Appendix: A Layered Dirichlet Process for Hierarchical Segmentation of Sequential Grouped Data

Adway Mitra[1], Ranganath B.N.[1], and Indrajit Bhattacharya[2]

[1] CSA Department, Indian Institute of Science, Bangalore, India,
{adway,ranganath}@csa.iisc.ernet.in,
[2] IBM India Research Lab, Bangalore, India,
indrajitb@gmail.com

1 Block Exchangeability

In this section, we provide proofs of the three theorems on Block Exchangeability which are stated without proofs in the paper.

Theorem 1 *If a model defining a joint distribution P is Completely Exchangeable then it is necessarily Block Exchangeable, but not the converse.*

Proof. Consider any assignment $\bar{z}$ of sequence z. Every time a particular value i is taken, the next value is either i or something other than i or it is the last entry, and so $n_i = n_{i,i} + n_{i,-i} + \delta(i, e)$, where n_i is the number of times the i-th value occurred in the sequence. So, the Exchangeability Statistic $S_C(z) = \{n_i\}$ can be uniquely determined from $S_B(z) = (\{n_{i,i}, n_{i,-i}\}, e)$.

Consider two assignments $\bar{z}_1$ and $\bar{z}_2$ of sequence z. From the above we know that $S_B(z_1) = S_B(z_2)$ implies $S_C(z_1) = S_C(z_1)$. But, since P is Completely Exchangeable by assumption, $S_C(z_1) = S_C(z_2)$ implies $P(z_1) = P(z_2)$. Combining, $S_B(z_1) = S_B(z_1)$ implies $P(z_1) = P(z_2)$. Hence, if a model is Completely Exchangeable then it is necessarily Block Exchangeable. But the entries in $S_B(z)$ cannot be obtained from $S_C(z)$ because any entry n_i cannot be uniquely split in to $n_{i,i}$ and $n_{i,-i}$. So, even though $S_C(z_1) = S_C(z_2)$, $S_B(z_1)$ need not be equal to $S_B(z_1)$. Hence, if a model is BE then it need not be CE.

Theorem 2 *If a model defining a joint distribution P is Block Exchangeable then it is necessarily Markov Exchangeable, but not the converse.*

Proof. Consider any assignment $\bar{z}$ of sequence z. The Exchangeability Statistic $S_B(z) = (\{n_{i,i}, n_{i,-i}\}, e)$ can be uniquely determined from $S_M(z) = (\{n_{ij}\}, s)$, since $n_{i,i} = n_{ii}$ and $n_{i,-i} = \sum_{j \neq i} n_{ij}$.

Consider two assignments x and y of sequence X. From the above we know that $S_M(x) = S_M(y)$ implies $S_B(x) = S_B(y)$. But, since P is Block Exchangeable by assumption, $S_B(x) = S_B(y)$ implies $P(x) = P(y)$. Combining, $S_M(x) = S_M(y)$ implies $P(x) = P(y)$. Hence, if a model is Block Exchangeable then it is necessarily Markov Exchangeable. But the entries in $S_M(z)$ cannot be obtained *from $S_B(z)$ because any entry $n_{i,-i}$ cannot be uniquely split in to $n_{i,j}$, for all $j \neq i$. So, even though $S_B(x) = S_B(y)$, $S_M(x)$ need not be equal to $S_M(y)$. Hence, if a model is ME then it need not be BE.*

Theorem 3 *The BE-DP prior distribution as well as the corresponding mixture model satisfy Block Exchangeability.*

Proof. *Consider the generative process in Eqn. 6. Given π and q, the joint probability of a sequence $\{z_i, g_i\}$ without the observations w_i can be factorized as*

$$p(z_{j_1}, z_{j_2}, \ldots, z_{j_n}|\pi, q) = p(z_{j_1}|\pi) \prod_{i=2}^{n} p(z_{j_i}|z_{j_{i-1}}, \pi, q)$$

where $\{j_1, j_2, \ldots, j_n\} = \{i : g_i = g\}$, for some unique value g. Now clearly, $p(z_{j_1}) = \prod_{k=1}^{K} \pi^{\delta(k, z_{j_1})}$ and

$$p(z_{j_i}|z_{j_{i-1}}) = \prod_{k=1}^{K} (\hat{\pi}_{gk}(k))^{\delta(k, z_{j_{i-1}})\delta(z_{j_i}, z_{j_{i-1}})} {q_{gk}}^{\delta(k, z_{j_{i-1}})(1-\delta(z_{j_i}, z_{j_{i-1}}))} \times \prod_{k'=1}^{K} \pi_g(k')^{\delta(k', z_{j_i})(1-\delta(z_{j_i}, z_{j_{i-1}}))}$$

