n-step transition

If A is a matrix we write
$A = (a_{ij})_{i,j \in S}$

Definition 7.

Let X be a Markov chain. We set

$$P(m, m+n) = (p_{ij}(m, m+n))_{i,j} ,$$

with

$$p_{ij}(m, m+n) = P\left(X_{m+n} = j \middle| X_m = i\right)$$

$p_{ij} = P(X_{m+1} = j \mid X_m = i)$ $\quad P = (p_{ij})$

Remark:

- P describes the short term behavior of X
- $P(m, m+n)$ describes the long term behavior of X

Chapman-Kolmogorov equations

homogeneous

Theorem 8.

Let X be a Markov chain with transition p. Then

1. For $m, n, r \geq 0$ we have
$$p_{ij}(m, m+n+r) = \sum_k p_{ik}(m, m+n) p_{kj}(m+n, m+n+r)$$
2. As a matrix,

matrix product
$$P(m, m+n+r) = P(m, m+n)P(m+n, m+n+r)$$
3. In particular,
$$P(m, m+n) = P^n$$

First step of the proof : General identity.

Claim : $P(A \cap B \mid C) = P(A \mid B \cap C)\, P(B \mid C)$

Write the rhs as

$P(A \mid B \cap C) \quad P(B \mid C)$

$$= \frac{P((A \cap B) \cap C)}{P(B \cap C)} \quad \frac{P(B \cap C)}{P(C)}$$

$$= P(A \cap B \mid C) \longrightarrow \text{lhs}$$

Step 2: compute

$$\boxed{P_{ij}(m, m+n+r)} = P(X_{m+n+r} = j \mid X_m = i)$$

$$= \sum_{k \in S} P(\underbrace{X_{m+n+r} = j}_{A} \,,\, {}_{\cap}\, \underbrace{X_{m+n} = k}_{B} \mid \underbrace{X_m = i}_{C})$$

$$= \sum_{k} P(X_{m+n+r} = j \mid X_{m+n} = k, X_m = i)\, P(X_{m+n} = k \mid X_m = i)$$

Markov

$$= \sum_{k} \underbrace{P(X_{m+n+r} = j \mid X_{m+n} = k)}_{P_{kj}(m+n,\, m+n+r)} \underbrace{P(X_{m+n} = k \mid X_m = i)}_{P_{ik}(m,\, m+n)}$$

$$\boxed{= \sum_{k} P_{ik}(m, m+n)\, P_{kj}(m+n, m+n+r)}$$

Proof of Theorem 8 (1)

Preliminary identity:

$$\mathbf{P}(A \cap B |\, C) = \mathbf{P}(A |\, B \cap C)\mathbf{P}(B |\, C)$$

Proof: Start from right hand side,

$$\begin{aligned}
\mathbf{P}(A |\, B \cap C)\mathbf{P}(B |\, C) &= \frac{\mathbf{P}(A \cap B \cap C)}{\mathbf{P}(B \cap C)} \frac{\mathbf{P}(B \cap C)}{\mathbf{P}(C)} \\
&= \frac{\mathbf{P}\left((A \cap B) \cap C\right)}{\mathbf{P}(C)} \\
&= \mathbf{P}(A \cap B |\, C)
\end{aligned}$$

Proof of Theorem 8 (2)

Computation: We have

$$
\begin{aligned}
p_{ij}(m, m+n+r) &= \mathbf{P}(X_{m+n+r} = j \,|\, X_m = i) \\
&= \sum_k \mathbf{P}(X_{m+n+r} = j,\, X_{m+n} = k \,|\, X_m = i) \\
&= \sum_k \mathbf{P}(X_{m+n+r} = j \,|\, X_{m+n} = k,\, X_m = i)\mathbf{P}(X_{m+n} = k \,|\, X_m = i) \\
&= \sum_k \mathbf{P}(X_{m+n+r} = j \,|\, X_{m+n} = k)\mathbf{P}(X_{m+n} = k \,|\, X_m = i) \\
&= \sum_k p_{ik}(m, m+n)p_{kj}(m+n, m+n+r)
\end{aligned}
$$

Law of X_n

$\mu^{(n)} = [\mu_1^{(n)}, \dots, \mu_N^{(n)}]$

Proposition 9.

