First and Second Moment Methods

1 First Moment Method

The *first moment method* refers to the application of the first moment (i.e. expectation) of a random variable in probabilistic reasoning. We have seen the argument from expectation, which is a type of first-moment method. This section discusses another first-moment method manifestation.

Markov inequality states that, if X is a non-negative random variable, then $\text{Prob}[X \geq a] \leq \text{E}[X]/a$ for all $a > 0$. In particular, $\text{Prob}[X \geq 1] \leq \text{E}[X]$. If X is a natural number then we can conclude that

$$\text{Prob}[X \neq 0] = \text{Prob}[X > 0] = \text{Prob}[X \geq 1] \leq \text{E}[X].$$

Thus, if $\text{E}[X] \to 0$ then $X = 0$ almost always. This simple idea can be used to prove many interesting results. In many applications of this idea X is a count of something, and thus it is certainly a natural number.

Problem 1. From the inequality $\text{Prob}[X \neq 0] \leq \text{E}[X]$ for natural number variable X, prove the union bound.

Proposition 1.1. *For any n-uniform hypergraph $\mathcal{H}$ with m edges, if $m < 2^{n-1}$ then the graph is 2-colorable.*

Proof. Color vertices of $\mathcal{H}$ with red/blue randomly with probability $1/2$. Let X be the number of monochromatic edges, then

$$\text{Prob}[\mathcal{H} \text{ is not properly colored}] = \text{Prob}[X > 0] \leq \text{E}[X] = \frac{m}{2^{n-1}} < 1.$$

Hence, with positive probability that coloring is valid and thus there always exists a valid 2-coloring if $m < 2^{n-1}$. □

The *chromatic number* $\chi(G)$ of a graph G is the minimum number of colors needed to color vertices of G so that adjacent vertices have different colors. An *independent set* of G is a subset of vertices no two of which are connected by an edge. Let $\alpha(G)$ denote the *maximum independent set* size of G. The following theorem is one of the first applications of the probabilistic method to graph theory [1]. Basically, Erdős showed that the chromatic number can not be inferred from "local" properties such as the smallest cycle size of the graph.

Theorem 1.2. *For any integer $k > 0$, there exists a triangle-free graph G whose chromatic number $\chi(G)$ is at least k.*

Proof. The proof strategy is as follows. I'm a little wordy here so that you understand the intuition behind the proof.

- Pick a random graph G from a graph distribution $\mathcal{G}(n,p)$, which denote the distribution of graphs on n vertices and each edge is present with probability p. This is also called the *Erdős-Rényi model*.
- Show that G has "small" maximum independent set with high probability. Small maximum independent set implies large chromatic number because each color class is an independent set. If each color class cannot be large then the number of colors has to be large. In particular, if G has n vertices then clearly $n \leq \chi(G) \cdot \alpha(G)$ and thus
$$\chi(G) \geq \frac{n}{\alpha(G)}.$$
- Show that G with high probability has few triangles. Remove one vertex from each triangle and we end up with G' which is triangle-free and still has "small" maximum independent set.

Specifically, pick G from $\mathcal{G}(n,p)$. The values of n and p shall be specified later.

First, we show that $\alpha(G) \leq \frac{n}{2k}$ with high probability using the first moment method. Let X be the number of independent sets of size $\frac{n}{2k}$ of G. Then,
$$\text{Prob}\left[\alpha(G) \geq \frac{n}{2k}\right] = \text{Prob}[X \neq 0] \leq \text{E}[X] = \binom{n}{n/2k}(1-p)^{\binom{n/2k}{2}} < 2^n \cdot e^{-\frac{pn(n-2k)}{8k^2}}.$$
The right hand side will be small as n gets large and p is not too small relative to n.

Next, let Y denote the number of triangles of G. Hence, if we remove one vertex from each triangle of G we end up with a graph G' with $n' \geq n - Y$ vertices. Note that $\alpha(G') \leq \alpha(G)$ because every independent set of G' is an independent set of G. Hence, if we are sure that $\alpha(G) \leq \frac{n}{2k}$ we have
$$\chi(G') \geq \frac{n'}{\alpha(G')} \geq \frac{n'}{\alpha(G)} \geq \frac{n-Y}{n/(2k)}.$$
In particular, if $n - Y \geq n/2$ then $\chi(G') \geq k$ and G' has no triangle as desired. In summary, we want *both* of the following two properties to hold with positive probability: $\alpha(G) \geq \frac{n}{2k}$ and $Y \leq n/2$.