Taking the product over i, we finally get the full probability as

$$p(z_{j_1}, \ldots, z_{j_n}) = \prod_{k=1}^{K} (\hat{\pi}_{gk}(k))^{\sum_{i=2}^{n} \delta(k, z_{j_{i-1}})\delta(z_{j_i}, z_{j_{i-1}})} \times \prod_{k=1}^{K} {q_{gk}}^{\sum_{i=2}^{n} \delta(k, z_{j_{i-1}})(1-\delta(z_{j_i}, z_{j_{i-1}}))} \times \prod_{k=1}^{K} \pi_g(k)^{\delta(k, z_{j_1}) + \sum_{i=2}^{n} \delta(k, z_{j_i})(1-\delta(z_{j_i}, z_{j_{i-1}}))}$$

The first term corresponds to the case that $z_{j_i} = z_{j_{i-1}}$, and the next two terms are for the case where they are not equal.

Now, the term $\sum_{i=2}^{n} \delta(k, z_{j_{i-1}})\delta(z_{j_i}, z_{j_{i-1}})$ equals the number of times the value k has been retained from the previous step, i.e. the number of self transitions from value k, which is denoted $n_{k,k}$.

Next, the term $\sum_{i=2}^{n} \delta(k, z_{j_{i-1}})(1 - \delta(z_{j_i}, z_{j_{i-1}}))$ is the number of transitions made from value k to some other value, which is denoted by $n_{k,-k}$

Finally, the term $\delta(k, z_{j_1}) + \sum_{i=2}^{n} \delta(k, z_{j_i})(1 - \delta(z_{j_i}, z_{j_{i-1}}))$ is the number of transitions made to k from some other value or the start. But a transition made to value k has a corresponding transition out of value k, unless k is the end-value e. If $k = e$, the above term is $n_{k,-k} + 1$. In other words, the above term is $n_{k,-k} + \delta(k, e)$.

So the probability of the given sequence is

$$p(z_{j_1}, \ldots, z_{j_n}) = \prod_{k=1}^{K} (\hat{\pi}_{gk}(k))^{n_{k,k}} {q_{gk}}^{n_{k,-k}} \pi_g(k)^{n_{k,-k} + \delta(k,e)}$$

Clearly, the probability of a sequence depends only on $(\{n_{k,k}, n_{k,-k}\}_k, e)$, which proves the model is Block-Exchangeable.

2 Mixed Exchangeability Model

In the paper we have defined generative processes satisfying CE,ME and BE for any particular layer. A particularly attractive property of the LaDP framework is that, it allows the use of any exchangeability property for any layer. To illustrate the generative process of such a mixed-exchangeability LaDP model, we describe in detail the generative process for a model with $L = 2$, where we have ME at $l = 2$ and BE at $l = 1$. We also have layer $l = 0$ to generate the observations, which uses CE. According to the nomenclature of Section 6 in the paper, this is called ME-BE-CE-LaDP.

$$\beta^2_{g_2} \sim GEM(\gamma^2);\ \pi^2_{g_2,g'_2} \sim DP(\alpha^2, \beta^2_{g_2}),\ g_2, g'_2 = 1\ldots\infty$$
$$\beta^1_{g_1} \sim GEM(\gamma^1);\ \pi^1_{g_1} \sim DP(\alpha^1, \beta^1_{g_1}),\ g_1 = 1\ldots\infty$$
$$q^1_{g_1,g'_1} \sim Beta(1,\kappa);\ \pi^1_{g_1,g'_1} = q^1_{g_1,g'_1}\pi^1_{g_1} + (1 - q^1_{g_1,g'_!})\delta_{g'_1},\ g_1, g'_1 = 1\ldots\infty$$
$$\beta^0_{g_0} \sim GEM(\gamma^0);\ \pi^0_{g_0} \sim DP(\alpha^0, \beta^0_g); g_0 = 1\ldots\infty$$

$$z^2_i \sim \pi^2_{z^3_i,p(i,2)},\ i = 1\ldots n$$
$$z^1_i \sim \pi^1_{z^2_i,p(i,1)},\ i = 1\ldots n$$
$$z^0_i \sim \pi^0_{z^1_i}; w_i = \mathcal{W}_{z^0_i},\ i = 1\ldots n$$

For the first datapoint of each group at layer $(l + 1)$, where $p(i,l)$ is undefined, we sample $z^l_i \sim \beta^l_{z^{l+1}_i}$, for $l = \{2, 1\}$.

z^3_i and w_i are observed for all $i \in [1, n]$.