

# A Clustering Bayesian Approach for Multivariate Non-Ordered Circular Data

Christophe Abraham    Rémi Servien    Nicolas Molinari

## 1 Appendix

### 1.1 Specification of the prior of $P_0$

Let us recall the notations of Section 2 :  $R$  is the  $2 \times 2$ -matrix of the rotation in  $\mathbb{R}^2$  with angle  $2\pi/k$  and center 0 and  $\varepsilon_{ij}$ ,  $j \in \{2, \dots, k\}$ , are  $k-1$  random variables with distribution  $\varepsilon_{ij} \sim N_2(0, I_2)$ . It is assumed that  $\mu_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{ik}$  are independent and we set  $\mu_{i1} \sim N_2(0, \rho I_2)$  and  $\mu_{ij} = R^{j-1}\mu_{i1} + \varepsilon_{ij}$  for  $j \in \{2, \dots, k\}$ . Then,  $P_0$  denotes the distribution of  $\mu_i = (\mu'_{i1}, \dots, \mu'_{ik})'$  and it is easily seen that  $P_0$  is centered, Gaussian with covariance matrix

$$\Sigma_0(\rho) = \begin{pmatrix} \rho I_2 & \rho R' & \rho R^{2'} & \dots & \rho R^{(k-1)'} \\ \rho R & (\rho + 1)I_2 & \rho R' & \dots & \rho R^{(k-2)'} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \rho R^{k-2} & \rho R^{k-3} & \dots & (\rho + 1)I_2 & \rho R' \\ \rho R^{k-1} & \rho R^{k-2} & \dots & \rho R & (\rho + 1)I_2 \end{pmatrix}.$$

where  $R'$  is the transposed matrix of  $R$ . If we denote by  $R_\alpha$  the matrix of the rotation with angle  $\alpha$ , it is easily seen that the distribution of  $((R_\alpha \mu_{i1})', \dots, (R_\alpha \mu_{ik})')'$  is still  $P_0$  by using that  $RR_\alpha = R_\alpha R$ . Hence,  $P_0$  is rotation invariant. It is worth noting that, thanks to the weak dissymmetry introduced in the definition of  $\mu_i$  between  $\mu_{i1}$  and the other components, the

inverse of  $\Sigma_0(\rho)$  has the following convenient expression:

$$\Sigma_0^{-1}(\rho) = \begin{pmatrix} (\rho^{-1} + (k-1))I_2 & -R' & -R^{2'} & \dots & -R^{(k-2)'} & -R^{(k-1)'} \\ -R & I_2 & 0 & \dots & \dots & 0 \\ -R^2 & 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ -R^{k-2} & \vdots & \ddots & \ddots & I_2 & 0 \\ -R^{k-1} & 0 & \dots & \dots & 0 & I_2 \end{pmatrix},$$

from which we obtain  $|\Sigma_0^{-1}(\rho)| = \rho^{-2}$  by Theorem 13.3.8 of Harville (1997).

Set  $\rho_j = \rho$  for  $j = 1$  and  $\rho_j = \rho + 1$  for  $j \geq 2$ , so that  $\mu_{ij} \sim N_2(0, \rho_j I_2)$  for all  $j \in \{1, \dots, k\}$ . Then,  $\rho_j^{-1} \|\mu_{ij}\|^2 \sim \chi^2(2)$  and  $\mathbb{E}\|\mu_{ij}\|^2 = 2\rho_j$ . Alternatively, using the polar coordinates, the density of  $\|\mu_{ij}\|$  can be obtained as well as its expected value  $\mathbb{E}\|\mu_{ij}\| = \sqrt{\rho_j \pi/2}$ . Correlations between the components of  $\mu$  can be studied from several definitions (see Mardia and Jupp, 1980, for a comparison of several correlation coefficients). Correlation coefficients introduced by Downs (1974); Mardia (1975); Johnson and Wehrly (1977) give the same following result in our situation:

$$\begin{aligned} \text{corr}^2(\mu_{i1}, \mu_{ij}) &= \rho/(\rho + 1), \quad j \geq 2, \\ \text{corr}^2(\mu_{ij}, \mu_{il}) &= \rho^2/(\rho + 1)^2, \quad j \geq 2, \quad l \geq 2, \quad l \neq j. \end{aligned}$$

For all the definitions cited above, the correlation coefficient is always between 0 and 1. We can see that the correlation coefficients between two different components of  $\mu_i$  are increasing functions of  $\rho$  ranging from 0 to 1; in other words, the components of  $\mu_i$  become more and more correlated as  $\rho$  increases. Table 1 gives the correlation coefficients for some values of  $\rho$ .

