Information Theory

Refresher on Logarithms

- Different expressions of log
  - Ln: Natural log = $\log_e a$
  - Lg: Binary log = $\log_2 a$
  - $\text{Log}_b a$ – no ambiguity here, log base b of a
  - Strictly evaluates to find what power *b* needs to be raised to exactly reach *a*
  - Examples

    $\text{Log}_{10} 100 = ?$

    $\text{Log}_2 64 = ?$

    $\text{Log}_2 1024 = ?$

Refresher on Logarithms

Log form

$log_b a = x$

Exponential form

$b^x = a$

- Examples

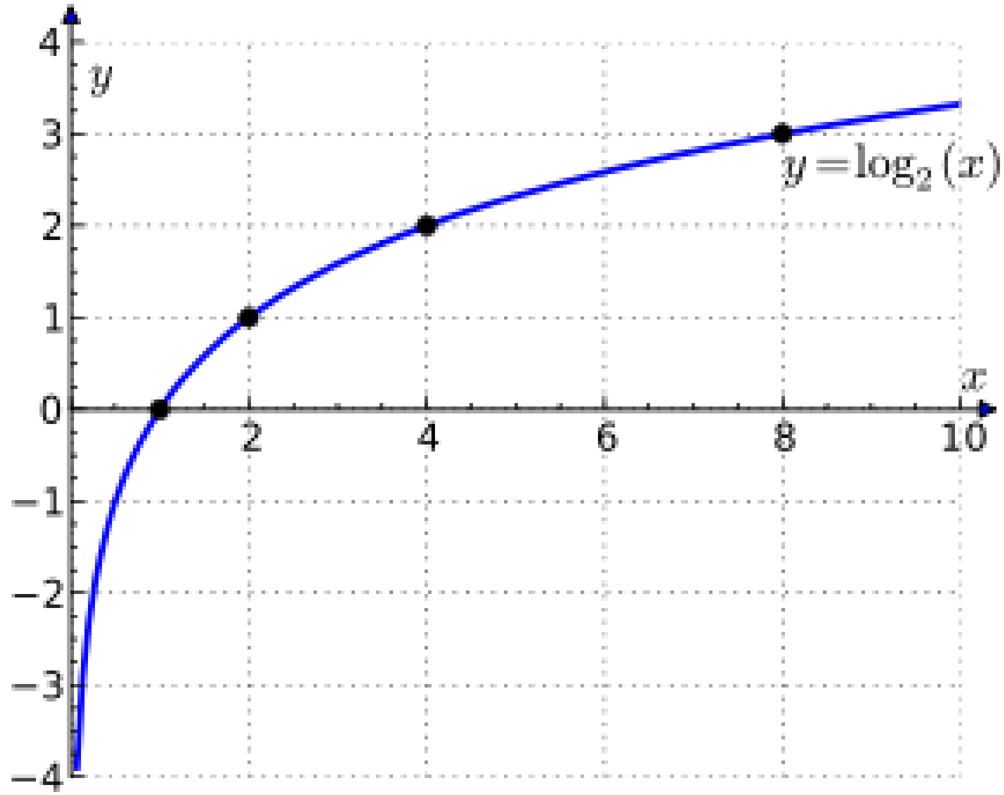
  $\log_2 1 = ?$

  $\log_2 1/2 = ?$

  $\log_2 1/4 = ?$

  $\log_2 1/8 = ?$

- *This will be useful for probabilities*



Refresher on Logarithms

- Important properties

Summation:

$\log_a m + \log_a n = \log_a m*n$

Powers:

$n \log_a m \;\; = \log_a m^n$

Conversion of bases:

