani. Stick-breaking construction for the indian buffet process. In *Proceedings of the Eleventh International Conference on Artifical Intelligence and Statistics (AISTAT 2007).*, 2007.