

## SUPPLEMENTARY MATERIAL: MIXTURE OF HIDDEN MARKOV MODELS FOR ACCELEROMETER DATA

BY MARIE DU ROY DE CHAUMARAY, MATTHIEU MARBAC, FABIEN NAVARRO

**1. Model identifiability.** The proof of Theorem 1 is split in two parts:

1. Identifiability of the parameters of the specific distribution per state is obtained using the approach of Teicher (1963). Hence  $\forall h = 1, \dots, M$

$$\boldsymbol{\lambda}_h = \tilde{\boldsymbol{\lambda}}_h, \quad \sum_{k=1}^K \delta_k \pi_{kh} (1 - \varepsilon_h) = \sum_{k=1}^K \tilde{\delta}_k \tilde{\pi}_{kh} (1 - \tilde{\varepsilon}_h).$$

2. Identifiability of the transition matrices and of the  $\varepsilon$  is shown using properties of Vandermonde matrices. Hence,

$$\forall k = 1, \dots, K, \quad \delta_k = \tilde{\delta}_k, \quad \mathbf{A}_k = \tilde{\mathbf{A}}_k, \quad \pi_k = \tilde{\pi}_k, \quad \varepsilon = \tilde{\varepsilon}.$$

1.1. *Identifiability of the parameters of the specific distribution per state.* Considering the marginal distribution at time  $t = 0$ , we have

$$\sum_{k=1}^K \sum_{h=1}^M \delta_k \pi_{kh} g(y_{i(0)}; \boldsymbol{\lambda}_h, \varepsilon_h) = \sum_{k=1}^K \sum_{h=1}^M \tilde{\delta}_k \tilde{\pi}_{kh} g(y_{i(0)}; \tilde{\boldsymbol{\lambda}}_h, \tilde{\varepsilon}_h).$$

Note that  $g(y_{i(0)}; \boldsymbol{\lambda}_h, \varepsilon_h) = (1 - \varepsilon_h)g_c(y_{i(0)}; \boldsymbol{\lambda}_h) + \varepsilon_h \mathbf{1}_{\{y_{i(0)}=0\}}$  is a pdf of a zero-inflated distribution, so it is a pdf of a bi-component mixture. We now use the same reasoning as Teicher (1963). We have

$$1 + \sum_{h=2}^M \frac{g(y_{i(0)}; \boldsymbol{\lambda}_h, \varepsilon_h) \sum_{k=1}^K \delta_k \pi_{kh}}{g(y_{i(0)}; \boldsymbol{\lambda}_1, \varepsilon_1) \sum_{k=1}^K \delta_k \pi_{k1}} = \frac{g(y_{i(0)}; \tilde{\boldsymbol{\lambda}}_1, \tilde{\varepsilon}_1) \sum_{k=1}^K \tilde{\delta}_k \tilde{\pi}_{k1}}{g(y_{i(0)}; \boldsymbol{\lambda}_1, \varepsilon_1) \sum_{k=1}^K \delta_k \pi_{k1}} + \sum_{h=2}^M \frac{g(y_{i(0)}; \tilde{\boldsymbol{\lambda}}_h, \tilde{\varepsilon}_h) \sum_{k=1}^K \tilde{\delta}_k \tilde{\pi}_{kh}}{g(y_{i(0)}; \boldsymbol{\lambda}_1, \varepsilon_1) \sum_{k=1}^K \delta_k \pi_{k1}}.$$

Considering  $y_{i(0)} \rightarrow \rho$ , by Assumption 2, we have

$$\boldsymbol{\lambda}_1 = \tilde{\boldsymbol{\lambda}}_1, \quad (1 - \varepsilon_1) \sum_{k=1}^K \delta_k \pi_{k1} = (1 - \tilde{\varepsilon}_1) \sum_{k=1}^K \tilde{\delta}_k \tilde{\pi}_{k1}.$$

Repeating the previous argument with  $h = 2, \dots, M$ , we conclude that, for  $h \in \{1, \dots, M\}$ ,

$$\boldsymbol{\lambda}_h = \tilde{\boldsymbol{\lambda}}_h, \quad (1 - \varepsilon_h) \sum_{k=1}^K \delta_k \pi_{kh} = (1 - \tilde{\varepsilon}_h) \sum_{k=1}^K \tilde{\delta}_k \tilde{\pi}_{kh}.$$

1.2. *Identifiability of the transition matrices.* First, we introduce two technical lemmas of which proofs are discussed in the next subsection. Second, we show that  $\mathbf{A}_k[1, 1] = \tilde{\mathbf{A}}_k[1, 1]$  then we extend the results to the whole transition matrices.

LEMMA 1. *Let  $N_0, N_1, \tilde{N}_0$  and  $\tilde{N}_1$  be four definite positive matrices of size  $K \times K$  such that for  $u \in \{1, \dots, K\}$  and  $k \in \{1, \dots, K\}$ ,*

$$N_0[u, k] = a_k^{u-1}, N_1[u, k] = a_k^{K+u-1}, \tilde{N}_0[u, k] = \tilde{a}_k^{u-1}, \tilde{N}_1[u, k] = \tilde{a}_k^{K+u-1},$$

*with  $a_k > a_{k+1} > 0$ ,  $\tilde{a}_k > \tilde{a}_{k+1} > 0$  and  $a_1 \geq \tilde{a}_1$ . If for any  $\tilde{w} \in \mathbb{R}_+^K$  there exists  $w \in \mathbb{R}_+^K$   $N_0 w = \tilde{N}_0 \tilde{w}$  and  $N_1 w = \tilde{N}_1 \tilde{w}$  then for  $k \in \{1, \dots, K\}$   $a_k = \tilde{a}_k$  and  $w = \tilde{w}$ .*

LEMMA 2. *Let  $N_0, \tilde{N}_0$  be two definite positive matrices of size  $K \times K$  such that for  $u \in \{1, \dots, K\}$  and  $k \in \{1, \dots, K\}$ ,*

$$N_0[u, k] = a_k^{u-1}, N_1[u, k] = a_k^{K+u-1},$$

*with  $a_k > a_{k+1} > 0$ ,  $\tilde{a}_k > \tilde{a}_{k+1} > 0$  and  $a_1 \geq \tilde{a}_1$ . Let  $D_u = \text{diag}(a_1^{Ku}, \dots, a_K^{Ku})$  and  $\tilde{D}_u = \text{diag}(\tilde{a}_1^{Ku}, \dots, \tilde{a}_K^{Ku})$ . If there exist  $\alpha \in ]0, 1[$ ,  $\tilde{\alpha} \in ]0, 1[$ ,  $w \in \mathbb{R}_+^K$  and  $\tilde{w} \in \mathbb{R}_+^K$  such that for  $u \in \{0, \dots, K-1\}$ , we have*

$$\alpha N_0 D_u w = \tilde{\alpha} \tilde{N}_0 \tilde{D}_u \tilde{w},$$

*then for  $k \in \{1, \dots, K\}$   $a_k = \tilde{a}_k$  and  $w = \tilde{w}$ .*

We consider the marginal distribution of  $(y_{i(0)}, \dots, y_{i(t-1)})$  with  $t = 1, \dots, 2K$ , where  $y_{i(0)} = y_{i(t')}$  for each  $t' = 1, \dots, t-3$ ,  $y_{i(t-2)} = y_{i(0)}^{\tau_1}$ ,  $y_{i(t-1)} = y_{i(0)}^{\tau_2}$  and  $y_{i(t)} = y_{i(0)}^{\tau_3}$ . Therefore, taking  $\tau_1 = \tau_2 = \tau_3 = 1$  and letting  $y_{i(0)}$  tend to  $\rho$  (see Assumption 1), we obtain, for  $t = 1, \dots, 2K$ , that

$$(1 - \varepsilon_1) \sum_{k=1}^K \delta_k \pi_{k1} (\mathbf{A}_k[1, 1](1 - \varepsilon_1))^{t-1} = (1 - \tilde{\varepsilon}_1) \sum_{k=1}^K \tilde{\delta}_k \tilde{\pi}_{k1} \left( \tilde{\mathbf{A}}_k[1, 1](1 - \tilde{\varepsilon}_1) \right)^{t-1}.$$

