EULERPAD WORKBOOKS

Probability & Statistics for Machine Learning

250 problems from counting and Bayes' rule to hypothesis tests, MLE, Markov chains and KL divergence, with every step worked

Free sample

Chapter 24: Markov chains: transition matrices, stationary distributions and absorption, problems and worked solutions

First edition, November 2026 · full book: 250 problems in 25 chapters

About this sample

This is Chapter 24 of *Probability & Statistics for Machine Learning*, complete: its 10 problems, then the worked solution to each one and the mistakes that most often go wrong on them. The full workbook has 250 problems in 25 chapters, 8 of them written only for the workbook, in two volumes: *Problems* and *Worked Solutions*.

About the workbook

This workbook covers the probability and statistics a data scientist or ML engineer uses every week, at the level of a first university course: counting, conditional probability, random variables and their moments, the normal distribution and the central limit theorem, estimation, testing, regression, Bayesian inference and information theory.

Problems are concrete: dice, defect rates, A/B tests, classifiers, sensor noise. Each asks for a number or a formula you can check, and each solution shows every line.

Chapters marked as workbook-only appear nowhere else. The rest also have free versions at eulerpad.app/practice; this edition collects them in teaching order, renumbers their cross-references, and adds a full answer key.

Every numerical claim in every solution is verified by a computer check (finite differences, simulation or exact arithmetic) before the book is built.

© 2026 Spencer Uresk · EulerPad · eulerpad.app
You are welcome to share this free sample unchanged.

Contents of the full workbook

PROBABILITY

1 Counting: permutations, combinations and the binomial theorem

2 Bayes' rule and conditional probability

3 Discrete random variables: expectation, variance and the binomial, geometric and Poisson

4 The negative binomial, hypergeometric and multinomial distributions WORKBOOK-ONLY

5 Continuous random variables: densities, CDFs, expectation and the uniform, exponential and normal

6 Poisson processes and the exponential distribution WORKBOOK-ONLY

7 Joint, marginal and conditional densities WORKBOOK-ONLY

8 Transformations of random variables: the CDF method, the Jacobian and sampling WORKBOOK-ONLY

9 Variance, covariance and correlation

10 Conditional expectation and the law of total variance

11 Moment generating functions: moments, sums and Chernoff bounds WORKBOOK-ONLY

THE NORMAL DISTRIBUTION AND LIMITS

12 The normal distribution and z-scores

13 The multivariate Gaussian: gradients and identities

14 The law of large numbers, Markov's inequality and Chebyshev's inequality

15 Sampling distributions and the central limit theorem: standard errors, confidence intervals and the √N rule

16 Order statistics and Monte Carlo estimation: maxima, minima, importance sampling and control variates WORKBOOK-ONLY

STATISTICAL INFERENCE

17 Maximum likelihood estimation: derivations by hand

18 Hypothesis tests and p-values: z and t tests, power and multiple comparisons

19 Two-sample tests and A/B testing: proportions, Welch's t, power, lift and peeking WORKBOOK-ONLY

20 Bootstrap and permutation tests: resampling for standard errors, intervals and exact p-values WORKBOOK-ONLY

21 Chi-square tests: goodness of fit, independence and the chi-square distribution

22 Simple linear regression: the least-squares line, residuals and R^2

23 Bayesian inference: conjugate priors, posterior means and MAP estimates

PROCESSES AND INFORMATION

24 Markov chains: transition matrices, stationary distributions and absorption IN THIS SAMPLE

25 Entropy, cross-entropy and KL divergence

CHAPTER 24

Markov chains: transition matrices, stationary distributions and absorption

A Markov chain is a random process that forgets its past: where it goes next depends only on where it is now. All of its behavior is stored in one matrix of transition probabilities, and nearly every question about it is a matrix computation. The distribution after n steps is a row vector times P^n, the long-run behavior is the left eigenvector of P for the eigenvalue 1, the speed at which the chain forgets its start is set by the next-largest eigenvalue, and the expected time until something happens comes from one matrix inverse. The same objects sit under reinforcement learning, where a policy turns an environment into a Markov chain, under Markov chain Monte Carlo, whose samplers are designed through detailed balance, and under PageRank. These ten problems compute each of them by hand on small chains.

