Needles, haystacks and optimal designs

Richard P. Brent

Australian National University
and University of Newcastle

15 October 2014

Introduction

In the abstract, I said that I would consider

- primality testing,
- finding bounds on Ramsey numbers, and
- finding almost-optimal designs.

These problems do not appear to have much in common, so why did I choose them?

The reason is that, in all three cases, probabilistic ideas are very helpful for their solution.

The key idea - Probability!

Lest men suspect your tail untrue,
Keep probability in view

John Gay, Fables (1727)

We can often use *probability* to our advantage in both mathematical proofs and algorithms, even when the problem does not seem to involve anything of a statistical or probabilistic nature.

I am sure that Professor Moyal would have appreciated this, since he was interested in both mathematics and statistics.

Primality testing

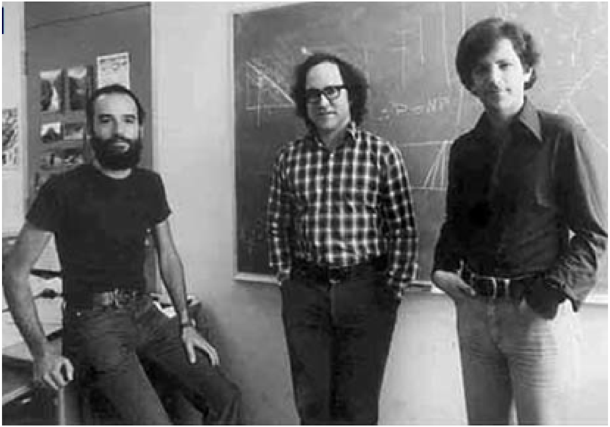
> *The problem of distinguishing prime numbers*
> *from composite numbers $\cdots$*
> *is known to be one of the most important*
> *and useful in arithmetic.*
>
> *Carl Friedrich Gauss (1801)*

Nowadays primality testing is even more important than it used to be, because several cryptographic systems need large prime numbers for their security.

An example is the widely-used Rivest-Shamir-Adleman (RSA) cryptosystem (1977). To generate a public key for RSA requires two large, “random” primes p and q, which are kept secret. The product $N = pq$ is made public – it is the “public key”.

The encryption and decryption algorithms only work correctly if p and q are primes.

SRA in 1977



Primality testing

Is this number prime?

520681969888053628793815844323088253114267574462263 91

No!

Proof. It is the product of

$$21532478004187212363498 6089$$

and

$$24181237746384863420617 6319.$$

(Multiply them to verify it! :-) □

How to test primality?

An obvious way to test if a given integer n is prime is to divide by all possible factors $2, 3, \ldots, \lfloor\sqrt{n}\rfloor$. If one of them divides n exactly, then n is composite. Otherwise n must be prime.

Unfortunately, this “brute force” algorithm is impractical – it takes “exponential time”.

“Exponential time” means an exponential function of the *length* of the input n. Usually the length ℓ is defined to be the number of *bits* required to encode n, i.e. $\ell \approx \log_2(n)$.

You can see that the *worst case time* for the brute force algorithm is about $2^{\ell/2}$, which is exponential in ℓ.

The *average time* (averaging over all ℓ-bit integers) is also exponential in ℓ.

Two theorems stated by Fermat

The "little" theorem. [Fermat (1640)]

If n is prime and $0 < b < n$, then

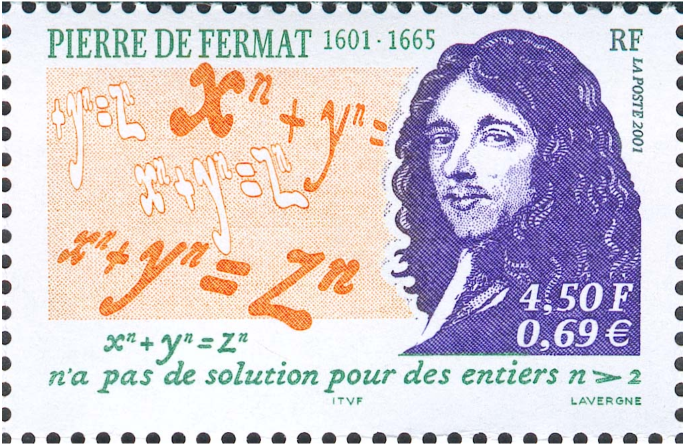
$$b^{n-1} = 1 \bmod n.$$

The "last" theorem. [Wiles and Taylor (1995)]

There are no positive integer solutions of

$$a^n + b^n = c^n$$

if n is an integer greater than 2.


