# Stochastic Processes and Financial Mathematics (37363)

# Chapter 5

# Markov processes, discrete-time Markov chains

Alex Novikov and Scott Alexander

School of Mathematical and Physical Sciences, UTS

Autumn 2025

# Chapter outline

Topics:

- Definition and general properties
- Gaussian Markov processes
- Chapman-Kolmogorov equations
  - Continuous Markov process
  - Discrete Markov process (Markov chain)
  - Homogenous Markov chain
- Discrete-time homogenous Markov chains

# Definition and general properties

**Definition 1 (Markov process)**

A SP $X_t$, $t \in D$, is a Markov process if for each

$$t_1 \leq t_2 \leq \ldots \leq t_n \leq t < t + s$$

the conditional probability

$$P(X_{t+s} \leq y | X_{t_1} = x_1, \ldots, X_{t_n} = x_n, X_t = x) = P(X_{t+s} \leq y | X_t = x).$$

If the distribution $P(X_{t+s} \leq y | X_t = x)$ is independent of $t$, then the Markov processes $X_t$ is said to have homogeneous transition probabilities.

Markov processes may be discrete-time or continuous-time processes.

They may also be discrete-space or continuous-space processes.

**Markov property**

$$\boxed{P(\text{“future”} \mid \text{“past and present”}) = P(\text{“future”} \mid \text{“present”})}$$

**Remark.**
In terms of conditional characteristic functions, the Markov property means that for each

$$t_1 \leq t_2 \leq \ldots \leq t_n \leq t < t + s$$

and $u \in \mathbb{R}$

$$E[e^{iuX_{t+s}} | X_{t_1}, \ldots, X_{t_n}, X_t] = E[e^{iuX_{t+s}} | X_t]$$

reflecting the one-to-one correspondence between distributions and CFs.

# Definition and general properties

A discrete-time Markov process can be written as

$$X_t = g_{t-1}(X_{t-1}, Y_t), \quad t = 1, 2, \ldots,$$

where the function $g_t(x, y)$ is non-random and $X_0$, $Y_1$, $Y_2$, ... are independent RVs.

**Examples.**
The AR(1) process

$$X_t = \lambda X_{t-1} + e_t$$

is Markov but AR($p$) processes with $p = 2, 3, \ldots$ are not.

However, we can convert an AR($p$) process into a $p$-dimensional Markov process.

For example, consider the AR(2) process

$$X_t = \lambda_1 X_{t-1} + \lambda_2 X_{t-2} + e_t$$

where $e_t$ are independent RVs and $\lambda_2 \neq 0$, which obviously is not Markov.

# Definition and general properties

However, the 2-dimensional process

$$\boldsymbol{Z}_t = \begin{pmatrix} X_t \\ X_{t-1} \end{pmatrix} = \begin{pmatrix} \lambda_1 X_{t-1} + \lambda_2 X_{t-2} + e_t \\ X_{t-1} \end{pmatrix}$$
$$= \begin{pmatrix} \lambda_1 & \lambda_2 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} X_{t-1} \\ X_{t-2} \end{pmatrix} + \begin{pmatrix} e_t \\ 0 \end{pmatrix}$$
$$= \begin{pmatrix} \lambda_1 & \lambda_2 \\ 1 & 0 \end{pmatrix} \boldsymbol{Z}_{t-1} + \begin{pmatrix} e_t \\ 0 \end{pmatrix}$$

is a 2-dimensional Markov process.

Other examples include white noise, random walks, Brownian motion (BM), geometric Brownian motion (gBM) etc., all of which have homogeneous transition probabilities.

# Gaussian Markov processes

Many of the Gaussian processes considered in Chapter 4 are Markov processes, for which many useful results are known.

The first presented here provides a practical means for checking whether a Gaussian process is Markov.

**Theorem 1 (Gaussian-Markov criteria)**

Let $X_t$ $t \in D$, be a Gaussian process with covariance function $Q(t, s)$. Then $X_t$ is a Markov process i.f.f. for any $t_1 < t_2 < t_3$, $t_i \in D$,

$$Q(t_1, t_2)Q(t_2, t_3) = Q(t_1, t_3)Q(t_2, t_2) \quad (1)$$

or equivalently

$$E[X_{t_3}|X_{t_1}, X_{t_2}] = E[X_{t_3}|X_{t_2}] \quad (2)$$

**Proof.**
We aim to show that that Markovian $\Longrightarrow$ (2) $\Longrightarrow$ (1) $\Longrightarrow$ Markovian under the assumption that $\text{var}(X_t, X_t) = Q(t,t) > 0$ for all $t$.

