Dynamic matrix-variate clustering of sport activities

CMStatistics 2020
Online

Stival Mattia - mattia.stival@phd.unipd.it
Bernardi Mauro - mauro.bernardi@unipd.it

Department of Statistical Sciences
University of Padova

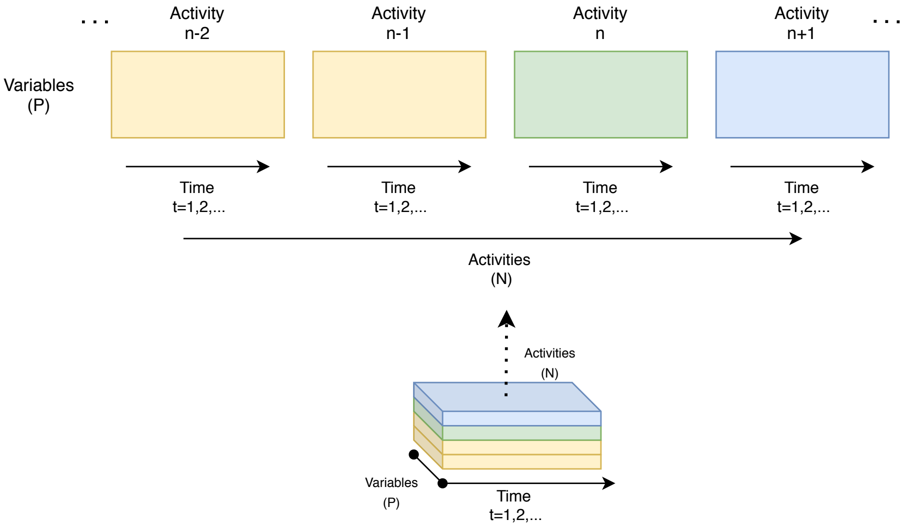
20 December, 2020

Introduction
Sport statistics and performance analysis

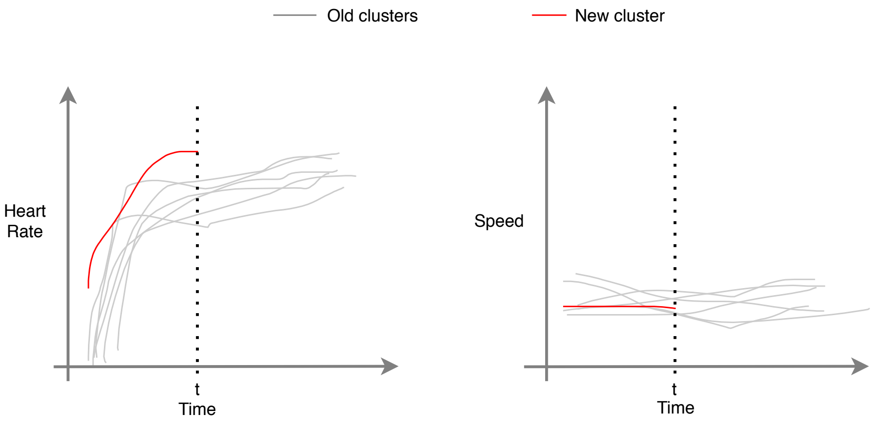
Performance analysis

- The monitoring of sport activities and the analysis of sport events are topics of increasing interest in several disciplines such as biology, medicine, statistics, engineering, material science, and mathematics.
- Improving the knowledge and individualizing the design of training activities allow to maximize the improvements, and avoid over-training, which may lead to impaired health, and typically under-performance (see, e.g., Cardinale and Varley, 2017).
- The use of GPS-enabled tracking devices and heart rate monitors is common in several sports, such as running, swimming, and cycling (Frick and Kosmidis, 2017).
- Cardinale and Varley (2017) focused on the validity and reliability of the use of such data, highlighting the importance of analysing them individually, for each athlete.

Performance analysis as a clustering problem



Performance analysis as a clustering problem



Notation and starting assumptions

- Let $y_{p,n,t}$ be a scalar denoting the observed value of the p–th variable, for the n–th activity, at time t, for $p = 1, \ldots, P$, $n = 1, \ldots, N$, and $t = 1, \ldots, T$.
- We assume that the N activities can be clustered into G different groups.
- Suppose to know that the n–th activity belongs to the group g, then its observations over time follow a dynamic linear model of the form

$$y_{p,n,t} = \mu_{p,t}^{(g)} + \upsilon_{p,n,t} = \sum_{g=1}^{G} \mathbb{I}(S_n = g)\mu_{p,t}^{(g)} + \upsilon_{p,n,t},$$

for $t = 1, \ldots, T$, $p = 1, \ldots, P$, some specifications of the trend $\mu_{p,t}^{(g)}$, and serially uncorrelated error terms.

Notation and starting assumptions

- $\boldsymbol{\mu}_t^{(g)} = \begin{bmatrix} \mu_{1,t}^{(g)} & \mu_{2,t}^{(g)} & \dots & \mu_{P,t}^{(g)} \end{bmatrix}^\intercal$ stores the trends at time t shared by activities belonging to the group g, for the P different observed variables.
- The membership of the n-th activity to the g-th group does not change over time.
- Activities belonging to the same group share the same mean trajectory, described by $\mathbf{M}_{1:T}^{(g)} = \begin{bmatrix} \boldsymbol{\mu}_1^{(g)} & \boldsymbol{\mu}_2^{(g)} & \dots & \boldsymbol{\mu}_T^{(g)} \end{bmatrix}$.
- The trends of one variable in different groups follow the same structural relations, but these relations change for different variables.

