MATH 423

Linear Algebra II

# Lecture 25: Markov chains (continued). The Cayley-Hamilton theorem (continued).

# Markov chain

**Stochastic** (or **random**) **process** is a sequence of experiments for which the outcome at any stage depends on a chance.

We consider a simple model, with a finite set $S$ of possible outcomes (called **states**) and discrete time. Then the stochastic process is a sequence $s_0, s_1, s_2, \ldots,$ where all $s_n \in S$ depend on chance.

**Markov chain** is a stochastic process with discrete time such that the probability of the next outcome depends only on the previous outcome.

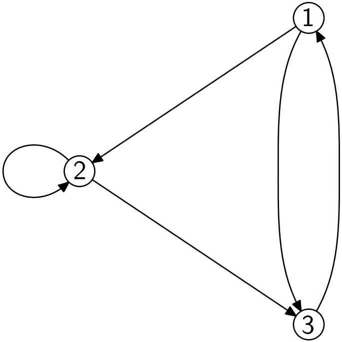
Let $S = \{1, 2, \dots, k\}$. The Markov chain is determined by **transition probabilities** $p_{ij}^{(t)}$, $1 \le i, j \le k$, $t \ge 0$, and by the **initial** probability distribution $q_i$, $1 \le i \le k$.

Here $q_i$ is the probability of the event $s_0 = i$, and $p_{ij}^{(t)}$ is the conditional probability of the event $s_{t+1} = j$ provided that $s_t = i$. By construction, $p_{ij}^{(t)}, q_i \ge 0$, $\sum_i q_i = 1$, and $\sum_j p_{ij}^{(t)} = 1$.

We shall assume that the Markov chain is **time-independent**, i.e., transition probabilities do not depend on time: $p_{ij}^{(t)} = p_{ij}$.

Then a Markov chain on $S = \{1, 2, \dots, k\}$ is determined by a **probability vector** $\mathbf{x}_0 = (q_1, q_2, \dots, q_k) \in \mathbb{R}^k$ and a $k \times k$ **transition matrix** $P = (p_{ij})$. The entries in each row of $P$ add up to 1.

# Example: random walk




$$\text{Transition matrix:} \quad P = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \\ 1 & 0 & 0 \end{pmatrix}$$

**Problem.** Find the (unconditional) probability distribution for any $s_n$, $n \geq 1$.

The probability distribution of $s_{n-1}$ is given by a probability vector $\mathbf{x}_{n-1} = (a_1, \ldots, a_k)$. The probability distribution of $s_n$ is given by a vector $\mathbf{x}_n = (b_1, \ldots, b_k)$.

We have

$$b_j = a_1 p_{1j} + a_2 p_{2j} + \cdots + a_k p_{kj}, \;\; 1 \leq j \leq k.$$

That is,

$$(b_1, \ldots, b_k) = (a_1, \ldots, a_k) \begin{pmatrix} p_{11} & \ldots & p_{1k} \\ \vdots & \ddots & \vdots \\ p_{k1} & \ldots & p_{kk} \end{pmatrix}.$$

$$\mathbf{x}_n = \mathbf{x}_{n-1}P \quad \Longrightarrow \quad \mathbf{x}_n^t = (\mathbf{x}_{n-1}P)^t = P^t\mathbf{x}_{n-1}^t.$$

Thus $\mathbf{x}_n^t = Q\mathbf{x}_{n-1}^t$, where $Q = P^t$ and the vectors are regarded as row vectors.

Then $\mathbf{x}_n^t = Q\mathbf{x}_{n-1}^t = Q(Q\mathbf{x}_{n-2}^t) = Q^2\mathbf{x}_{n-2}^t$.

Similarly, $\mathbf{x}_n^t = Q^3\mathbf{x}_{n-3}^t$, and so on.

Finally, $\boxed{\mathbf{x}_n^t = Q^n\mathbf{x}_0^t.}$

*Example.* Very primitive weather model:

Two states: "sunny" (1) and "rainy" (2).

