Hypercontractivity, take one

To have a first feeling of hypercontractivity and equivalence of norms we analyze a special case.

Theorem 1. *Let g be a multilinear polynomial of degree d in the variables $(x_1, \ldots, x_n)$ then*

$$\mathbb{E}[g^4] \leqslant 9^d (\mathbb{E}[g^2])^2 .$$

Proof. We work by induction over n. If $n=0$ the function g is constant so the claim is easy to verify (with $d=0$). Now

$$g(x_1, \ldots, x_{n+1}) = g_0(x_1, \ldots, x_n) + g_1(x_1, \ldots, x_n)\, x_{n+1}$$

with $\deg(g_0) \leqslant d$ and $\deg(g_1) \leqslant d-1$. Compute

$$\mathbb{E}[g^4] = \mathbb{E}[g_0^4 + 4\, g_0^3 g_1\, x_n + 6\, g_0^2\, g_1^2 + 4\, g_0\, g_1^3\, x_n^3 + g_1^4]$$

$$= \mathbb{E}[g_0^4] + 4 \underbrace{\mathbb{E}[g_0^3 g_1\, x_n]}_{=0} + 6\mathbb{E}\,[g_0^2\, g_1^2] + 4 \underbrace{\mathbb{E}[g_0\, g_1^3\, x_n^3]}_{=0} + \mathbb{E}[g_1^4]$$

$$= \mathbb{E}[g_0^4] + 6\mathbb{E}\,[g_0^2\, g_1^2] + \mathbb{E}[g_1^4]$$

Now by the induction hypothesis

$$\mathbb{E}[g_0^4] \leqslant 9^d \mathbb{E}[g_0^2]^2, \qquad \mathbb{E}[g_1^4] \leqslant 9^{d-1} \mathbb{E}[g_1^2]^2$$

$$\mathbb{E}[g_0^2 g_1^2] \leqslant \mathbb{E}[g_0^4]^{1/2} \mathbb{E}[g_1^4]^{1/2} \leqslant 3^d\, 3^{d-1} \mathbb{E}[g_0^2] \mathbb{E}[g_1^2]$$

so

$$\mathbb{E}[g^4] \leqslant 9^d \mathbb{E}[g_0^2]^2 + 6(3^d\, 3^{d-1} \mathbb{E}[g_0^2] \mathbb{E}[g_1^2]) + 9^d \mathbb{E}[g_1^2]^2$$

$$= 9^d (\mathbb{E}[g_0^2]^2 + 2\mathbb{E}[g_0^2] \mathbb{E}[g_1^2] + \mathbb{E}[g_1^2]^2)$$

$$= 9^d (\mathbb{E}[g_0^2] + \mathbb{E}[g_1^2])^2 = 9^d (\mathbb{E}[g_0^2 + 2 g_0 g_1 x_n + g_1^2 x_n^2])^2 = 9^d \mathbb{E}[g^2]^2$$

□

Remark 2. Note that the same proof works for more general distributions of the x_i. In particular it works for standard gaussian variables.

Hypercontractivity tells us that finite degree polynomials behave uniformly with respect to probability estimations, for example if probability of large values is uniformly bounded then moments are uniformly bounded.

Corollary 3. *Let q be a second order polynomial and assume that $p = \mathbb{P}(|q| > 1) < 1/81$. Then*

$$\mathbb{E}[q^2] \leqslant \frac{1}{1 - 9\, p^{1/2}} .$$

Proof. By the hypercontractivity estimate and Cauchy-Schwartz

$$\mathbb{E}[q^2] = \mathbb{E}[q^2 \mathbb{I}_{|q| \leqslant 1}] + \mathbb{E}[q^2 \mathbb{I}_{|q| > 1}] \leqslant 1 + \mathbb{E}[q^4]^{1/2} p^{1/2} \leqslant 1 + 9 p^{1/2} \mathbb{E}[q^2]$$

which gives the result. □

Exercise 1. Show that the above result cannot hold (even for different constants) for more general random variables.

The rationality of social choice and the majority function

Recall the rationality probability (sometimes we denote with $|S|$ the cardinality of S)

$$\mathbb{P}(\mathrm{Rational}(f)) = \mathbb{P}(\mathrm{NAE}(f(x), f(y), f(z)) = 1) = \frac{3}{4} - \frac{3}{4} \sum_{S \subseteq [\![n]\!]} \left(-\frac{1}{3}\right)^{|S|} \hat{f}(S)^2$$

