Shannon Entropy

Definition. Let (X, p) be a discrete probability space. The *entropy* of X is

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

The choice of logarithm base determines the unit: bits, digits, nats, etc. The *joint entropy* of spaces X and Y is then $H(X,Y) = -\sum_{x,y} p(x,y) \log p(x,y)$, and the *conditional entropy* of Y given X is

$$H(Y \mid X) = \mathbb{E}_X[H(Y)] = -\sum_{x,y} p(x,y) \log p(y \mid x).$$

We then have $H(X,Y) = H(X) + H(Y \mid X)$.

Consider a sequence of probability spaces $X_1, \ldots, X_k$ with common "alphabet" X and respective p.m.f.'s $p_1, \ldots, p_k$. Denote the ith joint space $(X_1, \ldots, X_i)$ by Y_i. Then we have

$$H(Y_k) = H(X_1) + \sum i = 2^k H(X_i \mid Y_{i-1}).$$

The infinite joint space $Y = (X_1, \ldots)$ is then said to have per-symbol entropy

$$H(Y) = \lim_{k \to \infty} \frac{H(Y_k)}{k},$$

provided this limit exists. In the special case when all allowed words of length k are equiprobable, we know that $H(Y_k) = \log |Y_k|$, so the limit is

$$H(Y) = \lim_{k \to \infty} \frac{\log |Y_k|}{k}.$$

For a sequence of events satisfying certain properties

Fractal Dimension

Definition. The *Box Dimension* (or Minkowski-Bouligand dimension) of a set $S \in \mathbb{R}^2$ is the limit

$$D_{\text{box}}(S) = \lim_{\epsilon \to 0} \frac{-\log N(\epsilon)}{\log \epsilon},$$

(provided the limit exists) where $N(\epsilon)$ is the number of square boxes of side length ϵ required to cover the set. For any finite sequence $\epsilon_0 > \cdots > \epsilon_n$, the slope of the best-fit line through the points $(-\log \epsilon_k, \log N(\epsilon_k))$ is an approximation to D_{box}.

Notice in particular that, for any $r > 0$, we must have

$$D_{\text{box}}(S) = \lim_{k \to \infty} \frac{-\log N(r^{-k})}{\log r^{-k}} = \lim_{k \to \infty} \frac{\log_r N(r^{-k})}{k}.$$

Accordingly, we define the *entropy of a fractal* to be

$$H(S) = D_{\mathrm{box}}(S) \text{ bits.}$$

This is the (limiting) amount of information produced each time we "zoom in" by a factor of two.

Image Analysis

Definition. An *image* is a real-valued function defined on a region in $\mathbb{R}^2$.

Often, our information about an image is limited to its (perhaps approximate) values at a set of regularly spaced lattice points on a rectangular region of the domain. It is then convenient to represent this grid of values in a matrix, called the *image matrix* (or simply the image, if the meaning is clear).

Definition. The *fractal spectrum* of an image I over a set of threshold values, $S \in \mathbb{R}$, is the function $\mathrm{FS} : S \to \mathbb{R}$ such that, for $t \in S$,

$$\mathrm{FS}(t) = D_{\mathrm{box}}(I_t^{-1}), \quad \text{where } I_t^{-1} = \{x : |I(x)| \leq t\}.$$

For an image matrix we take an approximation to D_{box}. If $S = \{t_1, \ldots, t_m\}$ is finite, we may write $\mathrm{FS}(I) = [D_{\mathrm{box}}(t_1), \ldots, D_{\mathrm{box}}(t_m)]^T \in \mathbb{R}^m$, where typically $t_1 < \cdots < t_m$.

Definition. Let I be an image matrix and let I' be a copy of I with scrambled entries (the entries of I' are a uniform random permutation of the entries of I). For $t \in \mathbb{R}$, the *fractal excess* of I at t is the expectation

$$\mathrm{fe}(t) = \mathbb{E}\left[\mathrm{FS}_{M'}(t) - \mathrm{FS}_M(t)\right].$$

The fractal excess FE of I over a threshold set S is simply the restriction of fe to S; again, for finite S, we may write FE as a vector in $\mathbb{R}^m$.