Forward Kolmogorov Eq.

$x_n = P^n x_0$.

Know the initial probability at time 0, we can find the distribution at any later time using powers of the matrix P.

Steady state vector

A vector x for which $Px = x$ is called a steady state of the Markov chain P.

Every Markov chain has at least one steady-state.

Strongly-connected

If in a Markov chain P, every state is reachable from every other state, we call the graph strongly-connected.

In this case, the Markov chain is irreducible, and it has only one steady-state.

Periodic

A Markov chain is not periodic if there is some power of P, say P^k, such that all entries in P^k are positive.

Primitive

A Markov chain P has some power k such that all entries in P^k are positive, then the chain is said to be Primitive.

A primitive (i.e. irreducible aperiodic) Markov chain has a single steady-state to which it always converges.

A primitive Markov chain is ergodic.

Ergodicity: the property of having identical time average and ensemble average.

Detailed Balance in MCMC

$\pi_i P_{ji} = \pi_j P_{ij}$ There are n^2 of these equations → detailed balance equations.

If a system is in detailed balance, it is in steady-state.

Existence of a Unique Steady State.

- A finite Markov Chain is guaranteed to have a unique steady state if it is both irreducible and aperiodic.
- If a chain is irreducible and aperiodic, it converges to this steady state regardless of the initial distribution.

Regular Markov Chain.

- A Markov Chain is regular if some power of its transition Matrix, P^n, has all positive entries. This implies irreducibility and usually suggests convergence to a unique steady state.

Bayes's Theorem

$$P(A|B) = \frac{P(B|A)\cdot P(A)}{P(B)}$$

Law of Total Probability.

$$P(B) = P(B|A_1)\cdot P(A_1) + P(B|A_2)\cdot P(A_2) + \dots + P(B|A_n)\cdot P(A_n)$$

Transition Matrix

Column sum to 1.

"Column-stochastic approach".

S0 → S0, S1 → S0, S2 → S0

S0 → S1, S1 → S1, S2 → S1

S0 → S2, S1 → S2, S2 → S2

To be primitive:

- Irreducible: Every state can be reached from any other state, not necessarily in one step, but in a finite number of steps. This means that the chain is a single communicating class.
- Aperiodic: Every state has a period of 1. The period of a state is defined as the greatest common divisor (GCD) of the steps in which it is possible to return to the state. An aperiodic state is one where GCD is 1.

This means that it's possible to return to the state at irregular intervals, not in multiples of some number greater than 1.

A primitive Markov Chain is one where, after some number of steps, there is a positive probability of being in any state regardless of the initial state, which implies that the chain is both irreducible and aperiodic. This ensures that the Markov chain has a unique stationary distribution to which it converges regardless of the initial state distribution.

MAP Estimator and MMSE Estimator

The MAP estimator for a Beta distribution is the mode of the distribution. For a Beta (α, β) distribution, the mode (when $\alpha > 1$ and $\beta > 1$) is:

$$\text{Mode} = \frac{\alpha - 1}{\alpha + \beta - 2}$$

MMSE: for a Beta distribution is its mean. The mean of a Beta (α, β) distribution is given by:

$$\text{Mean} = \frac{\alpha}{\alpha+\beta} = \text{MMSE}.$$

Steady State and Detailed Balance.

With a transition matrix and state distribution, determine if the state is a steady state.

$$P = \begin{bmatrix} 0.2 & 0.4 & 0.1 \\ 0.5 & 0.1 & 0.9 \\ 0.3 & 0.5 & 0 \end{bmatrix}$$

State distribution $x = (0.2, 0.5, 0.3)$

Check if $x \cdot P = x$

Different? → Not a steady state.

With a transition matrix and a steady state, determine if the steady state satisfies the detailed balance condition.

Check if the product πP return the same distribution π.

Metropolis-Hastings Algorithm.

- Any transition from a state with zero probability will be accepted if it moves to a state with non-zero probability.

Considering sampling a posterior distribution using MCMC.

Assume the target posterior is the uniform distribution on $[-2, 2]$ and 0 elsewhere. i.e. $p_T(x) = \frac{1}{4}$ for $-2 \le x \le 2$

$p_T(x) = 0$ for $x < -2$, $x > 2$.

Assume the candidate distribution is uniform on $[-4, 4]$. i.e. $p_c(x) = \frac{1}{8}$ for $-4 \le x \le 4$.

The Metropolis-Hastings rule is to accept a proposed transition from j to i with probability:

$$\alpha_{ij} = \min\left(1, \frac{\pi_i H_{ji}}{\pi_j H_{ij}}\right).$$

Consider a transition from $x = \frac{1}{2}$ to $x = -1$. What is the probability this transition will be accepted.

