Boolean function analysis

Yuval Filmus

June 18, 2019

1 Basics

1.1 Fourier expansion

In this class, we will study Boolean function analysis on various domains, starting with the *Boolean cube* $\{-1, 1\}^n$.

Question 1.1.1. Let $f\colon \{-1, 1\}^n \to \mathbb{R}$.

(a) Show that $f(x_1, \ldots, x_n)$ can be written as a polynomial in which no variable appears squared (such a polynomial is called *multilinear*).

Hint: There are many possible proofs. One option is induction on n.

(b) Show that this representation is unique.

Hint: Once again, there are many possible proofs. You can use a dimension argument, or induction.

The unique representation of a function on $\{-1, 1\}^n$ as a multilinear polynomial is known as its *Fourier expansion*:

$$f = \sum_{S \subseteq [n]} \hat{f}(S)\chi_S, \text{ where } \chi_S(x) = \prod_{i \in S} x_i \quad (\chi_\emptyset \equiv 1).$$

The coefficients $\hat{f}(S)$ are known as *Fourier coefficients*. The functions χ_S are known as *Fourier characters*, because they are characters of the group $\mathbb{Z}_2^n$. They form the *Fourier basis* for the vector space of real-valued functions on the Boolean cube.

The *degree* of f is the degree of its Fourier expansion.

Question 1.1.2. Show that $f\colon \{-1, 1\}^n \to \mathbb{R}$ depends only on the coordinates in S if and only if its Fourier expansion is supported on subsets of S, i.e., if $\hat{f}(T) \neq 0$ only for $T \subseteq S$.

Question 1.1.3. Given a function $f\colon \{-1, 1\}^n \to \mathbb{R}$, extend it to $\mathbb{R}^n$ using the unique multilinear representation, and define a function $g\colon \{-1, 1\}^{n-1} \to \mathbb{R}$ by

$$g(x_1, \ldots, x_{n-1}) = f(x_1, \ldots, x_{n-1}, 0).$$

(a) Show that

$$g(x_1, \ldots, x_{n-1}) = \frac{f(x_1, \ldots, x_{n-1}, -1) + f(x_1, \ldots, x_{n-1}, 1)}{2}.$$

(b) Give a formula for $\hat{f}(\emptyset)$.

1.2 Orthogonality

A crucial property of the Fourier characters is that they form an orthonormal basis with respect to the inner product

$$\langle f, g \rangle = \mathbb{E}[fg] = \frac{1}{2^n} \sum_{x \in \{-1,1\}^n} f(x)g(x).$$

The corresponding norm is

$$\|f\|^2 = \langle f, f \rangle = \mathbb{E}[f^2].$$

Question 1.2.1. In this question, we will show that the Fourier characters form an orthonormal basis. That is, $\|\chi_S\| = 1$ and $\langle \chi_S, \chi_T \rangle = 0$ for $S \neq T$.

(a) Show that $\chi_S \chi_T = \chi_R$ for an appropriate R.

(b) Show that $\mathbb{E}[\chi_\emptyset] = 1$ and $\mathbb{E}[\chi_S] = 0$ if $S \neq \emptyset$.

Hint: Choose $i \in S$, and consider pairs of inputs different only on the i'th coordinate.

(c) Show that the Fourier characters form an orthonormal basis.

Using linearity of the inner product, we can prove several useful formulas:

Question 1.2.2. (a) Show that $\hat{f}(S) = \langle f, \chi_S \rangle$.

(b) Reprove the formula for $\hat{f}(\emptyset)$.

(c) Show that $\langle f, g \rangle = \sum_S \hat{f}(S)\hat{g}(S)$.

(d) Prove *Parseval's identity:* $\|f\|^2 = \sum_S \hat{f}(S)^2$.

(e) Find a formula for $\mathbb{V}[f]$ (the variance of f).

A function whose range is $\{0,1\}$ or $\{-1,1\}$ is called *Boolean* (both conventions are used).

Question 1.2.3. Let $S \subseteq \{-1,1\}^n$, and let $f\colon \{-1,1\}^n \to \{0,1\}$ be the indicator function of S.

(a) Calculate $\hat{f}(\emptyset)$ in terms of $|S|$.

(b) Calculate $\sum_T \hat{f}(T)^2$ in terms of $|S|$.

1.3 Linearity testing

We can now present an application of the Fourier expansion, to property testing.

We consider the following scenario: we are given oracle access to a function $f\colon \{-1,1\}^n \to \{-1,1\}$, which somebody claims is equal to some character χ_S. We want to test this claim by sampling a few (correlated) values of f.

Question 1.3.1. Show that the Fourier characters satisfy the formula

$$\chi_S(x)\chi_S(y)\chi_S(xy) = 1, \text{ where } xy \triangleq (x_1 y_1, \ldots, x_n y_n).$$

This suggests the following text:

$$\boxed{\text{Sample } x, y \in \{-1,1\}^n, \text{ and test that } f(x)f(y) = f(xy).}$$

Suppose that a function $f\colon \{-1,1\}^n \to \{-1,1\}$ satisfies $f(x)f(y)f(xy) = 1$ with high probability, or at least with nontrivial probability (larger than 1/2). What can we say about f? The following exercises will answer this question.

Question 1.3.2. Show that

$$\sigma(f) := \mathop{\mathbb{E}}_{x,y}[f(x)f(y)f(xy)] = \sum_S \hat{f}(S)^3.$$

Hint: Expand $f(x)f(y)f(xy)$ into a triple sum using the Fourier expansion, and compute the expectation (with respect to x, y) of each summand.

(The quantity $\sigma(f)$ is similar to the Gowers U^2 norm.)

Question 1.3.3. Define

$$p(f) = \Pr[f(x)f(y)f(xy) = 1].$$

(a) Express $p(f)$ in terms of $\sigma(f)$.

(b) Show that if $p(f) \geq 1 - \epsilon$ then

$$\max_S \hat{f}(S) \geq 1 - 2\epsilon.$$

Hint: Use $\sum_S \hat{f}(S)^2 = 1$.

(c) Deduce that for the S from part (b),

$$\Pr[f = \chi_S] \geq 1 - \epsilon.$$

Hint: Show that $\Pr[f = \chi_S] = \frac{1}{2} + \frac{1}{2}\mathbb{E}[f\chi_S]$.

(d) Show that if $p(f) \geq 1/2 + \delta$ then $\mathbb{E}[f\chi_S] \geq 2\delta$ for some S (where $\delta \geq 0$).

(e) Show that $\mathbb{E}[f\chi_S] \geq 2\delta$ for at most $1/(2\delta)^2$ many characters χ_S.

1.4 Influences

For a function $f \colon \{-1,1\}^n \to \mathbb{R}$ and a coordinate $i \in [n]$, we define the Laplacian in direction i by

$$L_i f(x_1, \ldots, x_n) = \frac{f(x_1, \ldots, x_i, \ldots, x_n) - f(x_1, \ldots, -x_i, \ldots, x_n)}{2},$$

and the i'th influence by

$$\mathrm{Inf}_i[f] = \|L_i f\|^2.$$

Question 1.4.1. (a) When is $\mathrm{Inf}_i[f] = 0$?

(b) Show that if $f \colon \{-1,1\}^n \to \{-1,1\}$ then

$$\begin{aligned}\mathrm{Inf}_i[f] &= \Pr[f(x_1, \ldots, x_i, \ldots, x_n) \neq f(x_1, \ldots, -x_i, \ldots, x_n)] \\ &= \Pr[f(x_1, \ldots, 1, \ldots, x_n) \neq f(x_1, \ldots, -1, \ldots, x_n)].\end{aligned}$$

(c) Show that if $f \colon \{-1,1\}^n \to \{-1,1\}$ is a monotone function (that is, $x \leq y$ entrywise implies $f(x) \leq f(y)$) then

$$\mathrm{Inf}_i[f] = \mathbb{E}[x_i f] = \hat{f}(\{i\}).$$

Hint: Use $f(x_1, \ldots, 1, \ldots, x_n) \geq f(x_1, \ldots, -1, \ldots, x_n)$.

Question 1.4.2. (a) Compute $L_i \chi_S$.

Hint: Consider two cases: $i \in S$ and $i \notin S$.

(b) Compute the Fourier expansion of $L_i f$ in terms of the Fourier expansion of f.

(c) Deduce a formula for $\mathrm{Inf}_i[f]$ in terms of the Fourier coefficients:

$$\mathrm{Inf}_i[f] = \sum_{S \ni i} \hat{f}(S)^2.$$

We define the Laplacian of f by

$$Lf = \sum_{i=1}^{n} L_i f$$

and the total influence (or average sensitivity) by

$$\mathrm{Inf}[f] = \sum_{i=1}^{n} \mathrm{Inf}_i[f].$$

Question 1.4.3. Let $f \colon \{-1,1\}^n \to \{-1,1\}$.

The sensitivity of f at a point x is the number of neighbors y (i.e., points differing from x at a single coordinate) such that $f(x) \neq f(y)$. Show that $\mathrm{Inf}[f]$ is the average sensitivity of f.

Question 1.4.4. Let $f \colon \{-1,1\}^n \to \mathbb{R}$.

(a) Relate $\mathbb{V}[f]$ and $\mathbb{E}[(f(x) - f(y))^2]$, where the expectation is over two independent random points x, y in the hypercube.

(b) Relate $\mathrm{Inf}[f]$ and $\sum_{x \sim y}(f(x) - f(y))^2$, where the sum is over all edges of the hypercube $\{-1,1\}^n$, i.e., over all pairs of points differing in a single coordinate.

Question 1.4.5. (a) Give a formula for $\mathrm{Inf}[f]$ in terms of the Fourier coefficients.

(b) Give a formula for Lf in terms of the Fourier coefficients.

(c) Deduce that $\mathrm{Inf}[f] = \langle f, Lf \rangle$.

Bonus: Prove this directly from the spatial definitions of Lf and $\mathrm{Inf}[f]$ (the latter, in Question 1.4.4).

(d) Prove the (double-sided) *Poincaré's inequality*:

$$\mathbb{V}[f] \leq \mathrm{Inf}[f] \leq \deg f \cdot \mathbb{V}[f].$$

Can you prove $\mathbb{V}[f] \leq \mathrm{Inf}[f]$ combinatorially, using the definitions in Question 1.4.4? (hard)

Which monotone Boolean function has maximal total influence?

