Exercise_week_02

2025 年 10 月 1 日

Statistical Methods in Astrophysics Exercises

Week 01: Probability Theory

Personal Information

Name: [double click to insert your name here]

Student ID: [double click to insert your student ID here]

Exercise 1: Flood risk simulation

A dam is built in a region where days are sunny or rainy, each with a long-term stable probability of 0.5. A flood occurs if there are at least 7 consecutive rainy days in a row; a long run of rainy days is counted only once.

Question 1.1: A plain Monte-Carlo simulation **Task:** Simulate 10,000 years of weather data with independent and identically distributed probabilities of sunny and rainy days. Estimate the expected number of floods per 365-day year.

Hint: You may generate binary random numbers to represent sunny and rainy days, e.g., 0 for sunny and 1 for rainy.

[ ]:
```python
# NOTE: Run this cell in the first place
# Load packages for numerical calculations and plotting
import numpy as np
import matplotlib.pyplot as plt
# Enable inline plotting in Jupyter notebooks
%matplotlib inline
```

[ ]:
```python
# A helper function to return the lengths of consecutive runs of 1s in a␣
↪binary array
def consecutive_ones(arr):
    """Return the lengths of consecutive 1s in a binary array."""
    # Identify the start and end indices of sequences of 1s
    changes = np.diff(arr, prepend=0, append=0)
    start_indices = np.where(changes == 1)[0]
    end_indices = np.where(changes == -1)[0]
    # Calculate the lengths of these sequences
    lengths = end_indices - start_indices
    return lengths
```

[ ]:
```python
# Set the random seed
seed = ...      # NOTE: EDIT HERE to insert your student ID as the seed
nsample = 365 * 10000   # number of days to simulate

rng = np.random.default_rng(seed)
weather = rng.integers(0, 2, size=nsample)      # 0 for sunny, 1 for rainy

# Validate the overall fraction of rainy days
print(f'Fraction of rainy days: {weather.mean():.4%}')
```

[ ]:
```python
# Get the lengths of consecutive rainy days
rainy_lengths = consecutive_ones(weather)

# NOTE: EDIT BELOW to compute the expected number of floods per year
```

Question 1.2: Markov-chain Monte-Carlo (MCMC) simulation In reality, weather conditions are correlated in time. For example, a rainy day is more likely to be followed by another rainy day. We can model the weather as a two-state Markov chain with the following transition matrix:

$$T = \begin{bmatrix} P(S \to S) & P(S \to R) \\ P(R \to S) & P(R \to R) \end{bmatrix}$$

where S and R denote sunny and rainy days, respectively. $P(S \to S)$ is the probability of a sunny day being followed by another sunny day, and so on.

Assume that $P(S \to S) = P(R \to R) = 0.7$.

Tasks: 1. Complete the transition matrix and implement a Markov chain simulation to generate weather data in 100 years. Estimate the expected number of floods per 365-day year. 2. Plot the histogram of the lengths of consecutive rainy days obtained from both simulations in Questions 1.1 and 1.2. Choose appropriate bin sizes, ranges, and axes scales to clearly show the differences between the two distributions. 3. Comment on the differences between the two distributions.

[ ]:
```
def markov_chain_simulation(transition_matrix, n_steps, seed=42):
    """Markov-chain Monte-Carlo simulation of weather data.
    Args:
        transition_matrix (2x2 array): transition probabilities.
        n_steps (int): Number of steps to simulate.
        seed (int): Random seed for reproducibility.
    Returns: weather sequence
    """
    rng = np.random.default_rng(seed)
    current_state = rng.choice([0, 1])      # Start from a random state
    states= [current_state]

    for _ in range(n_steps - 1):
        prob = transition_matrix[current_state]
        next_state = rng.choice([0, 1], p=prob)
        states.append(next_state)
```

```
        current_state = next_state

    return np.array(states)
```

