## APEC MEETING
## UNIVERSITY OF TSUKUBA, TOKYO 2020, FEB 11—12

# THE COLLECTOR'S PROBLEM

## an example of inciting to thinking

Ivan Vysotskiy, Moscow, MCCME
http://ptlab.mccme.ru

## HISTORICAL REVIEW

All parents meet a math problem given by their kid having got first Kinder Surprise with a toy inside. The collection can contain 12 princesses or 10 hippos or something even more exiting.

The immediate question – how many eggs do we need to buy to complete the collection?

This leads to a curious statistical and probabilistic model which is being originally simple grows to a disproportionally and offensively sophisticated generalization as we need two collections for two kids.

# HISTORICAL REVIEW

The problem was probably set in de Moivre's *De Mensura Sortis* (1712); mentioned by Laplace in his *Théorie analytique des probabilité* (1812).

In 1930-s the American company Dixie Cup used a genial marketing trick with a coupons inside bottle lids. Of course, one who collects the entire collection was promised to win a prize.

That's why the problem is known not only as a *Coupon Collector's Problem*, but also as the *Dixie Cup Problem*, and its generalization to two or more collections is called the *Double Dixie Cup Problem*.

# HISTORICAL REVIEW

- The *Classic (Single) Collector's Problem* is easy and could be a support in teaching geometrics series for 9th-graders

$$E\xi = nH_n$$

And in high school as a support of integrating:

$$E\xi \approx n\ln n + 0,577$$

Solution of the *Double Dixie Cup Problem* was obtained as an integral by Newman and Shepp in 1960. From their solution they also drew out an asymptotic formula

$$E\xi_2 = n(\ln n + \ln\ln n) + O(n)$$

The nature of $O(n)$ was examined by Paul Erdős

# HISTORICAL REVIEW

In 2019, the exact solution was obtained as a determinant of a certain matrix.

At the same time, the upper estimatior

$$\mathrm{E}\,\xi_2 < nH_n + nH_{H_n}$$

was proved.

However, for many related tasks there is not yet satisfactory solutions.

## WHY THE COLLECTOR'S PROBLEM?

Many of you do think now 'Why did Ivan decide to blow the dust off a relic math problem and sell it here at the APEC meeting?'

The deeper I immerse to the problem the clearer I see that it is one of the rarest diamonds of math which deserve not only to mentioned but still has unsolved aspects and lot of didactical power.

The problem forces to work all mathematics taught in school. It's various subproblems could serve as easily apprehensible lifelike motivation for developing what we call 'Statistical Thinking' for 7h-graders and higher.

The most precious feature is that the statistical thinking here is scarcely working without some probabilitsitc intuition and math knowledge.

# INTRODUCTION

- Use a die. When rolling the die, count how many casts have you done to get all six faces (each at least once).

1 4 4 3 4 5 1 6 3 3 2

This time I got it (the full collection of 1 to 6) with 11 trial

# INTRODUCTION

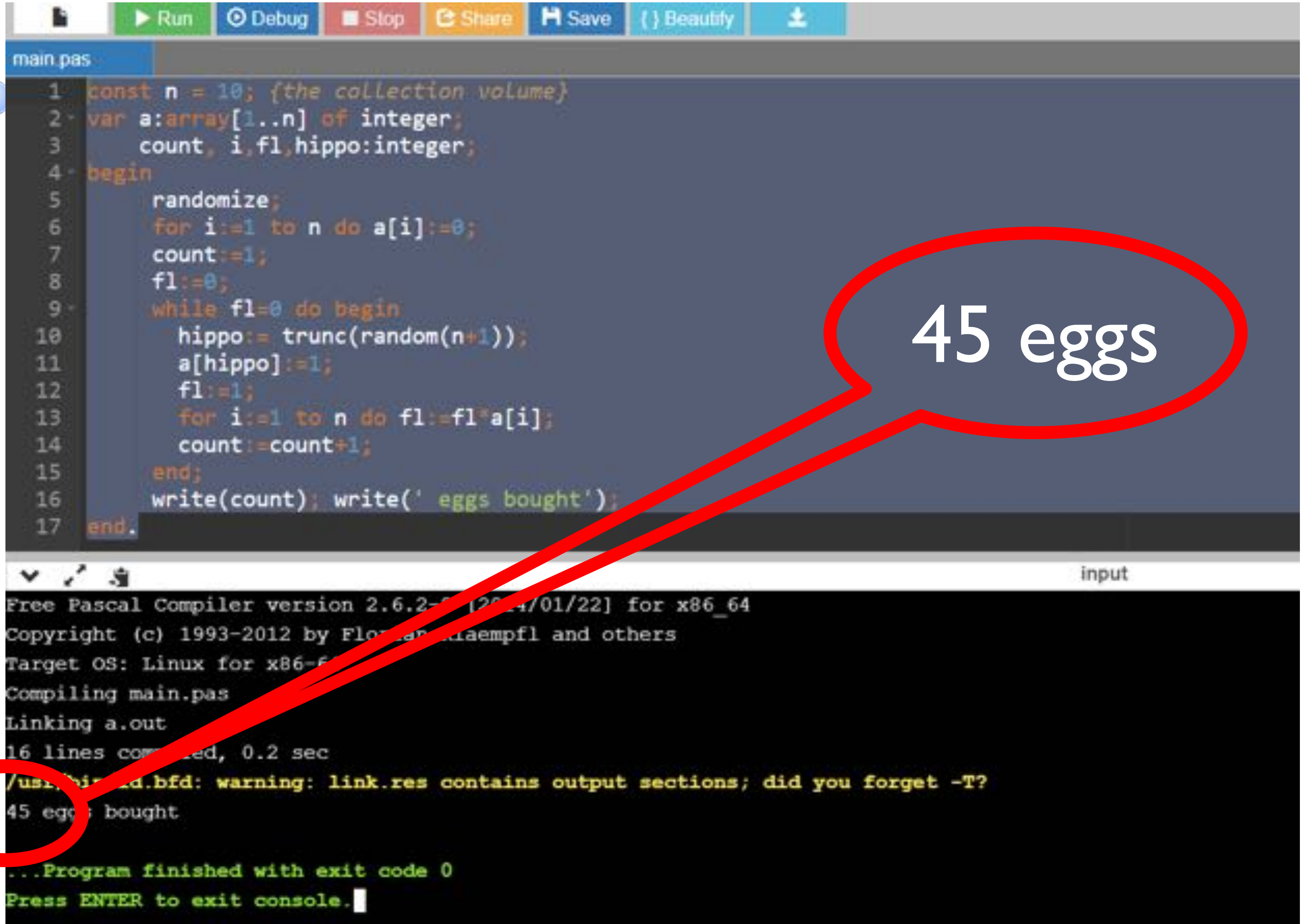
Now use a RNG (Excel or any other) with 10 random integers from 1 to 10.
Run it many times and find how many trials you have done before you got all ten possible

