LECTURE 3

Last time:

- Mutual Information.
- Convexity and concavity
- Jensen's inequality
- Information Inequality
- Data processing theorem

Lecture outline

- Fano's Inequality
- Stochastic processes, Entropy rate
- Markov chains
- Random walks on graphs
- Hidden Markov models

Reading: Chapter 4.

Quick Review

- Mutual Information:

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) \\ &= \sum_{x,y} P_{X,Y}(x,y) \log \frac{P_{X,Y}(x,y)}{P_X(x)P_Y(y)} \\ &= D(P_{X,Y}||P_X P_Y) \end{aligned}$$

- Chain Rule of Mutual Information.

$$I(X_1, X_2; Y) = I(X_1; Y) + I(X_2; Y|X_1)$$

- $D(p||q) \geq 0$.
- Entropy $H(X)$ is **concave** in P_X; Mutual information $I(X;Y)$ is **concave** in P_X for fixed $P_{Y|X}$, and **convex** in $P_{Y|X}$ for fixed P_X.
- $X \to Y \to Z \Rightarrow I(X;Y) \geq I(X;Z)$.

Fano's lemma

Suppose we have r.v.s X and Y, Fano's lemma bounds the error we expect when estimating X from Y

We generate an estimator of X that is $\widehat{X} = g(Y)$.

Probability of error $P_e = Pr(\widehat{X} \neq X)$

Indicator function for error $\mathbf{E}$ which is 0 when $X = \widehat{X}$ and 1 otherwise. Thus, $P_e = P(\mathbf{E} = 1)$

Fano's lemma:

$H(\mathbf{E}) + P_e \log(|\mathcal{X}| - 1) \geq H(X|Y)$

Proof of Fano's lemma

$$\begin{aligned}
H(\mathbf{E}, X|Y) &= H(X|Y) + H(\mathbf{E}|X, Y) \\
&= H(X|Y) \\
H(\mathbf{E}, X|Y) &= H(\mathbf{E}|Y) + H(X|\mathbf{E}, Y) \\
&\leq H(\mathbf{E}) + H(X|\mathbf{E}, Y) \\
&= H(\mathbf{E}) \\
&\quad +P_e H(X|\mathbf{E} = 1, Y) \\
&\quad +(1 - P_e) H(X|\mathbf{E} = 0, Y) \\
&= H(\mathbf{E}) + P_e H(X|\mathbf{E} = 1, Y) \\
&\leq H(\mathbf{E}) + P_e H(X|\mathbf{E} = 1) \\
&\leq H(\mathbf{E}) + P_e \log(|\mathcal{X}| - 1)
\end{aligned}$$

Works well (tight) when $|\mathcal{X}|$ is large.

Stochastic processes

- A stochastic process is an indexed sequence or r.v.s $X_0, X_1, \ldots$, a map from Ω to $\mathcal{X}^\infty$.
- A stochastic process is characterized by the joint PMF:

$$\begin{aligned} & P_{X_0,X_1,\ldots,X_n}(x_0, x_1, \ldots, x_n), \\ & \qquad (x_0, x_1, \ldots, x_n) \in \mathcal{X}^n, \text{ for } n = 0, 1, \ldots \end{aligned}$$

- The entropy of a stochastic process

$$\begin{aligned} & H(X_1, X_2, \ldots) \\ = \ & H(X_1) + H(X_2|X_1) + \ldots \\ & \quad + H(X_i|X_1, \ldots X_{i-1}) + \ldots \end{aligned}$$

Difficulties

- Sum to infinity
- all terms are different in general.

Entropy Rate

The entropy rate of a random process

$$\lim_{n\to\infty} \frac{1}{n} H(\underline{X}^n)$$

if it exists

Examples:

- i.i.d. sequence of r.v.s
- i.i.d. blocks of r.v.s

- A stochastic process is **stationary** if

$$P_{X_0,X_1,\ldots,X_n}(x_0,x_1,\ldots,x_n)$$

$$= P_{X_l,X_{l+1},\ldots,X_{l+n}}(x_0,x_1,\ldots,x_n)$$

  for every shift l and all $(x_0,x_1,\ldots,x_n) \in \mathcal{X}^n$.

For stationary processes, the limit exists.

Entropy Rate of Stationary Processes

Chain Rule:

$$\frac{1}{n}H(X_1, X_2, \ldots, X_n) \;=\; \frac{1}{n}\sum_{i=1}^{n} H(X_i|X_1, \ldots, X_{i-1})$$

The limit on LHS exists iff the individual terms on the RHS has a limit.

