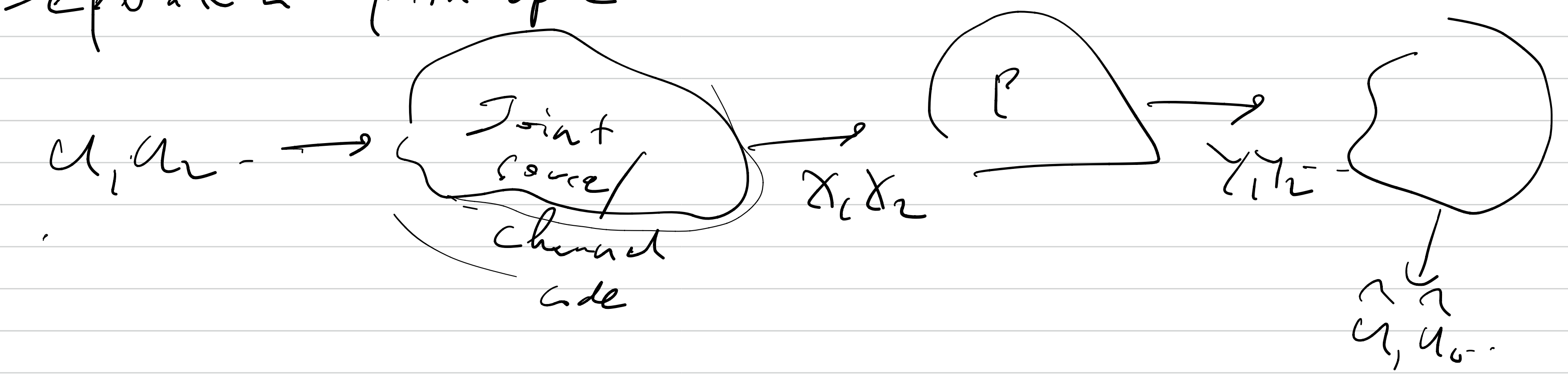
Information Theory & Coding

Nov 3rd, 2020

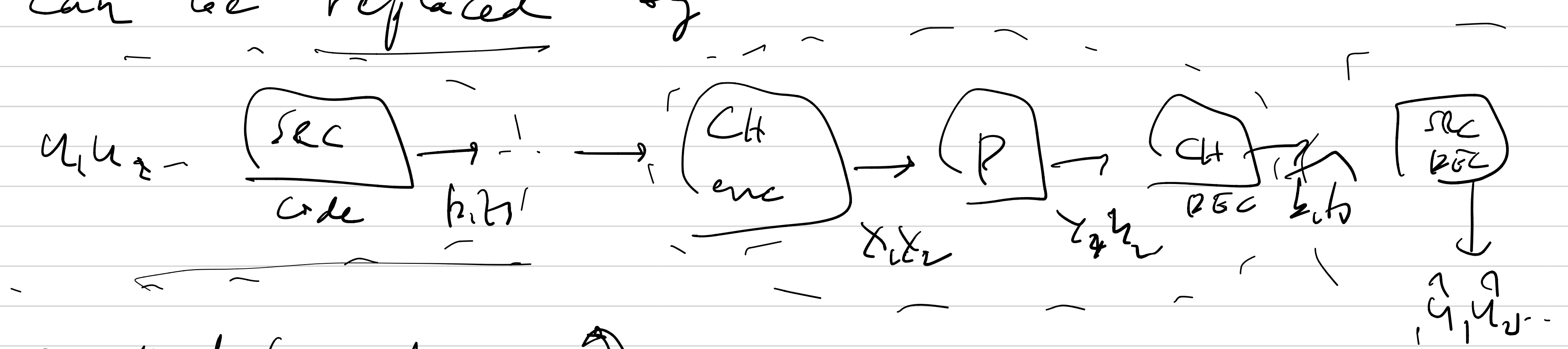
Yesterday: Channel capacity with cost constraints.

$$C(P,\beta) = \max_{\substack{P_X :\\ E[b(X)] \le \beta}} I(X;Y).$$

– Separation principle:



can be replaced by



a modular design.

Differential Entropy.