PIERRE DE FERMAT 1601 · 1665
RF
LA POSTE 2001
$x^n + y^n = z^n$
4,50 F
0,69 €
$x^n + y^n = z^n$
n'a pas de solution pour des entiers n > 2
ITVF
LAVERGNE

Can we use Fermat's little theorem?

Suppose *n* is a given (large) integer, and we want to determine if *n* is prime or composite.

We could choose some integer *b*, where $1 < b < n$, and compute

$$r = b^{n-1} \bmod n.$$

This can be done efficiently, in "polynomial time", by a "square and multiply mod *n*" algorithm that uses the binary representation of *n*.

If *n* is prime, then we know, from Fermat's theorem, that $r = 1$. Hence, if the computed $r \neq 1$, and we didn't make an error in our arithmetic, *n* must be composite.

However, a primality test based on this does not work.

If $r = 1$, we can't be *sure* that *n* is prime.

Carmichael numbers

There are some (not too rare) composite integers n, called *Carmichael numbers*, for which $r = 1$ for all b relatively prime to n. A primality test using Fermat's little theorem fails on Carmichael numbers.

For example, if $n = 561 = 3 \cdot 11 \cdot 17$ and b is any integer that is not divisible by $3, 11$ or 17, then

$$b^{n-1} = 1 \bmod n.$$

To see this, note that $p - 1$ divides $n - 1$ for $p \in \{3, 11, 17\}$. Thus, by Fermat's theorem, $b^{n-1} = 1 \bmod p$ for $p \in \{3, 11, 17\}$. If follows from the "Chinese remainder theorem" that $b^{n-1} = 1 \bmod n$.

The strong Fermat test

Suppose that $n - 1 = 2^k q$, where $k \geq 0$ and q is odd.
If n is prime and $0 < b < n$, then *either* (A)

$$b^q = 1 \bmod n$$

or (B) the sequence

$$S = (b^q, b^{2q}, b^{4q}, \ldots, b^{2^k q}) \bmod n$$

ends with 1, and the value just preceding the first appearance of 1 must be $-1 \bmod n$.

Conversely, if neither (A) nor (B) holds, then n must be composite.

We say that *b is a witness to the compositeness of n* because, given b, one can easily prove that n is composite (without finding any factors of n).

The Rabin-Miller probabilistic primality test

Rabin (1980) showed that, for any odd composite number $n > 4$, the number of witnesses to compositeness $b \in \{1, 2, 3, \ldots, n-1\}$ is *at least* $3(n-1)/4$.

This is true even if n is a Carmichael number!

Suppose $n > 4$ is composite and we choose 20 numbers $b_i \in \{1, 2, \ldots, n-1\}$ uniformly at random. The probability that *none* of the b_i is a witness to the compositeness of n is at most

$$(1 - 3/4)^{20} = 4^{-20} < 10^{-12}.$$

Thus, if we announce that n is composite if we find a witness, and n is prime otherwise, the probability that we make an error is less than one in 10^{12}.

Note that the errors are all in one direction – we may announce (with small probability) that a composite number is prime, but we never announce that a prime number is composite.

The AKS primality test

It was a surprise to most experts when, in 2002, a *deterministic polynomial-time* primality test was discovered by Agrawal, Kayal and Saxena (AKS).

Does this make the Rabin-Miller test obsolete? No!

Although the AKS test runs in polynomial time, the degree of the polynomial is quite large (originally 12, now reduced to about 6). Thus, the AKS test is *much slower* than one trial of the Rabin-Miller test. To test a 1024-bit number (about the size used in RSA) would take about 1000 years on a 1Ghz computer, whereas Rabin-Miller takes only a few seconds.

We could afford to use say 100 trials in the Rabin-Miller test, with probability of error less than 10^{-60}. This is considerably smaller than the probability that the AKS test would give the wrong answer because a cosmic ray passed through the computer and flipped a critical memory bit! Experience shows that such things, although rare, do happen.

Rabin, Miller, AKS

Michael Rabin in 1980; Gary L. Miller (date unknown);
Manindra Agrawal, Neeraj Kayal, and Nitin Saxena in 2002.

What about proofs?

