Lecture 3

Lecture date: Sep 10, 2007 *Scribe: Allan Sly*

1 Method of Dependency Graphs

We recall the definition of a dependency graph from the previous lecture. For a collection of random variables $\{X_i, i \in V\}$ indexed by the vertices V of a graph $G = (V, E)$ we say that G is a dependency graph if for any disjoint subsets $S, T \subseteq V$ with no edges between S and T we have $\{X_i, i \in S\}$ and $\{X_i, i \in T\}$ independent. Let $D = 1 +$ max degree G.

Lemma 1 *Suppose that $E(X_i) = 0, \sigma^2 = \mathrm{Var}(\sum X_i), W = \frac{\sum X_i}{\sigma}$ and $Z \sim N(0,1)$. Then*

$$\mathrm{Wass}(W, Z) \leq \frac{4}{\sqrt{\pi}\sigma^2}\sqrt{D^3 \sum E|X_i|^4} + \frac{D^2}{\sigma^3}\sum E|X_i|^3.$$

Proof: Let $W_i = \frac{1}{\sigma}\sum_{j \in N_i} X_j$ where $N_i = \{i\} \cup \{\text{neighbours of i}\}$. As in the case of iid random variables we have X_i and W_i independent but we do not in general have that W_i and $W - W_i$ are independent.

Take any f such that

$$|f| \leq 1, \quad |f'| \leq \sqrt{\frac{2}{\pi}}, |f''| \leq 2.$$

Then

$$EWf(W) = \frac{1}{\sigma}\sum E(X_i f(W)) = \frac{1}{\sigma}\sum E(X_i(f(W) - f(W_i))) = (I) + (II)$$

where

$$(I) = \frac{1}{\sigma}\sum E[X_i(f(W) - f(W_i) - (W - W_i)f'(W))]$$

and

$$(II) = \frac{1}{\sigma}\sum E[X_i(W - W_i)f'(W)].$$

Now

$$(I) \leq \frac{1}{\sigma}\sum \frac{1}{2}E|X_i(W - W_i)^2||f''|_\infty \leq \frac{1}{\sigma^3}\sum E|X_i(\sum_{j \in N_i} X_j)^2|$$

since $W - W_i = \frac{1}{\sigma}\sum_{j\in N_i} X_j$. Also

$$(II) = \frac{1}{\sigma}\sum EX_i(\sum_{j\in N_i} X_j f(W)) = E\Bigg(f'(W)\underbrace{\left[\frac{1}{\sigma^2}\sum X_i(\sum_{j\in N_i} X_j)\right]}_{T}\Bigg).$$

We will proceed by showing that T is concentrated. Note that since $EX_iW_i = 0$,

$$\frac{1}{\sigma}E\sum X_i(W - W_i) = \frac{1}{\sigma}E\sum X_iW = EW^2 = 1$$

and so

$$|(II) - f'(W)| = |E(f'(W)(T-1)| \le |f'|_\infty E|T-1| \le \sqrt{\frac{2}{\pi}}\sqrt{E(T-1)^2} = \sqrt{\frac{2}{\pi}}\sqrt{\mathrm{Var}(T)}$$

Combining these results we have

$$|EWf(W) - Ef'(W)| \le \sqrt{\frac{2}{\pi}}\sqrt{\underbrace{\mathrm{Var}(\frac{1}{\sigma^2}\sum X_i(\sum_{j\in N_i} X_j))}_{III} + \underbrace{\frac{1}{\sigma^3}\sum E|X_i(\sum_{j\in N_i} X_j)^2|}_{IV}}.$$

Now

$$\begin{aligned}
IV &\le \frac{1}{\sigma^3}\sum_i \sum_{j,k\in N_i} E|X_iX_jX_k| \\
&\le \frac{1}{\sigma^3}\sum_i \sum_{j,k\in N_i} \frac{1}{3}(E|X_i|^3 + E|X_i|^3 + E|X_i|^3) \\
&\le \frac{D^2}{\sigma^3}\sum E|X_i|^3
\end{aligned}$$

where the second inequality follows from the AM-GM inequality.

Next we need to estimate

$$\mathrm{Var}(\sum_{i,j\in N_i} X_iX_j).$$

The collection $\{X_iX_j, i \in V, j \in N_i\}$ is a collection with a dependency graph of maximum degree $2D^2$. This can be seen as follows: X_iX_j is independent of X_kX_l if neither k nor l belongs to $N_i \cup N_j$. Now $|N_i \cup N_j| \le 2D$ and each vertex in this set has at most D neighbors so the maximum degree of the new dependency graph is $2D^2$. Using the variance bound on sums of dependency graph variables derived in the previous lecture we have

$$\mathrm{Var}(\sum_{i,j\in N_i} X_iX_j) \le 2D^2\sum_{i\sim j}\mathrm{Var}(X_iX_j) \le 2D^3\sum EX_i^4$$

using the fact that

$$\mathrm{Var}(X_iX_j) \le E(X_i^2X_j^2) \le \frac{1}{2}EX_i^4 + \frac{1}{2}EX_j^4.$$

The proof is completed by substituting this estimate. $\square$

Example 2 *Let $Y_1, \ldots, Y_{n+1}$ be iid mean 0 variance 1 random variables and let $X_i = Y_iY_{i+1}$ A dependency graph for the X_i has edge set $\{(i, i+1) : 1 \le i \le n\}$ and $D = 3$. Then $\mathrm{Var}(\sum X_i) = \sigma^2 = Cn$ so*

$$\mathrm{Wass}\left(\frac{1}{\sigma}\sum X_i,\, Z\right) \le C\frac{1}{\sigma^2}\sqrt{D^3\sum EX_i^4} + \frac{D^2}{\sigma^3}\sum E|X_i|^3 \le \frac{c}{\sqrt{n}}$$

Exercise 3 *In an Erdos-Renyi random graph $G(n,p)$ let $T_{n,p}$ be the number of triangles. Using the method of dependency graphs show that for some absolute constant C*

$$\mathrm{Wass}\left(\frac{T_{n,p} - ET_{n,p}}{\sqrt{\mathrm{Var}(T_{n,p})}},\, Z\right) \le \frac{C}{np^{9/2}}$$

Exercise 4

- *Find out the best known result for the above problem.*
- *Show that the CLT can not hold if $np \not\to \infty$.*
- *Refine the method of dependency graphs to show a CLT when $np \to \infty$.*