Different types of Markovian models and applications in swimming and climbing

V. Barbu[1], E. Kalligeris[1,2], G. Hacques[2], J. Komar[2], I. Mavrogiannis[1], L. Seifert[2], and N. Vergne[1]

[1]Laboratoire de Mathématiques Raphaël Salem, Université de Rouen Normandie
[2]Laboratoire CETAPS, Université de Rouen Normandie

Summer school "ComplexSportsData", 26-29 June 2023, Caen

Sequence of random variables

```
gggcggcgacctcgcgggttttcgctatttatgaaattttccggtttaaggcgtttccgttcttcttcgt
cataacttaatgtttttatttaaaataccctctgaaaagaaaggaaacgacaggtgctgaaagcgaggctt
tttggcctctgtcgtttcctttctctgtttttgtccgtggaatgaacaatggaagtcaacaaaaagcagct
ggctgacattttcggtgcgagtatccgtaccattcagaactggcaggaacagggaatgcccgttctgcgag
gcggtggcaagggtaatgaggtgctttatgactctgccgccgtcataaaatggtatgccgaaagggatgct
gaaattgagaacgaaaagctgcgccgggaggttgaagaactgcggcaggccagcgaggcagatctccagcc
aggaactattgagtacgaacgccatcgacttacgcgtgcgcaggccgacgcacaggaactgaagaatgcca
gagactccgctgaagtggtggaaaccgcattctgtactttcgtgctgtcgcggatcgcaggtgaaattgcc
agtattctcgacgggctccccctgtcggtgcagcggcgttttccggaactggaaaaccgacatgttgattt
cctgaaacgggatatcatcaaagccatgaacaaagcagccgcgctggatgaactgataccggggttgctga
gtgaatatatcgaacagtcaggttaacaggctgcggcattttgtccgcgccgggcttcgctcactgttcag
gccggagccacagaccgccgttgaatgggcggatgctaattactatctcccgaaagaatccgcataccagg
aagggcgctgggaaacactgccctttcagcgggccatcatgaatgcgatgggcagcgactacatccgtgag
gtgaatgtggtgaagtctgcccgtgtcggttattccaaaatgctgctgggtgtttatgcctactttataga
```

Markov model

- Markov property (order 1) : $\mathbb{P}(X_t \mid X_u, u < t) = \mathbb{P}(X_t \mid X_{t-1})$
- Markov model : an initial distribution μ_0 and a transition matrix Π
- For $\mathscr{A} = \{a, c, g, t\}$, we have at order 1 :

$$\mu_0 = \begin{pmatrix} \mu_0(a) & \mu_0(c) & \mu_0(g) & \mu_0(t) \end{pmatrix}$$

$$\Pi = \begin{pmatrix} \pi_{aa} & \pi_{ac} & \pi_{ag} & \pi_{at} \\ \pi_{ca} & \pi_{cc} & \pi_{cg} & \pi_{ct} \\ \pi_{ga} & \pi_{gc} & \pi_{gg} & \pi_{gt} \\ \pi_{ta} & \pi_{tc} & \pi_{tg} & \pi_{tt} \end{pmatrix}$$

where $\pi_{uv} = \mathbb{P}(X_t = v \mid X_{t-1} = u)$, with $u \in \mathscr{A}$ and $v \in \mathscr{A}$ and where we can choose for example 0.25 for $\mu_0(u)$, $\forall u \in \mathscr{A}$.

Plan

Plan

Definition (linear drifting Markov chain of order 1 and of length n)

A sequence $X_0, X_1, \ldots, X_n$ with state space $E = \{1, 2, \ldots, s\}$ is called a *linear drifting Markov chain (of order 1)* of length n between the Markov transition matrices $\mathbf{\Pi}_0$ and $\mathbf{\Pi}_1$, if the distribution of X_t, $t = 1, \ldots, n$, is defined by

$$\mathbb{P}(X_t = v \mid X_{t-1} = u, X_{t-2}, \ldots) = \Pi_{\frac{t}{n}}(u, v), \; u, v \in E,$$

$$\text{where } \Pi_{\frac{t}{n}}(u, v) = \left(1 - \frac{t}{n}\right) \Pi_0(u, v) + \frac{t}{n} \Pi_1(u, v), \; u, v \in E.$$

Lemma

For a linear drifting Markov chain and $k_1 \leq k_2, k_1, k_2 \in \mathbb{N}$, *we have*

$$\mathbb{P}(X_{k_2} = j \mid X_{k_1 - 1} = i) = \left(\prod_{l=k_1}^{k_2} \mathbf{\Pi}_{\frac{l}{n}} \right) (i, j).$$

Definition (polynomial drifting Markov chain of order k and of length n)

A sequence $X_0, X_1, \ldots, X_n$ with state space $E = \{1, 2, \ldots, s\}$ is said to be a *polynomial drifting Markov chain of order k* and of length n if the distribution of X_t, $t = 1, \ldots, n$, is defined for $u_1, \ldots, u_k, v \in E$ by

$$\mathbb{P}(X_t = v \mid X_{t-1} = u_k, X_{t-2} = u_{k-1}, \ldots) = \mathbf{\Pi}_{\frac{t}{n}}(u_1, \ldots, u_k, v)$$

$$\text{where } \mathbf{\Pi}_{\frac{t}{n}}(u_1, \ldots, u_k, v) = \sum_{i=0}^{d} A_i(t)\mathbf{\Pi}_{\frac{i}{d}}(u_1, \ldots, u_k, v)$$

with A_i polynomials of degree d such as, for any $i, j \in \{0, 1, \ldots, d\}$, $A_i\left(\frac{nj}{d}\right) = \mathbb{1}_{\{i=j\}}$. Note : for $t = ni/d$, we have $\mathbf{\Pi}_{\frac{t}{n}} = \mathbf{\Pi}_{\frac{i}{d}}$.

Remark (A_i are Lagrange polynomials : chosen to have stochastic $\mathbf{\Pi}_{\frac{t}{n}}$.)

Indeed $\displaystyle\sum_{v \in E} \mathbf{\Pi}_{\frac{t}{n}}(u_1, \ldots, u_k, v) = \sum_{i=0}^{d} A_i(t) = A(t)$ *where* A *is a polynomial of degree* d *equal to one in* $(d+1)$ *points ; then* A *is a constant polynomial equal to one.*

Plan

One sample path : linear drifting Markov model

Proposition (least square estimators of $\Pi_0(u,v)$ and $\Pi_1(u,v)$)

$$\widehat{\Pi}_{0;n}(u,v) = \frac{\begin{pmatrix} \left(\sum_{t=1}^{n} \mathbb{1}_{\{X_{t-1}=u\}} \left(\frac{t}{n}\right)^2\right) \left(\sum_{t=1}^{n} \mathbb{1}_{\{X_{t-1}=u; X_t=v\}} \left(1-\frac{t}{n}\right)\right) - \\ \left(\sum_{t=1}^{n} \mathbb{1}_{\{X_{t-1}=u\}} \left(1-\frac{t}{n}\right)\left(\frac{t}{n}\right)\right) \left(\sum_{t=1}^{n} \mathbb{1}_{\{X_{t-1}=u; X_t=v\}} \left(\frac{t}{n}\right)\right) \end{pmatrix}}{\begin{pmatrix} \left(\sum_{t=1}^{n} \mathbb{1}_{\{X_{t-1}=u\}} \left(1-\frac{t}{n}\right)^2\right) \left(\sum_{t=1}^{n} \mathbb{1}_{\{X_{t-1}=u\}} \left(\frac{t}{n}\right)^2\right) - \\ \left(\sum_{t=1}^{n} \mathbb{1}_{\{X_{t-1}=u\}} \left(1-\frac{t}{n}\right)\left(\frac{t}{n}\right)\right) \left(\sum_{t=1}^{n} \mathbb{1}_{\{X_{t-1}=u\}} \left(1-\frac{t}{n}\right)\left(\frac{t}{n}\right)\right) \end{pmatrix}}$$

$$\widehat{\Pi}_{1;n}(u,v) = \frac{\begin{pmatrix} \left(\sum_{t=1}^{n} \mathbb{1}_{\{X_{t-1}=u\}} \left(1-\frac{t}{n}\right)^2\right) \left(\sum_{t=1}^{n} \mathbb{1}_{\{X_{t-1}=u; X_t=v\}} \left(\frac{t}{n}\right)\right) - \\ \left(\sum_{t=1}^{n} \mathbb{1}_{\{X_{t-1}=u\}} \left(1-\frac{t}{n}\right)\left(\frac{t}{n}\right)\right) \left(\sum_{t=1}^{n} \mathbb{1}_{\{X_{t-1}=u; X_t=v\}} \left(1-\frac{t}{n}\right)\right) \end{pmatrix}}{\begin{pmatrix} \left(\sum_{t=1}^{n} \mathbb{1}_{\{X_{t-1}=u\}} \left(1-\frac{t}{n}\right)^2\right) \left(\sum_{t=1}^{n} \mathbb{1}_{\{X_{t-1}=u\}} \left(\frac{t}{n}\right)^2\right) - \\ \left(\sum_{t=1}^{n} \mathbb{1}_{\{X_{t-1}=u\}} \left(1-\frac{t}{n}\right)\left(\frac{t}{n}\right)\right) \left(\sum_{t=1}^{n} \mathbb{1}_{\{X_{t-1}=u\}} \left(1-\frac{t}{n}\right)\left(\frac{t}{n}\right)\right) \end{pmatrix}}$$

