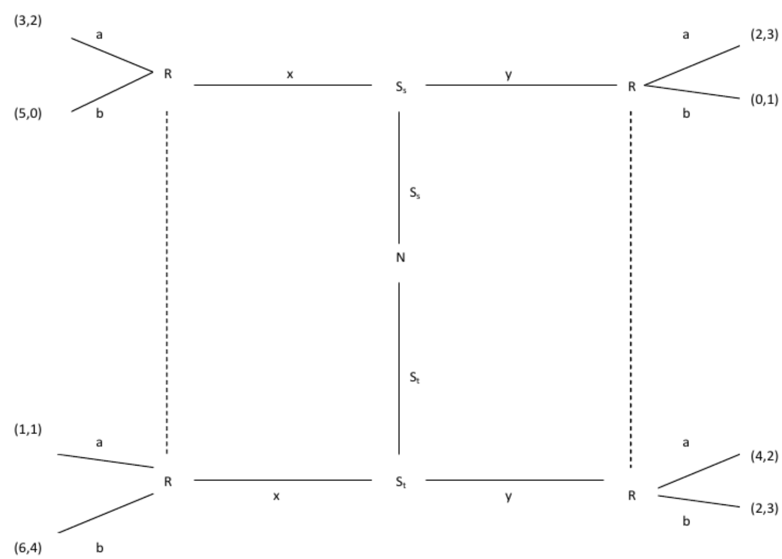
Note on finding equilibria in signalling games

Adlai Newson, adlai.newson@gmail.com

For this example I'm going to illustrate a general method for finding equilibria of a signalling game. I'm using question 6 from problemset 4 because it's the 'classic' signalling game setup, but these methods can apply to many types of signalling games.

Here's the game tree:



and the prior is $P(S_s) = \frac{3}{4}$.

Question 1: Find all the seperating equilibria, or show that no seperating equilibrium exists. Just a quick refresher, a seperating equilibrium is one where every different type of Sender (S) player will play a different action (send a different signal). This means that when the Reciever (R) player sees an action, he knows what type must have played that action. In general an equilibrium requires that:

> *Given the actions and beliefs of the other players, I can't be strictly better off if I play a different action.*

Generally, you will have to take every possible seperating equilibrium, then check to see if it satisfies the equilibrium condition above. So, what are the possible seperating equilibria? There are two:

(1) S_s plays x, S_t plays y
(2) S_s plays y, S_t plays x

Of course, it's also possible that there is no seperating equilibrium. So how do we find out which one (or both, or neither) is the seperating equilibrium? We're going to use this algorithm, which works not just for seperating equilibria, but *all kinds of equilibria* in signalling games. I'm putting it in a box because it's important:

Algorithm for checking an equilibrium in a signalling game:

(1) Assume the strategies of the sender players, S_s and S_t
(2) Construct the beliefs of the Receiver player (R) given the strategies in (1)
(3) Find the best responses of the R player, given the strategies in (1) and beliefs in (2)
(4) Given the best reponses of the R player, check to see if S_s or S_t have an incentive to deviate from the strategies we assumed in (1).

If anyone has an incentive to deviate in (4) then you're done - what you assumed in (1) is not an equilibrium. If instead you check all the S types (S_s and S_t) and neither has an incentive to deviate, then you've found an equilibrium.

Let's go through an example of the algorithm to see it in action. Let's pick the first proposed seperating equilibrium and see what happens.

(1) Assume that S_s plays x, S_t plays y
(2) Given (1), the beliefs of R will be $P(S_s|x) = 1$, $P(S_t|y) = 1$, and the other two are zero.
(3) Given R's beliefs, if R sees x, then he thinks he is playing against S_s, so his best reponse is a (because in the top-left game, 2 is bigger than 0). Again given beliefs, if R sees y, he thinks he's playing against S_t, so his BR is b (the bottom-right game). In math:

$$\begin{aligned} R^{BR}(x) &= a \\ R^{BR}(y) &= b \end{aligned}$$

(4) First consider S_s. If he plays x like we assume he does, he will get 3. If he deviates and plays y, then the R player will think that he is actually S_t, and R will play b, so S_s will get 0. In math,

$$\begin{aligned} \pi_{S_s}(x|R^{BR}(x)) &= 3 \\ \pi_{S_s}(y|R^{BR}(y)) &= 0 \end{aligned}$$

The first equality there says "The payoff to S_s from playing x, given the best response of R to x, is 3." So S_s prefers to stick to the equilibrium strategy, x. If we do the same for S_t, we get

$$\begin{aligned} \pi_{S_t}(y|R^{BR}(y)) &= 2 \\ \pi_{S_t}(x|R^{BR}(x)) &= 1 \end{aligned}$$

Since no player has an incentive to deviate, we have found a seperating equilibrium!

