

Summary: Absorbing Markov Chain Analysis

This problem illustrates the standard workflow for analyzing a finite Markov chain with transient states and multiple closed classes. The key steps and concepts are summarized below so the method can be applied to similar problems.

(a) Canonical Form of the Transition Matrix

The first step is to classify states as:

- **Transient states:** states that may be left permanently.
- **Closed (absorbing) classes:** once entered, the chain never leaves.

After reordering states so that all transient states come first, the transition matrix takes the **canonical form**:

$$P = \begin{pmatrix} Q & R \\ 0 & S \end{pmatrix},$$

where:

- Q contains transitions among transient states,
- R contains transitions from transient states to closed states,
- S contains transitions within closed classes.

This reordering does not change the chain, but exposes its long-term structure.

(b) Fundamental Matrix

The **fundamental matrix** is defined as

$$F = (I - Q)^{-1}.$$

Its interpretation is central:

F_{ij} is the expected number of visits to transient state j given the chain starts in transient state i .

Equivalently,

$$F = I + Q + Q^2 + Q^3 + \cdots,$$

which sums expected visits over all time before absorption.

(c) Expected Time to Absorption

The expected number of steps before entering a closed class is obtained by summing the entries of the corresponding row of F :

$$\mathbb{E}_i[T_{\text{abs}}] = \sum_j F_{ij}.$$

Each visit to a transient state corresponds to one unit of time, so row sums represent total expected time spent in transient states.

(d) Absorption Probabilities

Define

$$B = FR.$$

Then:

B_{ij} is the probability that the chain is eventually absorbed in closed state (or class) j , given it started in transient state i .

This formula follows from linearity of expectation:

- F_{ik} counts how many times state k is visited on average,
- R_{kj} gives the probability of exiting from k to j ,
- Multiplying and summing accounts for all possible exit times.

(e) Limiting Distribution

If the chain has multiple closed classes, there is **no unique global limiting distribution**. Instead, the limit depends on the starting state.

The limiting distribution is constructed as follows:

1. Transient states receive probability 0 in the limit.
2. Absorption probabilities (from B) determine which closed class is entered.
3. Within each closed class, the chain converges to that class's **stationary distribution**.

The final limiting distribution is a weighted mixture of the stationary distributions of the closed classes, with weights given by absorption probabilities.

Key Takeaway

Absorbing Markov chain analysis separates behavior into three phases: (1) movement among transient states, (2) probabilities of entering closed classes, and (3) long-run behavior within those classes.

This framework applies broadly to Markov chains with transient behavior followed by absorption.

Estimating and Smoothing Markov Transition Probabilities

Consider a time-homogeneous Markov chain with finite state space

$$\mathcal{S} = \{1, 2, \dots, K\},$$

and let n_{ij} denote the observed number of transitions from state i to state j . Define the row totals

$$n_{i\cdot} = \sum_{j=1}^K n_{ij}.$$

Maximum Likelihood Estimation (MLE)

Conditioning on being in state i , the next-state transitions follow a multinomial distribution. The maximum likelihood estimator of the transition probabilities is obtained by row normalization:

$$\hat{p}_{ij} = \Pr(X_{n+1} = j \mid X_n = i) = \frac{n_{ij}}{n_{i\cdot}}, \quad j = 1, \dots, K.$$

This estimator is simple and unbiased asymptotically, but can be unstable when $n_{i\cdot}$ is small.

Limitations of the MLE

If a particular state i is rarely observed (small $n_{i\cdot}$), then:

- Estimates have high variance.
- Zero counts produce zero probability estimates, even if the true transition probability is nonzero.

These issues motivate smoothing or regularization.

Laplace (Add-One) Smoothing

Laplace smoothing adjusts the counts by adding a constant (typically 1) to each possible transition:

$$\tilde{n}_{ij} = n_{ij} + 1.$$

The smoothed transition probabilities are then:

$$\tilde{p}_{ij} = \frac{n_{ij} + 1}{n_{i\cdot} + K}.$$

Laplace smoothing guarantees strictly positive probabilities but does not incorporate external structural information.

Dirichlet Smoothing with an Empirical Bayes Prior

A more principled approach places a Dirichlet prior on each transition row:

$$(p_{i1}, \dots, p_{iK}) \sim \text{Dirichlet}(\alpha_1, \dots, \alpha_K).$$

Let the prior mean be $m = (m_1, \dots, m_K)$ with $\sum_j m_j = 1$, and let $\alpha > 0$ denote the prior strength. Then:

$$\alpha_j = \alpha m_j.$$

After observing transition counts n_{ij} , the posterior distribution is:

$$(p_{i1}, \dots, p_{iK}) \mid \text{data} \sim \text{Dirichlet}(\alpha_1 + n_{i1}, \dots, \alpha_K + n_{iK}).$$

The posterior mean (used as a smoothed estimator) is:

$$\tilde{p}_{ij} = \frac{\alpha m_j + n_{ij}}{\alpha + n_i}.$$

Interpretation

- α controls the influence of the prior relative to the data.
- The estimator is a convex combination of the prior mean and the MLE.
- Empirical Bayes chooses m from pooled or related data, allowing information sharing across states.