SUPPLEMENT A TO "BAYESIAN COMPLEMENTARY CLUSTERING, MCMC AND ANGLO-SAXON PLACENAMES": ADDITIONAL CALCULATIONS AND DERIVATIONS

By Giacomo Zanella*

University of Warwick

We provide details of some calculations and derivations that were too long to be reported in Zanella (2014).

This supplementary material is divided in three sections. In Section 1 we provide the derivation of the likelihood function given in equation (3.2) of Zanella (2014). In Section 2 we derive equation (P4) of Zanella (2014) as an approximation of (P3) designed to require less computation at each Metropolis-Hastings step. Finally in Section 3 we prove that the MCMC algorithm described in Section 4.2 of Zanella (2014) is indeed targeting $\hat{\pi}(\rho)$, as claimed in that section. This fact follows from basic and well-known properties of the Euclidean square distance (presented in Lemma 1 of this supplementary material).

1. Derivation of likelihood function. Suppose that $x_1, ..., x_s$ are random vectors in $\mathbb{R}^2$ given by

$$x_l = z + y_l \ , \qquad l = 1, ..., s, \tag{1.1}$$

where z has probability density function $g(\cdot)$ on $\mathbb{R}^2$ and

$$y_l = w_l - \frac{1}{s}\sum_{i=1}^{s} w_i \ , \qquad l = 1, ..., s \ , \tag{1.2}$$

with $w_1, ..., w_s$ being s independent bivariate $N(0, \frac{\sigma^2}{\pi}\mathbb{I}_2)$ random vectors, where $\mathbb{I}_2$ is the 2×2 identity matrix. We need to prove that the probability density function (pdf) of $\mathbf{x} = (x_1, ..., x_s)$ on $\mathbb{R}^{2s}$ is

$$f_{(s,\sigma)}(x_1, ..., x_s) \quad = \quad \frac{g\left(\overline{x}\right)}{s(2\sigma^2)^{s-1}} \exp\left(-\frac{\pi\delta_{\mathbf{x}}^2}{2\sigma^2}\right), \tag{1.3}$$

*Supported by EPSRC through a PhD position under the CRiSM grant EP/D002060/1

where $\overline{x}$ is the Euclidean barycenter of $\mathbf{x}$ and $\delta_{\mathbf{x}}^2 = \sum_{i=1}^{s}(x_i - \overline{x})^2$. Expression (3.2) of Zanella (2014) can be obtained from multiplying (1.3) by the probability of obtaining a sequence of marks $(m_1, \dots, m_s) \in \{1, \dots, k\}^s$, that is $\prod_{i \neq j} \mathbb{1}(m_i \neq m_j) \cdot \frac{(k-s)!}{k!}$, and by a factor $s!$ arising because a given cluster $\{(x_1, m_1), \dots, (x_s, m_s)\}$ can be obtained in $s!$ different ways depending on the order of the points.

Let $y_i = \left(y_i^{(1)}, y_i^{(2)}\right)$ for i running from 1 to s. Note that the random vectors $\left(y_1^{(1)}, \dots, y_s^{(1)}\right)$ and $\left(y_1^{(2)}, \dots, y_s^{(2)}\right)$ are independent and identically distributed. Thus it suffices to consider $\left(y_1^{(1)}, \dots, y_s^{(1)}\right)$.

If we define $\mathbf{y} = \left(y_1^{(1)}, ..., y_s^{(1)}\right)^T$ and $\mathbf{w} = \left(w_1^{(1)}, ..., w_s^{(1)}\right)^T$ then (1.2) can be expressed as

$$\mathbf{y} = \mathbf{w} - \frac{1}{s}\mathbb{H}_s \mathbf{w},$$

where $\mathbb{H}_s$ is the $s \times s$ matrix with 1 in every position. Since the random vector $\mathbf{y}$ has zero mean then its covariance matrix Σ is

$$\Sigma = \mathbb{E}[\mathbf{y}^T\mathbf{y}] = \mathbb{E}[\mathbf{w}^T\mathbf{w}] - \frac{1}{s}\mathbb{E}[\mathbf{w}^T\mathbb{H}_s\mathbf{w}] - \frac{1}{s}\mathbb{E}[\mathbf{w}^T\mathbb{H}_s^T\mathbf{w}] + \frac{1}{s^2}\mathbb{E}[\mathbf{w}^T\mathbb{H}_s^T\mathbb{H}_s\mathbf{w}].$$

