This column is devoted to mathematics for fun. What better purpose is there for mathematics? To appear here, a theorem or problem or remark does not need to be profound (but it is allowed to be); it may not be directed only at specialists; it must attract and fascinate.

We welcome, encourage, and frequently publish contributions from readers—either new notes, or replies to past columns.

Please send all submissions to the Mathematical Entertainments Editor, **Alexander Shen,** Institute for Problems of Information Transmission, Ermolovoi 19, K-51 Moscow GSP-4, 101447 Russia; e-mail: shen@landau.ac.ru

Cliques, the Cauchy Inequality, and Information Theory

Here I present two rather unexpected solutions of a simple problem:

> Let $G = (V,E)$ be an undirected graph having n edges; to prove that the number of 3-cliques in G does not exceed $(\sqrt{2}/3)n^{3/2}$.

As usual, a 3-clique is a set of three vertices connected by three edges. It would be just as good to formulate the problem without any graph-theoretic terminology at all:

> Assume given a collection of n segments in the plane; to prove that there are at most $(\sqrt{2}/3)n^{3/2}$ triangles whose sides belong to the collection.

This theorem is an easy consequence of the following inequality. Let A be a finite subset of the Cartesian product $X \times Y \times Z$. Define the sets A_{xy}, A_{xz}, and A_{yz} as the projections of A onto $X \times Y$, $X \times Z$, and $Y \times Z$, respectively. Then

$$(\#A)^2 \leq \#A_{xy} \cdot \#A_{xz} \cdot \#A_{yz}, \qquad \text{(C)}$$

where $\#S$ stands for the cardinality of a finite set S.

How do we apply (C) to our problem? Denoting the set of vertices by V, the product $V \times V \times V$ is the set of ordered triples of vertices. We are estimating m, the number of 3-cliques; each of these corresponds to $6 = 3!$ ordered triples. Let A, then, be the subset of $V \times V \times V$ coming from 3-cliques. It has $6m$ elements, and we want to use (C) to estimate this. For this, we need to know the number of elements in the projections of A. Projection A_{xy} contains only pairs (x,y) that are connected by an edge, and each edge gives two pairs, so the cardinality of the projection does not exceed $2n$, where n is the number of edges. Therefore (C) says that

$$(6m)^2 \leq (2n) \cdot (2n) \cdot (2n),$$

in agreement with the stated conclusion.

Now I present two proofs of inequality (C).

(First proof) I prefer to use the following geometric version of (C): if A is a (measurable) subset of $\mathbb{R}^3$ having volume V, and S_1, S_2, S_3 are the areas of its two-dimensional projections (onto the three coordinate planes), then

$$V^2 \leq S_1 S_2 S_3. \qquad \text{(G)}$$

(If A is composed of unit cubes with integer vertices, this is exactly the statement (C), for the volume is the number of unit cubes and the area is the number of unit squares in the projection.)

To prove (G), first generalize it to say

$$\left(\iiint f(x,y)g(x,z)h(y,z)\,dx\,dy\,dz\right)^2 \leq \leq \iint f^2(x,y)\,dx\,dy \cdot \iint g^2(x,z)\,dx\,dz \cdot \iint h^2(y,z)\,dy\,dz \qquad \text{(I)}$$

for any non-negative f, g, and h. If f, g, and h are equal to 1 inside the corresponding projections of A and to 0 outside, then $f(x,y)g(x,z)h(y,z) = 1$ for all $(x,y,z) \in A$ (and maybe for some other points), so that (I) gives (G).

Now inequality (I) is a variation of the Cauchy inequality and may be reduced to it:

$$(\iiint f(x,y)g(x,z)h(y,z)\ dx\ dy\ dz)^2 \leq$$
$$\leq \iint f^2(x,y)\ dx\ dy \cdot \iint (\int g(x,z)h(y,z)\ dz)^2\ dx\ dy \leq$$
$$\leq \iint f^2(x,y)\ dx\ dy \cdot \iint (\int g^2(x,z)dz \int h^2(y,z)dz)\ dx\ dy =$$
$$= \iint f^2(x,y)\ dx\ dy \cdot \iint g^2(x,z)\ dx\ dz \cdot \iint h^2(y,z)\ dy\ dz.$$

(**Second proof**) This proof of (C) is completely different (and rather strange). It uses the notion of Shannon entropy of a random variable with finite range. If a random variable ξ takes n values with probabilities $p_1, \ldots, p_n$, then the Shannon entropy of ξ is defined as

$$H(\xi) = -\sum_i p_i \log_2 p_i$$

It does not exceed $\log_2 n$ and is equal to $\log_2 n$ when all values are equiprobable.

