LECTURE 9

Let us consider a stochastic process $\{X_k\}_k$ and let us introduce the following definitions:

DEF 1: $\{X_k\}_k$ - Discrete-Time Markov Process $\overset{DEF}{\Longleftrightarrow}$ $F_{X_k}(x|\mathcal{H}_k) = F_{X_k}(x|X_{k-1})$ $\forall k, \forall x$

where: $F_{X_k}(x|\mathcal{H}_k) \triangleq P(X_k \leq x | \mathcal{H}_k)$ and $\mathcal{H}_k \triangleq (x_1, x_2, \ldots, x_{k-1})$ realization of the history up to k.

DEF 2: $\{X_k\}_k$ - Discrete-Time Markov Chain $\overset{DEF}{\Longleftrightarrow}$ $\{X_k\}_k$ is a Markov Process, $X_k \in \{e_1, e_2, \ldots, e_n\}$ - countable set $\forall k$

Note: $\{X_k\}_k$ is a discrete-time process, hence the definitions $\Rightarrow$ Analogously, a continuous-time MP or MC can be defined $\Rightarrow$ The distinguished feature of a Markov entity is the fact that each RV X_k depends on the realization of just the previous one

$\Downarrow$ a generalization....

DEF 1': $\{X_k\}_k$ - m-th order Discrete-Time Markov Process $\overset{DEF}{\Longleftrightarrow}$ $F_{X_k}(x|\mathcal{H}_k) = F_{X_k}(x|X_{k-1}, X_{k-2}, \ldots, X_{k-m})$ $\forall k, \forall x$ with $m>1$

Note: Our interest is in def. 2 (Markov Chains) because of the specific formulation that can be developed:

$$\{e_1, e_2, \ldots, e_n\} - \text{FINITE} \Rightarrow P(X_k = e_i | X_{k-1} = e_j) = p_{ji}(k-1, k) \quad \forall i, j \in \{1, 2, \ldots, n\}$$

$\Rightarrow$ We can organize the transition probabilities in a $n \times n$ matrix

$$\hat{P}(k-1,k) = \begin{bmatrix} p_{11}(k-1,k) & \ldots & & p_{1n}(k-1,k) \\ p_{21}(k-1,k) & p_{22}(k-1,k) & \ldots & \vdots \\ \vdots & & \ddots & \\ p_{n1}(k-1,k) & \ldots & & p_{nn}(k-1,k) \end{bmatrix}$$

Moreover, we can determine a general solution to the evolution problem:

2

$$\cdots \longrightarrow X_{k-2} \longrightarrow X_{k-1} \longrightarrow X_k \longrightarrow \cdots$$

$$P(X_k = e_i, X_{k-1} = e_j) = P(X_k = e_i \mid X_{k-1} = e_j)\, P(X_{k-1} = e_j)$$

$$P(X_k = e_i, X_{k-1} = e_j, X_{k-2} = e_r) = P(X_k = e_i \mid X_{k-1} = e_j)\, P(X_{k-1} = e_j \mid X_{k-2} = e_r)\, P(X_{k-2} = e_r)$$

$$\cdots$$

If we define $P_r(k) \triangleq P(X_k = e_r) \quad \forall k, r$, the equations above become:

$$P(X_k = e_i, X_{k-1} = e_j) = P_{ji}(k-1, k)\, P_j(k-1)$$

$$P(X_k = e_i, X_{k-1} = e_j, X_{k-2} = e_r) = P_{ji}(k-1, k)\, P_{rj}(k-2, k-1)\, P_r(k-2)$$

$$\cdots$$

Moreover, if we consider two non-consecutive RVs, e.g.:

$$P(X_k = e_i, X_{k-2} = e_r) = \sum_j P(X_k = e_i, X_{k-1} = e_j, X_{k-2} = e_r) =$$

$$= \sum_j P_{ji}(k-1, k)\, P_{rj}(k-2, k-1)\, P_r(k-2)$$

$$= P_{ri}(k-2, k)\, P_r(k-2) \qquad (a)$$

where: $P_{ri}(k-2, k) \triangleq \sum_j P_{ji}(k-1, k)\, P_{rj}(k-2, k-1)$

Therefore, for any starting point k_1 and any end point $k_2 > k_1$, we can define all the possible transitions by constructing the matrix:

$$\hat{P}(k_1, k_2) = \begin{bmatrix} P_{11}(k_1, k_2) & \cdots & P_{1n}(k_1, k_2) \\ P_{21}(k_1, k_2) & & \vdots \\ \vdots & \ddots & \\ P_{n1}(k_1, k_2) & & P_{nn}(k_1, k_2) \end{bmatrix}$$ - Transition Probability Matrix

