Scaling and Entropy

How Can We Derive Models We Consider Appropriate –
i.e. Based on Scaling – From Ideas About Entropy

Michael Batty
University College London

m.batty@ucl.ac.uk
http://www.casa.ucl.ac.uk/

Outline

- What is Entropy
- Entropy-Maximising Reviewed Again
- The Mathematics
- The Classic Negative Exponential – as Good as an Inverse Power Law? Preferable Even
- A Phenomenological Demo – In Class
- Scaling: How Entropy Generates Power Laws
- Examples in Cities: Cost and Size

What is Entropy?

At one level, you don't need to know what it is. You just need to be familiar that there is a technique of maximising a quantity subject to known information – constraints

You could think of this quantity as Accessibility or as Utility – in fact many people do.

Maximising utility is easy enough to understand

Now there are some very useful insights if we think of entropy as information – and we will do so as Wilson (1970, 2010) does.

So we maximise information rather than entropy but there are some really interesting issues about entropy and thermodynamics that we don't have time to go into here. To give a taste of these, we need to look at the properties

So this is a bit of digression to begin with but let us not forget that this mysterious quantity called entropy is not widely understood even by physicists, perhaps especially by physicists.

Von Neumann’s to Shannon in 1948 says it all:

“You should call it entropy, for two reasons. In the first place your uncertainty function has been used in statistical mechanics under that name, so it already has a name. In the second place, and more important, no one really knows what entropy really is, so in a debate you will always have the advantage!”

Ok. Let me first state the formula for entropy as information which Shannon derived. It is

$$H = -\sum_{i=1}^{n} p_i \log p_i$$

How do we get this? Now we can get it many ways but the easiest in my view is this. We define information from the probability of an event occurring p_i. If the probability is low and the event occurs, the information gain is high and vice versa, so we define raw info as $\frac{1}{p_i}$

But if an event occurs and another event occurs which is independent, then the raw info is $\frac{1}{p_i p_j}$

Now information gained should be additive, we should be able to add the first info and the second info to get this but $\frac{1}{p_i p_j} \neq \frac{1}{p_i} + \frac{1}{p_j}$

The only function to do this is the log of $\dfrac{1}{p_i}$

And we thus write the information as follows

$$\left.\begin{aligned} F\left(\frac{1}{p_1 p_2}\right) &= F\left(\frac{1}{p_1}\right) + F\left(\frac{1}{p_2}\right) \\ -\log(p_1 p_2) &= -\log(p_1) - \log(p_2) \end{aligned}\right\}$$

And if we take the average or expected value of all these probabilities in the set, we multiply the info by the probability of each and sum

To get

$$H = -\sum_{i=1}^{n} p_i \log p_i$$

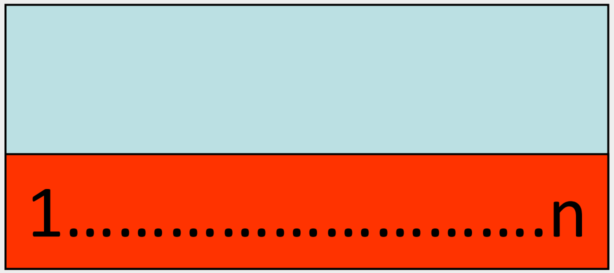
Now entropy or information is large – big – when all the probabilities are the same – uniform

$$p_i = \frac{1}{n}$$

And it is small – in fact 0 – when one probability is 1 and the rest are zero

$$p_i = 1, \quad and\ the\ rest\ are\ p_j = 0, \forall j \neq i$$

We can draw a graph of all these probabilities as follows – first when there are all equal



p = 1/n and Entropy H = max

And then when only one is equal to 1

p = 1, the rest 0, Entropy H = min

In the first case there is extreme homogeneity and in the second extreme order

Entropy-Maximising Reviewed Again

Now Alan Wilson, I think, introduced a method of maximising entropy which is equivalent to maximising uncertainty or information subject to a series of things we know about the distribution – like the fact that the probabilities must add to 1 and the average must be preserved – conserved and so on.