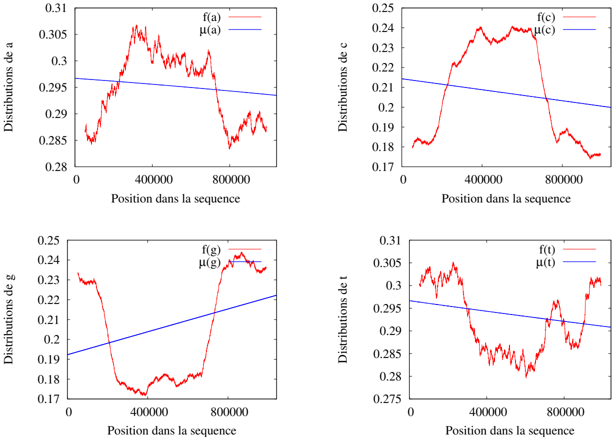
One sample path : polynomial drifting Markov model

Proposition

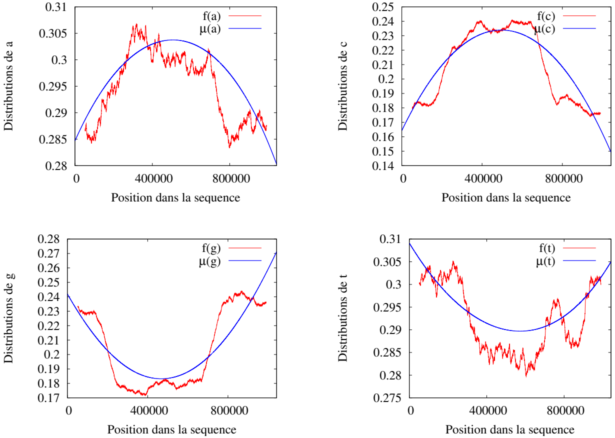
Let $\mathbb{1}_u := \mathbb{1}_{\{X_{t-k}\dots X_{t-1}=u_1\dots u_k\}}$ *and* $\mathbb{1}_{uv} := \mathbb{1}_{\{X_{t-k}\dots X_{t-1}=u_1\dots u_k, X_t=v\}}$. *For* $(X_0, X_1, \dots, X_n)$ *a sample path of a polynomial drifting Markov chain of order* k *and degree* d*, for any states* $u_1, \dots, u_k, v \in E$*, the estimators of* $\Pi_{\frac{i}{d}}(u_1, \dots u_k, v)$ *are given by solving the following linear system*

$$\begin{pmatrix} \sum_{t=k}^{n} \mathbb{1}_u A_0(t) A_0(t) & \cdots & \sum_{t=k}^{n} \mathbb{1}_u A_0(t) A_d(t) \\ \vdots & & \vdots \\ \sum_{t=k}^{n} \mathbb{1}_u A_d(t) A_0(t) & \cdots & \sum_{t=k}^{n} \mathbb{1}_u A_d(t) A_d(t) \end{pmatrix} \begin{pmatrix} \widehat{\Pi_{0;n}} \\ \vdots \\ \widehat{\Pi_{1;n}} \end{pmatrix} = \begin{pmatrix} \sum_{t=k}^{n} A_0(t) \mathbb{1}_{uv} \\ \vdots \\ \sum_{t=k}^{n} A_d(t) \mathbb{1}_{uv} \end{pmatrix}.$$

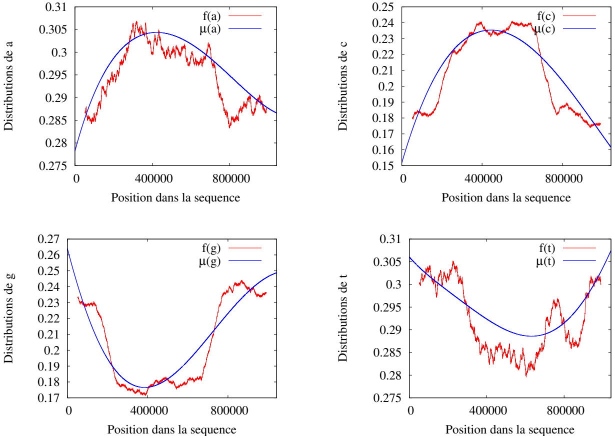
Frequencies / Stationnary distributions (degree 1) on *Chlamydia trachomatis*



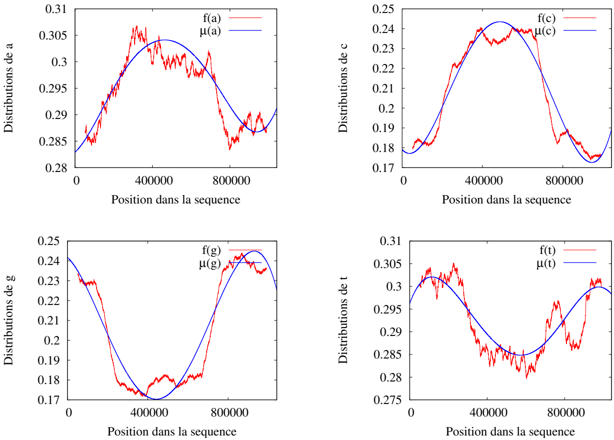
Frequencies / Stationnary distributions (degree 2)



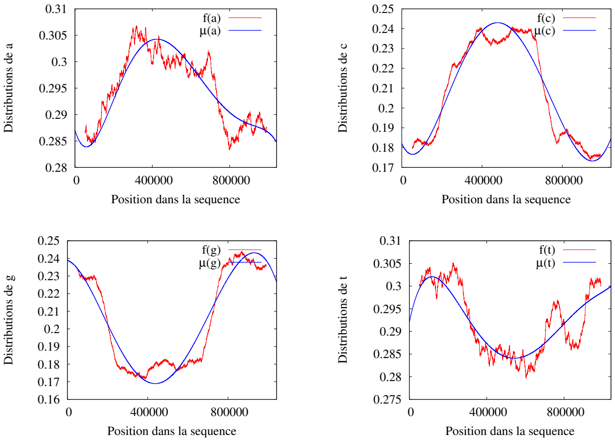
Frequencies / Stationnary distributions (degree 3)



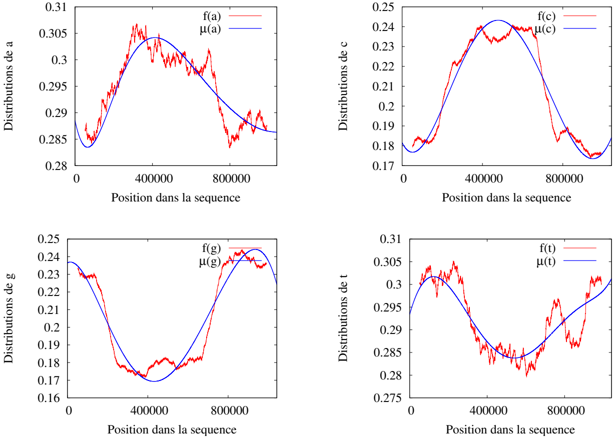
Frequencies / Stationnary distributions (degree 4)



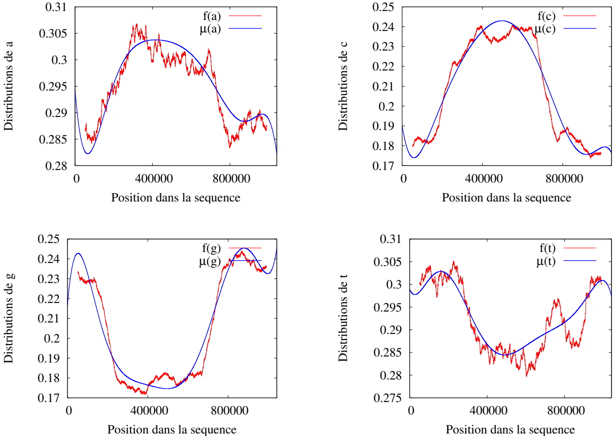
Frequencies / Stationnary distributions (degree 5)



