18.218 Topics in Combinatorics Spring 2021 – Lectures 13,14

Dor Minzer

1 Motivation and statement

In this lecture, we will begin discussing the invariance principle, which is a useful tool allowing one to transfer questions from the Boolean hypercube into Gaussian space. This is useful for several reasons: in Gaussian space, one may use several properties that are non-existent in the Boolean hypercube. One example is rotation invariance (i.e., a the Gaussian distribution over $\mathbb{R}^n$ is invariant under rotations) which is absent from the cube as rotations of Boolean vectors need not be Boolean vectors themselves.

An example of this phenomenon is already apparent in the well-known central-limit theorem. This theorem states that if $X_1, \ldots, X_n$ are "reasonable" random variables, independently distributed with mean 0 and variance 1, then the distribution of $(X_1 + \ldots + X_n)/\sqrt{n}$ approaches a standard Gaussian random variable $N(0, 1)$. This phrasing, while correct, is a bit misleading in a sense. The point here is that if the random variables $X_1, \ldots, X_n$ are reasonable and normalized, then the limiting distribution of $(X_1 + \ldots + X_n)/\sqrt{n}$ does not really depend on the specific distribution of $X_1, \ldots, X_n$, and will be the same. In other words, if we look at the linear function $f(z_1, \ldots, z_n) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_i$, then the asymptotic distribution of $f(X_1, \ldots, X_n)$ is the same for al reasonable $X_1, \ldots, X_n$. For example, we have that

$$f(X_1, \ldots, X_n) \approx f(G_1, \ldots, G_n),$$

where $X_1, \ldots, X_n$ are reasonable and normalized, and $G_1, \ldots, G_n$ are standard Gaussians.

The additional fact that $f(G_1, \ldots, G_n)$ is distributed as a standard Gaussian itself should be thought of as a "miracle" in this context; the way we have stated the statement suggests that perhaps one can prove such result for more general class of functions f. Indeed, the main question the invariance principle investigates, is what classes of functions we can prove such universality of the probability law of $f(X_1, \ldots, X_n)$ for.

To get some intuition into this question, we consider a few examples.

- $f(z_1, \ldots, z_n) = z_1$.
- $f(z_1, \ldots, z_n) = \prod_{i=1}^{100} z_i$.
- $f(z_1, \ldots, z_n) = \frac{1}{\sqrt{\binom{n}{3}}} \sum_{|S|=3} \prod_{i \in S} z_i$.

What goes wrong in the first 3 examples? How can you eliminate them? The issue with the first example is that there is a variable with large influence; this means that in a sense, f looks like a dictatorship, and for such functions it is clear that a uniform bit $\{-1, 1\}$ looks differently from a Gaussian random variable. This is also the issue with the second example. The issue with the third example is that the degree of f is high. The result, that will be the focus of this and next lecture, asserts that if one requires the function to not have influential variables and be of low-degree, then an invariance principle holds. More formally:

Theorem 1.1. *For all $d \in \mathbb{N}$, if $f(x_1, \dots, x_n) = \sum\limits_{|S| \leqslant d} \widehat{f}(S)\chi_S(x)$ is a function of degree at most d, and $\psi\colon \mathbb{R} \to \mathbb{R}$ is a smooth function with $\|\psi'''\|_\infty \leqslant C$, then*

$$\mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}[\psi(f(x))] - \mathop{\mathbb{E}}_{z\sim N(0,I_n)}[\psi(f(z))] \leqslant \frac{C}{2}2^{3d/2}\sum_{i=1}^{n} I_i[f]^{3/2}.$$

Corollary 1.2. *For all $C, \varepsilon > 0$, $d \in \mathbb{N}$ there is $\tau > 0$ such that if $f(x_1, \dots, x_n) = \sum\limits_{|S| \leqslant d} \widehat{f}(S)\chi_S(x)$ is a function of degree at most d, $\psi\colon \mathbb{R} \to \mathbb{R}$ is a smooth function with $\|\psi'''\|_\infty \leqslant C$ and $\mathsf{var}(f) \leqslant C$, then*

$$\mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}[\psi(f(x))] - \mathop{\mathbb{E}}_{z\sim N(0,I_n)}[\psi(f(z))] \leqslant \varepsilon.$$

Proof. Using the last theorem, we have that this difference is bounded by $\frac{C}{3}2^{3d/2}\tau I[f] \leqslant \frac{C}{3}2^{3d/2}\tau d\mathsf{var}(f) \leqslant C^2 9^d\sqrt{\tau}$, so choosing $\tau = \left(\frac{\varepsilon}{C^2 9^d}\right)^2$ finishes the proof. □

Thus, the theorem asserts that the distributions of $f(x)$ and $f(z)$ look very similar as far as *smooth test functions* are concerned. The above formulation of the invariance principle is the most basic version of it and there are extensions of it:

1. to non-smooth functions, such as $\psi(t) = 1_{t \leqslant 10}$. Proving these extensions requires smooth approximation to such functions, and the idea of anti-concentration in Gaussian space.
2. There is an extension of this result to functions that are not low-degree, but are close to low-degree functions and Lipshitz functions ψ.
3. The fact that z is distributed according to a standard Gaussian random variable is not very important, and similar statements can be made as long as: (1) the first and second moment of coordinates of x and z match, and (2) one has a hypercontractive inequality for both functions in x, and functions in z.

