Supplementary Material: Infinite Factorial Unbounded-State Hidden Markov Model

Isabel Valera, Francisco J. R. Ruiz and Fernando Perez-Cruz

APPENDIX A CULINARY METAPHOR

Following a similar procedure as in [1], we can derive the expression for $\lim_{M\to\infty} p([\mathbf{S}]|Q,\alpha,\beta_0,\beta,\gamma)$ given in the Section 2.2 of the main text from a culinary metaphor analogous to the IBP. In this process, T customers enter sequentially a restaurant with an infinitely long buffet of dishes. The first customer starts at the left of the buffet and takes a serving from each dish, taking (possibly different) quantities for each one and stopping after a Poisson(α) number of dishes. The number of units $q \in \{1,\ldots,Q-1\}$ she takes is independently sampled for each dish from a uniform distribution.

The t-th customer enters the restaurant and starts at the left of the buffet. At dish m, she looks at the customer in front of her to see how many units she has taken from that dish and proceeds as follows:

- If the $(t-1)$-th customer did not take the m-th dish, she serves herself that dish with probability $\frac{\sum_{i=1}^{Q-1} n_{0i}^m}{n_{0\bullet}^m+1}$, where n_{0i}^m is the number of previous customers who took i units from dish m when the person in front of them did not take the dish m. If she does, the number of units she takes is given by i with probability $\frac{\gamma+n_{0i}^m}{\sum_{j=1}^{Q-1}(\gamma+n_{0j}^m)}$ $(i=1,\ldots,Q-1)$.
- If the $(t-1)$-th customer took q units from the m-th dish, the t-th customer either serves herself i units with probability $\frac{\beta+n_{qi}^m}{\beta_0+(Q-1)\beta+\sum_{j=0}^{Q-1}(n_{qj}^m)}$ $(i=1,\ldots,Q-1)$, or she does not take that dish with probability $\frac{\beta_0+n_{q0}^m}{\beta_0+(Q-1)\beta+\sum_{j=0}^{Q-1}(n_{qj}^m)}$, where n_{qi}^m is the number of previous customers who took i units from dish m when the person in front of them took q units.

The t-th customer then moves on to the next dish and repeats the above procedure. After having passed all dishes people have previously served themselves from, she takes independent quantities $q \sim \text{Uniform}\left(\frac{1}{Q-1},\ldots,\frac{1}{Q-1}\right)$ from a Poisson($\frac{\alpha}{t}$) number of new dishes.

If we fill in the entries of the $T\times M$ matrix $\mathbf{S}$ with the number of units that every customer took from every dish, and we denote with $M_1^{(t)}$ the number of new dishes tried by the t-th customer, the probability of any particular matrix $\mathbf{S}$ being produced by this process is given in Eq. 1.

It is straightforward to check that there are $\frac{(Q-1)!}{(Q-N_Q)!N_f}\prod_{t=1}^{T} M_1^{(t)}!/\prod_{h=1}^{Q^T-1} M_h!$ matrices in the same equivalence class as $\mathbf{S}$, and therefore we can recover $\lim_{M\to\infty} p([\mathbf{S}]|Q,\alpha,\beta_0,\beta,\gamma)$ by summing over all possible matrices lying in the set $[\mathbf{S}]$ generated by this process.

APPENDIX B ASSIGNMENT PROBABILITIES FOR THE GIBBS SAMPLER

We now derive the probability $p(s_{tm}=k|\mathbf{S}_{\neg tm})$, needed in Section 4.1 of the main text. This expression can be expressed, up to a proportionality constant, as shown in Eq. 2. Let $n_{qi}^{\neg tm}$ be the number of transitions from state q to state i in chain m, excluding the transitions from state $s_{(t-1)m}$ to s_{tm} and from state s_{tm} to $s_{(t+1)m}$. Similarly, let $n_{q\bullet}^{\neg tm}$ be the total number of transitions from state q in chain m without taking into account state s_{tm}, namely, $n_{q\bullet}^{\neg tm}=\sum_{i=0}^{Q-1} n_{qi}^{\neg tm}$. The expression in (2) takes different forms depending on the values of $j=s_{(t-1)m}$ and $\ell=s_{(t+1)m}$, yielding Eq. 3 for $j=0$ and Eq. 4 for $j\neq 0$.

