# SSY316 Assignment 5

Edoardo Mangia
edoardom@student.chalmers.se

Nuree Kim
nuree@student.chalmers.se

December 2024

## Contents



# 1 Exercise 2

## 1. Verification that (1) Defines a Markov Chain

A Markov chain is defined by the property that the future state depends only on the present state and not on the sequence of events that preceded it.

Given the recursive equation:

$$x_{k+1} = 0.9x_k + v_k \tag{1}$$

where $v_k \sim \mathcal{N}(0, 0.19)$ are independent and identically distributed random variables.

To verify that this is a Markov chain:

- The state $x_{k+1}$ depends only on $x_k$ and $v_k$, not on any previous states $x_{k-1}, x_{k-2}, \ldots$.
- Given $x_k$, the distribution of $x_{k+1}$ is independent of $x_{k-1}, x_{k-2}, \ldots$.

Thus, (1) satisfies the Markov property and defines a first-order Markov chain.

## 2. Implementation and Histogram Plotting

To simulate the Markov chain defined by (1), we:

1. Start with $x_1 = 5$.

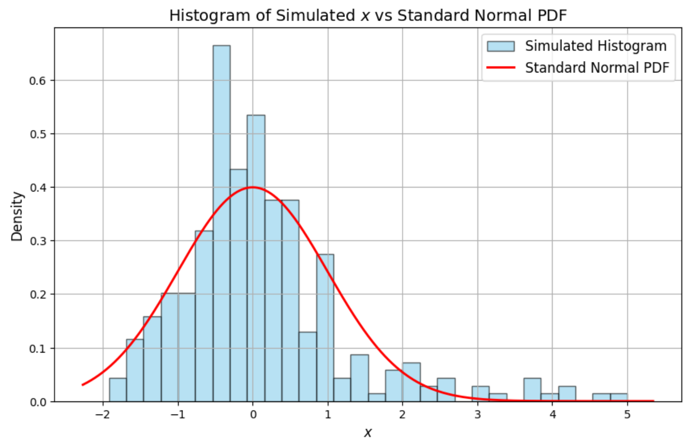
2. For each subsequent $k = 2, 3, \ldots, 300$,

$$x_k = 0.9x_{k-1} + v_{k-1}, \quad v_{k-1} \sim \mathcal{N}(0, 0.19) \tag{2}$$

3. Iterate the above equation to generate $x_2, x_3, \ldots, x_{300}$.
4. After the simulation, plot a histogram of the generated $x$ values.
5. Overlay the histogram with the Probability Density Function (PDF) of the standard normal distribution $\mathcal{N}(0, 1)$.

### Implementation

```python
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import norm

np.random.seed(42)

num_steps = 300           # Total number of steps in the Markov chain
initial_x = 5             # Initial value x[1] = 5
phi = 0.9                 # Coefficient for x[k]
sigma_v = np.sqrt(0.19)   # Standard deviation of v[k]

x = np.zeros(num_steps)
x[0] = initial_x

# Generate random noise v[k] ~ N(0, 0.19)
v = np.random.normal(0, sigma_v, num_steps - 1)

# Simulate the Markov chain
for k in range(1, num_steps):
    x[k] = phi * x[k-1] + v[k-1]

plt.figure(figsize=(10, 6))
plt.hist(x, bins=30, density=True, alpha=0.6, color='skyblue', edgecolor='black', label='Simulated Histogram')

# Overlay the standard normal PDF
xmin, xmax = plt.xlim()
x_vals = np.linspace(xmin, xmax, 1000)
plt.plot(x_vals, norm.pdf(x_vals, 0, 1), 'r', linewidth=2, label='Standard Normal PDF')

plt.title('Histogram of Simulated $x$ vs Standard Normal PDF', fontsize=14)
plt.xlabel('$x$', fontsize=12)
plt.ylabel('Density', fontsize=12)
plt.legend(fontsize=12)
plt.grid(True)
plt.show()
```




## 3. Time Series Plot and Burn-in Period

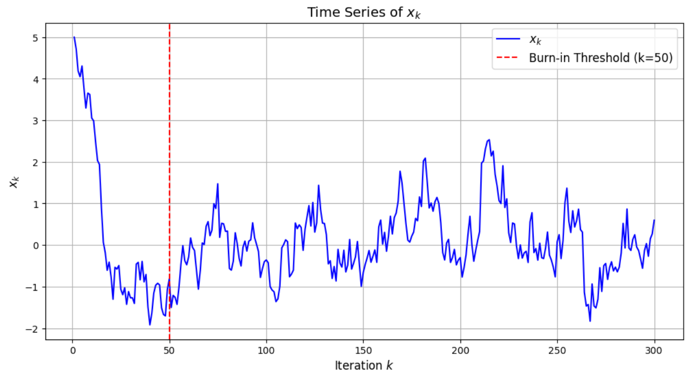
After simulating the Markov chain for $k = 1$ to $k = 300$, we plot the time series of $x_k$ to observe its convergence behavior.