BEFORE YOU START

- A chain on states $1, \dots, K$ has transition matrix P with $P_{ij} = \Pr(X_{t+1} = j \mid X_t = i)$. Every entry is ≥ 0 and every **row** sums to 1.
- Distributions are **row vectors**: if X_t has distribution π_t, then $\pi_{t+1} = \pi_t P$ and $\pi_n = \pi_0 P^n$. The n-step probability $\Pr(X_n = j \mid X_0 = i)$ is the entry $(P^n)_{ij}$.
- A **stationary distribution** satisfies $\pi P = \pi$ with $\pi_i \geq 0$ and $\sum_i \pi_i = 1$: it is a left eigenvector of P for the eigenvalue 1, that is an eigenvector of $P^\top$, scaled to sum to 1. If every state can reach every other and the chain is aperiodic, π is unique, $\pi_0 P^n \to \pi$ from any start, and the expected time to return to state i is $1/\pi_i$.
- Every eigenvalue of P has $|\lambda| \leq 1$. For a chain that converges, the distance from π shrinks roughly like $|\lambda_2|^n$, where λ_2 is the eigenvalue of next-largest size.
- **Detailed balance.** If $\pi_i P_{ij} = \pi_j P_{ji}$ for all i, j, then π is stationary and the chain is called reversible.
- **Absorbing chains.** Order the states transient first and write $P = \begin{pmatrix} Q & R \\ 0 & I \end{pmatrix}$. The fundamental matrix $N = (I - Q)^{-1}$ has N_{ij} = expected number of visits to transient j starting from i; the row sums $N\mathbf{1}$ are the expected numbers of steps before absorption, and $B = NR$ holds the absorption probabilities.

24.1 •○○ Each day is sunny (S) or rainy (R). After a sunny day the next is sunny with probability 0.9; after a rainy day it is sunny with probability 0.5. Write the transition matrix, and find the probability that it rains two days after a sunny day.

24.2 •○○

A walker on three states moves with

$$P = \begin{pmatrix} 0.5 & 0.5 & 0 \\ 0.25 & 0.5 & 0.25 \\ 0 & 0.5 & 0.5 \end{pmatrix}$$

and starts in state 1. Find its distribution after one, two and three steps.

24.3 ••○

Find the stationary distribution of the chain in Problem 24.2, and compare the distributions after one, two and three steps with it.

24.4 ••○

For the two-state chain $P = \begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix}$ with $0 < a + b < 2$, find the stationary distribution, the eigenvalues of P and a formula for $(P^n)_{12}$. Check it against Problem 24.1.

24.5 ••○

For the weather chain of Problem 24.1, how many days ahead must you look before the probability of rain is within 0.001 of its long-run value $\frac{1}{6}$, starting from a sunny day? Starting from a rainy day?

24.6 ••○

In the weather chain, find the expected number of days from one rainy day to the next, and the expected number of days from a sunny day until the first rainy day. Check the first answer by conditioning on the first step.

24.7 •••

A gambler has 1 chip and bets one chip at a time on fair coin flips, stopping at 0 or 4 chips. With transient states $1, 2, 3$ (the number of chips), compute the fundamental matrix, the expected number of bets, and the probability of reaching 4 chips, from each starting amount.

24.8 •••

Users on a site move between Browse and Cart each minute until they Buy or Leave. From Browse they stay with probability 0.5, go to Cart with 0.3 and leave with 0.2. From Cart they go back to Browse with 0.2, buy with 0.5 and leave with 0.3. For a user who starts browsing, find the probability of a purchase and the expected number of minutes before buying or leaving.