Then using the fact that $\mathbb{H}_s^T\mathbb{H}_s = s\mathbb{H}_s$ and $\mathbb{H}_s^T = \mathbb{H}_s$ we obtain

$$\Sigma = \frac{\sigma^2}{\pi}\mathbb{I}_s - \frac{2}{s}\mathbb{E}[\mathbf{w}^T\mathbb{H}_s\mathbf{w}] + \frac{1}{s}\mathbb{E}[\mathbf{w}^T\mathbb{H}_s\mathbf{w}] = \\ = \frac{\sigma^2}{\pi}\mathbb{I}_s - \frac{1}{s}\mathbb{E}[\mathbf{w}^T\mathbb{H}_s\mathbf{w}] = \frac{\sigma^2}{\pi}\left(\mathbb{I}_s - \frac{\mathbb{H}_s}{s}\right).$$

Note that $y_s^{(1)}$ equals $-\sum_{i=1}^{s-1} y_i^{(1)}$ because

$$\sum_{i=1}^{s} y_i^{(1)} = \sum_{i=1}^{s}\left(w_i^{(1)} - \frac{1}{s}\sum_{j=1}^{s} w_j^{(1)}\right) = \sum_{i=1}^{s} w_i^{(1)} - \sum_{j=1}^{s} w_j^{(1)} = 0.$$

Therefore we can focus on the distribution of $y_1^{(1)}, ..., y_{s-1}^{(1)}$ only. These form a Gaussian random vector $\mathbf{y}_{s-1} = (y_1^{(1)}, ..., y_{s-1}^{(1)})^T$ with zero mean and covariance matrix Σ_{s-1} which is the restriction of Σ to the first $s-1$ coordinates

$$\Sigma_{s-1} = \frac{\sigma^2}{\pi}\left(\mathbb{I}_{s-1} - \frac{\mathbb{H}_{s-1}}{s}\right). \tag{1.4}$$

Therefore the joint pdf of $\mathbf{y}_{s-1}$ in $\mathbb{R}^{s-1}$ is

$$(2\pi)^{-\frac{s-1}{2}}|\Sigma_{s-1}|^{-\frac{1}{2}}e^{-\frac{1}{2}(\mathbf{y}_{s-1})^T\Sigma_{s-1}^{-1}\mathbf{y}_{s-1}},$$

where $|\Sigma_{s-1}|$ denotes the determinant of Σ_{s-1}. Using the fact that $\mathbb{H}_{s-1}^2$ equals $(s-1)\mathbb{H}_{s-1}$ we can show that the inverse of Σ_{s-1} is $\frac{\pi}{\sigma^2}(\mathbb{I}_{s-1}+\mathbb{H}_{s-1})$. In fact

$$\frac{\sigma^2}{\pi}\left(\mathbb{I}_{s-1}-\frac{\mathbb{H}_{s-1}}{s}\right)\frac{\pi}{\sigma^2}\left(\mathbb{I}_{s-1}+\mathbb{H}_{s-1}\right)=\mathbb{I}_{s-1}+\mathbb{H}_{s-1}-\frac{\mathbb{H}_{s-1}}{s}-\frac{\mathbb{H}_{s-1}^2}{s}=$$
$$=\mathbb{I}_{s-1}+\frac{s-1}{n}\mathbb{H}_{s-1}-\frac{s-1}{n}\mathbb{H}_{s-1}=\mathbb{I}_{s-1}.$$

