4.4 Lecture 23 (26/Nov): cont. Khintchine-Kahane for 4-wise independence; begin MIS in NC

4.4.1 Paley-Zygmund: solution through an in-probability bound

Paley-Zygmund is usually stated as an alternative (to Cor. [14]) lower-tail bound for nonnegative rvs; i.e., it gives a way to say that a nonnegative rv $A$ is “often large”.

Let $\mu_i$ be the $i$th moment of $A$. Knowing only the first moment $\mu_1$ of $A$ is not enough, because for any value—even infinite—of the first moment, we can arrange, for any $\delta > 0$, a nonnegative rv $A$ which equals 0 with probability $1 - \delta$, yet has first moment $\mu_1$. We just have to move $\delta$ of the probability mass out to the point $\mu_1/\delta$, or, in the infinite $\mu$ case, spread $\delta$ probability mass out in a measure whose first moment diverges.

However, a finite second moment $\mu_2$ is enough to justify such a “usually large” statement. Actually PZ can be stated for rvs which are not necessarily nonnegative, so we’ll do it that way.

**Lemma 74 (Paley-Zygmund)** Let $A$ be a real rv with positive $\mu_1$ and finite $\mu_2$. For any $0 < \lambda \leq 1$,

$$\Pr(A > (1 - \lambda)\mu_1) > \frac{\lambda^2 \mu_1^2}{\mu_2}.\)

**Proof:** Let $\nu$ be the distribution of $A$. Let $p = \Pr(A > (1 - \lambda)\mu_1)$. (This is what we want to lower bound.) Decompose $\mu_1 = \int_{[\lambda, 1]} x \, d\nu(x) + \int_{(1 - \lambda)\mu_1, \infty} x \, d\nu(x)$. Now examine each of these terms.

$$\int_{[\lambda, (1 - \lambda)\mu_1]} x \, d\nu(x) \leq (1 - p)(1 - \lambda)\mu_1 \leq (1 - \lambda)\mu_1 \quad (4.13)$$

Apply Cauchy-Schwarz to the functions $x^2$ and $\mathbf{1}_{x > (1 - \lambda)\mu_1}$. These are not effectively proportional to each other w.r.t. $\nu$ (unless $\nu$ is supported on a single point, in which case $\mu_2^2 = \mu_2$ and the Lemma is immediate), so we get a strict inequality,

$$\int_{(1 - \lambda)\mu_1, \infty} x \, d\nu(x) \leq \sqrt{p \int_{-\infty}^{\infty} x^2 \, d\nu(x)} = p^{1/2} \mu_2^{1/2} \quad (4.14)$$

Putting (4.13), (4.14) together, $\mu_1 < (1 - \lambda)\mu_1 + p^{1/2} \mu_2^{1/2}$

$$\lambda \mu_1 < p^{1/2} \mu_2^{1/2}$$

as desired. (There’s not normally much to be gained by preserving the “$1 - p$” factor in (4.13), but it’s at least another reason for writing strict inequality in the Lemma.)

---

**Lemma 75** For functions $f$ and $g$ that are square-integrable w.r.t. a measure $\nu$, $\int f(x)g(x) \, d\nu(x) \leq \sqrt{\int f^2(x) \, d\nu(x) \cdot \int g^2(x) \, d\nu(x)}$.

**Proof:** Squaring and subtracting sides, it suffices to show: $0 \leq \int \int f^2(x)g^2(y) \, d\nu(x)d\nu(y) - \int f(x)g(x)f(y)g(y) \, d\nu(x)d\nu(y)$. This is equivalent (by swapping the dummy variables) to showing $0 \leq \int (f^2(x)g^2(y) + f^2(y)g^2(x) - 2f(x)f(y)g(x)g(y)) \, d\nu(x)d\nu(y) = \int (f(x)g(y) - f(y)g(x))^2 \, d\nu(x)d\nu(y)$ which is an integral of squares.

Say that $f$ and $g$ are effectively proportional to each other w.r.t. $\nu$ if this last integral is 0; this is the condition for equality in Cauchy-Schwarz.
Comment: This gives \( \Pr(A \leq 0) \leq \frac{\mu_2 - \mu_1^2}{\mu_2} \) which improves on the upper bound \( \frac{\mu_2 - \mu_1^2}{\mu_1} \) of Cor. [14]. It should be said though that PZ does not dominate Cor. [14] in all ranges (e.g., if the variance \( \mu_2 - \mu_1^2 \) is very small compared to the \( \mu_1^2 \), and \( \lambda \) is small).