[ ]:
```
nsample = 365 * 100      # number of days to simulate
# NOTE: EDIT BELOW to complete the transition matrix simulate
transition_matrix = np.array([[0.7, ...],
                              [..., 0.7]])

weather_mc = markov_chain_simulation(transition_matrix, nsample,␣
 ↪seed=seed)

# Validate the overall fraction of rainy days
print(f'Fraction of rainy days (Markov chain): {weather_mc.mean():.4%}')
```

[ ]:
```
# Get the lengths of consecutive rainy days
rainy_lengths_mc = consecutive_ones(weather_mc)

# NOTE: EDIT BELOW to compute the expected number of floods per year
```

[ ]:
```
# NOTE: EDIT BELOW to plot the histogram of lengths of rainy day runs
```

Discussions Add your discussion here (double click to edit, both Chinese and English are fine).

Question 1.3: Weather simulation with more states Suppose the weather can be sunny (S), cloudy (C), or rainy (R). The transition matrix is given by:

$$T = \begin{bmatrix} P(S \to S) & P(S \to C) & P(S \to R) \\ P(C \to S) & P(C \to C) & P(C \to R) \\ P(R \to S) & P(R \to C) & P(R \to R) \end{bmatrix} = \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.3 & 0.5 & 0.2 \\ 0.4 & 0.3 & 0.3 \end{bmatrix}$$

We start with a sunny day.

Tasks: 1. Simulate the weather for 5 days. Run the simulation with 100 different random seeds. What is the probability of having a rainy day on the 5th day? 2. Simulate the weather for 100 days. Run the simulation with 100 different random seeds. For each N in {2, 5, 10, 20, 50, 100}, plot the probabilities of being sunny, cloudy, and rainy on the N-th day as functions of N with appropriate axes scales. 3. Comment on the results.

[ ]:
```
# Modify the Markov-chain simulation for three states: sunny, cloudy,␣
↪and rainy
def markov_chain_simulation_3(transition_matrix, n_steps, seed=42):
    """Markov-chain Monte-Carlo simulation of weather data with three␣
↪states: sunny (0), cloudy (1), and rainy(2).
    Args:
        transition_matrix (3x3 array): transition probabilities.
        n_steps (int): Number of steps to simulate.
        seed (int): Random seed for reproducibility.
    Returns: weather sequence
    """
    rng = np.random.default_rng(seed)
    current_state = 0        # Start from a sunny day
    states= [current_state]

    for _ in range(n_steps - 1):
        prob = transition_matrix[current_state]
        next_state = rng.choice([0, 1, 2], p=prob)
        states.append(next_state)
        current_state = next_state

    return np.array(states)
```

[ ]:
```
transition_matrix_3 = np.array([[0.7, 0.2, 0.1],
                                [0.3, 0.5, 0.2],
                                [0.4, 0.3, 0.3]])
nsample = 5          # number of days to simulate
seeds = [seed + i for i in range(100)]   # Different seeds based on␣
↪student ID
```

```
is_rainy = []        # To store if the 5th day is rainy for each␣
↪simulation

for s in seeds:
    weather_3 = markov_chain_simulation_3(transition_matrix_3, nsample,␣
↪seed=s)
    is_rainy.append(weather_3[-1] == 2)      # Check if the 5th day is␣
↪rainy

# NOTE: EDIT BELOW to compute the probability of rainy on the 5th day
```

[ ]:
```
nsample = 100        # number of days to simulate
days = np.array([2, 5, 10, 20, 50, 100])
weather_states = [[] for _ in days]    # To store weather states for␣
↪different days

for s in seeds:
    weather_3 = markov_chain_simulation_3(transition_matrix_3, nsample,␣
↪seed=s)
    for i, day in enumerate(days):
        weather_states[i].append(...)    # NOTE: EDIT HERE to append the␣
↪weather state on the specified day

sunny_probs = []
cloudy_probs = []
rainy_probs = []

for states in weather_states:
    states = np.array(states)
    sunny_probs.append(...)    # NOTE: EDIT HERE and BELOW to compute the␣
↪probabilities of different weather states
```

[ ]:
```
# NOTE: EDIT BELOW to plot the probabilities of different weather states␣
↪on different days
```

Discussions Add your discussion here (double click to edit, both Chinese and English are fine).

