NOTE TO FILE:
Garvin H Boyle
R3:180210

A Comparison of The Formulae of Shannon and Boltzmann

Frontispiece

SHANNON

$$\left[{}_S S = -\sum_{i=1}^{k}(p_i \ln(p_i))\right]$$

VERSUS

BOLTZMANN

$$\left[{}_B S = k_B \ \ln(\Omega)\right]$$

A more sophisticated derivation, proposed by Boltzmann, uses the concept of entropy. Let us consider N particles with the total energy E. Let us divide the energy axis into small intervals (bins) of width $\Delta\varepsilon$ and count the number of particles N_k having the energies from ε_k to $\varepsilon_k + \Delta\varepsilon$. The ratio $N_k/N = P_k$ gives the probability for a particle to have the energy ε_k. Let us now calculate the multiplicity W, which is the number of permutations of the particles between different energy bins such that the occupation numbers of the bins do not change. This quantity is given by the combinatorial formula in terms of the factorials

$$W = \frac{N!}{N_1!\,N_2!\,N_3!\,\ldots}. \qquad (3)$$

The logarithm of multiplicity is called the entropy $S = \ln W$. In the limit of large numbers, the entropy per particle can be written in the following form using the Stirling approximation for the factorials

$$\frac{S}{N} = -\sum_k \frac{N_k}{N} \ln\left(\frac{N_k}{N}\right) = -\sum_k P_k \ln P_k. \qquad (4)$$

Table of Contents

1 - References

A. 2010 Yakovenko - Statistical Mechanics and Distribution of Money.pdf
B. 150527 PPR - Definition of EI R17.pdf
C. 180121 Boltzmann's entropy formula – Wikipedia.pdf
D. 180121 Shannon Entropy (information theory) – Wikipedia.pdf
E. 180121 Gibbs Entropy (statistical thermodynamics) – Wikipedia.pdf
F. 180119 Stirling's approximation – Wikipedia.pdf
G. 180122 XLS Shannon Vs Boltzmann R1.xlsm
H. 180118 NTF Custom Functions in Excel R3.pdf

2 - Background

At Ref A Dr Victor Yakovenko describes the formula for entropy that he uses to calculate entropy in his capital exchange models. He ascribes the basic formula to Ludwig Boltzmann, and uses the multinomial coefficient from combinatorics (Eq (3)) and then the formula that is very reminiscent of Shannon's formula (Eq (4)).

Figure 01 – Yakovenko's Formulae

A more sophisticated derivation, proposed by Boltzmann, uses the concept of entropy. Let us consider N particles with the total energy E. Let us divide the energy axis into small intervals (bins) of width $\Delta\varepsilon$ and count the number of particles N_k having the energies from ε_k to $\varepsilon_k + \Delta\varepsilon$. The ratio $N_k/N = P_k$ gives the probability for a particle to have the energy ε_k. Let us now calculate the multiplicity W, which is the number of permutations of the particles between different energy bins such that the occupation numbers of the bins do not change. This quantity is given by the combinatorial formula in terms of the factorials

$$W = \frac{N!}{N_1!\,N_2!\,N_3!\,\ldots}. \qquad (3)$$

The logarithm of multiplicity is called the entropy $S = \ln W$. In the limit of large numbers, the entropy per particle can be written in the following form using the Stirling approximation for the factorials

$$\frac{S}{N} = -\sum_k \frac{N_k}{N} \ln\left(\frac{N_k}{N}\right) = -\sum_k P_k \ln P_k. \qquad (4)$$

From the Ref A paper.

I refer to these two equations often, so I indicate them as (3) and (4) with round parentheses and my own equations are enclosed with square brackets like this: [3] or [4]. At Ref B I have analyzed the entropy levels associated with states of Model I of EiLab. I am now in the process of closely examining the tools used in that analysis. Refs C, D and E are Wikipedia articles discussing entropy as understood by three creative enunciators of the concept: Gibbs, Boltzmann, and Shannon. Ref F is a note from Wikipedia about Stirling's approximation for the function ln(N!) to be used when N is large. Stirling's approximation plays a key role in this note, and variations on it play a key role in the series of notes.

