

The pigeonhole principle


ICTS-RRI Maths Circle, Bengaluru

13, 27 April 2024, 10:00 am to 1:00 pm

The principle

Let n be a positive integer. If $n + 1$ balls are placed in n bins, then some bin must have at least two balls.

The principle is simple, yet it often leads to very pleasing conclusions¹.

Example. There are 20 guests at a party. Some guests shake hands with other guests, but no pair of guests shake hands twice. Show that there are two guests who shake hands exactly the same number of times. (To use the above pigeonhole principle, ask yourself the following: what are the balls, what are the bins?) 

It might help to visualize the pigeonhole principle using a picture. Draw a graph with vertices and edges as follows. On the left, keep a vertex for each ball; call the vertices a_1, a_2, \dots, a_{n+1} . On the right, keep a vertex for each bin; call these vertices b_1, b_2, \dots, b_n . If ball i is placed in bin j , draw a directed edge (an arrow) from a_i to b_j . The point is the following. The total number of arrows leaving the vertices on the left is exactly $n + 1$; so the total number of arrows landing on the vertices on the right must also be $n + 1$. Let d_j be the number of arrows that land on b_j . Then,

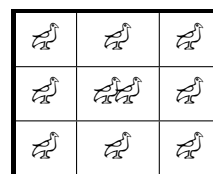
$$n + 1 = d_1 + d_2 + \dots + d_n \leq n \cdot \max_j d_j.$$

So $\max_j d_j \geq \frac{n+1}{n} = 1 + \frac{1}{n}$. But $\max_j d_j$ is an integer, so it must be at least 2. Notice that the quantity $1 + \frac{1}{n}$ is the average number of balls per bin. We simply combined two elementary facts: (i) in any collection of numbers, the maximum is always at least their average; (ii) an integer that is more than one must be at least two. In general, following this approach we arrive at a somewhat more general version of the pigeonhole principle.

Suppose m balls are placed in n bins, then

1. there is a bin containing at least m/n balls; and
2. there is a bin containing at most m/n balls.

Example. Each cell of a 5×41 table is coloured either *black* or *white*. Show that there are three rows and three columns so that all nine cells in their intersection have the same colour. This is a little tricky,



¹ <https://archive.org/details/mathematical-circles-russian-experience/page/31/mode/1up>

Practice problems

Attempt the problems in the book

Dmitri Fomin, Sergey Genkin, Ilya Itenberg:
Mathematical Circles (Russian Experience),
([click here](#)).

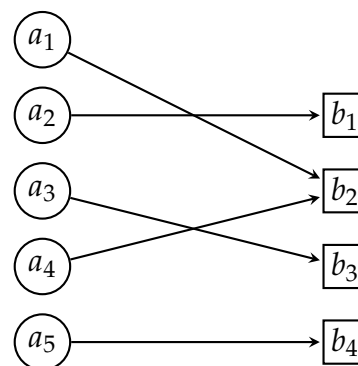


Figure 1: Five balls placed in four bins

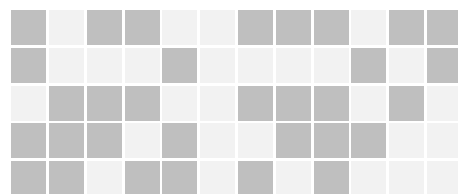


Figure 2: A 5×12 table

but please give it a try. (I don't want to spoil the challenge for you, but here are some imprecise hints, if you need them. First, focus on the columns. In this column, some three of the cells must be of the same colour (why?). Now, think of the columns as balls and put them in 10 bins. Why 10 bins? What are the bins for? When is a column placed in a bin? Conclude that some bin has at least . . . balls; then . . .)



Exploration I

Consider a sequence of distinct integers, e.g.,

$$s = 1, 5, 19, 4, 7, 22, 6, 8, 15, 11.$$

A subsequence of this sequence is obtained by removing some of the numbers in the sequence and retaining the rest. For example,

$$s_1 = 1, 4, 6, 8, 11$$

$$s_2 = 19, 15, 11$$

are subsequences of the sequence s . A sequence is *monotone increasing* if each element in the sequence is greater than the previous element (as in s_1); it is *monotone decreasing* if each element in the sequence is less than the previous element (as in s_2).

