Probability Theory 2 – Solutions III

1. Let $G \sim G(n,p)$, where $p = 10 \ln n / n$. Let δ denote the minimum degree in G and let Δ denote the maximum degree in G. Prove the following two claims:

   (a) $\lim_{n\to\infty} \mathbb{P}(\delta \geq \ln n) = 1$.

   (b) $\lim_{n\to\infty} \mathbb{P}(\Delta \leq 20 \ln n) = 1$.

   Solution: For every $1 \leq i \leq n$, let d_i denote the degree of vertex i in G. Note that $d_i \sim Bin(n-1, p)$. In particular, $\mathbb{E}(d_i) = (n-1)p$ and thus, for sufficiently large n we have $9 \ln n \leq \mathbb{E}(d_i) \leq 10 \ln n$.

   (a) Using Chernoff's inequality, for every $1 \leq i \leq n$ we obtain

   $$\mathbb{P}(d_i < \ln n) \leq \mathbb{P}(d_i \leq \mathbb{E}(d_i)/2) \leq e^{-(1/2)^2 \cdot \mathbb{E}(d_i)/2} \leq e^{-9 \ln n/8} = o(1/n).$$

   Therefore, a union bound argument implies that

   $$\mathbb{P}(\delta < \ln n) \leq \sum_{i=1}^{n} \mathbb{P}(d_i < \ln n) = o(1).$$

   (b) Using Chernoff's inequality, for every $1 \leq i \leq n$ we obtain

   $$\mathbb{P}(d_i > 20 \ln n) \leq \mathbb{P}(d_i \geq 2\mathbb{E}(d_i)) \leq e^{-1^2 \cdot \mathbb{E}(d_i)/3} \leq e^{-9 \ln n/3} = o(1/n).$$

   Therefore, a union bound argument implies that

   $$\mathbb{P}(\Delta > 20 \ln n) \leq \sum_{i=1}^{n} \mathbb{P}(d_i > 20 \ln n) = o(1).$$

2. Prove that $n^{-2/3}$ is a threshold for the property of containing a copy of K_4 (the complete graph on 4 vertices). That is, prove the following two claims:

   (a) If $p \ll n^{-2/3}$, then $\lim_{n\to\infty} \mathbb{P}(G(n,p) \text{ contains a copy of } K_4) = 0$.

   (b) If $p \gg n^{-2/3}$, then $\lim_{n\to\infty} \mathbb{P}(G(n,p) \text{ contains a copy of } K_4) = 1$.

   Solution: Let $G \sim G(n,p)$, let $r = \binom{n}{4}$ and let $A_1, \ldots, A_r$ denote all the subsets of $V(G)$ of size 4. For every $1 \leq i \leq r$, let $X_i = 1$ if the vertices of A_i induce a K_4 in G and $X_i = 0$ otherwise. For every $1 \leq i \leq r$ we have

   $$\mathbb{E}(X_i) = \mathbb{P}(X_i = 1) = p^6.$$

   Let $X = \sum_{i=1}^{r} X_i$. Then X counts the number of K_4's in G and, by the linearity of expectation, we have

   $$\mathbb{E}(X) = \sum_{i=1}^{r} \mathbb{E}(X_i) = \binom{n}{4} p^6.$$

Now, if $p \ll n^{-2/3}$, then $\mathbb{E}(X) = o\left(n^4(n^{-2/3})^6\right) = o(1)$. It follows by Markov's inequality that

$$\mathbb{P}(G \text{ contains a copy of } K_4) = \mathbb{P}(X \geq 1) \leq \mathbb{E}(X) = o(1).$$

Next, assume that $p \gg n^{-2/3}$; by monotonicity we can assume that $p \leq n^{-2/3} \log n$. In this case $\mathbb{E}(X) = \omega\left(n^4(n^{-2/3})^6\right) = \omega(1)$. We wish to apply Chebyshev's inequality and thus begin by bounding the variance of X from above. Recall that

$$Var(X) = Var(X_1 + \ldots + X_r) = \sum_{i=1}^{r} Var(X_i) + 2 \sum_{1 \leq i < j \leq r} Cov(X_i, X_j). \tag{1}$$

