Vesa Linja-aho, Patric Östergård

Definition of entropy:

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log_2 p(x) \tag{1}$$

Expectation (expected value):

$$E_p g(X) = \sum_{x \in \mathcal{X}} g(x) p(x) \tag{2}$$

Joint entropy:

$$H(X, Y) = -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log p(x, y) \tag{3}$$

Conditional entropy:

$$H(Y|X) = \sum_{x \in \mathcal{X}} p(x) H(Y|X = x) = \sum_{x \in \mathcal{X}} p(x) \sum_{y \in \mathcal{Y}} -p(y|x) \log p(y|x) \tag{4}$$

Mutual information:

$$I(X; Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) \tag{5}$$

Chain rule for entropy:

$$H(X, Y) = H(X) + H(Y|X) \tag{6}$$

Use a Venn diagram to memorize (5) and (6). The definitions given so far generalize to more than two variables.

Kraft inequality:

$$\sum_{i=1}^{m} q^{-l_i} \leq 1 \tag{7}$$

Channel capacity:

$$C = \max_{p(x)} I(X; Y) \tag{8}$$

Capacity of binary symmetric channel (BSC):

$$C = 1 - H(p, 1 - p) \tag{9}$$

Capacity of Gaussian channel:

$$C = \frac{1}{2} \log \left(1 + \frac{P}{N}\right) \tag{10}$$

Capacity of band-limited Gaussian channel:

$$C = W \log \left(1 + \frac{P}{N_0 W}\right) \tag{11}$$