Exercise 2: Stellar evolution simulation

A typical solar-mass star evolves through four stages: young (Y), main sequence (M), red giant (R), and white dwarf (W; resulting from supernova explosion). The transition matrix over a 10-million-year interval is given by (do not take the values too seriously; I made them up for this exercise):

$$T = \begin{bmatrix} P(Y \to Y) & P(Y \to M) & P(Y \to R) & P(Y \to W) \\ P(M \to Y) & P(M \to M) & P(M \to R) & P(M \to W) \\ P(R \to Y) & P(R \to M) & P(R \to R) & P(R \to W) \\ P(W \to Y) & P(W \to M) & P(W \to R) & P(W \to W) \end{bmatrix} = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 \\ 0 & 0.99 & 0.01 & 0 \\ 0 & 0 & 0.9 & 0.1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Question 2.1: stellar age **Task:** Simulate the life cycle of 1000 stars, and estimate the average age of a solar-mass star when it becomes a white dwarf using the 4-state Markov model.

[ ]:
```python
def markov_chain_simulation_star(transition_matrix, n_steps, seed=42):
    """Markov-chain Monte-Carlo simulation of stellar evolution with␣
↪four states: 0=Young, 1=Main Sequence, 2=Red Giant, 3=White Dwarf.
    Args:
        transition_matrix (4x4 array): transition probabilities.
        n_steps (int): Number of steps to simulate.
        seed (int): Random seed for reproducibility.
    Returns: stellar state sequence, with states represented as integers␣
↪0, 1, 2, 3.
    """
    rng = np.random.default_rng(seed)
    current_state = 0       # Start as young star
    # NOTE: EDIT BELOW to complete the function
```

[ ]:
```
def stellar_age(states):
    """Calculate the age of a star based on its state sequence.
    Args:
        states (array): Sequence of stellar states.
    Returns: age of the star in million years; -1 if it never becomes a␣
 ↪white dwarf
    """
    i = np.searchsorted(states, 3, side='left')     # First occurrence of␣
 ↪state 3 (white dwarf)
    if i < len(states) and states[i] == 3:
        return (i + 1) * 10    # Age in million years
    else:
        return -1               # Never becomes a white dwarf
```

[ ]:
```
stellar_transition_matrix = np.array([[0.5, 0.5, 0.0, 0.0],
                                      [0.0, 0.99, 0.01, 0.0],
                                      [0.0, 0.0, 0.9, 0.1],
                                      [0.0, 0.0, 0.0, 1.0]])


nsample = 1000          # number of stars to simulate
seeds = [seed + i for i in range(nsample)]  # Different seeds based on␣
 ↪student ID
max_steps = 1000        # Maximum number of steps to simulate for each␣
 ↪star
stellar_states = []     # To store the states of all stars

# Simulate the evolution of multiple stars
for s in seeds:
    states = markov_chain_simulation_star(stellar_transition_matrix,␣
 ↪max_steps, seed=s)
    stellar_states.append(states)

stellar_states = np.array(stellar_states)
```

[ ]:
```python
# Compute the ages of all stars
ages = np.array([stellar_age(states) for states in stellar_states])

# NOTE: EDIT BELOW to compute the average age; pay attention to stars that never become white dwarfs
```

Question 2.2: Stellar stage composition **Task:** For the 1000 stars simulated in Question 2.1, plot the porportion of stars in each stage (Y, M, R, W) as a function of time (up to 10 billion years). Use appropriate axes scales.