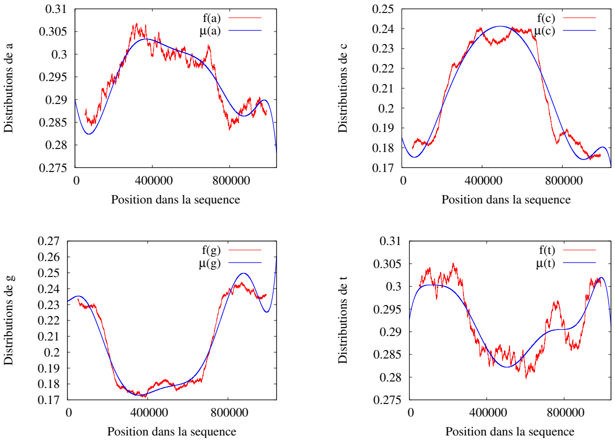
Frequencies / Stationnary distributions (degree 6)



Frequencies / Stationnary distributions (degree 7)



Frequencies / Stationnary distributions (degree 8)



Several sample paths

Let us consider H i.i.d. sample paths of a linear drifting Markov chain $(X_k)_{0 \le k \le n}$, $\mathscr{H}_1(n) := (X_0^1, X_1^1, \dots, X_n^1), \dots, \mathscr{H}_H(n) := (X_0^H, X_1^H, \dots, X_n^H)$.

Proposition

Under this setting, the estimator of $\Pi_0(u,v)$ *is given by :*

$$\widehat{\Pi}_{0;(n,H)}(u,v) =$$
$$\frac{\left(\begin{array}{c} \left(\sum_{t=1}^{n}\left(\sum_{h=1}^{H} \mathbb{1}_{\{X_{t-1}^h=u\}}\right)\left(\frac{t}{n}\right)^2\right)\left(\sum_{t=1}^{n}\left(\sum_{h=1}^{H} \mathbb{1}_{\{X_{t-1}^h=u, X_t^h=v\}}\right)\left(1-\frac{t}{n}\right)\right) - \\ \left(\sum_{t=1}^{n}\left(\sum_{h=1}^{H} \mathbb{1}_{\{X_{t-1}^h=u\}}\right)\left(1-\frac{t}{n}\right)\left(\frac{t}{n}\right)\right)\left(\sum_{t=1}^{n}\left(\sum_{h=1}^{H} \mathbb{1}_{\{X_{t-1}^h=u, X_t^h=v\}}\right)\left(\frac{t}{n}\right)\right) \end{array} \right)}{\left(\begin{array}{c} \left(\sum_{t=1}^{n}\left(\sum_{h=1}^{H} \mathbb{1}_{\{X_{t-1}^h=u\}}\right)\left(1-\frac{t}{n}\right)^2\right)\left(\sum_{t=1}^{n}\left(\sum_{h=1}^{H} \mathbb{1}_{\{X_{t-1}^h=u\}}\right)\left(\frac{t}{n}\right)^2\right) - \\ \left(\sum_{t=1}^{n}\left(\sum_{h=1}^{H} \mathbb{1}_{\{X_{t-1}^h=u\}}\right)\left(1-\frac{t}{n}\right)\left(\frac{t}{n}\right)\right)\left(\sum_{t=1}^{n}\left(\sum_{h=1}^{H} \mathbb{1}_{\{X_{t-1}^h=u\}}\right)\left(1-\frac{t}{n}\right)\left(\frac{t}{n}\right)\right) \end{array} \right)}.$$

Proposition

Under this setting, the estimator of $\Pi_1(u,v)$ *is given by :*

$$\widehat{\Pi}_{1;(n,H)}(u,v) = \frac{\left(\begin{array}{c} \left(\sum_{t=1}^{n}\left(\sum_{h=1}^{H}\mathbb{1}_{\{X_{t-1}^{h}=u\}}\right)\left(1-\frac{t}{n}\right)^{2}\right)\left(\sum_{t=1}^{n}\left(\sum_{h=1}^{H}\mathbb{1}_{\{X_{t-1}^{h}=u,X_{t}^{h}=v\}}\right)\left(\frac{t}{n}\right)\right) - \\ \left(\sum_{t=1}^{n}\left(\sum_{h=1}^{H}\mathbb{1}_{\{X_{t-1}^{h}=u\}}\right)\left(1-\frac{t}{n}\right)\left(\frac{t}{n}\right)\right)\left(\sum_{t=1}^{n}\left(\sum_{h=1}^{H}\mathbb{1}_{\{X_{t-1}^{h}=u,X_{t}^{h}=v\}}\right)\left(1-\frac{t}{n}\right)\right) \end{array} \right)}{\left(\begin{array}{c} \left(\sum_{t=1}^{n}\left(\sum_{h=1}^{H}\mathbb{1}_{\{X_{t-1}^{h}=u\}}\right)\left(1-\frac{t}{n}\right)^{2}\right)\left(\sum_{t=1}^{n}\left(\sum_{h=1}^{H}\mathbb{1}_{\{X_{t-1}^{h}=u\}}\right)\left(\frac{t}{n}\right)^{2}\right) - \\ \left(\sum_{t=1}^{n}\left(\sum_{h=1}^{H}\mathbb{1}_{\{X_{t-1}^{h}=u\}}\right)\left(1-\frac{t}{n}\right)\left(\frac{t}{n}\right)\right)\left(\sum_{t=1}^{n}\left(\sum_{h=1}^{H}\mathbb{1}_{\{X_{t-1}^{h}=u\}}\right)\left(1-\frac{t}{n}\right)\left(\frac{t}{n}\right)\right) \end{array} \right)}.$$

Method : minimize

$$\sum_{t=1}^{n}\sum_{u\in\mathscr{A}}\sum_{v\in\mathscr{A}}\sum_{h=1}^{H}\mathbb{1}_{\{X_{t-1}^{h}=u\}}\left(\Pi_{\frac{t}{n}}(u,v)-\mathbb{1}_{\{X_{t}^{h}=v\}}\right)^{2}.$$

Other types of data

Let $\mathbf{\Pi}_0$ and $\mathbf{\Pi}_1$ be two Markov transition matrices over $E = \{1, \ldots, s\}$.

1. $(X_0, X_1, \ldots, X_n)$.
2. $(X_0, X_1, \ldots, X_m), m \leq n$.
3. $\mathscr{H}_1(n) := (X_0^1, X_1^1, \ldots, X_n^1), \ldots, \mathscr{H}_H(n) := (X_0^H, X_1^H, \ldots, X_n^H)$.
4. $\mathscr{H}_1(n) := (X_0^1, X_1^1, \ldots, X_{m_1}^1), \ldots, \mathscr{H}_H(n) := (X_0^H, X_1^H, \ldots, X_{m_H}^H)$, with $m_i \leq n$, $i = 1, \ldots, H$.
5. $\mathscr{H}_1(n_1) := (X_0^1, X_1^1, \ldots, X_{n_1}^1), \ldots, \mathscr{H}_H(n_H) := (X_0^H, X_1^H, \ldots, X_{n_H}^H)$.
6. $\mathscr{H}_1(n_1) := (X_0^1, X_1^1, \ldots, X_{m_1}^1), \ldots, \mathscr{H}_H(n_H) := (X_0^H, X_1^H, \ldots, X_{m_H}^H)$; with $m_i \leq n_i$, $i = 1, \ldots, H$.

Publications, R Package

[1] **N. Vergne**, 2008. Drifting Markov models : Polynomial Drift. Applications to DNA sequences. *Statistical Applications in Genetics and Molecular Biology*, Vol 7, Iss 1, Article 6. Available at : http://www.bepress.com/sagmb/vol7/iss1/art6.

[2] **V. S. Barbu, N. Vergne**, 2019. Reliability and Survival Analysis for Drifting Markov Models : Modeling and Estimation. *Methodology and Computing in Applied Probability*, 1-23. Available at https://link.springer.com/article/10.1007/s11009-018-9682-8

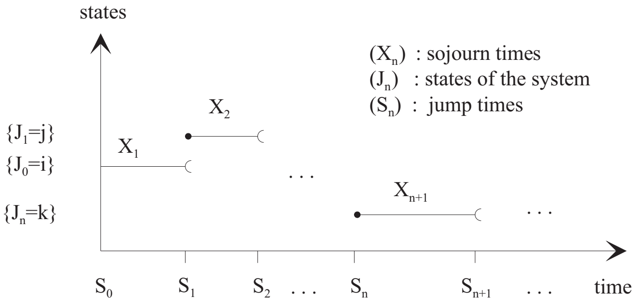
[3] **V. S. Barbu, G. Brelurut, A. Gilles, A. Lefebvre, C. Lothodé, V. Mataigne, A. Seiller** [1] **and N. Vergne**, 2021. drimmR : an R package for drifting Markov model estimation and reliability. Available at
https://cran.r-project.org/web/packages/drimmR/index.html.