The concept of entropy was first formulated in the fields of chemistry (Gibbs) and thermodynamics (Boltzmann) where the numbers of particles under consideration were of the order of 10^{23} (Avogadro's Number). To handle such inconceivably large numbers, certain mathematical techniques had to be developed. However, when I work on agent-based models (ABMs) I am working with numbers in the low hundreds, or teens, or often, zeros. I need to understand whether and how the various standard formulae perform with numbers at this low end of the domain. A first step, it seems, is to establish the mathematical connection between the classic formulae of Boltzmann [${}_{B}S = k_B \ \ln(\Omega)$] and Shannon $\left[{}_{S}S = -\sum_{i=1}^{k}(p_i \ln(p_i))\right]$.

Ref G is a spreadsheet in which I tested the two formulae on the same set of 1,000 histograms.

3 - Purpose

The purpose of this note is quite simply to outline the mathematical connection between Boltzmann's classic equation for entropy, and that of Shannon. The goal is to better understand which formula is most useful for my studies of entropy when working with ABMs.

4 - Discussion

I will be examining entropy from two traditions or logical regimes. I will be distinguishing the entropy as calculated in each regime with a left-subscript like this:

- ${}_{B}S$ for the "Entropy in the Boltzmann regime"
- ${}_{S}S$ for the "Entropy in the Shannon regime"
- S will be used for entropy in general

4.1 - Entropy – Boltzmann's Equation

Historically, Boltzmann defined entropy with the following formula:

$${}_{B}S = k_B \times ln(\Omega) \qquad [1]$$

By this he converted Clausius' empirical definition into a "statistical mechanics" definition. This definition seems somewhat arbitrary to me. Jaynes implies it was determined by guess and check method. I cannot challenge it. I cannot see the reason for it. It is opaque to me. So I accept it as probably reasonable, and use it.

Deciding that this measure of entropy is "correct" is somewhat similar, I think, to deciding that an arithmetic or geometric average is a correct measure. It is arbitrary, not having any innate characteristic that makes it more natural or more correct than any other measure. But, with consensus that others will use the same formula and interpret it the same way, it becomes pragmatically useful, and common agreement of all users makes it "correct".

So, my approach to entropy in ABMs is somewhat arbitrarily based on Boltzmann's interpretation of Clausius' work. Replacing Ω with its combinatorial equivalent in terms of using A and a_i, and replacing k_B with a general dimensionless scaling factor C, as suggested by Yakovenko's paper, I get:

$${}_{B}S = {}_{B}C \times ln(\Omega) = {}_{B}C \times ln\left(\frac{A!}{\prod_{i=1}^{K}(a_i!)}\right) \qquad [2]$$

which resolves to:

$${}_{B}S = {}_{B}C \times [ln(A!) - \sum_{i=1}^{K} \ln(a_i!)] \qquad [3]$$

This is a definition of entropy consistent with Boltzmann's regime:

- As described in the Ref C Wikipedia article;

- Augmented by the appropriate insertion of the multinomial coefficient shown in equation (3) in the Ref A paper (as shown in Figure 01); and
- With a change of variables suitable for use in agent-based models (ABMs).

The symbols and their meanings are:

- $_B$S is entropy, and the B means in the regime associated with Boltzmann's work;
- $_B$C is a dimensionless scaling factor – I don't know if this will be needed eventually, but it is there for now as a place-holder;
- ln() is the monadic function that calculates the natural logarithm of a number;
- ! is the monadic function that calculates the factorial of a number;
- A is the number of agents in the model;
- a_i is the number of agents that have been sorted into bin i of a histogram; and
- Σ is the summation function.