The determinant of Σ_{s-1} is $\frac{1}{s}\left(\sigma^2/\pi\right)^{s-1}$. This can be derived by the fact that the $s-1$ eigenvalues of $\frac{\pi}{\sigma^2}\Sigma_{s-1}=\mathbb{I}_{s-1}-\frac{\mathbb{H}_{s-1}}{s}$ are $\frac{1}{s},1,...,1$. An orthonormal basis of corresponding eigenvectors is given by the rows $\mathbf{r}_1,...,\mathbf{r}_{s-1}$ of an $(s-1)\times(s-1)$ Helmert matrix:

$$\begin{aligned}
\mathbf{r}_1 \quad &=(s-1)^{-1/2}\big(1,...,1\big),\\
\mathbf{r}_k \quad &=\big(k(k-1)\big)^{-1/2}\big(1,...,1,1-k,0,...,0\big) \qquad k=2,...,s-2,\\
\mathbf{r}_{s-1} \quad &=\big((s-1)(s-2)\big)^{-1/2}\big(1,...,1,1-(s-1)\big).
\end{aligned}$$

From $\mathbb{H}_{s-1}\mathbf{r}_1^T=(s-1)\mathbf{r}_1^T$ and $\mathbb{H}_{s-1}\mathbf{r}_k^T=0$ for k in $2,...,s-1$ it follows

$$\begin{aligned}
\left(\mathbb{I}_{s-1}-\frac{\mathbb{H}_{s-1}}{s}\right)\mathbf{r}_1^T&=\mathbf{r}_1^T-\frac{(s-1)}{s}\mathbf{r}_1^T=\frac{1}{s}\mathbf{r}_1^T,\\
\left(\mathbb{I}_{s-1}-\frac{\mathbb{H}_{s-1}}{s}\right)\mathbf{r}_k^T&=\mathbf{r}_k^T-0=\mathbf{r}_k^T \qquad k=2,...,s-1.
\end{aligned}$$

Therefore the joint pdf of $\mathbf{y}_{s-1}$ in $\mathbb{R}^{s-1}$ is

$$2^{-\frac{s-1}{2}}\frac{\sqrt{s}}{\sigma^{s-1}}\exp\left(-\frac{\pi}{2\sigma^2}(\mathbf{y}_{s-1})^T\big(\mathbb{I}_{s-1}+\mathbb{H}_{s-1}\big)\mathbf{y}_{s-1}\right). \tag{1.5}$$

Considering the exponent we have

$$-\frac{\pi}{2\sigma^2}(\mathbf{y}_{s-1})^T\big(\mathbb{I}_{s-1}+\mathbb{H}_{s-1}\big)\mathbf{y}_{s-1}=-\frac{\pi}{2\sigma^2}\left(\sum_{i=1}^{s-1}\left(y_i^{(1)}\right)^2+\sum_{i=1}^{s-1}\sum_{j=1}^{s-1}y_i^{(1)}y_j^{(1)}\right),$$

which equals

$$-\frac{\pi}{2\sigma^2}\left(\sum_{i=1}^{s-1}\left(y_i^{(1)}\right)^2+\left(\sum_{i=1}^{s-1}y_i^{(1)}\right)^2\right).$$

If we multiply together the joint pdfs of $\left(y_1^{(1)},\ldots,y_{s-1}^{(1)}\right)$ and $\left(y_1^{(2)},\ldots,y_{s-1}^{(2)}\right)$ we obtain the following expression for the pdf of the Gaussian family $y_1,...,y_{s-1}$ in $(\mathbb{R}^2)^{s-1}$, where $y_i=\left(y_i^{(1)},y_i^{(2)}\right)$

$$\frac{s}{(2\sigma^2)^{s-1}}\exp\left(-\frac{\pi}{2\sigma^2}\left(\sum_{i=1}^{s-1}y_i^2+\left(\sum_{i=1}^{s-1}y_i\right)^2\right)\right). \tag{1.6}$$