Note that $\text{E}[Y] = \binom{n}{3}p^3 < \frac{(np)^3}{6}$. Hence, suppose we set $p = n^{-2/3}$ then $\text{E}[Y] < n/6$. Markov inequality implies
$$\text{Prob}[Y > n/2] \leq \frac{n/6}{n/2} = 1/3.$$
And, when $n > 4k$ we have
$$\text{Prob}\left[\alpha(G) \geq \frac{n}{2k}\right] < 2^n \cdot e^{-\frac{pn(n-2k)}{8k^2}} < e^n \cdot e^{-\frac{pn^2}{16k^2}} = e^n \cdot e^{-n^{4/3}/(16k^2)} = e^{n-n^{4/3}/(16k^2)}.$$
Hence, when $n^{1/3} \geq 32k^2$ we have
$$\text{Prob}\left[\alpha(G) \geq \frac{n}{2k}\right] < e^{-n} < 2/3.$$
Overall,
$$\text{Prob}\left[\alpha(G) \geq \frac{n}{2k} \text{ or } Y > n/2\right] < 1.$$
□

Problem 2. The *clique number* $\omega(G)$ of a graph G is the maximum clique size in G. Show that for all sufficiently large n there is a graph G with n vertices for which $\chi(G) \geq n/2$ and $\omega(G) \leq n^{3/4}$. (Hint: think about the complement of a triangle-free graph.)

Problem 3. The length of a smallest cycle in a graph is called the *girth* of the graph. If the graph has no cycle then its girth is defined to be infinity. Now, use the exact same method as in the triangle-free case to show that, for any integers $g, k > 0$, there is a graph with girth $\geq g$ and chromatic number $\geq k$.

2 Second moment method

The inequality $\text{Prob}[X > 0] \leq \text{E}[X]$ implies that, when $\text{E}[X] \to 0$ we have $X = 0$ almost always. However, $\text{E}[X] \to \infty$ does not necessarily mean that $X > 0$ almost always. We need more information, in particular the variance $\text{Var}[X]$. Making use of $\text{Var}[X]$ is called the *second moment* method. (The kth moment is $\text{E}[X^k]$.) The simplest use of $\text{Var}[X]$ is probably Chebyshev inequality which states that, for any $a > 0$,

$$\text{Prob}[|X - \text{E}[X]| \geq a] \leq \frac{\text{Var}[X]}{a^2}.$$

From this, we can infer

$$\text{Prob}[X = 0] \leq \text{Prob}[|X - \text{E}[X]| \geq \text{E}[X]] \leq \frac{\text{Var}[X]}{\text{E}[X]^2}.$$

In fact, Shepp proved a slightly stronger inequality by applying Cauchy-Schwarz inequality which states that $\text{E}[XY]^2 \leq \text{E}[X]^2\text{E}[Y]^2$:

$$\begin{aligned} \text{E}[X]^2 &= (\text{E}[\mathbf{1}_{X\neq 0} \cdot X])^2 \\ &\leq \text{E}\left[\mathbf{1}_{X\neq 0}^2\right] \text{E}[X^2] \\ &= \text{Prob}[X \neq 0]\text{E}[X^2] \\ &= \text{E}[X^2] - \text{Prob}[X = 0]\text{E}[X^2]. \end{aligned}$$

Consequently, we obtain

$$\text{Prob}[X = 0] \leq \frac{\text{Var}[X]}{\text{E}[X^2]}.$$

(Under the assumption that $\text{E}[X^2] \neq 0$.) Let us summarize.

Proposition 2.1. *Let X be a random variable for which $\text{E}[X^2] \neq 0$ then*

$$\text{Prob}[X = 0] \leq \frac{\text{Var}[X]}{\text{E}[X^2]}.$$

If $\text{E}[X] \neq 0$, then

$$\text{Prob}[X = 0] \leq \frac{\text{Var}[X]}{\text{E}[X]^2}.$$

The second inequality is weaker, but it might be more convenient in some cases.

2.1 Erdős distinct sum problem

A set $A = \{a_1, \cdots, a_k\}$ of positive integers has *distinct subset sums* if the sums of all subsets of A are distinct. Let $f(n)$ be maximum k for which there's a k-subset of $[n]$ having distinct subset sums. For example, from $A = \{2^i \mid 0 \le i \le \lg n\}$ we know

$$f(n) \ge \lfloor \lg n \rfloor + 1$$

Can $f(n)$ be much larger than $\lg n$? That was Erdős' question which he offered 500 usd to anyone who can show that

$$f(n) \le \lg n + c?$$

for some constant c. How about a lower bound? Information theoretically we can derive the following. Suppose there was a k-subset with distinct sums. Since each sum is at most nk we have $2^k \le nk$, which implies

$$k \le \lg n + \lg k \le \lg n + \lg(\lg n + \lg k) \le \lg n + \lg(2 \lg n) = \lg n + \lg\lg n + O(1).$$

