ORF 418 - Optimal Learning
Emma Hubert

Midterm – Solution

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?

   Solution. The distribution of X is uniform on the interval $[0,1]$, so that $\mathbb{E}[X] = 1/2$. On the other hand, the distribution of Y is exponential with parameter $\lambda = 1$, so that $\mathbb{E}[Y] = 1$.

   $$1/2 = \mathbb{E}[X] \leq \mathbb{E}[Y] = 1.$$

   Therefore, if we want to maximise the expected reward, we should choose arm 2.

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).*

   Solution. The random variables X and Y are independent, therefore the joint density is given by the product of the two densities:

   $$f_{X,Y}(x,y) = \mathrm{e}^{-y}\mathbb{1}_{x\in[0,1]}\mathbb{1}_{y\geq 0},\ (x,y) \in \mathbb{R}^2.$$

   Then, we can compute $\mathbb{P}(X \geq Y)$ as follows:

   $$\begin{aligned}
   \mathbb{P}(X \geq Y) &= \int\int f_{X,Y}(x,y)\mathbb{1}_{x\geq y}\mathrm{d}x\mathrm{d}y = \int_{y=0}^{\infty}\int_{x=0}^{1} \mathrm{e}^{-y}\mathbb{1}_{x\geq y}\mathrm{d}x\mathrm{d}y = \int_{y=0}^{1}\mathrm{e}^{-y}\int_{x=y}^{1}\mathrm{d}x\mathrm{d}y \\
   &= \int_{y=0}^{1}\mathrm{e}^{-y}(1-y)\mathrm{d}y \\
   &= \big[-\mathrm{e}^{-y}(1-y)\big]_0^1 - \int_0^1 \mathrm{e}^{-y}\mathrm{d}y \ \ \text{(by integration by parts)} \\
   &= 1 - \big[-\mathrm{e}^{-y}\big]_0^1 = \mathrm{e}^{-1}.
   \end{aligned}$$

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 $A^\star = 2$.

   Solution. Since $A^\star = 1$, we will only use the first arm for the exploitation part. Therefore, the distribution of the reward r_t is uniform on $[0,1]$. By the law of large number, $J(A^\star)$ corresponds to the expected value of the rewards from the first arm, *i.e.*

$$J(A^\star) = \mathbb{E}\big[X\big] = 1/2.$$

   Similarly, if $A^\star = 2$, then we will only use the second arm for the exploitation part. Therefore, the distribution of the reward r_t is exponential with parameter $\lambda = 1$. By the law of large number, $J(A^\star)$ corresponds to the expected value of the rewards from the first arm, *i.e.*

$$J(A^\star) = \mathbb{E}\big[Y\big] = 1.$$

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.*

   Solution. Let X the reward given by arm 1, with uniform distribution on $[0,1]$, and let Y the reward given by arm 2, with exponential distribution with parameter $\lambda = 1$. Now, $A^\star = 1$ if and only if $X \geq Y$, and $A^\star = 2$ iff $Y > X$. To obtain the distribution of $A^\star$, we thus have to compute $\mathbb{P}(X \geq Y)$ and $\mathbb{P}(X \leq Y)$. By question 2, we know that $\mathbb{P}(X \geq Y) = \mathrm{e}^{-1}$. Therefore, the distribution of $A^\star$ is given by

$$A^\star = \begin{cases} 1 & \textbf{with probability } p = \mathrm{e}^{-1} \\ 2 & \textbf{with probability } 1-p = 1-\mathrm{e}^{-1}. \end{cases}$$

5. **[5pt]** Compute $\mathbb{E}[J(A^\star)]$.

   Solution. Given the previous computations, we have:

$$\mathbb{E}[J(A^\star)] = \mathbb{E}[J(1)|A^\star = 1]\mathbb{P}(A^\star = 1) + \mathbb{E}[J(2)|A^\star = 2]\mathbb{P}(A^\star = 2) = \frac{1}{2}p + 1 - p = 1 - \frac{1}{2}p = 1 - \frac{1}{2}\mathrm{e}^{-1}.$$

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)).

Solution. The general Dynamic Programming equation (Thm 2.2.1 of the Lecture notes) is:

$$v(x) = x^\top M x + \inf_{u\in\mathbb{R}^\ell} \{u^\top N u + \rho v(Ax+Bu)\}, \quad \textbf{for all} \quad x := (y,z)^\top \in \mathbb{R}^2.$$

Here we have $\rho = 1/2$, $u^\top N u = u^2$ since $N = 1$,

$$x^\top M x = \begin{pmatrix} y & z \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} y \\ z \end{pmatrix} = y^2, \quad \textbf{and} \quad \widetilde{x} := Ax + Bu = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y \\ z \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} u = \begin{pmatrix} y+u \\ z \end{pmatrix}.$$

Replacing in the previous DP equation, we obtain the desired result (0.1).