Question 1.4.6. Let $f \colon \{-1,1\}^n \to \{-1,1\}$ be monotone.

(a) Show that $\mathrm{Inf}[f] = \mathbb{E}[(x_1 + \cdots + x_n)f]$.

(b) Which monotone function maximizes $\mathrm{Inf}[f]$?

(c) *Bonus:* Estimate the total influence of this function.

1.5 Fourier levels

We can break up the Fourier expansion of a function into its homogeneous parts:

$$f^{=d} = \sum_{|S|=d} \hat{f}(S)\chi_S.$$

These are known as the *Fourier levels*.

Question 1.5.1. (a) Show that $f = \sum_{d=0}^n f^{=d}$.

(b) Show that $\|f\|^2 = \sum_{d=0}^n \|f^{=d}\|^2$.

Hint: Use Parseval's identity.

(c) Define $f^{\leq d}, f^{<d}, f^{\geq d}, f^{>d}$ and show that $f = f^{\leq d} + f^{>d} = f^{<d} + f^{\geq d} = f^{<d} + f^{=d} + f^{>d}$.

(d) Show that $\|f\|^2 = \|f^{\leq d}\|^2 + \|f^{>d}\|^2 = \|f^{<d}\|^2 + \|f^{\geq d}\|^2 = \|f^{<d}\|^2 + \|f^{=d}\|^2 + \|f^{>d}\|^2$.

There are succinct expressions for the Laplacian and total influence in terms of the Fourier levels. The underlying reason for the existence of these expressions is the fact that the Laplacian is "symmetric with respect to renaming variables".

Question 1.5.2. (a) Show that $Lf = \sum_{d=0}^n df^{=d}$.

(b) Show that $\mathrm{Inf}[f] = \sum_{d=0}^n d\|f^{=d}\|^2$.

Which balanced Boolean functions have minimal average sensitivity?

Question 1.5.3. Let $f\colon \{-1,1\}^n \to \{-1,1\}$ be balanced: $\mathbb{E}[f] = 0$.

(a) Show that $\mathrm{Inf}[f] \geq 1$.

Hint: Use Poincaré's inequality.

(b) Show that if $\mathrm{Inf}[f] = 1$ then $\deg f = 1$.

(c) Determine all balanced Boolean functions with minimal total influence.

Hint: If $f = \sum_{i=1}^n a_i x_i$ is Boolean then no two a_i, a_j can be non-zero.

1.6 Nisan–Szegedy*

Suppose that $f\colon \{-1,1\}^n \to \{-1,1\}$ has degree d. On how many variables can f depend?

Question 1.6.1. Suppose that $g\colon \{0,1\}^n \to \mathbb{Z}$. Show that $g(y_1, \ldots, y_n)$ can be expressed uniquely as a multilinear polynomial in $y_1, \ldots, y_n$ with integer coefficients.

Hint: Use induction on n. Alternatively, show how to express δ functions as multilinear polynomials.

Question 1.6.2. Suppose that $f\colon \{-1,1\}^n \to \{-1,1\}$ has degree $d \geq 1$.

(a) Let $y_i = \frac{1+x_i}{2}$, and define

$$g(y_1, \ldots, y_n) = \frac{1 + f(x_1, \ldots, x_n)}{2}.$$

Relate the Fourier expansion of f and the multilinear expansion of g.

(b) Show that $\deg g = d$ (as a polynomial).

(c) Deduce that all Fourier coefficients of f are integer multiples of 2^{1-d}.

(d) Show that $|\hat{f}(\{i\})| \leq \mathbb{E}[|L_i f|]$.

Hint: Use the triangle inequality.

(e) Show that $\mathrm{Inf}_i[f] = \mathbb{E}[|L_i f|]$, and deduce that for every i, either $\mathrm{Inf}_i[f] = 0$ or $\mathrm{Inf}_i[f] \geq 2^{1-d}$.

(f) Show that $\mathrm{Inf}[f] \leq d$, and deduce that f depends on at most $d2^{d-1}$ coordinates.

Recently, the upper bound has been improved to $O(2^d)$ by Chiarelli, P. Hatami and Saks. The hidden constant was subsequently improved by Wellens.

Question 1.6.3. (a) Given a degree $d-1$ function depending on C coordinates, construct a degree d function depending on $2C+1$ coordinates.

(b) Deduce that there exists a degree d function depending on $2^d - 1$ coordinates.

(c) Can you give a one-shot description of these functions?

We can extend the upper bound to functions attaining a constant number of values.

Question 1.6.4. Suppose that A is a finite set, and $f\colon \{-1,1\}^n \to A$ has degree d. Show that there exists a constant $C = C(d, A)$ such that f depends on at most C coordinates. (In fact, C need only depend on d and $|A|$.)

Hint: Write $f = \sum_{a \in A} f_a$, *where* $f_a(x) = 1$ *if* $f(x) = a$ *and* $f_a(x) = 0$ *otherwise.*

1.7 Noise operator

There are several ways to define the noise operator. We will see four equivalent ways.

Question 1.7.1. Let $0 \le \rho \le 1$, and $x \in \{-1,1\}^n$. Consider the following two distributions on $y, z \in \{-1,1\}^n$:

$$y_i = \begin{cases} +x_i & \text{w.p. } \frac{1+\rho}{2}, \\ -x_i & \text{w.p. } \frac{1-\rho}{2}, \end{cases} \qquad z_i = \begin{cases} x_i & \text{w.p. } \rho, \\ \text{random } \pm 1 & \text{w.p. } 1-\rho. \end{cases}$$

In both cases, the different coordinates are chosen independently.

(a) Show that y, z have the same distribution.

We denote the common distribution as $N_\rho(x)$.

(b) Calculate $\mathbb{E}[x_i y_i] = \mathbb{E}[x_i z_i]$.

(c) Let (x_1, y_1) be obtained by choosing x_1 uniformly from $\{\pm 1\}$ and y_1 according to $N_\rho(x_1)$. Let (x_2, y_2) be obtained by choosing y_2 uniformly from $\{\pm 1\}$ and x_2 according to $N_\rho(y_2)$.

Show that (x_1, y_1) and (x_2, y_2) are identically distributed, and describe their common distribution.

We denote this common distribution N_ρ; if $(x, y) \sim N_\rho$ then we say that x, y are *ρ-correlated.*

The noise operator, a linear operator on the vector space of functions $\{-1,1\}^n \to \mathbb{R}$ is defined by

$$(T_\rho f)(x) = \mathop{\mathbb{E}}_{y \sim N_\rho(x)}[f(y)].$$

Question 1.7.2. Show that

$$\langle T_\rho f, g \rangle = \mathop{\mathbb{E}}_{x,y \sim N_\rho}[f(x) g(y)].$$

Question 1.7.3. In this question we will find a spectral expression for $T_\rho f$.

(a) Calculate $T_\rho 1$.

(b) Calculate $T_\rho x_1$.

(c) Calculate $T_\rho \chi_S$.

Hint: Use the independence of the noise across the coordinates.

(d) Calculate $T_\rho f$ in terms of the Fourier expansion of f.

Hint: Use the linearity of T_ρ.

Question 1.7.4. Show that for every $\rho \in (0,1)$, the noise operator T_ρ is *contractive*, that is, $\|T_\rho f\| \le \|f\|$:

(a) Using the spatial definition given above. *Hint: Use* $\mathbb{E}[g]^2 \le \mathbb{E}[g^2]$.

(b) Using the spectral formula.

Applying noise significantly reduces the high degree part of a function.

Question 1.7.5. Show that for every $\epsilon > 0$ and $\rho \in (0,1)$ there exists a constant d such that

$$\|(T_\rho f)^{>d}\| \le \epsilon \|f\|.$$

There is a fourth way to conceive of the distribution $N_\rho(x)$, as a continuous time Markov chain. The following question is optional.

Question 1.7.6. Consider a continuous time Markov chain which starts at state $x \in \{-1,1\}^n$, and for each bit independently, at each infinitesimal interval of length ϵ, has an ϵ chance of flipping x_i.

(a) Fix i, let $t_0 = 0$, and let $t_1, t_2, \ldots$ be the random times at which bit x_i is flipped. Show that $t_1 - t_0, t_2 - t_1, \ldots$ are i.i.d. exponential random variables, and calculate their expectation.

Hint: To calculate $\Pr[t_1 > t]$, *divide* $[0,t]$ *into* t/ϵ *intervals of length* ϵ, *and take the limit* $\epsilon \to 0$.

(b) Show that the number of times that x_i is flipped up to time t has Poisson distribution, and calculate its expectation.

Hint: Calculate the probability that a bit is flipped k *times using a* k*-dimensional integral, and use the formula* $\int_{0<w_1<\cdots<w_k<1} dw_1 \cdots dw_k = 1/k!$.

(c) Determine the distribution of x_i at time t.

Hint: Calculate the Taylor expansion of $\cosh t = \frac{e^t+e^{-t}}{2}$.

(d) Express the distribution of the state at time t in the form $N_\rho(x)$.

2 Hypercontractivity

One of the major tools in Boolean function analysis is *hypercontractivity*, a mysterious property which states, in a formal way, that applying noise "smoothens" the function. This is expressed by relating different L_p norms:

$$\|f\|_p = \sqrt[p]{\mathbb{E}[|f|^p]}.$$

When $p \ge 1$, this is indeed a norm (i.e., it satisfies the triangle inequality), and

$$\|f\|_\infty = \lim_{p\to\infty} \|f\|_p = \max |f|.$$

Furthermore, $p \le q$ implies $\|f\|_p \le \|f\|_q$.

Question 2.0.1. Check that $\|f\|_2 = \|f\|$.

2.1 4-to-2

In this section we will prove Bonami's lemma

$$\|T_{\sqrt{1/3}} f\|_4 \le \|f\|_2$$

and draw some conclusions.

Question 2.1.1. For brevity, we will use T for $T_{\sqrt{1/3}}$ in this question. We will prove Bonami's lemma $\|Tf\|_4 \le \|f\|_2$ by induction.

(a) Show that the lemma is equivalent to $\mathbb{E}[(Tf)^4] \le \mathbb{E}[f^2]^2$.

(b) Prove the base case $n = 0$.

Given the lemma for $n-1$, we will prove it for n.

Write $f(x_1, \dots, x_n) = x_n g(x_1, \dots, x_{n-1}) + h(x_1, \dots, x_{n-1})$.

(c) Show that $\|f\|_2^2 = \|g\|_2^2 + \|h\|_2^2$.

