(5/14 Monday)

PROFILE HMMs :

are used for aligning protein sequences.

HMMs :
- match (~~mutation~~) states
- insert states
- delete states – silent

Adapt algorithms to deal with silent states
(Book by Durbin et al)

CONTINUOUS TIME MARKOV CHAINS

- Markov process
- Time set is $\mathbb{T} = [0, \infty)$ (sometimes $\mathbb{T} = [0, T]$)
- State space is discrete (finite or countably infinite) – subset of $\mathbb{N} = \{0, 1, 2, \dots\}$ typically

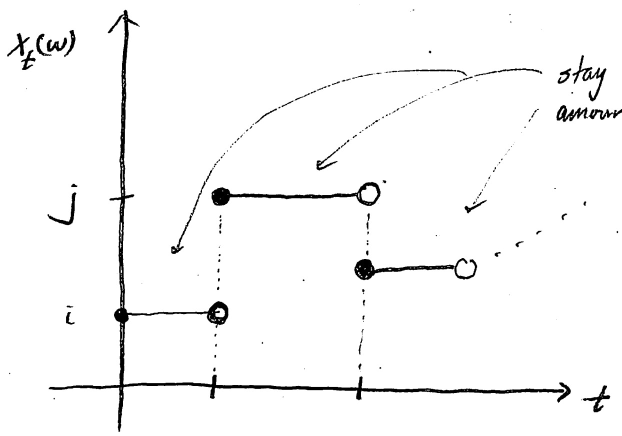
Markov Property :

$$\mathbb{P}(X_t = j \mid \underbrace{X_s = i,\ X_{s_n} = i_n, \dots, X_{s_1} = i_1}_{\text{past}}) \qquad \begin{cases} \forall\ 0 \le s_1 < \dots < s_n < s < t \\ i_1, \dots, i_n, i, j \in \mathbb{S} \end{cases}$$

$$= \mathbb{P}(X_t = j \mid X_s = i)$$

- Assume sample paths are right continuous.
- Assume time homogeneity:

$$\mathbb{P}(X_t = j \mid X_s = i) = \mathbb{P}(X_{t-s} = j \mid X_0 = i)$$



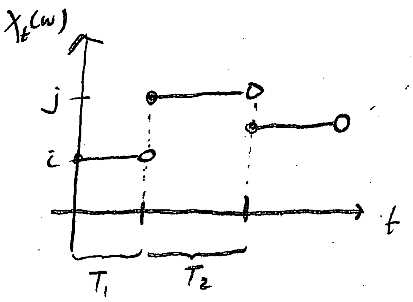
Notation $\mathbb{P}(X_t = j \mid X_0 = i) = P_{ij}(t)$

$$\{P_{ij}(t),\ t \geq 0\}$$

TWO WAYS OF DESCRIBING A CTS. TIME M.C.

① - Specify a transition matrix P for jumps of M.C. (on the diagonal, it only has zeroes & ones);
for each state i, specify $q(i) \in [0, \infty)$ that is the exponential rate of leaving state i
(if $P_{ii} = 0$, $q(i) > 0$; if $P_{ii} = 1$, $\underbrace{q(i)}_{\text{exponential holding rates}} = 0$)

Start in state i,
Hold in state i for an exponential amount of time with parameter $q(i)$ (if $q(i)=0$, stay at i forever)
If $q(i)>0$, when exponential holding time in i is up, jump to state j with Prob. P_{ij}.
When get to j, restart with new exponential holding time.