Two easy estimates are the following:

$$\mathbb{P}(\mathrm{Rational}(f)) \leqslant \frac{3}{4} + \frac{1}{4} W_1(f) \frac{3}{4} \sum_{S \subseteq [\![n]\!], \#(S) > 3} \left(-\frac{1}{3}\right)^{|S|} \hat{f}(S)^2$$

$$\leqslant \frac{3}{4} + \frac{1}{4} W_1(f) + \frac{1}{36} \sum_{S \subseteq [\![n]\!], \#(S) > 3} \hat{f}(S)^2$$

$$\leqslant \frac{3}{4} + \frac{1}{4} W_1(f) + \frac{1}{36} (1 - W_1(f)) = \frac{7}{9} + \frac{2}{9} W_1(f)$$

and if f is odd ($f(-x) = -f(x)$) which is the case when f is neutral, then $\hat{f}(S) = 0$ if $\#S$ is even and

$$\mathbb{P}(\mathrm{Rational}(f)) = \frac{3}{4} - \frac{3}{4} \sum_{S \subseteq [\![n]\!], \#S \text{ odd}} \left(-\frac{1}{3}\right)^{|S|} \hat{f}(S)^2$$

$$= \frac{3}{4} + \frac{3}{4} \sum_{S \subseteq [\![n]\!], \#S \text{ odd}} \left(\frac{1}{3}\right)^{|S|} \hat{f}(S)^2 \geqslant \frac{3}{4} + \frac{1}{4} W_1(f) \,.$$

If the function f is transitive then $\hat{f}(\{i\}) = \hat{f}(\{j\})$ for all $i, j \in [\![n]\!]$ and

$$W_1(f) = \sum_i \hat{f}(\{i\})^2 = n \hat{f}(\{1\})^2 = \frac{1}{n} \left(\sum_i |\hat{f}(\{i\})| \right)^2.$$

Lemma 4. *The quantity $\sum_i \hat{f}(\{i\})$ is maximized by the majority function.*

Proof.

$$\sum_i \hat{f}(\{i\}) = \mathbb{E}\left[\sum_i x_i f(x) \right] \leqslant \mathbb{E}\left[\left| \sum_i x_i \right| \right]$$

with equality only if $f(x) = \mathrm{sgn}(\sum_i x_i)$, the majority function. □

Remark 5. Note that as $n \to \infty$ by the CLT we have $n^{-1/2}\sum_i x_i \to G \sim \mathcal{N}(0,1)$ in law so that

$$n^{-1/2}\,\mathbb{E}\left[\left|\sum_i x_i\right|\right] \to \mathbb{E}[|G|] = \frac{2}{\sqrt{2\pi}}\int_0^\infty x e^{-x^2/2}\mathrm{d}x = \sqrt{\frac{2}{\pi}}\int_0^\infty e^{-z}\mathrm{d}z = \sqrt{\frac{2}{\pi}}\,.$$

The quantity $\sum_i \hat{f}(\{i\})$ has the following meaning

$$\mathbb{P}(x_i = f(x)) = \mathbb{E}\left[\frac{1}{2} + \frac{1}{2}x_i f(x)\right] = \frac{1}{2} + \frac{1}{2}\hat{f}(\{i\}).$$

Exercise 2. Show that majority is the only transitive social choice function between two alternatives which maximize

$$\mathbb{E}[\#(x_i = f(x))]$$

i.e. the expected number of voters which agree with the social choice.

Then we can conclude that at least asymptotically as $n \to \infty$

$$W_1(f) \leqslant \frac{1}{n}\left(\mathbb{E}\left[\left|\sum_i x_i\right|\right]\right)^2 \to \frac{2}{\pi}$$

and

$$\mathbb{P}(\text{Rational}(f)) \leqslant \frac{7}{9} + \frac{2}{9}W_1(f) \to 0.919$$

and for the majority function

$$\mathbb{P}(\text{Rational}(f)) \geqslant \frac{3}{4} + \frac{1}{4}W_1(f) \to 0.909$$

Actually it is possible to provide an exact asymptotic for $\mathbb{P}(\text{Rationality})$ in the case of the majority function. To do this we need to introduce the noise operator T_ρ which will also play an important rôle in the analysis of noise stability of Bfs.

The noise operator

Fix $\rho \in [-1, 1]$ the noise operator T_ρ is a linear operator defined on functions on Ω_n by the formula

$$T_\rho f(x) = \mathbb{E}[f(x\,y)]$$

where $y = (y_1, \ldots, y_n)$ is a vector of n independent random variables taking values in ± 1 such that $\mathbb{P}(y_i = +1) = (1+\rho)/2$.

Exercise 3. Show that $T_0 f(x) = \mathbb{E}[f(y)]$, $T_1 f = f$ and $T_\rho \varphi_S = \rho^{\#(S)}\varphi_S$ where φ_S is the parity function on $S \subseteq [\![n]\!]$. Show also that $\mathbb{E}[T_\rho f] = \mathbb{E}[f]$.

Lemma 6. *We have* $T_\rho \varphi_S = \rho^{|S|}\varphi_S$ *and*

$$T_\rho f(x) = \sum_{S \subseteq [\![n]\!]} \rho^{|S|}\hat{f}(S) x_S\,.$$

Proof.

$$\mathbb{E}[(x\,y)_S] = x_S\mathbb{E}[y_S] = x_S(\mathbb{E}[y_1])^{|S|} = x_S\left(\frac{1+\rho}{2} - \frac{1-\rho}{2}\right)^{|S|} = x_S\,\rho^{|S|}$$

and the second formula holds by linearity of the noise operator acting on the Fourier transform of f. □

Exercise 4. Use Lemma 6 to show that $T_\rho T_\sigma = T_{\rho\sigma}$ for all $\rho, \sigma \in [-1, 1]$.

