Hidden Markov models with financial applications

Jimmy Olsson

KTH Institute of Technology/Télécom ParisTech
jimmyol@kth.se

Lecture 1
Outline

1. A motivating example: S&P500
2. Hidden Markov models with discrete state space
3. Filtering and smoothing
4. Parameter estimation
Change in the schedule!

The computer session this afternoon is moved to next week! Tell your fellow students if they are not attending this lecture!

Slides!

These slides will be available at

http://www.maths.lth.se/matstat/staff/jimmy/

after the lecture.
Outline

1. A motivating example: S&P500
2. Hidden Markov models with discrete state space
3. Filtering and smoothing
4. Parameter estimation
A motivating example: S&P500

A brief look at the S&P500 index

Figure: Weekly log-returns of S&P500 from January 2, 2003 to September 28, 2012
Some characteristics of market data

- Roughly, **market data** can usually be characterized as
  - "bullish", i.e. increasing investor confidence and increased investing in anticipation of future price increases,
  - "neutral", or
  - "bearish", i.e. widespread investor fear and pessimism.

- In addition, equity market returns and volatility tend to move in opposite directions.
Some characteristics of market data (cont’d)

Figure: Statues of the two symbolic beasts of finance, the bear and the bull, in front of the Frankfurt Stock Exchange (Wikipedia).
A motivating example: S&P500 Hidden Markov models with discrete state space

Filtering and smoothing

Parameter estimation

A tentative stochastic model

■ Let \( \{ Y_t; t \in \mathbb{N} \} \) denote the sequence of (e.g.) weakly log-returns.

■ Let \((\pi(1), \pi(2), \pi(3))\) be a probability (row) vector.

■ Then a first attempt to model the returns could be to let

\[
Y_t \sim \begin{cases} 
N(\mu_1, \sigma_1^2) & \text{w. pr. } \pi(1), \\
N(\mu_2, \sigma_2^2) & \text{w. pr. } \pi(2), \\
N(\mu_3, \sigma_3^2) & \text{w. pr. } \pi(3),
\end{cases}
\]

where \( N \) denotes the normal distribution, \((\mu_1, \mu_2, \mu_3)\) are expected returns associated with the different trend types, and \((\sigma_1, \sigma_2, \sigma_3)\) are the corresponding volatilities.
A motivating example: S&P500

Hidden Markov models with discrete state space

Filtering and smoothing

Parameter estimation

A tentative stochastic model (cont’d)

Fitting such a mixture to the data yields

\((\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3) = (.005, -0.003, -0.002),\)

\((\hat{\sigma}_1, \hat{\sigma}_2, \hat{\sigma}_3) = (.013, .030, .082),\)

\((\pi(1), \pi(2), \pi(3)) = (.55, .42, .03).\)

Figure: Weekly log-returns of S&P500 from January 2, 2003 to September 28, 2012. Histogram superimposed with a mixture of three Gaussian distributions.
A motivating example: S&P500

Hidden Markov models with discrete state space  Filtering and smoothing  Parameter estimation

A tentative stochastic model (cont’d)

- However, estimating the autocorrelation shows clearly that the returns are serially correlated:

![Figure: Weekly log-returns of S&P500 from January 2, 2003 to September 28, 2012. Sample ACF and PACF of the squares of the log-returns.]

- Thus, this simple model fails to capture the “bullish” and the “bearish” structure of market data.
Goal of the day

Goal

To develop a more sophisticated modeling tool that allows also serial correlation to be captured, and to understand how to implement it on real data.
Outline

1. A motivating example: S&P500
2. Hidden Markov models with discrete state space
3. Filtering and smoothing
4. Parameter estimation
Markov chains

- In the following, let \( X = \{x_1, x_2, \ldots\} \) be a distinguished countable set referred to as the state space.

- A Markov transition matrix ("stochastic matrix") is a function \( M : X^2 \mapsto [0, 1] \) such that for all \( x \in X \),

\[
\sum_{x' \in X} M(x, x') = 1,
\]

i.e. each mapping \( x' \mapsto M(x, x') \), \( x \in X \), is a probability on \( X \).

