High-Dimensional Measures and Geometry

Lecture Notes from Jan 28, 2010

taken by Pankaj Singh

3 Measure Concentration in boolean cubes

3.1.1 Definition. *The boolean cube is the set* $I_n = \{0,1\}^n$ *endowed with distance* $d(x,y)$ *between* $x = (x_1, \cdots, x_n)$ *and* $y = (y_1, \cdots, y_n)$ *given by*

$$d(x,y) = \sum_{i=1}^{n} |x_i - y_i|$$

A natural probability measure on I_n *is the counting probability measure* μ_n *defined by*

$$\mu_n(\{x\}) = \frac{1}{2^n} \text{ for any } x \in I_n$$

Notation: We will denote $\frac{1}{2^n} \sum_{x \in I_n} f(x)$ by $\int_{I_n} f d\mu_n$ for any function f on I_n.

We will see that "almost all" points are "close" to any sufficiently large subset $A \subset I_n$. This will help us show concentration of measure on the cube. Again, we use the Laplace Transform.

3.1.2 Theorem. *Let* $A \subset I_n$ *be a non-empty, let* $f : I_n \to \mathbb{Z}^+$, $f(x) = d(x, A)$ *be the distance from* A. *For any* $t \geq 0$, *we then have*

$$\int_{I_n} e^{tf} d\mu_n \leq \frac{1}{\mu_n(A)} \left(\frac{1}{2} + \frac{e^t + e^{-t}}{4} \right)^n$$

We delay the proof of this theorem and first derive a consequence.

3.1.3 Corollary. *For* $\epsilon > 0$,

$$\mu_n \left(\{ x \in I_n : d(x, A) \geq \epsilon \sqrt(n) \} \right) \leq e^{-\epsilon^2 / \mu_n(A)}.$$

Proof of Corollary. By Taylor expansion of both sides, we get

$$\frac{1}{2} + \frac{e^t + e^{-t}}{4} \leq e^{t^2/4}$$

Using Laplace Transform, we obtain

$$\mu_n\left(\{x\in I_n : d(x,A)\geq \epsilon\sqrt(n)\}\right) \leq e^{-t\epsilon\sqrt{n}}\int_{I_n} e^{td(x,A)}d\mu_n$$

$$\leq e^{-t\epsilon\sqrt{n}}\frac{1}{\mu_n(A)}\left(\frac{1}{2}+\frac{e^t+e^{-t}}{4}\right)^n$$

[By preceeding Theorem]

Now, using inequalitiy from the Taylor expansion, we get

$$\mu_n\left(\{x\in I_n : d(x,A)\geq \epsilon\sqrt(n)\}\right) \leq e^{-t\epsilon\sqrt{n}}\frac{1}{\mu_n(A)}e^{t^2n/4}$$

Now choosing $t=2\epsilon/\sqrt{n}$, we get

$$\mu_n\left(\{x\in I_n : d(x,A)\geq \epsilon\sqrt(n)\}\right) \leq e^{-\epsilon^2}\frac{1}{\mu_n(A)}e^{\epsilon^2}\leq \frac{e^{-\epsilon^2}}{\mu_n(A)}$$

□

Proof of Theorem. We abbreviate

$$\left(\frac{1}{2}+\frac{e^t+e^{-t}}{4}\right)=c(t).$$

We use induction over n to prove this theorem. For $n=1$, we have either $|A|=1$ or $|A|=2$. When $|A|=1$, we have $\mu_1(A)=1/2$ and

$$\int_{I_1} e^{tf}d\mu_1 = \frac{1}{2}+\frac{e^t}{2}$$

$$\leq 2c(t)$$

[since $\mu_1(A)=1/2$]

When $|A|=2$, we have

$$\int_{I_1} e^{tf}d\mu_1 = \frac{1}{2}+\frac{1}{2}$$

$$=1\leq c(t)$$

[since $c(t)\geq 1$]

