# 1 Probability!

## 1.1 Foundations

### 1.1.1 Basic notions

Here's a quick glossary, since I don't actually want to teach these stuff:

- **event** - something that happens (e.g. a die rolling 6)
- **random variables** - a variable that randomly takes a value (e.g. the result of the die roll)
- **expected value** - the "average" value of a random variable
- **conditional probability** - the probability of an event happening given another event.
- **independence** - two events $A, B$ are independent if the probability of both event simultaneously happening is $P(A)P(B)$. Two random variables $A, B$ are independent if for any pair of values $a, b$, the events $A = a$ and $B = b$ are independent.

### 1.1.2 Linearity of Expectation

A lot of the strength of using expected values comes from the following fact:

**Theorem.** (Linearity of Expectation) Let $X_1, X_2, ..., X_n$ be *not necessarily independent* random variables, then

$$\mathbb{E}\left[X_1 + X_2 + ... + X_n\right] = \mathbb{E}\left[X_1\right] + \mathbb{E}\left[X_2\right] + ... + \mathbb{E}\left[X_n\right]$$

*Exercise.* Prove this! (Hint: you only have to do the 2-variable case.)

Here's two examples (from Evan Chen's notes) on how powerful this idea is:

**Example.** At MOP, there are $n$ people, each of who has a name tag. We shuffle the name tags and randomly give each person one of the name tags. Let $S$ be the number of people who receive their own name tag. Prove that the expected value of $S$ is 1.

*Solution.* The trick is to define *indicator variables* as follows: for each $i = 1, 2, \ldots, n$ let

$$S_i \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if person } i \text{ gets his own name tag} \\ 0 & \text{otherwise} \end{cases}$$

Obviously,

$$S = S_1 + S_2 + \cdots + S_n$$

Moreover, it is easy to see that $\mathbb{E}\left[S_i\right] = \mathbb{P}\left(S_i = 1\right) = \frac{1}{n}$ for each $i$ : if we look any particular person, the probability they get their own name tag is simply $\frac{1}{n}$. Therefore,

$$\mathbb{E}[S] = \mathbb{E}\left[S_1\right] + \mathbb{E}\left[S_2\right] + \cdots + \mathbb{E}\left[S_n\right] = \underbrace{\frac{1}{n} + \frac{1}{n} + \cdots + \frac{1}{n}}_{n \text{ times}} = 1.$$

□

The idea of indicator variables occurs so frequently that we have dedicated notation for it:

$$\mathbb{I}\left[\text{event}\right] = \begin{cases} 1 & \text{if event happens} \\ 0 & \text{otherwise} \end{cases}$$

In particular, $\mathbb{E}\left[\mathbb{I}\left[E\right]\right] = \mathbb{P}\left[E\right]$.

**Exercise.** (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 value of the number of unpoked babies?

### 1.1.3 Expectation Inequalities

You will likely not have to use these, but it's good to know I guess.

The expectation $\mathbb{E}$'s behave essentially like weighted sums, so all the inequalities involving weighted sums can essentially be re-written in expected value terms. For example:

1. Cauchy-Schwarz: $\mathbb{E}\left[X^2\right]\mathbb{E}\left[Y^2\right] \geq \mathbb{E}\left[XY\right]^2$
2. Jensen's inequality: $\mathbb{E}\left[f(X)\right] \geq f\left(\mathbb{E}\left[X\right]\right)$ for convex $f$

## 1.2 Showing Certain Existence with Probability (I)

How can arguments involving *probability* prove statements that are *certain*?

Usually, the statement we want to show is the existence of some object. However, sometimes direct construction is difficult, so instead we introduce randomness in the construction of a object (thus obtaining a random object). Thereafter, we make some indirect deduction that the desired object exists.

