Lecture 6

Lecturer: Irit Dinur *Scribe: Matthew Coudron*

1 The Invariance Principle

This lecture will explore a useful generalization of the Central Limit Theorem of probability, which is known as the the Invariance Principle. The Invariance Principle applies in the following setting:

Imagine that we have a function $f : \{-1,1\}^n \to \mathbb{R}$. Now imagine that we have i.i.d random variables $x_i \in \{-1,1\}$ and we compute $f(x_1,...,x_n)$. Now $f(x_1,...,x_n)$ itself may also be regarded as a random variable. We know that $f : \{-1,1\}^n \to \mathbb{R}$ can be expressed as a polynomial in the x_i's $f = \sum_S \hat{f}_S \prod_{i \in S} x_i$. Given this expression we can ask how much the random variable $f(x_1,...,x_n)$ differs from $f(y_1,...,y_n)$ where the y_i's are a different set of independent random variables. For example, the y_i's could be uniformly distributed on the interval $[-1,1]$.

It turns out that, just as in the Central Limit Theorem, Gaussian random variables play a central role in the Invariance Principle. We will reserve the notation g_i to refer to a sequence of independent random variables with each g_i distributed according to $\mathcal{N}(0,1)$ (the Normal distribution with mean 0 and variance 1).

Roughly speaking, we are interested in the question: Does f "notice" the difference between inputs $(x_1,...,x_n)$ and $(g_1,...,g_n)$. More precisely, we would like to bound the difference between the random variables $f(x_1,...,x_n)$ and $f(g_1,...,g_n)$. The canonical example of such a result is the Central Limit Theorem

Theorem 1. *The Lindeberg-Levy Central Limit Theorem*

Suppose $\{X_1,X_2,...\}$ is a sequence of i.i.d. random variables with $\mathbb{E}[X_i] = \mu$ and $Var[X_i] = \sigma^2 < \infty$. Then as $n \to \infty$, the random variables $\sqrt{n}\left(\frac{X_1+...+X_n}{n} - \mu\right)$ converge in distribution to a normal distribution, $\mathcal{N}(0,\sigma^2)$.

In particular, if we define $f(z_1,...,z_n) \equiv \frac{1}{n}(z_1,...,z_n)$. Then, since $\text{Var}[x_i] = \text{Var}[g_i] = 1$, Theorem 1 implies that $\sqrt{n}f(x_1,...,x_n)$, and $\sqrt{n}f(g_1,...,g_n)$ both converge in distribution to $\mathcal{N}(0,1)$ as $n \to \infty$. Roughly speaking, this means that $\forall t$

$$\lim_{n\to\infty} \text{Prob}(\sqrt{n}f(x_1,...,x_n) \leq t) = \lim_{n\to\infty} \text{Prob}(\sqrt{n}f(g_1,...,g_n) \leq t)$$

In this sense the function $f(z_1,...,z_n) \equiv \frac{1}{n}(z_1,...,z_n)$ does not "notice" the difference between inputs $(x_1,...,x_n)$ and $(g_1,...,g_n)$, at least for large n.

However, for arbitrary functions f, it is clear that we will need to be more careful about what question we ask. For example, note that with the simple function $f(z_1,...,z_n) \equiv z_1$ we have that $f(x_1,...,x_n)$ has a Bernoulli distribution, and $f(g_1,...,g_n)$ has a Gaussian distribution, so any reasonable norm will distinguish these two inputs. We revise our question as follows.

Revised Question: Given f that has small influences, that is:

$\max_{i\in[n]} \mathrm{Inf}_i(f) \leq \tau$

then does f notice the difference between $x_1, ..., x_n$ and $g_1, ..., g_n$?

In this case we have the following theorem:

Theorem 2. *The Mossel-O'Donnell-Oleszkiewicz Invariance Principle*

Let f be a degree d polynomial over $\mathbb{R}$, $f(z_1, ..., z_n) = \sum_{S\subset[n]} \hat{f}_S \prod_{i\in S} z_i$ such that $\sum_S \hat{f}_S^2 = 1$. Assume that $\forall i$ $Inf_i(f) \equiv \sum_{S:i\in S} \hat{f}_S^2 \leq \tau$. Then, for any smooth function $\phi : \mathbb{R} \to \mathbb{R}^+$ with bounded fourth derivative ($|\phi^{(4)}| \leq B$)

$$\left|\mathbb{E}_{x\in\{\pm1\}^n}[\phi(f(x))] - \mathbb{E}_{g\in\mathcal{N}(0,1)^n}[\phi(f(g))]\right| < small(\tau, \frac{1}{d})$$

Later in the lecture we will discuss a proof of Theorem 2 where $\text{small}(\tau, \frac{1}{d}) = \tau d$. Before we discuss the proof, let us consider an application.

