Á la carte Entropy

Derek M. Jones
<derek@knosof.co.uk>

Background

Researchers' go to topic when they have no idea what else to talk about

- http://shape-of-code.coding-guidelines.com/2015/04/04/entropy-software-researchers-go-to-topic-when-they-have-no-idea-what-else-to-talk-about/

Reasons to ignore a SE paper

- "...major indicators of clueless nonsense..."
- http://shape-of-code.coding-guidelines.com/2016/06/10/finding-the-gold-nugget-papers-in-software-engineering-research/

Problems entropy is used to solve

Source of pretentious techno-babble

Aggregating a list of probabilities

- $D_1 = (0.1, 0.3, 0.5, 0.7, 0.9)/2.5$
- $D_2 = (0.2, 0.4, 0.6, 0.8)/2$

Which aggregation algorithm is best?

Geometric mean: $\left(\prod_i^n p_i\right)^{\frac{1}{n}}$

- $D_1 = 0.16$
- $D_2 = 0.22$

Shannon entropy: $\sum_i^n p_i \log \frac{1}{p_i}$

- $D_1 = 1.43$
- $D_2 = 1.28$

$$\log \frac{1}{\prod_i^n p_i^{p_i}}$$

Shannon: leading brand of entropy

Figure 1. Buying the brand leader

Other brands of entropy are available

Generalized entropy

- Rényi entropy: $\frac{1}{1-q}\log\left(\sum_i^n p_i^q\right)$

- Tsallis entropy: $\frac{1}{q-1}\left(1-\sum_i^n p_i^q\right)$

Bespoke entropy

- "Generalised information and entropy measures in physics" by Christian Beck
- Quadratic entropy

Probability weights

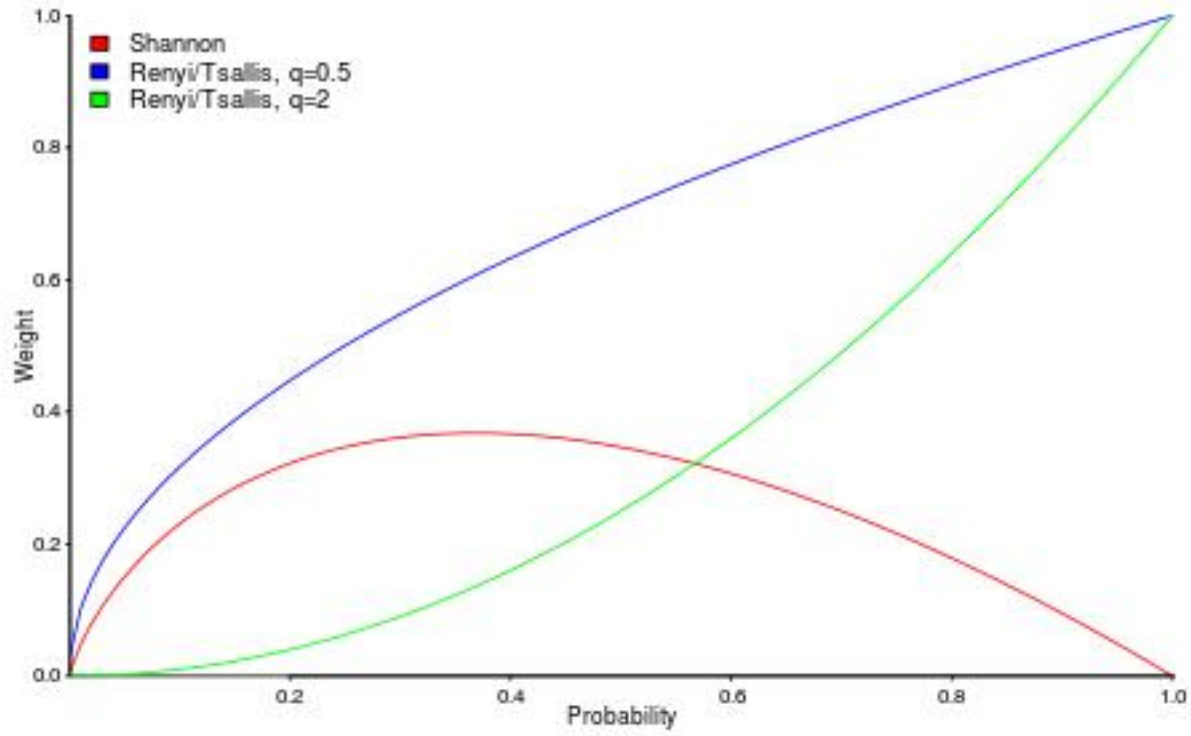

Figure 2. Weightings used by Shannon and Renyi/Tsallis

Shannon assumptions

Equilibrium state

Additive, i.e., $H(A,B) = H(A) + H(B)$

Other assumptions

Non-equilibrium state

Non-additive, i.e., $H(A+B) = H(A) + H(B) + (1-q)H(A)H(B)$

Not-Shannon processes

Long-range interactions

- memory usage
- "Initial Results of Testing Some Statistical Properties of Hard Disks Workload in Personal Computers in Terms of Non-Extensive Entropy and Long-Range Dependencies" by Dominik Strzalka

Preferential attachment

- not in equilibrium
- measurements showing a power law
- $1 < q \leq 2$

Password guessing

- $q = 2$ (collision entropy)

Rényi, Shannon or Tsallis?

Suck it and see

- "Using entropy measures for comparison of software traces" Miranskyy, Davison, Reesor, and Murtaza

Underlying characteristics of the problem

- data suggests a power law

Take-away

Entropy? Really nothing else to talk about?

Shannon mean-value may be non-optimal