Table 1: Correlation coefficients for several values of  $\rho$ .

$\rho$	0.1	0.2	0.5	1	2	3	5	10	20	30
$\rho/(\rho + 1)$	0.09	0.17	0.33	0.50	0.67	0.75	0.83	0.91	0.95	0.97
$\rho^2/(\rho + 1)^2$	0.01	0.03	0.11	0.25	0.44	0.56	0.69	0.83	0.91	0.94

The impact of  $\rho$  is highlighted by considering the radial projections of the components of  $\mu_i$ . Denote by  $v_{ij} = \mu_{ij}/\|\mu_{ij}\|$  the radial projection of

$\mu_{ij}$ . Then, we show below that  $(v_{i1}, \dots, v_{ik}) \rightarrow (v_{i1}, Rv_{i1}, \dots, R^{k-1}v_{i1})$  almost surely as  $\rho \rightarrow \infty$  and  $(v_{i1}, \dots, v_{ik}) \rightarrow (\varepsilon_{i1}/\|\varepsilon_{i1}\|, \dots, \varepsilon_{ik}/\|\varepsilon_{ik}\|)$  almost surely as  $\rho \rightarrow 0$  where  $\varepsilon_{i1}, \dots, \varepsilon_{ik}$  are independent and identically distributed random variables with distribution  $N_2(0, I_2)$ . Consequently, the radial projections of the components of  $\mu_i$  are highly correlated and equally spaced for large values of  $\rho$  but approximately independent and uniformly distributed for small values of  $\rho$ . Indeed, if we set  $\varepsilon_{i1} = \mu_{i1}/\sqrt{\rho}$ , we have

$$\mu_{ij} = R^{j-1}\mu_{i1} + \varepsilon_{ij} = \sqrt{\rho}R^{j-1}\varepsilon_{i1} + \varepsilon_{ij},$$

and

$$v_{ij} = \frac{\sqrt{\rho}R^{j-1}\varepsilon_{i1} + \varepsilon_{ij}}{\|\sqrt{\rho}R^{j-1}\varepsilon_{i1} + \varepsilon_{ij}\|}.$$

Then, by the continuous mapping theorem, it is easily seen that  $v_{ij} \rightarrow R^{j-1}\varepsilon_{i1}/\|R^{j-1}\varepsilon_{i1}\|$  almost surely as  $\rho \rightarrow \infty$ . By noting that

$$\frac{R^{j-1}\varepsilon_{i1}}{\|R^{j-1}\varepsilon_{i1}\|} = \frac{R^{j-1}\varepsilon_{i1}}{\|\varepsilon_{i1}\|} = R^{j-1}\frac{\varepsilon_{i1}}{\|\varepsilon_{i1}\|} = R^{j-1}v_{i1},$$

we conclude that  $(v_{i1}, \dots, v_{ik}) \rightarrow (v_{i1}, Rv_{i1}, \dots, R^{k-1}v_{i1})$  almost surely as  $\rho \rightarrow \infty$ . We prove similarly that  $(v_{i1}, \dots, v_{ik}) \rightarrow (\varepsilon_{i1}/\|\varepsilon_{i1}\|, \dots, \varepsilon_{ik}/\|\varepsilon_{ik}\|)$  almost surely as  $\rho \rightarrow 0$ .

## 1.2 SAMS and Gibbs Samplers

### 1.2.1 SAMS Sampler

The SAMS sampler is given in detail in Dahl (2003). Formula (12) of Dahl (2003) reduces to

$$P(l \in S_i | S_i, S_j) = \frac{|S_i|N_{2k}(y_l; \Sigma_i \sum_{h \in S_i} y_h / |S_i|, I_{2k} + \Sigma_i)}{|S_i|N_{2k}(y_l; \Sigma_i \sum_{h \in S_i} y_h / |S_i|, I_{2k} + \Sigma_i) + |S_j|N_{2k}(y_l; \Sigma_j \sum_{h \in S_j} y_h / |S_j|, I_{2k} + \Sigma_j)}$$

and formula (14) for the Metropolis-Hastings ratio is obtained with

$$p(y_{S_j}) = \prod_{h=1}^{|S_j|} N_{2k} \left( y_{i_h}; \Sigma_{-j} \sum_{h \in S_{-j}} y_h / |S_{-j}|, I_{2k} + \Sigma_{-j} \right)$$

where  $S_j = \{i_1, \dots, i_{|S_j|}\}$ ,  $S_{-j} = \{i_1, \dots, i_{j-1}\}$  and  $\Sigma_{-j} = (|S_{-j}|^{-1}I_{2k} + \Sigma_0^{-1}(\rho))^{-1}$ .

