Ralph's mutual information

Ralph knows Shannon developed a powerful communication (information) theory around a concept of entropy (roughly, a measure of uncertainty). Shannon's entropy measure is axiomatic

1. a measure $H(p)$ exists,
2. the measure is smooth,
3. the measure is monotonically increasing in uncertainty,
4. the measure is consistent in the sense that if different measures exist they lead to the same conclusion, and
5. the measure is additive.

The measure implied by the five conditions is[1]

$$H(p) = -\sum_i p_i \log p_i$$

where p_i is the probability belief associated with event i and $\sum_i p_i = 1$ for all $p_i \geq 0$.

Additivity implies

$$H(X,Y) = H(Y) + H(X \mid Y) = H(X) + H(Y \mid X)$$

where

$$H(X,Y) = -\sum_{i,j} p(x_i, y_j) \log p(x_i, y_j)$$

is **joint entropy** for the joint probability distribution over random variables X and Y (to simplify notation, probability is left implicit),

$$H(Y) = -\sum_j p(y_j) \log p(y_j)$$

$$H(X) = -\sum_i p(x_i) \log p(x_i)$$

refer to **entropy** for the marginal probability distributions over random variables Y and X, respectively, and

$$H(X \mid Y) = \sum_j p(y_j) H(X \mid Y = y_j)$$

$$= -\sum_j p(y_j) \sum_i p(x_i \mid y_j) \log p(x_i \mid y_j)$$

$$= -\sum_{i,j} p(x_i, y_j) \log p(x_i \mid y_y)$$

[1] As logarithms are monotone transformations, any base suffices. Shannon focused on bits and consequently worked with base two. We will typically employ natural logarithms.

is **conditional entropy** ($H(Y \mid X)$ is defined analogously). Additivity is akin to Bayes product rule

$$p(x_i, y_j) = p(x_i \mid y_j)\, p(y_j) = p(y_j \mid x_i)\, p(x_i)$$

but since we're working with logarithms it is additive.

Conditional entropy indicates uncertainty remaining say in X given information Y. **Mutual information** follows from conditional entropy.

$$I(X;Y) = H(X) - H(X \mid Y)$$

$$= H(X) - [H(X,Y) - H(Y)]$$

$$= H(X) + H(Y) - H(X,Y)$$

That is, mutual information is a measure of reduction in uncertainty due to information.

Further, **relative entropy** measures the disagreement between two distributions. Suppose we measure the average uncertainty is a reference random variable X by its entropy $H(X) = -\sum_j p(x_j) \log p(x_j)$ and hypothesize another random variable Y has the same distribution. Its average uncertainty is $-\sum_j p(x_j) \log p(y_j)$. The difference between these two is

$$-\sum_j p(x_j) \log p(y_j) - \left[-\sum_j p(x_j) \log p(x_j) \right]$$

$$= \sum_j p(x_j) \log p(x_j) - \sum_j p(x_j) \log p(y_j)$$

$$D_{KL}(X \parallel Y) = \sum_j p(x_j) \log \frac{p(x_j)}{p(y_j)}$$

This measure is called relative entropy or Kullback-Leibler divergence.

Since mutual information gauges the extent to which the (joint) distribution of dependent random variables differs from independent random variables, mutual information are relative entropy are intimately related.

$$D_{KL}(p(X,Y) \parallel p(X)\, p(Y)) = \sum_{i,j} p(x_i, y_j) \log \frac{p(x_i, y_j)}{p(x_i) p(y_j)}$$

$$= -\sum_{i,j} p(x_i, y_j) \log p(x_i) + \sum_{i,j} p(x_i, y_j) \log p(x_i \mid y_j)$$

$$= H(X) - H(X \mid Y) = I(X;Y)$$

Remarkably, Kelly's mutual information theorem gives mutual information economic content. Kelly's theorem says that if a decision maker maximizes long-run wealth (or equivalently, maximizes expected logarithmic returns) then mutual information of states and information identifies the expected gain from the information in logarithmic returns.

Conditions for the theorem are that a full complement of Arrow-Debreu assets (assets that pay off in exactly one state and zero in all other states) exist and there are no arbitrage opportunities (all state prices or Arrow-Debreu prices are positive). With these conditions in hand along with a long-run wealth maximization objective, Kelly's investment strategy is "to bet your beliefs." In other words, the fraction of wealth k_j ($\sum_j k_j = 1$) invested in the Arrow-Debreu asset that pays in state j equals your probability belief associated with state j, $k_j = p_j$. In equilibrium (Ross' recovery theorem), expected logarithmic returns from a Kelly strategy are equal to the riskless logarithmic return (riskless implies returns are the same across all states). Therefore, to get ahead an economic agent gathers information and expected (logarithmic) gains from the information are given by mutual information.

Ralph is considering the following investment options expressed in returns form (this implies the investment cost is normalized to one for each Arrow-Debreu asset).

$$A = \begin{bmatrix} & state_1 & state_2 \\ asset_1 & 2 & 0 \\ asset_2 & 0 & 2 \end{bmatrix}$$

States are uniformly distributed and state prices y solve $Ay = v$ where v is a vector of ones (the normalized investment cost).

Ralph can acquire information Z where the joint distribution of states S and information Z is

$$\begin{bmatrix} & s_1 & s_2 \\ z_1 & 0.4 & 0.1 \\ z_2 & 0.1 & 0.4 \end{bmatrix}$$

Suggested:

1. Determine state prices y.

2. Determine expected logarithmic returns, $E[r]$, from a Kelly investment strategy without information.

3. Determine $H(S)$, $H(Z)$, $H(S,Z)$, $H(S \mid Z)$, $I(S;Z)$, and $D_{KL}(p(S,Z) \parallel p(S)\, p(Z))$.

4. Determine expected logarithmic returns given information Z, $E[r \mid Z] = p(Z = z_1)\, E[r \mid z_1] + p(Z = z_2)\, E[r \mid z_2]$ from a Kelly investment strategy.

5. Compare the expected gain from the information, $E[gain] = E[r \mid Z] - E[r]$, with mutual information $I(S;Z)$.