2 Dictatorship vs. τ-Quasirandom Tests

Here a τ-Quasirandom function is a boolean function with all influences smaller than τ just as in 2.

The boolean hypercube is used as a gadget in inapproximability redutions, where proving a completeness, soundness gap of (c,s) for the approximation boils down to devising a (c,s)-Dictatorship vs. Quasirandom test (for boolean functions) which has the following properties:

Completeness: If f is a dictatorship then $\Pr[\text{Test passes}] \geq c - o(1)$

Soundness: If f is quasirandom then $\Pr[\text{Test passes}] \leq s + o(1)$.

Let's consider a test which corresponds to inapproximability of "Max-Cut" or Max-2LIN. This test was proposed by Kindler, Khot, Mossel and O'Donnell, "KKMO".

KKMO "max-cut" test:

Fix $0 < \rho < 1$ (we will also use a parameter δ defined by $\rho = 1 - 2\delta$.

1) Let $x \in \{\pm1\}^n$ be chosen uniformly at random.

2) Choose $y \in \{\pm1\}^n$ from the distribution $\mathcal{N}_\rho(x)$ (the distribution which is ρ-correlated with x). That is choose y such that for the i^{th} bit of y (for all i), we have $y_i = x_i$ with probability ρ, and with probability $1 - \rho$ y_i is uniformly random.

3) Accept if $f(x) = f(y)$.

KKMO Completeness: If f is a dictator function (on the i^{th} coordinate say), then

$$\text{Prob}[\text{Test passes}] = \text{Prob}[x_i = y_i] = 1 - \delta$$

KKMO Soundness: If f is quasi-random (meaning all influences are small), then

$$\text{Prob[Test passes]} = \text{Prob}[f(x) = f(y)] \equiv \text{NS}_\rho(f) \leq \text{NS}_\rho(\text{MAJ}) = \frac{\text{arcCos}(\rho)}{\pi}$$

Here we define $\text{NS}_\rho(f) \equiv \text{Prob}[f(x) = f(y)]$. We will now show how the Invariance Principle can be used to calculate $\text{NS}_\rho(\text{MAJ})$ and obtain the final equality above.

A Calculation

Note that

$$\text{MAJ}(x_1, ..., x_n) = \text{sgn}(\sum_{i \in [n]} x_i) \approx \text{sgn}(g)$$

and

$$\text{MAJ}(y_1, ..., y_n) = \text{sgn}(\sum_{i \in [n]} y_i) \approx \text{sgn}(g')$$

Where g and g' are defined to be normal distributions distributed "like" $\frac{1}{n}\sum_{i \in [n]} x_i$, and $\frac{1}{n}\sum_{i \in [n]} y_i$ respectively. In fact we can use $g \sim \mathcal{N}(0, 1)$, and $g' \equiv \rho g + \sqrt{1-\rho^2} g''$ where $g'' \sim \mathcal{N}(0, 1$ and g and g'' are independent. The Invariance Principle tells use that this g and g' will be very close to having the desired distributions.

With this definition of g an g', straightforward calculus gives

$$\text{NS}_\rho(\text{MAJ}) \approx \text{Prob}[\text{sgn}(g) = \text{sgn}(g')] = \frac{\text{arcCos}(\rho)}{\pi}$$

3 Proof of the Invariance Principle

We will now prove the Invariance Principle modulo a certain lemma (Lemma 3 below). We will also discuss the key idea for proving that lemma and use it to obtain a similar (though weaker) result.

Given f as in the statement of Theorem 2 we define

$$f_i \equiv f(g_1, ..., g_i, x_{i+1}, ..., x_n)$$

Note that $f_0 = f(x_1, ..., x_n)$, and $f_n = f(g_1, ..., g_n)$. Imagine that we have a $\phi : \mathbb{R} \to \mathbb{R}^+$ as in Theorem 2.

