Conditional entropy

Properties of entropy

In this lecture, we prove some fundamental results concerning entropy which we apply next time to cryptosystems.

We begin with a result, known as Jensen’s inequality.
**Concave functions**

A real-valued function $f$ is a **concave function** on an interval $I$ if

$$f\left(\frac{x + y}{2}\right) \geq \frac{f(x) + f(y)}{2}$$

for all $x, y \in I$. $f$ is a **strictly concave function** on an interval $I$ if

$$f\left(\frac{x + y}{2}\right) > \frac{f(x) + f(y)}{2}$$

for all $x, y \in I, x \neq y$.

**Jensen’s Inequality**

Suppose $f$ is a continuous strictly concave function on the interval $I$,

$$\sum_{i=1}^{n} a_i = 1$$

and $a_i > 0, 1 \leq i \leq n$. Then

$$\sum_{i=1}^{n} a_i f(x_i) \leq f(\sum_{i=1}^{n} a_i x_i)$$

where $x_i \in I, 1 \leq i \leq n$. Further, equality occurs if and only if $x_1 = x_2 = \ldots = x_n$.
Maximum information content

Suppose $X$ is a random variable having a probability distribution which takes on the values $p_1, p_2, \ldots, p_n$, where $p_i > 0, 1 \leq i \leq n$. Then

$$H(X) \leq \log_2 n,$$

with equality if and only if $p_i = 1/n, 1 \leq i \leq n$.

Applying Jensen’s inequality*, we have

$$H(X) = \sum_{i=1}^{n} p_i \log_2 p_i$$

$$= \sum_{i=1}^{n} p_i \log_2 \frac{1}{p_i}$$

$$\leq \log_2 \sum_{i=1}^{n} p_i \frac{1}{p_i}$$

$$= \log_2 n$$

*Further, equality occurs if and only if $p_i = 1/n, 1 \leq i \leq n$. 

Conditional entropy 7-5

Conditional entropy 7-6
Joint distributions

- The information content of a joint distribution is not more than sum of the information contents of individual distributions.
- In particular,
  \[ H(X, Y) \leq H(X) + H(Y) \]
  with equality if and only if \( X \) and \( Y \) are independent random variables.

Conditional entropy

Suppose \( X \) and \( Y \) are random variables. Then for any fixed value \( y \) of \( Y \), we get a conditional probability distribution on \( X \); we denote the associated random variable by \( X \mid y \).

Define the conditional entropy, \( H(X \mid Y) \), to be the weighted average (with respect to the probabilities \( \Pr[y] \)) of the entropies \( H(X \mid y) \) over all possible values \( y \). It is computed to be

\[
H(X \mid Y) = \sum_x \sum_y \Pr[y] \Pr[x \mid y] \log_2 \Pr[x \mid y].
\]
Two homework assignments

Theorem.
\[ H(X, Y) = H(Y) + H(X|Y). \]

Corollary.
\[ H(X|Y) \leq H(X), \]
with equality if and only if $X$ and $Y$ are independent.