16.36: Communication Systems Engineering

Lecture 2: Entropy

Eytan Modiano

Information content of a random variable

- **Random variable X**
  - **Outcome of a random experiment**
  - **Discrete R.V. takes on values from a finite set of possible outcomes**

    PMF: $P(X = y) = P_x(y)$

- **How much information is contained in the event X = y?**
  - **Will the sun rise today?**

    Revealing the outcome of this experiment provides no information

  - **Will the Celtics win the NBA championship?**

    Since this is unlikely, revealing yes provides more information than revealing no

- **Events that are less likely contain more information than likely events**

Measure of Information

- $I(x_i)$ = Amount of information revealed by an outcome $X = x_i$
- Desirable properties of $I(x)$:

1. If $P(x) = 1$ or $P(x) = 0$, then $I(x) = 0$
2. If $0 < P(x) < 1$, then $I(x) > 0$
3. If $P(x) < P(y)$, then $I(x) > I(y)$
4. If x and y are independent events then $I(x,y) = I(x)+I(y)$

- Above is satisfied by: $I(x) = Log_2(1/P(x))$
- Base of Log is not critical
  - Base 2 => information measured in bits

Entropy

- A measure of the information content of a random variable
- $X \in \{x_1, \ldots, X_M\}$
- $H(X) = E[I(X)] = \sum P(x_i) \, Log_2(1/P(x_i))$
- Example: Binary experiment
  - $X = x_1$ with probability p
  - $X = x_2$ with probability (1-p)
  - $H(X) = pLog_2(1/p) + (1-p)Log_2(1/(1-p)) = H_b(p)$
  - $H(X)$ is maximized with $p=1/2$, $H_b(1/2) = 1$

    Not surprising that the result of a binary experiment can be conveyed using one bit

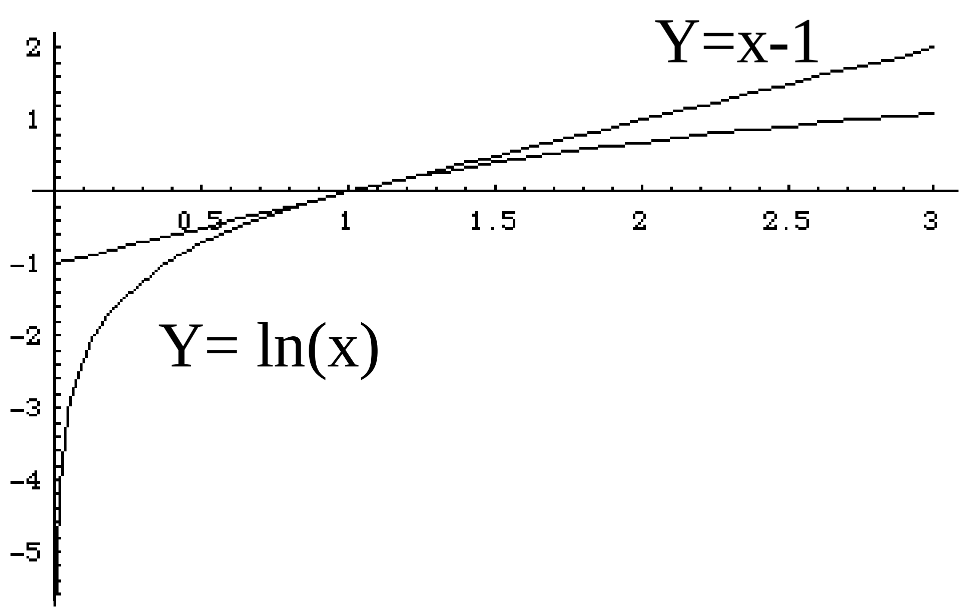
Simple bounds on entropy

- Theorem: Given a random variable with M possible values
  - $0 <= H(X) <= Log_2(M)$

  A) $H(X) = 0$ if and only if $P(x_i) = 1$ for some i

  B) $H(X) = Log_2(M)$ if and only if $P(x_i) = 1/M$ for all i

- Proof of A is obvious
- Proof of B requires
- the Log Inequality:
- if $x>0$ then $\ln(x) <= x-1$
- Equality if $x=1$



Proof, continued

Consider the sum $\sum_{i=1}^{M} P_i Log(\frac{1}{MP_i})$, by log inequality:

$$\leq \sum_{i=1}^{M} P_i(\frac{1}{MP_i} - 1) = \sum_{i=1}^{M}(\frac{1}{M} - P_i) = 0, \text{ equality when } P_i = \frac{1}{M}$$

Writing this in another way:

$$\sum_{i=1}^{M} P_i Log(\frac{1}{MP_i}) = \sum_{i=1}^{M} P_i Log(\frac{1}{P_i}) + \sum_{i=1}^{M} P_i Log(\frac{1}{M}) \leq 0, \text{equality when } P_i = \frac{1}{M}$$

$$\textit{That is, } \sum_{i=1}^{M} P_i Log(\frac{1}{P_i}) \leq \sum_{i=1}^{M} P_i Log(M) = Log(M)$$

Joint Entropy

Joint entropy: $H(X,Y) = \sum_{x,y} p(x,y)\log(\frac{1}{p(x,y)})$

Conditional entropy: H(X | Y) = uncertainty in X given Y

$$H(X|Y=y) = \sum_{x} p(x|Y=y)\log(\frac{1}{p(x|Y=y)})$$

$$H(X|Y) = E[H(X|Y=y)] = \sum_{y} \mathrm{p}(\mathrm{Y}=\mathrm{y})H(X|Y=y)$$

$$H(X|Y) = \sum_{x,y} p(x,y)\log(\frac{1}{p(x|Y=y)})$$

In General: $X_1,\ldots,X_n$ random variables

$$H(X_n | X_1,\ldots,X_{n-1}) = \sum_{x_1,\ldots,x_n} p(x_1,\ldots,x_n)\log(\frac{1}{p(x_n|x_1,\ldots,x_{n-1})})$$

Rules for entropy

1. Chain rule:

$$H(X_1, .., X_n) = H(X_1) + H(X_2|X_1) + H(X_3|X_2,X_1) + \ldots + H(X_n|X_{n-1}\ldots X_1)$$

2. $H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)$

3. If $X_1, .., X_n$ are independent then:

$$H(X_1, .., X_n) = H(X_1) + H(X_2) + \ldots + H(X_n)$$

If they are also identically distributed (I.I.d) then:

$$H(X_1, .., X_n) = nH(X_1)$$

4. $H(X_1, .., X_n) <= H(X_1) + H(X_2) + \ldots + H(X_n)$ (with equality if independent)

Proof: use chain rule and notice that $H(X|Y) < H(X)$
entropy is not increased by additional information

Mutual Information

- X, Y random variables
- Definition: $I(X;Y) = H(Y) - H(Y|X)$
- Notice that $H(Y|X) = H(X,Y) - H(X) \Rightarrow I(X;Y) = H(X)+H(Y) - H(X,Y)$
- $I(X;Y) = I(Y;X) = H(X) - H(X|Y)$
- Note: $I(X,Y) >= 0$ (equality if independent)
  - Because $H(Y) >= H(Y|X)$