ORIE 4742 - Info Theory and Bayesian ML

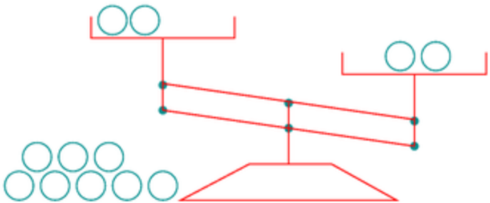
Lecture 3: Measuring Information

February 7, 2021

Sid Banerjee, ORIE, Cornell

Mackay's weighing puzzle

The weighing problem



You are given 12 balls, all equal in weight except for one that is either heavier or lighter.

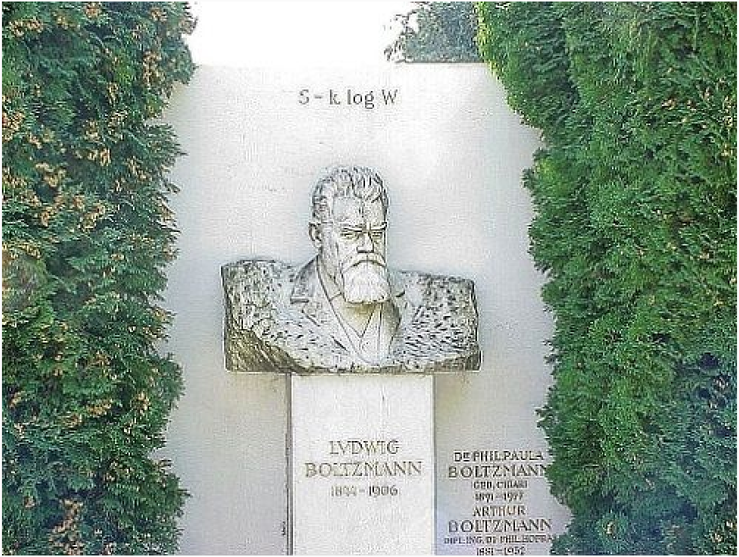
Design a strategy to determine

which is the odd ball

and whether it is heavier or lighter,

in as few uses of the balance as possible.

how much ‘information’ does a random variable have?



reading assignment: chapter 4 of Mackay

quantifying information content

measuring information

consider (discrete) rv X taking values $\mathcal{X} = \{a_1, a_2, \ldots, a_k\}$, with probability mass function $\mathbb{P}[X = a_i] = p_i \,\forall\, i, \sum_{i=1}^{k} p_i = 1$

Shannon's entropy function

- outcome $X = a_i$ has *information content*
$$h(a_i) = \log_2\left(\frac{1}{p_i}\right)$$
- random variable X has *entropy*
$$H(X) = \mathbb{E}[h(X)] = \sum_{i=1}^{k} p_i \log_2\left(\frac{1}{p_i}\right)$$

entropy: basic properties

Shannon's entropy function

- outcome $X = a_i$ has *information content*: $h(a_i) = \log_2\left(\frac{1}{p_i}\right)$
- random variable X has *entropy*: $H(X) = \mathbb{E}[h(X)] = \sum_{i=1}^{k} p_i \log_2\left(\frac{1}{p_i}\right)$

- only depends on distribution of X (i.e., $H(X) = H(p_1, p_2, \ldots, p_k)$)
- $H(X) \geq 0$ for all X
- if $X \perp\!\!\!\perp Y$, then $H(X,Y) = H(X) + H(Y)$
  where joint entropy $H(X,Y) \triangleq \sum_{(x,y)} p(x,y) \log_2 1/p(x,y)$

entropy: basic properties

Shannon's entropy function

- outcome $X = a_i$ has *information content*: $h(a_i) = \log_2\left(\frac{1}{p_i}\right)$
- random variable X has *entropy*: $H(X) = \mathbb{E}[h(X)] = \sum_{i=1}^{k} p_i \log_2\left(\frac{1}{p_i}\right)$

- if $X \sim$ uniform on $\mathcal{X}$, then $H(X) = \log_2 |\mathcal{X}|$; else, $H(X) \leq \log_2 |\mathcal{X}|$

designing questions to maximize information gain

the game of 'sixty three'

guess number $x \in \{0, 1, 2, \ldots, 62, 63\}$

designing questions to maximize information gain

the game of ’submarine’

player 1 hides a submarine in one square of an 8×8 grid

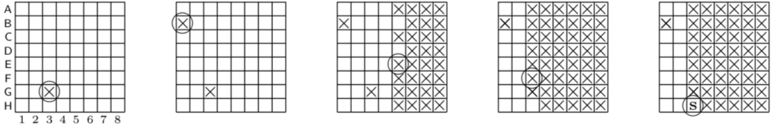
player 2 shoots at one square per round

designing questions to maximize information gain

the game of 'submarine'