In this lecture, we will first present prove a variant of Theorem 1.1 in the special case that f is a linear function. This is a basic result in probability theory called the Berry-Essen Theorem, and will help us in order to introduce the replacement method. We will then explain the difference and challenges that will arise when we try to adapt the argument to the setting of Theorem 1.1, and then briefly discuss hypercontractivity in Gaussian space.

2 The Berry-Essen Theorem

Theorem 2.1. *If $f(x_1, \dots, x_n) = \sum\limits_{i=1}^{n} a_i x_i$, and $\psi\colon \mathbb{R} \to \mathbb{R}$ is a smooth function with $\|\psi'''\|_\infty \leqslant C$, then*

$$\mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}[\psi(f(x))] - \mathop{\mathbb{E}}_{z\sim N(0,I_n)}[\psi(f(z))] \leqslant \frac{C}{2}\sum_{i=1}^{n} a_i^3.$$

Proof. Let $x \sim \{-1,1\}^n$, $z \sim N(0, I_n)$ be independent, and for each $0 \leqslant t \leqslant n$ consider the following hybrid distribution:

$$U_t = (x_1, \dots, x_t, z_{t+1}, \dots, z_n); \qquad U_{-(t+1)} = (x_1, \dots, x_t, z_{t+2}, \dots, z_n).$$

Note that $U_0 = z$, $U_n = x$, so our difference can be written as

$$\mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}[\psi(f(U_n))] - \mathop{\mathbb{E}}_{z\sim N(0,I_n)}[\psi(f(U_0))] = \sum_{t=0}^{n-1} \mathop{\mathbb{E}}_{x,z}[\psi(f(U_{t+1}))] - \mathop{\mathbb{E}}_{x,z}[\psi(f(U_t))]$$
$$\leqslant \sum_{t=0}^{n-1} \left| \mathop{\mathbb{E}}_{x,z}[\psi(f(U_{t+1}))] - \mathop{\mathbb{E}}_{x,z}[\psi(f(U_t))] \right| .$$

Our goal is to bound the summand corresponding to t by Ca_t^3. Fix t. Since f is linear, we may write $f(U_{t+1}) = g(U_{-(t+1)}) + a_{t+1}x_{t+1}$ and $f(U_t) = g(U_{-(t+1)}) + a_{t+1}z_{t+1}$, where g is a function on $n-1$ coordinates indexed by $i = 1, \dots, t, t+2, \dots, n$, and defined by $g(u) = \sum_{i\neq t+1} a_i u_i$. We may then write the tth-summand in the above sum as

$$\mathop{\mathbb{E}}_{x,z}\left[\psi\big(g(U_{-(t+1)}) + a_{t+1}x_{t+1}\big)\right] - \mathop{\mathbb{E}}_{x,z}\left[\psi\big(g(U_{-(t+1)}) + a_{t+1}z_{t+1}\big)\right] .$$

Fix $u = U_{-(t+1)}$, and expand g according to Taylor's theorem around the point $g(u)$. We get

$$\psi(g(u)+w) = \psi(g(u)) + \psi'(g(u))w + \frac{1}{2}\psi''(g(u))w^2 + \frac{1}{3!}\psi'''(\xi)w^3,$$

where $\xi \in (g(u), g(u)+w)$ is some point. Thus,

$$\mathop{\mathbb{E}}_{x,z}\left[\psi\big(g(U_{-(t+1)}) + a_{t+1}x_{t+1}\big)\right] =$$
$$\mathop{\mathbb{E}}_{x,z}\left[\psi(g(U_{-(t+1)})) + \psi'(g(U_{-(t+1)}))a_{t+1}x_{t+1} + \frac{1}{2}\psi''(U_{-(t+1)})a_{t+1}^2x_{t+1}^2 + \frac{1}{6}\psi'''(g(\xi_x(U_{-(t+1)})))a_{t+1}x_{t+1}^3\right],$$

where $\xi_x(U_{-(t+1)})$ is some random variable. Using the fact that $U_{-(t+1)}$ and x_{t+1} are independent and that the first and second moment of x_{t+1} are 0 and 1 respectively, we get that

$$\mathop{\mathbb{E}}_{x,z}\left[\psi\big(g(U_{-(t+1)}) + a_{t+1}x_{t+1}\big)\right]$$
$$= \mathop{\mathbb{E}}_{x,z}\left[\psi(g(U_{-(t+1)})) + \frac{1}{2}\psi''(\xi(U_{-(t+1)}))a_{t+1}^2 + \frac{1}{6}\psi'''(\xi_x(g(U_{-(t+1)})))a_{t+1}x_{t+1}^3\right].$$