$\log_a n * \log_b a = \log_b n$ ***or*** $\log_a n = \log_b n / \log_b a$

Refresher on Logarithms

- $b^x = a$, $log_b a = x$ when (a, b, x ∈ ℤ) can be seen in a tree

Always 1 root

b = branching factor

x = depth

a = nodes at this depth

$\log_2 1 = 0$

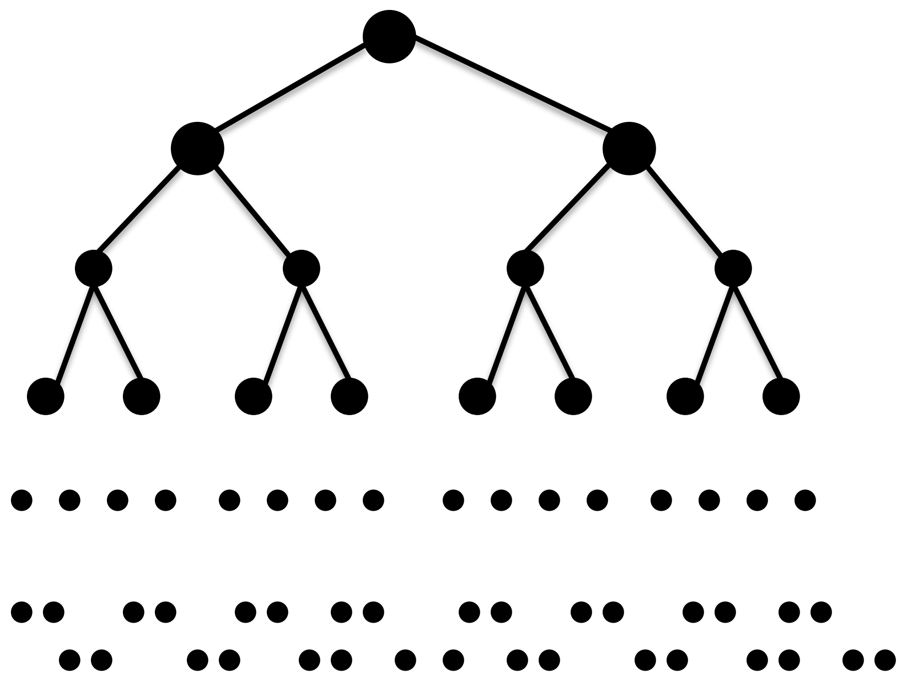
$\log_2 2 = 1$

$\log_2 4 = 2$

$\log_2 8 = 3$

$\log_2 16 = 4$

$\log_2 32 = 5$

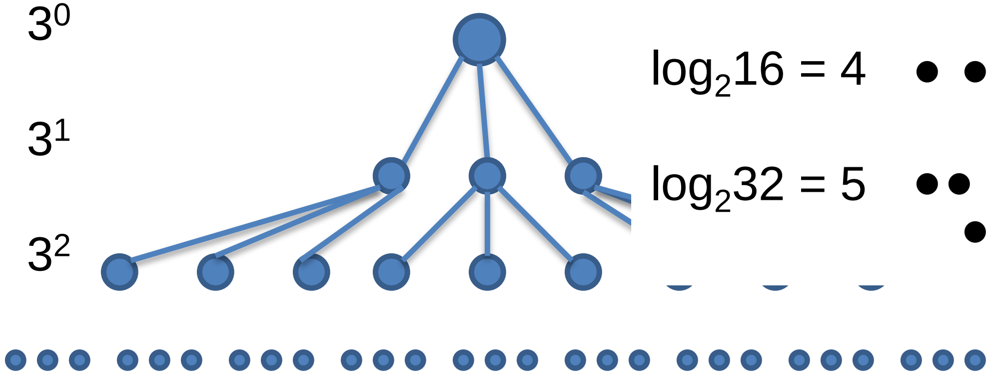


3^0

3^1

3^2

3^3



Refresher on Probability

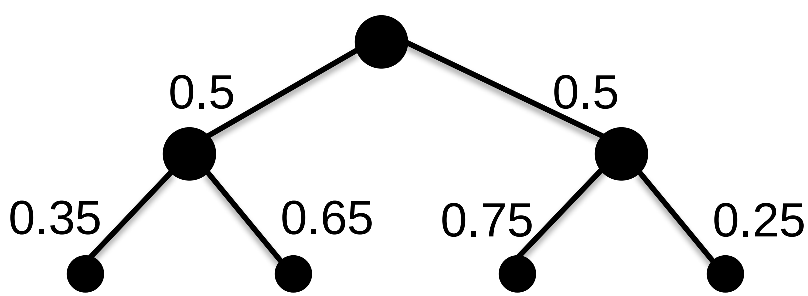
- Binary decision tree can express the number of yes or no questions needed to be answered in order to reach a conclusion.

