STOCHASTIC PROCESSES

Basic notions

Often the systems we consider evolve in time and we are interested in their dynamic behaviour, usually involving some randomness.

- the length of a queue
- the number of students passing the course S-38.143 each year
- the temperature outside
- the number of data packets in a network

A stochastic process X_t (or $X(t)$) is a family of random variables indexed by a parameter t (usually the time).

Formally, a stochastic process is a mapping from the sample space $\mathcal{S}$ to functions of t.

With each element e of $\mathcal{S}$ is associated a function $X_t(e)$.

- For a given value of e, $X_t(e)$ is a function of time ("a lottery ticket e with a plot of a function is drawn from a urn")

- For a given value of t, $X_t(e)$ is a random variable

- For a given value of e and t, $X_t(e)$ is a (fixed) number

The function $X_t(e)$ associated with a given value e is called the realization of the stochastic process (also trajectory or sample path).

State space: the set of possible values of X_t

Parameter space: the set of values of t

Stochastic processes can be classified according to whether these spaces are discrete or continuous:

	State space	
Parameter space	Discrete	Continuous
Discrete	$*$	$**$
Continuous	$***$	$****$

According to the type of the parameter space one speaks about discrete time or continuous time stochastic processes.

Discrete time stochastic processes are also called random sequences.

In considering stochastic processes we are often interested in quantities like:

- Time-dependent distribution: defines the probability that X_t takes a value in a particular subset of $\mathcal{S}$ at a given instant t

- Stationary distribution: defines the probability that X_t takes a value in a particular subset of $\mathcal{S}$ as $t \to \infty$ (assuming the limit exists)

- The relationships between X_s and X_t for different times s and t (e.g. covariance or correlation of X_s and X_t)

- Hitting probability: the probability that a given state is $\mathcal{S}$ will ever be entered

- First passage time: the instant at which the stochastic process first time enters a given state or set of states starting from a given initial state

The nth order statistics of a stochastic process X_t is defined by the joint distribution

$$F_{X_{t_1},\ldots,X_{t_n}}(x_1,\ldots,x_n) = P\{X_{t_1} \leq x_1,\ldots,X_{t_n} \leq x_n\}$$

for all possible sets $\{t_1,\ldots,t_n\}$.

A complete characterization of a stochastic process X_t requires knowing the stochastics of the process of all orders n.

1st order statistics:

Stationary distribution

$$F(x) = \lim_{t\to\infty} P\{X_t \leq x\}$$

Expectation (at time t)

$$\bar{X}_t = E[X_t]$$

2nd order statistics:

Covariance (autocovariance)

$$R_{t,s} = E[(X_t - \bar{X}_t)(X_s - \bar{X}_s)]$$

Stationary process

The statistics of all the orders are unchanged by a shift in the time axis

$$F_{X_{t_1+\tau},\ldots,X_{t_n+\tau}}(x_1,\ldots,x_n) = F_{X_{t_1},\ldots,X_{t_n}}(x_1,\ldots,x_n) \qquad \forall n, \ \forall t_1,\ldots,t_n$$

Stationarity in wide sense

$$\bar{X}_t = \text{constant}, \quad R_{t+\tau,s+\tau} = R_{t,s} \quad \forall \tau \ , \quad 1^{st} \text{ and } 2^{nd} \text{ order statistics are translation invariant}$$

Process of stationary increments

$$X_{t+\tau} - X_t \quad \text{is a stationarity process} \quad \forall \tau$$

Process of independent increments

$$X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}} \quad \text{are independent} \quad \forall t_1 < t_2 < \cdots < t_n$$

Ergodic process

The whole statistics of the process can be determined from a single (infinitely long) realization.

