"A Bayesian Semi-parametric Approach for the Differential Analysis of Sequence Counts Data" by Guindani et. *al.* – Supporting Information

Appendix I –Full conditionals for the one library case

We start outlining the relevant full conditionals for Markov chain Monte Carlo sampling in the one library case. For a given k, consider model (2) completed with $G^*(\lambda) = Ga(\alpha, \beta)$. We can integrate out the λ_i's and rewrite the likelihood as a function only of the cluster configurations $\mathbf{s}$, i.e. the components' assignments in the mixture. Thus,

$$p(\mathbf{y}|\mathbf{s}, k) = \frac{1}{\prod_{i=1}^{k} y_i!} \frac{\beta^{\alpha L}}{\Gamma(\alpha)^L} \prod_{j=1}^{L} \frac{\Gamma(\alpha + \bar{y}_j)}{(\beta + n_j)^{(\alpha+\bar{y}_j)}},$$

where $L = \max\{s_i, i = 1, \ldots, k\}$, $n_j = \sum_{i=1}^{k} I(s_i = j)$ and $\bar{y}_j = \sum_{i=1}^{k} I(s_i = j)\, y_i$ are, respectively, the number of clusters, the cluster frequencies and the total counts in cluster j at an arbitrary iteration of the posterior simulation. Therefore, posterior inference can be obtained by sampling the configuration indicators $\mathbf{s}$ and the unknown number of tags k in this reduced model.

More precisely, the full conditional for the configuration indicators $\mathbf{s}$ is Multinomial with probabilities

$$p(s_i = l|\mathbf{s}_{-i}, k, \mathbf{y}) \propto p(\mathbf{s}|k) \times p(\mathbf{y}|\mathbf{s}, k), \quad l = 1, \ldots, L^{-i} + 1,$$

where $\mathbf{s}_{-i} = (s_1, \ldots, s_{i-1}, s_{i+1}, \ldots, s_k)^T$, $L^{-i} = \max\{\mathbf{s}_{-i}\}$, and $p(\mathbf{s}|k)$ is as in (4) with $L = L^{-i}$ if $s_i = l$, $l = 1, \ldots, L^{-i}$ or $L = L^{-i} + 1$ if s_i is a newly sampled label (see MacEachern and Müller, 1998).

Inference on the number of distinct sequences k is obtained by means of a Metropolis-within-Gibbs step. At each iteration, we propose to increase (or decrease) the current value of k by a fixed amount with probability p (or $1 - p$). For example, in the application presented in section 3.4, we set $p = 1/2$. Accordingly, we add (or delete) a set of zero counts and corresponding labels s_i's to the current data. To be more specific, suppose we propose to move from the current k to $\tilde{k} = k + 1$. If the move is accepted, the data set has to be augmented to accommodate for the new tag with $y_{k+1} = 0$ and $s_{k+1} = \tilde{s}_{k+1}$. The value of $\tilde{s}_{k+1}$ is proposed by means of a single draw from the Pólya Urn. Let $\tilde{\mathbf{s}} = (s_1, \ldots, s_k, \tilde{s}_{k+1})^T$ and $L = \max\{s_i, i = 1, \ldots, k\}$. Then, it can be shown that $\tilde{s}_{k+1} \sim \text{Multin}(1; q_1, \ldots, q_{L+1})$, where

$$q_j \propto Prob(y = 0|\mathbf{s}, \mathbf{y}) = \left(\frac{\beta + n_j}{\beta + n_j + 1}\right)^{(\alpha+\bar{y}_j)}, \quad j = 1, \ldots, L$$

and

$$q_{L+1} \propto Prob(y = 0) = \left(\frac{\beta}{\beta + 1}\right)^{\alpha}.$$

In order to obtain the proper acceptance rate, we need to evaluate the probability of the reverse move, from $\tilde{k}$ to k. The move corresponds to the deletion of one of the zero counts previously added; hence, we sample the proposed deletion from a discrete uniform distribution on $\{k' + 1, \ldots, \tilde{k}\}$. Thus, the Metropolis-Hasting ratio for the upward move is given by

$$A = \frac{p(k+1)}{p(k)} \times \frac{p(\tilde{\mathbf{s}}|\tilde{k})}{p(\mathbf{s}|k)} \times \frac{p(\tilde{\mathbf{y}}|\tilde{\mathbf{s}}, \tilde{k})}{p(\mathbf{y}|\mathbf{s}, k)} \times \frac{1-p}{p} \times \frac{\frac{1}{\tilde{k}-k'}}{q_{\tilde{s}_{\tilde{k}}}},$$

and the move is accepted with probability $a = \min(1, A)$. Instead, the move from k to $k-1$ is accepted with probability $a = \min(1, A^{-1})$. In order to allow for the exploration of a large posterior support space, this step can be repeated multiple times in a single iteration. Alternatively, the previous steps can be easily modified to take into account a generic step ($m > 0$) up or down from the current state k.

