**Control Systems MSc**

**EEE Dept, Imperial College London**

GAME THEORY

COURSEWORK

Richard Vinter,

r.vinter@imperial.ac.uk

Updated: May 6th 2020

**Deadline for coursework submission: 2nd June 2020, 17.00 pm**

The coursework must be submitted by all MSc students taking the exam. It will count 25% towards the final grade.

**Coursework Grading Breakdown:**

**Task 1:** 20%

**Task 2:** 40%

**Task 3:** 40%**.**

The background theory for this homework is to be found in the Zero Sum Stochastic Dynamic Games Section of the Control Systems MSc Course Notes 'Game Theory'.

**Description of the Paint-Ball Shooting Contest**

Annie and Bruno are competing in a Paintball Contest. The rules are as follows:

Annie and Bruno stand with their paintguns, facing each other, and take turns to shoot at each other, if they wish to do so. If it is Annie's turn and they are $i$ steps away from each other and neither of them has yet fired, then Annie has two possible strategies:

1: Annie fires at Bruno, with probability $\alpha_i$ that she will hit him.

2: Annie does not fire and moves one step closer to Bruno.

If Annie fires and hits Bruno she has won the game, and if Annie fires and misses the target she has lost the game. If Annie does not fire, the outcome of the game is, so far, undecided and we must consider the next stage.

Likewise, if it is Bruno's's turn and they are both $i$ steps away from each other and neither of them has yet fired, then Bruno has two possible strategies:

1: Bruno fires at Annie, with probability $\beta_i$ that he will hit her.

2: Bruno does not fire and moves one step closer to Annie.

If Bruno fires and hits Annie he has won the game, and if Bruno fires and misses Annie he has lost the game. If Bruno does not fire, the game is, so far, undecided and we must consider what happens at the next stage.

(In this description of the game, $\alpha_i$ and $\beta_i$, $i = 1, 2, \ldots$, are given numbers in the interval $[0, 1]$, called the 'hitting probabilites'. We assume that $\alpha_0 = \beta_0 = 1$, i.e. the probability of hitting the target is 1 if the distance to the target is 0 steps.)

If the contestants start at time $t = 0$ at a distance of N, Annie aims to maximize the pay-off

$$\text{Probability[ Annie wins ]}, \tag{1}$$

while Bruno aims to minimize the pay-off (1) (which is equivalent to Bruno maximizing the probability of his winning, since the final outcome is always that one player wins and the other looses.) Thus the contest can be regarded as a zero-sum game.

*The coursework assignment is to answer the question 'What are the 'optimal' strategies for Annie and Bruno?' and to test your answer by means of MATLAB simulations. The answer is not obvious. Some calculations are required to determine the balance of advantage: should Annie should postpone firing because her hitting probability increases as she approaches Bruno, or should she fire immediately, to prevent Bruno from hitting her in the future?*

(You will notice that there is an equivalent formulation of the game involving slightly different, and more sensible-sounding, rules: Annie and Bruno take turns to choose to fire at each other (they may choose not to, in which case they move forward a step). They are allowed only one shot each. They continue until one of them has hit the target; the person who hits the target first wins the game. Observe that, if Annie (say) fires and misses, then Bruno can always choose not to fire, at each step, until he reaches the target when he fires and hits the target with probability 1. Thus, if Annie fires and misses, she looses the game, an outcome which is consistent with the original formulation. The original formulation is easier to mathematicize, because the constraints 'the competitors have one shot each' is automatically satisfied, whereas this constraint complicates the analysis of optimal strategies for the alternative formulation.)

*Mathematical Formulation*

Suppose Annie and Bruno start $N$ steps away from each other. Because they move forward one step at every stage of the game, increments in time (measured in 'steps') are the same as decriments in distance to the target. At time $t$, $0 \leq t \leq N$, we say that Annie is 'in' (meaning 'still in the game') if, at previous times, Bruno has not fired and hit Annie, and Annie has either not fired or has fired and hit Bruno. Otherwise we say Annie is 'out' (meaning 'out of the game'). The statements that Bruno is 'in' or 'out' have analogous meanings.