- $\pi_j = p_T(\frac{1}{2}) = \frac{1}{4}$, within range $[-2, 2]$
- $\pi_i = p_T(-1) = \frac{1}{4}$, within range $[-2, 2]$.
- $H_{ij} = p_c(\frac{1}{2}) = \frac{1}{8}$, within range $[-4, 4]$.
- $H_{ji} = p_c(-1) = \frac{1}{8}$, within range $[-4, 4]$.

$$\alpha_{ij} = \min\left(1, \frac{\frac{1}{4}\cdot\frac{1}{8}}{\frac{1}{4}\cdot\frac{1}{8}}\right).$$

∴ Accepted with a probability of 1.

Bayes's Factor Analysis.

A Bayes Factor (BF) measures the strength of evidence for one hypothesis over another. It's the ratio of the likelihoods for two hypotheses.

A BF > 1 indicates evidence in favor of the first hypothesis, a BF < 1 indicates evidence in favor of the second hypothesis, and BF = 1 indicates the data does not favor one hypothesis over the other.

Heads: 60, tails = 40.

Calculate the likelihood of the data under two hypotheses.

- The coin is fair: Heads and tails both have a probability of 0.5.
- The coin is biased towards Heads with a probability of 0.6.

Compute the Bayes Factor: Fair vs. Not.

Binomial distribution:

$$L(H_0) = \binom{100}{60}(0.5)^{60}(0.5)^{40}$$

$$L(H_1) = \binom{100}{60}(0.6)^{60}(0.4)^{40}$$

$$BF = \frac{L(H_0)}{L(H_1)} = \frac{(0.5)^{100}}{(0.6)^{60}(0.4)^{40}} = 0.133$$

The posterior odds are given by the prior odds times the BF.

$$\frac{P(H_1|\text{Data})}{P(H_0|\text{Data})} = \underbrace{\frac{P(H_1)}{P(H_0)}}_{=1} \cdot BF.$$

Prior odds are 1

Normalize odds to find the posterior probability.

$$P(H_1|\text{Data}) = \frac{BF}{BF+1}$$

$$P(H_0|\text{Data}) = \frac{1}{BF+1}$$

Markov Chain and # of steady-state.

- Every Markov Chain has at least one steady-state (a basic property of stochastic matrices).
- If a Markov Chain is irreducible (every node is reachable from every other node), then it has exactly one.

Problem B Type

Given three dice, a six-sided die, a ten-sided die, and twenty-sided die. Choose it at random, and observe that we get a nine (9).

Denote H_0 the hypothesis that we rolled the six-sided die. H_1 the hypothesis that we rolled the ten-sided die, and H_2 the hypothesis that we rolled the twenty-sided die.

hyp	$P(H)$	$P(D\mid H)$	$P(H)\cdot P(D\mid H)$	$P(H\mid D)$
H_0	$\frac{1}{3}$	0	0	0
H_1	$\frac{1}{3}$	$\frac{1}{10}$	$\frac{1}{30}$	$\frac{2}{3}$
H_2	$\frac{1}{3}$	$\frac{1}{20}$	$\frac{1}{60}$	$\frac{1}{3}$

$P(H\mid D)$: Probability of a hypothesis of given data D.

$P(D) = P(D\mid H_0)\cdot P(H_0) + P(D\mid H_1)\cdot P(H_1) + P(D\mid H_2)\cdot P(H_2)$

$$P(H\mid D) = \frac{P(D\mid H)\cdot P(H)}{P(D)} = 0.5$$

Problem C Type

Every exact pair of rolls has $\frac{1}{16}$ chance of occurring.

Count how many ways each outcome occurs.

Outcome	Number of Ways	Probability.
1	0	0
2	1 way (1,1)	$\frac{1}{16}$
3	2 ways (1,2 or 2,1)	$\frac{1}{16}\cdot 2 = \frac{2}{16}$
4	3 ways (1,3 or 2,2 or 3,1)	$\frac{1}{16}\cdot 3 = \frac{3}{16}$
5	4 ways (1,4 or 2,3 or 3,2 or 4,1)	$\frac{4}{16}$
6	3 ways (2,4 or 3,3 or 4,2)	$\frac{3}{16}$
7	2 ways (3,4 or 4,3)	$\frac{2}{16}$
8	1 ways (4,4)	$\frac{1}{16}$

Calculate the probability of superiority of of the sum of two four-sided over a single 6-sided dice.

- An outcome of 8 is always superior, add $\frac{1}{16}$ to our $P(S)$.
- 7 can always be superior, add $\frac{1}{8}$.
- 6 is superior to all rolls other than 6, so we can add $\frac{3}{16}\cdot\frac{5}{6}$
- 5 is superior to all rolls other than 5 or 6, so we can add $\frac{4}{16}\cdot\frac{4}{6}$
- And so on.

