Entropy

$$H = \sum_i p_i \log_2 \frac{1}{p_i}$$

Conditional Entropy

$$H(X \mid Y) = \sum_{x,y} p(x,y) \log \frac{1}{p(x \mid y)}$$

Mutual Information

$$I(X;Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$$

Independence Bound on Entropy

$$H(X_1, \ldots, X_n) \leq \sum_i H(X_i)$$

Fano’s Inequality

$$P_e \geq \frac{1 - H(X \mid Y)}{\log |X|}$$

Data Processing Inequality

If $X \to Y \to Z$ for some transformation $\to$, then $I(X;Y) \geq I(X;Z)$.

Mutual Information Distance

$$D(X,Y) = H(X,Y) - I(X;Y)$$

Kullback-Leibler Distance

$$D_{KL}(p \parallel q) = \sum_{x} p(x) \log \frac{p(x)}{q(x)}$$

Markov Process Entropy

Average of the entropy of each state weighted by occupancy probability.

Code Rate

R, weighted average cost of sending a symbol in bits/symbol.

Average Codeword Length

Same as Code Rate: bits per symbol.

Efficiency of a code

Information content of each bit: $\eta = \dfrac{H}{R}$

Prefix Property

No codeword is the prefix of a longer codeword: equivalent to unique decodability

Relationship between uniquely decodable and instantaneous

Instantaneous implies Uniquely Decodable

Shannon's Source-Coding Theorem

For a source with entropy H and $\varepsilon > 0$, there's a uniquely decodable code such that $R = H + \varepsilon$ as $\varepsilon \to 0$

Huffman Code

Optimal prefix code: binary tree, combine nodes and sum probabilities.

Kraft-McMillan Inequality

$$\sum_{i=1}^{N} \frac{1}{2^{c_i}} \leq 1 \text{ is necessary but not sufficient for instantaneous.}$$

Channel Matrix

$p(y_j \mid x_i)$, probability of getting output y_j given input x_i.

