Information theory and PRNG

Requirements (Shannon, 1948)

1) I(p) ≥ 0 (information is non-negative, p≥1)

2) I(1) = 0 (events that always occur carry no information)

3) I(p1 p2) = I(p1) + I(p2) (information due to independent events is additive)

Also, continuity, symmetry, and maximum when all possible events are equiprobable.

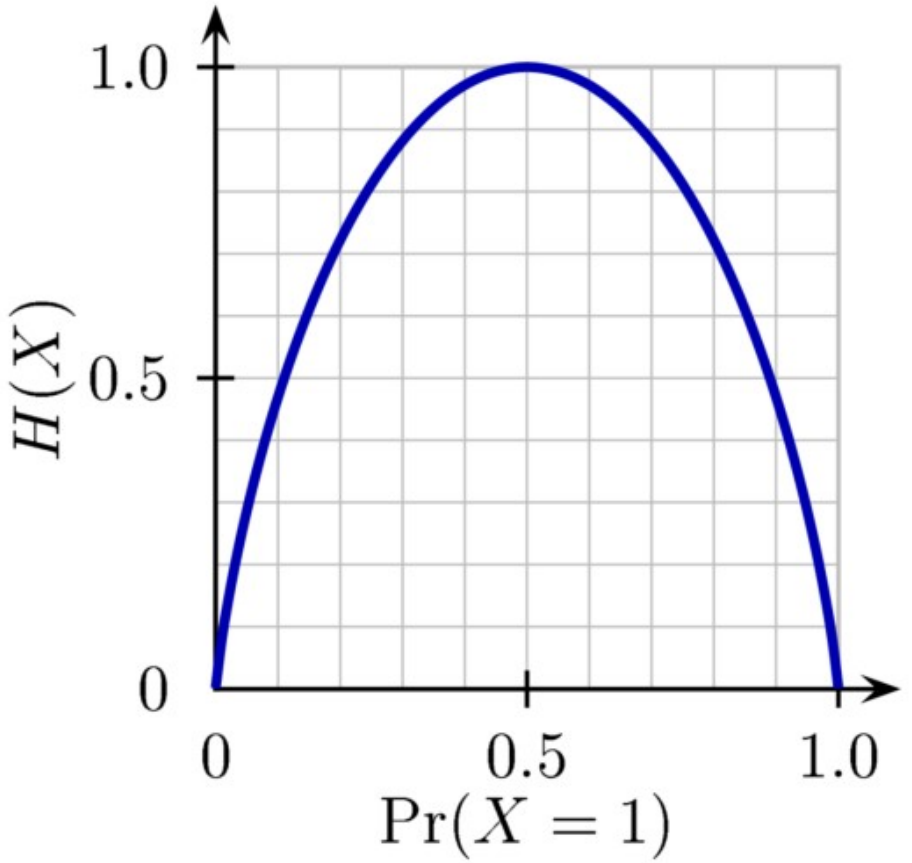
$$I(p) = \log(1/p)$$

1) I(1/2) = 0.30102999566...
2) I(1) = 0
3) I(1/2) + I(1/3) = log(2) + log(3) = 0.77815125038...
Joint probability: I(1/6) = 0.77815125038...

Continuity: I(1/2.01) = 0.30319605742...
Symmetry: log (3) + log(2) = 0.77815125038...
Maximum: log 3 + log 3 + log 3 = 1.43136376416...
log 2 + log 4 + log 4 = 1.20411998266...

Information = Entropy = Surpise

$$H[p] = -\sum_{i=1}^{k} p_i \log p_i$$



Pop quiz #1

- 4 times out of 10 the projector works
- 2 times out of 10 it's just blue
- 1 time out of 10 it won't turn on
- 1 time out of 10 it catches on fire
- 1 time out of 10 it flickers uncontrollably
- 1 time out of 10 sparks fly and the magic smoke escapes

What is the entropy (in bits) of one instance of trying to get the projector to work?

Answer

Input:

$$-0.4 \log_2(0.4) - 0.2 \log_2(0.2) - 4\,(0.1 \log_2(0.1))$$

Result:

$$2.32193\ldots$$

Pop quiz #2

- Abigail has three ways to make daddy say ouch (gouge his eye, dig her claws into his throat, or pull his hair). What is the maximum amount of entropy possible in each iteration of this random process?

Answer

Input:

$$-\frac{1}{3}\log_2\left(\frac{1}{3}\right)-\frac{1}{3}\log_2\left(\frac{1}{3}\right)-\frac{1}{3}\log_2\left(\frac{1}{3}\right)$$

Exact result:

$$\frac{\log(3)}{\log(2)}$$

$\log(x)$ is the natural logarithm

Decimal approximation:

More digits

1.584962500721156181453738943947816508759814407692481060455...