Lemma 3.

$$\forall i, \;\; |\mathbb{E}[\phi(f_{i-1})] - \mathbb{E}[\phi(f_i)]| < (\mathit{Inf}_i)^2$$

Using Lemma 3 and the triangle inequality we can now prove Theorem 2.

Proof. Proof of Theorem 2

$$|\mathbb{E}[\phi(f(x_1,...,x_n))] - \mathbb{E}[\phi(f(g_1,...,g_n))]| = |\mathbb{E}[\phi(f_0)] - \mathbb{E}[\phi(f_n)]| = \left|\sum_{i=0}^{n-1} (\mathbb{E}[\phi(f_i)] - \mathbb{E}[\phi(f_{i+1})])\right|$$

$$\leq \sum_{i=0}^{n-1} |\mathbb{E}[\phi(f_i)] - \mathbb{E}[\phi(f_{i+1})]| \leq \sum_{i=1}^{n} (\mathrm{Inf}_i)^2 \leq \max_i \mathrm{Inf}_i \left(\sum_{i=1}^{n} \mathrm{Inf}_i\right) \leq \tau d$$

Here the first inequality follows by the triangle inequality, the second by Lemma 3, the third inequality is straight forward, and the fourth follows from the fact that f is τ-Quasirandom and degree d by assumption ($\sum_{i=1}^{n} \mathrm{Inf}_i \leq d$ for degree-d f).

□

Clearly Lemma 3 plays a key role in the proof of the Invariance Principle. We will now give a proof of a weaker version of Lemma 3 which nonetheless contains the key idea of the full proof.

Lemma 4.

$$\forall i, \;\; |\mathbb{E}[\phi(f_{i-1})] - \mathbb{E}[\phi(f_i)]| < O(B \cdot 9^d)(Inf_i)^2$$

Here B is the bound on the fourth derivative of ϕ, and d is the degree of f, just as in Theorem 2.

Proof. Recall Taylor's formula

$$\phi(R+\epsilon) = \phi(R) + \epsilon\phi'(R) + \frac{\epsilon^2}{2!}\phi^{(2)}(R) + \frac{\epsilon^3}{3!}\phi^{(3)}(R) + \frac{\epsilon^4}{4!}\phi^{(4)}(\eta)$$

where η is some point in $\mathbb{R}$.

Now, for a given i, we divide $f(z_1,...,z_n)$ up as a polynomial of z_i, so $f(z_1,...,z_n) = R(z_1,...,z_{i-1},z_{i+1},...,z_n)+$ $z_iS(z_1,...,z_{i-1},z_{i+1},...,z_n)$. Recall that

$$f_i \equiv f(g_1,...g_i,x_{i+1},...,x_n) = R(g_1,...,g_{i-1},x_{i+1},...,x_n) + g_iS(g_1,...,g_{i-1},x_{i+1},...,x_n)$$

$$f_{i-1} \equiv f(g_1,...g_i,x_{i+1},...,x_n) = R(g_1,...,g_{i-1},x_{i+1},...,x_n) + x_iS(g_1,...,g_{i-1},x_{i+1},...,x_n)$$

So defining random variables $R \equiv R(g_1,...,g_{i-1},x_{i+1},...,x_n$, and $S \equiv S(g_1,...,g_{i-1},x_{i+1},...,x_n)$ we have that

$$\mathbb{E}[\phi(f_{i-1})] = \mathbb{E}[\phi(R+x_iS)] = \mathbb{E}\left[\phi(R)\right] + \mathbb{E}\left[x_iS\phi'(R)\right] + \mathbb{E}\left[\frac{(x_iS)^2}{2!}\phi^{(2)}(R)\right]$$

$$+\mathbb{E}\left[\frac{(x_iS)^3}{3!}\phi^{(3)}(R)\right]+\mathbb{E}\left[\frac{(x_iS)^4}{4!}\phi^{(4)}(\eta)\right]=\mathbb{E}\left[\phi(R)\right]+\mathbb{E}[x_i]\mathbb{E}\left[S\phi'(R)\right]$$

$$+\mathbb{E}[x_i^2]\mathbb{E}\left[\frac{S^2}{2!}\phi^{(2)}(R)\right]+\mathbb{E}[x_i^3]\mathbb{E}\left[\frac{S^3}{3!}\phi^{(3)}(R)\right]+\mathbb{E}[x_i^4]\mathbb{E}\left[\frac{S^4}{4!}\phi^{(4)}(\eta)\right]$$

