The Probabilistic Method

Using probabilistic techniques to prove deterministic results

Presented by: Gadi Aleksandrowicz

May 3, 2011

Outline

Outline

A silly example

- A coin is tossed.
- We know that there's a 50% chance to get "heads".
- What deterministic information does this knowledge give us?
- That the coin *has* a "heads" side, and a "tails" side (is not a total fake)

The basic idea

- Suppose we want to prove the existence of a combinatorial object.
- For example — a graph that has some desired properties.
- We define a probability space on *all* graphs, and show that a graph with the desired properties has a positive probability in that space.
- This proves (nonconstructively) the existence of at least one graph with the desired properties.
- What is the difference between this and simple counting arguments?
- It is possible to use many results and ideas from probability theory in order to handle the "show" part.

History

- First usage: Szele, 1943 (we'll see what he did soon enough).
- Main developer: Paul Erdős (first used the method in 1947)
- Initially used in Graph Theory, today used in many different fields: Number Theory, Complexity Theory, Game Theory, Computational Geometry. . .

Expectation

- Given a random variable X, taking values in $\mathbb{N}$, define $\mathrm{E}[X] \triangleq \sum_{n=0}^{\infty} n \cdot \mathrm{P}[X = n]$.
- "Linearity of Expectation": If $X_1, \ldots, X_n$ are random variables (not necessarily independent), then $\mathrm{E}[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} \mathrm{E}[X_i]$.
- If $\mathrm{E}[X] = A$ then there exists ω_1, ω_2 in the probability space such that

$$X(\omega_1) \leq A$$
$$X(\omega_2) \geq A$$

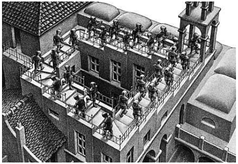
(Why?)

Szele's Result

- A *Tournament* is a directed graph such that for every $x, y \in V$ either $(x, y) \in E$ or $(y, x) \in E$ but not both.
- The idea: every two vertices play a match; $(x, y) \in E$ means x won.

Theorem

(Szele, 43)
There is a tournament T with n players and at least $n!2^{-(n-1)}$ Hamiltonian paths.



Szele's result (cont.)

Proof.

- Choose uniform distribution on all n-vertex tournaments.
- X is the random variable counting Hamiltonian paths.
- σ is a permutation on V. It defines a Hamiltonian path if and only if $(\sigma(i), \sigma(i+1)) \in E$ for all $1 \le i < n$.
- X_σ is the indicator of "σ defines a Hamiltonian path".
- $\mathrm{P}[X_\sigma] = 2^{-(n-1)}$ (why?)
- $X = \sum_\sigma X_\sigma$ (why?) so $\mathrm{E}[X] = \sum_\sigma \mathrm{E}[X_\sigma] = n!2^{-(n-1)}$.
- There is a tournament for which X is equal to at least $\mathrm{E}[X]$.

$\square$

Satisfiability problem

- Given boolean variables $x_1, \ldots, x_n$, a 3SAT formula is a CNF formula with 3 different literals in each of its M clauses.
- Example: $\varphi = (x_1 \vee \overline{x_2} \vee x_3) \wedge (\overline{x_3} \vee x_4 \vee \overline{x_2})$.
- Problem: is a given 3SAT formula satisfiable? (NP-complete).
- Easier problem: Are at least half the clauses satisfiable (simultaneously)? what about three-quarters?

Theorem

For every 3SAT formula there is an assignment which satisfies $\frac{7}{8}$ of the clauses.

Satisfiability problem (cont.)

Proof.

- Choose a random assignment.
- For clause i $(1 \leq i \leq M)$ define X_i as the indicator of the satisfiabilty of i.
- $X = \sum_{i=1}^{M} X_i$ is the number of satisfied clauses.
- $\mathrm{P}[X_i = 1] = \frac{7}{8}$ (why?)
- $\mathrm{E}[X] = \sum_{i=1}^{M} \mathrm{E}[X_i] = \frac{7}{8} M$.
- Therefore, there is an assignment satisfying at least $\frac{7}{8} M$ clauses.

□

Outline

Heilbronn's triangle problem

- S is a set of n points in the unit square.
- Define $T(S)$ as the minimum area of a triangle whose vertices are three distinct points from S.
- Define $T(n) = \max_{|S|=n} T(S)$.
- $T(n)$ is known exactly only for $n \leq 6$.