The Rabin-Miller probabilistic primality test is an example of a *randomized* algorithm, also called a *Monte Carlo* algorithm. There are many other useful randomized algorithms that I don't have time to talk about today. A good introduction is the book *Randomized Algorithms* by Motwani and Raghavan (1995).

Probabilistic ideas can also be used to give rigorous mathematical proofs. Paul Erdős (1913–1996) was a pioneer in this area of mathematics and gave many such proofs.

I will discuss an example due to Erdős, then a recent example from my own research.

For many more examples and references, see the book *The Probabilistic Method* by Alon and Spencer (2008).

The party problem

Suppose there are at least six people at a party. Then,

- either there are three guests who all know each other,
- or there are three guests no pair of whom know each other.

This is sometimes called the *theorem on friends and strangers*.

George Szekeres apparently looking at a “bath” party in Budapest.

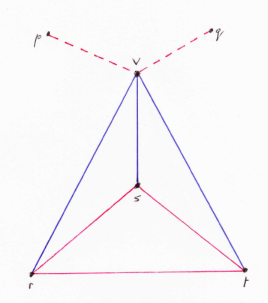
Translation into a problem in graph theory

Think of the people at a party as vertices in a graph. If two people, say u and v, know each other, draw a blue edge connecting vertices u and v in the graph. Otherwise, draw a red edge connecting vertices u and v.

Thus, if there are N people at the party, we obtain a complete graph K_N on N vertices, where the $N(N-1)/2$ edges are coloured blue or red.

The theorem on friends and strangers says that in any colouring of K_6 we can find a monochromatic K_3, i.e. a triangle whose edges are all the same colour.

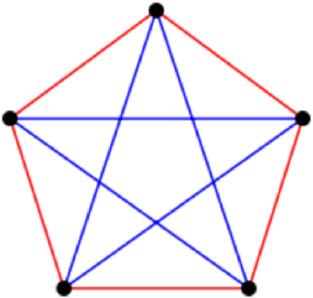
Proof



Suppose the edges of K_6 are coloured red and blue. Pick a vertex v. There are 5 edges incident to v, so at least 3 of them must be the same colour. Without loss of generality we can assume that these edges, connecting the vertex v to vertices r, s and t, are blue (see the illustration above).

If any of the edges (r, s), (s, t), (t, r) is blue, then there is an entirely blue triangle vrs, vst or vtr. If not, then those three edges are all red and we have an entirely red triangle rst. □

We need at least six people



This colouring of the complete graph K_5 does not have a monochromatic triangle. (In fact, each subgraph K_3 has two edges of one colour and one edge of the other colour.) Thus, the proof that we gave on the previous slide depends on having at least 6 vertices in the graph (or people at the party).

Ramsey's theorem and Ramsey numbers

Consider a *complete graph* K_N with N vertices. Suppose each edge is coloured red or blue.

Ramsey's theorem (1930) says that, for any positive integers r and s, there is a number $R(r, s)$ (now called a Ramsey number) such that, if $N \geq R(r, s)$, then K_N must have a subgraph K_r (with r vertices) that is coloured entirely red, *or* a subgraph K_s (with s vertices) that is coloured entirely blue.

Idea of proof [Erdős and Szekeres (1935)]: Use induction on $r + s$ to show that

$$R(r, s) \leq \binom{r+s-2}{r-1},$$

using

$$R(r, s) \leq R(r-1, s) + R(r, s-1). \quad \square$$

In particular, $R(3, 3) \leq \binom{4}{2} = 6$ (this gives another proof of the theorem on friends and strangers).

Frank Ramsey (1903–1930)

Upper and lower bounds on $R(k, k)$

I will consider the symmetric case $r = s = k$ say.
It is known that $R(2, 2) = 2$, $R(3, 3) = 6$, and $R(4, 4) = 18$.

The exact value of $R(k, k)$ is not known for $k > 4$.

For $k = 5$, all that we know is $43 \le R(5, 5) \le 49$ (although McKay and Radziszowski (1997) have given a convincing heuristic argument that $R(5, 5) = 43$).

Thus, it is interesting to obtain upper and lower bounds (as close together as possible) on $R(k, k)$.

From the inductive proof sketched above,

$$R(k, k) \le \binom{2k-2}{k-1} \sim \frac{4^{k-1}}{\sqrt{\pi k}} \text{ as } k \to \infty.$$

Thus, we have an upper bound that is exponential in k.

A lower bound on $R(k, k)$

There is no known deterministic construction that gives an exponential lower bound on $R(k, k)$. However, Erdős proved such a bound using the probabilistic method.