There are two specific versions of this argument that comes up pretty often (examples from David Arthur's notes):

*(1) To show that there exists a configuration where the quantity $X$ is at least $Y$, we can randomly select a configuration and show that $\mathbb{E}\left[X\right] \geq Y$. (This is essentially Pigeonhole but better.)*

**Example.** At the IMO, there are $n$ students, and $m$ pairs of these students are enemies. Prove that it is possible to divide the students into $k$ rooms so that there are at most $\frac{m}{k}$ pairs of enemies that are placed in the same room as each other.

*Solution.* Put every student into one of the $k$ rooms uniformly at random, then the probability of a given pair of enemies being in the same room is $1/k^2$. Hence, if $X$ is the total number of pairs of enemies that are placed in the same room, then $\mathbb{E}\left[X\right] = m/k$ (by linearity), so there exists some way to place students into rooms with $X \geq m/k$. □

#### Problems

1. (Canada 2009) Two circles of different radii are cut out of cardboard. Each circle is subdivided into 200 equal sectors. On each circle 100 sectors are painted white and the other sectors are painted black. The smaller circle is then placed on top of the larger circle, so that their centers coincide. Show that one can rotate the small circle so that the sectors on the two circles line up and at least 100 sectors on the small circle lie over sectors of the same color on the big circle.
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 $B$ .
3. Let $v_1, v_2, \ldots, v_n$ be unit vectors in $\mathbb{R}^d$. Prove that it is possible to pick $\varepsilon_i \in \{\pm 1\}$ such that
$$|\varepsilon_1 v_1 + \varepsilon_2 v_2 + \ldots + \varepsilon_n v_n| \leq \sqrt{n}$$
4. (Russia 1999) In a class every boy knows at least one girl. Prove there exist a group with at least half of the students such that each boy in the group knows an odd number of girls in the group.
5. (MOP 2007) In an $n \times n$ array, each of the numbers $1, 2, \ldots, n$ appears exactly $n$ times. Show that there is a row or a column in the array with at least $\sqrt{n}$ distinct numbers.
6. (ISL 1999/C4 ) Let $A$ be any set of $n$ residues mod $n^2$ . Show that there is a set $B$ of $n$ residues mod $n^2$ such that at least half of the residues mod $n^2$ can be written as $a+b$ with $a \in A$ and $b \in B$. (Hint: $1 - 1/n \leq e^{-1/n}$.)

7. ($*$) (Erdős) Prove that in any set $S$ of $n$ distinct positive integers we can always find a subset $T$ with $\frac{1}{3}n$ or more elements with the property that $a + b \neq c$ for any $a, b, c \in T$ (not necessarily distinct).

## 1.3 Showing Certain Existence with Probability (II)

*(2) To show that there exists an object satisfying various constraints, we can pick the object randomly and show that the total probability that the random object violates a constraint (across all constraints) is $< 1$.*

**Example.** At the IMO, there are $n$ people, some of whom are students and some of whom are guides. Each person brought $k > \log_2 n$ different colored shirts. To avoid confusion, the IMO wants to ensure that no guide is wearing the same colored shirt as a student. Prove that there is a choice of shirts which ensures this.

*Solution.* For each shirt color, we choose randomly and independently whether it will be allowed for students or allowed for guides. We then need to show that with positive probability, every person has at least one shirt they can wear.

Towards that end, consider a single person. The person owns $k$ shirt colors, and the probability that all of them are assigned to the other side is exactly $\left(\frac{1}{2}\right)^k$. It follows that the probability of somebody having no valid shirts is at most $n \cdot \left(\frac{1}{2}\right)^k < n \cdot \left(\frac{1}{2}\right)^{\log_2 n} = 1$. Therefore, the probability of everybody having at least one valid shirt is greater than 0, and we're done. $\square$

*Note.* Here we've implicitly used what's called the **union bound**: for two events $A, B$,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \leq P(A) + P(B)$$

### Problems

These are harder than usual:

1. (LYM inequality) Let $A_1, \ldots, A_s$ be subsets of $\{1, \ldots, M\}$, and suppose that none of the $A_i$ are subsets of each other. For each index $i$, let $a_i = |A_i|$. Prove that

$$\sum_{i=1}^{s} \frac{1}{\binom{M}{a_i}} \leq 1$$

Conclude also that $s \leq \binom{n}{\lfloor n/2 \rfloor}$ (which is Sperner's lemma).

