Time Series and Stochastic Processes

Peter Lukianchenko

1 October 2022

General result for randomly stopped sums:

Suppose $X_1, X_2, \ldots$ each have the same mean μ and variance σ^2, and $X_1, X_2, \ldots,$ and N are mutually independent. Let $T_N = X_1 + \ldots + X_N$ be the randomly stopped sum. By following similar working to that above:

$$\boxed{\begin{aligned} \mathbb{E}(T_N) = \mathbb{E}\left\{\sum_{i=1}^{N} X_i\right\} &= \mu\,\mathbb{E}(N) \\ \mathrm{Var}(T_N) = \mathrm{Var}\left\{\sum_{i=1}^{N} X_i\right\} &= \sigma^2\,\mathbb{E}(N) + \mu^2\,\mathrm{Var}(N). \end{aligned}}$$

Lecture

First-step analysis for probabilities:

The first-step analysis procedure for probabilities can be summarized as follows:

$$\mathbb{P}(\textit{eventual goal}) = \sum_{\substack{\textit{first-step}\\ \textit{options}}} \mathbb{P}(\textit{eventual goal} \mid \textit{option})\mathbb{P}(\textit{option})\,.$$

This is because the first-step options form a *partition of the sample space*.

Lecture

First-step analysis for expected reaching times:

The expression for expected reaching times is very similar:

$$\mathbb{E}(\textit{reaching time}) = \sum_{\substack{\textit{first-step}\\ \textit{options}}} \mathbb{E}(\textit{reaching time} \,|\, \textit{option})\,\mathbb{P}(\textit{option})\,.$$

Lecture

This follows immediately from the law of total expectation:

$$\mathbb{E}(X) = \mathbb{E}_Y\Big\{\mathbb{E}(X \,|\, Y)\Big\} = \sum_y \mathbb{E}(X \,|\, Y = y)\mathbb{P}(Y = y).$$

Let X be the reaching time, and let Y be the label for possible options:
i.e. $Y = 1, 2, 3, \ldots$ *for options 1, 2, 3, …*

We then obtain:

$$\begin{aligned} \mathbb{E}(X) &= \sum_y \mathbb{E}(X \,|\, Y = y)\mathbb{P}(Y = y) \\ \textit{i.e.} \qquad \mathbb{E}(\textit{reaching time}) &= \sum_{\substack{\textit{first-step}\\ \textit{options}}} \mathbb{E}(\textit{reaching time} \,|\, \textit{option})\mathbb{P}(\textit{option})\,. \end{aligned}$$

Example 1: Mouse in a Maze

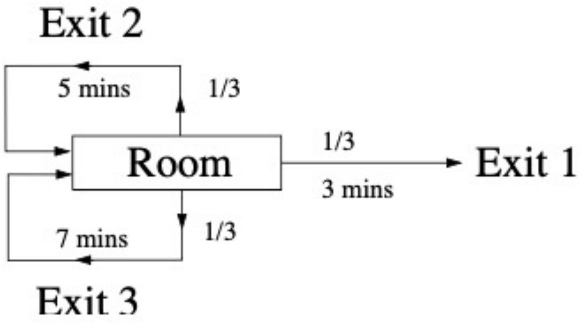
A mouse is trapped in a room with three exits at the centre of a maze.

- Exit 1 leads outside the maze after 3 minutes.
- Exit 2 leads back to the room after 5 minutes.
- Exit 3 leads back to the room after 7 minutes.

Every time the mouse makes a choice, it is equally likely to choose any of the three exits. What is the expected time taken for the mouse to leave the maze?

Let X = time taken for mouse to leave maze, starting from room R.

Let Y = exit the mouse chooses first $(1, 2, \text{or } 3)$.



Then:

$$\begin{aligned}
\mathbb{E}(X) &= \mathbb{E}_Y\Big(\mathbb{E}(X\,|\,Y)\Big) \\
&= \sum_{y=1}^{3} \mathbb{E}(X\,|\,Y=y)\mathbb{P}(Y=y) \\
&= \mathbb{E}(X\,|\,Y=1)\times\tfrac{1}{3} + \mathbb{E}(X\,|\,Y=2)\times\tfrac{1}{3} + \mathbb{E}(X\,|\,Y=3)\times\tfrac{1}{3}\,.
\end{aligned}$$

But:

$$\begin{aligned}
\mathbb{E}(X\,|\,Y=1) &= 3 \textit{ minutes} \\
\mathbb{E}(X\,|\,Y=2) &= 5+\mathbb{E}(X) \textit{ (after 5 mins back in Room, time } \mathbb{E}(X) \textit{ to get out)} \\
\mathbb{E}(X\,|\,Y=3) &= 7+\mathbb{E}(X) \textit{ (after 7 mins, back in Room)}
\end{aligned}$$

So

$$\begin{aligned}
\mathbb{E}(X) &= 3\times\tfrac{1}{3} + \Big(5+\mathbb{E}X\Big)\times\tfrac{1}{3} + \Big(7+\mathbb{E}X\Big)\times\tfrac{1}{3} \\
&= 15\times\tfrac{1}{3} + 2(\mathbb{E}X)\times\tfrac{1}{3} \\
\tfrac{1}{3}\,\mathbb{E}(X) &= 15\times\tfrac{1}{3} \\
\Rightarrow \qquad \mathbb{E}(X) &= 15 \textit{ minutes.}
\end{aligned}$$

Lecture

As for probabilities, first-step analysis for expectations relies on a good notation. The best way to tackle the problem above is as follows.

Define $\quad m_R = \mathbb{E}(\textit{time to leave maze} \mid \textit{start in Room}).$

First-step analysis:

$$\begin{aligned} m_R &= \tfrac{1}{3} \times 3 + \tfrac{1}{3} \times (5 + m_R) + \tfrac{1}{3} \times (7 + m_R) \\ \Rightarrow \quad 3m_R &= (3 + 5 + 7) + 2m_R \\ \Rightarrow \quad m_R &= 15 \textit{ minutes} \qquad \textit{(as before)}. \end{aligned}$$

Lecture

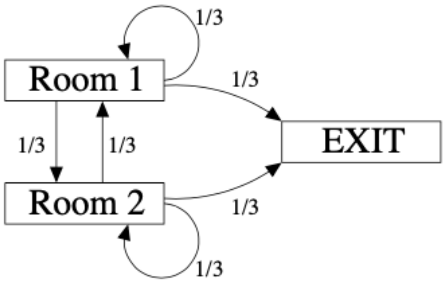
Example 2: Counting the steps

The most common questions involving first-step analysis for expectations ask for the ***expected number of steps before finishing.*** The number of steps is usually equal to the *number of arrows traversed from the current state to the end*.