For every $1 \leq i \leq r$ we have $X_i^2 = X_i$ and thus

$$Var(X_i) = \mathbb{E}(X_i^2) - (\mathbb{E}(X_i))^2 = p^6 - p^{12} \leq p^6. \tag{2}$$

Fix some $1 \leq i < j \leq r$. We distinguish between three cases:

(a) $|A_i \cap A_j| \leq 1$. In this case there are no potential edges in $A_i \cap A_j$ and thus X_i and X_j are independent. In particular, $Cov(X_i, X_j) = 0$.

(b) $|A_i \cap A_j| = 2$. In this case A_i and A_j share one potential edge. Therefore

$$Cov(X_i, X_j) = \mathbb{E}(X_i X_j) - \mathbb{E}(X_i)\mathbb{E}(X_j) = \mathbb{P}(X_i = 1, X_j = 1) - p^{12} = p^{11} - p^{12} \leq p^{11}.$$

Note that in this case $|A_i \cup A_j| = 6$. A straightforward calculation then shows that the number of such pairs is at most n^6.

(c) $|A_i \cap A_j| = 3$. In this case A_i and A_j share three potential edges (forming a triangle). Therefore

$$Cov(X_i, X_j) = \mathbb{E}(X_i X_j) - \mathbb{E}(X_i)\mathbb{E}(X_j) = \mathbb{P}(X_i = 1, X_j = 1) - p^{12} = p^9 - p^{12} \leq p^9.$$

Note that in this case $|A_i \cup A_j| = 5$. A straightforward calculation then shows that the number of such pairs is at most n^5.

Combining (1), (2), and the three cases above, we conclude that

$$\begin{aligned} Var(X) &\leq \binom{n}{4} p^6 + 2n^6 p^{11} + 2n^5 p^9 = \mathbb{E}(X)(1 + O(n^2 p^5 + np^3)) \\ &= \mathbb{E}(X)(1 + O(n^2(n^{-2/3}\log n)^5 + n(n^{-2/3}\log n)^3)) = (1 + o(1))\mathbb{E}(X). \end{aligned}$$

Using Chebyshev's inequality, we conclude that

$$\begin{aligned} \mathbb{P}(G \text{ does not contain a copy of } K_4) &= \mathbb{P}(X = 0) \leq \mathbb{P}(|X - \mathbb{E}(X)| \geq \mathbb{E}(X)) \\ &\leq \frac{Var(X)}{(\mathbb{E}(X))^2} \leq \frac{(1 + o(1))\mathbb{E}(X)}{(\mathbb{E}(X))^2} = o(1), \end{aligned}$$

where the last equality holds since $\mathbb{E}(X) = \omega(1)$.

3. Let $n \geq k \geq 2$ and $1 \leq m \leq 2^{k-2}$ be integers. Let $\{A_1, \ldots, A_m\}$ be a family of subsets of $\{1, \ldots, n\}$, each of size k. Devise a randomized algorithm with the following properties:

   (a) Its input is the number n and the family $\{A_1, \ldots, A_m\}$.

   (b) Its output is a red/blue colouring of the elements of the set $\{1, \ldots, n\}$.

   (c) With probability at least $1 - 2^{-100}$, the colouring produced by the algorithm is such that, for every $1 \leq i \leq m$, the set A_i contains at least one red element and at least one blue element.

   (d) The running time of the algorithm is $O(n + km)$.

Solution: We will present a randomized algorithm and then prove that it meets all the requirements of the question.

Algorithm:

(i) For $t = 1$ to 100 do:

(ii) For every $1 \leq i \leq n$ colour i red with probability 1/2 and blue with probability 1/2, where all of these decisions are mutually independent.

(iii) Check whether the resulting colouring is *good*, i.e., for every $1 \leq j \leq m$ check whether A_j contains at least one red element and at least one blue element. If the colouring is good, output it and terminate.

(iv) If the algorithm did not terminate after running for 100 iterations, then output the current colouring and terminate.