As $k$ increases, $x_k$ fluctuates around 0, indicating convergence towards the stationary distribution $\mathcal{N}(0, 1)$.

For this example, suppose the plot shows stabilization after approximately $k = 50$. Therefore, the burn-in period can be considered to be the first 50 iterations.

### Implementation

```
# Plot the time series of x[k]
plt.figure(figsize=(12, 6))
plt.plot(range(1, num_steps + 1), x, label='$x_k$', color='blue')

burn_in = 50
plt.axvline(x=burn_in, color='red', linestyle='--',
    label=f'Burn-in Threshold (k={burn_in})')

plt.title('Time Series of $x_k$', fontsize=14)
plt.xlabel('Iteration $k$', fontsize=12)
plt.ylabel('$x_k$', fontsize=12)
plt.legend(fontsize=12)
plt.grid(True)
plt.show()
```




## 4. Proof that the Stationary Distribution of (1) is $\mathcal{N}(0, 1)$

To prove that the stationary distribution $\pi$ of the Markov chain defined by:

$$x_{k+1} = 0.9x_k + v_k, \quad v_k \sim \mathcal{N}(0, 0.19)$$

is $\mathcal{N}(0, 1)$, we analyze the mean and variance in the stationary regime.

### Mean Calculation

Assume $x_k$ has a stationary distribution with mean $\mu$. Then,

$$\begin{aligned} \mu &= \mathbb{E}[x_{k+1}] \\ &= \mathbb{E}[0.9x_k + v_k] \\ &= 0.9\mathbb{E}[x_k] + \mathbb{E}[v_k] \\ &= 0.9\mu + 0 \quad (\text{since } \mathbb{E}[v_k] = 0) \end{aligned}$$

Solving for $\mu$:

$$\mu = 0.9\mu \implies \mu(1 - 0.9) = 0 \implies \mu = 0$$

### Variance Calculation

Let $\sigma^2$ be the stationary variance of $x_k$. Then,

$$\begin{aligned} \sigma^2 &= \mathrm{Var}(x_{k+1}) \\ &= \mathrm{Var}(0.9x_k + v_k) \\ &= 0.9^2\mathrm{Var}(x_k) + \mathrm{Var}(v_k) \quad (\text{since } x_k \text{ and } v_k \text{ are independent}) \\ &= 0.81\sigma^2 + 0.19 \end{aligned}$$

Solving for $\sigma^2$:

$$\begin{aligned} \sigma^2 - 0.81\sigma^2 &= 0.19 \\ 0.19\sigma^2 &= 0.19 \\ \sigma^2 &= 1 \end{aligned}$$

## 2 Exercise 3

### 1. KL Divergence Between Two Scalar Gaussians

Given two scalar Gaussian distributions:

$$p(x) = \mathcal{N}(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right),$$
$$q(x) = \mathcal{N}(x; m, s^2) = \frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(x-m)^2}{2s^2}\right).$$

The Kullback-Leibler (KL) divergence from $p(x)$ to $q(x)$ is defined as:

$$\mathrm{KL}(p \parallel q) = \int_{-\infty}^{\infty} p(x) \ln\left(\frac{p(x)}{q(x)}\right) dx.$$

**Derivation:**

$$\begin{aligned}
\mathrm{KL}(p \parallel q) &= \int_{-\infty}^{\infty} p(x) \ln\left(\frac{p(x)}{q(x)}\right) dx \\
&= \int_{-\infty}^{\infty} p(x) \left[\ln p(x) - \ln q(x)\right] dx \\
&= \underbrace{\int_{-\infty}^{\infty} p(x) \ln p(x) dx}_{\text{Entropy of } p} - \underbrace{\int_{-\infty}^{\infty} p(x) \ln q(x) dx}_{\text{Cross-Entropy}}.
\end{aligned}$$