Because, we consider  $2K$  marginal distributions, we can use Lemma 2 by setting  $\alpha = (1 - \varepsilon)$ ,  $\tilde{\alpha} = (1 - \tilde{\varepsilon})$ ,  $a_k = \mathbf{A}_k[1, 1](1 - \varepsilon_1)$ ,  $\tilde{a}_k = \tilde{\mathbf{A}}_k[1, 1](1 - \tilde{\varepsilon}_1)$ ,  $w_k = \delta_k \pi_{k1}$  and  $\tilde{w}_k = \tilde{\delta}_k \tilde{\pi}_{k1}$ . Therefore, we have  $\varepsilon = \tilde{\varepsilon}$ ,  $\mathbf{A}_k[1, 1] = \tilde{\mathbf{A}}_k[1, 1]$  and  $\delta_k \pi_{k1} = \tilde{\delta}_k \tilde{\pi}_{k1}$ . Using the previous approach, with  $\tau_1 = \tau_2 = 1$  and  $\tau_3 < 1$ , with  $h = 2, \dots, M$ , we have for  $t = 1, \dots, K$

$$(1 - \varepsilon_h)(1 - \varepsilon_1) \sum_{k=1}^K \delta_k \pi_{k1} (\mathbf{A}_k[1, 1](1 - \varepsilon_1))^{t-2} \mathbf{A}_k[1, h] = (1 - \tilde{\varepsilon}_h)(1 - \varepsilon_1) \sum_{k=1}^K \tilde{\delta}_k \tilde{\pi}_{k1} (\tilde{\mathbf{A}}_k[1, 1](1 - \varepsilon_1))^{t-2} \tilde{\mathbf{A}}_k[1, h],$$

and thus  $\mathbf{A}_k[1, h] = \tilde{\mathbf{A}}_k[1, h]$  and  $\varepsilon_h = \tilde{\varepsilon}_h$ . Similarly, taking  $\tau_2 < 1$  and  $\tau_1 = \tau_3 = 1$ , we have  $\mathbf{A}_k[h, 1] = \tilde{\mathbf{A}}_k[h, 1]$ . Finally, we have  $\mathbf{A}_k[h, h'] = \tilde{\mathbf{A}}_k[h, h']$  by increasing  $h$  and  $h'$ , by noting that with suitable choices of  $\tau_1, \tau_2$  and  $\tau_3$ , we have for  $t = 1, \dots, K$

$$\sum_{k=1}^K \delta_k \pi_{k1} (\mathbf{A}_k[1, 1](1 - \varepsilon_1))^{t-2} \mathbf{A}_k[1, h] \mathbf{A}_k[h, h'] \mathbf{A}_k[h', 1] =$$

$$\sum_{k=1}^K \tilde{\delta}_k \tilde{\pi}_{k1} (\mathbf{A}_k[1, 1](1 - \varepsilon_1))^{t-2} \mathbf{A}_k[1, h] \tilde{\mathbf{A}}_k[h, h'] \mathbf{A}_k[h', 1].$$

### 1.3. Proofs of the two technical lemmas.

PROOF OF LEMMA 1. Since  $a_k \neq a_{k'}$  and  $\tilde{a}_k \neq \tilde{a}_{k'}$ , then  $N_0, N_1, \tilde{N}_0$  and  $\tilde{N}_1$  are Vandermonde matrices and thus are invertible. Therefore, we have  $w = N_0^{-1} \tilde{N}_0 \tilde{w} = N_1^{-1} \tilde{N}_1 \tilde{w}$ , and thus

$$(N_0^{-1} \tilde{N}_0 - N_1^{-1} \tilde{N}_1) \tilde{w} = 0,$$

or similarly for  $u \in \{1, \dots, K\}$ ,

$$\sum_{k=1}^K a_k^u w_k = \sum_{k=1}^K \tilde{a}_k^u \tilde{w}_k.$$

Since the previous equation holds for any  $\tilde{w}$ , we have  $N_0^{-1} \tilde{N}_0 = N_1^{-1} \tilde{N}_1$ . Moreover, we have  $N_1 = N_0 D$  and  $\tilde{N}_1 = \tilde{N}_0 \tilde{D}$  where  $D = \text{diag}(a_1^K, \dots, a_K^K)$  and  $\tilde{D} = \text{diag}(\tilde{a}_1^K, \dots, \tilde{a}_K^K)$ . Denoting  $R = N_0^{-1} \tilde{N}_0$ ,  $DR = R\tilde{D}$  and then for  $u \in \{1, \dots, K\}$  and  $k \in \{1, \dots, K\}$

$$(1) \quad a_u^K R[u, k] = \tilde{a}_k^K R[u, k].$$

We now show that  $D = \tilde{D}$  and  $w = \tilde{w}$ , and hence  $R = I_K$  and  $\tilde{N}_0 = N_0$ , where  $I_K$  is the identity matrix of size  $K$ . First we show that  $a_1 = \tilde{a}_1$  and  $w_1 = \tilde{w}_1$ .

- If  $R[1, j] \neq 0$ , (1) implies that  $a_1^K R[1, j] = \tilde{a}_j^K R[1, j]$  and thus  $a_1 = \tilde{a}_j$ . However, this is impossible because  $a_1 \geq \tilde{a}_1 > a_j$  for  $j \in \{2, \dots, K\}$ . Hence, we have  $R[1, j] = 0$  for  $j = 2, \dots, K$ .
- Noting that  $R$  is a product of two invertible matrices,  $R$  is invertible. Therefore,  $R[1, 1] \neq 0$  because  $R[1, j] = 0$  for  $j = 2, \dots, K$ . Hence, we have  $a_1 = \tilde{a}_1$ .
- Note that  $R[1, 1] = \sum_{k=1}^K (N_0^{-1})[1, k] \tilde{N}_0[k, 1]$  and that  $\tilde{N}_0[k, 1] = \tilde{a}_1^k = a_1^k = N_0[k, 1]$ . Therefore, we have  $R[1, 1] = \sum_{k=1}^K (N_0^{-1})[1, k] N_0[k, 1] = (N_0^{-1} N_0)[1, 1] = 1$ .
- For  $j = 2, \dots, K$ ,  $a_1 > a_j$  so we have  $R[j, 1] = 0$ , because  $a_1 = \tilde{a}_1$ .
- Because  $w = R\tilde{w}$ , we have  $w_1 = \tilde{w}_1$ .

Equality of  $a_k = \tilde{a}_k$  and  $w_k = \tilde{w}_k$  can be shown recursively for  $k = 2, \dots, K$  using the same reasoning.  $\square$

## 2. Probabilities of misclassification.

2.1. *Technical lemmas.* This section presents some notations and three lemmas which are used for the proof of Theorem 2. The technical lemmas discuss the concentration of the frequency of the observation  $y_{i(t)}$  in a region of interest  $W$ , give an upper bound of  $p(\mathbf{y}_i \mid Z_{ik} = 1)$  and a concentration result of the ratio of  $\frac{p(\tilde{\mathbf{x}}_{ik}, \mathbf{y}_i \mid Z_{ik}=1)}{p(\tilde{\mathbf{x}}_{ik_0}, \mathbf{y}_i \mid Z_{ik_0}=1)}$ , where  $\tilde{x}_{ik} = \operatorname{argmax}_{\mathbf{x}_i \in \mathcal{X}} p(\mathbf{x}_i, \mathbf{y}_i \mid Z_{ik} = 1)$  is the estimator of the latent states conditionally on the observation  $\mathbf{y}_i$  and on component  $k$  obtained by applying the *maximum a posteriori* rule with the true parameter  $\boldsymbol{\theta}$ . The proof of the lemmas uses two concentration results for hidden Markov chains given by Kontorovich and Weiss (2014) and by León and Perron (2004).

**Preliminaries** Let  $\mathbf{v}_{ik(t)} = (v_{ik(t)h\ell}; h = 1, \dots, M; \ell = 1, \dots, M)$  with  $v_{ik(t)h\ell} = x_{ik(t-1)h} x_{ik(t)\ell}$  and  $\tilde{\mathbf{v}}_{ik(t)} = (\tilde{v}_{ik(t)h\ell}; h = 1, \dots, M; \ell = 1, \dots, M)$  with  $\tilde{v}_{ik(t)h\ell} = \tilde{x}_{ik(t-1)h} \tilde{x}_{ik(t)\ell}$ . In the following,  $\mathbb{P}_0(\cdot) = \mathbb{P}(\cdot \mid Z_{ik_0} = 1)$  by considering the true parameters.

