

Interpretations of $H(X)$

- Measures the “amount of information” in X
 - Think of each value of X as a “message”
 - Think of X as a random experiment (20 questions)
- Minimum average number of bits to compress values of X
 - The more random X is, the harder to compress

A fair coin has the maximum information, and is hardest to compress
A biased coin has some information, and can be compressed to <1 bit on average
A completely biased coin has no information, and needs only 0 bit

"Information of x " = "#bits to code x " = $-\log p(x)$ $H(X) = E_p[-\log p(x)]$