Similarly, we have

$$\mathop{\mathbb{E}}_{x,z}\left[\psi\big(g(U_{-(t+1)}) + a_{t+1}z_{t+1}\big)\right]$$
$$= \mathop{\mathbb{E}}_{x,z}\left[\psi(g(U_{-(t+1)})) + \frac{1}{2}\psi''(\xi(U_{-(t+1)}))a_{t+1}^2 + \frac{1}{6}\psi'''(g(\xi_z(U_{-(t+1)})))a_{t+1}z_{t+1}^3\right],$$

and taking the difference we get

$$\left| \mathop{\mathbb{E}}_{x,z}\left[\psi\big(g(U_{-(t+1)}) + a_{t+1}x_{t+1}\big)\right] - \mathop{\mathbb{E}}_{x,z}\left[\psi\big(g(U_{-(t+1)}) + a_{t+1}z_{t+1}\big)\right] \right|$$
$$\leqslant \frac{1}{6} \mathop{\mathbb{E}}_{x,z}\left[\left|\psi'''(g(\xi_x(U_{-(t+1)})))a_{t+1}^3x_{t+1}^3\right| + \left|\psi'''(g(\xi_z(U_{-(t+1)})))a_{t+1}^3z_{t+1}^3\right|\right] .$$

To finish the proof, we use the triangle inequality and bound each expectation separately. For the first one we have

$$\mathop{\mathbb{E}}_{x,z}\left[\psi'''(g(\xi_x(U_{-(t+1)})))a_{t+1}^3 x_{t+1}^3\right] \leqslant C\left|a_{t+1}\right|^3 \mathop{\mathbb{E}}_{x,z}\left[\left|x_{t+1}\right|^3\right] = C\left|a_{t+1}\right|^3$$

as $|x_{t+1}| \leqslant 1$. For the second one we have

$$\mathop{\mathbb{E}}_{x,z}\left[\psi'''(g(\xi_z(U_{-(t+1)})))a_{t+1}^3 z_{t+1}^3\right] \leqslant C\left|a_{t+1}\right|^3 \mathop{\mathbb{E}}_{x,z}\left[\left|z_{t+1}\right|^3\right] = C\left|a_{t+1}\right|^3 \frac{4}{\sqrt{2\pi}}.$$

Combining, we get that

$$\mathop{\mathbb{E}}_{x,z}\left[\psi(f(U_{t+1}))\right] - \mathop{\mathbb{E}}_{x,z}\left[\psi(f(U_t))\right] \leqslant \frac{1}{6}\left(1+\frac{4}{\sqrt{2\pi}}\right)C\left|a_{t+1}\right|^3 \leqslant \frac{C}{2}\left|a_{t+1}\right|^3. \qquad \square$$

Is this error bound even good? Note that in the central-limit theorem setting, we would have $a_i = \frac{1}{\sqrt{n}}$, so the error bound we have simplifies to $\frac{C}{2\sqrt{n}}$, which is very decent. In general, one can expect a bound on the sum of squares of the a_i's, say $\sum_{i=1}^{n} a_i^2 \leqslant 1$ (as is often the case in applications), and then we automatically get that the error can be further upper bounded by $\frac{C}{2}\max_i |a_i|$.

2.1 Generalizing the argument to low-degree polynomials

Can you see how to adapt the above argument to the setting of Theorem 1.1? What did we really do when we wrote $f(U_{t+1}) = g(U_{-(t+1)}) + a_{t+1}z_{t+1}$? What we really did here is check the influence of variable $t+1$ on the function at the point U_{t+1}. This can be generalized to low-degree polynomials by considering

$$g(U_{t+1}) = \sum_{S \not\ni t+1} \widehat{f}(S)\chi_S(U_{t+1}), \qquad \partial_{t+1} f(U_{t+1}) = \sum_{S \ni t+1} \widehat{f}(S)\chi_{S\setminus\{t+1\}}(U_{t+1}),$$

and then we can write $f(U_{t+1}) = g(U_{t+1}) + z_{t+1}\partial_{t+1}f(U_{t+1})$ and $f(U_t) = g(U_t) + x_{t+1}\partial_{t+1}f(U_t)$. Noting that both $g(U_{t+1})$ and $\partial_{t+1}f(U_{t+1})$ do not depend on the $t+1$ coordinate, we get that $g(U_{t+1}) = g(U_t)$, $\partial_{t+1}f(U_{t+1}) = \partial_{t+1}f(U_t)$. At this point, one may attempt to run the argument from the proof of Theorem 2.1, and everything goes through until the part where we need to bound the third powers of the remainder of Taylor's theorem. We will do that using hypercontractivity, but we should note here that we have a function that takes as input both Gaussian as well as bits, so we should first justify that the hypercontractive inequality holds for such functions.

3 Hypercontractivity in Gaussian space

Hypercontractivity can be abstracted and generalized beyond the Boolean hypercube and you can read about such formalization in Ryan O'Donnell's book. Our treatment here would be more specialized to the setting we are in.

Consider the Gaussian real line, i.e. $(\mathbb{R}, \mu)$ where $\mu(z) = \frac{1}{\sqrt{2\pi}}e^{-z^2/2}$ is the Gaussian density measure. We consider the space of functions $f\colon \mathbb{R} \to \mathbb{R}$ equipped with the inner product $\langle f, g\rangle = \int_{\infty}^{\infty} f(x)g(x)d\mu$.