- Probabilities can be expressed on the branches to express the likelihood of a decision being made

$\log_2 1 = 0$

$\log_2 2 = 1$

$\log_2 4 = 2$



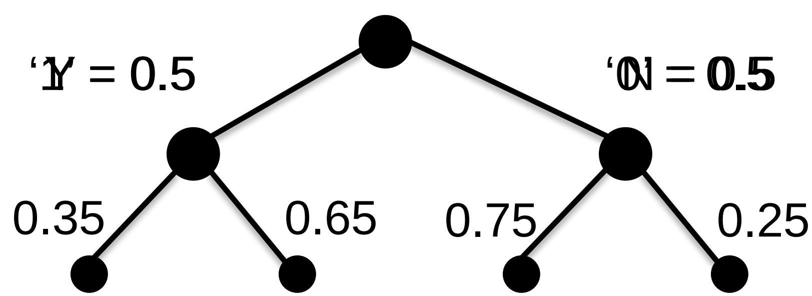
Refresher on Probability

- Let X be a random variable (RV) in a finite set of values $x_1 \ldots x_n$
- $P(X = x_i) = p_i$, $0 \leq p_i \leq 1$ is the probability that the R.V. $X = x_i$
- $P(X = \text{Yes}) + P(X = \text{No}) = 1$ for any decision if X is from {Yes, No} in the node of a tree (can express Yes as ‘1’ and No as ‘0’)

$\log_2 1 = 0$

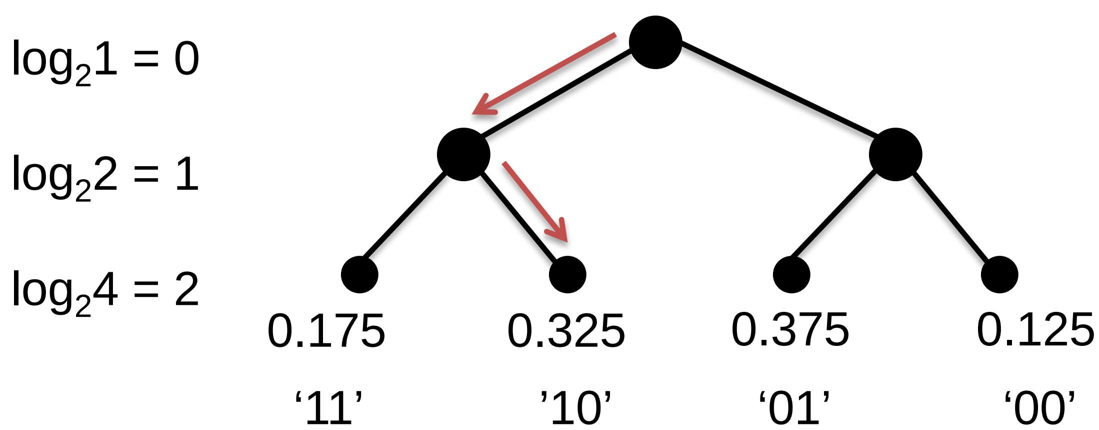
$\log_2 2 = 1$

$\log_2 4 = 2$



Refresher on Probability

- If RV X is from the **set of paths** through the tree, **π**
- $P(X = \pi_i) = 0 \leq p_i \leq 1$
- For all paths π_i $\sum_{i=1\ldots n} p_i = 1$
- What is the probability of path π_1 = '10'?
- Where did this value come from?



