2. Hidden Markov Models

2.1. **Definition and examples.**

Definition. A hidden Markov model (HMM) is comprised of two stochastic processes. A hidden state process

$$X_0, X_1, \ldots$$

which is Markov: given $X_0 = x_0, \ldots, X_{k-1} = x_{k-1}$, the conditional probability density function of the next state is

$$\begin{aligned} p(x_k \mid x_0, \ldots, x_{k-1}) &= p(x_k \mid x_{k-1}) \\ &= f(x_{k-1}, x_k). \qquad (2.1) \end{aligned}$$

In the second line we have called the transition probability density function of the Markov chain $f(\cdot, \cdot)$. Thus $f(\cdot, \cdot)$ is a probability density function in its second argument, that is for all x_{k-1} we have

$$\int f(x_{k-1}, x_k) dx_k = 1.$$

The second process is the observed process

$$Y_0, Y_1, \ldots$$

Observation Y_k depends only on value of the (hidden) state at time k:

$$\begin{aligned} p(y_k \mid x_0, y_0, \ldots, y_{k-1}, x_{k-1}, x_k) &= p(y_k \mid x_k) \\ &= g(x_k, y_k). \end{aligned}$$

We have given the conditional probability density function of Y_k given $X_k = x_k$ specific notation $g(x_k, y_k)$. Thus $g(x_k, y_k)$ is a probability

$$
\begin{tikzcd}
 & Y_{t-1} & Y_t & \\
X_{t-2} \arrow[r, "{f(X_{t-2}, x_{t-1})}"'] & X_{t-1} \arrow[u, "{g(X_{t-1}, y_{t-1})}"'] \arrow[r, "{f(X_{t-1}, x_t)}"'] & X_t \arrow[u, "{g(X_t, y_t)}"'] \arrow[r, "{f(X_t, x_{t+1})}"'] & X_{t+1}
\end{tikzcd}
$$

FIGURE 2.1. Evolution of the random variables of a hidden Markov model.

density function in the second argument, that is for all x_k

$$\int g(x_k, y_k) dy_k = 1. \tag{2.2}$$

Figure 2.1 illustrates how the random processes of the hidden Markov model are generated.

We can write the joint probability density function of $(X_0, Y_0, \ldots, X_k, Y_k)$ generically as follows,

$$
\begin{aligned}
p(x_0, y_0, \ldots, x_k, y_k) =& p(y_k \mid x_0, y_0, \ldots, x_k) p(x_k \mid x_0, y_0, \ldots, x_{k-1}, y_{k-1}) \\
& \vdots \\
& \times p(y_1 \mid x_0, y_0, x_1) p(x_1 \mid x_0, y_0) \\
& \times p(y_0 \mid x_0) p(x_0).
\end{aligned}
$$

Now use the limited memory properties in the definition of the hidden Markov model to arrive at the following fact.

Fact. *For a hidden Markov model with transition probability density function $f(x_{k-1}, x_k)$ and observation probability density function $g(x_k, y_k)$, the joint probability density function of $(X_0, Y_0, \ldots, X_k, Y_k)$ is*

$$p(x_0, y_0, \ldots, x_k, y_k) = p(x_0) g(x_0, y_0) f(x_0, x_1) g(x_1, y_1) \cdots f(x_{k-1}, x_k) g(x_k, y_k).$$

Example. The Gaussian state-space model is a state-space model where driving noises are Gaussian.

$$X_{k+1} = aX_k + bW_{k+1},$$

$$Y_k = cX_k + dV_k, \qquad k = 0, 1, \ldots \tag{2.3}$$

where $\{V_k\}$ and $\{W_k\}$ are independent and identically distributed $\mathcal{N}(0,1)$ and X_0 is $\mathcal{N}(\bar{\mu}_0, \bar{\sigma}_0)$.

$$f(x_{k-1}, x_k) = \frac{1}{\sqrt{2\pi}b} \exp\left\{-\frac{(x_k - ax_{k-1})^2}{2b^2}\right\},$$

$$g(x_k, y_k) = \frac{1}{\sqrt{2\pi}d} \exp\left\{-\frac{(y_k - cx_k)^2}{2d^2}\right\}.$$

The next example is the stochastic volatility model which is widely used in the statistical modeling of financial time-series data.