Theorem (Erdős, 1947). If

$$\binom{n}{k} < 2^{k(k-1)/2-1}$$

then $R(k, k) > n$.

Corollary. For all $k \geq 3$, we have $R(k, k) > 2^{k/2}$.

Proof of the theorem

Consider a random colouring of the edges of K_n, where each edge is coloured red or blue independently and with equal probability. For each induced subgraph G of k vertices, the probability that all $k(k-1)/2$ edges in G are monochromatic (i.e. all red or all blue) is $p = 2^{1-k(k-1)/2}$.

There are $\binom{n}{k}$ possible choices of G. Thus, the probability P that *some* such G is monochromatic satisfies

$$P \le p\binom{n}{k} < 1.$$

In other words, with positive probability $1 - P$, no such G is monochromatic. Thus, there must exist some colouring of K_n that has no monochromatic induced subgraph on k vertices. This implies that $R(k,k) > n$. □

Proof of the corollary

If $k \geq 3$ and $n = \lfloor 2^{k/2} \rfloor$, then

$$\begin{aligned}\binom{n}{k} \cdot 2^{1-k(k-1)/2} &< \frac{n^k}{k!} \cdot \frac{2^{1+k/2}}{2^{k^2/2}} \\ &\leq \frac{2^{1+k/2}}{k!} \cdot \frac{n^k}{2^{k^2/2}} \\ &\leq 1.\end{aligned}$$

Thus, $\binom{n}{k} < 2^{k(k-1)/2-1}$, and it follows from the Theorem that $R(k,k) > n$. Thus $R(k,k) \geq n+1 > 2^{k/2}$. □

Important points

Even if you didn't follow the details of the proofs, some points to note are:

- If the probability of something (e.g. a colouring with a certain property) is positive, then it must be possible.
  If it were impossible, then the probability would be zero.
- The condition that the probability is positive may be a rather unintuitive inequality that can be transformed into an easier-to-understand inequality (though this process may give away a little in precision).
- Since we were considering a finite graph K_n we could have given a proof that just involved a "counting argument" and did not mention probability. However, it would have been harder to understand and more likely to contain errors!

The upper and lower bounds on $R(k,k)$

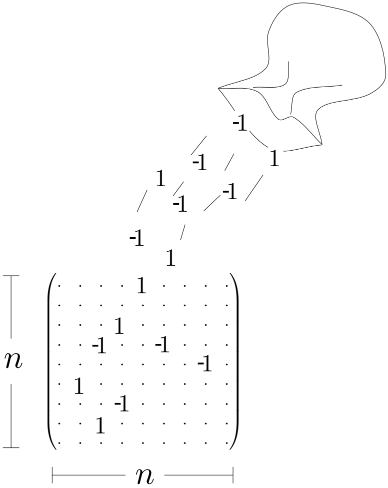
To summarise, we showed that, for $k \geq 3$,

$$2^{k/2} < R(k,k) \leq \binom{2k-2}{k-1} \sim \frac{4^{k-1}}{\sqrt{\pi k}} < 2^{2k}.$$

This is a wide range, but it is close to the best that is currently known – no one has improved significantly on the exponents $k/2$ and $2k$ (at least not on the coefficients $1/2$ and 2).

The maximal determinant problem

Suppose A is an $n \times n$ matrix with entries in $\{\pm 1\}$.
How large can the determinant $\det(A)$ be?



The maximal determinant problem

Hadamard (1893) partly answered the question by proving an upper bound

$$\det(A) \le n^{n/2}$$

that can be attained for infinitely many values of n. Such n are called *Hadamard orders* and the matrices attaining the bound $n^{n/2}$ (or $-n^{n/2}$) are called *Hadamard matrices*.

Jacques Hadamard (1865–1963)

A short proof of Hadamard's inequality

Consider the "Gram matrix" $G = A^T A$. Note that G is positive semi-definite, so has non-negative real eigenvalues λ_j. Also, the diagonal elements g_{jj} of G are n. Thus

$$\begin{aligned}
\det(A)^{2/n} = \det(G)^{1/n} &= \left(\prod_j \lambda_j\right)^{1/n} \\
&\le \frac{1}{n}\sum_j \lambda_j \quad \text{(by AGM inequality)} \\
&= \frac{\operatorname{trace}(G)}{n} = n.
\end{aligned}$$

Thus $\det(A) \le n^{n/2}$, and there is equality iff $G = nI$. □

The proof shows that Hadamard matrices are orthogonal, i.e. $A^T A = AA^T = nI$.

