

Introduction to Reinforcement Learning

Contents

1 Probability Basics	2
Introduction	2
Random variables and distributions	3
Expectation, variance, and covariance	4
Joint distributions and conditional expectation	5
Averages of repeated observations	7
2 Markov Chains - 1	9
Introduction	9
Accessibility, communication, and classes	11
Periodicity	11
Recurrence and transience	11
Chapman - Kolmogorov equations and classification of states	12
Limit theorems and stationary probabilities	13
Stationary distributions and equilibrium	13
Deducing stationary probabilities through invariance	15
3 Markov Chains - 2	16
Population model	16
Transitions among classes, gambler's ruin, and transient times	16
The gambler's ruin problem	17
The fundamental matrix and mean times in transient states	18
Applications of Markov chains	19
Examples	19
Analyzing a multiclass Markov chain	20
Waiting for two consecutive heads	21

1 Probability Basics

Introduction

An (RL) agent repeatedly acts without knowing exactly what will happen next. An action may lead to several possible states, a reward may be noisy, and even the agent's own policy may randomize its choice. Probability provides a language for describing these possibilities. It also tells us how to average experience without confusing one observed outcome with the behavior of the underlying system. There are three objects to keep separate. An **outcome** is one possible result of an experiment. The **sample space** Ω contains all possible outcomes. An **event** is a subset of Ω . If a die is rolled, then $\Omega = \{1, 2, 3, 4, 5, 6\}$, while the event “the result is even” is $\{2, 4, 6\}$. A probability measure assigns a number $\mathbb{P}(A)$ to each event A so that

$$\mathbb{P}(A) \geq 0, \quad \mathbb{P}(\Omega) = 1,$$

and probabilities add over disjoint events. These axioms are simple, but almost every rule used below follows from them. When information becomes available, probabilities change. If $\mathbb{P}(B) > 0$, the **conditional probability** of A given B is

$$\mathbb{P}(A | B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}. \quad (1.1)$$

Rearranging this definition gives the multiplication rule

$$\mathbb{P}(A \cap B) = \mathbb{P}(A | B)\mathbb{P}(B).$$

Events A and B are **independent** when $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Conditional probability and independence answer different questions. Conditioning asks how information changes a probability. Independence says that learning one event does not change the probability of the other. Suppose $B_1, \dots, B_m$ are disjoint and together cover Ω . Splitting A according to which B_i occurred gives the **law of total probability**,

$$\mathbb{P}(A) = \sum_{i=1}^m \mathbb{P}(A | B_i)\mathbb{P}(B_i). \quad (1.2)$$

Combining this identity with the multiplication rule gives Bayes' formula:

$$\mathbb{P}(B_j | A) = \frac{\mathbb{P}(A | B_j)\mathbb{P}(B_j)}{\sum_{i=1}^m \mathbb{P}(A | B_i)\mathbb{P}(B_i)}. \quad (1.3)$$

Example 1.1;(An alarm)

A component has prior failure probability 0.02. An alarm has sensitivity 0.90 and false-alarm probability 0.05. If F denotes failure and A denotes an alarm, then

$$\mathbb{P}(F | A) = \frac{0.90(0.02)}{0.90(0.02) + 0.05(0.98)} \approx 0.269.$$

The posterior failure probability is about 27%. The alarm is informative, but the low prior failure rate still matters.

Random variables and distributions

An outcome is often too detailed for the question being asked. A **random variable** X converts each outcome into a number. Formally, X is a function from Ω to $\mathbb{R}$. The randomness lies in the outcome selected by the experiment; once the outcome is fixed, the value of X is fixed as well. A random variable is **discrete** when its possible values form a finite or countable set. The number of faulty servers, the reward in a finite game, and the number of arrivals during one minute are discrete. A random variable is **continuous** when its possible values range over an interval or another uncountable set. Waiting time and temperature are standard examples.

Probability mass functions

For a discrete random variable, the **probability mass function** is

$$p_X(x) = \mathbb{P}(X = x).$$

It must satisfy $p_X(x) \geq 0$ and $\sum_x p_X(x) = 1$. A histogram places a bar of height $p_X(x)$ at every possible value. The bar height represents probability mass, not merely how often the value appeared in one finite data set. Three discrete families occur often. A Bernoulli random variable records one success and has $\mathbb{P}(X = 1) = p$. A binomial random variable counts successes in n independent Bernoulli trials:

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, \dots, n.$$

A Poisson random variable is commonly used for counts in a fixed interval:

$$\mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, \dots$$

The choice of a family is a modelling decision. Its assumptions should be checked against the mechanism that produces the data.

Probability density functions

For a continuous random variable, probability is described by a **density** f_X . It satisfies

$$f_X(x) \geq 0, \quad \int_{-\infty}^{\infty} f_X(x) dx = 1,$$

and probabilities are areas:

$$\mathbb{P}(a < X \leq b) = \int_a^b f_X(x) dx. \quad (1.4)$$

The value $f_X(x)$ is not itself a probability. A density can exceed one, provided its total area is one. Moreover, $\mathbb{P}(X = x) = 0$ for every fixed x when X has a density. This does not mean that X takes no value; it means that individual points have zero width and hence zero area. The uniform, exponential, normal, gamma, triangular, and truncated normal distributions shown in the lecture slides illustrate how different shapes encode different kinds of uncertainty. A uniform density treats every equal-length subinterval alike. An exponential model is often used for positive waiting times. A normal density is symmetric around its mean. No shape is universally correct; the context determines which model is reasonable.

The cumulative distribution function

The **cumulative distribution function**, or CDF, works for every random variable:

$$F_X(x) = \mathbb{P}(X \leq x). \quad (1.5)$$

It is nondecreasing, approaches 0 as $x \rightarrow -\infty$, approaches 1 as $x \rightarrow \infty$, and is right-continuous. It also gives interval probabilities through

$$\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a).$$

For a discrete variable, the CDF is a step function whose jump at x equals $p_X(x)$. For a continuous variable with density f_X , it is the accumulated area

$$F_X(x) = \int_{-\infty}^x f_X(t) dt,$$

and $F'_X(x) = f_X(x)$ wherever the derivative exists.

