Lecture 1

Lecture date: Aug 30 — *Scribe: Omid Etesami*

1 Polynomials of Random Variables

Elchanan started the first lecture by giving an informal overview of the course.

Consider the space of functions in the product space $L_2(\Omega_1, \mu_1) \times \ldots \times L_2(\Omega_n, \mu_n)$. Classical theory of probability mostly studies *linear* functions f of this space: $f(x_1, \ldots, x_n) = \sum_{i=1}^{n} g_i(x_i)$. Results such as Law of Large Numbers, Central Limit Theorems, and Invariance Principles show that the limiting behavior of linear functions f, as $n \to \infty$, can be described using normal random variables and Brownian motion.

In this course, we look at the space of nonlinear functions f. Thus we study a larger class of objects. This class is both richer and less "canonical" than the class of objects classically studied in probability. The limiting objects of this class are *polynomials of normal random variables*, rather than just normal random variables. This is clear if we consider the function

$$f(X) = (\frac{X_1 + \ldots + X_n}{\sqrt{n}})^2,$$

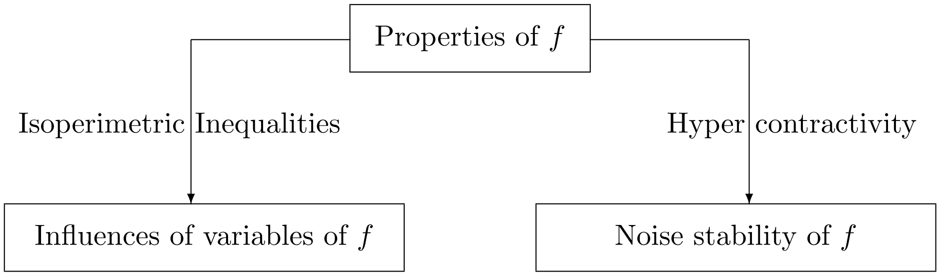
whose limit is the square of a normal random variable. This space of functions has been studied in *Gaussian chaos theory*.

In this course, general properties of f like

- the degree of f,
- the number of variables on which f depend, and
- monotonicity of f

will be related to

- the *influences* of the variables of f (the theory of perturbation of one single coordinate) through *isoperimetric inequalities*, and
- the *noise correlation* of f (the theory of perturbation of many coordinates) through *hypercontractivity*.



In the rest of this lecture, we will see a few applications.

2 Application: Learning Functions of a Few Variables

Assume a function $f : \{-1,1\}^n \to \{-1,1\}$ that only depends on $k \ll n$ coordinates: $f(x_1, \ldots, x_n) = g(x_{i_1}, \ldots, x_{i_k})$ for some function g and indices $i_1, \ldots, i_k$. We are given samples $(x^i, f(x^i))$ of f, where x^i are independent random inputs each uniformly chosen from $\{-1,1\}^n$. The goal is to efficiently find g and $i_1, \ldots, i_k$ using few samples.

The simplest way to learn f is to check all the $\binom{n}{k} 2^{2^k}$ possibilities for f, and see which is consistent with the given samples. We need at least $\log_2(\binom{n}{k} 2^{2^k}) \sim k \log_2(n) + 2^k$ samples to find f, since each sample gives at most 1 bit of information. It turns out that this number is more or less the correct number of samples required to get f. However, checking all possible functions takes $c_k n^k$ time.

Question 1 *How fast can one learn f?*

This is open. We will see, using methods of this course, that it is possible to learn f in time $c_k n^{0.71k}$.

3 Application: Threshold Phenomenon

Let $f : \{-1,1\} \to \{-1,1\}$ be a *monotone* Boolean function; that is, $x_i \leq y_i$ for all i implies $f(x) \leq f(y)$. Let $\{-1,1\}^n_\theta$ denote the biased product measure, where each coordinate has expected value θ; in other words, each coordinate is 1 with probability $(1+\theta)/2$ and -1 with probability $(1-\theta)/2$. Define $T(\theta) = \mathbf{E}_\theta[f]$.