Probability of getting a symbol y from a channel

$$\sum_{x} p(y \mid x) p(x)$$

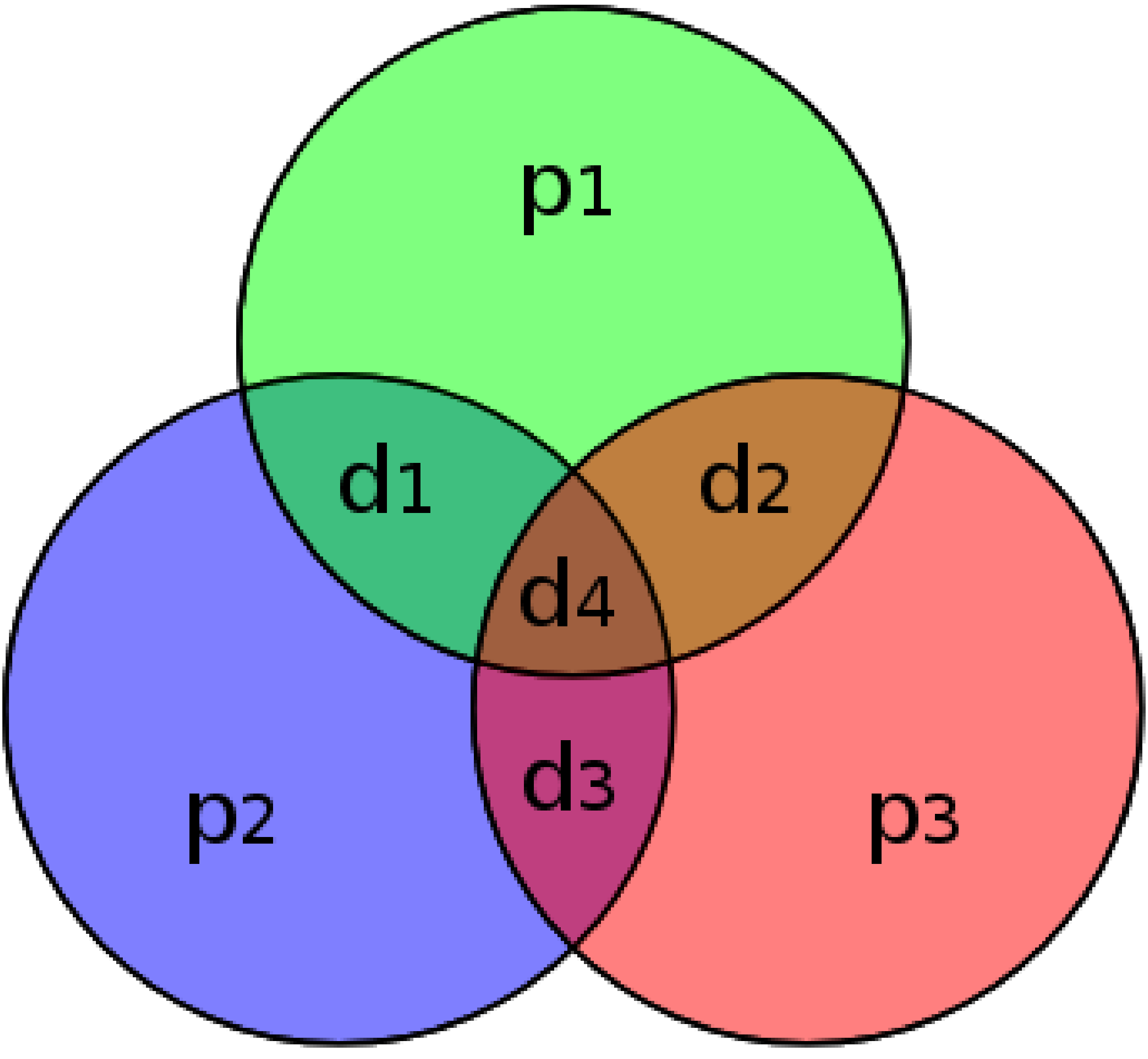
Channel Capacity

Maximum mutual information over all possible input distributions

Shannon's Channel-Coding Theorem

For a channel with capacity C and symbol source with entropy $H \leq C$, there's a coding scheme such that the source is reliably transmitted with arbitrarily small error.

7/4 Hamming Code


p1
d1
d2
d4
d3
p2
p3

Perfect Error Correcting Code

Code using m bits to correct $2^m - 1$ **error patterns**.

Fourier Series

$$\frac{a_0}{2} + \sum_{n=1}^{\infty} \left[a_n \cos(nx) + b_n \sin(nx) \right]$$

$$\left\{ \frac{1}{\sqrt{2}},\ \sin(x),\ \cos(x),\ \sin(2x),\ \cos(2x),\ \ldots \right\}$$

$$a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos(nx)\, dx, \qquad n = 0, 1, 2, \ldots$$

$$b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin(nx)\, dx, \qquad n = 1, 2, 3, \ldots$$

Lowpass, bandpass, highpass

Filter low, range, high frequencies

Distribution maximising channel entropy for fixed variance σ^2?

Gaussian: also maximises mutual information, so defines the channel capacity.

Maximum entropy of a Gaussian

$$H = \frac{1}{2}\log(2\pi e\sigma^2)$$

Capacity of an AWGN channel with spectral power density N_0

$$C = \int_{\omega_1}^{\omega_2} \log(1 + \frac{P}{WN_0})\, d\omega$$

Capacity of a gaussian-noise channel

$$C = \int_{\omega_1}^{\omega_2} \log(1 + SNR(\omega))\, d\omega$$

Power Spectral Density

$$N_0$$

Signal-to-Noise Ratio

$$\frac{P}{WN_0}$$

Fourier Transform

$$F(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt$$

Inverse Fourier Transform

$$f(t) = \int_{-\infty}^{\infty} F(\omega) e^{i\omega t} d\omega$$

Self-Fourier

Fourier transformation doesn't affect the form of the function.

Convolution

$$(f * g)(x) = \int_{-\infty}^{\infty} f(x - y)g(y)\,dy$$

Signal Modulation

Fourier transform of $f(t)e^{ict}$ is $F(\omega - c)$.

Nyquist's Sampling Theorem

Can completely reconstruct a function if we sample it at least $2W$.

Logan's Theorem

For a signal with $2W_L \geq W_H$, can reconstuct the function from just the zero-crossings.

Gabor's Information Diagram

Frequency vs Time, signal area is $> \dfrac{1}{4\pi}$.

Gabor Wavelet

Helical functions scaled by a Gaussian. Have optimal minimal area, but non-orthogonal.

RLE

Remove redundancy in strings of repeating values.

Predictive Coding

Record deviations from a prediction rather than actual sample values.

Dictionary Compression

Sparsity of language means number of useful strings is way fewer than address space.

LZW Algorithm

Dictionary compression algo, two passes.

Dyadic Wavelet

Mother wavelet $\Psi(x)$ which spawns daughters $\Psi_{jk}(x)$ which are transforms of the mother.

Kolmogorov Complexity

The length of the shortest program that can reproduce the string.

K-incompressible

Kolmogorov Complexity tends to the length of the string in the limit to ∞.