- For any subset \( A \subseteq X \) and any function \( f \) on \( X \) we will write, for \( x \in X \),

\[
M(x, A) := \sum_{x' \in A} M(x, x') \quad \text{and} \quad Mf(x) := \sum_{x' \in X} M(x, x') f(x').
\]
Recall the following definition.

**Definition (Markov chain)**

Let $(\Omega, \mathbb{P})$ be some probability space and let $M$ and $\xi$ be some Markov transition matrix and probability on $X$, respectively. A family $\{X_t; t \in \mathbb{N}\}$ of random variables on $\Omega$ (“stochastic process”) is called a **homogeneous Markov chain** with Markov transition matrix $M$ and initial distribution $\xi$ if for all $t \in \mathbb{N}$ and all $\{x_0, \ldots, x_{t+1}\} \subseteq X$,

(i) $\mathbb{P}(X_0 = x_0) = \xi(x_0),$

(ii) $\mathbb{P}(X_{t+1} = x_{t+1} \mid X_0 = x_0, \ldots, X_t = x_t) = M(x_t, x_{t+1}).$

Given $X$, $M$, and $\xi$, the existence of such a stochastic process follows from the Kolmogorov existence theorem.
Hidden Markov models (HMM): a first glance

- The term HMM refers to a bivariate stochastic process \( \{(X_t, Y_t); t \in \mathbb{N}\} \) with the following properties.
  1. The marginal process (“state sequence”) \( \{X_t; t \in \mathbb{N}\} \) is a Markov chain.
  2. The Markov chain \( \{X_t; t \in \mathbb{N}\} \) is only partially observed through the observation process \( \{Y_t; t \in \mathbb{N}\} \) taking values in some general state space \( Y \).
  3. Conditionally on \( \{X_t; t \in \mathbb{N}\} \), the observations \( \{Y_t; t \in \mathbb{N}\} \) are independent and such that the conditional distribution of \( Y_t \) depends on \( X_t \) only.

- The Markov property of \( \{X_t; t \in \mathbb{N}\} \) implies that also \( \{(X_t, Y_t); t \in \mathbb{N}\} \) is a Markov chain.
Hidden Markov models (HMM): a first glance (cont’d)

- HMMs are often described loosely as “Markov chains observed in noise”:

```
\[ \cdots \rightarrow X_t \rightarrow X_{t+1} \rightarrow \cdots \]
```

**Figure:** Representation of the dependence structure of a hidden Markov model, where \( \{ Y_t; t \in \mathbb{N}\} \) are the observations and \( \{ X_t; t \in \mathbb{N}\} \) is the state sequence.
A motivating example: S&P500 Hidden Markov models with discrete state space Filtering and smoothing Parameter estimation

HMMs: definition

We now define formally an HMM. For this purpose, let again X and Y be countable and general state spaces, respectively, and let

1. $\xi$ be a probability distribution on X,
2. $M$ be a Markov transition matrix on X,
3. $G$ be a Markov transition kernel from X to Y, i.e. for all $x \in X$, the mapping

$$ Y \ni A \mapsto G(x, A) $$

is a probability measure on Y; in addition, for all (measurable) $B \subseteq X$, the mapping

$$ X \ni x \mapsto G(x, B) $$

is a measurable function.
HMMs: definition (cont’d)

- Using these quantities, define, on $X \times Y$, the Markov transition kernel

$$K : ((x, y), A \times B) \mapsto \sum_{x' \in A} M(x, x') G(x', B)$$

and the probability

$$\xi_G : A \times B \mapsto \sum_{x' \in A} \xi(x') G(x', B)$$

- An HMM is now defined formally as follows.

**Definition (HMM with discrete state space)**

An HMM is the canonical Markov chain \( \{(X_t, Y_t); t \in \mathbb{N}\} \) induced by \( K \) and \( \xi_G \).
Using this definition, one may show that the following holds.