Comments on the maxdet problem

- We can ask the same question for $n \times n$ matrices that are allowed to have *real* entries in $[-1, 1]$. The answer is the same, since the maximum occurs at extreme points.
- We can ask the same question for $(n-1) \times (n-1)$ matrices whose entries are in $\{0, 1\}$. The answer is the same, except for a scaling factor of 2^{n-1} (see next slide).
- A more general problem that arises in the design of experiments is to maximise $\det(A^T A)$, where A is an $m \times n$ matrix with entries in $\{0, \pm 1\}$, and $m \geq n$.
  I will only consider the case $m = n$.

Determinants of $\{\pm 1\}$ matrices

An $n \times n$ $\{\pm 1\}$ matrix always has determinant divisible by 2^{n-1}, because of a well-known mapping from $\{0, 1\}$ matrices of order $n - 1$ to $\{\pm 1\}$ matrices of order n.

The mapping is reversible if we are allowed to normalise the first row and column of the $\{\pm 1\}$ matrix by changing the signs of rows/columns as necessary.

$$\begin{pmatrix}1&0&1\\1&1&0\\0&1&1\end{pmatrix} \xrightarrow{\text{double}} \begin{pmatrix}2&0&2\\2&2&0\\0&2&2\end{pmatrix}$$

$$\xrightarrow{\text{border}} \begin{pmatrix}1&1&1&1\\0&2&0&2\\0&2&2&0\\0&0&2&2\end{pmatrix} \xrightarrow[\text{first row}]{\text{subtract}} \begin{pmatrix}1&1&1&1\\-1&1&-1&1\\-1&1&1&-1\\-1&-1&1&1\end{pmatrix}$$

Design of experiments

The field of *Design of Experiments* was pioneered by C .S. Peirce (1877–1883) and R. A. Fisher (1926–1935).[1]

We might want to perform m experiments to find information about the effect of n variables, where $m \geq n$. For example, we could be trying to estimate the weights of n objects using m weighings, or estimate the effect of n different drugs on m patients. We can model the experiment by an $m \times n$ matrix A of $\{0, \pm 1\}$ entries.

Provided the outcomes are linear functions of the variables, a sensible criterion to choose the best experimental design is to maximize $\det(A^T A)$. Here $A^T A$ is called the *information matrix* of the design.

An $m \times n$ $\{\pm 1\}$-matrix A for which $\det(A^T A)$ is maximal is called a *D-optimal design*, and if $m = n$ it is called *saturated*.

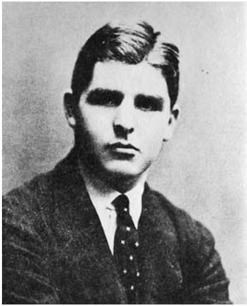
[1] The dates are dates of their most relevant publications.

Charles S. Peirce and Ronald A. Fisher

The Hadamard conjecture

> *It seems probable that, whenever m is divisible by 4, it is possible to construct an orthogonal matrix of order m composed of ± 1, but the general theorem has every appearance of difficulty.*

Paley, 1933



Raymond Paley (1907–1933)

The Hadamard conjecture

It is conjectured that Hadamard matrices exist for orders $n = 1, 2$, and $4k$ for all positive integers k. This conjecture is known as the *Hadamard conjecture*, although it seems to have been first explicitly stated by Paley. It is easy to prove that 1, 2 and $4k$ are the only possible orders.

Paley showed how to construct a Hadamard matrix of order $q + 1$ when $q \equiv 3 \bmod 4$ is a prime power, and of order $2(q + 1)$ when $q \equiv 1 \bmod 4$ is a prime power. Combined with a doubling construction of Sylvester, this shows that Hadamard matrices of order $n = 2^r(q + 1)$ exist whenever q is zero or an odd prime power, $r \geq 0$ and $4|n$.

Many other constructions have been found. Together, they show that all $n = 4k \leq 664$ are the orders of Hadamard matrices.

Hadamard and non-Hadamard orders

Recall that n is a *Hadamard order* if a Hadamard matrix of order n exists, and a *non-Hadamard order* otherwise.

For example, $1, 2, 4, 8, 12, 16, 20, 24$ are Hadamard orders; $3, 5, 6, 7, 9, 10, 11, 13$ are non-Hadamard orders.

