Probabilistic Methods in Combinatorics

Instructor: Oliver Janzer

Solutions to Assignment 6

Problem 1. Let $0 < p_1, p_2, \dots, p_n \leq 1$ be reals. For every $i \in [n]$, let X_i be a Bernoulli random variable of parameter p_i, such that $X_1, \dots, X_n$ are independent. Show that

$$\mathbb{P}\left(\left|\sum_{i=1}^n X_i - \sum_{i=1}^n p_i\right| \geq 10\sqrt{\sum_{i=1}^n p_i}\right) \leq 1/100.$$

Solution. As the random variables $X_1, \dots, X_n$ are independent, we have

$$\mathrm{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathrm{Var}\left(X_i\right) = \sum_{i=1}^n p_i(1-p_i) \leq \sum_{i=1}^n p_i.$$

Since $\mathbb{E}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n p_i$, it follows by Chebyshev's inequality that

$$\mathbb{P}\left(\left|\sum_{i=1}^n X_i - \sum_{i=1}^n p_i\right| \geq 10\sqrt{\sum_{i=1}^n p_i}\right) \leq \frac{\mathrm{Var}\left(\sum_{i=1}^n X_i\right)}{100\sum_{i=1}^n p_i} \leq \frac{\sum_{i=1}^n p_i}{100(\sum_{i=1}^n p_i)} \leq \frac{1}{100}.$$

Problem 2. Let X be a random variable taking nonnegative integer values. In the lectures we have seen that $\mathbb{P}(X=0) \leq \frac{\mathrm{Var}(X)}{\mathbb{E}[X]^2}$. Prove that in fact

$$\mathbb{P}(X=0) \leq \frac{\mathrm{Var}(X)}{\mathbb{E}[X^2]}.$$

Solution. Let $p = \Pr[X=0]$. We have $\mathbb{E}[X] = (1-p)\mathbb{E}[X \mid X>0]$ and $\mathbb{E}[X^2] = (1-p)\mathbb{E}[X^2 \mid X>0]$. Using $\mathbb{E}[X^2 \mid X>0] \geq \mathbb{E}[X \mid X>0]^2$, we get $\frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]} \leq 1-p$. Hence,

$$\Pr[X=0] = p \leq 1 - \frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]} = \frac{\mathrm{Var}[X]}{\mathbb{E}[X^2]}.$$

Problem 3. Let $G=(V,E)$ be a simple graph with n vertices and m edges, and let k be a positive integer. Prove that:

(a) There are at least $k^n \cdot (1 - \frac{m}{k})$ proper vertex-colourings of G with k colours.

(b) There are at most $k^n \cdot \frac{k-1}{m}$ proper vertex-colourings of G with k colours.

(c) The upper bound from (b) can be improved to $k^n \cdot \frac{k-1}{k+m-1}$.

Solution. Colour each vertex independently at random using one of the k colours. Let X be the number of edges whose two endpoints have the same colour. Note that the colouring is proper if and only if $X = 0$.

(a) For any fixed edge, this probability that both endpoints get the same colour is $1/k$. Since there are m edges in the graph, we have $\mathbb{E}[X] = m/k$. By Markov's inequality, $P(X \geq 1) \leq \mathbb{E}[X]$, so with probability at least $1 - m/k$, we have $X = 0$, i.e. the colouring is proper. Since there are k^n colourings in total, the proof is complete.

(b) Here we use the inequality $\mathbb{P}(X = 0) \leq \frac{\mathrm{Var}(X)}{\mathbb{E}[X]^2}$. Let us compute $\mathrm{Var}(X)$. Note that $X = \sum_{e \in E(G)} X_e$, where X_e is 1 if the endpoints of e get the same colour and 0 otherwise. Now $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \sum_{e,f \in E(G)} (\mathbb{E}[X_e X_f] - \mathbb{E}[X_e]\mathbb{E}[X_f])$. For any $e \in E(G)$, $\mathbb{E}[X_e]$ is the probability that the endpoints of e have the same colour, which is $1/k$. Also, for $e, f \in E(G)$, $\mathbb{E}[X_e X_f]$ is the probability that the endpoints of e have the same colour and the endpoints of f have the same colour, which is $1/k^2$ when $e \neq f$ (and $1/k$ when $e = f$). Hence, $\mathrm{Var}(X) = \sum_{e \in E(G)} (1/k - 1/k^2) = m \cdot \frac{k-1}{k^2}$. Therefore $\mathbb{P}(X = 0) \leq \frac{\mathrm{Var}(X)}{\mathbb{E}[X]^2} = \frac{m(k-1)/k^2}{m^2/k^2} = \frac{k-1}{m}$.

