COMPSCI 650 Applied Information Theory Feb 11, 2016

Lecture 7

Instructor: Arya Mazumdar *Scribe: Names Redacted*

1 Review

For Bernoulli variables in the sequence (p = p(1), 1 - p = p(0)),

$$P(\#\ of\ 1's\ in\ the\ sequence = qn) \sim \frac{1}{\sqrt{n}} 2^{-nD(q||p)} \tag{1}$$

When $q = p$, the probability of qn $1's$ in the sequence is $\frac{1}{\sqrt{n}}$.

$$\begin{aligned}
& P[|\#\ of\ 1's\ in\ the\ sequence - pn| > \epsilon_n] \\
& = \sum_{j>(p-\epsilon)n, j<(p+\epsilon)n} \frac{1}{\sqrt{n}} 2^{-nD(\frac{j}{n}||p)} \\
& \leq \frac{n}{\sqrt{n}} 2^{-nD((p\pm\epsilon)||p)} \\
& \approx \sqrt{n} 2^{-nD((p\pm\epsilon)||p)} < \delta \\
\Rightarrow \quad & 2^{-nD((p\pm\epsilon)||p)} < \frac{\delta}{\sqrt{n}} \\
\Rightarrow \quad & -nD((p\pm\epsilon)||p) < -\log\frac{\sqrt{n}}{\delta} \\
\Rightarrow \quad & D((p\pm\epsilon)||p) > \frac{\log\frac{\sqrt{n}}{\delta}}{n}
\end{aligned} \tag{2}$$

2 Fano's Inequality

2.1 An Example For Fano's Inequality

X: the temperature (in F) in Boston
Y: the temperature (in F) in New York
$\hat{X}$: the estimated temperature (in F) in Boston
$\hat{Y}$: the estimated temperature (in F) in New York
$X \to Y \to \hat{X}$ is a Markov Chain because $\hat{X}$ is only dependent on Y, and thus $P(\hat{X}|Y,X) = P(\hat{X}|Y)$ and $\hat{X} \neq X$

2.2 Fano's Inequality

For any estimator such that $X \to Y \to \hat{X}$:

$$P[\hat{X} \neq X] \geq \frac{H(X|\hat{X} - 1)}{\log|\mathcal{X}|}, \tag{3}$$

where $\mathcal{X}$ is the sample space of X.

3 Convex Functions and Concave Function

3.1 Convex Functions

Definition: a function is convex on an interval if for any two points and in anywhere:

$$
\begin{aligned}
&0 \le t \le 1 \\
&z = tx_1 + (1-t)x_2
\end{aligned} \tag{4}
$$

Function f() is convex if:

$$f(tx_1 + (1 = t)x_2) \le tf(x_1) + (1-t)f(x_2), \forall 0 \le t \le 1 \tag{5}$$

Also this definition is illustrated in figure 1.

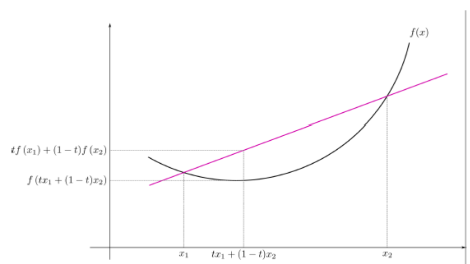

Figure 1: Definition of convex function

3.2 Concave Functions

Function f() is concave if:

$$f(tx_1 + (1 = t)x_2) \ge tf(x_1) + (1-t)f(x_2), \forall 0 \le t \ge 1 \tag{6}$$

4 Jensen's Inequality

4.1 Jensen's Inequality

Theorem: If f() is a convex function and X is a random variable then $E[f(x)] \ge f(E[X])$.

