American University of Beirut

MATH/STAT 238: Applied Probability Models

2025–2026 Fall

Weekly summary (last update: December 11, 2025)

Week 1

- Practical information about the course (see the Moodle page and the syllabus).
- What the course is about (see the syllabus).
- **Example 1** (Infinite monkey). A Monkey behind a keyboard repeatedly and independently typing random characters uniformly chosen from 32 characters (26 letters + space + new line + 4 punctuation marks).

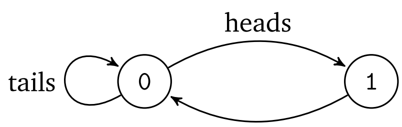
  Q What is the chance that the monkey eventually types: "`ABRACADABRA`"?

  $\boxed{\text{A}}$ 1. Proof sketch: Divide the sequence of characters into non-overlapping blocks of length 11. Each block has a probability of $(1/32)^{11}$ of being "`ABRACADABRA`". Apply the unfortunate pedestrian principle.

- Unfortunate Pedestrian Principle: If every time a pedestrian crosses a street, there is a positive probability ε of them getting hit by a car, then a pedestrian crossing the street over and over will eventually get hit by a car (in fact, infinitely many times).

  An alternative way to say this: Suppose a biased coin has a probability of ε to come up heads where $\varepsilon > 0$. If we repeat flipping the coin, it will eventually come up heads (in fact, infinitely many times).

- **Example 2** (A simple Markov chain).



  We do a random dance using using a fair coin. At each step, we are either standing at our original position (state $\mathtt{0}$) or one step to the right of the original position (state $\mathtt{1}$). We change positions according to the following rules:

  - If at time n we are in state $\mathtt{0}$, we flip the coin to decide whether at time $n+1$ we go to state $\mathtt{1}$ or stay in state $\mathtt{0}$.
  - If at time n we are in state $\mathtt{1}$, at time $n+1$ we simply move back to state $\mathtt{0}$ without flipping the coin.

  Q What will be the distribution of our position after 1 million steps?

  Intuitive expectation: It will reach an equilibrium.

  Q Does the system have an equilibrium distribution? If so, is it unique?

  It turned out there is a unique equilibrium distribution: $p(\mathtt{0}) = 2/3$ and $p(\mathtt{1}) = 1/3$.

  Q What is the probability distribution after 10 steps?

  We have a simple recursion:

$$\begin{cases} p_{n+1}(\mathtt{0}) = p_n(\mathtt{0}) \cdot \dfrac{1}{2} + p_n(\mathtt{1}) \,, & p_0(\mathtt{0}) = 1 \,. \\ p_{n+1}(\mathtt{1}) = p_n(\mathtt{0}) \cdot \dfrac{1}{2} \,. & p_0(\mathtt{1}) = 0 \,. \end{cases}$$

  which can be written in matrix form $p_{n+1} = p_n A$, where

$$p_n = \begin{bmatrix} p_n(\mathtt{0}) & p_n(\mathtt{1}) \end{bmatrix} \qquad \text{and} \qquad A = \begin{bmatrix} 1/2 & 1/2 \\ 1 & 0 \end{bmatrix} \,.$$

It follows that $p_n = p_0 A^n$ for every $n \geqslant 0$. Explicit computation using a computer gives

$$A = \frac{1}{1024}\begin{bmatrix} 683 & 341 \\ 682 & 342 \end{bmatrix} .$$

Hence, $p_{10}(\mathtt{0}) = 683/1024 = 0.6669921875 \approx {}^2\!/_3$. In other words, the distribution after 10 steps is already very close to the equilibrium distribution.

Q As time passes, does the distribution of our position approach its unique equilibrium distribution?

Diagonalizing A, we can compute A^n and use that to compute $p_n = p_0 A^n$. The end result:

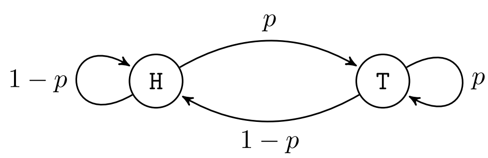
$$p_n(\mathtt{0}) = \big({}^2\!/_3 + ({}^1\!/_3)(-{}^1\!/_2)^n\big)p_0(\mathtt{0}) + \big({}^2\!/_3 - ({}^2\!/_3)(-{}^1\!/_2)^n\big)p_0(\mathtt{1}) .$$

From this formula, we can see that $p_n \to \begin{bmatrix} {}^2\!/_3 & {}^1\!/_3 \end{bmatrix}$ as $n \to \infty$, irrespective of the initial distribution. Furthermore, the convergence is exponentially fast!

Q What will be the empirical distribution of the system over a million steps?

Intuitive expectation: Approximately the same as the unique equilibrium distribution, by some sort of law of large numbers.

▶ **Example 3** (A Bernoulli process). A Bernoulli process (infinite coin flipping experiment) can be viewed as a a Markov chain with two states H and T in which the random choices do not depend on the current state.



Simpler description: A sequence $X_0, X_1, X_2, \dots$ of i.i.d. random variables with values from $\{\mathtt{H}, \mathtt{T}\}$ with $\mathbb{P}(X_n = \mathtt{H}) = p$.

Empirical distribution over n steps:

$$R_{\mathtt{H}}(n) := \frac{\mathbb{1}_{\mathtt{H}}(X_0) + \mathbb{1}_{\mathtt{H}}(X_1) + \cdots + \mathbb{1}_{\mathtt{H}}(X_{n-1})}{n}$$

The law of large numbers (LLN) says that $R_{\mathtt{H}}(n) \to p$ (in some sense) as $n \to \infty$.

Weak LLN: For every $\varepsilon > 0$, $\lim_{n\to\infty} \mathbb{P}\big(|R_{\mathtt{H}}(n) - p| \geqslant \varepsilon\big) = 0$.

(This kind of convergence is called **convergence in probability**.)

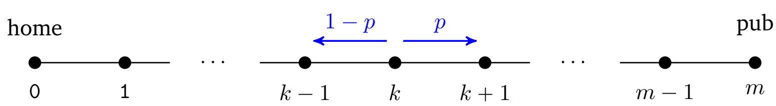
Strong LLN: $\mathbb{P}\left(\lim_{n\to\infty} R_{\mathtt{H}}(n) = p\right) = 1$.

(This kind of convergence is called **almost sure convergence**.)

The weak LLN has a simple proof using Chebyshev's inequality. We will prove the strong LLN later in the course.

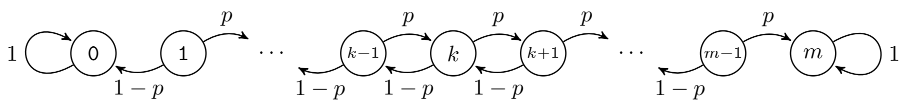
▶ In Example 2, the standard LLN does not apply because the random variables are neither independent nor identically distributed. Nevertheless, as we shall see, there is a sort of LLN for Markov chains.

▶ **Example 4** (Drunkard's walk). A drunkard is wandering on a street taking random steps one way or the other.



The length of the street is m steps. At one end of the street (position $\mathtt{0}$), there is the drunkard's home and at the other end (position m), there is a pub. At each time step, the drunkard goes either one step to towards the pub (with probability p) or one step towards home (with probability $1 - p$), independently of his previous steps. Here, $0 < p < 1$. Once the drunkard reaches either home or the pub, he stays there for the rest of the night.

The drunkard's movement can be viewed as a Markov chain.



The states 0 and m are **absorbing** states; once the system reaches either of these states, it stays there forever.

Ⓠ Does the system have an equilibrium distribution? If so, is it unique?

We observed that there are infinitely many equilibrium distributions. Namely, all distributions of the form $(1-r, 0, 0, \ldots, 0, r)$ (for $0 \leqslant r \leqslant 1$).

Exercise. Show that these are the only equilibrium distributions.

Note that 0 and m are themselves (non-statistical) equilibrium states. The equilibrium distribution $(1-r, 0, 0, \ldots, 0, r)$ (for $0 < r < 1$) indicates a situation in which we know the system is in one of the two (non-statistical) equilibrium states but we are uncertain which one.