Example

Suppose to observe $P = 2$ variables for N activities. One simple model for G groups can be described by the following relations

$$\texttt{First variable} \begin{cases} y_{1,n,t} & = \sum_{g=1}^{G} \mathbb{I}(S_n = g)\mu_{1,t}^{(g)} + \upsilon_{1,n,t} \\ \mu_{1,t+1}^{(g)} & = \mu_{1,t}^{(g)} + \epsilon_{1,t}^{(g)}, \end{cases}$$

$$\texttt{Second variable} \begin{cases} y_{2,n,t} & = \sum_{g=1}^{G} \mathbb{I}(S_n = g)\mu_{2,t}^{(g)} + \upsilon_{2,n,t} \\ \mu_{2,t+1}^{(g)} & = \mu_{2,t}^{(g)} + \beta_{2,t}^{(g)} + \epsilon_{2,t}^{(g)} \\ \beta_{2,t+1}^{(g)} & = \beta_{2,t}^{(g)} + \eta_{2,t}^{(g)}, \end{cases}$$

for given $\mu_{1,1}^{(g)}$, $\mu_{2,1}^{(g)}$, $\beta_{2,1}^{(g)}$,$g = 1, \dots, G$, $t = 1, \dots, T$, and $n = 1, \dots, N$.

Model representations: matrix-variate and vectorized form

After some matrix algebra tricks, the model can be represented in:

- matrix-variate form

$$\mathbf{Y}_t = \mathbf{Z}\mathbf{A}_t\mathbf{S}^\top + \boldsymbol{\Upsilon}_t, \quad \boldsymbol{\Upsilon}_t \sim \mathrm{MN}_{P,N}(\mathbf{0}, \boldsymbol{\Sigma}^R \otimes \boldsymbol{\Sigma}^C),$$
$$\mathbf{A}_{t+1} = \mathbf{T}\mathbf{A}_t + \boldsymbol{\Xi}_t, \quad \boldsymbol{\Xi}_t \sim \mathrm{MN}_{Q,G}(\mathbf{0}, \boldsymbol{\Psi}^R \otimes \boldsymbol{\Psi}^C),$$

  for some starting value $\mathbf{A}_1$.

- vectorized form

$$\mathbf{y}_t = (\mathbf{S} \otimes \mathbf{Z})\boldsymbol{\alpha}_t + \boldsymbol{\upsilon}_t, \quad \boldsymbol{\upsilon}_t \sim \mathrm{MVN}_{PN}(\mathbf{0}, \boldsymbol{\Sigma}^C \otimes \boldsymbol{\Sigma}^R),$$
$$\boldsymbol{\alpha}_{t+1} = (\mathbf{I}_G \otimes \mathbf{T})\boldsymbol{\alpha}_t + \boldsymbol{\xi}_t, \quad \boldsymbol{\xi}_t \sim \mathrm{MVN}_{QG}(\mathbf{0}, \boldsymbol{\Psi}^C \otimes \boldsymbol{\Psi}^R),$$

  with $\mathbf{y}_t = \mathrm{vec}(\mathbf{Y}_t)$, $\boldsymbol{\alpha}_t = \mathrm{vec}(\mathbf{A}_t)$, $\boldsymbol{\upsilon}_t = \mathrm{vec}(\boldsymbol{\Upsilon}_t)$, and $\boldsymbol{\xi}_t = \mathrm{vec}(\boldsymbol{\Xi}_t)$.

Matrix-variate form

The model in matrix-variate form

$$\begin{aligned}
\boxed{\mathbf{Y}_t} &= \mathbf{Z}\mathbf{A}_t\mathbf{S}^\intercal + \boldsymbol{\Upsilon}_t, \quad & \boldsymbol{\Upsilon}_t &\sim \mathrm{MN}_{P,N}(\mathbf{0}, \boldsymbol{\Sigma}^R \otimes \boldsymbol{\Sigma}^C),\\
\mathbf{A}_{t+1} &= \mathbf{T}\mathbf{A}_t + \boldsymbol{\Xi}_t, \quad & \boldsymbol{\Xi}_t &\sim \mathrm{MN}_{Q,G}(\mathbf{0}, \boldsymbol{\Psi}^R \otimes \boldsymbol{\Psi}^C),
\end{aligned}$$

with $\mathbf{A}_1 \sim \mathrm{MN}_{Q,G}(\hat{\mathbf{A}}_{1|0}, \mathbf{P}^R_{1|0} \otimes \mathbf{P}^C_{1|0})$.

$\boxed{\mathbf{Y}_t}$ is a matrix of dimension $P \times N$ that stores the observations of the P variables for the N activities, for all $t = 1, \ldots, T$.

Matrix-variate form

The model in matrix-variate form

$$\mathbf{Y}_t = \boxed{\mathbf{Z}}\, \mathbf{A}_t \mathbf{S}^\intercal + \boldsymbol{\Upsilon}_t, \qquad \boldsymbol{\Upsilon}_t \sim \mathrm{MN}_{P,N}(\mathbf{0}, \boldsymbol{\Sigma}^R \otimes \boldsymbol{\Sigma}^C)$$
$$\mathbf{A}_{t+1} = \boxed{\mathbf{T}}\, \mathbf{A}_t + \boldsymbol{\Xi}_t, \qquad \boldsymbol{\Xi}_t \sim \mathrm{MN}_{Q,G}(\mathbf{0}, \boldsymbol{\Psi}^R \otimes \boldsymbol{\Psi}^C)$$

with $\mathbf{A}_1 \sim \mathrm{MN}_{Q,G}(\hat{\mathbf{A}}_{1|0}, \mathbf{P}^R_{1|0} \otimes \mathbf{P}^C_{1|0})$.

$\boxed{\mathbf{Z}}$ is a $P \times Q$ design matrix, where Q denotes the number of latent states for each group, while $\boxed{\mathbf{T}}$ is a $Q \times Q$ transition matrix.

Working on the specification of $\boxed{\mathbf{Z}}$ and $\boxed{\mathbf{T}}$ allows us to consider different dynamics for different variables.

Matrix-variate form

The model in matrix-variate form

$$\mathbf{Y}_t = \mathbf{Z}\boxed{\mathbf{A}_t}\mathbf{S}^\intercal + \boldsymbol{\Upsilon}_t, \qquad \boldsymbol{\Upsilon}_t \sim \mathrm{MN}_{P,N}(\mathbf{0}, \boldsymbol{\Sigma}^R \otimes \boldsymbol{\Sigma}^C),$$
$$\mathbf{A}_{t+1} = \mathbf{T}\mathbf{A}_t + \boldsymbol{\Xi}_t, \qquad \boldsymbol{\Xi}_t \sim \mathrm{MN}_{Q,G}(\mathbf{0}, \boldsymbol{\Psi}^R \otimes \boldsymbol{\Psi}^C),$$

with $\mathbf{A}_1 \sim \mathrm{MN}_{Q,G}(\hat{\mathbf{A}}_{1|0}, \mathbf{P}^R_{1|0} \otimes \mathbf{P}^C_{1|0})$.

