27.3.2008 Recall that $\Pr[\text{NAE accepts } f] = \frac{3}{4} - \frac{3}{4}\sum_S (-\frac{1}{3})^{|S|}\hat{f}(S)^2 \le \frac{7}{9} + \frac{2}{9}W_1(f)$,

where $W_1(f) = \sum_{|S|=1}\hat{f}(S)^2$. So if NAE accepts w.p. $\ge 1-\varepsilon$ then $W_1(f) \ge 1 - \frac{9}{2}\varepsilon$.

We know that if $W_1(f)=1$ then f is dictator or antidictator (from homework).

Let's see another proof: $f = a_1 \chi_{\{1\}} + \dots + a_n \chi_{\{n\}}$.

$1\cdot\chi_\phi = \mathbb{1} = f^2 = (a_1\chi_{\{1\}} + \dots + a_n\chi_{\{n\}})^2 = (\sum_{i=1}^n a_i^2)\cdot\chi_\phi + 2\cdot\sum_{i<j} a_i a_j \chi_{\{i,j\}}$.

This implies that $\forall i\ne j.\ a_i\cdot a_j = 0$, hence at most one a_i is nonzero. ■

(02) Thm [FKN]: If $f:\{0,1\}^n \to \{-1,1\}$ has $\sum_{|S|>1}\hat{f}(S)^2 < \varepsilon$, then f is $O(\varepsilon)$-close to a 1-junta.

Remark: NAE test is a test for "dictator or antidictator". It also implies an "approximate Arrows theorem": the only election function having $1-\varepsilon$ prob. of reasonable outcomes are those close to (anti) dictators.

Proof: First we notice that wlog we can assume that $\hat{f}(\phi)=0$. Otherwise, we can define $g:\{0,1\}^{n+1}\to\{-1,1\}$ by $g(x,\overset{\{0,1\}}{y}) = \begin{cases} f(x) & y=0 \\ -f(x+\vec{1}) & y=1\end{cases}$. This transformation sends χ_S for $S\subseteq[n]$ of odd size to itself and sends χ_S of even size to $\chi_{S\cup\{n+1\}}$.

In particular, $\hat{g}(\phi)=0$ and $\sum_{|S|>1}\hat{g}(S)^2 < \varepsilon$. Moreover, if g is close to 1-junta then so is f. So from now on, assume that f is balanced, i.e., $\hat{f}(\phi)=0$.

Write $f = \ell + h$ with $\ell = \sum_{i=1}^n \hat{f}(\{i\})\chi_{\{i\}}$ and $h = \sum_{|S|>1}\hat{f}(S)\chi_S$. Then,

$\mathbb{1} = f^2 = \ell^2 + 2\ell h + h^2 = \ell^2 + h(2f-h)$.

Since $E_x[h(x)^2] < \varepsilon$, $\Pr[|h(x)| \ge 10\sqrt{\varepsilon}] \le 0.01$ (Markov). Hence, $\Pr[|h(x)\cdot(2f(x)-h(x))| > 21\sqrt{\varepsilon}] \le 0.01$

Moreover, $\ell^2 = \underbrace{(\sum_{i=1}^n \hat{f}(\{i\})^2)}_{\in(1-\varepsilon,1]}\chi_\phi + \underbrace{2\sum_{i<j}\hat{f}(\{i\})\cdot\hat{f}(\{j\})\cdot\chi_{\{i,j\}}}_{q}$.

Therefore, $\Pr[|q(x)| > 11\sqrt{\varepsilon}] \le 0.01$, or equivalently $\Pr[q(x)^2 > 121\varepsilon] \le 0.01$.

By the hypercontractive ineq., $E[q(x)^4] \le 81\cdot E[q(x)^2]^2$.

Using the following claim from the homework (with $X \leftarrow q^2$, $K\leftarrow 121\varepsilon$, $\delta\le 0.01$, $L\leftarrow E[q^2]$) shows that $E[q(x)^2] < 1000\varepsilon$.

Claim: [Paley-Zygmund] If X is a nonneg. r.v. with $\Pr[X>K]\le\delta$ and $E[X]\ge L > K > 0$ then $E[X^2] \ge \frac{(L-K)^2}{\delta}$.