The key point to remember is that when we take expectations, we are usually *counting something*.

You must remember to *add on whatever we are counting, to every step taken.*

The mouse is put in a new maze with two rooms, pictured here. Starting from Room 1, what is the expected number of steps the mouse takes before it reaches the exit?



1. Define notation: let

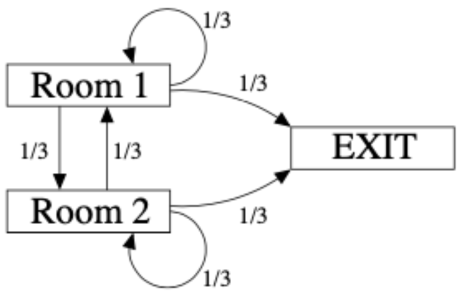
$$\begin{aligned} m_1 &= \mathbb{E}(\textit{number of steps to finish} \mid \textit{start in Room 1}) \\ m_2 &= \mathbb{E}(\textit{number of steps to finish} \mid \textit{start in Room 2}). \end{aligned}$$

2. First-step analysis:

$$\begin{aligned} m_1 &= \tfrac{1}{3} \times 1 + \tfrac{1}{3}(1 + m_1) + \tfrac{1}{3}(1 + m_2) \qquad (a) \\ m_2 &= \tfrac{1}{3} \times 1 + \tfrac{1}{3}(1 + m_1) + \tfrac{1}{3}(1 + m_2) \qquad (b) \end{aligned}$$

Incrementing before partitioning

In many problems, all possible first-step options incur the same initial penalty. The last example is such a case, because *every possible step adds 1 to the total number of steps taken.*



In a case where all steps incur the same penalty, there are two ways of proceeding:

1. *Add the penalty onto each option separately: e.g.*

$$m_1 = \tfrac{1}{3} \times 1 + \tfrac{1}{3}(1 + m_1) + \tfrac{1}{3}(1 + m_2)\,.$$

2. *(Usually quicker) Add the penalty once only, at the beginning:*

$$m_1 = 1 \quad + \quad \tfrac{1}{3} \times 0 + \tfrac{1}{3}m_1 + \tfrac{1}{3}m_2.$$

In each case, we will get the same answer (check). This is because the option probabilities sum to 1, *so in Method 1 we are adding* $(\frac{1}{3} + \frac{1}{3} + \frac{1}{3}) \times 1 = 1 \times 1 = 1$, *just as we are in Method 2.*

Lecture

Define the indicator random variable: $I_A = \begin{cases} 1 & \text{if event } A \text{ occurs,} \\ 0 & \text{otherwise.} \end{cases}$

Then $\mathbb{E}(I_A) = \mathbb{P}(I_A = 1) = \mathbb{P}(A)$.

We can refine this expression further, using the idea of conditional expectation. Let Y be any random variable. Then

$$\mathbb{P}(A) = \mathbb{E}(I_A) = \mathbb{E}_Y\Big(\mathbb{E}(I_A \,|\, Y)\Big).$$

But

$$
\begin{aligned}
\mathbb{E}(I_A \,|\, Y) &= \sum_{r=0}^{1} r\mathbb{P}(I_A = r \,|\, Y) \\
&= 0 \times \mathbb{P}(I_A = 0 \,|\, Y) + 1 \times \mathbb{P}(I_A = 1 \,|\, Y) \\
&= \mathbb{P}(I_A = 1 \,|\, Y) \\
&= \mathbb{P}(A \,|\, Y).
\end{aligned}
$$

Thus

$$\mathbb{P}(A) = \mathbb{E}_Y\Big(\mathbb{E}(I_A \,|\, Y)\Big) = \mathbb{E}_Y\Big(\mathbb{P}(A \,|\, Y)\Big).$$

This means that for ***any*** random variable X (discrete or continuous), and for any set of values S (a discrete set or a continuous set), we can write:

- for any ***discrete*** random variable Y,

$$\mathbb{P}(X \in S) = \sum_y \mathbb{P}(X \in S \,|\, Y = y)\mathbb{P}(Y = y).$$

- for any ***continuous*** random variable Y,

$$\mathbb{P}(X \in S) = \int_y \mathbb{P}(X \in S \,|\, Y = y) f_Y(y)\, dy.$$

Lecture

Define Y to be the number of OTHER matching tickets out of the OTHER 1 million tickets sold. (If you are lucky, $Y = 0$ so you have definitely won.)

If there are 1 million tickets and each ticket has a one-in-a-million chance of having the winning numbers, then

$$Y \sim \textit{Poisson}(1) \textit{ approximately.}$$

The relationship $Y \sim \text{Poisson}(1)$ arises because of the Poisson approximation to the Binomial distribution.

(a) What is the probability function of Y, $f_Y(y)$?

$$f_Y(y) = \mathbb{P}(Y = y) = \frac{1^y}{y!}e^{-1} = \frac{1}{e \times y!} \qquad \textit{for } y = 0, 1, 2, \ldots.$$

(b) What is the probability that yours is the only matching ticket?

$$\mathbb{P}(\textit{only one matching ticket}) = \mathbb{P}(Y = 0) = \frac{1}{e} = 0.368.$$

(c) The prize is chosen at random from all those who have matching tickets. What is the probability that you win if there are $Y = y$ OTHER matching tickets?

Let W be the event that I win.

$$\mathbb{P}(W \mid Y = y) = \frac{1}{y+1}.$$

Lecture

(d) Overall, what is the probability that you win, given that you have a matching ticket?

$$
\begin{aligned}
\mathbb{P}(W) &= \mathbb{E}_Y \left\{ \mathbb{P}(W \,|\, Y = y) \right\} \\
&= \sum_{y=0}^{\infty} \mathbb{P}(W \,|\, Y = y) \mathbb{P}(Y = y) \\
&= \sum_{y=0}^{\infty} \left(\frac{1}{y+1} \right) \left(\frac{1}{e \times y!} \right) \\
&= \frac{1}{e} \sum_{y=0}^{\infty} \frac{1}{(y+1)y!} \\
&= \frac{1}{e} \sum_{y=0}^{\infty} \frac{1}{(y+1)!} \\
&= \frac{1}{e} \left\{ \sum_{y=0}^{\infty} \frac{1}{y!} - \frac{1}{0!} \right\} \\
&= \frac{1}{e} \{ e - 1 \} \\
&= 1 - \frac{1}{e} \\
&= 0.632.
\end{aligned}
$$

Lecture

Expected hitting times

In the previous section we found the **probability** of hitting set A, starting at state i. Now we study **how long** it takes to get from i to A. As before, it is best to solve problems using first-step analysis and common sense. However, a general formula is also available.