Essentially we choose a probability distribution so that we let there be as much uncertainty as possible subject to what information we know which is certain.

This is not the easiest point to grasp – why would we want to maximise this kind of uncertainty – well because if we didn't we would be assuming more than we knew – if we know there is some more info then we put it in as constraints. If we know p=1, we say so in the constraints. Let us review the formal process

The Mathematics

Let me repeat the Wilson stuff which is standard statistical mechanics

Maximise $$H = -\sum_{i=1}^{n} p_i \log p_i$$

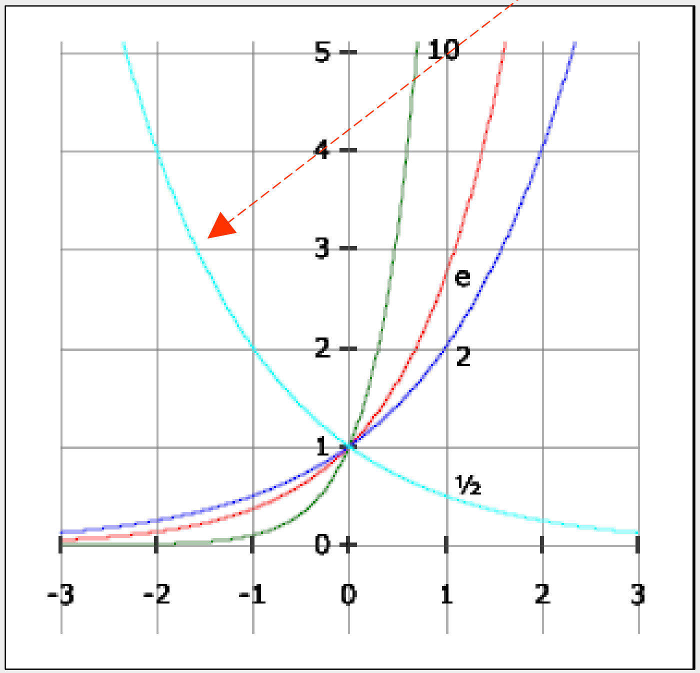
Subject to $$\sum_i p_i = 1$$ and $$\sum_i p_i c_i = \overline{C}$$

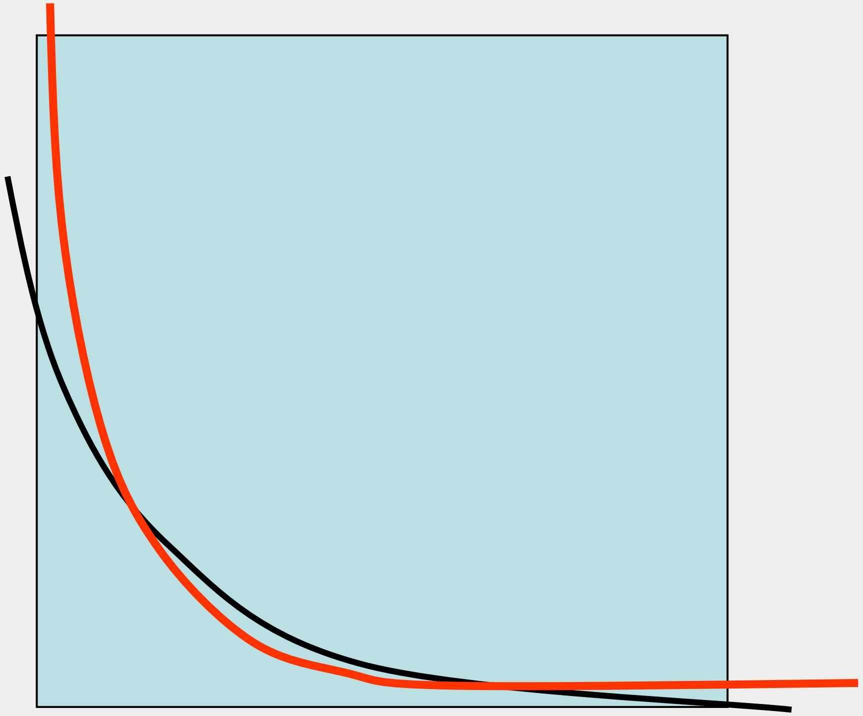
And we then get the classic negative exponential function which can be written as