Example 1.2;(The CDF of a fair die)

For a fair die, $F_X(x) = 0$ when $x < 1$, and

$$F_X(x) = \frac{k}{6} \quad \text{when } k \leq x < k+1, \quad k = 1, \dots, 5,$$

with $F_X(x) = 1$ for $x \geq 6$. The jump at each integer is $1/6$. In particular,

$$\mathbb{P}(2 < X \leq 5) = F_X(5) - F_X(2) = \frac{5}{6} - \frac{2}{6} = \frac{1}{2}.$$

Expectation, variance, and covariance

The distribution contains the full probabilistic description of a random variable. In practice, a few numerical summaries are often useful. The most important is the **expected value**,

$$\mathbb{E}[X] = \sum_x x p_X(x) \quad \text{or} \quad \mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x) dx, \quad (1.6)$$

according as X is discrete or continuous. Expectation is a probability-weighted average over hypothetical repetitions. It need not be a possible value of X . A fair die has mean 3.5, although no roll can equal 3.5. More generally, if g is a function, then

$$\mathbb{E}[g(X)] = \sum_x g(x) p_X(x) \quad \text{or} \quad \mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx. \quad (1.7)$$

This rule computes the expectation of a transformed variable without first deriving its new distribution. Expectation is linear. Whenever the expectations exist,

$$\mathbb{E}[aX + bY + c] = a\mathbb{E}[X] + b\mathbb{E}[Y] + c. \quad (1.8)$$

Independence is not required. This simple fact is especially useful with indicator variables. If $I_A = \mathbf{1}_{\{A\}}$ equals one when A occurs and zero otherwise, then $\mathbb{E}[I_A] = \mathbb{P}(A)$. Consequently, the expected number

of events that occur equals the sum of their individual probabilities, even when the events are dependent. The mean describes location but not spread. The **variance** is

$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2. \quad (1.9)$$

The standard deviation $\sqrt{\text{Var}(X)}$ has the same units as X . A variance of zero means that X is constant with probability one. A large variance indicates that values far from the mean receive appreciable probability, although variance alone does not describe the complete shape of the distribution.

Proof. Write $\mu = \mathbb{E}[X]$. Expanding the square and applying linearity gives

$$\mathbb{E}[(X - \mu)^2] = \mathbb{E}[X^2] - 2\mu\mathbb{E}[X] + \mu^2 = \mathbb{E}[X^2] - \mu^2.$$

This proves the second expression in (1.9). □

For constants a, b ,

$$\mathbb{E}[aX + b] = a\mathbb{E}[X] + b, \quad \text{Var}(aX + b) = a^2 \text{Var}(X).$$

Adding b shifts the distribution without changing its spread. Multiplying by a scales standard deviation by $|a|$ and variance by a^2 .

Example 1.3;(Mean and variance of a die)

For a fair die,

$$\mathbb{E}[X] = \frac{1 + 2 + 3 + 4 + 5 + 6}{6} = \frac{7}{2}, \quad \mathbb{E}[X^2] = \frac{1^2 + 2^2 + \dots + 6^2}{6} = \frac{91}{6}.$$

Therefore

$$\text{Var}(X) = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}, \quad \text{sd}(X) = \sqrt{\frac{35}{12}} \approx 1.708.$$

Two random variables may move together. Their **covariance** is

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

The correlation is $\text{Corr}(X, Y) = \text{Cov}(X, Y)/(\sigma_X \sigma_Y)$ when both standard deviations are positive. Independence implies zero covariance when the required moments exist, but zero covariance does not usually imply independence. The variance of a sum is

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2 \text{Cov}(X, Y). \quad (1.10)$$

Thus variances add for independent random variables, or more generally for uncorrelated ones.

Remark 1.1;(Averages and nonlinear functions)

In general, $\mathbb{E}[g(X)] \neq g(\mathbb{E}[X])$. For a convex function g , using Jensen's inequality gives $g(\mathbb{E}[X]) \leq \mathbb{E}[g(X)]$. This distinction matters whenever an objective contains a maximum, a square, a logarithm, or another nonlinear transformation.

Joint distributions and conditional expectation

Several random quantities are often observed together. For discrete X and Y , the **joint mass function** is

$$p_{X,Y}(x, y) = \mathbb{P}(X = x, Y = y).$$

Summing out one variable gives a marginal distribution:

$$p_X(x) = \sum_y p_{X,Y}(x, y), \quad p_Y(y) = \sum_x p_{X,Y}(x, y).$$

For continuous variables, a joint density $f_{X,Y}$ satisfies

$$\mathbb{P}((X, Y) \in C) = \iint_C f_{X,Y}(x, y) dx dy,$$

and the marginal density is obtained by integration, for example

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy.$$

The variables are independent exactly when their joint distribution factors into the product of their marginals. For discrete variables with $\mathbb{P}(Y = y) > 0$,

$$p_{X|Y}(x | y) = \frac{p_{X,Y}(x, y)}{p_Y(y)}, \quad \mathbb{E}[X | Y = y] = \sum_x x p_{X|Y}(x | y).$$

For continuous variables with $f_Y(y) > 0$, the corresponding formulas are

$$f_{X|Y}(x | y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}, \quad \mathbb{E}[X | Y = y] = \int_{-\infty}^{\infty} x f_{X|Y}(x | y) dx.$$

As y varies, the quantity $\mathbb{E}[X | Y = y]$ varies. Therefore $\mathbb{E}[X | Y]$ is itself a random variable: after Y is observed, it gives the appropriate conditional mean.

Theorem 1.1;(The law of iterated expectation)

Whenever $\mathbb{E}[|X|] < \infty$,

$$\mathbb{E}[\mathbb{E}[X | Y]] = \mathbb{E}[X]. \quad (1.11)$$

In words, averaging the conditional averages recovers the overall average.

Proof. For discrete variables,

$$\begin{aligned} \mathbb{E}[\mathbb{E}[X | Y]] &= \sum_y \mathbb{E}[X | Y = y] p_Y(y) \\ &= \sum_y \sum_x x p_{X|Y}(x | y) p_Y(y) \\ &= \sum_x x \sum_y p_{X,Y}(x, y) = \sum_x x p_X(x) = \mathbb{E}[X]. \end{aligned}$$

The continuous proof replaces the sums by integrals. □

Conditional expectation also separates two sources of variation:

$$\text{Var}(X) = \mathbb{E}[\text{Var}(X | Y)] + \text{Var}(\mathbb{E}[X | Y]). \quad (1.12)$$

The first term is the average uncertainty remaining after Y is known. The second measures how much the conditional mean changes across values of Y .

Example 1.4;(Two actions with the same mean reward)

Suppose an agent chooses a safe action with probability 0.6 and a risky action with probability 0.4. The safe action gives reward 1. The risky action gives 3 or -1 , each with conditional probability $1/2$. Both actions have conditional mean reward 1, so the tower property gives $\mathbb{E}[R] = 1$. They do not have the same risk. The safe action has conditional variance 0, while the risky action has conditional variance 4. Hence

$$\text{Var}(R) = 0.6(0) + 0.4(4) + \text{Var}(1) = 1.6.$$

An expected reward reports the centre of a distribution; it does not report the possible consequences around that centre.

Averages of repeated observations

Expectations become useful in data analysis because sample averages often approach them. Let $X_1, X_2, \dots$ be independent and identically distributed random variables with mean μ , and define

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i.$$

Theorem 1.2;(Weak law of large numbers)

If $\text{Var}(X_i) = \sigma^2 < \infty$, then for every $\varepsilon > 0$,

$$\mathbb{P}(|\bar{X}_n - \mu| \geq \varepsilon) \longrightarrow 0. \quad (1.13)$$

Thus the sample average converges in probability to the population mean.

