CX 4230 Project 3B: Conceptual Model

Karan Shah and Kartikeya Kumar
For Dr.Vuduc, Spring 2016

Conceptual Model

The conceptual model has been organized as follows:

- Introduction: What is it, who is it for?
- Structure of our world
- Initialization
    - Distribution
    - Properties
- How changes occur from each time step
    - Population change
    - Property change
    - Migration
- End of the simulation
- Shortcomings
- References

Introduction

Our project is a simulation of human population movement/migration both intra country and intercountry with a lot more parameters. We want to see how changing various factors that affect the quality of life of people makes them want to move to another place. What makes it different from existing studies is that it is a very dynamic system. The population as well as the conditions in different regions are not constant. Instead, they change according to mathematical models (differential equations and stochastic change). Also unlike some of the models we read about in the review above, we will be experimenting with political constructs like borders and wars.
If the model ends up simulating the real world, it can be used by sociologists to see how changing certain parameters can affect migration patterns. If it doesn't, it can be used to develop games (and alternate world histories).

Structure of our World

Our world can be thought of as a large 2D cell array. Each cell can be assumed as a sub-region and a block of contiguous cells can form a country. Each cell will have a randomly initialized population in it represented by a population counter variable. Also all cells have the same carrying capacity k which will be the max population limit of the cell.
Each cell will have a vector of properties such as a wealth score, climate score, political freedom score, a population growth rate, a death rate etc. These scores will be combined to form a composite 'desirability score' ranging from 0 (inhabitable) to 1(best place ever to live in) for that state.

Now, each country, as a whole, too will have the same list of properties that each of its cells or states within it have. Each country will have a state variable that determines if it's economy is booming or recessing and if it's at war or peace time. We can program these states to change randomly.

So, there will be a desirability score for both the states within a country as well as the country itself.
The desirability factor here has a population component such that a specific population in each cell is more desirable and desirability decreases as the population increases or decreases from the peak value. Also, each of these countries can be separated from each other by an ocean which is represented by empty cells with no properties.

Initialization of the World

(Still researching: We are looking for a suitable map generation algorithm. We plan on hard coding the structure right now, but we do have some ideas for the generation if time permits[1])

Distribution of properties:

Let's say that we have n countries with each country having k properties. Now, for initialization, we will hardcode the value of each of the k properties for the n countries. These macroproperties serve as means for the distribution of cell properties. The values of these properties go from 0 to 1.

Let p be the macroproperty. First we create a Gaussian distribution with mean p, standard deviation 1 and use all the values that fall in the range [1,0]. The function we created is:

$$P(x) = \frac{e^{-\frac{1}{2}(-6p+6x)^2}}{\sqrt{2\pi}}$$

The size of this distribution is determined by p. If there are n cells in a country, we generate $n' = \frac{n}{(1-|0.5-p|)}$. This is because we want atleast n values of the property in the region [0,1]. Here are some distributions with different values of p.

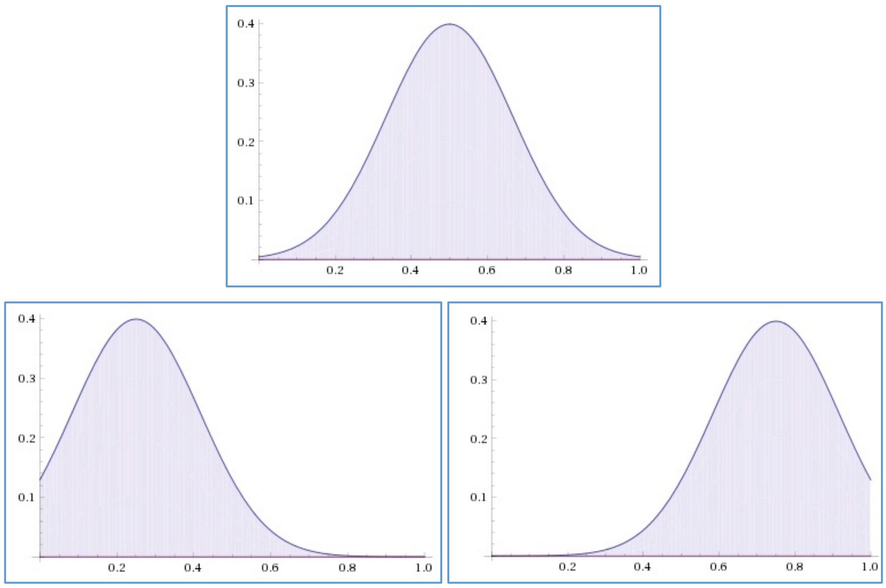

Figure 1 Distributions with different means p = 0.5, 0.75 and 0.25 clockwise from top. These distributions are limited to the range[0,1]. Images ploted in Mathematica with command “region ((e^(-1/2 (-6p+6x)^2))/sqrt(2 π))>0 from 0 to 1”.

