1 Expectation

Assume $\Omega$ is a sample space

$$\Omega = \{x_1, x_2, x_3, \ldots x_n\}$$

Assume $p$ is a probability distribution over $\Omega$. Assume $f$ is a function from $\Omega$ to real numbers:

$$f \in \mathbb{R}^\Omega$$

We define the expected value of $f$:

$$E(f) = p(x_1)f(x_1) + p(x_2)f(x_2) + p(x_3)f(x_3) + \ldots + p(x_n)f(x_n)$$

$$E(f) = \sum_{x \in \chi} p(x_i)f(x_i)$$

2 Entropy

We assume a random variable $X$ defined on an alphabet of symbols $\chi$ with pmf $p$. So the kinds of events we are now interested in are symbol occurrences. And we assume that the information measure of each symbol $x$, $x \in \chi$ is:

$$I(x) = -\log p(x)$$

We have:

$$E(I) = \sum p(x)(-\log p(x)) = -\sum p(x) \log p(x)$$

We call $E(I)$ the entropy of random variable $X$. It usually written $H$:

$$H(X) = -\sum_{x \in \chi} p(x) \log p(x)$$

It is often written directly as a function of the pmf $p$:

$$H(p) = -\sum_{x \in Dom(p)} p(x) \log p(x)$$

3 Entropy of Joint and Conditional Distributions

The entropy of a joint distribution is the same as the entropy of a distribution of a single random variable, except that the elements of the sample space are pairs:

$$H(X, Y) = -\sum_{x \in \chi} p(x, y) \log p(x, y)$$
Conditional entropy is somewhat different:

\[ H(Y|X) = \sum_{x \in \chi} p(x)H(p(Y|X = x)) \]

Recall that \( p(Y |X) \) is not a real pmf. What we do is sum up the entropies of each of the conditional pmfs of the form \( p(Y | X=x) \), weighting each by the probability of \( x \)

\[ H(Y|X) = \sum_{x \in \chi} p(x)[- \sum_{y \in \chi} p(y|x) \log p(y|x)] \]

Multiplying \( p(x) \) into the inner sum:

\[ H(Y|X) = - \sum_{x \in \chi} \sum_{y \in \chi} p(x)p(y|x) \log p(y|x) \]

Using the Chain rule:

\[ H(Y|X) = - \sum_{x \in \chi} \sum_{y \in \chi} p(x, y) \log p(y|x) \]

It turns out this is a rather natural definition of \( H(X|Y) \) because it leads to a chain rule for entropy (proof, p. 64 of text):

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