Transition matrix: $P = \begin{pmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{pmatrix}$.

Suppose that $\mathbf{x}_0 = (1, 0)$ (sunny weather initially).

**Problem.** Make a long-term weather prediction.

The probability distribution of weather for day $n$ is given by the vector $\mathbf{x}_n^t = Q^n \mathbf{x}_0^t$, where $Q = P^t$.

To compute $Q^n$, we need to diagonalize the matrix $Q = \begin{pmatrix} 0.9 & 0.5 \\ 0.1 & 0.5 \end{pmatrix}$.

$$\det(Q - \lambda I) = \begin{vmatrix} 0.9-\lambda & 0.5 \\ 0.1 & 0.5-\lambda \end{vmatrix} =$$

$$= \lambda^2 - 1.4\lambda + 0.4 = (\lambda - 1)(\lambda - 0.4).$$

Two eigenvalues: $\lambda_1 = 1$, $\lambda_2 = 0.4$.

$$(Q - I)\mathbf{v} = \mathbf{0} \iff \begin{pmatrix} -0.1 & 0.5 \\ 0.1 & -0.5 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\iff (x, y) = t(5, 1), \ t \in \mathbb{R}.$$

$$(Q - 0.4I)\mathbf{v} = \mathbf{0} \iff \begin{pmatrix} 0.5 & 0.5 \\ 0.1 & 0.1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\iff (x, y) = t(-1, 1), \ t \in \mathbb{R}.$$

$\mathbf{v}_1 = (5, 1)^t$ and $\mathbf{v}_2 = (-1, 1)^t$ are eigenvectors of $Q$ belonging to eigenvalues 1 and 0.4, respectively.

$$\mathbf{x}_0^t = \alpha \mathbf{v}_1 + \beta \mathbf{v}_2 \iff \begin{cases} 5\alpha - \beta = 1 \\ \alpha + \beta = 0 \end{cases} \iff \begin{cases} \alpha = 1/6 \\ \beta = -1/6 \end{cases}$$

Now $\mathbf{x}_n^t = Q^n \mathbf{x}_0^t = Q^n(\alpha \mathbf{v}_1 + \beta \mathbf{v}_2) =$

$$= \alpha(Q^n \mathbf{v}_1) + \beta(Q^n \mathbf{v}_2) = \alpha \mathbf{v}_1 + (0.4)^n \beta \mathbf{v}_2,$$

which converges to the vector $\alpha \mathbf{v}_1 = (5/6, 1/6)^t$ as $n \to \infty$.

The vector $\mathbf{x}_\infty = (5/6, 1/6)$ gives the **limit distribution**. Also, it is a **steady-state** vector.

*Remarks.* In this example, the limit distribution does not depend on the initial distribution, but it is not always so. However 1 is always an eigenvalue of the matrix $P$ (and hence $Q$) since $P\,(1, 1, \ldots, 1)^t = (1, 1, \ldots, 1)^t$.

# Multiplication of block matrices

**Theorem** Suppose that matrices $X$ and $Y$ are represented as block matrices: $X = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$, $Y = \begin{pmatrix} P & Q \\ R & S \end{pmatrix}$.

Then $XY = \begin{pmatrix} AP + BR & AQ + BS \\ CP + DR & CQ + DS \end{pmatrix}$ provided that all matrix products are well defined.

**Corollary 1** Suppose that $(m+n)\times(m+n)$ matrices $X$ and $Y$ are represented as block matrices:

$$X = \begin{pmatrix} A & U \\ O & B \end{pmatrix}, \qquad Y = \begin{pmatrix} A_1 & U_1 \\ O & B_1 \end{pmatrix},$$

where $A$ and $A_1$ are $m\times m$ matrices, $B$ and $B_1$ are $n\times n$ matrices, and $O$ is the $n\times m$ zero matrix. Then $XY = \begin{pmatrix} AA_1 & U_2 \\ O & BB_1 \end{pmatrix}$ for some $m\times n$ matrix $U_2$.

