

# Assignment 2: Discrete Time Discrete State Markov Chains

---

## Instructions

Complete the following problem set showing your work. Problems may be worked out "by hand" or in "python" or with the assistance of other analytical software (e.g., Mathematica, MatLab). You may use chatGPT to assist in coding.

Solutions must be type written (e.g., in Jupyter, markdown, or latex). Submit solutions as a PDF to Rebekah Hall (rah11@sfu.ca) by **11:59pm** on the **Saturday** of the corresponding week (see syllabus).

All problems are equally weighted within an assignment. Students in 468 may or may not choose to attempt the challenge question for a bonus 5pts. Students in 795 are required to complete the challenge question.

## Problem Set

### 1. Modelling Population Dynamics in an Ecological System

In population ecology, Markov chains are commonly used to model changes in the composition of populations over time. Consider a simplified ecological system with two competing species, Species A and Species B. Consider a markov chain modelling which species occupies the ecological survey site in a given year. The transition probabilities for the site is given by transition probability matrix:

$$P = \begin{bmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \end{bmatrix}$$

This matrix represents the transition probabilities for a one-year time step. For example,  $P[1, 1] = 0.8$  means that there is an 80% that the survey site will be dominated by individuals from Species A in two consecutive years. Given that the site currently is occupied by species A:

**Part A:** Calculate the probability that the population will consist of Species A after one year. How about two years?

**Part B:** Using the Chapman-Kolmogorov equation, calculate the probability that the population will consist of Species A after three years.

**Part C:** What is the stationary distribution of this system?

**Part D:** Interpret your answers in terms of ecosystem stability.

### 2. Modelling Species Occupancy in an Ecological Network

In the field of ecology, species occupancy in a network of interconnected habitats is a critical aspect to understand. Consider an ecological network consisting of three interconnected habitats: Habitat A, Habitat B, and Habitat C where  $A \leftrightarrow B \leftrightarrow C \leftrightarrow A$ . Ecologists want to model the occupancy dynamics of a particular species, "Species X," across these habitats over time using a discrete time discrete state stochastic process.

State Space:

- State 0: Species X is absent from all habitats.
- State 1: Species X is present in Habitat A
- State 2: Species X is present in Habitat B.
- State 3: Species X is present in Habitat C.
- State 4: Species X is present in Habitats A&B
- State 5: Species X is present in Habitats A&C.
- State 6: Species X is present in Habitats B&C.

- State 7: Species X is present in All Habitats.

*Transition Probabilities:*

- Rule 1: The probability of colonizing an adjacent habitat is 0.2.
- Rule 2: The probability of extinction from a habitat is 0.15.
- Rule 3: The probability of recolonization of a random habitat is 0.1

**Part A.** Draw a schematic diagram of the Markov Process illustrating states and transitions. Define the transition rate matrix,  $P$ , for this discrete time discrete state stochastic process.

**Part B.** Calculate the stationary distribution,  $\pi$ , for this process. Interpret the stationary distribution in the context of Species X's long-term occupancy in the ecological network.

**Part C.** Characterize the states of the system as *transient*, *absorbing* or *recurrent*.

**Part D.** Suppose that species  $X$  presents an important ecosystem service. If as a conservation biologist you can either support recolonization (upping the probability to 0.2) or reducing extinction probability to 0.1 what would be a better conservation strategy?

### 3. Genetic Drift

In this lecture we modelled genetic drift in two ways, using the Wright-Fisher model and using the Moran model. Assume that we have a population of size  $N = 50$  and two alleles 'A' and 'a' such that the initial frequency of the 'A' allele is initially  $p_A = 0.5$ . Assume that mutation from the  $A \rightarrow a$  and  $a \rightarrow A$  occurs with probability  $u = 0.001$  per reproduction event.

**Part A.** Write down the transition probabilities that describe each model of genetic drift

**Part B.** Calculate the stationary distribution in each model.

**Part C.** Simulate the dynamics of the 'A' allele in each model

**Part D.** Use your understanding of both models and your answers to A-C to compare and contrast the dynamics of these two models.

