Entropy for Data Science

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

12/4/2017

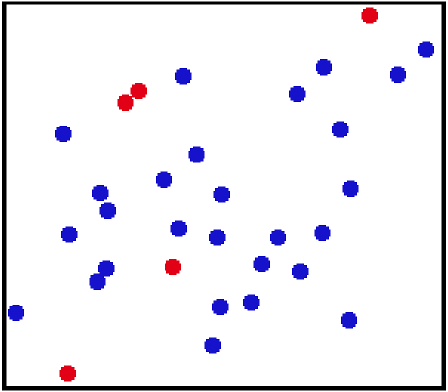
Outline

- History
  - Statistical Mechanics
- Shannon Entropy
  - Uniform Distribution
  - Normal Distribution
  - Tsallis Entropy
- Entropy for Data Science

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

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

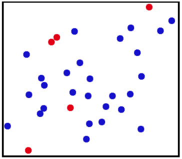
Statistical Mechanics - State Variables?

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

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

  - Characterize System Behaviors

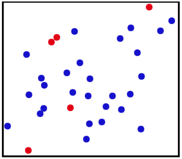
Statistical Mechanics - Ensemble Statistics



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

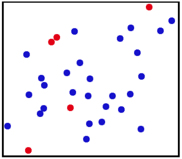
- v_0 and x_0 determines system
- Impractical for large N

James Maxwell's Kinetic Theory of Gases

- Consider Ensemble Statistics

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

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 Ensemble Statistics

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

Average Microscopic Behavior $\rightarrow$ Macroscopic Properties

Thermal Physics - Entropy

Rudolf Clausius' *equivalence-value* (1854)

> *If two transformations which, without necessitating any other permanent change, can mutually replace one another, be called equivalent, then the generations of the quantity of heat Q from work at the temperature T, has the equivalence-value:*

$$\frac{Q}{T}$$

Thermal Physics - Entropy

Rudolf Clausius' *equivalence-value* (1854)

> *If two transformations which, without necessitating any other permanent change, can mutually replace one another, be called equivalent, then the generations of the quantity of heat Q from work at the temperature T, has the equivalence-value:*

$$\frac{Q}{T} = S \tag{3}$$

Thermal Physics - Entropy

Rudolf Clausius' *equivalence-value* (1854)

> *If two transformations which, without necessitating any other permanent change, can mutually replace one another, be called equivalent, then the generations of the quantity of heat Q from work at the temperature T, has the equivalence-value:*

$$\frac{Q}{T} = S \tag{3}$$

Tranformation from T_1 to T_2 with Q is has an *equivalence-value*

$$Q\left(\frac{1}{T_2} - \frac{1}{T_1}\right)$$

Thermal Physics - Entropy

Rudolf Clausius' *equivalence-value* (1854)

> *If two transformations which, without necessitating any other permanent change, can mutually replace one another, be called equivalent, then the generations of the quantity of heat Q from work at the temperature T, has the equivalence-value:*

$$\frac{Q}{T} = S \tag{3}$$

Tranformation from T_1 to T_2 with Q is has an *equivalence-value*

$$Q\left(\frac{1}{T_2} - \frac{1}{T_1}\right) = \Delta S \tag{4}$$

Statistical Mechanics - Entropy

Ludwig Boltzmann's Entropy (1877)

$$S = k_b \ln(\Omega) \tag{5}$$

Ω is the multiplicity of a given macrostate

Statistical Mechanics - Entropy

Ludwig Boltzmann's Entropy (1877)

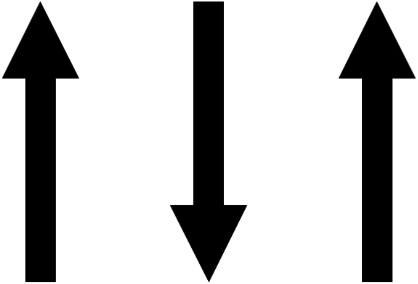
$$S = k_b \ln(\Omega) \tag{5}$$

Ω is the multiplicity of a given macrostate

- higher multiplicity $\rightarrow$ higher entropy

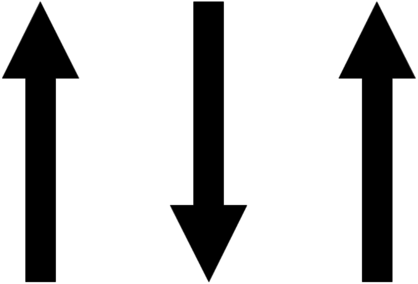
Macrostates, Microstates, and Multiplicity

Consider non-interacting paramagnet with 3 dipoles



Macrostates, Microstates, and Multiplicity

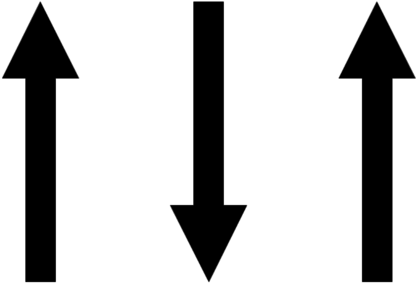
Consider non-interacting paramagnet with 3 dipoles



- Macrostate - 2 Up, 1 Down

Macrostates, Microstates, and Multiplicity

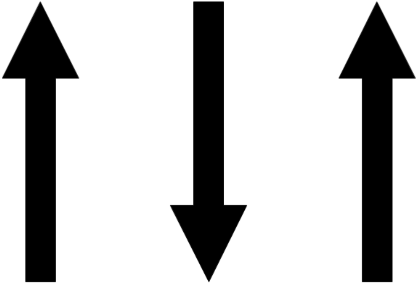
Consider non-interacting paramagnet with 3 dipoles



