

COMMONWEALTH OF AUSTRALIA

Copyright Regulations 1969

WARNING

This material has been reproduced and communicated to you by or on behalf of the University of Sydney pursuant to Part VB of the Copyright Act 1968 (**the Act**). The material in this communication may be subject to copyright under the Act. Any further copying or communication of this material by you may be the subject of copyright protection under the Act.

Do not remove this notice.

COMPx270: Randomised and Advanced Algorithms

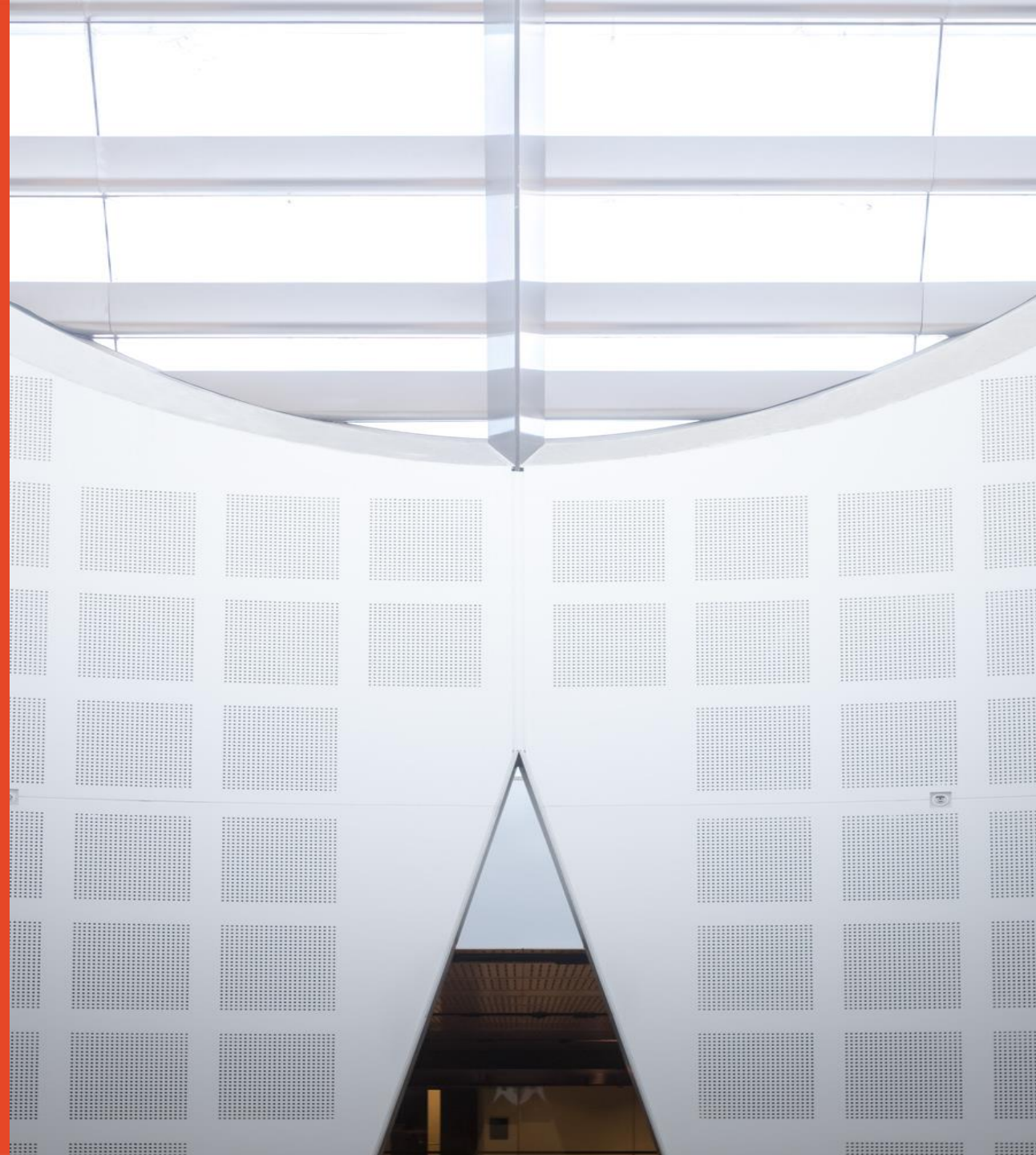
Lecture 3: Balls in Bins

Clément Canonne

School of Computer Science



THE UNIVERSITY OF
SYDNEY



A question 🎂

There are quite a few people in the classroom right now. What are the odds two of you (at least) have the same birthday?