(Since otherwise $E[q(x)^4] \ge \frac{(E[q^2]-121\varepsilon)^2}{\delta} > 81\cdot E[q^2]^2$ in contradiction).

Therefore, $1000\varepsilon > E[q(x)^2] = \sum_{i<j} \hat{f}(\{i\})^2 \cdot \hat{f}(\{j\})^2 = \frac{1}{2}\left((\sum \hat{f}(\{i\})^2)^2 - \sum_i \hat{f}(\{i\})^4 \right) \geq$

$\geq \frac{1}{2}(1-2\varepsilon - \sum \hat{f}(\{i\})^4)$. Hence, $(\sum \hat{f}(\{i\})^2) \cdot \max \hat{f}(\{i\})^2 \geq \sum \hat{f}(\{i\})^4 > 1-2002\varepsilon$

$\Rightarrow \exists i$. s.t. $|\hat{f}(\{i\})| > 1-1001\varepsilon$. ∎

Exponential Separation for one-way Quantum Communication

Goal: find tasks that quantum does better than classical.

Q. algorithms like Shor's factoring algorithm, only give conditional seperations.

Some previous results of unconditional seperations include [Raz99, Bar-Yossef Jayram Kerenedis '04]

Here we consider the one way communication model: $x \to \boxed{A} \xrightarrow{\text{message}} \boxed{B} \leftarrow y$, $\boxed{B} \to f(x,y)$.

Hidden Matching Problem

- Alice receives as input $x \in \{0,1\}^n$.
- Bob receives a sequence $M = (M_1, M_2, \dots, M_{n/4})$ of disjoint pairs from $[n]$. This induces a string $z = M(x) \in \{0,1\}^{n/4}$ obtained by XORing the two bits on each edge. Bob needs to "say something about z".

To make this a Boolean function we can give Bob an additional vector $w \in \{0,1\}^{n/4}$ that is either z or $\bar{z}$ and ask him which is the case.

Classical one-way protocol: Clearly we can solve using n bits of comm.

We can also do this using $\Theta(\sqrt{n})$ - Alice sends to Boo $\Theta(\sqrt{n})$ randomly chosen bit

With constant probability, Bob gets the two endpoints of one edge (birthday paradox) and gets a bit of z.

This can be done using $(\Theta(\sqrt{n}))$ bits of comm. (using shared randomness + Newman's thm).

We'll show that this is tight.

Q: What happens if we take a perfect matching (i.e., $n/2$ edges)?

Quantum one-way protocol: In a quantum one-way protocol with communication c, Alice sends to Bob a unit vector v in $\mathbb{R}^{2^c}$. Bob then "measures" this vector by choosing an orthonormal basis $w_1, w_2, \dots, w_{2^c}$ of $\mathbb{R}^{2^c}$ and then he obtains a number $i \in \{1, 2, \dots, 2^c\}$ with prob $\langle v, w_i \rangle^2$. We now show quantum protocol with only $O(\log n)$ communication.

Alice sends to Bob the vector $v = \frac{1}{\sqrt{n}}((-1)^{x_1}, (-1)^{x_2}, \dots, (-1)^{x_n})$. Bob, given his matching, measures in a basis that contains for each edge $\{i,j\} \in M$ the vectors $\frac{1}{\sqrt{2}}(e_i + e_j)$ and $\frac{1}{\sqrt{2}}(e_i - e_j)$ as well as vectors e_k for each $k \in [n]$ not in the matching.

With prob. $1/2$, Bob obtains an edge $\{i,j\}$ together with the parity on that edge giving him a bit of z, as required.

Classical lower bound of $\Omega(\sqrt{n})$

Assume we have a protocol in which Alice sends c bits to Bob. We consider their protocol when the input x and M are chosen uniformly. This allows us to assume wlog that the protocol is determinstic (since otherwise we can fix the random coins in a way that maximizes the success probability). Alice's message partitions $\{0,1\}^n$ into 2^c sets of average size 2^{n-c}. Upon receiving a message, Bob's view of x is as if it is uniformly distributed over some set A of size $\approx 2^{n-c}$.