We introduce also the notion of stability of a function on the cube Ω_n:

$$\mathbb{S}_\rho(f) = \mathbb{E}[f T_\rho f] = \mathbb{P}_\rho(f(x) = f(y)) - \mathbb{P}_\rho(f(x) \neq f(y))$$

where here $(x,\, y)$ is a random pair on Ω_n^2 such that x is uniform and $\mathbb{P}(x_i = y_i) = (1+\rho)/2$. Note that we have also

$$\mathbb{P}_\rho(f(x) = f(y)) = \frac{1}{2} + \frac{1}{2}\mathrm{Stab}_\rho(f).$$

In terms of the Fourier transform of f

$$\mathbb{S}_\rho(f) = \sum_{S \subseteq [\![n]\!]} \rho^{|S|}\hat{f}(S)^2$$

so the probability of rationality of f can be written as

$$\mathbb{P}(\mathrm{Rational}(f)) = \frac{3}{4} - \frac{3}{4}\sum_{S \subseteq [\![n]\!]} \left(-\frac{1}{3}\right)^{|S|}\hat{f}(S)^2 = \frac{3}{4} - \frac{3}{4}\mathbb{S}_{-1/3}(f).$$

Example 7. $\mathbb{S}_\rho(\mathrm{Dict}_i) = \rho$, $\mathbb{S}_\rho(\varphi_{[\![n]\!]}) = \rho^n$. This shows that the parity function is not stable since when $n \gg 1/(1-\rho)$ we have that the correlation between $\varphi_{[\![n]\!]}$ and $T_\rho\varphi_{[\![n]\!]}$ is quite small.

For the majority function we have the following formula.

Lemma 8. (Sheppard, 1899) *For all $\rho \in [0, 1]$, as $n \to \infty$,*

$$\mathbb{S}_\rho(\mathrm{Maj}_n) \to 1 - \frac{2}{\pi}\arccos(\rho).$$

Proof. We write $\mathbb{S}_\rho(\mathrm{Maj}_n) = 2\,\mathbb{P}_\rho(\mathrm{Maj}_n(x) = \mathrm{Maj}_n(y)) - 1$. Then we let $X_n = (x_1 + \cdots + x_n)/n^{1/2}$ and $Y_n = (y_1 + \cdots + y_n)/n^{1/2}$. By the two-dimensional CLT applied to the sequence of random vectors (X_n, Y_n) we have that

$$\mathbb{P}_\rho(\mathrm{Maj}_n(x) = \mathrm{Maj}_n(y)) \to \mathbb{P}_\rho(\mathrm{sgn}(X) = \mathrm{sgn}(Y))$$

where (X, Y) is a two-dimensional centered Gaussian random variable such that

$$\mathbb{E}[X^2] = \mathbb{E}[Y^2] = 1, \qquad \mathbb{E}[XY] = \rho.$$

But now

$$\mathbb{P}_\rho(\mathrm{sgn}(X)=\mathrm{sgn}(Y))=\mathbb{P}_\rho(X>0,Y>0)+\mathbb{P}_\rho(X<0,Y<0).$$

and the pair (X,Y) has the same law as the pair $(X,\cos(\theta)\,X+\sin(\theta)\,W)$ where (X,W) is a uncorrelated couple and $\cos(\theta)=\rho$. Then

$$\mathbb{P}_\rho(\mathrm{sgn}(X)=\mathrm{sgn}(Y))=\mathbb{P}(\cos(\theta)\,X+\sin(\theta)\,W>0,X>0)+\mathbb{P}(\cos(\theta)\,X+\sin(\theta)\,W<0,X<0)$$

and by a rotational symmetry argument this last probability is easily seen to correspond to the ratio $(\pi-\theta)/\pi$ so that at the end we get

$$\mathbb{P}_\rho(\mathrm{Maj}_n(x)=\mathrm{Maj}_n(y))\to 1-\frac{\theta}{\pi}=1-\frac{\arccos(\rho)}{\pi}$$

which gives the result. □

For $\rho=1-\varepsilon$ with $\varepsilon\ll 1$ we have $\mathbb{S}_\rho(\mathrm{Maj}_n)\sim 1-4\varepsilon^{1/2}/\pi$ so that with ε noise on the recording of votes the majority function gives only $2\varepsilon^{1/2}/\pi$ probability of having the wrong winner.

Moreover we can now state the precise asymptotics for the rationality with majority function

$$\mathbb{P}(\mathrm{Rational}(f))\to\frac{3}{4}-\frac{3}{4}\left(1-\frac{2}{\pi}\arccos\left(-\frac{1}{3}\right)\right)=\frac{3\arccos(-1/3)}{2\pi}\simeq 0.912.$$

This was first stated by Guilbaud [G. Guilbaud. Les théories de l'intérêt général et le problème logique de l'agrégration. Economie appliquée, 5:501–584, 1952].