Hint: Either use the Fourier expansion, or expand $\mathbb{E}[f^2]$.

(d) Show that $Tf = \frac{x_n}{\sqrt{3}} Tg + Th$.

(e) Deduce that

$$\mathbb{E}[(Tf)^4] = \frac{1}{9}\,\mathbb{E}[(Tg)^4] + 2\,\mathbb{E}[(Tg)^2(Th)^2] + \mathbb{E}[(Th)^4].$$

(f) Apply the Cauchy–Schwarz inequality to deduce

$$\mathbb{E}[(Tf)^4] \le \frac{1}{9}\,\mathbb{E}[(Tg)^4] + 2\sqrt{\mathbb{E}[(Tg)^4]}\sqrt{\mathbb{E}[(Th)^4]} + \mathbb{E}[(Th)^4].$$

(g) Apply the induction hypothesis on g, h and conclude the proof.

Hölder's inequality states that if $1 \le p, q \le \infty$ satisfy $1/p + 1/q = 1$ (we say that p, q are *conjugate*) then $\langle f, g \rangle \le \|f\|_p \|g\|_q$.

Using Hölder's inequality, we can obtain a similar statement with the L_2 norm on the left-hand side.

Question 2.1.2. In this exercise, we will show the "conjugate" Bonami's lemma.

(a) Show that $\langle T_\rho f, g \rangle = \langle f, T_\rho g \rangle$, and conclude that $\|T_\rho f\|_2^2 = \langle f, T_\rho^2 f \rangle$.

(b) Use Hölder's inequality to deduce $\|T_{1/\sqrt{3}} f\|_2^2 \le \|f\|_{4/3} \|T_{1/\sqrt{3}}^2 f\|_4$.

(c) Use Bonami's lemma to deduce

$$\|T_{1/\sqrt{3}} f\|_2 \le \|f\|_{4/3}.$$

Using this, we can show that every low-degree Boolean function is a junta (depends on a constant number of coordinates).

Question 2.1.3. Let $f\colon \{-1,1\}^n \to \{-1,1\}$ have degree d.

(a) Let $f_i \triangleq L_i f$. Show that $|f_i| \in \{0,1\}$, and so $\|f_i\|_p^p$ is the same for all $p \ge 1$.

(b) Show that $\|T_{1/\sqrt{3}} f_i\|_2^2 \ge 3^{-d} \|f_i\|_2^2$.

Hint: Use $\deg f_i \le \deg f$.

(c) Apply the conjugate Bonami's lemma to deduce $3^{-d}\|f_i\|_2^2 \le \|f_i\|_2^3$.

(d) Deduce that either $\mathrm{Inf}_i[f] = 0$ or $\mathrm{Inf}_i[f] \ge 9^{-d}$.

(e) Conclude that f depends on at most $d9^d$ coordinates.

Hint: Use the double-sided Poincaré's inequality.

Hypercontractivity shows that bounded-degree functions are "reasonable", in various ways.

Question 2.1.4 (Concentration)**.** Suppose that $\deg f = d$.

(a) Show that $\|f\|_4 \leq \sqrt{3}^d \|f\|_2$.

Hint: Apply hypercontractivity to $g = T_{\sqrt{3}} f$.

(b) Prove a concentration bound beating Chebyshev's inequality:

$$\Pr[|f - \mathbb{E}[f]| \geq s\sqrt{\mathbb{V}[f]}] \leq \frac{9^d}{s^4}.$$

Hint: Consider $f - \mathbb{E}[f]$.

Later we will see an even better concentration bound.

The Paley–Zygmund inequality states that if $X \geq 0$ then for $t \in [0,1]$,

$$\Pr[X > t\,\mathbb{E}[X]] \geq (1-t)^2 \frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]}.$$

If you haven't seen it before, try to prove this inequality.

Question 2.1.5 (Anti-concentration)**.** Show that if $\deg f = d$ then for $t \in [0,1]$,

$$\Pr[|f - \mathbb{E}[f]| \geq t\sqrt{\mathbb{V}[f]}] \geq (1-t^2)^2 9^{-d}.$$

2.2 Friedgut–Kalai–Naor

In this section, *Boolean* will mean ± 1-valued. We start by extending Question 1.5.3.

Question 2.2.1. Suppose that $f\colon \{-1,1\}^n \to \mathbb{R}$ has degree at most 1.

(a) Show that if at least two Fourier coefficients $\hat{f}(\{i\}), \hat{f}(\{j\})$ are non-zero then f cannot be Boolean.

Hint: Fix all other coordinates.

(b) Determine all Boolean functions of degree at most 1.

Now suppose that f *almost* has degree 1, in the sense that $\|f^{>1}\|$ is small. Must f be close to a Boolean degree 1 function? This is the statement of the FKN theorem.

Question 2.2.2. Suppose that $f\colon \{-1,1\}^n \to \{-1,1\}$ satisfies $\mathbb{E}[f] = 0$ (we will later get rid of this assumption) and $\|f^{>1}\|^2 = \epsilon$.

(a) Use anti-concentration to show that

$$\mathbb{E}[|(f^{=1})^2 - 1|] = \Omega\left(\frac{1}{2}\sqrt{\mathbb{V}[(f^{=1})^2]} - \epsilon\right).$$

Hint: Use $\mathbb{E}[(f^{=1})^2] = 1 - \epsilon$ and $|x-1| \geq |x - (1-\epsilon)| - \epsilon$.

(b) Show that $(f^{=1})^2 - 1 = -2ff^{>1} + (f^{>1})^2$, deduce

$$\mathbb{E}[|(f^{=1})^2 - 1|] = O(\sqrt{\epsilon}),$$

and conclude $\mathbb{V}[(f^{=1})^2] = O(\epsilon)$.

(c) Calculate this variance to be

$$\mathbb{V}[(f^{=1})^2] = 4 \sum_{1 \leq i < j \leq n} \hat{f}(\{i\})^2 \hat{f}(\{j\})^2 = 2(1-\epsilon)^2 - 2\sum_{i=1}^n \hat{f}(\{i\})^4.$$

Hint: Calculate the Fourier expansion of $(f^{=1})^2$ (first part) and complete the square (second part).

(d) Deduce that $\hat{f}(\{i\})^2 = 1 - O(\epsilon)$ for some $i \in [n]$.

Hint: Use $\sum_{i=1}^n \hat{f}(\{i\})^2 = 1 - \epsilon$.

(e) Conclude that $\Pr[f \neq g] = O(\epsilon)$ for some $g = \pm x_i$.

Hint: Relate $\Pr[f \neq g]$ *to* $\mathbb{E}[fg]$ *as in linearity testing.*

We get rid of the assumption $\mathbb{E}[f] = 0$ using a trick.

Question 2.2.3. Let $f\colon \{-1,1\}^n \to \{-1,1\}$ satisfy $\|f^{>1}\|^2 = \epsilon$. Define

$$g(x_0, x_1, \ldots, x_n) = x_0 f(x_0 x_1, \ldots, x_0 x_n).$$

(a) Show that g is Boolean, has zero mean, and satisfies $\|g^{>1}\|^2 = \epsilon$.

Hint: Calculate the Fourier expansion of g.

(b) Prove the *FKN theorem*: $\Pr[f \neq h] = O(\epsilon)$ for some *dictator* $h \in \{\pm 1, \pm x_1, \ldots, \pm x_n\}$ (a dictator is a function depending on at most one coordinate).

Hint: Show that $\Pr[f \neq x_i] = \Pr[g \neq x_i]$ *and* $\Pr[f \neq 1] = \Pr[g \neq x_0]$.

An equivalent formulation of the theorem relaxes the condition of being Boolean rather than the condition of having degree 1.

Question 2.2.4. Suppose that $f\colon \{-1,1\}^n \to \mathbb{R}$ has degree 1 and satisfies

$$\mathbb{E}[\operatorname{dist}(f, \{-1,1\})^2] \triangleq \mathbb{E}[\min(|f-1|, |f+1|)^2] = \epsilon.$$

(a) Let F result from rounding f to ± 1 (i.e., F is the sign of f). Show that $\|F^{>1}\|^2 \leq \epsilon$, and so $\|F - h\|^2 = O(\epsilon)$ for some dictator h.

Hint: Use $\|F - F^{\leq 1}\|^2 \leq \|F - f\|^2$*, which follows from* $F^{\leq 1}$ *being the projection of* F *to the vector space of functions of degree at most* 1.

(b) Use the inequality $(a+b)^2 \leq 2(a^2+b^2)$ (the L_2^2 *triangle inequality*) to conclude that $\|f - h\|^2 = O(\epsilon)$.

2.3 Kahn–Kalai–Linial theorem

The Kahn–Kalai–Linial theorem (1988) started the field of Boolean function analysis. It answers the following question. Suppose that $f\colon \{-1,1\}^n \to \{-1,1\}$ is a balanced Boolean function. How low can the maximal influence be? The Poincaré inequality shows that the maximal influence is at least $1/n$, but actually more is true.

Question 2.3.1. Let $f\colon \{01,1\}^n \to \{-1,1\}$. Define $\operatorname{MaxInf}[f] = \max_i \operatorname{Inf}_i[f]$.

(a) Show that $\operatorname{Inf}_i[f]^{3/2} \geq \|T_{1/\sqrt{3}} L_i f\|^2$ (using the conjugate Bonami lemma), and deduce

$$\sum_{S \neq \emptyset} 3^{-|S|} \hat{f}(S)^2 \leq \sum_S \frac{|S|}{3^{|S|}} \hat{f}(S)^2 \leq \sqrt{\operatorname{MaxInf}[f]} \operatorname{Inf}[f].$$

(b) Define a probability distribution $\mathcal{S}$ on non-empty subsets of $[n]$ by $\Pr[\mathcal{S} = S] \propto \hat{f}(S)^2$. Show that $\mathbb{E}[|S|] = \operatorname{Inf}[f] / \mathbb{V}[f]$.

(c) Use Jensen's inequality for the function 3^{-x} to deduce

$$\frac{3^{-\operatorname{Inf}[f]/\mathbb{V}[f]}}{\operatorname{Inf}[f]/\mathbb{V}[f]} \leq \sqrt{\operatorname{MaxInf}[f]}.$$

(d) Deduce the KKL theorem:

$$\operatorname{MaxInf}[f] = \Omega\left(\frac{\log n}{n}\,\mathbb{V}[f]\right).$$

Hint: Show that if $\operatorname{MaxInf}[f]/\mathbb{V}[f] \le C\frac{\log_3 n}{n}$ *for small enough* $C > 0$, *then the inequality from the preceding item is violated.*

The original motivation for proving the KKL theorem was an application to bribing.

