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?

- 2 state lotteries $S_1, S_2$, winning number is $X_1 = 1, X_2 = 1$
  Suppose $S_1 = \text{Vermont}, S_2 = \text{Texas}$ ($N_1 = \#\text{ of people in lottery } 1 << N_2$)
  - If we do not know $X_1, X_2$, then is $X_1 = 1$ or $X_2 = 1$ more surprising?
  - Is $X_1 = 1$ more/less surprising than $X_1 = 12793$?

axioms of ‘information’ - info exists only if uncertainty
  - the exact information does not matter
    (only the ‘surprise’ matters)
  - more ‘surprising’ r.v. have more info

(Shannon ’48)

Idea - Information of a r.v. $\equiv$ amount of uncertainty resolved by knowing the r.v.
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 $P[X = a_i] = p_i \forall i, \sum_{i=1}^{k} p_i = 1$

Shannon's entropy function

- outcome $X = a_i$ has information content
  $$h(a_i) = \log_2 \left( \frac{1}{p_i} \right)$$
  - $p_i$ large $\Rightarrow$ $h(a_i)$ is small
  - "convention" (bits)

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

<table>
<thead>
<tr>
<th>$X$</th>
<th>$h(X)$</th>
<th>$P(X)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>$\log_2 \left( \frac{1}{p_1} \right)$</td>
<td>$P_1$</td>
</tr>
<tr>
<td>$a_2$</td>
<td>$\log_2 \left( \frac{1}{p_2} \right)$</td>
<td>$P_2$</td>
</tr>
<tr>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
</tr>
<tr>
<td>$a_k$</td>
<td>$\log_2 \left( \frac{1}{p_k} \right)$</td>
<td>$P_k$</td>
</tr>
</tbody>
</table>
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 $X$ \( \because \log \left( \frac{1}{p_i} \right) \geq 0 \ \forall \ i \)
- if $X \perp Y$, then $H(X, Y) = H(X) + H(Y)$

where joint entropy $H(X, Y) \triangleq \sum_{(x,y)} p(x, y) \log_2 \frac{1}{p(x, y)}$

\[
= \sum_{(x,y)} p(x) p(y) \left( - \log_2 p(x) - \log_2 p(y) \right)
\]

\[
= \left( \sum_x - p(x) \log_2 p(x) \right) + \left( \sum_y - p(y) \log_2 p(y) \right)
\]
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: $H(X) = \mathbb{E}[h(X)] = \sum_{i=1}^{k} p_i \log_2 \left( \frac{1}{p_i} \right)$

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

\begin{align*}
1 & \quad - \sum_{i=1}^{\frac{|\mathcal{X}|}{k}} p_i \log p_i = - \sum_{i=1}^{\frac{|\mathcal{X}|}{k}} \frac{1}{|\mathcal{X}|} \log \frac{1}{|\mathcal{X}|} = \log |\mathcal{X}| \\
2 & \quad \forall \{p_i\} \text{ s.t. } p_i \geq 0, \sum_{i=1}^{|\mathcal{X}|} p_i = 1, \quad \max - \sum_{i=1}^{\frac{|\mathcal{X}|}{k}} p_i \log p_i \leq \log_2 |\mathcal{X}| \\
3 & \quad \text{Idw} \quad H(X) = \mathbb{E}[h(X)] \quad \text{where } h(x) = -\log p(x)
\end{align*}
\[
\begin{align*}
\mathbb{E}[h(x)] &= \mathbb{E} \left[ \log_2 \left( \frac{1}{p(x)} \right) \right] \\
\text{Jensen's Inequality: } \mathbb{E}\left[f(x)\right] &\geq f\left(\mathbb{E}[x]\right) \\
\implies \mathbb{E}\left[\log(g(x))\right] &\leq \log\left(\mathbb{E}[g(x)]\right) \\
\implies \mathbb{E}[h(x)] &= \mathbb{E} \left[ \log_2 \left( \frac{1}{p(x)} \right) \right] \\
&\leq \log_2 \left[ \mathbb{E} \left[ \frac{1}{p(x)} \right] \right] \\
&= \log_2 \left( \sum_{i=1}^{\infty} p_i \cdot \left( \frac{1}{p_i} \right) \right) = \log_2 |x| \\
&= \log_2 |x|
\end{align*}
\]
designing questions to maximize information gain *(heuristic)*

**the game of 'sixty three'**

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

Q: how many (and what) yes. No questions should you ask?

\[ Q_1 - \text{Is } X \geq 32 \ ? \begin{cases} \text{Yes} & \text{Is } X \geq 16 \ \text{-} \text{ -} \text{ -} \\ \text{No} & \end{cases} \]

**Ideal**

Binary search

\# of questions = 6 $\geq H(X)$ $(H(X) = 6$ if $X \sim \text{Unif}\{0, \ldots, 63\}$)

