American University of Beirut

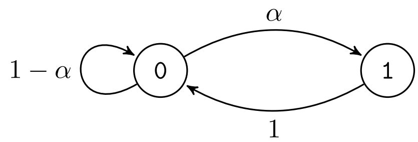
MATH/STAT 238: Applied Probability Models

2025–2026 Fall

Suggested exercises (last update: November 15, 2025)

added on: 2025-08-30

Problem 1 (A simple Markov chain). Consider a parametric version of the Markov chain discussed in class:



Namely, the system has two possible states `0` and `1`. The state of the system changes at discrete time steps according to random choices that depend only on the current state:

- From state `0`, the system jumps into state `1` with probability α and stays in state `0` with probability $1-\alpha$.
- From state `1`, the system jumps into state `0` with probability 1.

Here, α is a parameter with a value strictly between 0 and 1. Let $X_0, X_1, X_2, \ldots$ denote the sequence of the states of the Markov chain at consecutive time steps.

(a) Show that the Markov chain has exactly one equilibrium distribution and compute the probabilities it assigns to each state.

(b) Write recursive equations describing the distribution of X_{k+1} in terms of the distribution of X_n.

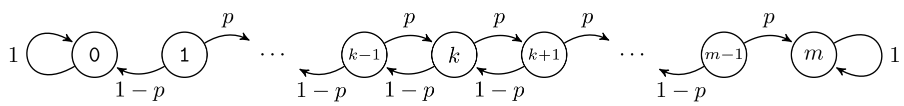
(c) Write the result of the previous part in matrix form, that is, in the form $p_{k+1} = p_k A$, where p_k and p_{k+1} are row vectors indicating the distributions at times k and $k+1$ respectively, and A is a 2-by-2 matrix.

(d) Find an explicit expression for the distribution of X_n as a function of the distribution of X_0.

Hint: One approach, as suggested in class, is to diagonalize A. You may want to try the special case $\alpha = 1/2$ first. Another approach is to use substitution to turn the system of recursions into a single recursive equation of order 2 for $p_k(\texttt{0})$, and then solve the equation using generating functions. You can use WolframAlpha or any similar computer algebra system to verify your answer.

(e) Use the result of the previous part to show that, irrespective of its initial distribution p_0, the distribution of the system converges to its unique equilibrium distribution; that is, $p_n \to p$ as $n \to \infty$, where p the distribution you find in part (a).

Problem 2 (Drunkard's walk). Recall the example of a drunkard walking on a street with $m+1$ blocks.



Assume that $0 < p < 1$. For $n = 0, 1, 2, \ldots$, let X_n denote the position of the drunkard at time n.

(a) Show that every equilibrium distribution for this chain is of the form $(1-r, 0, \ldots, 0, r)$ for some $r \in [0, 1]$.

For $k = 0, 1, 2, \ldots, m$, let $q(k) := \mathbb{P}(\text{reaching } m \text{ before } 0 \mid X_0 = k)$ be the probability of the drunkard ending up in the pub before reaching home if he starts at position k. During the lecture, we argued that

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

(b) Re-derive the above recurrence relation. Make sure to justify each step as rigorously as possible.

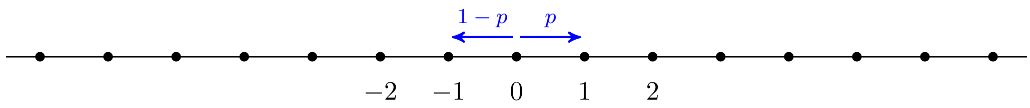
(c) Solve the above recurrence to find $q(k)$ for each k.

Hint: Define $\Delta(k) := q(k+1) - q(k)$. Write the recursion governing $\Delta(k)$ and identify its extreme values. Consider the cases $p = 1/2$ and $p \neq 1/2$ separately. As mentioned in class, the end result will be:

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

(d) Use Desmos or a similar software to plot $q(k)$ as a function of p, with k and m as parameters. (With Desmos, you can add a slider for each of the parameters k and m.) Vary the values of k and m to see how the graph of $p \mapsto q(k)$ changes. Interpret your observation in the case $m = 100$ and $k = 20$.

Problem 3 (Random walk on $\mathbb{Z}$)**.** Consider a particle performing a random walk on $\mathbb{Z}$. At every step, the particle moves either 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 $0 < p < 1$.



(a) Show that, irrespective of the value of p, the model has no equilibrium distribution.

(b) Use the result on the drunkard's walk to prove that when $p = 1/2$, the random walk is recurrent, that is, $\mathbb{P}\big(\text{the particle returns to } 0 \,\big|\, X_0 = 0\big) = 1$.

(c) Use the result on the drunkard's walk to prove that when $p \neq 1/2$, the random walk is transient, that is, $\mathbb{P}\big(\text{the particle returns to } 0 \,\big|\, X_0 = 0\big) < 1$.

(d) Show that when $p = 1/2$, almost surely the particle returns to its initial position infinitely many times.
Hint: Use the result of part (b). Every time the particle returns to its initial position, it starts afresh.

(e) Show that when $p \neq 1/2$, almost surely the particle returns to its initial position only finitely many times.
Hint: Use the result of part (c) and the hint for the previous part.

added on: 2025-09-10

Problem 4 (Galton-Watson process)**.** Consider a population of bacteria that evolves according to a discrete-time Galton-Watson process. 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. In other words, the number of offsprings of bacterium i at time step n is a random variable $Z_{n,i}$ with two possible values with $\mathbb{P}(Z_{n,i} = 2) = p$ and $\mathbb{P}(Z_{n,i} = 0) = 1 - p$.

Let $q(p) := \mathbb{P}(\text{extinction} \mid X_0 = 1)$. In class, it was claimed (without proof) that

1) $q(p) = 1$ when $p \leqslant 1/2$ and $q(p) < 1$ when $p > 1/2$.

2) $q(p)$ is a fixed point of the function $\varphi(s) := \mathbb{E}[s^{Z_{n,i}}] = 1 - p + ps^2$.

Assuming these,

(a) Find $q(p)$ for all values of $p \in [0, 1]$.