Given a matching M this induces a distribution on $z \in \{0,1\}^{n/4}$ given by $P_M(z) = \frac{2^{n/4}}{|A|} \cdot |\{x \in A \mid M(x) = z\}|$. Notice that the normalization is such that $\|P_M\|_1 = E[P_M] = 1$ and $\sum P_M(z) = 2^{n/4}$.

Thm: If $|A| \geq 2^{n-c}$ with $c < \frac{1}{10{,}000}\sqrt{n}$, then $\underset{M}{E}[\|P_M - \mathbb{1}\|_1] \leq \frac{1}{10}$ (the $\frac{1}{10}$ can be made arbitrarily close to 0 by choosing smaller c).

In words, this says that the statistical distance between P_M and the uniform distribution is $\leq \frac{1}{10}$ on average, implies that except with probability $\frac{1}{10}$, Bob behaves as if its input is uniform.

Lemma: Let $A \subseteq \{0,1\}^n$ be of size $\geq 2^{n-c}$ and let $f = \frac{2^n}{|A|} \cdot 1_A$ be the uniform "distribution" over A. Then $\forall \delta \in (0,1], k \geq 1$, $\sum_{|S|=k} \hat{f}(S)^2 \leq 2^{2\delta c} \cdot \delta^{-k}$.

Proof: Using the h.c. ineq. $\|T_{\sqrt{p-1}} f\|_2^2 \leq \|f\|_p^2$, we get
$$\sum \delta^{|S|} \cdot \hat{f}(S)^2 \leq \|f\|_{1+\delta}^2 = \left(\frac{2^n}{|A|}\right)^{2\delta/1+\delta} \leq \left(\frac{2^n}{|A|}\right)^{2\delta}.$$

Cor: If $|A| \geq 2^{n-c}$ and $k \in \{1,\dots,4c\}$ then $\sum_{|S|=k} \hat{f}(S)^2 \leq \left(\frac{4\sqrt{2}\,c}{k}\right)^k$.

Proof: Use $\delta = \frac{k}{4c} \in (0,1]$.

Example: Consider the set $A = \{x \in \{0,1\}^n \mid |x| \geq \frac{n}{2} + \sqrt{cn}\}$.

It is of size $2^{n-\Theta(c)}$ (because we moved $\sqrt{c}$ standard deviations, the probability that random x is in A is $2^{-\Theta(c)}$). If we choose $x \in A$ uniformly each bit is 1 w.p. $\frac{1}{2} + \sqrt{\frac{c}{n}}$ and therefore $\hat{f}(\{i\}) \approx \sqrt{\frac{c}{n}}$ for all i. This shows that the bound is tight for $k=1$. Similarly, if we take any k bits for small k, they are essentially independent and their XOR is 1 w.p. $\frac{1}{2} + \left(\sqrt{\frac{c}{n}}\right)^k$. This shows that the bound is tight for small k.

Lemma: For $S \subseteq [n]$ of even size $k \le n/2$, $\Pr_M[S \text{ is a union of edges of } M] = \frac{\binom{n/4}{k/2}}{\binom{n}{k}}$.

2.4.2008 Example (for the Cor.): Take $A = \{x \in \{0,1\}^n \mid x_1 = x_2 = \dots = x_c = 0\}$. $|A| = 2^{n-c}$

$\hat{f}(S) = 1$ if $S \subseteq \{1,\dots,c\}$ and 0 o.w. $\sum_{|S|=k} \hat{f}(S)^2 = \binom{c}{k} \approx \left(\frac{c}{k}\right)^k$.

Proof of Thm: Let $f = \frac{2^n}{|A|} \cdot 1\!\!1_A$. Then $p_M(z) = 2^{-\frac{3}{4}n} \cdot \sum_{x: M(x) = z} f(x)$. $\{0,1\}^{n/4}$

Then $\hat{p}_M(T) = \hat{f}(M^{-1}(T))$ where $M^{-1}(T)$ is the union of edges indicated by T.