*Proof that Markovian $\Longrightarrow$ (2) $\Longrightarrow$ (1).*
Note that

$$\begin{aligned} E[X_{t_3}|X_{t_1}, X_{t_2}] &= \int_{-\infty}^{\infty} x dF_{X_{t_3}|X_{t_1}, X_{t_2}}(x) \\ &= \int_{-\infty}^{\infty} x dF_{X_{t_3}|X_{t_2}}(x) \quad \text{(Makov property)} \\ &= E[X_{t_3}|X_{t_2}] \end{aligned}$$

which is (2).

By Part 2 of TNC (see Chapter 2) we know that

$$X_{t_3} - E[X_{t_3}|X_{t_1}, X_{t_2}] \text{ and } (X_{t_1}, X_{t_2})^T$$

are independent.

# Gaussian Markov processes

It follows from (2) that

$$X_{t_3} - E[X_{t_3}|X_{t_2}] \text{ and } (X_{t_1}, X_{t_2})^T$$

are independent so

$$X_{t_3} - E[X_{t_3}|X_{t_2}] \text{ and } X_{t_1}$$

must also be independent giving us

$$\mathrm{cov}(X_{t_3} - E[X_{t_3}|X_{t_2}], X_{t_1}) = 0$$

or more conveniently

$$\begin{aligned}
Q(t_1, t_3) &= \mathrm{cov}(X_{t_1}, X_{t_3}) = \mathrm{cov}(X_{t_1}, E[X_{t_3}|X_{t_2}]) \\
&= \mathrm{cov}\left(X_{t_1}, E[X_{t_3}] + \frac{Q(t_2, t_3)}{Q(t_2, t_2)}(X_{t_2} - E[X_{t_2}])\right) \\
&\text{(Part 3 TNC)} \\
&= \mathrm{cov}(X_{t_1}, X_{t_2})\frac{Q(t_2, t_3)}{Q(t_2, t_2)} = Q(t_1, t_2)\frac{Q(t_2, t_3)}{Q(t_2, t_2)}
\end{aligned}$$

which is (1).

# Gaussian Markov processes

*Proof that (1) $\implies$ Markovian.*

The first part of this requires showing that

$$X_{t+s} - E[X_{t+s}|X_t] \text{ and } (X_{t_1}, \ldots, X_{t_n})^T$$

are independent.

To do so note that for $t > t_1$ and $s > 0$

$$\begin{aligned}
\text{cov}(X_{t+s} - E[X_{t+s}|X_t], X_{t_1}) &= \text{cov}(X_{t+s}, X_{t_1}) - \text{cov}(E[X_{t+s}|X_t], X_{t_1}) \\
&= Q(t+s, t_1) - \text{cov}\left(E[X_{t+s}] + \frac{Q(t+s,t)}{Q(t,t)}(X_t - E[X_t]), X_{t_1}\right) \\
&\text{(Part 3 TNC)} \\
&= Q(t+s, t_1) - \frac{Q(t+s,t)}{Q(t,t)} Q(t, t_1) = 0
\end{aligned}$$

by (1) which by Part 1 TNC shows that

$$X_{t+s} - E[X_{t+s}|X_t] \text{ and } X_{t_1}$$

are independent.

# Gaussian Markov processes

Repeating this procedure (with suitable modification) we find

$$\operatorname{cov}(X_{t+s} - E[X_{t+s}|X_t], X_{t_i}) = 0$$

for $i \in \{1, \ldots, n\}$ giving us the independence of

$$X_{t+s} - E[X_{t+s}|X_t] \text{ and } (X_{t_1}, \ldots, X_{t_n})^T$$

as required. Then

$$\begin{aligned}
&E[e^{iuX_{t+s}}|X_{t_1}, \ldots, X_{t_n}, X_t] \\
&\quad = E\left[\frac{e^{iuE[X_{t+s}|X_t]}}{e^{iuE[X_{t+s}|X_t]}} e^{iuX_{t+s}} |X_{t_1}, \ldots, X_{t_n}, X_t\right] \\
&\quad = e^{iuE[X_{t+s}|X_t]} E[e^{iu(X_{t+s} - E[X_{t+s}|X_t])}|X_{t_1}, \ldots, X_{t_n}, X_t] \\
&\quad = e^{iuE[X_{t+s}|X_t]} E[e^{iu(X_{t+s} - E[X_{t+s}|X_t])}|X_t] \\
&\quad = e^{iuE[X_{t+s}|X_t]} e^{-iuE[X_{t+s}|X_t]} E[e^{iuX_{t+s}}|X_t] \\
&\quad = E[e^{iuX_{t+s}}|X_t].
\end{aligned}$$

# Gaussian Markov processes

By the one-to-one correspondence of CFs and distributions this can be restated as the Markov property

$$P(X_{t+s} \leq y | X_{t_1} = x_1, \ldots, X_{t_n} = x_n, X_t = x) = P(X_{t+s} \leq y | X_t = x).$$

---

Another important result provides the form of the covariance function for continuous, stationary Gaussian-Markov processes.