Remember for discrete valued X: $\quad H(X) = \sum_{x \in \mathcal{X}} p(x) \log_2 \frac{1}{p(x)} = E\left[\log \frac{1}{p(X)}\right].$

if X is $\mathbb{R}$-valued with a probability density function $f_X(x)$, a natural definition for entropy would be

$$h(X) = \int f_X(x) \log \frac{1}{f_X(x)}\, dx. \quad : \text{differential entropy.}$$

if $X_1, \dots, X_n$ are $\mathbb{R}$-valued RVs. with joint density

$f_{X^n}(x_1 \dots x_n)$, we define

$$h(X_1 \dots X_n) = \int \cdots \int f_{X^n}(x_1 \dots x_n) \log \frac{1}{f_{X^n}(x_1 \dots x_n)} \, dx_1 \dots dx_n$$

$$= E\left[\log \frac{1}{f_{X^n}(X^n)}\right]$$

similarly if X is $\mathbb{R}^n$ valued and Y is $\mathbb{R}^m$ valued

with joint density $f_{XY}(x,y) = f_X(x) f_{Y|X}(y|x)$

$$= f_Y(y) f_{X|Y}(x|y)$$

then $$h(X|Y) = E \log \frac{1}{f_{X|Y}(X|Y)}$$

$$= \int\int f_{XY}(x,y) \log \frac{1}{f_{X|Y}(x|y)} \, dx\, dy$$

$$= \int f_Y(y) \underbrace{\int f_{X|Y}(x|y) \log \frac{1}{f_{X|Y}(x|y)} \, dx}_{h(X|Y=y)} \, dy$$

Thm: $h(X^n) = h(X_1) + h(X_2|X_1) + \dots + h(X_n|X^{n-1})$

$h(X^n|Y) = h(X_1|Y) + \dots + h(X_n|X^{n-1}Y)$.

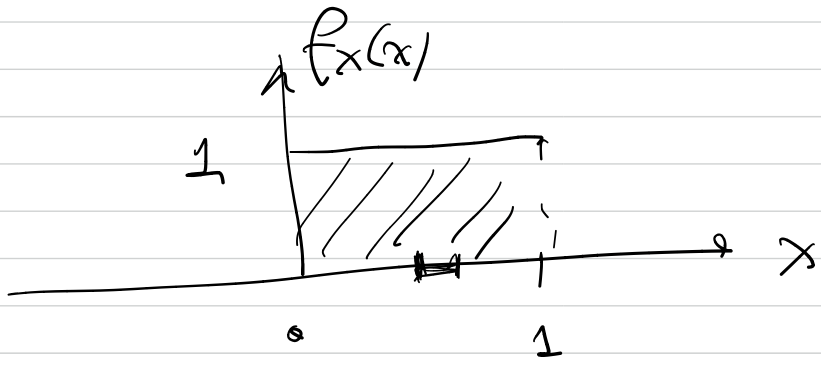
(Chain Rule)

Pf: $h(XY) = E \log \frac{1}{f_{XY}(XY)}$

$$= E \log \frac{1}{f_X(X) f_{Y|X}(Y|X)}$$

$$= h(X) + h(Y|X). \quad //$$

Example: X is Uniform $[0,1]$.



$$h(X) = \int_0^1 \log \frac{1}{1} \, dx = 0.$$

Example: Let c be a constant, X be a $\mathbb{R}^n$-valued RV.

Then $h(c+X) = h(X)$.

Pf: $(n=1)$ (general n is very similar):

$$f_X(x) = \lim_{\varepsilon \downarrow 0} \frac{\Pr(X \in (x, x+\varepsilon])}{\varepsilon}$$

$Y = c+X$, $f_Y(y) = \lim_{\varepsilon \downarrow 0} \frac{\Pr(Y \in (y, y+\varepsilon])}{\varepsilon}$

$$= \lim_{\varepsilon \downarrow 0} \frac{\Pr(X \in (y-c, y-c+\varepsilon])}{\varepsilon} = f_X(y-c)$$

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

$$= \int f_X(x) \log \frac{1}{f_X(x)}\, dx \qquad y = x+c, \quad dy = dx.$$

$$= h(X). \qquad //$$

<u>Thm</u>: if X is $\mathbb{R}$-valued RV & $c \neq 0$ constant.

$$h(cX) = h(X) + \log|c|$$

<u>Pf</u>: let $Y = cX$

$$f_Y(y) = \lim_{\varepsilon \downarrow 0} \frac{\Pr(Y \in (y, y+\varepsilon])}{\varepsilon}$$

$$= \lim_{\varepsilon \downarrow 0} \frac{\Pr(X \in (\frac{y}{c}, \frac{y+\varepsilon}{c}])}{c \cdot \varepsilon/c} = \frac{1}{c} f_X\left(\frac{y}{c}\right)$$

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

$$= \int f_X(x) \log \frac{c}{f_X(x)}\, dx \qquad x = \frac{y}{c} \quad dx = \frac{dy}{c}$$

$$= (\log c) + h(X) \qquad //$$

<u>Ex</u>: if X is uniform on $[a, b]$

$\rightarrow h(X) = \log_2 (b-a)$, because $X = a + (b-a)U$

U: uniform on $[0,1]$.