Let $D(n)$ be the maximum determinant of an $n \times n$ $\{\pm 1\}$-matrix, and

$$\mathcal{R}(n) := \frac{D(n)}{n^{n/2}} \le 1$$

be the ratio of $D(n)$ to the Hadamard bound.

For positive integers n, $\mathcal{R}(n) = 1$ iff n is a Hadamard order.

$\mathcal{R}(n)$ for small n

n	$\mathcal{R}$	n	$\mathcal{R}$	n	$\mathcal{R}$	n	$\mathcal{R}$
–	–	1	1	2	1	3	0.77
4	1	5	0.86	6	0.74	7	0.63
8	1	9	0.73	10	0.74	11	0.61
12	1	13	0.86	14	0.74	15	0.63
16	1	17	0.75	18	0.74	19	0.64
20	1	21	0.78	22	0.70?	23	0.61?

Table: $\mathcal{R}(n)$ for $n < 24$

Each block of two columns corresponds to a congruence class of n mod 4.

Known lower bounds on $\mathcal{R}(n)$

> *It appears plausible that there always exists such a matrix with determinant greater than $\frac{1}{2}h_n$, where $h_n = n^{n/2}$ is the Hadamard bound.*
>
> *Rokicki et al (2010)*

Tomas Rokicki

Will Orrick

Known lower bounds on $\mathcal{R}(n)$

What can we say about lower bounds on $\mathcal{R}(n)$?

Rokicki, Kazmenko, Meyrignac, Orrick, Trofimov and Wroblewski (2010) verified numerically that $\mathcal{R}(n) > 1/2$ for all $n \le 120$, and conjectured that this lower bound always holds.

However, the theoretical bounds are much weaker.

Until recently, the best result was

$$\mathcal{R}(n) \ge \frac{1}{\sqrt{3n}}\,.$$

This bound tends to zero as $n \to \infty$.

Improved lower bounds on $\mathcal{R}(n)$

Using the probabilistic method, we[2] recently showed that

$$\mathcal{R}(n) \geq c_d$$

for some $c_d > 0$ that depends only on $d = n - h$, where h is the largest Hadamard order $\leq n$.

Also, if the Hadamard conjecture is true, then $d \leq 3$ and

$$\mathcal{R}(n) > 1/9.$$

[2] Brent, Osborn and Smith, arXiv:1402.6817v2, 13 March 2014.

First try – A naive approach

How can we use the probabilistic method to give a lower bound on $\mathcal{R}(n)$?

An obvious approach is to consider a random $\{\pm 1\}$-matrix of order n, hoping that a random matrix often has a large determinant. It does, but not large enough!

Turán (1940) showed that the

$$\mathbb{E}[\det(A)^2] = n!$$

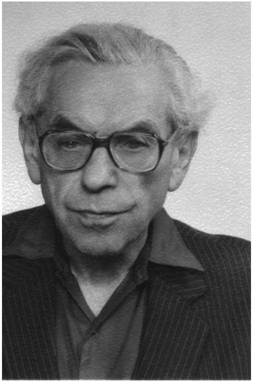
for $\{\pm 1\}$-matrices A of order n, chosen uniformly at random.

Compare this to the Hadamard bound $\det(A)^2 \le n^n$.

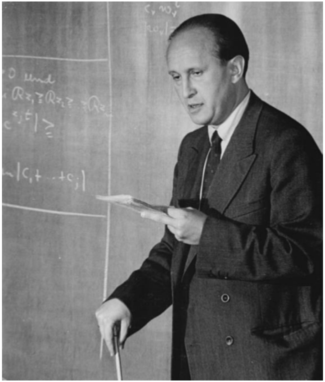
$$\mathbb{E}[\det(A)^2] = n! \approx \left(\frac{n}{e}\right)^n \sqrt{2\pi n} \ll n^n.$$

The difference is a factor of almost e^n.

Erdős and Turán



Paul Erdős (1913–1996)



Paul Turán (1910–1976)

Bordering a Hadamard matrix

Suppose $n = h + d$ where h is the order of a Hadamard matrix H, and d is small (if the Hadamard conjecture is true, we can assume that $0 \leq d \leq 3$).

We can start with H and add a “border” of d rows and columns. Since H has a large determinant (as large as possible for a $\{\pm 1\}$-matrix of order h), we might hope that the resulting order n matrix will often have a large determinant.