We prove a slightly better bound using the second-moment method. Specifically we will use Chebyshev inequality. The intuition is as follows. Fix n and a k-subset $A = \{a_1, \cdots, a_k\}$ with distinct subset sums. Let X be the sum of a random subset of A where each element of A is included in X with probability $1/2$. Let $\mu = \mathrm{E}[X]$, and $\sigma^2 = \mathrm{Var}[X]$. For any integer i, it is clear that $\mathrm{Prob}[X = i]$ is either 0 or $\frac{1}{2^k}$. By Chebyshev, for any $\alpha > 1$ we have

$$\mathrm{Prob}[|X - \mu| \ge \alpha\sigma] \le \frac{1}{\alpha^2} \;\Rightarrow\; \mathrm{Prob}[|X - \mu| < \alpha\sigma] \ge 1 - \frac{1}{\alpha^2}$$

Because there are at most $2\alpha\sigma + 1$ integers within $\alpha\sigma$ of μ, we conclude that

$$1 - \frac{1}{\alpha^2} \le \frac{1}{2^k}(2\alpha\sigma + 1).$$

As σ is a function of n and k, the inequality gives us a relationship between n and k. Specifically,

$$\sigma^2 = \frac{a_1^2 + \cdots + a_k^2}{4} \le \frac{n^2 k}{4} \;\Rightarrow\; \sigma \le n\sqrt{k}/2$$

which leads to

$$1 - \frac{1}{\alpha^2} \le \frac{1}{2^k}(\alpha n\sqrt{k} + 1).$$

This is equivalent to

$$n \ge \frac{2^k\left(1 - \frac{1}{\alpha^2}\right) - 1}{\alpha\sqrt{k}}.$$

For the best possible bound on n, we set $\alpha > 1$ to maximize the right hand side. In particular, $\alpha = \sqrt{3}$ implies

$$n \ge \frac{2}{3\sqrt{3}} \cdot \frac{2^k}{k}.$$

From here it straightforward that

$$k \le \lg n + \frac{1}{2}\lg\lg n + O(1).$$

Problem 4. Let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be n two-dimensional vectors each of whose cooridnates is an integer with absolute value at most $\frac{2^{n/2}}{100\sqrt{n}}$. Show that there are two discint subsets $I, J \subseteq [n]$ such that $\sum_{i \in I} \mathbf{v}_i = \sum_{j \in J} \mathbf{v}_j$.

2.2 Graph properties and threshold functions

Each graph in the distribution $\mathcal{G}(n,p)$ has n vertices, and there is an edge between any pair of vertices with probability p. This is called the Erdős-Rényi model of random graphs. Often we would like to know if a random graph has some property. For example, if the (random) graph represents a P2P network we'd like to know whether it is connected (with such and such probability). If the graph represents some social network, we'd like to know how large the clustering coefficient is, and so on.

Obviously, for most non-trivial properties the probability the property holds depends on p. For instance, consider the property of whether the random graph has a clique of size at least 4. As $p \to 0$, it is highly unlikely that the property holds. As $p \to 1$, it is extremely likely that the property does hold. It turns out that there's a "threshold" in the middle crossing which the property switches from "likely not hold" to "likely hold."

Definition 2.2 (Threshold function for graph properties)**.** A function $t(n)$ is called a threshold function for some graph property $\mathcal{P}$ if $t(n)$ satisfies one of the following.

$$\lim_{n\to\infty} \operatorname*{Prob}_{G\in\mathcal{G}(n,p)}[G \text{ has property } \mathcal{P}] = \begin{cases} 0 & \text{if } p = o(t(n)) \\ 1 & \text{if } p = \omega(t(n)) \end{cases}$$

or

$$\lim_{n\to\infty} \operatorname*{Prob}_{G\in\mathcal{G}(n,p)}[G \text{ has property } \mathcal{P}] = \begin{cases} 1 & \text{if } p = o(t(n)) \\ 0 & \text{if } p = \omega(t(n)) \end{cases}$$

It might be a little confusing to grasp the concept. Let us consider a concrete example. The *clique number* $\omega(G)$ of a graph G is the size of its maximum clique. We consider the property $\omega(G) \geq 4$.

Suppose we draw G from $\mathcal{G}(n,p)$. Let X be the number of cliques of size 4. Then, $\omega(G) \geq 4$ if and only if $X > 0$. For each $S \in \binom{[n]}{4}$, let $X_S = \mathbf{1}_{S \text{ is a clique of } G}$. Then

$$X = \sum_S X_S.$$

Hence,

$$\mathrm{E}[X] = \binom{n}{4} p^6 < \frac{n^4 p^6}{24}.$$

Hence, by the first moment method, if $p = o(n^{-2/3})$ we have

$$\mathrm{Prob}[X > 0] \leq \mathrm{E}[X] < \frac{n^4 p^6}{24} = o(1).$$

In other words, the property $\omega(G) \geq 4$ likely does not hold when $p = o(n^{-2/3})$. What about the case when $p = \omega(n^{-2/3})$? In this case, $\mathrm{E}[X] = \Theta(n^4 p^6) \to \infty$. However, the first moment inequality $\mathrm{Prob}[X > 0] \leq \mathrm{E}[X]$ does not imply that $\mathrm{Prob}[X > 0] \to 1$. We need the second moment inequalities of Proposition 2.1: $\mathrm{Prob}[X = 0] \leq \sigma^2/\mu^2$. In particular, if $\sigma^2 = o(\mu^2)$ then it is likely that $X = 0$ does not hold and thus the property $\omega(G) \geq 4$ likely holds. To this end, we need to estimate the variance of X. The trouble with computing the variance of X is that the variables X_S are not independent.