A question 🎂

Theorem. (The 🎂 paradox) If you gather 23 people in a room, then with probability at least 50% a pair will sharing their birthday.

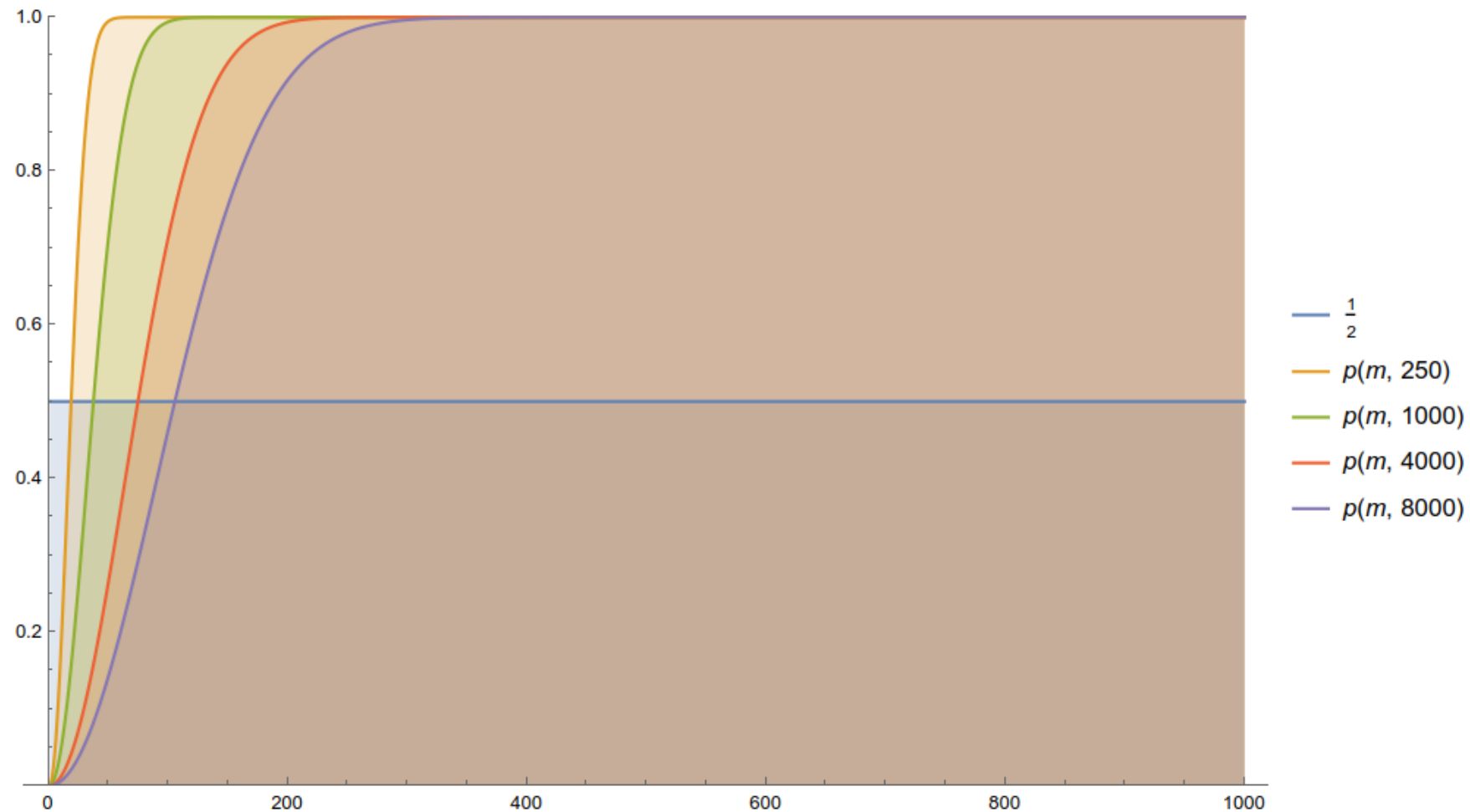
An answer? 🎂

Theorem. If you gather m people and give each a number uniform between 1 and n , then the probability $p(m, n)$ that at least two have the same number is...

