

The use of the probabilistic method for IMO-type problems - Atli FF:

This first section will contain only introductory definitions and results from probabilistics. We will consider only countable probability spaces, so we have no need of integrals or general measures. We can also consider only probabilities given by odds to simplify things. For those who are already familiar with probabilistics in a more general context, this first section might be redundant. The definitions given are equivalent to more general ones when restricted to the case of countably many possible outcomes.

Definition. Let Ω be a countable set of possible outcomes. For each outcome $\omega \in \Omega$ we assign a probability $P(\omega) \geq 0$ such that $\sum_{\omega \in \Omega} P(\omega) = 1$.

Definition. An event A is a subset of Ω . The probability of an event A is given by $P(A) = \sum_{\omega \in A} P(\omega)$.

A classic example would be the result of a dice roll. Consider for example a dice with six sides numbered 1 through 6. Then $\Omega = \{1, 2, 3, 4, 5, 6\}$ and $P(\omega) = 1/6$ for any ω . We could then define the event A as the result being odd, i.e. $A = \{1, 3, 5\}$. Then $P(A) = 1/2$. We could also consider a crooked dice where the odds of landing on 6 is $1/2$ and the other results have a chance of $1/10$. Then $P(A) = 3/10$.

Proposition (Union Bound). For any events A_1, \dots, A_n we have $P(A_1 \cup \dots \cup A_n) \leq P(A_1) + \dots + P(A_n)$.

The proof of this is left as an exercise to the reader and should not pose too much trouble. Despite the fairly trivial nature of this proposition it will often be quite useful going forward. Going forward I also won't prove all the results given because a lot of them are useful to an IMO contestant but their proofs are outside the scope of what an IMO contestant is expected to know.

A common trick to using the probabilistic method is to show existence by showing that the probability of something occurring is > 0 and the union bound is often helpful to achieve this. Let $R(k, l)$ be the smallest integer n so that every red-blue colouring of K_n has a red k -clique or blue l -clique. Some of you might recognize this construction as it is known as the Ramsey numbers. We can show that $R(k, k) \geq 2^{k/2}$ using this trick. Let $n = k/2$ and color each edge independently and randomly with probability $1/2$ for each outcome. For a subset S of size k we get that the event A_S of S being monochromatic has probability $2 \cdot 2^{-\binom{k}{2}}$ of occurring. By the union bound we get

$$P\left(\bigcup_S A_S\right) \leq \sum_S P(A_S) = \binom{n}{k} 2^{-\binom{k}{2}+1} \leq \frac{n^k}{k!} 2^{1+k/2-k^2/2} = \frac{2^{1+k/2}}{k!} \frac{n^k}{2^{k^2/2}}$$

Plugging in $n = 2^{k/2}$ we get that this is less than 1 for $k \geq 1$ so the probability that none of the A_S occur is positive, so there exists a colouring with no monochromatic k -clique.

Let us consider another example. A hypergraph is a graph where edges are subsets of vertices, possibly with more than 2 elements. If every edge contains n vertices it is called n -uniform. Suppose now that $n \geq 4$ and H is an n -uniform hypergraph with at most $4^{n-1}/3^n$ edges. We will show there exists a four-coloring of H such that every edge has all colors represented. We color each vertex independently and with uniform probability. For an edge e let A_e be the event that at least one color is missing from e . Then $P(A_e) < 4 \cdot 3^n \cdot (1/4)^n$. There are four colors that could be missing and for any one missing we have 3^n choices for the rest out of a total 4^n choices. The inequality is strict because we are considering the cases where more than one color is missing more than once. Union bound gives $P(\bigcup_e A_e) < 1$, so there is some such coloring.

Definition. A random variable is a function $X : \Omega \rightarrow \mathbb{R}$. The expected value of a random variable X is $\mathbb{E}[X] := \sum_{\omega \in \Omega} P(\omega)X(\omega)$.

Proposition (Linearity of expectation). \mathbb{E} is linear, that is to say for random variables X, Y and $\lambda \in \mathbb{R}$ we have $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$ and $\mathbb{E}[\lambda X] = \lambda \mathbb{E}[X]$.