Markov process

A stochastic process is called a Markov process when it has the Markov property:

$$\boxed{\mathrm{P}\{X_{t_n} \le x_n \,|\, X_{t_{n-1}} \le x_{n-1}, \ldots X_{t_1} \le x_1\} = \mathrm{P}\{X_{t_n} \le x_n \,|\, X_{t_{n-1}} \le x_{n-1}\}} \quad \forall n, \; \forall t_1 < \cdots < t_n$$

- The future path of a Markov process, given its current state $(X_{t_{n-1}})$ and the past history before t_{n-1}, depends only on the current state (not on how this state has been reached).

- The current state contains all the information (summary of the past) that is needed to characterize the future (stochastic) behaviour of the process.

- Given the state of the process at an instant its future and past are independent.

Example A process with independent increments is always a Markov process.

$$X_{t_n} = X_{t_{n-1}} + \underbrace{(X_{t_n} - X_{t_{n-1}})}_{\substack{\text{the increment is independent of} \\ \text{all the previous increments which} \\ \text{have given rise to the state } X_{t_{n-1}}}}$$

Markov chain

The use of the term Markov chain in the literature is ambiguous: it defines that the process is either a discrete time or a discrete state process.

In the sequel, we limit the use of the term for the case where the process is both discrete time and discrete state.

- Without loss of generality we can index the discrete instants of time by integers.
  - A Markov chain is thus a process X_n, $n = 0, 1, \ldots$.

- Similarly we can denote the states of the system by integers $X_n = 0, 1, \ldots$ (the set of states can be finite or countably finite).

In the following we additionally assume that the process is time homogeneous.

A Markov process of this kind is characterized by the (one-step) transition probabilities (transition from state i to state j):

$$\boxed{p_{i,j} = \mathrm{P}\{X_n = j \,|\, X_{n-1} = i\}}$$
     time homogeneity: the transition probability does not depend on n

The probability of a path

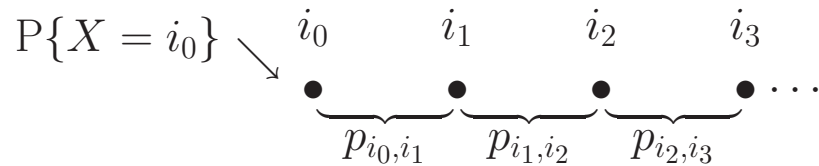The probability of a path $i_0, i_1, \ldots, i_n$ is

$$P\{X_0 = i_0, \ldots, X_n = i_n\} = P\{X_0 = i_0\}\, p_{i_0,i_1}\, p_{i_1,i_2} \cdots p_{i_{n-1},i_n}$$

Proof

$$P\{X_0 = i_0, X_1 = i_1\} = \underbrace{P\{X_1 = i_1 \mid X_0 = i_0\}}_{p_{i_0,i_1}} P\{X_0 = i_0\}$$

$$P\{X_0 = i_0, X_1 = i_1, X_2 = i_2\} = \underbrace{P\{X_2 = i_2 \mid X_1 = i_1, X_0 = i_0\}}_{p_{i_1,i_2}} \underbrace{P\{X_1 = i_1, X_0 = i_0\}}_{p_{i_0,i_1} P\{X_0 = i_0\}}$$

$$= P\{X_0 = i_0\}\, p_{i_0,i_1}\, p_{i_1,i_2}$$

Similarly, the proof can be continued for longer sequences.

The transition probability matrix of a Markov chain

The transition probabilities can be arranged as <u>transition probability matrix</u> $\mathbf{P} = (p_{i,j})$

$$
\begin{array}{c}
\text{final state} \rightarrow \\
\mathbf{P} = \begin{pmatrix} p_{0,0} & p_{0,1} & p_{0,2} & \cdots \\ p_{1,0} & p_{1,1} & p_{1,2} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \downarrow \text{initial state}
\end{array}
$$

- The row i contains the transition probabilities from state i to other states.

  – since the system always goes to some state, the sum of the row probabilities is 1

- A matrix with non-negative elements such that the sum of each row equals 1 is called a <u>stochastic matrix</u>.