The pdf of $\mathbf{x}$ given in (1.3) can be obtained by linear transformation from the pdf of z and $y_1,...,y_{s-1}$. Equations (1.1) and (1.2) can be expressed as

$$x_i = z + y_i \qquad i=1,...,s-1,$$
$$x_s = z-\sum_{i=1}^{s-1}y_i,$$

or equivalently as

$$\mathbf{x}^{(j)} = z^{(j)}\begin{pmatrix}1\\1\\\vdots\\1\end{pmatrix}+\begin{pmatrix} & \mathbb{I}_{s-1} & \\ \hline -1 & \cdots & -1\end{pmatrix}\begin{pmatrix}y_1^{(j)}\\y_2^{(j)}\\\vdots\\y_{s-1}^{(j)}\end{pmatrix}, \qquad j=1,2,$$

where the j-th superscript denotes the j-th coordinate and $\mathbb{I}_{s-1}$ denotes the $(s-1)\times(s-1)$ identity matrix. Thus $\mathbf{x}^{(j)}$ is a linear transformation of the random vector $(z^{(j)},y_1^{(j)},...,y_{s-1}^{(j)})^T$ through the matrix

$$J_s=\left(\begin{array}{c|ccc}1 & & & \\ \vdots & & \mathbb{I}_{s-1} & \\ 1 & & & \\ \hline 1 & -1 & \cdots & -1\end{array}\right). \tag{1.7}$$

Therefore the pdf of $\mathbf{x}$ in $\mathbb{R}^{2s}$ is equal to the pdf of $(z,y_1,...,y_{s-1})^T$ divided by the squared determinant $|J_s|^2$. Using Laplace's formula on the last row

$$\begin{aligned}|J_s| &= (-1)^{s+1}|M_{s,1}|+\sum_{j=2}^{s}(-1)^{s+j}(-1)|M_{s,j}| = \\ &= (-1)^{s+1}\Big(|M_{s,1}|+\sum_{j=2}^{s}(-1)^j|M_{s,j}|\Big),\end{aligned} \tag{1.8}$$

where $M_{i,j}$ is the matrix obtained from J_s by removing the i-th row and the j-th column. Note that $M_{s,1}$ is the identity matrix so its determinant is 1. Moreover

$$M_{s,2} = \left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 1 & & & \\ \vdots & & \mathbb{I}_{s-2} & \\ 1 & & & \end{array} \right),$$

and so its determinant is 1 too. For $j = 3, ..., s$, note that $M_{s,j}$ can be obtained from $M_{s,j-1}$ by switching the $(j-2)$-th row and the $(j-1)$-th one. Therefore

$$|M_{s,j}| = (-1)|M_{s,j-1}| = (-1)^{j-2}|M_{s,2}| = (-1)^{j-2}.$$

Plugging this results in (1.8) it follows

$$|J_s| = (-1)^{s+1}\Big(1 + \sum_{j=2}^{s}(-1)^j(-1)^{j-2}\Big) = (-1)^{s+1}s, \tag{1.9}$$

and therefore $|J_s|^2 = s^2$. Multiplying together the pdf of z and the pdf of $y_1, ..., y_{s-1}$ obtained in (1.6), and dividing by the jacobian term we obtain (1.3).

2. Derivation of equation (P4) of Zanella (2014).

2.1. *Framework and notation.* Let G be the bipartite graph induced by a configuration of n_1 red points and n_2 blue points (see Section 3.6.1 of Zanella (2014) for more details). Let ρ be a (random) partial matching contained in G distributed according to $\hat{\pi}(\rho)$, where $\hat{\pi}(\rho)$ is the full conditional posterior distribution of the cluster partition ρ under the Poisson model given in equation (3.3) of Zanella (2014). Such posterior can be written as $\hat{\pi}(\rho) \propto \prod_{(i,j)\in\rho} w_{ij}$, where w_{ij} is the weight of the edge (i, j) defined in equation (3.7) of Zanella (2014).