However, it's more straightforward to directly compute the KL divergence by expanding the logarithm:

$$\begin{aligned}
\mathrm{KL}(p \parallel q) &= \int_{-\infty}^{\infty} p(x) \ln\left(\frac{p(x)}{q(x)}\right) dx \\
&= \int_{-\infty}^{\infty} p(x) \left[\ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{(x-\mu)^2}{2\sigma^2} - \ln\left(\frac{1}{\sqrt{2\pi s^2}}\right) + \frac{(x-m)^2}{2s^2}\right] dx \\
&= \int_{-\infty}^{\infty} p(x) \left[\ln\left(\frac{s}{\sigma}\right) + \frac{(x-m)^2}{2s^2} - \frac{(x-\mu)^2}{2\sigma^2}\right] dx \\
&= \ln\left(\frac{s}{\sigma}\right) \underbrace{\int_{-\infty}^{\infty} p(x) dx}_{=1} + \frac{1}{2s^2} \int_{-\infty}^{\infty} p(x)(x-m)^2 dx - \frac{1}{2\sigma^2} \int_{-\infty}^{\infty} p(x)(x-\mu)^2 dx \\
&= \ln\left(\frac{s}{\sigma}\right) + \frac{1}{2s^2}\mathbb{E}_p[(x-m)^2] - \frac{1}{2\sigma^2}\mathbb{E}_p[(x-\mu)^2].
\end{aligned}$$

Since $x \sim \mathcal{N}(\mu, \sigma^2)$, we can compute the expectations using the provided identity:

$$\mathbb{E}_p[(x-a)^2] = (\mu - a)^2 + \sigma^2.$$

Applying this identity:

$$\begin{aligned}\mathbb{E}_p[(x-m)^2] &= (\mu-m)^2+\sigma^2,\\ \mathbb{E}_p[(x-\mu)^2] &= \sigma^2.\end{aligned}$$

Substituting back into the KL divergence expression:

$$\begin{aligned}\mathrm{KL}(p \parallel q) &= \ln\left(\frac{s}{\sigma}\right) + \frac{1}{2s^2}\left[(\mu-m)^2+\sigma^2\right] - \frac{1}{2\sigma^2}\cdot\sigma^2\\ &= \ln\left(\frac{s}{\sigma}\right) + \frac{(\mu-m)^2+\sigma^2}{2s^2} - \frac{1}{2}\\ &= \ln\left(\frac{s}{\sigma}\right) + \frac{(\mu-m)^2}{2s^2} + \frac{\sigma^2}{2s^2} - \frac{1}{2}\\ &= \ln\left(\frac{s}{\sigma}\right) + \frac{\sigma^2+(\mu-m)^2}{2s^2} - \frac{1}{2}.\end{aligned}$$

## 2. KL Divergence Between Two Multivariate Gaussians

Given two multivariate Gaussian distributions:

$$\begin{aligned}p(\mathbf{x}) &= \mathcal{N}(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma}),\\ q(\mathbf{x}) &= \mathcal{N}(\mathbf{x};\mathbf{m},\mathbf{S}),\end{aligned}$$

where $\boldsymbol{\mu},\mathbf{m} \in \mathbb{R}^d$ are mean vectors, and $\boldsymbol{\Sigma},\mathbf{S} \in \mathbb{R}^{d\times d}$ are positive definite covariance matrices.

The KL divergence from $p(\mathbf{x})$ to $q(\mathbf{x})$ is defined as:

$$\mathrm{KL}(p \parallel q) = \int_{\mathbb{R}^d} p(\mathbf{x})\ln\left(\frac{p(\mathbf{x})}{q(\mathbf{x})}\right)d\mathbf{x}.$$

**Derivation:**

$$\begin{aligned}\mathrm{KL}(p \parallel q) &= \int_{\mathbb{R}^d} p(\mathbf{x})\ln\left(\frac{p(\mathbf{x})}{q(\mathbf{x})}\right)d\mathbf{x}\\ &= \int_{\mathbb{R}^d} p(\mathbf{x})\left[\ln p(\mathbf{x}) - \ln q(\mathbf{x})\right]d\mathbf{x}\\ &= \underbrace{\int_{\mathbb{R}^d} p(\mathbf{x})\ln p(\mathbf{x})d\mathbf{x}}_{\text{Entropy of } p} - \underbrace{\int_{\mathbb{R}^d} p(\mathbf{x})\ln q(\mathbf{x})d\mathbf{x}}_{\text{Cross-Entropy}}.\end{aligned}$$

