Probabilistic techniques - tutorials

Classwork 4 – The second moment

1. Let $\omega(G)$ denote the size of the largest clique in G. Prove that the threshold function for the property "$\omega(G(n,p)) \geq 4$" is $n^{-2/3}$.

   Solution: (Alon-Spencer 4.4.1) The $S \in \binom{[n]}{4}$ and A_S be the event S is a clique, X_S the respective indicator random variable and set $X = \sum_{S \in \binom{[n]}{4}} X_S$. $E[X] = \sum_{A \in \binom{[n]}{4}} E[X_S] = \binom{n}{4} p^6$, **Case** $p << n^{-2/3}$**:** we have that $E[X] < \frac{n^4}{24} n^{-4-\epsilon} \to 0$. Then we can apply Markov's inequality $Pr[X \geq 1] \leq E[X] \to 0$. **Case** $p >> n^{-2/3}$**:** In this case $E[X] > n^{\epsilon} \to \infty$. We have the inequality $\Pr[X = 0] \leq \frac{\mathrm{Var}[X]}{E[X]^2}$. $\mathrm{Var}[X] \leq E[X] + \sum_{S,T:\ |S \cap T|=2} Cov[X_S, X_T] + \sum_{S,T:\ |S \cap T|=3} Cov[X_S, X_T]$. On the one hand $Cov[X_S, X_T] = E[X_S X_T] - E[X_S]E[X_T] \leq E[X_S X_T]$. For S, T such that $|S \cap T| = 2$ we have that $E[X_S X_T] = p^{11}$ and if $|S \cap T| = 3$, then $E[X_S X_T] = p^9$. There are $\binom{n}{4}\binom{4}{2}\binom{n-4}{2}$ pairs S, T such that $|S \cap T| = 2$ and $\binom{n}{4}\binom{4}{3}\binom{n-4}{1}$ pairs S, T such that $|S \cap T| = 3$. So approximatey $\mathrm{Var}[X] \leq E[X] + n^6 p^{11} + n^5 p^9$. Finally,

$$\begin{aligned}
\frac{\mathrm{Var}[X]}{E[X]^2} &\leq \frac{1}{E[X]} + \frac{n^6 p^1 1}{n^8 p^{12}} + \frac{n^5 p^9}{n^8 p^1 2} \\
&= \frac{1}{E[X]} + \frac{1}{n^2 p} + \frac{1}{n^3 p^3} \\
&\leq \frac{1}{E[X]} + \frac{1}{n^{4/3}} + \frac{1}{n} \to 0.
\end{aligned}$$

2. Let $n \geq 3$ be a positive integer and $x_1, \ldots, x_n$ with $x_i^2 \leq \frac{1}{10000} \frac{2^n}{n}$ for each $i \in [n]$. Prove that there exist two non-empty disjoint subsets $I, J \subset [n]$ such that

$$\sum_{i \in I} x_i = \sum_{j \in J} x_j.$$

   Solution: Put i in I with probability $1/2$ and consider $X = \sum_{i \in I} x_i$. Then $E[X] = \frac{1}{2} \sum_{i=1}^{n} x_i$ and $\mathrm{Var}[X] = \sum_{i=1}^{n} \mathrm{Var}[X_i]$ because the random variables are independent. $\mathbb{E}[X_i] = x_i/2$ and $E[X_i^2] = x_i^2/2$, so $\mathrm{Var}[X_i] = E[X_i^2] - E[X_i]^2 = x_i^2/2 - x_i^2/4 = x_i^2/4$. Conseuqently $\mathrm{Var}[X] = \frac{1}{4} \sum_{i=1}^{n} x_i^2 \leq 2^{n-2}/10000$. Finally, $\Pr[|X - E[X]| \leq 10^2 \frac{2^{n/2-1}}{10^2}] \geq \Pr[|X - E[X]| \leq 10^2 \sigma] \geq 1 - \frac{1}{10^4}$. Then the number of choices of I that lands inside the interval $(E[X] - 2^{n/2-1}, E[X] + 2^{n/2-1})$ is at least $2^n(1 - \frac{1}{10^4}) \geq 2^{n-1}$. Because there are $2^{n/2}$ possible values in the inverval we get that by pigeonhole principle that there are two choices that will end up with the same value by X.

3. A set of positive integers $\{x_1, \ldots, x_n\}$ is said to have distinct sums if all sums

$$\sum_{i \in S} x_i, S \subseteq \{1, \ldots, n\}$$

   are distinct. Let $f(n)$ = maximal k such that there exist a set

$$\{x_1, \ldots, x_k\} \subseteq \{1, \ldots, n\}$$

   with distinct sums. Show that

$$\lfloor \log_2(n) \rfloor \leq f(n) \leq \log_2(n) + \frac{\log_2(\log_2 n)}{2} + \mathcal{O}(1).$$