Proof. Independence gives

$$\text{Var}(\bar{X}_n) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{\sigma^2}{n}.$$

Chebyshev's inequality now yields

$$\mathbb{P}(|\bar{X}_n - \mu| \geq \varepsilon) \leq \frac{\text{Var}(\bar{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \longrightarrow 0.$$

□

The law of large numbers is a statement about accuracy, not about the shape of the sampling distribution. It says that large averages concentrate near μ . The **central limit theorem** describes the scale and approximate shape of the remaining error.

Theorem 1.3;(Central limit theorem)

If the X_i are i.i.d. with mean μ and variance $0 < \sigma^2 < \infty$, then

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{\mathcal{D}} N(0, 1). \quad (1.14)$$

Equivalently, for large n ,

$$\bar{X}_n \approx N\left(\mu, \frac{\sigma^2}{n}\right), \quad \sum_{i=1}^n X_i \approx N(n\mu, n\sigma^2).$$

Unlike the short proof of the weak law, a full proof of the central limit theorem ris beyond the scope of this chapter. The typical error in a sample average is of order $1/\sqrt{n}$, not $1/n$. Reducing the standard error by a factor of two therefore requires about four times as many independent observations. For repeated fair-die rolls, $\mu = 7/2$ and $\sigma^2 = 35/12$. Hence

$$\bar{X}_n \approx N\left(\frac{7}{2}, \frac{35}{12n}\right).$$

The number “30 or 40 observations” is only a rule of thumb. A highly skewed or heavy-tailed distribution may require a much larger sample, and dependence between observations can invalidate the i.i.d. calculation. If σ is estimated by the sample standard deviation s , the CLT motivates the approximate large-sample interval

$$\bar{X}_n \pm 1.96 \frac{s}{\sqrt{n}}.$$

The quantity $s/\sqrt{n}$ is the **standard error**. It measures uncertainty in the estimated mean, whereas s measures variability among individual observations.

2 Markov Chains

Introduction

Consider any system that evolve one observation at a time. For example a queue after successive departures, a gambler's fortune after successive bets, or the size of a population from one generation to the next. Loosely speaking, if we have observed some states of the system: $X_0, X_1, \dots, X_n$, then the most general prediction of X_{n+1} could depend on that entire history. Such a model may be realistic, but it can also be very hard to analyze. The Markov idea is to search for a state description that makes the present sufficient. we do not assume that the future is independent of the present. we assume that, once the present is known, the earlier past adds nothing to the next prediction.

Key Idea 2.1;(The present is a sufficient summary)

A stochastic process $\{X_n : n \geq 0\}$ on a finite or countable state space S is a time-homogeneous Markov chain when

$$\mathbb{P}(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = \mathbb{P}(X_{n+1} = j \mid X_n = i) = p_{ij}. \quad (2.1)$$

The current state is not independent of the next state; the past and the next state are conditionally independent given the current state.

The number p_{ij} is the **one-step transition probability** from i to j . Since a next state (j) must be chosen, the sum p_{ij} over all possible next states must be 1:

$$p_{ij} \geq 0, \quad \sum_{j \in S} p_{ij} = 1.$$

We arrange these numbers into a **transition matrix** $P = (p_{ij})$. Row i is the just a 'rule' the process consults in state i . If $\alpha^{(0)}$ is the row vector giving the law of X_0 , then $\alpha^{(1)} = \alpha^{(0)}P$.

Remark 2.1;(“memoryless”)

A state may itself contains memory: the age of a machine, the last two symbols in a sequence, etc. In the markovian realm, all predictive information from the past has been compressed into the ‘present’ state.

An example

Suppose a machine is inspected daily and labeled Good (G), Degraded (D), or Failed (F), with

$$P = \begin{pmatrix} 0.75 & 0.20 & 0.05 \\ 0.10 & 0.60 & 0.30 \\ 0.60 & 0.40 & 0 \end{pmatrix}.$$

Then, the probability of being failed if it is already degraded is 0.30. Starting from G , the probability of being failed after two days is found by inserting the intermediate state:

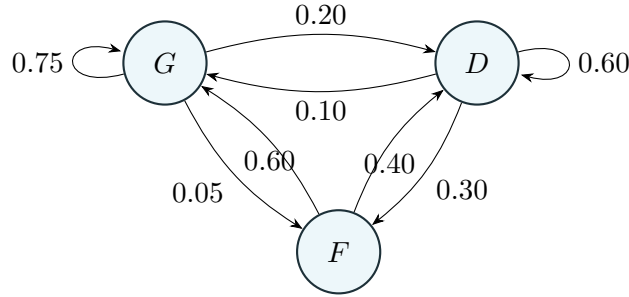
$$\mathbb{P}(X_2 = F \mid X_0 = G) = 0.75(0.05) + 0.20(0.30) + 0.05(0) = 0.0975.$$

This calculation, in a more general setting will later be formalised as the Chapman - Kolmogorov equations. One can think of the P as some sort of a simulator. If the machine is Good, we draw

its next label from the first row; after that draw, we forget which row we used and consult the row belonging to the newly drawn state. Repeating this experiment produces a sample path. Repeating the entire sample-path experiment many times produces empirical ‘state frequencies’. One purpose of the theory ahead would be to explain when if and when these frequencies stabilize. For instance, direct multiplication gives

$$P^2 = \begin{pmatrix} .6125 & .2900 & .0975 \\ .3150 & .5000 & .1850 \\ .4900 & .3600 & .1500 \end{pmatrix}.$$

Every entry now has a two-step meaning. The first, second and the third row of P^2 , for example, is the distribution two days after a Good, Bad and Degraded start respectively. The probability of a path can also be explained using matrix products.



A Random Walk¹

Let $Y_1, Y_2, \dots$ be independent and identically distributed (i.i.d.) integer-valued increments, and define the random walk S_n by:

$$S_0 = 0, \quad S_n = \sum_{r=1}^n Y_r \implies S_{n+1} = S_n + Y_{n+1}.$$

The transition probabilities are given by $p_{ij} = \mathbb{P}(Y_1 = j - i)$. For a simple random walk, each increment satisfies $Y_r = 1$ with probability p and $Y_r = -1$ with probability $q = 1 - p$. Fix an absolute-value trajectory ending at $R_n = |S_n| = i > 0$, and let r be its most recent visit to 0. After time r , the signed path cannot cross the origin. Thus, given the absolute history, there are exactly two compatible signed continuations: The **Positive Path** ($S_n = +i$): which contains u upward and d downward steps, where $u - d = i$. Its probability is $p^u q^d$. Or the **Reflected Path** ($S_n = -i$): that contains d upward and u downward steps, whose probability is $p^d q^u$. The ratio of these likelihoods simplifies to:

$$\frac{\mathbb{P}(S_n = +i)}{\mathbb{P}(S_n = -i)} = \frac{p^u q^d}{p^d q^u} = \left(\frac{p}{q}\right)^{u-d} = \left(\frac{p}{q}\right)^i.$$

Because all path history prior to the last zero visit at time r cancels out in this ratio, normalizing these two weights yields the conditional distribution for the hidden sign:

$$\mathbb{P}(S_n = i \mid R_0, \dots, R_n = i) = \frac{p^i}{p^i + q^i}. \quad (2.2)$$

¹Sections typeset in italics contain supplementary material for further reading and were not covered in class.