$\boxed{\mathbf{A}_t}$ is a $Q \times G$ matrix of latent states, where the g–th column stores the Q latent states of the g-th group.

Matrix-variate form

The model in matrix-variate form

$$\begin{aligned} \mathbf{Y}_t &= \mathbf{Z}\mathbf{A}_t \boxed{\mathbf{S}^\intercal} + \boldsymbol{\Upsilon}_t, & \boldsymbol{\Upsilon}_t &\sim \mathrm{MN}_{P,N}(\mathbf{0}, \boldsymbol{\Sigma}^R \otimes \boldsymbol{\Sigma}^C), \\ \mathbf{A}_{t+1} &= \mathbf{T}\mathbf{A}_t + \boldsymbol{\Xi}_t, & \boldsymbol{\Xi}_t &\sim \mathrm{MN}_{Q,G}(\mathbf{0}, \boldsymbol{\Psi}^R \otimes \boldsymbol{\Psi}^C), \end{aligned}$$

with $\mathbf{A}_1 \sim \mathrm{MN}_{Q,G}(\hat{\mathbf{A}}_{1|0}, \mathbf{P}^R_{1|0} \otimes \mathbf{P}^C_{1|0})$.

$\boxed{\mathbf{S}}$ is a (latent random) selection matrix of dimension $N \times G$, such that

$$\mathbf{S} = \begin{bmatrix} \mathbb{I}(S_1 = 1) & \mathbb{I}(S_1 = 2) & \dots & \mathbb{I}(S_1 = G) \\ \mathbb{I}(S_2 = 1) & \mathbb{I}(S_2 = 2) & \dots & \mathbb{I}(S_2 = G) \\ \vdots & \vdots & \dots & \vdots \\ \mathbb{I}(S_N = 1) & \mathbb{I}(S_N = 2) & \dots & \mathbb{I}(S_N = G) \end{bmatrix}.$$

The n-th row has value 1 in the g-th column if the n-th activity belongs to the g-th group, and zeros elsewhere.

Matrix-variate form

The model in matrix-variate form

$$
\begin{aligned}
\mathbf{Y}_t &= \mathbf{Z}\mathbf{A}_t\mathbf{S}^\intercal + \boldsymbol{\Upsilon}_t, \qquad \boxed{\boldsymbol{\Upsilon}_t} \sim \mathrm{MN}_{P,N}(\mathbf{0}, \boldsymbol{\Sigma}^R \otimes \boldsymbol{\Sigma}^C), \\
\mathbf{A}_{t+1} &= \mathbf{T}\mathbf{A}_t + \boldsymbol{\Xi}_t, \qquad \boxed{\boldsymbol{\Xi}_t} \sim \mathrm{MN}_{Q,G}(\mathbf{0}, \boldsymbol{\Psi}^R \otimes \boldsymbol{\Psi}^C),
\end{aligned}
$$

with $\boxed{\mathbf{A}_1} \sim \mathrm{MN}_{Q,G}(\hat{\mathbf{A}}_{1|0}, \mathbf{P}^R_{1|0} \otimes \mathbf{P}^C_{1|0})$.

The elements $\boxed{\boldsymbol{\Upsilon}_t}$, $\boxed{\boldsymbol{\Xi}_t}$, and $\boxed{\mathbf{A}_1}$ follow the matrix-variate normal distribution, with covariance matrix that can be decomposed by a Kronecker product (see, e.g., Gupta and Nagar, 2000).

If $\mathbf{S}$ is supposed to be known, the model can be intended as a special case of a general *matrix-variate state space model* of the form

$$\mathbf{Y}_t = \mathbf{C}_t + \mathbf{Z}_t \mathbf{A}_t \mathbf{S}_t^\intercal + \boldsymbol{\Upsilon}_t, \quad \boldsymbol{\Upsilon}_t \sim \mathrm{MN}_{P,N}(\mathbf{0}, \boldsymbol{\Sigma}^R \otimes \boldsymbol{\Sigma}^C), \tag{1}$$
$$\mathbf{A}_{t+1} = \mathbf{D}_t + \mathbf{T}_t \mathbf{A}_t \mathbf{U}_t^\intercal + \boldsymbol{\Xi}_t, \quad \boldsymbol{\Xi}_t \sim \mathrm{MN}_{Q,G}(\mathbf{0}, \boldsymbol{\Psi}^R \otimes \boldsymbol{\Psi}^C), \tag{2}$$

with $\mathbf{A}_1 \sim \mathrm{MN}_{Q,G}(\hat{\mathbf{A}}_{1|0}, \mathbf{P}_{1|0}^R \otimes \mathbf{P}_{1|0}^C)$, $\mathbf{Z}_t$ $(P \times Q)$ and $\mathbf{T}_t$ $(Q \times Q)$ known left design matrices, $\mathbf{S}_t$ $(N \times G)$ and $\mathbf{U}_t$ $(G \times G)$ known right design matrices, $\mathbf{C}_t$ and $\mathbf{D}_t$ known terms.

This specification includes many models proposed in literature for dealing with matrix-variate time series: West and Harrison (1997), Choukroun et al. (2006), Wang and West (2009), Wang et al. (2019), Chen et al. (2020), ...

Main results

Main results

Consider the model in matrix-variate form

$$
\begin{aligned}
\mathbf{Y}_t &= \mathbf{Z}\mathbf{A}_t\mathbf{S}^\intercal + \boldsymbol{\Upsilon}_t, \quad & \boldsymbol{\Upsilon}_t &\sim \mathrm{MN}_{P,N}(\mathbf{0}, \boxed{\boldsymbol{\Sigma}^R \otimes \boldsymbol{\Sigma}^C}), \\
\mathbf{A}_{t+1} &= \mathbf{T}\mathbf{A}_t + \boldsymbol{\Xi}_t, \quad & \boldsymbol{\Xi}_t &\sim \mathrm{MN}_{Q,G}(\mathbf{0}, \boxed{\boldsymbol{\Psi}^R \otimes \boldsymbol{\Psi}^C}),
\end{aligned}
$$

with $\mathbf{A}_1 \sim \mathrm{MN}_{Q,G}(\hat{\mathbf{A}}_{1|0}, \mathbf{P}^R_{1|0} \otimes \mathbf{P}^C_{1|0})$.