Solution: For the lower bound consider the set $\{2^i : i \leq \lfloor \log_2(n) \rfloor\}$.. The rest is devoted to proving the upper bound. Fix $\{x_1, \dots, x_k\} \subseteq \{1, \dots, n\}$ with distinct sums. Let $\epsilon_1, \dots, \epsilon_n$ random variables $\{0, 1\}$ taking each value with probability $1/2$. Set $X = \sum_{i=1}^k \epsilon_i x_i$, then $\mu = E[X] = \frac{1}{2}\sum_{i=1}^k x_i$ and $\sigma^2 = \text{Var}[X] = \sum_{i=1}^k \text{Var}[\epsilon_i x_i] = \frac{1}{4}\sum_{i=1}^k x_i^2 \leq \frac{n^2 k}{4}$, i.e. $\sigma \leq \frac{n}{2}\sqrt{k}$. By Chebyshev's inequality we have that $Pr[|X - \mu| \geq \lambda\sigma] \leq \lambda^{-2}$, then $Pr[|X - \mu| < \lambda\sigma] > 1 - \lambda^{-2}$. Notice that $Pr[X = a] \leq \frac{1}{2^k}$ for any $a \in \mathbb{N}$ (using that the sums are different). Then $Pr[|X - \mu| < \lambda\sigma] \leq 2^{-k}(\lambda\sigma + 1)$. So $n \geq \frac{2^{k+1}}{8\sqrt{k}}$, take logarithm and re-arrange.

4. Show that there is a positive constant c such that the following holds: For any n real numbers $a_1, \dots, a_n$ satifying $\sum_{i=1}^n a_i^2 = 1$, if $(\epsilon_1, \dots, \epsilon_n)$ is a $\{-1, +1\}$-random vector obtained by choosing each ϵ_i randomly and independently with uniform distribution to be either -1 or $+1$, then

$$\Pr\left[\left|\sum_{i=1}^n \epsilon_i a_i\right| \leq 1\right] \geq c.$$

Solution: To lower bound $Pr[|\sum_{i=1}^n \epsilon_i a_i| \leq 1]$ we will upper bound $Pr[|\sum_{i=1}^n \epsilon_i a_i| \geq 1]$ using Chebyshev. We have $\text{Var}[\epsilon_i a_i] = a_i^2$. Without loss of generality $a_1 \geq \dots \geq a_n \geq 0$. Distinguish two cases: **Case** $a_1 \geq 1/2$**:** First, we have that $\sum_{i=2}^n a_i^2 = 1 - a_1^2$, so

$$\Pr\left[\left|\sum_{i=2}^n \epsilon_i a_i\right| \geq 1 + a_1\right] \leq \frac{1 - a_1^2}{(1 + a_1)^2} = \frac{1 - a_1}{1 + a_1} \leq 1/3.$$

Then, $\Pr[|\sum_{i=2} \epsilon_i a_i| \leq 1 + a_1] \geq 2/3$. $Pr[-(1 + a_1) \leq \sum_{i=2}^n \epsilon_i a_i \leq 1 + a_1] = 2/3$ and since the number of ϵ_i that lands on the left and on the right is symmetric we have that $Pr[-(1 + a_1) \leq \sum_{i=2}^n \epsilon_i a_i \leq 0] \geq 1/3$, i.e. $Pr[-1 \leq \sum_{i=2} \epsilon_i a_i + a_1 \leq a_1] \geq 1/3$. Putting all together we have

$$\Pr[|\sum_{i=1}^n \epsilon_i a_i| \leq 1] \geq Pr[|\sum_{i=1}^n \epsilon_i a_i| \leq 1 | \epsilon_1 = 1] Pr[\epsilon_1 = 1] \geq Pr[-1 \leq \sum_{i=1}^n \epsilon_i a_i \leq a_1 | \epsilon_1 = 1] 1/2 \geq 1/6.$$

Case $a_1 \leq 1/2$**:** Split the a_i's into two groups b_i's and c_i's such that the sum of the squares is as similar as possible. Note that they differ at most by $1/4$ as $a_i^2 \leq 1/2$. Thus, so they lie in $[3/8, 5/8]$. There is at least $1/2$ probability that the sums $\sum \epsilon_i c_i$ and $\sum \epsilon_i b_i$ have different signs because for each choice of the first ones we can count how many there are of the other and for each wrong one just take the opposite signs. Finally

$$\begin{aligned}
\Pr[|\sum_{i=1}^n a_i| \leq 1] &= \Pr[|\sum \epsilon_i b_i + \sum \epsilon_i c_i| \leq 1] \geq \Pr[|\sum \epsilon_i b_i + \sum \epsilon_i c_i| \leq 1 \mid \text{ different signs}]/2 \\
&\geq \Pr[|\sum \epsilon_i b_i| \leq 1, |\sum \epsilon_i c_i| \leq 1 \mid \text{ different signs}]/2 \\
&\geq (1 - \sum b_i^2)(1 - \sum c_i^2)/2 \geq \frac{3}{8}\frac{3}{8}\frac{1}{2} = \frac{9}{128} = c,
\end{aligned}$$

where $\Pr[|\sum \epsilon_i b_i| \leq 1] \geq 1 - \sum b_i^2$ is by Chebyshev.