Transition Probabilities. Equation (2.2) shows that the hidden sign distribution depends strictly on the current distance i , independent of earlier absolute values. Consequently, $R_n = |S_n|$ retains the Markov property, with transition probabilities:

$$\mathbb{P}(R_{n+1} = i + 1 \mid R_n = i) = \left(\frac{p^i}{p^i + q^i} \right) p + \left(\frac{q^i}{p^i + q^i} \right) q = \frac{p^{i+1} + q^{i+1}}{p^i + q^i}, \quad (2.3)$$

$$\mathbb{P}(R_{n+1} = i - 1 \mid R_n = i) = \left(\frac{p^i}{p^i + q^i} \right) q + \left(\frac{q^i}{p^i + q^i} \right) p = \frac{pq(p^{i-1} + q^{i-1})}{p^i + q^i}, \quad (2.4)$$

and $\mathbb{P}(R_{n+1} = 1 \mid R_n = 0) = 1$.

Accessibility, communication, and classes

State j is **accessible** from state i , written $i \rightarrow j$, if $p_{ij}^{(n)} > 0$ for some $n \geq 0$. States i and j **communicate**, written $i \leftrightarrow j$, if $i \rightarrow j$ and $j \rightarrow i$.

Proposition 2.1;(Communication is an equivalence relation)

The relation $\leftrightarrow$ is reflexive, symmetric, and transitive. It therefore partitions the state space into disjoint communicating classes.

Proof. Reflexivity follows from $p_{ii}^{(0)} = 1$, and symmetry is part of the definition. For transitivity, suppose $i \leftrightarrow j$ and $j \leftrightarrow k$. There are integers m, n for which $p_{ij}^{(m)} > 0$ and $p_{jk}^{(n)} > 0$. Chapman - Kolmogorov gives

$$p_{ik}^{(m+n)} \geq p_{ij}^{(m)} p_{jk}^{(n)} > 0,$$

so $i \rightarrow k$. Reversing the roles gives $k \rightarrow i$, and hence $i \leftrightarrow k$. $\square$

A chain is **irreducible** if all states belong to one communicating class. A set C is **closed** if no one-step transition leaves it: $p_{ij} = 0$ whenever $i \in C$ and $j \notin C$. An absorbing state is a one-state closed class, equivalently $p_{ii} = 1$. These definitions separate two questions. Communication describes which states can reach one another. Closedness describes whether probability can escape. A reducible chain can have transient states that feed several different closed classes; its long-run behavior can then depend on the starting state.

Periodicity

The set of possible return times to i is $\mathcal{R}_i = \{n \geq 1 : p_{ii}^{(n)} > 0\}$. The **period** of i is

$$d(i) = \gcd \mathcal{R}_i.$$

If $d(i) = 1$, the state is aperiodic. A self-loop is sufficient for aperiodicity, although it is not necessary. States in the same communicating class have the same period. Indeed, suppose a path from j to i has length a and a path from i to j has length b . Whenever n is a possible return time to i , both $a + b$ and $a + n + b$ are possible return times to j . Thus $d(j)$ divides their difference n , and consequently $d(j)$ divides $d(i)$. Interchanging i and j gives the reverse divisibility, so $d(i) = d(j)$.

Recurrence and transience

Let

$$T_i^+ = \inf\{n \geq 1 : X_n = i\}, \quad f_{ii} = \mathbb{P}_i(T_i^+ < \infty).$$

State i is **recurrent** when $f_{ii} = 1$ and **transient** when $f_{ii} < 1$. Recurrence means eventual return is certain; it does not mean that the return must occur quickly.

Theorem 2.1;(The recurrence criterion)

A state i is recurrent if and only if

$$\sum_{n=0}^{\infty} p_{ii}^{(n)} = \infty. \quad (2.5)$$

If the series is finite, its value is the expected total number of visits to i , including the initial visit.

Proof. Let $N_i = \sum_{n \geq 0} \mathbf{1}_{\{X_n = i\}}$. By linearity of expectation,

$$\mathbb{E}_i N_i = \sum_{n \geq 0} \mathbb{P}_i(X_n = i) = \sum_{n \geq 0} p_{ii}^{(n)}.$$

After each visit to i , the strong Markov property makes the probability of another return equal to f_{ii} . Thus the number of visits has a geometric tail and mean $1/(1 - f_{ii})$ when $f_{ii} < 1$. Its mean is infinite exactly when $f_{ii} = 1$. $\square$

Recurrence and transience are class properties. Indeed, if $i \leftrightarrow j$, choose a, b such that $p_{ji}^{(a)} > 0$ and $p_{ij}^{(b)} > 0$. Then

$$p_{jj}^{(a+n+b)} \geq p_{ji}^{(a)} p_{ii}^{(n)} p_{ij}^{(b)}.$$

If the return series for i diverges, so does the return series for j . Symmetry gives the converse. For the simple random walk on $\mathbb{Z}$,

$$p_{00}^{(2n)} = \binom{2n}{n} p^n q^n \sim \frac{(4pq)^n}{\sqrt{\pi n}}.$$

The series diverges only when $p = q = 1/2$. Hence the symmetric walk is recurrent, while every biased simple random walk on $\mathbb{Z}$ is transient.

Chapman - Kolmogorov equations and classification of states

We begin by defining the n -step transition probabilities, denoted as P_{ij}^n , which represent the probability that a process currently in state i will be in state j after n additional transitions. Formally, this is expressed as $P_{ij}^n = \mathbb{P}(X_{n+m} = j \mid X_m = i)$ for any $n \geq 0$ and $i, j \geq 0$. Naturally, P_{ij}^1 is simply the one-step transition probability P_{ij} . To travel from state i to state j in $n + m$ steps, the chain must occupy some intermediate state k exactly at the n -th step. The Chapman - Kolmogorov equations provide a method for computing these n -step transition probabilities by formalizing this intuition.

Theorem 2.2;(Chapman - Kolmogorov equations)

For all integers $n, m \geq 0$ and all states i, j ,

$$P_{ij}^{n+m} = \sum_{k=0}^{\infty} P_{ik}^n P_{kj}^m. \quad (2.6)$$

Consequently, if we let $P^{(n)}$ denote the matrix of n -step transition probabilities, this equation asserts that $P^{(n+m)} = P^{(n)} P^{(m)}$. By induction, it follows that $P^{(n)} = P^n$, meaning the n -step transition matrix can be calculated by multiplying the one-step transition matrix P by itself n times.