We will deal with:

1. Identified prior for Kronecker product decomposable covariance matrices;

Main results

Consider the model in matrix-variate form

$$\mathbf{Y}_t = \mathbf{Z}\boxed{\mathbf{A}_t}\mathbf{S}^\intercal + \boldsymbol{\Upsilon}_t, \qquad \boldsymbol{\Upsilon}_t \sim \mathrm{MN}_{P,N}(\mathbf{0}, \boldsymbol{\Sigma}^R \otimes \boldsymbol{\Sigma}^C),$$
$$\mathbf{A}_{t+1} = \mathbf{T}\mathbf{A}_t + \boldsymbol{\Xi}_t, \qquad \boldsymbol{\Xi}_t \sim \mathrm{MN}_{Q,G}(\mathbf{0}, \boldsymbol{\Psi}^R \otimes \boldsymbol{\Psi}^C),$$

with $\mathbf{A}_1 \sim \mathrm{MN}_{Q,G}(\hat{\mathbf{A}}_{1|0}, \mathbf{P}^R_{1|0} \otimes \mathbf{P}^C_{1|0})$.

We will deal with:

1. Identified prior for Kronecker product decomposable covariance matrices;
2. Efficient Bayesian estimation of the states $\boxed{\mathbf{A}_{1:T}}$;

Main results

Consider the model in matrix-variate form

$$\begin{aligned}\mathbf{Y}_t &= \mathbf{Z}\mathbf{A}_t \boxed{\mathbf{S}^\top} + \boldsymbol{\Upsilon}_t, & \boldsymbol{\Upsilon}_t &\sim \mathrm{MN}_{P,N}(\mathbf{0}, \boldsymbol{\Sigma}^R \otimes \boldsymbol{\Sigma}^C),\\ \mathbf{A}_{t+1} &= \mathbf{T}\mathbf{A}_t + \boldsymbol{\Xi}_t, & \boldsymbol{\Xi}_t &\sim \mathrm{MN}_{Q,G}(\mathbf{0}, \boldsymbol{\Psi}^R \otimes \boldsymbol{\Psi}^C),\end{aligned}$$

with $\mathbf{A}_1 \sim \mathrm{MN}_{Q,G}(\hat{\mathbf{A}}_{1|0}, \mathbf{P}^R_{1|0} \otimes \mathbf{P}^C_{1|0})$.

We will deal with:

1. Identified prior for Kronecker product decomposable covariance matrices;
2. Efficient Bayesian estimation of the states $\mathbf{A}_{1:T}$;
3. Efficient Bayesian estimation of $\boxed{\mathbf{S}}$.

Efficient Bayesian states estimation

States estimation

- Consider the model in the vectorized form

$$
\begin{aligned}
\mathbf{y}_t &= (\mathbf{S} \otimes \mathbf{Z})\boldsymbol{\alpha}_t + \boldsymbol{\upsilon}_t, \quad \boldsymbol{\upsilon}_t \sim \text{MVN}_{PN}(\mathbf{0}, \boldsymbol{\Sigma}^C \otimes \boldsymbol{\Sigma}^R), \\
\boldsymbol{\alpha}_{t+1} &= (\mathbf{I}_G \otimes \mathbf{T})\boldsymbol{\alpha}_t + \boldsymbol{\xi}_t, \quad \boldsymbol{\xi}_t \sim \text{MVN}_{QG}(\mathbf{0}, \boldsymbol{\Psi}^C \otimes \boldsymbol{\Psi}^R),
\end{aligned}
$$

  with $\boldsymbol{\alpha}_1 \sim \text{MVN}_{QG}(\hat{\boldsymbol{\alpha}}_{1|0}, \mathbf{P}^C_{1|0} \otimes \mathbf{P}^R_{1|0})$.

- Conditionally on $\mathbf{S}$ and the model parameters, the estimation of the states can be obtained with standard routines for Bayesian state space models (see, e.g., Durbin and Koopman, 2012).

- Kalman filter becomes burdensome if N (and P) is large, since at each time point t it is required to invert a $NP \times NP$ matrix.

Reduction by transformation

- Consider the vector state space

$$
\begin{aligned}
\mathbf{y}_t &= \mathbf{Z}_{\text{vec}}\boldsymbol{\alpha}_t + \boldsymbol{\upsilon}_t, \quad \boldsymbol{\upsilon}_t \sim \text{MVN}_M(\mathbf{0}, \boldsymbol{\Sigma}),\\
\boldsymbol{\alpha}_{t+1} &= \mathbf{T}_{\text{vec}}\boldsymbol{\alpha}_t + \boldsymbol{\xi}_t, \quad \boldsymbol{\xi}_t \sim \text{MVN}_S(\mathbf{0}, \boldsymbol{\Psi}),
\end{aligned}
$$

with $\boldsymbol{\alpha}_1 \sim \text{MVN}_S(\hat{\boldsymbol{\alpha}}_{1|0}, \mathbf{P}_{1|0})$.