24.9 •••

A random walker on the graph with vertices A, B, C, D and edges AB, BC, CD, DA, AC moves to a uniformly chosen neighbor at each step. Show that $\pi_v = \deg(v)/(2|E|)$ satisfies detailed balance, and find the expected return time to A. Then show that the chain on three states with $P_{12} = P_{23} = P_{31} = \tfrac{2}{3}$ and $P_{21} = P_{32} = P_{13} = \tfrac{1}{3}$ has a uniform stationary distribution but is not reversible.

24.10 •••

Three web pages link as $1 \to 2$, $1 \to 3$, $2 \to 3$ and $3 \to 1$. A surfer follows a uniformly chosen link with probability $d = 0.85$ and otherwise jumps to a uniformly chosen page, so

$$G = dP + \frac{1-d}{3}\mathbf{1}\mathbf{1}^{\top}.$$

Find the PageRank vector $\pi G = \pi$, and compare it with the stationary distribution of the link chain P alone.

CHAPTER 24 · WORKED SOLUTIONS

Markov chains: transition matrices, stationary distributions and absorption

ANSWER KEY

24.1	$P = \begin{pmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{pmatrix}$; the probability of rain two days after a sunny day is 0.14
24.2	$\pi_1 = (\frac{1}{2}, \frac{1}{2}, 0)$, $\pi_2 = (\frac{3}{8}, \frac{1}{2}, \frac{1}{8})$, $\pi_3 = (\frac{5}{16}, \frac{1}{2}, \frac{3}{16})$
24.3	$\pi = (\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$; the distance from it halves each step
24.4	$\pi = \left(\frac{b}{a+b}, \frac{a}{a+b}\right)$, eigenvalues 1 and $1-a-b$, $(P^n)_{12} = \frac{a}{a+b}(1-(1-a-b)^n)$
24.5	6 days from a sunny start, 8 days from a rainy start
24.6	6 days between rainy days; 10 days from a sunny day to the first rain
24.7	From 1 chip: 3 bets on average and probability $\frac{1}{4}$ of reaching 4; from 2: 4 bets and $\frac{1}{2}$; from 3: 3 bets and $\frac{3}{4}$
24.8	$\Pr(\text{Buy}) = \frac{15}{44} \approx 0.341$; expected time $\frac{1+0.3}{0.44} = \frac{65}{22} \approx 2.95$ minutes
24.9	Detailed balance holds with $\pi = (0.3, 0.2, 0.3, 0.2)$ and the return time to A is $\frac{10}{3}$; the rotating chain has $\pi = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ but $\pi_1 P_{12} = \frac{2}{9} \neq \frac{1}{9} = \pi_2 P_{21}$
24.10	$\pi \approx (0.388, 0.215, 0.397)$, exactly $\frac{1}{1769}(686, 380, 703)$; without damping $\pi = (0.4, 0.2, 0.4)$

Problem 24.1

●○○

> Each day is sunny (S) or rainy (R). After a sunny day the next is sunny with probability 0.9; after a rainy day it is sunny with probability 0.5. Write the transition matrix, and find the probability that it rains two days after a sunny day.

1	$P = \begin{pmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{pmatrix}$, states in the order S, R	Row i is the distribution of tomorrow given today is i; each row sums to 1.
2	$\Pr(X_2 = R \mid X_0 = S) = P_{SS}P_{SR} + P_{SR}P_{RR}$	Condition on tomorrow's weather: sunny then rainy, or rainy then rainy.
3	$= 0.9 \cdot 0.1 + 0.1 \cdot 0.5 = 0.09 + 0.05 = 0.14$	The two paths are disjoint, so their probabilities add.
4	$P^2 = \begin{pmatrix} 0.86 & 0.14 \\ 0.70 & 0.30 \end{pmatrix}$	Step 2 is the (S, R) entry of P^2: row S of P times column R. Its rows still sum to 1.
5	$P = \begin{pmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{pmatrix}$; **the probability of rain two days after a sunny day is** 0.14	After a rainy day it is 0.30: the chain still remembers the start after two steps, but less than after one (0.1 against 0.5).