This gives us: $P(S) = \frac{1}{16} + \frac{1}{8} + \frac{3}{16}\cdot\frac{5}{6} + \frac{1}{4}\cdot\frac{4}{6} + \frac{3}{16}\cdot\frac{3}{6} + \frac{1}{8}\cdot\frac{2}{6} + \frac{1}{16}\cdot\frac{1}{6} = \frac{21}{32}$.

If the Markov Chain that have unique (only-one), check if them converge to that steady state regardless of the starting state?

1. Check if it is a Regular Matrix.
   - A Markov transition matrix is said to be regular if some power of the matrix, P^n (where n is a positive integer), has all positive entries. If a matrix is regular, it means that the Markov Chain is both irreducible and aperiodic.
   - P^2 test: if P^2 has no zero entries, it is a strong indication that the matrix is regular, and thus the Markov Chain is both irreducible and aperiodic.
   - Conversely, a Markov chain can be regular even if P^2 has zero entries, might require looking at a higher power of P to find all positive entries.

Steady-State Distribution.

- Normalization equation: The sum of probabilities must be 1:

$$\pi_1 + \pi_2 + \pi_3 = 1$$

Check steady-state example.

For the Markov Chain, assume that the chain is in state:

$\frac{6}{11}$, $\frac{3}{11}$, $\frac{2}{11}$ is the chain in steady-state?

Check $P\pi = \pi$.

Does the detailed balance hold?

Check $\pi_i P_{ji} = \pi_j P_{ij}$, for all i, j.

we have our transition matrix:

$$\begin{matrix} 0.8 & 0.2 & 0.3 \\ 0.1 & 0.6 & 0.3 \\ 0.1 & 0.2 & 0.4 \end{matrix}$$

For State 0 and State 1:

- $\pi(0)\cdot P(0\to 1) = \frac{6}{11}\cdot 0.1$
- $\pi(1)\cdot P(1\to 0) = \frac{3}{11}\cdot 0.2$

For State 0 and State 2:

- $\pi(0)\cdot P(0\to 2) = \frac{6}{11}\cdot 0.1$
- $\pi(2)\cdot P(2\to 0) = \frac{2}{11}\cdot 0.3$

For State 1 and State 2:

- $\pi(1)\cdot P(1\to 2) = \frac{3}{11}\cdot 0.2$
- $\pi(2)\cdot P(2\to 1) = \frac{2}{11}\cdot 0.3$

Check: The detailed balance condition holds if these pairs of products are equal for each case.

$\frac{6}{11}\cdot 0.1 = \frac{3}{11}\cdot 0.2$

DB holds for State 0 and 1.

$\frac{6}{11}\cdot 0.1 = \frac{2}{11}\cdot 0.3$

DB holds for State 0 and 2.

$\frac{3}{11}\cdot 0.2 = \frac{2}{11}\cdot 0.3$

DB holds for State 1 and 2.

Since the DB holds for all pairs of states, the Markov Chain is reversible.

Bayesian Updates with Coin Flips.

Uniform prior. (each value of p is equally likely.)

Compute posterior probabilities for $p = 0.1$, 0.5 and 0.7.

Calculate the likelihood of observing 7 heads in 10 flips

Likelihood $= \binom{10}{7}\cdot p^7\cdot(1-p)^3$

p	Prior	Likelihood	Prior · Likelihood	Posterior.
0.3	$\frac{1}{3}$	$\binom{10}{7}\cdot(0.3)^7\cdot(1-0.3)^3$		
0.5	$\frac{1}{3}$	$\binom{10}{7}\cdot(0.5)^7\cdot(1-0.5)^3$		
0.7	$\frac{1}{3}$	$\binom{10}{7}\cdot(0.7)^7\cdot(1-0.7)^3$		

Imagine we have two dice.

- A 6-sided die and a 4-sided die.

Calculate the distribution of outcomes.

75% chance for 6 die and 25% for 4 die. Mixed distribution of outcomes.

For $x=1$: $P(1) = \frac{1}{6}\cdot 0.75 + \frac{1}{4}\cdot 0.25$

$P(2) = \frac{1}{6}\cdot 0.75 + \frac{1}{4}\cdot 0.25$

$P(3) = \frac{1}{6}\times 0.75 + \frac{1}{4}\times 0.25$

$P(4) = \frac{1}{6}\cdot 0.75 + \frac{1}{4}\times 0.25$

$P(5) = P(6) = \frac{1}{6}\cdot 0.75 + 0\cdot 0.25$.

Calculate the probability of superiority of this mixed distribution over a single roll of a 4-sided die.

For a 4-sided die roll of 1:

$P(\text{Mixed outcome} > 1) = P(2) + P(3) + P(4) + P(5) + P(6)$ ....

Finally, average the probs calculated above, weighted by the probability of each outcome on the 4-sided side (which is $\frac{1}{4}$).

Absorbing Markov Chain

- This kind of Markov Chain, where at least one state is an absorbing state (a state that, once entered, cannot be left), have a steady-state solution.