Reduction by transformation

- Let $\mathbf{A}$ a $M \times M$ dimensional matrix

$$\mathbf{A} = \begin{bmatrix} \underset{m \times M}{\mathbf{A}^L} \\ \underset{(M-m) \times M}{\mathbf{A}^H} \end{bmatrix},$$

such that

$$\mathbf{y}_t^L = \mathbf{A}^L \mathbf{y}_t = \mathbf{A}^L \mathbf{Z}_{\text{vec}} \boldsymbol{\alpha}_t + \boldsymbol{\upsilon}_t^L, \quad \mathbf{y}_t^H = \mathbf{A}^H \mathbf{y}_t = \boldsymbol{\upsilon}_t^H,$$
$$\boldsymbol{\upsilon}_t^L = \mathbf{A}^L \boldsymbol{\upsilon}_t \sim \text{MVN}_m(\mathbf{0}, \mathbf{A}^L \boldsymbol{\Sigma} \mathbf{A}^{L\intercal}), \quad \boldsymbol{\upsilon}_t^H = \mathbf{A}^H \boldsymbol{\upsilon}_t \sim \text{MVN}_{M-m}(\mathbf{0}, \mathbf{A}^H \boldsymbol{\Sigma} \mathbf{A}^{H\intercal}),$$
$$\text{E}(\boldsymbol{\upsilon}_t^H \boldsymbol{\upsilon}_t^{L\intercal}) = \mathbf{0}.$$

- Let $\mathbf{A}$ a $M \times M$ dimensional matrix

$$\mathbf{A} = \begin{bmatrix} \overset{m\times M}{\mathbf{A}^L} \\ \overset{(M-m)\times M}{\mathbf{A}^H} \end{bmatrix},$$

such that

$$\boxed{\mathbf{y}_t^L = \mathbf{A}^L \mathbf{y}_t = \mathbf{A}^L \mathbf{Z}_{\text{vec}} \boldsymbol{\alpha}_t + \boldsymbol{\upsilon}_t^L}, \quad \mathbf{y}_t^H = \mathbf{A}^H \mathbf{y}_t = \boldsymbol{\upsilon}_t^H,$$
$$\boldsymbol{\upsilon}_t^L = \mathbf{A}^L \boldsymbol{\upsilon}_t \sim \text{MVN}_m(\mathbf{0}, \boldsymbol{\Sigma}^L), \quad \boldsymbol{\upsilon}_t^H = \mathbf{A}^H \boldsymbol{\upsilon}_t \sim \text{MVN}_{M-m}(\mathbf{0}, \boldsymbol{\Sigma}^H),$$
$$\text{E}(\boldsymbol{\upsilon}_t^H \boldsymbol{\upsilon}_t^{L\top}) = \mathbf{0}.$$

- By finding a suitable matrix $\mathbf{A}$, the state estimation can be obtained by considering a reduced form of the model without loss of information.

Reduction by transformation

- Jungbacker and Koopman (2008) provide both technical conditions and practical solutions to obtain $\mathbf{A}^L$. One possibility consists of considering

$$\mathbf{A}^L = \mathbf{Z}_\bullet^\intercal \boldsymbol{\Sigma}^{-1},$$

  where the columns of $\mathbf{Z}_\bullet (M \times m)$ form a basis of the column space of $\mathbf{Z}$.

- They show that it is not required to know $\mathbf{A}^H$ to estimate the states and to obtain the model likelihood.

- Suppose $\mathbf{Z}_{\text{vec}}$ is not full column rank. Then $\mathbf{Z}_\bullet$ can be obtained from the decomposition $\mathbf{Z}_{\text{vec}} = \mathbf{Z}_\bullet \mathbf{C}$, for any full rank $m \times S$ matrix $\mathbf{C}$ with $m \leqslant S$.

Reduction by transformation

- With the model in the vectorized form,

$$\mathbf{Z}_{\text{vec}} = (\mathbf{S} \otimes \mathbf{Z}) = (\mathbf{S}\mathbf{I}_G) \otimes (\mathbf{I}_P\mathbf{Z}) = (\mathbf{S} \otimes \mathbf{I}_P)(\mathbf{I}_G \otimes \mathbf{Z}),$$

hence

$$\mathbf{Z}_{\bullet} = \mathbf{S} \otimes \mathbf{I}_P \quad \text{and} \quad \mathbf{C} = \mathbf{I}_G \otimes \mathbf{Z},$$

both of rank GP, if $\text{rank}(\mathbf{S}) = G$ and $\text{rank}(\mathbf{Z}) = P$.

- In the Kalman routines, we can use an $m = GP$-dimensional vector of observations instead of the original $M = NP$-dimensional vector.

Efficient Bayesian estimation of $\mathbf{S}$

Estimation of cluster allocation

- Consider the model

$$\begin{aligned} \mathbf{Y}_t &= \mathbf{Z}\mathbf{A}_t\mathbf{S}^\intercal + \boldsymbol{\Upsilon}_t, \quad & \boldsymbol{\Upsilon}_t &\sim \mathrm{MN}_{P,N}(\mathbf{0}, \boldsymbol{\Sigma}^R \otimes \boldsymbol{\Sigma}^C), \\ \mathbf{A}_{t+1} &= \mathbf{T}\mathbf{A}_t + \boldsymbol{\Xi}_t, \quad & \boldsymbol{\Xi}_t &\sim \mathrm{MN}_{Q,G}(\mathbf{0}, \boldsymbol{\Psi}^R \otimes \boldsymbol{\Psi}^C), \end{aligned}$$

with $\mathbf{A}_1 \sim \mathrm{MN}_{Q,G}(\hat{\mathbf{A}}_{1|0}, \mathbf{P}^R_{1|0} \otimes \mathbf{P}^C_{1|0})$.

Estimation of cluster allocation

- Consider the model

$$\mathbf{Y}_t = \mathbf{Z}\mathbf{A}_t \boxed{\mathbf{S}^\top} + \boldsymbol{\Upsilon}_t, \quad \boldsymbol{\Upsilon}_t \sim \mathrm{MN}_{P,N}(\mathbf{0}, \boldsymbol{\Sigma}^R \otimes \boldsymbol{\Sigma}^C),$$
$$\mathbf{A}_{t+1} = \mathbf{T}\mathbf{A}_t + \boldsymbol{\Xi}_t, \quad \boldsymbol{\Xi}_t \sim \mathrm{MN}_{Q,G}(\mathbf{0}, \boldsymbol{\Psi}^R \otimes \boldsymbol{\Psi}^C),$$

with $\mathbf{A}_1 \sim \mathrm{MN}_{Q,G}(\hat{\mathbf{A}}_{1|0}, \mathbf{P}^R_{1|0} \otimes \mathbf{P}^C_{1|0})$.

1. In this form, the model can be considered already an augmented version, since $\boxed{\mathbf{S}}$ is a latent random selection matrix, storing cluster allocations.