- One can easily show that the product of two stochastic matrices is a stochastic matrix.

Many-step transition probability matrix

The probability that the system, initially in state i, will in
state j after two steps is

$$\sum_k p_{i,k} p_{k,j}$$

(takes into account all paths via an intermediate state k).
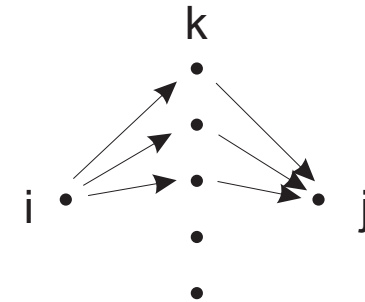
Clearly this the element $\{i, j\}$ of the matrix $\mathbf{P}^2$.

Similarly, one finds that the n-step transition probability matrix $\mathbf{P}^n$.

Denote its elements by $p_{i,j}^{(n)}$ (the subscript refers to the number of steps). Since it holds that
$\mathbf{P}^n = \mathbf{P}^m \cdot \mathbf{P}^{n-m}$ $(0 \le m \le n)$, we can write in component form

$$\boxed{p_{i,j}^{(n)} = \sum_k p_{i,k}^{(m)} p_{k,j}^{(n-m)}}$$ the Chapman-Kolmogorov equation

This simply expresses the law of total probability, where the transition in n steps from state
i to state j is conditioned on the system being in state k after m steps.

State probabilities

Denote

$$\pi_i^{(n)} = \mathrm{P}\{X_n = i\}$$ the probability that the process is in state i at time n

Arrange the state probabilities at time n in a state probability vector

$$\boldsymbol{\pi}^{(n)} = (\pi_0^{(n)}, \pi_1^{(n)}, \pi_2^{(n)}, \ldots)$$

By the law of total probability we have

$$\mathrm{P}\{X_1 = i\} = \sum_k \mathrm{P}\{X_1 = i \mid X_0 = k\}\mathrm{P}\{X_0 = k\}$$

or $\pi_i^{(1)} = \Sigma_k \, \pi_k^{(0)} p_{k,i}$ and in vector form $\boldsymbol{\pi}^{(1)} = \boldsymbol{\pi}^{(0)}\mathbf{P}$

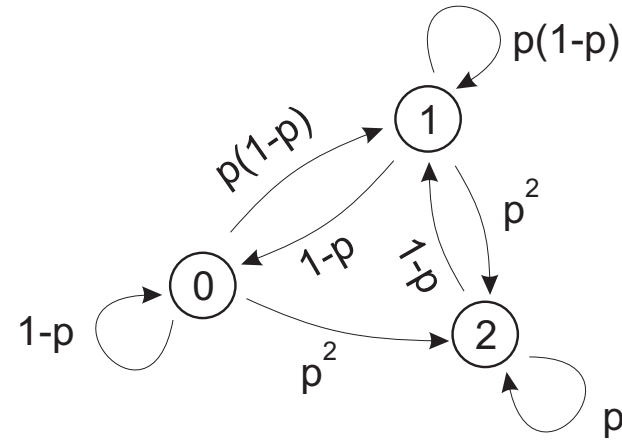As the process is Markovian and $\boldsymbol{\pi}^{(1)}$ represents the initial probabilities in the next step,

$$\boldsymbol{\pi}^{(2)} = \boldsymbol{\pi}^{(1)}\mathbf{P} \qquad \text{and generally} \qquad \boldsymbol{\pi}^{(n)} = \boldsymbol{\pi}^{(n-1)}\mathbf{P}$$

from which we have recursively

$$\boldsymbol{\pi}^{(n)} = \boldsymbol{\pi}^{(0)}\mathbf{P}^n$$ (Note, $\mathbf{P}^n$ is the n-step transition probability matrix.)

