CAS CS 599 B: Mathematical Methods for TCS

Lecturer: Mark Bun | Spring 2022

Lecture Notes 5:

Central Limit Theorem and Majority

Reading.

- O'Donnell, Analysis of Boolean Functions §5.1-5.2

1 Influence of Majority

When we talked about influences, we estimated the total influence of Majority as $\mathbf{I}[\mathrm{MAJ}_n] \approx \sqrt{2/\pi} \cdot \sqrt{n}$ using the combinatorics of the hypercube and Stirling's formula. Today we'll see a more flexible and general way to get the same result that illustrates some important general principles about Boolean Fourier analysis and beyond.

Lemma 1. *Let $f : \{-1,1\}^n \to \{-1,1\}$ be a monotone function. Then $\mathbf{Inf}_i[f] = \hat{f}(i)$, and hence $\mathbf{I}[f] = \sum_{i=1}^n \hat{f}(i)$.*

Proof. Recall the directional derivative operator

$$(D_i f)(x) = \frac{f(x^{(i\to 1)} - f(x^{i\to -1}))}{2}.$$

If f is monotone, then we have $(D_i f)(x) = 1$ if i is pivotal for x and $(D_i f)(x) = 0$ otherwise. So

$$\mathbf{Inf}_i[f] = \mathbb{E}[D_i f(x)] = \widehat{D_i f}(\emptyset) = \hat{f}(i).$$

□

We now estimate the total influence of Majority.

$$\begin{aligned}
\mathbf{I}[\mathrm{MAJ}_n] &= \sum_{i=1}^n \widehat{\mathrm{MAJ}_n}(i) \\
&= \sum_{i=1}^n \mathop{\mathbb{E}}_x [\mathrm{MAJ}_n(x) x_i] \\
&= \mathop{\mathbb{E}}_x [\mathrm{MAJ}_n(x)(x_1 + \cdots + x_n)] \\
&= \mathop{\mathbb{E}}_x [|x_1 + \cdots + x_n|] \\
&= \sqrt{n} \cdot \mathop{\mathbb{E}}_x \left[\frac{|x_1 + \cdots + x_n|}{\sqrt{n}} \right]
\end{aligned}$$

"Classical" Central Limit Theorem: If $X_1, \ldots, X_n$ is a sequence of i.i.d. random variables with mean μ and variance σ^2, then their sample average $\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n)$ resembles a Gaussian with mean μ and variance σ^2/n. Formally, as $n \to \infty$,

$$\sqrt{n}(\overline{X} - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2).$$

Applying this to our calculation with $X_i = x_i$, we get

$$\mathbf{I}[\mathrm{MAJ}_n] \approx \sqrt{n} \cdot \mathop{\mathbb{E}}_{Z \sim \mathcal{N}(0,1)}[|Z|] = \sqrt{n} \cdot \sqrt{2/\pi}$$

using the fact that

$$\mathop{\mathbb{E}}_{Z \sim \mathcal{N}(0,1)}[|Z|] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-z^2/2} |z| \, dz = \frac{2}{\sqrt{2\pi}} \int_0^{\infty} z e^{-z^2/2} \, dz = \frac{2}{\sqrt{2\pi}} \, e^{-z^2/2} \Big|_0^{\infty} = \sqrt{\frac{2}{\pi}}.$$

To make this approximation rigorous[1] we introduce a quantitative version of the CLT that is meaningful for finite n.

Theorem 2 (Berry-Esseen)**.** *Let $X_1, \ldots, X_n$ be random variables with $\mathbb{E}[X_i] = 0$ and $\mathrm{Var}\,[X_i] = \sigma_i^2$. Assume $\sum_{i=1}^n \sigma_i^2 = 1$. Let $S = \sum_{i=1}^n X_i$ and let $Z \sim \mathcal{N}(0,1)$ be a standard Gaussian. Then for all $t \in \mathbb{R}$,*

$$|\Pr[S \le t] - \Pr[Z \le t]| \le \gamma$$

where $\gamma = \sum_{i=1}^n \|X_i\|_3^3$.