### 1.2.2 Gibbs Sampler

Let us denote by  $\eta = \{S_1, \dots, S_q\}$  the current partition of the algorithm. For  $i = 1, \dots, n$ , the observation  $i$  is assigned to cluster  $S_j$ ,  $j \in \{1, \dots, q\}$  with probability proportional to

$$|S_j^-| \times N_{2k} \left( y_i; \Sigma_j \sum_{i \in S_j} y_i / |S_j^-|, I_{2k} + \Sigma_j \right)$$

where  $|S_j^-|$  is the cardinal of  $S_j \setminus \{i\}$ , or to (a new) cluster  $S_{q+1}$  with probability proportional to

$$n_0 \times N_{2k} (y_i; 0, I_{2k} + \Sigma_0(\rho)).$$

### 1.3 Full Conditional Distributions

**Full Conditional of  $r$**  Remember that  $x_i = (x'_{i1}, \dots, x'_{ik})' \in (\mathbb{R}^2)^k$ ,  $i \in \{1, \dots, n\}$ , are independent with distribution

$N_{2k}(\mu_i^{\tau_i}, I_{2k})$  with  $\mu_i^{\tau_i} = (\mu'_{i\tau_i(1)}, \dots, \mu'_{i\tau_i(k)})' \in (\mathbb{R}^2)^k$  and that  $x_{ij} = (x_{ij1}, x_{ij2})' = (r_{ij} \cos \theta_{ij}, r_{ij} \sin \theta_{ij})'$ .

Then, it is easy to see that  $(\theta_{ij}, r_{ij})$  are independent given  $\tau, \mu, \rho$  and  $n_0$ , with density:

$$p(\theta_{ij}, r_{ij} | \tau, \mu, \rho, n_0) = (2\pi)^{-1} e^{-\frac{1}{2} \mu'_{i\tau_i(j)} \mu_{i\tau_i(j)}} r_{ij} e^{-\frac{1}{2} (r_{ij}^2 - 2r_{ij} u'_{ij} \mu_{i\tau_i(j)})},$$

with  $u'_{ij} = (\cos \theta_{ij}, \sin \theta_{ij})$ . Then,

$$\begin{aligned} p(r | \theta, \tau, \mu, \rho, n_0) &\propto p(\theta, r | \tau, \mu, \rho, n_0) \\ &\propto \prod_{i=1}^n \prod_{j=1}^k p(\theta_{ij}, r_{ij} | \tau, \mu, \rho, n_0) \\ &\propto \prod_{i=1}^n \prod_{j=1}^k r_{ij} e^{-\frac{1}{2} (r_{ij} - u'_{ij} \mu_{i\tau_i(j)})^2}. \end{aligned}$$

### 1.4 Proof of Proposition 1

a) If we denote by  $\phi_{2k}$  the density of the  $N_{2k}(0, I_{2k})$  distribution, it can be shown after some calculations, that:

$$m(x_S) = \left( \prod_{i \in S} \phi_{2k}(x_i) \right) |\Sigma_0|^{-1/2} |\Sigma_S|^{-1/2} \exp \left( \frac{1}{2} \left\| \sum_{i \in S} x_i \right\|_S^2 \right).$$

Then, we have:

$$\prod_{j=1}^q m(x_{S_j}) = \left( \prod_{i=1}^n \phi_{2k}(x_i) \right) |\Sigma_0|^{-q/2} |\Sigma_{S_j}|^{-q/2} \exp \left( \frac{1}{2} \sum_{j=1}^q \left\| \sum_{i \in S_j} x_i \right\|_{S_j}^2 \right),$$

and

$$\prod_{j=1}^q m(x_{S_j}^\tau) = \left( \prod_{i=1}^n \phi_{2k}(x_i^{\tau_i}) \right) |\Sigma_0|^{-q/2} |\Sigma_{S_j}|^{-q/2} \exp \left( \frac{1}{2} \sum_{j=1}^q \left\| \sum_{i \in S_j} x_i^{\tau_i} \right\|_{S_j}^2 \right).$$