Definition: Let A be a subset of the state space S. The **hitting time** of A is the random variable T_A, where

$$T_A = \min\{t \geq 0 : X_t \in A\}.$$

T_A is the time taken before hitting set A ***for the first time.***

The hitting time T_A can take values $0, 1, 2, \ldots,$ ***and*** ∞.
If the chain *never* hits set A, then $T_A = \infty$.

Note: The hitting time is also called the **reaching time**. If A is a closed class, it is also called the ***absorption time.***

Definition: The **mean hitting time** for A, starting from state i, is

$$m_{iA} = \mathbb{E}(T_A \,|\, X_0 = i).$$

Note: If there is any possibility that the chain *never* reaches A, starting from i, ***i.e. if the hitting probability*** $h_{iA} < 1$, ***then*** $\mathbb{E}(T_A \,|\, X_0 = i) = \infty$.

Lecture

The vector of expected hitting times $\boldsymbol{m_A} = (m_{iA} : i \in S)$ is *the minimal non-negative solution to the following equations:*

$$m_{iA} = \begin{cases} 0 & \text{for} \quad i \in A, \\ 1 + \sum_{j \notin A} p_{ij} m_{jA} & \text{for} \quad i \notin A. \end{cases}$$

Proof (sketch):

Consider the equations $m_{iA} = \begin{cases} 0 & \text{for} \quad i \in A, \\ 1 + \sum_{j \notin A} p_{ij} m_{jA} & \text{for} \quad i \notin A. \end{cases} \quad (\star).$

We need to show that:

(i) the mean hitting times $\{m_{iA}\}$ collectively satisfy the equations $(\star)$;

(ii) if $\{u_{iA}\}$ is any other non-negative solution to $(\star)$, then the mean hitting times $\{m_{iA}\}$ satisfy $m_{iA} \leq u_{iA}$ for all i (minimal solution).

Lecture

Proof of (i): Clearly, $m_{iA} = 0$ if $i \in A$ (as the chain hits A immediately).

Suppose that $i \notin A$. Then

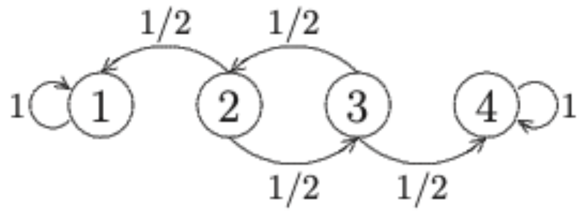
$$\begin{aligned}
m_{iA} &= \mathbb{E}(T_A \,|\, X_0 = i) \\
&= 1 + \sum_{j \in S} \mathbb{E}(T_A \,|\, X_1 = j)\mathbb{P}(X_1 = j \,|\, X_0 = i) \\
&\quad \text{(conditional expectation: take 1 step to get to state } j \\
&\quad \text{at time 1, then find } \mathbb{E}(T_A) \text{ from there)} \\
&= 1 + \sum_{j \in S} m_{jA}\, p_{ij} \qquad \text{(by definitions)} \\
&= 1 + \sum_{j \notin A} p_{ij}\, m_{jA}, \qquad \text{because } m_{jA} = 0 \text{ for } j \in A.
\end{aligned}$$

Thus the mean hitting times $\{m_{iA}\}$ must satisfy the equations $(\star)$.

Lecture

Example: Let $\{X_t : t \geq 0\}$ have the same transition diagram as before:

Starting from state 2, find the expected time to absorption.

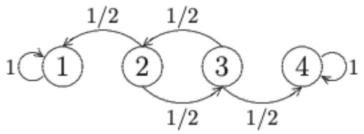


Solution:

Starting from state $i = 2$, we wish to find the expected time to reach the set $A = \{1, 4\}$ (the set of absorbing states).

Thus we are looking for $m_{iA} = m_{2A}$.

Now $$m_{iA} = \begin{cases} 0 & \textit{if} \quad i \in \{1,4\}, \\ 1 + \sum_{j \notin A} p_{ij} m_{jA} & \textit{if} \quad i \notin \{1,4\}. \end{cases}$$



Thus,

$$\begin{aligned} m_{1A} &= 0 \quad (\text{because } 1 \in A) \\ m_{4A} &= 0 \quad (\text{because } 4 \in A) \\ m_{2A} &= 1 + \tfrac{1}{2} m_{1A} + \tfrac{1}{2} m_{3A} \\ \Rightarrow \quad m_{2A} &= 1 + \tfrac{1}{2} m_{3A} \\ m_{3A} &= 1 + \tfrac{1}{2} m_{2A} + \tfrac{1}{2} m_{4A} \\ &= 1 + \tfrac{1}{2} m_{2A} \\ &= 1 + \tfrac{1}{2}\left(1 + \tfrac{1}{2} m_{3A}\right) \\ \Rightarrow \quad \tfrac{3}{4} m_{3A} &= \tfrac{3}{2} \\ \Rightarrow \quad m_{3A} &= 2. \end{aligned}$$

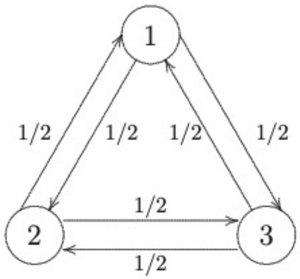
Thus, $$m_{2A} = 1 + \frac{1}{2} m_{3A} = 2.$$

The expected time to absorption is therefore $\mathbb{E}(T_A) = 2$ steps.

Lecture

Example: Glee-flea hops around on a triangle. At each step he moves to one of the other two vertices at random. What is the expected time taken for Glee-flea to get from vertex 1 to vertex 2?

Solution:



transition matrix, $P = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}$.

We wish to find m_{12}.

Now $m_{i2} = \begin{cases} 0 & \text{if } i = 2, \\ 1 + \sum_{j \neq 2} p_{ij} m_{j2} & \text{if } i \neq 2. \end{cases}$

Thus

$$\begin{aligned} m_{22} &= 0 \\ m_{12} &= 1 + \tfrac{1}{2}m_{22} + \tfrac{1}{2}m_{32} \;=\; 1 + \tfrac{1}{2}m_{32}. \\ m_{32} &= 1 + \tfrac{1}{2}m_{22} + \tfrac{1}{2}m_{12} \\ &= 1 + \tfrac{1}{2}m_{12} \\ &= 1 + \tfrac{1}{2}\left(1 + \tfrac{1}{2}m_{32}\right) \\ \Rightarrow \quad m_{32} &= 2. \end{aligned}$$

Thus $m_{12} = 1 + \frac{1}{2}m_{32} = 2$ *steps.*

Lecture

This raises the question: is there any distribution $\boldsymbol{\pi}$ such that $\boldsymbol{\pi}^T P = \boldsymbol{\pi}^T$?

