THE PROBABILISTIC METHOD

Exercise 1

Consider two disjoint events with non-null probability. Can they be independent?

Exercise 2

Give an example of a probability space with the events $A_1$, $A_2$ and $A_3$ that are two-by-two independent, but not mutually independent.

Exercise 3

What are the expectation and variance of the number of fixed points of a uniformly distributed permutation on $\{1, \ldots, n\}$?

Exercise 4

Let $G = (V, E)$ be a finite non-directed graph. An independent set $I$ is a subset of vertices no two elements of $I$ are adjacent:

$I$ independent set $\Rightarrow \forall u, v \in I, (u, v) \notin E.$

Set $\alpha(G)$ the maximal size of an independent set of $G$, and $d(v)$ the degree of vertex $v$.

We build a graph $G$ at random the following way: vertices are added one by one in a uniformly distributed order.

1. What is the probability that a vertex $v$ is added before all its neighbors?
2. Deduce that $\alpha(G) \geq \sum_{v \in V} \frac{1}{d(v)+1}.$

Exercise 5

Let is first recall a useful formula: $(1 - p) \leq e^{-p}$ for all $p \in [0, 1]$.

A dominating set $D$ is a set of vertices such that every other vertex is adjacent to a vertex of $D$:

$D$ dominating set $\Rightarrow \forall v \notin D, \exists u \in D$ such that $(u, v) \in E.$

1. Show that if $G$ is $k$-regular (all its vertices have degree exactly $k$), then for all dominating set, $D$, $|D| \geq \frac{|V|}{k+1}$.

We now assume that all the vertices of $G$ have degree at least $k$, and we build a dominating set the following way:

1. Select each vertex of $V$ in a set $S$ with probability $p = \frac{\ln(k+1)}{k+1}$;
2. $T$ is the subset of the vertices in $V \setminus S$ that have no neighbors in $S$;
3. $D = S \cup T$.

2. Show that $D$ is a dominating set.
3. What is the probability that a vertex \( v \) belongs to \( T \) ?

4. Deduce that there exists a dominating set of size at most \(|V|^{1 + \ln(k+1)} / k^{1+1} \).

**Exercice 6**

**Distinct sums**

A set of integers \( A \) is called a *distinct-sum set* if the sum \( \sum_{a \in S} a \) is different for each subset \( S \) of \( A \). For all non-negative integer \( n \), we define \( f(n) \) as the maximal size of a distinct-sum subset of \( \{1, 2, \ldots, n\} \).

1. Show that \( f(n) \geq 1 + \lfloor \log_2 n \rfloor \) (give an example of distinct-sum subset).

**Correction**: The set \( \{1, 2, \ldots, 2^{\lfloor \log_2 n \rfloor}\} \) is a distinct-sum set: a sum of elements corresponds to a binary decomposition.

2. Using a counting argument, show that \( f(n) \leq \log_2 n + \log_2 \log_2 n + O(1) \).

**Correction**: A distinct-sum set of size \( k \) returns \( 2^k \) different partial sums. A partial sum in necessarily less than \( nk \) (sum of at most \( k \) terms that are at most \( n \)). We then have \( 2^k \leq nk \). Then \( k \leq \log_2 (n^2) + 2 \leq \log_2 \log_2 (n^2) + \log_2 (2 \log_2 n) = \log_2 \log_2 (n) + O(1) \), where we use that \( k \leq n \). Then \( k \leq \log_2 n + \log_2 \log_2 n + O(1) \).

The goal of the rest of the exercise is to improve the coefficient of the term \( \log_2 \log_2 n \) by showing that

\[
f(n) \leq \log_2 n + \frac{1}{2} \log_2 \log_2 n + O(1) .
\]

Let \( A \subseteq \{1, 2, \ldots, n\} \) be a distinct-sum set. Set \( k = |A| \) and \( X \) a subset of \( A \) chosen uniformly at random, and define \( S_X \) as the sum of the elements of \( X \).

3. Let \( \lambda > 1 \). Show that

\[
\mathbf{P}( |S_X - \mathbf{E}[S_X]| \geq \lambda n \sqrt{k}/2 ) \leq 1/\lambda^2 .
\]

**Correction**: With the notations defined, we have : \( \mathbf{E}[S_X] = \frac{1}{2} \sum_{a \in X} a \). Indeed, let \( X_a = 0 \) if \( a \notin X \) (with probability \( 1/2 \)) and \( X_a = a \) otherwise. Then \( \mathbf{E}[X_a] = a/2 \), and we use the linearity of the expectation.