Quite useful, especially used in conjunction with indicator variables. An indicator variable is simply a random variable derived from an event A that takes value 1 on the elements of A and value 0 on elements outside A . This random variable is usually denoted $\mathbb{1}_A$. We consider an example. Color the edges of K_n red or blue independently with each color chosen with probability $1/2$. Let X be the number of red triangles in the graph, we wish to determine the expected value of X . Label the triangles in K_n by $1, 2, \dots, \binom{n}{3}$. Then we can define the event E_i as being the set of outcomes when triangle i is red. Clearly we have that E_i is true $1/8$ -th of the time so calculating the

expected value we get $\mathbb{E}[\mathbb{1}_{E_i}] = 1/8$. Then $X = \mathbb{1}_{E_1} + \dots + \mathbb{1}_{E_{\binom{n}{3}}}$ so linearity gives $\mathbb{E}[X] = \binom{n}{3}/8$.

Similarly we can consider another classical example. Consider a randomly chosen permutation (chosen uniformly) on n elements. Let X be the random variable giving the number of fixed points in the permutation. What is $\mathbb{E}[X]$? We can define the event F_i as the set of outcomes where the i -th value is fixed. What is $\mathbb{E}[F_i]$? Well exactly 1 out of every n permutations ($n!/(n-1)!$) fixes element i , so $\mathbb{E}[F_i] = 1/n$. Thus we get that $\mathbb{E}[X] = \mathbb{E}[F_1] + \dots + \mathbb{E}[F_n] = 1$.

Now we take an example that is more along the lines of what you'd use expectation for to solve IMO-type problems. We shall prove that there exists a tournament with n players containing at least $n!/2^{n-1}$ Hamiltonian paths. To do this we consider a tournament on n vertices where we choose the direction of each edge independently, choosing either direction with probability $1/2$. For a permutation σ on n elements let X_σ be indicator random variable for whether σ enumerates a Hamiltonian path in our tournament. Then $X = \sum_\sigma X_\sigma$ counts the number of Hamiltonian paths. We get that any such path enumerated by σ has $n-1$ edges so the odds that they all point in the right direction is 2^{-n+1} . Thus $\mathbb{E}[X] = \sum_\sigma 1/2^{n-1} = n!/2^{n-1}$. But the expected value is a weighted average, so there must be some choice of tournament giving us $X(\omega) \geq n!/2^{n-1}$ which completes the proof.

Proposition (Markov's inequality). For a random variable $X \geq 0$ we have for any $t > 0$ that $P(X \geq t) \leq \mathbb{E}[X]/t$.

Proof. $\mathbb{E}[X] = \sum_{\omega \in \Omega} P(\omega)X(\omega) \geq \sum_{\omega \in \Omega: X(\omega) \geq t} P(\omega)t = tP(X \geq t)$. □

Definition. The variance of a random variable X is defined to be $\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$. Calculating this out we get $\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.

Definition. The covariance of two random variables X, Y is defined to be $\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$.

Proposition (Variance of sums). For random variables X_1, \dots, X_k we have $\text{Var}(X_1 + \dots + X_k) = \sum \text{Var}(X_i) + \sum_{i \neq j} \text{Cov}(X_i, X_j)$.

Definition. We say two events A, B are independent if $P(A \cap B) = P(A)P(B)$. This definition can also be interpreted using conditional probability. The probability that A occurs assuming B occurs, denoted $P(A|B)$ is given by $P(A \cap B)/P(B)$ so this means the events are independent if $P(A|B) = P(A)$. Additionally for random variables X, Y we say they are independent if for all $t, s \in \mathbb{R}$ we have $P((X = t) \cap (Y = s)) = P(X = t)P(Y = s)$. Lastly we say that X_1, \dots, X_k are mutually independent if for every $t_1, \dots, t_k \in \mathbb{R}$ we have $P((X_1 = t_1) \cap \dots \cap (X_k = t_k)) = P(X_1 = t_1) \dots P(X_k = t_k)$. We note that mutual independence implies pairwise independence but not vice versa.

Proposition (Expectation of product). For mutually independent variables X_1, \dots, X_k we have $\mathbb{E}[X_1 \dots X_k] = \mathbb{E}[X_1] \dots \mathbb{E}[X_k]$. We get as a corollary that if X, Y are independent then $\text{Cov}(X, Y) = 0$. In a similar fashion we get that for pairwise independent X_i we have that Var is additive.

Proposition (Chebyshev's inequality). Let X be a random variable. Then for any $t > 0$ we get $P(|X - \mathbb{E}[X]| \geq t) \leq \text{Var}(X)/t^2$.

