STAT40810 — Stochastic Models

Brendan Murphy

Week 8

Stochastic Processes

Stochastic processes

Definition - stochastic process

A stochastic process (or a random process) is a collection of random variables indexed by time, $\{X_t, t \in \mathcal{T}\}$. The set of all possible values for the X_t is called the state space, $\mathcal{X}$.

A stochastic process can also be seen as a random function $\mathcal{T} \to \mathcal{X}$.

Reminder - discrete/continuous space/time

Discrete / continuous state space

If $\mathcal{X}$ is

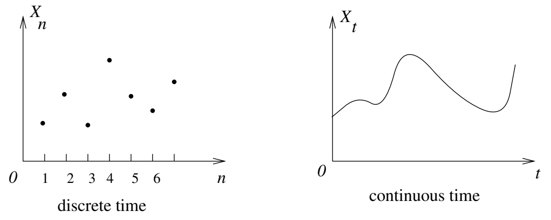
- an interval $[a, b]$, or the set of real numbers $\mathbb{R}$, or $\mathbb{R}^k$: "continuous state space".
- a finite or countable set: "discrete state space".

Discrete / continuous time

If $\mathcal{T}$ is

- an interval $[a, b]$ or the set of real numbers $\mathbb{R}$: "continuous time".
- a finite or countable set: "discrete time".

Illustration



This figure shows a sample path for each of two stochastic processes.

Example 1: Euro/dollar exchange rate

Figure : Euro/dollar exchange rate (ECB data).

- The state space is continuous, $\mathcal{X} = \mathbb{R}_+$.
- As the exchange rate is updated with very high frequency, we can view it as a continuous time process.

Example 2: Annual inflation rate in the USA

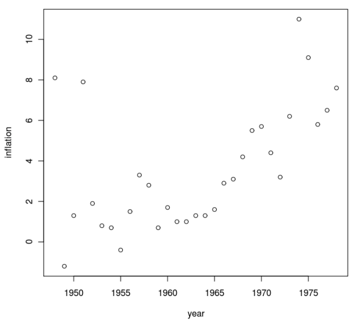


Figure : Inflation (source: Economic Report of the President 2004).

- The state space is continuous, $\mathcal{X} = \mathbb{R}$.
- Discrete time: $t \in \mathcal{T} = \{1948, 1949, \dots, 1978\}$.

Example 3: Annual number of major earthquakes

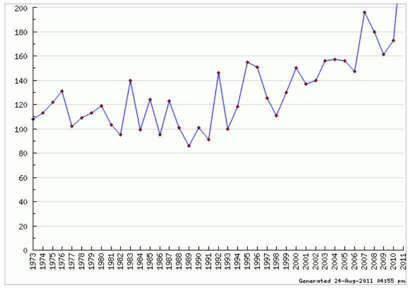


Figure : Magnitude ≥ 6. Source dlindquist.com.

- The state space is discrete, $\mathcal{X} = \mathbb{N}$.
- Discrete time: $t \in \mathcal{T} = \{1973, \dots, 2011\}$.

Example 4: Clickstream Data

- Heckerman et al. recorded the click patterns of users on MSNBC.
- The category of consecutive webpages visited was recorded:

```
1: frontpage
2: news
3: tech
4: local
5: opinion
6: on-air
7: misc
8: weather
9: msn-news
10: health
11: living
12: business
13: msn-sports
14: sports
15: summary
16: bbs
17: travel
```

- For example (user 1):

```
12 12 12 14 1 12 15 2 15 2 12 14 7 4 3 11 10 6 5 8 5 2 2 2 1 2 2 15 2 12 12 12 12 12 12
12 12 12 12 12 12 12 12 3 3 3 3 3 3 9 13 13 8 13 8 12 12 9 9 13 13 13 13 8 8 3 14
```

- For example (user 2):

```
9 9 13 13 13 13 8 8 3 14 9 13 13 8 3 11 2 2 2 10 9 13 7 9 9 9 9 1 8 2 12 7 4 1 11 5 6 9
9 9 4 4 7 13 12 11 13 13 13 13 13 13 13 14 13 14 14 14 13 7 8
```