(c) In this part we use the stronger bound $\mathbb{P}(X = 0) \leq \frac{\mathrm{Var}(X)}{\mathbb{E}[X^2]}$. Note that $\mathbb{E}[X^2] = \sum_{e,f \in E(G)} \mathbb{E}[X_e X_f] = m \cdot \frac{1}{k} + m(m-1) \cdot \frac{1}{k^2}$, so

$$\mathbb{P}(X = 0) \leq \frac{m(k-1)/k^2}{m/k + m(m-1)/k^2} = \frac{k-1}{k+m-1}.$$

Problem 4. Let $v_1 = (x_1, y_1), \ldots, v_n = (x_n, y_n)$ be n two-dimensional vectors, where each x_i and each y_i is an integer whose absolute value does not exceed $2^{n/2}/(100\sqrt{n})$. Show that there are two disjoint sets $I, J \subseteq \{1, 2, \ldots, n\}$ such that

$$\sum_{i \in I} v_i = \sum_{j \in J} v_j.$$

Solution. Let $t = \lfloor 2^{n/2}/(100\sqrt{n}) \rfloor$. Let $\epsilon_1, \ldots, \epsilon_n$ be independent random variables taking values 0 and 1 with probability 1/2. Let $X = \sum_{i \leq n} \epsilon_i x_i$. It was shown in Section 4.1 in lectures (in the proof of Theorem 4.3) that

$$\mathbb{P}\left(|X - \mathbb{E}[X]| \geq t\sqrt{n}\right) \leq \frac{1}{4}.$$

Similarly, if $Y = \sum_{i \leq n} \epsilon_i y_i$, then

$$\mathbb{P}\left(|Y - \mathbb{E}[Y]| \geq t\sqrt{n}\right) \leq \frac{1}{4}.$$

Thus, with probability at least 1/2, we have both $|X - \mathbb{E}[X]| < t\sqrt{n}$ and $|Y - \mathbb{E}[Y]| < t\sqrt{n}$. This means that for at least 2^{n-1} choices $(\epsilon_1, \dots, \epsilon_n) \in \{0,1\}^n$, we have

$$\sum_{i \leq n} \epsilon_i x_i \in \left[\mathbb{E}[X] - t\sqrt{n}, \mathbb{E}[X] + t\sqrt{n}\right]$$

and

$$\sum_{i \leq n} \epsilon_i y_i \in \left[\mathbb{E}[Y] - t\sqrt{n}, \mathbb{E}[Y] + t\sqrt{n}\right].$$

Observe that $(2t\sqrt{n}+1)^2 \leq (2^{n/2}/50+1)^2 < 2^{n-1}$ for $n \geq 2$. But both $[\mathbb{E}[X] - t\sqrt{n}, \mathbb{E}[X] + t\sqrt{n}]$ and $[\mathbb{E}[Y] - t\sqrt{n}, \mathbb{E}[Y] + t\sqrt{n}]$ contain at most $2t\sqrt{n} + 1$ integers, so there must exist distinct subsets $I, J \subset [n]$ such that $\sum_{i \in I} x_i = \sum_{j \in J} x_j$ and $\sum_{i \in I} y_i = \sum_{j \in J} y_j$. That is, $\sum_{i \in I} v_i = \sum_{j \in J} v_j$. By removing $I \cap J$ from both I and J, we can make sure that there are such subsets with I and J disjoint.

Remark. This is a two-dimensional analogue of sum-free set problem in the lecture notes: instead of $a_1, a_2, \dots \in [n]$ (in section 4.1), we have $v_1, v_2, \dots \in \{0, \dots, B\}^2$ here. What we proved is that the largest sum-free set in $\{0, \dots, B\}^2$ has size at most $2\log_2 B + \log_2 \log_2 B + O(1)$. In addition, the largest size is at least $2\log_2 B$ because of the following construction. Consider $(2^i, 0), (2^i, 2^i)$ for all $i \in \{0, 1, \dots, \lfloor \log_2 B \rfloor\}$. If $\sum_{i \in I_0}(2^i, 0) + \sum_{i \in I_1}(2^i, 2^i) = \sum_{i \in I_0'}(2^i, 0) + \sum_{i \in I_1'}(2^i, 2^i)$, then we know that $\sum_{i \in I_1} 2^i = \sum_{i \in I_1'} 2^i$, thereby $I_1 = I_1'$. Then, a similar argument shows $I_0 = I_0'$. This means these $2\log_2 B$ vectors form a sum-free set in $\{0, \dots, B\}^2$.