Example

$$\mathbf{P} = \begin{pmatrix} 1-p & p(1-p) & p^2 \\ 1-p & p(1-p) & p^2 \\ 0 & 1-p & p \end{pmatrix} \qquad p = 1/3$$

$$\mathbf{P} = \tfrac{1}{9}\begin{pmatrix} 6 & 2 & 1 \\ 6 & 2 & 1 \\ 0 & 6 & 3 \end{pmatrix} = \begin{pmatrix} 0.6666 & 0.2222 & 0.1111 \\ 0.6666 & 0.2222 & 0.1111 \\ 0 & 0.6666 & 0.3333 \end{pmatrix}$$

$$\mathbf{P}^2 = \tfrac{1}{9^2}\begin{pmatrix} 48 & 22 & 11 \\ 48 & 22 & 11 \\ 36 & 30 & 15 \end{pmatrix} = \begin{pmatrix} 0.5926 & 0.2716 & 0.1358 \\ 0.5926 & 0.2716 & 0.1358 \\ 0.4444 & 0.3704 & 0.1852 \end{pmatrix}$$

$$\mathbf{P}^3 = \tfrac{1}{9^3}\begin{pmatrix} 420 & 206 & 103 \\ 420 & 206 & 103 \\ 396 & 222 & 111 \end{pmatrix} = \begin{pmatrix} 0.5761 & 0.2826 & 0.1413 \\ 0.5761 & 0.2826 & 0.1413 \\ 0.5432 & 0.3045 & 0.1523 \end{pmatrix}$$

$$\vdots$$

$$\mathbf{P}^8 = \begin{pmatrix} 0.5714 & 0.2857 & 0.1429 \\ 0.5714 & 0.2857 & 0.1429 \\ 0.5714 & 0.2857 & 0.1429 \end{pmatrix}$$

Starting from an initial state i, the distribution of the final state can be read from the row i.

After 8 steps the final state distribution is independent of the initial state (to the accuracy of four digits): "the process forgets its initial state".
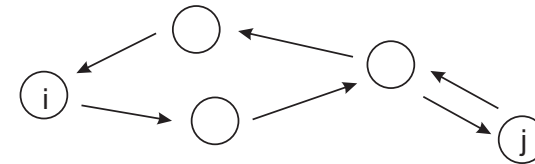
Classification of states of a Markov chain

<u>State i leads to state j</u> (written $i \to j$), if there is a path $i_0 = i, i_1, \ldots, i_n = j$ such that all the transition probabilities are positive, $p_{i_k, i_{k+1}} > 0, \ k = 0, \ldots, n-1$.
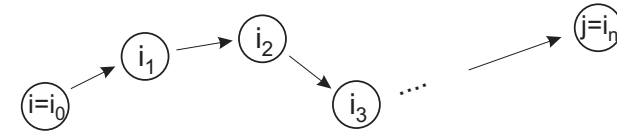
Then $(\mathbf{P}^n)_{i,j} > 0$.

<u>States i and j communicate</u> (written $i \leftrightarrow j$), if $i \to j$ and $j \to i$.

Communication is an <u>equivalence relation</u>: the states can be grouped into equivalent classes so that

- within each class all the states communicate with each other
- two states from two different classes never communicate which each other

The equivalence classes defined by the relation $\leftrightarrow$ are called the <u>irreducible</u> classes of states

> A Markov chain with a state space which is an irreducible class (the only one, i.e. all the states communicate) is called irreducible.

Classification of states (continued)

A set of states is closed, if none of its states leads to any of the states outside the set.

A single state which alone forms a closed set is called an absorbing state

  - for an absorbing state we have $p_{i,i} = 1$

  - one may reach an absorbing state from other states, but one cannot get out of it

Each state is either transient or recurrent.

  - A state i is transient if the probability of returning to the state is < 1.
    i.e. there is a non-zero probability that the system never returns to the state.

  - A state i is if the probability of returning to the state is $= 1$.
    i.e. with certainty, the system sometimes returns to the state.