If $\boldsymbol{\pi}^T P = \boldsymbol{\pi}^T$, then

$$
\begin{aligned}
X_t \sim \boldsymbol{\pi}^T \quad &\Rightarrow \quad X_{t+1} \sim \boldsymbol{\pi}^T P = \boldsymbol{\pi}^T \\
&\Rightarrow \quad X_{t+2} \sim \boldsymbol{\pi}^T P = \boldsymbol{\pi}^T \\
&\Rightarrow \quad X_{t+3} \sim \boldsymbol{\pi}^T P = \boldsymbol{\pi}^T \\
&\Rightarrow \quad \ldots
\end{aligned}
$$

In other words, if $\boldsymbol{\pi}^T P = \boldsymbol{\pi}^T$, and $X_t \sim \boldsymbol{\pi}^T$, then

$$X_t \sim X_{t+1} \sim X_{t+2} \sim X_{t+3} \sim \ldots$$

Thus, once a Markov chain has reached a distribution $\boldsymbol{\pi}^T$ such that $\boldsymbol{\pi}^T P = \boldsymbol{\pi}^T$, *it will stay there.*

If $\boldsymbol{\pi}^T P = \boldsymbol{\pi}^T$, we say that the distribution $\boldsymbol{\pi}^T$ is an *equilibrium distribution.*

Lecture

Equilibrium means a ***level position:*** there is *no more change* in the distribution of X_t as we wander through the Markov chain.

Note: Equilibrium does not mean that the **value** of X_{t+1} equals the **value** of X_t. It means that the **distribution** of X_{t+1} is the same as the **distribution** of X_t:

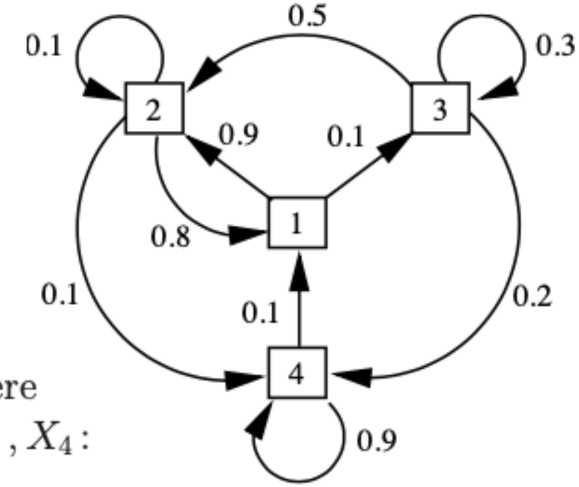
e.g. $\mathbb{P}(X_{t+1} = 1) = \mathbb{P}(X_t = 1) = \pi_1$;

$$\mathbb{P}(X_{t+1} = 2) = \mathbb{P}(X_t = 2) = \pi_2, \quad \textit{etc.}$$

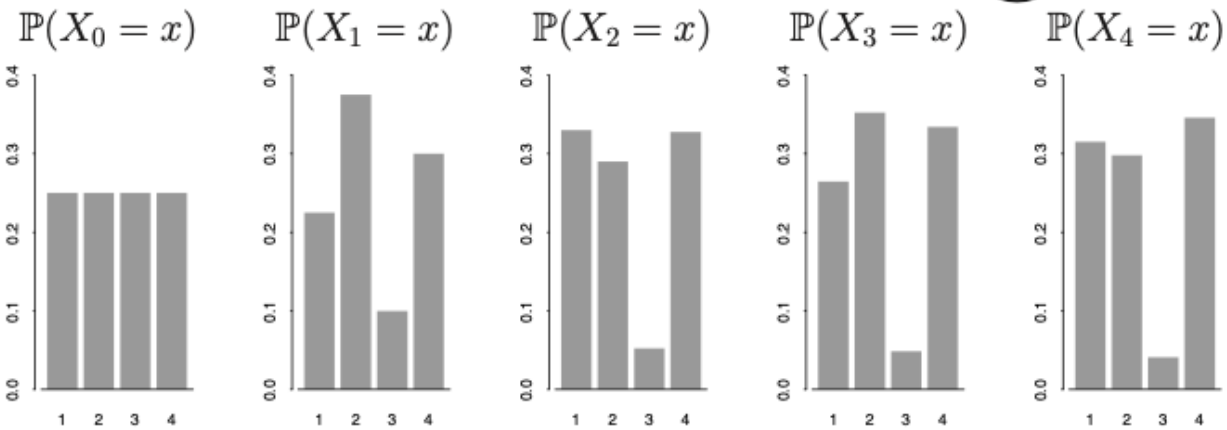
Lecture

Consider the following 4-state Markov chain:

$$P = \begin{pmatrix} 0.0 & 0.9 & 0.1 & 0.0 \\ 0.8 & 0.1 & 0.0 & 0.1 \\ 0.0 & 0.5 & 0.3 & 0.2 \\ 0.1 & 0.0 & 0.0 & 0.9 \end{pmatrix}$$

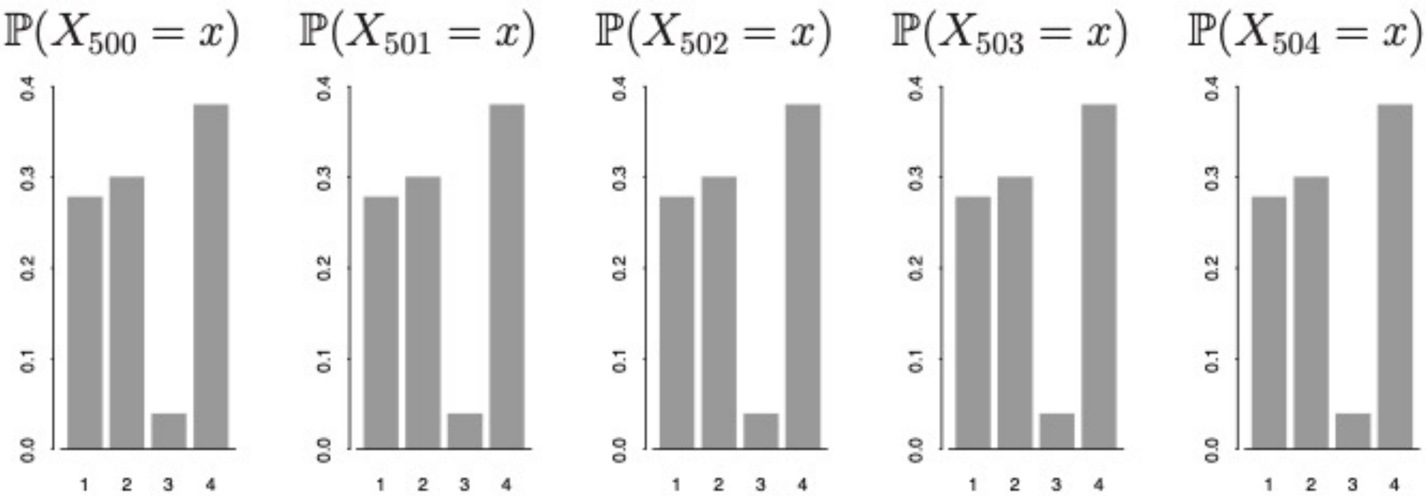


Suppose we start at time 0 with $X_0 \sim \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$: so the chain is equally likely to start from any of the four states. Here are pictures of the distributions of $X_0, X_1, \ldots, X_4$:



Lecture

The distribution starts off level, but quickly changes: for example the chain is least likely to be found in state 3. The distribution of X_t changes between each $t = 0, 1, 2, 3, 4$. Now look at the distribution of X_t 500 steps into the future:



The distribution has reached a steady state: it ***does not change*** between $t = 500, 501, \ldots, 504$. *The chain has reached equilibrium of its own accord.*

Lecture

Definition: Let $\{X_0, X_1, \ldots\}$ be a Markov chain with transition matrix P and state space S, where $|S| = N$ (possibly infinite). Let $\boldsymbol{\pi}^T$ be a row vector denoting a probability distribution on S: so each element π_i denotes the probability of being in state i, and $\sum_{i=1}^{N} \pi_i = 1$, where $\pi_i \geq 0$ for all $i = 1, \ldots, N$. The probability distribution $\boldsymbol{\pi}^T$ is an **equilibrium** distribution for the Markov chain if $\boldsymbol{\pi}^T P = \boldsymbol{\pi}^T$.

That is, $\boldsymbol{\pi}^T$ is an equilibrium distribution if

$$\left(\boldsymbol{\pi}^T P\right)_j = \sum_{i=1}^{N} \pi_i p_{ij} = \pi_j \quad \textit{for all } j = 1, \ldots, N.$$

By the argument given on page 174, we have the following Theorem:

Theorem 9.2: Let $\{X_0, X_1, \ldots\}$ be a Markov chain with transition matrix P. Suppose that $\boldsymbol{\pi}^T$ is an equilibrium distribution for the chain. If $X_t \sim \boldsymbol{\pi}^T$ for any t, then $X_{t+r} \sim \boldsymbol{\pi}^T$ *for all* $r \geq 0$. $\square$

Once a chain has hit an equilibrium distribution, *it stays there for ever.*

Note: There are several other names for an equilibrium distribution. If $\boldsymbol{\pi}^T$ is an equilibrium distribution, it is also called:

- **invariant:** *it doesn't change:* $\boldsymbol{\pi}^T P = \boldsymbol{\pi}^T$;
- **stationary:** *the chain 'stops' here.*

Lecture

Stationarity: the Chain Station

a BUS station is where a BUS stops

a train station is where a train stops

a **workstation** is where . . . ? ? ?

a stationary distribution is where a Markov chain stops

Lecture

Finding an equilibrium distribution

Vector $\boldsymbol{\pi}^T$ is an equilibrium distribution for P if:

1. $\boldsymbol{\pi}^T P = \boldsymbol{\pi}^T$;
2. $\sum_{i=1}^{N} \pi_i = 1$;
3. $\pi_i \geq 0$ *for all* i.

Conditions 2 and 3 ensure that $\boldsymbol{\pi}^T$ *is a genuine probability distribution.*

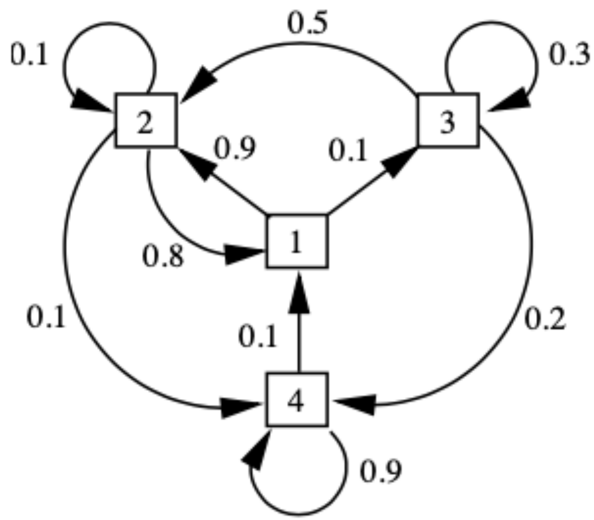
Condition 1 means that $\boldsymbol{\pi}$ is a row eigenvector of P.

Solving $\boldsymbol{\pi}^T P = \boldsymbol{\pi}^T$ by itself will just specify $\boldsymbol{\pi}$ up to *a scalar multiple.*
We need to include Condition 2 to scale $\boldsymbol{\pi}$ to a genuine probability distribution, and then check with Condition 3 that the scaled distribution is valid.

Lecture

Example: Find an equilibrium distribution for the Markov chain below.

$$P = \begin{pmatrix} 0.0 & 0.9 & 0.1 & 0.0 \\ 0.8 & 0.1 & 0.0 & 0.1 \\ 0.0 & 0.5 & 0.3 & 0.2 \\ 0.1 & 0.0 & 0.0 & 0.9 \end{pmatrix}$$

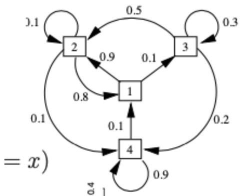


Solution:

Let $\boldsymbol{\pi}^T = (\pi_1,\ \pi_2,\ \pi_3,\ \pi_4)$.
The equations are $\boldsymbol{\pi}^T P = \boldsymbol{\pi}^T$ and $\pi_1 + \pi_2 + \pi_3 + \pi_4 = 1$.