$$p_{m,n} = 1 - \frac{n!}{n^m (n - m)!} = 1 - \frac{m!}{n^m} \binom{n}{m}$$

Proof.

```
DiscretePlot[{1/2, p[m, 250], p[m, 1000], p[m, 4000], p[m, 8000]}, {m, 1, 1000}, PlotRange -> {0, 1}, PlotLegends -> "Expressions"]
```



Let's start simple: $m=2$  

Now, for large values of "2"...

C: number of collisions when throwing m  into n . What is

$$c(m, n) = E[C]?$$

Number of collisions

... and what is $\text{Var}[C]$?



Number of collisions

... and what is $\text{Var}[C]$?



Number of collisions

$$\text{Var}[C] = \binom{m}{2} \frac{1}{n} \left(1 - \frac{1}{n} \right)$$

Number of collisions

$$\text{Var}[C] = \binom{m}{2} \frac{1}{n} \left(1 - \frac{1}{n} \right)$$

Now, we can use Chebyshev:

$$\Pr[X = 0] \leq 1/2$$

for $m = \Omega(\sqrt{n})$.

Number of collisions

$$\text{Var}[C] = \binom{m}{2} \frac{1}{n} \left(1 - \frac{1}{n} \right)$$

Now, we can use Chebyshev:

$$\Pr[X = 0] \leq 1/2$$

for $m = \Omega(\sqrt{n})$. Is it tight?

Number of collisions

$$\text{Var}[C] = \binom{m}{2} \frac{1}{n} \left(1 - \frac{1}{n} \right)$$

Now, we can use Chebyshev:

$$\Pr[X = 0] \leq 1/2$$

for $m = \Omega(\sqrt{n})$.

By Markov, we also have

$$\Pr[X \neq 0] = \Pr[X \geq 1] \leq E[X] \leq 1/2$$

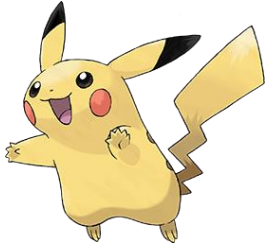
for $m = O(\sqrt{n})$.

Applications?

Bounding the variance: is it always that bad?

Two tricks (and even 3).

Coverage (Coupon collector)



