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. $X : \Omega \to \mathcal{X}$

$$H(X) := H(P_X) = \sum_{x \in \mathcal{X}} P_X(x) \log \frac{1}{P_X(x)} = \mathbb{E}\Big(\underbrace{\log \frac{1}{P_X(x)}}_{\text{real R.V.}}\Big)$$

$\log \frac{1}{P_X(x)}$: the suprisal 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): $X \in \{0,1\}$. $P(X=1) = p$ $\quad P(X=0) = 1-p$

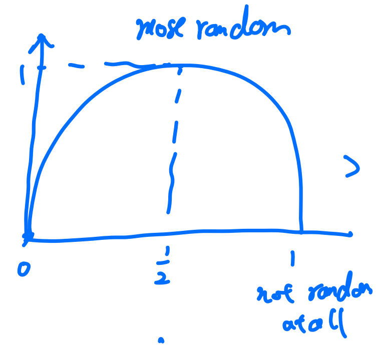
$$H(X) = p \log \frac{1}{p} + (1-p) \log \frac{1}{(1-p)} := h(p)$$

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 $p = \frac{1}{2}$

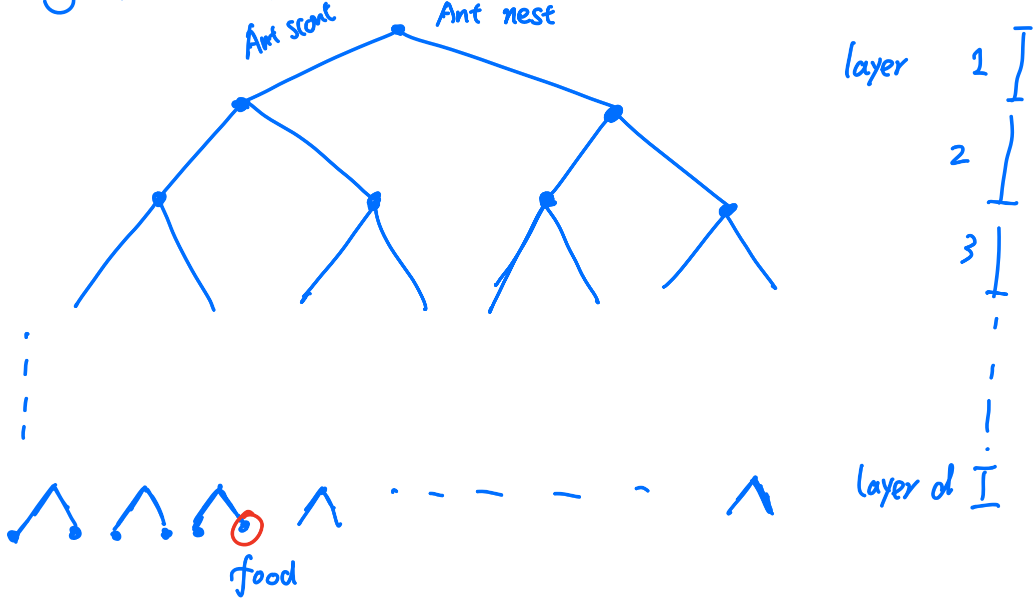


Example (Geometric): $X \in \{0,1,2,\dots\}$ $\qquad P(X=i) = p(1-p)^i \quad i = 0,1,2\dots$

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

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

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

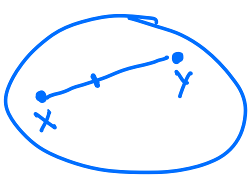
ant communication 0.7 - 1 bit / min

Convexity

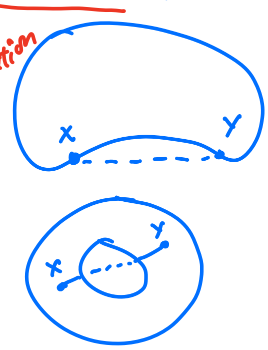
V a vector space ($V \cong \mathbb{R}^n$),

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

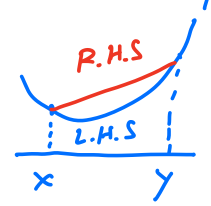
$\mathbb{E}(\lambda X+(1-\lambda)Y) = \lambda \mathbb{E}X+(1-\lambda)\mathbb{E}Y=0$

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

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

(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 $\Rightarrow f(\mathbb{E}X) \le \mathbb{E}f(X)$

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

constant R.V.

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

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

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

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

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

Properties of H

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

② If $\mathcal{X}$ is finite, $H(X) \le \log |\mathcal{X}|$ 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 $H(X) = \mathbb{E}\left[\log \frac{1}{P_X}\right] \ge 0$ $\qquad P_X(x) \le 1$, $\log \frac{1}{P_X(x)} \ge 0$

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

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

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

$= \lambda \sum_{w} f(P_1(w)) + 1-\lambda \sum_{w} f(P_2(w))$

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