*State Variables:* The state variables at time $t$ are $(A_t, B_t, T_t)$, whose components have the interpretations:

$$A_t = \begin{cases} 1 & \text{if Annie is 'in'} \\ 0 & \text{if Annie is 'out'} \end{cases} \quad B_t = \begin{cases} 1 & \text{if Bruno is 'in'} \\ 0 & \text{if Bruno is 'out'} \end{cases} \quad T_t = \begin{cases} T^A & \text{if it is Annie's turn} \\ T^B & \text{if it is Bruno's turn}\ . \end{cases}$$

*Control Variables:* The control variables at time $t$ are $u_t^A$ (for Annie) and $u_t^B$ (for Bruno).

$$u_t^A = \begin{cases} 1 & \text{Annie fires} \\ 0 & \text{Annie does not fire} \end{cases} \quad u_t^B = \begin{cases} 1 & \text{if Bruno fires} \\ 0 & \text{if Bruno does not fire}\ . \end{cases}$$

(Notice that Annie's control $u_t^A$ will have no affect when $T_t = T^B$, i.e. when it is not Annie's turn, and Bruno's control will affect no affect when $T_t = T^A$.)

*State Transition Relations:* Suppose the state $(A_t, B_t, T_t)$ at time $t$ is known. Then the state $(A_{t+1}, B_{t+1}, T_{t+1})$ at time $t+1$ is a discrete random variable, which depends on the competitors' control actions, and is described by the following relations:

1. If either $A_t = 0$ or $B_t = 0$, then
$$(A_{t+1}, B_{t+1}) = (A_t, B_t)$$
2. If $T_t = T^A$ and $u_t^A = 0$, then
$$(A_{t+1}, B_{t+1}, T_{t+1}) = (A_t, B_t, T^B)$$
3. If $T_t = T^A$, $A_t = 1$, $B_t = 1$ and $u_t^A = 1$, then
$$(A_{t+1}, B_{t+1}, T_{t+1}) = \begin{cases} (0, 1, T^B) & \text{w.p. } 1 - \alpha_{N-t} \\ (1, 0, T^B) & \text{w.p. } \alpha_{N-t} \end{cases}$$
4. If $T_t = T^B$ and $u_t^B = 0$, then
$$(A_{t+1}, B_{t+1}, T_{t+1}) = (A_t, B_t, T^A)$$

5. If $T_t = T^B$, $A_t = 1$, $B_t = 1$ and $u_t^B = 1$, then

$$(A_{t+1}, B_{t+1}, T_{t+1}) = \begin{cases} (1, 0, T^A) & \text{w.p. } 1 - \beta_{N-t} \\ (0, 1, T^A) & \text{w.p. } \beta_{N-t} \end{cases}$$

*The Pay-off.* The random variable $A_N$ (Annie is 'in' at time $N$) takes value 1 if Annie has won and takes value 0 if Annie has not won. The pay-off (see (1)) is therefore simply $E[A_N]$.

*Tasks*

Consider the Paintball Dual, formulated as a zero-sum stochastic game:

$$(G)_{(0,(A,B,T))}: \quad \underset{u_0^A(.),\dots,u_{N-1}^A(.)}{\text{Max}} \quad \underset{u_0^B(.),\dots,u_{N-1}^B(.)}{\text{Min}} \quad \left\{E[x_N^A] \mid (A_0, B_0, T_0 = (A, B, T))\right\}$$

in which $N$ is the number of steps between them at time $t = 0$, and in which the pay-off $E[x_N^A]$ is calculated for a fixed initial state $(A_0, B_0, T_0) = (A, B, T)$ and for any choices of state feedback control strategies

$$u_t^A(A_t, B_t, T_t) \ \text{ and } \ u_t^B(A_t, B_t, T_t), \qquad t = 0, \dots, N-1 \ .$$

For arbitrary states $(A, B, T)$ and times $t \le N$, let $V_t(A, B, T)$ be the value of the following game, with initial state $(A, B, T)$, over the time horizon $\{t, t+1, \dots, N\}$):

$$(G)_{(t,(A,B,T))}: \quad \underset{u_t^A(.),\dots,u_{N-1}^A(.)}{\text{Max}} \quad \underset{u_t^B(.),\dots,u_{N-1}^B(.)}{\text{Min}} \quad \left\{E[x_N^A] \mid (A_t, B_t, T_t = (A, B, T))\right\}$$