$\sum_i p_i = 1$

Claude Shannon

- Giant in computing
  - Devised boolean circuits (Master’s thesis)
  - Crypto (during WWII)
  - Channel communication (Bell labs)
- Foundational Works
  - 1949 Mathematical Theory of Communication
  - 1949 Communication Theory of Secrecy Systems
  - The founder of information theory
- Fundamental ideas for our course from Info Theory
  - Entropy: quantifying the uncertainty (security) in a message
  - Confusion and Diffusion: Properties of good Crypto Systems

Information Theory

- Mathematically answers fundamental questions about information
  - How is information ***quantified***?
  - How is information ***transmitted and understood***?
- Message space - # of all possible symbol sequences (*all paths through tree*)
- Early pioneers: Nyquist, Hartley, Shannon

Information Theory

- Transmission of binary digits, bits (0,1)
- Information in this medium is

  $H = n \log_2 s$ or

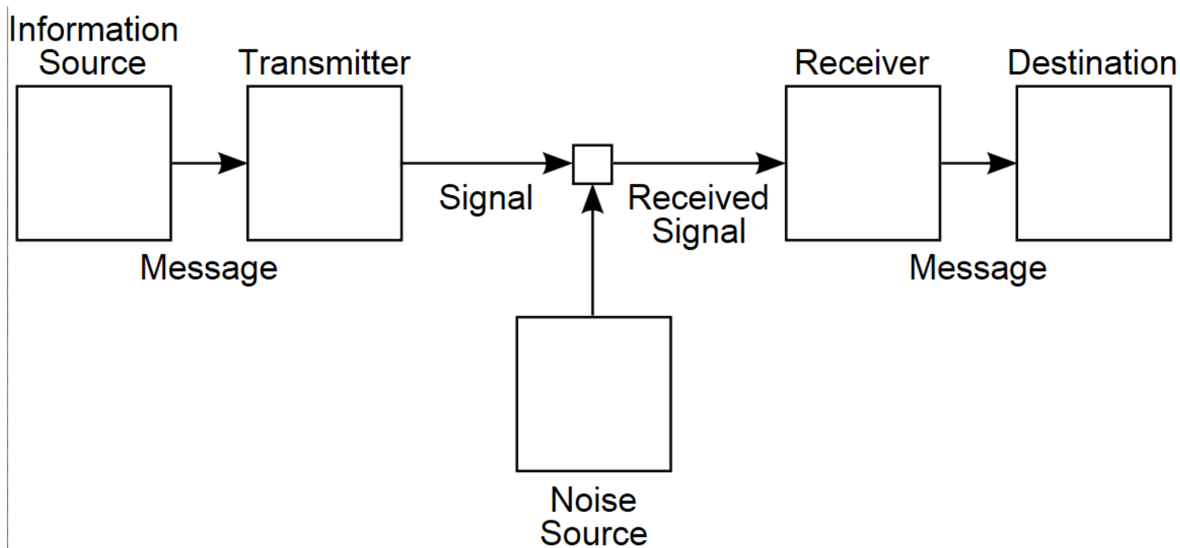
  $H = \log_2 s^n$ *by power property of logarithms*

  where H is information, n is # of symbols and

  s is the number of possible symbol sequences
- Basis for mathematical definition and theory, bits and entropy

Shannon's Communication Theory

- "The fundamental problem of communication is that of reproducing at one point, either exactly or approximately, a message selected at another point.“



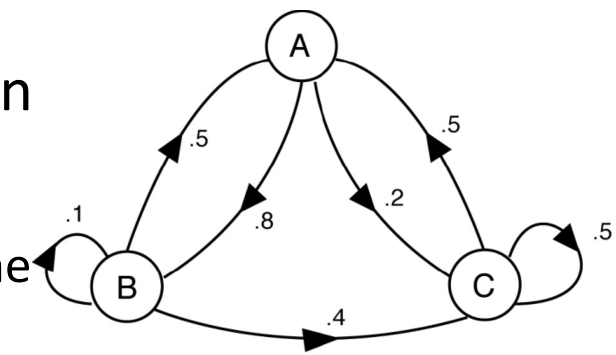
From: commons.wikimedia.org/w/index.php?curid=3573566