(b) Plot the graph of $1 - q(p) = \mathbb{P}(\text{survival} \mid X_0 = 1)$. You can use Desmos or any other computer software.

(c) Is $1 - q(p)$ a continuous function? Verify your answer mathematically.

(d) Is $1 - q(p)$ non-decreasing? Verify your answer mathematically.

(e) Where is $1 - q(p)$ differentiable? Justify your answer mathematically.

(f) For $p > 1/2$, is the function $1 - q(p)$ convex up, or down, or neither? Verify your answer mathematically.

Problem 5 (Exponential distribution; Review)**.** Let T and S be independent exponential random variables with rates λ and μ respectively.

(a) What is $\mathbb{P}(T > t)$ as a function of $t \in \mathbb{R}$?

(b) (memorylessness) Verify that for every $x, t > 0$, $\mathbb{P}(T > t + x \mid T > x) = \mathrm{e}^{-\lambda t}$. In other words, given $T > x$, the random variable $T - x$ is again exponentially distributed with the same rate.

(c) (minimum) Show that $W := \min(T, S)$ is again an exponential random variable and find its rate.

(d) Show that $\mathbb{P}(T < S) = \frac{\lambda}{\lambda+\mu}$.

added on: 2025-09-14 **Problem 6** (Implications of the Markov property). Let $(X_n)_{n=0}^{\infty}$ 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

$$(X_m, X_{m+1}, \ldots, X_{n-1}) \qquad \text{and} \qquad (X_{n+1}, X_{n+2}, \ldots, X_\ell)$$

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.

Problem 7 (Transition kernels). A *transition kernel* (a.k.a., *Markov kernel*, *stochastic kernel*, or *stochastic matrix*) from a countable set A to a countable set B is a map $K : A \times B \to [0,1]$ such that $\sum_{b \in B} K(a,b) = 1$ for every $a \in A$. Given an A-valued random variable X and a B-valued random variable Y, we write $X \xrightarrow{K} Y$ if $\mathbb{P}(Y = b \mid X = a) = K(a,b)$ for every $a \in A$ and $b \in B$.

(a) Show that if $X \xrightarrow{K} Y$ and X is distributed according to $p : A \to [0,1]$, then Y is distributed according to pK, where $(pK)(b) := \sum_{a \in A} p(a)K(a,b)$ for each $b \in B$.

(b) Show that if $X \xrightarrow{K} Y$ and $f : B \to \mathbb{R}$, then $\mathbb{E}[f(Y) \mid X = \cdot\,] = Kf$ where $Kf : A \to \mathbb{R}$ is defined by $(Kf)(a) := \sum_{b \in B} K(a,b)f(b)$ for each $a \in A$ provided that the expectation exists or the sum converges.

(*Note:* The existence of the expected value is guaranteed for instance if B is finite or if f is non-negative, in which case Kf may take value $+\infty$.)

Given two transition kernels $P : A \times B \to [0,1]$ and $Q : B \times C \to [0,1]$, we write $X \xrightarrow{P} Y \xrightarrow{Q} Z$ if $X \xrightarrow{P} Y$ and $Y \xrightarrow{Q} Z$, and in addition, the random variables X, Y, Z form a Markov chain (i.e., given Y, the random variables X and Z are independent).

(c) Show that if $X \xrightarrow{P} Y \xrightarrow{Q} Z$, then $X \xrightarrow{PQ} Z$, where $PQ : A \times C \to [0,1]$ is given by $(PQ)(a,c) := \sum_{b \in B} P(a,b)Q(b,c)$ for every $a \in A$ and $c \in C$. Explain why this may not be true if we only have $X \xrightarrow{P} Y$ and $Y \xrightarrow{Q} Z$.

Problem 8 (Set of stationary distributions). Let $K : S \times S \to [0,1]$ be the transition matrix of a Markov chain chain with countable state space S.

(a) Verify that any convex combination of two stationary distributions for K is again a stationary distribution for K.

(*Recall:* A *convex combination* of two functions π_1 and π_2 is a linear combination of the form $\lambda\pi_1 + (1-\lambda)\pi_2$ where $0 \leqslant \lambda \leqslant 1$. More generally, a convex combination of $\pi_1, \pi_2, \ldots, \pi_m$ is a linear combination of the form $\lambda_1\pi_1 + \lambda_2\pi_2 + \cdots + \lambda_m\pi_m$ where $0 \leqslant \lambda_i \leqslant 1$ for each i and $\sum_{i=1}^m \lambda_i = 1$.)

(b) Conclude that every Markov chain has either no, one, or uncountably many stationary distributions.

From part (a), it follows that the set of all stationary distributions for K is *convex* (i.e., closed under convex combinations). A stationary distribution π for K is said to be *extremal* if it cannot be written as a non-trivial convex combination of other stationary distributions for K, that is, if $\pi = \lambda\pi_1 + (1-\lambda)\pi_2$ for some stationary distributions π_1, π_2 and $0 < \lambda < 1$, then either $\pi_1 = \pi_2 = \pi$.

(c) (optional) Show that the extremal stationary distributions for K are linearly independent.

(d) (optional) Show that if S is finite, then K has at most $|S|$ extremal stationary distributions.

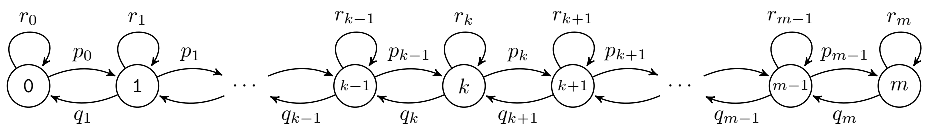
(*Remark:* In fact, the number of extremal stationary distributions for K is bounded by the number of irreducible components of K.)

(e) (optional; difficult) Show that if S is finite, then every stationary distribution for K can be written as a convex combination of the stationary ones in a unique fashion.

Hint: The existence follows from the Krein-Milman theorem. The uniqueness is a consequence of the linear independence of the extremal stationary distributions.

added on: 2024-09-28

Problem 9 (A birth-and-death process)**.** Consider a discrete-time Markov chain with state space $\{0, 1, \ldots, m\}$ and the following transition graph.