4.2 An Example

$f(x) = x^2$ is convex function. Therefore:

$$
\begin{aligned}
&E[f(x)] \ge f(E[X]) \\
\Rightarrow &E[x^2] \ge E[x]^2 \\
\Rightarrow &Var[X] = E[x^2] - E[x]^2 \ge 0
\end{aligned} \tag{7}
$$

4.3 Proof

Induction proof:

1. If $x = \{x_1, x_2\}$, $E[f(x)] = p_1 f(x_1) + p_2 f(x_2)$, $f(E(x)) = f(p_1 x_1 + p_2 x_2)$. From the definition of convex function: $p_1 f(x_1) + p_2 f(x_2) \geq f(p_1 x_1 + p_2 x_2)$, so $E[f(x)] \geq f(E(x))$ for $x \in \{x_1, x_2\}$.

2. Assume $E[f(x)] \geq f(E(x))$ for k points is true, then for $k+1$ points:

$$\begin{aligned}
E[f(X)] &= \sum_{i=1}^{k+1} f(x_i)p(x_i) \\
&= \sum_{i=1}^{k} f(x_i)p(x_i) + f(x_{k+1})p(x_{k+1}) \\
&= \sum_{i=1}^{k} p(x_i)(\sum_{i=1}^{k} f(x_i)\frac{p(x_i)}{\sum_{i=1}^{k} p(x_i)}) + f(x_{k+1})p(x_{k+1}) \\
&\geq (1 - p(x_{k+1}))f\left(\sum_{i=1}^{k} x_i \frac{p(x_i)}{\sum_{i=1}^{k} p(x_i)}\right) + f(x_{k+1})p(x_{k+1}) \\
&\geq f\left((1 - p(x_{k+1}))\frac{\sum_{i=1}^{k} x_i p(x_i)}{\sum_{i=1}^{k} x_i p(x_i)} + f(x_{k+1})p(x_{k+1})\right) \\
&= f(\sum_{i=1}^{k+1} x_i p(x_i)) \\
&= f(E[X])
\end{aligned} \tag{8}$$

5 Chain Rule For Mutual Information

5.1 Recall: Chain Rule

$$\begin{aligned}
&H(X|Y) = H(X) + H(Y|X) \\
&H(X_1, X_2, ..., X_n) = H(X_1) + H(X_2|X_1) + H(X_3|X_1, X_2) + \cdots + H(X_n|X_1, X_2, \ldots, X_n)
\end{aligned} \tag{9}$$

$$\begin{aligned}
I(X_1, X_2, \ldots, X_n; Y) =& H(X_1, X_2, \ldots, X_n) - H(X_1, X_2, \ldots, X_n|Y) \\
=& H(X_1) + H(X_2|X_1) + H(X_3|X_1, X_2) + \cdots + H(X_n|X_1, X_2, \ldots, X_{n-1}) \\
& - H(X_1|Y) - H(X_2|X_1, Y) - H(X_3|X_1, X_2, Y) - \cdots - H(X_n|X_1, X_2, \ldots, X_{n-1}, Y) \\
=& I(X_1; Y) + I(X_2; Y|X_1) + I(X_3; Y|X_1, X_2) + \cdots + I(X_n; Y|X_1, X_2, ..., X_{n-1})
\end{aligned} \tag{10}$$

6 Data Processing Inequality

6.1 Data Processing Inequality

Theorem: $X \to Y \to \hat{X}$ is a Markov Chain, then

$$I(X; Z) \leq I(X; Y) \tag{11}$$

6.2 Proof

From the chain rule:

$$\begin{aligned}
& I(X;Y,Z) = I(Y,Z;X) = I(Y;X) + I(Z;X|Y) \\
Since \quad & I(Z;X|Y) = 0 \\
thus \quad & I(X;Y,Z) = I(Y;X) \\
And \quad & I(Y,Z;X) = I(Z;X) + I(Y;X|Z) \\
so \quad & I(Y;X) = I(Z;X) + I(Y;X|Z) \geq 0 \\
& \Rightarrow I(Y;X) \geq I(Z;X)
\end{aligned} \tag{12}$$