\((X_a)_{a \in A}\) is a mutually independent family of random variables, so

\[
\text{Var}(S_X) = \sum_{a \in A} \text{Var}(X_a) = \frac{1}{4} \sum_{a \in X} a^2 \leq \frac{k n^2}{4} .
\]

By applying the Tchebychev inequality, one gets

\[
\mathbf{P}( |S_X - \mathbf{E}[S_X]| \geq \lambda n \sqrt{k}/2 ) \leq \frac{\text{Var}(S_X)}{(\lambda n \sqrt{k}/2)^2} \leq 1/\lambda^2 .
\]

4. Deduce that

\[
n \geq \frac{2^k (1 - 1/\lambda^2) - 1}{\sqrt{k} \lambda} .
\]

**Correction**: If \( A \) is a distinct-sum set, there are \( 2^k \) possible values for \( S_X \), and for all \( x \in \mathbb{N} \), \( \mathbf{P}(S_X = x) \in \{0, 2^{-k}\} \) (depending whether \( x \) is a possible sum or not). One then can bound \( \mathbf{P}( |S_X - \mathbf{E}[S_X]| < \lambda n \sqrt{k}/2 ) \leq (1 + 2\lambda n \sqrt{k}/2)^{-2^k} \), and one gets

\[
1 - \lambda^{-2} \leq \mathbf{P}( |S_X - \mathbf{E}[S_X]| < \lambda n \sqrt{k}/2 ) \leq (1 + 2\lambda n \sqrt{k}/2)^{-2^k} ,
\]

and \( n \geq \frac{2^k (1 - 1/\lambda^2) - 1}{\sqrt{k} \lambda} .\)
5. Conclude.

Corr ection: There exists \( C \) such that \( n \geq C \frac{2^k}{\sqrt{k}} \), and \( \sqrt{kn} \geq C 2^k \). Then \( k \leq \log_2(n) + \frac{1}{2} \log_2(k) - \log_2(C) \). But then \( \log_2(k) \leq \log_2 \log_2(n \sqrt{k}) + O(1) = \log_2 \log_2(n) + O(1) \), hence done \( k \leq \log_2 n + \frac{1}{2} \log_2 \log_2 n + O(1) \). This is true for all \( k \leq f(n) \).

Exercise 7 Analysis of the Quick-sort algorithm

Algorithm 1: Quick_sort

Input: A list \( S \) of \( n \) distinct numbers
Output: The sorted list of the elements of \( S \)
begin
\indent if \( S \) has 0 or 1 element then return \( S \);
\indent else
\indent \indent Choose an element \( x \) (pivot) of \( S \) and separate the other elements in two
\indent \indent sub-lists
\indent \indent \indent \( S_1 \), list of the elements of \( S \) that are \( < x \);
\indent \indent \indent \( S_2 \), list of the elements of \( S \) that are \( > x \);
\indent \indent Quick_sort(\( S_1 \)); Quick_sort(\( S_2 \));
\indent \indent Return the list \( S_1, x, S_2 \).
\end{algorithm}

1. Give an example of a list that requires \( \Omega(n^2) \) comparisons to sort the list with this algorithm.

The goal of the exercise is to show that if the pivots are chosen uniformly at random, then the expectation of the number of comparisons is \( 2n \ln n + O(n) \). We note \( y_1 < y_2 < \cdots < y_n \) the elements of the list.

2. What is the probability that two elements \( y_i \) and \( y_j \) are compared?

3. Deduce the result.

4. What happens if the first element is always chosen as pivot? What is the difference with the choice of a random pivot?

Exercise 8 Coloring

Let \( G = (V, E) \) be a non-directed graph and suppose that a set \( S(v) \) of \( 8r \) colors is assigned to each vertex \( v \in V \), with \( r \geq 1 \). Moreover, for each vertex \( v \) and each color \( c \in S(v) \), there are at most \( r \) neighbors \( u \) of \( v \) such that \( c \in S(u) \).

Show that there exists a proper coloring (no two adjacent vertices have the same color) such that for all \( v \), the color of \( v \) is chosen in \( S(v) \).