REMARK 1. For any  $k = 1, \dots, g$ ,  $\mathbf{V}_{ik(t)}$  is a finite, ergodic and reversible Markov chain with  $M^2$  states and transition matrix  $\mathbf{P}_k$  with general term defined for any  $(h_1, h_2, h_3, h_4) \in M^4$  by

$$\mathbf{P}_k[(h_1 - 1)M + h_2, (h_3 - 1)M + h_4] = \mathbb{P}(V_{ik(t)h\ell} = 1) = \begin{cases} 0 & \text{if } h_2 \neq h_3 \\ \mathbf{A}_k[h_2, h_4] & \text{otherwise} \end{cases}.$$

Moreover, the non-zero eigenvalues of  $\mathbf{P}_k$  are the non-zero eigenvalues of  $\mathbf{A}_k$  and the eigenvectors of  $\mathbf{P}_k$  are obtained from the eigenvectors of  $\mathbf{A}_k$ .

THEOREM 1 (Kontorovich and Weiss (2014)). Let  $U_{(1)}, U_{(2)}, \dots$  be a stationary  $\mathbb{N}$ -valued  $(G, \eta)$ -geometrically ergodic Markov or hidden Markov chain, and consider the occupation frequency

$$\hat{\rho}(E) = \frac{1}{T} \sum_{t=1}^T \mathbf{1}_{\{U_{(t)} \in E\}}, \quad E \subset \mathbb{N}.$$

When  $\sum_{u \in \mathbb{N}} \sqrt{\rho_u} < \infty$  with  $\rho_u = \mathbb{P}(U_{(1)} = u)$ , then for any  $\varepsilon > 0$

$$\mathbb{P} \left( \sup_{E \subset \mathbb{N}} |\rho(E) - \hat{\rho}(E)| > \varepsilon + \gamma_T(G, \eta) \sum_{y \in \mathbb{N}} \sqrt{\rho_y} \right) \leq e^{-\frac{T}{2G^2} (1-\eta)^2 \varepsilon^2},$$

where

$$\gamma_T(G, \eta) = \frac{1}{2} \sqrt{\frac{1 + 2G\eta}{T(1-\eta)}}.$$

**THEOREM 2** (León and Perron (2004)). *For all pairs  $(V, f)$ , such that  $V = (V_{(1)}, \dots, V_{(T)})$  is a finite, ergodic and reversible Markov chain in stationary state with the second-largest eigenvalue  $\lambda$  and  $f$  is a function taking values in  $[0, 1]$  such that  $\mathbb{E}[f(V_{(t)})] < \infty$ , the following bounds, with  $\lambda_0 = \max(0, \lambda)$ , hold for all  $s > 0$  such that  $\mathbb{E}[f(V_{(t)})] + s < 1$  and all time  $T$*

$$\mathbb{P}\left(\sum_{t=1}^T f(V_{(t)}) \geq (\mathbb{E}[f(V_{(1)})] + s)T \mid Z_{ik_0} = 1\right) \leq \exp\left(-2\frac{1-\lambda_0}{1+\lambda_0}Ts^2\right).$$

**Concentration of the frequency of the observations in  $W$**  Let  $W \subset \mathbb{R}^+$  be the subset of  $\mathbb{R}^+$  where the estimator of  $x_{i(t)}$  obtained by the *maximum a posteriori* rule is sensitive to  $x_{i(t-1)}$  and  $x_{i(t)}$  conditionally on  $y_{i(t)}$  and component  $k$ . We define

$$W = \{u \in \mathbb{R}^+ : \text{card}(\cup_{k=1}^g E_k(u)) \geq 2\},$$

where

$$E_k(u) = \{h_2 : \exists(h_1, h_3), h_2 = \text{argmax} e_k(u; h_1, h_2, h_3)\},$$

and

$$e_k(u; h_1, h_2, h_3) = \mathbf{A}_k[h_1, h_2] \mathbf{A}_k[h_2, h_3] g(u; \boldsymbol{\lambda}_{h_2}).$$

**LEMMA 3.** *Let  $\rho_{k_0} = \mathbb{P}_0(Y_{i(2)} \in W)$  and  $\hat{\rho}_{k_0} = \sum_{t=1}^T \mathbf{1}_{\{y_{i(t)} \in W\}}$ . For any  $\delta_1 > \frac{1}{\sqrt{2T}}$ ,*

$$\mathbb{P}_0(\hat{\rho}_{k_0} < \rho_{k_0} - \delta_1) \leq e^{-Tc_1},$$

$\delta_1 = \varepsilon + \frac{1}{\sqrt{2T}}$  and  $c_1 = \frac{1}{2}(\delta_1 - \frac{1}{\sqrt{2T}})^2 > 0$ . Moreover,  $\hat{\rho}_{k_0}$  is a consistent estimate of  $\rho_{k_0}$  because the marginal distribution of  $Y_{i(t)}$  is the same for any  $t$ , and thus  $\rho_{k_0} = \mathbb{P}_0(Y_{i(t)} \in W)$  for any  $t$ .

**PROOF OF LEMMA 3.** We have,

$$\mathbb{P}\left(|\rho_{k_0} - \hat{\rho}_{k_0}| > \varepsilon + \frac{1}{\sqrt{2T}}\right) \leq \mathbb{P}\left(|\rho_{k_0} - \hat{\rho}_{k_0}| > \varepsilon + \frac{1}{2\sqrt{T}}(\sqrt{\rho_{k_0}} + \sqrt{1 - \rho_{k_0}})\right).$$

Let  $U_{(t)} = \mathbf{1}_{\{y_{i(t)} \in W\}}$ . Then, for any  $k = 1, \dots, g$ ,  $U_{(1)}, \dots, U_{(T)}$  is a stationary  $\{0, 1\}$ -valued  $(1, 0)$ -geometrically ergodic hidden Markov chain conditionally on component  $k$ . Hence, by Theorem 1,

$$\mathbb{P}\left(|\rho_{k_0} - \hat{\rho}_{k_0}| > \varepsilon + \frac{1}{2\sqrt{T}}(\sqrt{\rho_{k_0}} + \sqrt{1 - \rho_{k_0}})\right) \leq e^{-\frac{T}{2}\varepsilon^2}.$$

We can conclude that

$$\mathbb{P}(\hat{\rho}_{k_0} < \rho_{k_0} - \delta_1) \leq e^{-Tc_1},$$

$$\delta_1 = \varepsilon + \frac{1}{\sqrt{2T}} \text{ and } c_1 = \frac{1}{2}(\delta_1 - \frac{1}{\sqrt{2T}})^2. \quad \square$$

**Upper-bound of the conditional probability of  $\mathbf{y}_i$  given  $Z_{ik} = 1$**  Let  $\gamma$  and  $\bar{\gamma}$  be upper-bounds of the ratio  $\frac{p(\tilde{\mathbf{x}}_{i(t-1)}, \mathbf{x}_{i(t)}, \tilde{\mathbf{x}}_{i(t+1)}, \mathbf{y}_{i(t)} | Z_{ik}=1)}{p(\tilde{\mathbf{x}}_{i(t-1)}, \tilde{\mathbf{x}}_{i(t)}, \tilde{\mathbf{x}}_{i(t+1)}, \mathbf{y}_{i(t)} | Z_{ik}=1)}$  when  $\mathbf{y}_{i(t)} \in W$  and  $\mathbf{y}_{i(t)} \notin W$  respectively. Thus,  $\gamma = \max_k \max_{h_1, h_2, h_3, h_4} \frac{A_k[h_1, h_2]}{A_k[h_3, h_4]}$  and  $\bar{\gamma}$  permit to upper bound the ratio between the likelihood computed for any  $(\mathbf{x}_i, \mathbf{y}_i)$  given  $Z_{ik} = 1$  and the likelihood computed with  $(\tilde{\mathbf{x}}_{ik}, \mathbf{y}_i)$  given  $Z_{ik} = 1$ . We have, if  $\mathbf{y}_{i(t)} \in W$ ,

$$\frac{p(\tilde{\mathbf{x}}_{i(t-1)}, \mathbf{x}_{i(t)}, \tilde{\mathbf{x}}_{i(t+1)}, \mathbf{y}_{i(t)} | Z_{ik} = 1)}{p(\tilde{\mathbf{x}}_{i(t-1)}, \tilde{\mathbf{x}}_{i(t)}, \tilde{\mathbf{x}}_{i(t+1)}, \mathbf{y}_{i(t)} | Z_{ik} = 1)} \leq \max_{u \in W} \max_{h_2 \in E_k(u), h_{2'} \in E_k(u), h_2 \neq h_{2'}} \frac{\max_{(h_1, h_3)} e_k(u; h_1, h_2, h_3)}{\min_{(h_{1'}, h_{3'}) \in \mathbf{e}_k(u; h_{2'})} e_k(u; h_{1'}, h_{2'}, h_{3'})} \leq \gamma,$$

where  $\mathbf{e}_k(u; h_2) = \{(h_1, h_3) : h_2 = \operatorname{argmax} e_k(u; h_1, h_2, h_3)\}$ . Moreover, we have, if  $\mathbf{y}_{i(t)} \notin W$ ,

$$\frac{p(\tilde{\mathbf{x}}_{i(t-1)}, \mathbf{x}_{i(t)}, \tilde{\mathbf{x}}_{i(t+1)}, \mathbf{y}_{i(t)} | Z_{ik} = 1)}{p(\tilde{\mathbf{x}}_{i(t-1)}, \tilde{\mathbf{x}}_{i(t)}, \tilde{\mathbf{x}}_{i(t+1)}, \mathbf{y}_{i(t)} | Z_{ik} = 1)} \leq \max_{u \notin W} \frac{\max_{h_2} \max_{(h_1, h_3) \notin \mathbf{e}_k(u; h_2)} e_k(u; h_1, h_2, h_3)}{\max_{h_2 \in E_k(u)} \min_{(h_{1'}, h_{3'}) \in \mathbf{e}_k(u; h_{2'})} e_k(u; h_{1'}, h_{2'}, h_{3'})} = \bar{\gamma}.$$

Note that  $\gamma \geq 1$  and  $\bar{\gamma} < 1$ .