Proof. The equations can be established by conditioning on the intermediate state at time n :

$$\begin{aligned}
 P_{ij}^{n+m} &= \mathbb{P}(X_{n+m} = j \mid X_0 = i) \\
 &= \sum_{k=0}^{\infty} \mathbb{P}(X_{n+m} = j, X_n = k \mid X_0 = i) \\
 &= \sum_{k=0}^{\infty} \mathbb{P}(X_{n+m} = j \mid X_n = k, X_0 = i) \mathbb{P}(X_n = k \mid X_0 = i) \\
 &= \sum_{k=0}^{\infty} P_{ik}^n P_{kj}^m.
 \end{aligned}$$

The important step happens in the third equality, where the Markovian property is invoked: the conditional distribution of any future state X_{n+m} , given the past state X_0 and the present state X_n , depends solely on the present state X_n . $\square$

Limit theorems and stationary probabilities

Let μ_{jj} denote the expected number of transitions needed to return to state j . If state j is recurrent, a process starting in j will return with probability 1; however, the expected time until this return occurs can still be infinite. A recurrent state j is classified as positive recurrent if $\mu_{jj} < \infty$, and it is classified as null recurrent if $\mu_{jj} = \infty$. Positive and null recurrence are class properties, meaning that if one state in a communicating class is positive (or null) recurrent, all states in that communicating class share the same property. Because transitions into a recurrent state j act as renewals, the sequence of visits forms a renewal process. If we let $N_j(n)$ denote the number of transitions into state j by time n , the strong law for renewal processes dictates that the long-run transition rate converges to the inverse of the expected return time. Therefore, if the process starts in state i and $i \leftrightarrow j$, we have

$$\lim_{n \rightarrow \infty} \frac{N_j(n)}{n} = \frac{1}{\mu_{jj}} \quad \text{with probability 1.} \quad (2.7)$$

In the case of null recurrence, this limit is identically zero. This indicates that while returns to state j are absolutely certain, they are so sparse that their long-run fractional frequency vanishes entirely.

Stationary distributions and equilibrium

A probability distribution $\{\pi_j, j \geq 0\}$ is said to be stationary for a Markov chain if it satisfies the system of equations defined by

$$\pi_j = \sum_{i=0}^{\infty} \pi_i P_{ij}, \quad \sum_{j=0}^{\infty} \pi_j = 1. \quad (2.8)$$

If the initial probability distribution of the process X_0 is chosen to be this stationary distribution, then the probability of being in state j at any future time n , denoted $\mathbb{P}(X_n = j)$, remains exactly π_j . While individual paths continue to transition dynamically from state to state, the ensemble probability law remains completely fixed over time, rendering the sequence a stationary process.

Key Idea 2.2;(Global flow balance)

The equation $\pi_j = \sum_i \pi_i P_{ij}$ establishes that, in stationarity, the long-run probability mass entering state j from all possible originating states exactly equals the total probability mass currently

residing in state j .

Theorem 2.3;(Positive recurrence and equilibrium)

An irreducible, aperiodic Markov chain possesses a stationary probability distribution if and only if all of its states are positive recurrent. When this condition is met, the stationary distribution is uniquely determined and directly related to the expected return time μ_{jj} by the equation

$$\pi_j = \frac{1}{\mu_{jj}}. \quad (2.9)$$

Furthermore, under these conditions, the chain is ergodic, meaning the n -step transition probabilities converge exactly to the stationary distribution, such that $\lim_{n \rightarrow \infty} P_{ij}^n = \pi_j$ for all initial states i .

Proof. Fix a state j and let T_j^+ be the first return time to j . Suppose first that the chain is positive recurrent, so that $\mu_{jj} = \mathbb{E}_j[T_j^+] < \infty$. During one complete cycle, beginning with a visit to j and ending immediately before the next visit to j , define

$$\nu_k = \mathbb{E}_j \left[\sum_{n=0}^{T_j^+-1} \mathbf{1}_{\{X_n=k\}} \right].$$

Thus ν_k is the expected number of visits to state k during one return cycle. The cycle contains exactly one initial visit to j , so $\nu_j = 1$, while summing over all states counts the expected cycle length:

$$\sum_k \nu_k = \mathbb{E}_j[T_j^+] = \mu_{jj}.$$

The vector ν is invariant under one transition. Indeed, using the Markov property and conditioning on the current state,

$$\begin{aligned} \sum_i \nu_i P_{ik} &= \mathbb{E}_j \left[\sum_{n=0}^{T_j^+-1} P_{X_n k} \right] \\ &= \mathbb{E}_j \left[\sum_{n=0}^{T_j^+-1} \mathbf{1}_{\{X_{n+1}=k\}} \right]. \end{aligned}$$

The last sum counts visits to k at times $1, \dots, T_j^+$. Because the cycle starts and ends at the same state j , this count is the same as the number of visits at times $0, \dots, T_j^+ - 1$. Hence $\sum_i \nu_i P_{ik} = \nu_k$. Normalizing by the finite cycle length therefore gives

$$\pi_k = \frac{\nu_k}{\mu_{jj}},$$

and $\pi P = \pi$, so π is a stationary probability distribution. In particular, since $\nu_j = 1$,

$$\pi_j = \frac{1}{\mu_{jj}}.$$

Because the chain is irreducible, the same cycle argument may be started from any state, so this identity holds for every state.

Conversely, suppose an irreducible chain has a stationary probability distribution π . A stationary chain with finite total mass cannot continually lose probability from a state without eventually bringing that mass back; equivalently, every state with positive stationary mass is recurrent. Irreducibility then makes every state recurrent. Starting the stationary chain and observing successive returns to a fixed state j , the visits again form a renewal process. By (2.7), their long-run frequency is $1/\mu_{jj}$. On the other hand, stationarity makes the expected fraction of time spent in j equal to π_j at every horizon. Since the empirical fractions are bounded between 0 and 1, bounded convergence gives

$$\pi_j = \frac{1}{\mu_{jj}}.$$

In an irreducible chain every stationary probability is strictly positive, so $\mu_{jj} = 1/\pi_j < \infty$. Thus every state is positive recurrent. The same identity also proves uniqueness: any stationary probability distribution must assign the value $1/\mu_{jj}$ to each state j .

It remains to explain the limiting statement. For a fixed state j , the successive return times to j form a renewal process whose inter-renewal mean is μ_{jj} . Aperiodicity says that the possible return times have greatest common divisor one. The discrete renewal theorem therefore gives

$$p_{jj}^{(n)} \longrightarrow \frac{1}{\mu_{jj}} = \pi_j.$$

Starting instead from any state i , irreducibility and recurrence ensure that the chain eventually hits j with probability one. After that first hit, the subsequent visits to j follow the same renewal structure. The corresponding delayed-renewal theorem consequently yields

$$p_{ij}^{(n)} \longrightarrow \frac{1}{\mu_{jj}} = \pi_j.$$

Hence the distribution of X_n converges to the unique stationary distribution, independently of the initial state. $\square$

Computationally, the stationary distribution is derived by solving the linear system $\pi_j = \sum_i \pi_i P_{ij}$ alongside the normalization constraint $\sum_i \pi_i = 1$. The resulting numerical value for π_j explicitly defines the long-run proportion of time the Markov chain spends in state j , and its reciprocal $1/\pi_j$ defines the mean number of transitions between successive visits to that state.