One may to find the analog of the Fourier expansion in this setting, and indeed there is such one. A good orthonormal set in this case is known as Hermite polynomials, given as $h_0(z) \equiv 1$, and for $k \geqslant 1$

$$h_k(z) = (-1)^k e^{z^2/2} \frac{d^k}{dz^k} e^{-z^2/2}.$$

The first few Hermite polynomials are $h_1(z) = z$, $h_2(z) = z^2 - 1$, $h_3(z) = z^3 - 3z$, and they satisfy a bunch of nice properties we will not discuss further here.

Thus, we get a basis for the space of functions $f\colon (\mathbb{R}^n, \mu^{\otimes n}) \to \mathbb{R}$ by $h_{\vec{k}}(z_1, \ldots, z_n)$ where $\vec{k} = (k_1, \ldots, k_n)$ and $h_{\vec{k}}(z_1, \ldots, z_k) = \prod_{i=1}^{n} h_{k_i}(z_i)$. The Hermite expansion of f is

$$f(z) = \sum_{\vec{k}} \widehat{f}(\vec{k}) h_{\vec{k}}(z).$$

Lastly, we need the notion of degrees. The degree of $h_{\vec{k}}$ is $k_1 + \ldots + k_n$, and the degree of f is the maximum degree of $h_{\vec{k}}$ such that $\widehat{f}(\vec{k}) \neq 0$.

Lemma 3.1 (Hypercontractivity for Gaussian space)**.** *Suppose $f\colon (\mathbb{R}^n, \mu^{\otimes n}) \to \mathbb{R}$ is a function of degree at most d, and $q \geqslant 2$. Then*

$$\|f\|_q \leqslant \sqrt{q-1}^d \|f\|_2.$$

Proof. Consider the sequence of functions g_r for $r = 1, \ldots, \infty$ where we have $x_{i,j}$ independent ± 1 bits for $i = 1, \ldots, n$ and $j = 1, \ldots, r$, defined by

$$g_r(x) = f\left(\frac{\sum_{j=1}^{r} x_{1,j}}{\sqrt{r}}, \ldots, \frac{\sum_{j=1}^{r} x_{n,j}}{\sqrt{r}}\right).$$

Note that as $\frac{\sum_{j=1}^{r} x_{1,j}}{\sqrt{r}}$ approach a standard Gaussian random variable, we have that

$$\lim_{r\to\infty} \mathop{\mathbb{E}}_{x}\left[|g_r(x)|^\ell\right] = \int_{-\infty}^{\infty} |f(z)|^\ell \, d\mu^{\otimes n}$$

for all $\ell \in \mathbb{N}$. Note that g_r has degree at most d, so combining this with hypercontractivity for bits we get that

$$\|f\|_4^4 = \lim_{r\to\infty} \|g_r\|_4^4 \leqslant \lim_{r\to\infty} \sqrt{q-1}^{4d} \|g_r\|_2^4 = \sqrt{q-1}^{4d} \|f\|_2^4,$$

finishing the proof. □

In a similar fashion, we may prove a hypercontractive inequality for functions that get as input both $\pm$ bits and Gaussians. For $f\colon \{-1,1\}^t \times \mathbb{R}^{n-t} \to \mathbb{R}$, we consider the natural orthonormal basis indexed by $(S, \vec{k})$ where $S \subseteq [t]$, $\vec{k} = (k_{t+1}, \ldots, k_n)$ and given as $\chi_{S,\vec{k}}(x, z) = \chi_S(x) h_{\vec{k}}(z)$. We define the degree of $\chi_{S,\vec{k}}$ as $|S| + k_{t+1} + \ldots + k_n$, and the degree of f as the maximal degree of $\chi_{S,\vec{k}}$ supported in its Fourier expansion.

Lemma 3.2. *Suppose $f\colon \{-1,1\}^t \times \mathbb{R}^{n-t} \to \mathbb{R}$ is a function of degree at most d, and $q \geqslant 2$. Then*

$$\|f\|_q \leqslant \sqrt{q-1}^d \|f\|_2.$$

4 Proof of Theorem 1.1

We are now in the position to prove Theorem 1.1. The proof is almost the same as the proof of Theorem 2.1, and as so we will be more brief and focus on the places where there is a difference.