- *The authors are with the Department of Signal Processing and Communications, University Carlos III in Madrid, Spain. E-mail: {ivalera, franrruiz, fernando}@tsc.uc3m.es*
- *Fernando Perez-Cruz is also member of the Technical Staff at Bell Labs (Alcatel-Lucent, NJ). E-mail: Fernando.Perez-Cruz@Alcatel-Lucent.com*

APPENDIX C
UPDATE EQUATIONS FOR THE VARIATIONAL ALGORITHM

The variational inference algorithm involves optimizing the variational parameters of $q(\Psi)$ to minimize the Kullback-Leibler divergence of $p_M(\Psi|\mathbf{X},\mathcal{H})$ from $q(\Psi)$, i.e., $D_{KL}(q||p_M)$. This optimization can be performed by iteratively applying the following fixed-point set of equations:

$$P_{jk}^m = \begin{cases} \exp\left\{\psi(\tau_{jk}^m) - \psi\left(\sum_{i=0}^{Q-1}\tau_{ji}^m\right)\right\}, \text{ if } j \neq 0 \\ \exp\left\{\psi(\nu_1^m) - \psi(\nu_1^m + \nu_2^m)\right\}, \text{ if } j = 0, k = 0 \\ \exp\left\{\psi(\nu_2^m) - \psi(\nu_1^m + \nu_2^m) + \psi(\varepsilon_k^m) \right. \\ \qquad \left. -\psi\left(\sum_{i=1}^{Q-1}\varepsilon_i^m\right)\right\}, \text{ if } j = 0, k \neq 0 \end{cases}, \tag{5}$$

$$b_{kt}^m = \exp\left\{ -\frac{1}{2\sigma_x^2}(\mathbf{L}_k)_m(\mathbf{L}_k)_m^\top + \frac{1}{\sigma_x^2}(\mathbf{L}_k)_m\left(\mathbf{x}_t - \sum_{\ell\neq m}\sum_{i=1}^{Q-1}(\mathbf{L}_i)_\ell q(s_{t\ell}=i)\right)^\top\right\}, \tag{6}$$

$$\tau_{jk}^m = \delta_{k0}\beta_0 + (1-\delta_{k0})\beta + \sum_{t=1}^{T} q(s_{(t-1)m}=j, s_{tm}=k), \tag{7}$$

$$\nu_1^m = 1 + \sum_{t=1}^{T} q(s_{(t-1)m}=0, s_{tm}=0), \tag{8}$$

$$\nu_2^m = \frac{\alpha}{M} + \sum_{t=1}^{T} q(s_{(t-1)m}=0, s_{tm}>0), \tag{9}$$

$$\varepsilon_k^m = \gamma + \sum_{t=1}^{T} q(s_{(t-1)m}=0, s_{tm}=k), \tag{10}$$

$$\mathbf{\Omega}_k = \left(\frac{1}{\sigma_0^2} + \frac{M}{\sigma_\phi^2}\right)^{-1}\mathbf{I}_D, \tag{11}$$

$$\boldsymbol{\omega}_k = \left(\frac{1}{\sigma_\phi^2}\mathbf{1}_M^\top\mathbf{L}_k + \frac{1}{\sigma_0^2}\boldsymbol{\mu}_0\right)\mathbf{\Omega}_k, \tag{12}$$

$$\mathbf{\Lambda}_k = \left(\frac{1}{\sigma_\phi^2}\mathbf{I}_M + \frac{1}{\sigma_x^2}\mathbf{C}_k\right)^{-1}, \text{ and} \tag{13}$$

$$\mathbf{L}_k = \mathbf{\Lambda}_k\left(\frac{1}{\sigma_\phi^2}\mathbf{1}_M\boldsymbol{\omega}_k + \frac{1}{\sigma_x^2}\mathbf{Q}_k^\top\left(\mathbf{X} - \sum_{j\neq k}\mathbf{Q}_j\mathbf{L}_j\right)\right), \tag{14}$$

where $(\mathbf{L}_k)_m$ denotes the m-th row of matrix $\mathbf{L}_k$, $\psi(\cdot)$ stands for the digamma function [2, p. 258–259], $\delta_{ii'}$ denotes the Kronecker delta function (which takes value one if $i = i'$ and zero otherwise), and the elements of the $T \times M$ matrices $\mathbf{Q}_k$ and $M \times M$ matrices $\mathbf{C}_k$ are, respectively, given by

$$(\mathbf{Q}_k)_{tm} = q(s_{tm}=k) \tag{15}$$

and

$$(\mathbf{C}_k)_{mm'} = \begin{cases} \sum_{t=1}^{T} q(s_{tm}=k)q(s_{tm'}=k), & \text{if } m \neq m' \\ \sum_{t=1}^{T} q(s_{tm}=k), & \text{if } m = m'. \end{cases} \tag{16}$$

The probabilities $q(s_{tm})$ and $q(s_{tm}, s_{(t-1)m})$ can be obtained through a standard forward-backward algorithm for HMMs within each chain, in which the variational parameters P_{jk}^m and b_{kt}^m play respectively the role of the transition probabilities and the observation probability associated with state variable s_{tm} taking value k in the Markov chain m [3].