Problem 24.2

●○○

> A walker on three states moves with
>
> $$P = \begin{pmatrix} 0.5 & 0.5 & 0 \\ 0.25 & 0.5 & 0.25 \\ 0 & 0.5 & 0.5 \end{pmatrix}$$
>
> and starts in state 1. Find its distribution after one, two and three steps.

1	$\pi_0 = (1, 0, 0)$	Certain to be in state 1; a row vector.
2	$\pi_1 = \pi_0 P = (0.5, 0.5, 0)$	Multiplying a one-hot row vector by P picks out a row of P.
3	$\pi_2 = 0.5\,(0.5, 0.5, 0) + 0.5\,(0.25, 0.5, 0.25) = (0.375, 0.5, 0.125)$	$\pi_1 P$ is the mixture of the rows of P, weighted by π_1.
4	$\pi_3 = 0.375\,(0.5, 0.5, 0) + 0.5\,(0.25, 0.5, 0.25) + 0.125\,(0, 0.5, 0.5) = (0.3125, 0.5, 0.1875)$	Each entry is still a sum of nonnegative terms, and the total stays 1.
5	$\pi_1 = (\frac{1}{2}, \frac{1}{2}, 0)$, $\pi_2 = (\frac{3}{8}, \frac{1}{2}, \frac{1}{8})$, $\pi_3 = (\frac{5}{16}, \frac{1}{2}, \frac{3}{16})$	Writing P times the column $(1, 0, 0)^\top$ instead gives the first column of P, which is not a distribution (Mistake 1).

Problem 24.3

●●○

> Find the stationary distribution of the chain in Problem 24.2, and compare the distributions after one, two and three steps with it.

1	$\pi_1 = 0.5\pi_1 + 0.25\pi_2$, $\pi_2 = 0.5\pi_1 + 0.5\pi_2 + 0.5\pi_3$, $\pi_3 = 0.25\pi_2 + 0.5\pi_3$	$\pi P = \pi$ entry by entry: column j of P gives the equation for π_j. Here π_i are the entries of π.
2	$\pi_2 = 2\pi_1$ and $\pi_2 = 2\pi_3$	From the first and third equations; the middle one then holds automatically, since one equation of $\pi P = \pi$ is always redundant.
3	$\pi_1 + 2\pi_1 + \pi_1 = 1$, so $\pi_1 = \frac{1}{4}$	The normalization replaces the redundant equation.
4	$\pi_0 P - \pi = (\frac{1}{4}, 0, -\frac{1}{4})$, $\pi_0 P^2 - \pi = (\frac{1}{8}, 0, -\frac{1}{8})$, $\pi_0 P^3 - \pi = (\frac{1}{16}, 0, -\frac{1}{16})$	The distributions of Problem 24.2 minus the stationary one. The gap halves at every step.
5	$\pi = (\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$; **the distance from it halves each step**	The eigenvalues of P are 1, $\frac{1}{2}$ and 0, and the halving is $\lambda_2 = \frac{1}{2}$. The eigenvector that `np.linalg.eig(P)` returns for $\lambda = 1$ is $(1, 1, 1)$, which is not π (Mistake 2).

Problem 24.4

●●○

> For the two-state chain $P = \begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix}$ with $0 < a + b < 2$, find the stationary distribution, the eigenvalues of P and a formula for $(P^n)_{12}$. Check it against Problem 24.1.