We assume that $p_k, q_k > 0$ for every $0 < k < m$ and $p_0, q_m > 0$.

(a) Find the unique stationary distribution of this Markov chain.

(b) Simplify your answer in the case where $r_k = 0$ and $p_k = q_k = 1/2$ for $0 < k < m$.

(c) Interpret your answer to the previous part in the case where p_0 and q_0 are very small by comparing this Markov chain with the Drunkard's walk.

Problem 10 (A topological proof of the existence theorem)**.** Let $K\colon S \times S \to [0, 1]$ be the transition matrix of a Markov chain $(X_n)_{n=0}^{\infty}$ with finite state space S. This exercise sketches an alternative proof of the fact that K has at least one stationary distribution, that is, a probability distribution $\pi\colon S \to [0, 1]$ such that $\pi K = \pi$.

Let $\mu\colon S \to [0, 1]$ be an arbitrary probability distribution on S. For $n \geqslant 0$, set

$$\overline{\mu}_n := \frac{1}{n}\left(\mu + \mu K + \mu K^2 + \cdots + \mu K^{n-1}\right) .$$

This is the uniform mixture of the distributions of $X_0, X_1, \ldots, X_{n-1}$ if $X_0 \sim \mu$.

(a) Verify that $\overline{\mu}_n$ is again a probability distribution on S.

There is no reason for $\overline{\mu}_n$ to be stationary, but when n is large, $\overline{\mu}_n$ is "close to being stationary". The "closeness" of two probability distributions $\nu_1, \nu_2\colon S \to [0, 1]$ can be measured by

$$\|\nu_2 - \nu_1\| := \sum_{a \in S} |\nu_2(a) - \nu_1(a)| .$$

(b) Show that, for every $n \geqslant 0$,

$$\|\overline{\mu}_n K - \overline{\mu}_n\| \leqslant \frac{2}{n} ,$$

which goes to 0 as $n \to \infty$.

Hint: Expand $\overline{\mu}_n K - \overline{\mu}_n$ and note that most of the terms cancel one another.

The sequence $\overline{\mu}_1, \overline{\mu}_2, \ldots$ need not converge, but it has a convergent subsequence. Namely, let

$$\mathscr{P}(S) := \{\nu\colon S \to [0, 1] : \sum_{a \in S} \nu(a) = 1\}$$

be the set of all probability distributions on S.

(c) (optional) Prove that $\mathscr{P}(S)$ is compact.

Hint: Show that, as a subset of $\mathbb{R}^S$, $\mathscr{P}(S)$ is closed and bounded, and apply the Heine-Borel theorem.

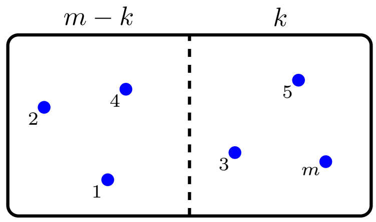
(d) (optional) Use the result of the previous part to show that there exists $n_1 < n_2 < \cdots$ such that $\overline{\mu}_{n_i}$ converges to some $\pi \in \mathscr{P}(S)$.

(e) Show that the distribution π obtained in the previous part is stationary for K.

(f) (optional) Explain why the above reasoning fails if S is countably infinite.

Problem 11 (Intermediate state)**.** Let K be the transition matrix of a Markov chain with state space S. Let $a, b \in S$ and $n_1, n_2 \in \mathbb{N}$. Show that $K^{n_1+n_2}(a, b) > 0$ if and only if there exists a state $c \in S$ such that $K^{n_1}(a, c) > 0$ and $K^{n_2}(c, b) > 0$.

Problem 12 (Two formulations of the Ehrenfest model)**.** Recall that in the Ehrenfest model, m identical particles are distributed between two boxes that sit next to each other, with the right box having n particles and the left one having $m-k$ particles.



Once we remove the barrier between the two boxes, the particles start jumping from one box into another. In the discrete-time version of the model, at each time step, one of the m particles is chosen uniformly at random and is moved to the opposite box. This model was proposed by Tatiana and Paul Ehrenfest as a toy model to give a microscopexplain the second law of thermodynamics at the microscopic level.

As we discussed in class, this model can be mathematically formulated in two different ways, depending on how much information we keep track of.

Formulation 1 (less detailed): The state of the system at time t is the number of particles in the right box at time t.

(a) Specify the state space and the transition probabilities of this model.

(b) Show that the Markov chain has a reversible stationary distribution. Find an explicit expression for this stationary distribution.

Hint: The answer is a binomial distribution. Specify its parameters.

(c) Verify that the Markov chain is irreducible, hence the reversible distribution obtained in the previous part is its unique stationary distribution.

(d) Is the Markov chain aperiodic?

Formulation 2 (more detailed): The state of the system at a given time is a sequence $(a_1, a_2, \dots, a_m)$ where

$$a_i := \begin{cases} 1 & \text{the } i\text{-th particle is in the right box,} \\ 0 & \text{otherwise.} \end{cases}$$

for instance, in the above illustration, the state of the system is $(0, 0, 1, 0, 1, \dots, 1)$.

(e) Specify the state space and the transition probabilities of this model.

(f) Show that the Markov chain has a reversible stationary distribution. Find an explicit expression for this stationary distribution.

(g) Verify that the Markov chain is irreducible, hence the reversible distribution obtained in the previous part is its unique stationary distribution.

(h) Is the Markov chain aperiodic?

(i) This Markov chain can be thought of as a random walk on a graph called the *hypercube* of dimension m. Identify the vertex set and the edge set of this graph.

Relationship between the two formulations: Let $(X_n)_{n=0}^{\infty}$ and $(Y_n)_{n=0}$ be the Markov chains according to the first and second formulations, respectively. Let π and ν be the corresponding stationary distributions.