**Theorem**

Let $\{(X_t, Y_t); t \in \mathbb{N}\}$ be an HMM defined as above. Then

(i) the state sequence $\{X_t; t \in \mathbb{N}\}$ is a Markov chain evolving according to $M$,

(ii) the components of any vector $(Y_0, \ldots, Y_t)$ of observations are conditionally independent given the corresponding states $(X_0, \ldots, X_t)$ such that the conditional distribution of $Y_t$ is $G(X_t, \cdot)$ (i.e. this distribution depends on $X_t$ only).
Consequently, the HMM can be viewed as a two-level stochastic model described by

1. **the state equation**

   \[ X_{t+1} \sim M(X_t, \cdot), \]

2. **the measurement equation**

   \[ Y_t \sim G(X_t, \cdot). \]

The state and measurement equations describe, together with the initial distribution \( \xi \), completely the system dynamics.
Marginal density of $Y_t$

In the case where each measure $G(x, \cdot)$, $x \in X$, admits a density $g$ with respect to some dominating measure $\mu$ (i.e. Lebesgue measure) we may write, for any $t$,

$$
\mathbb{E}_\xi(h(Y_t)) = \mathbb{E}_\xi(\mathbb{E}_\xi(h(Y_t) \mid X_t))
$$

$$
= \mathbb{E}_\xi\left(\int h(y_t)g(X_t, y_t) \mu(dy_t)\right)
$$

$$
= \int h(y_t) \sum_{x \in X} \mathbb{P}_\xi(X_t = x)g(x, y_t) \mu(dy_t).
$$

Thus, also the marginal distribution of $Y_t$ has a density (with respect to $\mu$) given by

$$
p_\xi(y_t) := \sum_{x \in X} \mathbb{P}_\xi(X_t = x)g(x, y_t).
$$
Autocorrelation of \( \{ Y_t; \ t \in \mathbb{N} \} \)

- In the case where \( \xi = \pi \) is a stationary distribution of \( M \) (if such a distribution exists), \( P_\xi(X_t = x) = \pi(x) \) for all time points \( t \), and for all \( t \in \mathbb{N} \),

\[
p_\xi(y_t) = \sum_{x \in X} \pi(x)g(x, y_t).
\]

- Let \( Gf(x) := \mathbb{E}(f(Y_0) \mid X_0 = x) = \int f(y)g(x, y)\mu(dy) \) and \( \Gamma(f) = \text{diag}(Gf(x); x \in X) \). Then the following holds.

**Theorem**

*If \( X \) is a finite set and \( \pi \) is a stationary distribution for \( M \), then*

\[
\text{Cov}_\pi(f(Y_t), f(Y_{t+h})) = \pi \Gamma(f) M^h \Gamma(f) \mathbf{1} - (\pi \Gamma(f) \mathbf{1})^2.
\]
The S&P500 index revisited

- Using HMMs, a refined model of the S&P500 index taking data dependence into account can be defined by letting
  \[ \{X_t; \ t \in \mathbb{N}\} \]
  be a Markov chain with transition matrix \( M \) on a state space \( X \) comprising two states corresponding to bullish and bearish markets. Assume that the chain starts in each state with equal probability.

- At each time step, a weekly log-return is generated as
  \[ Y_t \sim g(X_t, y) = \frac{1}{\sigma_{X_t} \sqrt{2\pi}} \exp \left( -\frac{(y - \mu_{X_t})^2}{2\sigma_{X_t}^2} \right), \]
  where the volatilities \( \sigma_i \) and the expected returns \( \mu_i \) need to be estimated along with the transition probabilities of \( \{X_t; \ t \in \mathbb{N}\} \).

- We will use this model as a working example throughout the first part of the course.
Outline

1. A motivating example: S&P500
2. Hidden Markov models with discrete state space
3. Filtering and smoothing
4. Parameter estimation
Optimal filtering

As we will see in the next, the following posterior distributions play a key role in practical implementations of HMMs. Let \((s, t) \in \mathbb{N}^2\) be such that \(s \leq t\) and let \(Y_{0:t} = (Y_0, \ldots, Y_t)\) be a given sequence of observations.