Proof. Just plug $Y = (X - \mathbb{E}[X])^2$ into Markov's inequality. □

These concepts and results can be useful to bound how far our random variables disperse around the expected value. For IMO-type problems this is usually useful to deduce that some certain fraction, say half, of the results are within close distance from the expected value. This often leads into using the pigeonhole principle or something of the sort. We shall look at an example of sum-free sets. Let A be a subset of $\{1, 2, \dots, n\}$. It is called sum-free if the $2^{|A|}$ different sums of subsets of A are all distinct. One might wonder what the maximum possible size of a sum-free subset of $\{1, 2, \dots, n\}$ is, we denote this size by $f(n)$. Choosing only powers of two we get that $f(n) \geq \lfloor \log_2(n) \rfloor$, but we can use probabilistic methods to get the upper bound $f(n) < \log_2(n) + \log_2(\log_2(n))/2 + C$ for some constant C . Let $A = \{a_1, \dots, a_k\}$ be a sum-free set of maximal size. Let $\epsilon_1, \dots, \epsilon_k$ be randomly chosen as either 0 or 1, each chosen independently. Let $X = \sum_i \epsilon_i a_i$. Then

$$\mathbb{E}[X] = \sum_{i=1}^k \mathbb{E}[\epsilon_i a_i] = \frac{1}{2} \sum_{i=1}^k a_i$$

$$\text{Var}(X) = \sum_{i=1}^k \text{Var}(\epsilon_i a_i) = \frac{1}{4} \sum_{i=1}^k a_i^2 \leq \frac{1}{4} k n^2$$

Thus Chebyshev's inequality gives us that

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

Thus the ratio of sums within distance $n\sqrt{k}$ from $\mathbb{E}[X]$ is at least $3/4$, so there are at least $3/4 \cdot 2^k$ of them within that distance. Thus we have $3/4 \cdot 2^k$ different sums in an interval of length $2n\sqrt{k}$ which gives us $3/4 \cdot 2^k \leq 2n\sqrt{k}$. Solving this inequality gives us our desired result.

This method can also be a nice complement to the method of alterations. The method of alterations revolves around using probabilistics to construct a solution that's close to good enough and making sure that you can change it to be fully correct without breaking anything. I could not find very many IMO-type problems where this method is applicable, so I left it out of this text. But those interested can take a look at the book *The Probabilistic Method*, which this text draws heavily from.

Proposition (Chernoff's inequality). Let X_1, \dots, X_n be independent random variables taking values -1 and $+1$ each with probability $1/2$. Let $X = \sum_i X_i$, then for every $t > 0$ we have

$$P(X \leq -t) = P(X \geq t) \leq e^{-t^2/2n}$$

As a corollary we can get the following variant when the X_i instead take values 0 and 1 each with probability $1/2$

$$P\left(X \leq \frac{n}{2} - t\right) = P\left(X \geq \frac{n}{2} + t\right) \leq e^{-2t^2/n}$$

The usefulness of this inequality is similar to the discussion of variances above, just that it allows for better bounds in the case where the X_i behave as in the statement of the theorem. It just so happens that we quite often construct our random variables so they can be expanded into such X_i , so it can come in handy. Though I wouldn't place great emphasis on it for IMO-type problems, it can be viewed as an occasionally handy tool.

Proposition (Lovász Local Lemma). Consider events A_1, \dots, A_n each occurring with probability at most $p < 1$. Assume each event A_i is mutually independent from all other events except at most d other events B_j . Then if $ep(d+1) \leq 1$ then the probability that none of the events occur is strictly positive.

This lemma is extremely powerful when applicable. Let us consider an example taken from Evan Chen's problem list. Let n be a positive integer. Suppose $11n$ points are arranged on a circle, each colored with with one of n colors such that each color appears exactly 11 times. Prove that one can select a point of every color such that no two are adjacent. We pick one vertex of every color, chosen uniformly at random from all vertices of that color and consider each color independently. We let E_i be the event that point i and its right-hand neighbour both get picked. If they are the same color the chance of this occurring is zero and if they are different we get that the odds are $1/11 \cdot 1/11$. If E_i and E_j are different such events that don't share any points or colors, they are independent. There are two E'_i that share vertices with E_i . E_i contains up to two colors and those correspond to up to 22 vertices which are then part of up to 44 events. But then we are counting the event itself and one of its neighbours again twice, so we have at most 42 other non-independent events. Thus we can take $p = 11^{-2}$ and $d = 42$ in the local lemma and get that there is a non-zero chance that none of the events occur since $43e/121 \leq 1$. But then there must be some choice of points satisfying our conditions. Lastly I will put a few results down below that I have used personally when using the probabilistic method but I have not found a lot of use for within IMO-type problems. They are here simply for the sake of completeness. Further reading about the results can be found in *The Probabilistic Method*. Additionally that book covers a lot of material not covered or mentioned here.