4.2 - Entropy – Shannon's Equation

On the other hand, Shannon's Equation, as described in the Ref D article, is very similar in form to equation (4) of Yakovenko's Ref A document. (See equation (4) in Figure 01.) As given in that Ref D article Shannon's version of entropy has this formula:

The entropy can explicitly be written as

$$\mathrm{H}(X) = \sum_{i=1}^{n} \mathrm{P}(x_i)\,\mathrm{I}(x_i) = -\sum_{i=1}^{n} \mathrm{P}(x_i)\log_b \mathrm{P}(x_i), \qquad [4]$$

Using my own notation, this would be:

$$_SS = {}_SC \times -\sum_{i=1}^{k}\left(\frac{a_i}{A} \times \ln\left(\frac{a_i}{A}\right)\right) \qquad [5]$$

Where $_S$C is a dimensionless scaling factor – a place-holder.

4.3 - Stirling's approximation – Classic form

I understand that Boltzmann's version of the formula, as shown in equations [1] through [3], can be converted to Shannon's version of the formula, as shown in equation [4], through a substitution of Stirling's approximation for ln(A!) with a formula that has several elaborations, each with improved accuracy. In fact, Stirling's formula for ln(A!) as described in Ref F can be written as the sum of an infinite series of terms in which each term provides additional accuracy.

According to the Wikipedia article referenced above, Stirling's approximation to A! is an infinite series, but is it often used in a basic truncated form, as per equation [6]. The most basic (i.e. inaccurate) version of Stirling's approximation looks like this:

$$\ln(A!) = A\ln(A) - A + Error \qquad [6]$$

4.4 - Derivation of Shannon from Boltzmann

So, working with a basic form of Stirling's approximation as shown in equation [6] in which the error is ignored, here I reconstruct the derivation of Yakovenko's equation (4) from my equation

[3]. Start with the exact expression stated in equation [3], and ignoring the dimensionless scaling factor C for the nonce:

$$ {}_{B}S = {}_{B}C \times [ln(A!) - \sum_{i=1}^{K} \ln(a_i!)\,] \quad [7]$$

Making the substitution of [6] into [7] for both ln(A!) and ln(a_i!):

$$ {}_{B}S \approx {}_{B}C \times \left(\left[\left(A \bullet ln(A)\right) - A\right] - \sum_{i=1}^{K}[a_i \bullet \ln(a_i) - a_i]\right) \quad [8]$$

$$ {}_{B}S \approx {}_{B}C \times \left[\left(A \bullet ln(A)\right) - A - \sum_{i=1}^{K} a_i \bullet \ln(a_i) + \sum_{i=1}^{K} a_i\right] \quad [9]$$

The second term and the fourth term cancel out, leaving:

$$ {}_{B}S \approx {}_{B}C \times \left[\left(A \bullet ln(A)\right) - \sum_{i=1}^{K} a_i \bullet \ln(a_i)\right] \quad [10]$$

Let

$$ p_i = \left(\frac{a_i}{A}\right) \quad \text{and} \quad a_i = A \bullet p_i \quad [11]$$

So, making a substitution and changing the sign on the sum by inverting the argument of ln():

$$ {}_{B}S \approx {}_{B}C \times \left[\left(A \bullet ln(A)\right) + \sum_{i=1}^{K} A \bullet p_i \bullet \ln\left(\frac{1}{A \bullet p_i}\right)\right] \quad [12]$$

Add and subtract a term within the brackets:

$$ {}_{B}S \approx {}_{B}C \times \left[\left(A \bullet ln(A)\right) + \sum_{i=1}^{K} A \bullet p_i \bullet \ln\left(\frac{1}{A \bullet p_i}\right) + \sum_{i=1}^{K} A \bullet p_i \bullet \ln(A) - \sum_{i=1}^{K} A \bullet p_i \bullet \ln(A)\right] \quad [13]$$

Combine the two middle terms within the brackets:

$$ {}_{B}S \approx {}_{B}C \times \left[\left(A \bullet ln(A)\right) + \sum_{i=1}^{K} A \bullet p_i \bullet \ln\left(\frac{A}{A \bullet p_i}\right) - \sum_{i=1}^{K} A \bullet p_i \bullet \ln(A)\right] \quad [14]$$

Combine the first and last terms within the brackets:

$$ {}_{B}S \approx {}_{B}C \times \left[\left(A \bullet ln(A)\right) \bullet [1 - \sum_{i=1}^{K} p_i] + \sum_{i=1}^{K} A \bullet p_i \bullet \ln\left(\frac{A}{A \bullet p_i}\right)\right] \quad [15]$$