In particular $(b-a) \geq 1$

$$h(X) \begin{matrix} \geq \\ \leq \end{matrix} 0 .$$

$(b-a) \leq 1$,

<u>Thm</u>: if X is $\mathbb{R}$-valued RV taking values in $[a,b]$ then $h(X) \leq \log(b-a)$. (i.e., the uniformly distributed RV has largest entropy among such RVs.)

<u>Pf</u>: Let
$$q(x) = \begin{cases} 0 & \text{if } x \notin [a,b] \\ \frac{1}{b-a} & \text{if } x \in [a,b]. \end{cases}$$

$$\log(b-a) = \int q(x) \log \frac{1}{q(x)} \, dx$$

$$= \int f_X(x) \log \frac{1}{q(x)} \, dx \quad \text{for any } f_X \text{ s.t. } f_X(x) = 0 \text{ outside } [a,b].$$

$$h(X) = \int f_X(x) \log \frac{1}{f_X(x)} \, dx.$$

$$h(X) - \log(b-a) = \int f_X(x) \log \frac{q(x)}{f_X(x)} \, dx.$$

$$\leq \log \int f_X(x) \frac{q(x)}{f_X(x)} \, dx \quad \text{(concavity of log)}$$

$$= \log 1 = 0. \quad //$$

Ex: entropy of a Gaussian RV. Suppose X is Gaussian with 0-mean, variance = 1,

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right).$$

$$h(X) = E\left[\log \frac{1}{f(X)}\right)$$

$$= E\left[\frac{1}{2}\log 2\pi + (\log e)\frac{1}{2}X^2\right]$$

$$= \frac{1}{2}\log 2\pi + \frac{1}{2}(\log e) = \frac{1}{2}\log(2\pi e).$$

Ex: Entropy of X which is Gaussian mean $=\mu$, var $=\sigma^2$

$$= \underline{\frac{1}{2}\log(2\pi e\sigma^2)}, \text{ because}$$

$\frac{X-\mu}{\sigma}$ is 0-mean, var = 1, ~~also~~ and Gaussian, so

$$h\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{2}\log(2\pi e).$$

Thus: if X has variance σ^2, then

$$h(X) \le \frac{1}{2}\log(2\pi e\sigma^2).$$ [I.e., gaussians have maximal entropy among such random variables.]

Pf: Let $Y = \frac{X - (\text{mean of } X)}{\sigma}$. Y has 0 mean
var$(Y) = 1$.

it suffices to prove $h(Y) \le \frac{1}{2}\log(2\pi e)$ since

$h(X) = (\log\sigma) + h(Y)$.

let $q(y) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}y^2)$ be the density of a

gaussian with same mean & variance and let $f(y)$ be

the prob. density of Y. note that

$$\frac{1}{2}\log(2\pi e) = \int q(y)\log\frac{1}{q(y)}\,dy$$

$$= \int f(y)\log\frac{1}{q(y)}\,dy \quad (**)$$

why $(**)$? because

$$\int f(y)\log\frac{1}{q(y)}\,dy = \int f(y)\left(\frac{1}{2}\log(2\pi) + \frac{1}{2}(\log e)y^2\right)dy$$

$$= \frac{1}{2}\log 2\pi + \frac{1}{2}(\log e)E(Y^2)$$

$$= \frac{1}{2}\log 2\pi + \frac{1}{2}\log e = \frac{1}{2}\log(2\pi e).$$

we have to show

$$\underbrace{\int f(y)\log\frac{1}{f(y)}\,dy - \int f(y)\log\frac{1}{q(y)}\,dy}\le 0.$$

concavity of $\log()$

but $$= \int f(y)\log\frac{q(y)}{f(y)}\,dy \le \log\int \cancel{f(y)}\frac{q(y)}{\cancel{f(y)}}\,dy$$

$$= \log 1 = 0. \quad //$$

- Relationship between differential entropy h and discrete entropy H.