Proposition (Azuma-Hoeffding inequality). Let us have a product probability space $\Omega = \prod_{i=1}^n \Omega_i$. An example would be a space where the outcomes are in S^n for some set S , then we can put $\Omega_i = S$. Let $X : \Omega \rightarrow \mathbb{R}$ be a random variable. We say that X is C -Lipschitz if changing the value of any one coordinate changes the value of X by at most C . If X is C -Lipschitz and $\lambda \geq 0$ we have

$$P(|X - \mathbb{E}[X]| \geq \lambda) \leq 2 \exp\left(-\frac{\lambda^2}{2C^2 n}\right)$$

Proposition (Correlation inequality). Let $A, B \subseteq \{0, 1\}^N$ be events. We consider A to be increasing if taking any element of A and increasing a coordinate from 0 to 1 does not take you outside A . We define events to be decreasing in a similar fashion. If A and B are both increasing or both decreasing then $P(A \cap B) \geq P(A)P(B)$. If one of them is increasing and the other decreasing then $P(A \cap B) \leq P(A)P(B)$.

The last thing to cover is that quite often using the probabilistic method involves using inequalities. I won't go over inequalities such as Jensen's inequality here and assume the reader either knows them or can look them up elsewhere. There are however some inequalities that one does not encounter particularly often outside this context. They are given below.

$$\binom{n}{k} \leq n^k/k!, \quad \binom{n}{k} \leq (en/k)^k, \quad \binom{n}{k} \geq (n/k)^k, \quad \binom{n}{k} \geq (n-k+1)^k/k!$$

$$1+x \leq e^x \text{ for } x \in \mathbb{R}, \quad 1-x \geq e^{-2x} \text{ for } x \in [0, 1/2]$$

We note that the inequality $1+x \leq e^x$ is quite often useful when setting $x = -p$ for some small probability p , so $1-p \leq e^{-p}$. This makes it easier to work with a large power of this value such as $(1-p)^N$.

Problems:

Problems rated on a scale of difficulty from 1 to 5. Problems with several subproblems are rated by their hardest subproblem. These difficulty levels are only my subjective experience with solving the problems, but I hope they give a rough idea at least. Starred problems require results from outside this lecture that the reader might not be familiar with or use material covered only briefly in the lecture. The source of the problems is given after the difficulty. I tried to give the earliest source I could find in each case. A lot of these problems can be solved non-probabilistically but they can all be solved probabilistically. I encourage the reader to try to solve them using probabilistic methods.

- (1, HMMT 2006) At a nursery 2006 babies sit in a circle. Suddenly each baby randomly pokes either the baby to its left or to its right. What is the expected number of unpoked babies?
- (1, Kraft Inequality) Let S be a finite set of finitely long binary strings such that no string in the set is a prefix of another. Denote the number of characters in a string s by $|s|$. Prove that $\sum_{s \in S} \frac{1}{2^{|s|}} \leq 1$.
- (2, Benny Sudakov) Let p be a prime number and $A \subset \mathbb{Z}_p$ be a set of $|A| < p^{2/3}$ residue classes modulo p . Show that there exist $x, y \in \mathbb{Z}_p$ such that $A \cap (A+x) \cap (A+y) = \emptyset$.
- (2, Bay Area Math Olympiad 2004) Consider n real numbers, not all zero, with sum zero. Prove that one can label the numbers as a_1, \dots, a_n such that $a_1a_2 + a_2a_3 + \dots + a_na_1 < 0$.
- (2, Russia 2006) At a tourist camp each person has at least 50 and at most 100 friends among the other people at the camp. Show that one can hand out a T-shirt to every person such that the T-shirts have at most 1331 colors and any person has 20 friends whose T-shirts have pairwise different colors.
- (2, IMC for University Students 2002) An olympiad has six problems and 200 contestants. Each problem is solved by at least 120 contestants. Prove that there exist two contestants such that each problem is solved by at least one of them.
- (2, Iran TST 2008) Suppose 799 teams participate in a tournament in which every pair of teams plays against each other exactly once. Prove that there exist two disjoint groups A and B of 7 teams each such that every team from A defeated every team from B .
- (2, The Probabilistic Method) Let G be a simple graph and assign each vertex v a list $S(v)$ of colors of size at least $10d$ where $d \geq 1$. Suppose additionally that for each $v \in V$ and $c \in S(v)$ there are at most d neighbors u of v such that $c \in S(u)$. Prove that G can be colored such that each vertex is given a color from its list and no edge is monochromatic.
- (2, The Probabilistic Method) We say a tournament T is k -dominated if for every set of k vertices there is some vertex that dominates all of them. Let $f(k)$ be the size of the smallest k -dominated tournament. Prove that $f(k) \geq 2^{k+1} - 1$.

