Assignment 1 - Information Theory and Inference

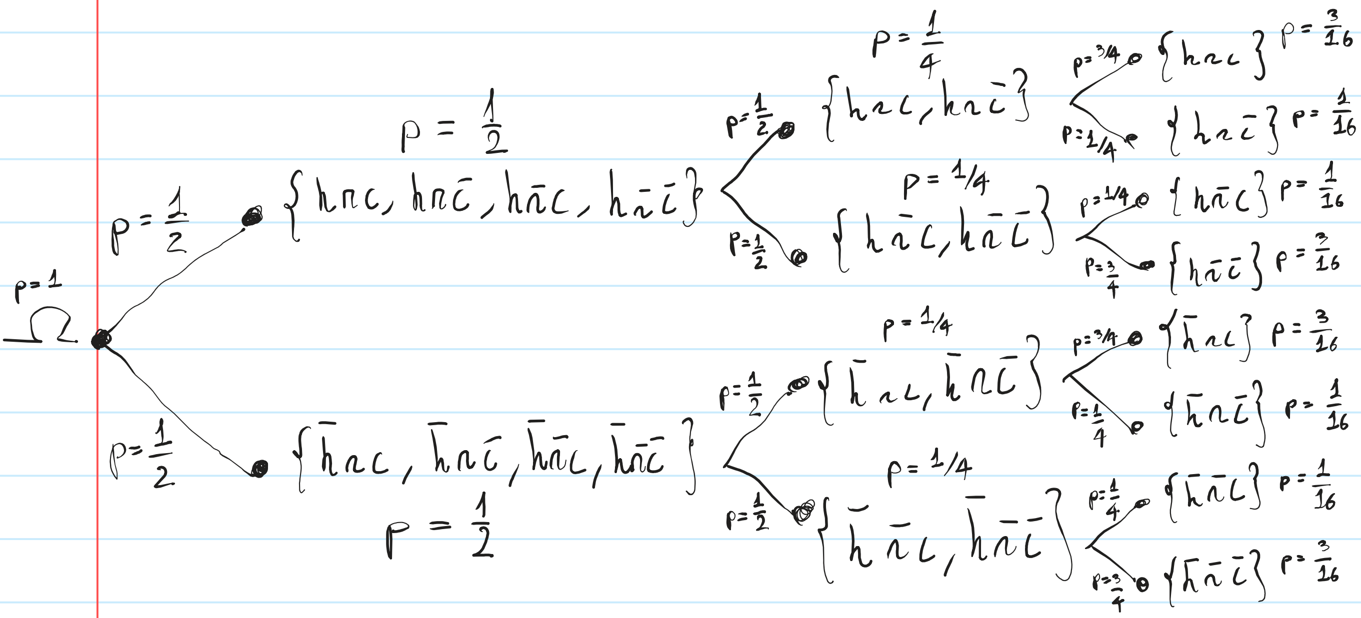
Gabriel Amorosetti - Matricola N. 2107530

We can characterize the weather as hot/cold $(h, \bar{h})$, rainy/non-rainy $(r, \bar{r})$, cloudy/non-cloudy $(c, \bar{c})$ with

$$P(h,r,c) = 3/16, \quad P(h,r,\bar{c}) = 1/16, \quad P(h,\bar{r},c) = 1/16, \quad P(h,\bar{r},\bar{c}) = 3/16$$

$$P(\bar{h},r,c) = 3/16, \quad P(\bar{h},r,\bar{c}) = 1/16, \quad P(\bar{h},\bar{r},c) = 1/16, \quad P(\bar{h},\bar{r},\bar{c}) = 3/16$$

Draw a probability tree for this probability space (including probabilities for nodes and edges) and compute the Shannon entropy at each level of the tree. How many questions are necessary (on average) to identify the outcome?



$$\boxed{\begin{array}{ll} H(S) = -\sum_{i=1,\dots,|\Omega|} P(\{\omega_i\}) \log P(\{\omega_i\}) & \\ S = (\Omega, \mathcal{F}, P) & \text{SHANNON ENTROPY} \end{array}}$$

$(\log \equiv \log_2)$

First level : $H(h) = -\sum_i P(\{h_i\}) \log P(\{h_i\})$

$$H(h) = -P(h) \log(h) - P(\bar{h}) \log(\bar{h})$$
$$= -\frac{1}{2} \log \frac{1}{2} - \frac{1}{2} \log \frac{1}{2}$$
$$= +\frac{1}{2} + \frac{1}{2}$$
$$= 1 \quad [bit]$$

Second level : $H(x) = -4 \cdot \frac{1}{4} \log \frac{1}{4}$

$= 2$ [bit]

Third level : $H(c) = -4 \frac{3}{16} \log \frac{3}{16}$

$-4 \frac{1}{16} \log \frac{1}{16}$

$= -\frac{3}{4} \log (3-4)$

$- \frac{1}{4} \times (-4)$

$= -\frac{3}{4} \log 3 + 4$

The total entropy $H(c) = -\frac{3}{4} \log 3 + 4$

$H(c) \approx 2{,}8$ gives us the

minimum average number of questions in

order to identify the outcome