Entropy

- A measure of the information of a **random process**
- Pop quiz #3: Based on the above definition, order the following binary sequences from most entropy to least entropy?:

  A) 11111111000000000000000011111111

  B) 10111001110011001100010000110100

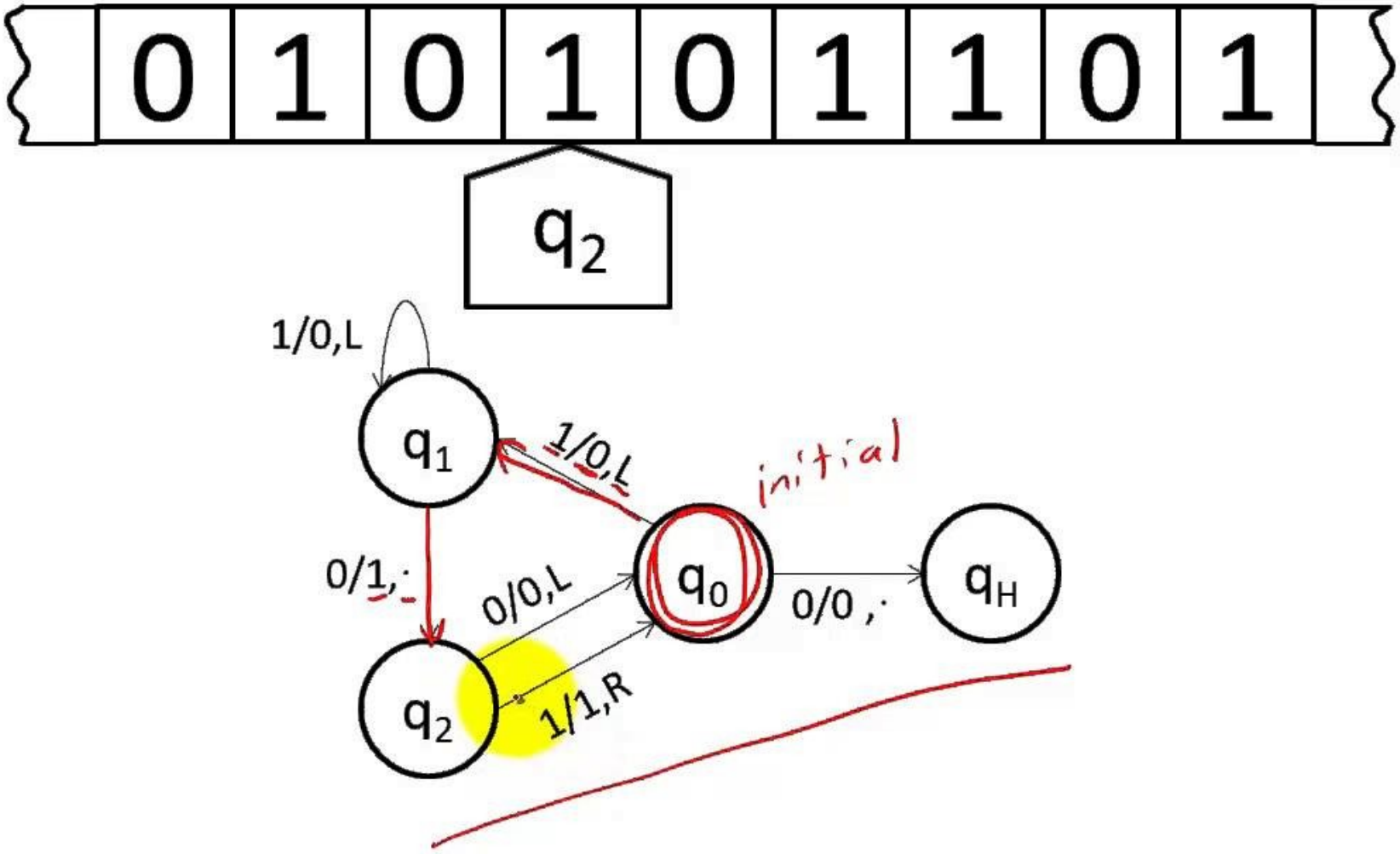
  C) 00000000000000000000000000000000

  D) 00010010000001000000010000000001

I pity the fool who uses the word “entropy” to describe a bit sequence or string without realizing that they are implicitly talking about algorithmic entropy (*a.k.a.*, Kolmogorov complexity) rather than the standard definition of entropy that Claude Shannon used to describe random processes!

Turing machines

Source: https://www.youtube.com/watch?v=gJQTFhkhwPA



Halting problem

- From a description of an arbitrary computer program and an input, determine whether the program will finish running or continue to run forever.