Estimation of cluster allocation

- Consider the model

$$\mathbf{Y}_t = \mathbf{Z}\mathbf{A}_t\mathbf{S}^\intercal + \boldsymbol{\Upsilon}_t, \quad \boldsymbol{\Upsilon}_t \sim \mathrm{MN}_{P,N}(\mathbf{0}, \boldsymbol{\Sigma}^R \otimes \boxed{\boldsymbol{\Sigma}^C}),$$
$$\mathbf{A}_{t+1} = \mathbf{T}\mathbf{A}_t + \boldsymbol{\Xi}_t, \quad \boldsymbol{\Xi}_t \sim \mathrm{MN}_{Q,G}(\mathbf{0}, \boldsymbol{\Psi}^R \otimes \boldsymbol{\Psi}^C),$$

with $\mathbf{A}_1 \sim \mathrm{MN}_{Q,G}(\hat{\mathbf{A}}_{1|0}, \mathbf{P}^R_{1|0} \otimes \mathbf{P}^C_{1|0})$.

1. In this form, the model can be considered already an augmented version, since $\mathbf{S}$ is a latent random selection matrix, storing cluster allocations.
2. Conditionally on the states, activities are dependent if $\boxed{\boldsymbol{\Sigma}^C}$ is not diagonal. Furthermore, $\boldsymbol{\Sigma}^C$ might depend on $\mathbf{S}$ (e.g., $\boldsymbol{\Sigma}^C = \mathbf{S}\boldsymbol{\Sigma}^G\mathbf{S}^\intercal + \boldsymbol{\Sigma}^I$).

Estimation of cluster allocation

- Consider the model

$$\mathbf{Y}_t = \mathbf{Z}\mathbf{A}_t\mathbf{S}^\intercal + \boldsymbol{\Upsilon}_t, \quad \boldsymbol{\Upsilon}_t \sim \mathrm{MN}_{P,N}(\mathbf{0}, \boldsymbol{\Sigma}^R \otimes \boldsymbol{\Sigma}^C),$$
$$\mathbf{A}_{t+1} = \mathbf{T}\mathbf{A}_t + \boldsymbol{\Xi}_t, \qquad \boldsymbol{\Xi}_t \sim \mathrm{MN}_{Q,G}(\mathbf{0}, \boldsymbol{\Psi}^R \otimes \boxed{\boldsymbol{\Psi}^C}),$$

with $\mathbf{A}_1 \sim \mathrm{MN}_{Q,G}(\hat{\mathbf{A}}_{1|0}, \mathbf{P}^R_{1|0} \otimes \mathbf{P}^C_{1|0})$.

1. In this form, the model can be considered already an augmented version, since $\mathbf{S}$ is a latent random selection matrix, storing cluster allocations.
2. Conditionally to the states, activities are dependent if $\boldsymbol{\Sigma}^C$ is not diagonal. Furthermore, $\boldsymbol{\Sigma}^C$ might depend on $\mathbf{S}$ (e.g., $\boldsymbol{\Sigma}^C = \mathbf{S}\boldsymbol{\Sigma}^G\mathbf{S}^\intercal + \boldsymbol{\Sigma}^I$).
3. States of different groups are dependent a priori if $\boxed{\boldsymbol{\Psi}^C}$ is not diagonal.

Estimation of cluster allocation

- Let $\mathbf{\Theta} = (\mathbf{\Sigma}^R, \mathbf{\Sigma}^C, \mathbf{\Psi}^R, \mathbf{\Psi}^C)$. The augmented posterior distribution can be written as

$$p(\mathbf{A}_{1:T}, \mathbf{\Theta}, \mathbf{S}, \boldsymbol{\pi} \mid \mathbf{Y}_{1:T}) \propto p(\mathbf{Y}_{1:T} \mid \mathbf{A}_{1:T}, \mathbf{\Theta}, \mathbf{S}) p(\mathbf{A}_{1:T} \mid \mathbf{\Theta}) \\ p(\mathbf{\Theta} \mid \mathbf{S}) p(\mathbf{S} \mid \boldsymbol{\pi}) p(\boldsymbol{\pi}).$$

- Adopting a fully conjugate approach, it is easy to derive a Gibbs sampler.
- $\mathbf{S}$ can be updated by taking each single row conditionally to the others (and the rest), where the states $\mathbf{A}_{1:T}$ can be marginalized out by use of the Kalman filter.
- Updating the whole $\mathbf{S}$ marginalizing out the states requires the use of $N(G-1)$ Kalman filters.

Proposal distribution for $\mathbf{S}$

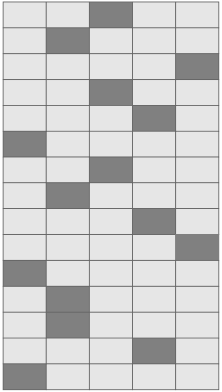
- We propose to simulate $\mathbf{S}$ in one-shot, via a Metropolis Hasting step. Let $\mathbf{S}^\star$ be the proposal for $\mathbf{S}$ at iteration it, and let $\mathbf{S}^{(it-1)}$ be $\mathbf{S}$ at the previous iteration.
- Given $\mathbf{S}^{(it-1)}$, each row $\mathbf{S}_n^\star = \begin{bmatrix} \mathbb{I}(S_n^\star = 1) & \cdots & \mathbb{I}(S_n^\star = G) \end{bmatrix}$ of $\mathbf{S}^\star$ is drawn independently on the others with probabilities

$$Pr(S_n^\star = k \mid S_n^{(it-1)}) = \frac{e^{\beta \mathbb{I}(S_n^{(it-1)} = k)}}{\sum_{g=1}^G e^{\beta \mathbb{I}(S_n^{(it-1)} = g)}}, \tag{3}$$

for $k = 1, \dots, G$, and β tuning parameter.