1	$\pi_1 a = \pi_2 b$, so $\pi = \left(\dfrac{b}{a+b}, \dfrac{a}{a+b}\right)$	In a two-state chain the flow from 1 to 2 must equal the flow back.
2	$\lambda_1 = 1$ and $\lambda_2 = \operatorname{tr} P - 1 = 1 - a - b$	Every stochastic matrix has $P\mathbf{1} = \mathbf{1}$, so 1 is an eigenvalue, and the two eigenvalues add to the trace $2 - a - b$.
3	$P^n = \dfrac{1}{a+b}\begin{pmatrix} b & a \\ b & a \end{pmatrix} + \dfrac{\lambda_2^n}{a+b}\begin{pmatrix} a & -a \\ -b & b \end{pmatrix}$	The first matrix has both rows equal to π; the second is the part that decays. At $n = 0$ the sum is I, and multiplying by P multiplies the second part by λ_2.
4	$(P^n)_{12} = \dfrac{a}{a+b}\left(1 - \lambda_2^n\right)$	The $(1,2)$ entry of step 3.
5	$\pi = \left(\frac{b}{a+b}, \frac{a}{a+b}\right)$, **eigenvalues** 1 **and** $1-a-b$, $(P^n)_{12} = \frac{a}{a+b}\left(1 - (1-a-b)^n\right)$	Problem 24.1 has $a = 0.1$, $b = 0.5$: $\pi = (\frac{5}{6}, \frac{1}{6})$, $\lambda_2 = 0.4$, and $(P^2)_{SR} = \frac{1}{6}(1 - 0.16) = 0.14$. At $a = b = 1$ the chain alternates, $\lambda_2 = -1$, and P^n never settles (Mistake 3).

Problem 24.5

●●○

> For the weather chain of Problem 24.1, how many days ahead must you look before the probability of rain is within 0.001 of its long-run value $\frac{1}{6}$, starting from a sunny day? Starting from a rainy day?

1	From sunny: $(P^n)_{SR} - \frac{1}{6} = -\frac{1}{6}(0.4)^n$	Problem 24.4, step 4, with $\lambda_2 = 0.4$.
2	$\frac{1}{6}(0.4)^n < 0.001$ needs $0.4^n < 0.006$, so $n > \dfrac{\ln 0.006}{\ln 0.4} \approx 5.58$	Dividing by $\ln 0.4 < 0$ flips the inequality.
3	From rainy: $(P^n)_{RR} - \frac{1}{6} = \frac{5}{6}(0.4)^n$	The $(2,2)$ entry of Problem 24.4, step 3: $\frac{a}{a+b} + \frac{b}{a+b}\lambda_2^n$.
4	$\frac{5}{6}(0.4)^n < 0.001$ needs $0.4^n < 0.0012$, so $n > \dfrac{\ln 0.0012}{\ln 0.4} \approx 7.34$	A larger starting gap, the same rate.
5	**6 days from a sunny start, 8 days from a rainy start**	Check: $\frac{1}{6}(0.4)^6 \approx 0.00068$ and $\frac{5}{6}(0.4)^8 \approx 0.00055$, while $\frac{1}{6}(0.4)^5 \approx 0.0017$ and $\frac{5}{6}(0.4)^7 \approx 0.0014$ are still too big. Each day multiplies the gap by $\lvert\lambda_2\rvert = 0.4$.

Problem 24.6

●●○

> In the weather chain, find the expected number of days from one rainy day to the next, and the expected number of days from a sunny day until the first rainy day. Check the first answer by conditioning on the first

step.

1	Return time to R $= 1/\pi_R = 6$ days	The chain spends a fraction $\frac{1}{6}$ of its days in R, so visits are 6 days apart on average.
2	From S, each day is rainy with probability 0.1, independently, until the first one: $h_{S\to R} = 1/0.1 = 10$	The number of steps until the first success is geometric with mean $1/p$.
3	Return time to R $= 1 + 0.5 \cdot 0 + 0.5 \cdot h_{S\to R}$	One step, then either it rains again at once (probability 0.5) or the chain is in S and must still reach R.
4	$= 1 + 0.5 \cdot 10 = 6$	Agrees with step 1.
5	**6 days between rainy days; 10 days from a sunny day to the first rain**	$1/\pi_R$ is a return time, not the time to get there from somewhere else: using 6 for the trip from S is Mistake 4. Likewise the return time to S is $1/\pi_S = \frac{6}{5} = 1 + 0.1 \cdot 2$.

Problem 24.7

A gambler has 1 chip and bets one chip at a time on fair coin flips, stopping at 0 or 4 chips. With transient states $1, 2, 3$ (the number of chips), compute the fundamental matrix, the expected number of bets, and the probability of reaching 4 chips, from each starting amount.

