ORF 418 - Optimal Learning
Emma Hubert

Midterm

October 12, in class (8:30am – 9:50am)

Exercise 1. [35pt] A multi-armed bandit problem.

Consider two **independent** random variables X and Y. X is uniform on the interval $[0, 1]$ and Y is exponential with parameter $\lambda = 1$, i.e. their densities are respectively given by:

$$f_X(x) = \mathbb{1}_{x \in [0,1]},\ x \in \mathbb{R}, \text{ and } f_Y(y) = \mathrm{e}^{-y} \mathbb{1}_{y \geq 0},\ y \in \mathbb{R}.$$

We consider a multi-armed bandit problem with two arms $A = 1, 2$. The reward of arm 1 is given by the random variable X, while the reward from arm 2 is described by the random variable Y.

1. **[5pt]** If we know the reward distributions of both arms and want to maximise the **expected reward**, which arm should we choose?
2. **[10pt]** Write down the joint density of the pair (X, Y) and compute $p = \mathbb{P}(X \geq Y)$ *(Hint: Use integration by parts)*.

We now assume that the reward distributions are **unknown**. We **first explore each arm one time** and obtain the (random) rewards $r^{(1)}$ from arm 1 and $r^{(2)}$ from arm 2. We let

$$A^\star := \underset{A \in \{1,2\}}{\arg\max}\, r^{(A)}.$$

We then **exploit** by using **only** $A^\star$ and receive independent rewards $(r_1, r_2, \ldots)$. The limiting reward is defined as

$$J(A^\star) := \lim_{T \to \infty} \frac{1}{T} \sum_{k=1}^{T} r_k.$$

3. **[5pt]** Compute $J(A^\star)$ for both $A^\star = 1$ and $A^\star = 2$.
4. **[10pt]** Compute the distribution of $A^\star$. *Hint: Use the result from question 2. If you do not have the value for p, state your result in terms of p.*
5. **[5pt]** Compute $\mathbb{E}[J(A^\star)]$.

Exercise 2. [35pt] Linear-Quadratic Problem.

Consider the **Linear-Quadratic control problem** with state dynamics $x_{k+1} = Ax_k + Bu_k$, with $x_0 := x$ fixed, and cost function (to be minimised)

$$J(x_0, u) := \sum_{k=0}^{\infty} \rho^k \Big(x_k^\top M x_k + u_k^\top N u_k \Big),$$

with the following parameters $d = 2$, $\ell = 1$, $\rho = 1/2$, $N = 1$ and

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},\ B = \begin{pmatrix} 1 \\ 0 \end{pmatrix},\ M = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

In the following, we will denote the system at time $k \in \mathbb{N}$ by:

$$x_k := \begin{pmatrix} y_k \\ z_k \end{pmatrix}.$$

We also denote by $v(x)$ the value function associated to this LQ problem, for any initial point $x := (y, z)^\top \in \mathbb{R}^2$.

1. **[5pt]** State the general Dynamic Programming Equation and use it to show that the value function satisfies:

$$v(x) = y^2 + \inf_{u\in\mathbb{R}} \left\{ u^2 + \frac{1}{2} v(\widetilde{x}) \right\}, \qquad \forall x := (y,z)^\top \in \mathbb{R}^2, \tag{0.1}$$

for some $\widetilde{x}$ to be determined ($\widetilde{x}$ is allowed to depend on (x,u)).

From now on, assume that there exists a matrix

$$V := \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix},$$

for some $a, b \in \mathbb{R}$, $a > -2$, such that $v(x) = x^\top V x$ for all $x = (y,z)^\top \in \mathbb{R}^2$. We also define

$$f(u) := u^2 + \frac{1}{2}\big(a(y+u)^2 + bz^2\big), \; u \in \mathbb{R}.$$

2. **[5pt]** Using the previous notations and assumptions, show that (0.1) is equivalent to:

$$ay^2 + bz^2 = y^2 + \inf_{u\in\mathbb{R}} f(u) \qquad \forall (y,z) \in \mathbb{R}^2. \tag{0.2}$$

3. **[5pt]** Show that for all $(y,z) \in \mathbb{R}^2$, there exists $u^\star$ such that:

$$\inf_{u\in\mathbb{R}} f(u) = f(u^\star) = \frac{ay^2}{2+a} + \frac{1}{2}bz^2. \tag{0.3}$$

4. **[10pt]** Using the previous result, show that the admissible solution to equation (0.2) is $(a,b) = (\sqrt{2}, 0)$.

5. **[5pt]** What is the optimal (feedback) control $u_k^\star$ at each step k? Also write out the equation for the evolution of the state process governed by this optimal control.

6. **[5pt]** Compute the controllability matrix and conclude whether this system is controllable. What can you say about the value function?

Exercise 3. [30pt] Bayes formula.

Emma uses her car 50% of the time, walks 10% of the time and rides the bus 40% of the time as she commutes to ORFE. She is late 30% of the time when walking; 10% of the time when driving; and 20% of the time when taking the bus.

We denote the events $B = \{\text{she takes the bus}\}$, $W = \{\text{she walks}\}$ and $L = \{\text{she is late}\}$.

1. **[5pt]** Give the values of $\mathbb{P}(B), \mathbb{P}(W), \mathbb{P}(L|B)$ and $\mathbb{P}(L|W)$ *(no justification needed)*.

2. **[10pt]** Compute the values of $\mathbb{P}((B \cup W)^c)$ and $\mathbb{P}(L|(B \cup W)^c)$. What meaning do these probabilities have?

3. **[15pt]** What is the probability she took the bus if she was late? *Give the formula used with appropriate notations and the exact result.*