Returning to Gale-Berlekamp: Lemma [74] is not directly usable for our purpose, i.e., we cannot plug in the rv \( A = |X| \), because all it will tell us is that \( \mu_1 \geq (1 - \lambda) \lambda^2 \mu_1^3 / \mu_2 \), i.e., \( \mu_2 \geq (1 - \lambda) \lambda \mu_1^3 / \mu_2 \), which follows already from Cauchy-Schwarz (with the better constant 1). Note, this shows how Paley-Zygmund serves as a more flexible, albeit slightly weaker, version of Cauchy-Schwarz.

Instead, we set \( B = |X| \) and \( A = B^2 \), and then apply Paley-Zygmund to \( A \).

This is not a technicality. It means that we are relying on 4-wise independence, not just 2-wise independence, of the \( X_i \)'s. And indeed, Exercise: There are for arbitrarily large \( n \), collections of \( n \) pairwise independent \( X_i \)'s, uniform in \( \pm 1 \), s.t. \( \Pr(|X| = 0) = 1 - 1/n \), \( \Pr(|X| = n) = 1/n \).

**Corollary 76** Let \( B \) be a nonnegative rv with finite fourth moment \( \mu_4(B) \). Then \( E(B) \geq \frac{16}{25 \sqrt{3}} \frac{\mu_2^{5/2}}{\mu_4(B)} \).

**Proof:** For any \( \theta, E(B) \geq \theta \Pr(B \geq \theta) = \theta \Pr(B^2 \geq \theta^2) \), so, applying Lemma [74] to \( A = B^2 \), with \( \theta = \sqrt{\mu_2(B)/5} \) and \( \lambda = 4/5 \),

\[
E(B) \geq \sqrt{\frac{\mu_2(B)}{5}} \Pr(B^2 \geq \mu_2(B)/5) \geq \frac{(4/5)^2}{\sqrt{5}} \frac{\mu_2(B)^{5/2}}{\mu_4(B)}.
\]

\( \square \)

### 4.4.2 Berger: a direct expectation bound

**Lemma 77 (Berger [12])** Let \( B \) be a nonnegative rv with \( \mu_4(B) < \infty \). Then \( \mu_1(B) \geq \frac{\mu_2(B)^{3/2}}{\mu_4(B)^{1/2}} \).

This is stronger than Cor. [76] for two reasons: the constant, and perhaps more importantly, because \( \mu_2 / \mu_4^{1/2} \leq 1 \) (power mean inequality).

Of course, this lemma does not give an in-probability bound, so it is incomparable with Lemma [74].

**Proof:** We start with an elementary inequality.

**Lemma 78** \( y \geq \frac{3^{3/2}}{2} (y^2 - y^4) \) for all \( y > 0 \).

**Proof:** The quartic

\[
\frac{3^{3/2}}{2} (y^2 - y^4) + y
\]

(which we need to be \( \geq 0 \) for all \( y \geq 0 \)) has positive leading coefficient, a double root at \( 1/\sqrt{3} \), and simple roots at 0 and \(-2/\sqrt{3}\). See Fig. [4.1]

Now to obtain a lower bound on \( \mu_1(B) \) we use a trick analogous to that we used when obtaining Chernoff bounds: we introduce an adjustable parameter, obtain a generic bound, and then optimize over the parameter. In this case we use a parameter \( a > 0 \) and apply the previous lemma (taking expectations of everything in sight) to conclude that

\[
\mu_1(B/a) \geq \frac{3^{3/2}}{2} (\mu_2(B/a) - \mu_4(B/a))
\]

81
\[ \mu_1(B)/a \geq \frac{3^{3/2}}{2}(\mu_2(B)/a^2 - \mu_4(B)/a^4) \]

We optimize this with the choice \( a = \sqrt{\mu_4(B)/\mu_2(B)} \), yielding

\[ \mu_1(B) \geq \frac{\mu_2(B)^{3/2}}{\mu_4(B)^{1/2}} \]

Figure 4.1: Lower bound on absolute value from a quartic inequality

4.4.3 cont. proof of Theorem 73

Since Lemma 77 is stronger than Cor. 76, we apply the lemma. Substituting our known moments for the rv \(|X|\),

\[ E(|X|) \geq \frac{n^{3/2}}{(3n^2 - 2n)^{1/2}} \geq \sqrt{n/3}. \]

Observe that we have lost only a small constant factor here compared with the precise value obtained for a fully-independent sample space from the CLT.

4.4.4 Maximal Independent Set in NC

Parallel complexity classes

\[ L = \text{log-space} = \text{problems decidable by a Turing Machine having a read-only input tape and a read-write work tape of size (for inputs of length } n \text{) } O(\log n) \].

\[ NC = \bigcup_k NC^k \], where \( NC^k \) = languages s.t. \( \exists c < \infty \) s.t. membership can be computed, for inputs of size \( n \), by \( n^c \) processors running for time \( \log^k n \).

\[ RNC = \text{same, but the processors are also allowed to use random bits. For } x \in L \Pr(\text{ error }) \leq 1/2, \text{ for } x \notin L \Pr(\text{ error }) = 0. \]

\[ L \subseteq NC^1 \subseteq \ldots \subseteq NC \subseteq RNC \subseteq RP \].