1	$Q = \begin{pmatrix} 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & 0 \end{pmatrix}, R = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & 0 \\ 0 & \frac{1}{2} \end{pmatrix}$	Rows of Q are moves among $1, 2, 3$ chips; the columns of R are absorption at 0 and at 4.
2	$N = (I - Q)^{-1} = \begin{pmatrix} 1.5 & 1 & 0.5 \\ 1 & 2 & 1 \\ 0.5 & 1 & 1.5 \end{pmatrix}$	Check: $(I - Q)N = I$; for example row 1 of $I - Q$ is $(1, -\frac{1}{2}, 0)$, and with the first column of N it gives $1.5 - 0.5 = 1$.
3	$N\mathbf{1} = (3, 4, 3)$	Expected number of bets. It matches the formula $k(4 - k)$ for a fair game.
4	$B = NR = \begin{pmatrix} 0.75 & 0.25 \\ 0.5 & 0.5 \\ 0.25 & 0.75 \end{pmatrix}$	Row k is (ruin, reach 4) from k chips; each row sums to 1.
5	**From 1 chip: 3 bets on average and probability $\frac{1}{4}$ of reaching 4; from 2: 4 bets and $\frac{1}{2}$; from 3: 3 bets and $\frac{3}{4}$**	A fair game keeps the expected fortune constant: $\frac{1}{4} \cdot 4 = 1$. Three straight wins, $\frac{1}{8}$, is only one of the ways to get there (Mistake 5).

Problem 24.8

Users on a site move between Browse and Cart each minute until they Buy or Leave. From Browse they stay with probability 0.5, go to Cart with 0.3 and leave with 0.2. From Cart they go back to Browse with 0.2, buy

with 0.5 and leave with 0.3. For a user who starts browsing, find the probability of a purchase and the expected number of minutes before buying or leaving.

1	$Q = \begin{pmatrix} 0.5 & 0.3 \\ 0.2 & 0 \end{pmatrix}, R = \begin{pmatrix} 0 & 0.2 \\ 0.5 & 0.3 \end{pmatrix}$	Transient states Browse, Cart; absorbing states Buy, Leave. Each row of $(Q \mid R)$ sums to 1.
2	$I - Q = \begin{pmatrix} 0.5 & -0.3 \\ -0.2 & 1 \end{pmatrix}$, $\det = 0.5 - 0.06 = 0.44$	A 2×2 inverse needs the determinant.
3	$N = \dfrac{1}{0.44}\begin{pmatrix} 1 & 0.3 \\ 0.2 & 0.5 \end{pmatrix} \approx \begin{pmatrix} 2.273 & 0.682 \\ 0.455 & 1.136 \end{pmatrix}$	Swap the diagonal, negate the off-diagonal, divide by the determinant.
4	$\Pr(\text{Buy}) = N_{11}R_{11} + N_{12}R_{21} = 0 + \dfrac{0.3}{0.44} \cdot 0.5 = \dfrac{15}{44}$	First column of $B = NR$. Only Cart leads to Buy, so the answer is the expected number of Cart visits times 0.5.
5	$\Pr(\text{Buy}) = \frac{15}{44} \approx 0.341$; **expected time** $\frac{1+0.3}{0.44} = \frac{65}{22} \approx 2.95$ **minutes**	The leave probability is $\frac{29}{44} \approx 0.659$, and the two add to 1. A user who has reached the cart buys with probability $\frac{25}{44} \approx 0.568$, which is the number to optimize.

Problem 24.9

•••

A random walker on the graph with vertices A, B, C, D and edges AB, BC, CD, DA, AC moves to a uniformly chosen neighbor at each step. Show that $\pi_v = \deg(v)/(2|E|)$ satisfies detailed balance, and find the expected return time to A. Then show that the chain on three states with $P_{12} = P_{23} = P_{31} = \frac{2}{3}$ and $P_{21} = P_{32} = P_{13} = \frac{1}{3}$ has a uniform stationary distribution but is not reversible.