A similar formulation can be derived when the set $\{e_1, e_2, \ldots, e_n, \ldots\}$ is not finite $\Rightarrow$ The transition matrices have infinite dimension

In general, the matrix $\hat{P}(k_1, k_2)$ depends on both k_1 and k_2. A special case is:

DEF 3: $\{X_k\}_k$ - Homogeneous Markov-Chain $\overset{DEF}{\iff}$ $\hat{P}(k_1, k_2) = \hat{P}(k_2 - k_1, 0) \quad \forall k_1, k_2 > k_1$

In particular, from this definition we derive:

$P_{ji}(X_k = e_i \mid X_{k-1} = e_j) = P_{ji}$ - constant $\forall k \Rightarrow$ The one-step transition matrix is time-invariant

Moreover, denoted with T the number of time step k s.t. $X_k = e_i$ for a given e_i, we have:

$$P(T > n+m \mid T > m) = \sum_{k=m+1}^{n+m} P(X_k = e_i \mid X_{k-1} = e_i) = n \cdot P_{ii} = P(T > n)$$

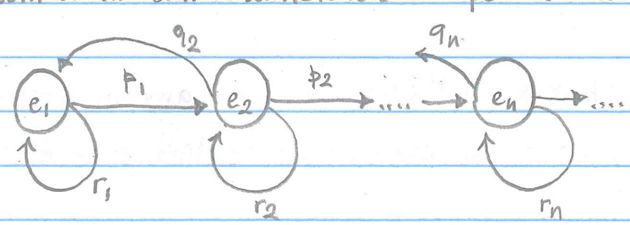
$\Rightarrow T$ is a geometric RV.

Finally, regardless of whether the MC is homogeneous or not, the following properties are satisfied:

$\sum_i P_{ji}(k_1, k_2) = \sum_i P(X_{k_2} = e_i \mid X_{k_1} = e_j) = 1 \Rightarrow$ The sum of the elements in a row of $\hat{P}(k_1, k_2)$ is always 1

$\sum_i P_{ji}(k_1, k_2) \neq \sum_j P_{ji}(k_1, k_2) \Rightarrow \hat{P}(k_1, k_2)$ is typically non-symmetric

Ex.: Random Walks can be envisioned as special Markov chains:



$p_i \triangleq P(X_k = e_{i+1} \mid X_{k-1} = e_i)$
$q_i \triangleq P(X_k = e_{i-1} \mid X_{k-1} = e_i)$
$r_i \triangleq P(X_k = e_i \mid X_{k-1} = e_i)$
$\forall k, \ i = 1, 2, \dots, n, \dots$

where X_k is the position on the graph above at time k

In this case, the transition matrix is:

$$\hat{P} = \begin{bmatrix} r_1 & p_1 & 0 & 0 & \cdots & 0 \\ q_2 & r_2 & p_2 & 0 & \cdots & 0 \\ 0 & q_3 & r_3 & p_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \end{bmatrix} \qquad \begin{array}{l} r_1 + p_1 = 1 \\ r_i + q_i + p_i = 1 \quad i > 1 \end{array}$$

← Infinite Tridiagonal Matrix

If the number n of states is finite, instead, we can have multiple scenarios at the first and last state (e_1, e_n). For instance:

$q_i = 0 \ \forall i$ (i.e., the evolution is in one direction only) $\Rightarrow$ $r_n = 1$, i.e., the last state is a sink

$$\left.\begin{array}{l} q_i = q \ \forall i \\ p_i = p \ \forall i \end{array}\right\} \Rightarrow r_i = 1 - q - p \quad 1 < i < n \quad \text{and} \quad \left\{\begin{array}{l} r_1 = 1 - p \\ r_n = 1 - q \end{array}\right.$$