Expanding the logarithms:

$$\begin{aligned}\ln p(\mathbf{x}) &= -\frac{1}{2}\ln\left((2\pi)^d|\boldsymbol{\Sigma}|\right) - \frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}),\\ \ln q(\mathbf{x}) &= -\frac{1}{2}\ln\left((2\pi)^d|\mathbf{S}|\right) - \frac{1}{2}(\mathbf{x}-\mathbf{m})^\top\mathbf{S}^{-1}(\mathbf{x}-\mathbf{m}).\end{aligned}$$

Substituting back into the KL divergence:

$$
\begin{aligned}
\mathrm{KL}(p \parallel q) = &\int_{\mathbb{R}^d} p(\mathbf{x}) \left[ -\frac{1}{2} \ln \left( (2\pi)^d |\boldsymbol{\Sigma}| \right) - \frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right] d\mathbf{x} \\
&- \int_{\mathbb{R}^d} p(\mathbf{x}) \left[ -\frac{1}{2} \ln \left( (2\pi)^d |\mathbf{S}| \right) - \frac{1}{2} (\mathbf{x} - \mathbf{m})^\top \mathbf{S}^{-1} (\mathbf{x} - \mathbf{m}) \right] d\mathbf{x} \\
= &-\frac{1}{2} \ln \left( \frac{|\boldsymbol{\Sigma}|}{|\mathbf{S}|} \right) - \frac{1}{2} \int_{\mathbb{R}^d} p(\mathbf{x}) \left[ (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) - (\mathbf{x} - \mathbf{m})^\top \mathbf{S}^{-1} (\mathbf{x} - \mathbf{m}) \right] d\mathbf{x}.
\end{aligned}
$$

To evaluate the remaining integral, we separate the quadratic terms:

$$
\begin{aligned}
&\int_{\mathbb{R}^d} p(\mathbf{x}) (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) d\mathbf{x} = \mathrm{Tr}(\boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}) = \mathrm{Tr}(\mathbf{I}_d) = d, \\
&\int_{\mathbb{R}^d} p(\mathbf{x}) (\mathbf{x} - \mathbf{m})^\top \mathbf{S}^{-1} (\mathbf{x} - \mathbf{m}) d\mathbf{x} = (\boldsymbol{\mu} - \mathbf{m})^\top \mathbf{S}^{-1} (\boldsymbol{\mu} - \mathbf{m}) + \mathrm{Tr}(\mathbf{S}^{-1} \boldsymbol{\Sigma}).
\end{aligned}
$$

Here, we have used the provided identity:

$$
\mathbb{E}_p \left[ (\mathbf{x} - \mathbf{a})^\top \mathbf{B} (\mathbf{x} - \mathbf{a}) \right] = (\boldsymbol{\mu} - \mathbf{a})^\top \mathbf{B} (\boldsymbol{\mu} - \mathbf{a}) + \mathrm{Tr}(\mathbf{B} \boldsymbol{\Sigma}).
$$

Substituting these results back into the KL divergence:

$$
\begin{aligned}
\mathrm{KL}(p \parallel q) &= -\frac{1}{2} \ln \left( \frac{|\boldsymbol{\Sigma}|}{|\mathbf{S}|} \right) - \frac{1}{2} \left[ d - \left( (\boldsymbol{\mu} - \mathbf{m})^\top \mathbf{S}^{-1} (\boldsymbol{\mu} - \mathbf{m}) + \mathrm{Tr}(\mathbf{S}^{-1} \boldsymbol{\Sigma}) \right) \right] \\
&= -\frac{1}{2} \ln \left( \frac{|\boldsymbol{\Sigma}|}{|\mathbf{S}|} \right) - \frac{1}{2} d + \frac{1}{2} (\boldsymbol{\mu} - \mathbf{m})^\top \mathbf{S}^{-1} (\boldsymbol{\mu} - \mathbf{m}) + \frac{1}{2} \mathrm{Tr}(\mathbf{S}^{-1} \boldsymbol{\Sigma}) \\
&= \frac{1}{2} \left[ \ln \left( \frac{|\mathbf{S}|}{|\boldsymbol{\Sigma}|} \right) - d + \mathrm{Tr} \left( \mathbf{S}^{-1} \boldsymbol{\Sigma} \right) + (\boldsymbol{\mu} - \mathbf{m})^\top \mathbf{S}^{-1} (\boldsymbol{\mu} - \mathbf{m}) \right].
\end{aligned}
$$