UC Berkeley
Department of Electrical Engineering and Computer Sciences

EECS 126: Probability and Random Processes

Discussion 8
Fall 2023

1. **Information Loss**

Suppose we have discrete random variables X and Y, which represent the input message and received message respectively. Let n be the number of distinct values X can take. Our estimate of X from Y is $\hat{X} = g(Y)$, where g is some decoding function. Now define $E = \mathbb{1}\{X \neq \hat{X}\}$ to be the indicator of estimation error, and define the probability of error $p_e := \mathbb{P}(X \neq \hat{X})$.

a. Show that $H(\hat{X} \mid Y) = 0$.
b. Show that $H(E, X \mid \hat{X}) = H(X \mid \hat{X})$.
c. Show that $H(X \mid Y) \leq p_e \log_2(n-1) + H(E)$.
(You may use the fact that $H(X \mid Y) \leq H(X \mid \hat{X})$.)

Hint. The chain rule for entropy can be generalized to three random variables:

$$H(A, B \mid C) = H(A \mid C) + H(B \mid A, C).$$

Solution:

a. Intuitively, $\hat{X} = g(Y)$ is a function of Y, so observing Y allows us to determine $\hat{X}$ with no remaining uncertainty. Formally,

$$\begin{aligned} H(\hat{X} \mid Y) &= \sum_z \sum_y p_{\hat{X},Y}(z, y) \log \frac{1}{p_{\hat{X}|Y}(z \mid y)} \\ &= \sum_z \sum_y p(y)\, \mathbb{1}\{z = g(y)\} \log \frac{1}{\mathbb{1}\{z = g(y)\}} = 0. \end{aligned}$$

b. By the chain rule for entropy,

$$H(E, X \mid \hat{X}) = H(X \mid \hat{X}) + H(E \mid X, \hat{X}) = H(X \mid \hat{X}).$$

$H(E \mid X, \hat{X}) = 0$ by the same reasoning as in part a: E is a function of $X, \hat{X}$.

c. Note that $H(X \mid Y) \leq H(X \mid \hat{X}) = H(E, X \mid \hat{X})$ by part b. Now, by another application of the chain rule,

$$\begin{aligned} H(E, X \mid \hat{X}) &= H(E \mid \hat{X}) + H(X \mid E, \hat{X}) \\ &= H(E \mid \hat{X}) + (1 - p_e)\, H(X \mid E = 0, \hat{X}) + p_e\, H(X \mid E = 1, \hat{X}). \end{aligned}$$

- $H(E \mid \hat{X}) \leq H(E)$ by problem 1d.
- $H(X \mid E = 0, \hat{X}) = 0$, as $E = 0$ implies $X = \hat{X}$.
- $H(X \mid E = 1, \hat{X}) \leq \log_2(n-1)$, as $X \neq \hat{X}$ means that X can take on $n-1$ possible values, so its conditional entropy is at most $\log_2(n-1)$.

Putting it all together, we have that

$$H(X \mid Y) \leq H(E) + p_e \log_2(n-1).$$

2. **Hitting Time with Coins**

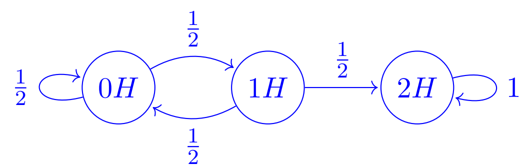
Consider a sequence of fair coin flips.

a. What is the expected number of flips until we first see two heads in a row?

b. What is the expected number of flips until we see a head followed immediately by a tail?

Solution:

a. We can create a Markov chain to compute the expected hitting time. $2H$ represents all sequences with HH as a subsequence, $1H$ all sequences that end in H but do not contain HH, and $0H$ all other sequences, including the initial empty sequence.

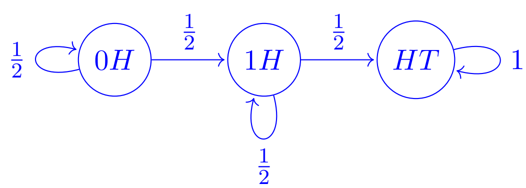


From here, we can set up our hitting-time equations, letting $\beta(i)$ denote the expected number of flips until two consecutive heads, given that we are in state i right now:

$$\begin{aligned}\beta(0H) &= 1 + \mathbb{P}(H)\cdot\beta(1H) + \mathbb{P}(T)\cdot\beta(0H)\\ &= 1 + \tfrac{1}{2}\beta(1H) + \tfrac{1}{2}\beta(0H)\\ \beta(1H) &= 1 + \mathbb{P}(H)\cdot\beta(2H) + \mathbb{P}(T)\cdot\beta(0H)\\ &= 1 + \tfrac{1}{2}\beta(2H) + \tfrac{1}{2}\beta(0H)\\ \beta(2H) &= 0.\end{aligned}$$

Solving this system of equations gives us $\beta(1H) = 4$ and $\beta(0H) = 6$. Thus, it takes 6 flips on average until we first see two heads in a row.

b. This part has a slightly different setup: if we flips heads after we just flipped a head, we do not need to reset to the initial state.



Letting $\beta(i)$ be the expected number of flips until we see HT, we have the equations

$$\begin{aligned}\beta(0H) &= 1 + \mathbb{P}(H)\cdot\beta(1H) + \mathbb{P}(T)\cdot\beta(0H)\\ &= 1 + \tfrac{1}{2}\beta(1H) + \tfrac{1}{2}\beta(0H)\\ \beta(1H) &= 1 + \mathbb{P}(H)\cdot\beta(HT) + \mathbb{P}(T)\cdot\beta(1H)\\ &= 1 + \tfrac{1}{2}\beta(HT) + \tfrac{1}{2}\beta(1H)\\ \beta(HT) &= 0.\end{aligned}$$

Solving this system gives $\beta(1H) = 2$ and $\beta(0H) = 4$.

3. **Before Absorption**

   Consider the Markov chain in Figure 1. Suppose that $X(0) = 1$. Calculate the expected number of times that the chain is in state 1 before being absorbed in state 3. ($X(0) = 1$ is included in this number.)

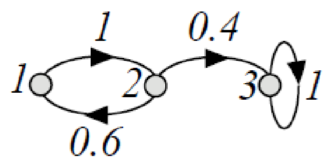


Figure 1: A Markov chain.

Solution: Let

$$\gamma(i) = \mathbb{E}\Big(\sum_{n=0}^{T_3} \mathbb{1}\{X(n) = 1\} \;\Big|\; X(0) = i\Big),$$

where and T_3 is the hitting time of state 3. We are interested in computing $\gamma(1)$. The first-step equations are:

$$\gamma(1) = 1 + \gamma(2)$$
$$\gamma(2) = 0.6\gamma(1)$$

Thus, $\gamma(1) = 1/0.4 = 2.5$.