- **Filter distribution:** \(\phi_{t|t}(A) \overset{\text{(not.)}}{=} \phi_t(A) := P_\xi(X_t \in A \mid Y_{0:t}),\)
- **predictor distribution:** \(\phi_{t+1|t}(A) := P_\xi(X_{t+1} \in A \mid Y_{0:t}),\)
- **marginal smoothing distribution:** \(\phi_{s|t}(A) := P_\xi(X_s \in A \mid Y_{0:t})\)
- **joint smoothing distribution:** \(\phi_{0:t|t}(A) := P_\xi(X_{0:t} \in A \mid Y_{0:t}).\)

(Note that the dependence on \(Y_{0:t}\) is kept implicit.)

The term “optimal filtering” refers to the problem of computing online, as new observations become available, the filter and predictor distributions.
Optimal filtering

In the presence of a transition density $g$, the next theorem provides a solution to the optimal filtering problem.

**Theorem**

*For all $t \in \mathbb{N}$ and $x \in X$ the following holds.*

(i) $\phi_{t+1|t}(x) = \sum_{x' \in X} \phi_t(x') M(x', x)$ *(prediction)*

(ii) $\phi_{t+1}(x) = \frac{\phi_{t+1|t}(x) g(x, Y_{t+1})}{\sum_{x' \in X} \phi_{t+1|t}(x') g(x', Y_{t+1})}$ *(correction)*

The prediction and correction steps can be implemented online as follows.
The forward filtering algorithm

Data: \( \{ Y_t; t \in \mathbb{N} \} \)

Result: \( \{ \phi_t; t \in \mathbb{N} \}, \{ \phi_{t+1|t}; t \in \mathbb{N} \} \)

Initialization: set \( \phi_{0|-1}(x) \leftarrow \xi(x), \forall x \in X; \)

for \( t \leftarrow 0, 1, 2, \ldots \) do

- set \( c_t \leftarrow \sum_{x' \in X} \phi_{t|-1}(x')g(x', Y_t); \)
- set \( \phi_t(x) \leftarrow \phi_{t|-1}(x)g(x, Y_t)/c_t, \forall x \in X; \)
- set \( \phi_{t+1|t}(x) \leftarrow \sum_{x' \in X} \phi_t(x')M(x', x), \forall x \in X; \)

end

**Algorithm 1:** The forward filtering algorithm
Marginal smoothing

To solve the marginal smoothing problem, one uses the fact that the hidden chain is still Markov when evolving conditionally on the observations \( Y_{0:t} \). This is also true in the backward direction; indeed,

\[
p_\xi(x_s \mid x_{s+1:t}, y_{0:t}) = \frac{p_\xi(y_{0:s}, x_s, x_{s+1}, y_{s+1:t}, x_{s+2:t})}{\sum_{x_s' \in X} p_\xi(y_{0:s}, x_s', x_{s+1}, y_{s+1:t}, x_{s+2:t})} \]

\( = \frac{p_\xi(y_{0:s}, x_s, x_{s+1})}{\sum_{x_s' \in X} p_\xi(y_{0:s}, x_s', x_{s+1})} \)

\( = \frac{\phi_s(x_s) M(x_s, x_{s+1})}{\sum_{x_s' \in X} \phi_s(x_s') M(x_s, x_{s+1})} \)

\( = p_\xi(x_s \mid x_{s+1}, y_{0:s}). \)
Define the **backward transition matrix**

\[ B_{\phi_s}(x_{s+1}, x_s) := \frac{\phi_s(x_s)M(x_s, x_{s+1})}{\sum_{x'_s \in \mathcal{X}} \phi_s(x'_s)M(x_s, x_{s+1})} \]

describing the conditional distribution of \( X_s \) given \( X_{s+1} \) and \( Y_{0:s} \) (or \( Y_{0:n} \)).

Using the Markov property described above one verifies immediately the following.