But, noting that

$$ \sum_{i=1}^{K} p_i = \sum_{i=1}^{K} \frac{a_i}{A} = \frac{1}{A}\sum_{i=1}^{K} a_i = 1 \quad [16]$$

So

$$ {}_{B}S \approx {}_{B}C \times \left[\sum_{i=1}^{K} A \bullet p_i \bullet \ln\left(\frac{1}{p_i}\right)\right] \quad [17]$$

or

$$ {}_{B}S \approx {}_{B}C \times [-A \sum_{i=1}^{K} p_i \bullet \ln(p_i)] \quad [18]$$

or

$$ {}_{B}S \approx \left[\frac{{}_{B}C}{{}_{S}C}\right] \times A \times {}_{S}S \quad [19]$$

Dividing by A we get:

$$\frac{{}_{B}S}{A} \approx \left[\frac{{}_{B}C}{{}_{S}C}\right] \times \left[-\sum_{i=1}^{K} p_i \bullet \ln(p_i)\right] \quad [20]$$

If the dimensionless scaling factors are equal, i.e. if ${}_{B}C = {}_{S}C$, then we get:

$$\frac{{}_{B}S}{A} \approx -\sum_{i=1}^{K} p_i \bullet \ln(p_i) \quad [21]$$

This replicates equation (4) of Yakovenko's Ref A paper. The symbol ≈ is meant to imply that this is an approximate answer, and its accuracy is dependent upon the accuracy of Stirling's approximation as used in the argument (at equation [6]). Equation [21] raised a question about proper units of measure for entropy in the two regimes. I will address that question in a separate diary note. But, the more important implication of equation [21] is this: ${}_{S}S$ must be multiplied by A to make it commensurable with ${}_{B}S$.

4.5 - Custom Functions – Surprisal(A, a_i) and LawnOfXFactorial(x)

At Ref G I made a spreadsheet in which I calculate the entropy using both the Boltzmann and Shannon regimes.

In order to evaluate the Shannon formula I decided to implement a custom function which I called Surprisal(A, a_i). A surprisal is one term of the sum – i.e. $[p_i \bullet \ln(p_i)]$ is a surprisal. This cannot be evaluated when pi = 0, but the contribution of such a surprisal to the total entropy is zero. So, using the techniques documented at Ref H, I implemented a custom function which checks for zeros as arguments and handles them appropriately.

```
'///////////////////////////////////////////////////////////////////////////////
'// Orrery Software, orrery@rogers.com, Garvin H Boyle.
'///////////////////////////////////////////////////////////////////////////////
'// This function computes a surprisal p*ln(1/p).
'// Its a component of Shannon's formula for entropy S=-A*SUM(pi*ln(pi)
Function Surprisal(A, b)
   If b = 0 Then
     Surprisal = 0
   Else
     Surprisal = (b / A) * Application.Ln(b / A)
   End If
End Function
```

Similarly, the formula using the multinomial coefficient requires the evaluation of ln(A!) for numbers greater than 170, and most computers have difficulty with this. I use a function called LawnOfXFactorial(x) to compute such numbers, when needed.

```
'/////////////////////////////////////////////////////////////////////////////////
'// Orrery Software, orrery@rogers.com, Garvin H Boyle.
'////////////////////////////////////////////////////////////////////////////////
'// This function computes ln(x!) using Stirling's approximation, but only when
'//   needed.
'// Stirling's approximation is: ln(x!) ~ ( x * ln(x) ) - x
'// MS Excel is not able to calculate x! if x > 170.
'// So, when it can, this function does the exact calculation (i.e. when x <= 170).
'// And when it cannot do that, it uses Stirling's approximation.
'/////////////////////////////////////////////////////////////////////////////////
Function LawnOfXFactorial(x)
   If x < 170 Then
     LawnOfXFactorial = Application.Ln(Application.Fact(x))
   Else
     LawnOfXFactorial = (x * Application.Ln(x) - x)
   End If
End Function
```

4.6 - Comparison of Shannon and Boltzmann

At Ref G I made a spreadsheet with 1,000 random histograms. For each bin in each histogram I generated a random number between 0 and 200. In each histogram I made five such bins (K=5), so the average bin size would be ~100, with some above 170. I therefore used custom functions to calculate both the ln(x!) numbers for the Boltzmann equation, and the surprisals for the Shannon equation.