- (2, IMO Shortlist 1987) Show that we can color the elements of the set $\{1, 2, \dots, 1987\}$ with 4 colors such that any arithmetic progression of ten terms from the set is not monochromatic.
- (2, IMO 1987) Let $p_n(k)$ be the number of permutations of the set $\{1, 2, \dots, n\}$ ($n \geq 1$) which have exactly k fixed points. Prove that $\sum_{k=0}^n kp_n(k) = n!$.
- (2, MathLinks.ro 2008) Let A_1, \dots, A_n and B_1, \dots, B_n be distinct finite subsets of \mathbb{N} such that for every i we have $A_i \cap B_i = \emptyset$ and for every $i \neq j$ we have $(A_i \cap B_j) \cup (A_j \cap B_i) \neq \emptyset$. Prove that for any real $0 \leq p \leq 1$ we have $\sum_i p^{|A_i|} (1-p)^{|B_i|} \leq 1$.
- (2*, Benny Sudakov) Let G be a graph with m edges and S a random set of vertices of G obtained by including any given vertex with probability $1/2$. Prove that the probability that S is an independent set in G is at least $(3/4)^m$.
- (3, IMO shortlist 1999) Let A be a set of N residues modulo N^2 . Prove that there exists a set B of N residues modulo N^2 such that $A + B$ contains at least half of the residue classes modulo N^2 .
- (3, IMO Shortlist 2006) Let S be a finite set of points in the plane such that no three of them are on a line. For each convex polygon P whose vertices are in S let $a(P)$ be the number of vertices of P and $b(P)$ be the number of points of S which are outside P . Prove that for every real number x we have $\sum_P x^{a(P)} (1-x)^{b(P)} = 1$ where the sum is taken over all convex polygons with vertices in S . A line segment, a point and the empty set are considered to be polygons of 2, 1 and 0 vertices respectively.
- (3, The Probabilistic Method) Prove that there is an absolute constant $c > 0$ with the following property. Let A be an n by n matrix with pairwise distinct entries. Then there is a permutation of the rows of A so that no column in the permuted matrix contains an increasing subsequence of length at least $c\sqrt{n}$.
- (3, Benny Sudakov) Let G be a directed graph on n vertices with at least $\lceil n \log_2(n) \rceil$ edges. Prove that there exists a tournament on n vertices that does not contain a subgraph isomorphic to G .
- (3, Russia 1996) Suppose we have 1600 students who have formed 16000 committees of 80 people each. Prove that one can find two committees having no fewer than 4 common members.
- (3, MOP 2007) In an $n \times n$ matrix each of the numbers $1, 2, \dots, n$ appears exactly n times. Show that there is a row or column in the array with at least \sqrt{n} distinct numbers.
- (3, Bollobás) Let A_1, \dots, A_n and B_1, \dots, B_n be distinct subsets of \mathbb{N} such that for every i we have $|A_i| = r$, $|B_i| = s$ and $A_i \cap B_i = \emptyset$. Additionally for every $i \neq j$ we have $A_i \cap B_j \neq \emptyset$. Prove that $n \leq \binom{r+s}{r}$.
- (3, The Probabilistic Method) Let G be a bipartite graph with 2^n vertices and assign each vertex v a list $S(v)$ of more than n colors. Prove that G can be colored by assigning each vertex a color from its list such that no edge is monochromatic.
- (3, The Probabilistic Method) Let $v_1, \dots, v_n \in \mathbb{R}^n$ be vectors of length 1. Show that there exist $\epsilon_1, \dots, \epsilon_n \in \{-1, 1\}$ such that $|\epsilon_1 v_1 + \dots + \epsilon_n v_n| \leq \sqrt{n}$.
- (3, Zarankiewicz) Show that there exists a partition of the set of positive integers into two classes such that neither class contains an infinite arithmetic progression and neither class contains 3 consecutive integers.
- (3, The Probabilistic Method) Let $v_1 = (x_1, y_1), \dots, v_n = (x_n, y_n)$ be n two-dimensional vectors where each x_i, y_i is an integer with absolute value at most $2^{n/2}/(100\sqrt{n})$. Show that there exist distinct disjoint sets $I, J \subseteq \{1, 2, \dots, n\}$ such that $\sum_{i \in I} v_i = \sum_{j \in J} v_j$.
- (3, IMO 1989) A permutation $(x_1, x_2, \dots, x_{2n})$ of $\{1, 2, \dots, 2n\}$ where n is a positive integer is said to have property P if $|x_i - x_{i+1}| = n$ for at least one i in $\{1, 2, \dots, 2n-1\}$. Show that for each n there are more permutations with property P than without.
- (3, Benny Sudakov) Let T be a tournament such that every vertex has outdegree at least 10. Show that the vertices of T can be 2-colored such that every vertex in T has at least one outneighbour in each color.
- (3, Benny Sudakov) Let H be a d -uniform hypergraph with $d \geq 6$. Show that the vertices of H can be colored red or blue such that the number of red and blue vertices in every edge differ by at most $\sqrt{6d \log(d)}$.