Hypercontractivity, take two

The aim of this paragraph is to prove the more general hypercontractivity estimate contained in the following theorem

Theorem 9. (Bonami-Gross-Beckner) *For all ρ, p, q such that $(q-1)/(p-1)\leqslant\rho^{-2}$ we have*

$$\|T_\rho f\|_q\leqslant\|f\|_p.$$

To prove the BGB inequality we first observe that the operator T_ρ can be obtained as the composition of the operators $T_\rho^{(i)}$ defined as

$$T_\rho^{(i)}f(x)=\mathbb{E}[f(x_1,\ldots,x_i\,y_i,\ldots,x_n)]$$

where $y_i=\pm1$ is such that $\mathbb{P}(y_i=1)=(1+\rho)/2$. Then

$$T_\rho f=T_\rho^{(1)}\cdots T_\rho^{(n)}f$$

and each operator $T_\rho^{(i)}$ acts only on the i-th coordinate. By elementary properties of the L^p spaces on product measures we have

$$L^p(\Omega_n;\mathbb{R})=L^p(\Omega;L^p(\Omega_{n-1};\mathbb{R}))$$

and if we assume that

$$\left\| T_\rho^{(i)} f_{i,x} \right\|_{L^q(\Omega;\mathbb{R})} \leqslant \| f_{i,x} \|_{L^p(\Omega;\mathbb{R})} \qquad (1)$$

where $f_{i,x}(y_i) = f(x_1, \ldots, y_i, \ldots, x_n)$ (the y_i is at the i-th position) is a function on $\Omega = \{\pm 1\}$ then the norm 1 property transfer to the whole T_ρ operator. Then we can compute

$$\| T_\rho f \|_{L^q(\Omega_n)} = (\mathbb{E}_{x_1,\ldots,x_n} [| \mathbb{E}_{y_1,\ldots,y_n} f(x_1 y_1, x_2\, y_2, \ldots, x_n\, y_n) |^q])^{1/q}$$

$$= \left\| \| \mathbb{E}_{y_1} \mathbb{E}_{y_2,\ldots,y_n} f(x_1 y_1, x_2\, y_2, \ldots, x_n\, y_n) \|_{L^q_{x_2,\ldots,x_n}} \right\|_{L^q_{x_1}}$$

by Minkonswki inequality (convexity of norm) we get

$$\leqslant \left\| \mathbb{E}_{y_1} \| \mathbb{E}_{y_2,\ldots,y_n} f(x_1 y_1, x_2\, y_2, \ldots, x_n\, y_n) \|_{L^q_{x_2,\ldots,x_n}} \right\|_{L^q_{x_1}}$$

by the Hypercontractivity inequality on the first coordinate:

$$\leqslant \left\| \| \mathbb{E}_{y_2,\ldots,y_n} f(x_1, x_2\, y_2, \ldots, x_n\, y_n) \|_{L^q_{x_2,\ldots,x_n}} \right\|_{L^p_{x_1}}$$

$$= \left\| \left\| \| \mathbb{E}_{y_2} \mathbb{E}_{y_3,\ldots,y_n} f(x_1, x_2\, y_2, \ldots, x_n\, y_n) \|_{L^q_{x_3,\ldots,x_n}} \right\|_{L^q_{x_2}} \right\|_{L^p_{x_1}}$$

$$\leqslant \left\| \left\| \| \mathbb{E}_{y_3,\ldots,y_n} f(x_1, x_2, \ldots, x_n\, y_n) \|_{L^q_{x_3,\ldots,x_n}} \right\|_{L^p_{x_2}} \right\|_{L^p_{x_1}}$$

$$= \left\| \| \mathbb{E}_{y_3,\ldots,y_n} f(x_1, x_2, \ldots, x_n\, y_n) \|_{L^q_{x_3,\ldots,x_n}} \right\|_{L^p_{x_1,x_2}}$$

by repeating inductively the same argument we finally obtain

$$\leqslant \| f(x_1, \ldots, x_n) \|_{L^p_{x_1,\ldots,x_n}} = \| f \|_p .$$

Note that in this argument for ease of notation we denoted with $L^p_{x_1,\ldots,x_n}$ the space and the L^p norm on functions over the variables $x_1, \ldots, x_n$ taken with uniform distribution over all the possible outcomes.

It remains only to prove eq. (1) for the one-dimensional operators. Note first that we can consider only positive functions since $\| T_\rho f \|_q \leqslant \| T_\rho |f| \|_q \leqslant \| |f| \|_p = \| f \|_p$. Any $f\colon \Omega \to \mathbb{R}_+$ can be written as (modulo a multiplicative constant)

$$f(x) = 1 + a\, x$$

where $|a| \leqslant 1$ so

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

and

$$\| f \|_p = \left(\frac{(1 + a)^p + (1 - a)^p}{2} \right)^{1/p}$$

So it remains to prove that for the right range of parameters and for all $a \in \mathbb{R}$:

$$\left(\frac{(1+a\,\rho)^q+(1-a\,\rho)^q}{2}\right)^{1/q} \leqslant \left(\frac{(1+a)^p+(1-a)^p}{2}\right)^{1/p}$$