Example. The stochastic volatility model is

$$X_k = aX_{k-1} + bW_k$$

$$Y_k = c \ \exp\left(X_k/2\right) V_k$$

where X_0 is $\mathcal{N}(\bar{\mu}_0, \bar{\sigma}_0)$ and $\{V_k\}$ and $\{W_k\}$ are both sequences of independent and identically distributed $\mathcal{N}(0, 1)$.

$$g(x_k, y_k) = \frac{1}{\sqrt{2\pi}c\exp(x_k/2)} \exp\left\{-\frac{y_k^2}{2c^2\exp(x_k)}\right\}.$$

Example. A discrete valued hidden process observed in Gaussian noise.

$X_k \in S = \{1, \ldots, n\}$ and

$$p(X_k = i_k \mid X_{k-1} = i_{k-1}) = P_{i_{k-1}, i_k}$$

for some transition probability matrix with row sum 1, i.e. $\sum_{j=1}^{n} P_{i,j} = 1$ for all rows i.

Given $X_k = i_k$, the observed process is

$$Y_k = c_{i_k} + d_{i_k} V_k$$

where $\{V_k\}$ are independent and identically distributed $\mathcal{N}(0,1)$ while $c_1, \ldots, c_n$ and $d_1, \ldots, d_n$ are real valued constants.

$$g(i_k, y_k) = \frac{1}{\sqrt{2\pi} d_{i_k}} \exp\left\{ -\frac{(y_k - c_{i_k})^2}{2 d_{i_k}^2} \right\}.$$

The law of $(X_0, Y_0, \ldots, X_k, Y_k)$ can be expressed using

$$p(i_0, y_0, \ldots, i_k, y_k) = p(i_0, i_1, \ldots, i_k) p(y_0, \ldots, y_k \mid i_0, \ldots, i_k)$$

where the hidden states is described by the probability mass function

$$p(i_0, i_1, \ldots, i_k) = p(i_0) P_{i_0, i_1} \cdots P_{i_{k-1}, i_k}$$

and the observed process by the probability density function

$$
\begin{aligned}
p(y_0, \ldots, y_k \mid i_0, \ldots, i_k) &= p(y_0 \mid i_0) \cdots p(y_k \mid i_k) \\
&= g(i_0, y_0) \cdots g(i_k, y_k).
\end{aligned}
$$

2.2. **Inference objectives.** Now that the hidden Markov model has been defined, we discuss some of the Bayesian inference objectives. Recall in a Bayesian treatment, we are interested in the conditional probability density function of the unobserved variables, conditioned on the observed ones.

Definition. Inference aims are as follows.

(i) Filtering: compute the conditional probability density function

$$p(x_k \mid y_0, \ldots, y_k)$$

if X_k are continuous random variables. If X_k are discrete valued compute the conditional probability mass function

$$p(i_k \mid y_0, \ldots, y_k).$$

(ii) Prediction: for $m > 0$, compute

$$p(x_{k+m} \mid y_0, \ldots, y_k) \qquad \text{or} \qquad p(i_{k+m} \mid y_0 \ldots, y_k).$$

(iii) Smoothing: for $m > 0$, compute

$$p(x_{k-m} \mid y_0, \ldots, y_k) \qquad \text{or} \qquad p(i_{k-m} \mid y_0 \ldots, y_k).$$

When X_k are discrete valued we can compute these exactly. When X_k is continuous valued, we can compute it exactly for the Gaussian state-space model. For all other HMMs with continuous valued X_k, these will be computed with the particle filter.