**Theorem 2 (continuous stationary Gaussian-Markov covariance function)**

Let $X_t$, $t \in \mathbb{R}$, be a continuous, stationary Gaussian-Markov process with covariance function $Q(s,t) = q(t-s)$. Then

$$Q(s,t) = q(t-s) = \sigma^2 e^{-\alpha|t-s|}, \;\; \sigma \geq 0, \;\; \alpha \geq 0,$$

where $\sigma^2 = var(X_t)$.

# Gaussian Markov processes

**Proof.**

Assume $\mathrm{var}(X_t) = Q(t,t) = q(0) > 0$ and note by stationarity

$$f(u) := \frac{q(u)}{q(0)} = \frac{\mathrm{cov}(X_t, X_{t+u})}{\sqrt{\mathrm{var}(X_t)}\sqrt{\mathrm{var}(X_{t+u})}} = \mathrm{corr}(X_t, X_{t+u})$$

so that

$$|f(u)| \leq 1$$

and further by stationarity

$$f(u) = \mathrm{corr}(X_t, X_{t+u}) = \mathrm{corr}(X_t, X_{t-u}) = f(-u).$$

Now let $u = t_2 - t_1$ and $v = t_3 - t_2$ and observe that

$$Q(t_1, t_2)Q(t_2, t_3) = Q(t_1, t_3)Q(t_2, t_2)$$

can be re-written as

$$q(u)q(v) = q(u+v)q(0)$$

or

$$\frac{q(u)}{q(0)}\frac{q(v)}{q(0)} = \frac{q(u+v)}{q(0)}$$

which is the functional equation

$$f(u)f(v) = f(u+v)$$

with solution $f(u) = e^{-\alpha u}$, unique among continuous functions.

But $f(u) = f(-u)$ so

$$f(u) = e^{-\alpha|u|}$$

where $\alpha \geq 0$ further restricts $0 < f(u) \leq 1$.

Then

$$Q(s,t) = q(t-s) = q(0)f(t-s) = \sigma^2 e^{-\alpha|t-s|}.$$

**Exercise.**
Show that a Brownian bridge $B_t^b$, $t \in [0,1]$, is a Markov process.

*Solution.* As class work.

# C-K eqns – continuous Markov process

Let $X_t$ be a Markov process (discrete or continuous-time).

If the values $X_t$ takes is an uncountable set (i.e. $X_t$ is continuous-space process) and $X_t$ has a density (i.e. $X_t$ is absolutely continuous), then the Markov property

$$P(X_{t+s} \leq y | X_{t_1} = x_1, \ldots, X_{t_n} = x_n, X_t = x) = P(X_{t+s} \leq y | X_t = x)$$

for

$$t_1 \leq t_2 \leq \ldots \leq t_n \leq t < t + s$$

is equivalent to

$$f_{X_{t+s}}(y | X_{t_1} = x_1, \ldots, X_{t_n} = x_n, X_t = x) = f_{X_{t+s}}(y | X_t = x)$$

where $f_{X_{t+s}}$ is conditional or transition density.

# C-K eqns – continuous Markov process

## Theorem 3 (C-K equation for continuous Markov process)

Let $X_t$ be a continuous Markov process (discrete or continuous-time) with transition density function $f(y, u|x, t)$.

Then

1. the $n$-dimensional density

$$f(x_1, t_1; \ldots; x_n, t_n) = f(x_1, t_1) \prod_{i=2}^{n} f(x_i, t_i | x_{i-1}, t_{i-1})$$

2. the Chapman-Kolmogorov equation

$$f(x_3, t_3 | x_1, t_1) = \int_{-\infty}^{\infty} f(x_2, t_2 | x_1, t_1) f(x_3, t_3 | x_2, t_2) dx_2$$

holds for any $t_1 < t_2 < t_3$.

# C-K eqns – discrete Markov process (Markov chain)

Let $X_t$ be a Markov process (discrete or continuous-time).

If the values $X_t$ takes is a countable set (i.e. $X_t$ is discrete-space process) then we call this a **Markov chain**, which satisfies the Markov property

$$P(X_{t+s} \leq y | X_{t_1} = x_1, \ldots, X_{t_n} = x_n, X_t = x) = P(X_{t+s} \leq y | X_t = x)$$

for

$$t_1 \leq t_2 \leq \ldots \leq t_n \leq t < t + s.$$

We shall obtain the equations for the transition probabilities of Markov chains

$$p(y, t|x, s) := P(X_t = y | X_s = x)$$

for both discrete and continuous-times cases in later sections.

# C-K eqns – discrete Markov process (Markov chain)

**Theorem 4 (C-K equation for Markov chain)**

Let $X_t$ be a Markov chain (discrete or continuous-time) with transition probabilities $p(y, u|x, t)$.

