Preliminary Quiz for Stochastic Control and Reinforcement Learning

Question 1: Expectation and Variance

Let X be a discrete random variable with probability mass function

$$P(X = 0) = 0.2, \qquad P(X = 1) = 0.5, \qquad P(X = 2) = 0.3.$$

Questions

(a) Compute $E[X]$.

(b) Compute $\mathrm{Var}(X)$.

(c) Compute $E[3X + 2]$.

Solution

The expectation is

$$E[X] = 0(0.2) + 1(0.5) + 2(0.3) = 1.1.$$

The second moment is

$$E[X^2] = 0^2(0.2) + 1^2(0.5) + 2^2(0.3) = 0 + 0.5 + 1.2 = 1.7.$$

Therefore,

$$\mathrm{Var}(X) = E[X^2] - E[X]^2 = 1.7 - (1.1)^2 = 1.7 - 1.21 = 0.49.$$

By linearity of expectation,

$$E[3X + 2] = 3E[X] + 2 = 3(1.1) + 2 = 5.3.$$

Answer

$$E[X] = 1.1, \qquad \mathrm{Var}(X) = 0.49, \qquad E[3X + 2] = 5.3.$$

Question 2: Best Mean-Square Prediction

Let X and Y be square-integrable random variables. Suppose there exists a square-integrable function $m(Y)$ such that

$$\mathbb{E}\big[(X - m(Y))g(Y)\big] = 0$$

for every square-integrable function $g(Y)$.

1. Show that for any square-integrable function $h(Y)$,

$$\mathbb{E}\big[(X - h(Y))^2\big] = \mathbb{E}\big[(X - m(Y))^2\big] + \mathbb{E}\big[(m(Y) - h(Y))^2\big].$$

2. Deduce that

$$\mathbb{E}\big[(X - m(Y))^2\big] \le \mathbb{E}\big[(X - h(Y))^2\big]$$

for every square-integrable function $h(Y)$.

3. Show that equality holds if and only if $h(Y) = m(Y)$. almost surely.

4. Briefly explain why $m(Y)$ can be interpreted as the best predictor of X based on Y.

Solution

Observe that

$$X - h(Y) = (X - m(Y)) + (m(Y) - h(Y)).$$

Squaring both sides gives

$$(X - h(Y))^2 = (X - m(Y))^2 + (m(Y) - h(Y))^2 + 2(X - m(Y))(m(Y) - h(Y)).$$

Taking expectations,

$$\begin{aligned}\mathbb{E}[(X - h(Y))^2] = \mathbb{E}[(X - m(Y))^2] &+ \mathbb{E}[(m(Y) - h(Y))^2] \\ &+ 2\mathbb{E}[(X - m(Y))(m(Y) - h(Y))].\end{aligned}$$

Since $m(Y) - h(Y)$ is itself a function of Y, the orthogonality assumption implies

$$\mathbb{E}[(X - m(Y))(m(Y) - h(Y))] = 0.$$

Therefore

$$\boxed{\mathbb{E}[(X - h(Y))^2] = \mathbb{E}[(X - m(Y))^2] + \mathbb{E}[(m(Y) - h(Y))^2]}.$$

Since the second term is non-negative,

$$\mathbb{E}[(X - m(Y))^2] \le \mathbb{E}[(X - h(Y))^2].$$

Thus $m(Y)$ minimises the mean-square error among all functions of Y. Equality holds if and only if

$$\mathbb{E}[(m(Y) - h(Y))^2] = 0,$$

which is equivalent to $h(Y) = m(Y)$ almost surely. Hence, the minimiser is unique.

Finally, $m(Y)$ is the unique function of the observation Y that minimises the mean-square prediction error. In probability theory, this function is precisely the conditional expectation $m(Y) = \mathbb{E}[X \mid Y]$.

Question 3: Markov Chains

Consider a Markov chain with state space $\{1, 2\}$ and transition matrix

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

Note that the entry P_{ij} is exactly the probability of moving from state i to state j in one step.

Questions

(a) If the chain starts in state 1, what is the probability that it is in state 2 after two steps?

(b) Find the stationary distribution π.

Solution

First compute

$$P^2 = \begin{bmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \end{bmatrix} \begin{bmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \end{bmatrix} = \begin{bmatrix} 0.72 & 0.28 \\ 0.56 & 0.44 \end{bmatrix}.$$

Hence,

$$P(X_2 = 2 \mid X_0 = 1) = 0.28.$$

Let the stationary distribution be

$$\pi = (\pi_1, \pi_2).$$

Then

$$\pi P = \pi, \qquad \pi_1 + \pi_2 = 1.$$

From $\pi P = \pi$,

$$\pi_1 = 0.8\pi_1 + 0.4\pi_2,$$

which implies

$$0.2\pi_1 = 0.4\pi_2,$$

and therefore

$$\pi_1 = 2\pi_2.$$

Using $\pi_1 + \pi_2 = 1$,

$$2\pi_2 + \pi_2 = 1,$$

so

$$\pi_2 = \frac{1}{3}, \qquad \pi_1 = \frac{2}{3}.$$

Answer

$$P(X_2 = 2 \mid X_0 = 1) = 0.28,$$

$$\pi = \left(\frac{2}{3}, \frac{1}{3} \right).$$

Question 4: Linear Systems with Constant Control

Consider the scalar controlled system

$$\dot{x}(t) = ax(t) + bu(t), \qquad x(0) = x_0,$$

where $a < 0$, $b \neq 0$, and the control is constant, $u(t) \equiv \bar{u}$.

Questions

(a) Solve for $x(t)$ explicitly.

