STO6 LECTURE 16

Note Title 4/11/2006

Today

Dif. Entropy (Contd.)

- Joint Entropy, Conditional Entropy,
- Mutual Information
- AEP
- Channel "capacity".

Last time

X with pdf $f = f_X$

Differential Entropy of X

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

Some examples

- Given $X \in [a,b]$ $h(X)$ maximized if $f(X) =$ uniform on $[a,b]$.

- Given X of mean 0, deviation 1,

  $h(x)$ maximized if $f(x) = N(0,1)$.

Proofs omitted

Entropy of collection of random variables $X_1 \ldots X_n$

$$h(X_1 \ldots X_n) = -\int_{x_1 \ldots x_n} f(x_1 \ldots x_n) \log f(x_1 \ldots x_n).$$

Ex

$X_1, \dots X_n$ i.i.d $\sim X$

$\Rightarrow \quad h(X_1, \dots X_n) = n \cdot h(X)$.

Example:

$\underline{X}$ i.i.d $\sim N(0, 1)$

$\underline{Y} = A \underline{X}$ for A known, invertible

then $Y \sim N(0, K_Y = AA^T)$

$$h(Y) = \frac{1}{2} \log\left[(2\pi e)^n \cdot |K_Y|\right]$$

"Entropy depends on coordinate system".

- Conditional entropy $h(X|Y) = -\int_y f(y) \int_x f(x|y) \log f(x|y) \, dx \, dy$

$$= \int_y \int_x f(x,y) \log f(x|y) \, dx \, dy$$

- Chain Rule: $h(X,Y) = h(X) + h(Y|X)$.
- Jensen $\Rightarrow$ $h(Y|X) \leq h(Y)$
- Mutual Information: $I(X;Y) = h(X) - h(X|Y)$

$$= h(Y) - h(Y|X).$$

- $I(X;Y) \geq 0$
- $D(f \| g) = \int f \log \frac{f}{g} \, dx \geq 0.$

Questions :

- $h(X) \geq 0$?
- $h(X+a) = h(X)$?
- $h(aX) = h(X)$?

$Y = g(X)$

$$\Rightarrow h(Y) = h(X) - E\left[\log\left|\frac{dx}{dy}\right|\right]$$

Example : Uniform dist max. diff. entropy.

$X \in [0,1]$

$Y \sim U(0,1)$

$Z = (X+Y) \pmod 1$; $Z \sim U(0,1)$

$h(Z) \leq h(Y)$

$h(Y,Z) = h(X,Y) = h(X) + h(Y)$

$h(Y,Z) \leq h(Y) + h(Z)$

$\Rightarrow h(X) \leq \underbrace{h(Z)} = h(U(0,1))$.

AEP

$X_1 \ldots X_n$ i.i.d. $\sim X$

Thm $-\frac{1}{n}\log f(X_1 \ldots X_n) \longrightarrow h(X)$

in probability.

Furthermore

if we define

$$A_n^{(\epsilon)} = \left\{ (x_1 \ldots x_n) \text{ s.t } \left| -\frac{1}{n}\log - \quad - h(x) \right| \leq \epsilon \right\}$$

& $\text{vol}(S) = \int_{x_1 \ldots x_n} 1_S \, dx_1 \ldots dx_n$

Then

Th. ① $\Pr\left[A_n^{(\epsilon)}\right] \longrightarrow 1$

② $\mathrm{Vol}\left(A_n^{(\epsilon)}\right) \leq 2^{(h(X)+\epsilon)n}$

③ $\mathrm{Vol}\left(A_n^{(\epsilon)}\right) \geq (1-\epsilon)\, 2^{(h(X)-\epsilon)n}$

Proof: ① LLN

② $$1 = \int f(x_1 \dots x_n)\, dx_1 \dots dx_n$$

$$\geq \int 1_A \cdot f(x_1 \dots x_n)\, dx_1 \dots dx_n$$

$$\geq \int 1_A \cdot 2^{-(h(x)+\epsilon)n}\, dx_1 \dots dx_n$$

$$= 2^{-(h(x)+\epsilon)n} \cdot \mathrm{Vol}\left(A_n^{(\epsilon)}\right)$$

③ $\Pr\left[A_n^{(\epsilon)}\right] = \int 1_A \cdot f(x_1 \dots x_n)\, dx_1 dx_n$

$$\leq \int 1_A \cdot 2^{-(h(x)-\epsilon)n} \, dx_1 \ldots dx_n$$

$$= 2^{-(h(x)-\epsilon)n} \cdot \mathrm{Vol}(A_n^{(\epsilon)}).$$

AEP : most compelling reason to define diff. entropy the way it is defined.

- Also explains why $h(x)$ being neg is unimportant

- Only relevance "$h(X)$ vs $h(Y)$".

Chennel Capacity

$$I(X;Y) = \max_{f_X} \{ h(X) - h(X|Y) \}$$

Implication

$$X \in [-1, 1]$$

Noise $\quad W \in_U [-\epsilon, \epsilon]$

$$Y = X + W$$

$$I(X;Y) = \max_{f_X} \{ h(Y) - h(Y|X) \}$$

$$= \max_{f_X} \{ h(Y) - h(W|X) \}$$

$$= \max_{f_X} \{h(Y)\} - h(W)$$

$$\le \log(2(1+\epsilon)) - \log 2\epsilon$$

$$\le \log\left(\frac{1}{\epsilon} + 1\right)$$

Question : is Channel Capacity a lower bound? is it an upper bound?

Answer : True ; but won't prove in general.

Only for simple $[-1, 1] \leftarrow$ input

$+ \; [-\epsilon, \epsilon] \leftarrow$ noise $\Big\}$ 6.441 Chan

channel

& for AWGN channel.

For 6.441 channel

$\log \frac{1}{\epsilon}$ lower bound on capacity

Upper bound ?

Fano's Inequality

$$H(X^n | Y^n) \leq 1 + P_e \cdot n \cdot R$$

$$\Rightarrow \quad I(X^n; Y^n) \geq H(X^n) - H(X^n | Y^n)$$

$$\geq nR(1 - P_e) - 1$$

But $I(X^n;Y^n) \leq n \cdot \text{Capacity}$

$\Rightarrow R \leq \text{Capacity}$ $\boxtimes$

<u>Sequence</u>

1. Recall $h(x)$
2. Introduce $h(X,Y)$, $D(f\|g)$, $h(X|Y)$, $I(X;Y)$

   AEP

3. <u>Unusual Properties</u>

   ① $h(x) < 0$!

   ② $h(ax)$!

③ $h(A \cdot \underline{X})$

④ h maximizes

⑤ $I(X;Y)$ discrete vs. continuous.

4. Capacity of Channel

(6.441 Channel ...)