*Task 1.* Using the Dynamic Programming method and the state transition relationships above to show that $V_t(A = 1, B = 1, T = T^A)$ and $V_t(A = 1, B = 1, T = T^B)$ satisfy the following 'backward' recursive relations:

$$\begin{cases} V_t(1, 1, T^A) = \alpha_{N-t} \vee V_{t+1}(1, 1, T^B) \\ V_t(1, 1, T^B) = (1 - \beta_{N-t}) \wedge V_{t+1}(1, 1, T^A)\,, \end{cases} \qquad t = N-1, N-2, \dots,$$

with starting conditions

$$V_N(1, 1, T^A) = 1 \ \text{ and } \ V_N(1, 1, T^B) = 0\,.$$

(Here $a \wedge b := \min\{a, b\}$ and $a \vee b := \max\{a, b\}$) .

Also show that, if the state is $(A = 1, B = 1, T^A)$ at time $t \le N - 1$, then the optimal strategy for Annie is

$$\begin{cases} \text{fire} & \text{if } \alpha_{N-t} > V_{t+1}(1, 1, T^B) \\ \text{don't fire} & \text{if } \alpha_{N-t} < V_{t+1}(1, 1, T^B) \end{cases}$$

and, if the state is $(A = 1, B = 1, T^B)$ at time $t \le N - 1$, then the optimal strategy for Bruno is

$$\begin{cases} \text{fire} & \text{if } (1 - \beta_{N-t}) < V_{t+1}(1, 1, T^A) \\ \text{don't fire} & \text{if } (1 - \beta_{N-t}) > V_{t+1}(1, 1, T^A)\,. \end{cases}$$

*Task 2.* Now assume that the 'hitting probabilities' are

$$\begin{cases} \alpha_i = \exp(-\rho_0 \times i) \\ \beta_i = 1 - \left(\frac{i}{\gamma_0 + i}\right) \end{cases}$$

in which $i \in \{0, 1, \ldots\}$ is 'distance from the target'. Here, $\rho_0 > 0$ and $\gamma_0 > 0$ are given constants. For choices of parameters

$$\rho_0 = 0.05 \text{ and } \gamma_0 = 5$$

write a MATLAB program to compute the sequences:

$$V_{N-i}(1, 1, T^A) \text{ and } V_{N-i}(1, 1, T^B)\,,$$

for $i = 0, 1, \ldots$, and plot them. For this choice of parameters, calculate

(a): the distance $i_A$ (between Annie and Bruno) when Annie should first shoot,

(b): the distance $i_B$ (between Annie and Bruno) when Bruno should first shoot,

(i.e. calculate the largest integer $i$ such that $u^A_{N-i} = 1$, when $T_{N-i} = T^A$ and the largest integer $i$ such that $u^B_{N-i} = 1$, when $T_{N-i} = T^B$).

*Task 3.* Simulate the game $G_{(0,(A,B,T))}$ for the above parameter values and $N = 10$, taking as starting values $(A, B, T) = (1, 1, T^A)$. You should assume that Annie and Bruno both use optimal strategies. Record $\hat{V}_0(1, 1, T^A)$ = the proportion of times that Annie wins over 100 simulations.

*Check that $\hat{V}_0(1, 1, T^A)$ agrees with $V_0(1, 1, T^A)$.*

A sensible, but not optimal, strategy for Bruno is the following

$$\begin{cases} \text{fire} & \text{if } \alpha_{N-t} > 0.5 \text{ and } \beta_{N-t} > 0.5 \\ \text{don't fire} & \text{otherwise}\,. \end{cases}$$

Repeat your simulations, now assuming that Annie follows her optimal strategy (see above), but Bruno follows his non-optimal strategy. *Calculate $\hat{\hat{V}}_0(1, 1, T^A)$ = the proportion of times that Annie now wins the game. Compare $\hat{\hat{V}}_0(1, 1, T^A)$ and $V_0(1, 1, T^A)$.*

*Express you opinion (without doing any further calculations) on the following question: if Annie knew that Bruno was following the sensible, but not optimal, strategy, could Annie improve her pay-off, by changing her strategy?*