Let's now sketch how to use the Berry-Esseen Theorem to estimate the total influence of Majority. We set

- $X_i = x_i/\sqrt{n}$
- $\sigma_i^2 = 1/n \implies \sum_{i=1}^n \sigma_i^2 = 1$
- $\|X_i\|_3^3 := \mathbb{E}[|X_i|^3] = n^{-3/2} \implies \gamma = 1/\sqrt{n}$

Thus, for any $T \ge 1$, we can estimate

$$\begin{aligned}
|\mathbb{E}[|S|] - \mathbb{E}[|Z|]| &= \left| \int_0^{\infty} \Pr[|S| \ge t] - \Pr[|Z| \ge t] \, dt \right| \\
&\le \int_0^T |\Pr[|S| \ge t] - \Pr[|Z| \ge t]| \, dt + \int_T^{\infty} \Pr[|S| \ge t] + \Pr[|Z| \ge t] \, dt \\
&\le T\gamma + \int_T^{\infty} 4e^{-t^2/2} \, dt \\
&\le O(T/\sqrt{n} + e^{-T^2/2}).
\end{aligned}$$

Setting $T = \sqrt{\ln n}$ makes this error bound $O(\sqrt{(\log n)/n})$. Hence we get that $\mathbf{I}[\mathrm{MAJ}_n] = \sqrt{n} \cdot \sqrt{2/\pi} \pm O(\sqrt{\log n})$. It turns out that a strengthening of Berry-Esseen lets one reduce the error bound to $O(1/\sqrt{n})$, hence $\mathbf{I}[\mathrm{MAJ}_n] = \sqrt{n} \cdot \sqrt{2/\pi} \pm O(1)$. (One can actually show that $\mathbf{I}[\mathrm{MAJ}_n] = \sqrt{n} \cdot \sqrt{2/\pi} \pm O(1/\sqrt{n})$.)

[1] $Y_n \xrightarrow{d} Y$ is not, by itself, enough to guarantee that $\mathbb{E}[Y_n] \to \mathbb{E}[Y]$

2 Noise Stability of Majority

We will now see how to use the CLT to study the noise stability of the Majority function.

Theorem 3. *For any* $\rho \in [-1, 1]$,

$$\lim_{n\to\infty} \mathbf{Stab}_\rho[\mathrm{MAJ}_n] = \frac{2}{\pi} \arcsin \rho.$$

As before, we'll heuristically justify this using the (multidimensional) CLT.

Multidimensional CLT: If $\vec{X}_1, \dots, \vec{X}_n$ is a sequence of i.i.d. random vectors with mean vector $\vec{\mu}$ and invertible covariance matrix Σ, then their sample average $\overline{X} = \frac{1}{n}(\vec{X}_1 + \cdots + \vec{X}_n)$ resembles a multivariate Gaussian with mean $\vec{\mu}$ and covariance matrix Σ/n. Formally, as $n \to \infty$,

$$\sqrt{n}(\overline{X} - \vec{\mu}) \xrightarrow{d} \mathcal{N}(0, \Sigma).$$

One can then make it fully rigorous using (multidimensional) Berry-Esseen. Recall by definition that

$$\mathbf{Stab}_\rho[\mathrm{MAJ}_n] = \mathop{\mathbb{E}}_{(x,x')\rho\text{-corr}} \left[\mathrm{MAJ}_n(x)\mathrm{MAJ}_n(x')\right] = \mathop{\mathbb{E}}_{(x,x')\rho\text{-corr}} \left[\mathrm{sgn}\left(\frac{x_1 + \cdots + x_n}{\sqrt{n}}\right) \mathrm{sgn}\left(\frac{x'_1 + \cdots + x'_n}{\sqrt{n}}\right)\right].$$

For each $i \in [n]$, let

$$\vec{X}_i = \begin{pmatrix} x_i \\ x'_i \end{pmatrix}$$

These are i.i.d. vectors with

$$\mathbb{E}\left[\vec{X}_i\right] = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\Sigma = \begin{pmatrix} \mathbb{E}[x_i^2] & \mathbb{E}[x_i x'_i] \\ \mathbb{E}[x_i x'_i] & \mathbb{E}[(x'_i)^2] \end{pmatrix} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} =: \Sigma_\rho.$$

Applying the multidimensional CLT, we get the approximation

$$\mathbf{Stab}_\rho[\mathrm{MAJ}_n] \approx \mathop{\mathbb{E}}_{(Z,Z')\sim\mathcal{N}(0,\Sigma_\rho)} \left[\mathrm{sgn}(Z)\,\mathrm{sgn}(Z')\right] = 1 - 2 \Pr_{(Z,Z')}[\mathrm{sgn}(Z) \neq \mathrm{sgn}(Z')].$$

The result follows from the next calculation ("Sheppard's Formula") about multivariate Gaussians, together with the complementary angle formula $\arcsin\rho + \arccos\rho = \pi/2$.