player 1 hides a submarine in one square of an 8×8 grid

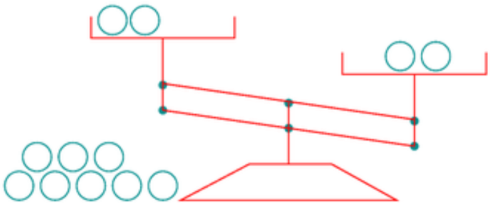
player 2 shoots at one square per round



move #	1	2	32	48	49
question	G3	B1	E5	F3	H3
outcome	$x = \mathtt{n}$	$x = \mathtt{n}$	$x = \mathtt{n}$	$x = \mathtt{n}$	$x = \mathtt{y}$
$P(x)$	$\frac{63}{64}$	$\frac{62}{63}$	$\frac{32}{33}$	$\frac{16}{17}$	$\frac{1}{16}$
$h(x)$	0.0227	0.0230	0.0443	0.0874	4.0
Total info.	0.0227	0.0458	1.0	2.0	6.0

Mackay’s weighing puzzle

The weighing problem



You are given 12 balls, all equal in weight except for one that is either heavier or lighter.

Design a strategy to determine

which is the odd ball

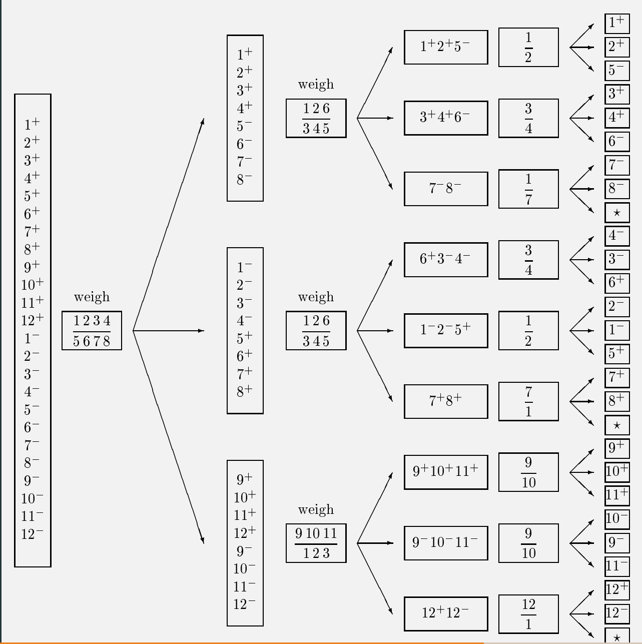
and whether it is heavier or lighter,

in as few uses of the balance as possible.

information acquisition in the weighing puzzle

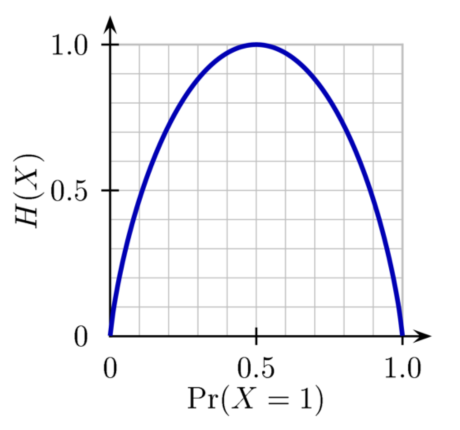
information acquisition in the weighing puzzle

weighing game: an optimal solution



binary entropy function

if X Bernoulli(p), then $H(X) \triangleq H_2(p) = -p\log_2(p) - (1-p)\log_2(1-p)$



- (useful formula) for any $k, N \in \mathbb{N}$, $k \leq N$: $\binom{N}{k} \approx 2^{NH_2(k/N)}$

conditional entropy

suppose $X \sim \{p_1, p_2, p_3, p_4\}$, and let $Y = \mathbb{1}_{[X \in \{a_1, a_2\}]}$; then we have

$$H(X) = H(Y) + (p_1 + p_2) H_2\left(\frac{p_1}{p_1 + p_2}\right) + (p_3 + p_4) H_2\left(\frac{p_3}{p_3 + p_4}\right)$$

conditional entropy

suppose $X \sim \{p_1, p_2, p_3, p_4\}$, and let $Y = \mathbb{1}_{[X \in \{a_1, a_2\}]}$; then we have

$$H(X) = H(Y) + (p_1 + p_2) H_2\left(\frac{p_1}{p_1 + p_2}\right) + (p_3 + p_4) H_2\left(\frac{p_3}{p_3 + p_4}\right)$$

conditional entropy

for any rvs X, Y: $H(X|Y) = \sum_{y \in \mathcal{Y}} p(y) H(X|Y = y)$
$= \sum_{y \in \mathcal{Y}} p(y) \sum_{x \in \mathcal{X}} p(x|y) \log_2(1/p(x|y))$

conditional entropy

conditional entropy

for any rvs X, Y: $H(X|Y) = \sum_{y \in \mathcal{Y}} p(y) H(X|Y = y)$
$= \sum_{y \in \mathcal{Y}} p(y) \sum_{x \in \mathcal{X}} p(x|y) \log_2(1/p(x|y))$

the chain rule

for any rvs X, Y:

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

blank

blank

blank