Let $\rho \circ (i, j)$ be a matching proposed according to the Metropolis-Hastings (MH) algorithm described in Section 4.1 of Zanella (2014) with proposal given by (P3) and target measure $\hat{\pi}(\rho)$. Such state is obtained by choosing an edge (i, j) according to a probability measure

$$q_\rho(i, j) \;\propto\; \frac{\hat{\pi}(\rho \circ (i, j))}{\hat{\pi}(\rho) + \hat{\pi}(\rho \circ (i, j))} \;=\; \frac{\frac{\hat{\pi}(\rho\circ(i,j))}{\hat{\pi}(\rho)}}{1 + \frac{\hat{\pi}(\rho\circ(i,j))}{\hat{\pi}(\rho)}}\,, \tag{2.1}$$

and then evaluating the corresponding state $\rho \circ (i, j)$ according to display (4.1) of Zanella (2014).

2.2. *Derivation of* (P4). Note that in order to evaluate $\frac{\hat{\pi}(\rho\circ(i,j))}{\hat{\pi}(\rho)}$, and thus $q_\rho(i,j)$, it is not enough to know whether $(i,j)\in\rho$ or $(i,j)\notin\rho$. For example, if $\rho\circ(i,j)$ is a switch move then $\frac{\hat{\pi}(\rho\circ(i,j))}{\hat{\pi}(\rho)}$ equals $\frac{w_{ij}}{w_{i'j}}$ or $\frac{w_{ij}}{w_{ij'}}$ and so one needs to know about i' or j' respectively, where i' or j' are defined in (4.1) of Zanella (2014). This increases the amount of computation needed at each MH step when using the proposal given by (P3).

We want to define a modification of $q_\rho(i,j)$, say $\tilde{q}_\rho(i,j)$, that depends on ρ only through whether $(i,j)\in\rho$ or $(i,j)\notin\rho$, meaning that it can be written in the following form

$$\tilde{q}_\rho(i,j)\propto\begin{cases}q^{(add)}(i,j) & \text{if}(i,j)\notin\rho\\ q^{(rem)}(i,j) & \text{if}(i,j)\in\rho\end{cases}\tag{2.2}$$

for some $q^{(add)}(i,j)$ and $q^{(rem)}(i,j)$. In such a way, one can evaluate $q^{(add)}(i,j)$ and $q^{(rem)}(i,j)$ for each edge (i,j) before running the MH algorithm and then, at each MH step, one would only need to update the value of $\tilde{q}_\rho(i,j)$ for the links that have been added or removed (at most 4) by switching from $q^{(add)}(i,j)$ to $q^{(rem)}(i,j)$ or the other way round. At the same time we want $\tilde{q}_\rho(i,j)$ to be similar to $q_\rho(i,j)$ in order to inherit some of its desirable properties in terms of acceptance rates and mixing. In order to do so we do not use $q_\rho(i,j)$ as defined in (2.1) but instead $q_\rho(i,j)\propto\sqrt{\frac{\hat{\pi}(\rho\circ(i,j))}{\hat{\pi}(\rho)}}$. This choice has similar mixing property to (2.1) (for example it satisfies detailed balance conditions in the asymptotic regime), while it allows some simplifications in the calculations below that would be harder with (2.1).

Given $q_\rho(i,j)\propto\sqrt{\frac{\hat{\pi}(\rho\circ(i,j))}{\hat{\pi}(\rho)}}$ and (2.2) a natural choice for $q^{(add)}$ and $q^{(rem)}$ is $\mathbb{E}\left[\sqrt{\frac{\hat{\pi}(\rho\circ(i,j))}{\hat{\pi}(\rho)}}\middle|(i,j)\notin\rho\right]$ and $\mathbb{E}\left[\sqrt{\frac{\hat{\pi}(\rho\circ(i,j))}{\hat{\pi}(\rho)}}\middle|(i,j)\in\rho\right]$ respectively, where the expectation are taken over $\rho\sim\hat{\pi}$. If $(i,j)\in\rho$ then $\frac{\hat{\pi}(\rho\circ(i,j))}{\hat{\pi}(\rho)}=\frac{1}{w_{ij}}$ and therefore we have

$$q^{(rem)}(i,j)\;=\;\mathbb{E}\left[\sqrt{\frac{\hat{\pi}(\rho\circ(i,j))}{\hat{\pi}(\rho)}}\middle|(i,j)\in\rho\right]\;=\;w_{ij}^{-1/2}\,.\tag{2.3}$$

Note that if $(i,j)\in\rho$ then $w_{ij}>0$ almost surely and so $q^{(rem)}(i,j)$ is well-defined.