Once we have n values that fit the distribution, we randomly assign them to the n cells in the country.
We repeat this process for all the properties that have this distribution.
We can also use the Pareto distribution[2] to make a more realistic model or even try different distributions like the skewed normal distribution.

This algorithm (by taking the value of the property of a country as its mean for states within that country) represents the real world notion that a poorer country, on average, will have poorer states within it as compared to states within a richer country. The simulation can eventually be repeated a certain number of times to average out the results as the initial state can make a lot of difference to the final results.

Properties of cells

Population: Population will be assigned by the Rank-Size distribution[3]. It is a variant of power law distribution. According to the rank size trend, if we plot the Population rank of countries and their Sizes on a log-linear scale, we get a line with slope = -1. We can have a global variable ρ which is the population density of the world. If there are n land cells on the map, the total initial population will be $n\rho$.

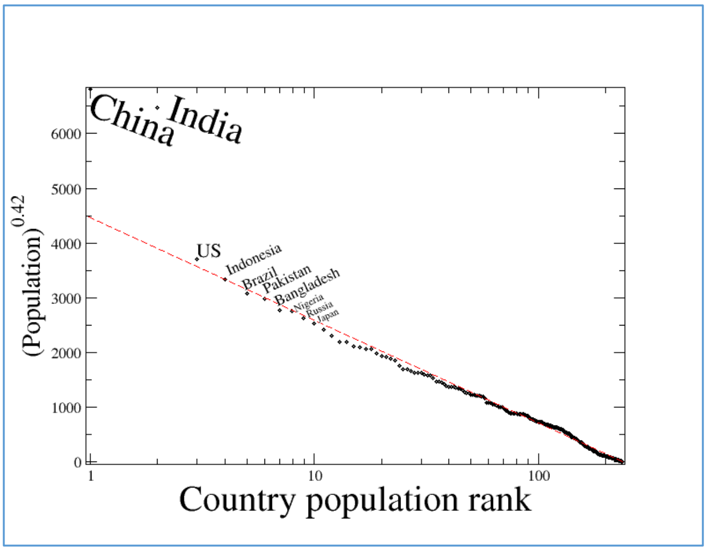

Figure 2 Country population rank (log scale) vs Population[3]

This has also been discussed in the literature review[4].

Another simpler option is to just assign a population of $n_i\rho$ where n_i is the number of land cells in country i such that

$$\text{Total population } n\rho = \sum_i n_i\rho$$

Once we determine the population for each country, we can use the normal distribution described above to assign initial populations to all cells.
The population of each cell also affects it desirability. People will find a cell with population k/2 the most desirable and desirability will decrease towards both 0 (nobody wants to live in an uninhabited desert) and k (overcrowding).

Wealth: Wealth can be assigned to countries with 1 being that the majority of the country will be super rich, 0.5 being the wealth follows a perfectly normal distribution and 0 being the majority of the country will be penniless. The wealth assigned to the country is used as a mean for the distribution above to set wealth for all cells. People will try to move so as to maximize the wealth difference if the destination cell is more wealthy.

Climate: Climate will be a linear function of "latitude", that is the row number. We assume that the lowest row is equator and the first row is the tundra. So, all the cells in the same row will have the same climate value. This can be assigned by setting a temperature of $30°C$ at the equator and $5°C$ at the poles and setting all the other values in between linearly. People will try to move so as to minimize the climate difference.

Birth and Death rate: We assign 4 variables, 2 for birth rate (b_1 and b_2) and 2 for death rate ($d_1$and d_2).

We have the contstraint: $\frac{b_1-d_1}{b_2+d_2} = k$, the carrying capacity of each cell. The meaning of these variables will be more clear in the next section.

Other factors that can be assigned are mobility (harder to move across borders) and less quantifiable factos such as social mobility (eg South Korea vs North Korea).

We set the economies of all countries to be booming and the world is at peace initially.

State Transitions

Our simulation can go through time steps with each time step representing a year. There are three main changes occuring at every time step:

Population growth

People are born and die at every time step. The change in the population, at each time step, for each region, is modelled by[5]:

$$B(N) = b_1 N - b_2 N^2$$
$$D(N) = d_1 N + d_2 N^2$$

Here, B(N) and D(N) are the number of births and deaths repectively for a population of N people. The coefficients are properties of the country. The N^2 in the equations shows that for higher N, there will be a decrease in birth rates because of resouce competition and higher death rates because of overcrowding and disease spread.

We now make a differential equation:

$$\frac{dN}{dt} = B(N) - D(N)$$
$$= (b_1 - b_1)N\left(1 - \frac{b_2 + d_2}{b_1 - d_1}N\right)$$

which is a logistic equation ($\frac{dN}{dt} = rN(1 - \frac{N}{K})$) with $r = (b_1 - b_1)$ and $K = \frac{b_1-d_1}{b_2+d_2}$

We perform this on each cell. We can make a matrix to solve this equation numerically.