2. (Sweden 2010, adapted). In a town with $n$ people, any two people either know each other, or they both know someone in common. Prove that one can find a group of at most $\sqrt{n \log n} + 1$ people, such that anyone else knows at least one person in the group.

3. 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 $c\sqrt{n}$. (Hint: you might find the following estimates useful:

$$\binom{n}{m} \leq \left(\frac{en}{m}\right)^m$$

$$m! \geq \sqrt{2\pi m}\left(\frac{m}{e}\right)^m$$

)

4. ($*$) (Erdős) Show that there exists a graph whose chromatic number and length of shortest cycle are both at least 2019.

## 1.4 Alteration (a.k.a. Random Algorithms)

The motivation for this technique is roughly based on the idea that randomly picking an object is decent on average but performs badly near the "tails":

**Problem.** (The stamp collector problem) McDonald's has decided to release a series of $n$ different Pokemon Happy Meal toys.

1. Show that the expected number of Happy Meals required to collect 90% of all the toys is around $n \ln 10$.
2. But if we truly "gotta catch 'em all", show that the expected number of Happy Meals required is instead around $n \ln n$.

This motivates the following approach: suppose that we want to find a set of objects $S$ of a certain size $N$ satisfying some property. Instead of picking a random set of size $N$ and hoping that it works (which is will not), we will

1. Initially pick each object with an undetermined probability $p$ to produce a set $S'$ that may not satisfy the property
2. Prune set $S'$ by removing objects until the desired property is satisfied (call the resulting set $S$)
3. Estimate $\mathbb{E}[|S|]$, and optimize the value of $p$.

Here's an example:

**Example.** (Weak Turán) A graph $G$ has $n$ vertices and average degree $d$. Prove that it is possible to select an independent set of size at least $\frac{n}{2d}$. (An *independent set* is a subset of vertices with no edges between them.)

*Solution.* Select each vertex with probability $p$. The resulting set (with $V$ vertices and $E$ edges) can be made into an independent set by sequentially taking each edge and deleting one of its endpoints. This removes at most $E$ vertices, so $|S| \geq V - E$. Usually, this would be a terrible bound, but

$$\mathbb{E}[|S|] \geq \mathbb{E}[V - E] = np - \frac{1}{2}ndp^2 = np\left(1 - \frac{1}{2}dp\right)$$

is actually not bad, and optimizing our $p$ (best at $p = 1/d$) gets us the bound. □

### Problems

1. Let $G$ be a connected graph with $v$ vertices and $e$ edges such that $e < 2v$. Show that $G$ has an independent set of size at least $\frac{1}{4}(2n - m)$.
2. (Crossing lemma) A graph with $V$ vertices and $E$ edges is drawn in the plane. Show that, as long as $E \geq 4V$, there will be at least $\frac{E^3}{64V^2}$ pairs of edges that cross.
3. (IMO 2014, weak version) A set of lines in the plane is in general position if no two are parallel and no three pass through the same point. A set of lines in general position cuts the plane into regions, some of which have finite area; we call these its finite regions. Prove that there exists a constant $c > 0$ such that for any set of $n$ lines in general position it is possible to colour at least $c\sqrt{n}$ lines blue in such a way that none of its finite regions has a completely blue boundary.
4. (Korea 2016) Let $U$ be a set of $m$ triangles. Prove that there exists a subset $W$ of $U$ which satisfies the following.

   (i). The number of triangles in $W$ is at least $0.45m^{\frac{4}{5}}$

   (ii) There are no points $A, B, C, D, E, F$ such that triangles $ABC$, $BCD$, $CDE$, $DEF$, $EFA$, $FAB$ are all in $W$.