- Trouble of modeling the production of symbols (signal) and eliminating noise
- Important to model signal generation statistically

Shannon's Communication Theory

- The law of large numbers says that performing many of the same experiment will converge towards an expected value
  - Extended to the Central Limit Theorem states that independent and identically distributed variables will be normally distributed
  - But the symbols in language aren't independent of each other

- Markov Chains can describe the production of symbols in a language
  - Each symbol is dependent on the prior outcome
  - Modeled through memoryless dependence
  - Showed that dependent variables can converge towards an equilibrium among states



Markov graph of transiton probabilites between states A, B and C

Shannon's Entropy

- Quantifies
  - Average ***Uncertainty*** in an unknown message
  - Information we do not have
  - Amount of surprise in a message

- The less likely an event is, the more information it provides when it occurs

Shannon's Entropy

$$H(X) = -\sum_{x \in X} p(x) \log_2 p(x)$$

- Fundamental idea: If Entropy of an information source drops, then fewer questions are required to guess the outcome
- What does this mean for cryptography?
- What does this mean for cryptanalysis?

Entropy – Practice Problems

- What is the entropy of a single fair coin toss?

$$
\begin{aligned}
H(X) &= -(0.5 \log_2 0.5 + 0.5 \log_2 0.5) \\
&= -(0.5\,(-1) + 0.5\,(-1)) \\
&= -(-1) \\
&= 1
\end{aligned}
$$

- What is the entropy of a single non-fair coin toss when heads occurs with probability 0.75?
- What about as probability of heads approaches 1?