Example 5: DNA Sequences

- DNA is a long polymer made from units called nucleotides.
- There are four nucleotides:

```
A: adenine
C: cytosine
G: guanine
T: thymine
```

```
TGGCGCTGGG CGCAATGCGC GCCATTACCG AGTCCGGGCT GCGCGTTGGT GCGGATATCT
CGGTAGTGGG ATACGACGAT ACCGAAGACA GCTCATGTTA TATCCCGCCG TTAACCACCA
TCAAACAGGA TTTTCGCCTG CTGGGGCAAA CCAGCGTGGA CCGCTTGCTG CAACTCTCTC
AGGGCCAGGC GGTGAAGGGC AATCAGCTGT TGCCCGTCTC ACTGGTGAAA AGAAAAACCA
CCCTGGCGCC CAATACGCAA ACCGCCTCTC CCCGCGCGTT GGCCGATTCA TTAATGCAGC
TGGCACGACA GGTTTCCCGA CTGGAAAGCG GGCAGTGAGC GCAACGCAAT TAATGTGAGT
TAGCTCACTC ATTAGGCACC CCAGGCTTTA CACTTTATGC TTCCGGCTCG TATGTTGTGT
GGAATTGTGA GCGGATAACA ATTTCACACA GGAAACAGCT A
```

- DNA can be viewed as a stochastic process where position in the sequence takes the role of time.

Example 1: white noise

Definition - white noise

The stochastic process $\{X_t, t \in \mathcal{T}\}$, with $\mathcal{T} = \mathbb{N}$ or $\mathcal{T} = \mathbb{Z}$, is said to be a white noise if the variable X_t's are iid (independent and identically distributed).

Definition - zero-mean white noise

A white noise with $\mathbb{E}(X_t) = 0$.

Definition - symmetric white noise

A white noise where the distribution of X_t is symmetric.

Example: $X_t \sim \mathcal{N}(0, \sigma^2)$.

Example 2: random walk

Definition - random walk

Let $\{Z_t, t \in \mathcal{T}\}$ be a white noise with $\mathcal{T} = \mathbb{N}$, put

$$X_t \equiv \sum_{i=1}^{t} Z_i = X_{t-1} + Z_t$$

The stochastic process $\{X_t\}$ is called a random walk, and the variables Z_t are called the increments or steps of the walk.

Definition - simple random walk

$Z_t = 1$ with probability p, $Z_t = -1$ with probability $1 - p$.

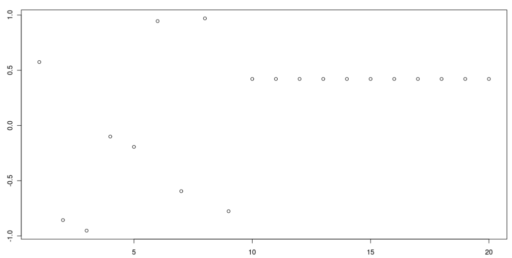
Definition - symmetric random walk

The distribution of Z_t is symmetric.

Notion of stationarity - main idea

Intuitively, a process is said to be stationary if its statistical properties do not change with time.

One could guess that imposing that all the X_t's have the same distribution is a good mathematical interpretation of this idea. However, it is not the case. E.g., X_1, ..., X_{10} are iid $\mathcal{N}(0, \sigma^2)$ and $X_{10} = X_{11} = X_{12} = \ldots$.



Stationarity

This is why the definition of stationarity is a bit more involved. It is, however, one of the most important notions about stochastic processes.

Definition - stationarity

The process $\{X_t\}$ is said to be stationary if for any t_1, ..., t_k and t,

$$(X_{t_1}, \dots, X_{t_k}) \text{ and } (X_{t_1+t}, \dots, X_{t_k+t})$$

have the same distribution.

Examples

- It is obvious that any white noise $\{X_t\}$ is stationary.
- One show that a random walk is never stationary (except in the degenerate case where the increments are not random $Z_t = 0$).

Markov property

Definition - Markov property

Let $\{X_t, t \in \mathcal{T}\}$ be a discrete time process with $\mathcal{T} = \mathbb{N}$ or $\mathcal{T} = \mathbb{Z}$. It is said to have the Markov property if for any $n \in \mathcal{T}$,

$$p(X_{n+1}|X_n, X_{n-1}, X_{n-2}, \dots) = p(X_{n+1}|X_n).$$

"If we know the current state of the process, what happened before is irrelevant with respect to the future."

Examples: a white noise and a random walk always satisfy the Markov property.