Question 2.3.2. Let $f\colon \{-1,1\}^n \to \{-1,1\}$. We think of f as a *voting scheme* for an election with n voters and two candidates. We assume that f is *monotone*: changing an input from -1 to 1 cannot change the result from 1 to -1.

(a) Show that if $-1+\epsilon \le \mathbb{E}[f] \le 1-\epsilon$ then there is a variable x_i such that $\mathbb{E}[f|_{x_i=1}] \ge \mathbb{E}[f] + \Omega(\frac{\log n}{n})$.

(b) Suppose that f is *candidate-oblivious*: $f(-x_1,\ldots,-x_n) = -f(x_1,\ldots,x_n)$. Show that there exists a set S of $O(n/\log n)$ voters such that $\mathbb{E}[f|_{x_S=1}] \ge 1-\epsilon$.

(c) How big should S be if f is the majority function?

Ajtai and Linial, improving on the majority function, constructed a function in which S has to contain at least $\Omega(n/\log^2 n)$ voters.

The following example shows that when $\mathbb{V}[f] = \Theta(1)$, the KKL inequality is tight.

Question 2.3.3. For parameters n, m such that $m \mid n$, the tribes function is a function on n variables x_{ij}, where $i \in [n/m]$ and $j \in [m]$, given by

$$\mathrm{Tribes}_{n,m} = \bigvee_i \bigwedge_j x_{ij}.$$

(Here $\vee$ is max and $\wedge$ is min).

(a) Find a value of m for which $\mathbb{E}[\mathrm{Tribes}_{n,m}] \approx 0$.

(b) Estimate $\mathrm{Inf}_{ij}[\mathrm{Tribes}_{n,m}]$ for that value of m, and compare to the KKL theorem.

Hint: Determine which inputs are such that flipping x_{ij} *changes the result.*

2.4 Friedgut's junta theorem

If $f\colon \{-1,1\}^n \to \{-1,1\}$ is a k-junta, that is, if f depends on k coordinates, then $\deg f \le k$ and so $\mathrm{Inf}[f] \le k$. Friedgut's junta theorem gives a kind of converse to this statement.

Question 2.4.1. Let $f\colon \{-1,1\}^n \to \{-1,1\}$ have total influence k. We will try to approximate f by a junta. The variables of the junta will be

$$J = \{i : \mathrm{Inf}_i[f] \ge \tau\},$$

for a suitable τ.

(a) Show that $|J| \le k/\tau$.

(b) Define g to be the J-junta obtained from f by averaging over all other variables. Show that

$$\|f-g\|^2 = \sum_{S \not\subseteq J} \hat{f}(S)^2.$$

Hint: Show that g *is obtained from* f *by retaining only the Fourier coefficients corresponding to subsets of* J.

(c) Given $\epsilon > 0$, show that

$$\sum_{|S| \geq k/\epsilon} \hat{f}(S)^2 \leq \epsilon.$$

Hint: Define a probability distribution $\mathcal{S}$ on subsets of $[n]$ by $\Pr[\mathcal{S} = S] \propto \hat{f}(S)^2$, and calculate $\mathbb{E}[|\mathcal{S}|]$.

(d) Using $\mathrm{Inf}_i[f]^{3/2} \geq \|T_{1/\sqrt{3}} L_i f\|^2$, show that when $i \notin J$,

$$\sum_{\substack{|S| \leq k/\epsilon \\ i \in S}} \hat{f}(S)^2 \leq 3^{k/\epsilon} \mathrm{Inf}_i[f]^{3/2} \leq 3^{k/\epsilon} \sqrt{\tau} \, \mathrm{Inf}_i[f].$$

(e) Conclude from the last three items that

$$\|f - g\|^2 \leq \epsilon + k 3^{k/\epsilon} \sqrt{\tau},$$

and calculate a value of τ for which the right-hand side is $O(\epsilon)$.

(f) Let G be the Boolean function resulting from rounding g. Show that G is a $2^{O(k/\epsilon)}$-junta satisfying $\Pr[f \neq G] = O(\epsilon)$.

Hint: Show that pointwise, $|f(x) - G(x)| = O(|f(x) - g(x)|)$, and use the L_2^2 triangle inequality.

The exponential dependence on k is necessary, as the example of Tribes shows: Question 2.3.3 shows that the (balanced) tribes function has total influence $O(\log n)$, yet depends in an essential way on all coordinates.

Bourgain's tail bound is the stronger result that a function either has mass $\Omega(\mathbb{V}[f]/\sqrt{k})$ beyond level k, or can be approximated by a junta

2.5 General hypercontractivity

So far we have considered hypercontractivity only for the norms $(4, 2)$ or $(2, 4/3)$. A similar result holds for a general pair (p, q) satisfying $1 \leq p \leq q < \infty$.

The proof will require a corollary of Hölder's inequality. It is well-known (and follows essentially from the Cauchy–Schwarz inequality) that

$$\|f\|_2 = \sup_{g \neq 0} \frac{\langle f, g \rangle}{\|g\|_2}.$$

In the same way, we can deduce from Hölder's inequality that if $1/q + 1/r = 1$ then

$$\|f\|_q = \sup_{g \neq 0} \frac{\langle f, g \rangle}{\|g\|_r}.$$

You might want to verify that this indeed follows from Hölder's inequality.

We will show that

$$\|T_\rho f\|_q \leq \|f\|_p, \text{ where } \rho = \sqrt{\frac{p-1}{q-1}}.$$

The proof is by induction on n. The case $n = 1$, known as the *two-point inequality* (why?), is not particularly enlightening, so we will just show the *tensorization* step.

Question 2.5.1. Let $f, g \colon \{-1, 1\}^{n+1} \to \mathbb{R}$, and suppose that we have proved hypercontractivity for n.

(a) Use Hölder's inequality to deduce that for functions $f', g' \colon \{01, 1\}^n \to \mathbb{R}$,

$$\langle T_\rho f', g' \rangle \leq \|f'\|_p \|g'\|_r,$$

where r is the conjugate of q (i.e., $1/q + 1/r = 1$).

(b) For $\sigma \in \{-1,1\}$, define $f_\sigma(x_1,\ldots,x_n) = f(x_1,\ldots,x_n,\sigma)$, and define g_σ similarly. Show that

$$\langle T_\rho f, g\rangle \le \mathop{\mathbb{E}}_{x_{n+1},y_{n+1}\sim N_\rho}[\|f_{x_{n+1}}\|_p \|g_{x_{n+1}}\|_r].$$

(c) Let $F(x) = \|f_x\|_p$ and $G(x) = \|g_x\|_r$. Show that

$$\langle T_\rho f, g\rangle \le \langle T_\rho F, G\rangle \le \|F\|_p \|G\|_r = \|f\|_p \|g\|_r.$$

(d) Use the consequence of Hölder's inequality to deduce $\|T_\rho f\|_q \le \|f\|_p$.

The two most important cases are when $p = 2$ (generalizing $(4,2)$) and when $q = 2$ (generalizing $(2,4/3)$).

Question 2.5.2. Let $p \le 2 \le q$, and suppose that $f\colon \{-1,1\}^n \to \mathbb{R}$ has degree d.

(a) Show that $\|f\|_q \le \sqrt{q-1}^d \|f\|_2$.
Hint: Apply hypercontractivity to $T_{\sqrt{q-1}} f$.

(b) Show that $\|f\|_2 \le \sqrt{p-1}^{-d} \|f\|_p$.

The sharp form of hypercontractivity has many applications.

Question 2.5.3 (Concentration)**.** Suppose that $f\colon \{-1,1\}^n \to \mathbb{R}$ has degree d. Show that there exist constants $C_1, C_2 > 1$ such that for $t \ge C_1^d$,

$$\Pr[|f - \mathbb{E}[f]| \ge t\sqrt{\mathbb{V}[f]}] \le C_2^{-t/d}.$$

Hint: Assume $\mathbb{E}[f] = 0$, and use $\|f\|_q \le \sqrt{q-1}^d \|f\|_2$ together with Markov's inequality for an appropriately chosen q depending on t, d.

Question 2.5.4 (Small set expansion)**.** Show that for every subset $A \subseteq \{-1,1\}^n$,

$$\Pr_{\substack{x\in A\\ y\sim N_\rho(x)}}[y \in A] \le (|A|/2^n)^{\frac{1-\rho}{1+\rho}}.$$

Hint: Consider $\langle T_\rho 1_A, 1_A\rangle = \|T_{\sqrt{\rho}} 1_A\|^2$.

Small set expansion states that if A is small and we run a short random walk starting at A (cf. the final definition we gave of the noise operator), then it is very likely that we escape A.

The following two questions are optional.

Question 2.5.5 (Level-k inequality)**.** Suppose that $f\colon \{-1,1\}^n \to \{0,1\}$ has mean $\mathbb{E}[f] = \alpha$. Show that for $k \le 2\ln(1/\alpha)$ we have

$$\|f^{\le k}\|^2 \le \left(\frac{2e}{k}\ln(1/\alpha)\right)^k \alpha^2.$$

Hint: Use $\|f^{\le k}\|^2 \le \rho^{-k}\langle T_\rho f, f\rangle$ for an appropriate ρ.

This inequality has recently been improved by Chin Ho Lee in *Fourier bounds and pseudorandom generators for product tests.*

Question 2.5.6. A generalization of Hölder's inequality states that

$$\langle f, g\rangle \le \|f\|_p^\alpha \|g\|_q^\beta$$

whenever $\alpha/p + \beta/q = 1$.

(a) Use the generalized Hölder inequality to show that

$$\|f\|_2^2 \le \|f\|_{2+\epsilon}^{(2+\epsilon)/(1+\epsilon)} \|f\|_1^{\epsilon/(1+\epsilon)}.$$

(b) Apply hypercontractivity to show that if $\deg f \le d$ then

$$\|f\|_2 \le \sqrt{1+\epsilon}^{d(2+\epsilon)/\epsilon} \|f\|_1,$$

and deduce $\|f\|_2 \le e^d \|f\|_1$ by taking the limit $\epsilon \to 0$.