5. (RMM 2019 Q3) Given any positive real number $\varepsilon$, prove that, for all but finitely many positive integers $v$, any graph on $v$ vertices with at least $(1+\varepsilon)v$ edges has two distinct simple cycles of equal lengths.

## 1.5 Probability is everywhere!

Actually, there are many situations that can be secretly interpreted in probabilistic terms...

**Example.** (APMO 1999) Let $a_1, a_2, \ldots$ be a sequence of real numbers satisfying $a_{i+j} \le a_i + a_j$ for all $i, j = 1, 2, \ldots$. Prove that

$$a_1 + \frac{a_2}{2} + \frac{a_3}{3} + \cdots + \frac{a_n}{n} \ge a_n$$

for each positive integer $n$.

*Solution.* Let $\mathrm{Cycle}(\pi, k)$ be the number of $k$-cycles in $\pi$. It is easy to get that $\sum_{k=1}^{n} \mathrm{Cycle}(\pi, k) = n$ for any permutation $\pi$. Now we let $\pi$ be a random permutation. Note that a fixed $k$-cycle appears in $\pi$ with probability $\frac{1}{n(n-1)\ldots(n-k+1)}$ and there are $(k-1)!\binom{n}{k}$ $k$-cycles in total, so $\mathbb{E}[\mathrm{Cycle}(\pi, k)] = \frac{1}{k}$. Hence:

$$\begin{aligned}
\mathbb{E}\left[a_n\right] &= \mathbb{E}\left[a_{\mathrm{Cycle}(\pi,1)\cdot 1 + \mathrm{Cycle}(\pi,2)\cdot 2 + \ldots}\right] \\
&\le \mathbb{E}\left[\mathrm{Cycle}(\pi,1)\cdot a_1 + \mathrm{Cycle}(\pi,2)\cdot a_2 + \ldots\right] \qquad \text{(applying } a_{i+j} \le a_i + a_j \text{ repeatedly)} \\
&= a_1 \mathbb{E}\left[\mathrm{Cycle}(\pi,1)\right] + a_2 \mathbb{E}\left[\mathrm{Cycle}(\pi,2)\right] + \ldots + a_n \mathbb{E}\left[\mathrm{Cycle}(\pi,n)\right] \\
&= a_1 + \frac{a_2}{2} + \frac{a_3}{3} + \cdots + \frac{a_n}{n}
\end{aligned}$$

□

### Problems

1. (CGMO 2012) Let $a_1, a_2, \ldots, a_n$ be non-negative real numbers. Prove that

$$\frac{1}{1+a_1} + \frac{a_1}{(1+a_1)(1+a_2)} + \frac{a_1 a_2}{(1+a_1)(1+a_2)(1+a_3)} + \cdots + \frac{a_1 a_2 \cdots a_{n-1}}{(1+a_1)(1+a_2)\cdots(1+a_n)} \le 1.$$

2. Let $x_1, x_2, ..., x_n, y_1, y_2, ..., y_n$ be positive reals where $x_i + y_i = 1$, and $m, n$ be natural numbers. Show that

$$(1 - x_1 x_2 ... x_n)^m + (1 - y_1^m)(1 - y_2^m)...(1 - y_n^m) \ge 1$$

3. (PIEs)

   (a) Let $A_1, ..., A_n$ be sets. Pick $x \in A_1 \cup A_2 \cup ... \cup A_n$ uniformly at random, then let $I_k = \mathbb{I}\left[x \in A_k\right]$. Show that for any $S \subseteq \{1, 2, ..., n\}$

$$\frac{\left|\bigcap_{s\in S} A_s\right|}{|A_1 \cup A_2 \cup ... \cup A_n|} = \mathbb{E}\left[\prod_{s \in S} I_s\right]$$