1. Thanks to FEDER DAISI

Plan

Definitions

- $Z = (Z_k)_{k \in \mathbb{N}}$, chain with state space $E = \{1, 2, \ldots, s\}$
- $S = (S_n)_{n \in \mathbb{N}}$, jump times
  $J = (J_n)_{n \in \mathbb{N}}$, visited states
  $X = (X_n)_{n \in \mathbb{N}}$, sojourn times of Z



Notation/Definitions

the initial distribution $\alpha(i) := \mathbb{P}(J_0 = i)$
the homogeneous SM kernel $\mathbf{q} = \big(q(i,j;k)\big)_{i,j\in E, k\in\mathbb{N}}$

$$q(i,j;k) := \begin{cases} \mathbb{P}(J_{n+1} = j, X_{n+1} = k \mid J_n = i), & k \in \mathbb{N}^* \\ 0, & k = 0 \end{cases}$$

the conditional sojourn time distributions $\mathbf{f} = \big(f(i,j;k)\big)_{i,j\in E, k\in\mathbb{N}}$

$$f(i,j;k) := \mathbb{P}(X_{n+1} = k \mid J_n = i, J_{n+1} = j),\ f(i,j;0) := 0$$

the transition matrix of the MC $(J_n)_{n\in\mathbb{N}}$, $\mathbf{p} = (p(i,j))_{i,j\in E}$

$$p(i,j) := \mathbb{P}(J_{n+1} = j \mid J_n = i),\ p(i,i) := 0$$

Note that $q(i,j;k) = p(i,j)\ f(i,j;k)$

R Packages

[4] **V. S. Barbu, C. Bérard, D. Cellier, M. Sautreuil, N. Vergne**, 2017. SMM : a R package for Simulation and Estimation of Multi-State Discrete-Time Semi-Markov and Markov Models. Available at https://cran.r-project.org/web/packages/SMM.

[5] **V. S. Barbu, C. Bérard, D. Cellier, M. Sautreuil, N. Vergne**, 2018. SMM : An R package for estimation and simulation of discrete-time semi-Markov models. *The R journal*. Available at https://rjournal.github.io/archive/2018/RJ-2018-050/index.html.

[6] **V. S. Barbu, F. Lecocq [2], C. Lothodé and N. Vergne**, 2021. smmR : a R package for Simulation, Estimation and Reliability of Semi-Markov Models. Available at https://cran.r-project.org/web/packages/smmR/index.html.

[7] **V. S. Barbu, F. Lecocq[2] C. Lothodé and N. Vergne**, 2023. smmR : A Semi-Markov R package. *Journal of Open Source Software*, 8(85), 4365, https://doi.org/10.21105/joss.04365.

2. Thanks to RIN Asterics and FEDER DAISI

Plan

Drifting semi-Markov models – 1

Definition (linear drifting semi-Markov chain of length n – Model 1)

A sequence $Z_0, Z_1, \ldots, Z_{N(n)}$ with state space $E = \{1, 2, \ldots, s\}$ is called a *linear drifting Markov chain of length* n *of Model 1* between the semi-Markov kernels $\boldsymbol{q}_0$ and $\boldsymbol{q}_1$, if for $t = 0, \ldots, n$, we have

$$\mathbb{P}(J_{t+1} = v, S_{t+1} - S_t = \ell | J_t = u) = q_{\frac{t}{n}}(u, v, \ell),$$

where $q_{\frac{t}{n}}(u, v, \ell) = \left(1 - \frac{t}{n}\right) q_0(u, v, \ell) + \frac{t}{n} q_1(u, v, \ell),\ u, v \in E, \ell \in \mathbb{N}.$

Drifting semi-Markov models – 2

Definition (linear drifting semi-Markov chain of length n – Model 2)

A sequence $Z_0, Z_1, \ldots, Z_{N(n)}$ with state space $E = \{1, 2, \ldots, s\}$ is called a *linear drifting Markov chain of length n of Model 2* between the semi-Markov kernels $\boldsymbol{q}_0$ and $\boldsymbol{q}_1$, if for $t = 0, \ldots, n$, we have

$$\mathbb{P}(J_{t+1} = v, S_{t+1} - S_t = \ell | J_t = u) = q_{\frac{t}{n}}(u, v, \ell), \text{ where}$$

$$q_{\frac{t}{n}}(u, v; \ell) = \left(1 - \frac{t}{n}\right) f(u, v; \ell) p_0(u, v) + \frac{t}{n} f(u, v; \ell) p_1(u, v), \ u, v \in E, \ell \in \mathbb{N},$$

with $\boldsymbol{p}_0$ and $\boldsymbol{p}_1$ Markov kernels, $f(u, v; \ell)$ the conditional distribution of the sojourn in state u before jumping to v equal to ℓ.

Drifting semi-Markov models – 3

Definition (linear drifting semi-Markov chain of length n – Model 3)

A sequence $Z_0, Z_1, \ldots, Z_{N(n)}$ with state space $E = \{1, 2, \ldots, s\}$ is called a *linear drifting Markov chain of length n of Model 3* between the semi-Markov kernels $\boldsymbol{q}_0$ and $\boldsymbol{q}_1$, if for $t = 0, \ldots, n$, we have

$$\mathbb{P}(J_{t+1} = v, S_{t+1} - S_t = \ell | J_t = u) = q_{\frac{t}{n}}(u, v, \ell), \text{ where}$$

$$q_{\frac{t}{n}}(u, v; \ell) = \left(1 - \frac{t}{n}\right) f_0(u, v, \ell) p(u, v) + \frac{t}{n} f_1(u, v, \ell) p(u, v), \; u, v \in E, \ell \in \mathbb{N},$$

with $\boldsymbol{p}$ a Markov kernel, $f_0(u, v; \ell)$ and $f_1(u, v; \ell)$ conditional distributions of the sojourn in state u before jumping to v equal to ℓ.

Polynomial DSMM

Model 1 : p **and** f **are drifting**. The DSM kernel is given by :

$$q^{(1)}_{\frac{t}{n}} = \sum_{i=0}^{d} A_i(t) q^{(1)}_{\frac{i}{d}}(u,v,l) = \sum_{i=0}^{d} A_i(t) p_{\frac{i}{d}}(u,v) f_{\frac{i}{d}}(u,v,l)$$

Model 2 : **Only** p **is drifting** (f is not drifting). The DSM kernel is given by :

$$q^{(2)}_{\frac{t}{n}}(u,v,l) = \sum_{i=0}^{d} A_i(t) q^{(2)}_{\frac{i}{d}}(u,v,l) = \sum_{i=0}^{d} A_i(t) p_{\frac{i}{d}}(u,v) f(u,v,l)$$

Model 3 : **Only** f **is drifting** (p is not drifting). The DSM kernel is given by :

$$q^{(3)}_{\frac{i}{d}}(u,v,l) = \sum_{i=0}^{d} A_i(t) q^{(3)}_{\frac{i}{d}}(u,v,l) = \sum_{i=0}^{d} A_i(t) p(u,v) f_{\frac{i}{d}}(u,v,l)$$

Plan

Non-parametric Estimation for model 1

Model 1

We can estimate the DSM kernel by Least Square Estimation (LSE),

$$\widehat{q}^{\,(1)}_{\frac{t}{n}}(u,v,l) = \sum_{i=0}^{d} A_i(t)\widehat{q}^{\,(1)}_{\frac{i}{d}}(u,v,l)$$

$\forall u, v \in E, l \in \{1, \dots, k_{max}\}$, where k_{max} is the maximum realized sojourn time in the sequence, we obtain the SM kernels $\widehat{q}^{\,(1)}_{\frac{i}{d}}(u,v,l), i = 0, \dots, d$.

