August 28th 201

lec: 6 Sufficient statistics
fanos inequality.

Recap: - log sum inequality and applications
- Data processing inequality

Aug 31 – Sep 8
Out of town

Start at 8:30 am
Coming Wednesday Aug 30

Data Processing inequalities
$X \to Y \to Z$
Markov chain if
$P(xyz) = P(x)P(y|x)P(z|y)$
Equivalently (X, Z) are independant given Y.
$P(xz|y) = P(x|y)P(z|y)$
Thus
$X \to Y \to Z$
$\Rightarrow Z \to Y \to X$

If $g(\cdot)$ is any deterministic function, we have
$X \to Y \to g(Y)$ is a MC.

Consequence:

$$I(X; YZ) = I(X; Y) + \underbrace{I(X; Z|Y)}_{\to 0}$$
$$= I(X; Z) + I(X; Y|Z)$$
$$\Rightarrow I(X; Y) \geq I(X; Z) \to (1)$$
and $I(X; Y) \geq I(X; Y|Z)$. $\to (2)$

In particular $I(X; Y) \geq I(X; g(Y))$

Eq (2) does not hold if $\{X, Y, Z\}$ do not form a markov chain

Eg: (slightly different from the text)
Suppose $Z = X \oplus Y$, $\{X, Y \in \{0, 1\}$. are binary independant random variables.
$I(X; Y) = 0$.
$I(X; Y|Z)$
$= H(X|Z) - H(X|Y, Z)$ ($H(X|Y,Z) \to 0$)
$= H(X|Z)$
$= H(X)$ {can show X, Z are independant}

TODAY
- Conclude data processing inequality
- Sufficient statistics
- fanos inequality.

Sufficient statistics
Let θ be a RV with finite alphabet. Let $\{P_\theta(\cdot)\}$ be a parameterized collection of prob mass functions (pmf).
We perform the following experiment
(a) select θ, (b) select sample vector (iid) $\underline{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}$ where $X_i \sim P_\theta(\cdot)$.

© let $T(\underline{X})$ be some statistical function of $\underline{X}$

Eg: $\mu(\underline{X}) = \frac{\sum_{i=1}^{n} X_i}{n}$, $\sigma(\underline{X}) = \frac{\sum_{i=1}^{n} X_i^2}{n}$

Then $\Theta \longrightarrow \underline{X} \longrightarrow T(\underline{X})$ form a markov chain.

Thus $I(\Theta; \underline{X}) \geq I(\Theta; T(\underline{X}))$. Suppose equality holds then this implies that Θ, X are independant given $T(\underline{X})$! Then $T(\underline{X})$ is said to be sufficient statistics for Θ.

Eg 1: let Θ be a random variable taking on values in $[0,1]$, let $\underline{X}$ be a 2nd random variable such that

$$Pr(X_i = 1) = \Theta, \quad Pr(X_i = 0) = 1 - \Theta$$

(X_i is binary $\in \{0,1\}$)

$\underline{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_m \end{bmatrix}$ X_i iid. set $W = \sum_{i=1}^{n} X_i = w_H(\underline{X})$ Hamming weight

Our interest is in $\Theta \longrightarrow \underline{X} \longrightarrow W$

Want to show that $\Theta \longrightarrow W \longrightarrow \underline{X}$ ie, $\underline{X}$ is independant of Θ given W.

$$p\left(\underline{x} \mid W = w, \Theta = \theta\right) = \begin{cases} 0 & w_H(\underline{x}) \neq w \\ \text{Constant} & w_H(\underline{x}) = w \end{cases}$$

$$\text{Constant} = \frac{1}{\binom{m}{w}}$$

$$p\left(\underline{x} \mid W = w\right) = \sum_{\theta} p\left(\underline{x} \mid W = w, \Theta = \theta\right) p\left(\theta \mid w\right)$$

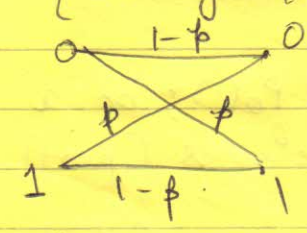
$$= \frac{1}{\binom{m}{w}} \sum_{\theta} p(\theta \mid w) = \frac{1}{\binom{m}{w}} = p\left(\underline{x} \mid W = w, \Theta = \theta\right)$$

$\therefore \Theta \longrightarrow W \longrightarrow X$ is a MC

W is sufficient statistic of Θ.

Eg 2: Binary code $\mathcal{C} \subseteq \{\underline{0}, \underline{1}\} \subseteq \mathbb{F}_2^n$ ← all 0, all 1 vectors.

{binary $\{0,1\}$ n-tuples}



Binary Symmetric channel

$$\boxed{\underline{Y} = \underline{X} \oplus \underline{Z}}$$

Alternate description of BSC.

$Z_i \in \{0,1\}$

$Z_i = \begin{cases} 1 & \text{prob } p \\ 0 & \text{prob } 1-p \end{cases}$

$\{Z_i\}$ iid

Let $W = \sum_{i=1}^{n} Y_i = w_H(\underline{Y})$

We have

$\underline{X} \to \underline{Y} \to W$. claim: $\underline{X} \to W \to \underline{Y}$

ie, $\underline{X}$ and $\underline{Y}$ are independant given W.

