Notes

Andrea Baisero

July 19, 2013

1 Rules

1.1 Food and Winning

Each player begins the game with $300(P-1)$ units of food, where P is the number of players.

If after any round you have zero food, you will die and no longer be allowed to compete. All players who survive until the end of the game will receive the survivor's prize.

The game can end in two ways. After a large number of rounds, there will be a small chance each additional round that the game ends. Alternatively, if there is only one person left with food then the game ends. In each case, the winner is the person who has the most food when the game ends.

The game is designed to encourage both cooperation and competition between players at different points in the game. Determining the right balance leads to a winning strategy.

1.2 Hunts

Each round is divided into hunts. A hunt is a game played between you and one other player. Each round you will have the opportunity to hunt with every other remaining player, so you will have $P-1$ hunts per round, where P is the number of remaining players.

In a hunt you can make the decision to hunt (become a hunter) or feign sickness and not hunt (become a slacker). Hunting is an active activity, so you will eat 6 units of food if you decide to hunt, and 2 units of food if you decide not to hunt. If both you and your partner do not hunt, then you get 0 food from the hunt. If only one of you decides to hunt then the hunt returns 6 units of food. If both of you decided to hunt then the hunt returns 12 units of food. The total food return of the hunt is split evenly between you and your partner. So, for example, if both of you decide to hunt, your individual net food gain/loss from the hunt would be $12/2-6=0$ units, whereas if both of you decided not to hunt your individual net food gain/loss is $0/2-2=-2$ food.

1.3 Reputation

Each player has a reputation R, which is defined as $R = \frac{H}{H+S}$, where H is the number of times you chose to hunt and S is the number of times you chose to slack (not hunt), both counted from the beginning of the game. Each player begins with a reputation of 0.

Each round, your tribe can save some time and hunt extra food if enough people opt to hunt. Each round, a random integer m, with $0 < m < P(P-1)$ will be chosen. If the total number of hunters is greater than or equal to m for the upcoming round, then everyone in the tribe gets $2(P-1)$ extra units of food after the round, which means that on average everyone gets 2 units of food per hunt. Before each round, you will find out the value of m.

Before each round you will be given: which number round you are in (from 1 to ∞), your food, your reputation, m, and an array with the reputation of each remaining member of the tribe when the round started. The location of each player in the array will be randomized after each round. You should decide your moves for that round for each player you will play against, and return an array of your decisions. After each hunt you will be given the food earned from each hunt you played.

1.4 Submission guidelines

The logic of your code and what game theory analyses you apply for the various steps must be either laid out as comments in the code or in a separate document. The algorithm (and separate document, if any) should be submitted by August 8, 2013 through your Brilliant account. Algorithms will be run tournament-style to determine the winner. The finalist teams will then have a short video interview with Brilliant staff, to verify that their work is their own.

Prizes for algorithms eligible for the final game will be awarded at two levels:

Grand Prize Winners – 5 teams will receive the Grand Prize of \$1000. The winning algorithm is guaranteed to receive the Grand Prize. 4 other winners will be selected based on the performance of their algorithm, logic, game theory analysis, and presentation. Finalists – Members of teams whose algorithm survives the game will receive an “I survived” t-shirt.

2 Recap of Rules

Annotation

Symbol	Meaning
A, A'	Actions of my agent and opponent agent respectively.
$U(A, A'), U(A', A)$	Utility function for my agent and opponent agent respectively.
T	Current number of turn. Starts with 1.
p_t	Number of players/rounds at turn t.
P_t	Number of total players/rounds until turn t, excluded.
h_t, s_t	Number of hunts/slacks (respectively) my agent performs at turn t.
H_t, S_t	Number of total hunts/slacks (respectively) until turn t excluded.
R_t, R'	My reputation (at turn t) and the opponent's reputation respectively.
pc, pc'	Probability of us and opponent cooperating.

Some Relationships

$$P_t = P_{t-1} + p_{t-1} = \sum_{\tau=1}^{t} p_\tau$$

$$H_t = H_{t-1} + h_{t-1} = \sum_{\tau=1}^{t} h_\tau$$

$$S_t = S_{t-1} + s_{t-1} = \sum_{\tau=1}^{t} s_\tau$$

$$P_t = H_t + S_t$$

$$p_t = h_t + s_t$$

$$R_t = \frac{H_t}{P_t}$$

The next one is important for reputation balancing (h_t would be our control variable),

$$\begin{aligned}
R_{t+1} &= \frac{H_{t+1}}{P_{t+1}} \\
&= \frac{H_t + h_t}{P_t + p_t} \\
&= \frac{R_t}{R_t} \frac{H_t + h_t}{P_t + p_t} \\
&= R_t \frac{P_t(H_t + h_t)}{H_t(P_t + p_t)} \\
&= R_t \frac{P_t}{P_t + p_t} \frac{H_t + h_t}{H_t} \\
&= R_t \frac{P_t}{P_t + p_t} + R_t \frac{P_t}{P_t + p_t} \frac{h_t}{H_t}
\end{aligned}$$

3 Strategy

3.1 Mixes Strategies

Should we make deterministic decisions?