For a stationary process

$$\begin{aligned} H(X_{n+1}|X_1^n) &\leq H(X_{n+1}|X_2^n) \\ &= H(X_n|X_1^{n-1}) \end{aligned}$$

Therefore the sequence $H(X_n|X_1^{n-1})$ is non-increasing and non-negative, so limit exists.

Theorem For stationary processes, the entropy rate

$$\lim_{n\to\infty} \frac{1}{n}H(X_1^n) = \lim_{n\to\infty} H(X_n|X_1^{n-1})$$

Markov Chain

- A discrete stochastic process is a Markov chain if

$$\begin{aligned} & P_{X_n|X_0,\ldots,X_{n-1}}(x_n|x_0,\ldots,x_{n-1}) \\ = \; & P_{X_n|X_{n-1}}(x_n|x_{n-1}) \end{aligned}$$

  for $n = 1, 2, \ldots$ and all $(x_0, x_1, \ldots, x_n) \in \mathcal{X}^n$.

X_n: **state** after n transitions

- belongs to a finite set, e.g., $\{1, \ldots, m\}$
- X_0 is either given or random

Time Invariant Markov Processes

The transition probability is time-invariant.

$$
\begin{aligned}
p_{i,j} &= \mathbf{P}(X_{n+1} = j \mid X_n = i) \\
&= \mathbf{P}(X_{n+1} = j \mid X_n = i,\ X_{n-1}, \ldots, X_0)
\end{aligned}
$$

Markov chain is characterized by probability transition matrix $\underline{P} = [p_{i,j}]$

Question: Stationary vs. Time Invariant.

Let $r_i(n) = P(X_n = i)$, condition on an initial condition or average over random initial state,

Key recursion

$$r_j(n+1) = \sum_i r_i(n) p_{i,j}$$

or $\vec{r}(n+1) = \vec{r}(n)P$

Review of Markov chains

- Is there always a solution of $\pi = \pi P$, which is a probability vector?
- Is that solution unique?
- Starting from any initial state (or random), does the state distribution converge to π?

A Markov chain with a single class of recurrent aperiodic states, there is a unique stationary distribution π.

Each row of P^n converges to π.

The Entropy Rate of Markov Chain

$$\begin{aligned} & \lim_{n\to\infty} \frac{1}{n} H(X_1^n) \\ = & \lim_{n\to\infty} H(X_n|X_{n-1}) \\ = & -\sum_{i,j} \pi_i p_{i,j} \log p_{i,j} \end{aligned}$$

Random walk on graph

Example: Random walk on a 3 × 3 chessboard

1	2	3
4	5	6
7	8	9

$$p_{2,j} = \left[\frac{1}{5}, 0, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, 0, 0, 0\right]$$

Condition on $X_{n-1} = 2$, observing X_n gives $\log_2 5(bit)$ information.

Entropy rate $4\pi_1 \log 3 + 4\pi_2 \log 5 + \pi_5 \log 8$

For $n \times n$ chessboard with $n \to \infty$, entropy rate approaches $\log 8$.

Random walk on graph

Consider undirected graph $\mathcal{G} = (\mathcal{N}, \mathcal{E}, \mathcal{W})$ where $\mathcal{N}, \mathcal{E}, \mathcal{W}$ are the nodes, edges and weights. With each edge there is an associated edge $W_{i,j}$

$$\begin{aligned} W_{i,j} &= W_{j,i}, \\ W_i &= \sum_j W_{i,j} \\ W &= \sum_{i,j: j>i} W_{i,j} \\ 2W &= \sum_i W_i \end{aligned}$$

We call a random walk the Markov chain in which the states are the nodes of the graph

$$p_{i,j} = \frac{W_{i,j}}{W_i}$$

$$\pi_i = \frac{W_i}{2W}$$

Random Walk on Graph

Check: $\sum_i \pi_i = 1$ and

$$\begin{aligned}
\sum_i \pi_i p_{i,j} &= \sum_i \frac{W_i}{2W}\frac{W_{i,j}}{W_i} \\
&= \sum_i \frac{W_{i,j}}{2W} \\
&= \frac{W_j}{2W} \\
&= \pi_j
\end{aligned}$$

Back to the Example: Random walk on 3×3 chessboard, $W_{i,j} = 1$ for all connected i, j, $2W = 40$.

