Dynamic Statistical Models with Hidden Variables

Benjamin Poignard

CREST & PSL (Paris Dauphine University, CEREMADE)

Chapter 3: Markov-switching models

Time series and breaks

Many economic time series occasionally exhibit dramatic breaks in their behavior. Such breaks may concern the level, the variability, the serial correlations..

They are often associated with events such as financial crises or abrupt changes in government policy. For instance, many economic variables tend to behave differently during economic downturns.

When models are fitted over sub-periods of long times series, one often detects significant changes in the estimated parameter values.

Old method

The main approach consisted in fitting different models over different subperiods. For instance, for AR(1) models :

$$
\begin{aligned}
y_t &= \Phi_1 y_{t-1} + \epsilon_t, \text{ if } t \leq t_0, \\
y_t &= \Phi_2 y_{t-1} + \epsilon_t, \text{ if } t > t_0.
\end{aligned}
$$

Problems:

Sub-models are not related.
Arbitrary choice of the break date t_0.
Assumption of non stationarity.

A simple way to model dynamic changes

AR(1) model with random AR coefficient :

$$y_t = \Phi(\Delta_t) y_{t-1} + \epsilon_t,$$

where (Δ_t) is a discrete random variable.

We need a structure over the latent variable to describe the probability law of the data.

Simple approach: Δ_t is the realization of a Markov chain (finite-state).

Such a model is called Markov-Switching (MS) AR(1) model.

References

The introduction of MS models in the econometrics literature is due to Hamilton:
Hamilton, J. D. (1989) *A new approach to the economic analysis of nonstationary time series and the business cycle.* Econometrica 57, 357384.

In the statistics literature, related models called Hidden Markov Models (HMM) had been studied much earlier :
Baum, L. E., et T. Petrie (1966) *Statistical inference for probabilistic functions of finite state Markov chains.* Annals of Mathematical Statistics 30, 1554-1563.

1. Hidden Markov Model
   - Finite-state Markov chains
   - Properties of Hidden Markov chains
2. MS-ARMA(p,q) process
3. Estimation of MS-AR models

Definition

Let $\Delta_0, \Delta_1, \cdots \in \mathcal{S} = \{1, \cdots, d\}$ a sequence of random variables. It is a Markov chain if

$$\begin{aligned}\mathbb{P}(\Delta_t = j | \Delta_{t-1} = i) &= \mathbb{P}(\Delta_t = j | \Delta_{t-1} = i, \Delta_{t-2} = e_{t-2} \cdots, \Delta_0 = e_0) \\ &= p(i,j),\end{aligned}$$

for any t and $(i, j, e_{t-2}, \cdots, e_0) \in \mathcal{S}^{t+1}$.

The set $\mathcal{S}$ is called the state space of the process and the $p(i,j)$ are the transition probabilities.

Law

The law of the Markov chain is entirely defined by

(i) the inital probabilities

$$\pi_0(i) = \mathbb{P}(\Delta_0 = i),\ \pi_0(i) \geq 0,\ i = 1, \cdots, d,\ \sum_{i=1}^{d} \pi_0(i) = 1.$$

(ii) the transition probability matrix

$$\mathbf{P} = (p(i,j))_{1 \leq i,j \leq d}.$$

High-order transitions

The k-th power of the transition matrix $\mathbf{P}^k = (p^{(k)}(i,j))_{1\leq i,j\leq d}$ provides the k-th step transition probabilities:

$$p^{(k)}(i,j) = \mathbb{P}(\Delta_t = j | \Delta_{t-k} = i),\ i,j \in \mathcal{S},\ k \geq 0.$$

Let

$$\pi_0 = \begin{pmatrix} \pi_0(1) \\ \vdots \\ \pi_0(d) \end{pmatrix},\ \text{and}\ \pi_n = \begin{pmatrix} \mathbb{P}(\Delta_n = 1) \\ \vdots \\ \mathbb{P}(\Delta_n = d) \end{pmatrix}.$$

We have

$$\pi_n = \mathbf{P}'\pi_{n-1},\ \pi_n = \mathbf{P}'^n\pi_0,\ n \geq 0.$$

Invariant probability

A probability π on $\mathcal{S}$ is called invariant probability if

$$\pi = \mathbf{P}'\pi,\ \pi'\iota = 1.$$

- If the limit law $\pi_\infty := \lim_{n\to\infty} \pi_n$ exists, then it is an invariant probability.
- An invariant probability always exists (for a finite state space).
- If $\pi_0 = \pi$ where π is an invariant probability, then $\pi_n = \pi$ for all $n \geq 0$.