To show $p(\underline{x} \mid y, w) = p(\underline{x} \mid w)$

$$p(\underline{0} \mid y, w) = \frac{p(y \mid \underline{0})\, p(\underline{0})}{p(y \mid \underline{0})\, p(\underline{0}) + p(y \mid \underline{1})\, p(\underline{1})}$$

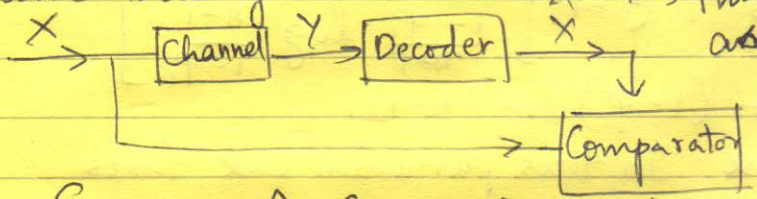
$p(\underline{0} \mid y)$

$$= \frac{p^w (1-p)^{m-w} \left(\frac{1}{2}\right)}{\frac{1}{2}\left[p^w (1-p)^{m-w} + (1-p)^w p^{m-w}\right]}$$

= a function of w only $= p(\underline{0} \mid w)$

Fanós inequality (FI)

Motivational setting



$$E = \begin{cases} 1 & X \neq \hat{X} \text{ (an error)} \\ 0 & \text{else} \end{cases}$$

$X \to Y \to \hat{X}$ a markov chain.

Idea: The probability of decoding error should reflect the residual uncertainity.

Set $P_e = \Pr\{E=1\}$ prob of error.

let $X \in \mathcal{X}$ $|\mathcal{X}| < \infty$ {finite alphabet}

Consider

$$H(X, E \mid \hat{X}) = H(X \mid \hat{X}) + \underbrace{H(E \mid X, \hat{X})}_{\to 0} \qquad H(X \mid Y) \leq H(X \mid \hat{X})$$

$$= H(E \mid \hat{X}) + H(X \mid E, \hat{X})$$

$$\therefore H(X \mid \hat{X}) = H(E \mid \hat{X}) + H(X \mid E, \hat{X}) \quad \text{(residual uncertainity)}$$

$$\leq H(E) + \underbrace{H(X \mid \hat{X}, E=0)}_{\to 0} P(E=0) + H(X \mid \hat{X}, E=1)\, P(E=1).$$

$$H(X|\hat{X}) \le H(P_e) + (P_{\bar{e}})\, H(X \mid E=1, \hat{X})$$
$$\le H(P_e) + P_e \log(|\mathcal{X}_0| - 1).$$
$$\Rightarrow \boxed{H(X|Y) \le H(X|\hat{X}) \le H(P_e) + P_e \log(|\mathcal{X}_0|-1)}$$

Let X, Y be 2 RVs; same alphabet.
Let $\Pr(X \neq Y) = p$, Then $H(X|Y) \le H(p) + p\log(|\mathcal{X}_0|-1)$

Eg: $X \longrightarrow x_1, x_2, \ldots x_n$
$p_1 > p_2 \ldots > p_n$
$Y = \hat{X} = x_1$!! (Constant)

Applying fano's inequality $P(\hat{X} \neq X) = P(Y \neq X) = 1 - p_1$

$$H(X|Y) \le H(1-p_1) + (1-p_1)\log(n-1)$$
$$= H(p_1) + (1-p_1)\log(n-1).$$

$$H(X|Y) = H(X|Y=x_1) = H(X)$$
$$= p_1 \log\frac{1}{p_1} + p_2 \log\frac{1}{p_2} + \ldots p_n \log\frac{1}{p_n}$$

When $p_2 = p_3 = \ldots = p_n = \dfrac{(1-p_1)}{n-1}$.

then $H(X|Y) = p_1 \log\frac{1}{p_1} + (n-1)\left[\frac{1-p_1}{n-1}\log\frac{n-1}{1-p_1}\right]$

$$= H(p_1) + (1-p_1)\log(n-1).$$

$\therefore$ Equality holds in this case. In this particular case fano's further inequality is tight.

Applications of Jensen's Inequality

$X \sim p(x)$, $Y \sim q(x)$ } Same alphabet $\mathcal{X}_0$

Want an estimate in entropic terms for $p(X=Y)$.

Note $p(X=Y) = \sum_{x \in \mathcal{X}_0} p(x)\, q(x)$ ⇒ assuming X, Y indep.

Consider

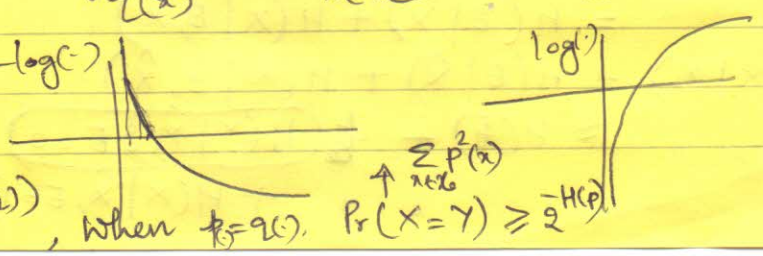
$$H(p) + D(p\|q) = \sum_{x \in \mathcal{X}_0} p(x)\log\frac{1}{p(x)} + \sum_{x \in \mathcal{X}_0} p(x)\log\frac{p(x)}{q(x)}$$
$$= \sum_{x \in \mathcal{X}_0} p(x)\log\frac{1}{q(x)} = \sum_{x \in \mathcal{X}_0} p(x)\left[-\log q(x)\right]$$

Using Jensen's Inequality:

$$\ge -\log \sum_x p(x)\, q(x).$$

$$= -\log(\Pr(X=Y))$$

$$\Rightarrow \Pr(X=Y) \ge 2^{-(H(p)+D(p\|q))}$$



, when $p = q(\cdot)$, $\Pr(X=Y) \ge 2^{-H(p)}$