Note that when $a=0$ the inequality is trivially verified and when $a \ll 0$ by series expansion we get

$$\frac{(1+a\,\rho)^q+(1-a\,\rho)^q}{2} = 1+\frac{q(q-1)}{2}a^2\rho^2+O(a^4)$$

and

$$\left(\frac{(1+a\,\rho)^q+(1-a\,\rho)^q}{2}\right)^{1/q} = 1+\frac{(q-1)}{2}a^2\rho^2+O(a^4)$$

and similarly

$$\left(\frac{(1+a)^p+(1-a)^p}{2}\right)^{1/p} = 1+\frac{(p-1)}{2}a^2+O(a^4)$$

so that the inequality can be verified only if

$$(q-1)\rho^2 \leqslant (p-1).$$

A more technical argument allows to prove that this is a sufficient condition for the validity of the inequality for all $|a| \leqslant 1$.

Influences and the Kahn-Kalai-Linial theorem

Kalai's robust form of Arrow's theorem tells us that we can't hope for a fair election rule that evades the Condorcet's paradox with probability close to 1. Given this we would like to look at fair election rules which give rational outcomes with the highest possible probability. In this case we should decide what means exactly "fair" in this context. Of course a reasonable notion of "fairness" should rule out dictators. A possibility would be to have a symmetric function under a transitive group of permutations. In this way no elector is preferred over the others. A weaker and more useful notion of fairness is provided by the condition of having "small influences".

Definition 10. *Given a Bf f we define the influence $\operatorname{Inf}_i(f)$ of the i-th variable as the quantity*

$$\operatorname{Inf}_i(f) = \mathbb{P}(f(x) \neq f(x\,e_i))$$

where $(e_i)_j = (-1)^{\mathbb{I}_{i=j}}$.

So $\operatorname{Inf}_i(f)$ is the probability that flipping the i-th input of f the outcome change. This notion of "influence" or "power" of a voter was first introduced by Penrose [L. Penrose. The elementary statistics of majority voting. J. of the Royal Statistical Society, 109(1):53-57, 1946] and later rediscovered by the lawyer Banzhaf [J. Banzahf. Weighted voting doesn't work: A mathematical analysis. Rutgers Law Review, 19(2):317–343, 1965] and is usually called the "Banzhaf power index" in the Social Choice literature. It has played a role in several United States court decisions [D. Felsenthal and M. Machover. The Measurement of Voting Power: Theory and Practice, Problem and Paradoxes. Edward Elgar, 1998].

A function with τ-small influences is a Bf such that $\max_{i \in [\![n]\!]} \operatorname{Inf}_i(f) \leqslant \tau$. One of the most important results for the class of functions with small influences is the Kahn-Kalai-Linial (KKL) theorem. The quantity $\mathcal{E}(f) = \sum_i \operatorname{Inf}_i(f)$ is called "energy" of f.

Theorem 11. (KKL) *No Bf f with zero bias has $o(\log n/n)$-small influences. More generally if f is unbiased and has τ-small influences then $\mathcal{E}(f) \geqslant O(\log(1/\tau))$.*

The bounds in the theorem are saturated (for different constants) by the Tribes_n function introduced by Ben-Or and Linial [Ben-Or and Linial, Collective Coin Flipping in "Randomness and Computation" (S. Micali ed.) Academic Press, New York, (1989), 91–115] which is defined by partitioning the n electors into $n/\log(n)$ "tribes" of $\log(n)$ electors and returning 1 if at least one of the tribes vote unanimously 1 and -1 otherwise.

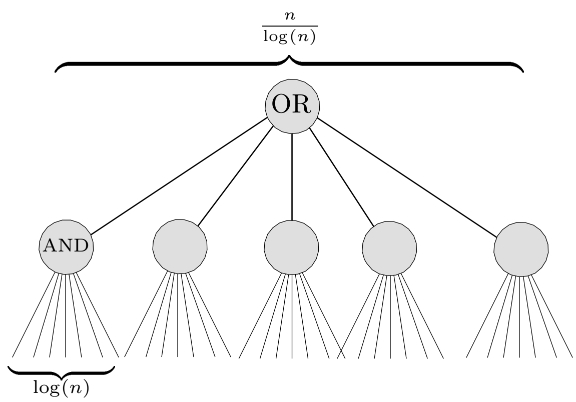

Figure 1. The Tribes_n function

Exercise 5. Show that $\mathbb{E}[\text{Tribes}_n] \simeq 0$ and that $\text{Inf}_i(\text{Tribes}_n) \simeq C \log n/n$.

To prove the KKL theorem we need some preliminary ingredients.

