Lec 04 01/24/2013

More properties of KL:

(4) Lemma: $p$ distribution on $\mathcal{U}$

$$\Rightarrow \quad H(p) = \log|\mathcal{U}| - \mathcal{D}(p \| u)$$

uniform

(5) KL-div vs. Chernoff Bounds:

Toss a fair coin $n$ times

$$\Pr(\text{we see } 0.6n \text{ heads}) \leq 2^{-\Omega(n)}.$$

$$(1/2+\varepsilon)n \qquad \leq 2^{-\frac{\varepsilon^2 n}{4}}$$

What is the constant?

uniform

$$\Pr(\text{we see } pn \text{ heads}) \leq 2^{-n\,\mathcal{D}(p \| u)}$$

(p-biased coin)

$$\frac{2^{-n\mathcal{D}(p\|u)}}{n^2} \leq$$

In general, experiment with distribution $q$ on some universe $\mathcal{U}$.

$X_1, \dots, X_n$ iid

$\Pr[$ freq. of symbols we see are according to $p] \leq 2^{-D(p\|q)}$.

$$\frac{2^{-D(p\|q)}}{(n+1)^{|\mathcal{U}|}}$$

(no proof)

Data processing Inequality:

Def. $X, Y, Z$ form a Markov Chain $(X \to Y \to Z)$ if the Conditional dist. of $Z$ only depends on $Y$. i.e.,

$\Pr(Z=z \mid Y=y, X=x) = \Pr(Z=z \mid Y=y)$.

E.g., $Z = g(Y)$.

Thm. If $X \to Y \to Z$, Then,

$$I(X;Y) \geq I(X;Z).$$

$$X \to Y \to Z \Rightarrow p(x,y,z) = p(x)\,p(y|x)\,p(z|y)$$

<u>Observation:</u>

$$p(x,z|y) = \frac{p(x,y,z)}{p(y)} = \frac{p(x).\,p(y|x)}{p(y)}\,p(z|y)$$

$$= p(x|y)\,p(z|y).$$

That is, given $Y$, $X \perp Z$.

<u>Proof of Thm:</u> $I(X;Y,Z) =$

$$I(X;Z) + I(X;Y|Z)$$

also $I(X;Y,Z) = I(X;Y) + \underbrace{I(X;Z|Y)}_{=\,0}$ (chain rule)

( 0 by observation )

$$\Rightarrow \quad I(X;Y) = I(X;Z) + \underbrace{I(X;Y|Z)}_{\geq 0}$$

$$\Rightarrow \quad I(X;Y) \geq I(X;Z). \qquad \square$$

---

**Coro:** $I(X;Y|Z) \leq I(X;Y)$.
when $X \to Y \to Z$.

**Coro':** $I(X;Y|\,g(Y)) \leq I(X;Y)$.
since $X \to Y \to g(Y)$.

---

## Fano's Inequality:

Suppose we know r.v. $Y$ and want to guess the value of a correlated $X$.

**Exercise:** $X = g(Y)$ (is a good case)

$$\Updownarrow$$

$$H(X|Y) = 0.$$

Fano is a quantitative version.

Say $Y \to g(Y) = \tilde{X}$ estimate of $X$.

$p_{err} := Pr(\tilde{X} \neq X)$.

(Best strategy outputs $\text{argmax}\, P(X \mid Y = y)$.)
(max. likelihood)
$\hookrightarrow$ (Exercise)

<u>Thm:</u> $h(p_{err}) + p_{err} \log(n-1) \geq H(X|Y)$.

where $n = |\text{supp}(X)|$.

<u>Proof.</u> $\mathcal{E}$ := event that we make an error.

$\Rightarrow P(\mathcal{E}) = p_{err}$.

& $H(\mathcal{E}) = h(p_{err})$. We know:

* $H(\mathcal{E} \mid X, Y) = 0$. add $H(X|Y)$.

