Reading Assignment

- Theorem 1.2 in the Csiszár-Körner book (this can be derived from our Little-Big lemmas for hypothesis testing)

Homework Questions Questions marked * are difficult and will not be tested in the quiz.

Q1 The total variation distance is defined as

\[ d(P_0, P_1) = \sup_{A \subset \mathcal{X}} P_0(A) - P_1(A). \]

Prove the following equivalent forms of the total variation distance for discrete distributions \( P_0 \) and \( P_1 \):

\[
\begin{align*}
    d(P_0, P_1) &= \sup_{A \subset \mathcal{X}} P_1(A) - P_0(A) \\
    &= \sup_{A \subset \mathcal{X}} |P_0(A) - P_1(A)| \\
    &= \sum_{x \in \mathcal{X} : P_1(x) \geq P_0(x)} P_1(x) - P_0(x) \\
    &= \sum_{x \in \mathcal{X} : P_0(x) \geq P_1(x)} P_0(x) - P_1(x) \\
    &= \frac{1}{2} \sum_{x \in \mathcal{X}} |P_0(x) - P_1(x)|.
\end{align*}
\]

Q2 Prove the following properties of the total variation distance:

(i) \( 0 \leq d(P_0, P_1) \leq 1 \).

(ii) \( d(P_0, P_1) = 0 \) if and only if \( P_0 = P_1 \).

(iii) \( d(P_0, P_1) = 1 \) if and only if \( P_0 \) and \( P_1 \) have disjoint supports.

Q3 Consider the hypothesis testing problem of deciding if a coin is unbiased or shows heads with probability 0.51 by tossing it 10,000 times.

(i) Use the Little-Big lemmas for hypothesis testing to give bounds for \( \beta_{0.01}(P, Q) \).

(ii) How do your bounds compare with the asymptotic behaviour given by Stein’s Lemma. (You need to estimate the KL divergence between the two hypothesis to answer this question.)

(iii) Repeat your calculations for the case when you want to test a coin which only shows heads with probability 0.01.
Q4 Calculate the entropy of the following rvs:

(i) \( X \) is the output of \( n \) independent tosses of a coin which shows heads with probability \( p \);
(ii) \( X \sim \text{Geometric}(p) \);
(iii) Consider a channel \((X, W, Y)\) with input \( X = \mathbb{R} \) and output \( Y = \mathbb{N} \) such that for every input \( x \in X \), the output \( Y \) is distributed as \( \text{Poisson}(x) \). Calculate the entropy of \( Y \) which is the output corresponding to a random input \( X \) distributed as \( \text{exponential}(\lambda) \). [Hint: Compute the distribution of \( Y \). Good news! It does not turn out to be Poisson.]

Q5 Calculate the KL divergence between the following distributions.

(i) \( P \equiv \mathcal{N}(\mu_1, \sigma^2) \) and \( Q \equiv \mathcal{N}(\mu_2, \sigma^2) \);
(ii) \( P \equiv \text{Poisson}(\lambda_1) \) and \( Q \equiv \text{Poisson}(\lambda_2) \);
(iii) \( P \equiv \text{Geometric}(p) \) and \( Q \equiv \text{Geometric}(q) \).

Q6 This question concerns total variation distance and KL divergence between two Bernoulli distributions.

(i) Calculate the total variation distance between \( P \equiv \text{Bernoulli}(p) \) and \( Q \equiv \text{Bernoulli}(q) \).
(ii)* For \( P \) and \( Q \) above, show that \( D(P\|Q) \geq \frac{2}{\ln 2} (p - q)^2 \).

Q7* Consider the \( M \)-ary hypothesis testing problem where you wish to decide the bias of the coin by observing \( n \) independent tosses from it, i.e., \( P_i \) corresponds to \( n \) i.i.d. samples from \( \text{Bernoulli}(p_i) \), \( 1 \leq i \leq M \).

(i) Describe the maximum likelihood rule for this problem.
(ii) Assuming that all \( p_i \)s are distinct and using a combination of the union bound, Stein's Lemma, and the lower bound in Q5(ii), obtain an estimate of the largest value of \( M \) for which the average probability of error corresponding to a uniform prior will remain below \( \epsilon \), when \( n \) is large.
(iii) Suppose now each \( p_i \) itself was selected randomly, independent of each other, uniformly in \([0, 1]\). Consider the following version of the hard-thresholding decision rule described in the class:

Detector \( i, 1 \leq i \leq M \), compares the \( i \)th coin with an unbiased coin by using the best binary hypothesis test with probability of false alarm less than \( \epsilon/2 \). It declares \( B_i = 1 \) if it chooses \( P_i \), \( B_i = 0 \) otherwise. The decision maker declares \( i \) if it is the unique index with \( B_i = 1 \); if no such \( i \) is found, it declares an error.

Repeat part (ii) and obtain an estimate of the largest value of \( M \) for which the expected average probability of error corresponding to a uniform prior on the message and the random choice of \( p_i \)s will remain below \( \epsilon \), when \( n \) is large.

Note: We are assuming that \( p_i \)s are selected randomly but are known to the decision rule. This corresponds to evaluating the expected error of your decision rule over randomly chosen hypotheses. This evaluation is very useful in theory, as we shall see later in the course. It gives us a handle over how good our designed decision rule is on average (over not only the uniform distribution used to select a particular hypothesis, but average over the choice of hypotheses themselves).