To analyse the effect of a border on the determinant, we need to look at the *Schur complement*.

The Schur complement

Let

$$A = \begin{bmatrix} H & B \\ C & D \end{bmatrix}$$

be an $n \times n$ matrix written in block form, where H is $h \times h$, and $n = h + d > h$. (Here H does not have to be Hadamard, any nonsingular $h \times h$ matrix will do.)

The *Schur complement* of H in A is the $d \times d$ matrix

$$D - CH^{-1}B.$$

The Schur complement is relevant to our problem because

$$\det(A) = \det(H) \cdot \det(D - CH^{-1}B).$$

Proof of the determinant identity

To prove the Schur complement identity

$$\det(A) = \det(H) \cdot \det(D - CH^{-1}B),$$

take determinants of each side in the identity

$$A = \begin{bmatrix} H & B \\ C & D \end{bmatrix} = \begin{bmatrix} I & 0 \\ CH^{-1} & I \end{bmatrix} \begin{bmatrix} H & B \\ 0 & D - CH^{-1}B \end{bmatrix} .$$

You can verify this identity directly by block matrix multiplication, or derive it by block Gaussian elimination.

Application of the Schur complement

Let H be an $h \times h$ Hadamard matrix that is a principal submatrix of an $n \times n$ matrix A, where $n = h + d$ as usual.

$$A = \begin{bmatrix} H & B \\ C & D \end{bmatrix} .$$

- Since H is Hadamard, $HH^T = hI$ and $\det(H) = h^{h/2}$, so

$$\det(A) = h^{h/2} \det(D - h^{-1} CH^T B) .$$

- To maximise $|\det(A)|$ we need to maximise

$$|\det(D - h^{-1} CH^T B)| .$$

(The sign of the determinant is not important, only the absolute value is of interest to us.)

A numerical example

Suppose we want to construct a large-determinant $\{\pm 1\}$-matrix of order 5. We could start with the order 4 Hadamard matrix

$$H = \begin{bmatrix} +1 & +1 & +1 & +1 \\ +1 & -1 & +1 & -1 \\ +1 & +1 & -1 & -1 \\ +1 & -1 & -1 & +1 \end{bmatrix}$$

which has $\det(H) = 16$, and add a border along the right and bottom.

Choosing B, C, D randomly

Suppose we randomly choose B, C and D to give

$$A = \left[\begin{array}{c|c} H & B \\ \hline C & D \end{array}\right] = \left[\begin{array}{cccc|c} +1 & +1 & +1 & +1 & -1 \\ +1 & -1 & +1 & -1 & +1 \\ +1 & +1 & -1 & -1 & +1 \\ +1 & -1 & -1 & +1 & +1 \\ \hline +1 & +1 & +1 & -1 & +1 \end{array}\right]$$

$B^T H = (H^T B)^T = [+2, -2, -2, -2]$,
$C = [+1, +1, +1, -1]$,
$C H^T B = 2 - 2 - 2 + 2 = 0$,
$\det(D - h^{-1} C H^T B) = \det(1) = 1$, and
$\det(A) = \det(H) \cdot 1 = 16.$ Disappointing!

Choosing only B randomly

Let's choose B randomly, but then choose C to avoid any cancellation in the inner product $C \cdot H^T B$, then choose D to maximise $|\det|$. This gives

$$A = \left[\begin{array}{c|c} H & B \\ \hline C & D \end{array}\right] = \left[\begin{array}{cccc|c} +1 & +1 & +1 & +1 & -1 \\ +1 & -1 & +1 & -1 & +1 \\ +1 & +1 & -1 & -1 & +1 \\ +1 & -1 & -1 & +1 & +1 \\ \hline +1 & -1 & -1 & -1 & -1 \end{array}\right]$$

$B^T H = (H^T B)^T = [+2, -2, -2, -2]$,
$C = [+1, -1, -1, -1]$, $CH^T B = 2 + 2 + 2 + 2 = 8$.
$\det(D - h^{-1} CH^T B) = \det(-1 - 2) = -3$, and
$\det(A) = \det(H) \cdot (-3) = -48$. Apart from the sign, which is easily fixed, we get the maximum possible determinant (48).

A good probabilistic construction

Choose the $h \times d$ $\{\pm 1\}$-matrix B uniformly at random.