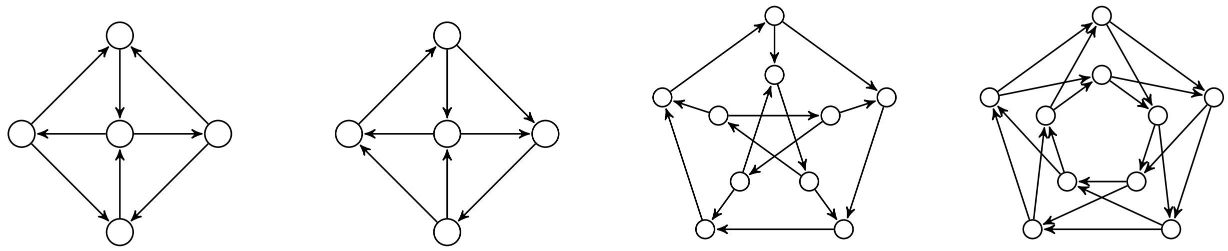
(j) Show that $X_n = f(Y_n)$ for some function f.

(k) Write the distribution of X_n in terms of the distribution of Y_n (i.e., write $\mathbb{P}(X_n = k)$ in terms of the probabilities of the form $\mathbb{P}(Y_n = a)$). Is the distribution of Y_n uniquely determined by the distribution of X_n?

(l) Use the previous part to derive an identity connecting π to ν. Verify that the answers to parts (b) and (f) satisfy this identity.

Problem 13 (Irreversible Markov chains)**.** Give an example of an irreducible Markov chains whose unique stationary distribution does *not* satisfy the detailed balance condition. Is there such an example with only two states?

Problem 14 (Irreducibility, aperiodicity, mixing)**.** The following are the transition graphs of three finite-state Markov chains. (*Recall:* The edges indicate transitions with positive probability.)



(i) (ii) (iii) (iv)

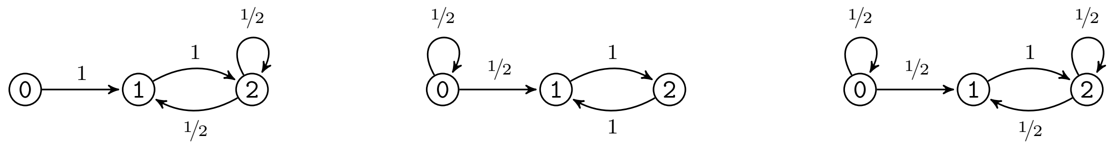
In each case,

(a) Verify that the Markov chain is irreducible.

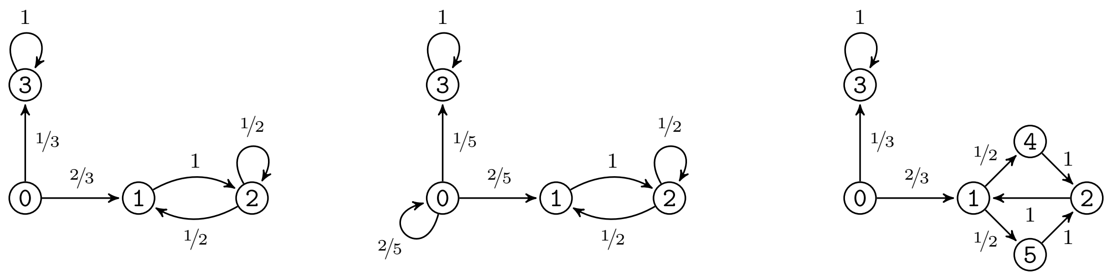
(b) Determine whether the Markov chain is periodic or aperiodic. If periodic, identify the largest period and a partitioning of the state space that witnesses the periodicity. If aperiodic, prove this by showing that the gcd of the cycle lengths is 1.

(c) Decide whether the Markov chain is mixing or not.

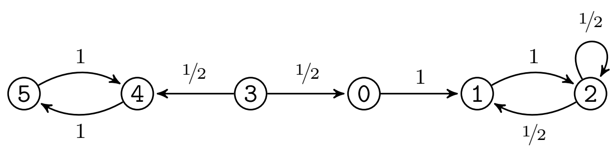
Problem 15 (Convergence)**.** Consider the Markov chains with the following transition graphs.



(i) (ii) (iii)



(iv) (v) (vi)



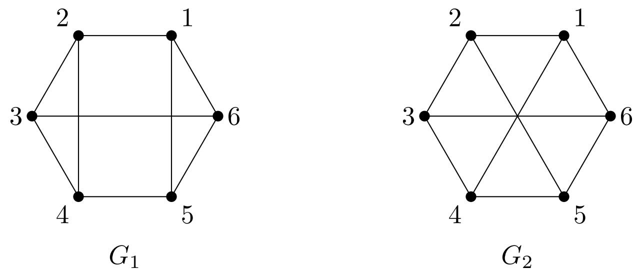
(vii)

In each case,

(a) Identify all the stationary distributions of the Markov chain.

(b) Determine if, starting from state 0, the distribution of the Markov chain converges to a stationary distribution. If so, identify the limiting distribution.

Problem 16 (Asymptotic behaviour; computer simulation; optional)**.** Consider a simple symmetric random walk on each of the following graphs.



(a) Use computer simulation to sample three realizations of the walk on G_1 from time 0 to 20 starting from vertex 1, and plot the results as a function of time. Repeat this for the walk on G_2.

(b) Now generate 200 independent realizations of the walk on G_1 from time 0 to 50 starting from vertex 1. For each time step $0 \leqslant t \leqslant 50$, calculate the empirical distribution of the position of these 200 walks, and plot this empirical distribution as a function of time. Repeat this for the walk on G_2. How do you explain the qualitative difference between the results of the experiments on G_1 and G_2?

(c) Generate a single realization of the walk on G_1 from time 0 to $N = 700$, starting from vertex 1. For each time $0 \leqslant t \leqslant N$, calculate the density of the time steps $0 \leqslant s \leqslant t$ in which the walk has been in position 1 (or $2, 3, \ldots, 6$). Plot this density as a function of time t. Does this density seem to converge to a constant? If necessary, vary the length of the experiment N to see the trend. Repeat the above for the walk on G_2.