LEMMA 4. *We have, for any  $k = 1, \dots, g$ ,*

$$\log p(\mathbf{y}_i | Z_{ik} = 1) \leq \log p(\tilde{\mathbf{x}}_{ik}, \mathbf{y}_i | Z_{ik} = 1) + T \log(\tilde{\gamma} + \bar{\gamma}) + T \hat{\rho}_{k_0} c_2 + \log \tilde{\gamma} + \log \left( 2M\gamma \max_{h, \ell} \frac{\pi k h}{\pi k \ell} \right),$$

where  $c_2 = 1 + \frac{\gamma}{1 + \bar{\gamma}}$  and  $\tilde{\gamma} = \max(2, \gamma)$ .

PROOF OF LEMMA 4. By definition, we have

$$p(\mathbf{y}_i | Z_{ik} = 1) = p(\tilde{\mathbf{x}}_{ik}, \mathbf{y}_i | Z_{ik} = 1) \sum_{\mathbf{x} \in \mathcal{X}} \frac{p(\mathbf{x}, \mathbf{y}_i | Z_{ik} = 1)}{p(\tilde{\mathbf{x}}_{ik}, \mathbf{y}_i | Z_{ik} = 1)}.$$

Let  $B_p(\tilde{\mathbf{x}}_{ik}) = \{\mathbf{x} : \|\mathbf{x} - \tilde{\mathbf{x}}_{ik}\|_0 = p\}$ , then

$$\sum_{\mathbf{x} \in \mathcal{X}} \frac{p(\mathbf{x}, \mathbf{y}_i | Z_{ik} = 1)}{p(\tilde{\mathbf{x}}_{ik}, \mathbf{y}_i | Z_{ik} = 1)} = \sum_{p=0}^{T+1} \sum_{\mathbf{x} \in B_p(\tilde{\mathbf{x}}_{ik})} \frac{p(\mathbf{x}, \mathbf{y}_i | Z_{ik} = 1)}{p(\tilde{\mathbf{x}}_{ik}, \mathbf{y}_i | Z_{ik} = 1)}.$$

Remark that

$$\frac{p(\mathbf{x}_{i(0)}, \mathbf{y}_{i(0)} | Z_{ik} = 1, \mathbf{x}_{i(1)}, \mathbf{y}_{i(1)})}{p(\tilde{\mathbf{x}}_{i(0)}, \mathbf{y}_{i(0)} | Z_{ik} = 1, \tilde{\mathbf{x}}_{i(1)}, \mathbf{y}_{i(1)})} < \gamma \max_{h, \ell} \frac{\pi k h}{\pi k \ell}.$$

Moreover, we observe  $T_W = T \hat{\rho}_{k_0}$  elements of the sequence  $\mathbf{y}_{i(1)}, \dots, \mathbf{y}_{i(T)}$  which belongs to

$W$ . We have

$$\begin{aligned} \sum_{\mathbf{x} \in \mathcal{X}} \frac{p(\mathbf{x}, \mathbf{y}_i | Z_{ik} = 1)}{p(\tilde{\mathbf{x}}_{ik}, \mathbf{y}_i | Z_{ik} = 1)} &\leq \left( M\gamma \max_{h,\ell} \frac{\pi_{kh}}{\pi_{k\ell}} \right) \sum_{p=0}^T \sum_{r=0}^p \binom{T_W}{\min(r, T_W)} \binom{T - T_W}{\min(u, T - T_W)} \gamma^r \bar{\gamma}^u \\ &= \left( M\gamma \max_{h,\ell} \frac{\pi_{kh}}{\pi_{k\ell}} \right) \left( \sum_{r=0}^{T_W} \binom{T_W}{r} \gamma^r \sum_{u=0}^{T-r} \binom{T - T_W}{u} \bar{\gamma}^u \right) \\ &\quad + \sum_{r=1+T_W}^T \sum_{u=0}^{T-r} \binom{T - T_W}{\min(u, T - T_W)} \gamma^r \bar{\gamma}^u. \end{aligned}$$

We have

$$\sum_{r=0}^{T_W} \binom{T_W}{r} \gamma^r \sum_{u=0}^{T-r} \binom{T - T_W}{u} \bar{\gamma}^u = (1 + \bar{\gamma})^T \left( 1 + \frac{\gamma}{1 + \bar{\gamma}} \right)^{T_W},$$

and

$$\sum_{r=1+T_W}^T \sum_{u=0}^{T-r} \binom{T - T_W}{\min(u, T - T_W)} \gamma^r \bar{\gamma}^u \leq (\tilde{\gamma} + \bar{\gamma})^T \tilde{\gamma} \left( \frac{\tilde{\gamma}}{\bar{\gamma} + \tilde{\gamma}} \right)^{T_W},$$

where  $\tilde{\gamma} = \max(2, \gamma)$ . Noting that  $1 + \bar{\gamma} < \tilde{\gamma} + \bar{\gamma}$  and  $1 + \frac{\gamma}{1 + \bar{\gamma}} > \frac{\tilde{\gamma}}{\bar{\gamma} + \tilde{\gamma}}$ , we have

$$\log p(\mathbf{y}_i | Z_{ik} = 1) \leq \log p(\tilde{\mathbf{x}}_{ik}, \mathbf{y}_i | Z_{ik} = 1) + T \log(\tilde{\gamma} + \bar{\gamma}) + T \hat{\rho}_{k_0} c_2 + \log \tilde{\gamma} + \log \left( 2M\gamma \max_{h,\ell} \frac{\pi_{kh}}{\pi_{k\ell}} \right),$$

where  $c_2 = 1 + \frac{\gamma}{1 + \bar{\gamma}}$ . Note that  $\gamma + 1 > c_2 > 1$ . □

### Concentration of the ratio of complete-data likelihood

LEMMA 5. For any  $k \neq k_0$  and for any  $\delta_3$  such that  $-\zeta < \delta_3 < u_{kk_0}$ , we have

$$\mathbb{P}_0 \left( \frac{1}{T} \sum_{t=1}^T \sum_{h=1}^M \sum_{\ell=1}^M v_{i(t)h\ell} \log \left( \frac{A_k[h, \ell]}{A_{k_0}[h, \ell]} \right) > \delta_3 \right) \leq \exp(-Tc_3),$$

where  $c_3 = 2 \frac{1 - \bar{\nu}_2(\mathbf{A}_{k_0})}{1 + \bar{\nu}_2(\mathbf{A}_{k_0})} s^2 > 0$  and  $s = \frac{\delta_3}{\omega_{kk_0}} + \frac{1}{\omega_{kk_0}} \sum_{h=1}^M \sum_{\ell=1}^M \pi_{k_0 h} A_{k_0}[h, \ell] \log \left( \frac{A_{k_0}[h, \ell]}{A_k[h, \ell]} \right)$ .

