1. Informal Description of Hidden Markov Models

Suppose we have a set of ***N*** urns, each with balls of ***M*** colours in different proportions. According to some random process, I randomly choose an urn and then select a ball from this urn with replacement. You get to see the colour of the ball but not which urn the ball comes from. Then, I choose another urn and select another ball. The process is repeated.

This is a Hidden Markov Model. The urns are the "***hidden states***" and the colours of the balls are the "***observed signals***". A ***Markov chain*** governs the successive choices of the urns, i.e., the transition matrix of the Markov chain dictates the choice of the urns.

The main characteristics of an HMM are

1. The ***N*** hidden states.

2. The ***M*** distinct types of observations, which are the colours of the balls in this example. We could have had continuous valued observations of course.

3. The state transition probability matrix $\boldsymbol{A} = \{a_{jk}\}$ giving the conditional probability that we are in state k at time $(t+1)$ given that we were in state j at time t, i.e.,
   $a_{jk}= P[Q_{t+1}= S_k|\ Q_t= S_j]$,
   where S and Q are both used for denoting the states; S_1, S_2,…, S_N are the N states and Q_t is the state at time t.

4. The probability distribution of the observations, conditional on a state. In our example, this is the conditional probability of choosing a ball of a given colour, after the urn has been selected. $\boldsymbol{B_{jv}}$ = probability of observing a ball of *v-th* colour from the *j-th* urn.

5. The initial state distribution $\boldsymbol{\Pi}$, which is the probability distribution governing the initial choice of the states. The probability that the j-th urn is the first urn to be selected is Π_j.

Together, these parameters are denoted by $\boldsymbol{\lambda}$ in the literature.

2. The Three Basic Problems for HMM.

2.1 **Evaluation Problem.** Given the parameters of an HMM, i.e. given $\boldsymbol{\lambda}$, calculate the probability of realisation of a sequence of observations $\boldsymbol{O}$. (**Forward-Backward algorithms**)

That is, compute $P[O|\lambda]$

2.2 **Decoding or Classification Problem.** Given an observation sequence ***O***, find the sequence of hidden states most likely to have occurred. That is, compute

$$\underset{Q}{argmax}\ P[\boldsymbol{Q}|\boldsymbol{O}]$$

where $\boldsymbol{Q} = q_1, q_2, \ldots, q_T$ is the series of states indexed by time. (**Viterbi Algorithm**)

2.3 **Estimation or Training Problem.** Given a sequence of observations, fit the HMM. That is estimate $\boldsymbol{\lambda} = (\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\Pi})$ the parameters of the HMM, that maximize $P[O|\lambda]$. (**Baum-Welch Algorithm**, a type of EM algorithm**)**

3.Application of Hidden Markov Models.

The majority of the state-of-the art Automatic Speech Recognition systems employ HMM's. More recently, HMM's have been used in the prediction of gene sequences. We first describe the attempts to use HMM's to model phenomena other than Speech.

HMM's have been used in modelling daily rainfall in a city. It is quite common to use the rainfall data for each day of the year going back several years to compile rainfall summary statistics like average rainfall and its variability. However, some questions regarding rainfall cannot be answered unless there is a stochastic model for the rainfall-generating process.

E.g., we may ask the question "***What is the probability that there is more than X cm of rainfall in a given week and the longest dry run is no longer than k days***?"

So to answer a question like the one we posed, one would fit a stochastic model to the time series, and then simulate generate different sample paths and get the quantities of interest using Monte Carlo methods. A good model will capture the serial relationship between successive observations.

The time series of wet-dry days can often exhibit persistence or anti-persistence. After ***two dry days in Singapore, it is quite likely we get rain on the third day***. This type of modelling can have straightforward and useful parallels in the financial markets.

Researchers have also fit HMM's to a time series of the Old Faithful geyser's ***waiting times between successive eruptions*** and duration of the eruptions.

4. Applications in Speech Recognition

In Automated Speech Recognition systems, the task is to recognise an utterance known to have come from a dictionary of ***V*** different words. Each utterance is an acoustic signal or observation sequence $o_1, o_2, \ldots, o_T$.

The process $\{\boldsymbol{O_T}\}$ generated by two probabilistic models:

1. The ***hidden*** Markov Chain $\{\boldsymbol{Q_T}\}$ representing the configuration of the vocal tract at successive instants in time.

2. A set of probability distributions one for each state $\boldsymbol{Q_t}$ that produces the observations $\boldsymbol{o_t}$ from a known finite set of observations.

The "training process" involves taking multiple instances of utterances of each word and then fitting an HMM to them. So we fit an HMM for each word. The multiple training instance of the same word also typically includes the utterances at different speeds. Speakers can also pause between syllables.

The classification or decoding problem involves taking the sequence of the observations and computing the probability of this sequence under each of the different HMM's. The word is classified as corresponding to that HMM under which this probability is a maximum.