(d) Pick your favorite function $f\colon \{1, 2, \ldots, 6\} \to \mathbb{R}$. As before, generate a single realization of the walk on either G_1 or G_2 from time 0 to $N = 700$, starting from vertex 1. For each time $0 \leqslant t \leqslant N$, calculate the average value of f over the position of the walk in times 0 up to t. Plot this average value as a function of time t. Does this average value seem to converge to a constant? Vary N to see the trend.

(e) Can you explain the observation in part (d) using the observation in part (c)?

Problem 17 (Alternative proof of the uniqueness theorem)**.** This problem sketches an alternative proof of the uniqueness theorem:

> **Theorem.** Every irreducible finite-state Markov chain has a unique stationary distribution.

Let $K\colon S \times S \to [0, 1]$ be the transition matrix of a Markov chain with finite state space S. In class, we proved the following lemmas:

- **Lemma 1.** If u is a left eigenvector of K corresponding to the eigenvalue 1, then so is $|u|$. (This was used in the proof of the existence theorem presented in class.)
- **Lemma 2.** If K is irreducible, then every stationary distribution for K is strictly positive. (This was used in the proof of the uniqueness theorem presented in class.)

(a) (optional) Reproduce the proof of Lemma 1.

(b) (optional) Reproduce the proof of Lemma 2.

Now, suppose that K is irreducible.

(c) Let u be a left eigenvector of K corresponding to the eigenvalue 1. Show that the elements of u are either all positive or all negative. *Hint:* Consider $u + |u|$.

(d) Show that K has at most one stationary distribution.
Hint: Let π and π' be two stationary distributions and consider $u := \pi' - \pi$.

Problem 18 (Periods and aperiodicity)**.** This exercise further explores the notion of periods and aperiodicity in Markov chains.

Let $(X_n)_{n=0}^{\infty}$ be a Markov chain with finite or countably infinite) state space S and (not necessarily irreducible) transition matrix K. The *period* of a state $a \in S$, denoted by $d(a)$, is the greatest common divisor of all positive integers ℓ such that $K^\ell(a, a) > 0$. Equivalently, $d(a)$ is the $\gcd$ of the lengths of all closed walks in the transition graph of K that pass through a. (We can use the convention $\gcd(\varnothing) = 0$ so that $d(a) = 0$ if there is no cycle passing through a.) A state is said to be *aperiodic* if it has period 1.

(a) Identify the periods of all the states in (some of) the Markov chains in Problems 14 and 15.

(b) Show that if $a, b \in S$ are strongly connected (i.e., there is a path from a to b and a path from b to a), then $d(a) = d(b)$.

Now, suppose that K is irreducible. By the last part, all states in S have the same period, which we denote by d.

(c) Show that d is the largest period of K, in the sense defined in class. More specifically, show that d is the largest integer such that S can be partitioned into d sets $S_0, S_1, \ldots, S_{d-1}$ with $\mathbb{P}(X_{n+1} \in S_{i+1} \,|\, X_n = a) = 1$ for all $i = 0, 1, \ldots, d-1$ and $a \in S_i$, and with the convention $S_d := S_0$.

(d) Conclude that K is aperiodic if and only if there exists a state $a \in S$ with $d(a) = 1$.

added on: 2025-10-14

Problem 19 (Random walks on groups; optional)**.** A number of important examples of Markov chains can be viewed as random walks on groups.

Let $\mathbb{G}$ be a finite or countably infinite group and $p : \mathbb{G} \to [0,1]$ a probability distributions. A *random walk* on $\mathbb{G}$ with *jump distribution* p is a Markov chain $(X_n)_{n=0}^{\infty}$ with state space S in which $X_n = Z_n * X_{n-1}$, where $Z_1, Z_2, Z_3, \ldots$ are i.i.d. random variables with distribution p. In other words, at every time step, the current state is multiplied from the left by a new random sample from $\mathbb{G}$ with distribution p.

As an example, the ordinary random walk on $\mathbb{Z}$ is a random walk on the additive group of the integers.

(a) Show that the Ehrenfest model (Problem 12, Formulation 2) can be viewed as a random walk on a group. Identify the underlying group and the jump distribution.

(b) Argue that every "iterative" card shuffling method is a random walk on the group S_{52} of all the permutations of the 52 cards. Identify the jump distribution for the top-to-random shuffling method.

(Recall that at every step of the *top-to-random* shuffling method, the top card is placed at one of the 52 possible positions relative to the other cards chosen uniformly at random.)

(c) Identify the transition probabilities of a random walk on a group $\mathbb{G}$ with jump distribution p.

(d) Verify that a random walk on a group $\mathbb{G}$ with jump distribution p is irreducible if and only if the elements $a \in \mathbb{G}$ with $p(a) > 0$ generate $\mathbb{G}$.

(e) Show that the uniform distribution on a finite group $\mathbb{G}$ is stationary for every random walk on $\mathbb{G}$.

(f) Argue that the top-to-random card shuffling method is mixing.

Problem 20 (Variance via an independent coupling)**.** Verify that the variance of every real-valued random variable X can be written as $\mathbb{V}\mathrm{ar}[X] = \frac{1}{2}\,\mathbb{E}[(X - X')^2]$, where X' is an independent copy of X.

Problem 21 (Total variation distance; coupling inequality)**.** Let S be a countable set. The *total variation distance* between two probability distributions $\mu, \nu : S \to [0,1]$ is defined as

$$d_{\mathsf{TV}}(\mu, \nu) := \sup_{A \subseteq S} \left|\mu(A) - \nu(A)\right| .$$

(The total variation distance between μ and ν is also denoted by $\|\mu - \nu\|_{\mathsf{TV}}$.)

(a) Compute the total variation distance between two Bernoulli distributions with parameters p and q.

(b) Show that $d_{\mathsf{TV}}(\mu, \nu) = \frac{1}{2}\sum_{a \in S}\left|\mu(a) - \nu(a)\right|$.

