1 Hidden Markov Models

Let $\{x_t, y_t\}_t$ be a hidden Markov model.

The following notations are used:

- $a_{m:n} = \{a_m, a_{m+1}, \dots, a_{n-1}, a_n\}$
- pdf: probability density function
- q_0 pdf of x_0
- $q_t(\cdot \mid x_{t-1})$ conditional pdf of x_t given x_{t-1}
- $q_{t',t}(\cdot \mid x_t)$ conditional pdf of $x_{t'}$ given x_t
- $g_{t_1,\dots,t_n}(\cdot \mid x_{t_1}, \dots, x_{t_n})$ conditional pdf of $y_{t_1}, \dots, y_{t_n}$ given $x_{t_1}, \dots, x_{t_n}$
- $p_t(\cdot \mid y_{0:t})$ conditional pdf of x_t given $y_{0:t}$
- $p_{t'|t}(\cdot \mid y_{0:t})$ conditional pdf of $x_{t'}$ given $y_{0:t}$
- $f_{T,h}(\cdot \mid y_{0:T})$ conditional pdf of y_{T+h} given $y_{0:T}$
- $\mathrm{p}(\cdot \mid \cdot)$ conditional pdf of its arguments.

Definition 1. *A hidden Markov model $\{x_t, y_t\}_t$ is a Markov process such that*

$$\mathrm{p}(x_t, y_t \mid x_{0:t-1}, y_{0:t-1}) = q_t(x_t \mid x_{t-1})\, g_t(y_t \mid x_t)$$

and $\mathrm{p}(x_0, y_0) = q_0(x_0)\, g_0(y_0 \mid x_0)$.

We use the convention $q_0(x_0 \mid x_{-1}) = q_0(x_0)$ in the sequel.

Lemma 1. *If two probability density functions are proportional, then they are equal.*

Proof. Let f and g be two pdf and c a real constant. If $f = c\,g$, then $1 = \int f = c \int g = c$, so $c = 1$. □

Proposition 1. *For all $t_1 < \dots < t_n$,*

$$g_{t_1,\dots,t_n}(y_{t_1}, \dots, y_{t_n} \mid x_{t_1}, \dots, x_{t_n}) = \prod_{k=1}^{n} g_{t_k}(y_{t_k} \mid x_{t_k}).$$

Proof.

$$
\begin{aligned}
& g_{t_1,\ldots,t_n}(y_{t_1},\ldots,y_{t_n} \mid x_{t_1},\ldots,x_{t_n}) \\
\propto\ & \mathrm{p}(x_{t_1},\ldots,x_{t_n},y_{t_1},\ldots,y_{t_n})
\end{aligned}
$$

according to Bayes' formula

$$
\begin{aligned}
&\propto \int\int \mathrm{p}(x_{0:t_n}, y_{0:t_n}) \prod_{t\notin\{t_1,\ldots,t_k\}} dx_t dy_t \\
&\propto \int\int \prod_{t=0}^{t_n} \mathrm{p}(x_t, y_t \mid x_{0:t-1}, y_{0:t-1}) \prod_{t\notin\{t_1,\ldots,t_k\}} dx_t dy_t
\end{aligned}
$$

according to the chain rule

$$
\propto \int\int \prod_{t=0}^{t_n} \mathrm{p}(x_t, y_t \mid x_{t-1}, y_{t-1}) \prod_{t\notin\{t_1,\ldots,t_k\}} dx_t dy_t
$$

according to the Markov property

$$
\propto \int\int \prod_{t=0}^{t_n} q_t(x_t \mid x_{t-1})\, g_t(y_t \mid x_t) \prod_{t\notin\{t_1,\ldots,t_k\}} dx_t dy_t
$$

according to Definition 1

$$
\begin{aligned}
\propto & [\int \prod_{t\notin\{t_1,\ldots,t_k\}} \underbrace{[\int g_t(y_t \mid x_t)\, dy_t]}_{=1} \prod_{t=0}^{t_n} q_t(x_t \mid x_{t-1}) \prod_{t\notin\{t_1,\ldots,t_k\}} dx_t] \\
& [\prod_{t\in\{t_1,\ldots,t_k\}} g_t(y_t \mid x_t)] \\
\propto & \prod_{t\in\{t_1,\ldots,t_k\}} g_t(y_t \mid x_t).
\end{aligned}
$$

□

Proposition 2. $\mathrm{p}(x_t \mid x_{t-1}, y_{0:t-1}) = q_t(x_t \mid x_{t-1})$.

