Apples, Oranges, and Bananas

Warut Suksompong

Puzzle 1 : Some apples and oranges are distributed among 99 boxes. Prove that we can choose 50 boxes so that together they contain at least half of all the apples and at least half of all the oranges.

This puzzle was given in the All-Russian Mathematical Olympiad in 2005. We invite the reader to try this puzzle before moving on.

First, if there are only apples, and we would like to choose 50 boxes so that together they contain at least half of the apples, that's easy. We just choose the 50 boxes with the greatest number of apples. Also, we can see that the bound 50 is tight. If every box contains one apple, and we are only allowed to choose 49 boxes, any effort will be fruitless – we will always get less than half of the apples no matter what we do.

But we have apples and oranges. What are we going to do ? We want to choose a subset with 50 boxes that satisfy two (symmetric) properties simultaneously. If we can show that the number of subsets that satisfy one property is greater than half of the number of all subsets of size 50, we would be done – some subset satisfies both properties at once. What we are guessing is stronger than what the puzzle asks for, so we are making a gamble. Nevertheless, the statement is true. To show that, we will make a little set-theoretic detour.

Recall that given a set of size $2n+1$, the number of subsets of size n is equal to the number of subsets of size $n+1$. That means we can pair up the subsets of size n with the subsets of size $n+1$. What is more interesting is that we can do it in a rather special way : Each subset of size n is a subset of its pair of size $n+1$.

Lemma 1 *Let $N = \{1, 2, \ldots, n\}$ and $0 \leq k < n/2$. Let A be the collection of all subsets of N of size k, and B be the collection of all subsets of N of size $k+1$. Construct a bipartite graph G with vertices $A \cup B$ in such a way that there is an edge joining $X \in A$ and $Y \in B$ if and only if $X \subseteq Y$. Then there exists a matching between all the elements of A and (not necessarily all) the elements of B.*

This Lemma is given as an exercise without solution in [**3**]. The following proof is based on the hint for the exercise.

Proof. We define a matching f. For any set $X \in A$ (which has size k), let $f(X)$ be a subset of N of size $k+1$ that contains X ; the single element that is added to $f(X)$ is determined in the following way :

Suppose $X = \{a_1, a_2, \ldots, a_k\}$, where $a_1 < a_2 < \ldots < a_k$, and define $a_0 = 0$. Let m be the largest among the indices $0 \leq i \leq k$ for which $2i - a_i$ is maximized. Then $f(X) = X \cup \{a_m + 1\}$.

First, we show that $f(X)$ is well-defined. To that end, we prove that a_m cannot be the largest element, n. Indeed, if that were the case, then $2m - a_m = 2m - n \leq 2k - n < 0 = 2 \cdot 0 - a_0$, contradicting the definition of m. We also prove that $a_m + 1$

cannot already be present in X. Indeed, if it were, then it would have index $m+1$. But then $2(m+1) - a_{m+1} = 2m + 2 - (a_m + 1) = 2m - a_m + 1 > 2m - a_m$, once again contradicting the definition of m.

Now we show that for any two distinct sets $X, Y \in A$, the sets $f(X)$ and $f(Y)$ are distinct. Equivalently, any subset Z of N of size $k+1$ can be an image of at most one subset of size k. Consider the elements of $Z = \{a'_1, a'_2, \ldots, a'_{k+1}\}$. We show that the added element corresponds to the *first* index that maximizes the value $2i - a_i$. Originally, the index m had the value $2m - a_m = r$ maximized. The values $2i - a'_i$ for $i = 1, 2, \ldots, m$ are unchanged. The new element, a'_{m+1}, has the corresponding value $2(m+1) - a_{m+1} = 2(m+1) - (a_m + 1) = r + 1$. For $a'_{m+2}, a'_{m+3}, \ldots, a'_{k+1}$, the corresponding values $2i - a'_i$ are $2(i+1) - a_i = 2i - a_i + 2$. By the definition of m, we have that $2i - a_i < 2m - a_m$, which is equivalent to $2i - a_i \leq 2m - a_m - 1$. Hence $2i - a_i + 2 \leq 2m - a_m + 1 = r + 1$. This means $m + 1$ is the first index that maximizes the value $2i - a_i$, as desired. Hence, Z is an image of at most one k-element set X with respect to the matching f. ■