Ⓠ Does the distribution of the system approach an equilibrium?

For now, without proof: Yes, but depending on the initial position, it will converge to different equilibrium distributions.

Ⓠ If the drunkard starts at position k, what is the chance that he reaches the pub before going home?

Let X_n denote the position of the drunkard after n steps. This is a random variable that takes its values from the state space $S := \{0, 1, \ldots, m\}$. The sequence $X_0, X_1, X_2, \ldots$ has the following important property:

Markov property. Given the current state X_n, the choice of X_{n+1} is independent of $X_0, X_1, \ldots, X_{n-1}$. In other words, for every $a_0, a_1, \ldots, a_{n+1} \in S$,

$$\mathbb{P}(X_{n+1} = a_{n+1} \mid X_0 = a_0, X_1 = a_1, \ldots, X_n = a_n) = \mathbb{P}(X_{n+1} = a_{n+1} \mid X_n = a_n) \ .$$

Let $q(k) := \mathbb{P}(\text{ending up in the pub} \mid X_0 = k)$. The function q satisfies the following recurrence relation:

$$\begin{cases} q(0) = 0 \ , \\ q(m) = 1 \ , \\ q(k) = pq(k+1) + (1-p)q(k-1) & \text{if } 0 < k < m. \end{cases}$$

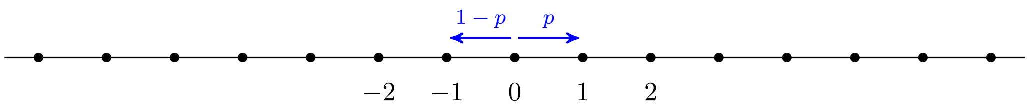
Solving this recursion, we get

- Case $p = 1/2$: $\quad q(k) = \dfrac{k}{m}$.
- Case $p \neq 1/2$: $\quad q(k) = \dfrac{1 - \left(\frac{1-p}{p}\right)^k}{1 - \left(\frac{1-p}{p}\right)^m}$.

Exercise. Derive the above formulas.

Week 2

▶ **Example 5** (Random walk on $\mathbb{Z}$). A particle is moving at random on $\mathbb{Z}$.



At every time step, the particle either moves one unit to the right (with probability p) or one unit to the left (with probability $1-p$), independently of the previous steps. We assume $0 < p < 1$.

The position of the Markov chain is described by a Markov chain with state space $\mathbb{Z}$. Let $X_0, X_1, X_2, \ldots$ denote the consecutive positions of the particle. These are integer-valued random variables satisfying the Markov property such that

$$\mathbb{P}(X_{n+1} = \ell \mid X_n = k) = \begin{cases} p & \text{if } \ell = k+1, \\ 1-p & \text{if } \ell = k-1, \\ 0 & \text{otherwise.} \end{cases}$$

(Q1) Does the model have an equilibrium distribution?

Exercise. No.

(Q2) Will the particle eventually return to its initial position?

The answer depends on the value of p.
Case 1: $p = 1/2$ (Symmetric random walk.)
Claim without proof: Almost surely (i.e., with probability 1) it returns. We say that the RW is **recurrent**.
Case 2: $p \neq 1/2$ (There is a drift either to the right or to the left.)
Claim: There is a positive probability that it never returns. We say that the RW is **transient**.

Proof of transience when $p \neq 1/2$. We can use the result about the drunkard's walk.
By symmetry, we may assume $p > 1/2$, that is, there is a drift towards right. We would like to show that $\mathbb{P}\big(\text{the particle never returns to } 0 \,\big|\, X_0 = 0\big)$ is strictly positive. To do so, let us examine the position of the particle after one step: we either have $X_1 = 1$ or $X_1 = -1$.
When $X_1 = 1$, we can use the result on the drunkard's walk that says that, for every integer $m \geqslant 2$, we have

$$\mathbb{P}\big(\text{the particle visits } m \text{ before } 0 \,\big|\, X_1 = 1\big) = \frac{1 - (\frac{1-p}{p})^1}{1 - (\frac{1-p}{p})^m} .$$

Taking the limit as $m \to \infty$ gives

$$\mathbb{P}\big(\text{the particle never returns to } 0 \,\big|\, X_1 = 1\big) \geqslant 1 - \frac{1-p}{p} .$$

To justify this rigorously, let $E_m := \{\text{the particle visits } m \text{ before } 0\} \cap \{X_1 = 1\}$. Note that $E_2 \supseteq E_3 \supseteq E_4 \supseteq \cdots$ (to reach $m+1$, the particle has to pass m) and

$$\{\text{the particle never returns to } 0\} \cap \{X_1 = 1\} \supseteq \bigcap_{m=2}^{\infty} E_m .$$

Hence, by the **monotone continuity** of the probabilities,

$$\begin{aligned}
\mathbb{P}\big(\text{the particle never returns to } 0 \,\big|\, X_1 = 1\big) &\geqslant \mathbb{P}\Big(\bigcap_{m=2}^{\infty} E_m \,\Big|\, X_1 = 1\Big) \\
&= \lim_{m\to\infty} \mathbb{P}(E_m \mid X_1 = 1) = 1 - \frac{1-p}{p} .
\end{aligned}$$

For the case $X_1 = -1$, we can use the crude bound $\mathbb{P}\big(\text{the particle never returns to } 0 \,\big|\, X_1 = -1\big) \geqslant 0$.
Now, conditioning on the first step (i.e., using the law of total probability), we can write

$$\begin{aligned}
&\mathbb{P}\big(\text{the particle never returns to } 0 \,\big|\, X_0 = 0\big) \\
&\quad = \mathbb{P}(X_1 = 1 \mid X_0 = 0)\,\mathbb{P}\big(\text{the particle never returns to } 0 \,\big|\, X_0 = 0, X_1 = 1\big) \\
&\qquad + \mathbb{P}(X_1 = -1 \mid X_0 = 0)\,\mathbb{P}\big(\text{the particle never returns to } 0 \,\big|\, X_0 = 0, X_1 = -1\big) \\
&\quad = p \cdot \mathbb{P}\big(\text{the particle never returns to } 0 \,\big|\, X_1 = 1\big) \\
&\qquad + (1-p) \cdot \mathbb{P}\big(\text{the particle never returns to } 0 \,\big|\, X_1 = -1\big) , \\
&\quad \geqslant p \cdot \left(1 - \frac{1-p}{p}\right) + (1-p) \cdot 0 \\
&\quad \geqslant 2p - 1 ,
\end{aligned}$$

which is strictly larger than 0 because $p > 1/2$. (For the second equality, we have used the Markov property to drop the condition $X_0 = 0$.) $\square$

Exercise. Use the result on the drunkard's walk to prove that the symmetric random walk on $\mathbb{Z}$ is recurrent.

(Q3) How many times will the particle return to its initial position?

A If $p = 1/2$ (the recurrent case), almost surely infinitely many times.
If $p \neq 1/2$ (the transient case), almost surely finitely many times.
Proof sketch: Every time the particle returns to its initial position, it starts afresh. Apply the unfortunate pedestrian principle.

Q4 If the particle starts at position 0, will it eventually visit a given position $m \neq 0$?

A If $p = 1/2$ (the recurrent case), it almost surely does.
If $p \neq 1/2$ (the transient case), the answer depends on whether $m > 0$ or $m < 0$. If $m > 0$, the particle almost surely visits m. If $m < 0$, there is a positive probability that the particle never visits m.
Exercise: Prove these.

Q5 In the recurrent case, what is the expected return time to the initial position?

Claim without proof: $+\infty$.

▶ **Example 5′** (Random walk on $\mathbb{Z}^d$). A particle is performing a random walk on $\mathbb{Z}^d$ (where $d = 1, 2, \ldots$). At each time step, it moves one unit to one of the $2d$ standard directions, each with probability $1/2d$.

Q Will the particle eventually return to its initial position?

Claim without proof: It depends on the dimension.