Proposal distribution for $\mathbf{S}$

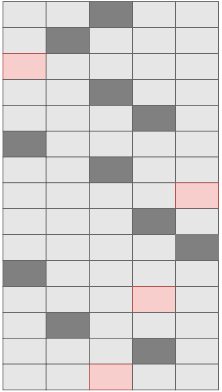
$\mathbf{S}^{(it-1)}$



$\beta = \log(4)$

0.125	0.125	0.5	0.125	0.125
0.125	0.5	0.125	0.125	0.125
0.125	0.125	0.125	0.125	0.5
0.125	0.125	0.5	0.125	0.125
0.125	0.125	0.125	0.5	0.125
0.5	0.125	0.125	0.125	0.125
0.125	0.125	0.5	0.125	0.125
0.125	0.5	0.125	0.125	0.125
0.125	0.125	0.125	0.5	0.125
0.125	0.125	0.125	0.125	0.5
0.5	0.125	0.125	0.125	0.125
0.125	0.5	0.125	0.125	0.125
0.125	0.5	0.125	0.125	0.125
0.125	0.125	0.125	0.5	0.125
0.5	0.125	0.125	0.125	0.125

$\mathbf{S}^\star$



Metropolis-Hasting step

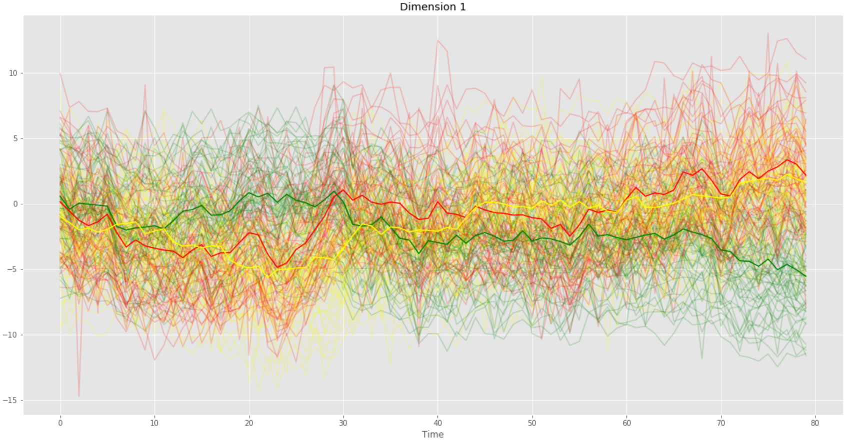
- We note that $\Pr(\mathbf{S}^\star \mid \mathbf{S}^{(it-1)}) = \Pr(\mathbf{S}^{(it-1)} \mid \mathbf{S}^\star)$.
- Then, we take $\mathbf{S}^{(it)} = \mathbf{S}^\star$ with probability

$$\rho(\mathbf{S}^\star, \mathbf{S}^{(it-1)}) = \min\left(\frac{p(\mathbf{Y}_{1:T} \mid \mathbf{S}^\star, \boldsymbol{\Theta})p(\boldsymbol{\Theta} \mid \mathbf{S}^\star)p(\mathbf{S}^\star \mid \boldsymbol{\pi})}{p(\mathbf{Y}_{1:T} \mid \mathbf{S}^{(it-1)}, \boldsymbol{\Theta})p(\boldsymbol{\Theta} \mid \mathbf{S}^{(it-1)})p(\mathbf{S}^{(it-1)} \mid \boldsymbol{\pi})}, 1\right),$$

  and $\mathbf{S}^{(it)} = \mathbf{S}^{(it-1)}$ with probability $1 - \rho(\mathbf{S}^\star, \mathbf{S}^{(it-1)})$.
- Obtaining $p(\mathbf{Y}_{1:T} \mid \mathbf{S}^\star, \boldsymbol{\Theta})$ and $p(\mathbf{Y}_{1:T} \mid \mathbf{S}^{(it-1)}, \boldsymbol{\Theta})$ requires the use of 2 Kalman filters, much less than $N(G-1)$.
- The move can be considered as a new move to include in the Allocation Sampler by Nobile and Fearnside (2007), that allows to change the unknown number of group G.

Examples & Conclusions

Synthetic data I



Example with running data

Consider $\mathbf{Y}_t = \begin{bmatrix} \mathbf{y}_t^{\cdot 1} & \dots & \mathbf{y}_t^{\cdot N} \end{bmatrix}$ with

$$\mathbf{y}_t^{\cdot n} = \begin{bmatrix} \texttt{HeartRate(bmp)}_{n,t} \\ \texttt{Speed(m/s)}_{n,t} \\ \texttt{Cadence(spm)}_{n,t} \end{bmatrix}$$

and $\mathbf{X}_t = \operatorname{diag}(x_{1,t}, x_{2,t}, \dots, x_{N,t})$, with

$$x_{n,t} = \texttt{Altitude(m)}_{n,t} - \texttt{Altitude(m)}_{n,t-1},$$

the data of N continuous running activities for one athlete.

Example with running data

- Consider the model

$$\mathbf{Y}_t = \mathbf{B}\mathbf{S}^\intercal\mathbf{X}_t + \mathbf{Z}\mathbf{A}_t\mathbf{S}^\intercal + \boldsymbol{\Upsilon}_t, \quad \boldsymbol{\Upsilon}_t \sim \mathrm{MN}_{P,N}(\mathbf{0}, \boldsymbol{\Sigma}^R \otimes \boldsymbol{\Sigma}^C),$$
$$\mathbf{A}_{t+1} = \mathbf{T}\mathbf{A}_t + \boldsymbol{\Xi}_t, \quad \boldsymbol{\Xi}_t \sim \mathrm{MN}_{Q,G}(\mathbf{0}, \boldsymbol{\Psi}^R \otimes \boldsymbol{\Psi}^C),$$

  with $\mathbf{A}_1 \sim \mathrm{MN}_{Q,G}(\hat{\mathbf{A}}_{1|0}, \mathbf{P}^R_{1|0} \otimes \mathbf{P}^C_{1|0})$, $\mathbf{B} \sim \mathrm{MN}_{P,G}(\hat{\mathbf{B}}_0, \boldsymbol{\Sigma}^R_{B,0} \otimes \boldsymbol{\Sigma}^C_{B,0})$.

- The model identifies group of activities that requires similar effort, after controlling for variation in altitude, which might impact on the outcome $\mathbf{Y}_t$.

- In the real data example, $P = 3$, $N = 90$, $G = 3$, $Q = 4$, and $T = 600$ seconds.

