Def: Let P be a prob. distribution on a discrete Ω. The (Shannon) entropy

$$H(P) := \sum_{\omega \in \Omega} P(\omega) \log \frac{1}{P(\omega)}$$

For a discrete R.V. $X: \Omega \to \mathcal{X}$

$$H(X) := H(P_X) = \sum_{x \in \mathcal{X}} P_X(x) \log \frac{1}{P_X(x)} = \mathbb{E}\left(\log \frac{1}{P_X(X)}\right)$$

real R.V.

$\log \frac{1}{P_X(x)}$: the surprisal of $X=x$ happens, $H(X)$: the uncertainty/Randomness of P_X.

Rem 1. Basis of log

$$\boxed{\log_2 \longleftrightarrow \text{bits}} \qquad \log_{256} \longleftrightarrow \text{bytes} = 8 \text{ bits}$$

2. We agree $0 \log \frac{1}{0} = 0$ by $\lim_{x \to 0} x \log \frac{1}{x} = 0$.

Example (Bernoulli): $X \in \{0, 1\}$. $P(X=1) = p$ $\quad P(X=0) = 1-p$

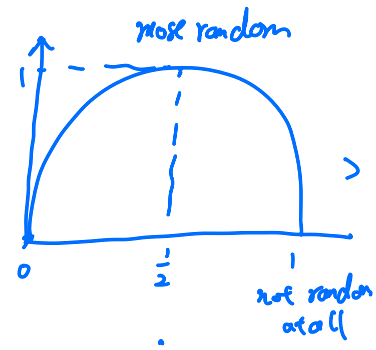
$$H(X) = p \log \frac{1}{p} + (1-p) \log \frac{1}{(1-p)} := h(p)$$

where $h(\cdot)$ is called binary entropy function

$h(\frac{1}{2}) = 1 \qquad h(0) = h(1) = 0$

In $\log_2$ basis, $\quad h(p) \le 1$ and

$h(p) = 1$ iff $p = \frac{1}{2}$



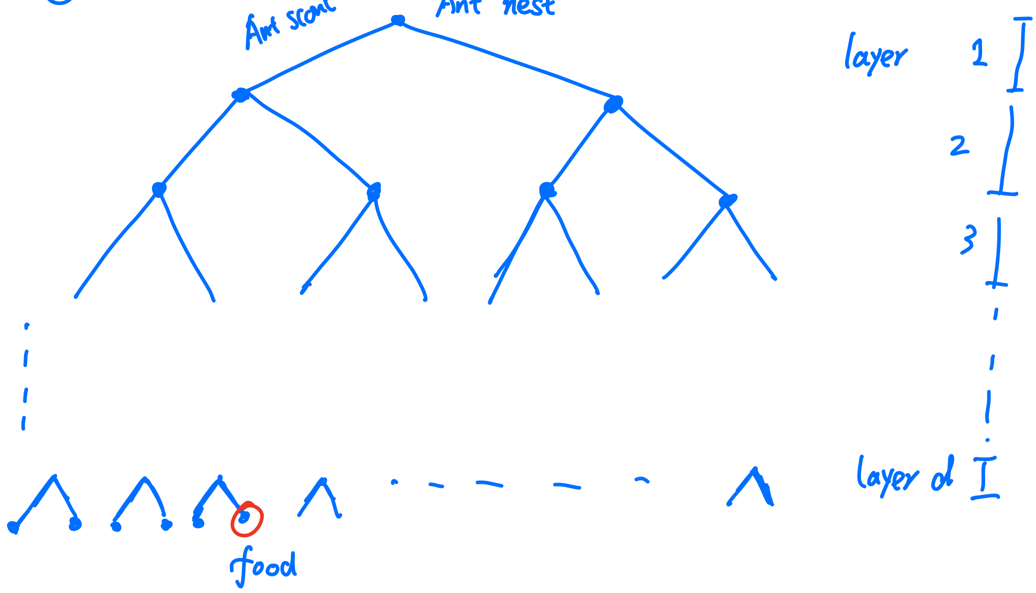
Example (Geometric): $X \in \{0, 1, 2, \dots\}$ $\qquad P(X=i) = p(1-p)^i \quad i = 0, 1, 2 \cdots$