Then

1. the $n$-dimensional distribution

$$P(X_{t_1} = x_1, \ldots, X_{t_n} = x_n) = P(X_{t_1} = x_1) \prod_{i=2}^{n} p(x_i, t_i | x_{i-1}, t_{i-1})$$

2. the Chapman-Kolmogorov equation

$$p(x_3, t_3 | x_1, t_1) = \sum_{x_2} p(x_2, t_2 | x_1, t_1) p(x_3, t_3 | x_2, t_2)$$

holds for any $t_1 < t_2 < t_3$.

# C-K eqns – homogenous Markov chain

A Markov chain is called "homogenous" if the transition probabilities $p(y, u|x, t)$ don't depend on $t$ and $u$ but rather on $u - t$.

Set

$$p_{ij}(s) := P(X_{t+s} = x_j | X_t = x_i).$$

Then the Chapman-Kolmogorov equation becomes

$$p_{ij}(s + u) = \sum_k p_{ik}(s) p_{kj}(u) \tag{3}$$

for any $s \geq 0$ and $u \geq 0$.

# C-K eqns – homogenous Markov chain

To see this, set $s = t_2 - t_1$ and $u = t_3 - t_2$ so $s + u = t_3 - t_1$ into the Chapman-Kolmogorov equation from Theorem 4 so

$$\begin{aligned} p_{ij}(s+u) = p(x_j, t_3 | x_i, t_1) &= \sum_{x_k} p(x_k, t_2 | x_i, t_1) p(x_j, t_3 | x_k, t_2) \\ &= \sum_k p_{ik}(s) p_{kj}(u). \end{aligned}$$

It is often convenient to collect the transition probabilities together into the "transition probability matrix"

$$\boldsymbol{P}(s) = \big(p_{ij}(s)\big)_{1 \le i,j \le n} = \begin{pmatrix} p_{11}(s) & p_{12}(s) & \cdots & p_{1n}(s) \\ p_{21}(s) & p_{22}(s) & \cdots & p_{2n}(s) \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1}(s) & p_{n2}(s) & \cdots & p_{nn}(s) \end{pmatrix} \tag{4}$$

where $n$ is the dimension of the state space.

# C-K eqns – homogenous Markov chain

Note that all rows of $\boldsymbol{P}(s)$ sum to one, i.e.

$$\sum_{j=1}^{n} p_{ij}(s) = 1 \text{ for all } i \in \{1, \dots, n\},$$

as $X_t$ has to move to some state.

We also need the probability of the Markov chain being in particular states, so set

$$p_i(t) := P(X_t = x_i) \tag{5}$$

and the state probability vector

$$\boldsymbol{p}(t) = \big(p_i(t)\big)_{1 \le i \le n} = \begin{pmatrix} p_1(t) \\ p_2(t) \\ \vdots \\ p_n(t) \end{pmatrix}$$

where $\sum_{i=1}^{n} p_i(t) = 1$ as $X_t$ has to be in some state.

# C-K eqns – homogenous Markov chain

The following proposition brings these ideas together in convenient matrix form.

**Proposition 1 (homogenous Markov chain probabilities)**

Let $X_t$ be a homogenous Markov chain (discrete or continuous-time) taking $n$ states with transition probability matrix $\boldsymbol{P}(s)$ given by (4) and state probability vector $\boldsymbol{p}(t)$ given by (5).

Then the Chapman-Kolmogorov equation can be written in matrix notation as

$$\boldsymbol{P}(s+u) = \boldsymbol{P}(s)\boldsymbol{P}(u) \tag{6}$$

and state probability vector calculated as

$$\boldsymbol{p}(t+s) = \boldsymbol{P}^T(s)\boldsymbol{p}(t). \tag{7}$$

# C-K eqns – homogenous Markov chain

**Proof.**
It is obvious that (6) follows from (3) and (4).

To see that (7) holds, we consider the case $n = 2$ and note that

$$\begin{aligned}
\boldsymbol{P}^T(s)\boldsymbol{p}(t) &= \begin{pmatrix} p_{11}(s) & p_{21}(s) \\ p_{12}(s) & p_{22}(s) \end{pmatrix} \begin{pmatrix} p_1(t) \\ p_2(t) \end{pmatrix} \\
&= \begin{pmatrix} p_{11}(s)p_1(t) + p_{21}(s)p_2(t) \\ p_{12}(s)p_1(t) + p_{22}(s)p_2(t) \end{pmatrix} \\
&= \begin{pmatrix} p_1(t+s) \\ p_2(t+s) \end{pmatrix} \\
&\quad \text{(total law of probability – see Chapter 1)} \\
&= \boldsymbol{p}(t+s)
\end{aligned}$$

which generalises to the case $n = 3, 4, \ldots$ without difficulty.