A *coupling* of two probability distributions $\mu, \nu : S \to [0,1]$ is a pair of random variables (X, Y) defined in the same probability space such that $X \sim \mu$ and $Y \sim \nu$.

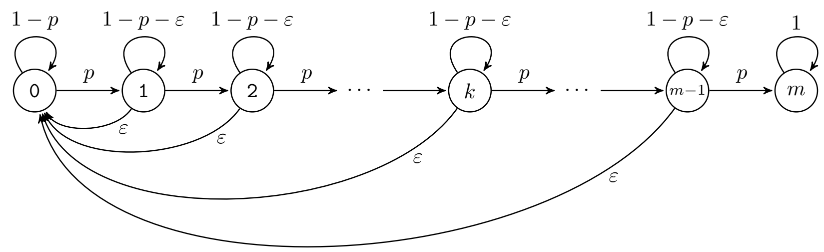
(c) (coupling inequality) Show that every coupling (X, Y) of two probability distributions μ and ν satisfies $d_{\mathsf{TV}}(\mu, \nu) \leqslant \mathbb{P}(X \neq Y)$.

Hint: Given $A \subseteq S$, compare $\mu(A) = \mathbb{P}(X \in A)$ with $\nu(A) = \mathbb{P}(Y \in A)$ by conditioning on whether $X = Y$ or $X \neq Y$.

(d) (optimal coupling; optional) Prove that every two probability distributions $\mu, \nu : S \to [0,1]$ have a coupling (X, Y) such that $d_{\mathsf{TV}}(\mu, \nu) = \mathbb{P}(X \neq Y)$.

(*Remark:* Such a coupling is called an *optimal* coupling.)

Problem 22 (Hitting times)**.** Consider a discrete-time Markov chain with state space $\{0, 1, \ldots, m\}$ and the following transition graph.



We may view m as the *target* state and 0 as the *reset* state. The states $1, 2, \ldots, m-1$ are the intermediate steps towards the target.

Let $(X_n)_{n \geqslant 0}$ denote a trajectory of the Markov chain.

(a) Find the probability that the Markov chain reaches m before 0 if the Markov chain starts from k.

(b) Find the expected hitting time of m if the Markov chain starts from 0.

(c) Give an intuitive explanation of your answers to the previous parts in the extreme case in which $\varepsilon = 0$?

Problem 23 (Uniqueness of harmonic functions). Let $(X_n)_{n=0}^{\infty}$ be a Markov chain with state space S and transition matrix $K\colon S \times S \to [0,1]$. Recall that a function $h\colon S \to \mathbb{R}$ is said to be *harmonic* (for K) at a state $x \in S$ if $h(x) = (Kh)(x)$. Suppose that S is finite and K is irreducible. Let $\varnothing \neq A \subsetneq S$.

(a) Prove that the only function that is harmonic on $S \backslash A$ and 0 on A is the constant 0.

(b) Show that for every boundary condition $h_0\colon A \to \mathbb{R}$, there exists one and only one function $h\colon S \to \mathbb{R}$ that is harmonic on $S \backslash A$ and satisfies $h(x) = h_0(x)$ for $x \in A$.

Hint: The set of functions that are harmonic at a given x is closed under linear combinations.

added on: 2025-10-29 **Problem 24** (Convergence of probability distributions). Let $\mu_1, \mu_2, \ldots, \mu$ be probability distributions on a countable set S. Show that the following conditions are equivalent:

(i) For every $a \in S$, we have $\mu_n(a) \to \mu(a)$ as $n \to \infty$.

(ii) We have $\|\mu_n - \mu\| \to 0$ as $n \to \infty$, where $\|\nu - \mu\| := \sum_{a \in S} |\nu(a) - \mu(a)|$.

Problem 25 (Notion of mixing). Let $K\colon S \times S \to [0,1]$ be the transition matrix of a Markov chain with (finite or countably infinite) state space S. Let π be a stationary distribution for this Markov chain. Verify that the following conditions are equivalent:

(i) For every $a, b \in S$, we have $K^n(a,b) \to \pi(b)$ as $n \to \infty$.

(ii) For every probability distribution $\mu : S \to [0,1]$, we have $\mu K^n \to \pi$ as $n \to \infty$.

(*Remark:* The convergence $\mu K^n \to \pi$ is defined using either of the equivalent conditions in Problem 24.)

Problem 26 (Lazy version of a Markov chain). Let $0 < \varepsilon < 1$. The *ε-lazy* version of a Markov chain with transition matrix $K : S \times S \to [0,1]$ is a Markov chain with transition matrix $(1-\varepsilon)K + \varepsilon I$, where I is the identity matrix on S. In other words, at every step, the lazy version flips a coin with parameter ε; if the coin comes up heads (with probability ε), the Markov chain remains in its current state; otherwise (with probability $1-\varepsilon$), it makes a random transition according to K.

(a) Verify that a Markov chain and its lazy versions all have the same stationary distributions.

(b) Verify that K is irreducible if and only if K_ε is.

(c) Verify that K_ε is aperiodic. Conclude that every lazy version of an irreducible finite-state Markov chain is mixing.

(d) Given two states $a, b \in S$, let $\eta_{a,b}$ denote the expected first hitting time of b starting from a in the original Markov chain. Show that the expected hitting time of b starting from a in the ε-lazy version is $\eta_{a,b}/(1-\varepsilon)$.

Hint: This is intuitively easy to see but writing down a rigorous argument for it is a good exercise. Let $(Y_n)_{n=0}^{\infty}$ be a Markov chain according to K_ε. Let $T_0 := 0$ and for $m > 0$, let T_m denote the time at which the coin comes up heads for the mth time. For $m \geqslant 0$, let $X_m := Y_{T_m}$. Show that $(X_m)_{m=0}^{\infty}$ is a Markov chain according to K that is independent of $T_0, T_1, T_2, \ldots$. Use conditioning to conclude.

Problem 27 (Transient states in finite and infinite Markov chains). Let $(X_n)_{n=0}^{\infty}$ be a Markov chain with state space $S = D \cup R$, where D denotes the set of transient states and R the set of recurrent states.