PROOF OF LEMMA 5. Let  $f(\cdot) \in [0, 1]$  defined by

$$f(\mathbf{v}_{i(t)}) = \frac{1}{\omega_{kk_0}} \left( \sum_{h=1}^M \sum_{\ell=1}^M v_{i(t)h\ell} \log \left( \frac{A_k[h, \ell]}{A_{k_0}[h, \ell]} \right) + u_{kk_0} \right),$$

where  $\omega_{kk_0} = u_{kk_0} + u_{k_0k}$ ,  $u_{kk_0} = \max_{(h,\ell)} \log \left( \frac{A_k[h,\ell]}{A_{k_0}[h,\ell]} \right)$ . Denoting  $\mathbb{E}_0[\cdot] = \mathbb{E}[\cdot \mid Z_{ik_0} = 1]$  the conditional expectation computed with the true parameters, we have, for  $t = 1, \dots, T$ ,

$$\mathbb{E}_0 [f(\mathbf{V}_{i(t)})] = \frac{1}{\omega_{kk_0}} \sum_{h=1}^M \sum_{\ell=1}^M \pi_{k_0h} A_{k_0}[h, \ell] \left( \log \left( \frac{A_k[h, \ell]}{A_{k_0}[h, \ell]} \right) + u_{k_0k} \right).$$

Therefore, we have

$$\begin{aligned} \mathbb{P}_0 \left( \sum_{t=1}^T \sum_{h=1}^M \sum_{\ell=1}^M v_{i(t)h\ell} \log \left( \frac{A_k[h, \ell]}{A_{k_0}[h, \ell]} \right) > \delta_2 \right) &= \mathbb{P}_0 \left( \sum_{t=1}^T f(\mathbf{v}_{i(t)}) > \frac{\delta_2 + T u_{k_0k}}{\omega_{kk_0}} \right) \\ &= \mathbb{P}_0 \left( \sum_{t=1}^T f(\mathbf{v}_{i(t)}) > T(\mathbb{E}[f(\mathbf{V}_{i(1)})] + s) \right), \end{aligned}$$

where  $s = \frac{\delta_2}{T\omega_{kk_0}} + \frac{1}{\omega_{kk_0}} \sum_{h=1}^M \sum_{\ell=1}^M \pi_{k_0h} A_{k_0}[h, \ell] \log \left( \frac{A_{k_0}[h, \ell]}{A_k[h, \ell]} \right)$ .

Note that  $\omega_{kk_0} > 0$  and that, by Assumption 3,  $\sum_{h=1}^M \sum_{\ell=1}^M \pi_{k_0h} A_{k_0}[h, \ell] \log \left( \frac{A_{k_0}[h, \ell]}{A_k[h, \ell]} \right) > \zeta > 0$  because it is a weighted sum of  $M$  Kullback-Leibler divergences. Thus, if  $-T\zeta < \delta_2$  then  $s > 0$ . Moreover, if  $\delta_2 < T u_{k_0k}$ , then  $\mathbb{E}[f(\mathbf{V}_{i(1)})] + s < 1$ . Assumption 1 and Remark 1 imply that  $\bar{\nu}_2(\mathbf{A}_{k_0})$  is the maximum between zero and the second-largest eigenvalue of reversible Markov chain of  $\mathbf{V}_{i(t)}$ . Therefore, using Theorem 2, we have for any  $\delta_3$  such that  $-\zeta < \delta_3 < u_{kk_0}$ ,

$$\mathbb{P}_0 \left( \frac{1}{T} \sum_{t=1}^T \sum_{h=1}^M \sum_{\ell=1}^M v_{i(t)h\ell} \log \left( \frac{A_k[h, \ell]}{A_{k_0}[h, \ell]} \right) > \delta_3 \right) \leq \exp(-T c_3),$$

where  $c_3 = 2 \frac{1 - \bar{\nu}_2(\mathbf{A}_{k_0})}{1 + \bar{\nu}_2(\mathbf{A}_{k_0})} s^2$  and  $s = \frac{\delta_3}{\omega_{kk_0}} + \frac{1}{\omega_{kk_0}} \sum_{h=1}^M \sum_{\ell=1}^M \pi_{k_0h} A_{k_0}[h, \ell] \log \left( \frac{A_{k_0}[h, \ell]}{A_k[h, \ell]} \right)$ .  $\square$

**2.2. Proof of Theorem 2.** Noting that  $\mathbb{P}(Z_{ik} = 1 \mid \mathbf{y}_i) \propto \delta_k p(\mathbf{y}_i \mid Z_{ik} = 1)$  and using Lemma 4, we have

$$\begin{aligned} \mathbb{P}_0 \left( \frac{\mathbb{P}(Z_{ik} = 1 \mid \mathbf{y}_i)}{\mathbb{P}(Z_{ik_0} = 1 \mid \mathbf{y}_i)} > a \right) &\leq \mathbb{P}_0 \left( \log \frac{p(\tilde{\mathbf{x}}_{ik}, \mathbf{y}_i \mid Z_{ik} = 1)}{p(\tilde{\mathbf{x}}_{ik_0}, \mathbf{y}_i \mid Z_{ik_0} = 1)} > -\log \frac{\delta_k}{a \delta_{k_0}} - \log \left( 2M \tilde{\gamma} \gamma \max_{h,\ell} \frac{\pi_{kh}}{\pi_{k\ell}} \right) \right. \\ &\quad \left. - T \log(\tilde{\gamma} + \bar{\gamma}) - T \hat{\rho}_{k_0} c_2 \right). \end{aligned}$$

Moreover,

$$\log \frac{p(\tilde{\mathbf{x}}_{ik}, \mathbf{y}_i \mid Z_{ik} = 1)}{p(\tilde{\mathbf{x}}_{ik_0}, \mathbf{y}_i \mid Z_{ik_0} = 1)} = \sum_{t=1}^T (d_{k1(t)} + d_{k2(t)}) + \sum_{h=1}^M \tilde{x}_{ik(1)h} \log \pi_{kh} - \tilde{x}_{ik_0(1)h} \log \pi_{k_0h},$$

where

$$d_{k1(t)} = \sum_{h=1}^M \sum_{\ell=1}^M (\tilde{v}_{ik(t)h\ell} - \tilde{v}_{ik_0(t)h\ell}) \log (\mathbf{A}_{k_0}[h, \ell] g_\ell(y_{i(t)}; \boldsymbol{\lambda}_\ell, \varepsilon_\ell)),$$



and

$$d_{k2(t)} = \sum_{h=1}^M \sum_{\ell=1}^M \tilde{v}_{ik(t)h\ell} \log \frac{\mathbf{A}_k[h, \ell]}{\mathbf{A}_{k_0}[h, \ell]}.$$

Therefore, we have

$$\mathbb{P}_0 \left( \frac{\mathbb{P}(Z_{ik} = 1 \mid \mathbf{y}_i)}{\mathbb{P}(Z_{ik_0} = 1 \mid \mathbf{y}_i)} > a \right) \leq \mathbb{P}_0 \left( \frac{1}{T} \sum_{t=1}^T (d_{k1(t)} + d_{k2(t)}) > -\frac{c_4}{T} - \log(\tilde{\gamma} + \bar{\gamma}) - \hat{\rho}_{k_0} c_2 \right),$$

with  $c_4 = \log \frac{\delta_k}{a\delta_{k_0}} + \log \left( 2M\tilde{\gamma}\gamma \max_{h,\ell} \frac{\pi_{kh}}{\pi_{k\ell}} \right) + \max_{k,k_0,h,\ell} \log \frac{\pi_{kh}}{\pi_{k_0\ell}}$ . By definition of  $W$ , we have  $\tilde{v}_{ik(t)h\ell} = \tilde{v}_{ik_0(t)h\ell}$  if  $\mathbf{y}_{i(t)} \notin W$ . Moreover, because  $\tilde{\mathbf{v}}_{ik_0}$  is the maximum *a posteriori* rule, if  $\mathbf{y}_{i(t)} \in W$ , then  $d_{k1(t)} < \gamma$ . Therefore, we have

$$\mathbb{P}_0 \left( \frac{\mathbb{P}(Z_{ik} = 1 \mid \mathbf{y}_i)}{\mathbb{P}(Z_{ik_0} = 1 \mid \mathbf{y}_i)} > a \right) \leq \mathbb{P}_0 \left( \frac{1}{T} \sum_{t=1}^T d_{k2(t)} > -\frac{c_4}{T} - (\gamma + c_2)\hat{\rho}_{k_0} - \log(\tilde{\gamma} + \bar{\gamma}) \right).$$

Hence, we have,

$$\mathbb{P}_0 \left( \frac{\mathbb{P}(Z_{ik} = 1 \mid \mathbf{y}_i)}{\mathbb{P}(Z_{ik_0} = 1 \mid \mathbf{y}_i)} > a \right) \leq \mathbb{P}_0(\hat{\rho}_{k_0} > \rho_{k_0} + \delta_1) + \mathbb{P}_0 \left( \frac{1}{T} \sum_{t=1}^T d_{k2(t)} > -\frac{c_4}{T} - \log(\tilde{\gamma} + \bar{\gamma}) - (\gamma + c_2)(\rho_{k_0} + \delta_1) \right).$$

Using Lemma 3, if  $\delta_1 > \frac{1}{\sqrt{2T}}$ , the first term of the right side of the previous equation can be upper bounded by  $e^{-Tc_1}$  with  $c_1 = \frac{1}{2}(\delta_1 - \frac{1}{\sqrt{2T}})^2$ .