- When $d \leqslant 2$, it almost surely does; the Markov chain is recurrent.
- When $d \geqslant 3$, there is a positive probability that it does not; the Markov chain is transient.

▶ Holiday

Week 3

▶ **Example 6** (Galton-Watson process). We have a population of bacteria that evolves at random at discrete time steps. At the beginning, there are m bacteria. At each time step, each bacterium either divides into two new bacteria (with probability p) or dies (with probability p), independently of the other bacteria in the current and previous generations. This is an example of a Galton-Watson process or a branching process.

The evolution of the population can be described with a discrete-time Markov process in which the "state" at time n represents the size of the population after n steps. As usual, we let X_n denote the state at time n.

Q Will the population survive indefinitely or eventually go extinct?

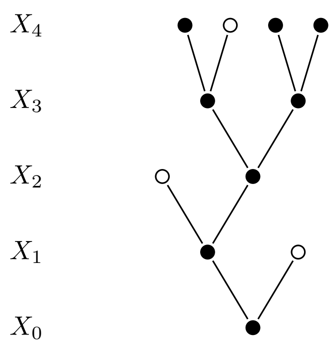
Observation 1: There is always a positive probability of extinction. Namely, with probability $(1-p)^m$, all the bacteria die in the first step, so the probability of eventual extinction is at least $(1-p)^m$.

Observation 2: It is enough to consider the case $X_0 = 1$. Indeed, since the fate of the family line of different bacteria from the same generation are independent, we have

$$\mathbb{P}(\text{extinction} \mid X_0 = m) = \mathbb{P}(\text{extinction} \mid X_0 = 1)^m \ .$$

So, let us assume that $m = 1$.

A nice way to mathematically represent the entire evolution history of a population is by its **family tree**, which is a random rooted binary tree.



Each bacterium is represented by a vertex, which is either black (if the bacterium self-replicates) or white (if the bacterium dies). From each self-replicating bacterium there is an edge to each of its children. The root corresponds to the initial bacterium and the vertices at level n correspond to the bacteria in generation n.

Intuitively, whether the population goes extinct or has a chance of indefinite survival has to do with the expected number of children of a single bacterium, which is $(1-p) \times 0 + p \times 2 = 2p$. If this expected value is less than 1, we expect extinction, and if it is larger than 1, we expect survival with exponential growth.

Claim without proof:

- If $2p < 1$, then $\mathbb{P}(\text{survival} \mid X_0 = 1) = 0$. This is the **sub-critical** regime.
- If $2p = 1$, then $\mathbb{P}(\text{survival} \mid X_0 = 1) = 0$. This is the **critical** regime.
- If $2p > 1$, then $\mathbb{P}(\text{survival} \mid X_0 = 1) > 0$. This is the **super-critical** regime.

The Galton-Watson process is one of the simplest models in probability theory that that exhibits such a **phase transition** as we vary its parameter.

Q In the super-critical regime, what is the probability of survival?

Let $q(p) := \mathbb{P}(\text{extinction} \mid X_0 = 1)$.
Claim without proof: $q(p)$ is a fixed point of the the function $\varphi(z) := 1-p+pz^2 = \mathbb{E}[z^{\#\text{ of children of a bacterium}}]$.

Q Is $q(p)$ continuous? Is it increasing? Where is it differentiable? Where is it convex upward or downward?

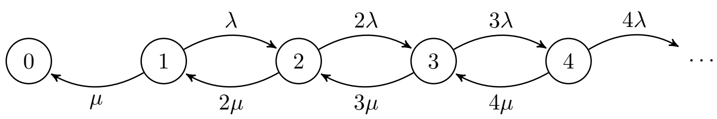
Exercise. Use Desmos or a similar mathematical software to explore how the fixed point of $\varphi(z)$ varies with p. Draw the graph of the probability of survival as a function of p.

▶ **Example 7** (Continuous-time Galton-Watson process). Again we have a population of bacteria, but the self-replication and death of different bacteria are not synchronized. Instead, they are triggered by exponential clocks. More explicitly, each bacterium has two independent exponential random variables attached to it:

- T_{b} (with rate λ): triggering self-replication,
- T_{d} (with rate μ): triggering death.

Suppose the bacterium is born at time t_0. If $T_{\mathsf{b}} < T_{\mathsf{d}}$, then at time $t_0 + T_{\mathsf{b}}$ the bacterium self-replicates and turns into two new bacteria. If $T_{\mathsf{d}} < T_{\mathsf{b}}$, then at time $t_0 + T_{\mathsf{d}}$ the bacterium dies. The birth and death exponential random variables of each bacterium are independent of each other and of the exponential random variables corresponding to all the other bacteria of all times.

The evolution of the population of bacteria is an example of a continuous-time Markov chain.



The state of the system indicates the current number of bacteria in the population. A transition from state k to state $k+1$ occurs if one of the k bacteria in the population self-replicates and likewise a transition from k to $k-1$ occurs if one of the k bacteria dies.

Exercise. The minimum of two independent exponential random variables with rates λ_1 and λ_2 is again an exponential random variable with rate $\lambda + \mu$. The minimum of k independent exponential random variables with rate λ is again an exponential random variable with rate $k\lambda$.

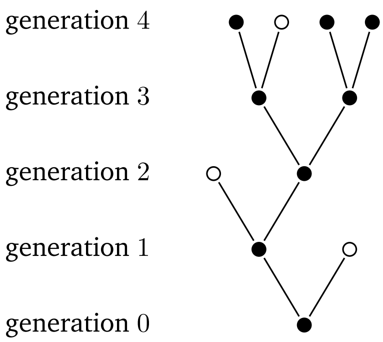
Thus, once in state k, the system takes an exponential time before it jumps to either state $k+1$ (with rate $k\lambda$) or state $k-1$ (with rate $k\mu$).

We denote the state at time $t \in [0, \infty)$ by X_t.

Q Will the population survive indefinitely or eventually go extinct?

Reasonable guess: $\mathbb{P}(\text{survival} \mid X_0 = 1) = 0$ if $\lambda < \mu$, and $\mathbb{P}(\text{survival} \mid X_0 = 1) > 0$ if $\lambda > \mu$. This would be analogous to the result about the discrete-time version of the process.

The analogy with the discrete-time Galton-Watson process can be made precise once we note that the evolution of the population starting from a single bacterium can still be described by its family tree.



The difference with the discrete-time model is that the family lines of the descendents of a bacterium do not evolve synchronously. Nevertheless, the vertices at the nth level still represent the nth generation descendents of the original bacterium, albeit these descendents might have lived in different periods. Thus, the survival/extinction problem in the continuous-time model is equivalent to the survival/extinction problem in the discrete-time model with parameter

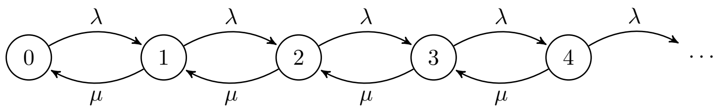
$$p := \mathbb{P}(\text{the bacterium self-replicates before it dies}) = \mathbb{P}(T_{\mathsf{b}} < T_\lambda) .$$

Exercise: If S and T are independent exponential random variables with rates λ and μ respectively, then $\mathbb{P}(S < T) = \lambda/(\lambda + \mu)$.

Therefore, assuming the result about the discrete-time Galton-Watson process, we obtain the following:

- If $\lambda < \mu$, then $\mathbb{P}(\text{survival} \mid X_0 = 1) = 0$. This is the **sub-critical** regime.
- If $\lambda = \mu$, then $\mathbb{P}(\text{survival} \mid X_0 = 1) = 0$. This is the **critical** regime.
- If $\lambda > \mu$, then $\mathbb{P}(\text{survival} \mid X_0 = 1) > 0$. This is the **super-critical** regime.

▶ **Example 8** (A Markovian queue). Customers arrive and line up at a service desk according to a Poisson process with rate λ. The service time of each customer is an exponential random variable with rate μ. We assume that the service times are independent of one another and of the arrival process. This is one of the simplest models of a queuing system. The evolution of the number of customers on the line (including the one at the desk) is a continuous time Markov chain similar to the continuous-time Galton-Watson process.



