Information Theory with Applications

MATH 6397 – Fall 2008

September 10, 2008

Homework Set 1, due Tu Sep 23, 2008

Unless noted, exercises are taken from the textbook. Additional hints may be available there.

1. **p. 52 Ex. 2.3 a.** Let $Z : \Omega \to \mathbb{R}$ be a real-valued random variable and $\tau > 0$. Show the bound
$$\mathbb{P}(\{Z \geq a\}) \leq e^{-\tau a}\mathbb{E}[e^{\tau Z}] .$$

2. **Additivity of entropy.** Let X, Y, Z be three discrete-valued random variables. Assume the random vector (X, Y, Z) has equal probability $1/4$ for each outcome $(0, 0, 0)$, $(0, 1, 0)$, $(1, 0, 0)$ and $(1, 0, 1)$.
   (a) Write down six ways of splitting $H(X, Y, Z)$ into three terms involving conditional entropy.
   (b) Compute $H(X)$, $H(Y|X)$ and $H(Z|X, Y)$ in bits.
   (c) Conclude the value of $H(X, Y, Z)$ and verify this result by a direct way of computing $H(X, Y, Z)$.

3. **p. 53, Ex. 2.12 a.** Given three random variables X, Y, Z which map to an at most countable alphabet, and the usual definition of conditional entropy, show the triangle-type inequality
$$H(X|Z) \leq H(X|Y) + H(Y|Z) .$$

4. **p. 58 Ex. 2.28.** Let P and Q be probability vectors in $\mathbb{R}^n$, $P = (p_1, p_2, \ldots p_m)$ and $Q = (q_1, q_2, \ldots q_m)$. Define the Chernoff-distance for parameter $\alpha \geq 0$ by
$$C_\alpha(P, Q) = -\log\left(\sum_{j=1}^{m} p_j^\alpha q_j^{1-\alpha}\right) .$$

Show the following:

(a) $C_0(P,Q) = C_1(P,Q) = C_\alpha(P,P) = 0$.

(b) If $0 \le \alpha \le 1$, $C_\alpha(P,Q) \ge 0$.

(c) For fixed P and Q, C_α is a concave function of α.

(d) If $\alpha \ge 1$, $C_\alpha(P,Q) \le 0$.

(e) If $\alpha \notin \{0,1\}$ then $C_\alpha(P,Q) = 0$ is equivalent to $P = Q$.

(f) Let $P \neq Q$ and define a family of probability vectors $\{S_\alpha = (s_{\alpha,1}, \dots s_{\alpha,m})\}$ by

$$s_{\alpha,j} = \frac{p_j^\alpha q_j^{1-\alpha}}{\exp(-C_\alpha(P,Q))}$$

then $C_\alpha(P,Q)$ achieves its maximum value

$$C(P,Q) = \max_{\alpha \ge 0} C_\alpha(P,Q)$$

at α^* where $D(S_{\alpha^*}||P) = D(S_{\alpha^*}||Q)$.

5. **p. 63, Ex. 2.32.** Let X_1^n and X_2^n be vector-valued random variables with n independent, identically distributed components $\{X_{1,j}\}$ and $\{X_{2,j}\}$ that take discrete values in $\mathbb{A}^n$ according to probability measures P_1 and P_2, respectively. By observing the realization X^n distributed according to either measure P_1 or P_2, we wish to decide which measure is present.

Assume, in contrast to the discussion in class, we know that P_1 or P_2 are chosen before our experiment by the outcome of a random variable Θ which picks P_1 with probability $\pi_1 = \mathbb{P}(\Theta = 1)$ and P_2 with probability $\pi_2 = \mathbb{P}(\Theta = 2)$ to generate X^n. For each outcome $x \in \mathbb{A}^n$, we choose the maximum-likelihood estimate $k = \phi(x)$ which satisfies

$$\mathbb{P}(\Theta = k, X^n = x)/P(X^n = x) = \max_{j=1,2} P(\Theta = j, X^n = x)/P(X^n = x)\,.$$

Show that the asymptotic error probability for this estimate is

$$-\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\Theta \neq \phi(X^n)) = C(P_1, P_2)\,.$$