# Discrete-time homogenous Markov chains

In this section we consider a discrete-time, homogeneous Markov chain $X_t$ with $n$ states and $t \in \{0, 1, 2, \ldots\}$.

First note from the Chapman-Kolmogorov equation (6) of Proposition 1 that the $s$-step transition probability matrix

$$\begin{aligned} \boldsymbol{P}(s) &= \boldsymbol{P}(1)\boldsymbol{P}(s-1) = \big(\boldsymbol{P}(1)\big)^2 \boldsymbol{P}(s-2) = \cdots = \big(\boldsymbol{P}(1)\big)^s \\ &= \boldsymbol{P}^s \end{aligned}$$

where $\boldsymbol{P} := \boldsymbol{P}(1)$ is the 1-step transition probability matrix.

So the state probability vector

$$\boldsymbol{p}(t+s) = (\boldsymbol{P}^T)^s \boldsymbol{p}(t) \quad \Longrightarrow \quad \boldsymbol{p}(t) = (\boldsymbol{P}^T)^t \boldsymbol{p}(0)$$

or in scalar form

$$p_j(t+s) = \sum_{i=1}^{n} p_i(t) p_{ij}(s) \quad \Longrightarrow \quad p_j(t) = \sum_{i=1}^{n} p_i(0) p_{ij}(t).$$

# Discrete-time homogenous Markov chains

**Example (a weather model).**
Let the Markov chain $X_t$ be a weather model with $n = 2$ states

1. $x_1 = 1$ (rain)
2. $x_2 = 2$ (no rain)

with 1-step transition probability matrix

$$\boldsymbol{P} = (p_{ij}) = \begin{pmatrix} \alpha & 1-\alpha \\ \beta & 1-\beta \end{pmatrix}.$$

---

Sometimes it makes more sense to number the states of the Markov chain from 0 to $n$, as the next example shows.

**Example (a gambling model).**
Let the Markov chain $X_t$ be the fortune of a gambler with states (wealth)

$$x_i = i \quad \text{for} \quad i \in \{0, 1, \ldots n\}$$

who quits when either $X_t = x_0 = 0$ or $X_t = x_n = n$.

# Discrete-time homogenous Markov chains

If the gambler gambles at each time step with probability of winning a dollar at each time step

$$p_{i,i+1}(1) = p$$

and of losing a dollar at each time step

$$p_{i,i-1}(1) = 1 - p,$$

then the 1-step transition probability matrix for the case $n = 3$ is

$$\boldsymbol{P} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1-p & 0 & p & 0 \\ 0 & 1-p & 0 & p \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

where $X_t = x_0 = 0$ and $X_t = x_n = n$ are **absorbing states.**

**LIMITING PROBABILITIES.**

**Theorem 5 (ergodic discrete-time Markov chain)**

A homogenous Markov chain $X_t$ is called ergodic if $\lim_{t\to\infty} \boldsymbol{p}(t)$ exists and this limit does not depend on the initial distribution $\boldsymbol{p}(0)$.

If a Markov chain is ergodic, then for all initial states $i$

$$\lim_{s\to\infty} p_{ij}(s) = \lim_{t\to\infty} p_j(t) = \pi_j \quad \text{or} \quad \lim_{t\to\infty} \boldsymbol{p}(t) = \boldsymbol{\pi}$$

so that

$$\lim_{t\to\infty} \boldsymbol{p}(t+1) = \lim_{t\to\infty} \left(\boldsymbol{P}^T \boldsymbol{p}(t)\right)$$

becomes

$$\boldsymbol{\pi} = \boldsymbol{P}^T \boldsymbol{\pi} \quad \text{subject to} \quad \sum_{j=1}^{n} \pi_j = 1, \tag{8}$$

where $\boldsymbol{\pi}$ is the stationary distribution of $X_t$.

# Discrete-time homogenous Markov chains

Note that if $X_t$ starts from the stationary distribution, i.e $\boldsymbol{p}(0) = \boldsymbol{\pi}$, then

$$\boldsymbol{p}(1) = \boldsymbol{P}^T \boldsymbol{p}(0) = \boldsymbol{P}^T \boldsymbol{\pi} = \boldsymbol{\pi}$$

and so

$$\begin{aligned}
\boldsymbol{p}(t) &= \left(\boldsymbol{P}^T\right)^t \boldsymbol{p}(0) \\
&= \left(\boldsymbol{P}^T\right)^{t-1} \boldsymbol{P}^T \boldsymbol{p}(0) = \left(\boldsymbol{P}^T\right)^{t-1} \boldsymbol{\pi} \\
&= \left(\boldsymbol{P}^T\right)^{t-2} \boldsymbol{P}^T \boldsymbol{\pi} = \left(\boldsymbol{P}^T\right)^{t-2} \boldsymbol{\pi} \\
&= \left(\boldsymbol{P}^T\right)^{t-3} \boldsymbol{P}^T \boldsymbol{\pi} = \left(\boldsymbol{P}^T\right)^{t-3} \boldsymbol{\pi} \\
&\vdots \\
&= \boldsymbol{P}^T \boldsymbol{\pi} = \boldsymbol{\pi}
\end{aligned}$$

for all $t \geq 0$.