We start by analyzing the running time of the algorithm. It consists of at most 100 rounds. In every round, each integer between 1 and n flips a coin once to determine its colour – this takes time $O(n)$. Once a colouring is produced, we check whether this colouring is good, by checking each of the given m sets to see if it contains a red element and a blue element. This takes time $O(km)$. Altogether, the running time of every round and of the entire algorithm is $O(n + km)$ as required.

Finally, we wish to show that the algorithm fails to produce a colouring in which every given k-set contains both colours with probability at most 2^{-100}. Consider any one of the 100 iterations of the algorithm. Since all coin flips are independent, for every $1 \leq j \leq m$, the probability that all elements of A_j have the same colour is $2 \cdot 2^{-k} = 2^{1-k}$. Applying a union bound over all given sets then yields

$$\mathbb{P}(\exists 1 \leq j \leq m \text{ such that all elements of } A_j \text{ have the same colour}) \leq m2^{1-k} \leq 1/2, \quad (3)$$

where the last inequality follows from our assumption that $m \leq 2^{k-2}$. If the algorithm fails to produce a good colouring, then none of the 100 random colourings it produced were good. By independence and by (3), we conclude that

$$\mathbb{P}(\text{the algorithm fails to produce a good colouring}) \leq (1/2)^{100}.$$

4. Let $\mathcal{F} \subseteq \mathbb{Z}_2^n$ be a family of binary vectors of length n. For any two binary vectors $\bar{x} = (x_1, \ldots, x_n)$ and $\bar{y} = (y_1, \ldots, y_n)$, define the distance between them to be

$$dist(\bar{x}, \bar{y}) = |\{1 \leq i \leq n : x_i \neq y_i\}|.$$

Assume that the distance between any two vectors in $\mathcal{F}$ is at least $n/10$. Devise a randomized algorithm with the following properties:

(a) Its input are vectors $\bar{x}, \bar{y} \in \mathcal{F}$.

(b) Its output is either $\bar{x} = \bar{y}$ or $\bar{x} \neq \bar{y}$.

(c) If $\bar{x} = \bar{y}$, then the algorithm will output $\bar{x} = \bar{y}$.

(d) If $\bar{x} \neq \bar{y}$, then the algorithm will output $\bar{x} \neq \bar{y}$ with probability at least $1 - 2^{-100}$.

(e) The running time of the algorithm is constant (i.e., independent of n).

Solution: We will present a randomized algorithm and then prove that it meets all the requirements of the question.

Algorithm:

(i) For 1000 times, draw an element of $\{1, \ldots, n\}$ independently and uniformly at random (with replacement); denote the set of all drawn numbers by J.

(ii) For every $i \in J$, compare x_i and y_i.

(iii) If $x_i = y_i$ for every $i \in J$, then output $\bar{x} = \bar{y}$. Otherwise, output $\bar{x} \neq \bar{y}$.

First, it is evident that the running time of the algorithm is constant as we make 1000 random element draws and then compare 1000 pairs of bits.

Next, we prove the (randomized) correctness of the algorithm. If $\bar{x} = \bar{y}$, then for any random subset $J \subseteq \{1, \ldots, n\}$, we will always have $x_i = y_i$ for every $i \in J$. Hence, in this case, with probability 1 the algorithm will output $\bar{x} = \bar{y}$. Assume then that $\bar{x} \neq \bar{y}$. Let $D = \{1 \leq i \leq n : x_i \neq y_i\}$; by assumption $|D| \geq n/10$. Hence, for each of the 1000 random number draws, the probability of this number being in $\{1, \ldots, n\} \setminus D$ is at most $9/10$. It follows by independence that

$$\mathbb{P}(J \subseteq \{1, \ldots, n\} \setminus D) \leq (9/10)^{1000} < 2^{-100}.$$

By the description of the algorithm we conclude that

$$\mathbb{P}(\text{the algorithm outputs } \bar{x} = \bar{y}) = \mathbb{P}(J \subseteq \{1, \ldots, n\} \setminus D) < 2^{-100}.$$