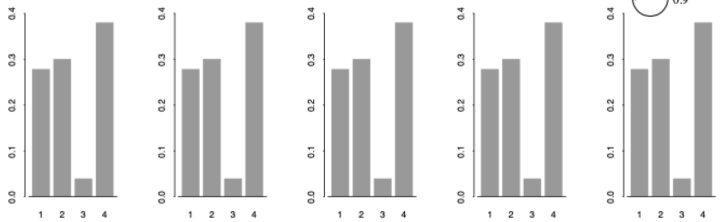
$$\boldsymbol{\pi}^T P = \boldsymbol{\pi}^T \quad \Rightarrow \quad (\pi_1\ \pi_2\ \pi_3\ \pi_4) \begin{pmatrix} 0.0 & 0.9 & 0.1 & 0.0 \\ 0.8 & 0.1 & 0.0 & 0.1 \\ 0.0 & 0.5 & 0.3 & 0.2 \\ 0.1 & 0.0 & 0.0 & 0.9 \end{pmatrix} = (\pi_1\ \pi_2\ \pi_3\ \pi_4)$$

Long-term behaviour



In Section 9.1, we saw an example where the Markov chain wandered of its own accord into its equilibrium distribution:

$\mathbb{P}(X_{500} = x)$ $\mathbb{P}(X_{501} = x)$ $\mathbb{P}(X_{502} = x)$ $\mathbb{P}(X_{503} = x)$



This will always happen for this Markov chain. In fact, the distribution it converges to (found above) does not depend upon the starting conditions: *for ANY value of X_0, we will always have $X_t \sim (0.28,\ 0.30,\ 0.04,\ 0.38)$ as $t \to \infty$.*

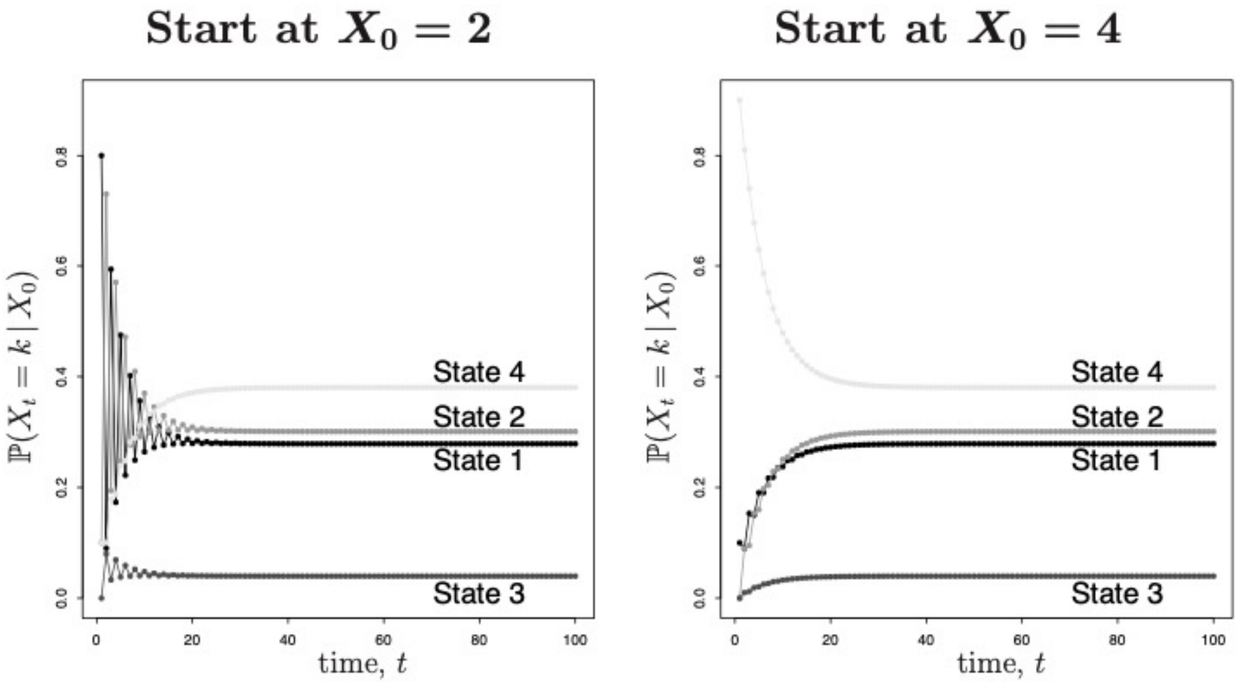
What is happening here is that *each row of the transition matrix P^t converges to the equilibrium distribution* $(0.28,\ 0.30,\ 0.04,\ 0.38)$ *as $t \to \infty$:*

$$P = \begin{pmatrix} 0.0 & 0.9 & 0.1 & 0.0 \\ 0.8 & 0.1 & 0.0 & 0.1 \\ 0.0 & 0.5 & 0.3 & 0.2 \\ 0.1 & 0.0 & 0.0 & 0.9 \end{pmatrix} \quad \Rightarrow \quad P^t \to \begin{pmatrix} 0.28 & 0.30 & 0.04 & 0.38 \\ 0.28 & 0.30 & 0.04 & 0.38 \\ 0.28 & 0.30 & 0.04 & 0.38 \\ 0.28 & 0.30 & 0.04 & 0.38 \end{pmatrix} \quad \text{as } t \to \infty.$$

(If you have a calculator that can handle matrices, try finding P^t for $t = 20$ and $t = 30$: you will find the matrix is already converging as above.)

This convergence of P^t means that *for large t, no matter WHICH state we start in, we always have probability*

- about *0.28* of being in State *1* after t steps;
- about *0.30* of being in State *2* after t steps;
- about *0.04* of being in State *3* after t steps;
- about *0.38* of being in State *4* after t steps.



The **left graph** shows the probability of getting from state 2 to state k in t steps, as t changes: $(P^t)_{2,k}$ for $k = 1, 2, 3, 4$.

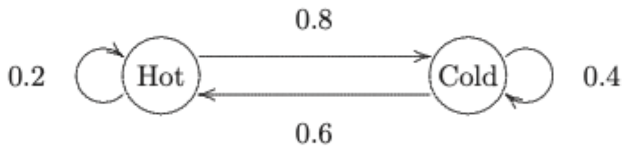
The **right graph** shows the probability of getting from state 4 to state k in t steps, as t changes: $(P^t)_{4,k}$ for $k = 1, 2, 3, 4$.

The ***initial behaviour*** differs greatly for the different start states.

The ***long-term behaviour*** (large t) is the same for both start states.