Proof. Let $x \sim \{-1,1\}^n$, $z \sim N(0, I_n)$ be independent, and for each $0 \leqslant t \leqslant n$ consider the following hybrid distribution:

$$U_t = (x_1, \ldots, x_t, z_{t+1}, \ldots, z_n).$$

Note that $U_0 = z$, $U_n = x$, so our difference can be bounded as before by

$$\sum_{t=0}^{n-1} \mathop{\mathbb{E}}_{x,z}\left[\psi(f(U_{t+1}))\right] - \mathop{\mathbb{E}}_{x,z}\left[\psi(f(U_t))\right] \ .$$

Fix t, and recall the functions

$$g(U_{t+1}) = \sum_{S \not\ni t+1} \widehat{f}(S)\chi_S(U_{t+1}), \qquad \partial_{t+1} f(U_{t+1}) = \sum_{S \ni t+1} \widehat{f}(S)\chi_{S\setminus\{t+1\}}(U_{t+1}),$$

We may write $f(U_{t+1}) = g(U_{t+1}) + x_{t+1}\partial_{t+1} f(U_{t+1})$ and $f(U_{t+1}) = g(U_{t+1}) + z_{t+1}\partial_{t+1} f(U_{t+1})$, and then write the tth-summand in the above sum as

$$\mathop{\mathbb{E}}_{x,z}\left[\psi\big(g(U_{t+1}) + x_{t+1}\partial_{t+1} f(U_{t+1})\big)\right] - \mathop{\mathbb{E}}_{x,z}\left[\psi\big(g(U_{t+1}) + z_{t+1}\partial_{t+1} f(U_{t+1})\big)\right] \ .$$

We use Taylor's theorem to get that

$$\begin{aligned}\mathop{\mathbb{E}}_{x,z}\left[\psi\big(g(U_{t+1}) + x_{t+1}\partial_{t+1} f(U_{t+1})\big)\right] = \mathop{\mathbb{E}}_{x,z}\Big[&\psi(g(U_{t+1})) + \psi'(g(U_{t+1}))x_{t+1}\partial_{t+1} f(U_{t+1}) \\ &+ \frac{1}{2}\psi''(g(U_{t+1}))x_{t+1}^2\partial_{t+1} f(U_{t+1})^2 \\ &+ \frac{1}{6}\psi'''(g(\xi_x(U_{t+1})))x_{t+1}^3\partial_{t+1} f(U_{t+1})^3\Big],\end{aligned}$$

and

$$\begin{aligned}\mathop{\mathbb{E}}_{x,z}\left[\psi\big(g(U_{t+1}) + z_{t+1}\partial_{t+1} f(U_{t+1})\big)\right] = \mathop{\mathbb{E}}_{x,z}\Big[&\psi(g(U_{t+1})) + \psi'(g(U_{t+1}))z_{t+1}\partial_{t+1} f(U_{t+1}) \\ &+ \frac{1}{2}\psi''(g(U_{t+1}))z_{t+1}^2\partial_{t+1} f(U_{t+1})^2 \\ &+ \frac{1}{6}\psi'''(g(\xi_z(U_{t+1})))z_{t+1}^3\partial_{t+1} f(U_{t+1})^3\Big].\end{aligned}$$

Thus, the first three terms match, and taking the difference we get

$$\begin{aligned}&\mathop{\mathbb{E}}_{x,z}\left[\psi\big(g(U_{t+1}) + a_{t+1}x_{t+1}\big)\right] - \mathop{\mathbb{E}}_{x,z}\left[\psi\big(g(U_{t+1}) + a_{t+1}z_{t+1}\big)\right] \\ &\leqslant \frac{1}{6} \mathop{\mathbb{E}}_{x,z}\left[\psi'''(g(\xi_x(U_{t+1})))x_{t+1}^3\partial_{t+1} f(U_{t+1})^3 + +\psi'''(g(\xi_z(U_{t+1})))z_{t+1}^3\partial_{t+1} f(U_{t+1})^3\right] \ .\end{aligned}$$

To bound the first expectation, we note that it is at most

$$C \cdot \mathop{\mathbb{E}}_{x,z}\left[|\partial_{t+1} f(U_{t+1})|^3\right] = C \cdot \|\partial_{t+1} f\|_3^3 \leqslant C(\sqrt{2}^d \|\partial_{t+1} f\|_2)^3 \leqslant C 2^{3d/2} I_{t+1}[f]^{3/2}.$$

For the second expectation, we bound it by

$$C \cdot \mathop{\mathbb{E}}_{x,z}\left[|z_{t+1}|^3 \, |\partial_{t+1} f(U_{t+1})|^3\right] = C\frac{4}{\sqrt{2\pi}} \cdot \|\partial_{t+1} f\|_3^3 \leqslant \frac{4C}{\sqrt{2\pi}} 2^{3d/2} I_{t+1}[f]^{3/2},$$

and combining these bounds finishes the proof. □

5 Extensions of the invariance principle

We shall now see several extensions of the invariance principle. These are by no way extensive.

5.1 Invariance principle for non-smooth test functions

In this section, we show that the invariance principle continues to hold for some non-smooth functions. We will consider cutoff functions, i.e. $\psi_t(y) = 1_{y \geqslant t}$, and for simplicity we consider the case $t = 0$.

Theorem 5.1. *For all $d \in \mathbb{N}$, $\varepsilon > 0$ there is $\tau > 0$ such that if $f(x_1, \ldots, x_n) = \sum_{|S| \leqslant d} \widehat{f}(S)\chi_S(x)$ is a function of degree at most d, and $\max_i I_i[f] \leqslant \tau$, then*

$$\mathop{\mathbb{E}}_{x \sim \{-1,1\}^n}[\psi_0(f(x))] - \mathop{\mathbb{E}}_{z \sim N(0, I_n)}[\psi_0(f(z))] \leqslant \varepsilon.$$

5.1.1 Smooth approximation of ψ_0

To prove this statement, we use smooth approximations. Namely, we fix a parameter δ and find a function ψ_δ such that:

1. $\psi_\delta \colon \mathbb{R} \to [0, 1]$ has continuous third derivative and $\|\psi_\delta'''\|_\infty \leqslant O\left(\frac{1}{\delta^3}\right)$.
2. $\psi_\delta(y) = 0$ for $y \leqslant 0$ and $\psi_\delta(y) = 1$ for $y \geqslant \delta$.