$$\begin{aligned} H(X) &= \sum_{i=0}^{\infty} p(1-p)^i \log \frac{1}{p(1-p)^i} \\ &= \sum_{i=0}^{\infty} p(1-p)^i \left(\log \frac{1}{p} + i \log \frac{1}{(1-p)}\right) \\ &= \log \frac{1}{p} \sum_{i=0}^{\infty} p(1-p)^i + p \log \frac{1}{1-p} \sum_{i=0}^{\infty} i p(1-p)^i \\ &= \log \frac{1}{p} + p \log \frac{1}{(1-p)} \cdot \frac{1-p}{p^2} = \frac{h(p)}{p} \to +\infty \end{aligned}$$

as $p \to 0$

Example (∞ entropy): Can $H(X) = +\infty$? Yes, $P(X=k) = \frac{c}{k \ln^2 k}$, $k = 2, 3 \cdots$

Why "log": A experiment



Time for ant scout to describe the location of food $\sim \log_2 2^d = d$

left, right left ⋯ d binary digit

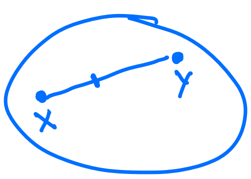
ant communication ≈ 1 bit/min

Convexity

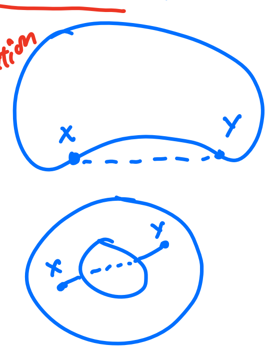
V a vector space ($V \cong \mathbb{R}^n$),

A subset $S \subseteq V$ is convex if

$\forall\ x, y \in S$, $\underbrace{\lambda x + (1-\lambda) y \in S}_{\text{Convex combination}}$ for $\lambda \in [0,1]$



Convex



not convex

Example : ① $\mathbb{R}^n$ is convex

$[0,1] \subseteq \mathbb{R}$, $(a.b) \subseteq \mathbb{R}$

② $\mathcal{P}(\mathcal{X}) = \{$ prob. distribution on $\mathcal{X} \}$

③ $\mathcal{P}_0(\mathbb{R}) = \{ P_X \mid \mathbb{E}(X) = 0 \} \subseteq \mathcal{P}(\mathbb{R})$

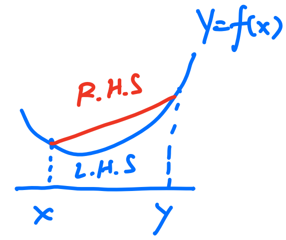
$\mathbb{E}(\lambda X + (1-\lambda) Y) = \lambda \mathbb{E}X + (1-\lambda)\mathbb{E}Y = 0$

A function $f : S \to \mathbb{R}$ is

(i) convex if $f(\lambda x + (1-\lambda) y) \leq \lambda f(x) + (1-\lambda) f(y)$, $\forall x.y \in S. \lambda \in [0.1]$

(ii) strictly convex if $f(\lambda x + (1-\lambda)) < \lambda f(x) + (1-\lambda) f(y)$, $\forall x \neq y \in S$, $\lambda \in (0.1)$

(iii) (strictly) concave if $-f$ is (strictly) convex



Example: ① $x \longmapsto x \log x$ convex strictly

$x \longmapsto \log x$ concave strictly

② $X \longmapsto \mathbb{E}X$ convex but not strictly (proof?)

Jensen inequality : $\forall X : \Omega \to S \subseteq \mathbb{R}^n$ vector valued R.V.

If f convex $\Rightarrow f(\mathbb{E}X) \leq \mathbb{E}f(X)$

strictly convex , then $f(\mathbb{E}X) = \mathbb{E}f(X)$ iff $X = \mathbb{E}X$ a.s.

constant R.V.