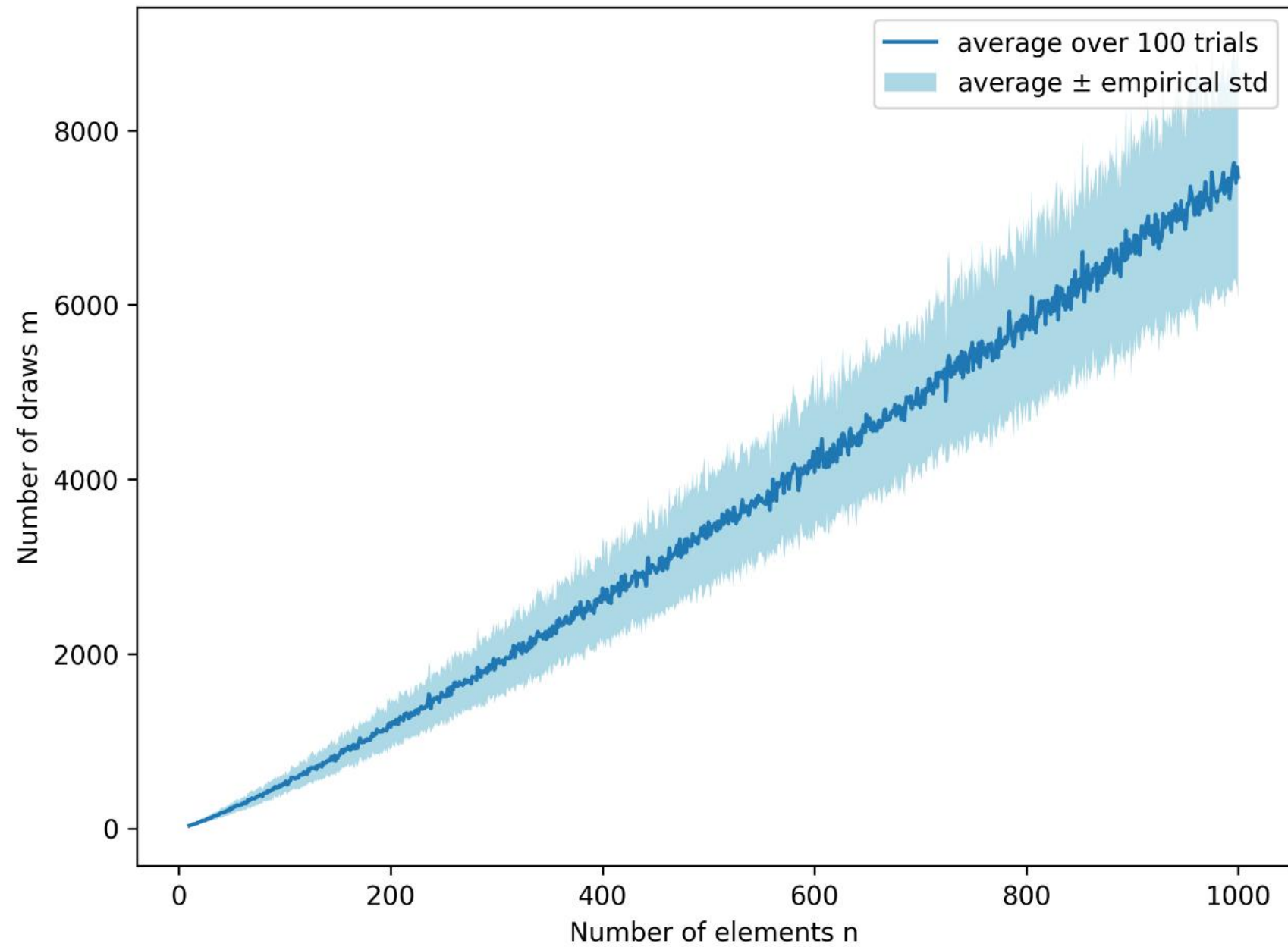
"What is the expected number of balls $M(n)$ we need to throw before each of the n bins contains **at least one** ball?"

Coverage (Coupon collector)



"What is the expected number of balls $M(n)$ we need to throw before each of the n bins contains **at least one** ball?"

- $\Theta(n)$?
- $\Theta(n \log n)$?
- $\Theta(n^2)$?
- Something else?



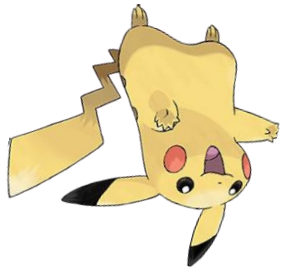
Coverage (Coupon collector)



"What is the expected number of balls $M(n)$ we need to throw before each of the n bins contains **at least one** ball?"

Theorem. In expectation, $M(n) = \Theta(n \log n)$ balls. (Even more precisely: $n \ln n + O(1)$.)

Proof.



What about the variance?