Estimators of $\widehat{p}_{\frac{i}{d}}(u,v)$ and $\widehat{f}_{\frac{i}{d}}(u,v,l)$ are obtained as in SMM:

$$\widehat{p}_{\frac{i}{d}}(u,v) = \sum_{l=1}^{k_{max}} \widehat{q}^{\,(1)}_{\frac{i}{d}}(u,v,l),$$

$$\widehat{f}_{\frac{i}{d}}(u,v,l) = \frac{\widehat{q}^{\,(1)}_{\frac{i}{d}}(u,v,l)}{\sum_{l=1}^{k_{max}} \widehat{q}^{\,(1)}_{\frac{i}{d}}(u,v,l)} = \frac{\widehat{q}^{\,(1)}_{\frac{i}{d}}(u,v,l)}{\widehat{p}_{\frac{i}{d}}(u,v)}.$$

Non-parametric Estimation for model 1

Model 1 : we solve $MJ = P$ to obtain $\widehat{q}_{\frac{i}{d}}^{(1)}(u,v,l)$

$$M = \begin{pmatrix} \sum_{t=1}^{n} \mathbb{1}_u(t)A_0(t)A_0(t) & \cdots & \sum_{t=1}^{n} \mathbb{1}_u(t)A_0(t)A_d(t) \\ \vdots & \ddots & \vdots \\ \sum_{t=1}^{n} \mathbb{1}_u(t)A_d(t)A_0(t) & \cdots & \sum_{t=1}^{n} \mathbb{1}_u(t)A_d(t)A_d(t) \end{pmatrix}$$

$$J = \begin{pmatrix} \widehat{q}_0^{(1)}(u,v,l) \\ \vdots \\ \widehat{q}_{\frac{i}{d}}^{(1)}(u,v,l) \\ \vdots \\ \widehat{q}_1^{(1)}(u,v,l) \end{pmatrix} \text{ and } P = \begin{pmatrix} \sum_{t=1}^{n} \mathbb{1}_{uvl}(t)A_0(t) \\ \vdots \\ \sum_{t=1}^{n} \mathbb{1}_{uvl}(t)A_i(t) \\ \vdots \\ \sum_{t=1}^{n} \mathbb{1}_{uvl}(t)A_1(t) \end{pmatrix}$$

Where $\mathbb{1}_u(t) = \mathbb{1}_{\{J_{t-1}=u\}}(t)$, and $\mathbb{1}_{uvl}(t) = \mathbb{1}_{\{J_t=v, J_{t-1}=u, X_t=l\}}(t)$.

Non-parametric Estimation for model 2

Model 2

$$\widehat{p}_{\frac{i}{d}}(u,v) = \sum_{l=1}^{k_{max}} \widehat{q}_{\frac{i}{d}}^{(1)}(u,v,l) \textbf{ (as in Model 1)}$$

$$\widehat{f}(u,v,l) = \frac{\sum_{i=0}^{d} \widehat{q}_{\frac{i}{d}}^{(1)}(u,v,l)}{\sum_{i=0}^{d} \sum_{l=1}^{k_{max}} \widehat{q}_{\frac{i}{d}}^{(1)}(u,v,l)}, \quad \text{with} \quad \sum_{l=1}^{k_{max}} \widehat{f}(u,v,l) = 1$$

This leads to the estimated SM kernel for **Model 2** $\widehat{q}_{\frac{i}{d}}^{(2)}(u,v,l)$ being described through model 1 :

$$\widehat{q}_{\frac{i}{d}}^{(2)}(u,v,l) = \frac{\left(\sum_{l=1}^{k_{max}} \widehat{q}_{\frac{i}{d}}^{(1)}(u,v,l)\right)\left(\sum_{i=0}^{d} \widehat{q}_{\frac{i}{d}}^{(1)}(u,v,l)\right)}{\sum_{i=0}^{d} \sum_{l=1}^{k_{max}} \widehat{q}_{\frac{i}{d}}^{(1)}(u,v,l)}$$

Non-parametric Estimation for model 3

Model 3

$$\widehat{p}(u,v) = \frac{\sum_{i=0}^{d} \sum_{l=1}^{k_{max}} \widehat{q}_{\frac{i}{d}}^{(1)}(u,v,l)}{d+1}, \quad \text{with} \sum_{v \in E} \widehat{p}(u,v) = 1$$

$$\widehat{f}_{\frac{i}{d}}(u,v,l) = \frac{\widehat{q}_{\frac{i}{d}}^{(1)}(u,v,l)}{\sum_{l=1}^{k_{max}} \widehat{q}_{\frac{i}{d}}^{(1)}(u,v,l)} \quad \textbf{(as in model 1)}.$$

This leads to the estimated SM kernel for **Model 3** $\widehat{q}_{\frac{i}{d}}^{(3)}(u,v,l)$ being described through model 1 :

$$\widehat{q}_{\frac{i}{d}}^{(3)}(u,v,l) = \frac{\widehat{q}_{\frac{i}{d}}^{(1)}(u,v,l) \sum_{i=0}^{d} \sum_{l=1}^{k_{max}} \widehat{q}_{\frac{i}{d}}^{(1)}(u,v,l)}{(d+1) \sum_{l=1}^{k_{max}} \widehat{q}_{\frac{i}{d}}^{(1)}(u,v,l)}$$

In conclusion :

Model 1

$$\widehat{q}_{\frac{t}{n}}^{(1)}(u,v,l) = \sum_{i=0}^{d} A_i \widehat{q}_{\frac{i}{d}}^{(1)}(u,v,l)$$

Model 2

$$\widehat{q}_{\frac{t}{n}}^{(2)}(u,v,l) = \sum_{i=0}^{d} A_i(t) \left(\frac{\left(\sum_{l=1}^{k_{max}} \widehat{q}_{\frac{i}{d}}^{(1)}(u,v,l)\right)\left(\sum_{i=0}^{d} \widehat{q}_{\frac{i}{d}}^{(1)}(u,v,l)\right)}{\sum_{i=0}^{d}\sum_{l=1}^{k_{max}} \widehat{q}_{\frac{i}{d}}^{(1)}(u,v,l)} \right)$$

Model 3

$$\widehat{q}_{\frac{t}{n}}^{(3)}(u,v,l) = \sum_{i=0}^{d} A_i(t) \left(\frac{\widehat{q}_{\frac{i}{d}}^{(1)}(u,v,l)\sum_{i=0}^{d}\sum_{l=1}^{k_{max}} \widehat{q}_{\frac{i}{d}}^{(1)}(u,v,l)}{(d+1)\sum_{l=1}^{k_{max}} \widehat{q}_{\frac{i}{d}}^{(1)}(u,v,l)} \right)$$

Parametric Estimation

Estimators of parameters for the discrete distribution of sojourn time

We note $\widehat{m_1} = \sum_{x=1}^{k_{max}} x\widehat{f}(x)$ and $\widehat{m_2} = \sum_{x=1}^{k_{max}} x^2\widehat{f}(x) - \widehat{m_1}^2$.

- Geometric(p) : $f(x) = p(1-p)^{x-1}, x = 1, \dots, k_{max}$. Then $\widehat{p} = 1/\widehat{m_1}$.
- Poisson(λ) : $f(x) = \frac{\lambda^{x-1} \exp -\lambda}{(x-1)!}, x = 1, \dots, k_{max}$. Then $\widehat{\lambda} = \widehat{m_1}$.
- Negative Binomial (α, p) : $f(x) = \frac{\Gamma(x+\alpha-1)}{\Gamma(\alpha)(x-1)!} p^\alpha (1-p)^{x-1}$, $x = 1, \dots, k_{max}$. Therefore:

$$\widehat{p} = \frac{\widehat{m_1}}{\widehat{m_2}}, \qquad \widehat{\alpha} = \widehat{m_1}\frac{\widehat{p}}{1-\widehat{p}} = \frac{\widehat{m_1}^2}{\widehat{m_2} - \widehat{m_1}}$$

- Discrete Weibull (q, β) : $f(x) = q^{(x-1)^\beta} - q^{x^\beta}$, $x = 1, \dots, k_{max}$. Therefore:

$$\widehat{q} = 1 - \widehat{f}(1), \qquad \widehat{\beta} = \frac{\sum_{i=2}^{k_{max}} \log_i(\log_{\widehat{q}}(\sum_{j=1}^{i} \widehat{f}(j))}{k_{max} - 1}$$

Plan

dsmmR - Main Functions

It constitutes of the main functions :

1. `fit_dsmm()`:

   Fit a DSMM on a given sequence. Parametric or non-parametric estimation is possible (**Model 1**, **2** or **3**).

2. `parametric_dsmm()` & `nonparametric_dsmm()`:

   Define a parametric or non-parametric DSMM (**Model 1**, **2** or **3**)

3. `simulate.dsmm()`:

   Generate a sequence of states with a maximum number of simulations equal to $n+1$ (model size).

4. `get_kernel()`:
   Compute the DSM kernel $q_{\frac{t}{n}}$.