**Corollary 2** Suppose that a square matrix $X$ is represented as a block matrix: $X = \begin{pmatrix} A & U \\ O & B \end{pmatrix}$, where $A$ and $B$ are square matrices and $O$ is a zero matrix. Then for any polynomial $p(x)$ we have $p(X) = \begin{pmatrix} p(A) & U_p \\ O & p(B) \end{pmatrix}$, where the matrix $U_p$ depends on $p$.

**Corollary 3** Using notation of Corollary 2, if $p_1(A) = O$ and $p_2(B) = O$ for some polynomials $p_1$ and $p_2$, then $p(X) = O$, where $p(x) = p_1(x)p_2(x)$.

*Proof:* We have $p(X) = p_1(X)p_2(X)$. By Corollary 2,

$$p_1(X) = \begin{pmatrix} O & U_{p_1} \\ O & p_1(B) \end{pmatrix}, \quad p_2(X) = \begin{pmatrix} p_2(A) & U_{p_2} \\ O & O \end{pmatrix}.$$

Multiplying these block matrices, we get the zero matrix.

# Cayley-Hamilton Theorem

**Theorem** If $A$ is a square matrix, then $p(A) = O$, where $p(x)$ is the characteristic polynomial of $A$, $p(\lambda) = \det(A - \lambda I)$.

*Proof for a complex matrix A:* The proof is by induction on the number $n$ of rows in $A$. The **base** of induction is the case $n = 1$. This case is trivial as $A = (a)$ and $p(x) = a - x$.

For the **inductive step**, we are to prove that the theorem holds for $n = k + 1$ assuming it holds for $n = k$ ($k$ any positive integer). Let $a_0$ be any complex eigenvalue of $A$ and $\mathbf{v}_0$ a corresponding eigenvector. Then $p(x) = (a_0 - x)p_0(x)$ for some polynomial $p_0$. Let us extend vector $\mathbf{v}_0$ to a basis for $\mathbb{C}^n$ (denoted $\alpha$). We have $A = UXU^{-1}$, where $U$ changes coordinates from $\alpha$ to the standard basis and $X$ is a block matrix of the form $X = \begin{pmatrix} a_0 & C \\ O & B \end{pmatrix}$.

# Cayley-Hamilton Theorem

We have $A = UXU^{-1}$, where $U$ changes coordinates from $\alpha$ to the standard basis and $X$ is a block matrix of the form

$$X = \left(\begin{array}{c|ccc} a_0 & c_1 & \dots & c_k \\ \hline 0 & & & \\ \vdots & & B & \\ 0 & & & \end{array}\right).$$

The characteristic polynomial of $X$ is $p$ since the matrix $X$ is similar to $A$. We know from the previous lecture that $p(x) = p_1(x)p_2(x)$, where $p_1$ and $p_2$ are characteristic polynomials of $(a_1)$ and $B$, resp. Since $p_1(x) = a_0 - x$ and $p(x) = (a_0 - x)p_0(x)$, we obtain $p_2(x) = p_0(x)$.

By the inductive assumption, $p_0(B) = O$. By Corollary 3, $p(X) = O$. Finally, $p(A) = Up(X)U^{-1} = UOU^{-1} = O$.

*Example.* $A = \begin{pmatrix} 2 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$.

Characterictic polynomial:

$$p(\lambda) = \det(A - \lambda I) = (2 - \lambda)(1 - \lambda)^2$$
$$= (2 - \lambda)(1 - 2\lambda + \lambda^2) = 2 - 5\lambda + 4\lambda^2 - \lambda^3.$$

By the Cayley-Hamilton theorem,

$$2I - 5A + 4A^2 - A^3 = O$$
$$\implies \tfrac{1}{2}A(A^2 - 4A + 5I) = I$$
$$\implies A^{-1} = \tfrac{1}{2}(A^2 - 4A + 5I).$$