We need a method for bounding the variance of a sum of dependent indicator variables. To be general, suppose $X = \sum_{i=1}^{n} X_i$. Then,

$$\begin{aligned}
\text{Var}[X] &= \text{E}\left[\left(\sum_i X_i\right)^2\right] - \left(\text{E}\left[\sum_i X_i\right]\right)^2 \\
&= \sum_i \text{E}[X_i^2] + \sum_{i\neq j} \text{E}[X_i X_j] - \sum_i \text{E}[X_i]^2 - \sum_{i\neq j} \text{E}[X_i]\text{E}[X_j] \\
&= \sum_{i=1}^{n} \text{Var}[X_i] + \sum_{i\neq j} \text{Cov}[X_i, X_j]
\end{aligned}$$

Where the *covariance* of any two variables X, Y is defined to be

$$\text{Cov}[X, Y] = \text{E}[XY] - \text{E}[X]\text{E}[Y].$$

If X_i is an indicator for event A_i and $\text{Prob}[X_i = 1] = p_i$, then

$$\text{Var}[X_i] = p_i(1 - p_i) \leq p_i = \text{E}[X_i].$$

If A_i and A_j are independent, then

$$\text{Cov}[X_i, X_j] = \text{E}[X_i X_j] - \text{E}[X_i]\text{E}[X_j] = 0.$$

If A_i and A_j are *not* independent (denoted by $i \sim j$)

$$\text{Cov}[X_i, X_j] \leq \text{E}[X_i X_j] = \text{Prob}[A_i \cap A_j].$$

Thus, we can rewrite the variance of X as

$$\text{Var}[X] \leq \text{E}[X] + \sum_i \sum_{j\neq i, j\sim i} \text{Prob}[A_i \cap A_j].$$

Noting that $\text{Prob}[A_i \cap A_j] = \text{Prob}[A_i]\,\text{Prob}[A_j \mid A_i]$, we have the following theorem.

Theorem 2.3. *Suppose $X = \sum_{i=1}^{n} X_i$, where X_i is an indicator for event A_i. Then,*

$$\text{Var}[X] \leq \text{E}[X] + \sum_i \text{Prob}[A_i] \underbrace{\sum_{j:j\sim i} \text{Prob}[A_j \mid A_i]}_{\Delta_i}$$

Corollary 2.4. *If $\Delta_i \leq \Delta$ for all i, then*

$$\text{Var}[X] \leq \text{E}[X](1 + \Delta)$$

Let's get back to the $\omega(G) \geq 4$ property. We need to estimate $\Delta_S = \sum_{T\sim S} \text{Prob}[A_T \mid A_S]$. The events A_T and A_S are dependent iff $T \cap S \geq 2$, because only then T and S share edges.

$$\begin{aligned}
\Delta_S &= \sum_{T\sim S} \text{Prob}[A_T \mid A_S] \\
&= \sum_{|T\cap S|=2} \text{Prob}[A_T \mid A_S] + \sum_{|T\cap S|=3} \text{Prob}[A_T \mid A_S] \\
&= \binom{n-4}{2}\binom{4}{2}p^5 + \binom{n-4}{1}\binom{4}{3}p^3 \\
&= \Delta.
\end{aligned}$$

So,

$$\sigma^2 \leq \mu(1+\Delta)$$

Recall we wanted $\sigma^2/\mu^2 = o(1)$, which holds as long as $\Delta = o(\mu)$. Better yet, when $p = \omega\left(n^{-2/3}\right)$, certainly

$$\Delta = \binom{n-4}{2}\binom{4}{2}p^5 + \binom{n-4}{1}\binom{4}{3}p^3 = o\left(n^4p^6\right).$$

Theorem 2.5. *$t(n) = n^{-2/3}$ is a threshold function for the $\omega(G) \geq 4$ property.*

Problem 5. (a) Derive the threshold function for the property $\omega(G) \geq 5$.

(b) Guess the general form of the threshold function for property $\omega(G) \geq k$ for a given integer k?

References

[1] P. ERDŐS, *Graph theory and probability*, Canad. J. Math., 11 (1959), pp. 34–38.