STAT 206A: Polynomials of Random Variables 7

Lecture 7

Lecture date: Sep 20 *Scribe: Alexandra Kolla*

In this lecture, we will prove the following theorem :

Theorem 1 *Let $f, g \in L^2(\prod \mu_i)$ such that f is computed by an algorithm T that queries x_i w.p δ_i and $g = \sum_{S \in W} g_S$ where W is an anti - chain. The following relation holds:*

$$(\mathbf{Cov}[f,g])^2 \leq \mathbf{Var}[f] \cdot \sum_{i=1}^{n} \delta_i(T) \cdot I_i(g)$$

Proof: If $W = \{\emptyset\}$ then $\mathbf{Cov}[f,g] = 0 \leq$ (any positive number) and the theorem trivially holds. We therefore make, w.l.o.g. the following assumptions:

- Since W is an anti - chain $\emptyset \notin W$ therefore $\mathbf{E}[g] = \hat{g}_\emptyset = 0$.
- $\mathbf{E}[f] = 0$ since neither $\mathbf{Var}[f]$ nor $\mathbf{Cov}[f,g]$ or $\delta_i(T)$ change when shifted by a constant.
- $\mathbf{E}[fg] \geq 0$ (if not consider $-f$ in the place of f).

With the above assumptions, $\mathbf{Cov}[f,g] = \mathbf{E}[(f - \mathbf{E}[f]) \cdot (g - \mathbf{E}[g])] = \mathbf{E}[fg]$ so it suffices to show

$$\mathbf{E}[fg] \leq |f|_2 \cdot \sqrt{\sum_{i=1}^{n} \delta_i(T) \cdot I_i(g)}$$

We consider our input x to be chosen in to steps :

1. First the algorithm chooses coordinates that determine the value of f. Let P be the random subset of $[n]$ that specifies the coordinates that are questioned by T and let x_P denote those coordinates.
2. Let x_R denote the rest of the coordinates.

$x = (x_P, x_R)$ is chosen according to the given probability measure.
For the purposes of the proof, we will use the following notation. Let f be a function on our space depending on u and v:

$$\mathbf{E}_u[f(u,v)] = E[f(u,v)|v]$$

We can now rewrite the expected value :

$$\mathbf{E}[fg] = \mathbf{E}_{x_P}[\mathbf{E}_{x_R}[f(x_P, x_R)g(x_P, x_R)]] =$$
$$\mathbf{E}_{x_P}[f(x_P)\mathbf{E}_{x_R}[g(x_P, x_R)]] \leq$$
$$|f|_2\sqrt{\mathbf{E}_{x_P}[\mathbf{E}^2_{x_R}(g(x_P, x_R))]}$$

where the first line follows from the fact that f is independent of the values of coordinates in R and the second line from the Cauchy -Schwartz inequality.
What remains to show in order for the proof to be complete is

$$\mathbf{E}_{x_P}[\mathbf{E}^2_{x_R}(g(x_P, x_R))] \leq \sum_{i=1}^{n} \delta_i(T) \cdot I_i(g)$$

Define $G^{x_P}(x_R) = g(x_P, x_R)$ to be a random function conditioned on x_P. I.e. we fix value for x_P and for this value G^{x_P} is a function depending only on x_R. By definition, $\mathbf{E}^2_{x_R}[(g(x_P, x_R))] = \mathbf{E}^2_{x_R}[(G^{x_P}(x_R))]$.