Asymptotic bounds

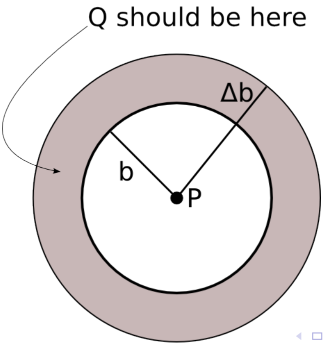
- Heilbronn conjectured that $T(n) = O\left(n^{-2}\right)$.
- However, in 1982 it was proved that $T(n) = \Omega\left(n^{-2}\log n\right)$.
- We show by a simpler argument: $T(n) = \Omega\left(n^{-2}\right)$.

Theorem

For every n there is a set S of n points in the unit square such that $T(S) \geq \left(100n^2\right)^{-1}$.

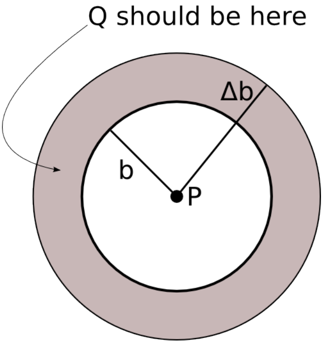
Part 1: The area of a "random triangle"

- Let P, Q, R be independently, uniformly selected from U.
- Let $\mu = \mu(PQR)$ be the area of the triangle PQR.
- We bound $\mathrm{P}[\mu \leq \varepsilon]$ using some calculus.
- Let x be the distance between P and Q.
- Then we have $\mathrm{P}[b \leq x \leq b + \Delta b] \leq \pi(b + \Delta b)^2 - \pi b^2$



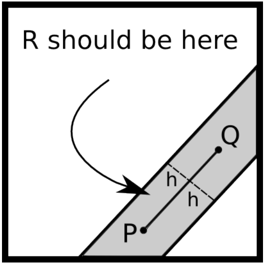
Part 1: The area of a "random triangle" (cont.)

- When we take db to be infinitesimally small, we get:
- $\mathrm{P}[b \leq x \leq b+db] \leq \pi(b+db)^2 - \pi b^2 = 2\pi b db.$



Part 1: The area of a "random triangle" (cont.)

- Given P, Q at distance b, where can we place R while keeping $\mu \leq \varepsilon$?
- Let h be the distance of R from the line PQ.
- The area of PQR is $\frac{hb}{2}$.
- Therefore we have $h \leq \frac{2\varepsilon}{b}$.
- This means R lies in a strip of width $\frac{4\varepsilon}{b}$ and length at most $\sqrt{2}$.



Part 1: The area of a "random triangle" (cont.)

- Therefore the probability for a triangle with $\mu \leq \varepsilon$ for a given b is at most $\frac{4\sqrt{2}\varepsilon}{b}$.
- We have that the total probability of a triangle with $\mu \leq \varepsilon$ is bounded by:

$$\int_0^{\sqrt{2}} (2\pi b)\left(\frac{4\sqrt{2}\varepsilon}{b}\right) db = 16\pi\varepsilon$$

Part 2: Choosing our points

- We use an "alteration" trick.
- We won't prove directly that our desired set exists, but the existence of a "slightly flawed" set.
- Then we'll show that the "flaws" can be fixed by removing some of the "bad" elements of the set.
- So, instead of choosing n points, we choose $2n$ points uniformly, $P_1, P_2, \ldots, P_{2n}$.

Part 2: Choosing our points

- Let X be the number of "bad" triangles - $P_iP_jP_k$ with area less than $\left(100n^2\right)^{-1}$.
- As before, we write X as a sum of indicators, for each triplet i, j, k.
- The probability that $P_iP_jP_k$ is of area less than $\varepsilon = \left(100n^2\right)^{-1}$ is smaller than $\frac{16\pi}{100n^2} < \frac{16\cdot 3.5}{100n^2} < 0.6n^{-2}$.
- And so we have $\mathrm{E}[X] \leq \binom{2n}{3}\left(0.6n^2\right) = \frac{6(2n)(2n-1)(2n-2)}{6\cdot 10n^2} < \frac{(2n)^3}{10n^2} = \frac{8n^3}{10n^2} < n$

Part 3: Alteration

- Since $\mathrm{E}[X] < n$, there is a choice of points such that $X < n$, i.e. no more than n "bad" triangles.
- Choose a point P_i from each of the "bad" triangles and remove it from the set.
- The same point may be deleted more than once, but this only helps.
- We end up with a set of at least n points, which induces no bad triangles.