Pf: Convexity $\Rightarrow$ $f(\lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_n x_n)$

$$\leq \lambda_1 f(x_1) + \lambda_2 f(x_2) + \cdots \lambda_n f(x_n)$$

$$\lambda_i \geq 0 \qquad \sum_{i=1}^{n} \lambda_i = 1$$

$$f(\mathbb{E}X) = f\left(\sum_w P(w) X(w)\right) \leq \sum_w P(w) \, f(X(w))$$

$$\boxed{P(w) \geq 0 \quad \Sigma P(w) = 1} \qquad = \mathbb{E} f(X)$$

Properties of H

① $H(X) \geq 0$. $H(X) = 0$ iff X is constant

② If $\mathcal{X}$ is finite, $H(X) \leq \log |\mathcal{X}|$ with equality iff P_X is uniform on $\mathcal{X}$

③ For any bijective f, $H(X) = H(f(X))$

④ $P \longmapsto H(P)$ is strictly concave

Pf: ① $H(X) = \mathbb{E}\left[\log \frac{1}{P_X}\right] \geq 0$ $\qquad P_X(x) \leq 1$, $\log \frac{1}{P_X(x)} \geq 0$

② $H(X) = \mathbb{E}\left[\log \frac{1}{P_X}\right] \leq \log \mathbb{E}\left(\frac{1}{P_X}\right)$

$$= \log \sum_x P(x) \frac{1}{P(x)} = \log |\mathcal{X}|$$

equality iff $\log \frac{1}{P_X}$ is constant

$\Leftrightarrow$ P_X constant

$$\sum_{x \in \mathcal{X}} P_X(x) = 1 \Rightarrow P_X(x) = \frac{1}{|\mathcal{X}|}$$

③ $P_X(x) = P(\{w \mid X(w) = x\}) = P(\{w \mid f \circ X(w) = f(x)\}) = P_{f(X)}(f(x))$

$$H(X) = \sum_x P_X(x) \log \frac{1}{P_X(x)} = \sum_x P_{f(x)}(f(x)) \log \frac{1}{P_{f(x)}(f(x))} = H(X)$$

④: $H(\lambda P_1 + (1-\lambda) P_2) = \sum_w f(\lambda P_1(w) + (1-\lambda) P_2(w))$ $\quad f(t) = t\log\frac{1}{t}$

$\geq \sum_w \lambda f(P_1(w)) + (1-\lambda) f(P_2(w))$ $\quad = -t\log t$

concave

$= \lambda \sum_w f(P_1(w)) + 1-\lambda \sum_w f(P_2(w))$

$= \lambda H(P_1) + \lambda H(P_2)$

Random vector

Let $X_1 \cdots X_n : \Omega \to \mathcal{X}$ be R.V.s

Define $X^n = (X_1, \cdots X_n) : \Omega \to \mathcal{X}^n$ n-dim random vector.

Entropy: $H(X^n) = H(X_1, X_2, \cdots X_n)$

$$= \mathbb{E}\left[\frac{1}{\log P_{X_1 X_2 \cdots X_n}}\right]$$

In particular, for two R.V. X and Y

$$H(XY) = \mathbb{E}\left[\frac{1}{\log P_{XY}}\right] = \sum_{x,y} P_{XY}(X=x, Y=y) \log \frac{1}{P_{XY}(X=x, Y=y)}$$

Definition (Condition Entropy).

$$H(X|Y) = \mathbb{E}_{y \sim P_Y}\left[H(P_{X|Y=y})\right] = \mathbb{E}\left[\log \frac{1}{P_{X|Y}}\right]$$

Expected uncertainty $H(P_{X|Y=y})$ over $y \sim P_Y$.

Notation ① $P_{X|Y=y}$ is a distribution on X $\quad P_{X|Y=y}(x) = P(X=x | Y=y)$

$$= \frac{P(X=x, Y=y)}{P(Y=y)}$$

② $P_{X|Y}$ is a R.V. on $\mathcal{X} \times \mathcal{Y}$: $P_{X|Y}(x,y) = P(X=x | Y=y)$