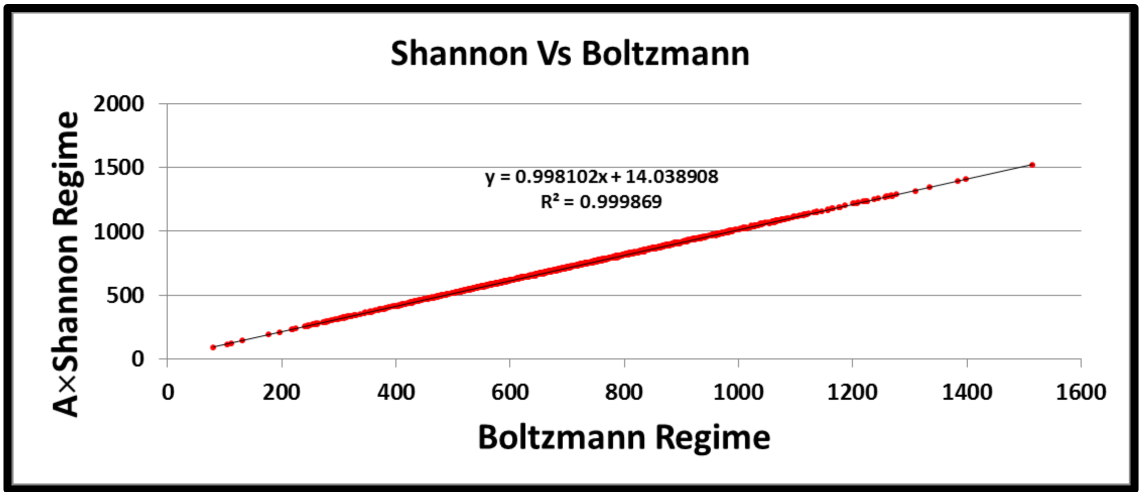
The results were interesting. Define the following spreadsheet labels:

- Let the results from the Boltzmann calculation $_BS$.
- Let the results from the Shannon calculation $_SS$.
- Let the error be $E = {}_BS – [A \times {}_SS]$.
- Let the error squared be $E^2 = E*E$.
- Let the absolute error be ABS(E).
- **Summary**: Let the RMS error be $SUM(E^2)^{1/2}$ for all E.
- **Summary**: Let the average error be AVE(E) for all E.
- **Summary**: Let the average absolute value of error be AVE(ABS(E)).

The summary results of one run plus the first few detailed rows of the table of 1,000 sets of results is:

	RMS Error	Ave Error	Ave ABS(E)	
	404.67	-12.49	12.49	
$_BS$	$A \times {}_SS$	E	ABS(E)	E^2
716.60	732.44	-15.84	15.84	250.8234
678.47	693.96	-15.48	15.48	239.7764
1192.20	1198.63	-6.43	6.43	41.36311
712.98	723.83	-10.86	10.86	117.8419

And, here is a scatter graph of $[A \times {}_SS]$ VS $_BS$:



The equation of the linear trend line was rather consistent across twelve trials of 1,000 histograms per trial. Here's the descriptive statistics for the 12 trials.