P-Complete = problems that are in P, and that are complete for P w.r.t. reductions from a lower complexity class (usually, log-space).
Maximal Independent Set

MIS is the problem of finding a Maximal Independent Set. That is, an independent set that is not strictly contained in any other. This does not mean it needs to be a big, let alone a maximum cardinality set. (It is NP-complete to find an independent set of maximum size. This is more commonly known as the problem of finding a maximum clique, in the complement graph.)

There is an obvious sequential greedy algorithm for MIS: list the vertices \( \{1, \ldots, n\} \). Use vertex 1. Remove it and its neighbors. Use the smallest-index vertex which remains. Remove it and its neighbors, etc.

The independent set you get this way is called the Lexicographically First MIS. Finding it is P-complete w.r.t. L-reductions [24]. So it is interesting that if we don’t insist on getting this particular MIS, but are happy with any MIS, then we can solve the problem in parallel, specifically, in NC^2.

We’ll see an RNC, i.e., randomized parallel, algorithm of Luby [72] for MIS. Then, we’ll see how to derandomize the algorithm. (Some of the ideas we’ll see also come from the papers [61, 7]).

Notation: \( D_v \) is the neighborhood of \( v \), not including \( v \) itself. \( d_v = |D_v| \).

Luby’s MIS algorithm:

Given: a graph \( G = (V, E) \) with \( n \) vertices.

Start with \( I = \emptyset \).

Repeat until the graph is empty:

1. Mark each vertex \( v \) pairwise independently with probability \( \frac{1}{2d_v+1} \).
2. For each doubly-marked edge, unmark the vertex of lower degree (break ties arbitrarily).
3. For each marked vertex \( v \), append \( v \) to \( I \) and remove the vertices \( v \cup D_v \) (and of course all incident edges) from the graph.

An iteration can be implemented in parallel in time \( O(\log n) \), using a processor per edge.

We’ll show that an expected constant fraction of edges is removed in each iteration (and then we’ll show that this is enough to ensure expected logarithmically many iterations).

Definition 79 A vertex \( v \) is good if it has \( \geq d_v/3 \) neighbors of degree \( \leq d_v \). (Let \( G \) be the set of good vertices, and \( B \) the remaining ones which we call bad.) An edge is good if it contains a good vertex.

Lemma 80 If \( d_v > 0 \) and \( v \) is good, then \( \Pr(\exists \text{ marked } w \in D_v \text{ after step (1)} \) \( \geq \frac{1}{18} \).

This follows immediately from the following, using \( \sum_{w \in D_v} \Pr(w \text{ marked }) \geq \frac{d_v}{3} \frac{1}{2d_v+1} \geq \frac{1}{9} \).

Lemma 81 If \( \{X_i\} \) are pairwise independent events s.t. \( \Pr(X_i) = p_i \) then \( \Pr(\bigcup X_i) \geq \frac{1}{2} \min(\frac{1}{2}, \sum p_i) \).

Compare with the pairwise-independent version of the second Borel-Cantelli lemma. Of course, that is about guaranteeing that infinitely many events occur, here we’re just trying to get one to occur, but the lemmas are nonetheless quite analogous.

Proof: If \( \sum p_i < 1/2 \) then consider all events, otherwise there is a subset s.t. \( 1/2 \leq \sum p_i \leq 1 \) (consider two cases depending on whether any \( p_i \) exceeds 1/2); apply the following argument just to that subset.
\[ \Pr(\bigcup X_i) \geq \sum p_i - \sum_{i<j} \Pr(X_i \cap X_j) \quad \text{Bonferroni level 2} \]
\[ = \sum p_i - \sum_{i<j} p_i p_j \]
\[ \geq \sum p_i - \frac{1}{2} \sum_i p_i \sum_j p_j \]
\[ = (\sum p_i)(1 - \frac{1}{2} \sum p_i) \]
\[ \geq \frac{1}{2} \sum p_i. \]

So we can run the algorithm using a pairwise independent space, with the bits having various biases \( \frac{1}{2^{d/v}} \). It’s ok for the analysis if we round these probabilities by a factor of \( 1 \pm 0.1 \), so we can get these biased bits by sampling at each vertex an integer label between 1 and \( 10n \), and marking the vertex if its label is less than \( 5n/d_v \). Below we show that this space can be poly-size; we can therefore put the problem in MIS by implementing a separate batch of \(|E|\) processors per point in the sample space, running the mark-unmark process in each batch of processors, and then using the result from a sample point that deleted the most edges.