Computer Science 294 Lecture 15 Notes

Daniel Raban

March 7, 2023

1 LTFs and Central Limit Theorems

1.1 Fourier coefficients of linear threshold functions

Recall the following definition.

Definition 1.1. A **linear threshold function (LTF)** is a function $f : \{\pm 1\}^n \to \{\pm 1\}$ of the form

$$f(x) = \operatorname{sgn}(a_0 + a_1 x_1 + \cdots + a_n x_n), \qquad a_0, \dots, a_n \in \mathbb{R}.$$

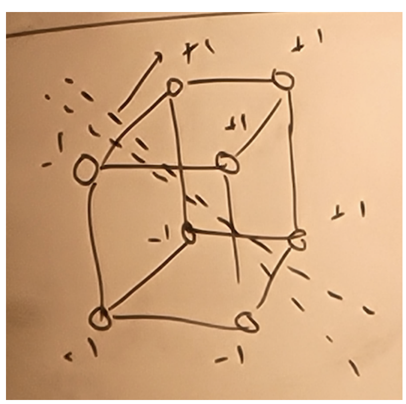
Example 1.1.

$$\mathrm{MAJ}_n(x) = \operatorname{sgn}(x_1 + \cdots + x_n).$$

Example 1.2. Dictator functions, OR, and AND are all LTFs.

Theorem 1.1 (Chow's parameters)**.** *Let $f : \{\pm 1\}^n \to \{\pm 1\}$ be a linear threshold function, and let $g : \{\pm 1\}^n \to \{\pm 1\}$ be any other function. If for every $|S| \leq 1$, $\widehat{f}(S) = \widehat{g}(S)$, then $f = g$.*

Here is a geometric interpretation of LTFs: A LTF defines a half-space, and everything on one side get the value $+1$, while everything on the other side gets -1. In fact, you can perturb the hyperplane by a little bit, so we can assume without loss of generality that $a_0 + a_1 x_1 + \cdots + a_n x_n \neq 0$ for all $x \in \{\pm 1\}^n$.



Remark 1.1. This theorem tells us that there are at most $(2^n)^{n+1}$

Proof. Write $f(x) = \operatorname{sgn}(\ell(x))$ with $\deg \ell \leq 1$, $\ell(x) = \sum_{|S|\leq 1} \widehat{\ell}(S) \prod_{i\in S} x_i$, and $\ell(x) \neq 0$ for all $x \in \{\pm 1\}^n$. By Plancherel's theorem,

$$\langle f, \ell \rangle = \sum_{|S|\leq 1} \widehat{f}(s)\widehat{\ell}(s).$$

On the other hand,

$$\begin{aligned}\langle f, \ell \rangle &= \mathbb{E}[f(X)\ell(X)] \\ &= \mathbb{E}[|\ell(X)|] \\ &\geq \mathbb{E}[g(X)\ell(X)] \\ &= \sum_{|S|\leq 1} \widehat{g}(S)\widehat{\ell}(S).\end{aligned}$$

Since $\widehat{f}(S) = \widehat{g}(S)$ for all $|S| \leq 1$, this inequality must have been an equality. So we must have $g(x) = \operatorname{sgn}(\ell(x))$ for all x. □

Definition 1.2. A **degree** d **polynomial threshold function (LTF)** is a function $f : \{\pm 1\}^n \to \{\pm 1\}$ of the form

$$f(x) = \operatorname{sgn}(p), \qquad p \in \mathbb{R}[x_1, \ldots, x_n].$$

Remark 1.2. You can show that if f is a degree d PTF and $\widehat{f}(S) = \widehat{g}(S)$ for all $|S| \leq d$, then $f = g$.