$\Rightarrow H(X, \mathcal{E} \mid Y) = H(X|Y)$.

expand in 2 ways, the one before & :

$$H(X, \mathcal{E} \mid Y) = H(\mathcal{E} \mid Y) + H(X \mid \mathcal{E}, Y).$$

$$\Rightarrow H(X \mid Y) = H(\mathcal{E} \mid Y) + H(X \mid \mathcal{E}, Y)$$

$$\leq \underbrace{H(\mathcal{E})}_{h(P_{err})} + \underbrace{H(X \mid \mathcal{E}, Y)}_{?}$$

$$\star \quad H(X \mid \mathcal{E}, Y) = p(\mathcal{E}=0)\, \overbrace{H(X \mid Y, \mathcal{E}=0)}^{0}$$

$$+ \underbrace{p(\mathcal{E}=1)}_{P_{err}}\, \underbrace{H(X \mid Y, \mathcal{E}=1)}_{\leq \log(n-1)} .$$

since we know $X \neq Y$.

✓

$\square$ .

# Asymptotic Equipartition Property (AEP)

$X$ with dist. $p$ (source)

Sample $X_1, \ldots, X_n$ iid. $\leftarrow X$.

Suppose outcome is $a_1, \ldots, a_n$.

$$\Rightarrow \Pr_{\vec{a}}\left[ p(a_1, \ldots, a_n) \approx 2^{-H(X)\cdot n} \right] \longrightarrow 1.$$

" almost all events are almost equally surprising " !

<u>Proof</u> (by weak law of large numbers)

$\Downarrow$

* Suppose $Z_1, \ldots, Z_n$ iid integers

$$\begin{pmatrix} \forall \varepsilon, \delta, \exists n_0, \\ \forall n \geq n_0 \end{pmatrix} \Rightarrow \Pr\left[ \left| \frac{Z_1 + \cdots + Z_n}{n} - \overset{\text{mean}}{\mu} \right| > \varepsilon \right] < \delta.$$

Here we set $Z := \log \frac{1}{p(a)}$ w.p. $p(a)$.

$$\Rightarrow \mathbb{E}(Z) = H(X).$$

$$\Rightarrow \Pr\left[ \underbrace{\left| \frac{1}{n} \sum \log \frac{1}{p(a_i)} - H(X) \right| > \varepsilon} \right] \le \varepsilon.$$

$$\Downarrow$$

$$\left| + \frac{\log p(a_1, \ldots, a_n)}{n} + H(X) \right| > \varepsilon$$

$$\Downarrow$$

$$p(a_1, \ldots, a_n) \le 2^{-n(H(X) - \varepsilon)}$$

&

$$2^{-n(H(X)+\varepsilon)} \le p(a_1, \ldots, a_n)$$

$\square$

---

AEP: For $(a_1, \ldots, a_n) \sim p(a_1, \ldots, a_n)$

$$\Downarrow$$

$$\Pr\left[ 2^{-n(H(X)+\varepsilon)} \le p(a_1, \ldots, a_n) \le 2^{-n(H(X)-\varepsilon)} \right] \ge 1-\varepsilon.$$

Corollary (typicality) $X$ r.v. on $\Sigma$,

Typical set $A_\varepsilon^{(n)}$ with respect to $p(x)$ is the set $\{(x_1, \dots, x_n) \in \Sigma^n \mid$
$$2^{-n(H(X)+\varepsilon)} \le p(x_1, \dots, x_n) \le 2^{-n(H(X)-\varepsilon)}\}.$$

Lemma: $a_1, \dots, a_n$ i.i.d $\leftarrow X$,

① $\Pr[(a_1, \dots, a_n) \in A_\varepsilon^{(n)}] \ge 1-\varepsilon$.

② $(1-\varepsilon)2^{n(H(X)-\varepsilon)} \le |A_\varepsilon^{(n)}| \le 2^{n(H(X)+\varepsilon)}$

$\square$