Markov property: Further result

Theorem

Let $\{X_t\}$ satisty the Markov property then for any $n \in \mathcal{T}$ and $k \in \mathbb{N}$,

$$p(X_{n+k}|X_n, X_{n-1}, X_{n-2}, \dots) = p(X_{n+k}|X_n).$$

STAT40810 — Stochastic Models

Brendan Murphy

Week 8

(Finite) Markov Chains

Andrey Markov

Andrey Andreyevich Markov (1856-1922), Russian mathematician, famous for Markov inequality, Markov processes, Markov chains, ... and his political opinions - among others, his support for the writer Leo Tolstoy led to his excommunication in 1912. His brother and his son were famous mathematicians too.

Reminder - Markov property

Definition - Markov property

Let $\{X_t, t \in \mathcal{T}\}$ be a discrete time process with $\mathcal{T} = \mathbb{N}$. It is said to have the Markov property if for any $n \in \mathcal{T}$,

$$p(X_{n+1}|X_n, X_{n-1}, X_{n-2}, \dots) = p(X_{n+1}|X_n).$$

In other words, to make predictions about the future of the process, it suffices to consider the present state, and not the past history.

In other other words, the state of the system is important but now how it arrived at that state.

Definition of a finite Markov chain

Definition - finite Markov chain

The process $\{X_t, t \in \mathbb{N}\}$ is said to be a finite Markov chain if:

1. it satisfies the Markov property,
2. the state space $\mathcal{X}$ is finite. We will usually take $\mathcal{X} = \{1, ..., M\}$.

Markov property for a finite Markov chain

The Markov property can be written:

$$\mathbb{P}\left(X_{n+1} = i_{n+1}|X_n = i_n, \dots, X_1 = i_1, X_0 = i_0\right) = \mathbb{P}\left(X_{n+1} = i_{n+1}|X_n = i_n\right).$$

Homogeneous finite Markov chain

Definition - homogeneous finite Markov chain

The Markov chain is said to be homogeneous if:

$$\mathbb{P}\left(X_{n+1} = j | X_n = i\right) = p(i,j) \text{ does not depend on } n.$$

We will only consider homogeneous finite Markov chains, which we abbreviate to "Markov chains".

Note that $p(i,j) \geq 0$ and

$$\sum_{j=1}^{M} p(i,j) = 1.$$

Basic properties

Definition

We put $\pi_n(i) = \mathbb{P}(X_n = i)$.

We have:

$$\begin{aligned}\mathbb{P}\left(X_n = i_n, \dots, X_0 = i_0\right) \\ &= \mathbb{P}(X_n = i_n | X_{n-1} = i_{n-1}, \dots, X_0 = i_0) \\ &\quad \times \mathbb{P}\left(X_{n-1} = i_{n-1}, \dots, X_0 = i_0\right) \\ &= p(i_{n-1}, i_n)\mathbb{P}\left(X_{n-1} = i_{n-1}, \dots, X_0 = i_0\right).\end{aligned}$$

By recurrence:

$$\mathbb{P}\left(X_n = i_n, \dots, X_0 = i_0\right) = \pi_0(i_0)p(i_0, i_1)\dots p(i_{n-1}, i_n).$$

Transition matrix

Definition - Transition matrix

$$P = \begin{pmatrix} p(1,1) & \dots & p(1,M) \\ \vdots & \dots & \vdots \\ p(M,1) & \dots & p(M,M) \end{pmatrix}.$$

Remember that any transition matrix satisfies $\sum_{j=1}^{M} p(i,j) = 1$ and $p(i,j) \geq 0$. Such a matrix is called a *stochastic matrix*.

Notation

It is also usual to represent the probability distribution of X_n by a row vector

$$\pi_n = (\pi_n(1), \dots, \pi_n(M)).$$

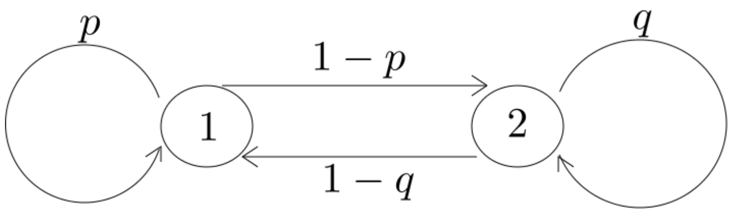
Example 1 - two state Markov chain

Example: the weather in Dublin is either "sunny" or "rainy". For the sake of brevity, we say that the weather is in state 1 when it is sunny, 2 when it is rainy. We assume that when it is sunny, the probability that is is still sunny the next day is p. When it is rainy, the probability that it is still rainy the next day is q.
This gives a Markov chain with transition matrix:

$$P = \begin{pmatrix} p & 1-p \\ 1-q & q \end{pmatrix}.$$

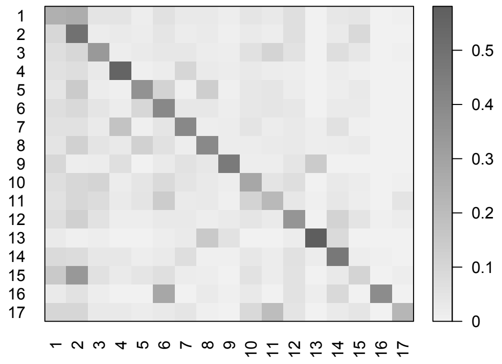
Example 1 - transition graph of the chain