Irreducibility, aperiodicity

It is possible for a chain starting in i to reach j if and only if

$$p^{(n)}(i,j) > 0, \text{ for some } n.$$

If it is true for all i and j, the Markov chain is called irreducible.

A state i is called aperiodic if

$$1 = gcd\{n; p^{(n)}(i,i) > 0\}.$$

If all states verify this condition, the chain is aperiodic.

Exponential convergence to the stationary law and ergodicity

Proposition

If the chain is **irreducible** *and* **aperiodic***, there is a stationary distribution π and there exists $K \geq 0$ and $0 < \rho < 1$ such that*

$$|p^{(n)}(i,j) - \pi(j)| \leq K\rho^n,$$

for all states i and j.

Then under these conditions, we have the ergodicity property

$$\frac{1}{n}\sum_{t=1}^{n} f(\Delta_t) \underset{n\to\infty}{\longrightarrow} \mathbb{E}_\pi[f(\Delta_t)] = \sum_{i=1}^{d} \pi(i) f(i), \; a.s.$$

for any function $f(.)$. An irreducible, aperiodic and stationary Markov chain is called ergodic.

1. Hidden Markov Model
   - Finite-state Markov chains
   - Properties of Hidden Markov chains
2. MS-ARMA(p,q) process
3. Estimation of MS-AR models

Definition

A process (X_t) follows a HMM if

(i) conditionally to a hidden Markov chain (Δ_t), the variables $X_0, , X_1, \cdots$ are independent.

(ii) the conditional law of X_s given (Δ_t) only depends on Δ_s.

Canonical HMM:

$$\epsilon_t = \sigma(\Delta_t)\eta_t,$$

such that

- $0 < \sigma(1) < \cdots < \sigma(d)$.
- (η_t) is an iid sequence of variables $\mathbb{E}[\eta_t] = 0$ and $\mathrm{Var}(\eta_t) = 1$.
- (Δ_t) is an ergodic Markov chain on $\mathcal{S}$.
- the sequences (η_t) and (Δ_t) are independent.

Unconditional law

If $\mathbb{P}_{\eta_t} = \mathcal{N}_{\mathbb{R}}(0,1)$, then the distribution of ϵ_t is a mixture of centered Gaussian variables such that its density corresponds to

$$f(x) = \sum_{k=1}^{d} \pi(k) \frac{1}{\sigma(k)} \varphi(\frac{x}{\sigma(k)})$$

For any law of η_t, the marginal moments of (ϵ_t) can be obtained as

$$\mathbb{E}[\epsilon_t^r] = \mathbb{E}[\sigma^r(\Delta_t)]\mathbb{E}[\eta_t^r] = \sum_{k=1}^{d} \sigma^r(k)\pi(k)\mathbb{E}[\eta_t^r].$$

Definition

The MS-ARMA(p,q) model is defined as

$$\begin{cases} X_t &= c(\Delta_t) + \sum_{i=1}^{p} a_i(\Delta_t) X_{t-i} + \epsilon_t + \sum_{j=1}^{q} b_j(\Delta_t)\epsilon_{t-j}. \\ \epsilon_t &= \epsilon(\Delta_t) = \sigma(\Delta_t)\eta_t,\ \mathbb{P}_{\eta_t} = \mathcal{L}(0,1) \text{ i.i.d.,} \end{cases}$$

with (Δ_t) an ergodic Markov chain on $\mathcal{S}$ and independent of (η_t). Besides $a_i(.), b_j(.), c(.) \in \mathbb{R}$ and $\sigma(.) > 0$.

This model contains the HMM as a particular case.

Except the case $p = 0$, the existence of stationary solutions require additional conditions.

Notations

For any functions $f : \mathcal{S} \to \mathcal{M}_{n\times m}(\mathbb{R})$ and for all non-negative intergers i, n, m, let

$$\mathbf{P}^{(i)}(f) = \begin{pmatrix} p^{(i)}(1,1)f(1) & \cdots & p^{(i)}(d,1)f(1) \\ \vdots & & \vdots \\ p^{(i)}(1,d)f(d) & \cdots & p^{(i)}(d,d)f(d) \end{pmatrix}, \Pi(f)) = \begin{pmatrix} \pi(1)f(1) \\ \vdots \pi(d)f(d) \end{pmatrix}$$

- If $i = 1$, then we denote $\mathbf{P}(f) = \mathbf{P}^{(1)}(f)$.
- for $f = 1$, let $\mathbf{P} = \mathbf{P}(1) = (p(j,i))$ the transpose of the transition matrix.