We are now ready to show that the number of subsets with 50 boxes that contain at least half of all the apples is greater than half of the number of all subsets with 50 boxes. First, each subset with 50 boxes can be uniquely paired with its complement, which is a subset with 49 boxes. Of the two subsets, at least one has at least half of all the apples. If at least half of the subsets with 50 boxes have at least half of the apples, we are done. Otherwise, at least half of the subsets with 49 boxes have at least half of all the apples. Using Lemma 1, the subsets with 50 boxes that are matched with these subsets with 49 boxes also have at least half of the apples, which means we are done.

But are we ? We want to show that the number of subsets with 50 boxes that contain at least half of all the apples is *greater than* half of the number of all subsets with 50 boxes. We have only shown that it is *at least* half, so there is more work to do. The only case in which it might be *exactly* half is when exactly half of the subsets with 49 boxes and exactly half of the subsets with 50 boxes contain at least half of all the apples, and they correspond to one another in the matching induced by Lemma 1. Assume that the boxes contain $a_1 \geq a_2 \geq \ldots \geq a_{99}$ apples, and assume without loss of generality that one pair in the matching is $\{a_3, a_5, \ldots, a_{99}\}$ and $\{a_1, a_3, a_5, \ldots, a_{99}\}$. The former set contains less than half of all the apples. Indeed, $a_3 + a_5 + \ldots + a_{99} \leq a_2 + a_4 + \ldots + a_{98} < a_1 + a_2 + a_4 + \ldots + a_{98}$. (Unless $a_1 = 0$, but that would imply $a_1 = a_2 = \ldots = a_{99} = 0$, and the situation can be handled separately.) On the other hand, the latter set contains at least half of all the apples. Indeed, $a_1 + a_3 + a_5 + \ldots + a_{99} \geq a_1 + a_3 + \ldots + a_{97} \geq a_2 + a_4 + \ldots + a_{98}$. So we are really done this time.

The good news is that we have solved the puzzle. The bad news is that the Russian Olympiad is organized for students of several class years separately, and it is sometimes the case that the jury proposes similar puzzles to different class years, adjusting the difficulty accordingly. This is in fact the case here, and the puzzle we have solved is for 8th grade students. Here is the one for 9th graders :

Puzzle 2 : Some apples and oranges are distributed among 100 boxes. Prove that we can choose 34 boxes so that together they contain at least one-third of all the apples and at least one-third of all the oranges.

Unfortunately, it seems hard to apply the same method when the concerned sets include only around one-third of the boxes. Let us start by again sorting the boxes according to the number of apples, say, $a_1 \geq a_2 \geq \ldots \geq a_{100}$. The crucial observation is that choosing boxes $1, 4, 7, \ldots, 100$ is enough to guarantee at least one-third of all the apples. Indeed, $a_1 + a_4 + a_7 + \ldots + a_{100} \geq a_2 + a_5 + \ldots + a_{98}$ and $a_1 + a_4 + a_7 + \ldots + a_{100} \geq a_3 + a_6 + \ldots + a_{99}$. This hints at splitting the boxes into 34 groups : $\{1\}, \{2, 3, 4\}, \{5, 6, 7\}, \cdots, \{98, 99, 100\}$. By the observation we just made, choosing any 34 boxes, one from each group, will guarantee us at least one-third of all the apples. Consequently, we can choose the box with the greatest number of oranges from each group and guarantee ourselves at least one-third of all the oranges as well.

This solution can also be applied to our first puzzle quite easily. Moreover, it has the merit of being constructive – if we actually have to choose the boxes rather than just proving their existence, it gives us a method to do it within a reasonable amount of time.