  P: if (halts(P)) then: while true {}; else halt;

Gödel's Incompleteness Theorem

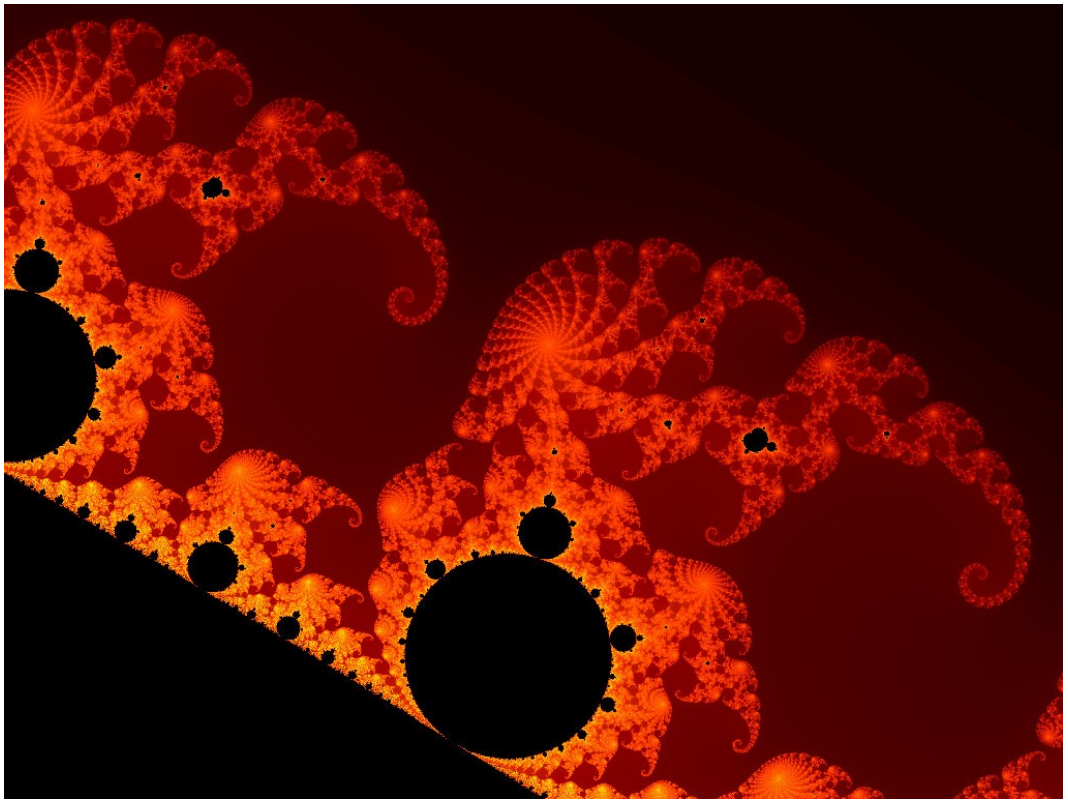
- In axiomatic systems capable of arithmetic, can't separate mathematics and meta-mathematics, can always form sentences of the form:

  This sentence is false.

- As Einstein put it:
  - “As far as the laws of mathematics refer to reality, they are not certain, as far as they are certain, they do not refer to reality.”
  - “We can't solve problems by using the same kind of thinking we used when we created them.”

Kolmogorov complexity

- Defined as the length of the shortest computer program that produces the object as output.



Chaitin's incompleteness theorem

- “The fact that a specific string is complex cannot be formally proven, if the complexity of the string is above a certain threshold.” [Wikipedia]
- Berry's paradox:
  - “the smallest positive integer not definable in fewer than twelve words”

PRNG

- Example:

  $X_{i+1} = X_i * a + b \bmod m$

- Seed
- Period
- Should be cryptographically strong
  - Assume attacker knows the algorithm and lots of past bits
  - *E.g.*, take the above and encrypt it with AES and a well chosen key (unknown to the attacker).
  - For more detailed examples, look up Yarrow or Fortuna.

PRNG problems

- Seeding problems
  - E.g., when you first boot an embedded device
  - QQ browser
- Periodicity problems
  - Never use srand(time) and rand() for crypto
- Backdoors
  - Dual_EC_DRBG
- Predictability
  - Witty worm

Sources

- https://en.wikipedia.org/wiki/Entropy_%28information_theory%29
- http://www.amazon.com/Elements-Information-Theory-Telecommunications-Processing/dp/0471241954/ref=sr_1_8?ie=UTF8&qid=1457368787&sr=8-8&keywords=information+theory
- http://www.amazon.com/Mathematical-Theory-Communication-Claude-Shannon/dp/0252725484/ref=sr_1_1?s=books&ie=UTF8&qid=1457368808&sr=1-1&keywords=information+theory
- https://en.wikipedia.org/wiki/Turing_machine
- https://www.youtube.com/watch?v=gJQTFhkhwPA
- https://en.wikipedia.org/wiki/Halting_problem
- http://www.icir.org/vern/papers/witty-imc05.pdf