# Classical case

Given $f: \mathcal{X} \to \mathbb{C}$, $\quad \mathrm{supp}(f) := \{x \in \mathcal{X} : f(x) \neq 0\}$

Def (Relative entropy). Let $P, Q$ be prob distributions on $\mathcal{X}$.

$$D(P\|Q) = \begin{cases} \sum_{x\in\mathcal{X}} P(x) \log \frac{P(x)}{Q(x)} & \text{if } \mathrm{supp}(P) \subseteq \mathrm{supp}(Q) \\ +\infty & \text{else} \end{cases}$$

- Also called Kullback–Liebler Divergence, 1954
- $D(P\|Q) = \mathop{\mathbb{E}}_{X\sim P}\left(\log \frac{P(X)}{Q(X)}\right) = \sum Q(x) \frac{P(x)}{Q(x)} \log \frac{P(x)}{Q(x)}$

  $= \mathop{\mathbb{E}}_{X\sim Q}\left(\frac{P}{Q} \log \frac{P}{Q}\right)$

  why this is well defined $\mathrm{supp}(P) \subseteq \mathrm{supp}(Q) \Rightarrow \frac{P(x)}{Q(x)}$ could be $\frac{c_1}{c_2}$

  or $\frac{0}{c_2} = 0$, $0 \log 0 = 0$

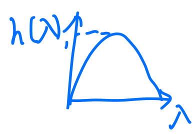
  Not $\frac{c_2}{0}$ (undefined)

- Describe How well we can distinguish $P$ from $Q$ (Hypothesis Testing)

Example: $P, Q \in \{0,1\}$ $\quad P(0) = \lambda$, $P(1) = 1-\lambda$, $Q(0) = \frac{1}{2}$, $Q(2) = \frac{1}{2}$

$$D(P\|Q) = \lambda \log \frac{\lambda}{0.5} + (1-\lambda) \log \frac{1-\lambda}{0.5} = \lambda \log\lambda + (1-\lambda)\log(1-\lambda) - 1$$

$$= 1 - h(\lambda) \to 0 \text{ if } \lambda \to \frac{1}{2}$$



Properties of $D(\cdot\|\cdot)$

① $D(P\|Q) \geq 0$ , with equality iff $P=Q$

② $D(\sum \lambda_i P_i \| \sum \lambda_i Q_i) \leq \sum \lambda_i D(P_i \| Q_i)$ Joint Convex

Pf: ① $D(P\|Q) = \mathbb{E}_Q \left( \frac{P}{Q} \log \frac{P}{Q} \right) \geq \mathbb{E}_Q \left(\frac{P}{Q}\right) \log \mathbb{E}\left(\frac{P}{Q}\right)$ $\left( t \mapsto t \log t \right.$ convex $\left.\right)$

$= 1 \cdot \log 1 = 0$

$\mathbb{E}_Q \frac{P}{Q} = \sum Q(x) \frac{P(x)}{Q(x)} = \sum P(x) = 1$

② Define an extension of $D$

$$\bar{D}(P\|Q) = \sum P(x) \log \frac{P(x)}{Q(x)} + (Q(x) - P(x))$$

$\bar{D}(P\|Q) = D(P\|Q)$ for $P, Q \in \mathcal{P}(\mathcal{X})$

We show $\bar{D} : \mathbb{R}_+^{\mathcal{X}} \times \mathbb{R}_+^{\mathcal{X}} \to \mathbb{R}_+$ joint convex

$\bar{D}(P\|Q) = \sum_{x \in \mathcal{X}} g(P(x), Q(x))$ $\quad g(a,b) = a \log \frac{a}{b} + (b-a)$

Hessien $\nabla^2 g(a,b) = \begin{bmatrix} 1/a & -1/b \\ -1/b & a/b^2 \end{bmatrix} \succeq 0 \Rightarrow g$ is convex over $\mathbb{R}^2$

$\Rightarrow \bar{D}$ is joint convex

From Relative entropy to Entropies

① For $P$ over $\mathcal{X}$, $\quad H(P) = \log|\mathcal{X}| - D(P \| \frac{1}{|\mathcal{X}|})$

$$D(P\|\tfrac{1}{|\mathcal{X}|}) = \sum p(x) \log \frac{p(x)}{1/|\mathcal{X}|} = \sum p(x)\left(\log p(x) + \log|\mathcal{X}|\right)$$
$$= \log$$

② For $P_{XY}$ over $\mathcal{X} \times \mathcal{Y}$,

$$H(X|Y) = \log|\mathcal{Y}| - D(P_{XY} \| P_X \times \tfrac{1}{|\mathcal{Y}|})$$
$$I(X:Y) = D(P_{XY} \| P_X \times P_Y)$$

Indeed, $D(P_{XY}\|P_X P_Y) = \mathop{\mathbb{E}}_{P_{XY}}\left(\log \frac{P_{XY}}{P_X \times P_Y}\right) = \mathop{\mathbb{E}}_{P_{XY}} \log P_{XY} + \mathop{\mathbb{E}}_{P_{XY}} \log\frac{1}{P_X} + \mathop{\mathbb{E}}_{P_{XY}} \log\frac{1}{P_Y}$

$$= H(X) + H(Y) - H(XY)$$

③ For $P_{XYZ}$ over $\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$

$$I(X:Y|Z) := H(XZ) + H(YZ) - H(XYZ) - H(Z)$$
$$= D(P_{XYZ} \| \tfrac{1}{|\mathcal{X}|} \times P_{YZ}) - D(P_{XZ} \| \tfrac{1}{|\mathcal{X}|} \otimes P_Z)$$
$$= (\log|\mathcal{X}| - H(X|YZ)) - (\log|\mathcal{X}| - H(X|Z))$$
$$= H(X|Z) - H(X|YZ)$$
$$= H(XZ) - H(Z) - H(XYZ) + H(YZ)$$

Can we prove $I(X:Y|Z)$ using relative entropy? Yes.