Recurrent states are further classified according to the expectation of the time $T_{i,i}$ [1] it takes to return to the state:

positive recurrent   null recurrent
expectation of first return time $< \infty$ expectation of first return time $= \infty$

---

[1]The first return time $T_{i,i}$ of state i is the time at which the Markov chain first returns to state i when $X_0 = i$.
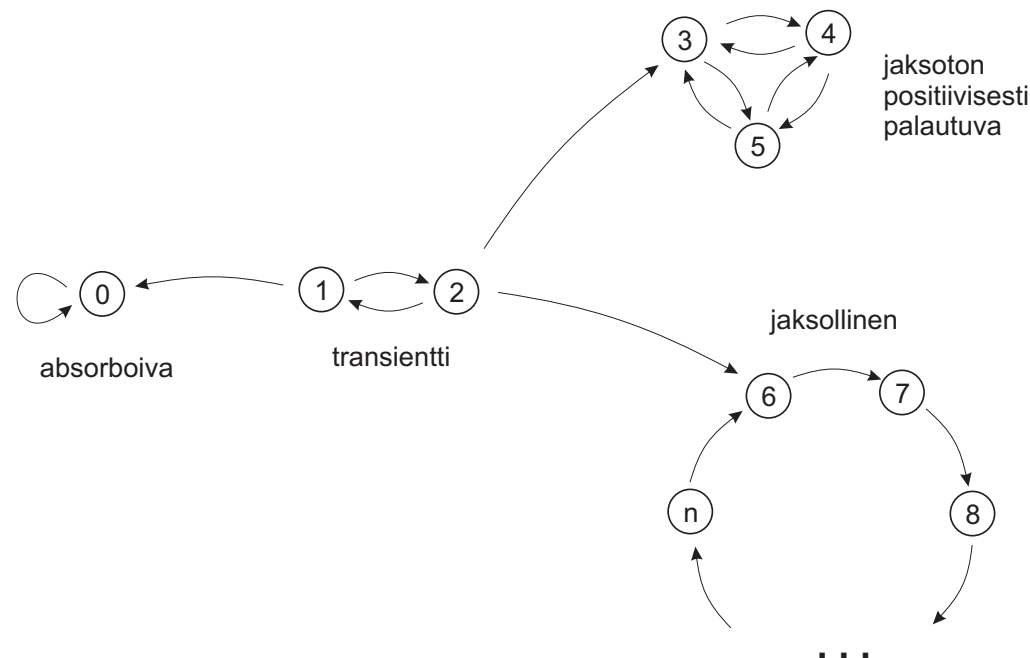
Classification of states (continued)

Type	# of visits	$\mathrm{E}[T_{i,i}]$
Transient	$< \infty$	∞
Null recurrent	∞	∞
Positive recurrent	∞	$< \infty$

If the first return time of state i can only be a multiple of an integer $d > 1$ the state i is called periodic. Otherwise the state is aperiodic.

An aperiodic positive recurrent state is ergodic
A Markov chain is ergodic, iff all its states are ergodic.

Classification of states (continued)

<u>Proposition:</u> In an irreducible Markov chain either

- all the states are transient, or

- all the states are null recurrent, or

- all the states are positive recurrent

<u>Remarks on the life time of a state and the first return time</u>

The number of steps the system consecutively
stays in state i is geometrically distributed
$$\sim \mathrm{Geom}(1 - p_{i,i})$$
because the exit from the state occurs with the
probability $1 - p_{i,i}$.
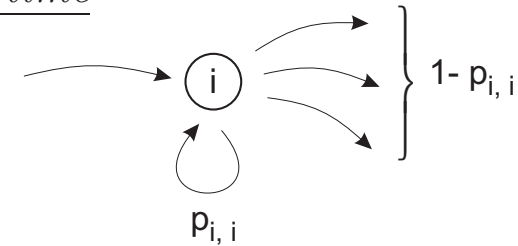
After each visit of state i, the first return time $T_{i,i}$ back to state i is independent of the first
return times after any of the other visits to the state (follows from the Markov property).