This is a standard construction from calculus, and we quickly outline it below. Consider $h \colon \mathbb{R} \to [0, \infty)$ defined by $h(y) = \alpha e^{-1/(1-y^2)}$ for $|y| \leqslant 1$ and $h(y) = 0$ otherwise, where α is chosen so that the integral of h is 1; the function h is called a mollifier. Then h is smooth and $\|h'''\|_\infty = O(1)$. Consider

$$\psi(y) = (1_{(-\infty,0]} * h)(y).$$

1. If $y \leqslant -1$, then

$$\psi(y) = \int_{-\infty}^{\infty} 1_{(-\infty,0]}(w) h(y - w) dw = \int_{-\infty}^{0} h(y - w) dw = 1.$$

2. If $y > 1$, then

$$\psi(y) = \int_{-\infty}^{\infty} 1_{(-\infty,0]}(w) h(y - w) dw = \int_{-\infty}^{0} h(y - w) dw = 0.$$

Additionally, ψ is smooth with $\|\psi\|_\infty = O(1)$. Take $\psi_2(y) = \psi_2(1-y)$, so that $\psi_2 = 0$ on $y \leqslant 0$, and $\psi_2 = 1$ on $y \geqslant 2$. Take $\psi_\delta(y) = \psi_2(\frac{y}{\delta})$ so that $\psi_\delta = 0$ for $y \leqslant 0$ and $\psi_3 = 1$ for $y \geqslant \delta$. We have by the chain rule that $\|\psi_\delta'''\|_\infty \leqslant O(1/\delta^3) \cdot \|\psi\|_\infty = O(1/\delta^3)$.

5.1.2 An anticoncetration bound in Gaussian space

If $G \sim N(0,1)$, and $I \subseteq \mathbb{R}$ is an interval of length ε, then one can easily show that $\Pr\left[|G| \leqslant \varepsilon\right] \leqslant O(\varepsilon)$. The following theorem, due to Carbery and Wright, generalizes this fact to multi-linear polynomials.

Theorem 5.2. *Suppose $f(x) = \sum\limits_{0<|S|\leqslant d} a_S\chi_S$ is a multi-linear polynomial such that $\sum\limits_S a_S^2 \leqslant 1$, and $I \subseteq \mathbb{R}$ is an interval of length at most ε. Then*

$$\Pr_{z\sim N(0,1)}\left[|f(z)| \leqslant \varepsilon\right] \leqslant O(d\varepsilon^{1/d}).$$

5.1.3 Proof of Theorem 5.1

We prove that

$$\mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}\left[\psi_0(f(x))\right] - \mathop{\mathbb{E}}_{z\sim N(0,I_n)}\left[\psi_0(f(z))\right] \leqslant \varepsilon,$$

and the proof of the other inequality is analogous. Let $\delta > 0$ to be determined, and pick ψ_δ from the previous section. Then

$$\mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}\left[\psi_0(f(x))\right] \leqslant \mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}\left[\psi_\delta(f(x))\right] \leqslant \mathop{\mathbb{E}}_{z\sim N(0,I_n)}\left[\psi_\delta(f(z))\right] + O\left(\frac{1}{\delta^3}\right)2^{3d/2}d\sqrt{\tau}.$$

where we used Theorem 1.1. Note that $\psi_\delta(f(z)) = \psi_0(f(z))$ if $f(z) \geqslant \delta$ or $f(z) \leqslant 0$, and otherwise it is at most 1, so

$$\mathop{\mathbb{E}}_{z\sim N(0,I_n)}\left[\psi_\delta(f(z))\right] \leqslant \mathop{\mathbb{E}}_{z\sim N(0,I_n)}\left[\psi_0(f(z))\right] + \Pr_{z\sim N(0,I_n)}\left[0 \leqslant f(z) \leqslant \delta\right].$$

Combining the two inequalities and using Theorem 5.2 we get that

$$\mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}\left[\psi_0(f(x))\right] \leqslant \mathop{\mathbb{E}}_{z\sim N(0,I_n)}\left[\psi_0(f(z))\right] + O\left(\frac{1}{\delta^3}\right)2^{3d/2}d\sqrt{\tau} + O(d\delta^{1/d}).$$

We choose $\delta = 2^{-C\cdot d\log(d/\varepsilon)}$ for large enough $C > 0$ so that the second error term is at most $\varepsilon/2$, and then τ small enough so that the first term is at most ε, and the proof is concluded.