Guided by our numerical examples, choose $C = (c_{ij})$, where

$$c_{ij} = \mathrm{sgn}(H^T B)_{ji} \;\text{ for }\; 1 \le i \le d,\; 1 \le j \le h$$

so there is no cancellation in the inner products defining the diagonal elements of $C \cdot H^T B$.

In the case $d = 1$ this construction is due to Brown and Spencer (1971) and also (independently) Best (1977).

Results obtained using the probabilistic construction

Write $F = h^{-1}CH^TB$, so the Schur complement is $D - F$.

By studying the probability distribution of elements of F we find that, with a positive probability, F is close to a diagonal matrix.

The diagonal elements of F have mean $\mu \approx (2h/\pi)^{1/2}$ and variance $\sigma^2 \leq 1/4$.

The off-diagonal elements have mean 0 and variance 1.

Thus, F is usually close to the diagonal matrix μI.

Using these facts about F, we[3] can prove Theorems 1–2 (see next slide).

[3] Brent, Osborn and Smith, arXiv:1402.6817v2, 13 March 2014.

Lower bound for $d \leq 3$

Theorem 1 applies if $d \leq 3$, which is always the case if the Hadamard conjecture is true.

Theorem 1. If $0 \leq d \leq 3$, $n = h + d$, where h is the order of a Hadamard matrix, then

$$\mathcal{R}(n) \geq \left(\frac{2}{\pi e}\right)^{d/2}.$$

Corollary. If the Hadamard conjecture is true, then

$$\mathcal{R}(n) > 1/9.$$

General lower bound

Without assuming the Hadamard conjecture, we have

Theorem 2. If $d \geq 0$, $n = h + d$ as above, then

$$\mathcal{R}(n) \geq \left(\frac{2}{\pi e}\right)^{d/2} \left(1 - d^2\sqrt{\frac{\pi}{2h}}\right).$$

To prove Theorem 2, we need some additional ingredients:

- Chebyshev's inequality (1867) for the tail of a probability distribution, and
- Ostrowski's inequality (1938) for $\det(I - E)$ where the elements of E are small.

The $O(d^2/h^{1/2})$ term

By a result of Livinsky (2012) on gaps between Hadamard orders,

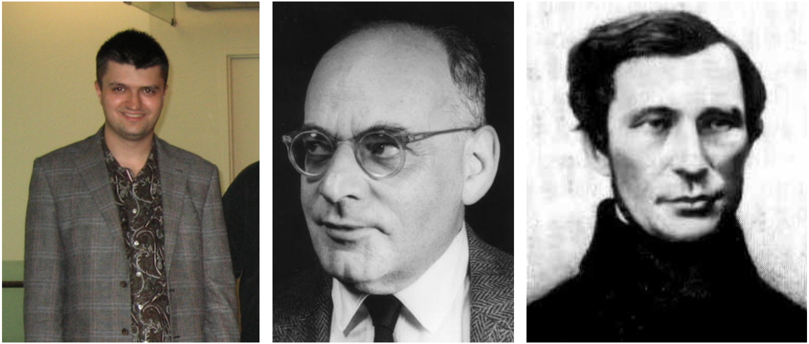
$$d^2/h^{1/2} \to 0 \;\text{ as }\; n \to \infty,$$

so the lower bound given by Theorem 2 is close to

$$\left(\frac{2}{\pi e}\right)^{d/2}$$

when n is large.

Livinskyi, Ostrowski, Chebyshev



Ivan Livinskyi (recent); Alexander Ostrowski (1893–1986); Pafnuty Chebyshev (1821–1894)

A randomised algorithm

The probabilistic construction can easily be used to give a randomised algorithm for finding large-determinant matrices, i.e. nearly D-optimal designs.

The algorithm actually works better than the theory suggests. In all the cases that we have tried, there is no difficult in finding an $n \times n$ $\{\pm 1\}$-matrix A with

$$\frac{\det(A)}{n^{n/2}} \geq \left(\frac{2}{\pi e}\right)^{d/2}.$$

Conclusion

We've seen that probabilistic ideas can be used for

- practical primality testing,
- finding lower bounds on Ramsey numbers,
- finding large-determinant $\{\pm 1\}$-matrices (almost-optimal designs), and
- proving lower bounds that are close to Hadamard's upper bound on the largest-possible determinants of $\{\pm 1\}$-matrices.

There are many other examples that I could have given if we had more time.

I hope that I have convinced you that probabilistic ideas are relevant even for problems that do not appear to involve any randomness!