Denote $\bar{T}_i = \mathrm{E}[T_{i,i}]$ $\bar{T}_i = \begin{cases} \infty & \text{if the state is transient} \\ < \infty & \text{if the state is positive recurrent} \end{cases}$

Kolmogorov's theorem

In an irreducible, aperiodic Markov chain there always exist the limits

$$\pi_j = \lim_{n \to \infty} \pi_j^{(n)} = 1/\bar{T}_j$$

and these are independent of the initial state.

Furthermore, either

i) all the states of the chain are transient or all of the states are null recurrent; in either case $\pi_j = 0, \ \forall j$,

     or

ii) all the states of the chain are positive recurrent, and there exists a unique stationary distribution $\boldsymbol{\pi}$ which is obtained as the solution of the equations

$$\boxed{\begin{array}{c} \boldsymbol{\pi} = \boldsymbol{\pi} \cdot \mathbf{P} \\[2mm] \boldsymbol{\pi} \cdot \mathbf{e}^{\mathrm{T}} = 1 \end{array}}$$

    or

$$\pi_j = \sum_i \pi_i \, p_{i,j} \quad \text{and} \quad \sum_j \pi_j = 1$$

($\mathbf{e}$ is a row vector with all the components equal to 1, and $\mathbf{e}^{\mathrm{T}}$ is the corresponding column vector)

Remarks on the stationary distribution

If the limit probabilities (the components of the vector) $\boldsymbol{\pi}$ exist, they must satisfy the equation $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$, because

$$\boldsymbol{\pi} = \lim_{n\to\infty} \boldsymbol{\pi}^{(n)} = \lim_{n\to\infty} \boldsymbol{\pi}^{(n+1)} = \lim_{n\to\infty} \boldsymbol{\pi}^{(n)} \cdot \mathbf{P} = \boldsymbol{\pi}\mathbf{P}$$

The equation $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$ can also be expressed in the form: $\boldsymbol{\pi}$ is the (left) eigenvector of the matrix $\mathbf{P}$ belonging to the eigenvalue 1 (or belonging to the eigenvalue 0 of the matrix $(\mathbf{P}-\mathbf{I})$).

π_j defines which proportion of time (steps) the system stays in state j.

The limit distribution $\boldsymbol{\pi}$ is called the <u>stationary distribution</u> or the <u>equilibrium distribution</u>

Note. An equilibrium does not mean that nothing happens in the system, but merely that the information on the initial state of the system has been "forgot" or "washed out" because of the stochastic development.

Global balance

Equation $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$ or $\pi_j = \sum_i \pi_i\, p_{i,j},\ \forall j$, is often called the (global) balance condition.

Since $\sum_i p_{j,i} = 1$ (the transition takes the system to some state), one can write

$$\underbrace{\sum_i \pi_j\, p_{j,i}}_{} \qquad = \qquad \underbrace{\sum_i \pi_i\, p_{i,j}}_{} \qquad\qquad \text{One equation for each state } j.$$

prob. that the system is in state j and makes a transition to another state

prob. that the system is in another state and makes a transition to state j

Balance of probability flows: there are as many exits form state j as there are entries to it.
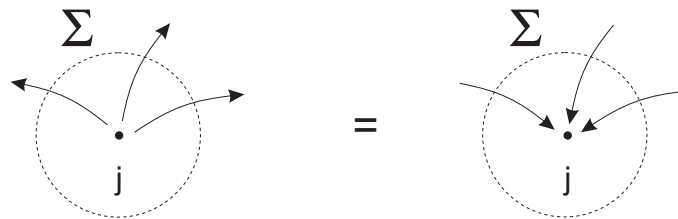
- If the balance equations are known to be satisfied for all but one of the states, they are automatically satisfied also for that particular state (due to conservation of probability flows)