Computation of expectations

Lemma

Let $f_0, \cdots, f_k$ functions defined on $\mathcal{S} \to \mathcal{M}_{n\times n}(\mathbb{R})$. For $k > 0$, then

$$\mathbb{E}[f_0(\Delta_t) f_1(\Delta_{t-1}) \cdots f_k(\Delta_{t-k})] = \vec{I}\mathbf{P}(f_0) \cdots \mathbf{P}(f_{k-1}) \Pi(f_k),$$

with $\vec{I} = (I_n, \cdots, I_n)$ a $n \times nd$ matrix with I_n the identity matrix of size n.

1. Hidden Markov Model

2. MS-ARMA(p,q) process
   - Stationarity of the MS-AR(1) model
   - Stationarity of the MS-ARMA(p,q) model

3. Estimation of MS-AR models

MS-AR(1) without constant

$$X_t = a(\Delta_t)X_{t-1} + \sigma(\Delta_t)\eta_t.$$

Issues:

- Existence of a strictly sationary solution.
- Existene of a 2*nd* order statioanry solution.

We look for non-anticipative solutions, idest solutions as

$$X_t = f(\eta_t, \Delta_t, \eta_{t-1}, \Delta_{t-1}, \cdots).$$

By iteration, for $k \geq 1$, then

$$\begin{aligned} X_t &= a(\Delta_t)\cdots a(\Delta_{t-k+1})X_{t-k} \\ &+ \sum_{i=0}^{k-1} a(\Delta_t)\cdots a(\Delta_{t-i+1})\sigma(\Delta_{t-i})\eta_{t-i}. \end{aligned}$$

By convention, $a(\Delta_t)\cdots a(\Delta_{t-i+1}) = 1$ if $i = 0$.

Solutions should be given by

$$\tilde{X}_t = \sum_{i=0}^{k-1} a(\Delta_t)\cdots a(\Delta_{t-n+1})\sigma(\Delta_{t-n})\eta_{t-n},$$

under the condition the series convergences almost surely.

Cauchy root test

We use the n-th root Cauchy test to derive an absolute convergence condition. To do so, let

$$u_n = a(\Delta_t)\cdots a(\Delta_{t-n+1})\sigma(\Delta_{t-n})\eta_{t-n}.$$

We have

$$|u_n|^{\frac{1}{n}} = \exp\{\frac{1}{n}\sum_{k=1}^{n}\log|a(\Delta_{t-k+1})| + \frac{1}{n}\log(\sigma(\Delta_{t-n})|\eta_{t-n}|)\}.$$

Since lim sup $n^{-1}\log(\sigma(\Delta_{t-n})|\eta_{t-n}|) = 0$ a.s., we obtain by the ergodic theorem

$$\text{lim sup } |u_n|^{\frac{1}{n}} = \exp\{\mathbb{E}[\log|a(\Delta_t|]\}.$$

This provides the condition

$$\mathbb{E}[\log|a(\Delta_t)|] < 0.$$

Uniqueness

Under the previous condition

$$\tilde{X}_t = \sum_{k=1}^{\infty} a(\Delta_t) \cdots a(\Delta_{t-n+1}) \sigma(\Delta_{t-n}) \eta_{t-n}, \text{ a.s.}$$

is properly defined and is solution of the MS-AR(1) model.
Suppose there exists another strictly stationary solution (X_t^*):

$$X_t^* = a(\Delta_t) \cdots a(\Delta_{t-k+1}) X_{t-k}^* + \sum_{k=1}^{k-1} a(\Delta_t) \cdots a(\Delta_{t-i+1}) \sigma(\Delta_{t-i}) \eta_{t-i}.$$

Then

$$|X_t - X_t^*| \leq |a(\Delta_t) \cdots a(\Delta_{t-k+1})| |X_{t-k}^*| + \mathcal{R}_{t,k},$$

such that $\mathcal{R}_{t,k} \to 0$ and $|a(\Delta_t) \cdots a(\Delta_{t-k+1})| \underset{k \to \infty}{\longrightarrow} 0$ a.s..
Consequently

$$X_t = X_t^* \text{ a.s.}$$

Strict stationary condition

Proposition

There exists a unique striclty stationary solution if

$$\mathbb{E}[\log|a(\Delta_t)|] = \sum_{i=1}^{d} \log|a(i)|\pi(i) < 0.$$

This solution is nonanticipative.