How does function $T(\theta)$ look like?

Proposition 2 *When f is monotone, $T(\theta)$ is monotonically increasing.*

Proof: Suppose $\theta_1 \leq \theta_2$. One can choose $(x_1, \ldots, x_n) \in \{-1,1\}_{\theta_1}$ and $(y_1, \ldots, y_n) \in \{-1,1\}_{\theta_2}$ such that $x_i \leq y_i$ is always true for all i. (The vectors x and y are said to be coupled.) This way, we always have $f(x) \leq (y)$, and hence $T(\theta_1) \leq T(\theta_2)$. $\square$

When f is nontrivial, i.e. when f assumes both values -1 and 1, $T(\theta)$ increases from -1 to 1 as θ increases from -1 to 1.

Question 3 *How quickly does $T(\theta)$ change from -1 to 1?*

We will see in this course that there are essentially two extremes:

1. Course threshold: A function with course threshold is $F(x_1, \ldots, x_n) = x_1$, where $T(\theta) = \theta$.
2. Sharp threshold: A function with sharp threshold is the majority function $\text{Maj}(x_1, \ldots, x_n) = \text{sign}(\sum_{i=1}^n x_i)$. For majority on n bits, we have $T(\pm 100/\sqrt{n}) \sim \pm(1 - e^{-100})$, which shows that T suddenly jumps from close to -1 to close to 1 over the narrow interval $[-100/\sqrt{n}, +100/\sqrt{n}]$.

4 Application: First Passage Percolation

Consider the lattice Z^2. To each edge of the lattice, we give a random independent weight, which is 1 with probability 1/2 and 2 with probability 1/2. Define the distance $d(x, y)$ between two points x and y to be the weight of the least-weight path between x and y. Let $L(n) = d(x, y)$ be the distance between the two points $x = (0, 0)$, $y = (n, 0)$. We know that $n \leq L(n) \leq 2n$.

Question 4 *What is the growth rate of* $\mathbf{Var}(L(n))$*?*

Using some non-rigorous but illuminating arguments from statistical physics, $\mathbf{Var}(L(n))$ is surprisingly of order of $n^{2/3}$; however, it is open to prove this rigorously. We will show that $\mathbf{Var}(L(n)) = o(n)$.

5 Example: Isoperimetric Inequality

This example has nothing to do with polynomials of random variables, but is useful for the next application.

Consider the n-dimensional sphere S^n with the uniform (Haar) measure.

Question 5 *Among all subsets $A \subset S^n$ with a fixed measure $\mu(A)$, which one has the minimum boundary?*

Answer 6 *The cap of given measure.*

When $\mu(A) = 1/2$, the cap is a hemisphere, and the boundary is the equator.

6 Application: Paradoxes in Social Choice

Consider everyone in the class vote a bit on a certain issue, and we have a function $f : \{-1,1\}^n \to \{-1,1\}$ that aggregates all the votes into a single bit representing the class's opinion on that issue. The function can for example be a dictatorship (only depend on one person), or be a democracy (majority function).

Now assume that we want to find the favorite day in the week based on the opinions of the class. There are three candidate days: FRI, SAT, and SUN. Everyone in the class has an ordering of these three days. First, we compare FRI with SAT by aggregating, using f, the individual comparisons of FRI with SAT. Next, we similarly compare SAT with SUN, and FRI with SUN.

A paradoxical situation may happen, e.g. the overall ordering may turn out to be FRI > SAT > SUN > FRI. For any function f other than dictatorship, this paradoxical situation is inevitable.

Question 7 *Suppose that everyone has an independent and uniformly random ordering. Among all functions f, where everyone in the class has a bounded influence, which function leads to minimum probability of paradox?*

Answer 8 *We will see in this course that the answer is the majority function.*