An alternative way to represent the transition matrix P:



MSNBC Clickstream

- A Markov chain transition matrix was estimated from the MSNBC data.



- There are interesting patterns in the transitions.

Example 3 - random walks

Example: random walk with reflecting boundaries

$$p(i, i+1) = p, \quad p(i, i-1) = 1 - p,$$

$$p(0, 1) = 1 \text{ and } p(M, M-1) = 1.$$

Example: random walk with absorbing boundaries

$$p(i, i+1) = p, \quad p(i, i-1) = 1 - p,$$

$$p(0, 0) = 1 \text{ and } p(M, M) = 1.$$

Looking n steps ahead

Definition

Let us define

$$p_n(i, j) = \mathbb{P}(X_n = j | X_0 = i).$$

Theorem

$$p_{m+n}(i, j) = \sum_{k=1}^{N} p_m(i, k) p_n(k, j).$$

Theorem

$$\Big(p_n(i, j)\Big)_{1 \le i, j \le N} = P^n.$$

Looking n steps ahead

Theorem

$$\pi_n = \pi_0 P^n.$$

Example: Dublin weather example with $p = 0.2$, $q = 0.7$,

$$P = \begin{pmatrix} 0.2 & 0.8 \\ 0.3 & 0.7 \end{pmatrix}, \quad P^2 = \begin{pmatrix} 0.28 & 0.72 \\ 0.27 & 0.73 \end{pmatrix}, \quad P^3 = \begin{pmatrix} 0.272 & 0.728 \\ 0.273 & 0.727 \end{pmatrix},$$

$$\cdots \quad P^n \simeq \begin{pmatrix} 0.272727\ldots & 0.727272\ldots \\ 0.272727\ldots & 0.727272\ldots \end{pmatrix}$$

Note that this implies that $\pi_n \simeq (0.2727\ldots, 0.7272\ldots)$ does not depend on π_0 when n is large!

Some questions and a first answer

1. Do we have, for some probability distribution $\pi = (\pi(1), \ldots, \pi(N))$,
$$\pi_n = \pi_0 P^n \to \pi,$$
as $n \to \infty$?
2. If this is the case, does π depend on π_0?

Some questions and a first answer

1. Do we have, for some probability distribution $\pi = (\pi(1), \ldots, \pi(N))$,
$$\pi_n = \pi_0 P^n \to \pi,$$
as $n \to \infty$?
2. If this is the case, does π depend on π_0?

Theorem

If $\pi_n \to \pi$, as $n \to \infty$, then π satisfies $\pi P = \pi$.

Definition

Any π such that $\pi P = \pi$ is called a "stationary" or "equilibrium" or "invariant" distribution.

Reminder on eigenvalues and eigenvectors

For a matrix M, if there is $\lambda \in \mathbb{R}$ and a column vector $v \in \mathbb{R}^n$ such that

$$Mv = \lambda v$$

then λ is an "eigenvalue" and v is a "eigenvector".

Reminder on eigenvalues and eigenvectors

For a matrix M, if there is $\lambda \in \mathbb{R}$ and a column vector $v \in \mathbb{R}^n$ such that

$$Mv = \lambda v$$

then λ is an "eigenvalue" and v is a "eigenvector".
If we take transposes of the definition of "stationary distribution" we get:

$$P^T \pi^T = \pi^T$$

Reminder on eigenvalues and eigenvectors

For a matrix M, if there is $\lambda \in \mathbb{R}$ and a column vector $v \in \mathbb{R}^n$ such that

$$Mv = \lambda v$$

then λ is an "eigenvalue" and v is a "eigenvector".
If we take transposes of the definition of "stationary distribution" we get:

$$P^T \pi^T = \pi^T$$

So, the eigenvector of P^T (with eigenvalue 1) gives the stationary distribution.