Without loss of generality, we should probably start working directly using mixed strategies, where we assign a probability of cooperation to both our agent and the opponent's.

3.2 Opponent strategy

A naive idea is to have the opponent's probability of cooperation approximated as his own reputation, i.e. $pc' \approx R'$. This is very simple, but unfortunately it is also too suboptimal. If we take this assumption, then the optimal strategy becomes to always deflect.

More generally, we should aim at modelling his probability of cooperation as a function of not only his reputation, but also mine, i.e. $pc' \approx f(R', R_t)$.

3.3 Utilities

Utility Table

$$U = \begin{array}{|c||c|c|}
\hline
A \setminus A' & \text{H} & \text{S} \\
\hline
\text{H} & 0 \setminus 0 & -3 \setminus +1 \\
\hline
\text{S} & +1 \setminus -3 & -2 \setminus -2 \\
\hline
\end{array}$$

Expected Utility

Let's try to analyze our expected utilities, taking the naive approximation of pc'.

$$\begin{aligned}
E[U(A, A')/A = H] &= E[U(H, A')] \\
&= U(H, H)pc' + U(H, S)(1 - pc') \\
&= 0pc' + -3(1 - pc') \\
&= -3 + 3pc'
\end{aligned}$$

$$\begin{aligned}
E[U(A, A')/A = S] &= E[U(S, A')] \\
&= U(S, H)pc' + U(S, S)(1 - pc') \\
&= 1pc' + -2(1 - pc') \\
&= -2 + 3pc'
\end{aligned}$$

Notice how the second is always higher than the first. This simply confirms that in the non-iterated prisoner's dilemma (or in the case we assume the opponent doesn't base his strategy on our own reputation) the best strategy is always to slack.

When (if) we manage to model pc' also as a function of our own reputation, we might try this approach again..

3.4 Relative interpretation of Utilities

When is it worth it to risk by cooperating?

It can be argued that it is mostly unimportant exactly how much we win or lose in absolute terms, but rather it is important only how much we win or lose w.r.t. all the other possible outcomes.

This is pragmatically true only as long as we can avoid having very low stacks of food, in which the importance of considering the absolute terms arises. Consider, in the game at hand, that your action may change the outcomes either from 1 to -2 or from 0 to -3. In both cases, the relative change is always -3, and as such the analysis is simplified. However, if you actually do have few items of food left, then the difference between the absolute terms actually starts to gain importance. However, I find two reasons why this can be largely ignored: *a*) If you have so few items of food left, you are practically already doomed. You will die inevitably unless you find a very long streak of collaborators in front of you, a very unlikely event. Even so, you will only either stay at the same food level (if you collaborate with them), or gain 1 very puny food item, at the cost of decreasing your reputation for each puny gain. *b*) The game starts with a food stack in an order of magnitude which I estimate to be between 10^4 and 10^6. As such, the absolute difference of 1 between the outcomes is largely ignorable. *c*) More importantly, if we ever do get at such low levels of food, it means that our strategy was a losing one to start with, and that can not be attributed to the change of P.O.V from absolutist to relativist utility interpretations.

So let's assume that we are only interested in a relative interpretation of utilities. As such, the game's utilities can be translated as:

- For each turn, we lose 2 items of food consistently. This is the term which can be pragmatically ignored.
- We lose 1 extra item of food per session if we choose to cooperate.
- We gain 3 items of food per session if our opponent chooses to cooperate.

In a table representation, the relativist P.O.V of utility simply means that we can adjust all the utilities by adding/subtracting a constant:

$$U = \begin{array}{|c||c|c|} \hline A \setminus A' & \text{H} & \text{S} \\ \hline \text{H} & +2 \setminus +2 & -1 \setminus +3 \\ \hline \text{S} & +3 \setminus -1 & 0 \setminus 0 \\ \hline \end{array}$$

It is thus easy to postulate that any agent should only be interested in adopting strategies which include cooperation component if and only if these allow the agent to receive at least 1 cooperation for every 3 which he gives away. This aspect is fundamentally true, and should be included in the final strategy.

If we port this in the framework of mixed strategies, then we can say that an agent should adopt cooperation-inclusive strategies if and only if the probability of receiving cooperation is at least one third of the probability of giving out cooperation.

Is this correct? Not sure.. Think more about it.

