ECE 231A Discussion 2

TA: Hengjie Yang

Email: hengjie.yang@ucla.edu

04/10/2020

Preliminary: weak law of large numbers and convergence of r.v.'s

Weak law of large numbers (WLLN): For $n \in \mathbb{N}$, let $S_n = \sum_{i=1}^{n} X_i$ be the sum of n i.i.d. r.v.'s satisfying $\mathbb{E}[|X|] < \infty$. Then, $\forall \epsilon > 0$,

$$\lim_{n\to\infty} \Pr\left\{\left|\frac{S_n}{n} - \mathbb{E}[X]\right| > \epsilon\right\} = 0.$$

$\frac{S_n}{n} \to \mathbb{E}[X]$

Convergence of random variables: Given a sequence of r.v.'s $X_1, X_2, \ldots$, we say that the sequence $X_1, X_2, \ldots$ converge to a random variable X:

1. *In probability* if $\forall \epsilon > 0, \lim_{n\to\infty} \Pr\{|X_n - X| > \epsilon\} = 0$
2. *In mean-square* if $\lim_{n\to\infty} \mathbb{E}[|X_n - X|^2] = 0$ $\quad Z_n = |X_n - X|^2 \to 0$
3. *With probability* 1 *(also called almost surely)* if $\Pr\{\lim_{n\to\infty} X_n = X\} = 1$, (equivalently, $\forall \epsilon > 0$, $\lim_{m\to\infty} \Pr\{|X_n - X| > \epsilon, \forall n \geq m\} = 1$)

mean-square $\Rightarrow$ conv. in prob.

$\Pr\{w \in \Omega : \lim_{n\to\infty} X_n(w) = X(w)\} = 1$

Asymptotic equipartition property (AEP) and typical sets

AEP: If $X_1, X_2, \ldots$ are i.i.d. $\sim p(x)$, then $\forall \epsilon > 0$

$$\lim_{n\to\infty} \Pr\left\{ \left| -\frac{\log p(X_1, X_2, \ldots, X_n)}{n} - H(X) \right| > \epsilon \right\} = 0$$

$$= -\frac{\sum_{i=1}^{n} \log p(X_i)}{n} \to \mathbb{E}[-\log p(X)]$$

Proof: the weak law of large numbers (WLLN) on $-\frac{\log p(X_1, X_2, \ldots, X_n)}{n}$.

Typical sets: Given $n \in \mathbb{N}$, $\epsilon > 0$, the typical set $A_\epsilon^{(n)}$ w.r.t. $p(x)$ is defined by

$$A_\epsilon^{(n)} = \left\{ (x_1, x_2, \ldots, x_n) \in \mathcal{X}^n : \left| -\frac{\log p(x_1, x_2, \ldots, x_n)}{n} - H(X) \right| < \epsilon \right\}.$$

$$\lim_{n\to\infty} \Pr\{A_\varepsilon^{(n)}\} = 1$$

Properties of typical sets

Theorem (Properties of typical sets)

(i) *If* $(x_1, \ldots, x_n) \in A_\epsilon^{(n)}$, $2^{-n(H(X)+\epsilon)} \leq p(x_1, x_2, \ldots, x_n) \leq 2^{-n(H(X)-\epsilon)}$ *and* $H(X) - \epsilon \leq -\frac{\log p(x_1,\ldots,x_n)}{n} \leq H(X) + \epsilon$.

(ii) $\Pr\{A_\epsilon^{(n)}\} > 1 - \epsilon$ *for* n *sufficiently large.*

(iii) $|A_\epsilon^{(n)}| \leq 2^{n(H(X)+\epsilon)}$

(iv) $|A_\epsilon^{(n)}| \geq (1-\epsilon)2^{n(H(X)-\epsilon)}$ *for* n *sufficiently large.*

Proof:

(i) apply the definition of $A_\epsilon^{(n)}$.

(ii) apply the definition of limit on $\Pr\{A_\epsilon^{(n)}\}$.

(iii) apply $1 \geq \sum_{x^n \in A_\epsilon^{(n)}} p(x^n)$ and property (i).