**Example (a weather model) cont.**
Recall the weather model example with 1-step transition probability matrix

$$\boldsymbol{P} = (p_{ij}) = \begin{pmatrix} \alpha & 1-\alpha \\ \beta & 1-\beta \end{pmatrix}.$$

The stationary distribution $\boldsymbol{\pi} = (\pi_1, \pi_2)^T$ is the solution of the problem

$$\boldsymbol{\pi} = \boldsymbol{P}^T \boldsymbol{\pi}, \quad \sum_{i=1}^{2} \pi_i = 1 \quad \text{and} \quad \pi_1, \pi_2 \geq 0.$$

Note that this is an eigenvector problem corresponding to an eigenvalue of one, but with the added condition that the eigenvector components must be non-negative and sum to one as they are probabilities.

In scalar form we solve

$$
\begin{aligned}
\pi_1 &= \alpha\pi_1 + \beta\pi_2 \\
\pi_2 &= (1-\alpha)\pi_1 + (1-\beta)\pi_2 \\
1 &= \pi_1 + \pi_2
\end{aligned}
$$

which gives the unique stationary distribution

$$
\pi_1 = \frac{\beta}{1-\alpha+\beta} \quad \text{and} \quad \pi_2 = \frac{1-\alpha}{1-\alpha+\beta}
$$

so long as

$$
1-\alpha+\beta > 0
$$

which is when $\alpha \neq 1$ and $\beta \neq 0$.

# Discrete-time homogenous Markov chains

The case $\alpha = 1$ and $\beta = 0$ is considered separately.

If $\alpha = 1$ and $\beta = 0$ then

$$\boldsymbol{P} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

and in scalar form we solve

$$\begin{aligned} \pi_1 &= \pi_1 \\ \pi_2 &= \pi_2 \\ 1 &= \pi_1 + \pi_2 \end{aligned}$$

and, hence, any vector $(\pi_1, \pi_2)^T$ with $\pi_1 + \pi_2 = 1$, $\pi_1, \pi_2 \geq 0$, is a stationary distribution as

$$\boldsymbol{P}^T \boldsymbol{\pi} = \boldsymbol{\pi}$$

irrespective of $\boldsymbol{\pi}$.

---

We should not always expect the stationary distribution to be unique, but the following theorem tells us when we may.

# Discrete-time homogenous Markov chains

**Proposition 2 (ergodicity and stationarity)**

If a Markov chain is ergodic then the stationary distribution is unique.

**Proof.**
Assume there exist two different stationary distributions of a Markov chain, say $\boldsymbol{\pi}$ and $\boldsymbol{\pi}'$.

If we take $\boldsymbol{p}(0) = \boldsymbol{\pi}$, then $\boldsymbol{p}(t) = \boldsymbol{\pi}$ for all $t$ and so $\lim_{t\to\infty} \boldsymbol{p}(t) = \boldsymbol{\pi}$.

By the same reasoning, if we take $\boldsymbol{p}(0) = \boldsymbol{\pi}'$, then $\lim_{t\to\infty} \boldsymbol{p}(t) = \boldsymbol{\pi}'$.

But this contradicts the definition of ergodicity, which states that this limit should not depend on $\boldsymbol{p}(0)$.

# Discrete-time homogenous Markov chains

**Example**.
Consider a Markov chain with 1-step transition probability martrix

$$\boldsymbol{P} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

which has unique stationary distribution $\boldsymbol{\pi}^T = (\pi_1, \pi_2)^T = (\frac{1}{2}, \frac{1}{2})^T$.

But if $\boldsymbol{p}^T(0) = (1, 0)^T$ then

$$\begin{aligned} \boldsymbol{p}(1) = \boldsymbol{P}^T \boldsymbol{p}(0) &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \\ \boldsymbol{p}(2) = \boldsymbol{P}^T \boldsymbol{p}(1) &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \\ \boldsymbol{p}(3) = \boldsymbol{P}^T \boldsymbol{p}(2) &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \end{aligned}$$

etc., and we see that the limit does not exist contradicting ergodicity.

# Discrete-time homogenous Markov chains

The next theorem provides simple but sufficient conditions for ergodicity.

