# 140412 NTF Entropy In a Histogram

Note To File:

G. H. Boyle

14 April 2014

## References

1. Yakovenko, “Applications of statistical mechanics to economics: Entropic origin of the probability distributions of money, income, and energy consumption.”, 2012

arXiv:1204.6483v1 [q-fin.ST] 29 Apr 2012

1. ??

## Background

In discussion with Dr Yakovenko he suggested that my use of Stirling’s approximation in small histograms may cause distortions. I went back and researched that, essentially by re-reading his paper and checking a description of Stirling’s approximation on Wikipedia. He is, of course, right. So, I need to re-think the formulae I use.

This is the first of several notes. In this one, I just lay out the basic formula, as per the referenced article.

## Standard Definition of Entropy

This is the standard equation for the definition of entropy:

where Ω is the total number of possible microstates (or configurations of the histogram):

in a 4-bin histogram, or:

in a K-bin histogram.

## Translation to ABM histogram variable names

Using my variable set:

* N is replaced by A = number of agents.
* K = number of bins.
* i is an index into K, so 1 ≤ i ≤ K.
* And, the number of agents in bin i is ai.

## Definition of entropic measure

Let h(a1, a2, a3, a4) be a 4-bin histogram where ai is the number of agents in bin i. A is the sum of the ai. K is 4.

where

so

then

For example, if a configuration (a histogram) has values h(2, 3, 1, 2), then there are 2 agents in bin 1, 3 agents in bin 2, 1 agent in bin 3 and 2 agents in bin 4. All agents are distinguishable, apart from bin location. Ω is the number of permutations among agents that produce this configuration. In this example there are eight agents, so A = 8. Suppose you select agents randomly from among the 8, and place them into bins, left to right, as selected. There are 8 ways to select the first agent and put it into bin 1, there are 7 ways to choose the second and put it into bin 1, there are 6 ways to choose the third and place it into bin 2; etc. So there are 8! ways in which the agents can be placed, but, in each bin, the order by which the agents entered is not important. So permutations within each bin are divided out, giving the factorials on the bottom.

So, Ω is a characteristic of a configuration of a histogram. It is the number of permutations of agents that can produce this particular configuration. I think, then, in the language of thermodynamics, h(2, 3, 1, 2) is a macrostate, and Ω is the number of otherwise indistinguishable microstates that present themselves as this macrostate. Each of the microstates is equally probable.

## Definition of entropic index

To normalize this definition of entropy to exist in the interval [0, 1] divide by a maximal value.

S is at maximal value when all of the ai are equal to some constant α. But the sum of all ai is A.

or:

Then:

or:

Then:

This is an accurate formula. It has the problem that, when A! is too large, computers cannot evaluate it.

I have the additional problem that I want to verify my model (EiLab Model I) using MS Excel spreadsheets, and they only work for A <= 170. Now, this is not a serious practical limit, as Model I does not go above that limit usually, however, if I am to be consistent in all my implementations in ABMs, many will. So, to save rework later, I want to be consistent across all models, whether in MS Excel or in C++.