Lecture 4

1 The first and second moment methods

Let X be a non-negative random variable whose values are integers (sometimes, we refer to such random variables as *counting* random variables). Assume that $\mathbb{E}(X)$ tends to 0 as some parameter n tends to infinity. It then follows by Markov's inequality that

$$\mathbb{P}(X \geq 1) \leq \frac{\mathbb{E}(X)}{1} = \mathbb{E}(X) \ \Rightarrow \ \lim_{n \to \infty} \mathbb{P}(X = 0) = 1.$$

This useful observation is known as the first moment method.

Now, assume that as n tends to infinity, $\mathbb{E}(X)$ tends to 1 or even to infinity. Does that imply that $\lim_{n \to \infty} \mathbb{P}(X = 0) = 0$? As the following example shows, it does not. For every positive integer n let X_n be a random variable such that $\mathbb{P}(X_n = n^2) = 1/n$ and $\mathbb{P}(X_n = 0) = 1 - 1/n$. Then

$$\lim_{n \to \infty} \mathbb{E}(X_n) = \lim_{n \to \infty} n = \infty$$

but

$$\lim_{n \to \infty} \mathbb{P}(X_n = 0) = 1.$$

What else do we need to assume so that we can infer that $\lim_{n \to \infty} \mathbb{P}(X = 0) = 0$? Applying Chebyshev's inequality to X we have that

$$\mathbb{P}(X = 0) \leq \mathbb{P}(|X - \mathbb{E}(X)| \geq \mathbb{E}(X)) \leq \frac{Var(X)}{(\mathbb{E}(X))^2}.$$

Hence, it suffices to prove that $\frac{Var(X)}{(\mathbb{E}(X))^2}$ tends to 0 as n tends to infinity (we often denote this by $Var(X) = o((\mathbb{E}(X))^2)$). This useful tool is known as the second moment method.

1.1 The appearance of a triangle

Claim 1.1. *Construct a graph G with vertex set $[n]$ as follows. For every $1 \leq i < j \leq n$ flip a biased coin, where all coin flips are mutually independent. If the outcome of the coin is heads, which happens with probability p, connect i and j by an edge of G. Prove that*

(a) *If $p = o(1/n)$, then $\lim_{n \to \infty} \mathbb{P}(G \text{ contains a triangle}) = 0$.*

(b) *If $p = \omega(1/n)$, then $\lim_{n \to \infty} \mathbb{P}(G \text{ contains a triangle}) = 1$.*

Proof. Let $t = \binom{n}{3}$ and let $A_1, \ldots, A_t$ be an enumeration of all subsets of $[n]$ of size 3. For every $1 \leq i \leq t$, let X_i be the indicator random variable for the event "the vertices of A_i form a triangle in G". Let $X = \sum_{i=1}^{t} X_i$; then X counts the number of triangles in G. Clearly

$$\mathbb{E}(X_i) = \mathbb{P}(X_i = 1) = p^3$$

for every $1 \leq i \leq t$. It thus follows by the linearity of expectation that

$$\mathbb{E}(X) = \sum_{i=1}^{t} \mathbb{E}(X_i) = \binom{n}{3} p^3. \tag{1}$$

In particular, if $p = o(1/n)$, then $\lim_{n\to\infty} \mathbb{E}(X) = 0$. It thus follows by the first moment method that $\lim_{n\to\infty} \mathbb{P}(G \text{ contains a triangle}) = 0$; this proves (a).

Assume now that $p = \omega(1/n)$. It then follows from (1) that $\lim_{n\to\infty} \mathbb{E}(X) = \infty$. In order to prove (b) we will use the second moment method; thus we need to bound $Var(X)$ from above. We will use the formula for the variance of a sum of random variables, namely

$$Var(X) = \sum_{i=1}^{t} Var(X_i) + 2 \sum_{1 \leq i < j \leq t} Cov(X_i, X_j). \tag{2}$$

Note first that

$$Var(X_i) = \mathbb{E}(X_i^2) - (\mathbb{E}(X_i))^2 \leq \mathbb{E}(X_i) = p^3 \tag{3}$$

holds for every $1 \leq i \leq t$. Fix some $1 \leq i < j \leq t$ and let $\ell = |A_i \cap A_j|$. Observe that $\ell \leq 2$ as $i \neq j$. If $\ell \leq 1$, then X_i and X_j are determined by disjoint sets of coin flips and are thus independent; in particular $Cov(X_i, X_j) = 0$. Assume then that $\ell = 2$. We then have

$$Cov(X_i, X_j) = \mathbb{E}(X_i X_j) - \mathbb{E}(X_i)\mathbb{E}(X_j) = \mathbb{P}(X_i = 1, X_j = 1) - p^6 = p^5 - p^6 \leq p^5. \tag{4}$$