Outline

Introduction

- How large can a set of points in $\mathbb{R}^d$ be such that:
- Every angle determined by a triplet of points from the set is strictly less than 90°.
- Danzer and Grünbaum proved in 1962 that an upper bound is 2^d.
- The also conjectured that a much better upper bound is $2d-1$.
- This was disproven in 1983 by Erdős and Füredi using the probablistic method.

Theorem

For every $d \geq 1$ there is a set of at least $\left\lfloor \frac{1}{2}\left(\frac{2}{\sqrt{3}}\right)^d \right\rfloor$ points in $\mathbb{R}^d$ satisfying the 90° condition.

Basic observations

- We'll choose points from the vertices of the d dimensional unit cube.
- First observation - the angles they can form are at most 90°, so it suffices to find a condition as to when they are exactly 90° (proof - next slide).
- We can think of the vertices as vectors in $\mathbb{F}_2^d$.
- Associate with $(a_1, \dots, a_d)$ the set $A = \{i \mid 1 \leq i \leq d, a_i = 1\}$.
- Second observation (proof - the slide after the next) - the vertices a, b, c determine right angle at c if and only if for the corresponding sets A, B, C :

$$A \cap B \subseteq C \subseteq A \cup B$$

Proof for the 1st observation

- Remember that in an inner product space (such as $\mathbb{R}^d$) one can define angle between vectors u, v by:

$$\cos(\theta) = \frac{\langle v, u \rangle}{\|v\| \|u\|}$$

- Therefore $0^\circ \leq \theta \leq 90^\circ$ if $\langle v, u \rangle$ is nonnegative.

Proof for the 1st observation (cont.)

- For a, b, c, to determine the angle at c, one consideres $\langle (a-c),(c-b)\rangle$. We have:

$$\begin{aligned}
\langle (a-c),(b-c)\rangle &= \langle a,b\rangle - \langle a,c\rangle - \langle c,b\rangle + \langle c,c\rangle \\
&= \langle a,b\rangle - \langle a+b-c,c\rangle = \sum_{i=1}^{d} a_i b_i - \sum_{i=1}^{d} (a_i+b_i-c_i)\,c_i \\
&= \sum_{i\in A\cap B} 1 - \sum_{i\in C} (a_i+b_i-1) \\
&= \sum_{i\in A\cap B} 1 - \sum_{i\in A\cap B\cap C} 1 + \sum_{i\in C-(A\cup B)} 1 \\
&= |A\cap B| - |A\cap B\cap C| + |C-(A\cup B)|
\end{aligned}$$

- Since $A\cap B\cap C \subseteq A\cap B$, this is obviously nonnegative.

Proof for the 2nd observation

- By the previous calculation, we have that $\theta = 90^\circ$ (and so, $\cos(\theta) = 0$) if and only if:

$$|A \cap B| - |A \cap B \cap C| + |C - (A \cup B)| = 0$$

- Since $|A \cap B \cap C| \leq |A \cap B|$ we have that this equation holds if and only if:

1. $|A \cap B| = |A \cap B \cap C| \iff A \cap B = A \cap B \cap C \iff A \cap B \subseteq C$
2. $|C - (A \cup B)| = 0 \iff C - (A \cup B) = \emptyset \iff C \subseteq A \cup B$

And we are done.

What now?

- We remain with the following goal:
- Prove the existance of a set of $m = \left\lfloor \frac{1}{2}\left(\frac{2}{\sqrt{3}}\right)^d \right\rfloor$ points such that no three points satisfy $A \cap B \subseteq C \subseteq A \cup B$.
- As with Heilbronn's triangle problem, we use the "alteration" technique - choosing "too many" points and proving the problems can be fixed by removing a "small" amount of points.
- Therefore, we choose randomly and independently $2m$ vectors in $\mathbb{F}_2^d$.
- Given a, b, c, what is the probability that they satisfy the set condition?
- Turns out it's exactly $\left(\frac{3}{4}\right)^d$.