Example 2.1:(Exercise)

Consider a two-state Markov chain with states 0 and 1, governed by the transition probabilities $P_{00} = \alpha$, $P_{01} = 1 - \alpha$, $P_{10} = \beta$, and $P_{11} = 1 - \beta$. The steady-state balance equations are formulated as $\pi_0 = \pi_0\alpha + \pi_1\beta$ along with the normalization requirement $\pi_0 + \pi_1 = 1$. By expressing π_1 as $1 - \pi_0$ and substituting it into the balance equation, the unique stationary probabilities can be isolated as

$$\pi_0 = \frac{\beta}{1 - \alpha + \beta}, \quad \pi_1 = \frac{1 - \alpha}{1 - \alpha + \beta}. \quad (2.10)$$

Deducing stationary probabilities through invariant

In certain stochastic models, stationary probabilities can be determined not by algebraically solving the steady-state equations, but by hypothesizing an initial distribution and proving that the Markov chain structurally preserves it over time.

Example 2.2;(A population model)

Consider a population where, during each time period, every existing member independently dies with probability p , and new members arrive according to a Poisson distribution with mean λ . If X_n denotes the population size at the beginning of period n , the sequence explicitly forms a Markov chain. Suppose we assume the initial population X_0 is distributed as a Poisson random variable with parameter α . Because each individual independently survives with probability $1 - p$, the number of surviving individuals is a Poisson random variable with mean $\alpha(1 - p)$. Since the newly arriving members form an independent Poisson random variable with mean λ , the total population at the next step, X_1 , follows a Poisson distribution with a combined mean of $\alpha(1 - p) + \lambda$. For the Markov chain to be perfectly stationary, the distribution parameter must remain invariant across steps, enforcing the equality $\alpha = \alpha(1 - p) + \lambda$. Solving this algebraically yields $\alpha = \lambda/p$. By the uniqueness property of stationary distributions for ergodic chains, we conclude that the limiting probability distribution is exactly Poisson with mean λ/p .

Transitions among classes, gambler's ruin, and transient times

A recurrent class operates as a closed system; once the Markov chain enters it, the process can never leave. i.e.,

Proposition 2.2;(A recurrent class has no exit)

Let R be a recurrent class of states. If $i \in R$ and $j \notin R$, then the one-step transition probability is $P_{ij} = 0$.

Proof. Suppose, for the sake of contradiction, that $P_{ij} > 0$. Because states i and j do not communicate (as j is strictly outside the recurrent class R), the process cannot travel from j back to i , meaning the n -step transition probability $P_{ji}^n = 0$ for all n . Consequently, if the process starts in state i , there is a strictly positive probability of at least P_{ij} that the process transitions to j and never returns to i . This directly contradicts the mathematical definition of recurrence, which requires the process to return to state i with probability 1. Therefore, P_{ij} must equal 0. $\square$

When analyzing transitions from transient states into a recurrent class, the objective is often to determine the probability of eventually reaching a specific target set of states, alongside the expected time required to arrive there. By conditioning on the state immediately following the initial transition, we establish a system of linear equations that act mathematically as boundary-value problems. Let A denote a target set of states. We define the hitting probability h_i as the probability that the process, starting in state i , will eventually enter the target set A . Similarly, let t_i denote the mean hitting time, representing the expected number of transitions required to reach A from state i . By conditioning on the outcome of the first transition, these quantities satisfy the following systems of equations:

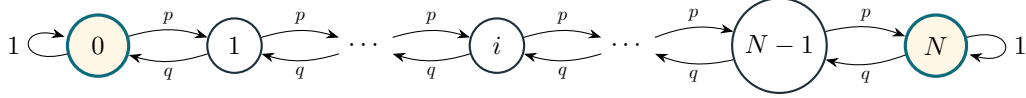
$$h_i = \sum_j P_{ij} h_j \quad (i \notin A), \quad h_i = 1 \quad (i \in A), \quad (2.11)$$

$$t_i = 1 + \sum_j P_{ij} t_j \quad (i \notin A), \quad t_i = 0 \quad (i \in A). \quad (2.12)$$

The boundary conditions strictly dictate that $h_i = 1$ and $t_i = 0$ when the process begins already inside the target set A . The time equation inherently requires an initial constant of 1 because the first conditional step consumes exactly one time unit before evaluating the remaining expected transitions, whereas the probability equation partitions the total probability space without accumulating a unit cost.

The Gambler's Ruin Problem

Consider a gambler starting with an initial fortune of i units, who wins 1 unit with probability p and loses 1 unit with probability $q = 1 - p$ on each independent play. The game concludes when the gambler's fortune reaches either a predetermined goal N or 0. This process constitutes a Markov chain over the state space $\{0, 1, \dots, N\}$, with transition probabilities $P_{i,i+1} = p$ and $P_{i,i-1} = q$ for the transient states $1 \leq i \leq N-1$, and absorbing barriers defined by $P_{00} = 1$ and $P_{NN} = 1$. Because the intermediate states are transient and visited only finitely often, the process will eventually reach either N or 0.



Let h_i denote the probability that the gambler's fortune eventually reaches N before hitting 0, given an initial starting state of i . By conditioning on the outcome of the initial play, we establish the recursive relation $h_i = ph_{i+1} + qh_{i-1}$ subject to the boundary conditions $h_0 = 0$ and $h_N = 1$. Because $p+q = 1$, this equation can be algebraically rearranged into successive geometric differences: $h_{i+1} - h_i = \frac{q}{p}(h_i - h_{i-1})$. Summing these differences and enforcing the upper boundary condition $h_N = 1$ yields the exact hitting probability:

$$h_i = \begin{cases} \frac{1 - (q/p)^i}{1 - (q/p)^N}, & p \neq \frac{1}{2}, \\ \frac{i}{N}, & p = \frac{1}{2}. \end{cases} \quad (2.13)$$

As the target wealth N approaches infinity, the probability of the gambler's fortune diverging to infinity rather than resulting in ruin depends entirely on the win probability p . If $p \leq 1/2$, the limiting probability is strictly 0, meaning that a gambler playing a fair or disadvantageous game against an infinitely rich adversary will eventually go broke with probability 1. Conversely, if $p > 1/2$, the limit evaluates to $1 - (q/p)^i$, signifying a strictly positive probability that the gambler's fortune escapes to infinity. Let e_i denote the expected number of bets the gambler makes before reaching either 0 or N , starting from state i . For an asymmetric game where $p \neq q$, this expectation can be efficiently derived using [Wald's equation](#). Let X_j represent the winnings on the j -th independent bet, with an expected value of $E[X_j] = p(1) + q(-1) = 2p - 1$. If B represents the total number of bets until the game stops, the overall accumulated change in wealth is $\sum_{j=1}^B X_j$. By Wald's equation, the expected total change is $(2p - 1)E[B]$. However, this total change strictly equals $N - i$ with probability h_i , and $-i$ with probability $1 - h_i$. Equating these expectations yields $(2p - 1)e_i = (N - i)h_i + (-i)(1 - h_i) = Nh_i - i$. This directly isolates the expected duration:

$$e_i = \frac{1}{2p - 1} \left(\frac{N[1 - (q/p)^i]}{1 - (q/p)^N} - i \right), \quad p \neq \frac{1}{2}. \quad (2.14)$$

In the strictly fair case ($p = q = 1/2$), the drift vanishes ($2p - 1 = 0$), meaning Wald's equation cannot directly isolate the expectation. Instead, the duration must be derived from the standard recursive time equation $e_i = 1 + \frac{1}{2}e_{i+1} + \frac{1}{2}e_{i-1}$. Rearranging this gives $e_{i+1} - 2e_i + e_{i-1} = -2$. Because the second discrete difference is a constant, e_i must take the form of a quadratic function, $e_i = ai^2 + bi + c$. Applying the absorbing boundary condition $e_0 = 0$ dictates that $c = 0$, while the boundary condition $e_N = 0$ dictates that $b = -aN$. Substituting these parameters back into the recursion confirms that $a = -1$, yielding the exact parabolic solution $e_i = i(N - i)$. This geometric shape confirms that games beginning near either boundary terminate quickly, whereas games starting near the midpoint of the state space endure the longest expected durations.