1	Degrees A: 3, B: 2, C: 3, D: 2, and $2\|E\| = 10$, so the candidate is $\pi = (0.3, 0.2, 0.3, 0.2)$	Each edge adds 1 to two degrees.
2	For an edge uv: $\pi_u P_{uv} = \dfrac{\deg u}{10} \cdot \dfrac{1}{\deg u} = \dfrac{1}{10} = \pi_v P_{vu}$	The degree cancels; for non-neighbors both sides are 0.
3	Detailed balance implies $\pi P = \pi$	Sum $\pi_i P_{ij} = \pi_j P_{ji}$ over i: the right side gives $\pi_j \sum_i P_{ji} = \pi_j$.
4	Return time to $A = 1/0.3 = \frac{10}{3}$	The busier vertices are visited more often.
5	**Detailed balance holds with** $\pi = (0.3, 0.2, 0.3, 0.2)$ **and the return time to** A **is** $\frac{10}{3}$**; the rotating chain has** $\pi = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ **but** $\pi_1 P_{12} = \frac{2}{9} \neq \frac{1}{9} = \pi_2 P_{21}$	Its columns also sum to 1, so the uniform π is stationary, yet probability circulates $1 \to 2 \to 3 \to 1$ on net. Detailed balance is sufficient for stationarity, not necessary; MCMC samplers are built to satisfy it because it is easy to check one pair at a time.

Problem 24.10

•••

Three web pages link as $1 \to 2$, $1 \to 3$, $2 \to 3$ and $3 \to 1$. A surfer follows a uniformly chosen link with probability $d = 0.85$ and otherwise jumps to a uniformly chosen page, so

$$G = dP + \frac{1-d}{3}\mathbf{1}\mathbf{1}^\top.$$

Find the PageRank vector $\pi G = \pi$, and compare it with the stationary distribution of the link chain P alone.

1	$P = \begin{pmatrix} 0 & 0.5 & 0.5 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$	Page 1 splits its probability over two links.
2	$\pi_j = 0.05 + 0.85 \sum_i \pi_i P_{ij}$	$\pi \frac{1-d}{3}\mathbf{1}\mathbf{1}^\top = 0.05\,\mathbf{1}^\top$ because $\sum_i \pi_i = 1$.
3	$\pi_1 = 0.05 + 0.85\pi_3$, $\pi_2 = 0.05 + 0.425\pi_1$, $\pi_3 = 0.05 + 0.85(0.5\pi_1 + \pi_2)$	Column by column.
4	Substitute π_2 and $\pi_3 = (\pi_1 - 0.05)/0.85$ into $\pi_1 + \pi_2 + \pi_3 = 1$: $2.21125\,\pi_1 = 0.8575$	Multiply through by 0.85 first.
5	$\pi \approx (0.388, 0.215, 0.397)$, **exactly** $\frac{1}{1769}(686, 380, 703)$; **without damping** $\pi = (0.4, 0.2, 0.4)$	Page 3 ranks first because it collects links from both other pages. The jumps pull every score toward $\frac{1}{3}$, and they make the chain converge at the rate $0.85 \cdot \lvert\lambda_2(P)\rvert = 0.85 \cdot 0.707 \approx 0.601$ per step, since P's other eigenvalues are $-\frac{1}{2} \pm \frac{1}{2}i$.

Where this goes wrong

1. Multiplying P by a column vector

Most linear algebra applies a matrix to a column vector, and the start distribution gets written as one.

1	$\pi_0 = (1, 0)$, sunny today, and $P = \begin{pmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{pmatrix}$	Right so far: Problem 24.1.
2	"Tomorrow's distribution is $P\pi_0^\top$."	With rows indexed by today, the distribution must multiply from the left: $\pi_1 = \pi_0 P$.
3	$P\pi_0^\top = (0.9, 0.5)^\top$ ✗	Its entries add to 1.4, so it is not a distribution: it is the first column of P, the probability of sun tomorrow from each state today. The correct $\pi_1 = (0.9, 0.1)$.