Hence, $\left(E_M[\|p_M - 1\!\!1\|_1]\right)^2 \le E_M[\|p_M - 1\!\!1\|_1^2] \le E_M[\|p_M - 1\!\!1\|_2^2] = E_M\left[\sum_{T \ne \phi} \hat{p}_M(T)^2\right] =$

$$= E_M\left[\sum_{T \ne \phi} \hat{f}(M^{-1}(T))^2\right] = \sum_{S \ne \phi} \Pr_M\left[\begin{smallmatrix}S \text{ is a union of} \\ \text{edges of } M\end{smallmatrix}\right] \cdot \hat{f}(S)^2 =$$

$$= \sum_{\substack{k=2 \\ \text{even}}}^{n/2} \frac{\binom{n/4}{k/2}}{\binom{n}{k}} \cdot \sum_{|S|=k} \hat{f}(S)^2 \le \sum_{\substack{k=2 \\ \text{even}}}^{n/2} \left(\sqrt{\frac{e}{2}} \cdot \sqrt{\frac{k}{n}}\right)^k \cdot \sum_{|S|=k} \hat{f}(S)^2$$

$\downarrow$ Stirling: $\left(\frac{n}{k}\right)^k \le \binom{n}{k} \le \left(\frac{en}{k}\right)^k$

Terms with $k > 4c$: Using $\sum_S \hat{f}(S)^2 = \frac{2^n}{|A|} \le 2^c$ and $k \le n/2$,

$$2^c \cdot \sum_{k=4c}^{\infty} \left(\sqrt{\frac{e}{4}}\right)^k < 0.99^c < \frac{1}{200}.$$

For $k \le 4c$, $\sum_{\substack{k=2 \\ \text{even}}}^{4c} \left(\sqrt{\frac{ek}{2n}}\right)^k \cdot \left(\frac{4\sqrt{2}c}{k}\right)^k = \sum_{\substack{k=2 \\ \text{even}}}^{4c} \left(\frac{4\sqrt{e} \cdot c}{\sqrt{nk}}\right)^k \le \sum \left(\frac{4\sqrt{e}}{10{,}000} \frac{1}{\sqrt{k}}\right)^k < \frac{1}{200}$. ∎

Def: For two distributions $p, q : D \to \mathbb{R}^+$ with $\sum p(x) = \sum q(x) = 1$ we define their statistical distance/variational distance as $\Delta(p,q) = \frac{1}{2} \sum |p(x) - q(x)|$.

Remark: $0 \le \Delta \le 1$ with $\Delta = 0$ iff $p = q$.

Claim: $\Delta(p,q) = \max_{S \subseteq D} \left|\sum_S (p(x)) - \sum_S (q(x))\right|$

Claim: For any (possibly randomized) function $f: D \to D'$, $\Delta(f(x), f(y)) \le \Delta(x,y)$, $\begin{smallmatrix}x \sim p \\ y \sim q\end{smallmatrix}$.

In our case, take $f: \{0,1\}^{n/4} \to \{0,1\}$ to be the output of the following process.

Choose z according to p_M, give Bob the set $\{z, \bar{z}\}$ and ask him to tell which one is z. Output 1 iff Bob is successful. If we chose z according to the uniform distribution, then the output of f would be 1 w.p. $1/2$. Since we choose z from p_M the prob. of outputting 1 is at most $\frac{1}{2} + \Delta(p_M, U) = \frac{1}{2} + \frac{1}{2}\|p_M - 1\!\!1\|_1$, so Bob cannot win w.p. above 55%.

Proof of the Hyper. Ineq.

Thm: $\forall f: \{0,1\}^n \to \mathbb{R}$, $1 \le p \le q \le \infty$, $\|T_\rho f\|_q \le \|f\|_p$ for $\rho = \sqrt{\frac{p-1}{q-1}}$.

Proof: By induction on n.

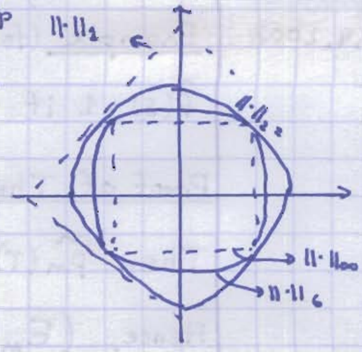
Base case (sketch) - $n=1$: Let $a, b \in \mathbb{R}$ such that $f(0) = a+b$, $f(1) = a-b$.