(c) Show that if $\mathbb{E}[f] = 0$ then $\mathbb{E}[f 1_{f>0}] = \|f\|_1/2$, and deduce using Cauchy–Schwarz that

$$\frac{1}{2}\|f\|_1 \le e^d \|f\|_1 \sqrt{\Pr[f > 0]}.$$

(d) Conclude that if $\mathbb{E}[f] = 0$ and $f \neq 0$ then

$$\Pr[f > 0] \ge \frac{e^{-2d}}{4}.$$

3 Biased hypercube

3.1 Skewed hypercube

So far we have considered functions on the Boolean cube $\{-1, 1\}^n$ with respect to the uniform measure. In many situations, it is natural to consider a biased measure. For example, a $G(n, p)$ random graph is formed by putting in each edge with probability p. When $p = 1/2$, we get the uniform measure on all graphs, but we are often interested in much smaller p.

It will be more natural to consider $\{0, 1\}^n$ instead of $\{-1, 1\}^n$. The measures we consider will be

$$\mu_p(x_1, \ldots, x_n) = p^{x_1+\cdots+x_n}(1-p)^{(1-x_1)+\cdots+(1-x_n)}.$$

Question 3.1.1. Let S be a random subset of $[n]$ obtained by putting in each element with probability p. Show that $1_S \sim \mu_p$, where 1_S is the characteristic function of S, which we can also think of as a vector of length n.

When $n = \binom{m}{2}$, the distribution μ_p is the law of $G(m, p)$.

We define the inner product and the various norms just as in the case of the Boolean cube, taking expectation with respect to μ_p rather than with respect to the uniform measure.

When $p \neq \frac{1}{2}$, the Fourier characters are no longer orthonormal, and so we have to use different functions.

Question 3.1.2. Find all functions $\omega\colon \{0, 1\} \to \mathbb{R}$ that, together with 1, form an orthonormal basis for the functions on $\{0, 1\}$ with respect to μ_p.

Hint: Write $\omega(x) = \frac{x-\alpha}{\beta}$, *and find* α *and* β.

The function ω is unique up to sign, and there is no common convention regarding the sign. We will choose the expression in which $\beta > 0$.

Question 3.1.3. For a subset $S \subseteq [n]$, define

$$\omega_S(x) = \prod_{i \in S} \omega(x_i).$$

Show that the ω_S form an orthonormal basis for all functions on $\{0, 1\}^n$.

Since the ω_S form a basis, every function has a unique Fourier expansion of the form

$$f = \sum_{S \subseteq [n]} \hat{f}(S)\omega_S.$$

One big difference between the uniform case and the skewed case is that we no longer have $\omega_S \omega_T = \omega_{S \triangle T}$ (check!).

There are two natural ways to define influence. One is spectral:

$$\mathrm{Inf}_i[f] = \sum_{S \ni i} \hat{f}(S)^2.$$

The other is spatial:

Question 3.1.4. (a) Show that

$$\mathrm{Inf}_i[f] = p(1-p)\,\mathbb{E}[(f(x) - f(x \oplus e_i))^2],$$

where $x \oplus e_i$ is obtained from x by flipping the ith input.

(b) Show that

$$\mathrm{Inf}_i[f] = \frac{1}{2}\,\mathbb{E}[(f(x) - f(y))^2],$$

where y is obtained from x by resampling the ith coordinate.

(c) Show that when f is monotone and Boolean (that is $\{0,1\}$-valued) then $\mathrm{Inf}_i[f] = \sqrt{p(1-p)}\hat{f}(\{i\})$.

Hint: Look up the corresponding question in the uniform case.

Some people prefer the normalization $\mathrm{Inf}_i[f] = \mathbb{E}[(f(x) - f(x \oplus e_i))^2]$, in which case the spectral formula will be $\mathrm{Inf}_i[f] = \frac{1}{p(1-p)} \sum_{S \in i} \hat{f}(S)^2$.

Similarly, there are two ways to define the noise operator. Spectrally, we use the exact same formula:

$$T_\rho f = \sum_{d=0}^{n} \rho^d f^{=d},$$

where $f^{=d}$ is defined analogously to the case $p = \frac{1}{2}$. The corresponding spatial definition is:

Question 3.1.5. (a) Given $x \sim \mu_p$, let y be the random vector in which each coordinate is resampled with probability $1 - \rho$. Show that

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

(b) Calculate the joint distribution of (x_i, y_i) in this experiment.

When p is constant, or at least bounded away from 0 and 1, hypercontractivity holds, with a very similar proof.

3.2 Erdős–Ko–Rado

The Erdős–Ko–Rado theorem is a basic result in extremal combinatorics. Although usually stated for subsets of $\binom{[n]}{k}$, it has an analog in the μ_p setting.

A family $F \subseteq 2^{[n]}$ is *intersecting* if any two sets in F have at least one element in common. The *μ_p-measure* of F, denoted $\mu_p(F)$, is the probability that a random $x \sim \mu_p$ belongs to F.

Question 3.2.1. (a) Give an example of an intersecting family F with $\mu_p(F) = p$.

(b) Show that when $p = 1/2$, every intersecting family has measure at most $1/2$.

Hint: A set cannot intersect its complement.

(c) Show that when $p > 1/2$, as $n \to \infty$ there are intersecting families with measure arbitrarily close to 1.

Hint: Take all sets of size larger than $n/2$.

The interesting setting is thus $p < 1/2$. The Erdős–Ko–Rado theorem states that $\mu_p(F) \le p$ for all intersecting families F. We will prove this using a spectral method.

Question 3.2.2. In this question, $p < 1/2$.

(a) Construct a $2^{[1]} \times 2^{[1]}$ matrix B such that

- (i) $B(\{1\},\{1\}) = 0$.
- (ii) $\omega_\emptyset$ is an eigenvector with eigenvalue 1.
- (iii) $\omega_{\{1\}}$ is an eigenvector with eigenvalue λ (you have to determine λ).

(b) Construct a $2^{[n]} \times 2^{[n]}$ matrix A such that

- (i) $B(S,T) = 0$ whenever S,T are intersecting.
- (ii) $\omega_\emptyset$ is an eigenvector with eigenvalue 1.
- (iii) All other ω_S are also eigenvectors.

Hint: Let $A = B^{\otimes n}$, the tensor *or* Kronecker *power of B.*

(c) Suppose F is intersecting. Show that $\langle 1_F, A1_F \rangle = 0$. *(This is the inner product with respect to μ_p.)*

(d) Expand $\langle 1_F, A1_F \rangle$ in terms of the Fourier coefficients of 1_F.

(e) Use $\widehat{1_F}(\emptyset)^2 = \mu_p(F)^2$ and $\sum_{S \neq \emptyset} \widehat{1_F}(S)^2 = \mu_p(F)(1-\mu_p(F))$ to conclude $\mu_p(F) \le p$.

(f) What can you conclude when $\mu_p(F) = p$?

Katona came up with a simple argument which is also worthwhile to know. The following questions are optional.

Question 3.2.3. Let $p < 1/2$, and let $F \subseteq 2^{[n]}$ be an intersecting family.

Consider the experiment in which n points are thrown on a unit circumference circle. Consider a "window" of length p on the circumference, and let S be the set of points falling inside it.

(a) Show that $\mu_p(F) = \Pr[S \in F]$.

(b) Show that the same formula holds if we choose the location of the window randomly.

(c) Show that for every fixed location of the points, the measure of windows such that $S \in F$ is at most p. (This is a bit subtle.)

(d) Conclude that $\mu_p(F) \le p$.

Question 3.2.4. Suppose that $f(x) = \sum_{i=1}^n w_i x_i$ and $\mathbb{E}[f] \ge 0$. Define $q = \Pr[f \ge 0]$.

(a) When p is the reciprocal of an integer, show that $q \ge p$.

(b) Can you extend the argument to other values of p?

(c) Can you find values of p for which $q < p$ for some f?

(d) What is the minimal value of q for these values of p?

3.3 Friedgut–Kalai sharp threshold theorem

Suppose that $f\colon \{0,1\}^n \to \{0,1\}$ is monotone and non-constant. Our goal is to understand how fast f increases from 0 to 1.

Question 3.3.1. Let $f\colon \{0,1\}^n \to \{0,1\}$ be a non-constant monotone Boolean function, and let $F(p)$ be the expectation of f under μ_p.

(a) Show that $F(p)$ is increasing.

Hint: Given $p \leq q$, construct a coupling *(x,y) such that $x \sim \mu_p$, $y \sim \mu_q$, and $x \leq y$. This shows that $F(p) \leq F(q)$. Since F is a polynomial, it is strictly increasing.*

(b) Show that there exists a unique critical probability p_c such that $F(p_c) = 1/2$.

(c) Show that $F(\min(Cp,1)) \geq 1 - (1-F(p))^C$ for integer $C \geq 1$.

Hint: If we OR C many μ_p vectors then we get a μ_q vector, for $q \leq Cp$.

(d) Deduce that for every $\epsilon > 0$ there exists $C > 0$ such that $F(\min(Cp,1)) \geq 1-\epsilon$ and $F(p/C) \leq \epsilon$.

We now give a formula for the derivative of $\mu_p(f) := \mathbb{E}[f(\mu_p)]$ in terms of the total influence of f, for monotone f.

Question 3.3.2. Let $f\colon \{0,1\}^n \to \mathbb{R}$. Extend f to a function on $[0,1]^n \to \mathbb{R}$ using its multilinear expansion.

(a) Use multilinearity to show that $f(p,\ldots,p) = \mathbb{E}_{x\sim\mu_p}[f(x)]$.

Hint: Show this first for x_i, then for general monomials.

(b) Use multilinearity to show that

$$\frac{\partial f}{\partial x_i}(x) = f(x|_{x_i=1}) - f(x|_{x_i=0}).$$

(c) Conclude the formula

$$\mathop{\mathbb{E}}_{\mu_p}\left[\frac{\partial f}{\partial x_i}\right] = \frac{1}{\sqrt{p(1-p)}}\hat{f}(\{i\}).$$

Hint: The foregoing shows that $\mathbb{E}[\partial f/\partial x_i] = \mathbb{E}[f(x|_{x_i=1}) - f(x|_{x_i=0})]$.