**Theorem**

*For all \((s, t) \in \mathbb{N}^2 \) such that \( s < t \) it holds that*

\[ \phi_{s|t}(x_s) = \sum_{x_{s+1} \in \mathcal{X}} B_{\phi_s}(x_{s+1}, x_s) \phi_{s+1|t}(x_{s+1}). \]
The backward marginal smoothing algorithm

Consequently, having computed the filter distributions \( \{\phi_0, \phi_1, \ldots, \phi_t\} \) in a prefatory filtering pass, the marginal smoothing distributions can be computed in a backward manner using the following algorithm.

**Data:** \( \{\phi_0, \phi_1, \ldots, \phi_t\} \)

**Result:** \( \{\phi_0|t, \phi_1|t, \ldots, \phi_{t-1}|t\} \)

**for** \( s \leftarrow t - 1, t - 2, \ldots, 0 \)** **do**

- set \( \phi_{s|t}(x_s) = \sum_{x_{s+1} \in X} B_{\phi_s}(x_{s+1}, x_s) \phi_{s+1|t}(x_{s+1}), \forall x_s \in X; \)

**end**

**Algorithm 2:** The backward marginal smoothing algorithm
Note that for $s < t$,

$$p_\xi(x_s, x_{s+1} | y_0:t) = p_\xi(x_s | x_{s+1}, y_0:s)p_\xi(x_{s+1} | y_0:t)$$

$$= B_{\phi_s}(x_s, x_{s+1})\phi_{s+1|t}(x_{s+1}),$$

yielding the conditional distribution of consecutive states $X_s$ and $X_{s+1}$ given $Y_{0:t}$.

We will denote this bivariate distribution by $\phi_{s:s+1|t}$.

As we will see next, computation of distributions of this kind is crucial when calibrating HMMs.
Outline

1. A motivating example: S&P500
2. Hidden Markov models with discrete state space
3. Filtering and smoothing
4. Parameter estimation
An HMM contains unknown model parameters that need to be estimated (“calibrated”) on the basis of the observed data before the model can be used in practice (e.g. for prediction).

We let $\theta \in \Theta \subseteq \mathbb{R}^d$ be a vector containing these parameters and write $M_\theta$ and $g_\theta$ (and similarly for other quantities) to emphasize the dependence of the model on $\theta$.

We fix a distinguished element $\theta_* \in \Theta$ and assume that we have access to a single realization $Y_{0:t}$ of the observation process sampled under $\theta_*$. Thus, $\theta_*$ is interpreted as the “true” parameter, which is not known a priori.

Our aim is to estimate $\theta_*$ using the maximum likelihood method.
Given \( \{ Y_t; t \in \mathbb{N} \} \), the maximum likelihood estimator (MLE) is given by

\[
\hat{\theta}_t := \arg \max_{\theta \in \Theta} \ell_t(\theta),
\]

where

\[
\ell_t(\theta) := \log p_{\xi, \theta}(Y_{0:t})
\]

is the log-likelihood function.

The MLE can be shown to be strongly consistent, i.e.

\[
\hat{\theta}_t \to \theta_* \quad \text{as} \quad t \to \infty,
\]

under weak model assumptions.
Maximum likelihood estimation in HMMs

As

$$p_\xi(y_s \mid y_{0:s-1}) = \sum_{x_s \in X} p_{\xi,\theta}(y_s \mid x_s, y_{0:s-1}) p_{\xi,\theta}(x_s \mid y_{0:s-1})$$

$$= \sum_{x_s \in X} g_\theta(x_s, y_s) \phi_{s \mid s-1, \theta}(x_s)$$

$$= c_{s,\theta},$$

the log-likelihood may be computed pointwise as

$$\ell_t(\theta) = \log \left( p_\xi(y_0) \prod_{s=1}^{t} p_\xi(y_s \mid y_{0:s-1}) \right)$$

$$= \sum_{s=0}^{t} \log c_{s,\theta}.$$
A motivating example: S&P500 Hidden Markov models with discrete state space Filtering and smoothing Parameter estimation

The expectation-maximization (EM) algorithm

- Still, maximizing $\ell_t(\theta)$ is a complicated task. Nevertheless, as HMMs involve unobserved data, the problem of computing the MLE can be cast efficiently into the framework of the expectation-maximization (EM) algorithm.