Theorem 4 (Sheppard's Formula)**.** *Let* Z, Z' *be* ρ*-correlated standard Gaussians. Then*

$$\Pr[\mathrm{sgn}(Z) \neq \mathrm{sgn}(Z')] = \frac{\arccos\rho}{\pi}.$$

Proof. Fix any pair of unit vectors $u, v \in \mathbb{R}^2$ such that $\langle u, v\rangle = \rho$. Then the pair (Z, Z') is identically distributed to $(\langle G, u\rangle, \langle G, v\rangle)$ where $G \sim \mathcal{N}(0,1)^2$ is a standard Gaussian. Then

$$\Pr[\mathrm{sgn}(Z) \neq \mathrm{sgn}(Z')] = \Pr[\mathrm{sgn}\langle G, u\rangle \neq \mathrm{sgn}\langle G, v\rangle] = \Pr[\mathrm{sgn}\langle G/\|G\|, u\rangle \neq \mathrm{sgn}\langle G/\|G\|, v\rangle].$$

Since standard Gaussians are rotationally symmetric, $G/\|G\|$ is just a uniformly random unit vector. So this is just the probability that a random unit vector is positively correlated (makes an acute angle with) u, but negatively correlated (makes an obtuse angle with) v, or vice versa. A bit of geometry (complementary angles and such) shows that this is θ/π where $\theta = \arccos\rho$ is the angle between u and v. ☐

Note that in this argument, the only property we used of Majority was that it could be expressed as the sign of a "nice" linear combination of independent bits. The argument indeed generalizes to the class of *linear threshold functions* or LTFs.

Definition 5. A function $f : \{-1,1\}^n \to \{-1,1\}$ is a linear threshold function if it can be expressed as $f(x) = \mathrm{sgn}(a_0 + a_1x_1 + \cdots + a_nx_n)$. It is *unbiased* if $a_0 = 0$.

Using multidimensional Berry-Esseen, one can show that

Theorem 6. *Let $f : \{-1,1\}^n \to \{-1,1\}$ be an unbiased LTF with $\sum_{i=1}^n a_i^2 = 1$ and $|a_i| \le \varepsilon$ for all i. Then $|\mathbf{Stab}_\rho[f] - (2/\pi)\arcsin\rho| \le O(\varepsilon/\sqrt{1-\rho^2})$.*

In the special case of Majority, $a_i = 1/\sqrt{n}$ for every i, so one can take $\varepsilon = 1/\sqrt{n}$.

The results we've discussed illustrate the "invariance principle" for Boolean Fourier analysis: A sufficiently "nice" (e.g., low degree or noise stable, with small influences) function on the hypercube is well-approximated by a function on Gaussian space. Thus, one can use analytic tools like rotation invariance and calculus on Gaussian r.v.'s. Chapter 11 of O'Donnell gives many more illustrations, culminating in the proof of the "Majority is Stablest" theorem:

Theorem 7. *Let $\rho \in (0,1)$. Then for any $f : \{-1,1\}^n \to [-1,1]$ with $\mathbb{E}[f] = 0$ and $\mathbf{Inf}_i[f] \le \tau$ for all i,*

$$\mathbf{Stab}_\rho[f] \le \frac{2}{\pi}\arcsin\rho + o_\tau(1).$$

3 More on LTFs

Linear threshold functions are quite basic objects, arising in learning theory (where they are often referred to as halfspaces) and circuit complexity (where they correspond to threshold circuits). Chow's Theorem is a beautiful result about the Fourier spectra of LTFs.

Theorem 8. *Let $f : \{-1,1\}^n \to \{-1,1\}$ be an LTF. Let $g : \{-1,1\}^n \to \{-1,1\}$ be any function. If $\hat{g}(S) = \hat{f}(S)$ for all $|S| \le 1$, then $g = f$.*

That is, the level-0 and level-1 Fourier coefficients of an LTF completely characterize it within the class of all Boolean functions.

Proof. Let $f(x) = \mathrm{sgn}(\ell(x))$ where $\ell(x)$ is a degree-1 polynomial. Then for every $x \in \{-1,1\}^n$, we have $f(x)\ell(x) = |\ell(x)| \ge g(x)\ell(x)$, with equality iff $f(x) = g(x)$. Thus (using Plancharel twice),

$$\sum_{|S|\le 1} \hat{f}(S)\ell(S) = \mathbb{E}[f(x)\ell(x)] \ge \mathbb{E}[g(x)\ell(x)] = \sum_{|S|\le 1} \hat{g}(S)\ell(S).$$

By assumption, the left-most and right-most expressions are equal. This can only happen if $f(x)\ell(x) = g(x)\ell(x)$ for every x, which can only happen if $f(x) = g(x)$ for every x. □

Another highlight is Peres's Theorem, which you can read about in Chapter 5.5 of O'Donnell.

Theorem 9. *For any LTF $f : \{-1,1\}^n \to \{-1,1\}$, we have $\mathbf{NS}_\delta[f] \le O(\sqrt{\delta})$.*

One could achieve the same conclusion for all "regular" halfspaces from our noise stability calculation via the CLT. It turns out there's a nice elementary proof that works for all halfspaces.