(d) Use the chain rule to deduce

$$\frac{\mathrm{d}}{\mathrm{d}p}\mathop{\mathbb{E}}_{\mu_p}[f] = \frac{1}{\sqrt{p(1-p)}}\sum_{i=1}^n \hat{f}(\{i\}).$$

(e) When f is monotone and Boolean, use Question 3.1.4 to show that

$$\frac{\mathrm{d}}{\mathrm{d}p}\mathop{\mathbb{E}}_{\mu_p}[f] = \frac{\mathrm{Inf}[f]}{p(1-p)} = \sum_{i=1}^n \mathop{\mathbb{E}}_{\mu_p}[(f(x) - f(x\oplus e_i))^2].$$

Above we have proved the KKL theorem only for $p = 1/2$, but the proof goes through for arbitrary p. When taking into account the hypercontractive parameters for μ_p, we obtain the following result: If $f\colon \{0,1\}^n \to \{0,1\}$ is a Boolean function then with respect to μ_p,

$$\max_i \mathrm{Inf}_i[f] = \Omega\left(\frac{1}{\log(1/\min(p,1-p))}\cdot\frac{\log n}{n}\mathbb{V}[f]\right).$$

Question 3.3.3. Let $f\colon \{0,1\}^n \to \{0,1\}$ be a monotone Boolean function, and suppose that f is *transitive-symmetric*: f is invariant under some transitive permutation group.[1]

(a) Show that all influences of f are identical.

(b) Show that if $p \le 1/2$ then
$$\frac{\mathrm{d}}{\mathrm{d}p} \mathop{\mathbb{E}}_{\mu_p}[f] = \Omega\left(\frac{\log n}{p\log(1/p)}\mathbb{V}[f]\right).$$

Question 3.3.4 (Friedgut–Kalai)**.** Let $f\colon \{0,1\}^n \to \{0,1\}$ be a non-constant monotone Boolean function which is transitive-symmetric, and define $F(p) = \mathbb{E}_{\mu_p}[f]$. Let $p_c \le 1/2$ be the critical probability of f.

(a) Let $\epsilon \in (0,1/3)$, and define a parameter η by
$$\eta = B\log(1/\epsilon)\cdot\frac{\log(1/p_c)}{\log n},$$
where B is a large enough constant. Define
$$p_- = (1-\eta)p_c, \quad p_+ = (1+\eta)p_c.$$
Assuming $\eta < 1/2$, show that the bound in the preceding question holds for all $p \in [p_-,p_+]$ even when p is replaced by p_c.

(b) Show that for an appropriate choice of B, for all $p \in [p_-,p_+]$ we have
$$F'(p) \ge \frac{2\ln(1/2\epsilon)}{\eta p_c}F(p)(1-F(p)).$$

(c) For $p \in [p_-,p_c]$, obtain a lower bound on $F'(p)/F(p)$ and so on $(\ln F(p))'$.

(d) Deduce that $F(p_-) \le \epsilon$. Similarly deduce that $F(p_+) \ge 1-\epsilon$.
Hint: For the latter, consider $(\ln(1-F(p))'$.

This shows that all transitive-symmetric functions have *threshold window* of scale $\frac{\log(1/p_c)}{\log n}$. The theory of sharp thresholds, developed by Friedgut, Bourgain, Kalai, and Hatami, studies better bounds of this form when the functions have more symmetries or when they are known not to correlate with "local structures".

4 Invariance principle

4.1 Gaussian space

So far we have considered functions on the Boolean cube $\{-1,1\}^n$. The invariance principle relates function on the Boolean cube with functions on *Gaussian space*, which is $\mathbb{R}^n$ with the Gaussian measure
$$\gamma(x_1,\ldots,x_n) = \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-(x_1^2+\cdots+x_n^2)/2}.$$
It is a standard fact that this is indeed a probability measure. We define the inner product and norm according to this measure:
$$\langle f,g\rangle = \mathop{\mathbb{E}}_{\gamma}[fg], \quad \|f\|_p = \left(\mathop{\mathbb{E}}_{\gamma}[|f|^p]\right)^{1/p}.$$

Question 4.1.1. Show that the Fourier characters x_S are orthogonal and have unit norm.

The rest of this section is somewhat beyond the scope of this class, so feel free to skip some questions.

[1] A permutation group is a subset of S_n. It operates on $\{0,1\}^n$ by permuting the coordinates. A permutation group is *transitive* if for each i,j, there is a permutation sending i to j.

4.1.1 Hermite basis

A technicality which doesn't occur for finite domains is that we cannot consider arbitrary functions. We will only be interested in *measurable* functions with *finite L^2 norm*, a space we denote by $L^2(\mathbb{R}^n, \gamma)$. Henceforth we will not always state that our functions are measurable and have finite norms.

The Fourier characters form an orthonormal basis for the space of all functions on the Boolean cube. The same doesn't hold for Gaussian space, even when $n = 1$. Indeed, the space of functions $\mathbb{R} \to \mathbb{R}$ is infinite-dimensional. We therefore need to expand the Fourier basis to a larger basis known as the *Hermite basis*.

Question 4.1.2. (a) Suppose that $h_2(x) = ax^2 + bx + c$ has unit norm and is orthogonal to the two one-dimensional Fourier characters $1, x$. Determine h_2 (up to sign).

(b) Suppose that $h_3(x) = ax^3 + bx^2 + cx + d$ has unit norm and is orthogonal to $1, x, h_2$. Determine h_3 (up to sign).

The general formula is

$$h_i(x) = \frac{(-1)^i}{\sqrt{i!} \cdot e^{-x^2/2}} \frac{\mathrm{d}}{\mathrm{d}x^i} e^{-x^2/2}.$$

Question 4.1.3. Check that this formula agrees with $h_0 = 1$, $h_1 = x$, and the value of h_2 which you calculated above.

Question 4.1.4. (a) Show that

$$e^{tx-t^2/2} = \sum_{i=0}^{\infty} h_i(x) \frac{t^i}{\sqrt{i!}}.$$

Hint: Write $e^{tx-t^2/2} = e^{x^2/2} e^{-(t-x)^2/2}$, and use the Taylor expansion of the latter around $t = 0$.

(b) Show that $\mathbb{E}_\gamma[e^{(t+s)x}] = e^{(t+s)^2/2}$, and conclude that $\mathbb{E}[e^{tx-t^2/2} e^{sx-s^2/2}] = e^{ts}$.

Hint: Complete the square and use $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2} = 1$.

(c) Conclude that the h_i have unit norm and are orthogonal.

Hint: Compare the Taylor expansion of e^{ts} to the expansion of $e^{tx-t^2/2} e^{sx-s^2/2}$ from part (a).

The Hermite polynomials are a basis for $L^2(\mathbb{R}, \gamma)$ in the following sense: for every function f there exist coefficients c_i such that

$$f = \sum_{i=0}^{\infty} c_i h_i \text{ a.e.}$$

This is known as the *Hermite expansion* of f.

Since the proof involves basic measure theory, we only sketch it here:

(a) From general principles, it suffices to show that if $\langle f, h_i \rangle = 0$ for all i then $f = 0$ a.e.

(b) If $\langle f, h_i \rangle = 0$ for all i then the Fourier transform of f, which is $\hat{f}(t) = \langle f, e^{-itx} \rangle$, vanishes.

(c) The Fourier transform is invertible, hence $f = 0$.

We extend the Hermite basis to $\mathbb{R}^n$ in the natural way:

$$h_{i_1,\ldots,i_n}(x_1, \ldots, x_n) = h_{i_1}(x_1) \cdots h_{i_n}(x_n).$$

Question 4.1.5. Show that n-dimensional Hermite basis is an orthonormal collection of functions.

4.1.2 Noise

We can extend all the setup of Boolean function analysis from the Boolean cube to Gaussian space. For our purposes, it will suffice to consider only the noise operator and hypercontractivity.

The noise operator can be defined in at least three different ways. We will only present two of them. The third is via Brownian motion.

Question 4.1.6. For a function $f \in L^2(\mathbb{R}, \gamma)$ and parameter ρ, define a function $U_\rho f$ by

$$U_\rho f(x) = \mathop{\mathbb{E}}_{z \sim \gamma}[f(\rho x + \sqrt{1-\rho^2} z)].$$

Show that

$$\langle f, U_\rho g \rangle = \mathop{\mathbb{E}}_{x,y}[f(x)g(y)],$$

where x, y are sampled from a Gaussian distribution with zero mean and covariance matrix $\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. (We say that x, y are *ρ-correlated Gaussians.*)

Hint: Show that if $x, z \sim \gamma$ and $y = \rho x + \sqrt{1-\rho^2} z$, then x, y are ρ-correlated Gaussians.

Question 4.1.7. Calculate $U_\rho h_i$ for $i = 0, 1, 2$.

The pattern persists.

Question 4.1.8. (a) Fix two unit vectors $a, b \in \mathbb{R}^2$ such that $\langle a, b \rangle = \rho$, and let $g \sim \gamma$ be a two-dimensional Gaussian. Show that $x = \langle a, g \rangle$ and $y = \langle b, g \rangle$ are ρ-correlated Gaussians.

(b) Show that $\mathbb{E}[e^{sx - s^2/2} e^{ty - t^2/2}] = e^{st\rho}$.

Hint: Write $e^{sx+ty} = e^{sa_1 g_1 + tb_1 g_1} e^{sa_2 g_2 + tb_2 g_2}$ and use $\mathbb{E}_{w \sim \gamma}[e^{rw}] = e^{r^2/2}$.

(c) Conclude that $\langle U_\rho h_i, h_j \rangle = 0$ if $i \neq j$ and ρ^i otherwise, and so $U_\rho h_i = \rho^i h_i$.

Hint: Use Question 4.1.4(a).

Question 4.1.9. (a) Extend the definition of U_ρ to functions on $\mathbb{R}^n$.

(b) Compute $U_\rho h_{i_1, \ldots, i_n}$.

(c) Explain how to decompose $f = \sum_{i=0}^\infty f^{=i}$ so that $U_\rho f = \sum_{i=0}^\infty \rho^i f^{=i}$.

4.1.3 Rotation sensitivity

The invariance principle, which we discuss below, reduces questions about the Boolean cube to questions on Gaussian space. Here is an example of the kind of reasoning which is possible on Gaussian space.

Question 4.1.10. Let $f \colon \mathbb{R}^n \to \mathbb{R}$. For an angle θ, define

$$\mathrm{RS}_\theta(f) = \frac{1}{2} \mathop{\mathbb{E}}_{x,y}[(f(x) - f(y))^2],$$

where x, y are $\cos\theta$-correlated.