Proof.

$$\begin{aligned}
\mathrm{p}(x_t \mid x_{t-1}, y_{0:t-1}) &= \int \mathrm{p}(x_t, y_t \mid x_{t-1}, y_{0:t-1})\, dy_t \\
&= \int \mathrm{p}(x_t, y_t \mid x_{t-1}, y_{t-1})\, dy_t
\end{aligned}$$

according to the Markov property

$$= q_t(x_t \mid x_{t-1}) \underbrace{\int g_t(y_t \mid x_t)\, dy_t}_{=1}$$

according to Definition 1. ◻

Proposition 3. *For all s,* $\mathrm{p}(y_t \mid x_t, y_{0:s}) = g_t(y_t \mid x_t)$.

Proof.

$$\mathrm{p}(y_t \mid x_t, y_{0:s}) \propto \mathrm{p}(y_t, y_{0:s} \mid x_t)$$

according to Bayes' formula

$$\propto \int \mathrm{p}(y_t, y_{0:s} \mid x_t, x_{0:s})\, \mathrm{p}(x_{0:s} \mid x_t)\, dx_{0:s}$$

according to the law of total probability

$$\propto g_t(y_t \mid x_t) \int \mathrm{p}(y_{0:s} \mid x_{0:s})\, \mathrm{p}(x_{0:s} \mid x_t)\, dx_{0:s}$$

according to Proposition 1

$$\propto g_t(y_t \mid x_t).$$

◻

Proposition 4 (Prediction)**.**

$$p_{t|t-1}(x_t \mid y_{0:t-1}) = \int q_t(x_t \mid x_{t-1})\, p_{t-1}(x_{t-1} \mid y_{0:t-1})\, dx_{t-1}.$$

Proof.

$$p_{t|t-1}(x_t \mid y_{0:t-1}) = \int \mathrm{p}(x_t \mid x_{t-1}, y_{0:t-1})\, p_{t-1}(x_{t-1} \mid y_{0:t-1})\, dx_{t-1}$$

according to the law of total probability

$$= \int q_t(x_t \mid x_{t-1})\, p_{t-1}(x_{t-1} \mid y_{0:t-1})\, dx_{t-1}$$

according to Proposition 2. □

Proposition 5 (Update)**.**

$$p_t(x_t \mid y_{0:t}) = \frac{g_t(y_t \mid x_t)\, p_{t|t-1}(x_t \mid y_{0:t-1})}{\int g_t(y_t \mid x_t)\, p_{t|t-1}(x_t \mid y_{0:t-1})\, dx_t}.$$

Proof.

$$p_t(x_t \mid y_{0:t}) \propto \mathrm{p}(y_t \mid x_t, y_{0:t-1})\, p_{t|t-1}(x_t \mid y_{0:t-1})$$

according to Bayes' formula

$$\propto g_t(y_t \mid x_t)\, p_{t|t-1}(x_t \mid y_{0:t-1})$$

according to Proposition 3. □

Proposition 6 (State forecast)**.**

$$p_{T+h|T}(x_{T+h} \mid y_{0:T}) = \int q_{T+h,T}(x_{T+h} \mid x_T)\, g_T(x_T \mid y_{0:T})\, dx_T.$$

Proof.

$$p_{T+h|T}(x_{T+h} \mid y_{0:T})$$
$$= \int \mathrm{p}(x_{T+h} \mid x_{T+h-1}, y_{0:T})\, \mathrm{p}(x_{T+h-1} \mid y_{0:T})\, dx_{T+h-1}$$

according to the total law of probability

$$= \int q_{T+h}(x_{T+h} \mid x_{T+h-1})\, \mathrm{p}(x_{T+h-1} \mid y_{0:T})\, dx_{T+h-1}$$

according to Proposition 2

$$= \int \cdots \int q_{T+h}(x_{T+h} \mid x_{T+h-1}) \cdots q_{T+1}(x_{T+1} \mid x_T)\, g_t(x_T \mid y_{0:T})\, dx_{T:T+h-1}$$

by recursion

$$= \int [\int q_{T+h}(x_{T+h} \mid x_{T+h-1}) \cdots q_{T+1}(x_{T+1} \mid x_T)\, dx_{T+1:T+h-1}]\, g_T(x_T \mid y_{0:T})\, dx_T.$$