# Data processing inequality

Denote $C(\mathcal{X}) = \{ f : \mathcal{X} \to \mathbb{C} \}$ $\qquad C^{\mathcal{X}} \quad C^{\mathcal{Y}}$

A classical channel from $\mathcal{X}$ to $\mathcal{Y}$ is a linear map $N : C(\mathcal{X}) \to C(\mathcal{Y})$

① Positive: $f \geq 0 \Rightarrow Nf \geq 0$

② Measure preserving: $\mathbb{E}_X f = \mathbb{E}_Y Nf$

So $N : \mathcal{P}(X) \to \mathcal{P}(\mathcal{Y})$ send p.d.f to p.d.f

$N : \mathbb{C}^{\mathcal{X}} \to \mathbb{C}^{\mathcal{Y}}$ is given by a matrix $N(y|x)$

$p \in \mathbb{C}^{\mathcal{X}} \to \quad q(y) = \sum_{x \in \mathcal{X}} N(y|x) p(x)$

① $\Rightarrow N(y|x) \geq 0 \quad \forall y, x$

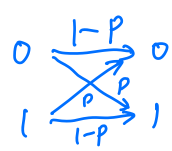
② $\Rightarrow \sum_y N(y|x) = 1$

Such $N(Y|X)$ is called a stochastic matrix

actually the coditional prob. $P_{Y|X}$ for $P_{XY} = N(Y|X) P_X$

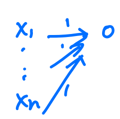
$X$ sender $\xrightarrow{N(Y|X)}$ $Y$ Reciever $\qquad N(y|x)$ prob. receiving $y$ given input $X$.

Example: (Bit flip Channel) Fix $p \in [0,1]$:




Model for noisy communication.

Example: (Forgetting Channel): $\Phi : \mathbb{C}^{\mathcal{X}} \to \mathbb{C}$.



Data processing Inequality

For any channel $N$,

$$D(Np \| Nq) \le D(p\|q)$$

Pf: $$D(Np\|Nq) = \sum_y Np(y) \log \frac{Np(y)}{Nq(y)}$$

$$= \sum_{x,y} N(y|x)\,p(x) \log \frac{Np(y)}{Nq(y)}$$

$$= \sum_x p(x) \left( \sum_y N(y|x) \log \frac{Np(y)}{Nq(y)} \right)$$

$$= \sum_x p(x) \log \exp\left( \sum_y N(y|x) \log \frac{Np(y)}{Nq(y)} \right)$$

$$D(p\|q) = \sum_x p(x) \log \frac{p(x)}{q(x)}$$

$$D(p\|q) - D(Np\|Nq) = \sum p(x) \log \frac{p(x)}{q(x) \exp(\cdot \text{---} )}$$

$$= \sum_x p(x) \left[ \log p(x) - \log\left( q(x) \exp \sum_y N(y|x) \log \frac{Np(y)}{Nq(y)} \right) \right]$$

$$\ge \sum_x p(x) \log p(x) - \log\left( q(x) \sum_y N(y|x) \frac{Np(y)}{Nq(y)} \right)$$

$$= \sum_x p(x) \log \frac{p(x)}{\left( q(x) \sum_y N(y|x) \frac{Np(y)}{Nq(y)} \right)}$$

Claim $\quad r(x) = q(x) \sum_{y} N(y|x) \dfrac{Np(y)}{Nq(y)}$ is a prob distribution

Indeed $\quad r = R \quad Np$

$R : \mathbb{C}^{Y} \to \mathbb{C}^{X}$ is the channel

$$Ru(x) = \sum_{y} \frac{N(y|x)q(x)}{Nq(y)} u(y) \qquad R(x|y) = \frac{N(y|x)q(x)}{Nq(y)}$$

$$\sum_{x} R(x|y) = \sum_{x} \frac{N(y|x)q(x)}{Nq(y)} = \frac{Nq(y)}{Nq(y)} = 1 \qquad \text{Bayes Rule!}$$

Thus $\quad D(p\|q) - D(Np\|Nq) \geq D(p\|RNp) \geq 0$

If $\quad D(p\|q) = D(Np\|Nq) \Rightarrow \quad p = RNp$

Moreover $\quad RNq = q$

$D(p\|q) = D(Np\|Nq)$ iff $\exists\, R : \mathbb{C}^{X} \to \mathbb{C}^{Y}$,

$R \circ N(p) = p \qquad R \circ N(q) = q.$