Information Theory
Practice Set 9

Lakshmi Prasad Natarajan

Solutions are not to be returned

Practice Set

1. Exercise problems from Cover & Thomas, Chapter 11:
   11.1

2. *Applying the Stein's Lemma in Large Deviations Analysis.*

   `Hints available further below in this document.`

   Large deviations problems analyse the probabilities of rare events: such as asking what is the probability that we observe between 890 and 910 heads in 1000 tosses of a fair coin. One of the main ideas in large deviations analysis is 'change of measure': if an event B has high probability under distrbution P_1, what is the probability of B under distribution P_2? We will try to understand this question from the perspective of hypothesis testing.

   Suppose P_2 is the probability distribution of the toss outcome of a fair coin, and P_1 is that of an unfair coin: $P_1[1] = \theta + \epsilon$, $P_1[0] = 1 - \theta - \epsilon$, where $\theta > 0.5$ and $\epsilon > 0$ is a small number. The utility of ϵ will become clearer later.

   Let $X_1, \ldots, X_n$ be iid observations according to P_2. Suppose we are interested in the question: what is the probability that the number of 1's in $X_1, \ldots, X_n$ is greater that $n\theta$, where $0.5 < \theta$?

   Let $B = \left\{(x_1, \ldots, x_n) \ : \ \frac{1}{n}\sum_i x_i > \theta\right\}$.

   (a) Argue that $\lim_{n\to\infty} P_1[B] = 1$. Is B a typical set for the distribution P_1? Why or why not?

   (b) Look at this problem from the perspective of Neyman-Pearson hypothesis testing, and derive a lower bound on
   $$\lim_{n\to\infty} \frac{\log P_2[B]}{n}$$
   assuming that this limit exists.

   (c) In part (b), note that neither P_2 nor B is dependent on the choice of ϵ. Hence, the lower bound on the limit actually holds for any choice of $\epsilon > 0$. Can you improve the lower bound you derived in part (b) by somehow 'getting rid' of ϵ from the expression (in a loose sense, by choosing the ϵ that gives you the best lower bound).

   `Remark`: The lower bound derived in part (c) is in fact tight. This is the actual value of $\lim_{n\to\infty} \frac{1}{n} \log P_2[B]$.

   `Remark`: For an introduction to large deviations, you can read the method of types (Section 11.1 in Cover & Thomas), followed by Sanov's theorem (Section 11.4). This is not a part of the exam/quiz syllabus.

3. Exercise Problems in Cover & Thomas, Chapter 5:
   5.6 (assume $D = 2$), 5.11 (show that all suffix-free codes are uniquely decodable; show that any set of lengths obtainable via suffix-free codes can also be obtained using prefix-free codes), 5.18 (a) and (b), 5.23(a).

Hints for selected questions.

Hints

- Question 2

  Part (a). B is not a typical set for P_1. However, B contains all sequences with roughly $n(\theta + \epsilon)$ number of 1's. Use the weak law of large numbers to show that B is a high probability set.

  Part (b). Think of B as the acceptance region of a Neyman-Pearson hypothesis test. Then, the type-I error is small as $n \to \infty$. Then, show that

$$\lim_{n\to\infty} \frac{1}{n} P_2[B] \geq -D\left(P_1 || P_2\right).$$

  Part (c). Note that $-D(P_1||P_2)$ is a continuous decreasing function of ϵ and it tends to $D\left((1-\theta, \theta) \ || \ (0.5, 0.5)\right)$ as $\epsilon \to 0^+$. Argue that

$$\lim_{n\to\infty} \frac{1}{n} P_2[B] \geq -D\left((1-\theta, \theta) \ || \ (0.5, 0.5)\right).$$