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| 1 | | | | | | |
| 2 | 1 | 2 | 11 | 8 | 21 | 5 |
| 3 | 2 | 1 | 12 | 10 | 22 | 2 |
| 4 | 3 | 6 | 13 | 9 | 23 | 8 |
| 5 | 4 | 7 | 14 | 4 | 24 | 1 |
| 6 | 5 | 4 | 15 | 6 | 25 | 6 |
| 7 | 6 | 9 | 16 | 7 | 26 | 6 |
| 8 | 7 | 3 | 17 | 3 | 27 | 1 |
| 9 | 8 | 4 | 18 | 5 | 28 | 10 |
| 10 | 9 | 3 | 19 | 10 | 29 | 9 |
| 11 | 10 | 5 | 20 | 7 | 30 | 7 |

# INTRODUCTION

```
const n = 10; {the collection volume}
var a:array[1..n] of integer;
   count, i,fl,hippo:integer;
begin
    randomize;
    for i:=1 to n do a[i]:=0;
    count:=1; fl:=0;
    while fl=0 do begin
      hippo:= trunc(random(n+1));
      a[hippo]:=1;
      fl:=1;
      for i:=1 to n do fl:=fl*a[i];
      count:=count+1;
    end;
    write(count); write(' eggs bought');
end.
```

# INTRODUCTION



# INTRODUCTION

- Find the average after many (let's say 100) such experiments.

You'll likely get a number close to

$$10\cdot\left(1+\frac{1}{2}+\frac{1}{3}+\ldots+\frac{1}{10}\right)\approx 29{,}3.$$

In the two experiments above

$$\frac{12+45}{2}=28{,}5.$$

# HUNTING HIPPOS. GRADE 7 - 8

You've got already 4 hippos out of 10. What is the probability to get a new hippo (#5) in the next egg bought?

You and your friend decided to collect to entire sets of hippos. The moment you totally harvested the first collection, some hippos still aren't in the second one. Prove:

$$\mathrm{P}\left(1\, hippo's\ missing\right) = \mathrm{P}\left(2\, hippos\ are\right)$$

Show that

$$\mathrm{P}(1) = \mathrm{P}(2) > \mathrm{P}(3) > \mathrm{P}(4) > \ldots > \mathrm{P}(10)$$

# HUNTING HIPPOS. GRADE 9 – 12

Find the math expectation of the number of missing hippos when the first collection is completed.

How many spare hippos # 1 in average will you have after the 1st collection is completed?

How many spare hippos # 3?

Solve the classic Collector's problem: show that to hold the full collection you need to buy in average

$$\mathrm{E}\xi = 10 \cdot \left(1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{10}\right) = 10H_{10}$$

chocolate eggs.

# HOW MANY TAXI METER CABS IN KHON KAEN?

*Russian doctor of math in Khon Kaen*
*Was of greatest sagacity man…*

One mathematician uses to use taxi meter in Khon Kaen as he often visits this city.

Being incredibly observant and having a memory like an elephant, during the 22nd trip, he noticed that he had already used this car once with the same driver.
But the cars have not been repeated before. How many Taxi Meters cars are there in Khon Kaen?

Can we model this problem using Python or Pascal?

# HOW MANY TAXI METER CABS IN KHON KAEN?

- For advanced ones, we can pose this task as a project or a mathematical study research. MM gives the equation

$$\sqrt{\frac{\pi x}{2}+\frac{1}{2}}=22 \quad \text{from which} \quad x \approx 294$$

# CONCLUSION

To get the estimator 294 for the Khon Kaen Taxi problem by the way we need to use combinatorics, calculus, deep ideas about Poisson distribution and Euler gamma-function, Stirling formula, Taylor series, chain fractions theory and something more.
That's why I call such problems 'generous' and didactically rich ones.

We saw only a small part of possible mathematical constructions connected to the Collector's problem.
I think that it's possible to built the whole mathematic course for 7 – 12 using exclusively questions being arisen from the 'Hunting Hippos'.

**Don't invent math problems.**
**Take them from the life. For free!**

# THANK YOU

What is it, Newman??
Shepp, old chap! It's the rarest one for my collection