Ⓠ Will the size of the queue blow up (i.e., increase indefinitely) or reach an equilibrium distribution?

Later, we will see a correspondence with a simple random walk on $\mathbb{N} := \{0, 1, 2, \ldots\}$ with reflection at the origin.

Claim without proof:

- If $\lambda < \mu$, then the system approaches an equilibrium distribution. This is the **sub-critical** regime.
- If $\lambda = \mu$, then the system is recurrent but has no equilibrium distribution. This is the **critical** regime.
- If $\lambda > \mu$, then the system is transient (it *blows up*). This is the **super-critical** regime.

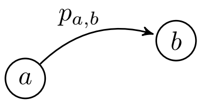
Other questions we may want to answer:

Ⓠ In the sub-critical regime, what is the average queue length at equilibrium?

Ⓠ What is the average waiting time for each customer?

▶ Informal description of a (finite-state, discrete-time) Markov chain:
A **Markov chain** (MC) is a system whose state randomly changes at discrete time steps according to random choices that depend only on the current state.

▶ Mathematical description of a Markov chain:
The **state space** is a finite or countably infinite set S. For each pair $a, b \in S$ of state, there is a **transition probability** $p_{a,b}$ that indicates the probability that, in the next step, the system jumps to state b if its current state is a.



We often “store” these transition probabilities in a function $K : S \times S \to [0,1]$, where $K(a,b) := p_{a,b}$.

Note: For each $a \in S$, the function $K(a,\cdot)$ is a probability distribution, thus $\sum_{b\in S} K(a,b) = 1$.

Mathematically, (the trajectory of) a **Markov chain** (MC) with state space S and transition probabilities K is a sequence

$$X_0, X_1, X_2, \dots$$

of S-valued random variables such that

$$\begin{aligned}
&\mathbb{P}(X_{n+1} = a_{n+1} \mid X_0 = a_0, X_1 = a_1, \dots, X_n = a_n) \\
&\quad = \mathbb{P}(X_{n+1} = a_{n+1} \mid X_n = a_n) && (\bigstar) \\
&\quad = K(a,b) && (☾)
\end{aligned}$$

for every $n \in \mathbb{Z}^+$ and $a_0, a_1, \dots, A_{n+1} \in S$.

Equation ($\bigstar$) is called the **Markov property**. It means that, given X_n, the random variables X_{n+1} and $(X_0, X_1, \dots, X_{n-1})$ are independent.

Equation (☾) implicitly implies that $\mathbb{P}(X_{n+1} = a_{n+1} \mid X_n = a_n)$ does not depend on n. This is called **time homogeneity**.

- More generally, a Markov chain refers to a (finite or infinite) sequence of discrete random variables that satisfy the Markov property. A Markov chain (X_n) in which $\mathbb{P}(X_{n+1} = a_{n+1} \mid X_n = a_n)$ does not depend on n is said to be **time-homogeneous**, and otherwise **time-inhomogeneous**. In this course, unless otherwise stated, all Markov chains are assumed to be time-homogeneous.
- *Remark:* Markov processes with uncountable state spaces also show up in some applications but mathematically defining and analyzing such systems requires more technical background.
- *Exercise* (Implications of the Markov property):
  Let $X_0, X_1, \dots$ be a Markov chain with state space S and transition probabilities K.
  (a) Show that $\mathbb{P}(X_0 = a, X_1 = b, X_2 = c) = \mathbb{P}(X_0 = a)K(a,b)K(b,c)$ for every $a, b, c \in S$. Generalize!
  (b) Let $a, b, c \in S$. Show that, given the event $\{X_n = b\}$, the events $\{X_{n-1} = a\}$ and $\{X_{n+1} = c\}$ are independent.
  (c) Let $b \in S$ and $A, C \subseteq S$. Show that, given the event $\{X_n = b\}$, the events $\{X_{n-1} \in A\}$ and $\{X_{n+1} \in C\}$ are independent.
  (d) Let $m < n < \ell$. Show that, given the value of X_n, the random variables
  $$(X_m, X_{m+1}, \dots, X_{n-1}) \qquad \text{and} \qquad (X_{n+1}, X_{n+2}, \dots, X_\ell)$$
  are independent.
  (e) Let $B \subseteq S$. Given the event $\{X_n \in B\}$, are the random variables X_{n-1} and X_{n+1} necessarily independent? If not, provide a counter-example.

Week 4

- Consider a Markov chain $(X_n)_{n=0}^{\infty}$ with state space S and transition function $K\colon S \times S \to [0,1]$. By the distribution of the Markov chain at time n we mean the probability distribution (pmf) of X_n, that is,
  $$\nu_n\colon S \to [0,1] \qquad\qquad \nu_n(a) := \mathbb{P}(X_n = a)\ .$$
  Q How can we find ν_{n+1} if we know ν_n?

  A For each $b \in S$,
  $$\begin{aligned}
  \nu_{n+1}(b) &= \mathbb{P}(X_{n+1} = b) \\
  &= \mathbb{E}[\mathbb{P}(X_{n+1} = b \mid X_n)] && \text{(conditioning on } X_n\text{)} \\
  &= \sum_{a\in S} \mathbb{P}(X_n = a)\, \mathbb{P}(X_{n+1} = b \mid X_n = a) \\
  &= \sum_{a\in S} \nu_n(a) K(a,b)\ .
  \end{aligned}$$

▶ *Remark*: Make sure you understand the notation $\mathbb{P}(X_{n+1} = b \mid X_n)$ if you have not seen it before. The **conditional probability** $\mathbb{P}(E \mid X)$ of an event E **given a** (discrete) **random variable** X is itself a random variable of the form $g(X)$ where $g(a) := \mathbb{P}(E \mid X = a)$. This is a powerful notation as it allows to write the law of total probability in a concise form $\mathbb{P}(E) = \mathbb{E}[\mathbb{P}(E \mid X)]$. In this special case, the law of total probability is simply referred to as **conditioning on** X.

▶ Note that the expression $\sum_{a \in S} \nu_n(a) K(a, b)$ can be thought of in terms of matrix multiplication. Namely, if we arbitrarily order the elements of S so that $S = \{a_1, a_2, \ldots, a_m\}$, view ν_n as a row vector and K as a square matrix, then the above recursion reads

$$
\begin{aligned}
&\begin{bmatrix} \nu_{n+1}(a_1) & \nu_{n+1}(a_2) & \cdots & \nu_{n+1}(a_m) \end{bmatrix} \\
&\qquad = \begin{bmatrix} \nu_n(a_1) & \nu_n(a_2) & \cdots & \nu_n(a_m) \end{bmatrix}
\begin{array}{cc}
\begin{array}{ccc} & \text{column } j & \\ & \downarrow & \end{array} & \\
\begin{bmatrix} \ddots & \vdots & \iddots \\ \cdots & K(a_i, a_j) & \cdots \\ \iddots & \vdots & \ddots \end{bmatrix} & \leftarrow \text{row } i
\end{array}
\end{aligned}
$$

The function K viewed as a matrix is called a **stochastic matrix**:

(i) Its entries are from $[0, 1]$, and

(ii) The entries in each row add up to 1:

$$\sum_{b \in S} K(a, b) = 1 \quad \text{for all } a \in S.$$

A **doubly stochastic** matrix has both row-sums and column-sums equal to 1.

▶ Q For $\ell \geqslant 0$, how can we find $\nu_{n+\ell}$ if we know ν_n?

A For $\ell \geqslant 2$, iterating the above gives

$$\nu_{n+\ell} = \nu_{n+\ell-1} K = (\nu_{n+\ell-2} K) K = \cdots = \nu_n K^\ell$$

(We have used the fact that matrix multiplication is associative.)
Clearly, $\nu_{n+\ell} = \nu_n K^\ell$ also holds for $\ell = 0, 1$.

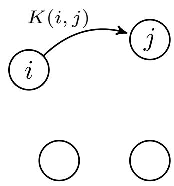
Exercise: Derive the identity $\nu_{n+2} = \nu_n K^2$ from scratch by conditioning on X_n and X_{n+1}.