   (b) Derive the Principle of Inclusion-Exclusion:

$$|A_1 \cup A_2 \cup ... \cup A_n| = \sum_{S \subseteq \{1,...,n\}} (-1)^{|S|} \left|\bigcap_{s\in S} A_s\right|$$

   (c) (∗) If $A_1, A_2, ..., A_n$ are non-empty, prove the following quantity is always non-negative:

$$\sum_{i=1}^{n} \frac{1}{|A_i|} - \sum_{1\le i<j\le n} \frac{2}{|A_i \cup A_j|} + \sum_{1 \le i<j<k\le n} \frac{3}{|A_i \cup A_j \cup A_k|} + ... + (-1)^n \frac{1}{|A_1 \cup ... \cup A_n|}$$

4. (a) Consider $X, Y$, which are two identically distributed and independent random vectors in $\mathbb{R}^2$. Denote $\mu = \mathbb{E}[X]$ (does this make sense?). Show that the following quantities are equal:

$$\mathbb{E}\left[|X-\mu|^2\right] = \mathbb{E}\left[|X|^2\right] - |\mu|^2 = \frac{1}{2}\mathbb{E}\left[|X-Y|^2\right]$$

This quantity is also called the *variance* of $X$.

(b) Prove Stewart's theorem: in $\triangle ABC$, let $D$ be a point on side $\overline{BC}$. Let $d = AD$, $m = BD$, and $n = DC$. Then

$$mb^2 + nc^2 = ad^2 + amn.$$

5. (ISL 2006 C3) 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 let $b(P)$ be the number of points of $S$ which are outside $P$. Prove that for every real number $x$ :

$$\sum_P x^{a(P)}(1-x)^{b(P)} = 1$$

where the sum is taken over all convex polygons with vertices in $S$.

*Note: a line segment, a point, and the empty set are considered to be convex polygons of* $2, 1,$ *and 0 vertices, respectively.*

6. Suppose $a, b, c$ are positive real numbers such that for every positive integer $n$, $\lfloor an \rfloor + \lfloor bn \rfloor = \lfloor cn \rfloor$.

Prove that at least one of $a, b, c$ is an integer.

7. (Kürschák 2003) Prove that the following inequality holds with the exception of finitely many positive integers $n$:

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \gcd(i,j) > 4n^2.$$

(Hint: $\gcd(i,j) \geq \sum_p p \cdot \mathbb{I}\left[p \mid (i,j)\right]$)

## 1.6 Misc Problems

Probability doesn't always immediately destroy a problem. Here are some problems where you need to consider the right approach (in addition to using probability at some point):

1. (a) (Caro-Wei Theorem). Consider a graph $G$ with vertex set $V$. Prove that one can find an independent set with size at least

$$\sum_{v \in V} \frac{1}{\deg v + 1}$$

(b) (Kömal). Show that one can find a subset with size at least

$$\sum_{v \in V} \frac{2}{\deg v + 1}$$

whose induced subgraph contains no cycles.

2. Let $G$ be a graph with $m$ edges. Prove that $G$ has a bipartite subgraph with $m/2 + c\sqrt{m}$ edges, for some constant $c > 0$.

3. (USAMO 2010, modified) A blackboard contains $n$ pairs of nonzero integers. Suppose that for each positive integer $k$ at most one of the pairs $(k, k)$ and $(-k, -k)$ is written on the blackboard. A student erases some of the $2n$ integers, subject to the condition that no two erased integers may add to 0. Show that the student can always erase at least one integer from at least
$$\left(\frac{\sqrt{5}-1}{2}\right) n$$
pairs.

4. (ISL 2017 A5, bound only) Let $x_1, x_2, ..., x_n$ be real numbers. Show that
$$\sum_{1 \le i < j \le n} x_i x_j \ge \frac{n-1}{2} \min_{\sigma} \left\{ \sum_{i=1}^{n-1} x_{\sigma(i+1)} x_{\sigma(i)} \right\}$$
where $\sigma$ ranges across all permutations of $\{1, 2, ..., n\}$.