5.1.4 Piecewise smooth functions

Using Theorem 5.1, it is not hard now to show that invariance holds for all piecewise smooth test functions ψ, i.e. test functions for which there is a partition of the real line into intervals $\mathbb{R} = I_1 \cup \ldots \cup I_r$ such that ψ is smooth in the interior of each I_i. We omit the proof.

5.2 Invariance principle for functions with small Fourier tails

Next, we extend the invariance principle to functions that are not low-degree, but almost low degree.

Theorem 5.3. *For all $C, \varepsilon > 0$, $d \in \mathbb{N}$ there is $\tau > 0$ such that if $f(x_1, \ldots, x_n) = \sum \widehat{f}(S)\chi_S(x)$ is a function such that $\max_i I_i[f^{\leqslant d}] \leqslant \tau$, and $\psi \colon \mathbb{R} \to \mathbb{R}$ is a piecewise smooth function C-Lipshitz function with $\|\psi'''\|_\infty \leqslant C$,*

$$\mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}[\psi(f(x))] - \mathop{\mathbb{E}}_{z\sim N(0,I_n)}[\psi(f(z))] \leqslant \varepsilon + 2C\|f^{\geqslant d}\|_2.$$

Proof. Write $f = f^{\leqslant d} + f^{>d}$. Then since ψ is C-Lipshitz

$$\mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}[\psi(f(x))] - \mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}\left[\psi(f^{\leqslant d}(x))\right] \leqslant \mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}\left[C \ f^{\leqslant d}(x) \ \right] \leqslant C\|f^{\leqslant d}\|_2.$$

Similarly,

$$\mathop{\mathbb{E}}_{z\sim N(0,I_n)}[\psi(f(z))] - \mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}\left[\psi(f^{\leqslant d}(z))\right] \leqslant \mathop{\mathbb{E}}_{z\sim N(0,I_n)}\left[C \ f^{\leqslant d}(z) \ \right] \leqslant C\|f^{\leqslant d}\|_2.$$

The result now follows from Theorem 5.1. □

5.3 Other extensions of the invariance principle

There are other extensions of the invariance principle: multi-dimensional versions, more relaxed requirements, general product domains and more. We will not elaborate on these points further.

6 Majority is stablest

We finish this lecture by showing one prominent application of the invariance principle, which was actually the original motivation for it. The Gaussian analog of the majority is stablest theorem was already known in the 19th century, and the idea of Mossel, O'Donnell and Oleszkiewicz was to deduce the Boolean case from it. We will show this reduction, starting with presenting the theorem in the Gaussian case.

Definition 6.1. *For $\rho \in [0, 1]$, the operator T_ρ acting on functions $f \colon \mathbb{R}^n \to \mathbb{R}$ is defined as*

$$\mathrm{U}_\rho(z) = \mathop{\mathbb{E}}_{w\sim N(0,I_n)}\left[f(\rho z + \sqrt{1-\rho^2}w)\right].$$

Note that the distribution of $\rho z + \sqrt{1-\rho^2}w$ is standard Gaussian that is ρ-correlated with z, so this is the analog of the noise operator from the Boolean case. It is easy to check that $\mathrm{U}_\rho\chi_S(z) = \rho^{|S|}\chi_S(z)$ for all monomials χ_S.

Definition 6.2. *Given $\rho \in [0, 1]$ and $f \colon \mathbb{R}^n \to \mathbb{R}$, the noise stability of f with parameter ρ is $\mathsf{Stab}_\rho(f) = \langle f, \mathrm{U}_\rho f\rangle$.*

The Gaussian analog of the majority is stablest theorem states that half-spaces maximize the noise stability of balanced, bounded functions:

Theorem 6.3. *[Borel's theorem] Let $\rho \in [0,1]$, and $f\colon \mathbb{R}^n \to [-1,1]$ with $\mathbb{E}[f] = 0$. Then $\mathsf{Stab}_\rho(f) \leqslant 1 - \frac{2}{\pi}\mathsf{Arccos}(\rho)$.*

We will not prove this theorem here, though at least for many values of ρ there is a relatively simple proof due to Kindler and O'Donnell, and in general there are several known proofs which are not too hard. Instead, we will show how to deduce the Majority is Stablest theorem from it.

Theorem 6.4. *For all $\varepsilon > 0$, $\rho \in (0,1)$ there are $d \in \mathbb{R}$ and $\tau > 0$ such that if $f\colon \{-1,1\}^n \to [-1,1]$ is balanced and $\max_i I_i[f^{\leqslant d}] \leqslant \tau$, then*

$$\mathsf{Stab}_\rho(f) \leqslant 1 - \frac{2}{\pi}\mathsf{Arccos}(\rho) + \varepsilon.$$

Proof. Let $\delta > 0$ small to be determined, and let $f' = \mathrm{T}_{1-\delta} f$. In the homework you will show that $\mathsf{Stab}_\rho(f) \leqslant \mathsf{Stab}_\rho(f') + O_\rho(\delta)$, and in the rest of the proof we will upper bound $\mathsf{Stab}_\rho(f')$.