- Let $p$ and $q$ be two probability densities on some common state space $E$ with respect to some reference measure $\lambda$. The EM algorithm uses the fact that the Kullback-Leibler divergence

$$K(p\|q) := \int \log \left( \frac{p(x)}{q(x)} \right) p(x) \lambda(dx) \geq 0$$

is always positive and zero only if and only if $p = q$ ($\lambda$-almost surely).
The EM algorithm (cont’d)

The algorithm goes as follows.

Data: Initial value $\theta_0$

Result: $\{\theta_\ell; \ell \in \mathbb{N}\}$

for $\ell \leftarrow 0, 1, 2, \ldots$ do

set $Q_t(\theta_\ell; \theta) \leftarrow \mathbb{E}_{\xi, \theta_\ell}(\log p_{\xi, \theta}(X_{0:t}, Y_{0:t}) \mid Y_{0:t}), \forall \theta$;

set $\theta_{\ell+1} \leftarrow \arg \max_{\theta \in \Theta} Q_t(\theta_\ell; \theta)$

end

Algorithm 3: The EM algorithm

The two steps within the main loop are referred to as expectation and maximization steps, respectively.
A motivating example: S&P500 Hidden Markov models with discrete state space
Filtering and smoothing
Parameter estimation

The EM inequality

- In order to understand the EM algorithm, define the entropy

\[ H_t(\theta'; \theta) := \ell_t(\theta) - Q_t(\theta'; \theta) \]
\[ = \log p_{\xi,\theta}(Y_{0:t}) - \sum_{x_{0:t} \in X_{t+1}} \log p_{\xi,\theta}(x_{0:t}, Y_{0:t}) \phi_{0:t|t,\theta'}(x_{0:t}) \]
\[ = - \sum_{x_{0:t} \in X_{t+1}} \log \phi_{0:t|t,\theta}(x_{0:t}) \phi_{0:t|t,\theta'}(x_{0:t}). \]

- Consequently,

\[ H_t(\theta; \theta') - H_t(\theta'; \theta') = \sum_{x_{0:t} \in X_{t+1}} \log \left( \frac{\phi_{0:t|t,\theta'}(x_{0:t})}{\phi_{0:t|t,\theta}(x_{0:t})} \right) \phi_{0:t|t,\theta'}(x_{0:t}) \]
\[ = K(\phi_{0:t|t,\theta'} \| \phi_{0:t|t,\theta}) \geq 0. \]
The EM inequality (cont’d)

- By rearranging the terms we obtain the following.

**Theorem (EM inequality)**

For all \((\theta, \theta') \in \Theta^2\) it holds that

\[
\ell_t(\theta) - \ell_t(\theta') \geq Q_t(\theta'; \theta) - Q_t(\theta'; \theta'),
\]

where the equality is strict unless \(\phi_{0:t|t, \theta'} = \phi_{0:t|t, \theta}\).

- Thus, by the very construction of \(\{\theta_\ell; \ell \in \mathbb{N}\}\) it is made sure that \(\{\ell_t(\theta_\ell); \ell \in \mathbb{N}\}\) is non-decreasing. Hence, the EM algorithm is a monotone optimization algorithm.
Convergence of EM

Under additional differentiability assumptions one may prove that
\[ \nabla_\theta \ell_t(\theta') = \nabla_\theta Q_t(\theta'; \theta) \big|_{\theta = \theta'} . \]

Thus, if the algorithm ever stops at \( \tilde{\theta} \), then the mapping \( \theta \mapsto Q_t(\tilde{\theta}; \theta) \) must be maximal at \( \tilde{\theta} \), which implies that \( \nabla_\theta \ell_t(\tilde{\theta}) = 0 \), i.e. \( \tilde{\theta} \) is a stationary point of the likelihood.

