Numerical Politics

Numerical experiments in the politics of interacting, autonomous agents

Daniel Tang

May 3, 2023

Contents

Chapter 1

Introduction

Numerical politics is a new approach to the study of politics in which we perform numerical experiments on simulated societies in order to gain an understanding of how organised, collective behaviour can emerge among interacting agents. This github repo (`https://github.com/danftang/NumericalPolitics`) presents a set of numerical experiments that allows us to build an understanding of this emergent behaviour while also presenting the software and numerical techniques necessary to perform these experiments. Our approach will be to start with very simple agents in simple environments and gradually introduce more complex behaviours and complex environments. This will allow us to build up our theoretical understanding and introduce new numerical techniques in a logical order.

Eventually, this repo is intended to provide the software necessary to allow anyone to set up a numerical "laboratory" and start studying simulated societies. In these pages we'll also discuss how the practice of numerical politics can contribute to our understanding of how to effectively structure and govern real-world societies.

1.1 The framework: thinking clearly about collective behaviour

The subject matter of numerical politics is the collective behaviour of many interacting agents. Agents interact by passing messages between eachother. Each agent has a number of "channels" for receiving messages and any agent with the channel's ID can send a message down the channel. When an agent receives a message, it can respond by updating its internal state and/or sending yet more messages. More formally, the behaviour of an agent can be defined as a probability distribution, $P(a\|m, c, \psi)$, which is the probability that an agent in state ψ performs action a in response to receiving message m in channel c. The action a defines the agent's new internal state and/or a set of messages passed to other agent's channels at some time after receipt of the message. At first sight this may seem a bit abstracted from real-world application, but we choose it because it's a very flexible formalism that can easily be adapted to all applications.

Our particular interest here will be to make predictive statements about individual and collective wellbeing of the agents. For this we assume there exists a function, $W(\psi)$, that is some measure of the wellbeing of an agent in state ψ and define the collective wellbeing of a set of agents, S, as the sum of individual wellbeings $\Omega = \sum_{\psi \in S} W(\psi)$. It should be the subject of much debate exactly what the function $W(\psi)$ ought to be.

Notice that in this definition there is no mention of government. A goverment, if there is one, is encoded within the behaviours of the agents and is part of the model, as opposed to it being an exogenous actor imposing "policy interventions" on the agents. In this way, a government is best thought of as an emergent property of the agents' behaviours. We choose to make government endogenous because our interest here is *not* to simulate specific policy interventions but to understand the fundamental principles of organised, collective behaviour.

Chapter 2

Sugar and spice world (version 1)

We begin with an anarchic world with two commodities: sugar and spice. Suppose half the agents can farm only sugar and half can farm only spice, each at a rate of one unit per unit time. Suppose the agents randomly encouter eachother, whereupon each agent can either offer to trade or try to steal the other agent's crop. If both agents offer to trade then half the crop of each agent is swapped, however, if one agent offers to trade and the other tries to steal then the stealing agent gets half the others crop and the other is left with only half its original crop. If both try to steal then they are both unsuccessful and no food is transferred. This is the classic prisoner's dilemma, an agent's wellbeing after an interaction is a function of the two agent's actions and is given in table 2.1.

This world is simple enough that we can see immediately that the optimal collective wellbeing occurs when all agents trade. In this case, the average wellbeing of all agents is 3. However, under what circumstances will agents reach this optimum?

If we assume the agents are Q-learning then we can ask what kinds of society do the agents create for themselves using their learning. More mathematically we can look at the society as a dynamic system and ask about the distribution of wellbeing on the attractors. In the special case where the attractor is a point, we have a stable society where no amount of learning from further encounters will change any agent's policy.

Agent 1	Agent 2	Agent 1 wellbeing
trade	trade	3
trade	steal	0
steal	trade	4
steal	steal	1

Table 2.1: The wellbeing of an agent after an interaction

2.1 Zero memory agents

If agents have no memory of previous encounters then each encounter is a simple prisoner's dilemma situation. The state of a Q-learning agent is just the Q-values of trade, Q_t and steal, Q_s. Equilibrium is when

$$\begin{aligned} Q_t &= 3P(t) + r\max(Q_t, Q_s) \\ Q_s &= 4P(t) + P(s) + r\max(Q_t, Q_s) \end{aligned}$$

but

$$Q_s - Q_t = P(t) + P(s) = 1$$

so $Q_s > Q_t$ irrespective of the other agent's behaviour so a zero memory Q-learning agent will always learn to steal, leading to a society where all agents try to steal and every agent is much worse off than in a trading society, with an average wellbeing of 1.

So, memoryless Q-learning agents get stuck in an equilibrium that is far from optimal both collectively and individually.

What needs to change in order to improve these agent's lives?

2.2 One step memory agents

If we give the agents the ability to remember the last encounter they had with another agent (if this isn't the first encounter) then the dynamics of the society gets much more interesting.

2.2.1 Experiment 1: two agents

We start with the simple case of a society of just two agents. Experiment 1 in the accompanying repo shows that there is more than one attractor in this world (i.e. the society is non-ergodic) and that all attractors are points in policy space. The society where both agents always try to steal from eachother is still an attractor, but their memory of the previous encounter introduces other attractors.

Another attractor has agents adopting the policy of always trying to steal unless they both traded last time. In this society the agents spend most of the time trying to steal from eachother, but if they both happen to explore the trade option at the same time, they will enter a period of sustained trading until one of them explores the steal option. So, average wellbeing is slightly better than the always-steal case, but still far from optimal.

Interestingly, there is also a third attractor where agents can learn to both adopt the policy described in table 2.2. The first three entries in the table have analogues in human behaviour, but the last is a little unintuitive: if we both tried to steal from eachother last time, then this time I'll try to trade. This is key to the success of the policy as it means that whatever state the agents get into they quickly revert to mutual trading. As the exploration probability tends

My last move	Your last move	My next move	Human trait
trade	trade	trade	mutual-benefit
trade	steal	steal	revenge
steal	trade	steal	exploitation
steal	steal	trade	?

Table 2.2: The optimum behaviour of an agent with one-step memory

to zero, this society tends to the optimum of always trading, while remaining unexploitable (if I always try to steal from an agent with this policy, we'll flip between mutual stealing and me stealing from the agent, but my average wellbeing will be 2.5, less than if I take on the policy in table 2.2).

Note that the agents do not learn the tit-for-tat policy: if you tried to steal from me last time, I'll try to steal from you this time, but if you traded with me last time, I'll try to trade with you again. Mutual adoption of this strategy has three stable states: mutual trading, mutual stealing and alternating stealing. So, the agents will get into long runs of non-optimal behaviour. When faced with this behaviour, a Q-learner will learn to avoid this non-optimality by unilaterally adopting the policy in table 2.2, which leads to better average wellbeing for both agents.

2.2.2 Experiment 2: many agents

Experiment 2 shows what happens as this society gorws.