Proof of the probability

- When do we have $A \cap B \subseteq C \subseteq A \cup B$?
- Only if none of the following conditions hold for every $1 \leq i \leq d$:

1. $a_i = b_i = 0$ and $c_i = 1$
2. $a_i = b_i = 1$ and $c_i = 0$

- There are 8 possible choices of a_i, b_i, c_i, so we have $\frac{1}{4}$ probability this will happen for i.
- Therefore, the probability it will not happen for any i is $\left(\frac{3}{4}\right)^d$.

Using expectation

- There are at most $3 \cdot \binom{2m}{3}$ angles created by points in the set (maybe less because some points might repeat).
- Thus, using the standard expectation calculation we have that the expected number of right angles is:

$$3\binom{2m}{3}\left(\frac{3}{4}\right)^d = \frac{(2m)(2m-1)(2m-2)}{2}\left(\frac{3}{4}\right)^d < \frac{(2m)^3}{2}\left(\frac{3}{4}\right)^d$$

$$\leq \frac{1}{2}\left(\frac{8}{\sqrt{3^3}}\right)^d\left(\frac{3}{4}\right)^d = \frac{1}{2}\left(\frac{8\cdot 3}{4\cdot\sqrt{3^3}}\right)^d = \frac{1}{2}\left(\frac{2}{\sqrt{3}}\right)^d = m$$

- Therefore, there is a set of $2m$ points for which at most m of the angles are 90°.

Summing it up

- As with Heilbronn, we remove an offending point from each right angle triplet.
- We remain with m points and no right angles.
- The points are all distinct, since if $a = c$ then $A \cap B \subseteq C \subseteq A \cup B$ (for any b).

Outline

Statement

- Suppose we have n mutually independent events, each holds with probability at least $p > 0$.
- The probability that *all* of the events hold simultaneously is at least $p^n > 0$.
- Therefore, knowing that every event has a chance to happen tells us there's a chance all of them will happen.
- Sounds trivial? It is... But what if the events are not independent? Is there a "weak" dependency such that a similar result holds?
- Given events $A_1, \ldots, A_n$, define a digraph $D = (V, E)$ to model their dependencies: A_i is mutually independent of all the events $\{A_j | (i,j) \notin E\}$.

Statement (general case)

Theorem

(The Local Lemma: General Case)
If there exists real number $x_1, \ldots, x_n$ *such that* $0 \leq x_i < 1$ *and for all* i *we have* $P[A_i] \leq x_i \prod_{(i,j)\in E}(1-x_j)$*, then*
$P\left[\bigwedge_{i=1}^{n} \overline{A_i}\right] \geq \prod_{i=1}^{n}(1-x_i) > 0$

First, a lemma:

Theorem

For any $S \subseteq \{1, \ldots, n\}$*,* $|S| = s < n$ *and any* $i \notin S$ *we have*
$P\left[A_i | \bigwedge_{j\in S} \overline{A_j}\right] \leq x_i$

Proof of the lemma

- We prove by induction on $|S| = s$. For $s = 0$ it follows trivially from the Local Lemma's hypothesis.
- Given S, assume correctness for each set of size $s' < s = |S|$.
- Put $S_1 = \{j \in S | (i,j) \in E\}$ and $S_2 = S - S_1$. Then we have:

$$P\left[A_i | \bigwedge_{j \in S} \overline{A_j}\right] = \frac{\mathrm{P}\left[A_i \wedge \left(\bigwedge_{j \in S_1} \overline{A_j}\right) | \bigwedge_{k \in S_2} \overline{A_k}\right]}{\mathrm{P}\left[\bigwedge_{j \in S_1} \overline{A_j} | \bigwedge_{k \in S_2} \overline{A_k}\right]}$$

This follows from the general equation $\frac{\mathrm{P}[A \cap B | C]}{\mathrm{P}[B|C]} = \frac{\mathrm{P}[A \cap B \cap C]}{\mathrm{P}[B \cap C]} = \mathrm{P}\left[A | B \cap C\right]$

Proof of the lemma (cont.)