Theorem 1.2 (Gotsman-Linial)**.** *Let $f(x) = \operatorname{sgn}(\ell(x))$ be a LTF. Then $W^{\leq 1}(f) \geq 1/2$.*

Proof.

$$\begin{aligned}\|\ell\|_1 &= \mathbb{E}_X[|\ell(X)|] \\ &= \mathbb{E}_X[\ell(X)f(X)] \\ &= \langle \ell, f \rangle \\ &= \sum_{|S|\leq 1} \widehat{\ell}(S)\widehat{f}(S)\end{aligned}$$

Let $f^{\leq 1}(x) = \sum_{|S|\leq 1} \widehat{f}(S)\chi_S(x)$ be the truncated version of f.

$$= \langle \ell, f^{\leq 1} \rangle$$

Using Cauchy-Schwarz,

$$= \sqrt{\mathbb{E}[\ell(X)^2]}\sqrt{\mathbb{E}[f^{\leq 1}(X)^2]}$$

$$= \|\ell\|_2 \sqrt{\sum_{|S|\le 1} \widehat{f}(S)^2}.$$

Rearranging, we get

$$W^{\le 1}(f) \ge \frac{\|\ell\|_1^2}{\|\ell_2\|^2}$$

By the KK inequality for linear functions,

$$\ge \frac{1}{2}.$$

□

Example 1.3.

$$W^{\le 1}(\chi_i) = 1.$$

Example 1.4.

$$W^{\le 1}(\mathrm{MAJ}) = \sum_{i=1}^{n} \widehat{f}(\{i\})^2 \approx \sum_{i=1}^{n} \left(\sqrt{\frac{2}{\pi}} \frac{1}{\sqrt{n}} \right)^2 = \frac{2}{\pi}$$

1.2 The central limit theorem and influence of the majority function

Recall the Central Limit Theorem from probability theory.

Theorem 1.3. *Let $X_1, \dots, X_n$ be iid with mean μ and variance σ^2. Then*

$$\frac{X_1 + \cdots + X_n - n\mu}{\sqrt{n}} \xrightarrow{d} N(0, \sigma^2).$$

First, let's use this to analyze the influence of the majority function.

Proposition 1.1.

$$\mathbb{I}(\mathrm{MAJ}_n) = \sum_i \mathrm{Inf}(\mathrm{MAJ}_n) \approx \sqrt{\frac{2}{\pi}} \sqrt{n}.$$

Proof.

$$\mathbb{I}(\mathrm{MAJ}_n) = \sum_i \mathrm{Inf}(\mathrm{MAJ}_n)$$

Since the majority function is monotone,

$$= \sum_{i=1}^{n} \mathrm{MAJ}_n(\{i\})$$

$$= \sum_{i=1}^{n} \mathbb{E}_{X \sim U_n}[\mathrm{MAJ}_n(X) X_i]$$

$$= \mathbb{E}_X \left[\mathrm{MAJ}_n(X) \sum_{i=1}^{n} X_i \right]$$

The majority function is the sign of this sum.

$$= \mathbb{E}_X \left[\left| \sum_{i=1}^{n} X_i \right| \right]$$
$$= \sqrt{n}\, \mathbb{E}_X \left[\left| \frac{\sum_{i=1}^{n} X_i}{\sqrt{n}} \right| \right]$$
$$\approx \sqrt{n}\, \mathbb{E}_{Z \sim N(0,1)}[|Z|]$$

Z is a continuous random variable with probability density function $\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$.

$$= \sqrt{n} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} |z| \, dz$$
$$= 2\sqrt{n} \int_{0}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} z \, dz$$
$$= \frac{2}{\sqrt{2\pi}} (-e^{-z^2/2}) \Big|_0^\infty$$
$$= \sqrt{n} \frac{2}{\sqrt{2\pi}}$$
$$= \sqrt{n} \sqrt{\frac{2}{\pi}}. \qquad \square$$

How precise is this approximation? The following theorem gives us an answer.