Using Lemma 5, the second term of the right-hand side of the previous equation can be upper bounded by  $e^{-Tc_3}$  with  $c_3 = 2 \frac{1-\bar{\nu}_2(\mathbf{A}_{k_0})}{1+\bar{\nu}_2(\mathbf{A}_{k_0})} s^2$ , where  $s = \frac{\delta_3}{\omega_{kk_0}} + \frac{1}{\omega_{kk_0}} \sum_{h=1}^M \sum_{\ell=1}^M \pi_{k_0h} \mathbf{A}_{k_0}[h, \ell] \log \left( \frac{\mathbf{A}_{k_0}[h, \ell]}{\mathbf{A}_k[h, \ell]} \right)$  and  $\delta_3 = -\frac{c_4}{T} - \log(\tilde{\gamma} + \bar{\gamma}) - (\gamma + c_2)(\rho_{k_0} + \delta_1)$ , if  $\delta_3$  is such that  $-\zeta < \delta_3 < u_{kk_0}$ . Thus, we have the following condition on  $\delta_1$

$$\frac{\zeta - \frac{c_4}{T} - \log(\tilde{\gamma} + \bar{\gamma})}{\gamma + c_2} - \rho_{k_0} > \delta_1 > -\frac{u_{kk_0} + \frac{c_4}{T} + \log(\tilde{\gamma} + \bar{\gamma})}{\gamma + c_2} - \rho_{k_0}.$$

Noting that  $\gamma + c_2 < 1 + 2\gamma$ , the previous upper bound can be satisfied under the following assumption

ASSUMPTION 1. *It holds that*

$$\frac{\zeta - c_4 - \log(\tilde{\gamma} + \bar{\gamma})}{1 + 2\gamma} - \rho_{k_0} - \frac{1}{\sqrt{2}} > 0,$$

with  $c_4 = \log \frac{\delta_k}{a\delta_{k_0}} + \log \left( 2M\tilde{\gamma}\gamma \max_{h,\ell} \frac{\pi_{kh}}{\pi_{k\ell}} \right) + \max_{k,k_0,h,\ell} \log \frac{\pi_{kh}}{\pi_{k_0\ell}}$ .

For any  $a$  such that Assumption 1 holds and for any  $\delta_1$  with  $\frac{1}{\sqrt{2T}} < \delta_1 < \frac{\zeta - c_4 - \log(\tilde{\gamma} + \bar{\gamma})}{\gamma + c_2} - \rho_{k_0}$ , we have

$$\mathbb{P}_0(\hat{\rho}_{k_0} > \rho_{k_0} + \delta_1) \leq \mathcal{O}(e^{-Tc_1}),$$

and

$$\mathbb{P}_0\left(\frac{1}{T} \sum_{t=1}^T d_{k2}(t) > \delta_3\right) \leq \mathcal{O}(e^{-Tc_3}),$$

with  $\delta_3 = -\frac{c_4}{T} - \log(\tilde{\gamma} + \bar{\gamma}) - (\gamma + c_2)(\rho_{k_0} + \delta_1)$ . Therefore, there exists  $c > 0$  such that

$$\mathbb{P}_0\left(\frac{\mathbb{P}(Z_{ik} = 1 \mid \mathbf{y}_i)}{\mathbb{P}(Z_{ik_0} = 1 \mid \mathbf{y}_i)} > a\right) \leq \mathcal{O}(e^{-Tc}).$$

If the misclassification error is studied, we should consider  $a = 1$ . Then, a sufficient condition to have the exponential decreasing of the probability of misclassifying an observation is obtained on the basis of Assumption 1 with  $a = 1$ .

### 3. Details about the conditional distribution. Forward formula

$$\alpha_{ikhs(t)}(\boldsymbol{\theta}) = \mathbb{P}(X_{is(t)} = h \mid Z_{ik} = 1; \boldsymbol{\theta}) p(y_{is(0)}, \dots, y_{is(t)} \mid X_{is(t)} = h, Z_{ik} = 1; \boldsymbol{\theta}),$$

which measures the probability of the partial sequence  $y_{is(0)}, \dots, y_{is(t)}$  and ending up in state  $h$  at time  $t$  under component  $k$ . For any  $(i, k, h, s)$ , we can define  $\alpha_{ikhs(t)}$  recursively, as follows,

$$\begin{aligned} \alpha_{ikhs(t)}(\boldsymbol{\theta}) &= \pi_{kh} p(y_{is(t)}; \boldsymbol{\lambda}_h) \\ \forall t \in \{0, \dots, T_{is} - 1\} \quad \alpha_{ikhs(t+1)}(\boldsymbol{\theta}) &= \left( \sum_{h=1}^M A_k[h, \ell] \alpha_{ikhs(t)}(\boldsymbol{\theta}) \right) p(y_{is(t+1)}; \boldsymbol{\lambda}_h). \end{aligned}$$

Considering independence between the  $S_i$  sequences  $\mathbf{y}_{is}$ , the pdf of  $\mathbf{y}_i$  under component  $k$  is

$$p(\mathbf{y}_i \mid Z_{ik} = 1; \boldsymbol{\theta}) = \prod_{s=1}^{S_i} \sum_{h=1}^M \alpha_{ikhs(T_{is})}(\boldsymbol{\theta}).$$

Therefore,

$$p(\mathbf{y}_i; \boldsymbol{\theta}) = \sum_{k=1}^K \delta_k \left( \prod_{s=1}^{S_i} \sum_{h=1}^M \alpha_{ikhs(T_{is})}(\boldsymbol{\theta}) \right).$$

### Backward formula

$$\beta_{ikhs(t)}(\boldsymbol{\theta}) = p(y_{is(t+1)}, \dots, y_{is(T_{is})} \mid X_{is(t)} = h, Z_{ik} = 1; \boldsymbol{\theta}),$$

which measures the probability of the ending partial sequence  $y_{is(t+1)}, \dots, y_{is(T_{is})}$  given a start in state  $h$  at time  $t$  under component  $k$ . We can define  $\beta_{ikhs(t)}(\boldsymbol{\theta})$  recursively, for any  $(i, k, h, s)$ , as

$$\beta_{ikhs(T_{is})}(\boldsymbol{\theta}) = 1$$

$$\forall t \in \{0, \dots, T_{is} - 1\} \quad \beta_{ikhs(t)}(\boldsymbol{\theta}) = \sum_{\ell=1}^M A_k[h, \ell] p(y_{i(t+1)}; \boldsymbol{\lambda}_\ell) \beta_{ik\ell s(t+1)}(\boldsymbol{\theta}).$$

Considering independence between the  $S_i$  sequences  $\mathbf{y}_{is}$ , the pdf of  $\mathbf{y}_i$  under component  $k$  is

$$p(\mathbf{y}_i | Z_{ik} = 1; \boldsymbol{\theta}) = \prod_{s=1}^{S_i} \sum_{h=1}^M \pi_{kh} \beta_{ikhs(0)}(\boldsymbol{\theta}) p(y_{i(0)}; \boldsymbol{\lambda}_h).$$

$$p(\mathbf{y}_i; \boldsymbol{\theta}) = \sum_{k=1}^K \delta_k \left( \prod_{s=1}^{S_i} \sum_{h=1}^M \pi_{kh} \beta_{ikhs(0)}(\boldsymbol{\theta}) p(y_{i(0)}; \boldsymbol{\lambda}_h) \right).$$