Take $d \in \mathbb{N}$ to also be determined later, and define the function $\mathsf{Square}(t) = t^2$ for $t \in [0,1]$ and $\mathsf{Square}(t) = 0$ for $t \leqslant 0$, and otherwise 1. Then Square is 2-Lipshitz and piecewise smooth, so we may apply the invariance principle on it. Now that

$$\mathsf{Stab}_\rho(f') = \langle f', \mathrm{T}_\rho f'\rangle = \langle \mathrm{T}_{\sqrt{\rho}} f', \mathrm{T}_{\sqrt{\rho}} f'\rangle = \mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}\Big[\mathsf{Square}(\mathrm{T}_{\sqrt{\rho}} f'(x))\Big].$$

Thus, by Theorem 5.3 we have

$$\mathsf{Stab}_\rho(f') \leqslant \mathop{\mathbb{E}}_{z\sim N(0,I_n)}\Big[\mathsf{Square}(\mathrm{T}_{\sqrt{\rho}} f'(z))\Big] + \frac{\varepsilon}{2} + 4\|(\mathrm{T}_{\sqrt{\rho}} f')^{\geqslant d}\|_2$$

for $\tau(d,\varepsilon) > 0$ small enough. Note that

$$\|(\mathrm{T}_{\sqrt{\rho}} f')^{\geqslant d}\|_2^2 \leqslant \sum_{|S|\geqslant d} \widehat{f}'(S)^2 \leqslant (1-\delta)^{2d},$$

so the second error term is at most $4(1-\delta)^d$. Next, we would like to apply Theorem 6.3. Towards this end, note first that as f is multilinear, $\mathrm{T}_{\sqrt{\rho}} f' = \mathrm{U}_{\sqrt{\rho}} f'$. It may not necessarily be the case that f' is bounded on $\mathbb{R}^n$ (in fact it is most likely not), and to get around this issue we will argue that it is "mostly bounded".

Define $\mathsf{trunc}(s) = s$ if $|s| \leqslant 1$, and otherwise 1 if $s > 1$ or -1 if $s < -1$, and consider the function $F(z) = \mathsf{trunc}(f'(z))$. By Theorem 5.3

$$\mathop{\mathbb{E}}_{z\sim N(0,I_n)}\left[\,|F(z) - f'(z)|\,\right] = \mathop{\mathbb{E}}_{z\sim N(0,I_n)}\Big[\mathsf{dist}(f'(z),[0,1])\Big] \leqslant \mathop{\mathbb{E}}_{x\sim\{-1,1\}^n}\Big[\mathsf{dist}(f'(x),[0,1])\Big] + 4\|(f')^{\geqslant d}\|_2,$$

and the first expectation is 0 whereas the error term is at most $4(1-\delta)^d$. In particular, it follows that

$$\mathop{\mathbb{E}}_{z\sim N(0,I_n)}\Big[\mathsf{Square}(\mathrm{T}_{\sqrt{\rho}} f'(z))\Big] \leqslant \mathop{\mathbb{E}}_{z\sim N(0,I_n)}\Big[\mathsf{Square}(\mathrm{T}_{\sqrt{\rho}} F(z))\Big] + 4(1-\delta)^d = \mathsf{Stab}_\rho(F) + 4(1-\delta)^d$$

Finally, to apply Theorem 6.3 we would like F to be balanced. Note that

$$\mathop{\mathbb{E}}_z[F] \;=\; \mathop{\mathbb{E}}_z[F] - \mathop{\mathbb{E}}_z[f'] \;\leqslant \mathop{\mathbb{E}}_z[\,|F(z) - f'(z)|\,] \leqslant 4(1-\delta)^d,$$

so F is nearly balanced. It is not hard to show that in that case, the conclusion of Theorem 6.3 holds with bit of an error bound. For example, letting $F' = \frac{F - \mathbb{E}[F]}{1+4(1-\delta)^d}$, we have that F' is balanced and bounded so $\mathsf{Stab}_\rho(F') \leqslant 1 - \frac{2}{\pi}\mathsf{Arccos}(\rho)$, and

$$\mathsf{Stab}_\rho(F) - \mathsf{Stab}_\rho(F') \leqslant 4\|F - F'\|_1 \leqslant 4 \cdot 4(1-\delta)^d(\|F\|_1 + 1) \leqslant 32(1-\delta)^d.$$

Combining everything, we get that

$$\begin{aligned}\mathsf{Stab}_\rho(f) \leqslant \mathsf{Stab}_\rho(f') + O_\rho(\delta) &\leqslant \mathop{\mathbb{E}}_{z \sim N(0,I_n)}\left[\mathsf{Square}(\mathrm{T}_{\sqrt{\rho}} f'(z))\right] + O_\rho(\delta) + O((1-\delta)^d) \\ &\leqslant \mathsf{Stab}_\rho(F) + O_\rho(\delta) + O((1-\delta)^d) \\ &\leqslant 1 - \frac{2}{\pi}\mathsf{Arccos}(\rho) + O_\rho(\delta) + O((1-\delta)^d).\end{aligned}$$

Choosing $\delta(\rho) > 0$ now so that the first error bound is at most ε, and then d so that the second error bound is at most $\varepsilon/2$, finishes the proof. $\square$