REFERENCES

[1] J. Van Gael, Y. W. Teh, and Z. Ghahramani, "The infinite factorial hidden Markov model," in *Advances in Neural Information Processing Systems (NIPS)*, vol. 21, 2009.
[2] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. New York: Dover Publications, 1972.
[3] Z. Ghahramani and M. I. Jordan, "Factorial hidden Markov models," *Machine Learning*, vol. 29, no. 2-3, pp. 245–273, 1997.

$$p(\mathbf{S}|Q,\alpha,\beta_0,\beta,\gamma)=\frac{\alpha^{M_+}}{\prod_{t=1}^{T} M_1^{(t)}!}e^{-\alpha H_T} \times \prod_{m=1}^{M_+}\left[\frac{\Gamma(n_{00}^m+1)\Gamma\left(\sum_{i=1}^{Q-1}n_{0i}^m\right)}{\Gamma(n_{0\bullet}^m+1)}\frac{\Gamma\left((Q-1)\gamma\right)\prod_{i=1}^{Q-1}\Gamma(n_{0i}^m+\gamma)}{\Gamma\left(\sum_{i=1}^{Q-1}(n_{0i}^m+\gamma)\right)\left(\Gamma(\gamma)\right)^{Q-1}}\prod_{q=1}^{Q-1}\left(\frac{\Gamma\left(\beta_0+(Q-1)\beta\right)}{\Gamma(\beta_0)\left(\Gamma(\beta)\right)^{Q-1}}\frac{\Gamma(n_{q0}^m+\beta_0)\prod_{i=1}^{Q-1}\Gamma(n_{qi}^m+\beta)}{\Gamma\left(n_{q\bullet}^m+\beta_0+(Q-1)\beta\right)}\right)\right]. \tag{1}$$

$$p(s_{tm}=k|\mathbf{S}_{\neg tm})\propto\begin{cases}\int_{\mathbf{a}_j^m}p(s_{tm}=k|\mathbf{a}_j^m)p(\mathbf{a}_j^m|\{s_{\tau m}|s_{(\tau-1)m}=j,\tau\neq t,t+1\})d\mathbf{a}_j^m\times \\ \quad\times\int_{\mathbf{a}_k^m}p(s_{(t+1)m}=\ell|\mathbf{a}_k^m)p(\mathbf{a}_k^m|\{s_{\tau m}|s_{(\tau-1)m}=k,\tau\neq t,t+1\})d\mathbf{a}_k^m, & \text{if } k\neq j\\ \int_{\mathbf{a}_k^m}p(s_{(t+1)m}=\ell,s_{tm}=k|\mathbf{a}_k^m)p(\mathbf{a}_k^m|\{s_{\tau m}|s_{(\tau-1)m}=k,\tau\neq t,t+1\})d\mathbf{a}_k^m, & \text{if } k=j.\end{cases} \tag{2}$$

a) If $j=0$:

$$p(s_{tm}=k|\mathbf{S}_{\neg tm})\propto\begin{cases}\frac{n_{00}^{\neg tm}+1}{(n_{0\bullet}^{\neg tm}+1)(n_{0\bullet}^{\neg tm}+2)}\left(\delta_{\ell 0}(n_{00}^{\neg tm}+2)+(1-\delta_{\ell 0})\frac{(\gamma+n_{0\ell}^{\neg tm})\sum_{i=1}^{Q-1}n_{0i}^{\neg tm}}{\sum_{i=1}^{Q-1}(\gamma+n_{0i}^{\neg tm})}\right), & \text{if } k=0\\ \frac{(\delta_{\ell 0}\beta_0+(1-\delta_{\ell 0})\beta+n_{k\ell}^{\neg tm})\,(\gamma+n_{0k}^{\neg tm})\left(\sum_{i=1}^{Q-1}n_{0i}^{\neg tm}\right)}{(1+n_{0\bullet}^{\neg tm})\,(\beta_0+(Q-1)\beta+n_{k\bullet}^{\neg tm})\left(\sum_{i=1}^{Q-1}(\gamma+n_{0i}^{\neg tm})\right)}, & \text{if } k=1,\ldots,Q-1.\end{cases} \tag{3}$$

b) If $j\neq 0$:

$$p(s_{tm}=k|\mathbf{S}_{\neg tm})\propto\begin{cases}\frac{(\beta_0+n_{j0}^{\neg tm})}{(n_{0\bullet}^{\neg tm}+1)\left(\beta_0+(Q-1)\beta+n_{j\bullet}^{\neg tm}\right)}\left(\delta_{\ell 0}(n_{00}^{\neg tm}+1)+(1-\delta_{\ell 0})\frac{(\gamma+n_{0\ell}^{\neg tm})\sum_{i=1}^{Q-1}n_{0i}^{\neg tm}}{\sum_{i=1}^{Q-1}(\gamma+n_{0i}^{\neg tm})}\right), & \text{if } k=0\\ \frac{(\delta_{\ell 0}\beta_0+(1-\delta_{\ell 0})\beta+n_{k\ell}^{\neg tm}+\delta_{k\ell}\delta_{kj})\,(\beta+n_{jk}^{\neg tm})}{(\beta_0+(Q-1)\beta+n_{k\bullet}^{\neg tm}+\delta_{kj})\left(\beta_0+(Q-1)\beta+n_{j\bullet}^{\neg tm}\right)}, & \text{if } k=1,\ldots,Q-1.\end{cases} \tag{4}$$