### 4. Extinction Probabilities in a Branching Process

In biology, branching processes are often used to model the population growth of species. Extinction probabilities are crucial in understanding the long-term survival of a population. Consider a simple branching process where each individual has, on average, 2.5 offspring.

*Model Parameters:*

Initial Population: Start with a single individual (the founding member).

Offspring Distribution: Each individual produces, on average, 2.5 offspring per generation.

Probability Distribution: Assume a Poisson distribution for simplicity.

**Part A.** Suppose that we start with a single individual calculate the probability that the population goes extinct after one generation. What is the extinction prob. in a single generation if there are  $n > 1$  individuals?

**Part B.** Let's explore how sensitive your answer is to the number of individuals per parent. What is the probability of extinction in 1 generation from 1 initial individual if the mean number of offspring is  $\lambda$ . Calculate the extinction probability in exactly two generations given that there is a single initial individual across various values of  $\lambda$ . What is the probability that the population goes extinct in the long term across various values of  $\lambda$ ?

Note you can work this problem out either numerically or analytically you choose.

**Part C.** The distribution of offspring per parent in human populations is not Poisson but rather negative binomial. The negative binomial distribution has two parameters  $0 < r$  and  $0 < p < 1$  with a mean of  $\frac{r(1-p)}{p}$  and variance  $\frac{r(1-p)}{p^2}$  (hence the variance is greater than the mean whereas in the Poisson distribution the variance equals the mean). Setting  $p = 0.5$  and  $r = 2.5$  (such that the mean is the same as in part A) and starting with a single individual, what is the probability of extinction in a single generation? How does this compare to your result from A?

**Part D.** Use your answers to the above parts to describe how extinction probability depends on the average number of offspring per parent, the mean and variance of the offspring distribution, and how these features influence the population's short-term and long-term survival.

## 5. Wright-Fisher Model with Selection

In class we discussed the Wright-Fisher model of genetic drift both with and without mutation.

**Part A:** What is the state space in this model? What are the transition probabilities in the Wright-Fisher Model with and without mutation for a population with  $N$  gene copies? In each of these models what is the probability that a particular offspring individual has a 'A' parent and what is the probability that they have a 'a' parent?

**Part B:** Suppose that there are currently  $i$  focal 'A' individuals in the population what is the frequency  $p$  of that allele? What is the frequency of the alternate 'a' allele? Rewrite the transition probabilities in part A in terms of this allele frequency  $p$ .

**Part C:** Now, let's introduce selection favouring the focal 'A' allele. Suppose that haploid individuals with allele 'A' have fitness  $W_A = 1$  and that individuals with the other 'a' allele have reduced fitness  $W_a = 1$  such that the probability that a particular offspring is a 'A' is given by its 'relative fitness':

$$w_A = \frac{pW_A}{pW_A + (1-p)W_a}$$

What is the probability that a particular offspring carries the 'a' allele?

**Part D:** Modify the transition probabilities to incorporate selection. Consider how the fitness values affect the probabilities of different genotype combinations in the next generation.

**Part E:** Use this transition probability to express the probability of having allele frequency  $p(t+1)$  in the next generation in terms of the current allele frequencies  $p(t)$ .

## 6. Challenge Question: Life History Evolution

Life history theory is the study of how organisms are born, age, reproduce, and die. Stochastic processes are important in formulating predictions and understanding of life history evolution. Consider a population with synchronous (aka discrete-time) reproduction and survival (e.g., plants). Suppose that the focal plant can exhibit one of two reproductive strategies:

Strategy 1: Each adult plant produces 10 seeds every year each of which have a 15% chance of surviving to become a reproductive adult in the following year.

Strategy 2: Each adult plant produces 50 seeds every year each of which has a 2% chance of surviving to become a reproductive adult in the following year.

**Part A:** Propose a branching process describing the number of adult plants each year. What are the relevant model parameters for Strategy 1 and 2. What is the mean # of offspring per parent in each strategy.

**Part B:** What is the probability of extinction for each strategy?

**Part C:** Simulate 100 trajectories for each strategy. How does the probability of extinction compare to your prediction in Part B? How does the distribution in population sizes compare between the two models?

**Part D:** Use your answers to part B and part C to formulate a hypothesis about which strategy is favoured by natural selection.