Weighing the Odds: A Probabilistic Approach to Coin Weighing

Davyd Poplyonkin, Artem Chernii, Andrii Smutchak
Mentored by Anton Trygub

May 15, 2023

Overview

Introduction

Coin weighing problems have intrigued mathematicians and puzzle enthusiasts for centuries.

The problem has a lot of variations but usually involves a set of identical-looking coins, some of which are fake. The goal of the problem is usually to identify some (or all) fake coins by doing some weighings.

In some problems, there are two scales, and in one weighing we can put two subsets of coins on these scales. Then, we will learn which subset has a larger total weight, or that they have equal total weight.

In others, there is only one scale, and in one weighing we can put some subset of coins on this scale. Then, we will learn the total weight of this subset of coins.

Some examples of coin problems

Peru 2004

There are 100 identical-looking coins, at least one of them is fake. The fake coins have equal weight and are lighter than real coins. Show that we can determine the number of fake coins in 51 weighing.

Germany 2014

There are 9 identical-looking coins. One of them is fake and thus lighter. We are given 3 indistinguishable balance scales to find the fake coin; however, one of the scales is defective and shows a random result each time. Show that the fake coin can still be found with 4 weighings.

Israel 2020

There are 7 identical-looking coins. Three of them are fake: they have equal weight and are lighter than real coins. How many weighings are needed to identify at least one fake coin?

Problem statement

Problem statement

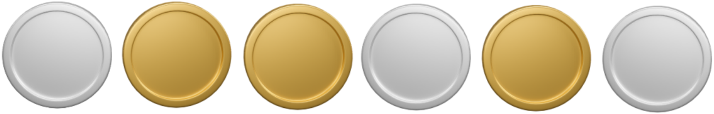
You are given n identical-looking coins. Some of them might be fake. The weights of all real coins are equal, and the weights of all fake coins are equal and smaller than the weight of real coins. Both weights are known to you.

You can use a scale that outputs the combined weight of any subset of the coins. You must decide in advance which subsets $S_1, \ldots, S_k \subseteq [n]$ of the coins to weigh.

What is the minimum number of weighings (k) needed to identify the weight of every coin?

Illustration of the problem statement

Let $n = 6$. In the picture below we have n coins, golden coins denote real coins, each weighing x, and silver coins denote fake coins, each weighing y. Here coins $2, 3, 5$ are real and $1, 4, 6$ are fake, but we don't know this.



We can try to do two weightings, with $S_1 = \{1, 3, 5, 6\}$ and $S_2 = \{2, 3, 4, 5\}$. The answer for weighing S_1 is $2x + 2y$ and for weighing S_2 is $3x + y$. So, we learned that there are 2 real and 2 fake coins among $\{1, 3, 5, 6\}$, and 3 real and 1 fake among $S_2 = \{2, 3, 4, 5\}$.

Unfortunately, this doesn't let us determine all weights, as if coins $2, 3, 4, 6$ were real, and $1, 5$ were fake, we would get the same results.

Rephrasing

Let S denote the set of real coins. We provide sets $S_1, S_2, \ldots, S_k$, and learn sizes $|S_1 \cap S|, |S_2 \cap S|, \ldots, |S_k \cap S|$.

Our goal is to be able to always determine set S from these sizes. So, the question becomes:

Rephrasing

What is the smallest k, for which exist k sets $S_1, S_2, \ldots, S_k \in [n]$, so that there are no distinct subsets $X, Y \subset [n]$ such that $|X \cap S_i| = |Y \cap S_i|$ for all $i \in [k]$?

Easy bounds

Finding a precise value of k is really hard. What bounds can we provide?

Upper bound 1

$k = n$ weighings are enough.

Proof.

Just set $S_i = \{i\}$ for all i: learn the weights of all coins separately. □

Hard to do much better with our current knowledge.

Easy bounds

Lower bound 1

We must do at least $k \geq \frac{n}{\log_2(n+1)}$ weighings.

Proof.

What is the total number of different results that we can get with k weighings?

$|S_i| \leq n$ for $1 \leq i \leq n \Rightarrow$ we can get at most $n+1$ different results for each weighing. Therefore, the total number of possible results is at most $(n+1)^k$.

From the other side, since to every configuration of real/fake coins corresponds at least one result of weighing, we must have $(n+1)^k \geq 2^n \Leftrightarrow k \geq \frac{n}{\log_2(n+1)}$.

□

How to do better?

We learned that the optimal value of k lies in $[\frac{n}{\log_2(n+1)}, n]$. Can we do better? If so, how to obtain better bounds?

Here, probabilistic methods come to play. We will start with the following question:

Question

Can we somehow provide a bound on the probability that a non-negative random variable is greater than or equal to a certain threshold?