▶ Two representations of the transition probabilities:

<u>transition matrix</u>

$$
\begin{array}{cc}
 & \begin{array}{ccc} & \text{column } j & \\ & \downarrow & \end{array} \\
\text{row } i \rightarrow & \begin{bmatrix} \ddots & \vdots & \iddots \\ \cdots & K(a_i, a_j) & \cdots \\ \iddots & \vdots & \ddots \end{bmatrix}
\end{array}
$$

<u>transition graph</u>



▶ Interpretation of K^ℓ:

$$
\begin{aligned}
K(a, b) &= \mathbb{P}(X_{n+1} = b \mid X_n = a) \\
K^2(a, b) &= \sum_x K(a, x) K(x, b) = \mathbb{P}(X_{n+2} = b \mid X_n = a) \\
&\ \,\vdots \qquad\qquad \vdots \\
K^\ell(a, b) &= \mathbb{P}(X_{n+\ell} = b \mid X_n = a)
\end{aligned}
$$

▶ A probability distribution $\pi \colon S \to [0, 1]$ is an **equilibrium distribution** (or an **stationary distribution**, or a **steady-state distribution**) for the Markov chain if

$$X_n \sim \pi \qquad \text{implies} \qquad X_{n+1} \sim \pi\ .$$

In matrix notation, this is equivalent to the identity $\pi K = \pi$. More explicitly, π is a stationary distribution if

$$\sum_{x \in S} \pi(x) K(x, a) = \pi(a) \qquad \text{for every } a \in S.$$

The latter equation (or its matrix form) is referred to as the **balance equation**.

If $X_0 \sim \pi$, we say that the Markov chain $(X_n)_{n=0}^{\infty}$ is **stationary** (or **in equilibrium**).

Note that stationary distributions are in fact properties of the transition function K rather than a trajectory $(X_n)_{n=0}^{\infty}$.

- *Remark*: A Markov chain (or a transition function) can have either 0, or 1, or infinitely many stationary distributions. This is because any convex combination

$$\pi'' = c\pi + (1-c)\pi' \qquad \text{(for } 0 \leqslant c \leqslant 1\text{)}$$

of two stationary distributions π and π' is again a stationary distribution. (Recall the drunkard's walk.)

- Three fundamental questions regarding stationary distributions:

(Q1) Do stationary distributions exist?

If so,

(Q2) Is the stationary distribution unique?

(Q3) Does the distribution of the Markov chain converge to a stationary distribution?

We first discuss these questions for finite-state MCs and later address them in the case of general countable-state MCs.

- **Existence Theorem** (for finite-state MCs). Every finite-state MC has at least one stationary distribution.
- Various proofs of the existence theorem:
  - Using linear algebra (what we did in class)
  - Topological (see the suggested exercises)
  - Probabilistic (we will late see it in the setting of countable-state MCs)
  - ...
- Uniqueness can be recognized from the transition graph $\longrightarrow$ examples
- Irreducibility

Week 5

- Midterm 1
- Uniqueness theorem + general version for countable-state MCs
- Detailed balance and reversibility
- Various proofs:
  - "Extremal analysis" (works only for finite-state MCs)
  - Similar but for harmonic functions (exam)
  - Probabilistic: later for countable-state MCs
- Simple proof of uniqueness for random walk on a cycle
- Proof of the uniqueness theorem for finite-state irreducible MCs
- Mixing/ergodic MCs
- For finite-state MCs, mixing can be recognized by looking at the transition graph only $\longrightarrow$ examples
- Various proofs of convergence:
  - Using linear algebra
  - Using couplings
  - Using information theory
- Mixing/ergodic MCs
- For finite-state MCs, mixing can be recognized by looking at the transition graph only $\longrightarrow$ examples

Week 6

- Periods and aperiodicity
- For infinite-state MCs, there is another obstacle: probability mass “escaping to infinity”
- Various proofs of convergence:
  - Using linear algebra
  - Using couplings
  - Using information theory
- Proof of the convergence theorem for finite-state irreducible MCs
- Idea of coupling: thought experiment
- Proof modulo the lemma $\mathbb{P}(T < \infty) = 1$.
- Proof of $\mathbb{P}(T < \infty) = 1$ using the other lemma.
- Reducing the other lemma to a number theory lemma
- Laziness makes the MC aperiodic
- Lazy Ehrenfests model; interpretation of special case $\varepsilon = 1$
- Hard-core model → irreducibility, aperiodicity, reversible distribution
- Interpretation of special case $\lambda = 1$: uniform sample from independent sets of a graph
- Idea of Markov chain Monte Carlo method for sampling from a distribution
- Iterative card shuffling methods are of this type

Week 7

- Markov chain Monte Carlo
- Gibbs sampler for graph colourings
- Decision problem
- Counting problem
- Sampling problem
- Jerrum: Fast convergence when $k \geqslant 2\Delta + 1$; fast sampling
- Randomized approximate counting using Gibbs sampler
- Hitting times and return times: motivation + definitions
- Order of hitting probability: interpretation in terms of betting
- Order of hitting probability: harmonic functions
- Order of hitting probability: 1st example

Week 8

- Midterm 2
- Order of hitting probabilities: 1st example + alternative argument (based on random time)
- Order of hitting probabilities: 2nd example
- Mean hitting times of a target set A: $u(x) = 1 + (Ku)(x)$ for $x \notin A$.
- Mean hitting times: example 1 + alternative solution
- Mean hitting times: example 2 + alternative solution
- Mean hitting times: example 3
- Sample space of a MC
- Stopping times and the strong Markov property
- Transience and recurrence: definition
- Let $q_{a,b} := \mathbb{P}(T_b < \infty \mid X_0 = a)$. Proposition: $\mathbb{P}(T_b^k < \infty \mid X_0 = a) = q_{a,b}q_{b,b}^{k-1}$ for every $k \geqslant 2$.

Week 9

- Recurrence or transience of a state
- Examples in finite-state MCs: can be recognized from the transition graph
- Examples in countable-state MCs: random walk on $\mathbb{Z}$ (unbiased: each state recurrent; biased: each state transient)
- Statement of decomposition theorem
- N_x: number of visits to x.
- Corollary of the proposition from last time:
  Starting from $x \in S$,
  If x is recurrent, then $N_x = \infty$ almost surely (apply the proposition).
  If x is transient, then N_x is geometric with parameter $1 - q_{x,x}$.
- Corollary: For each $x \in S$, either $\mathbb{P}(N_x = \infty \mid X_0 = x)$ or $\mathbb{E}[N_x \mid X_0] < \infty$.
- Remark: For an arbitrary non-negative RV M, there are four possibilities:
  - $M = \infty$ almost surely.
  - $0 < \mathbb{P}(M = \infty) < 1$ (In particular, $\mathbb{E}[M] = \infty$).
  - $M < \infty$ almost surely but $\mathbb{E}[M] = \infty$.
  - $\mathbb{E}[M] < \infty$ (In particular, $M < \infty$ almost surely).
- Notation: $a \rightsquigarrow b$ means $K^\ell(a,b) > 0$ for some $\ell > 0$ (i.e., there is a path from a to b).
- Proposition: Let $x, y \in S$ be distinct. If x is recurrent and $x \rightsquigarrow y$, then
  (a) $q_{x,y} = 1$ (i.e., starting from x, almost surely we visit y).
  (b) $q_{y,x} = 1$.
  (c) y is also recurrent.

  Proofs: (a) by MP. (b) by unfortunate pedestrian. (c) by strong MP.
- Proposition: Let $C \subseteq S$ be an irreducible component (i.e., irreducible and closed). Then,
  (a) Either every $x \in C$ is transient, or every $x \in C$ is recurrent.
  (b) If C is finite, then every $x \in C$ is recurrent.

Observe: (a) is a corollary of the previous proposition.
Proof 1 of (b): Unfortunate pedestrian.
Proof 2 of (b): Use the following general way of expanding the expectation.