(a) Show that $\mathrm{RS}_\theta(f) = \|f\|^2 - \langle f, U_{\cos\theta} f \rangle$.

(b) When f is Boolean (± 1-valued), show that $\mathrm{RS}_\theta(f) = 2 \Pr[f(x) \neq f(y)]$.

Recall that we can generate two ρ-correlated Gaussians using the method of Question 4.1.8. We can use the same method to couple two correlated pairs.

Question 4.1.11. Let α, β be angles, and consider a Boolean function $f : \mathbb{R}^n \to \{\pm 1\}$.

(a) Show that there exist two-dimensional vectors a, b, c such that $\measuredangle(a,b) = \alpha$, $\measuredangle(b,c) = \beta$ and $\measuredangle(a,c) = \alpha + \beta$.

(b) Use the method of Question 4.1.8 to show that $\mathrm{RS}_{\alpha+\beta}(f) \le \mathrm{RS}_\alpha(f) + \mathrm{RS}_\beta(f)$.

Hint: Let $g \sim \gamma$, and define $x = \langle a, g\rangle$, $y = \langle b, g\rangle$, $z = \langle c, g\rangle$. Use $\Pr[f(x) \neq f(z)] \le \Pr[f(x) \neq f(y)] + \Pr[f(y) \neq f(z)]$.

(c) Show that $\mathrm{RS}_{\pi/2}(f) = \mathbb{V}[f]$.

Hint: Two $\pi/2$-correlated Gaussians are independent.

(d) Conclude that if $\rho = \cos(\pi/2k)$ for integer k then

$$\langle f, U_\rho f\rangle \le 1 - \frac{2\cos^{-1}\rho}{\pi}\,\mathbb{V}[f].$$

(e) Show that the inequality is tight for $f = \mathrm{sgn}(x_1)$.

Hint: Let x, y be sampled as before (with $\measuredangle(a,b) = \theta$). The value of $f(x), f(y)$ depends only on the angle γ of g: for example, $\mathrm{sgn}(x_1) > 0$ *iff* $\gamma \in (\alpha - \pi/2, \alpha + \pi/2)$*. Figure out for which values of γ we have $f(x) \neq f(y)$.*

Note $\langle f, U_\rho f\rangle = \|U_{\sqrt{\rho}} f\|^2$, and so $\sqrt{\langle f, U_\rho f\rangle}$ satisfies the triangle inequality. This implies that the maximum of $\langle f, U_\rho f\rangle$ over functions $f : \mathbb{R}^n \to [-1,1]$ is obtained on "extreme functions", which are Boolean functions. In other words, the results of the preceding question are valid for all functions $f : \mathbb{R}^n \to \{-1,1\}$. We skip the formal argument.

4.1.4 Hypercontractivity

We can deduce hypercontractivity for Gaussian space from hypercontractivity on the Boolean cube using the central limit theorem.

Question 4.1.12. (a) Let $x_1, x_2, \ldots$ be infinitely many independent random variables distributed uniformly over $\{-1,1\}$. Define $\sigma^{(m)} = \frac{x_1 + \cdots + x_m}{\sqrt{m}}$. Show using the central limit theorem that $\sigma \to \gamma$, in an appropriate sense.

(b) Let $(x_1, y_1), (x_2, y_2), \ldots$ be infinitely many independent ρ-correlated ± 1 random variables, that is, $\mathbb{E}[x_i y_i] = \rho$. Define $\sigma^{(m)} = \frac{x_1 + \cdots + x_m}{\sqrt{m}}$ and $\tau^{(m)} = \frac{y_1 + \cdots + y_m}{\sqrt{m}}$. Show that $(\sigma^{(m)}, \tau^{(m)})$ tends to ρ-correlated Gaussians.

Hint: Show that $\mathbb{E}[\sigma^{(m)}\tau^{(m)}] = \rho$.

(c) For $f \in L^2(\mathbb{R}, \gamma)$, define $f^{(m)} : \{-1,1\}^m \to \mathbb{R}$ by $f^{(m)}(x_1, \ldots, x_m) = f(\sigma^{(m)})$. Show that

$$\lim_{m\to\infty} \langle T_\rho f^{(m)}, g^{(m)}\rangle = \langle U_\rho f, g\rangle \text{ and } \lim_{m\to\infty} \|f^{(m)}\|_p = \|f\|_p.$$

(For the latter statement, we need $f \in L^p(\mathbb{R}, \gamma)$, i.e., we need that $\|f\|_p$ exists.)

(d) Conclude hypercontractivity for Gaussian space, with the same parameters as the Boolean cube.

4.2 Invariance principle

The central limit theorem states that for $x_1, \dots, x_n \sim \{-1, 1\}$,

$$\frac{x_1 + \cdots + x_n}{\sqrt{n}} \longrightarrow N(0,1).$$

Here $N(0,1)$ is the standard Gaussian distribution. More generally,

$$a_1x_1 + \cdots + a_nx_n \approx N(0, a_1^2 + \cdots + a_n^2),$$

as long as none of the a_i is too large compared to the rest. The Berry–Esseen theorem states this in a quantitative way. Let us normalize so that $\sum_{i=1}^n a_i^2 = 1$. Then for all t,

$$|\Pr[a_1x_1 + \cdots + a_nx_n < t] - \Pr[N(0,1) < t]| = O\left(\sum_{i=1}^n |a_i|^3\right) = O(\max_i |a_i|).$$

The invariance principle generalizes this statement from *linear forms* to *low-degree polynomials.*

In stating the invariance principle, we need to figure out two ingredients:

1. Given a function on the Boolean cube, what is the corresponding distribution?
2. What is the correct error term generalizing $\max_i |a_i|$?

Whereas for linear forms there is a "universal" limiting distribution, the same cannot be true even for quadratic polynomials. Indeed,

$$\frac{x_1x_2 + \cdots + x_{2n-1}x_2n}{\sqrt{n}} \longrightarrow N(0,1)$$

whereas

$$\left(\frac{x_1 + \cdots + x_n}{\sqrt{n}}\right)^2 \longrightarrow N(0,1)^2.$$

The solution is simple but ingenious.

Question 4.2.1. Let $z_1, \dots, z_n$ be independent standard Gaussians. Show that

$$\sum_{i=1}^n a_iz_i \sim N(0, a_1^2 + \cdots + a_n^2).$$

Therefore the Berry–Esseen can be seen as stating that

$$a_1x_1 + \cdots + a_nx_n \approx a_1z_1 + \cdots + a_nz_n,$$

where the quality of the approximation depends on $\max_i |a_i|$ (assuming $\sum_{i=1}^n a_i^2 = 1$). This shows how to lift an *arbitrary* function from the Boolean cube to Gaussian space. It turns out that the correct error term depends on the maximum *influence* of the function.

Whereas the Berry–Esseen theorem is stated in terms of CDFs (cumulative distribution functions), it will be easier to do the calculations for *test functions* satisfying appropriate constraints.

We start with the argument for linear forms, and then show how to generalize it to arbitrary low-degree functions.

Question 4.2.2. Let $x_1, \dots, x_n$ be i.i.d. random ± 1 variables, let $g_1, \dots, g_n$ be i.i.d. standard Gaussians, and let Ψ satisfy $|\Psi''''(z)| \le B$ for all z.

Define $f(y_1, \dots, y_n) = a_1y_1 + \cdots + a_ny_n$, and for $0 \le i \le n$, let

$$E_i = \mathbb{E}[\Psi(f(x_1, \dots, x_i, g_{i+1}, \dots, g_n))].$$

(a) Let $1 \le i \le n$. Show that for an appropriate random variable C,

$$E_i - E_{i-1} = \mathbb{E}[\Psi(C + a_i x_i) - \Psi(C + a_i g_i)].$$

(b) Use the Taylor expansion of Ψ around C to show that

$$|E_i - E_{i-1}| = O(Ba_i^4).$$

Hint: Use $\mathbb{E}[x_i^k] = \mathbb{E}[z_i^k]$ *for* $k = 0, 1, 2, 3$.

(c) Conclude that

$$|\mathbb{E}[\Psi(f(x_1, \dots, x_n))] - \mathbb{E}[\Psi(f(g_1, \dots, g_n))]| = O\left(B \sum_{i=1}^{n} a_i^4\right).$$

In order to deduce the Berry–Esseen theorem as stated above, we need the following rather technical result. The following question is optional.

Question 4.2.3. (a) Define a function μ by

$$\mu(x) = \begin{cases} Ke^{-1/(1-x^2)}, & \text{if } -1 \le x \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

Show that some choice of K makes μ a probability density function.

(b) Prove by induction that for $k \ge 0$, $\mu^{(k)}(x) = p_k(x)(1 - x^2)^{-2k}\mu(x)$, where p_k is a polynomial.

(c) Show that for every k there exists a constant C_k such that $|\mu^{(k)}(x)| \le C_k$ for all x.

(d) Let $s(x) = [x \le 0]$, that is, $s(x) = 1$ if $x \le 0$, and $s(x) = 0$ otherwise. For $\eta > 0$, define

$$s_\eta(x) = \mathop{\mathbb{E}}_{y \sim \mu}[s(x + \eta y)] = \mathop{\mathbb{E}}_{y \sim \mu}[s(x/\eta + y)].$$

Show that for $k \ge 1$ and $|x| \le \eta$, $|s_\eta^{(k)}(x)| \le C_k/\eta^k$ for all x. (Here $s_\eta^{(k)}$ is the kth derivative of s_η.)

Hint: Write $s_\eta(x) = \int_{-1}^{-x/\eta} \mu(y)\, \mathrm{d}y$.

(e) Show that $s_\eta(x) = s(x)$ for $|x| \ge \eta$, and s_η is decreasing for $|x| \le \eta$.

(f) Show that for every random variable X,

$$\Pr[X < -\eta] \le \mathbb{E}[s_\eta(X)] \le \Pr[X < \eta].$$

Using the function s_η, we can obtain a version of Berry–Esseen for CDFs.

Question 4.2.4. Let $f(y_1, \dots, y_n) = a_1 y_1 + \dots + a_n y_n$, where $\sum_{i=1}^n a_i^2 = 1$, and let $\delta = \sum_{i=1}^n a_i^4$.

(a) Applying Berry–Esseen for a shift of $s_{\eta/2}$, show that for every t,

$$\Pr[f(x_1, \dots, x_n) < t] \le \Pr[f(g_1, \dots, g_n) < t + \eta] + O(\delta/\eta^4).$$

(b) Show that $\Pr[t \le f(g_1, \dots, g_n) \le t + \eta] = O(\eta)$.