Lemma 12.

$$\text{Inf}_i(f) = \sum_{S \ni i} \hat{f}(S)^2 \qquad \mathcal{E}(f) = \sum_{S \subseteq [\![n]\!]} |S|\, \hat{f}(S)^2$$

Proof. Define the discrete derivative $L_i f = (f(x) - f(x\, e_i))/2$. We start by computing

$$f(x\, e_i) = \sum_{S \subseteq [\![n]\!]} \hat{f}(S)(-1)^{\mathbb{I}_{i \in S}} x_S$$

and

$$L_i f(x) = \frac{f(x) - f(x\, e_i)}{2} = \frac{1}{2} \sum_{S \subseteq [\![n]\!]} \hat{f}(S) x_S((-1)^{\mathbb{I}_{i \in S}} - 1) = \sum_{S \subseteq [\![n]\!], S \ni i} \hat{f}(S) x_S\,.$$

Then we observe that

$$\text{Inf}_i(f) = \mathbb{E}[|L_i f|^2] = \frac{1}{4}\mathbb{E}[(f(x\, e_i) - f(x))^2] = \sum_{S \subseteq [\![n]\!], S \ni i} \hat{f}(S)^2$$

and

$$\mathcal{E}(f) = \sum_{i \in [\![n]\!]} \sum_{S \subseteq [\![n]\!], S \ni i} \hat{f}(S)^2 = \sum_{S \subseteq [\![n]\!]} \sum_{i \in [\![n]\!]} \mathbb{I}_{i \in S} \hat{f}(S)^2 = \sum_{S \subseteq [\![n]\!]} |S|\, \hat{f}(S)^2.$$

□

Proof. (of the KKL theorem) I found this argument in an unpublished note of P. Pansu. Let $H = T_{e^{-1}}$, then

$$\mathcal{E}(Hf) = \sum_{S \subseteq [\![n]\!], |S| \geqslant 1} |S| e^{-2|S|} \hat{f}(S)^2 = \sum_{S \subseteq [\![n]\!], |S| \geqslant 1} g(|S|) \hat{f}(S)^2$$

where $g(x) = x\, e^{-2x}$ is a concave function for $x \geqslant 1$. By Jensen's inequality (using the fact that $\sum_{S \subseteq [\![n]\!], |S| \geqslant 1} \hat{f}(S)^2 = \mathrm{Var}(f)$)

$$\mathcal{E}(Hf) \geqslant \mathrm{Var}(f) g\left(\frac{\sum_{S \subseteq [\![n]\!]} |S| \hat{f}(S)^2}{\mathrm{Var}(f)} \right) = \mathcal{E}(f) e^{-2\mathcal{E}(f)/\mathrm{Var}(f)}.$$

$$\geqslant \mathcal{E}(f) \exp(-2\mathcal{E}(f)/\mathrm{Var}(f)).$$

Now we apply the hypercontractivity estimate with $q = 2$ and $p = 2/(1+\varepsilon)$ with ε small to the function $L_i f = (f(x) - f(x\, e_i))/2 \in \{-1, 0, 1\}$. Then

$$\mathbb{E}[|H L_i f|^2] \leqslant (\mathbb{E}[|L_i f|^p])^{2/p} = (\mathbb{E}[|L_i f|^2])^{1/p} = (\mathrm{Inf}_i(f))^{2/p} = \mathrm{Inf}_i(f)\, (\mathrm{Inf}_i(f))^{\varepsilon}$$

(Show that $L_i H = H L_i$). Which gives

$$\mathcal{E}(Hf) = \sum_i \mathbb{E}[|L_i H f|^2] = \sum_i \mathbb{E}[|H L_i f|^2] \leqslant \sum_i \mathrm{Inf}_i(f)\, (\mathrm{Inf}_i(f))^{\varepsilon}$$

$$\leqslant \mathcal{E}(f) \Big(\max_i \mathrm{Inf}_i(f) \Big)^{\varepsilon}$$

Then combining these two inequalities we get

$$\exp(-2\mathcal{E}(f)) \leqslant \Big(\max_i \mathrm{Inf}_i(f) \Big)^{\varepsilon}$$

If the function f has τ-small influences then

$$\mathcal{E}(f) \geqslant O(\log(1/\tau)).$$

and if we let $M = \max_i \mathrm{Inf}_i(f)$ then

$$e^{-2n M/\mathrm{Var}(f)} \leqslant M^{\varepsilon} \Rightarrow -n M/\mathrm{Var}(f) \leqslant \varepsilon \log(M) \Rightarrow M \geqslant O(\log n/n) \mathrm{Var}(f)$$

Indeed fix $0 < \alpha < 1$. Then if $M > D\, \mathrm{Var}(f)\, n^{-\alpha}$ we obviously have $M \geqslant O(\log n/n)\mathrm{Var}(f)$ while if $M \leqslant D\, \mathrm{Var}(f) n^{-\alpha}$ we obtain

$$n M/\mathrm{Var}(f) \geqslant -\varepsilon \log(M) \geqslant -\varepsilon \log(D) - \varepsilon\, \mathrm{Var}(f) + \alpha \log(n)$$

and then

$$M \geqslant \mathrm{Var}(f) O(\log n/n).$$

□

Another interesting property of Boolean functions related to total influence is the fact that it measures the "support" of the function, that is, the size of the set of variables that essentially determines the outcome. This is a theorem of Friedgut. A k-junta is a Boolean function with depends at most on a subset of $k \leqslant n$ variables (among the n possible)