Theorem 1.4 (Berry-Esseen)**.** *Let $X_1, \ldots, X_n$ be independent random variables with $\mathbb{E}[X_i] = 0$ and $\mathrm{Var}(X_i) = \sigma_i^2$. Assume that $\sum_{i=1}^n \sigma_i^2 = 1$. Let $S = X_1 + \cdots + X_n$, so that $\mathbb{E}[S] = 0$ and $\mathrm{Var}(S) = 1$. Then for every $u \in \mathbb{R}$,*

$$|\mathbb{P}(S \le u) - \mathbb{P}_{Z \sim N(0,1)}(Z \le u)| \le \mathrm{const} \cdot \beta, \qquad \beta = \sum_{i=1}^{n} \mathbb{E}[|X_i|^3].$$

To apply this to LTFs, let $X_i = a_i x_i$, where $x_i \in \{\pm 1\}$. Then $\mathrm{Var}(X_i) = \mathbb{E}[(a_i x_i)^2] = a_i^2$, so we want $\sum_{i=1}^n a_i^2 = 1$. Then $\sum_{i=1}^n a_i x_i \approx N(0,1)$ with error

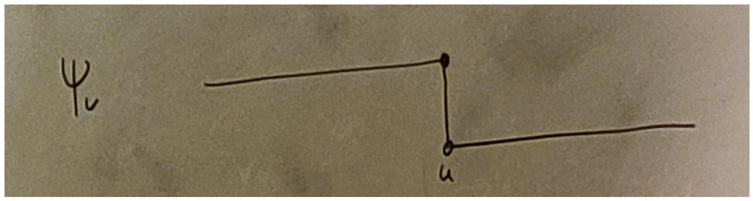
$$\beta = \sum_{i=1}^{n} \mathbb{E}[|X_i|^3] = \sum_{i=1}^{n} |a_i|^3 \le (\max_i |a_i|) \underbrace{\sum_{i=1}^{n} a_i^2}_{=1} = \max_i(|a_i|).$$

The majority function is the case of $a_1 = a_2 = \cdots = a_n = 1/\sqrt{n}$.

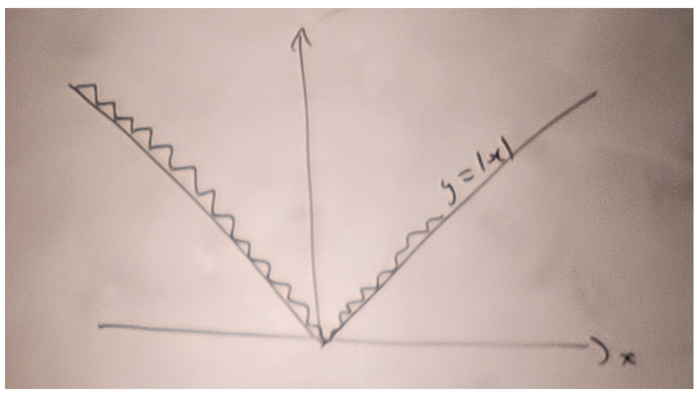
We can think of the Berry-Esseen theorem more generally than just in terms of cumulative distribution functions. The condition

$$\mathbb{P}(S \leq u) \approx \mathbb{P}(Z \leq u)$$

is the same as saying $\mathbb{E}[\psi_u(S)] \approx \mathbb{E}[\psi_u(Z)]$ for the test function ψ_u:



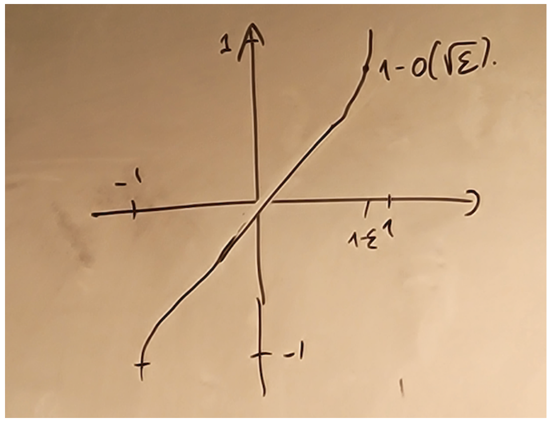
The test function that we actually care about is the absolute value function. We can try to approximate this by step functions near 0 (and use a Chernoff bound away from 0).