Then, $T_\rho f(0) = a + \rho b$, $T_\rho f(1) = a - \rho b$. Our goal is to prove that $\forall a,b$,

$$\|(a+\rho b, a-\rho b)\|_q \le \|(a+b, a-b)\|_p$$

Without loss of generality assume $a=1$. Let us prove the case $b=\varepsilon\to 0$.

Then we wish to prove $\left(\frac{(1+\rho\varepsilon)^q+(1-\rho\varepsilon)^q}{2}\right)^{1/q} \le \left(\frac{(1+\varepsilon)^p+(1-\varepsilon)^p}{2}\right)^{1/p}$.



See figure on page 7

Consider the RHS, $(1+\varepsilon)^p = 1+p\varepsilon+\frac{p(p-1)}{2}\varepsilon^2+O(\varepsilon^3)$.

$(1-\varepsilon)^p = 1-p\varepsilon+\frac{p(p-1)}{2}\varepsilon^2+O(\varepsilon^3)$.

So, $\frac{(1+\varepsilon)^p+(1-\varepsilon)^p}{2} = 1+\frac{p(p-1)}{2}\varepsilon^2+O(\varepsilon^3)$.

Since $(1+\delta)^{1/p} = 1+\frac{1}{p}\delta+O(\delta^2)$, $RHS = 1+\frac{p-1}{2}\varepsilon^2+O(\varepsilon^3)$.

Similarly, $LHS = 1+\frac{q-1}{2}(\rho\varepsilon)^2+O(\varepsilon^3)$. For $LHS \le RHS$ we need $\rho \le \sqrt{\frac{p-1}{q-1}}$.

Inductive step: Assume that the ineq. holds for $1,2,\dots,n-1$. We write $f:\{0,1\}^n\to\mathbb{R}$

$$\underset{\in\{0,1\}^{n-1}}{f(\underset{\downarrow}{x},x_n)} = \begin{cases} f_0(x) & x_n=0 \\ f_1(x) & x_n=1\end{cases}$$ where f_0,f_1 are restrictions of f.

$$\text{So, } T_\rho f(x,x_n) = \begin{cases} \left(\frac{1+\rho}{2}\right)T_\rho f_0(x) + \left(\frac{1-\rho}{2}\right)T_\rho f_1(x) & x_n=0 \\ \left(\frac{1-\rho}{2}\right)T_\rho f_0(x) + \left(\frac{1+\rho}{2}\right)T_\rho f_1(x) & x_n=1\end{cases}$$

$$\|T_\rho f\|_q = \left(\frac{\|(\frac{1+\rho}{2})T_\rho f_0+(\frac{1-\rho}{2})T_\rho f_1\|_q^q + \|(\frac{1-\rho}{2})T_\rho f_0+(\frac{1+\rho}{2})T_\rho f_1\|_q^q}{2}\right)^{1/q}$$

induction hyp.
$$\le \left(\frac{\|(\frac{1+\rho}{2})f_0+(\frac{1-\rho}{2})f_1\|_p^q + \|(\frac{1-\rho}{2})f_0+(\frac{1+\rho}{2})f_1\|_p^q}{2}\right)^{1/q}$$

Minkowski Ineq.
$$\le \left(\frac{1}{2^{n-1}}\cdot\sum_{x\in\{0,1\}^{n-1}}\left(\frac{|(\frac{1+\rho}{2})f_0(x)+(\frac{1-\rho}{2})f_1(x)|^q + |(\frac{1-\rho}{2})f_0(x)+(\frac{1+\rho}{2})f_1(x)|^q}{2}\right)^{\frac{1}{q}\cdot p}\right)^{1/p}$$

Ind. hyp for $n=1$
$$\le \left(\frac{1}{2^{n-1}}\sum_{x\in\{0,1\}^{n-1}}\left(\left(\frac{|f_0(x)|^p+|f_1(x)|^p}{2}\right)^{1/p}\right)^p\right)^{1/p} = \left(\frac{1}{2^n}\sum_{x\in\{0,1\}^n}|f(x)|^p\right)^{1/p} = \|f\|_p.$$

Gaussian Random Variables

<u>Def</u>: $g=(g_1,\dots,g_n)\in\mathbb{R}^n$ is a (n-dimensional standard) <u>Gaussian random variable</u> if $g_1,\dots,g_n$ are i.i.d. Gaussian / normal $N(0,1)$.