Hint: Notice that $f(g_1, \dots, g_n) \sim N(0, 1)$, *and use the fact that the density of* $N(0, 1)$ *is bounded.*

(c) Conclude that

$$\Pr[f(x_1, \dots, x_n) < t] \le \Pr[f(g_1, \dots, g_n) < t] + O(\eta + \delta/\eta^4).$$

(d) Prove the same inequality in the other direction, and choose the best value of η to obtain a version of Berry–Esseen for CDF distance.

The proof of the invariance principle is very similar.

Question 4.2.5. Let $x_1, \ldots, x_n$ be i.i.d. random ± 1 variables, let $g_1, \ldots, g_n$ be i.i.d. standard Gaussians, and let Ψ satisfy $|\Psi''''(z)| \leq B$ for all z.

Let f be a degree d multivariate polynomial in n variables, and for $0 \leq i \leq n$, define

$$E_i = \mathbb{E}[\Psi(f(x_1, \ldots, x_i, g_{i+1}, \ldots, g_n))].$$

(a) Write $f(y_1, \ldots, y_n) = y_i k(y_1, \ldots, y_n) + h(y_1, \ldots, y_n)$. Show that

$$|E_i - E_{i-1}| = O(B\, \mathbb{E}[k^4]).$$

(b) Show that $y_i k = L_i f$, and conclude that $\mathbb{E}[k^2] = \mathrm{Inf}_i[f]$.

(c) Use hypercontractivity to bound $\mathbb{E}[k^4] = O_d(\mathrm{Inf}_i[f]^2)$.

(d) Conclude that

$$|\, \mathbb{E}[\Psi(f(x_1, \ldots, x_n))] - \mathbb{E}[\Psi(f(g_1, \ldots, g_n))]| = O_d(B\|f\|^2 \max_i \mathrm{Inf}_i[f]).$$

Hint: Use the double-sided Poincaré inequality.

Question 4.2.6. Carbery and Wright showed that if f is a degree d multivariate polynomial of unit norm, then the probability that $f(\gamma)$ lies in an interval of length η is $O(d\eta^{1/d})$.

Use this to give a bound on

$$|\Pr[f(x_1, \ldots, x_n) < t] - \Pr[f(g_1, \ldots, g_n) < t]|$$

in terms of the maximum influence of f.

So far our invariance principle applies for functions Ψ with bounded fourth moment. It is often better to consider Lipschitz functions instead. A function Ψ is *C-Lipschitz* if $|\Psi(x) - \Psi(y)| \leq C|x - y|$. The following question is optional.

Question 4.2.7. Let Ψ be C-Lipschitz, and fix $\eta > 0$.

Define $\Psi_\eta(x) = \mathbb{E}[\Psi(x + \eta y)]$, where y is a standard Gaussian.

(a) Let $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$. Prove by induction that $\phi^{(k)}(z) = P_k(z)\phi(z)$, where $P_k(z)$ is a polynomial.

(b) Conclude that for every $k \geq 0$ there exists a constant C_k such that $|\phi^{(k)}(z)| \leq C_k$ for all z.

(c) Show that $|\Psi_\eta(x) - \Psi(x)| = O(\eta C)$ for all x.
Hint: Use the fact that $\mathbb{E}[|y|] = O(1)$.

(d) Show that

$$\Psi_\eta(x) = \frac{1}{\eta} \int_{-\infty}^{\infty} \Psi(y) \phi\left(\frac{x - y}{\eta}\right) \mathrm{d}y.$$

(e) Deduce that for $k \geq 1$,

$$\Psi_\eta^{(k)}(x) = \frac{1}{\eta^{k+1}} \int_{-\infty}^{\infty} \Psi(y) \phi^{(k)}\left(\frac{x - y}{\eta}\right) \mathrm{d}y.$$

(f) Conclude that $|\Psi_\eta^{(k)}(x)| = O_k(C/\eta^{k-1})$ for all x.
Hint: Without loss of generality $\Psi(x) = 0$, and then $|\Psi(y)| \leq C|x - y|$. Now make a change of variables to get something independent of η.

Question 4.2.8. Prove an invariance principle for Lipschitz functions Ψ.
Hint: Apply the invariance principle to Ψ_η, incuring an extra error of $2C\eta$, and optimize η.

4.3 Majority is stablest

Consider a balanced function $f\colon \{-1,1\}^n \to \{-1,1\}$. How stable can f be to noise? That is, how large can $\langle T_\rho f, f\rangle$ be?

Question 4.3.1. Let $f\colon \{-1,1\}^n \to \{-1,1\}$ be balanced ($\mathbb{E}[f] = 0$), and let $\rho \in (0,1)$.

(a) Show that $\langle T_\rho f, f\rangle \le \rho$.

(b) Show that equality holds iff f is a dictator.

A dictator is noise-stable because it's a junta. What if we rule out juntas? One natural way is to put a bound on the individual influences. When the maximum influence is small, the invariance principle implies (as we will work out) that the function behaves as if it lived in Gaussian space.

In Gaussian space there are no preferred directions, so dictators don't make sense (this is because Gaussian space is rotation-invariant). Indeed, Borell's theorem states that the most stable balanced function is a halfspace through the origin. In a slightly stronger form, it states that if $f\colon \mathbb{R}^n \to [-1,1]$ and $h\colon \mathbb{R} \to \{-1,1\}$ is a threshold function with the same mean then for all $\rho \in (0,1)$,

$$\langle U_\rho f, f\rangle \le \langle U_\rho h, h\rangle.$$

We proved a special case of this result for balanced f in Question 4.1.11.

In the rest of this section, we will show that if all influences of $f\colon \{-1,1\}^n \to [-1,1]$ are small, then $\langle T_\rho f, f\rangle$ is not much larger than the corresponding answer in Gaussian space. The proof is a bit technical, so you can skip some of the calculations to get the gist of the argument.

We start with the case in which f has small degree, say $\deg f = d$.

Question 4.3.2. Let $f\colon \{-1,1\} \to [-1,1]$ have degree d, and extend f to $\mathbb{R}^n$ using its Fourier expansion.

(a) Show that $\langle T_\rho f, f\rangle = \mathbb{E}[S(T_{\sqrt{\rho}} f)]$, where $S(x) = \min(x^2, 1)$.

Hint: Since $T_{\sqrt{\rho}}$ is an averaging operator, $T_{\sqrt{\rho}} f \in [-1,1]$.

(b) Show that S is 2-Lipschitz, and so obtain a bound on

$$\left| \mathop{\mathbb{E}}_{\{-1,1\}^n}[S(T_{\sqrt{\rho}} f)] - \mathop{\mathbb{E}}_{\gamma}[S(U_{\sqrt{\rho}} f)] \right|.$$

(c) It might seem that we can bound the latter expectation using Borell's theorem, but actually $U_{\sqrt{\rho}} f$ need not be $[-1,1]$-valued. Let clip be the function that clips its argument to $[-1,1]$. Bound $|\mathbb{E}[U_{\sqrt{\rho}} f] - \mathbb{E}[\mathrm{clip}(U_{\sqrt{\rho}} f)]|$.

Hint: Apply the invariance principle to the 1-Lipschitz function $\mathrm{dist}_{[-1,1]}$, which measures the L_1-distance of its argument to the interval $[-1,1]$; use $\mathbb{E}_{\{-1,1\}^n}[\mathrm{dist}_{[-1,1]}(T_{\sqrt{\rho}} f)] = 0$.

(d) Let $\Lambda_\rho(\mu)$ be the bound given by Borell's theorem (i.e., $\langle U_\rho h, h\rangle$ where h is a hyperplane with $\mathbb{E}[h] = \mu$). Show that Λ_ρ is Lipschitz.

Hint: Let x, y be ρ-correlated Gaussians. Since $\langle U_\rho h, h\rangle = \|h\|^2 - 2\Pr[h(x) \ne h(y)] = 1 - 4\Pr[x < \mu < y]$, it suffices to show that $\Pr[x < \mu < y]$ is Lipschitz in μ.

(e) Apply Borell's theorem to $\mathrm{clip}(U_{\sqrt{\rho}} f)$ to obtain a bound on $\mathbb{E}_\gamma[S(\mathrm{clip}(U_{\sqrt{\rho}} f))]$.

(f) Deduce a bound on $\mathbb{E}_\gamma[S(U_{\sqrt{\rho}} f)]$ using the fact that S is Lipschitz, and conclude a bound on $\langle T_\rho f, f\rangle$.

When f doesn't have low degree, we make it low degree by adding a small amount of noise.

Question 4.3.3. Let $f\colon \{-1,1\}^n \to [-1,1]$, and extend f to $\mathbb{R}^n$ using its Fourier expansion.

(a) Let $g = T_{1-\delta} f$. Show that $\|g^{>d}\|_1 \le \|g^{>d}\|_2 \le e^{-d\delta}$.

(b) Find a value of ρ' such that $\langle T_{\rho'} g, g\rangle = \langle T_\rho f, f\rangle$.

(c) Show that if Ψ is C-Lipschitz then

$$\Big|\mathop{\mathbb{E}}_{\{-1,1\}^n}[\Psi(g)] - \mathop{\mathbb{E}}_{\gamma}[\Psi(g)]\Big| \le \Big|\mathop{\mathbb{E}}_{\{-1,1\}^n}[\Psi(g^{\le d})] - \mathop{\mathbb{E}}_{\gamma}[\Psi(g^{\le d})]\Big| + 2C\|g^{>d}\|_1.$$

(d) Apply the preceding question to obtain a bound on $\langle T_\rho f, f\rangle$ in terms of ρ' and the maximum influence of f.

(e) Show that

$$|\langle T_\rho f, f\rangle - \langle T_{\rho'} f, f\rangle| \le \frac{|\rho - \rho'|}{1-\rho'}.$$

Hint: Note $\frac{\mathrm{d}}{\mathrm{d}\rho}\langle T_\rho f, f\rangle = \sum_{d=1}^{\infty} d\rho^{d-1}\|f^{=d}\|^2$. *Use* $d\rho^{d-1} \le \frac{1}{1-\rho}$.

(f) Deduce a bound on $\langle T_\rho f, f\rangle$ in terms of ρ and the maximum influence of f when $\mathbb{E}[f] = 0$.