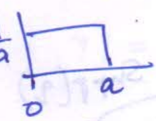
Lec 21 : Relative entropy and mutual information (Continuous RV Case)

Recap

1) $h(X) = -\int_S f(x) \log f(x)\, dx$

Eg: $\frac{1}{a}$  $\log a$

Eg:  $\frac{1}{2} \log 2\pi e \sigma^2$

$N(0, \sigma^2)$

2) $A_\epsilon^{(n)} \Rightarrow \left| h(X) - \log \frac{1}{f(\underline{x})} \right| < \epsilon$

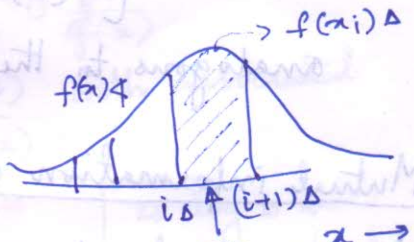
$P_r\left(\underline{x} \in A_\epsilon^{(n)} \right) \geq 1-\epsilon$ for n sufficiently large

$(1-\epsilon)\, 2^{n(h(X)-\epsilon)} \leq Vol(A_\epsilon^{(n)}) \leq 2^{n(h(X)+\epsilon)}$

3) Quantization

$X^\Delta \triangleq x_i$ with probability $P_i = f(x_i)\Delta$

$H(X^\Delta) + \log \Delta \xrightarrow[\Delta \to 0]{} h(X)$



$\Rightarrow$ If. $\Delta = 2^{-n}$

$H(X^\Delta) \longrightarrow \infty$ as $n \longrightarrow \infty$!!

4) Joint and Conditional differential entropy

$\underline{X} = (X_1, X_2, \ldots X_n)$

$h(\underline{X}) \triangleq \int f(\underline{x}) \log \frac{1}{f(\underline{x})}\, d\underline{x}$

$h(X|Y) \triangleq \int f(x,y) \log \frac{1}{f(x|y)}\, dx\, dy$

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

⑤ $\underline{X} \sim N(t, K)$

$$h(\underline{X}) = \frac{1}{2}\log(2\pi e)^n |K|$$

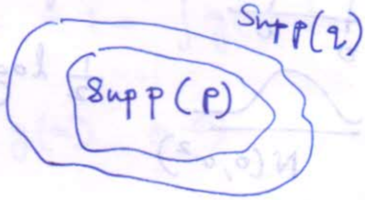
Relative entropy

$$D(p(x) \| q(x)) = D(P\|q)$$

$$= \int p(x) \log \frac{p(x)}{q(x)}\, dx \longrightarrow \text{density function}$$

If $q(x) = 0$, $p(x) \neq 0$

$\Rightarrow \infty$



Mutual Information

$$I(X;Y) = D\left(f(x,y) \,\|\, f(x) f(y)\right)$$

$$I(X;Y) = \begin{cases} h(X) - h(X|Y) \\ h(Y) - h(Y|X) \\ h(X) + h(Y) - h(X,Y) \end{cases}$$

analogous to the discrete case.

Mutual information and Quantization

$$H(X^\Delta) + \log \Delta \longrightarrow h(X)$$

$$H(Y^\Delta) + \log \Delta \longrightarrow h(Y)$$

$$H(X^\Delta, Y^\Delta) + \log \Delta^2 \longrightarrow h(X,Y)$$

$$\therefore\ I(X^\Delta; Y^\Delta) = H(X^\Delta) + H(Y^\Delta) - H(X^\Delta, Y^\Delta)$$

$$\longrightarrow I(X;Y)$$

Mutual information

Correlated Guassian RVs

Let $(X, Y) \sim \mathcal{N}(0, K)$

$$K = \begin{bmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{bmatrix} \qquad \rho = \text{correlation coefficient}$$

$$I(X;Y) = h(X) + h(Y) - h(X,Y)$$

$$= \frac{1}{2}\log 2\pi e\sigma^2 + \frac{1}{2}\log 2\pi e\sigma^2 - \frac{1}{2}\log (2\pi e)^2(\sigma^4 - \rho^2\sigma^4)$$

$$= \frac{1}{2}\log (2\pi e)^2\sigma^4 - \frac{1}{2}\log(2\pi e)^2\sigma^4(1-\rho^2)$$

$$= \frac{1}{2}\log\frac{1}{1-\rho^2} = \begin{cases} 0 & \rho = 0 \\ \infty & \rho \in \{\pm 1\} \end{cases}$$

Jensen's Inequality Application

<u>Claim</u>: $D(f\|g) \geq 0$

<u>Proof</u>: $D(f\|g) \triangleq \int f(x)\log\frac{f(x)}{g(x)}\,dx$

(let Y be RV taking on value $\frac{g(x)}{f(x)}$ with associated density $f(x)$)

Jensen's inequality continuos case.

$$-D(f\|g) = E[\log Y] \leq \log EY \qquad \left[\text{as log is a concave function}\right]$$

$$= \log\left\{\int f(x)\frac{g(x)}{f(x)}\,dx\right\}$$

$$= \log 1 = 0$$

$$\Rightarrow D(f\|g) \geq 0.$$

J.I also says that when $h(\cdot)$ is convex, in the inequality

$$E[h(X)] \geq h(E[X])$$ with equality iff

X is a constant a.e.

