CIE6020/MAT3350
Selected Topics in Information Theory

Lecture 3: Inequalities

17 Janurary 2019

The Chinese University of Hong Kong, Shenzhen

Basic Inequalities

Lemma (Fundamental Inequality)
For any $a > 0$,

$$\ln a \leq a - 1$$

with equality if and only if $a = 1$.

Lemma (Log-sum inequality)
For arbitrary non-negative numbers a_i, b_i, $i = 1, \ldots, n$ we have

$$\sum_{i=1}^{n} a_i \log \frac{a_i}{b_i} \geq a \log \frac{a}{b}$$

where $a = \sum_i a_i$ and $b = \sum_i b_i$. The equality holds iff $a_i b = b_i a$ for $i = 1, \ldots, n$.

Proof.

- We may assume that a_i are positive and b_i are positive.
- Further, it is sufficient to prove the lemma for $a = b$.
- For this case, the statement becomes $\sum_i a_i \log \frac{b_i}{a_i} \leq 0$ and follows from the inequality $\ln x \leq x - 1$.

$\square$

Theorem (Information inequality)
Let p and q be two PMF over the same alphabet. Then

$$D(p||q) \geq 0$$

with equality iff $p(x) = q(x)$ for all x.

Corollary
For any two random variables X and Y,

$$I(X;Y) \geq 0$$

with equality iff X and Y are independent.

Corollary
For any two random variables X and Y,

$$I(X;Y) \geq 0$$

with equality iff X and Y are independent.

Corollary

$$I(X;Y|Z) \geq 0,$$

with equality iff X and Y are conditional independent given Z.

Theorem
$H(X) \leq \log|\mathcal{X}|$ where $\mathcal{X}$ is the alphabet of X. The equality holds iff X is uniformly distributed on $\mathcal{X}$.

Proof.
Let u be the uniform distribution on $\mathcal{X}$, i.e., $u(x) = |\mathcal{X}|^{-1}$, $x \in \mathcal{X}$. Then

$$\begin{aligned}
\log|\mathcal{X}| - H(X) &= \sum_x p(x) \log|\mathcal{X}| + \sum_x p(x) \log p(x) \\
&= \sum_x p(x) \log 1/u(x) + \sum_x p(x) \log p(x) \\
&= \sum_x p(x) \log \frac{p(x)}{u(x)} \\
&\geq 1 \cdot \log \frac{1}{1} \qquad (1) \\
&= 0,
\end{aligned}$$

where inequality follows from the log-sum inequality. The equality in (1) holds if and only $p(x) = u(x)$ for all $x \in \mathcal{X}$. $\square$

Theorem (Conditioning reduces entropy)

$$H(X|Y) \leq H(X)$$

with equality iff X and Y are independent.

Theorem (Independence bound on entropy)

$$H(X_1, X_2, \ldots, X_n) \leq \sum_{i=1}^{n} H(X_i)$$

with equality iff X_i are independent.

Convextiy of Information Measures

Theorem (Convexity of relative entropy)
$D(p||q)$ *is convex in the pair* (p, q).

Theorem (Convexity of relative entropy)
$D(p||q)$ *is convex in the pair* (p, q).

Proof.
The theorem says that if (p_1, q_1) and (p_2, q_2) are two pairs of probability mass functions, then

$$D(\lambda p_1 + \bar{\lambda} p_2 || \lambda q_1 + \bar{\lambda} q_2) \leq \lambda D(p_1||q_1) + \bar{\lambda} D(p_2||q_2)$$

for all $0 \leq \lambda \leq 1$. $\square$

Theorem (Concavity of entropy)
$H(p)$ *is a concave function of* p.

Theorem (Concavity of entropy)
$H(p)$ is a concave function of p.

Proof.
$H(p) = \log |\mathcal{X}| - D(p||u)$ $\square$

where u is the uniform distribution on $|\mathcal{X}|$.

Theorem
The mutual information $I(X;Y)$ is a concave function of $p(x)$ for fixed $p(y|x)$ and a convex function of $p(y|x)$ for fixed $p(x)$.

Theorem
The mutual information $I(X;Y)$ is a concave function of $p(x)$ for fixed $p(y|x)$ and a convex function of $p(y|x)$ for fixed $p(x)$.

Proof.
To prove the first part, we write

$$I(X;Y) = H(Y) - H(Y|X) = H(Y) - \sum_x p(x) H(Y|X = x).$$

Theorem
The mutual information $I(X;Y)$ is a concave function of $p(x)$ for fixed $p(y|x)$ and a convex function of $p(y|x)$ for fixed $p(x)$.

Proof.
To prove the first part, we write

$$I(X;Y) = H(Y) - H(Y|X) = H(Y) - \sum_x p(x)H(Y|X = x).$$

To prove the second part, we use $I(X,Y) = D(p(x,y)||p(x)p(y))$, where the latter is a convex function of $(p(x,y), p(x)p(y)) = (p(x)p(y|x), p(x)p(y))$, which is a linear function of $p(y|x)$ when $p(x)$ is fixed. $\square$

Data-Processing Inequality

Markov Chain

Definition (Markov chain)
Random variables X, Y and Z form a Markov chain $X \to Y \to Z$ if

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

1. $X \to Y \to Z$ iff X and Z are conditional independent given Y.
2. $X \to Y \to Z$ iff $I(X; Z|Y) = 0$.
3. $X \to Y \to Z$ implies $Z \to Y \to X$.

Theorem

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

with equality iff $X \to Y \to Z$ *forms a Markov chain.*

Theorem (Data processing inequality)
If $X \to Y \to Z$, then $I(X;Y) \geq I(X;Z)$.

Theorem (Data processing inequality)
If $X \to Y \to Z$, then $I(X;Y) \geq I(X;Z)$.

Proof.
Using the chain rule of mutual information to expend $I(X;Y|Z)$ in two different ways:

$$\begin{aligned} I(X;Y,Z) &= I(X;Z) + I(X;Y|Z) \\ &= I(X;Y) + I(X;Z|Y). \end{aligned}$$

See that $I(X;Z|Y) = 0$ and $I(X;Y|Z) \geq 0$. $\square$

Corollary
$I(X;Y) \geq I(X;g(Y))$, *where* g *is any function.*

Corollary
$I(X;Y) \geq I(X;g(Y))$, *where* g *is any function.*

Corollary
If $X \to Y \to Z$, *then* $I(X;Y|Z) \leq I(X;Y)$.

Information Diagram for Three Random Variables