Reminder on eigenvalues and eigenvectors

For a matrix M, if there is $\lambda \in \mathbb{R}$ and a column vector $v \in \mathbb{R}^n$ such that

$$Mv = \lambda v$$

then λ is an "eigenvalue" and v is a "eigenvector".
If we take transposes of the definition of "stationary distribution" we get:

$$P^T \pi^T = \pi^T$$

So, the eigenvector of P^T (with eigenvalue 1) gives the stationary distribution.

Theorem

Any transition matrix P has an invariant distribution.

Example: MSNBC

- The stationary distribution is:

```
1: frontpage (0.093)
2: news (0.197)
3: tech (0.059)
4: local (0.075)
5: opinion (0.048)
6: on-air (0.081)
7: misc (0.059)
8: weather (0.052)
9: msn-news (0.026)
10: health (0.053)
11: living (0.040)
12: business (0.070)
13: msn-sports (0.021)
14: sports (0.071)
15: summary (0.041)
16: bbs (0.005)
17: travel (0.010)
```

Example: social mobility

Data: 3500 pairs (father, adult son) from England and Wales early 1950's (a time of low unemployment). They have been classified according to 7 categories:

1. Professional and high administrative
2. Managerial and executive
3. Higher-grade supervisory and non-manual
4. Lower-grade supervisory and non-manual
5. Skilled manual and routine non-manual
6. Semi-skilled manual
7. Unskilled manual

The results are given in table in form of probabilities $p(i,j) = n_{ij} / \sum_k n_{ik}$, where n_{ij} is the number of pairs with father being from category i and son from j.

Example: social mobility

Data: 3500 pairs (father, adult son) from England and Wales early 1950's (a time of low unemployment). They have been classified according to 7 categories:

1. Professional and high administrative
2. Managerial and executive
3. Higher-grade supervisory and non-manual
4. Lower-grade supervisory and non-manual
5. Skilled manual and routine non-manual
6. Semi-skilled manual
7. Unskilled manual

The results are given in table in form of probabilities $p(i,j) = n_{ij} / \sum_k n_{ik}$, where n_{ij} is the number of pairs with father being from category i and son from j.

Social mobility: the data

F\S	1	2	3	4	5	6	7
1	0.388	0.146	0.202	0.062	0.140	0.047	0.015
2	0.107	0.267	0.227	0.120	0.206	0.053	0.020
3	0.035	0.101	0.188	0.191	0.357	0.067	0.061
4	0.021	0.039	0.112	0.212	0.430	0.124	0.062
5	0.009	0.024	0.075	0.123	0.473	0.171	0.125
6	0.000	0.013	0.041	0.088	0.391	0.312	0.155
7	0.000	0.008	0.036	0.083	0.364	0.235	0.274

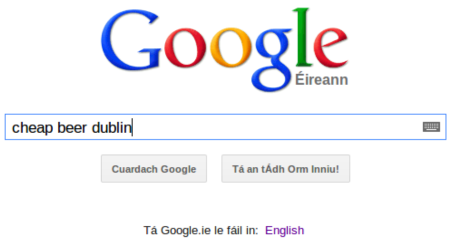
Social mobility - limiting and actual distributions

This table shows the calculated limiting distribution π for the corresponding Markov chain and the actual distributions for fathers and sons.

Class	π	Actual distribution: Father	Actual distribution: Son
1	0.023	0.037	0.029
2	0.042	0.043	0.046
3	0.088	0.098	0.094
4	0.127	0.148	0.131
5	0.409	0.432	0.409
6	0.182	0.131	0.170
7	0.129	0.111	0.121

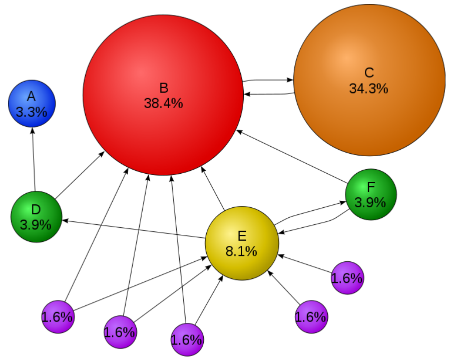
It is tempting (although probably incorrect) to conclude that English/Welsh society was stable in 1950's, for so many generations the limiting distribution was already being approached.
This might be not true if subsequent developments on the job market are taken into account.

Example 2



Example 2: Google's PageRank algorithm

How does Google rank web pages?



Source: http://en.wikipedia.org/wiki/PageRank