Spring 2017
STAT 364 - Information Theory
Homework 5
Due: Mar 30, 2017
`Prof. Yihong Wu`

1 Exercises

1. Consider the hypothesis testing problem:

$$H_0 : X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} P = \mathcal{N}(0,1)\,,$$
$$H_1 : X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} Q = \mathcal{N}(\mu,1)\,.$$

Questions:

(a) Compute the Stein exponent.

(b) Compute the tradeoff region $\mathcal{E}$ of achievable error-exponent pairs (E_0, E_1). Express the optimal boundary in explicit form (eliminate the parameter).

(c) Compute the Chernoff exponent.[1]

2. A binary signal detector is being built. When the signal A is being sent a sequence of i.i.d. $X_j \sim \mathcal{N}(-1,1)$ is received. When the signal B is being sent a sequence of $X_j \sim \mathcal{N}(+1,1)$ is being received. Given a very large number n of observations $(X_1, \ldots, X_n)$ propose a detector for deciding between A and B. Consider two separate design cases:

(a) Misdetecting A for B or B for A are equally bad.

(b) Misdetecting A for B in 10^{-3} cases is ok, but the opposite should be avoided as much as possible.

Estimate the performance of your detector for $n = 20$ in either case.

3. *Baby version of Sanov's theorem.* Let $\mathcal{X}$ be a finite set. Let $\mathcal{E}$ be a *convex* subset of the simplex of probability distributions on $\mathcal{X}$. Assume that $\mathcal{E}$ has non-empty interior. Let $X^n = (X_1, \ldots, X_n)$ be iid drawn from some distribution P and let π_n denote the empirical distribution, i.e., $\pi_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$, which is a function of X^n. Our goal is to show that

$$E \triangleq \lim_{n\to\infty} \frac{1}{n} \log \frac{1}{P(\pi_n \in \mathcal{E})} = \inf_{Q\in\mathcal{E}} D(Q\|P). \tag{1}$$

a) Define the following set of joint distributions $\mathcal{E}_n \triangleq \{Q_{X^n} : Q_{X_i} \in \mathcal{E}\}$. Show that

$$\inf_{Q_{X^n}\in\mathcal{E}_n} D(Q_{X^n}\|P_{X^n}) = n \inf_{Q\in\mathcal{E}} D(Q\|P),$$

where $P_{X^n} = P^n$.

b) Consider the conditional distribution $\tilde{P}_{X^n} = P_{X^n|\pi_n\in\mathcal{E}}$. Show that $\tilde{P}_{X^n} \in \mathcal{E}_n$.

[1] Recall the Chernoff exponent is the exponent with which the optimal average probability of error of type I and II vanishes as $n \to \infty$.

c) Show that

$$P(\pi_n \in \mathcal{E}) \le \exp\Big(-n \inf_{Q\in\mathcal{E}} D(Q\|P)\Big), \quad \forall n.$$

d) For any Q in the interior of $\mathcal{E}$, show that

$$P(\pi_n \in \mathcal{E}) \ge \exp(-nD(Q\|P) + o(n)), \quad n \to \infty.$$

(Hint: Use data processing as in the proof of the large deviation theorem.)

e) Conclude (1).

4. *Error exponents of data compression.* Let X^n be iid according to P on a finite alphabet $\mathcal{X}$. Let $\epsilon_n^*(R)$ denote the minimal probability of error achieved by fixed-length compressors and decompressors for X^n of compression rate R. We have learned that the if $R < H(P)$, then $\epsilon_n^*(R)$ tends to zero. The goal of this exercise is to show it converges exponentially fast and find the best exponent.

   (a) For any sequence x^n, denote by $\pi(x^n)$ its empirical distribution and by $\hat{H}(x^n)$ its empirical entropy, i.e., the entropy of the empirical distribution.[2] For each $R > 0$, define the set $T = \{x^n : \hat{H}(x^n) < R\}$. Show that

$$|T| \le \exp(nR)(n+1)^{|\mathcal{X}|}.$$

   (b) Show that for any $R > H(P)$,

$$\epsilon_n^*(R) \le \exp\Big(-n \inf_{Q:H(Q)>R} D(Q\|P)\Big).$$

   Specify the achievable scheme. (Hint: Use Sanov's theorem in Exercise 3.)

   (c) Prove that the above exponent is asymptotically optimal:

$$\limsup_{n\to\infty} \frac{1}{n} \log \frac{1}{\epsilon_n^*(R)} \le \inf_{Q:H(Q)>R} D(Q\|P).$$

   (Hint: Recall that any compression scheme for memoryless source with rate below the entropy fails with probability tending to one. Use data processing inequality.)

[2] For example, for the binary sequence $x^n = (010110)$, the empirical distribution is Bern(1/2) and the empirical entropy is 1 `bit`.