On the other hand if $(i,j)\notin\rho$ then $\sqrt{\frac{\hat{\pi}(\rho\circ(i,j))}{\hat{\pi}(\rho)}}$ can have different values depending on ρ and we cannot compute $\mathbb{E}\left[\sqrt{\frac{\hat{\pi}(\rho\circ(i,j))}{\hat{\pi}(\rho)}}\middle|(i,j)\notin\rho\right]$ in closed form. Thus approximations are needed. First we fix $(i,j)\in\{1,\dots,n_1\}\times$

$\{1,\ldots,n_2\}$ and we define the following probabilities

$$p_i^{(r)} = \Pr\left((i,j') \notin \rho \ \forall j' = 1,\ldots,n_2 \middle| (i,j) \notin \rho\right) \ ,$$
$$p_j^{(r)} = \Pr\left((i',j) \notin \rho \ \forall i' = 1,\ldots,n_1 \middle| (i,j) \notin \rho\right) \ ,$$
$$p_{i'j'} = \Pr\left((i',j') \in \rho \middle| (i,j) \notin \rho\right) \ .$$

Then we use the following approximation

$$\mathbb{E}\left[\sqrt{\frac{\hat{\pi}(\rho \circ (i,j))}{\hat{\pi}(\rho)}}\middle|(i,j) \notin \rho\right] \ \approx \ p_i^{(r)} p_j^{(b)} \sqrt{w_{ij}} + \tag{2.4}$$
$$p_j^{(b)} \sum_{j' \neq j} \sqrt{\frac{w_{ij}}{w_{ij'}}} p_{ij'} + p_i^{(r)} \sum_{i' \neq i} \sqrt{\frac{w_{ij}}{w_{i'j}}} p_{i'j} + \sum_{i' \neq i} \sum_{j' \neq j} \sqrt{\frac{w_{ij} w_{i'j'}}{w_{i'j} w_{ij'}}} p_{ij'} p_{i'j},$$

Equation (2.4) is an approximation because it factorizes probabilities of non-independent events, like $(i,j') \in \rho$ and $(i',j) \in \rho$.

Then we drop the terms $w_{i'j'}$ in the RHS of (2.4) (thus introducing a further approximation). By doing so the RHS of (2.4) becomes

$$\sqrt{w_{ij}} \left(p_i^{(r)} + \sum_{j' \neq j} \frac{p_{ij'}}{\sqrt{w_{ij'}}}\right) \left(p_j^{(b)} + \sum_{i' \neq i} \frac{p_{i'j}}{\sqrt{w_{i'j}}}\right) = \tag{2.5}$$
$$= \sqrt{w_{ij}} \left(1 - \sum_{j' \neq j} p_{ij'} + \sum_{j' \neq j} \frac{p_{ij'}}{\sqrt{w_{ij'}}}\right) \left(1 - \sum_{i' \neq i} p_{i'j} + \sum_{i' \neq i} \frac{p_{i'j}}{\sqrt{w_{i'j}}}\right) =$$
$$= \sqrt{w_{ij}} \left(1 - \sum_{j' \neq j} p_{ij'} \left(\frac{\sqrt{w_{ij'}} - 1}{\sqrt{w_{ij'}}}\right)\right) \left(1 - \sum_{i' \neq i} p_{i'j} \left(\frac{\sqrt{w_{i'j}} - 1}{\sqrt{w_{i'j}}}\right)\right) \ .$$

