COMPSCI 650 Applied Information Theory 2/18/2016

Lecture 8

Instructor: Arya Mazumdar *Scribe:* [redacted]

1 Fano's inequality

Recall: If you have a Markov chain :

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

Statistical inference:

$$X \to Y \to \hat{X}$$
$$\hat{X} = g(y)$$
$$I(X;Y) \geq I(X;\hat{X})$$
$$P(X \neq \hat{X}) \geq \frac{H(X|Y) - 1}{\log(|X| - 1)} \qquad Fano's\ Inequality$$

Fano's inequality shows the bottom line for the probability of error. We are going to prove that: **Proof** Fano's Inequality

$$E = \begin{cases} 0 & x = \hat{x} \\ 1 & x \neq \hat{x} \end{cases} \tag{1}$$
$$P(E = 1) = P(X \neq \hat{X}) = P_e \tag{2}$$
$$H(X, E|\hat{X}) = H(X|\hat{X}) + H(E|X, \hat{X}) \tag{3}$$
$$= H(E|\hat{X}) + H(X|E, \hat{X}) \tag{4}$$

$H(E|X, \hat{X})$ in line (3) is equal to zero because if we know both $X, \hat{X}$ then we can compute the E for sure then where is no uncertainty in E given $X, \hat{X}$, therefore the entropy is zero.

$$H(X|\hat{X}) = H(E|\hat{X}) + H(X|E, \hat{X}) \tag{5}$$
$$By\ Removing\ Conditions\ H(X|\hat{X}) \leq H(E) + H(X|E, \hat{X}) \tag{6}$$

We know E is binary random variable so we can use binary entropy formula.

$$H(X|\hat{X}) \leq H(E) + H(X|E, \hat{X}) = H_2(P_e) + P(E = 1)H(X|E = 1, \hat{X}) + P(E = 0)H(X|E = 0, \hat{X}) \tag{7}$$

There is going to be no uncertainty in X if we know the estimation of X and also know there is no error. So, the entropy $H(X|E = 0, \hat{X})$ equals to zero.

$$\sup_x H(X|E = 1, \hat{X}) = P_e \log(|X| - 1) \tag{8}$$
$$H(X|\hat{X}) \leq H_2(P_e) + P_e \log(|X| - 1) \tag{9}$$

We also know from the Markov chain that $H(X|\hat{X}) \geq H(X|Y)$. So we have :

$$H(X|\hat{X}) \geq H(X|Y) \tag{10}$$
$$H(X|\hat{X}) \leq H_2(P_e) + P_e \log(|X| - 1) \tag{11}$$
$$(10), (11) \to H_2(P_e) + P_e \log(|X| - 1) \geq H(X|Y) \tag{12}$$

The equation (12) is the lower bound for multi-hypothesis testing.

2 Binary Hypothesis Testing

In Binary Hypothesis Testing we have two probability distribution P_1, P_2. The nature(V) will choose a sample from one of the distributions with a specific probability. We want to estimate which distribution the sample is chosen from.

$$V = \begin{cases} 1 & with\ probability\ \frac{1}{2} \\ 2 & with\ probability\ \frac{1}{2} \end{cases} \tag{13}$$

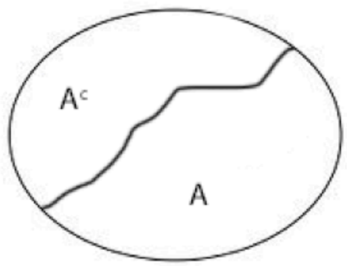
$g(X)$ is an estimator that operates on X and tells about whether X is from distribution P_1 or P_2.

We can define the probability of error as following:

$$Pr(Error) = P_1(g(x) = 2)P(V = 1) + P_2(g(x) = 1)P(V = 2) \tag{14}$$

$$\overset{(13)}{=} \frac{1}{2}\left(P_1(g(x) = 2) + P_2(g(x) = 1)\right) \tag{15}$$

Any estimator g will divide $r.v$ X into two parts.



A is the region where g(x) is equal to 1 and A^c is the region where g(x) is equal to 2. Next, we found the best probability of error (P_e^*).

$$P_e^* = \frac{1}{2}\inf_{A\subset\mathcal{X}}[P_1(A^c) + P_2(A)] \tag{16}$$

$$= \frac{1}{2}\inf_{A\subset\mathcal{X}}[1 - P_1(A) + P_2(A)] \tag{17}$$

$$= \frac{1}{2}\left[1 - \sup_{A\subset\mathcal{X}}[P_1(A) - P_2(A)]\right] \tag{18}$$

$$\tag{19}$$

We define $||P_1 - P_2||_{TV} = \sup_{A\subset\mathcal{X}}[P_1(A) - P_2(A)]$ (Total Variation Distance). Which is Le Cam's equation.