3 Sojourn Time Distributions (degree 2) : f_0, $f_{\frac{1}{2}}$ and f_1

$f_0(u,v,l=1) = \begin{pmatrix} 0 & 0.2 & 0.7 \\ 0.3 & 0 & 0.4 \\ 0.2 & 0.8 & 0 \end{pmatrix}$	$f_{\frac{1}{2}}(u,v,l=1) = \begin{pmatrix} 0 & 0.3333333 & 0.4 \\ 0.3 & 0 & 0.4 \\ 0.2 & 0.1 & 0 \end{pmatrix}$	$f_1(u,v,l=1) = \begin{pmatrix} 0 & 0.3 & 0.3 \\ 0.3 & 0 & 0.5 \\ 0.05 & 0.1 & 0 \end{pmatrix}$
$f_0(u,v,l=2) = \begin{pmatrix} 0 & 0.3 & 0.2 \\ 0.2 & 0 & 0.5 \\ 0.1 & 0.15 & 0 \end{pmatrix}$	$f_{\frac{1}{2}}(u,v,l=2) = \begin{pmatrix} 0 & 0.3333333 & 0.4 \\ 0.4 & 0 & 0.2 \\ 0.3 & 0.4 & 0 \end{pmatrix}$	$f_1(u,v,l=2) = \begin{pmatrix} 0 & 0.2 & 0.6 \\ 0.3 & 0 & 0.35 \\ 0.9 & 0.2 & 0 \end{pmatrix}$
$f_0(u,v,l=3) = \begin{pmatrix} 0 & 0.5 & 0.1 \\ 0.5 & 0 & 0.1 \\ 0.7 & 0.05 & 0 \end{pmatrix}$	$f_{\frac{1}{2}}(u,v,l=3) = \begin{pmatrix} 0 & 0.3333333 & 0.2 \\ 0.3 & 0 & 0.4 \\ 0.5 & 0.5 & 0 \end{pmatrix}$	$f_1(u,v,l=3) = \begin{pmatrix} 0 & 0.5 & 0.1 \\ 0.4 & 0 & 0.15 \\ 0.05 & 0.7 & 0 \end{pmatrix}$

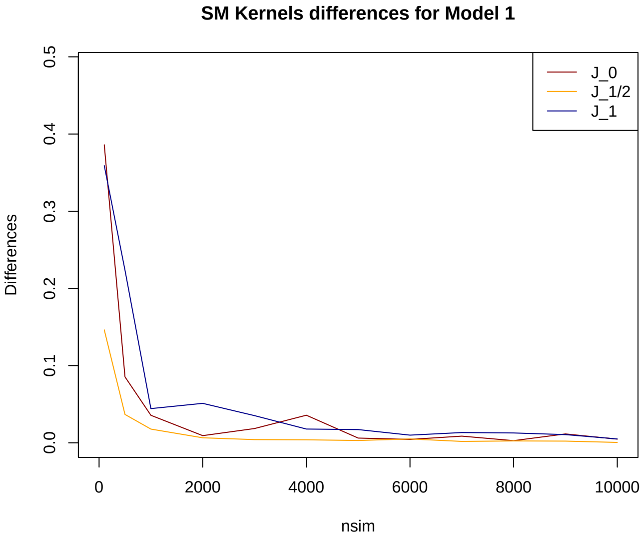
3 Transition Matrices (degree 2) : $p_0, p_{\frac{1}{2}}$ and p_1 and metric

$p_0(u,v) = \begin{pmatrix} 0 & 0.1 & 0.9 \\ 0.5 & 0 & 0.5 \\ 0.3 & 0.7 & 0 \end{pmatrix}$	$p_{\frac{1}{2}}(u,v) = \begin{pmatrix} 0 & 0.6 & 0.4 \\ 0.7 & 0 & 0.3 \\ 0.6 & 0.4 & 0 \end{pmatrix}$	$p_1(u,v) = \begin{pmatrix} 0 & 0.2 & 0.8 \\ 0.6 & 0 & 0.4 \\ 0.7 & 0.3 & 0 \end{pmatrix}$

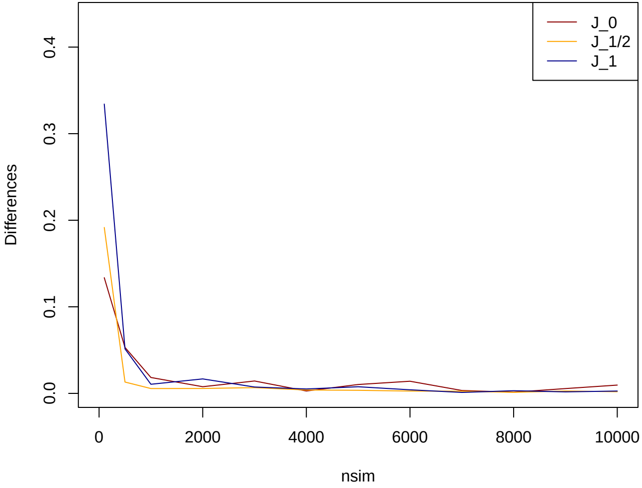
We are going to use the following metric, defining the distance between the $d+1$ theoretical kernels $q_{\frac{i}{d}}^{(M)}$ with the estimated ones $\widehat{q}_{\frac{i}{d}}^{(M)}$, for all 3 models $M = 1, 2, 3$:

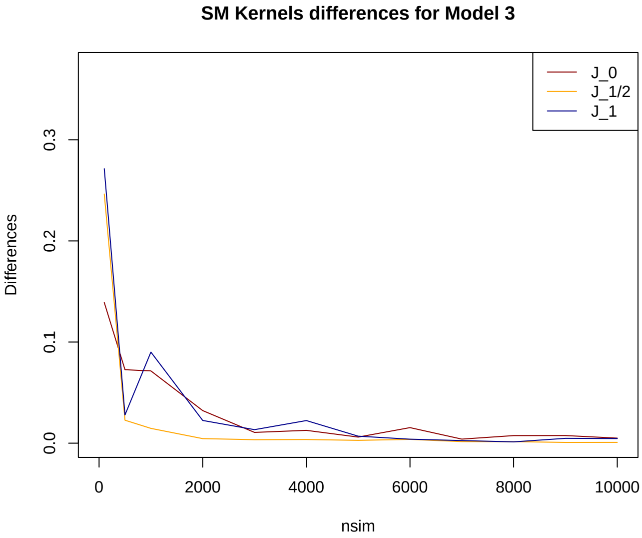
$$d\left(q_{\frac{i}{d}}^{(M)}, \widehat{q}_{\frac{i}{d}}^{(M)}\right) = \sum_{uvl} \left(q_{\frac{i}{d}}^{(M)} - \widehat{q}_{\frac{i}{d}}^{(M)}\right)^2,$$

where $u, v \in E, l \in \{1, \ldots, k_{max}\}$.


SM Kernels differences for Model 1
Differences
nsim
0.0
0.1
0.2
0.3
0.4
0.5
0
2000
4000
6000
8000
10000
J_0
J_1/2
J_1

SM Kernels differences for Model 2




SM Kernels differences for Model 3
J_0
J_1/2
J_1
Differences
0.0
0.1
0.2
0.3
0
2000
4000
6000
8000
10000
nsim

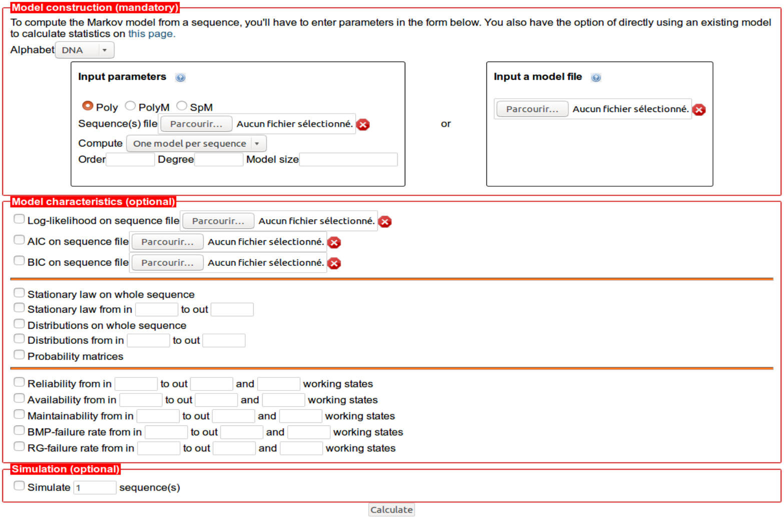
R Package

[8] **V. S. Barbu, I. Mavrogiannis [3] and N. Vergne**, 2022. `dsmmR` : a R package for Estimation and Simulation of Drifting Semi-Markov Models. Available at `https://cran.r-project.org/web/packages/dsmmR/index.html`.