Besides,

$$q_{T+h,T}(x_{T+h} \mid x_T) = \int q_{T+h}(x_{T+h} \mid x_{T+h-1}) \cdots q_{T+1}(x_{T+1} \mid x_T)\, dx_{T+1:T+h-1}$$

by recursion. □

Proposition 7 (Observation forecast)**.**

$$f_{T,h}(y_{T+h} \mid y_{0:T}) = \int g_{T+h}(x_{T+h} \mid y_{0:T+h})\, p_{T+h|T}(x_{T+h} \mid y_{0:T})\, dx_{T+h}.$$

Proof.

$$f_{T,h}(y_{T+h} \mid y_{0:T}) = \int \mathrm{p}(y_{T+h} \mid x_{T+h}, y_{0:T})\, p_{T+h|T}(x_{T+h} \mid y_{0:T})\, dx_{T+h}$$

according to the law of total probability

$$= \int g_{T+h}(y_{T+h} \mid x_{T+h})\, p_{T+h|T}(x_{T+h} \mid y_{0:T})\, dx_{T+h}$$

according to Proposition 3. □

2 Filtering in Hidden Markov Models

Bayesian filtering consists in computing at each time step $p_{t|t-1}(x_t \mid y_{0:t-1})$, called the predictor, and $p_t(x_t \mid y_{0:t})$, called the filter. The predictor and the filter obey to the recursive relation defined in Propositions 4 and 5.

The particle filter approximates the predictor and the filter recursively with a weighted sum of Dirac measures.

Suppose we have an approximation $\hat{p}_{t-1}$ of the filter at time $t-1$ as

$$\hat{p}_{t-1} = \sum_{i=1}^{N} w_{t-1}^i \, \delta_{\xi_{t-1}^i}$$

where the weights sum up to 1.

The approximation $\hat{p}_{t|t-1}$ of the predictor at time t is obtained by applying the Markov kernel to $\hat{p}_{t-1}$ as

$$\int q_t \, \hat{p}_{t-1} \approx \sum_{i=1}^{N} w_{t-1}^i \, \delta_{\xi_t^i} = \hat{p}_{t|t-1}$$

where $\xi_t^i \sim q_t(\cdot \mid \xi_{t-1}^i)$ independently for all $i \in \{1, \ldots, n\}$.

Then the approximation $\hat{p}_t$ of the new filter at time t is obtained by updating $\hat{p}_{t|t-1}$ as

$$g_t \, \hat{p}_{t|t-1} = \sum_{i=1}^{N} w_{t-1}^i \, g_t(\xi_t^i) \, \delta_{\xi_t^i} \propto \sum_{i=1}^{N} w_t^i \, \delta_{\xi_t^i} = \hat{p}_t$$

where $w_t^i = \dfrac{w_{t-1}^i \, g_t(\xi_t^i)}{\sum_{i=1}^{N} w_{t-1}^i \, g_t(\xi_t^i)}$ for all $i \in \{1, \ldots, n\}$.

With the recursive algorithm described above, it is commonly observed that after some iterations most particles get a zero weight and only a few particles get a significant weight. This phenomenon is called weight degeneracy and degrades the particle approximations. To avoid it, the particles are regenerated by multinomial resampling: a new sample is drawn from the particle set so that each particle has a probability to be drawn equal to its weight. Thus particles with a large weight tend to be duplicated whereas those with a small weight are discarded. The weights are then reset to one.

The particle filter is detailed in Algorithm 2. In this algorithm, multinomial resampling is performed when the effective sample size N_{eff} falls below a predefined threshold (typically $N_{\text{th}} = 2/3\, N$).

3 Forecasting in Hidden Markov Models

Forecasting consists in computing the distribution of y_{T+h} given the observed time series $y_{0:T}$. Propositions 6 and 7 are used to compute $p_{T,h}(x_{T+h} \mid y_{0:T})$, called the state forecast, then $f_{T,h}(y_{T+h} \mid y_{0:T})$, called the forecast.