- (3, Benny Sudakov) Let G be a d -regular graph with $d \geq 2$. Show that there exists a subset U of the vertex set of G such that for every vertex v the neighbourhood $N(v)$ of v satisfies $1 \leq |N(v) \cap U| \leq 50 \log(d)$.
- (3, IMO 1970) In a plane there are 100 points no three of which are collinear. Consider all possible triangles having these points as vertices. Prove that no more than 70% of these triangles are acute-angled.
- (3*, Chinese Olympiad 1986) Let z_1, \dots, z_n be complex numbers. Show that there is a set $S \subseteq \{1, 2, \dots, n\}$ such that $\pi \left| \sum_{j \in S} z_j \right| \geq \sum_{j=1}^n |z_j|$.
- (3*, The Probabilistic Method) Let G be the graph whose vertices are all 7^n vectors of length n with coordinate values in \mathbb{Z}_7 in which two vertices are adjacent iff they differ in precisely one coordinate. Let $U \subseteq V$ be a set of 7^{n-1} vertices of G and let W be the set of all vertices of G whose distance from U exceeds $(c+2)\sqrt{n}$ where $c > 0$ is a constant. Prove that $|W| \leq 7^n e^{-c^2/2}$.
- (3*, The Probabilistic Method) A family of subsets \mathcal{G} is called intersecting if $G_1 \cap G_2 \neq \emptyset$ for all $G_1, G_2 \in \mathcal{G}$ let $\mathcal{F}_1, \mathcal{F}_2, \dots, \mathcal{F}_k$ be k intersecting families of subsets of $\{1, 2, \dots, n\}$. Prove that $|\bigcup_{i=1}^k \mathcal{F}_i| \leq 2^n - 2^{n-k}$.
- (4, The Probabilistic Method) Let X be a set of pairwise orthogonal unit vectors in \mathbb{R}^n such that the projection of each of these vectors onto the first k coordinates has length at least ϵ . Show that $|X| \leq k/\epsilon^2$ and that this is tight when $\epsilon^2 = k/2^r < 1$ for some integer r .
- (4, Erdős) Prove that in any set S of n distinct positive integers we can always find a subset T with more than $n/3$ elements with the property that $a + b \neq c$ for any $a, b, c \in T$ (not necessarily distinct).
- (4, Benny Sudakov) Let G be a simple graph with n vertices and m edges and let k be a positive integer. Prove: a) There are at least $k^n(1 - m/k)$ k -colourings of G . b) There are at most $k^n(k - 1)/m$ k -colourings of G . c) Improve the bound in (b) to $k^n(k - 1)/(k + m - 1)$.
- (4, Sperner) Let \mathcal{A} be a family of subsets of $\{1, 2, \dots, n\}$ such that there exist no distinct $A, B \in \mathcal{A}$ with $A \subseteq B$. Prove that $|\mathcal{A}| \leq \binom{n}{\lfloor n/2 \rfloor}$.
- (4, Benny Sudakov) Let S_1, \dots, S_k be subsets of $\{1, 2, \dots, n\}$. Prove that if $k < 1.99n/\log_2(n)$ then there are two distinct subsets X, Y of $\{1, 2, \dots, n\}$ such that $|X \cap S_i| = |Y \cap S_i|$ for all i .
- (4, Benny Sudakov) Define two vectors to be ϵ -almost orthogonal if their inner product has absolute value at most ϵ . Show that: a) For every $0 < \epsilon < 1$ there exists an ϵ -almost orthogonal set in \mathbb{R}^n of size at least $\lfloor \exp(n\epsilon^2/4) \rfloor$. b) For every $0 < \epsilon < 1$ there exists an ϵ -almost orthogonal set in \mathbb{R}^n of size at least $\exp(n\epsilon^2/2)/4$. c) There is a $2\sqrt{\log(n)}/n$ -almost orthogonal set in \mathbb{R}^n of size at least $n^2/4$.
- (4, Benny Sudakov) You have an $n \times n$ grid where each cell is white or black. You can invert the color in a row or column as often as you like and your goal is to maximize the number of black cells. Prove that there exists an initial configuration such that you can never turn more than a ratio of $1/2 + \sqrt{\ln(2)/n}$ of them black.
- (4, Ravi Boppana) Let $p_n(k)$ be the number of permutations of the set $\{1, 2, \dots, n\}$ ($n \geq 1$) which have exactly k fixed points. Determine the value of $\sum_{k=0}^n k^2 p_n(k)$.
- (4, IMO 1998) In a competition there are a contestants and b judges where $b \geq 3$ is an odd integer. Each judge rates each contestant as either pass or fail. Suppose k is a number such that for any two judges their ratings coincide for at most k contestants. Prove that $k/a \geq (b - 1)/(2b)$.
- (4, IMO 1989) Let n and k be positive integers and let S be a set of n points in the plane such that no three points in S are collinear and for any point $P \in S$ there are at least k points of S equidistant from P . Prove that $k \leq 1/2 + \sqrt{2n}$.
- (4*, The Probabilistic Method) Let G be a graph with chromatic number 1000 and let U be a random subset of the vertex set of G chosen uniformly among all such subsets. Let H be the subgraph induced by G on the vertex set U . Prove that there is less than a 1% chance that H has chromatic number ≤ 400 .
- (4*, MOP Test 2008) Suppose that a, b, c are positive real numbers such that for every integer n we have $\lfloor an \rfloor + \lfloor bn \rfloor = \lfloor cn \rfloor$. Prove that at least one of a, b, c is an integer.