, i.e., the boundary states act like "reflecting barriers"

$$\left.\begin{array}{l} r_i = 0 \ \forall i \\ p_n \triangleq P(X_k = e_1 \mid X_{k-1} = e_n) \\ q_1 \triangleq P(X_k = e_n \mid X_{k-1} = e_1) \\ p_i = p \quad q_i = q = 1 - p \quad \forall i \end{array}\right\} \Rightarrow \hat{P} = \begin{bmatrix} 0 & p & 0 & 0 & \cdots & q \\ q & 0 & p & 0 & \cdots & 0 \\ 0 & q & 0 & p & \cdots & 0 \\ \vdots & & \vdots & & \ddots & p \\ p & 0 & 0 & 0 & q & 0 \end{bmatrix}$$

i.e., a circular boundary is allowed

□

An important observation stems from example (a):

$$P_{ri}(k-2, k) = \left[P_{r1}(k-2, k-1) \ \ P_{r2}(k-2, k-1) \cdots P_{rn}(k-2, k-1) \right] \begin{bmatrix} P_{1i}(k-1, k) \\ P_{2i}(k-1, k) \\ \vdots \\ P_{ni}(k-1, k) \end{bmatrix}$$

$\Rightarrow \hat{P}(k-2, k) = \hat{P}(k-2, k-1)\, \hat{P}(k-1, k)$, i.e., any m-step-long evolution $(m > 1)$ can be computed by using the one-step transition probability matrices

⇓

For a homogeneous MC: $\hat{P}(k_1, k_2) = \hat{P}^{\,k_2 - k_1}$

Hence, a homogeneous MC is uniquely defined by $\hat{P}$ and the vector of initial conditions $\pi_0 \triangleq \begin{bmatrix} \pi_{01} \\ \pi_{02} \\ \vdots \\ \pi_{0n} \end{bmatrix}$, where $\pi_{0i} \triangleq P(X_0 = e_i)$ $i = 1, 2, 3, \ldots, n$

Let us now consider a homogeneous MC $\{X_k\}_k$ but let us assume that X_k is NOT accessible. At any time $k > 0$, a measure Y_k (which depends on X_k) is available, instead $\Rightarrow$ Y_k is a probabilistic function of X_k

$$\begin{tikzcd} \cdots \arrow[r] & X_k \arrow[r] \arrow[d] & X_{k+1} \arrow[r] \arrow[d] & \cdots \\ & Y_k & Y_{k+1} & \end{tikzcd}$$

DEF 4: $\{(X_k, Y_k)\}_k$ - Hidden Markov Model (HMM) $\overset{DEF}{\Longleftrightarrow}$ $\{X_k\}_k$ - Markov Chain, $F_{Y_k}(y \mid \mathcal{H}_k) = F_{Y_k}(y \mid X_k)$ $\forall y, k$

where $F_{Y_k}(\cdot)$ - conditional cdf of Y_k and $\mathcal{H}_k \triangleq \{(X_1, Y_1), (X_2, Y_2), \ldots, (X_{k-1}, Y_{k-1})\}$

In this definition, it is important that: (1) X_k is a latent variable (i.e., hidden variable) and (2) Y_k only depends on the current value of X_k at time k

Note: We will mainly deal with HMMs that satisfy:

- $\{X_k\}_k$ - homogeneous MC with $N > 1$ (finite) possible values (a.k.a. "states")
- $\{Y_k\}_k$ - can assume $L > 1$ (finite) possible values $\{y_1, y_2, \ldots, y_L\}$ (a.k.a. "alphabet")

Under these two additional conditions, the HMM is univocally defined by three constitutive models:

- π_0 ($N \times 1$ vector of initial probabilities for the state X for $k \leq 0$)
- $\hat{P}$ ($N \times N$ one-step transition probability matrix for X_k, $k \geq 1$)
- $\hat{Q}$ ($N \times L$ matrix of atomic probabilities of Y_k, $k \geq 1$)

i.e., $\hat{Q} = \{q_{ij}\}$ and $q_{ij} \triangleq P(Y_k = y_j \mid X_k = e_i)$

Note: The specific class of HMMs defined above is TIME-INVARIANT. More general variations can be obtained to model realistic scenarios. For instance:

Y_K - continuous variable $\Rightarrow$ One can choose: $f_{Y_K}(y|X_K) \sim$ Gaussian Mixture with parameters depending on X_K

Y_K is affected by an exogenous RV U_K $\Rightarrow$ One can augment the conditional probability model: $P(Y_K = y | X_K, U_K)$

Note: From the definition of HMM, the following multiplicative structure is obtained:

$$P(X_1, X_2, \dots, X_n, Y_1, Y_2, \dots, Y_n) = P(Y_1, Y_2, \dots, Y_n | X_1, X_2, \dots, X_n) P(X_1, X_2, \dots, X_n)$$

Probability of a specific sequence of states and output values

$$= \left(\prod_{k=1}^{n} P(Y_k|X_k) P(X_k|X_{k-1}) \right) P_0$$

initial Condition on X for k≤0

Ex.: Let us assume that, at each time k, Y_K is the outcome of a coin-tossing experiment, i.e., $Y_K \in \{T, H\}$. However, we do not know the rules of the experiment run at time k (e.g., if it requires tossing the coin many times and only reporting the last outcome; or, if there is a bias in the experiment; etc.) $\Rightarrow$ X_K is the type of experiment run to obtain Y_K

$\Downarrow$

In this case, even though we can't access X_K, we can assume that - for each type of experiment X_K - a probability function is defined for Y_K $\Rightarrow$ We want to infer the actual sequence of experiments $X_1, X_2, X_3, \dots$, from the sequence of observed head and tail values $Y_1, Y_2, Y_3, \dots$

Ex.: Biomed. applications for the concept of HMM can be found in the literature:

$Y_K \triangleq$ respiratory rate

$X_K \triangleq$ level of breathing regularity

$\Rightarrow$ The HMM is used to study apnea events in infants

$Y_k \triangleq$ DNA sequencing data
$X_k \triangleq$ methylation of DNA
$\Rightarrow$ The HMM is used to describe and detect cancer-related variations

$Y_k \triangleq$ sounds
$X_k \triangleq$ words
$\Rightarrow$ In speech recognition, HMMs are used to decode words from a sequence of sounds. □

Three problems are of interest when dealing with HMMs:

Evaluation Problem

P1: $\Sigma_i \triangleq (\pi_0, \hat{P}, \hat{Q})$ - known
$(y_1, y_2, \dots, y_n)$ - sequence of observations at time $k = 1, 2, \dots, n$
How to efficiently compute the joint probability $f(y_1, y_2, \dots, y_n | \Sigma_i)$ conditioned to the model Σ_i?

Decoding Problem

P2: $\Sigma_i \triangleq (\pi_0, \hat{P}, \hat{Q})$ - known
$(y_1, y_2, \dots, y_n)$ - known
How to estimate the best sequence of states $(x_1, x_2, \dots, x_n)$ to explain $(y_1, y_2, \dots, y_n)$?

Estimation Problem

P3: $(y_1, y_2, \dots, y_n)$ - known
How to estimate the best model Σ to explain the sequence $(y_1, y_2, \dots, y_n)$?

Note: Problem P1 is relevant since, for any given sequence $(y_1, y_2, \dots, y_n)$ of observations, the number of possible combinations of state values rapidly grows with the length n and the size N

Problem P2 tries to estimate the hidden part $\Rightarrow$ There is no "correct" solution but only an "optimal" one (which depends on the chosen optimality criteria)

Problem P3 deals with finding the values of the model parameters in Σ_i that maximize the likelihood to observe the specific sequence $(y_1, y_2, \dots, y_n)$ given Σ_i □

References:

L.R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," Proc. of the IEEE, vol. 77 (2), pp. 257-286, 1989

Z. Ghahramani, "An Introduction to Hidden Markov Models and Bayesian Networks," Int. J. Pattern Recog. & Artif. Intel., vol. 15 (1), pp. 9-42, 2001

A copy of both articles is available on Husky CT