Second-order stationarity of the MS-AR(1) model

Existence in L^2 of

$$\tilde{X}_t = \sum_{k=0}^{\infty} \mathcal{R}_{t,k},\ \mathcal{R}_{t,k} = a(\Delta_t)\cdots a(\Delta_{t-n+1})\sigma(\Delta_{t-n})\eta_{t-n}.$$

Using the L^2 norm, we have

$$\|\tilde{X}_t\| \leq \sum_{k=0}^{\infty} \|\mathcal{R}_{t,k}\|,$$

and

$$\begin{aligned} \mathbb{E}[\mathcal{R}_{t,k}^2] &= \mathbb{E}[a^2(\Delta_t)\cdots a^2(\Delta_{t-k})] \\ &= (1,\cdots,1)\mathbf{P}^n(a^2)\Pi(\sigma^2). \end{aligned}$$

Second-order stationarity of the MS-AR(1) model

Proposition

If

$$\rho(\mathbf{P}(a^2)) = \rho\begin{pmatrix} p(1,1)a^2(1) & \cdots & p(d,1)a^2(1) \\ \vdots & & \vdots \\ p(1,d)a^2(d) & \cdots & p(d,d)a^2(d) \end{pmatrix} < 1,$$

$(\tilde{X}_t)$ *is the unique second order stationary and non-anticipative solution of the MS-AR(1) model.*
Conversly, if $\rho(\mathbf{P}(a^2)) \geq 1$*, there is no second order stationary and non-anticipative solution.*

Vectorial representation

Vectorial dynamic: $Z_t = A_t Z_{t-1} + B_t$, with $B_t = C_t + \Sigma_t \eta_t$,

$$
\begin{aligned}
C_t &= (c(\Delta_t), 0, \cdots, 0, \cdots, 0)' \in \mathbb{R}^{p+q}, \\
Z_t &= (X_t, \cdots, X_{t-p+1}, \epsilon_t, \cdots, \epsilon_{t-q+1})' \in \mathbb{R}^{p+q}, \\
\Sigma_t &= (\sigma(\Delta_t), \cdots, 0, \sigma(\Delta_t), \cdots, 0).
\end{aligned}
$$

$$
A_t = \begin{pmatrix} a(\Delta_t) & b(\Delta_t) \\ 0 & J \end{pmatrix} \in \mathcal{M}(\mathbb{R})_{(p+q)\times(p+q)}.
$$

Vectorial dynamic: $Z_t = A_t Z_{t-1} + B_t$.

$$a(\Delta_t) = \begin{pmatrix} a_1(\Delta_t) & & \cdots & & a_p(\Delta_t) \\ 1 & 0 & \cdots & & 0 \\ 0 & 1 & \cdots & & 0 \\ \vdots & \ddots & \ddots & & \vdots \\ 0 & & \cdots & 1 & 0 \end{pmatrix},$$

$$b(\Delta_t) = \begin{pmatrix} b_1(\Delta_t) & & \cdots & & b_q(\Delta_t) \\ 0 & 0 & \cdots & & 0 \\ 0 & 0 & \cdots & & 0 \\ \vdots & \ddots & \ddots & & \vdots \\ 0 & & \cdots & 1 & 0 \end{pmatrix}, J = \begin{pmatrix} 0 & 0 & \cdots & & 0 \\ 1 & 0 & \cdots & & 0 \\ 0 & 1 & \cdots & & 0 \\ \vdots & \ddots & \ddots & & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}$$

Top Lyapunov exponent

The top Lyapunov exponent of the sequence $(a(\Delta_t))$ is defined as

$$\begin{aligned}\gamma &= \inf_{t>0} \mathbb{E}[\frac{1}{t}\log \|a(\Delta_t)a(\Delta_{t-1})\cdots a(\Delta_1)\|] \\ &\overset{a.s.}{=} \lim_{t\to\infty} \frac{1}{t}\log \|\prod_{i=1}^{t} a(\Delta_{t-i})\|,\end{aligned}$$

for any norm on $\mathcal{M}(\mathbb{R})_{p\times p}$.