Example: A fair die $\Omega = \{1, 2, \cdots, 6\}$

$$X = \begin{cases} \text{large} \\ \text{small} \end{cases} \qquad Y = \begin{cases} \text{even} \\ \text{odd} \end{cases}$$

$$H(X) = 1 \qquad H(Y) = 1$$

$P(S\&E) = \frac{1}{6}$ $\quad P(L\&E) = \frac{1}{3}$ $\quad P(S\&O) = \frac{1}{3}$ $\quad P(L\&O) = \frac{1}{6}$

$$H(XY) = \frac{1}{6}\log 6 + \frac{1}{3}\log 3 + \frac{1}{3}\log 3 + \frac{1}{6}\log 6$$

$$= \log 3 + \frac{1}{3}\log 2$$

$P_{X|Y}(S|E) = \frac{1}{3}$ $\quad P_{X|Y}(L|E) = \frac{2}{3}$

$P_{X|Y}(S|O) = \frac{2}{3}$ $\quad P_{X|Y}(L|O) = \frac{1}{3}$

$$H(X|Y) = P_{XY}(SE)\log\frac{1}{P_{X|Y}(S|E)} + \cdots$$

$$= \frac{1}{6}\log 3 + \frac{1}{3}\log\frac{3}{2} + \frac{1}{3}\log\frac{3}{2} + \frac{1}{6}\log 3$$

$$= \log 3 + \frac{2}{3}\log\frac{1}{2}$$

Properties of $H(X|Y)$

① $H(X|Y) \leq H(X)$ with "=" iff X and Y independent

② $H(XY) = H(Y) + H(X|Y) \leq H(Y) + H(X)$

with "=" iff X and Y independent

③ $H(XY) \geq \max\{H(X), H(Y)\}$

Pf: ① Using concavity of $P \mapsto H(P)$

$$H(X|Y) = \mathbb{E}_{Y \sim P_Y} H(P_{X|Y=y}) \leq H(\mathbb{E}_{Y \sim P_Y} P_{X|Y=y})$$

$$= H(P_X) = H(X)$$

② $H(XY) = \mathbb{E}\left[\log \frac{1}{P_{XY}}\right]$ $\qquad P_{XY}(x, y) = P_{X|Y}(x|y) P_Y(y)$

$$= \mathbb{E}\left[\log \frac{1}{P_{X|Y} \cdot P_Y}\right]$$

$$= \mathbb{E}\left[\log \frac{1}{P_{X|Y}} + \log \frac{1}{P_Y}\right]$$

$$= \mathbb{E}\left[\log \frac{1}{P_{X|Y}}\right] + \mathbb{E}\left[\log \frac{1}{P_Y}\right]$$

$$= H(X|Y) + H(Y)$$

③ $H(XY) = H(X) + H(Y|X) \geq H(X)$

Corollary. For any function f,

① $H(X) = H(X, f(X))$, ② $H(f(X)|X) = 0$

① $H(X) \geq H(f(X))$

with equality iff f injective

Pf: ① $P_{X f(X)}(x, y) = \begin{cases} P_X(x), & \text{if } y = f(x) \\ 0. & \text{otherwise} \end{cases}$

② $H(X) = H(X, f(X)) = H(X) + H(f(X)|X)$

More explicitly. $P_{f(X)|X}(y|x) = \begin{cases} 1 & \text{if } y = f(x) \\ 0 & \text{otherwise} \end{cases}$

$$H(f(X)|X) = \mathbb{E}_{x \sim P_X}\left(H(f(X)|X=x)\right) = 0$$

③ $H(X) = \underset{\overset{\vee\!|}{0}}{H(X|f(X))} + H(f(X))$

Equality $\Rightarrow$ $H(X|f(X)) = 0$

$\Rightarrow$ $\mathbb{E}_{y \sim P_{f(X)}} H(X|f(X)=y) = 0$

$\Rightarrow$ $H(X|f(X)=y) = 0 \ \forall y$

$\Rightarrow$ $X = g f(x) \ \forall y \in f(\mathcal{X}) \Rightarrow f$ injective.

History : Thermo dynamics

No Perpetual Motion Machine by conservation of energy.

1st law

2nd law: No machine can produce work by only drawing heat from a warm body but without expend heat to environment.

heat work

(No free conversion from heat to work)

3rd law: Entropy cannot reduce.

Boltzmann & Gibbs

Entropy of ideal gas $$S = kn \sum_{j=1}^{L} p_j \log \frac{1}{p_j}$$

k Boltzmann constant

n # of particle