If these guidelines are not guaranteed, then the agent performs best to simply avoid all sorts of cooperation at all.

Conclusion

Choosing a policy which maximises the previous expectation leads to a static policy of always slacking, which overall would lead to a consistent loss of 2 items of food for every turn.

However, we will see that maybe we are not really using all the information we have. Up to now, we have assumed that the opponent's action is only dependent on his responsibility, but we have left out that his actions in front of *us* are very likely to depend on our own responsibility.

The next section will try to analyse why and how our reputation changes our opponents' actions, and how to make use of this to increase our expected food value.

3.5 Reputation Balancing

Why?

We have already seen that slacking is the Nash Equilibrium solution, i.e. the local optimum when we try to maximize the expected utilities. Unfortunately, this is far from a global optimum: Being this a prisoner dilemma problem, we have that both players gain by cooperating, but only if they both cooperate. This is the problem we are trying to address.

It is reasonable to assume that *some* players will try to cooperate so that they all gain by hunting together. If such a group forms, then I would rather be in it than out of it. The intended goal is to avoid being classified as a "slacker": If you are a slacker, then everyone will slack against you.

The problem is that there is no means of communication with the community so as to organize the up-coming of such a group: The only information we can exchange is our reputation. The goal of reputation balancing should be to send a message to your opponent: "look: you can trust me".

On the other hand the creation of such a group opens ground for all sorts of treason and backstabbing. Confirming yourself as a strong hunter is also not a good strategy: If you are a compulsive hunter then other people will assume they can slack their way into profit at your expenses. All in all, we just want to keep our reputation just high enough for that to the message to pass. We gain the maximum when we cooperate just enough to let the message pass. The question now is: *what is the threshold?*

One other interesting thing to consider is that the message we pass is only effective if there is someone listening on the other end. If we are surrounded by compulsive slackers, then there is little we can do but slack ourselves[1].

What value to aim for?

This is a very tough question.

Assuming we would simply want to avoid the opponent to predict our own decision, then we would want to maximize the future entropy

$$H^{entropy}(R_{t+1}) = -R_{t+1}\log(R_{t+1}) - (1 - R_{t+1})\log(1 - R_{t+1})$$

To maximize, we simply aim at having $R_{t+1} = 0.5$, which is quite intuitive considering our goal of avoiding possible predictions by our opponent.

However, there's still problems with this approach: *a*) We are only maximizing entropy under the assumption that your opponent takes your reputation as a naive approximation of your probability of cooperating. *b*) What are exactly the advantages of maximizing your opponent's entropy? This is not poker, if two people want to cooperate it is not enough to come out as a guy who might back-stab you 50% of the times. I strongly suspect that the value of reputation we should aim for is higher than .5.

How?

This highly depends on what the relationship is.

[1] This is the reason why might want to keep a global reputation measure at hand, but more on that later.

3.6 Public Goods

Are Public Goods good or bad?

> “In many realistic situations there are benefits to cooperation that accrue to all players, but only if a certain number of players cooperate. These benefits are called public goods. A typical public good is a community playground built by donations. The playground can be built only if enough people help, but the playground benefits everyone equally. Public goods tend to increase the chance that people will cooperate.”

from *Training Exercise*

The only serious observation I can make regarding the issue of public goods is that rich people don’t want to achieve public goods, whereas poor people do. The reason is simple: Public goods affect everyone positively. As such, rich people would be losing their dominance against poorer people, and the other way round. For example, consider the difference between two players who have 100 and 5 respectively and who have 900 and 805 respectively. In the former case, the richer guy is dominating; in the latter, the dominance is daft.

3.7 TODO

What is missing?

The most crucial thing which is still missing is how to model how our reputation influences the opponent’s chance of cooperation..

4 Final Strategy

Given the above considerations, the idea seems to arise that the only important thing to decide at each turn is h_t. Given the relativist P.O.V of utility, it doesn’t matter exactly with *who* we collaborate or not, but only *how many times* we do. This is enforced and assured also by the fact that our opponents can not identify us, and never even receive information about which action ours was.

Thus the idea of the final algorithm is the following:

- Determine what kind of reputation maximizes the number of collaborations you receive, and determine if that kind of reputation is too expensive to get (more than 3 of your collaborations to get 1 of theirs).
- Determine how many collaborations you need to give out to achieve that reputation.
- Scatter the collaborations across the opponents. There is only one aspect w.r.t. which it does matter to who you do or you do not collaborate: Try to reward the slackers rather than the hunters, because the (successful)

hunters are likely to already have more food than the slackers. To successfully undermine the opposition, only give advantage to who doesn't pose a threat[2].

[2] Funny observation: If too many people reason like me, then the slackers would totally win.