(b) Compute $\lim_{t \to \infty} x(t)$.

(c) Given a target state x^*, find the constant control $\bar{u}$ such that the steady state equals x^*.

Solution

The equation is linear with constant forcing. Multiplying by the integrating factor e^{-at},

$$\frac{d}{dt}\left(e^{-at}x(t)\right) = b\bar{u}\,e^{-at},$$

and integrating from 0 to t,

$$e^{-at}x(t) - x_0 = \frac{b\bar{u}}{a}\left(1 - e^{-at}\right).$$

Therefore

$$x(t) = x_0e^{at} + \frac{b\bar{u}}{a}\left(e^{at} - 1\right).$$

Since $a < 0$, we have $e^{at} \to 0$ as $t \to \infty$, hence

$$\lim_{t\to\infty} x(t) = -\frac{b\bar{u}}{a}.$$

Setting the steady state equal to the target,

$$-\frac{b\bar{u}}{a} = x^* \quad\Longrightarrow\quad \bar{u} = -\frac{ax^*}{b}.$$

Answer

$$x(t) = x_0e^{at} + \frac{b\bar{u}}{a}\left(e^{at} - 1\right), \qquad \lim_{t\to\infty} x(t) = -\frac{b\bar{u}}{a}, \qquad \bar{u} = -\frac{ax^*}{b}.$$

Question 5: One-Step Optimal Control

The state undergoes a single transition

$$x_1 = ax_0 + bu,$$

where u is the control. The cost is

$$J(u) = qx_1^2 + ru^2, \qquad q \geq 0,\ r > 0.$$

Questions

(a) Find the optimal control u^* minimising $J(u)$.

(b) Show that $u^* = -Kx_0$ is a linear feedback, and give the gain K.

(c) Compute the optimal cost $J(u^*)$ and show that it is quadratic in x_0, i.e. $J(u^*) = Px_0^2$, giving P explicitly.

Solution

The cost is

$$J(u) = q(ax_0 + bu)^2 + ru^2,$$

which is a strictly convex quadratic in u since $J''(u) = 2(qb^2 + r) > 0$. Setting the derivative to zero,

$$J'(u) = 2qb(ax_0 + bu) + 2ru = 0,$$

gives

$$(qb^2 + r)\,u = -qab\,x_0, \qquad \text{so} \qquad u^* = -\frac{qab}{r + qb^2}\,x_0.$$

Hence $u^* = -Kx_0$ with feedback gain

$$K = \frac{qab}{r + qb^2}.$$

Substituting back, the optimally controlled state is

$$x_1 = ax_0 + bu^* = ax_0 \left(1 - \frac{qb^2}{r + qb^2}\right) = \frac{ar}{r + qb^2}\, x_0,$$

and the optimal cost is

$$J(u^*) = q\left(\frac{ar}{r + qb^2}\right)^2 x_0^2 + r\left(\frac{qab}{r + qb^2}\right)^2 x_0^2 = \frac{qa^2 r(r + qb^2)}{(r + qb^2)^2}\, x_0^2 = \frac{qa^2 r}{r + qb^2}\, x_0^2.$$

Answer

$$u^* = -Kx_0, \qquad K = \frac{qab}{r + qb^2}, \qquad J(u^*) = Px_0^2 \text{ with } P = \frac{qa^2 r}{r + qb^2}.$$

(This is one step of the discrete Riccati recursion appearing in the LQR problem.)

Question 6: Simulating a Feedback-Controlled System (Python/Matlab)

Consider the closed-loop system

$$\dot{x} = ax + bu, \qquad u = -kx, \qquad a = 1,\ b = 1,\ x_0 = 1.$$

Questions

(a) Using the Euler discretisation

$$x_{n+1} = x_n + \Delta t\,(a - bk)x_n, \qquad \Delta t = 0.01,\ t \in [0, 3],$$

write a program (Python or Matlab) that simulates the trajectory for $k = 0.5$ and $k = 2$, and plots both.

(b) Compare with the exact solution $x(t) = x_0 e^{(a-bk)t}$, and explain how the value of k affects the stability of the system.

Solution

A Python implementation:

```
import numpy as np
import matplotlib.pyplot as plt

a, b, x0 = 1.0, 1.0, 1.0
dt, T = 0.01, 3.0
n = int(T / dt)
t = np.linspace(0, T, n + 1)

plt.figure()
for k in [0.5, 2.0]:
    x = np.zeros(n + 1)
    x[0] = x0
    for i in range(n):
        x[i + 1] = x[i] + dt * (a - b * k) * x[i]
    exact = x0 * np.exp((a - b * k) * t)
    plt.plot(t, x, label=f"Euler, k={k}")
    plt.plot(t, exact, "--", label=f"Exact, k={k}")

plt.xlabel("t"); plt.ylabel("x(t)"); plt.legend(); plt.show()
```

The closed-loop dynamics are $\dot{x} = (a - bk)x$, so the sign of $a - bk$ determines stability.

For $k = 0.5$: $a - bk = 0.5 > 0$, and the trajectory grows exponentially (unstable); at $t = 3$ the Euler scheme gives $x \approx 4.47$ against the exact value $e^{1.5} \approx 4.48$.

For $k = 2$: $a - bk = -1 < 0$, and the trajectory decays to zero (stable); at $t = 3$ the Euler scheme gives $x \approx 0.049$ against the exact value $e^{-3} \approx 0.050$.

Answer

The feedback stabilises the system if and only if $k > a/b = 1$: the closed-loop rate is $a - bk$, which is negative precisely when $k > a/b$. The Euler trajectories match the exact solutions closely for $\Delta t = 0.01$.