Given an imputed cluster structure $\mathbf{s}$, it is always possible to sample from the posterior of the cluster-specific abundances λ_j^*, $j = 1, \ldots, L$ (hence, the tag specific λ_i). As a matter of fact, $\lambda_j^*|k, \mathbf{s}, \mathbf{y} \sim Ga(\alpha + \bar{y}_j, n_j + \beta)$, where $n_j = \sum_{i=1}^{k} I(s_i = j)$ is the cluster frequency and $\bar{y}_j = \sum_{i=1}^{k} y_i\, I(s_i = j)$ represents the cluster mass, $j = 1, \ldots, L$.

Appendix II –Full conditionals for the class comparison.

We follow the discussion in Section 3 and denote with, $x_t = 1, \ldots, C$, the tissue collected in sample t. Then, $y_{ix} = \sum_{t:x_t=x} y_{it}$ is the observed count of sequence i under condition x and m_x is the number of samples drawn under condition x, $x = 1, \ldots, C$. Again, we can integrate out the random probability measure G and the parameters of the base measure and consider only the cluster configuration indicators s_{ix} such that $s_{ix} = j$ iff $\lambda_{ix} = \lambda_j^*$. Then, the marginal likelihood is

$$p(\mathbf{y}|\mathbf{s}, k) = h(\mathbf{y})\, \frac{\beta^{\alpha L}}{\Gamma(\alpha)^L} \prod_{j=1}^{L} \frac{\Gamma(\alpha + \tilde{y}_j)}{(\beta + M_j)^{\alpha + \tilde{y}_j}}, \tag{9}$$

where $h(\mathbf{y}) = 1/(\prod_{i=1}^{K} \prod_{x=1}^{C} y_{ix}!)$, $\tilde{y}_j = \sum_{i=1}^{k} \sum_{x=1}^{C} y_{ix}\, I(s_{ix} = j)$ is the sum of counts in cluster j, $M_j = \sum_{x=1}^{C} N_{j,x}\, m_x$, with $N_{j,x} = \sum_{i=1}^{k} I(s_{ix} = j)$, is a measure of the cluster size, and L denotes the number of clusters.

In order to update the cluster configurations $\{s_{ix}\}$, it is convenient to rewrite (8) explicitly in terms of the latent indicators of differential expression, w_{ix}. We need to introduce further notation. Let $x > 1$, since $x = 1$ denotes the reference condition. Also, let $\mathbf{w}_{-i,x}$ denote the vector of w's with the exclusion of the single element w_{ix}; analogously, define $\mathbf{s}_{-i,x}$. Recall that if $w_{ix} = 0$, then $s_{ix} = s_{i1}$. Hence, the full conditional $p(s_{ix}, x > 1 | \, w_{ix} = 0, \mathbf{w}_{-i,x}, \mathbf{s}_{-i,x}, \mathbf{y})$ is a point mass at s_{i1}. Instead, conditional on $w_{ix} = 1$,the distribution of s_{ix} is Multinomial with probabilities

$$p(s_{ix} = j \mid w_{ix} = 1, w_{-i,x}, \mathbf{s}_{-i,x}, \mathbf{y}, k) \propto \begin{cases} \frac{\nu}{\nu+W} \, p(\mathbf{y} \mid \mathbf{s}, k) & \text{if } j = L^{-i,x} + 1 \\ \frac{W_j}{\nu+W} \, p(\mathbf{y} \mid \mathbf{s}, k) & \text{if } j = 1, \ldots, L^{-i,x}, \end{cases} \tag{10}$$

where $W = \sum_{i=1}^{k} \sum_{x=2}^{C} I(w_{ix} = 1)$ is the overall frequency of differentially abundant sequences and $W_j = \sum_{i=1}^{k} \sum_{x=2}^{C} I(s_{ix} = j) \, I(w_{ix} = 1)$ is the cluster specific frequency ($W = \sum_j W_j$), $j = 1, \ldots, L$, at any given iteration. In addition, $L^{-i,x} = \max\{s_{r,z}, (r,z) \neq (i,x), r = 1, \ldots, k, z = 1, \ldots, C\}$ and