(iv) apply property (ii) and (i).

$$1 \geq \sum_{x^n \in A_\varepsilon^{(n)}} p(x^n) \geq \sum_{x^n \in A_\varepsilon^{(n)}} 2^{-n(H(X)+\varepsilon)} = |A_\varepsilon^{(n)}| \, 2^{-n(H(X)+\varepsilon)}$$

Stationary process, Markov process and entropy rate

Stationary process: A random process $\{X_i\}_{i\in\mathbb{N}}$ is stationary if $p(x_1,\ldots,x_k) = p(x_{1+\tau},\ldots,x_{k+\tau})$, for $\forall k$, $\forall \tau$, and $\forall x_1,\ldots,x_k \in \mathcal{X}$.

$p(X_k \mid X_{k-1} \cdots X_1) = p(X_{k+\tau} \mid X_{k+\tau-1} \cdots X_{1+\tau})$

Corollary: $H(X_1,\ldots,X_k) = H(X_{1+\tau},\ldots,X_{k+\tau})$ for $\forall k$, $\forall \tau$, and $\forall x_1,\ldots,x_k \in \mathcal{X}$.

Markov process (Markov chain): a random process $\{X_i\}_{i\in\mathbb{N}}$ is a Markov process if

$$p(X_{n+1}|X_n,\ldots,X_1) = p(X_{n+1}|X_n), \quad \forall n,\ \forall X_1,\ldots,X_{n+1}$$

Entropy rate: the entropy of a stochastic process $\{X_i\}_{i\in\mathbb{N}}$ is given by

$$H(\mathcal{X}) \triangleq \lim_{n\to\infty} \frac{1}{n} H(X_1, X_2, \ldots, X_n)$$

if the limit exists.

Examples:

- If $\{X_i\}_{i\in\mathbb{N}}$ is an i.i.d. process, $H(\mathcal{X}) = \lim_{n\to\infty} \frac{nH(X_1)}{n} = H(X_1)$.
- If $\{X_i\}_{i\in\mathbb{N}}$ is a sequence of independent, but not identically distributed r.v.'s, since each $H(X_i)$ is distinct, the limit of $\frac{1}{n}\sum_i H(X_i)$ may not exist.

Entropy rate for stationary schocastic process

Theorem

If $\{X_i\}$ is a stationary process, then the entropy rate $H(\mathcal{X})$ exists and

$$H(\mathcal{X}) = \lim_{n\to\infty} \frac{1}{n} H(X_1, \ldots, X_n) = \underbrace{\lim_{n\to\infty} H(X_n | X_{n-1}, X_{n-2}, \ldots, X_1)}_{\triangleq H'(\mathcal{X})}$$

Proof:

$\{b_n\}_{n\in\mathbb{N}}$ $b_n \to b$ $a_n = \frac{1}{n}\sum_{i=1}^{n} b_n \to b$

1. Cesáro mean
2. $0 \le H(X_n|X^{n-1}) \le H(X_i|X^{i-1})$, $1 \le i \le n$, for stationary process.

$$H(X_n|X^{n-1}) = H(X_n | X_{n-1} \cdots \cancel{X_1})$$

$X^{n-1} \Downarrow (X_{n-1}, \cdots X_1)$

$$\le H(X_n | X_{n-1} \cdots X_2)$$

conditioning reduces entropy

$$= H(X_{n-1} | X_{n-2} \cdots X_1)$$

$H(X_2|X_1) \le H(X_2) = H(X_1)$

$$H(X_1) \ge H(X_2|X_1) \ge H(X_3|X^2) \cdots \ge H(X_n|X^{n-1}) \cdots \ge 0$$

Entropy rate for stationary Markov chains

Theorem: For a stationary Markov chain $\{X_i\}_{i\in\mathbb{N}}$, $H(\mathcal{X}) = H(X_2|X_1)$.
Proof:

$$\begin{aligned} H(\mathcal{X}) &= \lim_{n\to\infty} H(X_n|X_{n-1},\dots,X_1) \\ &= \lim_{n\to\infty} H(X_n|X_{n-1}) \\ &= \lim_{n\to\infty} H(X_2|X_1) = H(X_2|X_1) \end{aligned}$$

Stationary distribution for Markov chains: Let $P = [P_{i,j}]$, $i,j\in\{1,\dots,m\}$, denote the transition matrix and $\mu^T\in\mathbb{R}^m$. Thus, solving

$X_0 \sim \mu$ $\quad X_1 \sim p(X_1|X_0)$, $\sim \mu$

$$\mu P = \mu$$

$\mu = \{\mu_1 \cdots \mu_m\}$ $\quad \mu$

gives the stationary distribution.