- $\mathrm{P}\left[A_i | \bigwedge_{j\in S} \overline{A_j}\right] = \frac{\mathrm{P}[A_i \wedge (\bigwedge_{j\in S_1} \overline{A_j}) | \bigwedge_{k\in S_2} \overline{A_k}]}{\mathrm{P}[\bigwedge_{j\in S_1} \overline{A_j} | \bigwedge_{k\in S_2} \overline{A_k}]}$
- We bound the numerator by noting that since A_i is independent of the events in S_2, we have:

$$\mathrm{P}\left[A_i \wedge \left(\bigwedge_{j\in S_1} \overline{A_j}\right) | \bigwedge_{k\in S_2} \overline{A_k}\right] \leq \mathrm{P}\left[A_i | \bigwedge_{k\in S_2} A_k\right]$$
$$= P[A_i] \leq x_i \prod_{(i,j)\in E} (1 - x_j)$$

- For the denominator we'll have to use the induction hypothesis in a technical way.

Bounding the denominator

- Write $S_1 = \{j_1, j_2, \ldots, j_r\}$.
- If $r = 0$ we are done (the denominator is 1). Otherwise:

$$\mathrm{P}\left[\overline{A_{j_1}} \wedge \overline{A_{j_2}} \wedge \cdots \wedge \overline{A_{j_r}} | \bigwedge_{l \in S_2} \overline{A_l}\right]$$

$$= \left(1 - \mathrm{P}\left[A_{j_1} | \bigwedge_{l \in S_2} \overline{A_l}\right]\right) \cdot \left(1 - \mathrm{P}\left[A_{j_2} | \overline{A_{j_1}} \wedge \bigwedge_{l \in S_2} \overline{A_l}\right]\right) \cdots$$

$$\cdots \left(1 - \mathrm{P}\left[A_{j_r} | \bigwedge_{l \in S_2} \overline{A_{j_1}} \wedge \cdots \wedge \overline{A_{j_{r-1}}} \wedge \overline{A_l}\right]\right)$$

$$\geq \prod_{t=1}^{r} (1 - x_{j_t}) \geq \prod_{(i,j) \in E} (1 - x_j)$$

- And so we have $\frac{\mathrm{P}[A_i \wedge (\bigwedge_{j \in S_1} \overline{A_j}) | \bigwedge_{k \in S_2} \overline{A_k}]}{\mathrm{P}[\bigwedge_{j \in S_1} \overline{A_j} | \bigwedge_{k \in S_2} \overline{A_k}]} \leq \frac{x_i \prod_{(i,j) \in E} (1 - x_j)}{\prod_{(i,j) \in E} (1 - x_j)} = x_i$

Finishing the proof

- Back to the local lemma itself - remember we want to bound $\mathrm{P}\left[\bigwedge_{i=1}^{n} \overline{A_i}\right]$.
- Remember also what we just proved: $\mathrm{P}\left[A_i | \bigwedge_{j \in S} \overline{A_j}\right] \leq x_i$.
- And so:

$$\mathrm{P}\left[\bigwedge_{i=1}^{n} \overline{A_i}\right] = (1 - \mathrm{P}[A_i])\left(1 - \mathrm{P}\left[A_2 | \overline{A_1}\right]\right) \cdots \left(1 - \mathrm{P}\left[A_n | \bigwedge_{i=1}^{n-1} \overline{A_i}\right]\right)$$
$$\geq \prod_{i=1}^{n} (1 - x_i)$$

The Local Lemma - symmetric case

Theorem

(The Local Lemma - symmetric case)
Let $A_1, \ldots, A_n$ be events such that A_i is mutually independent of all the other events except for at most d events, and that $P[A_i] \leq p$ for all events.
Denote $e = 2.71828183\ldots$ as usual.
If $ep(d+1) \leq 1$ holds, then $P\left[\bigwedge_{i=1}^{n} \overline{A_i}\right] > 0$

Proof.

We use the general local lemma with $x_i = \frac{1}{d+1}$ and using the known inequality $\left(1 - \frac{1}{d+1}\right)^d > \frac{1}{e}$ □

Simple use - 2-Colorability of hypergraphs

- A hypergraph $H = (V, E)$ is a set V of vertices, and set $E \subseteq 2^V$ of edges (a graph is the special case when all edges are of size 2).
- H is 2-colorable if there is a coloring of V by two colors such that there are no monochromatic edges.