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:

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

Solution. Using the assumption $v(x) = x^\top V x$ for all $x = (y,z)^\top \in \mathbb{R}^2$ in (0.1), we obtain:

$$x^\top V x = y^2 + \inf_{u\in\mathbb{R}} \left\{ u^2 + \frac{1}{2}\widetilde{x}^\top V \widetilde{x} \right\}.$$

Using the previous notation for V, we have:

$$\begin{pmatrix} y & z \end{pmatrix} \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \begin{pmatrix} y \\ z \end{pmatrix} = y^2 + \inf_{u\in\mathbb{R}} \left\{ u^2 + \frac{1}{2} \begin{pmatrix} y+u & z \end{pmatrix} \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \begin{pmatrix} y+u \\ z \end{pmatrix} \right\},$$

which is equivalent to (0.2).

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

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

Solution. To compute $\inf_{u\in\mathbb{R}} f(u)$, we compute the FOC:

$$f'(u) = 2u + a(y+u) = u(2+a) + ay = 0,$$

and the optimal is therefore $u^\star = \dfrac{-ay}{2+a}$ (since $f''(u) = 2 + a > 0$ by assumption on a). We now compute:

$$\begin{aligned} \inf_{u\in\mathbb{R}} f(u) = f(u^\star) &= \left(\frac{-ay}{2+a}\right)^2 + \frac{1}{2}\left(a\left(y + \frac{-ay}{2+a}\right)^2 + bz^2\right) = \frac{a^2y^2}{(2+a)^2} + \frac{1}{2}a\left(\frac{2y}{2+a}\right)^2 + \frac{1}{2}bz^2 \\ &= \frac{a^2y^2}{(2+a)^2} + \frac{2ay^2}{(2+a)^2} + \frac{1}{2}bz^2 = \frac{a^2y^2 + 2ay^2}{(2+a)^2} + \frac{1}{2}bz^2 \\ &= \frac{ay^2}{2+a} + \frac{1}{2}bz^2. \end{aligned}$$

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

   Solution. By the previous question, we know that:

$$\inf_{u\in\mathbb{R}} f(u) = \frac{ay^2}{2+a} + \frac{1}{2}bz^2, \quad \forall (y,z) \in \mathbb{R}^2.$$

   Plugging this result in (0.2), we obtain:

$$ay^2 + bz^2 = y^2 + \frac{ay^2}{2+a} + \frac{1}{2}bz^2, \quad \forall (y,z) \in \mathbb{R}^2.$$

   which is equivalent to

$$y^2\left(a - 1 - \frac{a}{2+a}\right) - \frac{1}{2}bz^2 = 0, \quad \forall (y,z) \in \mathbb{R}^2.$$

   Therefore, we have the following system:

$$\begin{cases} a - 1 - \dfrac{a}{2+a} = 0, \\ b = 0. \end{cases}$$

   The first line gives $(a-1)(2+a) = a$, *i.e.* $a^2 = 2$. Since the value function is positive, we have $a = \sqrt{2}$ (and $b = 0$).

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.

   Solution. By the question 3, we know that the optimal control, given by the minimiser of f, is $u^\star = \dfrac{-a}{2+a}y$. By the result of question 4, $a = \sqrt{2}$, so that the feedback control is $u_k^\star = \dfrac{-\sqrt{2}}{2+\sqrt{2}}y_k$ at each step $k \in \mathbb{N}$. The dynamics of the system at time $k \in \mathbb{N}$ with optimal control are given by:

$$x_{k+1} = \begin{pmatrix} y_{k+1} \\ z_{k+1} \end{pmatrix} = Ax_k + Bu_k^\star = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} y_k \\ z_k \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} u_k^\star = \begin{pmatrix} \frac{2}{2+\sqrt{2}}y_k \\ z_k \end{pmatrix}.$$

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

   Solution. Given the matrices A and B, the controllability matrix is defined by:

$$\mathcal{C} := \begin{pmatrix} B & AB \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}.$$

   The previous matrix has rank 1, so the system is not controllable. Recall that in class, we have seen that if the system is controllable, then the value function is finite. Here, the system is not controllable, but the value function is finite (we computed it previously).

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).*

   Solution. $\mathbb{P}(B) = 0.4, \mathbb{P}(W) = 0.1, \mathbb{P}(L|B) = 0.2$ **and** $\mathbb{P}(L|W) = 0.3$**.**

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?

   Solution. $\mathbb{P}((B \cup W)^c) = 1 - \mathbb{P}(B) - \mathbb{P}(W) = 0.5$ **is the probability that Emma does not take the bus or walk, *i.e.* the probability that Emma uses her car.** $\mathbb{P}(L|(B \cup W)^c) = 0.1$ **is the probability that Emma is late when driving by car.**

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.*

   Solution. We want to compute the probability $\mathbb{P}(B|L)$**. We use the following Bayes formula:**

$$\begin{aligned}\mathbb{P}(B|L) &= \frac{\mathbb{P}(L|B)\mathbb{P}(B)}{\mathbb{P}(L|B)\mathbb{P}(B) + \mathbb{P}(L|W)\mathbb{P}(W) + \mathbb{P}(L|(B \cup W)^c)(1 - \mathbb{P}(B) - \mathbb{P}(W))} \\ &= \frac{0.2 \times 0.4}{0.2 \times 0.4 + 0.3 \times 0.1 + 0.1 \times 0.5} = \frac{0.08}{0.08 + 0.03 + 0.05} = \frac{8}{16} = \frac{1}{2}.\end{aligned}$$