$$p_i = \frac{\exp(-\lambda c_i)}{\sum_i \exp(-\lambda c_i)} \quad , \quad \sum_i p_i = 1$$

Now we don't know this is a negative function, it might be positive – it depends on how we set up the problem but in working out probabilities wrt to costs, it implies the higher the cost, the lower the probability of location

The Classic Negative Exponential – as Good as an Inverse Power Law? Preferable Even





Note the tails of the two distributions

A Phenomenological Demo – In Class

I don't think we can do this in class but let us try – assume that you all start with a location cost of 100 pounds, and then you have to choose someone at random and swap a small fraction say 1 pounds so that one of you wins and one loses 1 pounds. We choose randomly

I will try it with five people and see what we get

Ok – I will chose five people adjacent to each other

A, B, C, D, and E

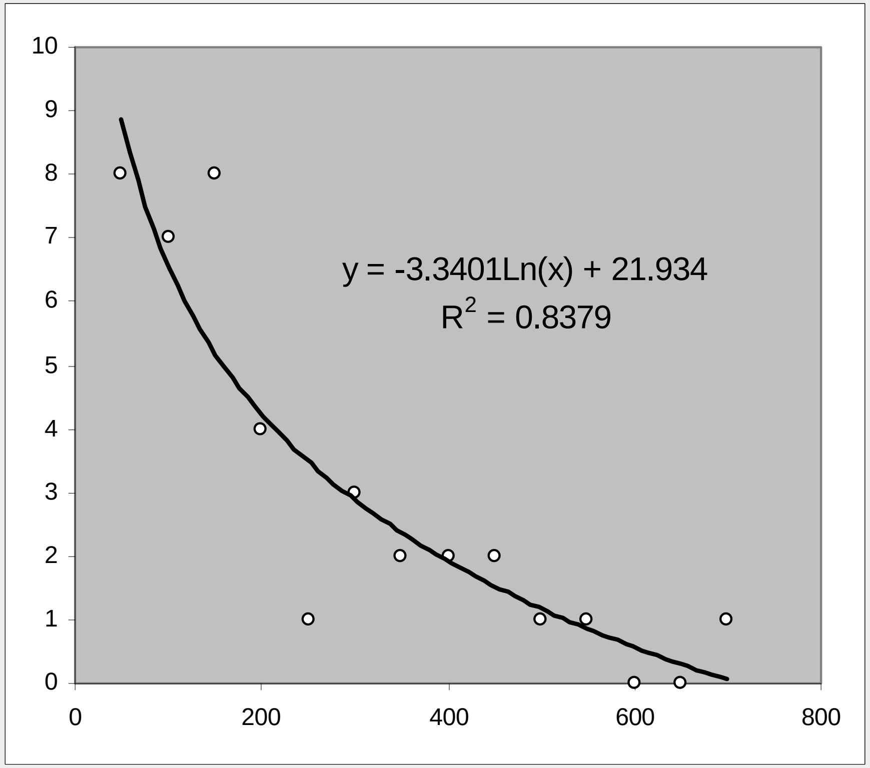
Now I will read the following choices from my random number generator, and then also tell you who to choose, randomly and what you will swap randomly

You then do it each time and keep a track of your totals. Oh let us do it 10 times - what do we get

The answer is – I can't do it in class – because I wrote a little computer program in VB to do it and I needed over a million runs and some 40 people to be able to get near to a negative exponential as I will show you