Now, assume that the result is true on I_{n-1}. Consider

$$I_n^1=\{x\in I_n : x_n=1\} \text{ and } I_n^o=\{x\in I_n : x_n=0\}.$$

Also, we let

$$A_0=\{x\in I_n^0 : x\in A\} \text{ and } A_1=\{x\in I_n^1 : x\in A\}.$$

Define μ_{n-1} on I_n^1 and I_n^0. Then, by definition, we have

$$\mu_n(A) = \frac{\mu_{n-1}(A_1) + \mu_{n-1}(A_0)}{2}.$$

Now, we pick $x \in I_n$ and let $x' = (x_1, \cdots, x_{n-1}) \in I_{n-1}$. Since $x \in I_n^1$ or $x \in I_n^0$, we have

$$d(x, A) = min\{d(x', A_1), d(x', A_0) + 1\} \text{ if } x \in I_n^1$$

or

$$d(x, A) = min\{d(x', A_0), d(x', A_1) + 1\} \text{ if } x \in I_n^0,$$

because a closest point is in $A_1 \subset I_n^1$ or in $A_0 \subset I_n^0$.
Define

$$f_0(x) = d(x, A_0) \text{ and } f_1(x) = d(x, A_1).$$

Now, we have

$$\begin{aligned}
\int_{I_n} e^{tf} d\mu_n &= \int_{I_n^1} e^{tf} d\mu_n + \int_{I_n^0} e^{tf} d\mu_n \\
&= \int_{I_n^1} e^{tmin\{d(x',A_1),d(x',A_0)+1\}} d\mu_n(x') + \int_{I_n^0} e^{tmin\{d(x',A_0),d(x',A_1)+1\}} d\mu_n(x') \\
&= \int_{I_n^1} min\{e^{td(x',A_1)}, e^t e^{d(x',A_0)}\} d\mu_n(x') + \int_{I_n^0} min\{e^{td(x',A_0)}, e^t e^{d(x',A_1)}\} d\mu_n(x') \\
&\le min\left\{\int_{I_n^1} e^{td(x',A_1)} d\mu_n(x'), \int_{I_n^1} e^t e^{td(x',A_0)} d\mu_n(x')\right\} \\
&\quad + min\left\{\int_{I_n^0} e^{td(x',A_0)} d\mu_n(x'), \int_{I_n^0} e^t e^{td(x',A_1)} d\mu_n(x')\right\} \\
&= min\left\{\frac{1}{2}\int_{I_n^1} e^{td(x',A_1)} d\mu_{n-1}(x'), \frac{1}{2}\int_{I_n^1} e^t e^{td(x',A_0)} d\mu_{n-1}(x')\right\} \\
&\quad + min\left\{\frac{1}{2}\int_{I_n^0} e^{td(x',A_0)} d\mu_{n-1}(x'), \frac{1}{2}\int_{I_n^0} e^t e^{td(x',A_1)} d\mu_{n-1}(x')\right\} \\
&\quad \text{[Using induction assumption, we have]} \\
&\le \frac{1}{2} min\left\{\frac{c^{n-1}(t)}{\mu_{n-1}(A_1)}, e^t \frac{c^{n-1}(t)}{\mu_{n-1}(A_0)}\right\} + \frac{1}{2} min\left\{\frac{c^{n-1}(t)}{\mu_{n-1}(A_0)}, e^t \frac{c^{n-1}(t)}{\mu_{n-1}(A_1)}\right\} \\
&= \frac{1}{2}\frac{c^{n-1}(t)}{\mu_n(A)}\left[min\left\{\frac{\mu_n(A)}{\mu_{n-1}(A_1)}, \frac{e^t\mu_n(A)}{\mu_{n-1}(A_0)}\right\} + min\left\{\frac{\mu_n(A)}{\mu_{n-1}(A_0)}, \frac{e^t\mu_n(A)}{\mu_{n-1}(A_1)}\right\}\right] \qquad (*)
\end{aligned}$$

We abbreviate

$$a_0 = \frac{\mu_{n-1}(A_0)}{\mu_n(A)} \text{ and } a_1 = \frac{\mu_{n-1}(A_1)}{\mu_n(A)}.$$

So, we have $a_0 + a_1 = 2$. The R. H. S. of the inequality $(*)$ is maximal if we choose a_0, a_1 to minimize the expression $min\{a_1^{-1}, e^t a_0^{-1}\} + min\{a_0^{-1}, e^t a_1^{-1}\}$.