Finally by approximating $p_{ij'}$ with $\frac{w_{ij'}}{1+\sum_{s \neq i} w_{sj'} + \sum_l w_{il}}$ and similarly $p_{i'j}$ with $\frac{w_{i'j}}{1+\sum_{s \neq j} w_{i's} + \sum_l w_{lj}}$ we obtain

$$q^{(add)}(i,j) \ = \ \sqrt{w_{ij}} \left(1 - \sum_{j' \neq j} \frac{w_{ij'} - \sqrt{w_{ij'}}}{1 + \sum_{s \neq i} w_{sj'} + \sum_l w_{il}}\right) \left(1 - \sum_{i' \neq i} \frac{w_{i'j} - \sqrt{w_{i'j}}}{1 + \sum_{s \neq j} w_{i's} + \sum_l w_{lj}}\right) \ . \tag{2.6}$$

3. Correctness of the k-dimensional algorithm. Given a k-type point configuration $\mathbf{x}$, a partial matching ρ contained in the induced k-partite hypergraph and a non-trivial subset of the k types A we construct $\mathbf{x}^{2D} = \big((x_1^{2D}, m_1^{2D}, d_1), ..., (x_{n^{2D}}^{2D}, m_{n^{2D}}^{2D}, d_{n^{2D}})\big)$ and $\rho^{2D} = \big(C_1^{2D}, ..., C_{N(\rho^{2D})}^{2D}\big)$ according to step 1 of the transition kernel described in Section 4.2 of Zanella (2014). In step 2 of the same transition kernel we target $\hat{\pi}^{2D}$, that is the two-dimensional version of equation (3.3) of Zanella (2014) modified in order to take into account of the multiplicities $d_1, ..., d_{n^{2D}}$. The explicit expression for $\hat{\pi}^{2D}$ is

$$\hat{\pi}^{2D}(\rho^{2D}) = \pi^{2D}(\rho^{2D} \mid \mathbf{x}^{2D}, \sigma, \lambda, \mathbf{p}^{(c)}) \quad \propto \tag{3.1}$$
$$\prod_{j=1}^{N(\rho^{2D})} \left(\frac{g\left(\overline{x}_{C_j^{2D}}\right) \lambda\, p^{(c)}_{s_j^{2D}}}{c_{s_j^{2D}}} \exp\left(-\frac{\pi \delta^2_{C_j^{2D}}}{2\sigma^2}\right) \prod_{i,l \in C_j^{2D},\, i \neq l} \mathbb{1}(m_i^{2D} \neq m_l^{2D}) \right),$$

where the modified multiplicities, barycenters and intra-cluster square distances are defined as $s_j^{2D} = \sum_{i \in C_j^{2D}} d_i$, $\overline{x}_{C_j^{2D}} = \frac{\sum_{i \in C_j^{2D}} d_i x_i^{2D}}{\sum_{i \in C_j^{2D}} d_i}$ and $\delta^2_{C_j^{2D}} = \sum_{i \in C_j^{2D}} d_i (x_i^{2D} - \overline{x}_{C_j^{2D}})^2$.

Remark 1. *Note that $\hat{\pi}^{2D}(\rho^{2D})$ given in (3.1) can be expressed as a probability distribution on the space of matchings contained in a weighted bipartite graph where the probability of each matching ρ^{2D} is proportional to its total weight (i.e. $\hat{\pi}^{2D}(\rho^{2D}) \propto \prod_{(i,j) \in \rho^{2D}} w_{ij}^{2D}$) for some suitably defined weights w_{ij}^{2D}. Here i and j refer to two points of $\boldsymbol{x}^{2D}$ having different types m^{2D} and w_{ij}^{2D} is the weight of the edge connecting i and j and is given by* (3.1).

We need to prove that $\hat{\pi}^{2D}$ is proportional to $\hat{\pi}$ on the collection of possible moves of $P^{(A)}$ (see Section 4.2 of Zanella (2014) for the definition of $P^{(A)}$). We need the following result.