$\therefore$ $D(f\|g) = 0$ iff $\frac{f(x)}{g(x)} = \text{constant} = a$ a.e

By integrating $\int f(x)\,dx = 1 = \int g(x)\,dx$

$\Rightarrow a = 1 \quad \therefore f(x) = g(x)$ (a.e)

(a.e) $\Longleftrightarrow$ except for a set of measure zero, we will ignore this)

Hence $I(X;Y) = D(f(x,y) \| f(x) f(y)) \geq 0$

with equality iff $f(x,y) = f(x) f(y)$ a.e i.e.,

X, Y are independant

Hence $h(X|Y) = h(X) - I(X;Y)$
$\leq h(X)$

Hence
$$h(X_1, \ldots X_n) = \sum_{i=1}^{n} h(X_i | X_{[i-1]}) \leq \sum_{i=1}^{n} h(X_i)$$

with equality iff $\{X_1, \ldots X_n\}$ are independant

An unexpected fallout

(Hadamard inequality)

$$K = \begin{bmatrix} K_{11} & K_{12} & K_{13} \\ K_{21} & K_{22} & K_{23} \\ K_{31} & K_{32} & K_{33} \end{bmatrix}$$

let $\underline{X} \sim \eta(0, K)$

$h(\underline{X}) = \frac{1}{2} \log (2\pi e)^n |K|$

$h(X_i) = \frac{1}{2} \log (2\pi e) K_{ii}$

$\rightarrow$ Applying $h(\underline{X}) \leq \sum_{i=1}^{n} h(X_i) \Rightarrow$

For any K positive semidefinite

$$\boxed{|K| \leq \prod_{i=1}^{n} K_{ii}}$$

Differential entropy under translation

Claim: $h(X+c) = h(X)$ [c is some constant]

Pf: Let $Y = X + c$.

$$f_Y(y) = f_X(y-c)$$

$$h(Y) = \int f_Y(y) \log \frac{1}{f_Y(y)}\, dy$$

$$= \int f_X(y-c) \log \frac{1}{f_X(y-c)}\, dy$$

$$= h(X) \quad \text{[after change of variables]}$$

Differential entropy under scaling

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

let $Y = aX$

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y}{a}\right)$$

$$\therefore h(Y) = \int_y \frac{1}{|a|} f_X\left(\frac{y}{a}\right) \log \frac{|a|}{f_X\left(\frac{y}{a}\right)}\, dy$$

$$= \int \frac{1}{|a|} f_X(x) \log \frac{|a|}{f_X(x)}\, |a|\, dx$$

$$= \int f_X(x) \log \frac{1}{f_X(x)}\, dx + \int \log|a|\, f_X(x)\, dx$$

$$= h(X) + \log|a|.$$

Analogously if $\underset{n\times 1}{\underline{Y}} = A \underset{n\times 1}{\underline{X}}$, A non singular, then

$$h(\underline{Y}) = h(\underline{X}) + \log(|A|)$$

$$I(X;Y) = \sup_{P,Q} I([X]_P ; [Y]_Q)$$

where P, Q is a partition

Consider P', Q' a refinement of partition P, Q

Then it can be shown that

$$I([X]_{P'} ; [Y]_{Q'}) \geq I([X]_P ; [Y]_Q)$$

$\underline{X}$ zero mean, $E[\underset{n\times 1}{\underline{X}}\;\underset{1\times n}{\underline{X}^T}] = K$

$$Y \sim \eta(\underline{0}, \underline{K}) \rightarrow E[X_i X_j] = E[Y_i Y_j]$$

$$\boxed{h(X) \leq h(Y)} \quad \text{--- ①}$$

Pf of ①: Consider $D(f\|g) = \int_x f(x) \log \frac{f(x)}{g(x)}\, dx$

$X \Leftrightarrow f(x)$

$Y \Leftrightarrow g(x) = \frac{1}{(2\pi)^{n/2}|K|^{1/2}} e^{-\frac{x^T K^{-1} x}{2}}$

$$= -h(X) - \int_x f(x) \log g(x)\, dx$$

$$= -h(X) + \int f(x)\left[\log (2\pi)^{n/2}|K|^{1/2} + \cancel{\log}\frac{x^T K^{-1} x}{2}\right] dx$$

$$= -h(X) + \frac{1}{2}\log(2\pi)^n |K| + \int \log_e f(x)\, \frac{x^T K^{-1} x}{2}\, dx$$

$$= -h(X) + h(Y)$$

$\Rightarrow h(Y) \geq h(X)$

Equality holds iff $\underline{X} \sim \eta(0, K)$

$$\int f(x) \log \frac{1}{g(x)} \Leftarrow \begin{cases} \frac{\log e}{2} E_f[x^T K^{-1} x] \\ = \frac{\log e}{2} E_g[x^T K^{-1} x] \end{cases}$$

$$= \int g(x) \log \frac{1}{g(x)}$$

$\Downarrow$ as this involves only 2nd moments.