WLOG, we may assume $a_0 \geq a_1$. Now, at the end points, we have

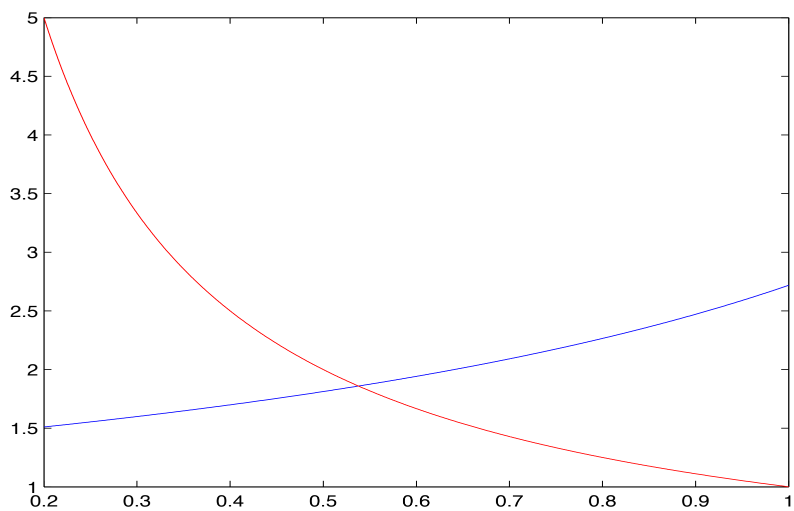
1. $a_0 = 2, a_1 = 0$ and therefore

$$\text{R. H. S. } \leq \frac{1}{2}\frac{c^{n-1}(t)}{\mu_n(A)}\left(\frac{e^t}{2}+\frac{1}{2}\right) = \frac{c^{n-1}(t)}{\mu_n(A)}\left(\frac{1}{4}+\frac{e^t}{4}\right)$$

2. $a_0 = 1 = a_1$ and then

$$\text{R. H. S. } \leq \frac{1}{2}\frac{c^{n-1}(t)}{\mu_n(A)}\left(min\{1,e^t\}+min\{1,e^t\}\right) = \frac{c^{n-1}(t)}{\mu_n(A)} \leq \frac{c^{n}(t)}{\mu_n(A)}$$

Now, when $a_0 > a_1$, $a_0^{-1} < a_1^{-1} < e^t a_1^{-1}$. So, the minimum is relevent in the first term only. The following is the sketch of the graph of a_1^{-1} [in red] and $e^t(2-a_1)^{-1}$ [in blue, for $t=1$] for $a_1 \in [0,1]$.



One is increasing and the other is decreasing. Therefore, $min\{a_1^{-1}, e^t a_0^{-1}\}$ is maximized when $a_1^{-1} = e^t(2-a_1^{-1}) \Rightarrow \frac{2}{a_1} - 1 = e^t \Rightarrow a_1 = \frac{2}{1+e^t}$ and $a_0 = \frac{2e^t}{1+e^t}$. Inserting these gives us the maximum of $min\{a_1^{-1}, e^t a_0^{-1}\} + min\{a_0^{-1}, e^t a_1^{-1}\}$

$$= \frac{1+e^t}{2} + \frac{1+e^t}{2e^t} = 1 + \frac{1}{2}(e^t + e^{-t}) = 2c(t).$$

Therefore, from $(*)$, we get

$$\int_{I_n} e^{tf} d\mu_n \leq \frac{1}{\mu_n(A)}\left(\frac{1}{2} + \frac{e^t + e^{-t}}{4}\right)^n.$$

$\square$