Lemma 1. *For any $x_1, ..., x_s, z \in \mathbb{R}^n$, let $\overline{x} = s^{-1} \sum_{i=1}^s x_i$. It holds*

$$\sum_{i=1}^{s} (x_i - z)^2 = \sum_{i=1}^{s} (x_i - \overline{x})^2 + s(\overline{x} - z)^2.$$

PROOF. Given $x_1, \dots, x_s, z$ and $\overline{x}$ as above it holds

$$\sum_{i=1}^{s}(x_i - z)^2 = \sum_{i=1}^{s}(x_i - \overline{x} + \overline{x} - z)^2 =$$
$$= \sum_{i=1}^{s}\Big((x_i - \overline{x})^2 + (\overline{x} - z)^2 + (x_i - \overline{x})(\overline{x} - z)\Big) =$$
$$= \sum_{i=1}^{s}(x_i - \overline{x})^2 + s(\overline{x} - z)^2 + 2(\overline{x} - z)\sum_{i=1}^{s}(x_i - \overline{x}) = \sum_{i=1}^{s}(x_i - \overline{x})^2 + s(\overline{x} - z)^2.$$

□

Let C_j be a cluster of $(\mathbf{x}, \rho)$ and C_j^{2D} be the corresponding cluster in the projected two-color configuration $(\mathbf{x}^{2D}, \rho^{2D})$. We define $C_j^A = \{i \in C_j : m_i \in A\}$, $s_j^A = \#\{i \in C_j : m_i \in A\}$ and $\overline{x}_{C_j^A} = \frac{\sum_{i \in C_j^A} x_i}{s_j^A}$. Also $C_j^{A^c}$, $s_j^{A^c}$ and $\overline{x}_{C_j^{A^c}}$ are defined analogously for A^c. Then it follows from Lemma 1 that

$$\delta^2_{C_j} = \sum_{i \in C_j}(x_i - \overline{x}_{C_j})^2 = \sum_{i \in C_j^A}(x_i - \overline{x}_{C_j})^2 + \sum_{i \in C_j^{A^c}}(x_i - \overline{x}_{C_j})^2 =$$
$$\sum_{i \in C_j^A}\left(x_i - \overline{x}_{C_j^A}\right)^2 + s_j^A\left(\overline{x}_{C_j^A} - \overline{x}_{C_j}\right)^2 +$$
$$+ \sum_{i \in C_j^{A^c}}\left(x_i - \overline{x}_{C_j^{A^c}}\right)^2 + s_j^{A^c}\left(\overline{x}_{C_j^{A^c}} - \overline{x}_{C_j}\right)^2 =$$
$$\sum_{i \in C_j^A}\left(x_i - \overline{x}_{C_j^A}\right)^2 + \sum_{i \in C_j^{A^c}}\left(x_i - \overline{x}_{C_j^{A^c}}\right)^2 + \delta^2_{C_j^{2D}}. \tag{3.2}$$

From (3.1), (3.2) and equation (3.3) of Zanella (2014) it follows that

$$\hat{\pi}(\rho) = \hat{\pi}^{2D}(\rho^{2D}) \cdot \exp\left\{\sum_{j=1}^{N(\rho)}\left(\sum_{i \in C_j^A}\left(x_i - \overline{x}_{C_j^A}\right)^2 + \sum_{i \in C_j^{A^c}}\left(x_i - \overline{x}_{C_j^{A^c}}\right)^2\right)\right\}. \tag{3.3}$$

The second factor of the right-hand side of (3.3) is constant with respect to the action of $P^{(A)}$. It follows that $\hat{\pi^{2D}}$ is proportional to $\hat{\pi}$ on the set of allowed moves of $P^{(A)}$, as desired.

References.

Zanella, G. (2014). Bayesian Complementary Clustering, MCMC and Anglo-Saxon placenames. *arXiv preprint arXiv:1409.6994*.

Department of Statistics
University of Warwick
Coventry, CV4 7AL
United Kingdom
E-mail: g.zanella@warwick.ac.uk