Shannon's Entropy

- H(X) is max when all possibilities are equally likely

- H(X) = 0 if for one event p = 1 and all others p = 0

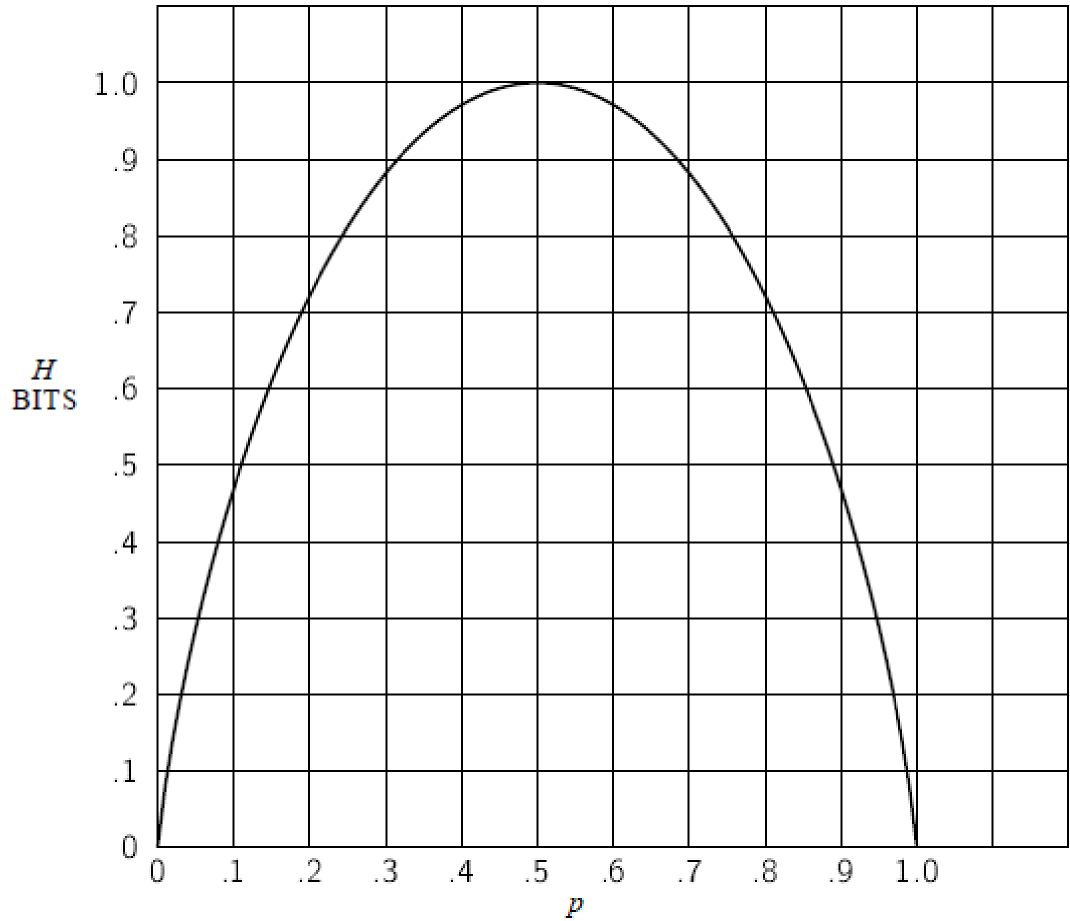


Fig. 7—Entropy in the case of two possibilities with probabilities p and $(1 -$

Entropy – Practice Problems (cont)

- What is the entropy of a character chosen from single case alphabetic symbols when all symbols are equally likely?

  $H(X) = -\ 1/26\ (\log_2 1/26) + \ldots + 1/26\ (\log_2 1/26))$

  $H(X) = -\sum_{i=1\ldots26} ((1/26)*\text{-}4.70044))$

  $H(X) = -\ (26 * ((1/26)*\text{-}4.70044))$

  $H(X) = 4.7044$

Shannon's Entropy
vs
Hartley's Information*

- What is the relationship between entropy and information?
  - Informally Entropy is the amount of information you don't have
  - Hartley: n-length sequences
  - Entropy (so far) measured per character
- When all probabilities are equal:

$$H(x) = -p \sum \log_2 p = \log_2 |A|$$

where $|A|$ is the size of the alphabet

Entropy – Practice Problems (cont)

- What is the entropy of sequence of single case alphabetic symbols when symbols follow the frequencies (a priori probabilities) of English?

$H(X) = - \sum_{i = 1 \ldots 26} ((p(ltr(i))*log(p(ltr(i))$

$H(X) = 4.18$

Entropy

- Entropy was reduced when frequency analysis was incorporated 4.70 → 4.18

- Reduced entropy allows for increased compression in encoding

  - Encoding exploits statistical redundancy

  - Encryption seeks to eliminate statistical redundancy

Should you encrypt before encoding? Why?

- Reduced entropy also means reduced security

Redundancy

- Shannon also termed ***redundancy*** as the difference between the quantity of bits used to represent information and the quantity of bits that the representation actually holds
- Redundancy (D) related to how much a language can be compressed
- D(x) = $H(x) - H(x^*)$
- *where H(x*) is a measured rate of Entropy over a set of messages using symbols in x*

Redundancy

- What is the redundancy for English per character just taking statistical frequencies into account?

- $H(x) = 4.70$
- $H(x^*) = 4.18$
- $R = 4.7 – 4.18 = 0.52$

Unicity Distance

- Minimum ciphertext required to allow an adversary to
  - Recover a unique encryption key
  - Reduce number of spurious keys to 0
- Small unicity distance → theoretical attack (with unlimited computational resources)

$$U = H(k)/D$$

Unicity Distance

- What is the Unicity Distance for a key using alphabetic substitution on English text just considering statistical frequencies?
- Key space = 26!   (probabilities equal)
- Entropy of a key in key space $H(k) = \log_2(26!)$

  $U = \log_2(26!) / 0.52$

  $U = 88.4/0.52$

  U = 170 char