Where the second equality follows from the independence of x_i and the other variables. Very similarly,

$$\mathbb{E}[\phi(f_i)]=\mathbb{E}[\phi(R+g_iS)]=\mathbb{E}\left[\phi(R)\right]+\mathbb{E}\left[g_iS\phi'(R)\right]+\mathbb{E}\left[\frac{(g_iS)^2}{2!}\phi^{(2)}(R)\right]$$

$$+\mathbb{E}\left[\frac{(g_iS)^3}{3!}\phi^{(3)}(R)\right]+\mathbb{E}\left[\frac{(g_iS)^4}{4!}\phi^{(4)}(\eta')\right]=\mathbb{E}\left[\phi(R)\right]+\mathbb{E}[g_i]\mathbb{E}\left[S\phi'(R)\right]$$

$$+\mathbb{E}[g_i^2]\mathbb{E}\left[\frac{S^2}{2!}\phi^{(2)}(R)\right]+\mathbb{E}[g_i^3]\mathbb{E}\left[\frac{S^3}{3!}\phi^{(3)}(R)\right]+\mathbb{E}[g_i^4]\mathbb{E}\left[\frac{S^4}{4!}\phi^{(4)}(\eta')\right]$$

Again the second equality follows from the independence of g_i and the other variables. Now we note that $\mathbb{E}[g_i]=\mathbb{E}[x_i]=0$, $\mathbb{E}[g_i^2]=\mathbb{E}[x_i^2]=1$, and $\mathbb{E}[g_i^3]=\mathbb{E}[x_i^3]=0$. The first two equalities follow by definition of g_i. The third follows because the function $y\to y^3$ is odd, and both x_i and g_i are have distributions which are symmetric about the origin (the are negative exactly as often as they are positive...). It follows that

$$|\mathbb{E}[\phi(f_{i-1})]-\mathbb{E}[\phi(f_i)]|=\left|\mathbb{E}[x_i^4]\mathbb{E}\left[\frac{S^4}{4!}\phi^{(4)}(\eta)\right]-\mathbb{E}[g_i^4]\mathbb{E}\left[\frac{S^4}{4!}\phi^{(4)}(\eta')\right]\right|$$

$$\le\left|\mathbb{E}[x_i^4]\mathbb{E}\left[\frac{S^4}{4!}\phi^{(4)}(\eta)\right]\right|+\left|\mathbb{E}[g_i^4]\mathbb{E}\left[\frac{S^4}{4!}\phi^{(4)}(\eta')\right]\right|\le O(B)\left|\mathbb{E}\left[\frac{S^4}{4!}\right]\right|=O(B)\mathbb{E}\left[S^4\right]$$

Where the last inequality uses the fact that $\left|\phi^{(4)}(\eta')\right|,\left|\phi^{(4)}(\eta)\right|\le B$ by assumption, and the fact that $\mathbb{E}[x_i^4]$ and $\mathbb{E}[g_i^4]$ are both finite constant (not hard to calculate).

For the final step we will apply Hypercontractivity. Note that $S=\frac{d}{dx_i}f_{i-1}$, so that $\mathbb{E}[S^2]=\mathrm{Inf}_i(f_{i-1})=\sum_{S:i\in S}\hat{f}_S^2$. Clearly, the degree of S is at most d, so by Hypercontractivity we get

$$\mathbb{E}\left[S^4\right]\le\left(\mathbb{E}\left[S^2\right]\right)^2 9^d$$

Using the previous work gives

$$|\mathbb{E}[\phi(f_{i-1})]-\mathbb{E}[\phi(f_i)]|\le O(B)\mathbb{E}\left[S^4\right]\le O(B)\left(\mathbb{E}\left[S^2\right]\right)^2 9^d\le O(B\cdot 9^d)(\mathrm{Inf}_i)^2$$

This is the desired result. □

In fact, with a more careful use of Hypercontractivity, and the noisy hypercube function T_ρ the proof of Lemma 4 can be tightened to obtain Lemma 3.