Theorem 13. (Friedgut) *Any Boolean function f is ε-close to a $2^{O(\mathcal{E}(f)/\varepsilon)}$-junta.*

Proof. We start by estimating the Fourier spectrum of f using the size of $\mathcal{E}(f)$ as in KKL. Fix $0 \leqslant d \leqslant n$, then

$$\sum_{S \subseteq [\![n]\!], |S| > d} \hat{f}(S)^2 \leqslant \frac{1}{d} \sum_{S \subseteq [\![n]\!], |S| > d} |S| \hat{f}(S)^2 \leqslant \frac{\mathcal{E}(f)}{d}$$

Then let $L = \{i \in [\![n]\!] : \mathrm{Inf}_i(f) \leqslant \tau\}$ for some fixed $0 \leqslant \tau \leqslant 1$. Then

$$\sum_{S \subseteq [\![n]\!], |S| \leqslant d, S \cap L \neq \varnothing} \hat{f}(S)^2 \leqslant \sum_{S \subseteq [\![n]\!], |S| \leqslant d, S \cap L \neq \varnothing} |S \cap L| \hat{f}(S)^2$$

$$\leqslant \sum_{i \in [\![n]\!]} \sum_{S \subseteq [\![n]\!], |S| \leqslant d} \mathbb{I}_{i \in S \cap L} \hat{f}(S)^2 = \sum_{i \in L} \sum_{S \subseteq [\![n]\!], |S| \leqslant d, S \ni i} \hat{f}(S)^2$$

$$\leqslant e^{2d} \sum_{i \in L} \sum_{S \subseteq [\![n]\!], |S| \leqslant d, S \ni i} e^{-2|S|} \hat{f}(S)^2$$

$$\leqslant e^{2d} \sum_{i \in L} \sum_{S \subseteq [\![n]\!], S \ni i} e^{-2|S|} \hat{f}(S)^2 = e^{2d} \sum_{i \in L} \|L_i H f\|_2^2 = e^{2d} \sum_{i \in L} \mathbb{E}[|L_i H f|^2]$$

$$\leqslant e^{2d} \sum_{i \in L} \mathrm{Inf}_i(f) \, (\mathrm{Inf}_i(f))^{\sigma} \leqslant e^{2d} \tau^{\sigma} \mathcal{E}(f)$$

Then we can choose $\tau = (\varepsilon e^{-2d}/2)^{1/\sigma}$ and $d = 2\mathcal{E}(f)/\varepsilon$ to have

$$\sum_{|S| \leqslant d, S \subseteq J} \hat{f}(S)^2 \geqslant 1 - \varepsilon.$$

At this point it is enough to show that the function $g = \mathrm{sgn}(h)$ where

$$h(x) = \sum_{|S| \leqslant d, S \subseteq L^c} \hat{f}(S) x_S$$

is ε-close to f. When $f \neq g$ we have that f, h do not have the same sign so $(f - h)^2 > 1$ but if $\mathbb{P}(f = g) > \varepsilon$ then $\mathbb{E}[(f - h)^2] \geqslant \mathbb{P}(f \neq g) > \varepsilon$. So it is enough to prove that $\mathbb{E}[(f - h)^2] \leqslant \varepsilon$ which means that

$$\mathbb{E}[(f - h)^2] = \sum_{S : |S| > d \text{ or } S \cap L \neq \varnothing} \hat{f}(S)^2 \leqslant \varepsilon$$

which is exactly what we have. The support of the function h (and thus of the function g) is of size $M = |L^c|$. This means that $\mathcal{E}(f) \geqslant M\tau$ which gives

$$M \leqslant \mathcal{E}(f)/\tau \leqslant (2/\varepsilon)^{1/\sigma} \mathcal{E}(f) e^{4\mathcal{E}(f)/\varepsilon\sigma} \leqslant 2^{O(\mathcal{E}(f)/\varepsilon)}$$

as claimed. □

We will apply now eq. (2) to show the following remarkable property of balanced election schemes:

Theorem 14. *In any balanced election scheme there is a coalition of fractional size at most* $O(1/\log n)$ *which controls the election with probability* 0.99.

In order to formulate properly this fact we introduce the notion of influence $\mathrm{Inf}_J(f)$ of a coalition $J \subseteq [\![n]\!]$ on the Boolean function f as the probability that the random function

$$\{\pm 1\}^J \ni x \mapsto f_J(x) = f(y \otimes_J x)$$

is not constant when y is chosen uniformly in Ω_n and $(y \otimes_J x)_i = y_i$ if $i \notin J$ and $(y \otimes_J x)_i = x_i$ if $i \in J$ is the insertion of the values $x \in \{\pm 1\}^J$ in the vector $y \in \{\pm 1\}^n$ at the positions specified by the set J. So

$$\mathrm{Inf}_J(f) = \mathbb{P}_y(\exists x, x' \in \{\pm 1\}^J \text{ such that } f(y \otimes_J x) \neq f(y \otimes_J x')).$$

Note that $\mathrm{Inf}_{\{i\}}(f) = \mathrm{Inf}_i(f)$ (the influence of the i-th individual).

