Problem Set 2

Issued: Monday 28th September, 2020 **Due:** Monday 12th October, 2020

Notations: We use Bern(p) to denote the Bernoulli distribution with the parameter p, and use Binom(n, p) to denote the binomial distribution with parameters n and p.

2.1. Please use Chain Rule for mutual information to derive $I(X_1, \ldots, X_n; Y_1, \ldots, Y_m)$.

2.2. *Conditional mutual information vs. unconditional mutual information.* Give examples of joint random variables X, Y, and Z such that

(a) $I(X;Y|Z) < I(X;Y)$.

(b) $I(X;Y|Z) > I(X;Y)$

2.3. *Information measures.* Suppose $Z_1, \ldots, Z_n$ are i.i.d. Bern$\left(\frac{1}{2}\right)$ random variables, and let $X_A \triangleq (Z_i)_{i \in A}$ be the random vector consisting of the bits with indices in A. Prove that

(a) For all non-empty $A \subset \{1, \ldots, n\}$, we have $H(X_A) = |A|$.

(b) For all non-empty $A_1, A_2 \subset \{1, \ldots, n\}$, we have

$$H(X_{A_1}, X_{A_2}) = |A_1 \cup A_2|, \tag{1a}$$

$$H(X_{A_1}|X_{A_2}) = |A_1 \setminus A_2|, \tag{1b}$$

$$I(X_{A_1}; X_{A_2}) = |A_1 \cap A_2|. \tag{1c}$$

2.4. Let (X, Y) be uniformly distributed in the unit l_p-ball $B_p \triangleq \{(x, y) : |x|^p + |y|^p \le 1\}$, where $p \in (0, \infty)$. Also define the l_∞-ball $B_\infty \triangleq \{(x, y) : |x| \le 1, |y| \le 1\}$.

(a) Are X and Y independent for $p = 1$?

(b) Compute $I(X;Y)$ for $p = \frac{1}{2}$, $p = 1$ and $p = \infty$.

(c) What do you think $I(X;Y)$ converges to as $p \to 0$. Explain it.

2.5. Let $\mathcal{N}(\boldsymbol{m}, \boldsymbol{\Sigma})$ be the Gaussian distribution on $\mathbb{R}^n$ with mean $\boldsymbol{m} \in \mathbb{R}^n$ and covariance matrix $\boldsymbol{\Sigma}$.

(a) Under what conditions on $\boldsymbol{m}_0, \boldsymbol{\Sigma}_0, \boldsymbol{m}_1, \boldsymbol{\Sigma}_1$ is

$$D\left(\mathcal{N}(\boldsymbol{m}_1, \boldsymbol{\Sigma}_1) \| \mathcal{N}(\boldsymbol{m}_0, \boldsymbol{\Sigma}_0)\right) < \infty \tag{2}$$

(b) Compute $D\left(\mathcal{N}(\boldsymbol{m}, \boldsymbol{\Sigma}) \| \mathcal{N}(0, \boldsymbol{I}_n)\right)$, where $\boldsymbol{I}_n$ is the $n \times n$ identity matrix.

(c) Compute $D\left(\mathcal{N}(\boldsymbol{m}_1, \boldsymbol{\Sigma}_1) \| \mathcal{N}(\boldsymbol{m}_0, \boldsymbol{\Sigma}_0)\right)$ for a non-singular $\boldsymbol{\Sigma}_0$.

2.6. There are two probability distribution P and Q over a finite alphabet $\mathcal{X}$ with cardinality k. Let us use $P_1 \ge P_2 \ge \cdots \ge P_k$ and $Q_1 \ge Q_2 \ge \cdots \ge Q_k$ to denote the non-increasing ordering of p.m.f P and Q respectively ($\sum_{i=1}^k P_i = \sum_{i=1}^k Q_i = 1$). We say that P is *more uniform* then Q if

$$\forall l \in [1:k], \sum_{i=1}^{l} P_i \le \sum_{i=1}^{l} Q_i \tag{3}$$

In this problem, we would like to prove that if P is more uniform then Q in the sense of (3), then

$$H(P) \ge H(Q) \tag{4}$$

(a) Prove that for convex function $f(\cdot)$, $\sum_{i=1}^{k} f(P_i) \leq \sum_{i=1}^{k} f(Q_i)$.

(b) Use (a) to prove (4)

2.7. *Total correlation.* For a given set of n random variables $X_1, \ldots, X_n$, the total correlation $C(X_1, \ldots, X_n)$ is defined as the K-L divergence from the joint distribution to the product distribution, i.e.,

$$C(X_1, \ldots, X_n) \triangleq D\left(P_{X^n} \middle\| \prod_{i=1}^{n} P_{X_i}\right).$$

(a) Prove that

$$C(X_1, \ldots, X_n) = \sum_{i=1}^{n} H(X_i) - H(X^n) \tag{5a}$$

$$= \sum_{i=1}^{n-1} I(X^i; X_{i+1}). \tag{5b}$$

(b) When will the total correlation be zero?

2.8. *Divergence of order statistics.* Given $x^n = (x_1, \ldots, x_n) \in \mathbb{R}^n$, let $x_{(1)} \leq \cdots \leq x_{(n)}$ denote the ordered entries. Let P, Q be distributions on $\mathbb{R}$ and $P_{X^n} = P^n, Q_{X^n} = Q^n$.

(a) Prove that

$$D(P_{X_{(1)} \ldots X_{(n)}} \| Q_{X_{(1)} \ldots X_{(n)}}) = nD(P \| Q). \tag{6}$$

(b) Show that

$$D(\mathrm{Binom}(n, p) \| \mathrm{Binom}(n, q)) = nD(\mathrm{Bern}(p) \| \mathrm{Bern}(q)). \tag{7}$$