Entropy for Data Science

Charlie Edelson, Caleb Dowdy, Chris Leonard, Nicole Navarro, Aaron Niskin, Lance Price

12/4/2017

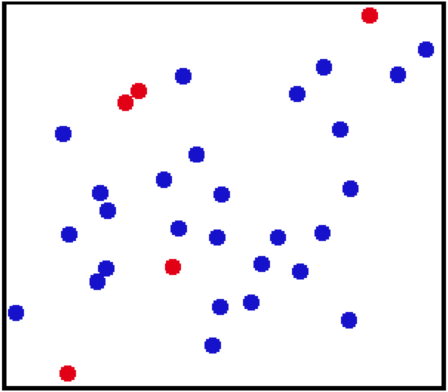
Outline

- History
  - Statistical Mechanics
- Shannon Entropy
  - Uniform Distribution
  - Normal Distribution
- Tsallis Entropy

History

Statistical Mechanics

Consider a box with N particles of a monatomic gas



How would you model this?

Statistical Mechanics - State Variables?

- We can talk about state variables: P, T, N, and V

Statistical Mechanics - State Variables?

- We can talk about state variables: P, T, N, and V
  - Ideal Gas Law

$$PV = nRT \tag{1}$$

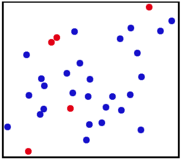
Statistical Mechanics - State Variables?

- We can talk about state variables: P, T, N, and V
  - Ideal Gas Law

$$PV = nRT \tag{1}$$

- Characterize System Behaviors

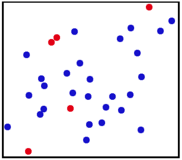
Statistical Mechanics - Ensemble Statistics



Assume each particle obeys Newton's Law

- v_0 and x_0 determines system
- Impractical for large N

Statistical Mechanics - Ensemble Statistics



Assume each particle obeys Newton's Law

- v_0 and x_0 determines system
- Impractical for large N

James Maxwell's Kinetic Theory of Gases

- Consider Ensamble Statistics

$$PV = \frac{Nm\bar{v}^2}{3} \tag{2}$$

Statistical Mechanics Entropy

Average Behaviour $\rightarrow$ Macroscopic Properties

Statistical Mechanics Entropy

Average Behaviour $\rightarrow$ Macroscopic Properties

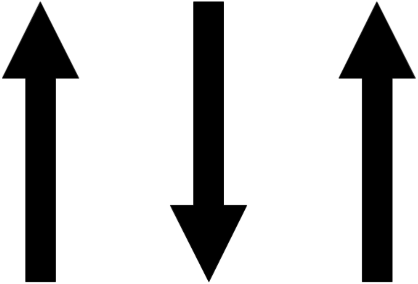
Ludwig Boltzman's statistical mechanical entropy (1877)

$$S = k_b \ln(\Omega) \tag{3}$$

Ω is the multiplicity of a given macrostate

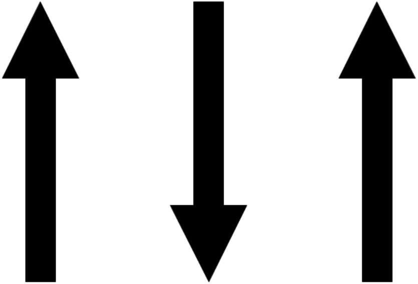
Macrostates, Microstates, and Multiplicity

Consider non-interacting paramagnet with 3 dipoles



Macrostates, Microstates, and Multiplicity

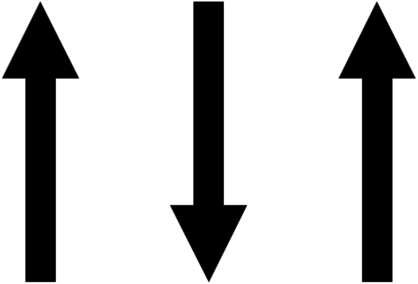
Consider non-interacting paramagnet with 3 dipoles



- Macrostate - 2 Up, 1 Down

Macrostates, Microstates, and Multiplicity

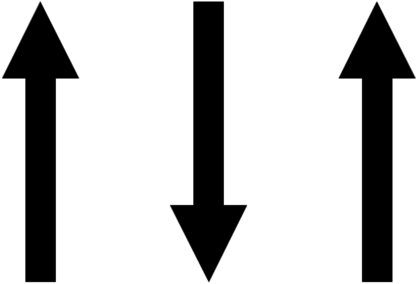
Consider non-interacting paramagnet with 3 dipoles