If ξ and η are both random variables with finite range, then so is the pair $\langle \xi,\eta \rangle$, and its Shannon entropy $H(\langle \xi,\eta \rangle)$ is given by the general definition.

The conditional entropy $H(\xi \mid \eta)$ of ξ when η is known can be defined as

$$H(\xi \mid \eta) = H(\langle \xi, \eta \rangle) - H(\eta).$$

You can easily check that this agrees with the natural definition of conditional entropy of ξ given η: namely, fix any value of η and compute $H(\xi)$ using the conditional probabilities of the ξ values in place of the p_i; and then take the weighted average of the results, weighted by the probabilities of the various values of η.

It is a standard fact that

$$H(\langle \xi,\eta \rangle) \leq H(\xi) + H(\eta).$$

The reader not acquainted with these matters will probably enjoy tackling this by straightforward analysis. So now we know that

$$0 \leq H(\xi \mid \eta) \leq H(\xi) \qquad \text{(L)}$$

for any ξ and η.

Now it is easy to prove that

$$2H(\langle \xi,\eta,\tau \rangle) \leq H(\langle \xi,\eta \rangle) + H(\langle \xi,\tau \rangle) + H(\langle \eta,\tau \rangle). \qquad \text{(E)}$$

Indeed, (E) can be rewritten as

$$H(\tau \mid \langle \xi,\eta \rangle) + H(\eta \mid \langle \xi,\tau \rangle) \leq H(\langle \eta,\tau \rangle)$$

where the right-hand side equals $H(\tau) + H(\eta \mid \tau)$. It remains to note that $H(\tau \mid \ldots) \leq H(\tau)$, and $H(\eta \mid \langle \xi,\tau \rangle) \leq H(\eta \mid \tau)$. (The first of these is fact (L), the second is the "conditionalized" version of it. The intuitive content of these is that any conditional entropy is smaller the more we know.)

Now to prove (C) using (E). Consider the random variable that is uniformly distributed in the set $A \subset X \times Y \times Z$. It can be considered as a triple of (dependent) variables $\langle \xi,\eta,\tau \rangle$, where $\xi \in X$, $\eta \in Y$, and $\tau \in Z$. The entropy of the triple $\langle \xi,\eta,\tau \rangle$ equals $\log_2 \#A$. Using (E), we get that

$$2 \log_2 \#A \leq H(\langle \xi,\eta \rangle) + H(\langle \xi,\tau \rangle) + H(\langle \eta,\tau \rangle).$$

The pair $\langle \xi,\eta \rangle$ takes values in A_{xy}, therefore its entropy does not exceed $\log_2 \#A_{xy}$. For the same reasons $H(\langle \xi,\tau \rangle) \leq \log_2 \#A_{xz}$ and $H(\langle \eta,\tau \rangle) \leq \log_2 \#A_{yz}$. Therefore

$$2 \log_2 \#A \leq \log_2 \#A_{xy} + \log_2 \#A_{xz} + \log_2 \#A_{yz},$$

and we get (C) by exponentiation.

***

I have received the following letter, completing the picture sketched in an earlier column.

> Your column in *The Mathematical Intelligencer* for Spring, 2000, never mentions the name of the problem discussed. It is called the "majority problem" in the theoretical computer science literature, and the two most interesting papers on the subject (he says with a blush) are
>
> L. Alonso, E. M. Reingold, and R. Schott, "Determining the Majority," *Info. Proc. Let.* 47 (1993), 253–255.
>
> L. Alonso, E. M. Reingold, and R. Schott, "The Average-Case Complexity of Determining the Majority," *SIAM J. Computing* 26 (1997), 1–14.
>
> Both papers deal with (exact, achievable) lower bounds.

Edward M. Reingold
Department of Computer Science
University of Illinois at Urbana-Champaign
Urbana, IL 61801-2987
USA
e-mail: reingold@cs.uiuc.edu