The “if the algorithm ever stops”-part has to be established rigorously and some more analysis is thus needed to proof the convergence. This is however possible.
In order to be practically useful, the E- and M-steps of EM have to be feasible. A rather general context in which this is the case is the following.

**Definition (Exponential family)**

The family \( \{ p_{\xi,\theta}(x_0:t, y_0:t); \theta \in \Theta \} \) defines an exponential family if the complete data likelihood is of form

\[
p_{\xi,\theta}(x_0:t, y_0:t) = \exp(\psi(\theta)^T S(x_0:t) - c(\theta)) h(x_0:t),
\]

where \( S \) and \( \psi \) are (possibly) vector-valued functions on \( X^{t+1} \) and \( \Theta \), respectively, and \( h \) is a non-negative real-valued function on \( X^{t+1} \). (All these quantities may depend on \( y_{0:n} \).)
EM in exponential families (cont’d)

- The intermediate quantity becomes

\[ Q_t(\theta'; \theta) = \psi(\theta) \mathbf{E}_{\xi,\theta'} (S(x_{0:n}) \mid Y_{0:t}) - c(\theta) \]
\[ + \mathbf{E}_{\xi,\theta'} (\log h(x_{0:n}) \mid Y_{0:t}), \]

where \((*)\) does not depend on \(\theta\) and may thus be ignored.

- Consequently, in order to be able to apply EM we need

1. to be able to compute the smoothed sufficient statistics

\[ s_t = \mathbf{E}_{\xi,\theta'} (S(x_{0:n}) \mid Y_{0:t}) = \sum_{x_{0:t} \in X^{t+1}} s(x_{0:t}) \phi_{0:t|t,\theta'} (x_{0:t}), \]

2. maximization of \(\theta \mapsto \psi(\theta) \mathbf{E}_{s_t} - c(\theta)\) to be feasible for all \(s_t\).
EM in exponential families (cont’d)

- As we will see, each component $S^{(i)}$ of the function $S$ is typically of additive form, i.e.

$$S^{(i)}(x_{0:t}) = \sum_{s=0}^{t-1} h_s(x_s, x_{s+1}),$$

where \(\{h_0, \ldots, h_{t-1}\}\) depend generally on $y_{0:t}$.

- The backward marginal smoother thus saves the day, as

$$\sum_{x_{0:t} \in \mathcal{X}^{t+1}} S^{(i)}(x_{0:t}) \phi_{0:t|t, \theta'}(x_{0:t})$$

$$= \sum_{s=0}^{t-1} \left( \sum_{x_{s:s+1} \in \mathcal{X}^2} h_s(x_s, x_{s+1}) \phi_{s:s+1|t, \theta'}(x_s, x_{s+1}) \right).$$
The S&P500 index revisited

- During the first computer session (next week!) we will, in order to estimate market trends, calibrate the S&P500 index model (with two states) using recent data.
- This model forms an exponential family (why?).
- In order to prepare for the computer session, do:

  Preparation, Computer session 1

1. Identify the functions \( S, h, \) and \( \psi; \)
2. think of how to estimate the vector of smoothed expectations \( s_t = \mathbb{E}_{\xi, \theta'}(S(x_0:n) | Y_{0:t}) \) given a data record \( Y_{0:t}; \)
3. given \( s_t \), think of how to maximize the mapping \( \theta \mapsto \psi(\theta)^\top s_t - c(\theta) \) and derive the updating formulas.

- More information will come!
What’s next?

- Implement an HMM with discrete state space in practice.
- Extension to models with continuous state space, such as the stochastic volatility model

\[
X_{t+1} = \alpha X_t + \sigma \epsilon_{t+1},
\]

\[
Y_t = \beta \exp(X_t/2) \epsilon_t,
\]

where \( \{\epsilon_t; t \in \mathbb{N}\} \) and \( \{\epsilon_t; t \in \mathbb{N}\} \) are white noise;

- particle filters;
- ...