We define also the partial influences toward ± 1 as

$$\mathrm{Inf}_J^{\pm}(f) = \mathbb{P}_y(\exists x \in \{\pm 1\}^J \text{ such that } f(y \otimes_J x) = \pm 1) - \mathbb{P}(f = \pm 1).$$

Define $A^{\pm} = \{\exists x \in \{\pm 1\}^J \text{ such that } f(y \otimes_J x) = \pm 1\}$ and $B^{\pm} = \{f(y) = \pm 1\}$ then it is clear that $B^{\pm} \subseteq A^{\pm}$ and this implies that

$$\mathrm{Inf}_J^{\pm}(f) = \mathbb{P}_y(A^{\pm} \backslash B^{\pm}) \in [0, 1].$$

Moreover we can easily show that

$$\{\exists x, x' \in \{\pm 1\}^J \text{ such that } f(y \otimes_J x) \neq f(y \otimes_J x')\} = C^+ \cup C^-$$

where $C^{\pm} = A^{\pm} \backslash B^{\pm} = \{f(y) = \mp 1, \exists x \in \{\pm 1\}^J \text{ such that } f(y \otimes_J x) = \pm 1\}$ giving

$$\mathrm{Inf}_J^{+}(f) + \mathrm{Inf}_J^{-}(f) = \mathrm{Inf}_J(f).$$

Exercise 6. Prove that for a monotone function f we have

$$\mathbb{P}(f_{+,i} = 1) = \mathbb{P}(f = 1) + \frac{1}{2}\mathrm{Inf}_i(f) \tag{2}$$

where $f_{+,i}(x) = f(x_1, \ldots, x_{i-1}, +1, x_{i+1}, \ldots, x_n) = f(x \otimes_{\{i\}} (+1))$.

Theorem 14 readily follows form the next lemma.

Lemma 15. *If f is such that* $\mathrm{Var}(f) \geqslant O(1)$ *then for all* $\varepsilon > 0$ *there exists a coalition* $J \subseteq [\![n]\!]$ *of size at most* $O(\log 1/\varepsilon)\, n/\log n$ *such that* $\mathrm{Inf}_J(f) \geqslant 1 - \varepsilon$.

Proof. We will assume that f is monotone (an independent argument will allow to lift this hypothesis). Consider the following algorithm.

Given f with $\mathbb{P}(f = 1) = p \in [1 - 2\delta, 1 - \delta]$ then $\mathrm{Var}(f) = 1 - \mathbb{E}[f]^2 = 1 - p^2 \geqslant 1 - (1 - \delta)^2 \geqslant 2\delta$. By KKL in the form of eq. (2) we know that there exists a coordinate i with influence $\gamma \geqslant C\delta \log n/n$ at least. If we look at the restriction $f_{+,i}: \Omega_{n-1} \to \{\pm 1\}$ where we have fixed the i-th coordinate to $+1$, by eq. (2) we have

$$\mathbb{P}(f_{+,i} = 1) = p + \gamma/2 \geqslant p + C\delta \log n/n \geqslant 1 - 2\delta + C\delta \log n/n$$

Substitute $f_{+,i} \rightarrow f$ and repeat the process until $\mathbb{P}(f = 1) \geqslant 1 - \delta$. So if we repeat the process r times where

$$1 = O(1) \sum_{k=n-r}^{r} \frac{\log k}{k} = O(1)(\log^2 n - \log^2 (n - r)).$$

But setting $\xi = r/n$,

$$\log^2 n - \log^2 (n - r) = \log^2 n - (\log n + \log (1 - \xi))^2 = -2 \log n \log (1 - \xi) - \log^2 (1 - \xi)$$

$$\sim 2 \log n \, \xi + O(\xi^2)$$

so that we must require $\xi \sim 1/\log n$ and we get that repeating the process $n/\log n$ times we pass from a function which has $\mathbb{P}(f = 1) = 1 - 2\,\delta$ to a function such that $\mathbb{P}(f = 1) = 1 - \delta$. Iterating this block of selections n times we can pass for a function with $\mathbb{P}(f = 1) = 1 - 2^n\,\delta$ to a function with $\mathbb{P}(f = 1) = 1 - \delta$. This means that in at most $\log(1/\varepsilon)$ iterations we can take any Boolean function to a function with $\mathbb{P}(f = 1) = 1 - \varepsilon$. Along the way we have selected a coalition J containing $O(1)\ (n/\log n)\ \log 1/\varepsilon$ individuals. Since we have started from $\mathbb{P}(f = 1) = 1/2$ (balanced function we got a function f with $\mathrm{Inf}_J^+(f) = 1/2 - \varepsilon$. In the same way we can work with $\{f = -1\}$ to select a coalition J' with as many individuals as J such that $\mathrm{Inf}_{J'}^-(f) = 1/2 - \varepsilon$. Summing these together we get $\mathrm{Inf}_{J \cup J'}(f) \geqslant 1 - 2\varepsilon$, completing the proof. □