- Macrostate - 2 Up, 1 Down
- Microstate - $\uparrow, \downarrow, \uparrow$

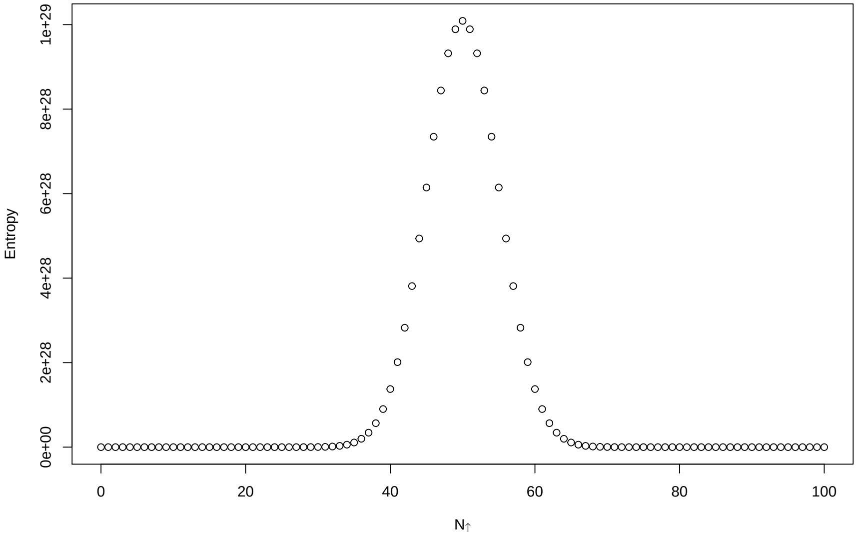
Macrostates, Microstates, and Multiplicity

Consider non-interacting paramagnet with 3 dipoles



- Macrostate - 2 Up, 1 Down
- Microstate - $\uparrow, \downarrow, \uparrow$
- $\Omega = \binom{N}{N_\uparrow} = \binom{3}{2}$

Entropy of 100 Dipole Paramagnet



Interpretation

Features of paramagnet entropy

- Minimum at 0 and 100 $\rightarrow$ 1 microstate each
- Maximum at 50 $\rightarrow 10^{29}$ microstates!

Interpretation

Features of paramagnet entropy

- Minimum at 0 and 100 $\rightarrow$ 1 microstate each
- Maximum at 50 $\rightarrow 10^{29}$ microstates!

Measure of "mixed-up-ness" of a physical system

- Higher entropy $\rightarrow$ more mixing (randomness)
- Lower entropy $\rightarrow$ less mixing

Shannon Entropy

Telephone Line Information Loss

Claude Shannon at Bell Telephone (1939)

- Quantify "lost information" in phone-line signals
- "Information Uncertainty"

$$H = -K \sum_{i=1}^{k} p(i) \log(p(i)) \tag{4}$$

Difficult time naming H...

Naming “Information Uncertainty”

. . . until he visited John von Neumann

> *My greatest concern was what to call it. I thought of calling it ‘information’, but the word was overly used, so I decided to call it ‘uncertainty’. When I discussed it with John von Neumann, he had a better idea. Von Neumann told me, ‘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, nobody knows what entropy really is, so in a debate you will always have the advantage.*

Shannon Entropy

Same properties as statistical mechanic entropy

- Low entropy $\rightarrow$ low randomness
- High entropy $\rightarrow$ high randomness

Shannon Entropy

Same properties as statistical mechanic entropy

- Low entropy $\rightarrow$ low randomness
- High entropy $\rightarrow$ high randomness

Directely related to information content of random variable X

- Unpredicatbility of state
- Average information content

Shannon Entropy

Same properties as statistical mechanic entropy

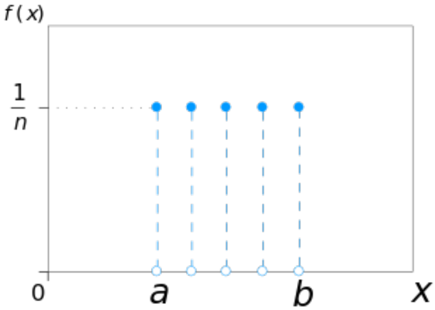
- Low entropy $\rightarrow$ low randomness
- High entropy $\rightarrow$ high randomness

Directely related to information content of random variable X

- Unpredicatbility of state
- Average information content

$$\mathrm{H}(X) = \mathrm{E}[-\ln(X)] \tag{5}$$

Shannon Entropy - Uniform Distribution



- Maximum Entropy
  - $\mathrm{H}(X) = \ln(N)$
- Boltzman Statisical Mechanics Entropy