<u>Observation</u>:
- g's distribution is spherically symmetrix.

  Indeed, its prob. density function is $g(x_1,\dots,x_n) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}}\cdot e^{-x_i^2/2} = \frac{1}{(\sqrt{2\pi})^n}e^{-(\sum x_i^2)/2}$

  $= \frac{1}{(\sqrt{2\pi})^n}\cdot e^{-\|x\|_2^2/2}$ which only depends on $\|x\|_2$.
- $\|g\|_2^2 = \sum_{i=1}^n g_i^2$ is the sum of n r.v.s with mean 1 and variance 3.

  So by the central limit theorem (CLT) it is with high prob. in $n\pm O(\sqrt{n})$

  Hence, $\|g\|_2 = \sqrt{n}(1\pm O(1/\sqrt{n}))$. So we can think of g as distributed uniformly on a sphere of radius $\sqrt{n}$.

Def: $(x,y) \in \mathbb{R}\times\mathbb{R}$ are <u>ρ-correlated normal variables</u> (or, more precisely, (x,y) is distributed like the bivariate normal $N\Big(\underbrace{\binom{0}{0}}_{\text{mean}}, \underbrace{\begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}}_{\text{cov. matrix}}\Big)$) if x is chosen according to $N(0,1)$ and then $y = \rho\cdot x + \sqrt{1-\rho^2}\cdot z$ where z is an independent $N(0,1)$ r.v.

Remark: • y's marginal is $N(0,\rho^2) + N(0, 1-\rho^2) = N(0,1)$.

In fact, the definition is symmetric

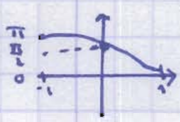
• $E[xy] = E[\rho x^2] + E[\sqrt{1-\rho^2}\, x\cdot z] = \rho\cdot E[x^2] = \rho$.

See figure on page 8

Def: $g, h \in \mathbb{R}^n$ are <u>n-dimensional ρ-correlated Gaussians</u> if each coordinate (g_i, h_i) is chosen independently from a ρ-correlated 1-dim Gaussians $N\Big(\binom{0}{0}, \begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}\Big)$.

Observation: $E[\langle g,h\rangle] = \sum_{i=1}^{n} E[g_i h_i] = \rho\cdot n$, and in fact $\langle g,h\rangle$ is concentrated around ρn.

Therefore, the angle between g and h is concentrated on $\arccos\rho$.



Noise Stability

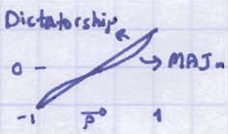
Def: For a function $f:\{0,1\}^n \to \mathbb{R}$ define its noise stability by $S_\rho(f) = \langle f, T_\rho f\rangle = \sum_S \rho^{|S|} \hat{f}(S)^2$.

For ±1 functions, this is $1 - 2\Pr_{x\sim_\rho y}[f(x) \neq f(y)]$ (i.e., $y \sim x + \mu_{\frac{1-\rho}{2}}$).

Remark: $NS_\delta(f) = \frac{1}{2}(1 - S_{1-2\delta}(f))$.

