Lecture 3

- Differential Entropy
- Asymptotic Equipartition Property (AEP)
- Source Coding

Differential Entropy

- For continuous r.v. X with pdf f(x)

$$h(X) = -\int f(x) \log f(x) dx$$

- Properties:

$h(X)$ not necessarily non-negative

Translation: $h(X+a) = h(X)$

Scaling: $h(aX) = h(X) + \log |a|$

Conditioning: $h(X|Y) \leq h(X)$

Mutual Inf: $I(X;Y) = h(X) - h(X|Y)$

Asymptotic Equipartition Property

$X_1, X_2, \ldots$ are iid ~ p(x)

AEP: $-\frac{1}{n} \log p(X_1, \ldots, X_n) \to H(X)$

Typical set $A_\varepsilon^{(n)}$: sequences $(x_1, \ldots, x_n)$ satisfying

$2^{-n(H(X)+\varepsilon)} \leq p(x_1, \ldots, x_n) \leq 2^{-n(H(X)-\varepsilon)}$

- As n gets large, almost all sequences are typical
- Roughly $2^{nH(x)}$ typical sequences, each with probability roughly equal to $2^{-nH(x)}$

Properties of Typical Set

Typical set $A_\varepsilon^{(n)}$: sequences $(x_1, \ldots, x_n)$ satisfying

$2^{-n(H(X)+\varepsilon)} \leq p(x_1, \ldots, x_n) \leq 2^{-n(H(X)-\varepsilon)}$

Properties:

1. $P(A_\varepsilon^{(n)}) > 1 - \varepsilon$ for sufficiently large n
2. $|A_\varepsilon^{(n)}| \leq 2^{n(H(X)+\varepsilon)}$
3. $|A_\varepsilon^{(n)}| \geq (1-\varepsilon) 2^{n(H(X)-\varepsilon)}$

AEP Intuition

- AEP <u>does not</u> imply sequence $x_1, x_2, \ldots x_n$ converges, but implies probability of sequence (i.e. $p(x_1, x_2, \ldots x_n)$) converges to $2^{-nH(X)}$
- X_i ~ Bern(1/4)
- Typical set is sequences which have roughly ¼ one's, ¾ zero's
- (0,0,…) is the <u>most likely</u> sequence (Prob = $(3/4)^n$), but is <u>not in</u> typical set
- Prob(typical set) -> 1 as n -> infinity, so sufficient to only consider typical sequences

Source Coding

- Thm: Let X ~ p(x) be a finite alphabet source. If R > H(X), then there exist fixed length block codes (n, 2^{nR}) with P(e) -> 0 as n -> 0. Conversely, any sequence of codes (n, 2^{nR}) with P(e) -> 0 satisfies R >= H(X).
- Achievability: Uniquely code typical sequences, assign all atypical sequences to a single codeword
- Converse: Uses Fano's inequality

Block Codes

- Encoder: Assigns length k (normally 2^{nR}) codeword to each length n vector of source symbols
- Decoder: Maps from length k code to length n vector
- Prob. Error: probability decoded vector not equal to input vector

Fano's Inequality

$$(X,Y) \sim p(x,y), \quad P_e = \Pr(X \neq Y)$$

$$\text{Thm}: \quad H(X|Y) \leq H(P_e) + P_e(\log|X|-1)$$
$$\leq 1 + P_e(\log|X|-1)$$

- If P_e=0, then need H(X|Y)=0
- Fano's proves that if P_e small (but not necessarily zero), then H(X|Y) must be small
- Used in almost every converse: X message, Y is output of decoder

Joint Typicality

$(X, Y) \sim p(x, y)$

$$\text{AEP:}\quad -\frac{1}{n}\log p((X_1, Y_1), \ldots, (X_n, Y_n)) \to H(X, Y)$$

Typical set $A_\varepsilon^{(n)}$: sequences $((x_1, y_1), \ldots, (x_n, y_n))$:

$$2^{-n(H(X)+\varepsilon)} \le p(x_1, \ldots, x_n) \le 2^{-n(H(X)-\varepsilon)}$$

$$2^{-n(H(Y)+\varepsilon)} \le p(y_1, \ldots, y_n) \le 2^{-n(H(Y)-\varepsilon)}$$

$$2^{-n(H(X,Y)+\varepsilon)} \le p((x_1, y_1), \ldots, (x_n, y_n)) \le 2^{-n(H(X,Y)-\varepsilon)}$$

Properties

1. $\mathbf{P((X^n, Y^n) \in A_\varepsilon^{(n)}) \to 1}$ **as** $\mathbf{n \to \infty}$

2. $|\mathbf{A}_\varepsilon^{(n)}| \le 2^{n(H(X,Y)+\varepsilon)}$

3. If $(\tilde{X}^n, \tilde{Y}^n) \sim p(x^n)p(y^n)$, **then**

$$\Pr((\tilde{X}^n, \tilde{Y}^n) \in A_\varepsilon^{(n)}) \le 2^{-n(I(X;Y)-3\varepsilon)}$$

$$\Pr((\tilde{X}^n, \tilde{Y}^n) \in A_\varepsilon^{(n)}) \ge (1-\varepsilon)2^{-n(I(X;Y)+3\varepsilon)}$$

4. $x^n \in A_\varepsilon^{(n)}, \quad A_\varepsilon^{(n)}(Y^n, x^n) = \{y^n : (x^n, y^n) \in A_\varepsilon^{(n)}\}$

$$|A_\varepsilon^{(n)}(Y^n, x^n)| \le 2^{n(H(Y|X)+2\varepsilon)}$$