Pertubations in properties

Each country, at a given time step, will in a ‘state’ of either –prosperous, recession, war″. These states can randomly be changed after a certain number of time steps for any of the countries. We add a random factor pertubation factor ε to different properties depending on this state. For example, the population growth rates change, birth rate increases by ε during prosperity and death rates go up during war. Similarly, wealth score increases in prosperous times, is stagnant in recession and is negative during war.

Theses tiny pertubations change the desirability of all cells in a country and lead to more chaotic migrations. We have to take care that the properties stay within bounds and follow required contraints.

Migration

Let the population of cell i be P_i. We define a function desirability of a cell j to the population of cell i:

$$\delta_{ji} = \frac{1}{d_{ij}} \sum_k w_k F_k(p_{ki}, p_{kj})$$

The function $F_k(p_{ki}, p_{kj})$ gives the difference between the kth peoperty of p of cells i and j. So, for a property like wealth, F will return a the difference between the wealth of j and i. This means if the wealth at j is higher, F will be higher. Similarly, for a property like population, F will return a higher value if the population at j is near the peak value $k/2$. For a property like climate, F will return a higher value for the least difference in i and j. w_k is a weight of the property in order of importance. In this model, we assume all people have the same priorities and so the weights will be same for all people. The summation over k simply creates a score that is composite of all the properties. We define $\sum_k w_k F_k(p_{ki}, p_{kj})$ such that it give a value between 0 and 1.We divide by d_{ij}, the distance between i and j because it is harder to move as the distance increases.

Ideally at every time step, each cell i in a country checks the desirability of all other cells j of that country and we move a small proportion of the population to the cell with the highest index. We can use the desirability index itself to move that population. So at every time step, a population of $P_i d_{ij*}$ moves from i to $j*$ where $j*$ is the place with the highest desirability factor in the country.

Also, each cell also checks with the desirability of countries by creating a desirability score out of country properties that is independent of the distance. We can use a small random number ϵ and transfer $P_i\epsilon$ from i to a random cell in the country with highest desirability for that cell.

Our migration modeling is similar to the Push Pull[6] factor paper but it also takes into account dynamic changes in properties and population.

So, at each time step, we will first do the population update (births/deaths) and other for each cell. Using the updated population, we will then calculate the desirability of each cell using the formula given above. After this, for a given cell, we look at the desirability of all the other cells within the same country and on the desirability of all the other countries.

End of the simulation

Because of the number of factors involved in determining the desirability, we don't think that it will reach a steady state easily. Also, random events like war make it difficult to reach a steady state. One way to reach a steady state is if the population reaches the carrying capacity of the system or if everybody dies for some reason. Otherwise, the simulation can be ended by fixing a certain number of time steps. At the end, a visualization across different time steps of a heat map of the world representing size of populations at different time steps for all the countries can be shown to see how migration changed the world population.

Shortcomings

The biggest shortcoming with this model is that it is computationally very intensive. At each time step, each cell in the system calculates the desirability of every other cell in it's country. This is like an N-body problem. We will try implementing N-body problem solutions to this. There are no other papers that we know of that have used N-body approach to migration. We might have to implement a simpler version of this model because of the time constraints.

We have tried to implement our system on an aggregate basis and not at an individual level. By looking at the desirability of a location, we move a certain number of people from one region to another, incrementing and decrementing the population counters accordingly. So, we are never really looking at an individual making a migration choice. The reason we did not want to implement that was because keeping track of thousands of individuals moving across our world was way too computationally intensive even with some of the faster data structures we researched.

We are not experts on human behavior. We used simple math concepts to model human behavior whereas human behavior has a much stronger random component. The model also assumes that everybody has the same choice in places to move and has equal access to information, which is not true of the real world. Thus, we don't expect this model to work very well in the real world.

There are thousands of other factors that affect migration. In that sense, this is a much simplified model. Also, we don't know how the model we proposed will behave when we add a lot of variables (huge number of countries, big grid, many timesteps).

One thing we fear is that the presence of randomly generated high desirability places will serve as attractors and everybody will try moving there. We have added some counter measures in the model (extreme population densities decrease desirability) but this could still happen when we actually run the system.

References

[1] Anon. Looking for a good world map generation algorithm.
Retrieved April 4, 2016 from
http://stackoverflow.com/questions/2520131/looking-for-a-good-world-map-generation-algorithm

[2]Anon. Pareto Distribution. Retrieved April 4, 2016 from http://mathworld.wolfram.com/paretodistribution.html

[3]Anon. Rank-size Distribution. Retrieved April 4, 2016 from https://en.wikipedia.org/wiki/rank-size˙distribution?oldformat=true

[4]David A. Young. 2002. A New Space-Time Computer Simulation Method for Human Migration.*American Anthropologist* 104, 1 (2002), 138–158.

[5]Anon. Logistic population growth. Retrieved April 4, 2016 from http://raven.iab.alaska.edu/˜ntakebay/teaching/programming/logistic/node1.html

[6]Guido Dorigo and Waldo Tobler. 1983. Push-Pull Migration Laws. *Annals of the Association of American Geographers* 73, 1 (1983), 1–17.