Before we move on, it is worth asking what happens in the first puzzle when we change the number 99 to 100 (or any other even number). What is the least number for which there always exists some subset with that number of boxes that together contain at least half of the apples and at least half of the oranges ? If we choose 50 out of 99 boxes using our previous method, and then choose the 100th box, we achieve our goal using 51 boxes. Is that the best we can do ? It is, when 49 boxes contain one apple each and no oranges and the remaining 51 boxes contain one orange each and no apples.

Now comes a remarkable thing. Even if we have a third type of fruit, we can still choose 51 out of the 100 boxes so that they together contain at least half of each type of fruit. This is exactly what 11th graders were tasked in the Olympiad. Here is the formal statement :

Puzzle 3 : Some apples, oranges, and bananas are distributed among 100 boxes. Prove that we can choose 51 boxes so that together they contain at least half of all the apples, at least half of all the oranges, and at least half of all the bananas.

This puzzle is quite difficult, and none of the students managed to solve it in the actual Olympiad. The methods we have so far don't seem to extend easily to accommodate for the third type of fruit. Focusing on just one type of fruit leaves us in trouble in satisfying the guarantee for the other two types. We would like to somehow deal with two types of fruit at once. We begin with the following cute lemma.

Lemma 2 *Let $(a_1, b_1), (a_2, b_2), \ldots, (a_{2n}, b_{2n})$ be pairs of positive integers. They can be partitioned into two groups of n pairs each so that if A_1 and A_2 denote the sum of the a_i in the first and second group respectively, then $|A_1 - A_2| \leq \max_i a_i$,* *and an analogous statement holds for b_i.*

Proof. We sort the pairs according to a_i in descending order, and split the pairs into n groups : $\{(a_1, b_1), (a_2, b_2)\}, \{(a_3, b_3), (a_4, b_4)\}, \ldots, \{(a_{2n-1}, b_{2n-1}), (a_{2n}, b_{2n})\}$. If we choose n pairs, one from each group, the difference in a_i will not exceed a_1. Indeed, the maximum difference is $a_1 - a_2 + a_3 - a_4 + \ldots + a_{2n-1} - a_{2n} \leq a_1$.

We first choose any n pairs, one from each group. Suppose that the difference in b_i exceeds $\max_i b_i$. Assume without loss of generality that the sum of b_i in the first group is greater than that in the second group. This means there exists two pairs, say (a_{2i-1}, b_{2i-1}) and (a_{2i}, b_{2i}), such that $b_{2i-1} > b_{2i}$, the first pair belongs to the first group, and the second pair belongs to the second group. We switch these two pairs. The difference in b_i changes by no more than $2b_1$, so the absolute value strictly decreases. We continue this process as long as the difference in b_i exceeds $\max_i b_i$. Since there are only finitely many partitions, the process necessarily stops, giving us the desired partition. ■

We are now ready to solve the 11th grade puzzle. In fact, the solution will be short and sweet. We first choose a box with the greatest number of apples, and from the remaining boxes we choose a box with the greatest number of oranges. Now, thanks to Lemma 2, we can partition the remaining 98 boxes into two groups of 49 boxes so that the difference between the apples does not exceed the box with the greatest number of apples, and likewise with oranges. We choose the 49 boxes with more bananas. Combined with the first two boxes, we have our desired set of 51 boxes.

Finally, what if there are more types of fruit ? If there are apples, oranges, bananas, and pears distributed among 100 boxes, what is the least number for which there always exists some subset with that number of boxes that together contain at least half of each type of fruit ? We leave this as an open puzzle for the reader.

Note : Credit for the puzzles from the Olympiad is due to I. Bogdanov, G. Chelnokov, and E. Kulikov. The solutions are also due to them, except for the non-constructive solution to the first puzzle, which the author came up with.

References

[**1**] *Problem 110178* (in Russian),
`http://problems.ru/view\_problem\_details\_new.php?id=110178`

[**2**] *Problem 110198* (in Russian),
`http://problems.ru/view\_problem\_details\_new.php?id=110198`

[**3**] László Lovász and Michael D. Plummer, *Matching Theory*, Akadémiai Kiadó, 1986.