5. (ELMO 2015) Let $m, n, k > 1$ be positive integers. For a set $S$ of positive integers, define $S(i, j)$ for $i < j$ to be the number of elements in $S$ strictly between $i$ and $j$. We say two sets $(X, Y)$ are a fat pair if
$$X(i, j) \equiv Y(i, j) \pmod{n}$$
for every $i, j \in X \cap Y$. (In particular, if $|X \cap Y| < 2$ then $(X, Y)$ is fat.)

If there are $m$ distinct sets of $k$ positive integers such that no two form a fat pair, show that $m < n^{k-1}$.

## 1.7 Optional Reading: Entropy

*Based of my final presentation for Math 159 at Stanford.*

*Motivation.* How can we measure the *surprise* of an event? If we assume this only depends on the probability of the event $p$, we want to find $S : [0, 1] \to \mathbb{R}_{\ge 0}$ satisfying the following:

- $S(1) = 0$ (there is no surprise on seeing a certain event)
- $S$ is strictly decreasing (rare events have more surprise)
- $S$ is continuous
- $S(pq) = S(p) + S(q)$ (given two independent events $E, F$, the surprise from $E \cap F$ should be the surprise from $E$ and the surprise of $F$ conditioned on $E$.)
- $S(1/2) = 1$ (normalizing condition)

In fact, the unique function $S$ that satisfies all of the above is $S(p) = -\log_2(p)$ [1]. With that, we can define entropy to be the expected surprise:

**Definition.** The **entropy** $H[X]$ of a random variable $X$ is the expected amount of surprise from revealing $X$: i.e.
$$H[X] = \sum_x -\mathbb{P}[X = x] \log \mathbb{P}[X = x] = \mathbb{E}_X[-\log \mathbb{P}[X]].$$ [2]

where $x$ varies over the range of $X$. (This can also be interpreted as the *expected amount of information* contained in $X$, or a measure of randomness of $X$.)

This definition generalizes in a straightforward way for the **joint entropy** of multiple random variables $H[X_1, ..., X_n]$ (by considering the tuple $(X_1, ..., X_n)$ as a random variable). For example, for two variables $X, Y$ ranging over $(x, y)$:
$$H[Y|X] = \sum_{(x,y)} -\mathbb{P}[X = x, Y = y] \log \mathbb{P}[X = x, Y = y]$$

[1] All logs will be base 2 from here on out.

[2] Inside $\mathbb{E}_X[\cdot]$, we will use $\mathbb{P}[X]$ to denote $\mathbb{P}[X = x]$.

The **conditional entropy** of $Y$ given $X$ is

$$H[Y|X] := \mathbb{E}_X\left[H[Y_X]\right]$$

where $Y_X$ is a random variable identically distributed to $Y$ conditioned on $X$: i.e.

$$\mathbb{P}\left[Y_{X=x} = y\right] = \mathbb{P}\left[Y = y|X = x\right] \qquad \text{for all } y$$

Here are some properties of entropy you should try to prove:

1. For a discrete random variable $X$ taking $n$ possible values, $H[X] \le \log n$. In particular, the entropy is maximized when $X$ is uniformly distributed.
2. Joint and conditional entropies are related by
$$H[X,Y] = H[X] + H[Y|X]$$
3. $(*)$ Revealing information decreases entropy, i.e.
$$H[X|Y,Z] \le H[X|Y]$$
In particular, equality holds when $Z$ is independent of $(X,Y)$.
4. Entropy is subadditive, i.e.
$$H[X,Y] \le H[X] + H[Y]$$

### 1.7.1 Counting with Entropy

Entropy can used to count objects! Specifically, there are two main approaches to estimate the size of a set $S$:

*(1) Select a random element $X$ of the set $S$ by some process, then estimate $H[X]$. By property 1, $H[X] \le \log|S|$ which is a lower bound.*