Example with running data

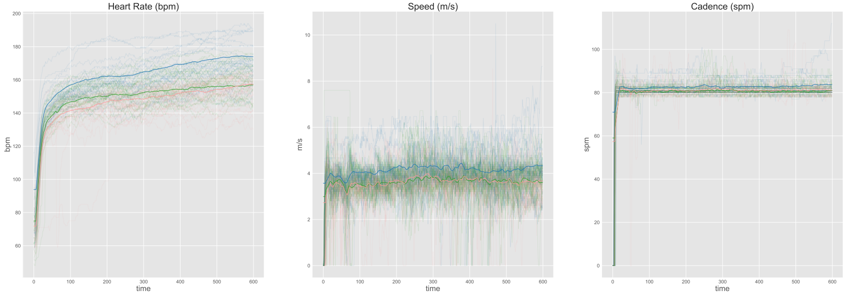


Figure: Running activities clustered according to MAP (background) and posterior mean of the signals $\mathbf{M}_{1:T} = \mathbf{Z}\mathbf{A}_{1:T}$ (thicker line).

Example with running data

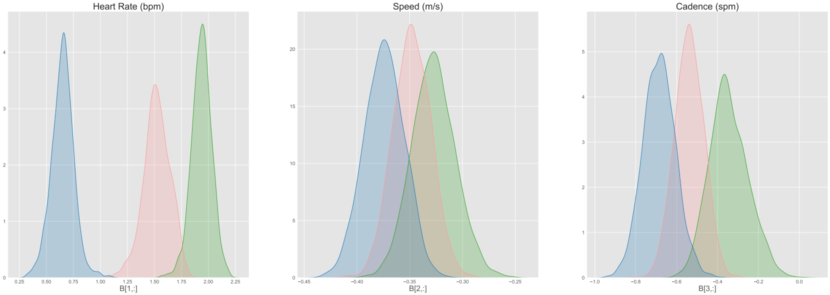


Figure: The effect of 1 meter variation in Altitude on Heart Rate, Speed and Cadence for the different groups.

Conclusions

- **On the model**

1. The model is developed for unsupervised classification of a large panel of multivariate time series by means of few common latent states (but it could also be used for the "partially" unsupervised case).
2. Working on the specification of $\mathbf{\Sigma}^C$, $\mathbf{\Psi}^C$, and $\mathbf{S}$ allows to include particular kinds of dependence across groups and time series.
3. The state space specification of the model is thought for dealing with streams of data.
4. The use of matrix-variate normal distributions allows to derive easily a framework in which new time series (activities) are observed conditional on past observations.
5. *Ad hoc* modifications of the algorithm can be derived for dealing with missing values on the matrix of observations.

Conclusions

- **On the number of groups**
  1. If the number of groups is known, the model performances in term of accuracy are good.
  2. If the number of groups is unknown, an estimate can be obtained by comparing different models with some information criteria, or by considering a prior distribution on the number of groups (Allocation Sampler, Nobile and Fearnside, 2007).
- **On the online learning methodology**
  1. As only few time series are observed simultaneously, the online learning methodology should consider these series conditionally on the past activities.
  2. The obtained MCMC sample of the posterior distribution may be used as a "new" prior for new time series... *But it is still a work in progress...*

Some useful reference by arguments

Performance analysis

- Cardinale, M. and Varley, M.C. (2017) Wearable training-monitoring technology: applications, challenges, and opportunities. *International Journal of Sports Physiology and Performance*, **12**(s2), 55 – 62.
- Frick, H. and Kosmidis, I. (2017) trackeR: Infrastructure for Running and Cycling Data from GPS-Enabled Tracking Devices in R. *Journal of Statistical Software*, **82**(7), 1 – 29.
- Santos-Fernandez, E., Wu, P. and Mengersen, K. L. (2019) Bayesian statistics meets sports: a comprehensive review. *Journal of Quantitative Analysis in Sports* **15**(4), 289–312.

Matrix time series

- Bai, J. and Wang, P. (2015) Identification and Bayesian estimation of dynamic factor models. *J. Bus. Econom. Statist.* **33**(2), 221–240.
- Chen, E. Y., Tsay, R. S. and Chen, R. (2020) Constrained factor models for high-dimensional matrix-variate time series. *J. Amer. Statist. Assoc.* **115**(530), 775–793.
- Jungbacker, B. and Koopman, S.J. (2008) Likelihood-based analysis for dynamic factor models. Discussion paper Vrije Universiteit, Amsterdam.
- Wang, D., Liu, X. and Chen, R. (2019) Factor models for matrix-valued high-dimensional time series. *Journal of econometrics* **208**(1), 231–248.
- Wang, H. and West, M. (2009) Bayesian analysis of matrix normal graphical models. *Biometrika* **96**(4), 821–834.

Some useful reference by arguments

Matrix-variate regression and envelope models

- Ding, S. and Cook, R. D. (2018) Matrix variate regressions and envelope models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80**(2), 387–408.
- Hoff, P. D. (2015) Multilinear tensor regression for longitudinal relational data. *Ann. Appl. Stat.* **9**(3), 1169–1193.
- Li, L. and Zhang, X. (2017) Parsimonious tensor response regression. *J. Amer. Statist. Assoc.* **112**(519), 1131–1146.
- Viroli, C. (2012) On matrix-variate regression analysis. *J. Multivariate Anal.* **111**, 296–309.

Some useful reference by arguments

Others

- McCulloch, R.E., Polson, N.G. and Rossi, P.E. (2000) A Bayesian analysis of the multinomial probit model with fully identified parameters. *Journal of econometrics*, **99**(1), 173 – 193.
- Nobile, A. and Fearnside, A. T. (2007). Bayesian finite mixtures with an unknown number of components: the allocation sampler. *Stat. Comput.*, **17**(2):147–162.
- Titsias, M. K. and Yau, C. (2017) The Hamming ball sampler. *J. Amer. Statist. Assoc.* **112**(520), 1598–1611.
- Viroli, C. (2011) Finite mixtures of matrix normal distributions for classifying three-way data. *Stat. Comput.* **21**(4), 511–522.
- Zanella, G. (2020) Informed proposals for local MCMC in discrete spaces. *J. Amer. Statist. Assoc.* **115**(530), 852–865.

Thanks for your attention!