So we found one equilibrium, and now let's check to see if the other one is also a seperating equilibrium.

(1) Assume that S_s plays y, S_t plays x
(2) Given (1), the beliefs of R will be $P(S_t|x) = 1$, $P(S_s|y) = 1$, and the other two are zero.
(3) Given these beliefs, here are the best responses:

$$\begin{aligned} R^{BR}(x) &= b \\ R^{BR}(y) &= a \end{aligned}$$

(4) Given these best responses, let's check for deviations for S_s. If S_s plays y like we assumed he would, he would get 2. If he deviates and plays x, he gets 5. That means that there's an incentive to deviate, so we can stop here - this can't be an equilibrium!

It's also important to note that we could have noticed that playing x for player S_s is a **dominant strategy**. Why is that true? Because the minimum possible payoff that S_s could get by playing x is bigger than the maximum possible payoff from playing y. Since when a player has a dominant strategy, every equilibrium must have that player playing his dominant strategy, we could have noticed this, then concluded that the only possible seperating equilibrium is the first one, and only checked whether S_t had an incentive to deviate from y. **In general, it is a good idea to check for dominant strategies before you start checking all the cases.** But the four steps we used above will work whether or not there is a dominant strategy, which is why I wanted to go through it - on the exam, there may or may not be a dominant strategy, it just happens that in this game there is one.

ANSWER: The unique seperating equilibrium is when S_s plays x, S_t plays y, and R plays strategy $s_R = (s_x, s_y) = (a, b)$, and $P_R(S = S_s|x) = 1$, $P_R(S = S_t|y) = 1$.

Question 2: Find all pooling equilibria, or show that no pooling equilibrium exists. Now we can again use the algorithm, but there's a slight problem: how do we construct the beliefs for the R player when considering the off-path action? What do I mean? Let's try the algorithm and see what happens:

(1) Assume that both S_s and S_t play x
(2) Given this, by Bayes rule, we have $P_R(S_s|x) = \frac{3}{4}$ and $P_R(S_t|x) = \frac{1}{4}$. But what does R think when he sees someone play y? We can't use Bayes rule (try it and see what happens). Let's skip this and try to move on.
(3) Ignoring the best reponse of R when he sees y, let's figure out what he should do if he sees x:

$$\begin{aligned} EU_R(a|x) &= \frac{3}{4}(2) + \frac{1}{4}(1) = \frac{7}{4} \\ EU_R(b|x) &= \frac{3}{4}(0) + \frac{1}{4}(4) = 1 \end{aligned}$$

Since $\frac{7}{4} > 1$, we have that $R^{BR}(x) = a$.
(4) Now we check for deviations. Since S_s always does better by playing x, we don't need to worry about him. What about S_t? Given that R will play a, his equilibrium payoff will be 1. But what will R do if S_t deviates and plays y? In this example, *it doesn't matter whether R plays a or b when he sees y, S_t has an incentive to deviate.* Just check the payoffs to see that this is true. That means that there can't be a pooling equilibrium.

So that's lucky for us that in this example we didn't need to worry about what the R player would do when he saw something 'off the path.' But in general what's the rule about off-path beliefs? It may sound a bit strange, but you can assume any off path beliefs you want in order to support (make rational) the equilibrium. For example, if I needed to, I could have assumed that $P(S_s|y) = 1$ or $P(S_s|y) = 0$; whatever would make the equilibrium work. That's a bit of a side note, probably you won't run into this for the rest of this course, but it's something to be aware of.

ANSWER: By the proof above, no pooling equilibrium exists.

Question 3: Find all semi-pooling equilibria, or show that no semi-pooling equilibria exist. Again we're going to use the four-step algorithm. A semi-pooling equilibrium is one where one type of S player plays a pure strategy and the other plays a mixed strategy (you saw at least one example in your homework). The R player can either play a mixed or pure strategy.

