Plan

- Entropy methods
- Statistical physics

Lecture 1 Entropy

- Information theory : efficiently encode complicated sets A by simpler ones

  Ex: Sending messages (complicated set) through some channel (w./ noise) by $\{0,1\}$-strings (simpler sets)

  Here · encoding : injection
  · efficiently : want it to be "as bijective as possible"

- We talk about entropy of a discrete random variable X defined over a finite set A.

  $H(X) \in \mathbb{R}$ non-negative real measuring the amount of
  - information
  - surprise
  - randomness

Formally, let X disc. r.v. over A.

define $$H(X) = \sum_{x \in A} \Pr(X=x) \cdot \log \frac{1}{\Pr(X=x)}$$

- All log is base 2.
- $0 \log 0 = 0$ convention.

Example. Let X be the (r.v.) outcome of a coin flip.

1. Coin completely biased: two sides = heads

$$H(X) = 1 \cdot \log \frac{1}{1} = 0$$

As outcome is a certainty $\Rightarrow$ $H(X)=0$ means zero infor/surprise/randomness

2. Coin is balanced & lands in head/tail w/ prob. $\frac{1}{2}$ each.

$$H(X) = \frac{1}{2} \log \frac{1}{1/2} + \frac{1}{2} \log \frac{1}{1/2} = 1$$

- In general, X is over A, w/ $|A| = n$.

$\Rightarrow H(X) \leq \log n$ w/ equality iff X is uniform over A.

- For combinatorial problems, say we want estimate $|A|$.
  - define r.v. $X \sim A$
    X is uniformly distributed.
  - $H(X) = \log |A|$
    ↑ bounds on $H(X)$

- Let X be a r.v. over A
  Y be a r.v. over B

the joint entropy

$$H(X,Y) = \sum_{a \in A, b \in B} \Pr(X=a, Y=b) \log \frac{1}{\Pr(X=a, Y=b)}$$

- the conditional entropy of Y given X is

$$H(Y|X) = \mathop{\mathbb{E}}_{a \in A} \left[H(Y|X=a) \right]$$

$$= \sum_{a \in A} \Pr(X=a) \cdot H(Y|X=a).$$

1.1 Shannon - Khinchin axioms for entropy.

We would like to "forget" entropy function, and do things axiomatically.

1. Invariance. $H(X)$ entropy depends solely on the prob. distribution of X.

I.e. if $Y = f(X)$, for some function f

$\Rightarrow H(X,Y) = H(X)$

- if f is bijective

$\Rightarrow H(X) = H(Y)$

2. Maximality. If X takes values over A, then $H(X)$ is maximised when $X \sim A$.

I.e. $H(X)$ higher $\Leftrightarrow^{X}$ more unpredictable.

3. Extensibility. If X is over A
Y is over $B \supseteq A$,

and if $Pr(X=a) = Pr(Y=a) \Rightarrow H(Y) = H(X)$

4. Additivity. $H(X,Y) = H(X) + H(Y|X)$

info carried by (X,Y)

info given by X

additional info we gain from Y after knowing X.

5. Continuity. $H(X)$ depends continuously on $\Pr(X=a)$.

6. Normalisation. $X \sim [2] \Rightarrow H(X) = 1$

Rmk All six axioms $\Rightarrow H(X) = \sum_{a \in A} \Pr(X=a) \log \frac{1}{\Pr(X=a)}$

1.2 Basic properties of entropy.

Lemma 1.1 (Independence) If X, Y are indep. r.v.

$\Rightarrow H(Y|X) = H(Y)$

and $H(X,Y) = H(X) + H(Y)$

In general, X_i indep. copies of X

$\Rightarrow H(X_1, \cdots, X_n) = n \cdot H(X)$ + induction

Pf: For any x, the distribution of Y given $X=x$ is the same as that of Y as X, Y are indep.

$H(Y)$ by invariance

$\Rightarrow H(Y|X) = \sum_x \Pr(X=x) \cdot H(Y|X=x)$

$$= \sum_x \Pr(X=x)\ H(Y)$$

$$= H(Y).$$

- $H(X,Y) \overset{\text{additivity}}{=} H(X) + H(Y|X)$

$$= H(X) + H(Y). \qquad \blacksquare$$

Lem 1.2 (Abs. certainty has no info, c.f. biased coin)

If X is a r.v. taking only one value

$$\Rightarrow H(X)=0.$$

Pf: • Take X_1, X_2 indep. copies of X.

By independence (Lem 1.1) $\Rightarrow H(X_1, X_2) = 2H(X)$

• As X (hence also X_i) takes only one value

by invariance $H(X_1, X_2) = H(X) = 2H(X)$

$$\Rightarrow H(X)=0. \qquad \blacksquare$$

Lem 1.3 Let $X \sim A$, $Y \sim B$. If $A \subseteq B$

$\Rightarrow H(X) \leq H(Y)$ w/ equality iff $A=B$.

