Information and Coding Theory **Autumn 2014**

Homework 2

Due: November 25, 2014

Note: *You may discuss these problems in groups. However, you must write up your own solutions and mention the names of the people in your group. Also, please do mention any books, papers or other sources you refer to. It is recommended that you typeset your solutions in LaTeX.*

1. **Biased coins strike back.** In class we cosidered the problem of distinguishing coins distributed according to the following two distributions:

$$P = \begin{cases} 1 & \text{w.p. } \frac{1}{2} - \varepsilon \\ \\ 0 & \text{w.p. } \frac{1}{2} + \varepsilon \end{cases} \qquad \text{and} \qquad Q = \begin{cases} 1 & \text{w.p. } \frac{1}{2} \\ \\ 0 & \text{w.p. } \frac{1}{2} \end{cases}$$

We derived matching upper and lower bounds (up to constants) of the form $\Theta(1/\varepsilon^2)$ on the number of coin tosses required to distinguish the two distributions. Consider now the problem of distinguishing two extremely biased coins with slightly differing biases:

$$P' = \begin{cases} 1 & \text{w.p. } \varepsilon \\ \\ 0 & \text{w.p. } 1 - \varepsilon \end{cases} \qquad \text{and} \qquad Q' = \begin{cases} 1 & \text{w.p. } 2\varepsilon \\ \\ 0 & \text{w.p. } 1 - 2\varepsilon \end{cases}$$

Find tight upper and lower bounds (up to constants) on the number of independent coin tosses required to distinguish coins distributed according to P' and Q'.

2. **Counting using method of types (Problem 11.5 from the book).** Let U be a finite universe with $|U| = m$ and let $g : U \to \mathbb{R}$ be a real valued functions. Let $S \subseteq U^n$ be the set sequences $x_1, \ldots, x_n$ with each $x_i \in U$ defined as

$$S = \left\{ (x_1, \ldots, x_n) \in U^n \;\middle|\; \frac{1}{n} \sum_{i=1}^{n} g(x_i) \geq \alpha \right\}.$$

Let $\Pi = \left\{ P \;\middle|\; \sum_{a \in U} P(a) g(a) \geq \alpha \right\}$. Show that

$$|S| \leq (n+1)^m \cdot 2^{nH^*},$$

where $H^* = \max_{P \in \Pi} H(P)$.

3. **Loaded dice.** Consider the following game played using a dice: a single dice is rolled and we gain a dollar if the outcome is 2, 3, 4 or 5, and lose a dollar if it's 1 or 6.

   (a) What is our expected gain assuming all outcomes in $\{1, 2, 3, 4, 5, 6\}$ are equally likely.

   (b) Find the maximum entropy distribution over the universe $U = \{1, 2, 3, 4, 5, 6\}$ such that the expected gain is at least α (say α is greater than the expected gain for the uniform distribution).

4. **Finding one in many hidden coins.** We considered algorithms which tried to find a biased coin among N coins, where in a position j (unknown to the algorithm) we have a coin with probability of heads equal to $1/2 - \varepsilon$, and in the remaining positions we have fair coins which come up heads and tails with equal probability. The algorithm A outputs a pair $(a_t, b_t) \in [N]^2$ at each time t. Here, b_t represents the algorithm's guess at time t for the position of the biased coin and a_t is the position for which it asks to see the output of the toss at time t. We showed that for $T \leq 60N/\varepsilon^2$, there exists a set of at least $N/3$ positions (depending on A) such that if the biased coin in hidden in one of these positions, then the algorithm finds it with probability at most $1/2$.

   Here, we consider a generalization of the above setup where we have many biased coins and the algorithm succeeds if it manages to find any one of the biased coins. Let $Z_1, \ldots, Z_N$ be independent random variables with the following distribution:

$$Z_i \;=\; \begin{cases} 1 & \text{w.p. } \frac{k}{N} \\ \\ 0 & \text{w.p. } 1 - \frac{k}{N} \end{cases} .$$

   Given a sequence of values $\mathbf{z} = (z_1, \ldots, z_N)$ for the above random variables, we take distribution of the coin in the i^{th} position to be

$$P \;=\; \begin{cases} 1 & \text{w.p. } \frac{1}{2} - \varepsilon \\ \\ 0 & \text{w.p. } \frac{1}{2} + \varepsilon \end{cases} \qquad \text{if } z_i = 1$$

   and

$$Q \;=\; \begin{cases} 1 & \text{w.p. } \frac{1}{2} \\ \\ 0 & \text{w.p. } \frac{1}{2} \end{cases} \qquad \text{if } z_i = 0$$

   Coins in all N positions are independent. Also, the values $z_1, \ldots, z_N$ are only chosen once at the beginning and remain fixed through the run of the algorithm A. At each step t, A outputs a pair $(a_t, b_t) \in [N]^2$ and sees the output of the coin in position a_t as before. The algorithm succeeds after T steps, if the guess b_{T+1} made after seeing T tosses indeed contains the location of a biased coin i.e., b_{T+1} is such that $z_{b_{T+1}} = 1$.

   For a fixed $\mathbf{z} \in \{0,1\}^N$, let $D_{\mathbf{z}}$ denote the distribution for the view of the algorithm when the biased coins are located according to $\mathbf{z}$. Let $B_{\mathbf{z}}$ denote the set $\{i \in [N] | z_i = 1\}$. Let $D_{\mathbf{0}}$ denote the distribution for $\mathbf{z} = (0, 0, \ldots, 0)$.