2.3. **Exact computations for the finite state hidden Markov model.** Given $p(i_k \mid y_0, \ldots, y_k)$, we can compute the filter at time $k+1$, i.e. $p(i_{k+1} \mid y_0, \ldots, y_{k+1})$. This will be done by first computing the intermediate conditional probability mass function $p(i_{k+1} \mid y_0, \ldots, y_k)$. Using Bayes' law,

$$
\begin{aligned}
p(y_{0:k}, i_k, i_{k+1}) &= p(y_{0:k}, i_k) p(i_{k+1} \mid y_{0:k}, i_k) \\
&= p(y_{0:k}, i_k) P_{i_k, i_{k+1}}.
\end{aligned}
$$

1 Thus

$$
\begin{aligned}
p(i_{k+1} \mid y_{0:k}) &= \frac{p(y_{0:k}, i_{k+1})}{p(y_{0:k})} \\
&= \frac{\sum_{i_k=1}^{n} p(y_{0:k}, i_k, i_{k+1})}{p(y_{0:k})} \\
&= \frac{\sum_{i_k=1}^{n} p(y_{0:k}, i_k) P_{i_k, i_{k+1}}}{p(y_{0:k})}.
\end{aligned}
$$

$$\boxed{p(i_{k+1} \mid y_{0:k}) = \sum_{i_k=1}^{n} p(i_k \mid y_{0:k}) P_{i_k, i_{k+1}} \qquad \text{(prediction step)}}$$

Define $\pi_k(i_k) = p(i_k \mid y_{0:k})$ and then the vector

$$\pi_k = [p(i_k = 1 \mid y_{0:k}), \ldots, p(i_k = n \mid y_{0:k})]^{\mathrm{T}}.$$

The prediction step can be expressed tersely with vectors and matrices as

$$p(i_{k+1} \mid y_{0:k}) = \left[\pi_k^{\mathrm{T}} P\right]_{i_{k+1}}.$$

To compute $\pi_{k+1}(i_{k+1}) = p(i_{k+1} \mid y_{0:k+1})$ of the update step use Bayes' law again,

$$p(i_{k+1}, y_{0:k}, y_{k+1}) = p(y_{0:k}, i_{k+1}) p(y_{k+1} \mid i_{k+1}) = p(y_{0:k}, i_{k+1}) g(i_{k+1}, y_{k+1})$$

$$p(i_{k+1} \mid y_{0:k+1}) = \frac{p(i_{k+1}, y_{0:k}, y_{k+1})}{\sum_{i_{k+1}=1}^{n} p(i_{k+1}, y_{0:k}, y_{k+1})}$$

$$= \frac{p(y_{0:k}, i_{k+1})g(i_{k+1}, y_{k+1})}{\sum_{i_{k+1}=1}^{n} p(y_{0:k}, i_{k+1})g(i_{k+1}, y_{k+1})}$$

Replace $p(y_{0:k}, i_{k+1})$ in numerator and denominator with

$$p(y_{0:k}, i_{k+1})/p(y_{0:k}) = p(i_{k+1} \mid y_{0:k})$$

to get the filter at time $k+1$.

$$\boxed{p(i_{k+1} \mid y_{0:k+1}) = \frac{p(i_{k+1} \mid y_{0:k})g(i_{k+1}, y_{k+1})}{\sum_{i_{k+1}=1}^{n} p(i_{k+1} \mid y_{0:k})g(i_{k+1}, y_{k+1})} \qquad \text{(update step)}}$$

Define the diagonal matrix (with all non-diagonal elements equal to zero)

$$B_{k+1} = \begin{bmatrix} g(1, y_{k+1}) & & \\ & \ddots & \\ & & g(n, y_{k+1}) \end{bmatrix}$$

then the update step (with the prediction step embedded within) can be expressed compactly as

$$\pi_{k+1}^{\mathrm{T}} = \frac{\pi_k^{\mathrm{T}} P B_{k+1}}{\pi_k^{\mathrm{T}} P B_{k+1} \mathbf{1}}$$

where $\mathbf{1} = [1, \ldots, 1]^{\mathrm{T}}$.