#### 4. Additional simulation results.

##### 4.1. Additional tables for the analysis of simulated data.

TABLE 1

Convergence of the estimators with 1000 replicates: ARI between estimated and true partition, ARI between estimated and true latent states and MSE between the MLE and the true parameters

| $n$   | $T$ | Adjusted Rand index |        |                | Mean square error |       |       |            |
|---|-----|---------------------|--------|----------------|-------------------|-------|-------|------------|
|   |     | partition           | states | $\mathbf{A}_k$ | $\varepsilon_h$   | $a_h$ | $b_h$ | $\delta_k$ |
| Case hard ( $e = 0.75$ and $a_2 = 3$ )        |     |                     |        |                |                   |       |       |            |
| 10  | 100 | 0.812               | 0.350  | 0.097          | 0.002             | 0.285 | 0.085 | 0.057      |
| 10  | 500 | 1.000               | 0.385  | 0.031          | 0.000             | 0.077 | 0.024 | 0.052      |
| 100   | 100 | 0.870               | 0.377  | 0.032          | 0.000             | 0.087 | 0.028 | 0.007      |
| 100   | 500 | 1.000               | 0.391  | 0.018          | 0.000             | 0.045 | 0.017 | 0.005      |
| Case medium-easy ( $e = 0.75$ and $a_2 = 5$ ) |     |                     |        |                |                   |       |       |            |
| 10  | 100 | 0.996               | 0.661  | 0.018          | 0.001             | 0.253 | 0.031 | 0.052      |
| 10  | 500 | 0.999               | 0.669  | 0.004          | 0.000             | 0.050 | 0.006 | 0.052      |
| 100   | 100 | 0.998               | 0.669  | 0.002          | 0.000             | 0.027 | 0.003 | 0.005      |
| 100   | 500 | 1.000               | 0.671  | 0.000          | 0.000             | 0.005 | 0.001 | 0.005      |
| Case easy ( $e = 0.90$ and $a_2 = 5$ )        |     |                     |        |                |                   |       |       |            |
| 10  | 100 | 1.000               | 0.832  | 0.006          | 0.000             | 0.159 | 0.017 | 0.048      |
| 10  | 500 | 1.000               | 0.839  | 0.001          | 0.000             | 0.030 | 0.003 | 0.046      |
| 100   | 100 | 1.000               | 0.836  | 0.001          | 0.000             | 0.015 | 0.002 | 0.005      |
| 100   | 500 | 1.000               | 0.839  | 0.000          | 0.000             | 0.003 | 0.000 | 0.005      |

TABLE 2

*Convergence of the estimators obtained on 1000 replicates with and without missing data when data are sampled from case hard: ARI between estimated and true partition, ARI between estimated and true latent states and MSE between the MLE and the true parameters*

| $n$ | $T$ | missingness    | Adjusted Rand index |        |                | Mean square error |       |       |            |
|-----|-----|----------------|---------------------|--------|----------------|-------------------|-------|-------|------------|
|     |     |                | partition           | states | $\mathbf{A}_k$ | $\varepsilon_h$   | $a_h$ | $b_h$ | $\delta_k$ |
| 10  | 100 | no missingness | 0.812               | 0.350  | 0.097          | 0.002             | 0.285 | 0.085 | 0.057      |
|     |     | MCAR-1         | 0.753               | 0.342  | 0.103          | 0.002             | 0.360 | 0.092 | 0.061      |
|     |     | MCAR-2         | 0.690               | 0.333  | 0.120          | 0.003             | 0.320 | 0.091 | 0.068      |
|     |     | MNAR           | 0.542               | 0.303  | 0.152          | 0.004             | 0.456 | 0.109 | 0.080      |
| 10  | 500 | no missingness | 1.000               | 0.385  | 0.031          | 0.000             | 0.077 | 0.024 | 0.052      |
|     |     | MCAR-1         | 1.000               | 0.386  | 0.029          | 0.000             | 0.072 | 0.024 | 0.052      |
|     |     | MCAR-2         | 0.999               | 0.385  | 0.032          | 0.000             | 0.074 | 0.024 | 0.052      |
|     |     | MNAR           | 0.991               | 0.347  | 0.042          | 0.003             | 0.177 | 0.076 | 0.052      |
| 100 | 100 | no missingness | 0.870               | 0.377  | 0.032          | 0.000             | 0.087 | 0.028 | 0.007      |
|     |     | MCAR-1         | 0.837               | 0.373  | 0.033          | 0.000             | 0.091 | 0.030 | 0.008      |
|     |     | MCAR-2         | 0.801               | 0.369  | 0.035          | 0.000             | 0.099 | 0.032 | 0.008      |
|     |     | MNAR           | 0.664               | 0.337  | 0.038          | 0.003             | 0.099 | 0.083 | 0.011      |
| 100 | 500 | no missingness | 1.000               | 0.391  | 0.018          | 0.000             | 0.045 | 0.017 | 0.005      |
|     |     | MCAR-1         | 1.000               | 0.391  | 0.018          | 0.000             | 0.050 | 0.017 | 0.005      |
|     |     | MCAR-2         | 1.000               | 0.391  | 0.018          | 0.000             | 0.044 | 0.017 | 0.005      |
|     |     | MNAR           | 0.996               | 0.355  | 0.023          | 0.003             | 0.109 | 0.073 | 0.005      |

TABLE 3

*Convergence of the estimators obtained on 1000 replicates with and without missing data when data are sampled from case medium-easy: ARI between estimated and true partition, ARI between estimated and true latent states and MSE between the MLE and the true parameters*

| $n$ | $T$ | missingness    | Adjusted Rand index |        |                | Mean square error |       |       |            |
|-----|-----|----------------|---------------------|--------|----------------|-------------------|-------|-------|------------|
|     |     |                | partition           | states | $\mathbf{A}_k$ | $\varepsilon_h$   | $a_h$ | $b_h$ | $\delta_k$ |
| 10  | 100 | no missingness | 0.996               | 0.661  | 0.018          | 0.001             | 0.253 | 0.031 | 0.052      |
|     |     | MCAR-1         | 0.993               | 0.659  | 0.020          | 0.001             | 0.276 | 0.034 | 0.052      |
|     |     | MCAR-2         | 0.986               | 0.657  | 0.023          | 0.001             | 0.308 | 0.038 | 0.051      |
|     |     | MNAR           | 0.966               | 0.621  | 0.052          | 0.004             | 0.561 | 0.100 | 0.052      |
| 10  | 500 | no missingness | 0.999               | 0.669  | 0.004          | 0.000             | 0.050 | 0.006 | 0.052      |
|     |     | MCAR-1         | 0.999               | 0.669  | 0.004          | 0.000             | 0.051 | 0.006 | 0.052      |
|     |     | MCAR-2         | 0.999               | 0.668  | 0.004          | 0.000             | 0.052 | 0.007 | 0.052      |
|     |     | MNAR           | 1.000               | 0.636  | 0.021          | 0.003             | 0.217 | 0.085 | 0.052      |
| 100 | 100 | no missingness | 0.998               | 0.669  | 0.002          | 0.000             | 0.027 | 0.003 | 0.005      |
|     |     | MCAR-1         | 0.996               | 0.668  | 0.002          | 0.000             | 0.030 | 0.003 | 0.005      |
|     |     | MCAR-2         | 0.993               | 0.667  | 0.002          | 0.000             | 0.033 | 0.004 | 0.005      |
|     |     | MNAR           | 0.978               | 0.641  | 0.013          | 0.003             | 0.160 | 0.074 | 0.005      |
| 100 | 500 | no missingness | 1.000               | 0.671  | 0.000          | 0.000             | 0.005 | 0.001 | 0.005      |
|     |     | MCAR-1         | 1.000               | 0.671  | 0.000          | 0.000             | 0.005 | 0.001 | 0.005      |
|     |     | MCAR-2         | 1.000               | 0.671  | 0.000          | 0.000             | 0.005 | 0.001 | 0.005      |
|     |     | MNAR           | 1.000               | 0.644  | 0.011          | 0.003             | 0.140 | 0.076 | 0.005      |