Strict stationarity condition

Proposition

Suppose $\gamma_a < 0$. *Then, for any* $t \in \mathbb{Z}$, *the sequence*

$$Z_t = B_t + \sum_{k=1}^{\infty} A_t \cdots A_{t-k+1} B_{t-k}$$

converges almost surely and the process (X_t) *defined as the first component of* (Z_t) *is the unique strictly stationary solution of the MS-ARMA(p,q) model.*

MS-AR(p) model

$$X_t = \sum_{i=1}^{p} a_i(\Delta_t) X_{t-i} + \sigma(\Delta_t)\eta_t, \ \mathbb{P}_{\eta_t} = \mathcal{N}_{\mathbb{R}}(0,1).$$

Parameters of interest

$$\begin{aligned}\theta &= (p(1,1), \cdots, p(1,d-1), \cdots, p(d,d-1), \\ &\quad a_1(1), \cdots, a_1(d), \cdots, \sigma(1), \cdots, \sigma(d))'.\end{aligned}$$

The likelihood can be written by conditioning with respect to all possible paths

$$(e_1, \cdots, e_T)$$

of the Markov chain, with $e_i \in \mathcal{S}$. The probability of this path is

$$\mathbb{P}(e_1, \cdots, e_T) = \mathbb{P}(\Delta_1 = e_1, \cdots, \Delta_T = e_T) = \pi(e_1)p(e_1, e_2) \cdots p(e_{T-1}, e_T).$$

Likelihood

For any path, we have a likelihood

$$L_T^{(e_1,\cdots,e_T)}(X_1,\cdots,X_T) = \prod_{t=1}^{T} \phi_{e_t}(X_t - \sum_{k=1}^{p} a_k(e_t) X_{t-k}),$$

with $\phi_i(.)$ the density of $\mathcal{N}_{\mathbb{R}}(0, \sigma^2(i))$.
The likelihood of the observations is

$$L_T(X_1,\cdots,X_T;\theta) = \sum_{(e_1,\cdots,e_T)\in\mathcal{S}^T} L_T^{(e_1,\cdots,e_T)}(X_1,\cdots,X_T)\mathbb{P}(e_1,\cdots,e_T).$$

This likelihood is not tractable in practice.

There are several methods to compute the likelihood: algorithm based on a matrix product; the forward-backward algorithm (Baum, 1972); Hamilton filter (1989).

Hamilton filter

The method is based on the log-likelihood and provides filtered probabilities of the regimes.

Neglecting the distribution of X_1, we have

$$\log L_T(X_1, \cdots, X_T; \theta) = \sum_{t=1}^{T} \log \ f_t(X_t | X_{t-1}, \cdots, X_1),$$

with

$$f_t(X_t | X_{t-1}, \cdots, X_1) = \sum_{j=1}^{d} f_t(X_t | X_{t-1}, \cdots, X_1, \Delta_t = j) \mathbb{P}(\Delta_t = j | X_{t-1}, \cdots, X_1).$$

Hamilton filter

Let

$$\begin{aligned} \pi_{t|t-1}(j) &= \mathbb{P}(\Delta_t = j | X_{t-1}, \cdots, X_1), \\ \pi_{t|t}(j) &= \mathbb{P}(\Delta_t = j | X_t, \cdots, X_1). \end{aligned}$$

We have

$$\begin{aligned} \pi_{t+1|t}(j) &= \sum_{j=1}^{d} \mathbb{P}(\Delta_{t+1} = j | \Delta_t = i, X_{t-1}, \cdots, X_1) \pi_{t|t}(i) = \sum_{j=1}^{d} p(i,j) \pi_{t|t}(i), \\ \pi_{t|t}(j) &= \frac{g(X_t, \cdots, X_1 | \Delta_t = j) \pi(j)}{g(X_t, \cdots, X_1)}, \end{aligned}$$

using the formula $\mathbb{P}(A|X = x) = \frac{f(x|A)}{f(x)} \mathbb{P}(A)$.