- General trick: The expectation of a random variable taking values from non-negative integers can be written as follows:
$$\mathbb{E}[M] = \sum_{n=1}^{\infty} \mathbb{P}(M \geqslant n)$$
This holds even if one of the two sides is infinite, in which case it becomes $\infty = \infty$. It also continues to hold if M is allowed to take value $+\infty$.

  Proof when $M < \infty$ almost surely: change the order of summation. The case $\mathbb{P}(M = +\infty) > 0$ can be verified separately.

  A similar expression for non-negative continuous random variables.

- Stationary distributions and positive recurrence

- Two theorems that:

  I. Generalize the existence and uniqueness theorems to infinite-state MCs (with probabilistic proofs).
  II. Give explicit interpretations of the stationary distributions in terms of return times and number of visits.
  III. Make connection between stationarity and recurrence.

- Informal discussion

- Notation: $\mathbb{P}_x(\cdot) = \mathbb{P}(\cdot \mid X_0 = x)$ and $\mathbb{E}_x[\cdot] = \mathbb{E}[\cdot \mid X_0 = x]$.

- Notation: Given a state a and a stopping time T, we let
$$\begin{aligned} N_a^T &:= \#\text{ of visits to } a \text{ before } T \\ &= \sum_{n=0}^{T-1} \mathbb{1}_a(X_n) \\ &= \sum_{n=0}^{\infty} \mathbb{1}_a(X_n) \cdot \mathbb{1}_{T>n} \end{aligned}$$

- **Theorem 1**. Let $(X_n)_{n\geqslant 0}$ be a MC with a countable state space S. Let $x \in S$ be such $\mathbb{E}_x[T_x^+] < \infty$. Then,
$$\pi_x(y) := \frac{1}{\mathbb{E}_x[T_x^+]}\,\mathbb{E}_x[N_y^{T_x^+}] \qquad (\text{for } y \in S)$$
is a stationary distribution.

- **Theorem 2** (Kac's recurrence theorem). Let $(X_n)_{n\geqslant 0}$ be an irreducible MC with a countable state space S. Let π be a stationary distribution. Then, for every $x \in S$, ($\mathbb{E}_x[T_x^+] < \infty$ and)
$$\pi(x) = \frac{1}{\mathbb{E}_x[T_x^+]}\ .$$

- Definition: A recurrent state $x \in S$ is said to be
  - **positive recurrent** if $\mathbb{E}_x[T_x^+] < \infty$.
  - **null recurrent** if $\mathbb{E}_x[T_x^+] = \infty$.

- **Corollary 1** (existence theorem). A MC has a stationary distribution if and only if it has a positive recurrent state.

- **Corollary 2** (uniqueness theorem). An irreducible MC has at most one stationary distribution.

- **Corollary 3**. Let $C \subseteq S$ be an irreducible component (i.e., irreducible and closed). Then,
  - Either all $x \in S$ are positive recurrent,
  - Or all $x \in S$ are null recurrent,
  - Or all $x \in S$ are transient.

- Proof of Theorem 1 (one case was left for next time).

Week 10

- The rest of the proof of Theorem 1
- "Proof" of Theorem and the spot-the-error game
- **Theorem 3** (Convergence theorem). Let $K\colon S \times S \to [0,1]$ be the transition matrix of a countable state MC. Suppose that K is irreducible and aperiodic, and has a stationary distribution π. Then, for every initial distribution μ, we have $\mu K^n \to \pi$ as $n \to \infty$.

  In other words, for every trajectory $(X_n)_{n\geqslant 0}$, the distribution of X_n converges to π as $n \to \infty$.

  Proof. The coupling proof given in the finite-state case can be adapted to this more general setting.
- *Remark 1*: An equivalent statement of the theorem: An irreducible and aperiodic countable-state MC is mixing if and only if one (and hence all) of its states are positive recurrent.
- *Remark 2*:

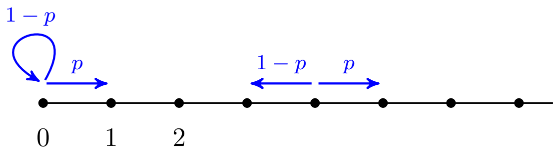
  Q In what sense $\mu K^n \to \pi$?

  $\boxed{\text{A1}}$ $(\mu K^n)(a) \to \pi(a)$ for every $a \in S$.

  $\boxed{\text{A2}}$ $\left\|\mu K^n - \pi\right\|_{\mathsf{TV}} \to 0$.

  Exercise. Verify that the above two notions of convergence are equivalent.
- *Remark 3*. Unlike in the finite-state case, the convergence in the case of mixing countable-state MCs is not necessarily uniform in the starting distribution.

  Example. Consider a random walk on $\mathbb{N} = \{0, 1, 2, \ldots\}$ with a drift towards the origin and lazy reflection at the origin.

  $0 < p < 1/2$

  This is an irreducible and aperiodic Markov chain.

  Exercise. The MC has a stationary distribution that satisfies detailed balance.

  By the theorem, this Markov chain is mixing. However, depending on where it starts, it can take arbitrarily long for its distribution to converge to its unique stationary distribution.
- *Remark 4*. If a stationary distribution does not exist, it may still be the case that the MC "forgets its initial condition" over time.

  Exercise. Consider a simple, symmetric random walk on $\mathbb{Z}$. Let K denote its transition function. Adapt the coupling proof of the convergence theorem to show that for every two initial distributions μ and ν, we have $\left\|\mu K^n - \nu K^n\right\|_{\mathsf{TV}} \to 0$ as $n \to \infty$.
- **Branching process** (a.k.a., the Galton-Watson process): We have a population of bacteria. At every step, each bacterium is replaced with a random number of offsprings. We assume that the number of offsprings of different bacteria are independent of one another have a common distribution $p\colon \mathbb{N} \to [0,1]$.
- *Remark*: In the example we considered earlier, each bacterium can have either two or no offsprings, hence $p(2) = \alpha$ and $p(0) = 1 - \alpha$ for some $0 \leqslant \alpha \leqslant 1$. In that example, the family tree of a bacterium had a neat representation as a random binary tree. For a general p, the tree representation does not work very well.
- Mathematical formulation: Let X_n denote the population size at generation n. Then,

  $$X_n = Z_1^{(n)} + Z_2^{(n)} + \cdots + Z_{X_{n-1}}^{(n)}$$

  where $Z_k^{(n)}$ indicates the number of offsprings of the kth bacterium in generation $n-1$. Here, the random variables $Z_k^{(n)}$ (for $n \geqslant 1$ and $k \geqslant 1$) are i.i.d. with common distribution p. We also assume that the family $\big(Z_k^{(n)}\big)_{n,k\geqslant 1}$ is independent of the initial population. With this formulation, $(X_n)_{n\geqslant 0}$ is a Markov chain with state space $\mathbb{N} := \{0, 1, 2, \ldots\}$.

- Fundamental question:

  Q Does the population have a chance of surviving indefinitely, or will it almost surely go extinct?

  Note: As long as $p(0) > 0$, the population has a strictly positive probability of going extinct. For instance, there is a positive probability that none of the bacteria in generation 0 has an offspring.

  As discussed earlier, the answer depends on the average number of offsprings per bacterium:

$$\mu := \mathbb{E}[Z_k^{(n)}] = \sum_{m=0}^{\infty} p(m)m \ .$$

- **Theorem**. Consider a branching process in which the average number of offsprings per individual is μ.
  - If $\mu < 1$, then the population almost surely goes extinct (the **sub-critical** regime).
  - If $\mu = 1$, then the population almost surely goes extinct (the **critical** regime).
  - If $\mu > 1$, then there is a positive probability that the population survives indefinitely (the **super-critical** regime).