2. The stationary distribution read off a right eigenvector

`np.linalg.eig(P)` returns eigenvectors, and the one for $\lambda = 1$ looks like the answer.

1	P from Problem 24.2 has eigenvalue 1	Right so far: every stochastic matrix does.
2	"Take the eigenvector of P for $\lambda = 1$ and normalize it."	That solves $Pv = v$; the stationary distribution solves $\pi P = \pi$, an eigenvector of $P^\top$.

3	$\pi = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ ✗	$P\mathbf{1} = \mathbf{1}$ for every stochastic matrix, so this is always the uniform vector. Check: $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})P = (0.25, 0.5, 0.25) \neq (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. Use `np.linalg.eig(P.T)`, which gives $\pi = (\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$.

3. Expecting every chain to converge to π

A stationary distribution exists, so the chain is assumed to approach it.

1	$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ has $\pi = (\frac{1}{2}, \frac{1}{2})$	Right so far: $\pi P = \pi$, and it is Problem 24.4 with $a = b = 1$.
2	"So $\pi_0 P^n \to (\frac{1}{2}, \frac{1}{2})$ from $\pi_0 = (1, 0)$."	Convergence needs the chain to be aperiodic as well; this one returns to each state only at even times.
3	$\pi_{100} \approx (\frac{1}{2}, \frac{1}{2})$ ✗	π_n alternates between $(0, 1)$ and $(1, 0)$, so $\pi_{100} = (1, 0)$. The eigenvalues are 1 and -1, and $\lvert\lambda_2\rvert = 1$ means the gap never shrinks. Only the long-run average of the π_n is $(\frac{1}{2}, \frac{1}{2})$. Adding a self-loop, as PageRank's jumps do, removes the period.

4. Using 1/π_j as the time to reach j

The return time $1/\pi_j$ is the one expected-time formula that needs no linear algebra.

1	$\pi_R = \frac{1}{6}$ in the weather chain	Right so far: Problem 24.4.
2	"Starting from a sunny day, rain comes after $1/\pi_R$ days on average."	$1/\pi_R$ is the mean time from R back to R. From S the chain must first leave S, which it does slowly.
3	$h_{S \to R} = 6$ days ✗	From S the wait is geometric with success probability 0.1, so $h_{S \to R} = 10$. The return time 6 averages that with the immediate repeat: $1 + 0.5 \cdot 10 = 6$ (Problem 24.6).

5. Multiplying the win probabilities in gambler's ruin

Reaching 4 chips from 1 takes three more wins than losses, and three wins in a row is the obvious way.

1	Each bet is won with probability $\frac{1}{2}$, independently	Right so far: Problem 24.7.
2	"Reaching 4 chips from 1 needs three wins, so the probability is $(\frac{1}{2})^3$."	That counts only the path win–win–win. A loss followed by two wins and so on can still reach 4, as long as the gambler never hits 0 first.

3	$\Pr(\text{reach } 4 \mid \text{start at } 1) = \frac{1}{8}$ ✗	The fundamental matrix gives $\frac{1}{4}$: the first row of N is $(1.5, 1, 0.5)$, and $\Pr = N_{13} \cdot \frac{1}{2} = 0.25$. Fairness confirms it: the expected final fortune $4 \cdot \frac{1}{4} = 1$ equals the starting 1 chip, while $4 \cdot \frac{1}{8} = 0.5$ would lose chips on average in a fair game.

THE FULL WORKBOOK

Probability & Statistics for Machine Learning

250 problems from counting and Bayes' rule to hypothesis tests, MLE, Markov chains and KL divergence, with every step worked

- 250 problems in 25 chapters: probability, the normal distribution and limits, statistical inference and processes and information
- 8 chapters written only for the workbook
- A worked solution for every problem, one step per line with the reason beside it
- An answer key per chapter, and the common mistakes for every topic
- Two PDFs, *Problems* and *Worked Solutions*, US Letter, ready to print

Get it at eulerpad.app/workbooks/