The fundamental matrix and mean times in transient states

Consider a finite-state Markov chain where the states are ordered such that the set of transient states $T = \{1, 2, \dots, t\}$ appears first. Let Q denote the matrix composed exclusively of the transition probabilities between these transient states. Because the transient states cannot form a closed class, some row sums of Q are strictly less than 1. For any two transient states i and j , let m_{ij} represent the expected total number of time periods the chain spends in state j , given that the process begins in state i . By conditioning on the initial transition from state i , this expectation is expressed algebraically as $m_{ij} = \delta(i, j) + \sum_{k=1}^t P_{ik} m_{kj}$, where $\delta(i, j)$ evaluates to 1 if $i = j$ and 0 otherwise. Defining M as the matrix of these expected values m_{ij} , this system of linear equations translates directly into the matrix relation $M = I + QM$. Algebraic rearrangement yields $(I - Q)M = I$, demonstrating that the expected time spent in transient states is determined by inverting the matrix $I - Q$. Expanding this inversion geometrically produces the fundamental matrix equation:

$$M = (I - Q)^{-1} = I + Q + Q^2 + \dots \quad (2.15)$$

The fundamental matrix M provides the foundation for answering several questions regarding transient behavior simultaneously. The expected time the process survives in the transient region before eventual absorption is calculated by summing the expected visits across all transient states, achieved computationally by the matrix-vector product $M\mathbf{1}$. Furthermore, if a matrix R is constructed to record the direct one-step transition probabilities from transient states into absorbing recurrent states, the product $B = MR$ determines the exact absorption probabilities. Before absorption occurs, the Markov chain occupies a specific transient state j an expected M_{ij} times. Every visit to state j provides an independent probability R_{ja} of transitioning directly into an absorbing state a . Summing these expected transition opportunities over all intermediate transient states j yields the total expected number of entrances into state a , represented by B_{ia} . Because absorption into a recurrent class acts as a terminal event for the transient process, the expected number of entrances into state a is mathematically identical to the absolute probability of the chain ever entering it.

Example 2.3; (Interpreting the fundamental matrix)

Suppose a Markov chain has two transient states and two absorbing states, with transition blocks Q (transient-to-transient) and R (transient-to-absorbing) defined as:

$$Q = \begin{pmatrix} 1/4 & 1/2 \\ 1/3 & 1/3 \end{pmatrix}, \quad R = \begin{pmatrix} 1/4 & 0 \\ 0 & 1/3 \end{pmatrix}.$$

Inverting the matrix $I - Q$ yields the fundamental matrix M , which is then used to compute the expected total survival time $M\mathbf{1}$ and the exact absorption probabilities MR :

$$M = \begin{pmatrix} 2 & 3/2 \\ 1 & 9/4 \end{pmatrix}, \quad M\mathbf{1} = \begin{pmatrix} 7/2 \\ 13/4 \end{pmatrix}, \quad MR = \begin{pmatrix} 1/2 & 1/2 \\ 1/4 & 3/4 \end{pmatrix}.$$

By calculating a single matrix inverse, we extract three distinct physical interpretations of the system. The matrix M explicitly counts the expected visits to specific states; for instance, a process starting in the first transient state expects to visit that same state exactly 2 times. The vector $M\mathbf{1}$ sums these visits, revealing that the process will survive an expected total of 3.5 transitions before absorption. Finally, the matrix MR maps the final destinations, indicating that a process starting in the first transient state has exactly a 1/2 probability of ending up in the first absorbing state.

An applications of Markov chains

Certain algorithms in computer science and operations research function by successively moving from one element to a better element until an optimal configuration is reached. Suppose an algorithm is evaluating N ordered elements, where state 1 represents the strictly optimal choice. From any current state $i > 1$, the process transitions uniformly to any of the strictly better states, meaning the one-step probability of transitioning to state j (where $j < i$) is exactly $1/(i-1)$. Let T_i denote the total number of transitions required to move from state i to state 1. By conditioning on the outcome of the initial transition, we can show a recursive formulation for the expected number of steps:

$$E[T_i] = 1 + \frac{1}{i-1} \sum_{j=1}^{i-1} E[T_j]$$

Using the base condition $E[T_1] = 0$ and the method of mathematical induction one can show that the expected time from state N evaluates to the sum:

$$E[T_N] = \sum_{j=1}^{N-1} \frac{1}{j} \quad (2.16)$$

While the deterministic worst-case scenario dictates that the algorithm might require $N-1$ separate improvements, this probabilistic expectation demonstrates that the required number of steps grows only logarithmically as N increases. To extract the variance and limiting distribution of this algorithmic process, the total time T_N is decomposed into a sum of indicator variables, $T_N = \sum_{j=1}^{N-1} I_j$, where I_j equals 1 if the algorithm ever visits state j and 0 otherwise. These indicator variables are statistically independent, and the unconditional probability of visiting any state j is exactly $P(I_j = 1) = 1/j$. This structural independence is established by conditioning on the sequence of states visited prior to reaching j . Given the specific outcomes of the indicators $I_{j+1}, \dots, I_N$, let n denote the smallest visited state index that is strictly greater than j . Conditional on the process occupying state n , the probability that state j is ever visited evaluates to the ratio of the probability of jumping directly from n to j over the probability of jumping from n to any state index less than or equal to j . This ratio is precisely $(1/(n-1))/(j/(n-1))$, which algebraically simplifies to $1/j$. Because this probability depends solely on the target state j and remains entirely independent of the prior history n , the indicator variables are proven to be independent. Because T_N is expressed as a linear sum of independent Bernoulli random variables, its variance is straightforwardly the sum of their individual variances:

$$\text{Var}(T_N) = \sum_{j=1}^{N-1} \frac{1}{j} \left(1 - \frac{1}{j}\right)$$

For sufficiently large values of N , the sum of a large sequence of independent Bernoulli random variables—each possessing a relatively small probability of success—is mathematically well-approximated by a Poisson distribution. Consequently, T_N follows an approximately Poisson distribution with a mean scaling asymptotically as $\log N$.