- Proof of the sub-critical case.
  - Let $c := \mathbb{E}[X_0]$. Use conditioning to show that $\mathbb{E}[X_n] = c\mu^n$, which goes to 0 as $n \to \infty$.
  - Apply Markov's inequality to get $\mathbb{P}(X_n \neq 0) = \mathbb{P}(X_n \geqslant 1) \leqslant c\mu^n$.
  - Let

$$N := \langle \text{the number of } n\text{s for which } X_n \neq 0 \rangle = \sum_{n=0}^{\infty} \mathbb{1}_{X_n \neq 0}$$

    which takes its values in $\mathbb{N} \cup \{\infty\}$. Then, by countable additivity (since the terms are non-negative),

$$\mathbb{E}[N] = \sum_{n=0}^{\infty} \mathbb{E}[\mathbb{1}_{X_n \neq 0}] = \sum_{n=0}^{\infty} \mathbb{P}(X_n \neq 0) \leqslant \sum_{n=0}^{\infty} c\mu^n < \infty \ .$$

    Therefore, $\mathbb{P}(N = \infty) = 0$. In other words, almost surely, there exists an n_0 such that $X_n = 0$ for all $n \geqslant n_0$.

Week 11

- Continuation of branching process
- Law of large numbers

Week 12

- Proof of the Strong LLN when the random variables have finite 4th moments.
- Law of large numbers for Markov chains and its proof
- Poisson processes: various characterizations

Week 13

- Poisson processes: superposition theorem + colouring theorem
- Finite-state continuous-time MCs: the transitions are triggered by independent Poisson clocks
- Transition semigroup $(K^t)_{t \geqslant 0}$, infinitesimal generator L
- Kolmogorov's forward and backward equations
- Exercise: differentiability of K^t for $t > 0$.
- Example with two states: explicit computation
- In general: $K^t = \mathrm{e}^{tL}$ for $t \geqslant 0$.

Week 14

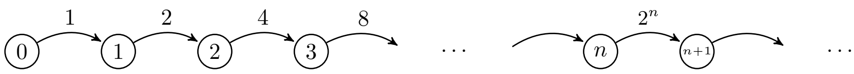
- Two generic constructions of continuous-time MCs:
  - (I) Prescribing transition rates (i.e., the generator L)
  - (II) From a discrete-time MC clocked by a Poisson precess
- **Theorem** (finite-state continuous-time MCs). Every finite-state continuous-time MC of type I has a description of type II and vice versa. Moreover, every transition semigroup on a finite state space that is right-continuous at $t = 0$ describes a MC of type I/II.
- Translations between type I and type II descriptions.
- When the state space is infinite, a type II model runs into no complications. However, constructing type I models becomes more subtle.

  Two things that can go wrong:

  (1) *Instability*: $\lambda_a := \sum_{\substack{b\in S\\ b\neq a}} \lambda_{a,b}$ diverges to $+\infty$. (This means instantaneous exit!)

  (2) *Explosion*: "Escaping to infinity" in finite time!

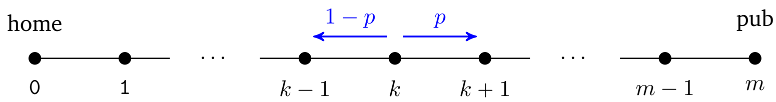
  Example:



  Exercise. In the above example, show that $\mathbb{P}(T_n < 4 \text{ for every } n) \geqslant 1/2$.

- A brief introduction to **martingales**
- Interpretation of probabilities in terms of betting: You bet on an event E. The ticket price is α, but you win \$1 if E occurs.

  Q What value of α makes this game fair?
- Reasoning about probabilities in terms of betting
- **Example** (Drunkard's walk).



  Q What is the probability that the drunkard reaches the pub before getting home?

  Case 1: $p = 1/2$

  At every step, you bet on whether the drunkard goes forward or backward:

  - If he goes forward, you win \$1.
  - If he goes backward, you lose \$1.

  The game is fair because

$$\text{expected payoff per step} = \frac{1}{2} \times 1 + \frac{1}{2} \times (-1) = 0\,.$$

  Let

$$W_n := \text{your net wealth after } n \text{ steps}$$

  Fairness translates to

$$\mathbb{E}\big[W_{n+1} \mid \underbrace{X_0, X_1, \ldots, X_n}_{\text{whatever is known up until the } n\text{th step}}\big] = W_n\,.$$

  In particular, $\mathbb{E}[W_n \mid X_0 = k] = W_0$.

Let $T \coloneqq T_{\{0,m\}}$ the hitting time of $\{0, m\}$. *Presumably*, fairness also implies

$$\mathbb{E}[W_T \mid X_0 = k] = W_0 \ .$$

But

$$\begin{aligned}\mathbb{E}[W_T \mid X_0 = k] &= \mathbb{P}(T_0 < T_m \mid X_0 = k) \cdot (W_0 - k) + \mathbb{P}(T_m < T_0 \mid X_0 = k) \cdot (W_0 + n - k) \\ &= W_0 - k + n \cdot \mathbb{P}(T_m < T_0 \mid X_0 = k) \ ,\end{aligned}$$

which gives

$$W_0 - k + n \cdot \mathbb{P}(T_m < T_0 \mid X_0 = k) = W_0 \ .$$

It follows that

$$\mathbb{P}(T_m < T_0 \mid X_0 = k) = \frac{k}{n} \ .$$

Recall that, earlier you found this probability by solving a recursion (namely, the harmonicity equation).

Q Can we do something similar when $p \neq 1/2$?

General case $0 < p < 1$

We need to come up with a fair game similar to the one in the symmetric case.

Attempt 1

Let

$$W_n \coloneqq \begin{cases} W_{n-1} + q & \text{if } X_n = X_{n-1} + 1, \\ W_{n-1} - p & \text{if } X_n = X_{n-1} - 1. \end{cases}$$

In other words, at every step,

- If the drunkard goes forward, you win q dollars.
- If the drunkard goes backward, you lose p dollars.

Following the same reasoning is slightly more complicated because W_T depends not only on X_T but also on the number of forward and backward moves. In the end, we do not get an answer to the original question (i.e., the value of $\mathbb{P}(T_m < T_0 \mid X_0 = k)$) but find the following interesting identity:

$$(2p - 1)\, \mathbb{E}[T_{\{0,m\}} \mid X_0 = k] = m\, \mathbb{P}(T_m < T_0 \mid X_0 = k) - k \ .$$

▶ The sequence $W_0, W_1, \ldots$ used above is an example of what we call a martingale.

▶ A real-valued process $(W_n)_{n\geqslant 0}$ is a **martingale** relative to another process $(X_n)_{n\geqslant 0}$ if

(i) W_n is uniquely determined by $(X_0, X_1, \ldots, X_n)$.
Common terminology: We say that $(W_n)_{n\geqslant 0}$ is **adapted** to $(X_n)_{n\geqslant 0}$.

(ii) $\mathbb{E}[W_n]$ exists and is finite.
In other words, $\mathbb{E}[|W_n|] < \infty$.

(iii) $\mathbb{E}[W_{n+1} \mid X_0, X_1, \ldots, X_n] = W_n$.
This is the **fairness** condition.

▶ Interpretation: A martingale describes the evolution of a fair game:

- X_n is the new information learned at time n.
- W_n is your net wealth at time n.
- $\delta W_n \coloneqq W_n - W_{n-1}$ is your net gain in the nth step.

▶ Changing the fairness condition, we get the following more general concepts:

- **sub-martingale**: $\mathbb{E}[W_{n+1} \mid X_0, X_1, \ldots, X_n] \geqslant W_n$. (The game can be favourable to you.)
- **super-martingale**: $\mathbb{E}[W_{n+1} \mid X_0, X_1, \ldots, X_n] \leqslant W_n$. (The game can be unfavourable to you.)

▶ **Examples**.

(1) (Additive martingale)
Let $Z_1, Z_2, \ldots$ be i.i.d. RVs with values in S, and let $f\colon S \to \mathbb{R}$ be such that $\mathbb{E}[f(Z_i)] = 0$. Then,

$$W_n \coloneqq f(Z_1) + f(Z_2) + \cdots + f(Z_n)$$

defines a martingale relative to $(Z_n)_{n\geqslant 1}$.
The net gains $\delta w_n = f(Z_n)$ are i.i.d. In the example of the Drunkard's walk, $Z_i = \pm 1$ and f is the identity.