Theorem

Let $H = (V, E)$ be a hypergraph in which every edge has at least k elements and each edge intersects at most d other edges. If $e(d+1) \leq 2^{k-1}$, then H is 2-colorable.

2-colorability of hypergraphs, cont.

Theorem

Let $H = (V, E)$ be a hypergraph in which every edge has at least k elements and each edge intersects at most d other edges. If $e(d+1) \leq 2^{k-1}$, then H is 2-colorable.

Proof.

Choose a random coloring. For $f \in E$, the probability of the event A_f in which f is monochromatic is $\mathrm{P}[A_f] = 2 \cdot 2^{-|f|} \leq 2^{-(k-1)}$. Every A_f is mutually independent of all other events save for at most d others. Now use the symmetric case of the local lemma. $\square$

Not-so-simple use - k-coloring of $\mathbb{R}$

- The following beautiful result is due to Erdős and Lovász (1975). In this paper they also proved the local lemma.
- A k-coloring of $\mathbb{R}$ is a function $c : \mathbb{R} \to \{1, \ldots, k\}$.
- We say that a subset $T \subseteq \mathbb{R}$ is multicolored if $c(T) = \{1, \ldots, k\}$.

Theorem

Let m and k be positive integers such that

$$e(m(m-1)+1)k\left(1-\frac{1}{k}\right)^m \leq 1$$

Then for any $S \subseteq \mathbb{R}$ of size m there is a k-coloring of $\mathbb{R}$ such that each translation $x + S, x \in \mathbb{R}$ is multicolored.

Proof, part 1: Using the local lemma

- First we prove the result not for all translations, but only translations from a finite set X.
- Fix X, and define $Y = \bigcup_{x \in X}(x+S)$. Randomally choose a coloring $c : Y \to \{1, \ldots, k\}$.
- Let A_x be the event "the set $x+S$ is not multicolored".
- $\mathrm{P}[A_x] \leq k\left(1-\frac{1}{k}\right)^m$ (union bound on "the color i does not participate in the coloring").
- Each A_x is independent of all other $A_{x'}$ unless $(x+S)\cap(x'+S) \neq \emptyset$. There are at most $m(m-1)$ such x'.
- By the symmetric case of the local lemma, we are done.

Proof, part 2: A reminder from topology

- For a fixed $x \in \mathbb{R}$, denote by C_x the set of all colorings of $\mathbb{R}$ such that $x + S$ is multicolored.
- What we saw: For every finite X, we have $\bigcap_{x \in X} C_x \neq \emptyset$.
- We want to show: $\bigcap_{x \in \mathbb{R}} C_x \neq \emptyset$.
- This reminds us (?) of a well-known result in point-set topology: The finite intersection property definition of compact spaces.

Theorem

A topological space X is compact if and only if for every family of closed sets such that the intersection of finitely many sets from the family is nonempty, the intersection of the whole family is nonempty.

Showing the space is compact

- Remember that our space is the space of all colorings of $\mathbb{R}$, i.e. functions $c : \mathbb{R} \to \{1, \dots, k\}$.
- We can think of this space as the product space $\{1, \dots, k\}^{\mathbb{R}}$ of discrete k-points spaces.
- Each discrete space $\{1, \dots, k\}$ is compact, being finite.
- Hence, by Tychonoff's theorem, we have that $\{1, \dots, k\}^{\mathbb{R}}$ is compact (with respect to the product topology).

Showing that each set is closed

- We wish to show that each C_x is closed (in the standard product topology).
- Recall C_x is the set of all colorings of $\mathbb{R}$ such that $x+S$ is multicolored.
- The complement of C_x is $\bigcup_{1\le i\le k} D_x^i$ where D_x^i is the set of colorings of $\mathbb{R}$ for which $i \notin c(x+S)$ for $c \in D_x^i$.
- In other words, $D_x^i = \{1,\ldots,i-1,i+1,\ldots,k\}^{(x+S)} \times \{1,\ldots,k\}^{\mathbb{R}-(x+S)}$.
- By definition, this implies D_x^i is open, hence C_x is closed.

For Further Reading I

N. Alon, J. Spencer,
The Probabilistic Method, 2nd edition.
New York: Wiley-Interscience, 2000.