3. Thanks to FEDER DATALAB

Plan

Model



Model construction (mandatory)

To compute the Markov model from a sequence, you'll have to enter parameters in the form below. You also have the option of directly using an existing model to calculate statistics on this page.

Alphabet DNA

Input parameters

Poly PolyM SpM

Sequence(s) file Parcourir... Aucun fichier sélectionné.

Compute One model per sequence

Order Degree Model size

or

Input a model file

Parcourir... Aucun fichier sélectionné.

Model characteristics (optional)

Log-likelihood on sequence file Parcourir... Aucun fichier sélectionné.

AIC on sequence file Parcourir... Aucun fichier sélectionné.

BIC on sequence file Parcourir... Aucun fichier sélectionné.

Stationary law on whole sequence

Stationary law from in to out

Distributions on whole sequence

Distributions from in to out

Probability matrices

Reliability from in to out and working states

Availability from in to out and working states

Maintainability from in to out and working states

BMP-failure rate from in to out and working states

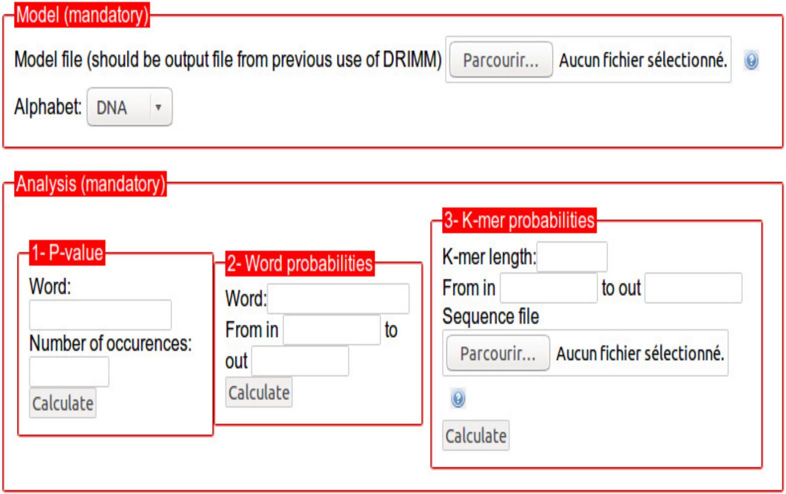
RG-failure rate from in to out and working states

Simulation (optional)

Simulate 1 sequence(s)

Calculate

Analysis

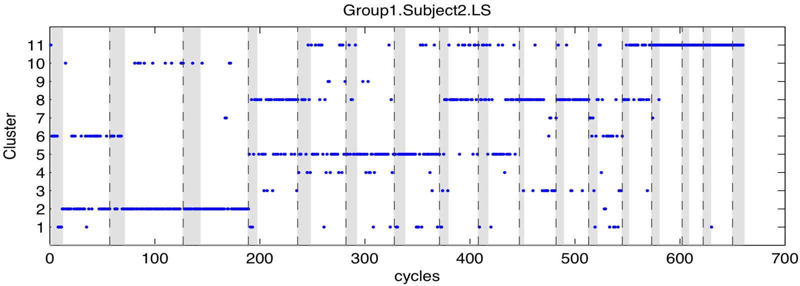

Model (mandatory)
Model file (should be output file from previous use of DRIMM)
Parcourir...
Aucun fichier sélectionné.
Alphabet:
DNA
Analysis (mandatory)
1- P-value
Word:
Number of occurences:
Calculate
2- Word probabilities
Word:
From in
to
out
Calculate
3- K-mer probabilities
K-mer length:
From in
to out
Sequence file
Parcourir...
Aucun fichier sélectionné.
Calculate

Download

[9] **J. Komar, A. Lefebvre, H. Mayeur and N. Vergne**, 2021. WebDRIMM : a web interface for drifting Markov model estimation and reliability. Available at http://bioinfo.univ-rouen.fr/WebDRIMM/download.php.

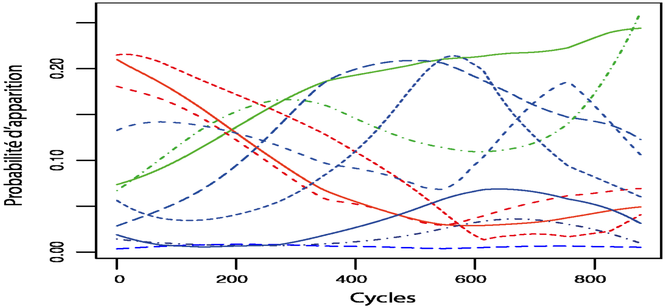
Plan

Swimming : learning



- 30 swimmers, 20 sessions of 250m : arm-leg coordination
- 11 states : 11 characteristic motor behaviors
- Here, coordination 11 has been progressively installed in this swimmer

Swimming : 3 phenomena of learning



- Disappearance (red) of beginner behaviors
- Appearance (green) of expert behavior
- Motor exploration (blue) : appearance and disappearance of behaviors (necessity ?)
- Perspective : study of these behaviors as a function of learning conditions

Publication

[10] **J. Komar, L. Seifert, N. Vergne and K.M. Newell**, 2023. Narrowing the coordination solution space during motor learning standardizes individual patterns of search strategy but diversifies learning rates. *Scientific Reports*, 13. Available at : https://www.nature.com/articles/s41598-023-29238-z

UNIVERSITE DE ROUEN   cnrs   LMRS   ETAPS

Plan

Visual Motor skill

The **effective** coordination of one's movements, within physically demanding environments, is a major challenge in several activities such as that of climbing.

It is without a doubt that climbers rely on multiple sources of information to scan their environment and aid in their actions. These sources of information encompass visual, auditory, tactile as well as prior experience.

Visual-motor skill is defined as the ability to **identify** important visual cues in the environment and **coordinate** movements to achieve a desired outcome.

Visual Motor skill

When it comes to the modeling of visual-motor skill data, **Drifting Markov Models (DMMs)** could be proved a flexible tool due to their ability to model sequence **heterogeneities** as opposed to the "traditional" Markov chains or hidden Markov models.

Why to assume heterogeneity ?

1. It is usually involved to the evolution of dynamical systems, and ;
2. Interactions may occur between the system and the environment that can potentially lead to a heterogeneity in the functioning of the system.

In this work, we study climber´s dynamic of learning across several sessions, i.e., on a long time scale, through the DMMs mechanism based on a real case study.

Experimental Data

The experiment was conducted at the University of Rouen Normandy. By utilizing specialized equipment such as eye-tracking glasses and action cameras, $\mathbf{10}^{*}$ climbing Sessions of $\mathbf{11}$ individuals were recorded. Each session consisted by several trials, specifically :

- Sessions 1 and 10 : $\mathbf{6}$ Trials ;
- Sessions 2-9 : $\mathbf{9}$ Trials.

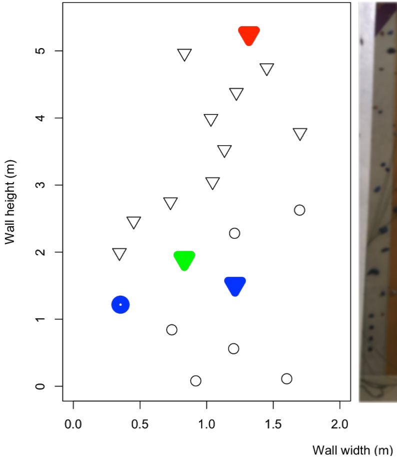
After a thorough preparation/investigation/cleaning of the raw data, the final amount of trials remained, from the starting 854, were 729.

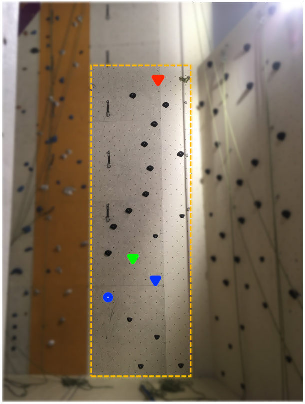
*Please note that due to medical issues, individual No2 was submitted only to 4 Sessions in total.

Guidelines

1. Climb as fluently as possible, i.e., minimizing pauses and jerky movements of the body ;
2. Use all the handholds in a specific order from bottom to top ;
3. Use all the handholds/footholds with a single limb contact at a time (participants were prohibited to use a hold with both hands/feet simultaneously).