Chernoff Bounds

Chernoff Bounds

Let $X = \sum_{i=1}^{n} X_i$ where $X_i = 1$ with probability p_i and $X_i = 0$ with probability $1 - p_i$, and all X_i are independent. Let $\mu = \mathbb{E}(X) = \sum_{i=1}^{n} p_i$. Then

Upper Tail: $\mathbb{P}(X \geq (1+\delta)\mu) \leq e^{-\frac{\delta^2}{2+\delta}\mu}$ for all $\delta > 0$

Lower Tail: $\mathbb{P}(X \leq (1-\delta)\mu) \leq e^{-\frac{\delta^2}{2}\mu}$ for all $0 < \delta < 1$

Corollary

For $0 < \delta \leq 1$:

$$\mathbb{P}(X \geq (1+\delta)\mu) \leq e^{-\frac{\delta^2}{3}\mu} \tag{1}$$

Example problem

Problem

You have n bins and n balls ($n > 1$). You throw each ball into a randomly chosen bin. Show that with probability at least $1 - \frac{1}{n}$, every bin contains at most $100 \ln n$ balls.

Let us denote the number of balls in i^{th} bin as X_i. Let's use Chernoff bounds for $\delta = 100 \ln n - 1$ and $\mu = 1$. Note that for $n \geq 2$ we have $\delta \geq 2$, and therefore $2 + \delta \leq 2\delta$.

$$P[X \geq 100 \ln n] = P[X \geq (1 + \delta)) \cdot 1] \leq e^{-\frac{\delta^2}{2+\delta} \cdot \frac{1}{3}} \leq e^{-\frac{\delta^2}{2\delta} \cdot \frac{1}{3}} = e^{-\frac{100}{6} \ln n}$$

Now, the probability that at least one bin contains $100 \ln n$ balls is $\leq$

$$ne^{-\frac{100}{6} \ln n} = n \cdot n^{-\frac{100}{6}} \leq \frac{1}{n}$$

Improved lower bound

Using these probabilistic methods, we can get a better estimate on k!

Lower bound 2

For sufficiently large n, we must do at least $k \geq \frac{1.99n}{\log_2 n}$ weighings.

Main idea: Before, we were considering all possible configurations of fake/real coins, and noticed that there could be at most $(n+1)^k$ possible results corresponding to them, as all sizes $|S_i \cap S|$ have to be between 0 and n.

But for a large portion of these configurations, the sizes $|S_i \cap S|$ are centered around $\frac{|S_i|}{2}$.

Improved lower bound

Assume such sets exist for $k < \frac{1.99n}{\log_2 n}$. Let's consider a random subset X of $[n]$, each element is added to X with probability $\frac{1}{2}$ independently. Then let's choose constant $t = 0.5001$ and consider a particular subset S_i.

Consider the probability that $||X \cap S_i| - \frac{|S_i|}{2}| \geq 1000n^t$. Note that $\mu = \mathbb{E}[|X \cap S_i|] = \frac{|S_i|}{2}$. By Chernoff bound with $\delta = \frac{2000n^t}{|S_i|}$ we get:

$$\mathbb{P}[||X \cap S_i| - \frac{|S_i|}{2}| \geq 1000n^t] \leq 2e^{-\frac{10^6 n^{2t}}{|S_i|^2}\frac{|S_i|}{2}\frac{1}{3}} = 2e^{-\frac{n^{2t}}{|S_i|}\frac{10^6}{6}}$$

$$2e^{-\frac{n^{2t}}{|S_i|}\frac{10^6}{6}} \leq 2e^{-\frac{n^{2t}}{n}\frac{10^6}{6}} = 2e^{-n^{0.0002}\frac{10^6}{6}}$$

Clearly, for large enough n, this is smaller than $\frac{1}{4n}$, and therefore $||X \cap S_i| - \frac{|S_i|}{2}| \geq 1000n^t$ holds for some i with probability at most $\frac{1}{2}$.

Improved lower bound

So, for at least 2^{n-1} possible X, all of these inequalities fail. Let's show that for some two of them, say X, $Y, |X \cap S_i| = |Y \cap S_i|$ for all i. Indeed, look at all these tuples of numbers $|X \cap Y|$. There are at most $2001n^t$ candidates for each value, so there are at most $(2001n^t)^{1.99\frac{n}{\log_2 n}}$ such tuples. Let's rewrite:

$$(2001n^t)^{1.99\frac{n}{\log_2 n}} = 2001^{1.99\frac{n}{\log_2 n}} \cdot 2^{t \log_2 n \cdot 1.99\frac{n}{\log_2 n}} \leq 2001^{\frac{2n}{\log_2 n}} \cdot 2^{0.996n}$$

Improved lower bound

And it's enough to show that starting from large enough n, this is less than 2^{n-1}. Need:

$$2001^{\frac{2n}{\log_2 n}} \cdot 2^{0.996n} < 2^{n-1} \Leftrightarrow 2001^{\frac{2}{\log_2 n}} < 2^{\frac{0.004n-1}{n}}$$

This is clear as for large n the right part goes to $2^{0.004}$, and the left to 1. $(1 < 2^{0.004})$

Improved upper bound

Upper bound 2

There exists some constant c such that for large enough n, $\frac{cn}{\log_2 n}$ weighings are sufficient.