Lecture

Example 1:



$$P = \begin{pmatrix} 0.2 & 0.8 \\ 0.6 & 0.4 \end{pmatrix}$$

We can show that the general solution for P^t is:

$$P^t = \frac{1}{7}\left\{ \begin{pmatrix} 3 & 4 \\ 3 & 4 \end{pmatrix} + \begin{pmatrix} 4 & -4 \\ -3 & 3 \end{pmatrix} (-0.4)^t \right\}$$

As $t \to \infty, (-0.4)^t \to 0,$ *so*

$$P^t \to \frac{1}{7}\begin{pmatrix} 3 & 4 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} \frac{3}{7} & \frac{4}{7} \\ \frac{3}{7} & \frac{4}{7} \end{pmatrix}$$

This Markov chain will therefore converge to the equilibrium distribution $\boldsymbol{\pi}^T = \left(\frac{3}{7}, \frac{4}{7}\right)$ *as* $t \to \infty$*, regardless of whether the flea starts in state 1 or state 2.*

Introduction: Probability Space

A probability space is a triplet $(\boldsymbol{\Omega}, \mathcal{F}, \mathbb{P})$ that can be described informally as follows:

- $\boldsymbol{\Omega}$ is *the sample space*. We can think of Ω as the set of all possible outcomes in "nature" or in a "random experiment" that we want to model. In this context, "nature" chooses exactly one point $\omega \in \boldsymbol{\Omega}$, but we do not know which one, otherwise, we would have no uncertainty and we would know exactly what is going to happen.
- $\mathcal{F}$ is *a collection of event of interests*. An event is a subset of Ω, so $\mathcal{F}$ is a set of subsets of Ω. We can think of $\mathcal{F}$ as all the information that "nature" has or all the information that is relevant to the modelling of a "random experiment".
- $\mathbb{P}$ is *a function that assigns a probability $P(A)$ to each event $A \in \mathcal{F}$*. In particular, given an event $A \in \mathcal{F}$, $P(A)$ is a number in the interval $[0, 1]$ that represents our belief on how likely the event A is to occur.

Definition

Mathematically, *a probability space* is a triplet $(\boldsymbol{\Omega}, \mathcal{F}, \mathbb{P})$ such that

- $\boldsymbol{\Omega}$ is a set,
- $\mathcal{F}$ is σ-algebra on $\boldsymbol{\Omega}$
- $\mathbb{P}$ is a probability measure on $(\boldsymbol{\Omega}, \mathcal{F})$

Introduction: Generating σ-algebras

Lemma

Let $\{\mathcal{F}_i, i \in \boldsymbol{I}\}$ be a family of σ-algebras on $\boldsymbol{\Omega}$ indexed by a set $\boldsymbol{I} \neq \emptyset$.
The collection $\bigcup_{i \in I} \mathcal{F}_i$ is a σ-algebra on Ω.

Proof:

We have to check the defining properties of a σ-algebra. To this end, we note that the family of events $\bigcap_{i \in I} \mathcal{F}_i$

$$
\begin{aligned}
A_1, A_2, \ldots, A_n, \ldots \in \bigcap_{i \in I} \mathcal{F}_i \; & \Rightarrow A_1, A_2, \ldots, A_n, \ldots \in \mathcal{F}_i \text{ for all } i \in I \\
& \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}_i \text{ for all } i \in I \text{ (because each } \mathcal{F}_i \; is \; a \; \sigma\text{−algebra)} \\
& \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \bigcap_{i \in I} \mathcal{F}_i
\end{aligned}
$$

Introduction: Measurable space and Countably Additive Measure

Definition

A pair $(\Omega, \mathcal{F})$, where Ω is a set and $\mathcal{F}$ is a σ-algebra on Ω, is called *measurable space.*

Definition

Let $(\boldsymbol{S}, \boldsymbol{\mathcal{S}})$ be a measurable space, so that $\boldsymbol{\mathcal{S}}$ is a σ-algebra on the set $\boldsymbol{S}$.
A measure defined on $(\boldsymbol{S}, \boldsymbol{\mathcal{S}})$ is a function $\mu : \boldsymbol{\mathcal{S}} \to [0, \infty]$ that is *countably additive*, i.e., it is such that

(i) $\mu(\emptyset) = 0$, and
(ii) if $A_1, A_2, \ldots, A_n, \ldots \in \boldsymbol{\mathcal{S}}$ is any sequence of pairwise disjoint sets (i.e., $A_i \cap A_j = \emptyset$ for all $i \neq j$), then

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n)$$

Sigma Algebra

- $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space
- X, Z two random variables
- elementary conditional probability :

$$\mathbb{P}[X = x \mid Z = z] = \mathbb{P}[X = x, Z = z]/\mathbb{P}[Z = z]$$

- elementary conditional expectation :

$$\mathbb{E}[X \mid Z = z] = \sum_x x\mathbb{P}[X = x \mid Z = z]$$

- $Y = \mathbb{E}[X \mid \sigma(Z)]$?
  - Y is measurable with respect to $\sigma(Z)$
  - $\mathbb{E}[Y 1_{Z=z}] = \mathbb{E}[X 1_{Z=z}]$

Sigma Algebra

- $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space
- X is a random variable on the probability space with $\mathbb{E}[|X|] < \infty$
- $\mathcal{A} \subset \mathcal{F}$ is a sub σ-algebra

Then there exists a random variable Y such that

- Y is $\mathcal{A}$-measurable with $\mathbb{E}[|Y|] < \infty$
- for any $A \in \mathcal{A}$, we have $\mathbb{E}[Y 1_A] = \mathbb{E}[X 1_A]$.

Moreover, if $\tilde{Y}$ also satisfies the above two properties, then $\tilde{Y} = Y$ a.s.
A random variable Y with the above two properties is called the **conditional expectation** of X given $\mathcal{A}$, and we denote it by $\mathbb{E}[X \mid \mathcal{A}]$.

Remark :

- If $\mathcal{A} = \{\emptyset, \Omega\}$, then $\mathbb{E}[X \mid \mathcal{A}] = \mathbb{E}[X]$.
- If X is $\mathcal{A}$-measurable, then $\mathbb{E}[X \mid \mathcal{A}] = X$.
- If $Y = \mathbb{E}[X \mid \mathcal{A}]$, then $\mathbb{E}[Y] = \mathbb{E}[X]$

Sigma Algebra

- **Probability space** is triple $(\Omega, \mathcal{F}, P)$ where Ω is sample space, $\mathcal{F}$ is set of events (the σ-algebra) and $P : \mathcal{F} \to [0, 1]$ is the probability function.
- σ-**algebra** is collection of subsets closed under complementation and countable unions. Call $(\Omega, \mathcal{F})$ a measure space.
- **Measure** is function $\mu : \mathcal{F} \to \mathbb{R}$ satisfying $\mu(A) \geq \mu(\emptyset) = 0$ for all $A \in \mathcal{F}$ and countable additivity: $\mu(\cup_i A_i) = \sum_i \mu(A_i)$ for disjoint A_i.
- Measure μ is **probability measure** if $\mu(\Omega) = 1$.