From (3.1), it can be seen that  $\phi_{2k}(x_i) = \phi_{2k}(x_i^{\tau_i})$  and we conclude that:

$$\frac{\prod_{j=1}^q m(x_{S_j}^\tau)}{\prod_{j=1}^q m(x_{S_j})} = \exp \left( \frac{1}{2} \sum_{j=1}^q \left( \left\| \sum_{i \in S_j} x_i^{\tau_i} \right\|_{S_j}^2 - \left\| \sum_{i \in S_j} x_i \right\|_{S_j}^2 \right) \right),$$

hence the result.

b) If we denote by  $g$  the density of  $G$  with respect to the counting measure, we have:

$$p_I(\eta|x) \propto g(\eta) p_I(x|\eta),$$

and

$$p_{II}(\eta|x) \propto g(\eta) \frac{1}{(k!)^n} \sum_{\tau \in \mathcal{P}^n} p_I(x^\tau|\eta).$$

and we deduce that:

$$\frac{p_{II}(\eta|x)}{p_I(\eta|x)} = B_G f(x, \eta),$$

where

$$B_G = \frac{\sum_{\eta} g(\eta) p_I(x|\eta)}{\sum_{\eta} g(\eta) \frac{1}{(k!)^n} \sum_{\tau \in \mathcal{P}^n} p_I(x^\tau|\eta)}.$$

c) From Lemma 1 below, we have:

$$\begin{aligned} \min_G B_G &= \min_{\eta} \frac{p_I(x|\eta)}{\frac{1}{(k!)^n} \sum_{\tau \in \mathcal{P}^n} p_I(x^\tau|\eta)}, \\ &= \min_{\eta} \frac{1}{f(x, \eta)} \\ &= \frac{1}{\max_{\eta} f(x, \eta)}. \end{aligned}$$

□

**Lemma 1.** Let  $h \in \mathbb{R}^n$  and  $f \in \mathbb{R}^n$  such that  $f_i > 0$  and  $h_i > 0$  for all  $i \in \{1, \dots, n\}$ . Write  $\mathcal{D} = \{p \in \mathbb{R}^n, \sum_{i=1}^n p_i = 1, p_i \geq 0 \text{ for all } i\}$ . We have:

$$\inf_{p \in \mathcal{D}} \frac{\sum_{i=1}^n p_i f_i}{\sum_{i=1}^n p_i h_i} = \min_{1 \leq i \leq n} \frac{f_i}{h_i}.$$

**Proof of Lemma 1** Assume without loss of generality that  $\min_i f_i/h_i = f_1/h_1$ . Then we have:

$$\begin{aligned} \frac{f_i}{g_i} \geq \frac{f_1}{g_1} \text{ for all } i \in \{1, \dots, n\} &\iff f_i g_1 - f_1 g_i \geq 0 \text{ for all } i \in \{1, \dots, n\} \\ &\iff \sum_{i=1}^n p_i (f_i g_1 - f_1 g_i) \geq 0 \text{ for all } p \in \mathcal{D} \\ &\iff g_1 \sum_{i=1}^n p_i f_i \geq f_1 \sum_{i=1}^n p_i g_i \text{ for all } p \in \mathcal{D} \\ &\iff \frac{\sum_{i=1}^n p_i f_i}{\sum_{i=1}^n p_i g_i} \geq \frac{f_1}{g_1} \text{ for all } p \in \mathcal{D}. \end{aligned}$$

We conclude by noting that

$$\frac{\sum_{i=1}^n p_i f_i}{\sum_{i=1}^n p_i g_i} = \frac{f_1}{g_1}$$

for  $p = (1, 0, \dots, 0)$ . □

## References

- Dahl, D. B. (2003). An improved merge-split sampler for conjugate Dirichlet process mixture models. Technical Report, Univ. of Wisconsin - Madison, **1086**, 1–32.
- Downs, T. D. (1974). Rotational angular correlations. In Ferin, M., Halberg, F., Richart, M., and van der Wiele, L., editors, Biorhythms and Human Reproduction, pages 97–104, New York: Wiley, 1974.
- Harville, D. A. (1997). Matrix algebra from a statistician's perspective. Springer, New York.

- Johnson, R. A. and Wehrly, T. (1977). Measures and models for angular correlation and angular-linear correlation. Journal of the Royal Statistical Society: Series B, **39**, 222–229.
- Mardia, K. V. (1975). Statistics of directinal data (with discussion). Journal of Royal Statistical Society: Series B, **37**(1), 349–393.
- Mardia, K. V. and Jupp, P. E. (1980). A general correlation coefficient for directional data and related regerssion problems. Biometrika, **67**(1), 163–173.