**Theorem 6 (sufficient condition for ergodicity)**

Let $X_t$ be a Markov chain taking a finite number of states. If there exists $s_0$ such that

$$\min_{i,j\geq 1} p_{ij}(s_0) > 0,$$

then $X_t$ is ergodic with unique stationary distribution $\boldsymbol{\pi}$ and for all $i$ and $j$

$$\lim_{s\to\infty} p_{ij}(s) = \lim_{t\to\infty} p_j(t) = \pi_j > 0.$$

Also, there exists $\lambda > 0$ and $c > 0$ such that for all $t$ and $j$

$$|p_{ij}(s) - \pi_j| < ce^{-\lambda s}.$$

**Proof**.
Rather technical so omitted.

# Discrete-time homogenous Markov chains

The next theorem provides another means for checking if a Markov chain is ergodic.

**Theorem 7 (sufficient condition for ergodicity)**

The Markov chain is ergodic when

1. all states communicate and the chain is aperiodic (see Definition 2 page 46).
2. the Markov chain is positively recurrent; i.e., starting in any state the mean time to return to that state is finite.

**Proof.**
Omitted.

**Example**.
Let $X_t$, $t \in \{0, 1, \ldots\}$, be a Markov chain with states $\{1, 2, 3\}$ and 1-step transition probability matrix

$$\boldsymbol{P} = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 \\ \frac{1}{6} & \frac{1}{3} & \frac{1}{2} \end{pmatrix}.$$

Then the 100-step transition probability matrix is

$$\boldsymbol{P}(100) = \boldsymbol{P}^{100} = \begin{pmatrix} 0.352941 & 0.294118 & 0.352941 \\ 0.352941 & 0.294118 & 0.352941 \\ 0.352941 & 0.294118 & 0.352941 \end{pmatrix}$$

and stationary distribution

$$\boldsymbol{\pi} = \begin{pmatrix} 6/17 \\ 5/17 \\ 6/17 \end{pmatrix} = \begin{pmatrix} 0.352941 \\ 0.294118 \\ 0.352941 \end{pmatrix}$$

The Mathematica example below shows how to calculate $\boldsymbol{P}^{100}$.

```
PP = {{0, 1/2, 1/2}, {1, 0, 0}, {1/6, 1/3, 1/2}};
MatrixPower[PP, 100] // N // MatrixForm
```

$$\begin{pmatrix} 0.352941 & 0.294118 & 0.352941 \\ 0.352941 & 0.294118 & 0.352941 \\ 0.352941 & 0.294118 & 0.352941 \end{pmatrix}$$

The Mathematica example below shows how to calculate $\boldsymbol{\pi}$.

```
PP = {{0, 1/2, 1/2}, {1, 0, 0}, {1/6, 1/3, 1/2}};
eigensys = Eigensystem[Transpose[PP]];
eigensys[[2, 1]] / Total[eigensys[[2, 1]]] // N // MatrixForm
```

$$\begin{pmatrix} 0.352941 \\ 0.294118 \\ 0.352941 \end{pmatrix}$$

The last line re-weights the eigenvectors so they are non-negative and sum to one.

# Discrete-time homogenous Markov chains

**CLASSIFICATION OF STATES.**

State $x_i$ is said to be **absorbing** if once entered it is never left, i.e. $p_{ii}(s) = 1$ and $p_{ij}(s) = 0$, $i \neq j$, for all $s > 0$.

State $x_j$ is said to be **accessible** from state $x_i$ if $p_{ij}(s) > 0$ for some $s > 0$.

Two states $x_i$ and $x_j$ that are accessible to each other are said to **communicate** and we write $i \longleftrightarrow j$.

Obviously, if $i \longleftrightarrow j$ and $j \longleftrightarrow k$ then $i \longleftrightarrow k$.

Two states that communicate are said to be the same **class**.

The Markov chain is said to be **irreducible** if there is only one class, that is all states communicate with each other.

# Discrete-time homogenous Markov chains

For any state $x_i$ we let $f_i$ denote the probability that, starting in state $x_i$, the process will reenter state $x_i$

State $x_i$ is said to be **recurrent** if $f_i = 1$ and **transient** if $f_i < 1$.

Denote

$$\xi_i = \sum_{t=0}^{\infty} I(X_t = x_i),$$

$I$ the indicator function, as the number of time periods that the process is in state $x_i$.

If state $x_i$ is transient then, starting in $x_i$, the RV $\xi_i$ has the geometric distribution; i.e.

$$P(\xi_i = k | X_0 = x_i) = f_i^{k-1}(1 - f_i), \quad k = 1, 2, \ldots,$$

and therefore

$$E[\xi_i | X_0 = x_i] = \frac{1}{1 - f_i} < \infty.$$

# Discrete-time homogenous Markov chains

If state $x_i$ is recurrent then, starting in $x_i$, the process $X_t$ visits the state $x_i$ infinitely many times and therefore

$$E[\xi_i | X_0 = x_i] = \infty.$$

The following results assist in identifying recurrent states.