From previous lectures, given x_P is fixed, we can write G^{x_P} as an orthogonal sum $g(x_P, x_R) = G^{x_P} = \sum_{S\subset[n], S\subset R} G^{x_P}_S$. Therefore,

$$\mathbf{E}^2_{x_R}(G^{x_P}(x_R)) = |G^{x_P}_\emptyset|^2_2 = \sum_{S\subset[n], S\subset R} |G^{x_P}_S|^2_2 - \sum_{\emptyset\neq S\subset[n], S\subset R} |G^{x_P}_S|^2_2 \leq \sum_{S\subset[n], S\subset R} |G^{x_P}_S|^2_2 - \sum_{S\subset W, S\subset R} |G^{x_P}_S|^2_2 \tag{1}$$

since $\emptyset \notin W$.
Now, we take expected values over x_P and we observe:

$$\mathbf{E}_{x_P}[\sum_{S\subset[n], S\subset R} |G^{x_P}_S|^2_2] = \mathbf{E}_{x_P}[\mathbf{E}_{x_R}[g^2(x_P, x_R)]] = \|g\|^2_2 \tag{2}$$

We will now try to express the quantity $\sum_{S\subset W, S\subset R} |G^{x_P}_S|^2_2$ in terms of the orthogonal projections g_S. For the function $g(x_P, x_R)$ it holds $g = \sum_{S\subset[n]} g_S$. Fixing x_P, we get:

$$g(x_P, x_R) = \sum_{S\subset[n]} g_S(x_P, x_R) = \sum_{S\subset[n], S\cap P=\emptyset} g_S(x_P, x_R) + \sum_{S\subset[n], S\cap P\neq\emptyset} g_S(x_P, x_R)$$

Now, observe that for $S \cap P \neq \emptyset$ the terms $g_S(x_P, x_R)$ are functions that live in the space $L^2_{S\cap R}$ therefore, they can be decomposed into an orthogonal sum $g_S = \sum_{S'\subset S, S'\subset R} h'_S$ where h'_S denote the projections onto the corresponding subspace. Altogether, from the expression of g (fixing x_P) we get:

$$g(x_P, x_R) = \sum_{S\subset[n], S\cap P=\emptyset} g_S(x_P, x_R) + \sum_{S\subset[n], S\cap P\neq\emptyset} \sum_{S'\subset S, S'\subset R} h'_S$$

By the assumption of the theorem, $g = \sum_{S \in W} g_S$ so we only need to look at subsets $S \subset [n]$ such that $S \in W$. We distinguish the following two cases:

1. $S \cap P = \emptyset$. Then S appears in the decomposition of $g(x_P, \cdot)$ and contributes the term $g_S(x_P, \cdot)$.
2. $S \cap P \neq \emptyset$. Then S contributes in the decomposition the term $\sum_{S' \subset S, S' \subset R} h'_S$. None of the $S' \subset S$ are in W so the total contribution is 0.

Altogether now we have :

$$\sum_{S \subset W, S \subset R} G_S^{x_P} = \sum_{S \in W, S \cap P = \emptyset} g_S(x_P, x_R),$$

and therefore

$$\sum_{S \in W, S \subset R} |G_S^{x_P}|_2^2 = \sum_{S \in W, S \cap P = \emptyset} |g_S|_2^2 \tag{3}$$

Let A_S be the event that we query none of the coordinates in S and B_S the event that we query at least one coordinate in S. From (3) we get

$$\mathbf{E}_{x_P}[\sum_{S \in W, S \subset R} |G_S^{x_P}|_2^2] = \sum_{S \in W} \mathbf{P}[A_S] |g_S|_2^2 \tag{4}$$

From (1) using (2) and (4) we get :

$$\mathbf{E}_{x_P}[\mathbf{E}_{x_R}^2[G^{x_P}(x_R)]] \leq \sum_{S \in W} |g_S|_2^2 - \sum_{S \in W} \mathbf{P}[A_S] |g_S|_2^2 =$$

$$\sum_{S \in W} \mathbf{P}[B_S] |g_S|_2^2 \leq \sum_{S \in W} |g_S|_2^2 \cdot (\sum_{i \in S} \delta_i) = (\sum_{i \in S} \delta_i) \cdot \sum_{S : i \in S} |g_S|_2^2 = \sum \delta_i I_i(g)$$

where the last equation follows from the definition of influence.

□