Q1 - Is $X$ even? \[ \begin{cases} \text{Yes} & \text{Is } X/2 \text{ odd or even?} \ \text{-} \text{ -} \text{ -} \\ \text{No} & \text{Is } X+1/2 \text{ odd or even?} \ \text{-} \text{ -} \text{ -} \end{cases} \]

Claim - Amount of entropy in each answer = 1 bit
designing questions to maximize information gain

the game of 'submarine'

player 1 hides a submarine in one square of an 8 \times 8 grid
player 2 shoots at one square per round

\[ X = \{(x, y) ; x \in \{1, \ldots, 8\}, y \in \{1, \ldots, 8\}\} \]

If \( X \sim \text{Unif}(X) \), then \( H(X) = 6 \) \((=3+3)\)

- Question \( = (Q_x, Q_y) \)

\[ Q_1 \equiv \text{Is } (x, y) = (1, 1) \ ? \]

\[ Q_2 \equiv \text{Is } (x, y) = (1, 2) \ ? \]

\[ h(Y_1) = -\frac{1}{64} \log_2 \frac{1}{64} - \frac{63}{64} \log_\frac{63}{64} \]

\[ h(Y_2) = -\frac{1}{63} \log_3 \frac{1}{63} - \frac{62}{63} \log_\frac{62}{63} \]
designing questions to maximize information gain

the game of 'submarine'

player 1 hides a submarine in one square of an $8 \times 8$ grid
player 2 shoots at one square per round

<table>
<thead>
<tr>
<th>move #</th>
<th>question</th>
<th>outcome</th>
<th>$P(x)$</th>
<th>$h(x)$</th>
<th>Total info.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>G3</td>
<td>$x = n$</td>
<td>$\frac{63}{64}$</td>
<td>0.0227</td>
<td>0.0227</td>
</tr>
<tr>
<td>2</td>
<td>B1</td>
<td>$x = n$</td>
<td>$\frac{62}{63}$</td>
<td>0.0230</td>
<td>0.0458</td>
</tr>
<tr>
<td>32</td>
<td>E5</td>
<td>$x = n$</td>
<td>$\frac{32}{33}$</td>
<td>0.0443</td>
<td>1.0</td>
</tr>
<tr>
<td>48</td>
<td>F3</td>
<td>$x = n$</td>
<td>$\frac{16}{17}$</td>
<td>0.0874</td>
<td>2.0</td>
</tr>
<tr>
<td>49</td>
<td>H3</td>
<td>$x = y$</td>
<td>$\frac{1}{16}$</td>
<td>4.0</td>
<td>6.0</td>
</tr>
</tbody>
</table>
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.
What is the best you can do? \( X \equiv \text{true outcome} \)

- \( X \equiv \text{set of outcomes} = \{(1, h), (1, l), (2, h), (2, l), \ldots (24, l)\} \)

\[ |X| = 24 \Rightarrow H(X) = \log_3 24 \text{ trits} = \log_2 24 \text{ bits} \]

- Consider each weighing - 3 outcomes - LH, RH, E

max info per weighing = \( \log_3 3 = 1 \) in 'trits'

\( \log_2 3 \) bits

\[ \Rightarrow \text{Need } k \text{ questions s.t. } k \log_3 3 \geq \log_2 24 \]

\[ \Rightarrow k \geq 3 \]
information acquisition in the weighing puzzle
weighing game: an optimal solution

- Ch 2 of Mackay
- Ch 1, Ch 4 = source coding
binary entropy function

If $X \sim \text{Bernoulli}(p)$, then

$$H(X) \triangleq H_2(p) = -p \log_2(p) - (1 - p) \log_2(1 - p)$$

- (useful formula) for any $k, N \in \mathbb{N}, k \leq N$:

$$\binom{N}{k} \approx 2^{NH_2(k/N)}$$
Suppose $X \sim \{p_1, p_2, p_3, p_4\}$, and let $Y = 1_{X \in \{a_1, a_2\}}$; then we have

$$H(X) = H(Y) + (p_1 + p_2)H_2\left(\frac{p_1}{p_1 + p_2}\right) + (p_3 + p_4)H_2\left(\frac{p_3}{p_3 + p_4}\right)$$
suppose $X \sim \{p_1, p_2, p_3, p_4\}$, and let $Y = 1_{\{X \in \{a_1, a_2\}\}}$; then we have

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

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

$$= \sum_{y \in Y} p(y) \sum_{x \in X} p(x|y) \log_2(1/p(x|y))$$
the chain rule

the chain rule (information content)

for any rvs $X, Y$ and realizations $x, y$:

$$h(x, y) = h(x) + h(y|x) = h(y) + h(x|y)$$
the chain rule

the chain rule (entropy)

for any rvs $X, Y$:

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