High-Dimensional Measures and Geometry

Lecture Notes from Feb 9, 2010

taken by Pankaj Singh

5.1.1 Question. Could we have better concentration results for I_n?

To study this, we first define "The Hamming Ball".

5.1.2 Definition. *The Hamming ball $B(r)$ of radius $r \geq 0$ in I_n is defined by*

$$B(r) = \{x \in I_n : d(x, 0) \leq r\} = \{x \in I_n : \sum_{i=1}^{n} x_i \leq r\}.$$

It has volume (with respect to un-normalized counting measure)

$$|B(r)| = \sum_{k=0}^{\lfloor r \rfloor} \binom{n}{k}.$$

We need an asymptotic way to compute the volume of this ball.

5.1.3 Lemma. *Let $B(r)$ be as above, and $H(t) = -t \ln t - (1-t)\ln(1-t)$ for $0 \leq t \leq 1$. If $0 \leq \lambda \leq \frac{1}{2}$ and $B_n = B(\lambda n) \subset I_n$, then*

1. $\ln |B_n| \leq nH(\lambda)$ *and*
2. $\lim_{n\to\infty} \frac{1}{n}|B_n| = H(\lambda)$

Proof. For (1), we recall that

$$\begin{aligned} 1 &= [\lambda + (1-\lambda)]^n \\ &= \sum_{k=0}^{n} \binom{n}{k} \lambda^k (1-\lambda)^{n-k} \\ &\geq \sum_{k=0}^{\lfloor \lambda n \rfloor} \binom{n}{k} \lambda^k (1-\lambda)^{n-k} \end{aligned}$$

If $\lambda \leq \frac{1}{2}$, then

$$\frac{\lambda}{1-\lambda} \leq 1$$

and so, we have

$$\lambda^k(1-\lambda)^{-k} \geq \lambda^{\lambda n}(1-\lambda)^{\lambda n}.$$

Thus, we have from above

$$\begin{aligned} 1 &\geq \sum_{0\leq k\leq \lfloor \lambda n\rfloor} \binom{n}{k} \lambda^{\lambda n}(1-\lambda)^{n-\lambda n} \\ &= |B_n| e^{-nH(\lambda)} \\ \Rightarrow \ln|B_n| &\leq nH(\lambda). \end{aligned}$$

To show (2), we use Stirling's formula

$$\ln(n!) = n\ln n - n + c_n \ln n, \text{ where } c_n \text{ stays bounded.}$$

We use this for binomial co-effiecients in

$$n^{-1}\ln|B_n| \geq n^{-1}\ln\binom{n}{m} = n^{-1}[\ln(n!) - \ln(m!) - \ln((n-m)!)] \text{ for } \lambda n - 1 < m \leq \lambda n, m \in \mathbb{Z}.$$

Stirling approximation gives us

$$n^{-1}\ln|B_n| \geq n^{-1}[n\ln n - n - m\ln m + m - (n-m)\ln(n-m) + (n-m) + C_n(\ln n + \ln m + \ln(n-m))],$$

where C_n stays bounded.
Now, re-expressing with λn instead of m, at the cost of C_n to c_n', we have

$$\begin{aligned} n^{-1}\ln|B_n| &\geq n^{-1}[n\ln n - \lambda n\ln\lambda n - (n-\lambda n)\ln(n-\lambda n) + c_n'(\ln n + \ln\lambda n + \ln(n-\lambda n))] \\ &= \ln n - \lambda(\ln\lambda + \ln n) - (1-\lambda)\left(\ln n + \ln(1-\lambda)\right) + \frac{c_n'}{n}\left(\ln n + \ln\lambda n + \ln(n-\lambda n)\right) \\ &= H(\lambda) + \frac{c_n'}{n}\left(\ln n + \ln\lambda n + \ln(n-\lambda n)\right) \end{aligned}$$

This lower bound establishes (2). □

5.2 Hamming Ball and Coin Toss

Consider I_n and μ_n(the normalized counting measure) on I_n as before. Let $f: I_n \to \mathbb{R}$,

$$f(x) = \sum_{i=1}^{n} x_i,$$

then $E[f] = \frac{n}{2}$ and f is 1-Lipschitz. By concentration result for I_n, for any $t \geq 0$, we have

$$\mu_n(\{x \in I_n : f(x) - \frac{n}{2} \leq -t\}) \leq e^{-2t^2/n}.$$

Letting $t = \lambda n$, for $0 < \lambda < 1/2$, we obtain

$$\mu_n(\{x \in I_n : f(x) - \frac{n}{2} \leq -\lambda n\}) \leq e^{-2\lambda n}.$$

We compare this with our more precise estimate based on the volume of B_n.
If n is even, $n = 2m$ and $\lambda n \in \mathbb{N}$, then $\{x \in I_n : f(x) - m + t \le 0\}$ contains all the points with at most $(m-t)$ non-zero co-ordinates. Thus, we have

$$\mu_n(\{x \in I_n : f(x) - m + t \le 0\}) = \frac{1}{2^n}\sum_{k=0}^{m-t}\binom{n}{k} = 2^{-n}|B(m-t)|.$$

By bound from above lemma, we have

$$\mu_n(\{x \in I_n : f(x) - m + t \le 0\}) \le 2^{-n}e^{nH(\frac{1}{2}-\lambda)}.$$

If λ is small, then we have

$$\begin{aligned} H(\frac{1}{2}-\lambda) &= -(\frac{1}{2}-\lambda)H(\frac{1}{2}-\lambda) - (\frac{1}{2}+\lambda)H(\frac{1}{2}+\lambda) \\ &= -(\frac{1}{2}-\lambda)H(1-2\lambda) - (\frac{1}{2}+\lambda)H(1+2\lambda) + (\frac{1}{2}-\lambda+\frac{1}{2}+\lambda)\ln 2 \\ &= \ln 2 - 2\lambda^2 + \text{ higher order terms }. \end{aligned}$$

So, we have

$$\mu_n(\{x \in I_n : f(x) - \frac{n}{2} \le -\lambda n\}) \le e^{-2n\lambda^2}e^{k_n\lambda^3},$$

where k_n is a constant.
For larger values of λ, bounds are different, for example, $\lambda = 1/2$ givies us one point set and so

$$\mu_n(\{x \in I_n : f(x) - \frac{n}{2} \le -\lambda n\}) \le \frac{1}{2^n} = e^{-n\ln 2} << e^{-n/2}.$$

5.2.4 Question. What about an unfair coin?

Pick $0 < p < 1$, and let $\mu_n(\{x\}) = p^k(1-p)^{n-k}$ with $k = \sum_{i=1}^n$. Define $f(x) = k = \sum_{i=1}^n$. Then, $E[f] = np$, f is 1-Lipschitz with respect to the Hamming distance. From martingale technique, we have

$$\mu_n(\{x \in I_n : f(x) - np \le -t\}) \le e^{-t^2/2n}$$

and

$$\mu_n(\{x \in I_n : f(x) - np \ge t\}) \le e^{-t^2/2n}.$$

Natural scaling would be $t = \alpha\sqrt{np(1-p)}$, but then $\frac{t^2}{n} = \frac{\alpha^2 p(1-p)}{2}$ and as $n \to \infty, p \to 0$ and $np = \beta^2$, we obtain a trivial estimate for $t \to \alpha\beta$.

5.2.5 Question. Is it possible to get a non-trivial bound with exponential $e^{-\alpha^2/2}$ on R. H. S. ?

We will try to address this question in next class.