Information Theory and Networks

Lecture 33: Information Theory, the Universe and Everything

Matthew Roughan
<matthew.roughan@adelaide.edu.au>
http://www.maths.adelaide.edu.au/matthew.roughan/Lecture_notes/InformationTheory/

School of Mathematical Sciences,
University of Adelaide

October 31, 2013

Part I

Information Theory, the Universe, and Everything

If you take a pack of cards as it comes from the maker and shuffle it for a few minutes, all traces of the original systematic order disappears. The order will never come back however long you shuffle. There is only one law of nature — the second law of thermodynamics — which recognises a distinction between the past and the future. Its subject is the random element in a crowd. A practical measure of the random element which can increase in the universe but never decrease is called entropy.

Arthur Eddington, The Nature of the Physical World, 1928

Section 1

More on the Second Law

The second law of thermodynamics

In a closed system, entropy cannot decrease.

- An operation is dissipative if it turns useful forms of energy into useless ones, such as heat energy
- Arrow of time implicit in this
  - most physical laws are reversible
    - they don't have a "natural" direction for time
    - you couldn't tell if a "video" of physical events at the microscopic level was running forward or backwards
  - yet most macroscopic processes are not
    - largely due to the 2nd law
  - 2nd law creates idea of causality?
    - one things in "caused" by another in linear time
    - our consciousness perceives it that way because we are also subject to the second law

Some problems

- Large-scale Universe
  - big-bang
    - where does local organization come from?
  - heat death might be OK, but big crunch reverses it
    - so long-term entropy works out the same?
- Black holes have no hair
  - if black holes evaporate (Hawking's radiation)
  - what happens to information that drops into a black hole?
  - so black holes have entropy

$$S_{BH} = \frac{k_B A}{4\ell_P^2}$$

  - where
    - A is the area of the event horizon
    - ℓ_P is the Plank length

The Second Law and Markov Chains

- Model an isolated system as a Markov chain
  - transitions according to physical laws governing the system
  - future of system independent of past (except through current state)
- 2nd law (in naive form) doesn't work
  - entropy can decrease
  - e.g., consider a case where the
    - initial distribution is uniform (max entropy)
    - stationary distribution is non-uniform
  - we could just chalk this up to Markov chains not really being covered by thermodynamics, but
- Four different interpretations of 2nd law [CT91, pp.34-36]
  1. relative entropy decreases with n
  2. relative entropy decreases RE stationary distribution
  3. entropy increases if the stationary distribution is uniform
  4. conditional entropy increases with n

The Second Law and Markov Chains: 1

- For the Markov chain discussed above
- Consider relative entropy $D(\mu_n \| \mu'_n)$ where
  1. μ_n and μ'_n are two probability distributions at time n
  2. μ_{n+1} and μ'_{n+1} are corresponding distributions at time $n+1$
- Then

$$D(\mu_n \| \mu'_n) \geq D(\mu_{n+1} \| \mu'_{n+1})$$

  1. relative entropy decreases with time
- Think of $D(\cdot)$ as a distance
  1. the two probability distributions get closer together as the system evolves
  2. remember $D(\cdot)$ has a lower-bound of zero, so a limit must exist

The Second Law and Markov Chains: 2

- For the Markov chain discussed above
- Consider $D(\mu_n \| \mu)$
  1. μ_n is a probability distributions at time n
  2. μ is the stationary distribution
- Then

$$D(\mu_n \| \mu) \geq D(\mu_{n+1} \| \mu)$$

  1. relative entropy decreases with time
- Think of $D(\cdot)$ as a distance
  1. the probability distribution gets closer and close to the stationary distribution
  2. remember $D(\cdot)$ has a lower-bound of zero, so a limit must exist
  3. if the stationary distribution is unique, the limit is 0

The Second Law and Markov Chains: 3

- For the Markov chain discussed above
  1. consider the case where the stationary distribution is uniform
- We can write the relative entropy as

$$D(\mu_n \| \mu) = \log |\Omega| - H(\mu_n)$$

  1. We know $D(\mu_n \| \mu)$ can't increase
  2. so $H(\mu_n)$ can't decrease
- So for this case $H(\mu_n) \leq H(\mu_{n+1})$
  1. this makes sense for physical systems
  2. in equilibrium, microstates are equally likely (uniform stationary distribution)
  3. so this kind-of handles the transition to equilibrium
- A nice example is a shuffle
  - a crude idea of a shuffle is a random permutation, with ultimately uniform distribution of all cards

The Second Law and Markov Chains: 4

- For the Markov chain discussed above
  1. make the additional assumption that it is stationary
  2. consider the conditional entropy $H(X_n|X_1)$

$$H(X_n|X_1) \leq H(X_{n+1}|X_1)$$

- The conditional uncertainty about the future increases
  - we know less and less about the state, the further we try to see into the future

Section 2

Landauer's Principle

Landauer's Principle

- Any calculation must involve some exchange of energy
  - so there is a lower bound on per bit calculation
  - any logically irreversible manipulation (e.g., erasure of a bit) is accompanied by an increase in entropy
- Landauer limit
  - minimum possible energy required to change one bit

    $$= k_B T \ln 2$$

    where k_B =Boltzmann's constant and T is temperature (in K)
  - modern computers use millions of times this energy
- Practical lower bound given by $T = 3K$ cosmic background radiation

Explanation

- Imagine a hypothetical efficient computer
  - never wastes energy
  - it's isolated (no energy comes in or out)
  - any
    - logical state (binary bits in computer) is a macrostate
    - represented by some number of microstates (physical states of electrons, magnetic particles, etc.)
  - we can imagine either
    - keep track of logical state
    - of not
- Irreversible computation
  - two or more logical states map to a single state
  - not invertible

Explanation

(1) Don't keep track of logical state

- Irreversible calculation implies that the number of possible logical states of the computer decreases
  - in erasing a bit, we have reduced no. of states by factor of 2
  - All else equal (equal probabilities)

$$H(X) = \log_2 |\Omega|$$

    - so if we reduce state space by factor of two
    - $H(X)$ is reduced by 1 bit
- But entropy can't decrease in isolated system
  - there must be some other increase
  - number of physical microstates corresponding to the macrostate (or the logical bits), must have increased to compensate
  - energy is dissipated into heat

Explanation

(2) Keep track of logical state

- Irreversible calculation doesn't change the number of possible states (just the actual state)
- But, from previous argument the number of microstates increased
- So from the point of view of the computer's user
  - entropy just increased by 1 bit

- Importance
  - logical (information) operations have real physical consequence
- All this is a bit on the speculative side
  - some claims it is just wrong [Ben03, She01]
  - but it does fix some problems, e.g.,
    - Maxwell's demon
- 2012 there is a claim that the release of heat has been measured http://spectrum.ieee.org/computing/hardware/landauer-limit-demonstrated
- Maybe we need "reversible" computing
- Some other related issues
  - Bekenstein bound on entropy/information that can be contained within finite region of space
  - Black hole information paradox

Further reading I

- Charles H. Bennett, *Notes on landauer's principle, reversible computation and maxwell's demon*, arXiv:physics/0210005v2, 2003, `http://arxiv.org/abs/physics/0210005`.
- Thomas M. Cover and Joy A. Thomas, *Elements of information theory*, John Wiley and Sons, 1991.
- Neri Merhav, *Information theory and statistical physics – lecture notes*, arXiv: 1006.1565v1, June 2010.
- Orly R. Shenker, *Logic and entropy*, `http://philsci-archive.pitt.edu/115/`, 2001.