First, suppose that S is finite

(a) Show that, starting from any state $a \in S$, the first hitting time T_R of the set of recurrent states is almost surely finite.

(b) Conclude that, for every $a \in D$, $\mathbb{P}(X_n = a \mid X_0 = a) \to 0$ as $n \to \infty$.

Next, let us remove the assumption of the finiteness of S.

(c) Give an example of a Markov chain with no recurrent states.

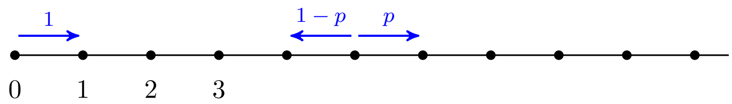
(d) Give an example of a Markov chain such that (i) D and R are both non-empty, (ii) R is reachable from every state $a \in D$, (iii) the first hitting time T_R of the set of recurrent states starting from some transient state $a \in S$ is *not* almost surely finite.

(e) Show that, for every $a \in D$, we still have $\mathbb{P}(X_n = a \mid X_0 = a) \to 0$ as $n \to \infty$.

Hint: Show the stronger property that $\sum_{n=0}^{\infty} \mathbb{P}(X_n = a \mid X_0 = a) < \infty$.

(f) Is it true that, for every $a \in D$, $\mathbb{P}(X_n \in D \mid X_0 = a) \to 0$ as $n \to \infty$?

Problem 28 (Random walk reflected at the origin)**.** Consider a particle performing a random walk on $\mathbb{N} = \{0, 1, 2, \ldots\}$. When the particle is at a position $a \geqslant 1$, it moves either one unit to the right (with probability p) or one unit to the left (with probability $1 - p$), independently of its previous steps. When the particle is at position 0, it moves one unit to the right (with probability 1). We assume that $0 < p < 1$.



(a) Argue that the Markov chain has at most one stationary distribution.

(b) For which values of p does the Markov chain have a stationary distribution?

(c) Depending on the value of p, determine if each of the states is transient, null recurrent, or positive recurrent. Compute the expected return time of the positive recurrent states.

added on: 2025-11-15

Problem 29 (Borel-Cantelli lemma: Alternative proof)**.** Let $A_1, A_2, A_3, \ldots$ be a sequence of events in a probability space Ω.

(a) Verify that

$$\{\omega \in \Omega : \omega \in A_n \text{ for infinitely many values of } n\} = \bigcap_{n_0=1}^{\infty} \bigcup_{n=n_0}^{\infty} A_n \,.$$

(b) Give a similar interpretation of $\bigcup_{n_0=1}^{\infty} \bigcap_{n=n_0}^{\infty} A_n$.

(c) Argue that $\mathbb{P}\left(\bigcap_{n_0=1}^{\infty} \bigcup_{n=n_0}^{\infty} A_n\right) = \lim_{n_0\to\infty} \mathbb{P}\left(\bigcup_{n=n_0}^{\infty} A_n\right)$.

(d) Argue that $\mathbb{P}\left(\bigcup_{n=n_0}^{\infty} A_n\right) \leqslant \sum_{n=n_0}^{\infty} \mathbb{P}(A_n)$.

(e) Show that, if $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, then $\mathbb{P}\left(A_n \text{ occurs for infinitely many values of } n\right) = 0$.

Problem 30 (Second Borel-Cantelli lemma)**.** Let $A_1, A_2, A_3, \ldots$ be a sequence of independent events in a probability space Ω. Show that, if $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$, then $\mathbb{P}\left(A_n \text{ occurs for infinitely many values of } n\right) = 1$.

Hint: (i) Use part (b) for the complements. (ii) Use the fact that $1 - a \leqslant \mathrm{e}^{-a}$ for every $a \in \mathbb{R}$.

Problem 31 (Chernoff–Hoeffding bound; strong LLN)**.** The Chernoff-Hoeffding theorem states that:

Theorem. Let $X_1, X_2, \ldots, X_n$ be independent Bernoulli random variables with parameter $p \in (0, 1)$. For $n \in \mathbb{Z}_+$, let $S_n := X_1 + X_2 + \cdots + X_n$. Then, for every $n \in \mathbb{Z}_+$, we have

$$\mathbb{P}\left(\frac{S_n}{n} \geqslant p + \varepsilon\right) \leqslant \mathrm{e}^{-nD(p,p+\varepsilon)} \qquad \text{for every } \varepsilon \geqslant 0 \text{ with } p + \varepsilon < 1,$$
$$\mathbb{P}\left(\frac{S_n}{n} \leqslant p - \varepsilon\right) \leqslant \mathrm{e}^{-nD(p,p-\varepsilon)} \qquad \text{for every } \varepsilon \geqslant 0 \text{ with } p - \varepsilon > 0,$$

where $D(p, q) := q \ln \frac{q}{p} + (1 - q) \ln \frac{1-q}{1-p}$ for $p, q \in (0, 1)$.

(a) Verify that $D(p, q) \geqslant 0$ with equality if and only if $q = p$.

Hint (brute force approach): Show that the function $q \mapsto D(p, q)$ takes it minimum precisely when $q = p$.

Hint (simple approach): Use Jensen's inequality and the fact that $x \mapsto \ln(x)$ is convex.

(b) Use the Chernoff–Hoeffding theorem and the Borel-Cantelli lemma to give a proof of the strong law of large numbers in the case of i.i.d. Bernoulli random variables.

Problem 32 (Proof of the Chernoff-Hoffdding bound). Recall that, for every $\mathbb{R}$-valued random variable M and every $\theta \geqslant 0$, applying Markov's inequality to the positive random variable $\mathrm{e}^{\theta M}$ gives

$$\mathbb{P}(M \geqslant c) \leqslant \mathbb{E}[\mathrm{e}^{\theta M}]\mathrm{e}^{-\theta c} \tag{§}$$

for every $c \in \mathbb{R}$. (Note that if $\mathbb{E}[\mathrm{e}^{\theta M}] = \infty$, the inequality still holds, though it is trivial.) In concrete examples, we can look for a value for θ that minimizes the right-hand side hence providing the sharpest upper bound of this form for the probability $\mathbb{P}(M \geqslant c)$.

