TUTORIAL IV

1 Typical sets

Let $X^n = X_1 \dots X_n$ be independent and identically distributed bits with $X_1 \sim \text{Ber}(p)$, i.e., $P_{X_1}(0) = 1 - p$ and $P_{X_1}(1) = p$ (assume that $0 < p < 1/2$) and let $\delta > 0$ with $p + \delta \leq 1/2$.

1. Recall that $h_2(p) = -p \log_2 p - (1-p)\log_2(1-p)$. Show that for $k \leq n/2$ the following inequality holds:

$$1 + \binom{n}{1} + \cdots + \binom{n}{k} \leq 2^{h_2(k/n)n} .$$

A: *The left hand side of the inequality counts the number of subsets of $\{1, \dots, n\}$ of size at most k, including the empty set. Thus we define X the random variable obtained by choosing a subset of size at most k uniformly at random. We have:*

$$\sum_{i=0}^{k} \binom{n}{i} = 2^{H(X)}$$

Moreover since X represents a subset of $\{1, \dots, n\}$, so we can write $X = (X_1, \dots, X_n)$ where $X_i = 1$ if and only if i is in X. Then, if $p = \Pr[1 \in X]$:

$$H(X) \leq H(X_1) + H(X_2) + \cdots + H(X_n) = nH(X_1) = nh_2(p)$$

Since we choose at most k elements, $k \leq n/2$ and h_2 is non-decreasing on $(0, 1/2]$, we obtain $H(X) \leq nh_2(k/n)$. This give:

$$\sum_{i=0}^{k} \binom{n}{i} \leq 2^{nh_2(n/k)}$$

2. Using the previous inequality show that there exists a set $S_\delta \subseteq \{0,1\}^n$ with $|S_\delta| \leq 2^{n \cdot h_2(p+\delta)}$ satisfying the property that

$$\lim_{n\to\infty} \mathbf{P}\{X^n \in S_\delta\} = 1 .$$

A: *The set to define is intuitively the set that gives the typical elements that appear. Since $p < 1/2$, the elements with high probability will have many 0 and few 1. Let $N_1(w)$ be the number of 1's in a word w. Define, for some k fixed later:*

$$S_\delta = \{w \in \{0,1\}^n : N_1(w) \leq 1\}$$

Then:

$$|S_\delta| = \sum_{i=0}^{k} \binom{n}{i} \leq 2^{nh_2(k/n)}$$

Because we want the cardinality to be at most $2^{nh_2(p+\delta)}$, and h_2 is non decreasing, we set $k = \lfloor n(p+\delta) \rfloor$. This gives $|S_\delta| \leq 2^{nh_2(p+\delta)}$.

Now let's show that $\lim_{n\leftarrow n} \Pr[X^n \in S_\delta] = 1$. *Let* Y_n *be distributed with a binomial distribution on* (n,p). *We have* $\Pr[X^n \in S_\delta] \leq \Pr[Y_n \leq k]$. *By Chernoff bound, for* $\epsilon > 0$*:*

$$\Pr[Y_n > (1+\epsilon)np] \leq \left(\frac{e^\epsilon}{(1+\epsilon)^{1+\epsilon}}\right)^{np}.$$

To obtain the good inequality, set $\epsilon = \delta/p$. *Then:*

$$\Pr[Y_n \geq k] = \Pr[Y_n \geq \lfloor n(p+\delta)\rfloor + 1] \leq \Pr[Y_n \geq n(p+\delta)] \leq \left(e^{\delta/p}\left(\frac{\delta+p}{p}\right)^{-\frac{\delta+p}{p}}\right)^{np}$$

Define $f(p,\delta) = e^{\delta/p}\left(\frac{\delta+p}{p}\right)^{-\frac{\delta+p}{p}}$. *By studing f we prove* $f(p,\delta) < 1$, *which implies that* $\Pr[Y_n \geq k]$ *tends to* 0. *This gives the wanted limit.*