**Example.** (CTST 2018, rephrased) Let $G(U \cup V, E)$ be a bipartite graph with $|U| = m$ and $|V| = n$. If $P$ is the set of length 3 paths (edges may be repeated), then

$$|P| \ge \frac{|E|^3}{mn}$$

*Solution.* Let $e_1, e_2, e_3$ be the edges of the path, randomly selected as follows: $e_2 = uv$ is uniformly selected among all edges, then $e_1$ is uniformly selected among all edges at $u$ and $e_3$ is uniformly selected among all edges at $v$ independent of $e_1$.

First we compute $H[e_1|e_2]$:

$$\begin{aligned} H[e_1|e_2] &= \sum_{u\in U} \frac{\deg u}{|E|} \log(\deg u) \\ &\ge \frac{m}{|E|}\left(\frac{|E|}{m}\log\left(\frac{|E|}{m}\right)\right) = \log\left(\frac{|E|}{m}\right) \end{aligned}$$

Similarly $H[e_3|e_2] \ge \log\left(\frac{|E|}{n}\right)$. Then:

$$\begin{aligned} \log|P| &\ge H[e_1, e_2, e_3] \\ &= H[e_2] + H[e_1|e_2] + H[e_3|e_2] \\ &\ge \log\left(\frac{|E|^3}{mn}\right) \end{aligned}$$

□

*(2) Select $X$ from $S$ uniformly at random, then bound $H(X)$ from above using the properties. Since $\log|S| = H(X)$ is bounded above, this is an upper bound.*

**Example.** (IMO 1992 / Discrete Loomis-Whitney) Let $S$ be a finite subset of $\mathbb{Z}^3$, and let $S_{xy}$ be the projection of $S$ onto the $xy$-plane (with $S_{yz}, S_{zx}$ defined similarly). Then:

$$|S|^2 \leq |S_{xy}| \cdot |S_{yz}| \cdot |S_{zx}|$$

*Solution.* First we show

$$2H[X, Y, Z] \leq H[X, Y] + H[Y, Z] + H[Z, X]$$

Indeed,

$$\begin{aligned} H[X, Y] + H[Y, Z] + H[Z, X] - 2H[X, Y, Z] &= H[X, Y] - H[X|Y, Z] - H[Y|X, Z] \\ &\geq H[X, Y] - H[X|Y] - H[Y] = 0 \end{aligned}$$

Now select $(P_x, P_y, P_z) \in S$ uniformly at random, then

$$\begin{aligned} 2\log|S| &= 2H[P_x, P_y, P_z] \\ &\leq H[P_x, P_y] + H[P_y, P_z] + H[P_z, P_x] \\ &\leq \log|S_{xy}| + \log|S_{yz}| + \log|S_{zx}| \end{aligned}$$

which is exactly log of the inequality we needed to show. □

**Problems**

1. Given finite sets $X, A_1, ..., A_{n-1}$ and functions $f_i : X \to A_i$, a vector $(x_1, ..., x_n) \in X^n$ is called nice if $f_i(x_i) = f_i(x_{i+1})$ for each $i = 1, 2, .., n-1$. Show that the number of nice vectors is at least
$$\frac{|X|^n}{\prod_{i=1}^{n-1} |A_i|}$$

2. Let $\mathcal{F}$ be a family of graphs on the labeled set of vertices $\{1, 2, ..., t\}$ and suppose that for any two members of $\mathcal{F}$ there is a triangle contained in both of them. Then
$$|\mathcal{F}| < \frac{1}{4} 2^{\binom{t}{2}}.$$

3. ($*$) (Bregman) Let $A$ be an $n \times n$ matrix whose entries are either 0 or 1, and let $r_i$ be the number of 1's in row $i$. Show that the number of ways to select $n$ 1's, no two from the same row or column, is at least
$$\prod_{i=1}^{n} (r_i!)^{1/r_i}$$