$$\rightarrow P_e^* = \frac{1}{2}\left[1 - ||P_1 - P_2||_{TV}\right] \tag{20}$$

Example: Calculating the total variation distance.

$$\mathcal{X} = \{1, 2, 3\}$$

$$P_1 : P_1(1) = 0.5 \qquad P_1(2) = 0.25 P_1(3) = 0.25$$

$$P_2 : P_2(1) = 0.3 \qquad P_2(2) = 0.3 P_2(3) = 0.4$$

subset	$P_1(A) - P_2(A^c)$
{1}	0.2
{2}	-0.05
{3}	-0.15
{1,2}	0.15
{1,3}	0.05
{2,3}	-0.2

Total variation distance is the maximum distance of the all possible subsets of $\mathcal{X}$.

$$||P_1 - P_2||_{TV} = \sup_{A \subset \mathcal{X}} [P_1(A) - P_2(A)] = 0.2$$

2.1 Symmetry of Total Variation distance

Proof

$$\begin{aligned} ||P_1 - P_2||_{TV} &= \sup_{A \subset \mathcal{X}} [P_1(A) - P_2(A)] && (21) \\ &= \sup_{A \subset \mathcal{X}} [1 - P_1(A^c) - 1 + P_2(A^c)] && (22) \\ &= \sup_{A^c \subset \mathcal{X}} [P_2(A^c) - P_1(A^c)] && (23) \\ &= \sup_{B \subset \mathcal{X}} [P_2(B) - P_1(B)] && (24) \\ &= ||P_2 - P_1||_{TV} && (25) \end{aligned}$$

Note: Total Variation distance also satisfies triangularity property.
Note: Total Variation distance is a valid distance.

2.2 l_1 distance

$$X = (x_1, x_2, ..., x_n) \qquad Y = (y_1, y_2, ..., y_n) \tag{26}$$

$$||X - Y||_{l_1} = \sum_{i=1}^{n} |x_i - y_i| \tag{27}$$

$$||P_1 - P_2||_{TV} = \frac{1}{2}||P_1 - P_2||_{l_1} \tag{28}$$

Example: Calculate the l_1 distance from the previous example.

$$||P_1 - P_2||_{l_1} = 0.4$$

$$||P_1 - P_2||_{TV} = \frac{1}{2}0.4 = 0.2$$

Proof $||P_1 - P_2||_{TV} = \frac{1}{2}||P_1 - P_2||_{l_1}$
Suppose $A \triangleq \{x \in \mathcal{X} : P_1(x) > P_2(x)\}$. We have :

$$\begin{aligned} ||P_1 - P_2||_{TV} &= P_1(A) - P_2(A) && (29) \\ &= \sum_{x \in A} P_1(x) - P_2(x) && (30) \\ &= \sum_{x \in A} |P_1(x) - P_2(x)| && (31) \end{aligned}$$

Also We have

$$\begin{aligned} ||P_1 - P_2||_{TV} &= P_1(A) - P_2(A) && (32)\\ &= 1 - P_1(A^c) - 1 + P_2(A^c) && (33)\\ &= P_2(A^c) - P_1(A^c) && (34)\\ &= \sum_{x \in A^c} P_2(x) - P_1(x) && (35)\\ &= \sum_{x \in A^c} |P_2(x) - P_1(x)| && (36) \end{aligned}$$

We know $x = a = b \rightarrow x = \frac{a+b}{2}$,So we have:

$$\begin{aligned} ||P_1 - P_2||_{TV} &= \frac{\sum_{x \in A} |P_1(x) - P_2(x)| + \sum_{x \in A^c} |P_1(x) - P_2(x)|}{2} && (37)\\ &= \frac{1}{2} \sum_{x \in \mathcal{X}} |P_1(x) - P_2(x)| = \frac{1}{2} ||P_1 - P_2||_{l_1} && (38) \end{aligned}$$

2.3 Pinsker's Inequality

$$\frac{2}{\ln 2} ||P_1 - P_2||_{TV}^2 \leq D(P_1||P_2) \ (Pinsker's\ Inequality) \quad (39)$$

$$||P_1 - P_2||_{TV}^2 = \frac{1}{4} ||P_1 - P_2||_{l_1}^2 \quad (40)$$

$$D(P_1||P_2) \geq \frac{1}{2 \ln 2} ||P_1 - P_2||_{l_1}^2 \quad (41)$$

Example:

$$\begin{aligned} &1.P(x = 1) = p && \{p, 1 - p\}\\ &2.P(x = 1) = p + \epsilon && \{p + \epsilon, 1 - p - \epsilon\}\\ &D(p||p + \epsilon) \geq \frac{1}{2 \ln 2} 4\epsilon^2 = \frac{2\epsilon^2}{\ln 2} \end{aligned}$$