The Wall





Clustering

The first step of the analysis is to **determine** the state space A, i.e., the clusters required for applying the DMM procedure. To that end, an unsupervised clustering technique (k-Means) is applied based on the following variables :

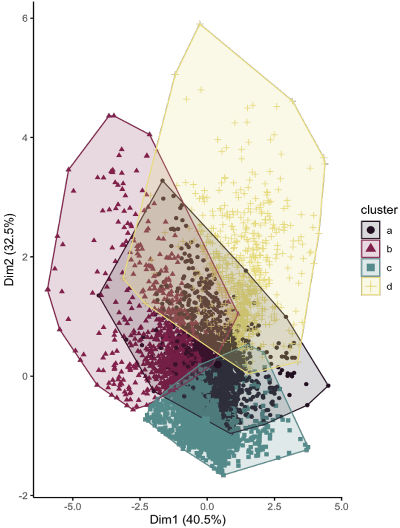
- **Gaze Offset** (time difference between last gaze visit within the Area of Interest (AOI $= 30cm$ around the hold) of the previous hold and touching the next hold) – GO ;
- **Time Difference Between Touching Two Holds** – TD2H, and ;
- **Duration of Last Gaze Visit Within the AOI** – DLGV.

After testing several values for k and based on Fisher's Information (FI, the higher the better) defined as :

$$FI = \frac{\text{Inter cluster effect}}{\text{Intra cluster effect}} = \frac{SSB}{SST_w},$$

the chosen number of clusters is 4, namely a, b, c and d.

Clustering



Clustering

Table – Centroid **values** of each cluster

Cluster	GO	TD2H	DLGV
a (medium effective)	-0.047	0.936	2.928
b (less effective)	-0.241	2.100	0.960
c (most effective)	-0.059	0.798	0.786
d (less effective)	-0.002	1.060	6.506

Table – Centroid **SDs** of each cluster

Cluster	GO	TD2H	DLGV
a	0.231	0.518	0.750
b	0.344	0.751	0.570
c	0.242	0.326	0.419
d	0.239	0.591	1.694

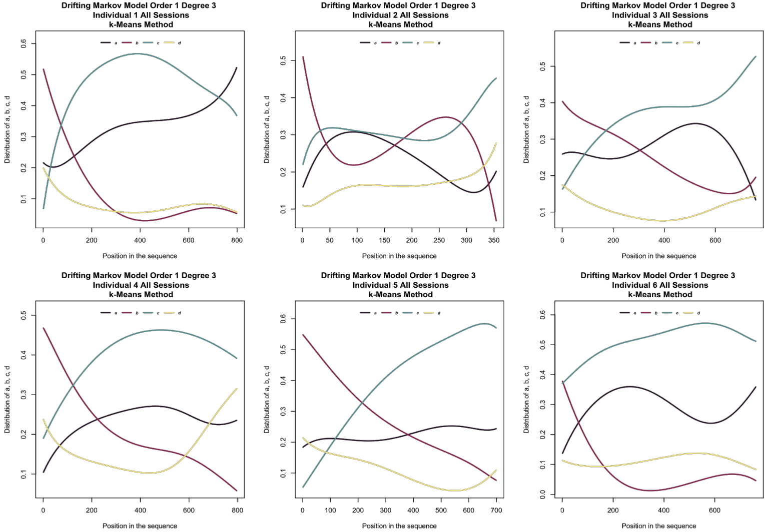
Drifting Markov Model

Having concluded with the state space $A = \{a, b, c, d\}$, we can move onto the DMM procedure. After several combinations of orders and degrees, we concluded that the **optimal** selection (in terms of both complexity and information gained and AIC) is the DMM of O(1) and D(3) :

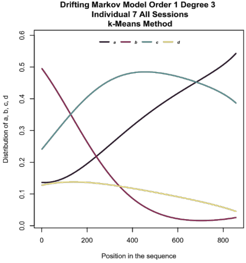
$$\Pi_{\frac{t}{n}}(u,v) = \left(-\frac{9}{2}\frac{t^3}{n^3} - 9\frac{t^2}{n^2} - \frac{11}{2}\frac{t}{n} + 1\right)\mathbf{\Pi}_0 + \left(\frac{27}{2}\frac{t^3}{n^3} - \frac{45}{2}\frac{t^2}{n^2} + 9\frac{t}{n}\right)\mathbf{\Pi}_{\frac{1}{3}}$$
$$+ \left(-\frac{27}{2}\frac{t^3}{n^3} + 18\frac{t^2}{n^2} - \frac{9}{2}\frac{t}{n}\right)\mathbf{\Pi}_{\frac{2}{3}} + \left(\frac{9}{2}\frac{t^3}{n^3} - \frac{9}{2}\frac{t^2}{n^2} + \frac{t}{n}\right)\mathbf{\Pi}_1 .$$

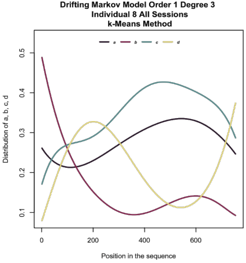
Through the *drimmR* package in *R*, the above DMM was fitted to each of the 11 individuals on the sequence of all of their trials with results as follows.

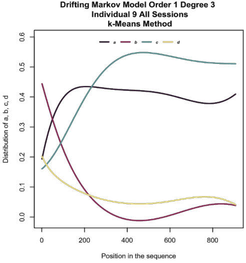
Distributions of states

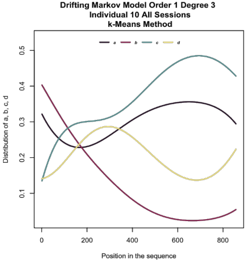


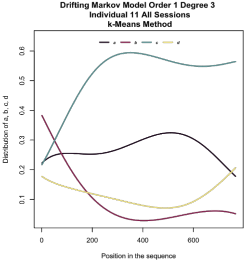
Distributions of states











First conclusions of the modeling

In this work our **goal** was to study climber´s **dynamic of learning on a long time scale** through DMMs. To that end, a real case study was conducted at University of Rouen Normandy concerning 729 trials of 11 individuals. By applying the k-Means clustering technique for obtaining the state space A = {a,b,c,d} three effectiveness-related behaviors/patterns were discriminated :

1. **Less effective**
   Clusters "d" (quite long TD2H and longest DLGV) and "b" (most negative GO and longest TD2H) ;
2. **Medium effective**
   Cluster "a" (quite long TD2H and second longest DLGV) ;
3. **Most effective**
   Cluster "c" (shortest DLGV and shortest TD2H).

First conclusions of the modeling

Furthermore, by fitting the DMM of order 1-degree 3 on the whole sequence of trials of each individual, we observed that :

1. All individuals "struggled" at the beginning of learning since they used (with high probability) the less effective cluster "b".
2. Across the learning process, the behavior of individuals was positively evolved leading to the adoption (by the majority) of the most effective cluster "c".
3. There were three individuals that appeared a different behavior at the end of the process.

Conferences and bibliography

[11] **V. S. Barbu, G. Hacques, E. Kalligeris[4], L. Seifert and N. Vergne**, 2023. Assessing the Dynamic of Visual-Motor Skill in Climbing via Drifting Markov Modeling, *35rd Panhellenic Statistical Conference, Athens, Greece*.

[12] **V. S. Barbu, G. Hacques, E. Kalligeris [4], L. Seifert and N. Vergne**, 2023. Investigating the Dynamic of Visual-Motor Skill Acquisition in Climbing through Drifting Markov Modeling, *A 24-a Conferinta a societat II de probabilitati si statistica din Romania (SPSR), Bucharest, Romania*.

[13] **T.N Nguyen, L. Seifert, G. Hacques, M.H. Kölbl and Y. Chahir**, 2023. Vision-based Global Localization of Points of Gaze in Sport Climbing, *International Journal of Pattern Recognition and Artificial Intelligence*.

[14] **Seifert L., R. Cordier, D. Orth, Y. Courtine and J.L. Croft**, 2017. Role of Route Previewing Strategies on Climbing Fluency and Exploratory Movements, *PLOS ONE*.

UNIVERSITE DE ROUEN cnrs LMRS CETAPS

4. Thanks to RIN EYE-TRACKING RÉGION NORMANDIE UNION EUROPÉENNE Fonds européen de développement régional

Plan

Concluding remarks

Conclusions

- drimmR : drifting Markov models, reliability, different types of datas.
- smmR : semi-Markov models, reliability, different types of sojourn time, censoring, non-parametric and parametric.
- dsmmR : drifting semi-Markov models, models 1, 2 and 3, non-parametric and parametric.

Future directions

- hsmmR : hidden semi-Markov models, reliability, different types of sojourn time, censoring, non-parametric and parametric (thanks to ANR HSMM-INCA).
- gdrimmR : generalized drifting Markov models, reliability (thanks to CNRS for E. Kalligeris contract).
- Eye-Tracking : work in collaboration with CETAPS to analyse climbing data. Clustering of models.