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

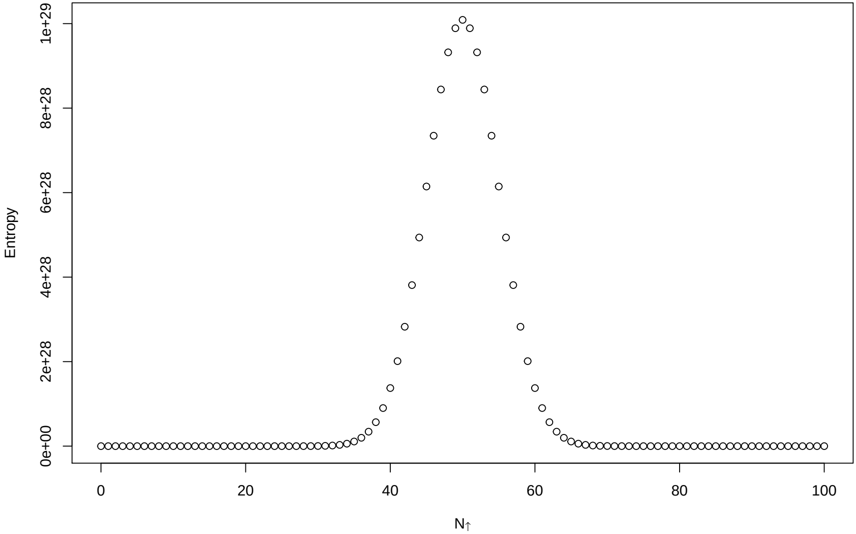
Macrostates, Microstates, and Multiplicity

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"

$$H = -K \sum_{i=1}^{k} p(i) \log(p(i)) \tag{6}$$

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

Directly related to information content of random variable X

- Unpredictability of state
- Average information content

Shannon Entropy

Same properties as statistical mechanic entropy

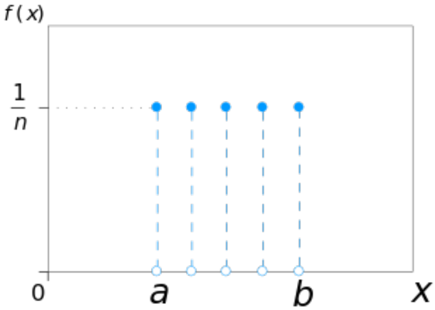
- Low entropy $\rightarrow$ low randomness
- High entropy $\rightarrow$ high randomness

Directly related to information content of random variable X

- Unpredictability of state
- Average information content

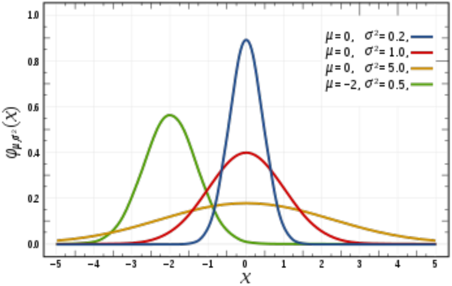
$$\mathrm{H}(X) = \mathrm{E}[-\ln(X)] \tag{7}$$

Shannon Entropy - Uniform Distribution



- Maximum Entropy
  - $\mathrm{H}(X) = \ln(N)$
- Boltzmann Statistical Mechanics Entropy
  - $\mathrm{S} = k_b \ln(\Omega)$

Shannon Entropy - Normal Distribution



- Maximum Entropy for fixed mean and variance
  - $\mathrm{H}(X) = \frac{1}{2}\log(2\pi e\sigma^2)$
- Part of the reason *why the normal distribution*

Tsallis Entropy

Attempt to generalize statistical mechanics

$$S_q(p_i) = \frac{k}{q-1}\left(1 - \sum_i p_i^q\right) \tag{8}$$

- q is entropic-index
  - $q \to 1$, recover Shannon Entropy
- Experimental Verification
  - Physical significant still debated

Entropy for Data Science

Anamoly Detection

- Entropy for detecting changes in “randomness”
  - Machine Learning and Metric Entropy

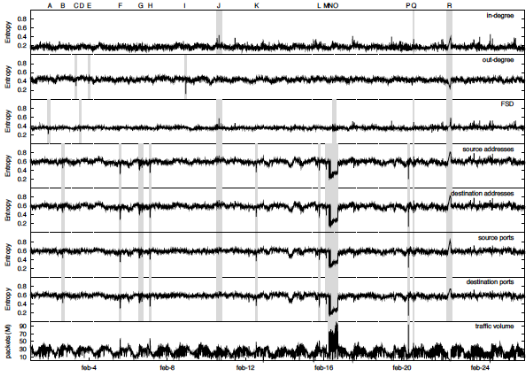


Figure 1: Nychis et. al. 2008

Anamoly Detection - Tsallis Entropy

- Similar work with Tsallis Entropy
  - q values for specific anamolies

Anamoly Detection - Tsallis Entropy

- Similar work with Tsallis Entropy
  - q values for specific anamolies

Currently in the works with Akamai!

Summary

- Entropy from Statistical Mechanics
  - Measure of “mixedness” or macrostate
- Adopted by Shannon $\rightarrow$ Infromation Loss
  - Average Information of State
  - Unpredictability
- Applicaitons in Anamoly Detection