   (a) For an appropriate constant c, show that

$$\mathop{\mathbb{P}}_{D_{\mathbf{z}}}[b_{T+1} \in B_{\mathbf{z}}] \;\leq\; \mathop{\mathbb{P}}_{D_{\mathbf{0}}}[b_{T+1} \in B_{\mathbf{z}}] + c \cdot \varepsilon \cdot \left(\mathop{\mathbb{E}}_{D_{\mathbf{0}}} \left[|\{t \in [T] \;\mid\; a_t \in B_{\mathbf{z}}\}| \right] \right)^{1/2} .$$

   (b) Use the above to show that

$$\mathop{\mathbb{E}}_{\mathbf{z}} \left[\mathop{\mathbb{P}}_{D_{\mathbf{z}}}[A \text{ finds a biased coin}] \right] \;=\; \mathop{\mathbb{E}}_{\mathbf{z}} \left[\mathop{\mathbb{P}}_{D_{\mathbf{z}}}[b_{T+1} \in B_{\mathbf{z}}] \right] \;\leq\; \frac{k}{N} + c \cdot \varepsilon \cdot \left(\frac{kT}{N} \right)^{1/2} .$$

5. **Chernoff bound for read-k families.** We used Sanov's theorem to derive the Chernoff bound for independent random variables $X_1, \ldots, X_n$ taking values uniformly in $\{0,1\}$. In particular, we showed that

$$\mathbb{P}\left[X_1 + \cdots + X_n \geq \left(\frac{1}{2} + \varepsilon\right) n\right] \leq (n+1)^2 \cdot 2^{-n \cdot D\left(\frac{1}{2}+\varepsilon \| \frac{1}{2}\right)},$$

where $D\left(\frac{1}{2} + \varepsilon \| \frac{1}{2}\right)$ denotes the KL-divergence of two distributions on $\{0,1\}$, with probabilities $(\frac{1}{2}+\varepsilon, \frac{1}{2}-\varepsilon)$ and $(\frac{1}{2}, \frac{1}{2})$. In this problem, we will consider functions $f_1, \ldots, f_r$ depending on the variables $X_1, \ldots, X_n$ and prove a concentration bound on the expression $f_1 + \cdots + f_r$.

Let $S_1, \ldots, S_r$ be subsets of $[n]$ for each $i \in [r]$, let $f_i : \{0,1\}^{S_i} \to \{0,1\}$ be a function which depends only on the variables in S_i. We use the shorthand X_{S_i} to denote the variables $\{X_j\}_{j \in S_i}$. Moreover, we have the property that each variable is involved in only k functions i.e., $\forall j \in [n],\ |\{i \in [r] \mid j \in S_i\}| = k$. Such a family of functions is called a read-k family (it is not too hard to see that the lower bound extends to the case when each vartiable is in *at most* k functions).

(a) Recall that for two random variables Z_1 and Z_2 distributed on *same universe* U, we also use $D(Z_1 \| Z_2)$ to mean $D(P_1 \| P_2)$. Let $Y_1, \ldots, Y_n$ be (not necessarily independent) random variables jointly distributed on $\{0,1\}^n$ and let $X_1, \ldots, X_n$ be random variables as above, distributed uniformly and independently on $\{0,1\}^n$. Let the sets $\{S_i\}_{i \in [r]}$ be as above. Use Shearer's lemma to show that

$$k \cdot D(Y_1, \ldots, Y_n \| X_1, \ldots, X_n) \geq \sum_{i \in [r]} D(Y_{S_i} \| X_{S_i}).$$

(b) Let $A = \left\{(a_1, \ldots, a_n) \in \{0,1\}^n \mid \sum_{i \in [r]} f_i\left(\{a_j\}_{j \in S_i}\right) \geq t\right\}$. Let $(Y_1, \ldots, Y_n)$ be uniformly distributed over the set A (note that $Y_1, \ldots, Y_n$ are not necessarily independent). Prove that

$$\mathbb{P}_{X_1, \ldots, X_n}\left[\sum_{i \in [r]} f_i(X_{S_i}) \geq t\right] = 2^{-D(Y_1, \ldots, Y_n \| X_1, \ldots, X_n)},$$

where the probability is over the uniform distribution for $X_1, \ldots, X_n$.

(c) For each $i \in [r]$, let $\mathbb{E}[f_i(X_{S_i})] = \mu_i$ and $\mathbb{E}[f_i(Y_{S_i})] = \nu_i$. Prove that

$$D(Y_{S_i} \| X_{S_i}) \geq D(\nu_i \| \mu_i),$$

where $D(\nu_i \| \mu_i)$ denotes the divergence of two distributions on $\{0,1\}$ with probabilities $(\nu_i, 1-\nu_i)$ and $(\mu_i, 1-\mu_i)$.

(d) Use the above bounds and the convexity of KL-divergence in both its arguments to show that for $\mu = \frac{1}{r} \cdot (\mu_1 + \cdots + \mu_r)$,

$$\mathbb{P}_{X_1, \ldots, X_n}\left[f_1(X_{S_1}) + \cdots + f_r(X_{S_r}) \geq (\mu + \varepsilon) \cdot r\right] \leq 2^{-(r/k) \cdot D(\mu + \varepsilon \| \mu)}.$$