This is what I get – next page

```
Private Sub Command1_Click()
Dim People(100) As Single
Money = 100
SwapMoney = 1
n = 40
For i = 1 To n
People(i) = Money
'Print i, People(i)
Next i
t = 1
For i = t To 1000000
ii = Int((Rnd(1) * n) + 1)
jj = Int((Rnd(1) * n) + 1)
If ii = jj Then GoTo 777
If People(ii) = 0 Then People(ii) = 1: GoTo 777
If People(jj) = 0 Then People(jj) = 1: GoTo 777
d = Rnd(1)
If d > 0.5 Then
fid = SwapMoney
fjd = -fid
End If
People(ii) = People(ii) + fid
People(jj) = People(jj) + fjd
Total = 0
For iz = 1 To n
Total = Total + People(iz)
Next iz
'Print ii, jj, fid, fjd, People(ii), People(jj), Total
777 Next i
For i = 1 To n
Print i, People(i)
Next i
NewFile = "Money.txt"
Open NewFile For Output As #2
For i = 1 To n
Print #2, i, People(i)
Next i
Close #2
End Sub
```



I am not sure Excel has fitted a negative exponential

- Here is one where I have run it 10 million times with 1000 people
- In fact.. If I haven't been able to do it, this slide will be blank but note that what happens is the uniform distribution changes to something like a normal distribution and then to a negative exponential
- And it takes a little bit of experimentation to know how to run these hypothetical problems

Scaling: How Entropy Generates Power Laws

In essence I can modify the random model a little bit to show that if we let people accumulate more and more wealth – relax the conservation law then what we get is an inverse power law but the immediate way is to maximise entropy subject now to the following

We maximise entropy subject to a normalisation constraint on probabilities and now a logarithmic cost constraint of the form

Max $$H = -\sum_{i=1}^{n} p_i \log p_i$$

Subject to $$\sum_i p_i = 1 \qquad \sum_i p_i \log c_i = \bar{C}$$

Note the meaning of the log cost constraint. If we do all this we get

If we do all this we get the following model where we could simply put $\log c_i$ into the negative exponential getting

$$p_i = \frac{\exp(-\lambda \log c_i)}{\sum_i \exp(-\lambda \log c_i)} \quad \Rightarrow \quad p_i = \frac{c_i^{-\lambda}}{\sum_i c_i^{-\lambda}}$$

A power law.

But this is not the rank size relation as in the sort of scaling we looked at last week. Why not?

Cost and Size

Entropy-maximising location models tend to look at location probabilities as functions of cost and benefit of the locations.

Scaling models of city size or firm size or income size tend to look at probabilities of those sizes which have nothing to do with costs

Thus the problems are different

We must maximise entropy with respect to average city size not average locational cost and then we get the probabilities of locating in small cities much higher than in large cities as city size is like cost.

It is entirely possible of course for probabilities of locating in big cities to be higher than in small cities but as there are so many more small cities than big cities, small ones dominate.

So to look at the city size problem, we must substitute for cost with size as

$$p_i = \frac{\exp(-\lambda \log P_i)}{\sum_i \exp(-\lambda \log P_i)} \quad \Rightarrow \quad p_i = \frac{P_i^{-\lambda}}{\sum_i P_i^{-\lambda}}$$

And then we take the frequency as p_i and then the size as P_i, form the counter cumulative which is the rank and then twist the equation round to get the rank size rule – and hey presto we can connect up with slide 17 of last week

Last things

Basically I have actually worked out some of these equations for Greater London population data. These are in the working paper I sent the link around and in my chapter

Refs: http://www.casa.ucl.ac.uk/publications/workingPaperDetail.asp?ID=154

and Batty, M. (2009) Cities as Complex Systems: Scaling, Interactions, Networks, Dynamics and Urban Morphologies, In R. Meyers (Editor) **Encyclopaedia of Complexity and Systems Science**, Volume 1, pp 1041-1071, Springer, Berlin, DE.