$$\begin{aligned}
\pi_1 &= \frac{3}{40} \\
\pi_2 &= \frac{5}{40}, \\
\pi_5 &= \frac{8}{40}
\end{aligned}$$

Random walk on graph

$$
\begin{aligned}
& H(X_2|X_1) \\
= & -\sum_i \pi_i \sum_j p_{i,j} \log(p_{i,j}) \\
= & -\sum_i \frac{W_i}{2W} \sum_j \frac{W_{i,j}}{W_i} \log\left(\frac{W_{i,j}}{W_i}\right) \\
= & -\sum_{i,j} \frac{W_{i,j}}{2W} \log\left(\frac{W_{i,j}}{W_i}\right) \\
= & -\sum_{i,j} \frac{W_{i,j}}{2W} \log\left(\frac{W_{i,j}}{2W}\right) \\
+ & \sum_{i,j} \frac{W_{i,j}}{2W} \log\left(\frac{W_i}{2W}\right) \\
= & -\sum_{i,j} \frac{W_{i,j}}{2W} \log\left(\frac{W_{i,j}}{2W}\right) + \sum_i \frac{W_i}{2W} \log\left(\frac{W_i}{2W}\right)
\end{aligned}
$$

Entropy rate is difference of two entropies

Hidden Markov models

Consider an ALOHA wireless model

$\mathcal{M}$ users sharing the same radio channel to transmit packets to a base station

During each time slot, a packet a arrives to a user's queue with probability p, independently of the other $\mathcal{M} - 1$ users

Also, at the beginning of each time slot, if a user has at least one packet in its queue, it will transmit a packet with probability q, independently of all other users

If two packets collide at the receiver, they are not successfully transmitted and remain in their respective queues

Hidden Markov models

Let $X_i = (n_1, n_2, \ldots, n_{\mathcal{M}})$ denote the random vector at time i where n_m is the number of packets that are in user m's queue. X_i is a Markov chain.

Consider the random vector $Y_i = (y_1, y_2, \ldots, y_{\mathcal{M}})$ where $y_i = 1$ if user i transmits during time slot i and $y_i = 0$ otherwise

Is Y_i Markov?

Hidden Markov processes

Let $\ldots, X_1, X_2, \ldots$ be a stationary Markov chain and let $Y_i = \phi(X_i)$ be a process, each term of which is a function of the corresponding state in the Markov chain

$\ldots, Y_1, Y_2, \ldots$ form a hidden Markov chain, which is not always a Markov chain, but is still stationary

What is its entropy rate?

We can compute $H(Y_n|Y_1^{n-1})$, which monotonically decreases with n.

Need a lower bound to the entropy rate.

Genie Trick

- Want to construct another sequence b_n, which is lower bound of the entropy rate $\lim_{n\to\infty} H(Y_n|Y_1^{n-1})$, yet has the same limit.
- Lower bound of entropy: give a genie. Genie information has to be small, but enough to flip the scale.
- Choose to look at $H(Y_n|Y_1^{n-1}, X_1)$.

Claim

$$H(Y_n|Y_1^{n-1}, X_1) \leq \lim_{n\to\infty} H(Y_n|Y_1^{n-1})$$

$$\begin{aligned} H(Y_n|Y_1^{n-1}, X_1) &= H(Y_n|Y_1^{n-1}, X_1, X_0, ..., X_{-k}) \\ &= H(Y_n|Y_1^{n-1}, X_{-k}^1, Y_{-k}^0) \\ &\leq H(Y_n|Y_{-k}^{n-1}) \end{aligned}$$

All the information about the history is captured in X_1.

Hidden Markov processes

Claim

$$\begin{aligned} & H(Y_n|Y_1^{n-1}) - H(Y_n|Y_1^{n-1}, X_1) \\ = \; & I(X_1; Y_n|Y_1^{n-1}) \to 0 \end{aligned}$$

Indeed,

$$\begin{aligned} \lim_{n\to\infty} I\left(X_1; Y_1^n\right) &= \lim_{n\to\infty} \sum_{i=1}^{n} I\left(X_1; Y_i|Y_1^{i-1}\right) \\ &= \sum_{i=1}^{\infty} I\left(X_1; Y_i|Y_1^{i-1}\right) \end{aligned}$$

since we have an infinite sum in which the terms are non-negative and which is upper bounded by $H(X_1)$, the terms must tend to 0