Consider the row vector

$$\mu_i^{(n)} = \mathbf{P}(X_n = i)$$

Then

$$\mu^{(m+n)} = \mu^{(m)} P^n$$

In particular,

$$\overset{\mathbb{R}^{1,N}}{\mu^{(n)}} = \overset{\mathbb{R}^{1,N}}{\mu^{(0)}} \, \overset{\mathbb{R}^{N,N}}{P^n}$$

$$\underbrace{[\quad]}_{\mu^{(n)}} = \underbrace{[\quad]}_{\mu^{(0)}} \underbrace{[\quad]}_{P^n}$$

Proof of Proposition 9

Computation: Write

$$
\begin{aligned}
\mu_j^{(m+n)} &= \mathbf{P}\left(X_{m+n} = j\right) \\
&= \sum_i \mathbf{P}\left(X_{m+n} = j \,|\, X_m = i\right) \mathbf{P}\left(X_m = i\right) \\
&= \sum_i \mu_i^{(m)} p_{ij}(m, m+n) \\
&= \mu^{(m)} P^n
\end{aligned}
$$

Rmk about the examples. One can get Markov chains in 2 ways

(i) Given a random dynamics (ex: at every step, flip a coin to know if we are going up or down). Then prove that we get a Markov chain

(ii) We are directly given some data, and assume that they come from a Markov chain

↳ P is given directly (see next example)

Example: weather in West Lafayette (1)

Model: We choose $S = \{1, \ldots, 6\} := \{VN, N, SN, SG, G, VG\}$.

Transition: from empirical data, we have found

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0.4 & 0.6 & 0 & 0 & 0 & 0 \\ 0.3 & 0 & 0.4 & 0.2 & 0.1 & 0 \\ 0 & 0 & 0 & 0.3 & 0.7 & 0 \\ 0 & 0 & 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0 & 0.8 & 0 & 0.2 \end{pmatrix}$$

ex: P(VN on Th | SN on W) = 30%
P(G on Th | SN on W) = 10%

Example: weather in West Lafayette (2)

Model: We choose $S = \{1, \ldots, 6\} := \{VN, N, SN, SG, G, VG\}$.

Prediction for J+2:

$$P^2 = \begin{pmatrix} 0.4 & 0.6 & 0 & 0 & 0 & 0 \\ 0.24 & 0.76 & 0 & 0 & 0 & 0 \\ 0.12 & 0.3 & 0.16 & 0.19 & 0.18 & 0.05 \\ 0 & 0 & 0 & 0.44 & 0.21 & 0.35 \\ 0 & 0 & 0 & 0.55 & 0.35 & 0.1 \\ 0 & 0 & 0 & 0.4 & 0.56 & 0.04 \end{pmatrix}$$

ex: P(VN on F | SN on W) = 12%
P(G on F | G on W) = 35%

Example: weather in West Lafayette (3)

Model: We choose $S = \{1, \ldots, 6\} := \{VN, N, SN, SG, G, VG\}$.

Prediction for J+28:

$$P^{28} = \begin{pmatrix} 0.29 & 0.71 & 0 & 0 & 0 & 0 \\ 0.29 & 0.71 & 0 & 0 & 0 & 0 \\ 0.14 & 0.36 & 7.2 \times 10^{-12} & 0.23 & 0.16 & 0.10 \\ 0 & 0 & 0 & 0.47 & 0.33 & 0.20 \\ 0 & 0 & 0 & 0.47 & 0.33 & 0.20 \\ 0 & 0 & 0 & 0.47 & 0.33 & 0.20 \end{pmatrix}$$

We see

(i) Initial data is lost after 28 days (Markov is short memory)

(ii) SN is a transient state

(iii) 2 clusters : {VN, N} and {SG, G, VG} → minimal classes