TABLE 4

Convergence of the estimators obtained on 1000 replicates with and without missing data when data are sampled from case easy: ARI between estimated and true partition, ARI between estimated and true latent states and MSE between the MLE and the true parameters

| $n$ | $T$ | missingness    | Adjusted Rand index |        |       | Mean square error |       |       |            |
|-----|-----|----------------|---------------------|--------|-------|-------------------|-------|-------|------------|
|     |     |                | partition           | states | $A_k$ | $\varepsilon_h$   | $a_h$ | $b_h$ | $\delta_k$ |
| 10  | 100 | no missingness | 1.000               | 0.832  | 0.006 | 0.000             | 0.159 | 0.017 | 0.048      |
|     |     | MCAR-1         | 1.000               | 0.829  | 0.007 | 0.000             | 0.185 | 0.019 | 0.048      |
|     |     | MCAR-2         | 1.000               | 0.826  | 0.008 | 0.001             | 0.207 | 0.022 | 0.048      |
|     |     | MNAR           | 1.000               | 0.771  | 0.017 | 0.003             | 0.315 | 0.042 | 0.048      |
| 10  | 500 | no missingness | 1.000               | 0.839  | 0.001 | 0.000             | 0.030 | 0.003 | 0.046      |
|     |     | MCAR-1         | 1.000               | 0.838  | 0.001 | 0.000             | 0.031 | 0.003 | 0.046      |
|     |     | MCAR-2         | 1.000               | 0.837  | 0.001 | 0.000             | 0.031 | 0.003 | 0.046      |
|     |     | MNAR           | 1.000               | 0.781  | 0.007 | 0.003             | 0.119 | 0.030 | 0.046      |
| 100 | 100 | no missingness | 1.000               | 0.836  | 0.001 | 0.000             | 0.015 | 0.002 | 0.005      |
|     |     | MCAR-1         | 1.000               | 0.833  | 0.001 | 0.000             | 0.018 | 0.002 | 0.005      |
|     |     | MCAR-2         | 1.000               | 0.831  | 0.001 | 0.000             | 0.019 | 0.002 | 0.005      |
|     |     | MNAR           | 1.000               | 0.779  | 0.006 | 0.003             | 0.084 | 0.026 | 0.005      |
| 100 | 500 | no missingness | 1.000               | 0.839  | 0.000 | 0.000             | 0.003 | 0.000 | 0.005      |
|     |     | MCAR-1         | 1.000               | 0.838  | 0.000 | 0.000             | 0.003 | 0.000 | 0.005      |
|     |     | MCAR-2         | 1.000               | 0.838  | 0.000 | 0.000             | 0.003 | 0.000 | 0.005      |
|     |     | MNAR           | 1.000               | 0.783  | 0.005 | 0.003             | 0.072 | 0.025 | 0.005      |

#### 4.2. Additional figures for the analysis of the data from Huang et al. (2018).

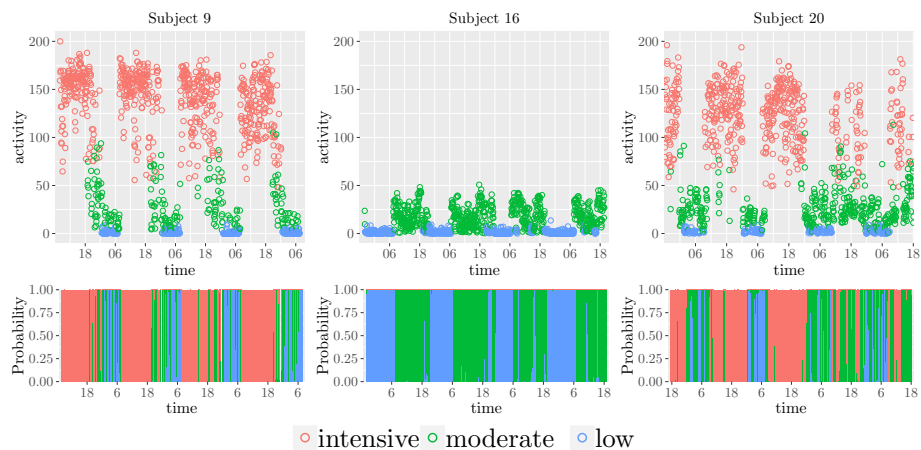


FIG 1. State estimation for the three subjects: (top) accelerometer data where color indicates the expected value of  $Y_{i(t)}$  conditionally the most likely state and on the most likely component; (bottom) probability of each state at each time.

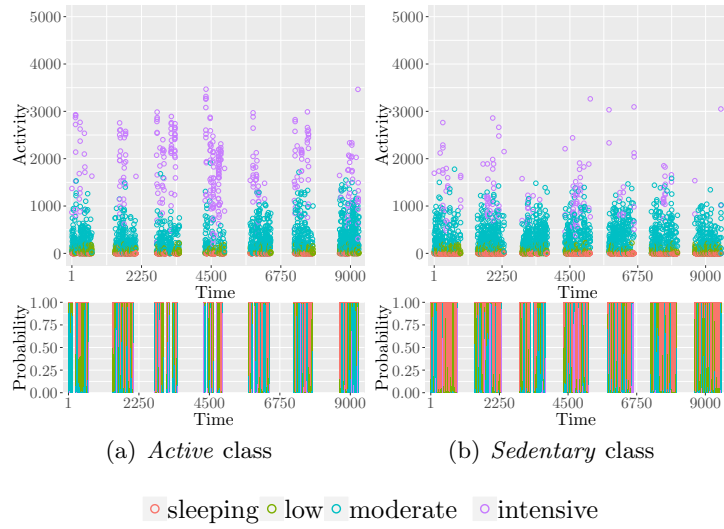


FIG 3. Examples of observation assigned into the five classes with the probabilities of the states.

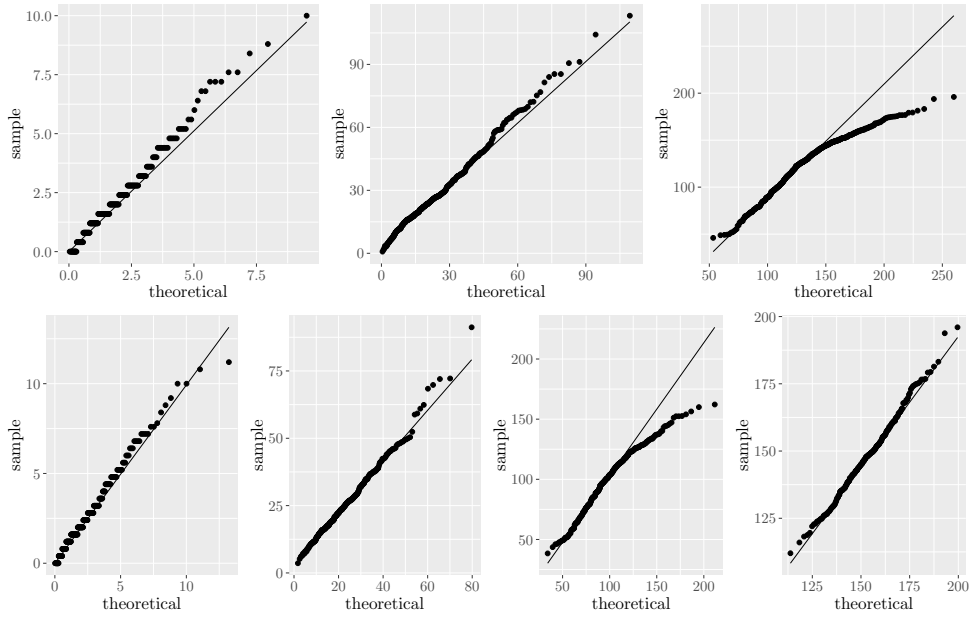


FIG 2. QQ-plots of subject 20. Top: from left to right, state 1 to 3. Bottom: Top: from left to right, state 1 to 4.

4.3. Additional figures for the analysis of the PAT data.

TABLE 5  
*Transition matrix for the two remaining classes*

|                 | Class <i>active</i>    |           |                |                 |
|-----------------|------------------------|-----------|----------------|-----------------|
|                 | sleeping               | low-level | moderate-level | intensive-level |
| sleeping        | 0.87                   | 0.12      | 0.01           | 0.00            |
| low-level       | 0.17                   | 0.73      | 0.10           | 0.00            |
| moderate-level  | 0.04                   | 0.30      | 0.66           | 0.01            |
| intensive-level | 0.08                   | 0.08      | 0.18           | 0.66            |
|                 | Class <i>sedentary</i> |           |                |                 |
|                 | sleeping               | low-level | moderate-level | intensive-level |
| sleeping        | 0.79                   | 0.16      | 0.05           | 0.00            |
| low-level       | 0.17                   | 0.66      | 0.16           | 0.01            |
| moderate-level  | 0.05                   | 0.14      | 0.79           | 0.03            |
| intensive-level | 0.01                   | 0.02      | 0.15           | 0.82            |

## References.

- Huang, Q., Cohen, D., Komarzynski, S., Li, X.-M., Innominato, P., Lévi, F., and Finkenstädt, B. (2018). Hidden markov models for monitoring circadian rhythmicity in telemetric activity data. *J. Royal Soc. Interface*, 15(139):20170885.
- Kontorovich, A. and Weiss, R. (2014). Uniform Chernoff and Dvoretzky-Kiefer-Wolfowitz-type inequalities for Markov chains and related processes. *J. Appl. Probab. Stat.*, 51(4):1100–1113.
- León, C. A. and Perron, F. (2004). Optimal Hoeffding bounds for discrete reversible Markov chains. *Ann. Appl. Probab.*, 14(2):958–970.
- Teicher, H. (1963). Identifiability of finite mixtures. *Ann. Math. Stat.*, pages 1265–1269.