$H(X_2|X_1) = \sum_x p(X_1) \cdot \underbrace{H(X_2|X_1 = x)}_{P}$

Corollary: For a stationary Markov chain with m states,

$$H(\mathcal{X}) = H(X_2|X_1) = \sum_{i=1}^m \left[\mu_i \left(-\sum_{j=1}^m P_{i,j}\log(P_{i,j})\right)\right]$$

Remark: If a Markov chain is irreducible and aperiodic, then the stationary distribution μ is unique. Moreover, $\lim_{n\to\infty}\mu_n = \mu$, $\forall \mu_0 \in \Delta_{n-1}$.

Exercise 1: AEP

Let $X_1, X_2, \ldots$ be i.i.d. r.v.'s drawn according to the PMF $p(x), x \in \{1, 2, \ldots, m\}$. Thus, $p(x_1, \ldots, x_n) = \prod_{i=1}^{n} p(x_i)$. We know that $-\frac{1}{n} p(x_1, \ldots, x_n) \to H(X)$ in probability. Let $q(x_1, \ldots, x_n) = \prod_{i=1}^{n} q(x_i)$, where q is another PMF on $\{x_1, \ldots, x_n\}$.

$q(x_1)\ q(x_2) \cdots q(x_n)$ iid

(i) Evaluate $\lim_{n\to\infty} -\frac{1}{n} \log q(X_1, \ldots, X_n)$, where $X_1, X_2, \ldots$ are i.i.d. $\sim p(x)$.

(ii) Now evaluate the limit of the log likelihood ratio $\frac{1}{n} \log \frac{q(X_1,\ldots,X_n)}{p(X_1,\ldots,X_n)}$, when $X_1, X_2, \ldots$ are i.i.d. $\sim p(x)$.

(i) $$-\frac{1}{n}\log\left(\frac{q(x^n)}{p(x^n)}\cdot p(x^n)\right)$$

$$= -\frac{1}{n}\log\left(\frac{q(x^n)}{p(x^n)}\right) - \frac{1}{n}\log p(x^n)$$

$$\Rightarrow -E_p\left[\log\frac{q(x)}{p(x)}\right] - E_p[\log p(x)]$$

$$= D(p\|q) + H(p)$$

(ii) $$\frac{1}{n}\log\frac{q(x^n)}{p(x^n)}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\log\frac{q(x_i)}{p(x_i)}$$

$$\to E_p\left[\log\frac{q(x)}{p(x)}\right]$$

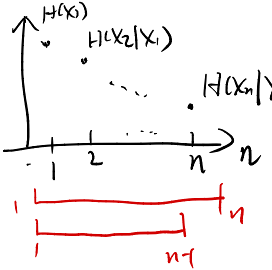
$$= -D(p\|q)$$

Exercise 2: Monotonicity of entropy per element

For a stationary stochastic process $X_1, X_2, \ldots, X_n$, show that

$$\frac{H(X_1, X_2, \ldots, X_n)}{n} \leq \frac{H(X_1, X_2, \ldots, X_{n-1})}{n-1}.$$

Hint: For a stationary process, $H(X_n|X^{n-1}) \leq H(X_i|X^{i-1}), 1 \leq i \leq n$.



$$\frac{H(X_1 \cdots X_n)}{n} = \frac{\sum_{i=1}^{n} H(X_i|X^{i-1})}{n} \quad (1)$$

$$= \frac{H(X_n|X^{n-1}) + \sum_{i=1}^{n-1} H(X_i|X^{i-1})}{n}$$

$$H(X_n|X^{n-1}) \leq H(X_{n-1}|X^{n-2}) \leq H(X_{n-2}|X^{n-1}) \cdots \leq H(X_1)$$

$$\Downarrow$$

$$H(X_n|X^{n-1}) \leq \frac{1}{n-1} \sum_{i=1}^{n} H(X_i|X^{i-1}) \quad (2)$$

Apply (2) to (1)

$$\leq \frac{\frac{1}{n-1} H(X_1 \cdots X_{n-1}) + H(X_1 \cdots X_{n-1})}{n}$$

$$= \frac{H(X_1 \cdots X_{n-1})}{n-1}$$