Examples

Let us check out a small collection of longer examples; a first encounter with Markov chains becomes much easier once the same ideas are seen cooperating inside a single problem. In each example we shall resist the temptation to calculate immediately. we shall first ask what the states mean, then inspect the geometry of the chain, and only then choose the algebra.

Analyzing a Multiclass Markov Chain

Consider a five-state Markov chain defined by the following transition probability matrix:

$$P = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 \\ 1/4 & 3/4 & 0 & 0 & 0 \\ 1/4 & 0 & 1/4 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

Analyzing the communication structure of this matrix reveals its fundamental geometry without requiring the explicit calculation of matrix powers. States 0 and 1 communicate with each other, and because there are no positive transition probabilities leading outside this pair, they form a closed recurrent class. Similarly, states 3 and 4 communicate exclusively with one another, forming a second closed recurrent class. State 2, however, can transition toward state 0 or toward state 3. Because neither of these closed classes allows a return path, state 2 is strictly transient. The presence of multiple isolated classes precludes the existence of a single, universal long-run distribution that operates independently of the initial state. Furthermore, the first recurrent class $\{0, 1\}$ possesses self-transitions, ensuring it is aperiodic. The second recurrent class $\{3, 4\}$, by contrast, transitions deterministically between its two states at every step, yielding a strict period of two. The existence of periodicity within a recurrent class necessitates a careful distinction between stationary probability distributions and pointwise convergence. The stationary equations for the full Markov chain directly reflect this structural division. Within the recurrent class $\{0, 1\}$, the equilibrium balance equation is expressed as $\pi_0(1/2) = \pi_1(1/4)$, which produces a normalized stationary distribution of $(1/3, 2/3)$. For the periodic class $\{3, 4\}$, the process spends exactly half its time in each state, resulting in a stationary distribution of $(1/2, 1/2)$. Because the chain contains multiple disjoint recurrent classes, any stationary distribution for the entire state space can be formulated as a convex combination of the independent class distributions:

$$\pi^{(a)} = a \left(\frac{1}{3}, \frac{2}{3}, 0, 0, 0 \right) + (1 - a) \left(0, 0, 0, \frac{1}{2}, \frac{1}{2} \right), \quad 0 \leq a \leq 1.$$

The probability mass assigned to the transient state 2 is strictly zero. A stationary distribution cannot assign positive probability mass to a transient state, because mass would continuously leak into the absorbing classes without ever returning, permanently violating the balance requirement. If the process is initialized in the transient state 2, it will eventually enter one of the two recurrent classes with probability 1. By conditioning on the first transition that strictly leaves state 2, the process moves toward the $\{0, 1\}$ class with proportional probability $1/4$ and toward the $\{3, 4\}$ class with proportional probability $1/2$. Consequently, the absolute probability of eventual absorption into the first recurrent class is evaluated as:

$$\mathbb{P}_2\{\text{eventual entrance into } \{0, 1\}\} = \frac{1/4}{1/4 + 1/2} = \frac{1}{3}.$$

The probability of eventual entrance into $\{3, 4\}$ is the complement, $2/3$. The self-transition at state 2 delays the ultimate exit time but exercises no mathematical influence over the likelihood of the final destination. It is mathematically imprecise to declare that the limiting distribution starting from state 2 is simply the combined stationary distribution with $a = 1/3$. The probability mass that absorbs into $\{0, 1\}$ converges pointwise to $(1/3, 2/3)$ due to the aperiodicity of the class. Conversely, the mass entering $\{3, 4\}$ oscillates indefinitely depending on whether the observed time step is even or odd, meaning the standard pointwise limit P^n fails to converge. However, applying Cesàro averaging smooths out this strict period-two oscillation, allowing the time-averaged limits to converge properly to the

expected mixture:

$$\frac{1}{3} \left(\frac{1}{3}, \frac{2}{3}, 0, 0, 0 \right) + \frac{2}{3} \left(0, 0, 0, \frac{1}{2}, \frac{1}{2} \right).$$

Key Idea 2.3;(Stationarity $\nRightarrow$ convergence)

A probability vector may perfectly satisfy the stationary balance equation $\pi = \pi P$ even when the explicit transition matrix powers P^n oscillate without reaching a pointwise limit. Stationarity merely dictates that a specifically prepared initial distribution remains completely invariant under the application of the transition matrix. Pointwise convergence, by contrast, dictates that a distribution produced from an arbitrary starting state strictly approaches a limit, a property that inherently requires the Markov chain to be irreducible, positive recurrent, and aperiodic.

Waiting for two consecutive heads

Consider a sequence of independent tosses of a fair coin, and let T denote the number of tosses required until the pattern "HH" (two consecutive heads) first appears. While the progress toward this pattern inherently depends on preceding tosses, the sequence can be rigorously modeled as a Markov chain by constructing a state space that strictly encodes the relevant history. Let state 0 represent a history ending in a tail (or the initial state), providing no progress toward the target pattern. Let state 1 denote a history ending in exactly one head, representing partial progress. Let state 2 represent the completion of the "HH" pattern, operating as an absorbing state. The one-step transition probability matrix is then formulated as:

$$P = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 0 & 1 \end{pmatrix}.$$

The transition from state 1 to state 0 with probability $1/2$ reflects the occurrence of a tail, which breaks the sequence of consecutive heads and resets the progress toward the pattern. Let t_i denote the expected number of additional tosses required to reach the absorbing state 2, given the current state is i . By definition, $t_2 = 0$. Applying first-step analysis, the expected hitting times satisfy the following linear equations:

$$t_0 = 1 + \frac{1}{2}t_0 + \frac{1}{2}t_1, \quad t_1 = 1 + \frac{1}{2}t_0 + \frac{1}{2}t_2.$$

The constant 1 in these equations accounts for the immediate transition step. Substituting $t_2 = 0$ isolates the variables, yielding the system $t_1 = 1 + \frac{1}{2}t_0$ and $t_0 = 2 + t_1$. Solving this system directly provides the expected waiting times:

$$\boxed{t_0 = 6, \quad t_1 = 4}.$$

The expected waiting time for two consecutive heads from the initial state is therefore six tosses. This exceeds the intuitive expectation of four (the reciprocal of the probability of "HH" occurring in a fixed, independent two-toss block). The disparity arises because overlapping attempts in a continuous sequence are not statistically independent; the appearance of a tail following a head completely nullifies the partial progress, forcing the process to restart from state 0. This result can also be derived systematically using the fundamental matrix. Restricting the transition matrix to the transient states 0 and 1 yields the submatrix Q :

$$Q = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 0 \end{pmatrix}, \quad (I - Q)^{-1} = \begin{pmatrix} 4 & 2 \\ 2 & 2 \end{pmatrix}.$$

The first row of the fundamental matrix $(I - Q)^{-1}$ dictates that, prior to absorption, a process starting in state 0 expects to visit state 0 exactly four times and state 1 exactly two times. The sum of these expected

visits is six, which confirms the expected absorption time t_0 . The fundamental matrix formulation implicitly resolves the first-step analysis equations by calculating the expected state visits simultaneously.