Surprisingly, probabilistic methods again come to the rescue!

Improved upper bound

Let's choose $k = \frac{cn}{\log_2 n}$ for some c.

The sets S_i provide a matrix A with dimensions $k \times n$, such that if x is the indicator vector of real coins, then Ax is the result of weighings.

We have to show that we can choose such A with entries in $\{0, 1\}$ that $Ax \neq Ay$ for any $x \neq y \in \{0, 1\}^n$. That is, we need to show that there exists such A for which $Ax \neq 0$ for all $x \neq 0 \in \{-1, 0, 1\}^n$.

Main idea: Let's choose all entries of A randomly independently, and show that the probability that A satisfies all these constraints is positive!

Improved upper bound

Lemma

Consider vector x with a ones and b minus ones. Choose vector v with the same length, with entries chosen from $\{0, 1\}$ uniformly independently. Then

$$\mathbb{P}[v^T x = 0] = \frac{\binom{a+b}{b}}{2^{a+b}}$$

Proof.

Only elements of v at positions where x is ± 1 matter. For $x^T v = 0$, we need to "take" an equal number of ones and minus ones. If we take t ones, we take t minus ones and therefore do **not** take $b - t$ minus ones. So, we are just choosing some $t + (b - t) = b$ elements out of $a + b$. The probability of this happening is $\frac{\binom{a+b}{b}}{2^{a+b}}$. □

Improved upper bound

Let's bound $\frac{\binom{a+b}{b}}{2^{a+b}}$.

Consequence of Stirling approximations

For any $n \geq 1$ and $0 \leq a \leq n$

$$\frac{\binom{n}{a}}{2^n} \leq min(\frac{1}{2}, \frac{1}{\sqrt{n}})$$

So, the probability that A will fail for given x with t nonzeros is at most $\min(\frac{1}{2}, \frac{1}{\sqrt{t}})^k$.

Improved upper bound

For a fixed number t, there are $\binom{n}{t}2^t$ vectors x with exactly t nonzeros. Therefore, the probability that $Ax = 0$ at least for one x doesn't exceed

$$\sum_{t=1}^{n} \binom{n}{t} 2^t \min\left(\frac{1}{2}, \frac{1}{\sqrt{t}}\right)^k$$

We want this value to be less than 1.

Improved upper bound

Let's try to bound it.

Note that $\binom{n}{t}2^t = n(n-1)\cdot\ldots\cdot(n-t+1)\frac{2^t}{t!} \le 2n^t$. Choose $M = \sqrt{n}$. We get:

$$\sum_{t=1}^{M}\binom{n}{t}2^t\min(\frac{1}{2},\frac{1}{\sqrt{t}})^k \le \sum_{t=1}^{M}2n^t2^{-k} \le 4n^M2^{-k}$$

On other side,

$$\sum_{t=M+1}^{n}\binom{n}{t}2^t\min(\frac{1}{2},\frac{1}{\sqrt{t}})^k \le \sum_{t=M+1}^{n}2^n2^tM^{-\frac{k}{2}} \le n4^nM^{-\frac{k}{2}}$$

Improved upper bound

It's enough to show that for this M

$$4n^M 2^{-k} < \frac{1}{2}; \quad n4^n M^{-\frac{k}{2}} < \frac{1}{2}$$

$$\textbf{First:} \quad 4n^M 2^{-k} = 2^{2+(M\log_2 n)-k} = 2^{2+n^{\frac{1}{2}}\log_2 n - \frac{cn}{\log_2 n}}$$

$$\textbf{Second:} \quad n4^n M^{-\frac{k}{2}} = 2^{\log_2 n + 2n - \frac{\log_2 n}{3}\frac{cn}{2\log_2 n}} = 2^{\log_2 n + 2n - \frac{cn}{6}}$$

It's easy to see that both $2 + n^{\frac{1}{2}}\log_2 n - \frac{cn}{\log_2 n}$ and $\log_2 n + 2n - \frac{cn}{6}$ are less than -1 starting from some n for some choice of c. So, the statement is proven.

Conclusion

In conclusion, a probabilistic approach is a powerful tool for solving coin-weighing problems (and a lot of others!).

In this presentation, we hope to have demonstrated the power of probabilistic techniques in combinatorics.

References

1. Noga Alon and Joel H. Spencer. 2016. The Probabilistic Method (4th. ed.). Wiley Publishing.
2. Stirling's approximation, Wikipedia
3. MIT 18.310 lecture notes on Chernoff Bounds (https://math.mit.edu/ goemans/18310S15/chernoff-notes.pdf)


THANK YOU FOR YOUR ATTENTION