**Proposition 3 (recurrent communicating states)**

If state $x_i$ of a Markov chain is recurrent and $i \longleftrightarrow j$, then state $x_j$ is recurrent.

**Proposition 4 (recurrent states in irreducible Markov chain)**

All states in a finite-state irreducible Markov chain are recurrent.

**Proofs.**
Omitted.

---

**Example.**
Consider a Markov chain consisting of four states $\{1,2,3,4\}$ with the 1-step transition probability matrix

$$\boldsymbol{P} = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

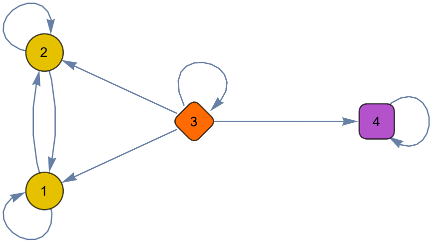
The **classes** of this Markov chain are $\{1,2\}$, $\{3\}$ and $\{4\}$.

State 4 is an **absorbing** state.

# Discrete-time homogenous Markov chains

We can use Mathematica to plot a network diagram of the Markov chain, which can assist in classifying states and identifying classes.

```
PP = {{1/2, 1/2, 0, 0}, {1/2, 1/2, 0, 0}, {1/4, 1/4, 1/4, 1/4}, {0, 0, 0, 1}};
Graph[DiscreteMarkovProcess[4, PP]]
```




We see that states 1, 2 and 3 are transient states and state 4 is recurrent.

The class $\{3\}$ is transient and classes $\{1, 2\}$ and $\{4\}$ are recurrent.

**Example.**
Consider a Markov chain consisting of five states {1,2,3,4,5} with the 1-step transition probability matrix

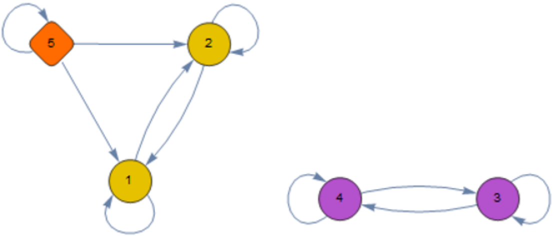
$$\boldsymbol{P} = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 \\ 1/4 & 1/4 & 0 & 0 & 1/2 \end{pmatrix}$$

The **classes** of this Markov chain are $\{1, 2\}$, $\{3, 4\}$ and $\{5\}$.

# Discrete-time homogenous Markov chains

Using Mathematica to visualise the Markov chan state transitions.

```
PP = {{1 / 2, 1 / 2, 0, 0, 0}, {1 / 2, 1 / 2, 0, 0, 0}, {0, 0, 1 / 2, 1 / 2, 0}, {0, 0, 1 / 2, 1 / 2, 0},
   {1 / 4, 1 / 4, 0, 0, 1 / 2}};
Graph[DiscreteMarkovProcess[5, PP]]
```




All states are transient.

The class $\{5\}$ is transient and classes $\{1, 2\}$ and $\{3, 4\}$ are recurrent.

# Discrete-time homogenous Markov chains

To wrap up this section we need the concept of the **period** of states of a Markov chain.

**Definition 2 (period of Markov chain state)**

State $x_i$ is said to have period $d_i$ if $p_{ii}(s) = 0$ whenever $s$ is not divisible by $d_i$.

The period of state $x_i$, $d_i$, is the greatest common divisor of $s$ with this property.

A state $x_i$ with period $d_i = 1$ is said to be aperiodic.

**Proposition 5 (period as class property)**

Periodicity is a class property, i.e. if $i \longleftrightarrow j$ and state $x_i$ has period $d_i$, then state $j$ also has period $d_j = d_i$.

**Example.**
Consider the Markov chain with 1-step transition probability matrix

$$\boldsymbol{P}(1) = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

Let's use Mathematica for the calculations.

```
PP = {{0, 0, 1}, {1, 0, 0}, {0, 1, 0}};
Table[MatrixPower[PP, n] // MatrixForm, {n, {1, 2, 3, 4, 5, 6}}]
```

$$\left\{ \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right\}$$

Then $d_i = 3$ for $i \in \{1, 2, 3\}$.

# Discrete-time homogenous Markov chains

**Theorem 8 (irreducible, aperiodic finite-state Markov chain)**

For an irreducible aperiodic finite-state Markov chain,

$$\lim_{t \to \infty} \boldsymbol{p}(t) = \boldsymbol{\pi} = (\pi_1, \ldots, \pi_n)^T$$

exists, is independent of initial distribution (Markov chain is ergodic) and can be found by solving (8).

Also, for any initial distribution,

$$\lim_{s \to \infty} p_{ij}(s) = \pi_j$$

exists and is independent of $i$.

# References I