Sigma Algebra

- **monotonicity**: $A \subset B$ implies $\mu(A) \leq \mu(B)$
- **subadditivity**: $A \subset \cup_{m=1}^{\infty} A_m$ implies $\mu(A) \leq \sum_{m=1}^{\infty} \mu(A_m)$.
- **continuity from below:** measures of sets A_i in increasing sequence converge to measure of limit $\cup_i A_i$
- **continuity from above:** measures of sets A_i in decreasing sequence converge to measure of intersection $\cap_i A_i$

Why not all Subsets are Sigma-Algebra?

- Uniform probability measure on $[0,1)$ should satisfy **translation invariance:** If B and a horizontal translation of B are both subsets $[0,1)$, their probabilities should be equal.
- Consider **wrap-around translations** $\tau_r(x) = (x + r) \mod 1$.
- By translation invariance, $\tau_r(B)$ has same probability as B.
- Call x, y "equivalent modulo rationals" if $x - y$ is rational (e.g., $x = \pi - 3$ and $y = \pi - 9/4$). An **equivalence class** is the set of points in $[0,1)$ equivalent to some given point.
- There are uncountably many of these classes.
- Let $A \subset [0,1)$ contain **one** point from each class. For each $x \in [0,1)$, there is **one** $a \in A$ such that $r = x - a$ is rational.
- Then each x in $[0,1)$ lies in $\tau_r(A)$ for **one** rational $r \in [0,1)$.
- Thus $[0,1) = \cup \tau_r(A)$ as r ranges over rationals in $[0,1)$.
- If $P(A) = 0$, then $P(S) = \sum_r P(\tau_r(A)) = 0$. If $P(A) > 0$ then $P(S) = \sum_r P(\tau_r(A)) = \infty$. Contradicts $P(S) = 1$ axiom.

Sigma Algebra

- The **Borel σ-algebra** $\mathcal{B}$ is the smallest σ-algebra containing all open intervals.
- Say that $\mathcal{B}$ is "generated" by the collection of open intervals.
- Why does this notion make sense? If $\mathcal{F}_i$ are σ-fields (for i in possibly uncountable index set I) does this imply that $\cap_{i \in I} \mathcal{F}_i$ is a σ-field?

Sigma Algebra

A **filtration** is a non-decreasing family of sub σ–algebras of $\mathcal{F}$ indexed by time, i.e. a family $\mathbb{F} := (\mathcal{F}_t)_{t\in\mathbb{T}}$ such that

$$\mathcal{F}_s \subseteq \mathcal{F}_t,$$

for $s \leq t$, where $t, s \in \mathbb{T}$.

Sigma Algebra

Let $\mathbb{F}$ be a (continuous time) filtration. We say that $\mathbb{F}$ is the **right-continuous filtration** if for any $t \in \mathbb{T}$ we get

$$\mathcal{F}_t = \mathcal{F}_{t_+} \ ,$$

where $\mathcal{F}_{t_+} := \bigcap_{s>t, s\in\mathbb{T}} \mathcal{F}_s$.

Sigma Algebra

process X is said to be **adapted** to filtration $\mathbb{F}$ (or $\mathbb{F}$**-adapted**) if X_t is $\mathcal{F}_t$–measurable for any $t \in \mathbb{T}$.

st. pr $X_t \sim$ predictable if X_{t+1} is $\mathcal{F}_t$- measurable

Sigma Algebra

Let X be a stochastic process. We say that $\mathbb{F}^X := (\mathcal{F}_t^X)_{t \in \mathbb{T}}$, where

$$\mathcal{F}_t^X = \sigma(X_s, s \leq t, s \in \mathbb{T})$$

is a filtration **generated** by stochastic process X.

Sigma Algebra

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a standard probability space with $\Omega = [0,1]$.[9] Let

$$\mathcal{A} := \sigma(N \subset [0,1] : \#N < \infty)$$

denote the σ–algebra of countable sets (and their complements). For time horizon $\mathbb{T} = [0, +\infty)$ we define filtration $\mathbb{F} = (\mathcal{F}_t)_{t \in \mathbb{T}}$ by setting

$$\mathcal{F}_t := \begin{cases} \mathcal{A} & \text{for } t \in [0,1); \\ \mathcal{F} & \text{for } t \in [1,\infty). \end{cases}$$

Next, we define a stochastic process $X = (X_t)_{t \in \mathbb{T}}$ by setting

$$X_t(\omega) := \mathbb{1}_{\Delta}(t,\omega) = \begin{cases} 1 & \text{if } t = \omega \text{ and } t \leq 1/2 \\ 0 & \text{otherwise} \end{cases}, \qquad t \in \mathbb{T}, \omega \in \Omega.$$

where $\Delta := \{(t,t) : t \in [0, \frac{1}{2}]\}$ is a subset of $\mathbb{T} \times \Omega$.

Definition (Moments)

Let X be a discrete random variable, and let $n \geqslant 1$ be an integer. The number

$$E(X^n) = \sum_x x^n p_X(x)$$

is called the n-th moment of X. Notice that the first moment is the mean.

Definition

Let X be a discrete random variable. The function

$$M_X(t) = E(e^{tX})$$

is called the moment generating function (MGF) of X.

Problem (Geometric)

Let X be geometric with parameter p. Show that

$$M_X(t) = \frac{pe^t}{1 - qe^t}.$$

Problem (Binomial)

Let X be binomial with parameters n and p. Show that

$$M_X(t) = (pe^t + q)^n.$$

Problem (Poisson)

Let X be Poisson with parameter λ. Show that

$$M_X(t) = e^{\lambda(e^t - 1)}.$$

Sigma Algebra

$$M_x(t) = E[e^{tx}]$$

$$M'_x = E[\cancel{t}x\, e^{tx}]$$

Theorem

Let X be a discrete random variable. Then

$$M'_X(0) = E(X).$$

In general, for each $n \geqslant 1$,

$$M_X^{(n)}(0) = E(X^n).$$

Sigma Algebra

Theorem

1. *Let X be geometric with parameter p. Then*

$$E(X) = 1/p \quad \textit{and} \quad \operatorname{var}(X) = q/p^2.$$

2. *Let X be binomial with parameters n and p. Then*

$$E(X) = np \quad \textit{and} \quad \operatorname{var}(X) = npq.$$

Sigma Algebra

Theorem (Change of Scale Theorem)

Let $Y = aX + b$, where a and b are real numbers and X is a random variable. Then $M_Y(t) = e^{tb} M_X(at)$.

Sigma Algebra

Theorem (Uniqueness Theorem)

Let X and Y be random variables. If $M_X(t) = M_Y(t)$ for all $t \in [-a, a]$ for some positive real number a, then X and Y have the same distribution, that is, $F_X = F_Y$.

Sigma Algebra

Thank you for your attention!

See next week!