- (5, The Probabilistic Method) We say a tournament T is k -dominated if for every set of k vertices there is some vertex that dominates all of them. Let $f(k)$ be the size of the smallest k -dominated tournament. Prove that $f(k) \geq Ck2^k$ for some constant C .
- (5, The Probabilistic Method) Let X be a set of m distinct residue classes modulo a prime p . An interval of length l in $\{0, 1, \dots, p-1\}$ is a set of the form $\{i, i+1, \dots, i+l-1\}$ with all values considered modulo p for some i . Show that if $m \geq 4k^2$ then there is an integer a such that the set aX intersects every interval of length at least p/k in $\{0, 1, \dots, p-1\}$.
- (5*, The Probabilistic Method) Show that there is a finite n_0 such that any directed graph on $n > n_0$ vertices in which each outdegree is at least $\log_2(n) - \log_2(\log_2(n))/10$ contains an even simple directed cycle.
- (5*, The Probabilistic Method) Let G be a graph on $n \geq 10$ vertices and suppose that if we add any edge to G that is not already present then the number of copies of K_{10} in G increases. Show that the number of edges of G is at least $8n - 36$.

References:

- Alon and Spencer's *The Probabilistic Method*.
- Teaching material from Benny Sudakov's course on the probabilistic method at ETH Zürich.
- Ravi Boppana's problem set.
<https://artofproblemsolving.com/community/c6h23507p1943887>
- Po-Shen Loh's notes on the probabilistic method.
<https://www.math.cmu.edu/~ploh/olympiad.shtml>
- Evan Chen's notes on probability.
<https://web.evanchen.cc/handouts/ProbabilisticMethod/ProbabilisticMethod.pdf>