Pf: • If $A=B$, invariance.

• By extensibility, we can view X as

defined on B: $\begin{cases} \Pr(X=a) = \frac{1}{|A|} & \text{for } a \in A \\ \Pr(X=b) = 0 & \text{for } b \in B \setminus A \end{cases}$

$\Rightarrow$ by maximality $H(X) \leq H(Y)$.

- For strict ineq, say $|A| < |B|$.
  - If $|A| = 1 \Rightarrow H(X) = 0$ by Lem 1.2.
  - Let $B' \subseteq B$ w/ $|B'| = 2$, take $Z \sim B'$, $Y \sim B'$

$$\Rightarrow H(Y) \geq H(Z) = 1 > 0 = H(X)$$

(↑ normalisation)

- Suppose then $|A| \geq 2$, $H(X) \geq 1$
  - Take indep copies X_i of X & Y_i of Y
  - Choose n suff. large so that

$$|A|^n \leq |B|^{n-1} \quad (\text{as } |A| < |B|)$$

  - By independence (Lem 1.1)

$$n H(X) = H(X_1, \cdots, X_n) \leq H(Y_1, \cdots, Y_{n-1}) = (n-1) H(Y)$$

$$\Rightarrow H(X) < H(Y). \qquad \blacksquare$$

- For a r.v. X w/ rational atom prob.

(say X defined over A, $\forall a \in A$, $\Pr(X=a) \in \mathbb{Q}$)

we can link it to uniform dist. as follows.

Construction 1.4 We may assume $\forall a \in A$

$$\Pr(X=a) = \frac{m_a}{n} \quad \text{for some } m_a \in \mathbb{N}$$

Let $U \sim [n]$. Can think of X is determined by U as follows.

Partition $[n] = (V_a, a \in A)$ w/

$|V_a| = m_a \quad \left(\sum_{a \in A} \frac{m_a}{n} = 1\right)$

Let X' be r.v. s.t.

$$X' = a \quad \text{if} \quad U \in V_a$$

$$\Pr(X'=a) = \Pr(U \in V_a) = \frac{m_a}{n}$$

So X & X' are identically distributed.

- We can also think of U in terms of X.

Let U' be r.v. over $[n]$ s.t.

if $X=a \Rightarrow U'$ is unif over V_a

i.e. $U' \mid X=a \sim V_a$

$U' \sim [n] \longleftarrow \frac{1}{m_a} \cdot \frac{m_a}{n} = \frac{1}{n}$.

Lem 1.5 (Non-negativity) Let X be a r.v. taking values in a finite set A.

$\Rightarrow H(X) \geq 0$.

Pf: · Suppose X takes values in A w/ rational probabilities: $\forall a \in A$, $P_r(X=a) = \frac{m_a}{n}$, $m_a \in \mathbb{N}$.

· Let $U \sim [n]$ and let X' be as Construction 1.4, so

$$\begin{cases} H(X') = H(X) \\ H(X', U) = H(U) \end{cases}$$

X' is determined by U (by invariance)

· Additivity $\Rightarrow H(X', U) = H(X') + H(U \mid X')$

$\Rightarrow H(X) = H(X') = H(X', U) - H(U \mid X')$

$= H(U) - H(U \mid X') \geq 0$

NTS

NTS: $H(U \mid X') \leq H(U)$

- Recall that $\forall a \in A$,

$$(U \mid X' = a) \sim V_a \subseteq [n]$$

$$U \sim [n]$$

Lem 1.3 $\Rightarrow$ $H(U \mid X' = a) \leq H(U)$

$\Rightarrow$ $H(U \mid X') \leq H(U)$ as a is arbitrary.

$$H(U \mid X') = \sum_{a \in A} \Pr(X' = a)\, H(U \mid X' = a)$$

$$\leq \sum_{a \in A} \Pr(X' = a)\, H(U)$$

$$= H(U).$$

$\Rightarrow$ $H(X) \geq 0$ when X has rational atom prob.

- general case follows from { • rationals dense in $\mathbb{R}$ • continuity . ▨

- if Y is determined by X, then Y carries no more info than X.

Lem 1.6 (monotonicity) Given r.v. X, Y.

if $Y = f(X)$ for some function f.

$\Rightarrow \quad H(Y) \leq H(X)$

Pf: • As Y is determined by X,

invariance $\Rightarrow \quad H(X,Y) = H(X)$

• Additivity $\quad H(X,Y) = H(Y) + H(X|Y)$

non-neg. $\geq H(Y)$ ▨