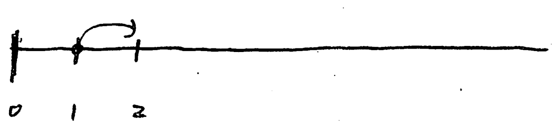


$T_1 \sim \exp(q(i))$
$T_2 \sim \exp(q(j))$

$\mathbb{P}(T_1 > x) = e^{-q(i)x}, \; x \geq 0$

If T is an exponential r.v.
$\mathbb{P}(T > t+s \mid T > s) = \mathbb{P}(T > t)$ <u>memoryless property</u>
(key to Markov property of cts time MC)

<u>One Problem</u>:



$P_{i,i+1} = 1$ for all $i = 0, 1, 2, \ldots$

$q(i) < \infty$,
it's possible that could only take a finite amount of time to reach ∞ (that's bad; called explosion)

Let T_i = amount of time spent in state $i \sim \exp(q(i))$

$i = 0, 1, 2, \ldots$

T = Time to get to ∞ = $\sum_{i=0}^{\infty} T_i$

$$E[T] = E\left[\sum_{i=0}^{\infty} T_i\right] = \sum_{i=0}^{\infty} E[T_i] = \sum_{i=0}^{\infty} 1/q(i)$$

If $\sum_{i=0}^{\infty} 1/q(i) < \infty$ then $E[T] < \infty \Rightarrow T < \infty$ a.s.

Running off to infinity is called <u>explosion</u>
If it occurs, minimal process X has $X(t) = \Delta$ for $t \geq T$
(Δ is a cemetery state not in $\mathbb{S}$)

Sufficient conditions that rule out explosion

(i) $\mathbb{S}$ is finite;
(ii) $\sup_i q(i) < \infty$
(iii) If start in i & i recurrent under P,
then no explosion from i.

P matrix is a transition matrix for skeleton or jump chain.

② <u>Infinitesimal description</u>

$Q = (Q_{ij})_{i,j \in \mathbb{S}}$ $\quad$ <u>(i)</u> $0 \leq -Q_{ii} < \infty$ for all $i \in \mathbb{S}$

(ii) $Q_{ij} \geq 0$ for all $i \neq j$

(iii) $\sum_{j \neq i} Q_{ij} = -Q_{ii}$ for all $i \in S$

(row sums are 0)

Infinitesimally: h small but arbitrary ($h>0$)

$$P_{ij}(h) = \mathbb{P}(X_h = j \mid X_0 = i) \qquad i \neq j$$

$$= Q_{ij} h + o(h) \qquad \frac{o(h)}{h} \to 0 \text{ as } h \to 0$$

$$P_{ij}(h) = 1 - \sum_{j \neq i} Q_{ij} h + o(h) \qquad i = j$$

$$= 1 + Q_{ii} h + o(h)$$

Connection to ①: h small

From ①: $P_{ij}(h) = q(i) h \cdot P_{ij} + o(h) \qquad i \neq j$

$$\Rightarrow Q_{ij} = q(i) P_{ij}$$

$$\& \; Q_{ii} = -\sum_{j \neq i} Q_{ij} = -q(i) \sum_{j \neq i} P_{ij} = \begin{cases} -q(i) & \text{if } q(i) > 0 \\ -q(i) & \text{if } q(i) = 0 \end{cases}$$

In matrix form (finitely many states; "good" cts time M.C.)

$P(h) = I + Qh + o(h)$ for h small.

$P(0) = I$

$$\frac{P(h) - P(0)}{h} = Q + \frac{o(h)}{h}, \text{ let } h \to 0, \quad \left.\frac{dP}{dt}\right|_{t=0} = Q$$

$$P(t) = \left(P_{ij}(t)\right)_{i,j \in S}$$

Q is called infinitesimal generator.

Markov Property

$\Rightarrow P(t+s) = P(s) \cdot P(t)$ (Chapman-Kolmogorov eqn.)

$P(t+h) = P(h)\, P(t) = P(t)\, P(h) \quad \forall\, t \geq 0,\ h > 0$

$$\Rightarrow \frac{dP(t)}{dt} = \underbrace{Q\,P(t)}_{\text{backward Kolmogorov eqn.}} = \underbrace{P(t)\,Q}_{\text{forward Kolm eqn.}}$$

$P(0) = I$

$P(t) = e^{Qt} = \sum_{n=0}^{\infty} \frac{Q^n t^n}{n!}$ (might or might not converge)
(converges if have finite S or "good" MC)

Read Ch II