Combining (2), (3) and (4) implies that

$$Var(X) \leq n^3 p^3 + 2n^4 p^5 = o(n^6 p^6) = o((\mathbb{E}(X))^2),$$

where the first equality holds by our assumption that $p = \omega(1/n)$. By the second moment method, we conclude that $\lim_{n\to\infty} \mathbb{P}(G \text{ contains a triangle}) = 1$ as claimed. $\square$

1.2 Distinct sums

A set of positive integers $\{x_1, \ldots, x_k\}$ is said to have *distinct sums* if the 2^k sums $\sum_{i \in S} x_i : S \subseteq \{1, \ldots, k\}$ are all distinct. For a positive integer n, let $f(n)$ denote the largest integer k for which there exist integers $1 \leq x_1, \ldots, x_k \leq n$ such that $\{x_1, \ldots, x_k\}$ has distinct sums. An example of such a set is $\{2^i : 0 \leq i \leq \lfloor \log_2 n \rfloor\}$; this shows that $f(n) \geq 1 + \lfloor \log_2 n \rfloor$. We will prove that this lower bound is asymptotically tight.

Theorem 1.2. *$f(n) \leq \log_2 n + \frac{1}{2} \log_2 \log_2 n + O(1)$.*

Proof. Let $k = f(n)$ and let $1 \leq x_1, \ldots, x_k \leq n$ be integers for which $\{x_1, \ldots, x_k\}$ has distinct sums. Let $I_1, \ldots, I_k$ be mutually independent random variables such that $\mathbb{P}(I_i = 0) = \mathbb{P}(I_i = 1) = 1/2$ for every $1 \leq i \leq k$. Let $X = x_1 I_1 + \ldots + x_k I_k$. By the linearity of expectation we have

$$\mathbb{E}(X) = \mathbb{E}(x_1 I_1 + \ldots + x_k I_k) = \sum_{i=1}^{k} x_i \cdot \mathbb{E}(I_i) = \frac{x_1 + \ldots + x_k}{2}.$$

Since, moreover, $I_1, \ldots, I_k$ are mutually independent, it follows that

$$Var(X) = Var(x_1 I_1 + \ldots + x_k I_k) = \sum_{i=1}^{k} x_i^2 \cdot Var(I_i) = \frac{x_1^2 + \ldots + x_k^2}{4} \leq \frac{n^2 k}{4}. \tag{5}$$

Applying Chebyshev's inequality to X with $t = 2\sqrt{Var(X)}$ implies that

$$\mathbb{P}(|X - \mathbb{E}(X)| > t) \leq \frac{Var(X)}{t^2} = \frac{1}{4}$$

and thus

$$\mathbb{P}(|X - \mathbb{E}(X)| \leq t) \geq 1 - \frac{1}{4} = \frac{3}{4}. \tag{6}$$

On the other hand, since, by assumption, all 2^k sums $\sum_{i \in S} x_i : S \subseteq \{1, \ldots, k\}$ are distinct, the probability that $X - \mathbb{E}(X) = s$ for some arbitrary real number s is either 0 or 2^{-k}. In particular, $\mathbb{P}(X - \mathbb{E}(X) = s) \leq 2^{-k}$ holds for every $s \in [-t, t]$. Since there are at most $2t+1$ values $s \in [-t, t]$ for which $\mathbb{P}(X - \mathbb{E}(X) = s) > 0$, a union bound argument implies that

$$\mathbb{P}(|X - \mathbb{E}(X)| \leq t) \leq 2^{-k}(2t + 1) \leq 2^{-k}(2n\sqrt{k} + 1), \tag{7}$$

where the last inequality holds by (5) and the choice of t. Comparing (6) and (7) shows that

$$3/4 \leq 2^{-k}(2n\sqrt{k} + 1) \;\Rightarrow\; 3/4 \cdot 2^k \leq 2n\sqrt{k} + 1 \;\Rightarrow\; 2^k/\sqrt{k} \leq Cn,$$

where $C > 0$ is some constant. It follows that $f(n) = k \leq \log_2 n + \frac{1}{2}\log_2 \log_2 n + O(1)$ as claimed. $\square$