Slopes	Intercepts
0.997713	14.316782
0.998397	13.743965
0.997805	14.217808
0.998146	14.004098
0.997926	14.076571
0.997875	14.189667
0.997553	14.297488
0.997877	14.128388
0.997312	14.389961
0.997932	14.148565
0.998198	13.935794
0.997401	14.370837

	Slopes	*Intercepts*
Mean	0.997844583	14.15166
Standard Error	9.2174E-05	0.0550451
Median	0.997876	14.169116
Mode	#N/A	#N/A
Standard Deviation	0.0003193	0.1906817
Sample Variance	1.01953E-07	0.0363595
Kurtosis	-0.33758518	0.3833366
Skewness	-0.08077165	-0.7783157
Range	0.001085	0.645996
Minimum	0.997312	13.743965
Maximum	0.998397	14.389961
Sum	11.974135	169.81992
Count	12	12
Confidence Level(95%)	0.000202874	0.1211534

Based on this it seems that the connection between $_SS$ and $_BS$ is:

$$A \times {}_SS \approx [0.9978 \pm 0.0003]\ {}_BS + [14.15 \pm 0.19] \qquad [21]$$

Or, alternately:

$${}_BS \approx \frac{[A \times {}_BS] - [14.15 \pm 0.19]}{[0.9978 \pm 0.0003]} \qquad [22]$$

This introduces a bias of 14 that has a relatively large impact when $_SS$ or $_BS$ is small, and that is the part of the domain of S in which ABMs will normally be operational. That is a concern.

4.7 - Conclusions

I have learned several notable things from this exercise:

- There are several versions of "Stirling's Approximation" formula, and the connection between Boltzmann and Shannon uses one of the weaker (less accurate) versions. This implies that whatever the weakness is, it is transferred in some fashion to Shannon's formula.
- Shannon's formula is unable to cope with bins having a zero in them. The natural logarithm of zero is undefined, but the surprisal of an empty bin is zero. Special processing is needed to avoid the undefined value.
- Of the two formulae, Boltzmann's seems to be more directly applicable to ABMs in which small numbers of agents are involved, and the partitions include a lot of bins with small numbers of agents (e.g. tens rather than trillions).
- However, Boltzmann's formula also needs to depend on Stirling's formula when calculating larger values of ln(A!), so special knowledge of that formula is needed whichever of these regimes I decide to use.

Somewhat arbitrarily, I have decided that I will use the formula that more accurately reflects the "true" unbiased entropy when doing calculations for ABMs. I.e. I will use the formula that is directly derived from (a) Boltzmann's formula and (b) the multinomial coefficient, without the intercession of (c) Stirling's approximation. More specifically, I will use my equation (3). However, I will still need to use Stirling's approximation (some version of it) to actually evaluate the entropy within an ABM – specifically when large numbers are involved. I am making a distinction between the use of Stirling's approximation as a means to produce a derivative formula analytically, as I did in equations [6] through [21], and the use of it as a means to evaluate a specific measure of a system. The derivative formula is biased, whereas a specific measurement will introduce error that can be managed. I am thinking here specifically of the problem of using Stirling's approximation when A or a_i are small. This can be easily avoided when using my equation [3]. It cannot be so easily avoided when using equation [20] since the approximate nature of Stirling's truncated series is now innate to equation [20].

Furthermore, the interpretation of equation [20] in the context of an ABM is curious. It calculates the entropy per agent. In the context of information theory and bits which are either on (a 1) or off (a 0), I would presume that A is the number of 1s in the bit stream. This raises two questions for me, for which I do not currently have an answer:

- Does this imply, in any way, that entropy can be localized to a single agent?
- Does a better understanding of the usual units of measure of entropy in information theory provide insight into the meaning of equation [20]?

5 - Summary

At Ref A Yakovenko suggested that Ω as used in Boltzmann's equation (see my equation [1]) for entropy can be replaced by the combinatorial multinomial coefficient (see Yakovenko's equation (3)). He then presented another equation (see Yakovenko's equation (4)) looking very much like Shannon's equation for information entropy.

Stirling developed a sophisticated sum of an infinite series which evaluates to ln(A!) when A is large. A truncated version of this infinite series is often used for pragmatic estimates.

Using a severely truncated version of Stirling's formula I showed how Yakovenko's cited equation (4) derives from my equation [1].

I considered what insights this exercise has given me with respect to calculation and interpretation of entropy in an ABM.

6 - Yet-To-Do

To understand this better, I need to investigate:

- The impact of the error inherent in Stirling’s approximation;
- The units of measure for entropy in information theory and thermodynamics; and
- The implication for interpretation of economic entropy, as described in my Ref B paper.