Since we know that S_s is always going to play x, the only possible semi-pooling is when S_s plays x and S_t mixes, but we don't know whether R will play a pure or mixed strategy.

(1) Assume that S_s plays x, S_t mixes between x and y with probabilities α and $1-\alpha$

(2) Given these, by Bayes rule, the beliefs of R are

$$\begin{aligned} P(S_s|x) &= \frac{\frac{3}{4}}{\frac{3}{4}+\alpha\frac{1}{4}} = \frac{3}{3+\alpha} \\ P(S_t|x) &= 1-\frac{3}{3+\alpha} \\ P(S_t|y) &= 1 \\ P(S_s|y) &= 0 \end{aligned}$$

(3) Given these beliefs,

$$\begin{aligned} EU_R(a|x) &= \left(\frac{3}{3+\alpha}\right)(2)+\left(1-\frac{3}{3+\alpha}\right)(1) \\ &= 1+\frac{3}{3+\alpha} \\ EU_R(b|x) &= 0+\left(1-\frac{3}{3+\alpha}\right)(4) \\ &= 4\left(1-\frac{3}{3+\alpha}\right) \end{aligned}$$

So that it's rational for R to play a when he sees x if

$$\begin{aligned} 1+\frac{3}{3+\alpha} &\geq 4\left(1-\frac{3}{3+\alpha}\right) \\ &\implies \alpha \leq \frac{3}{4} \end{aligned}$$

NOTE: I made a mistake in this calculation; the inequality should be $\alpha < 2$. Since $\alpha < 1$ by assumption, that means R will always play a (when he sees x), so R will never mix, which in turn means there is no semi-pooling equilibrium. I'm going to leave the mistake in, though, because it shows you how to find a semi-pooling equilibrium (but the real answer is that one doesn't exist).[a]

Notice that if $\alpha = \frac{3}{4}$, R might play a mixed strategy.

From the beliefs we also know that when R sees y, he believes it must be S_t, so $R^{BR}(y) = b$

(4) Given the best responses, let's see if S_t finds the mixing strategy we assumed in (1) to be rational. What's the condition we have to check to see if S_t is being rational by mixing? He must be getting equal payoffs from playing x and y. Let's suppose that R plays a when he sees x with probability β. Then the condition for S_t to mix is

$$\begin{aligned} EU_{S_t}(x|R^{BR}(x)) &= \beta(3)+(1-\beta)(1) \\ &= 1+2\beta \\ EU_{S_t}(y|R^{BR}(y)) &= 2 \end{aligned}$$

so S_t will mix when $\beta = \frac{1}{2}$. Again, we don't have to check whether S_s is being rational because he is playing his dominant strategy. And that's the equilibrium!

[a] Thanks to student Youcheong for spotting the mistake here in the old version

ANSWER: Here's the equilibrium (the $\mathbf{s}$ notation just denotes 'strategy'):

$$\begin{aligned}
\mathbf{s}_{S_t} &= x \text{ wp } \frac{3}{4},\ y \text{ wp } \frac{1}{4} \\
\mathbf{s}_{S_s} &= x \\
\mathbf{s}_R &= \left(\left(a \text{ wp } \frac{1}{2}, b \text{ wp } \frac{1}{2}\right), \quad b\right) \\
P_R(S = S_t \big| x) &= \frac{4}{5} \\
P_R(S = S_t | y) &= 1
\end{aligned}$$

Let's sum up what we did for this more complicated case. We used the algorithm and found that the best response of R depended on the mixing probability α. If α is big, R plays b, if it's small, R plays a, but in both of these cases[1] it won't be rational for S_t to play a mixed strategy, so we would end up with a contradiction of our first assumption: that S_t is mixing. Therefore, we set α so that R will mix, then found β that rationalizes S_t playing a mixed strategy.

This may seem like a fairly complicated procedure, but the good news is that it will always work. Probably any question you will get will be similar to this example or a little bit harder; I probably went into more detail than you will need. The important thing is that you understand how to check whether something is an equilibrium or not.. an exam question probably won't be "Find all of the semi-pooling equilibria (Hint: there are infinitely many)" but it might be something like "Find the semi-pooling equilibrium where player two always plays Left." ie: you will be given some information, and you will have to show that there is an equilibrium there... if you figure out the possible cases and apply the algorithm above, you will ace it :)

[1] Check to see if this is true