1.3 Stability of the majority function

Theorem 1.5. *For all $\rho \in [-1, 1]$,*

$$\mathrm{Stab}_\rho(\mathrm{MAJ}_n) \xrightarrow{n\to\infty} \frac{2}{\pi}\arcsin\rho.$$



Proof. Recall that a ρ-correlated pair of inputs is given by independent (X_i, Y_i) with $\mathbb{E}[X_i] = \mathbb{E}[Y_i] = 0$ with $\mathbb{E}[X_i Y_i] = \rho$.

$$\begin{aligned}
\mathrm{Stab}_\rho(\mathrm{MAJ}_n &= \mathbb{E}_{(X,Y)\ \rho\text{-corr.}}[\mathrm{MAJ}_n(X)\,\mathrm{MAJ}_n(Y)] \\
&= \mathbb{E}_{(X,Y)\ \rho\text{-corr.}}\left[\mathrm{sgn}\left(\frac{X_1 + \cdots + X_n}{\sqrt{n}}\right)\mathrm{sgn}\left(\frac{Y_1 + \cdots + Y_n}{\sqrt{n}}\right)\right]
\end{aligned}$$

What is the correlation of the Gaussians that approximate these sums? Let

$$S^{[i]} = \begin{bmatrix} \frac{1}{\sqrt{n}} X_i \\ \frac{1}{\sqrt{n}} Y_i \end{bmatrix} \in \mathbb{R}^2, \qquad S = \begin{bmatrix} \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \\ \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i \end{bmatrix}.$$

The CLT over $\mathbb{R}^2$ tells us that S is approximately a multivariate Gaussian $Z = \begin{bmatrix} Z_i \\ Z_2 \end{bmatrix}$, which is determined only by its mean and covariance matrix. The mean is

$$\mathbb{E}[S] = \sum_{i=1}^n \mathbb{E}[S^{[i]}] = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

and the covariances are

$$\begin{aligned}
\mathbb{E}[S_X S_Y] &= \mathbb{E}\left[\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i\right)\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i\right)\right] \\
&= \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i Y_i] \\
&= \rho.
\end{aligned}$$

Similarly, we see that $\mathbb{E}[S_X^2] = 1$ and $\mathbb{E}[S_Y^2] = 1$. So we get the covariance matrix

$$\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}.$$

So we get that

$$\begin{aligned} \mathrm{Stab}_\rho(\mathrm{MAJ}_n) &\approx \mathbb{E}_{Z=\left[\begin{smallmatrix} Z_1 \\ Z_2 \end{smallmatrix}\right]}[\mathrm{sgn}(Z_1)\,\mathrm{sgn}(Z_2)] \\ &= \frac{2}{\pi}\arcsin\rho. \end{aligned}$$ □

The last equality is given by Sheppard's Lemma, which we will see next time. Later, we will prove the following result.

Theorem 1.6 ("Majority is stablest")**.** *Fix $0 < \rho < 1$, and let $f : \{\pm 1\}^n \to \{\pm 1\}$ with $\mathbb{E}[f] = 0$ and $\mathrm{Inf}_i(f) \le \varepsilon$. Then*

$$\mathrm{Stab}_\rho(f) \le \mathrm{Stab}_\rho(\mathrm{MAJ}_n) + o_{\varepsilon \to 0}(1).$$

Remark 1.3. The reason we need a condition on the influence of each voter is because dictators function are the most stable, in general.

Next time, we will also see Peres' theorem about sensitivity of linear threshold functions.

Theorem 1.7 (Peres)**.** *For every LTF f, $\mathrm{NS}_\delta(f) \le O(\sqrt{\delta})$.*