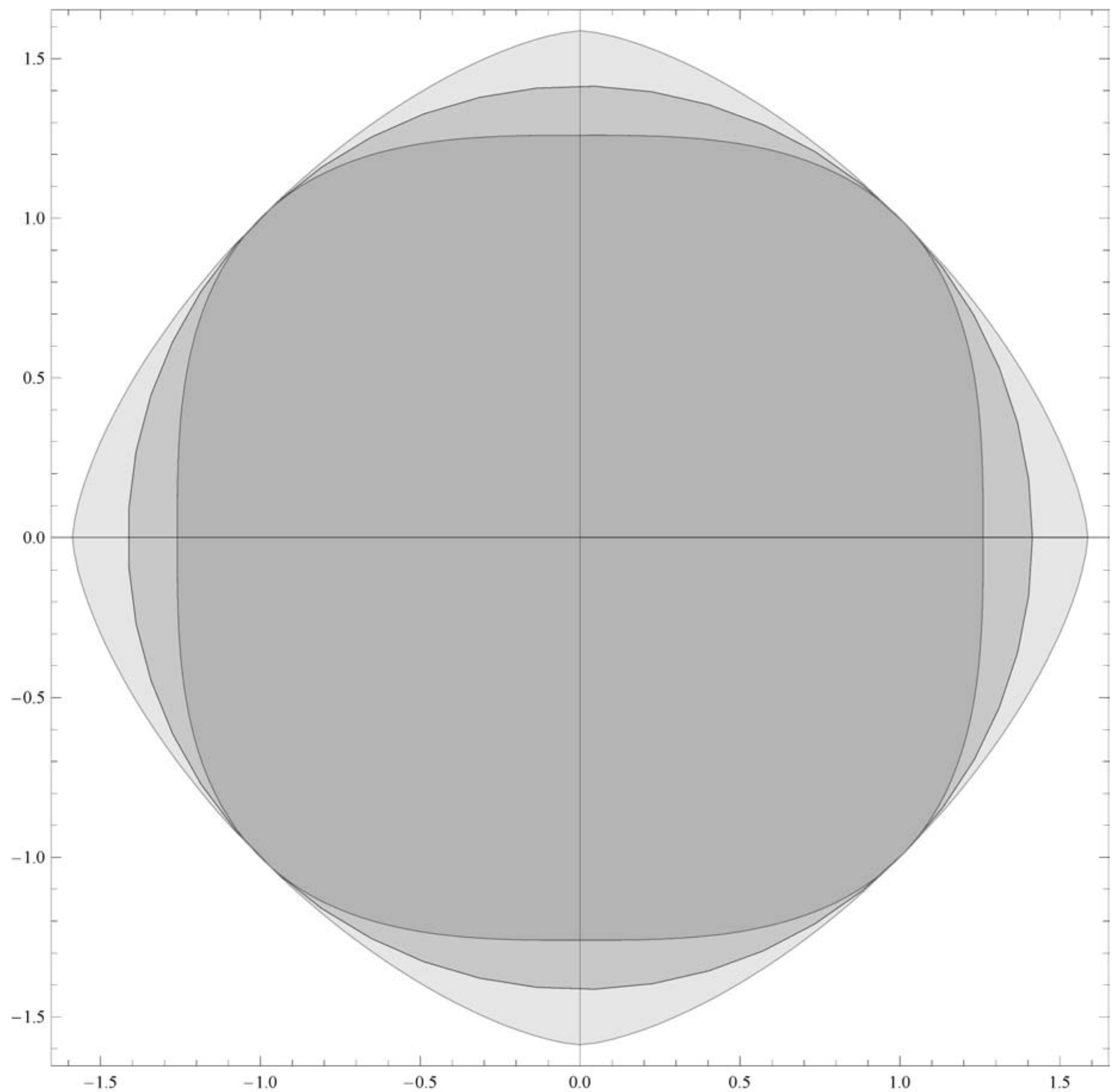
Examples: • Constant function ±1 : 1.

• Dictatorship: $S_\rho = \rho$.



Prop: $S_\rho(MAJ_n) = \frac{2}{\pi}\arcsin\rho + O(\frac{1}{n})$.

```
ParametricPlot[
 {
  r*{Cos[θ] , Sin[θ]} / ((Abs[Cos[θ]]^p + Abs[Sin[θ]]^p) / 2)^(1 / p)  /. p → 2,
  r*{Cos[θ] , Sin[θ]} / ((Abs[Cos[θ]]^p + Abs[Sin[θ]]^p) / 2)^(1 / p)  /. p → 3,
  r*{Cos[θ] , Sin[θ]} / ((Abs[Cos[θ]]^p + Abs[Sin[θ]]^p) / 2)^(1 / p)  /. p → 1.5
 },
 {r, 0, 1}, {θ, 0, 2 π}, Mesh → False, PlotPoints → 50]
```



```
Needs["MultivariateStatistics`"];

GaussSam[ϵ_] := Random[MultinormalDistribution[{0, 0}, {{1, 1 - ϵ}, {1 - ϵ, 1}}]];

GaussSams[k_, ϵ_] := Transpose[Table[GaussSam[ϵ], {k}]];

A = Table[GaussSams[2, 0.001], {3000}];

ListLinePlot[A, PlotStyle → Black]
```