- – the balance equations are linearly dependent
    ($\Rightarrow$ the homogeneous equation $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$ has a non-zero solution)
  – the solution is determined up to a constant factor
  – in order to determine the unknown factor, one needs the normalization condition
    $\sum_j \pi_j = 1$

Example. We revisit the previous example (now with a general p).

$$
\begin{pmatrix} \pi_0 & \pi_1 & \pi_2 \end{pmatrix} = \begin{pmatrix} \pi_0 & \pi_1 & \pi_2 \end{pmatrix} \begin{pmatrix} 1-p & p(1-p) & p^2 \\ 1-p & p(1-p) & p^2 \\ 0 & 1-p & p \end{pmatrix}
$$

Write the first two equations (equalities of the first two components of the vectors on the lhs and rhs)

$$
\begin{aligned}
\pi_0 &= (1-p)\pi_0 + (1-p)\pi_1 & \Rightarrow \quad \pi_0 = \tfrac{1-p}{p}\pi_1 \\[2mm]
\pi_1 &= p(1-p)\pi_0 + p(1-p)\pi_1 + (1-p)\pi_2 \\[2mm]
&= (1-p)^2\pi_1 + p(1-p)\pi_1 + (1-p)\pi_2 \\[2mm]
&= (1-p)\pi_1 + (1-p)\pi_2 & \Rightarrow \quad \pi_1 = \tfrac{1-p}{p}\pi_2 \qquad \Rightarrow \quad \pi_0 = \left(\tfrac{1-p}{p}\right)^2 \pi_2
\end{aligned}
$$

or

$$
\boldsymbol{\pi} = \left(\left(\tfrac{1-p}{p}\right)^2 \quad \tfrac{1-p}{p} \quad 1 \right) \pi_2 \, .
$$

By the normalization condition $\pi_0 + \pi_1 + \pi_2 = 1$ one gets

$$
\boldsymbol{\pi} = \left(\tfrac{(1-p)^2}{1-p(1-p)} \quad \tfrac{p(1-p)}{1-p(1-p)} \quad \tfrac{p^2}{1-p(1-p)} \right) \qquad \text{With } p = \tfrac{1}{3}: \qquad \boldsymbol{\pi} = \begin{pmatrix} 0.5714 & 0.2857 & 0.1429 \end{pmatrix}
$$

On solving the balance (or equilibrium) equations

In general, the equilibrium equations are solved as follows (assume a finite state space with n states):

- Write the balance condition for all but one of the states ($n - 1$ equations)

  - the equations fix the relative values of the equilibrium probabilities
  - the solution is determined up to a constant factor

- The last balance equation (automatically satisfied) is replaced by the normalization condition $\Sigma_j \, \pi_j = 1$.

For calculation by a computer, it's often more convenient to use the following method (usually, one construct the whole transition probability matrix $\mathbf{P}$):

Write n copies of the normalization condition $\boldsymbol{\pi} \cdot \mathbf{e}^{\mathrm{T}} = 1$:

$$\boldsymbol{\pi} \cdot \mathbf{E} = \mathbf{e} \quad \text{where } \mathbf{E} \text{ is an } n \times n\text{-matrix with all the elements equal to 1, } \mathbf{E} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix}$$

Add this equation and the balance equation $\boldsymbol{\pi} \cdot \mathbf{P} = \boldsymbol{\pi} \Rightarrow \boldsymbol{\pi} \cdot (\mathbf{P} + \mathbf{E} - \mathbf{I}) = \mathbf{e}$

$$\boxed{\boldsymbol{\pi} = \mathbf{e} \cdot (\mathbf{P} + \mathbf{E} - \mathbf{I})^{-1}}$$

The inhomogeneous equation has a unique solution. The solution automatically satisfies both the balance and the normalization conditions.