$$\begin{aligned} p(\mathbf{y} \mid \mathbf{s}, k) &\propto p(y_{r,t}, (r, x(t)) = (i,x) \mid y_{r,t}, (r, x(t)) \neq (i,x), \mathbf{s}) \\ &= \frac{\Gamma(\alpha + \tilde{y}_j)}{(\beta + M_j)^{(\alpha + \tilde{y}_j)}} \frac{(\beta + M_j^-)^{(\alpha + \tilde{y}_j^-)}}{\Gamma(\alpha + \tilde{y}_j^-)}, \quad j = 1, \ldots, L, \end{aligned} \tag{11}$$

where the quantities $M_j^- = M_j - N_{j,x}^- \, m_x$ with $N_{j,x}^- = \sum_{r \neq i} I(s_{r,x} = j)$, and $\tilde{y}_j^- = \sum_{(r,z) \neq (i,x)} y_{r,z} \, I(s_{r,z} = j)$ denote, respectively, measures of the size and the total counts of each cluster, with the ith observations in condition x excluded. Since the algorithm just described relies on repeated draws of the single elements of each vector $\mathbf{s}_x$ and $\mathbf{w}_x$, it is not efficient when applied to large datasets. We can improve the mixing of the chain and decrease computation time by employing a merge-split move such as the one devised by the SAMS sampler by Dahl (2003). This algorithm was used in the data example presented in section 3.4. Conditional on the observations $y_{i,x}$ with $w_{i,x} = 1$, the algorithm can be described as follows:

1. At any given iteration, uniformly select a pair of distinct observations, say i and j.

2. (a) If i and j belong to the same cluster, say S, then propose a new cluster configuration by splitting the common cluster as follows:

   – Start by forming singleton sets, say $S_i = \{i\}$ and $S_j = \{j\}$;

- Consider a uniformly selected permutation of the remaining elements in S;
- Any remaining element l in S is added to either S_i or S_j with probabilities

$$p(l \in S_i | \ldots) = \frac{W_{S_i}\, p(\mathbf{y}|\mathbf{s}_{(i)}, k)}{W_{S_i}\, p(\mathbf{y}|\mathbf{s}_{(i)}, k) + W_{S_j}\, p(\mathbf{y}|\mathbf{s}_{(j)}, k)}, \qquad (12)$$

  where W_{S_r} is the cardinality of cluster S_r, $r \in \{i, j\}$, $p(\mathbf{y}|\mathbf{s}', k)$ is as in (11) and $\mathbf{s}_{(r)}$ is the vector of cluster configurations obtained assuming $s_l \in S_r$, $r \in \{i, j\}$.
- Let $\mathbf{s}^*$ denote the proposed partition. Accept $\mathbf{s}^*$ over the current partition $\mathbf{s}$ with probability

$$a_{\text{split}} = \min\left[1, \frac{p(\mathbf{s}^*|\mathbf{y})}{p(\mathbf{s}|\mathbf{y})} \frac{p(\mathbf{s}|\mathbf{s}^*)}{p(\mathbf{s}^*|\mathbf{s})}\right],$$

  where $p(\mathbf{s}|\mathbf{y})$ and $p(\mathbf{s}^*|\mathbf{y})$ are the partition posterior distributions evaluated, respectively, at $\mathbf{s}$ and $\mathbf{s}^*$, $p(\mathbf{s}^*|\mathbf{s})$ is the product of the probabilities in (12) and $p(\mathbf{s}|\mathbf{s}^*) = 1$.

2. (b) If i and j belong to two different clusters, propose to merge them in a new partition s^*. The Metropolis Hasting ratio for the proposed move is $a_{\text{merge}} = \min[1, \frac{1}{a_{\text{split}}}]$, which requires the computation of a product of probabilities (12) to take into account the reverse split of the merged partition s^* back into the current $\mathbf{s}$. We refer to Dahl (2003) for further details.

Finally, the full conditional for the indicators of differential abundance $\{w_{ix}, i > 1\}$ is obtained as follows. Note that if $s_{ix} \neq s_{i1}$ then $w_{ix} = 1$ with probability one. On the other hand, if $s_{ix} = s_{i1}$ then $w_{ix} = 1$ or $w_{ix} = 0$ with probability

$$p(w_{ix} \mid s_{i1} = s_{ix}, \mathbf{s}_{-i,x}, \mathbf{w}_{-i,x}, \mathbf{y}) \propto p(s_{ix} \mid w_{ix}, s_{i1} = s_{ix}, w_{-i,x}, \mathbf{y}) \times p(w_{ix}),$$

where the conditional distribution on the right side is given by (10). Alternatively, it would be possible to update the pairs (w_{ix}, s_{ix}) jointly, for $x > 1$. Finally, the update of k mimicks the one described in Appendix I with some minor adjustments to take into account the presence of multiple samples; hence, it is omitted.