Let

$$\eta_t(i) = \frac{1}{\sigma_i}(X_t - \sum_{k=1}^{p} a_k(i) X_{t-k})$$

Hamilton filter

$$
\begin{aligned}
\pi_{t|t}(j) &= \frac{g(X_t,\cdots,X_1|\Delta_t=j)\pi(j)}{g(X_t,\cdots,X_1)} \\
&= \frac{\frac{\phi(\eta_t(j))}{\sigma(j)} g(X_{t-1},\cdots,X_1|\Delta_t=j)\pi(j)}{g(X_t,\cdots,X_1)} \\
&= \frac{\frac{\phi(\eta_t(j))}{\sigma(j)} \mathbb{P}(\Delta_t=j|X_{t-1},\cdots,X_1)g(X_{t-1},\cdots,X_1)}{g(X_t|X_{t-1},\cdots,X_1)g(X_{t-1},\cdots,X_1)} \\
&= \frac{\frac{\phi(\eta_t(j))}{\sigma(j)} \pi_{t|t-1}(j)}{g(X_t|X_{t-1},\cdots,X_1)} \\
&= \frac{\frac{\phi(\eta_t(j))}{\sigma(j)} \pi_{t|t-1}(j)}{\sum_{i=1}^{d} \frac{\phi(\eta_t(i))}{\sigma(i)} \pi_{t|t-1}(i)}
\end{aligned}
$$

Hamilton filter

Finally, the sequences

$$\pi_{t|t-1}(j) = \mathbb{P}(\Delta_t = j | X_{t-1}, \cdots, X_1),\ \pi_{t|t}(j) = \mathbb{P}(\Delta_t = j | X_t, \cdots, X_1),$$

are obtained recursively by

$$\begin{cases} \pi_{t|t}(j) & = & \dfrac{\frac{\phi(\eta_t(j))}{\sigma(j)}\pi_{t|t-1}(j)}{\displaystyle\sum_{i=1}^{d} \frac{\phi(\eta_t(i))}{\sigma(i)}\pi_{t|t-1}(i)} \\ \pi_{t+1|t}(j) & = & \displaystyle\sum_{i=1}^{d} p(i,j)\pi_{t|t}(i). \end{cases}$$

with initial values $\pi_{1|0}(i) = \pi(i)$.

Hamilton filter in matrix form

Let

$$\pi_{t|t} = \begin{pmatrix} \pi_{t|t}(1) \\ \vdots \\ \pi_{t|t}(d) \end{pmatrix}, \pi_{t+1|t} = \begin{pmatrix} \pi_{t+1|t}(1) \\ \vdots \\ \pi_{t+1|t}(d) \end{pmatrix}, \Phi_t = \begin{pmatrix} \frac{\phi(\eta_t(1))}{\sigma(1)} \\ \vdots \\ \frac{\phi(\eta_t(d))}{\sigma(d)} \end{pmatrix}$$

Let $\odot$ the Hadamard product. We have

$$\pi_{t|t} = \frac{\pi_{t|t-1} \odot \Phi_t}{\iota' \{\pi_{t|t-1} \odot \Phi_t\}}, \pi_{t+1|t} = \mathbf{P}\pi_{t|t}.$$

Hamilton filter in matrix form

We finally obtain the conditional log-likelihood

$$\mathcal{L}_T(X_1, \cdots, X_T; \theta) = \sum_{t=1}^{T} \log\ f_t(X_t | X_{t-1}, \cdots, X_1),$$

with

$$f_t(X_t | X_{t-1}, \cdots, X_1) = \iota' \{\pi_{t|t-1} \odot \Phi_t\}.$$

Maximizing the likelihood

Maximizing the (log-)likelihood can be done by a classic optimization procedure or the EM algorithm. The intuition is as follows:

If in addition to the observations $(X_1, \cdots, X_T)$, one could observe $(\Delta_1, \cdots, \Delta_T)$, we could easily estimate θ and the initial law π_0 by MLE.

The EM algorithm alternates as follows: the E-step evaluates the likelihood given the current value of the parameters; the M-step maximizes the objective function computed in the E-step.

The EM algorithm

E-step Suppose we have an estimate $(\theta^{(k)}, \pi_0^{(k)})$ of (θ_0, π_0). Then it makes sense to approximate the unknown log-likelihood by its expectation given the observations $(X_1, \cdots, X_T)$ computed under the law $(\theta^{(k)}, \pi_0^{(k)})$.

M-step We aim at maximizing in (θ, π_0) the log-likelihood $\mathcal{L}_T(X_1, \cdots, X_T; \theta, \pi_0 | \theta^{(k)}, \pi_0^{(k)})$.