[ ]:
```python
# Compute the fractions of stars in each stage over time
fractions = (stellar_states[..., None] == np.arange(4)).mean(axis=0)
fracY, fracM, fracR, fracW = fractions.T

# NOTE: EDIT BELOW to plot the fractions of stars in different stages over time
```

Question 2.3: Stellar population with continuous formation Assume that 1000 stars have formed at a constant rate over the past 10 billion years, i.e., the birth times of the stars can be drawn from a uniform distribution between 0 and 10 billion years.

Task: Compute the present-day (10 billion years after the first stellar birth) proportions of stars in each stage (Y, M, R, W) in this population.

Hint: Can you complete this task without running a new simulation?

[ ]:
```python
rng = np.random.default_rng(seed - 1)
birth_times = rng.integers(0, 1000, size=nsample)   # Birth times of stars (unit: 10 million years)
today = 1000        # Present time (unit: 10 million years)

# Use the birth times to determine the current states of the stars
mask = birth_times < today
ages = today - birth_times[mask]
```

```
present_states = np.array([stellar_states[i, age - 1] if age - 1 <␣
↪max_steps else 3 for i, age in zip(np.where(mask)[0], ages)])

# NOTE: EDIT BELOW to print the present-day proportions of stars in␣
↪different stages
```

Exercise 3: Transient detection

A special type of astronomical transient event (e.g., a supernova explosion) starts on a random night of an observing season with 180 nights in total. The transient event lasts for D consecutive nights and is only **confirmed** if it is observed in two consecutive nights within its duration.

Question 3.1: Detection probability with independent nights Assume each night is independently clear (observable) with a probability of 0.6.

Task: For each D in $\{2, 4, 8, 12, 16, 20\}$, simulate 10,000 observing seasons and estimate the probability of confirming the transient event. Plot the detection probability as a function of D with appropriate axes scales.

Hint: You may generate a binary random number sequence to represent clear and unusable nights, e.g., 0 for unusable and 1 for clear, then check if there are two consecutive 1s within the D nights starting from a random night.

[ ]:
```
durations = [2, 4, 8, 12, 16, 20]
```

Question 3.2: Detection probability with correlated weather The weather conditions are correlated in time. For example, a clear night is more likely to be followed by another clear night. We can model the weather as a two-state Markov chain with the following transition matrix:

$$T = \begin{bmatrix} P(U \to U) & P(U \to C) \\ P(C \to U) & P(C \to C) \end{bmatrix} = \begin{bmatrix} 0.7 & 0.3 \\ 0.2 & 0.8 \end{bmatrix}$$

where U and C denote unusable and clear nights, respectively. $P(U \to U)$ is the probability of an unusable night being followed by another unusable night, and so on.

Task: Repeat Question 3.1 using the Markov chain model to simulate the weather. Confirm that the steady-state probability of clear nights is still 0.6. Compare your transient detection probability results with those from Question 3.1.

```
[ ]:
```

Question 3.3: Detection probability with cadence gaps Not all clear nights are usable for transient detection due to various reasons (e.g., the moon is too bright). Assume that in every 29 nights, the first 5 nights are unusable for transient detection (even if the skies are clear).

Task: Repeat Question 3.2 with this additional constraint. Compare your results with those from Question 3.2. Comment on all your results.

Hint: You may use the modulo operator (`%`) to identify the nights that are unusable for transient detection, or use a predefined mask array to represent the cadence gaps. You do not have to rerun the weather simulation; just apply the cadence gaps to the results from Question 3.2.

```
[ ]:
```

Discussions Add your discussion here (double click to edit, both Chinese and English are fine).

Note: steps for submitting the exercise

1. In the menu bar, select `File > Download` to download your notebook as a `.ipynb` file.
2. Select `File > Save and Export Notebook As > PDF` to export your notebook as a PDF file.
3. Combine the `.ipynb` and `.pdf` files into a single `.zip` or `.tar.gz` archive.
4. Upload your archive to the web learning platform (网络学堂).