Explore: Construct a sequence t by reordering the numbers $1, 2, \dots, 16$, so that t contains neither a monotone increasing nor a monotone decreasing subsequence of five elements. Do you think every such sequence obtained by reordering $1, 2, \dots, 17$ always has a monotone subsequence with 5 elements? Why? Is it true that every infinite sequence of distinct integers has an infinite monotone subsequence.

Exploration II

Consider a graph where the vertices represent locations in a city and the edges represent two-way streets connecting adjacent locations. There is a robber in the city. We wish to determine the minimum number of cops needed to catch the robber. The rules are the following. First, the cops occupy some vertices in the graph. Then, the robber places herself on some vertex of her choice.

The cops and robber make their moves in alternate time-steps. The cops traverse one edge at a time; the robber can run through several edges in one step. If it is the cops' turn, one of the cops moves to an adjacent vertex (we allow dummy moves, where no cop moves). If it is the robber's turn, she can move to another vertex along a path on which *none of the intermediate vertices is occupied by a cop*, or she can stay put. If at any time a cop and a robber are at the same vertex, the robber is caught.

The game is played with perfect information, the cops know where the robber is at each point, and the robber knows where the cops are.

- (a) How many cops does one need to catch a robber on the path P_5 ? How many does one need for the cycle C_8 ? What about paths and cycles with more vertices?
- (b) Consider the 8×8 grid graph $G_{8 \times 8}$. Show that with eight cops there is a strategy for the cops to catch the robber no matter how the robber tries to evade them. What if we have only seven cops? What if we play the game on an $n \times n$ grid graph $G_{n \times n}$?

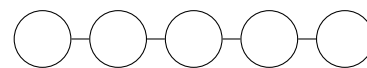


Figure 3: The path graph P_5

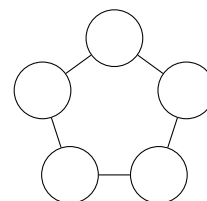


Figure 4: The cycle graph C_5

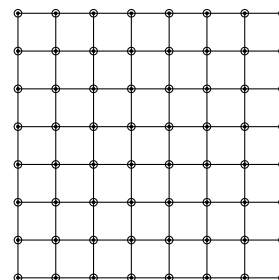


Figure 5: The grid graph $G_{8 \times 8}$

Exploration III

I asked ChatGPT the following question:

Write the numbers $1, 2, \dots, 64$ in the cells of the 8×8 board, different numbers in different cells, in such a way that adjacent numbers differ by as little as possible.

After thinking a little, ChatGPT said this:

To minimize the difference between adjacent numbers, we can arrange them in a serpentine pattern. Here's one possible arrangement. (It gave the arrangement on the right.)

Then it said:

In this arrangement, adjacent numbers typically differ by 1, except for a few places where they differ by 8. This arrangement ensures that adjacent numbers have the least possible difference on average.

```
01 02 03 04 05 06 07 08
16 15 14 13 12 11 10 09
17 18 19 20 21 22 23 24
32 31 30 29 28 27 26 25
33 34 35 36 37 38 39 40
48 47 46 45 44 43 42 41
49 50 51 52 53 54 55 56
64 63 62 61 60 59 58 57
```

Figure 6: ChatGPT's arrangement

Question: Is ChatGPT bluffing? What is the maximum difference between adjacent numbers in the above arrangement? Is there a better arrangement? How much better? Give it a try!

Definitions for our discussion 13 April: We may think of this problem in the language of graphs with vertices and edges. Suppose each vertex of a graph is labelled with a distinct integer. The *stretch* of this labelling is the *maximum* difference between the labels of adjacent vertices. Our goal is to find a labelling with *minimum* stretch. The stretch of the best labelling is the *stretch number* of the graph; we write $sn(G)$ for the stretch number of the graph G . Both the words, *minimum* and *maximum*, are used above. The order in which they appear is important to the definition; make sure you understand it.

- (a) Consider the graphs P_5 and C_5 described on the previous page. What is $sn(P_5)$? What is $sn(C_5)$? What is the stretch number of the path graph P_n with n vertices; what is the stretch number of the cycle graph C_n with n vertices?
- (b) If the graph has N vertices, we may assume that in the optimal labelling the labels are $1, 2, \dots, N$. (Why?)
- (c) What is the stretch number of the $n \times n$ grid graph $G_{n \times n}$?

We will meet on 13, 27 April 2024, 10:00 am to 1:00 pm.