Algorithm 1 Particle Filter

for $i \leftarrow 1 \cdots N$ **do**
  $\xi_0^i \sim q_0$
  $w_0^i \leftarrow g_0(y_0 \mid \xi_0^i)$
end for
normalize weights
for $t \leftarrow 1 \cdots h$ **do**
  for $i \leftarrow 1 \cdots N$ **do**
    $\xi_t^i \sim q_t(\cdot \mid \xi_{t-1}^i)$
    $w_t^i \leftarrow g_t(y_t \mid \xi_t^i)$
  end for
  normalize weights
  $N_{\text{eff}} \leftarrow 1/\sum_{i=1}^N w_t^i$
  if $N_{\text{eff}} < N_{\text{th}}$ **then**
    for $i \leftarrow 1 \cdots N$ **do**
      $\xi_t^i \sim \text{Multinomial}(\{\xi_t^1, \dots, \xi_t^N\}, \{w_t^1, \dots, w_t^N\})$
      $w_t^i \leftarrow 1/N$
    end for
  end if
end for

Suppose we have the particle approximation $\hat{p}_T$ of the filter at time T as

$$\hat{p}_T = \sum_{i=1}^N w_T^i \, \delta_{\xi_T^i}.$$

The approximation $\hat{p}_{T,h}$ of the state forecast is obtained by applying h times the Markov kernel to $\hat{p}_T$ as

$$\hat{p}_{T,h} = \int q_{T+1,T+h} \, \hat{p}_T \approx \sum_{i=1}^N w_T^i \, \delta_{\xi_{T+h}^i}$$

where $\xi_{T+1}^i \sim q_{T+1}(\cdot \mid \xi_T^i)$, ..., $\xi_{T+h}^i \sim q_{T+h}(\cdot \mid \xi_{T+h-1}^i)$.

The pdf approximation of the forecast is obtained by integrating $g_{T+h}(\cdot \mid y_{0:T+h})$ with respect to $\hat{p}_{T,h}$ as

$$\int g_T \, \hat{p}_{T,h} = \sum_{i=1}^N w_T^i \, g_{T+h}(\cdot \mid \xi_{T+h}^i).$$

The particle approximation $\hat{f}_{T,h}$ of the forecast is obtained by resampling the particles according to their weights, then drawing one observation per particle, so that

$$\hat{f}_{T,h} = \frac{1}{N} \sum_{i=1}^{N} \delta_{\zeta^i_{T+h}}$$

where $\zeta^i_{T+h} \sim g_{t+h}(\cdot \mid \xi^i_{T+h})$ and $\xi^i_{T+h} \sim \text{Multinomial}(\{\xi^1_{T+h}, \ldots, \xi^N_{T+h}\}, \{w^1_T, \ldots, w^N_T\})$ independently for all $i \in \{1, \ldots, N\}$.

We can then compute the empirical quantiles of the sample $\{\zeta^1_{T+h}, \ldots, \zeta^N_{T+h}\}$.

The particle forecasting algorithm is displayed in Algorithm 3.

Algorithm 2 Forecasting Algorithm

for $t \leftarrow 1 \cdots h$ **do**
  for $i \leftarrow 1 \cdots N$ **do**
    $\xi^i_{T+t} \sim q_{T+t}(\cdot \mid \xi^i_{T+t-1})$
  end for
end for
for $i \leftarrow 1 \cdots N$ **do**
  $\xi^i_{T+h} \sim \text{Multinomial}(\{\xi^1_{T+h}, \ldots, \xi^N_{T+h}\}, \{w^1_T, \ldots, w^N_T\})$
  $y^i_{T+h} \sim g_{t+h}(\cdot \mid \xi^i_{T+h})$
end for

4 Stochastic Volatility Model

Consider the model

$$\begin{aligned} y_t &= \mu + e^{x_t/2}\, \varepsilon_t \\ x_t &= \alpha + \beta\, x_{t-1} + \gamma\, \eta_t \end{aligned}$$

for $t \geq 1$, where $\{\varepsilon_t, \eta_t\}_t$ is a Gaussian white noise with covariance matrix

$$\text{var}[(\varepsilon_t, \eta_t)] = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix},$$

and with initial condition $x_0 \sim q_0$.