1st Semester 2011/12

Homework Set #1
Method of types, Sanov's Theorem, Strong typicality

1. **Sanov's theorem:**
Prove the simple version of Sanov's theorem for the binary random variables, i.e., let $X_1, X_2, \ldots, X_n$ be a sequence of binary random variables, drawn i.i.d. according to the distribution:

$$\Pr(X = 1) = q, \quad \Pr(X = 0) = 1 - q. \tag{1}$$

Let the proportion of 1's in the sequence $X_1, X_2, \ldots, X_n$ be $p_{\mathbf{X}}$, i.e.,

$$p_{X^n} = \frac{1}{n}\sum_{i=1}^{n} X_i. \tag{2}$$

By the law of large numbers, we would expect $p_{\mathbf{X}}$ to be close to q for large n. Sanov's theorem deals with the probability that p_{X^n} is far away from q. In particular, for concreteness, if we take $p > q > \frac{1}{2}$, Sanov's theorem states that

$$-\frac{1}{n}\log \Pr\left\{(X_1, X_2, \ldots, X_n) : p_{X^n} \geq p\right\} \rightarrow p\log\frac{p}{q} + (1-p)\log\frac{1-p}{1-q} \tag{3}$$

Justify the following steps:

- $$\Pr\left\{(X_1, X_2, \ldots, X_n) : p_{\mathbf{X}} \geq p\right\} \leq \sum_{i=\lfloor np \rfloor}^{n} \binom{n}{i} q^i (1-q)^{n-i} \tag{4}$$
- Argue that the term corresponding to $i = \lfloor np \rfloor$ is the largest term in the sum on the right hand side of the last equation.
- Show that this term is approximately 2^{-nD}.
- Prove an upper bound on the probability in Sanov's theorem using the above steps. Use similar arguments to prove a lower bound and complete the proof of Sanov's theorem.

2. **Strong Typicality**
   Let X^n be drawn i.i.d.$\sim P(x)$. Prove that for each $x^n \in T_\epsilon^{(n)}(X)$,

$$2^{-n(H(X)+\delta')} \leq P^n(x^n) \leq 2^{-n(H(X)-\delta')}$$

   for some $\delta' = \delta'(\delta)$ that vanishes as $\delta \to 0$.

3. **Weak Typicality vs. Strong Typicality**
   In this problem, we compare the weakly typical set $A_\epsilon(P)$ and the strongly typical set $T_\delta(P)$. To recall, the definition of two sets are following.

$$A_\epsilon(P) = \left\{ x^n \in \mathcal{X}^n : \left| -\frac{1}{n} \log P^n(x^n) - H(P) \right| \leq \epsilon \right\}$$
$$T_\delta(P) = \left\{ x^n \in \mathcal{X}^n : \|P_{x^n} - P\|_\infty \leq \frac{\delta}{|\mathcal{X}|} \right\}$$

   (a) Suppose P is such that $P(a) > 0$ for all $a \in \mathcal{X}$. Then, there is an inclusion relationship between the weakly typical set $A_\epsilon(P)$ and the strongly typical set $T_\delta(P)$ for an appropriate choice of ϵ. Which of the statement is true: $A_\epsilon(P) \subseteq T_\delta(P)$ or $A_\epsilon(P) \supseteq T_\delta(P)$? What is the appropriate relation between δ and ϵ?

   (b) Give a description of the sequences that belongs to $A_\epsilon(P)$, vs. the sequences that belongs to $T_\delta(P)$, when the source is uniformly distributed, i.e. $P(a) = \frac{1}{|\mathcal{X}|}, \forall a \in \mathcal{X}$. (Assume $|\mathcal{X}| < \infty$.)

   (c) Can you explain why $T_\delta(P)$ is called **strongly** typical set and $A_\epsilon(P)$ is called **weakly** typical set?

4. **The probability of being jointly strongly when drawn dependently**
   Let Y^n be distributed according to the conditional distribution $p(y^n|x^n) = \prod_{i=1}^n P_{Y|X}(y_i|x_i)$. Then for every $x^n \in T_\epsilon^{(n)}(X)$, $\Pr(x^n, Y^n) \in T_{\epsilon'}^{(n)}(X,Y) \to 1$ as $n \to \infty$ and $\lim_{\epsilon=0} \epsilon'(\epsilon) = 0$.

5. **The probability of being jointly strongly typical when drawn independently**

Given $(x^n, y^n) \in T_\epsilon^{(n)}(X,Y)$. Let Z^n be distributed according to $\prod_{i=1}^n P_{Z|X}(z_i|x_i)$ (instead of $P_{Z|X,Y}$. Then,

$$\begin{aligned}
\Pr\{(x^n, y^n, Z^n) \in T_\epsilon^{(n)}(X,Y,Z)\} &\leq 2^{-n(I(Y;Z|X)-\delta(\epsilon))} \\
\Pr\{(x^n, y^n, Z^n) \in T_\epsilon^{(n)}(X,Y,Z)\} &\geq (1-\delta_{\epsilon,n})2^{-n(I(Y;Z|X)+\delta(\epsilon))},
\end{aligned}$$

where $\delta(\epsilon)$ goes to zero when ϵ goes to zero and $\delta_{\epsilon,n}$ goes to zero for any ϵ as n goes to infinity.

6. **The size of the conditional type** Prove that given $x^n \in T_\epsilon^{(n)}(X)$, then

$$(1-\delta_{\epsilon,n})2^{nH(Y|X)(1+\epsilon)} \leq |T_\epsilon^{(n)}(Y|x^n)| \leq 2^{nH(Y|X)(1-\epsilon)}.$$

7. **Simple version of Markov Lemma**

   Suppose X, Y, Z form a Markov chain $X - Y - Z$. Let $(x^n, y^n) \in T_\epsilon^{(n)}(X,Y)$ and Z^n is drawn i.i.d. according to $P(z|y)$, i.e., $P(z^n|y^n) = \prod_{i=1}^n P(z_i|y_i)$. Show that

$$\Pr\{(x^n, y^n, Z^n) \in T_\epsilon^{(n)}(X,Y,Z)\} \to 1 \qquad (5)$$

   as $n \to \infty$.

   (a) Is it true that for any $X - Y - Z$ and every sequence x^n, y^n, z^n such that if $(x^n, y^n) \in T_\epsilon^{(n)}(X,Y)$ and $(y^n, z^n) \in T_\epsilon^{(n)}(Y,Z)$, then $(x^n, y^n, z^n) \in T_\epsilon^{(n)}(X,Y,Z)$

8. **Large deviations.**

   Let $X_1, X_2, \dots$ be i.i.d. random variables drawn according to the Bernoulli distribution

$$\Pr\{X_i = 1\} = \Pr\{X_i = -1\} = \frac{1}{2}.$$

   Let S_n be the random walk defined by

$$S_n = \sum_{i=1}^n X_i.$$

   Find the function $f(\alpha)$ such that, for all $\alpha > 0$,

$$\lim_{n\to\infty} -\frac{1}{n} \log \Pr\{S_n \geq n\alpha\} = f(\alpha).$$

9. **Counting.**
Let $\mathcal{X} = \{1, 2, \ldots, m\}$. Show that the number of sequences $x^n \in \mathcal{X}^n$ satisfying $\frac{1}{n}\sum_{i=1}^{n} g(x_i) \geq \alpha$ is approximately equal to 2^{nH^*}, to first order in the exponent, where

$$H^* = \max_{P: \sum P(i)g(i) \geq \alpha} H(P).$$