(2) (Multiplicative martingale)
Let $Z_1, Z_2, \ldots$ be i.i.d. RVs with values in S, and let $g\colon S \to \mathbb{R}$ be such that $\mathbb{E}[g(Z_i)] = 1$. Then,

$$W_n \coloneqq g(Z_1)g(Z_2)\cdots g(Z_n)$$

defines a martingale relative to $(Z_n)_{n\geqslant 1}$.
This describes a scenario in which the stake at each step is proportional to the current wealth; namely, $\delta W_{n+1} = (g(Z_{n+1}) - 1) \cdot W_n$.

(3) (Doob/Levy martingale)
Let $X_0, X_1, \ldots$ be an arbitrary process and Y a real-valued RV with finite expectation. Then,

$$W_n \coloneqq \mathbb{E}[Y \mid X_0, X_1, \ldots, X_n]$$

defined a martingale relative to $(X_n)_{n\geqslant 0}$.
Interpretation: We are learning more and more about Y.

- W_n is the best estimate for Y at time n.

(4) (Relation to harmonic functions)
Let $X_0, X_1, \ldots$ be a MC with state space S and let $h\colon S \to \mathbb{R}$. Then,

$$W_n \coloneqq h(X_n)$$

defines a martingale if and only if h is harmonic.

▶ Back to the Drunkard's walk.

Case 2: $p \neq {}^1\!/_2$

Attempt 2

Let us try the multiplicative martingale. Let $q \coloneqq 1 - p$. Let $Z_1, Z_2, \ldots$ be the sequence of moves. Namely,

$$Z_i \coloneqq \begin{cases} 1 & \text{with probability } p \text{ (forward move)}, \\ -1 & \text{with probability } q \text{ (backward move)}. \end{cases}$$

So, $X_n = X_0 + Z_1 + Z_2 + \cdots Z_n$. As before, we let $T \coloneqq T_{\{0,m\}}$ be the first time the drunkard reaches either home or the pub.

Consider a function of the form

$$g\colon \begin{array}{r} 1 \longmapsto \alpha > 0 \\ -1 \longmapsto \beta > 0 \end{array}$$

and define $W_0 \coloneqq 1$ and recursively $W_n \coloneqq W_{n-1} \cdot g(Z_n)$. We want $(W_n)_{n\geqslant 0}$ to be a martingale.

The fairness condition requires $\mathbb{E}[g(Z_n)] = 1$, which means $p\alpha + q\beta = 1$. There are infinitely many solutions but the following two

$$\alpha = \beta = 1 \qquad \text{and} \qquad \alpha = {}^q\!/_p,\ \beta = {}^p\!/_q$$

have the following useful property: they ensure that W_T depends only on X_T and not on the entire trajectory up until time T (compare with the first attempt). The first solution is trivial. Let us proceed with the second one.

As before, fairness *presumably* implies

$$\mathbb{E}[W_T \mid X_0 = k] = W_0 = 1 \,.$$

But

$$\mathbb{E}[W_T \mid X_0 = k] = \mathbb{P}(T_0 < T_m \mid X_0 = k) \cdot ({}^p\!/_q)^k + \mathbb{P}(T_m < T_0 \mid X_0 = k) \cdot ({}^q\!/_p)^{m-k} \,.$$

We obtain that

$$\mathbb{P}(T_m < T_0 \mid X_0 = k) = \frac{1 - (q/p)^k}{1 - (q/p)^m} \,,$$

which is consistent with we earlier found by solving the "harmonicity" recursion.

Combined with the result of the first attempt, as a bonus we also find

$$\mathbb{E}[T_{\{0,m\}} \mid X_0 = k] = \frac{1}{2p-1}\left[m \cdot \left(\frac{1 - (q/p)^k}{1 - (q/p)^m}\right) - k\right] .$$

▶ Let $(W_n)_{n \geqslant 0}$ be a martingale relative to $(X_n)_{n \geqslant 0}$, and let T be a stopping time for $(X_n)_{n \geqslant 0}$.

Q Does the fairness hypothesis really imply $\mathbb{E}[W_T \mid X_0] = W_0$?

Note that

$$\mathbb{E}[W_{n+1} \mid X_0] = \mathbb{E}\big[\,\mathbb{E}[W_{n+1} \mid X_0, X_1, \dots, X_n] \,\big|\, X_0\big] = \mathbb{E}[W_n \mid X_0] \,,$$

from which it follows $\mathbb{E}[W_n \mid X_0] = W_0$ for every n.

▶ **Counter-examples**.

(1) (Bet-doubling strategy). You play a game with a fair coin. Before each flip, you can make a bet c dollars on *heads*:

- If the coin comes up heads, you win c dollars.
- If the coin comes up tails, you lose c dollars.

You can change your bet from flip to flip, and you can stop playing whenever you wish.

Consider the following strategy: You start by betting \$1 on the first flip. If the coin comes up heads, you stop playing; if it comes up tails, you bet \$2 on the second flip. You continue playing as long the coin comes up tails, each time doubling your bet, and stop as soon the coin comes up heads. With this strategy, your total payoff once you stop the game is \$1, irrespective of the number of rounds! Namely, if the first heads appears on the nth flip,

$$\text{total payoff} = -1 - 2 - 2^2 - \cdots - 2^{n-2} + 2^{n-1} = 1$$

Your wealth at time n can be described as $W_n = W_{n-1} + c_n \cdot Z_n$, where $Z_1, Z_2, \dots$ are i.i.d. with $\mathbb{P}(Z_i = 1) = \mathbb{P}(Z_i = -1) = 1/2$ and c_n is the amount you bet on the nth flip. Let T denote the time you stop playing, which is a stopping time (unless you have premonitions).

Clearly, $(W_n)_{n \geqslant 0}$ is a martingale relative to $(Z_n)_{n \geqslant 1}$, hence the fairness condition suggests that we should have $\mathbb{E}[W_T] = W_0$. However, with the above strategy, namely choosing

$$c_n := 2^{n-1} \qquad \text{and} \qquad T := \inf\{n : Z_n = 1\}$$

we have $W_T = W_0 + 1$ almost surely.

(2) (Null recurrent RW on $\mathbb{Z}$). Let $Z_1, Z_2, \dots$ be i.i.d. with $\mathbb{P}(Z_i = 1) = \mathbb{P}(Z_i = -1) = 1/2$ and $X_n := Z_1 + Z_2 + \cdots + Z_n$, so that $(X_n)_{n \geqslant 0}$ is a symmetric RW on $\mathbb{Z}$ starting at the origin. Then, $W_n := X_n$ is a martingale, but for the stopping time $T := T_1$ (the hitting time of 1), we almost surely have $W_T = 1 \neq 0 = W_0$.

Conclusion: In general, the fairness condition of a martingale $(W_n)_{n \geqslant 0}$ does not imply $\mathbb{E}[W_T \mid X_0] = W_0$ when T is a stopping time.

Q Is our hand-wavy computation in the drunkard example doomed?

▶ **Optional stopping theorem** (Doob). Let $(W_n)_{n \geqslant 0}$ be a martingale relative to $(X_n)_{n \geqslant 0}$, and let T be a stopping time for $(X_n)_{n \geqslant 0}$. Then,

$$\mathbb{E}[W_T \mid X_0] = W_0$$

provide that one of the following holds:

(v1) T is bounded.

(v2) T is almost surely finite and W is bounded.

(v3) $\mathbb{E}[T] < \infty$ and $(W_n)_{n \geqslant 0}$ has **bounded increments**.

Explanation:

- T being bounded means that there exists a $t_{\max} < \infty$ such that $\mathbb{P}(T \leqslant t_{\max}) = 1$.
- $(W_n)_{n \geqslant 0}$ having bounded increments means there exists a $c \in \mathbb{R}$ such that $|W_n - W_{n-1}| \leqslant c$ almost surely for every $n \geqslant 1$.

▶ *Note*: In the above counter-examples, the hypotheses of none of the versions of the optional stopping theorem are satisfied. On the other hand, versions 2 and 3 are applicable to the drunkard's walk example (all three martingales)