Suppose X is a $\mathbb{R}$-valued RV. Let $\delta > 0$,

define $X_\delta = \delta \left\lfloor \frac{X}{\delta} \right\rfloor$.

$$X_\delta = \begin{cases} \vdots \\ -\delta & \text{if } -\delta \leq X < 0 \\ 0 & \text{if } 0 \leq X < \delta \\ \delta & \text{if } \delta \leq X < 2\delta \\ \vdots \end{cases}$$

$$X_\delta = n\delta \iff n\delta \leq X < (n+1)\delta.$$

X_δ is obtained by rounding X down to the nearest multiple of δ.

$$H(X_\delta) = \sum_n \Pr(X_\delta = n\delta) \log \frac{1}{\Pr(X_\delta = n\delta)}$$

note $\Pr(X_\delta = n\delta) = \int_{n\delta}^{(n+1)\delta} f(x)\, dx \qquad f = f_X.$

So $H(X_\delta) = \sum_n \int_{n\delta}^{(n+1)\delta} f(x) \log \frac{1}{\Pr(X_\delta = n\delta)} dx$

Also observe that

$$\Pr(X_\delta = n\delta) = \int_{n\delta}^{n\delta+\delta} f(x)\,dx \approx f(n\delta)\cdot\delta$$

for δ small so

$$H(X_\delta) \approx \sum_n \int_{n\delta}^{(n+1)\delta} f(x) \log \frac{1}{\delta f(n\delta)}\, dx$$

$$\approx \sum_n \int_{n\delta}^{n\delta+\delta} f(x) \log \frac{1}{\delta f(x)}\, dx$$

$$= \int f(x) \log \frac{1}{\delta f(x)}\, dx$$

$$= \log \frac{1}{\delta} + h(X).$$

So: $$h(X) = \lim_{\delta \downarrow 0} \left(H(X_\delta) + \log \delta \right) \quad //$$

Ex: X is uniform in $[0,1]$.

with $\delta = 2^{-k}$ $\quad X_\delta \in \{0, \delta, \ldots, (2^k-1)\delta\}$

with $\Pr(X_\delta = n\delta) = 2^{-k}$ $\quad n = 0, \ldots, 2^k-1.$

$H(X_\delta) = k$

We also see: X and Y are $\mathbb{R}$-valued,

$$X_\delta = \delta \left\lfloor \frac{X}{\delta} \right\rfloor \text{ and } Y_\varepsilon = \varepsilon \left\lfloor Y/\varepsilon \right\rfloor$$ be quantized

versions of X and Y. By the same reasoning,

$$h(X,Y) = \lim_{\delta,\varepsilon \downarrow 0} \left[H(X_\delta, Y_\delta) + \log(\delta\varepsilon) \right],$$

Now define

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

as the mutual information between X & Y.

and observe
$$h(X) = \lim_{\delta \to 0} H(X_\delta) + \log \delta$$

$$h(Y) = \lim_{\varepsilon \to 0} H(Y_\varepsilon) + \log \varepsilon$$

$$h(XY) = \lim_{\varepsilon,\delta \downarrow 0} H(X_\delta Y_\varepsilon) + \log \delta + \log \varepsilon$$

So
$$I(X;Y) = \lim_{\varepsilon,\delta \downarrow 0} I(X_\delta ; Y_\varepsilon),$$

Hence no need for "new" notation for differential mutual information.

Example: let $X \sim N(0, \sigma_X^2)$
$Z \sim N(0, \sigma_Z^2)$ } independent.

let $Y = X + Z \sim N(0, \sigma_X^2 + \sigma_Z^2) = \sigma_Y^2$

$I(X; Z) = 0$ (because X & Z indep.)

$$I(X; Y) = h(X) + h(Y) - h(X, Y)$$
$$= h(Y) - h(Y|X)$$
$$= \frac{1}{2}\log(2\pi e \sigma_Y^2) - h(Y|X)$$
$$= \frac{1}{2}\log(2\pi e \sigma_Y^2) - h(Y - \alpha(X)|X) \text{ for any } \alpha()$$
$$= \frac{1}{2}\log \text{ '' } - h(Y - X|X)$$
$$= \frac{1}{2}\log(2\pi e \sigma_Y^2) - h(Z|X)$$
$$= \frac{1}{2}\log(2\pi e \sigma_Y^2) - h(Z)$$
$$= \frac{1}{2}\log 2\pi e \sigma_Y^2 - \frac{1}{2}\log 2\pi e \sigma_Z^2$$
$$= \frac{1}{2}\log \frac{\sigma_Y^2}{\sigma_Z^2} = \frac{1}{2}\log\left(1 + \frac{\sigma_X^2}{\sigma_Z^2}\right).$$