Midterm 1 Review

- Entropy: $H(X) = -\sum_{x\in\mathcal{X}} p(x)\log p(x)$ bits, ≥ 0
- Maximum entropy distribution
- Joint entropy: $H(X,Y)$
- Conditional entropy: $H(Y|X)$, conditioning reduces entropy $H(Y|X) \leq H(Y)$
- Chain rule for entropy: $H(X,Y) = H(X) + H(Y|X)$
- Relative entropy: $D(p||q) \geq 0$
- Mutual information: $I(X;Y) \geq 0$
- $D(p||q)$
  $\Rightarrow H(X) = \log|\mathcal{X}| - D(p||u)$
  $\Rightarrow I(X;Y) = H(X) - H(X|Y) = H(X) - H(X|Y)$, $I(X;Y) = D(p(x,y)||p(x)p(y))$
- Jensen's inequality: f convex, then $Ef(X) \geq f(EX)$.
- Data processing inequality
- Fano's inequality: $P_e \geq (H(X|Y) - 1)/\log|\mathcal{X}|$
- Law of large number for product of random variables: $\sqrt[n]{\prod_{i=1}^n X_i} \to e^{E\log X}$
- AEP: $-\frac{1}{n}\log p(x_1, \cdots, x_n) \approx H(X)$
- Entropy rate for Markov chain: $H(\mathcal{X}) = -\sum_{ij}\mu_i p_{ij}\log p_{ij}$
- Random walk on graph: $\mu_i = W_i/2W$
- Uniquely decodable code, instantaneous (prefix) codes
- Kraft inequality: $\sum_{i=1}^m D^{-l_i} \leq 1$
- Optimal code length: $l_i = -\log_D p_i$
- How to construct Huffman code