Let $X_1, X_2, \ldots, X_n$ be independent Bernoulli random variables with parameter $p \in (0,1)$, and for each $n \in \mathbb{Z}_+$, let $S_n := X_1 + X_2 + \cdots + X_n$.

(a) Compute $\psi(\theta) := \mathbb{E}[\mathrm{e}^{\theta X_1}]$.

(b) Verify that $\mathbb{E}[\mathrm{e}^{\theta(S_n/n)}] = \psi(\theta/n)^n$. Apply (§) to show that $\mathbb{P}(S_n/n \geqslant q) \leqslant \psi(\theta/n)^n \mathrm{e}^{-\theta q}$ for every $q \in \mathbb{R}$.

Now, consider $q \in (0,1)$ and let $g(\theta) := \psi(\theta/n)^n \mathrm{e}^{-\theta q}$.

(c) Find a value $\theta_0 \in \mathbb{R}$ that minimizes $g(\theta)$.

Hint: Write $g(\theta) = \big(\psi(\theta/n)\mathrm{e}^{-q\theta/n}\big)^n = h(\theta/n)^n$, where $h(t) := \psi(t)\mathrm{e}^{-qt}$. Note that $x \mapsto x^n$ is increasing on $(0,\infty)$. Find $t_0 > 0$ such that $h'(t_0) = 0$. Verify that $h(t)$ is convex, hence it is minimized at $t = t_0$.

(d) Verify that $\theta_0 \geqslant 0$ if and only if $q \geqslant p$.

(e) Verify that $g(\theta_0) = \mathrm{e}^{-nD(p,q)}$ where $D(p,q) := q \ln \frac{q}{p} + (1-q)\ln\frac{1-q}{1-p}$.

(f) Conclude that $\mathbb{P}(S/n \geqslant p + \varepsilon) \leqslant \mathrm{e}^{-nD(p,p+\varepsilon)}$ for every $\varepsilon \geqslant 0$ such that $p + \varepsilon < 1$.

(g) Show that $\mathbb{P}(S/n \leqslant p - \varepsilon) \leqslant \mathrm{e}^{-nD(p,p-\varepsilon)}$ for every $\varepsilon \geqslant 0$ such that $p - \varepsilon > 0$.

Hint: Apply the result of the previous part to the random variables $Y_i := 1 - X_i$.

Problem 33 (LLN for Markov chains). In class, we proved the following version of the law of large numbers for Markov chains (also known as the *ergodic theorem* of Markov chains):

> **Theorem** (LLN for frequencies). Let $(X_n)_{n=0}^{\infty}$ be an irreducible Markov chain with state space S and a stationary distribution π. Then, almost surely,
>
> $$\frac{N_n(a)}{n} \to \pi(a) \qquad \text{(for every } a \in S\text{)}$$
>
> where $N_n(a)$ is the number of visits to a before time n.

More generally, we have

> **Theorem** (LLN for averages). Let $(X_n)_{n=0}^{\infty}$ be an irreducible Markov chain with state space S and a stationary distribution π. Let $h\colon S \to \mathbb{R}$ be such that $\sum_{a \in S} \pi(a)|h(a)| < \infty$. Then, almost surely,
>
> $$\frac{1}{n}\sum_{k=0}^{n-1} h(X_k) \to \pi(h)$$
>
> where $\pi(h) := \sum_{a \in S} \pi(a)h(a)$ is the mean of h with respect to π.
>
> (*Remark:* The hypothesis $\sum_{a \in S} \pi(a)|h(a)| < \infty$ simply means that the series defining $\pi(h)$ is absolutely convergent.)

(a) Argue that the "LLN for frequencies" is a special case of the "LLN for averages".

(b) Assuming that S is finite, use the "LLN for frequencies" to prove the "LLN for averages".

Hint: Write the average of h in terms of $N_n(a)$ for $a \in S$.

(c) Explain what goes wrong in the argument for the previous part if S is not finite.

(d) Prove the general version of "LLN for averages" by imitating the proof of the "LLN for frequencies" discussed in class. (Namely, assuming $X_0 = a$, start by breaking the time interval from 0 to $n-1$ into cycles based on the consecutive return times to a.)

Problem 34 (Poisson distribution; Review I). Recall that a Poisson random variable is meant as a model of the number of successes in a very large number of independent trials each successful with a very small probability. Verify the consistency of this interpretation with that of a binomial random variable. Namely, let $0 < \mu < \infty$. Show that, as $n \to \infty$, the distribution of a binomial random variable with parameters n and μ/n converges to a Poisson distribution. (Recall Problem 24 regarding the meaning of convergence.)

Problem 35 (Poisson distribution; Review II). You may want to do this exercise alongside Problem 36.

(a) Show that the sum of two independent Poisson random variables is again a Poisson random variable.

(b) Extend the previous result to countable sums.

Hint: For a finite sum, use induction. For an infinite sum, take a limit. Beware of the case in which the sum of the means diverges.

(c) (Raikov's theorem; optional) Let X and Y be independent random variables with values from the set of non-negative integers. Suppose that $N = X + Y$ has a Poisson distribution. Prove that X and Y also have Poisson distributions.

(d) Let $Z_1, Z_2, \ldots$ be a sequence of i.i.d. Bernoulli random variables with parameter p. Let N be a Poisson random variable with mean μ, and suppose that N is independent of Z_is. Let $N_1 := \sum_{i=1}^{N} Z_i$ and $N_0 := N - N_1$.

(i) Show that N_1 and N_0 are Poisson random variables and find their means.

(ii) Show that N_1 and N_0 are independent.

(e) Let M and N be independent Poisson random variables with means μ and η respectively. Let $k \in \mathbb{Z}^+$. Show that, given $M + N = k$, M has a binomial distribution and find its parameters.

Problem 36 (Bernoulli vs. Poisson processes). For each of the statements in Problems 5 and 35,

(a) Give an interpretation in the context of Poisson processes.

(b) Find an analogous statement in the context of Bernoulli processes.