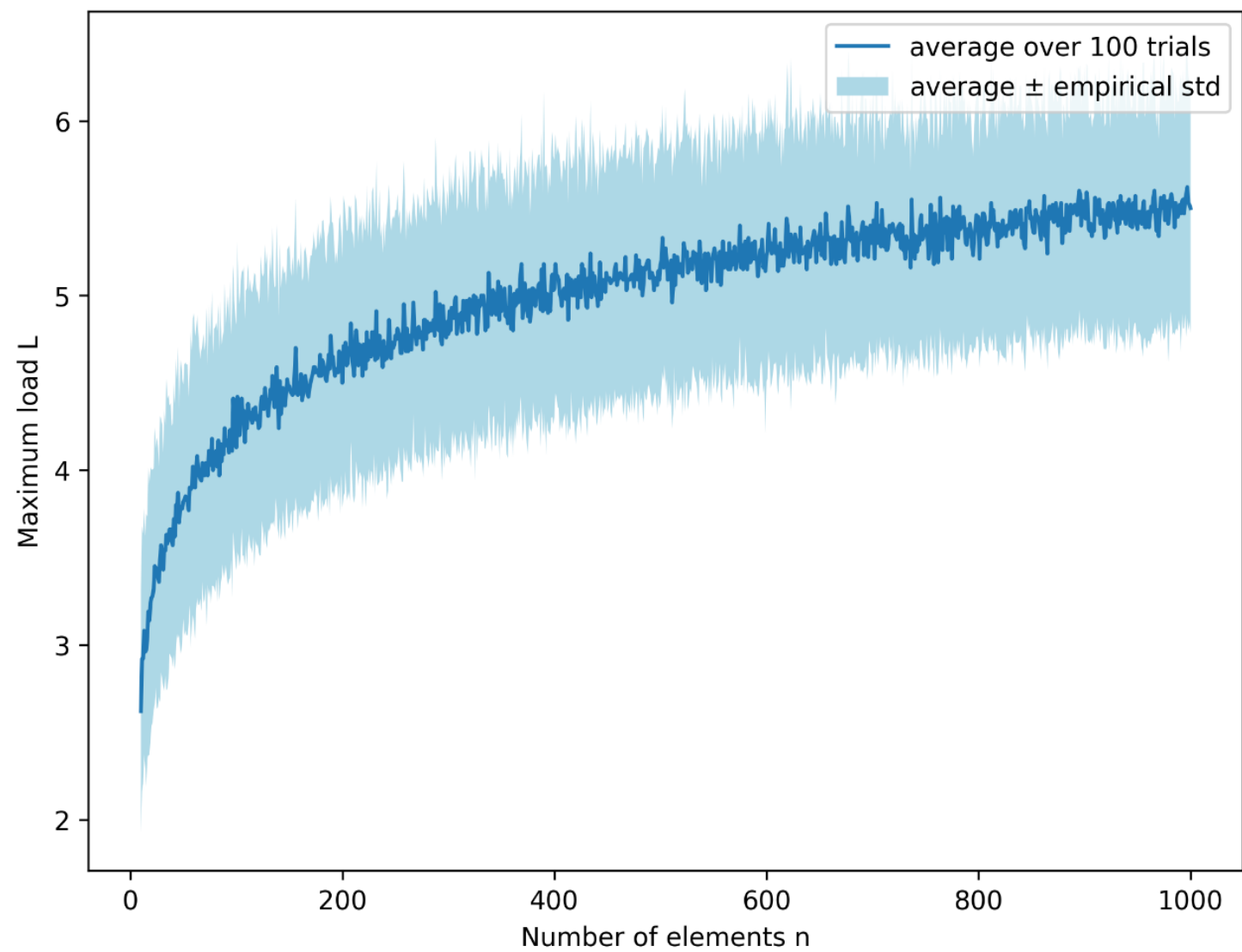
Load balancing

"What is the expected number of balls $L(n)$ the fullest of the n bins contains after throwing n balls?"

Load balancing

"What is the expected number of balls $L(n)$ the fullest of the n bins contains after throwing n balls?"

- $\Theta(1)$?
- $\Theta(\log n)$?
- $\Theta(\sqrt{n})$?
- Something else?



Load balancing

"What is the expected number of balls $L(n)$ the fullest of the n bins contains after throwing n balls?"

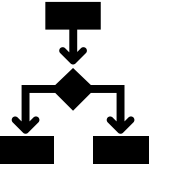
Theorem. The expected maximum load is $L(n) = \Theta(\log n / \log \log n)$.

Proof.

Proof.

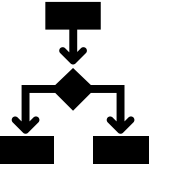
$$\binom{n}{k} \leq \left(\frac{n}{k}\right)^k \leq \left(\frac{e \cdot n}{k}\right)^k$$

Load balancing (a twist)



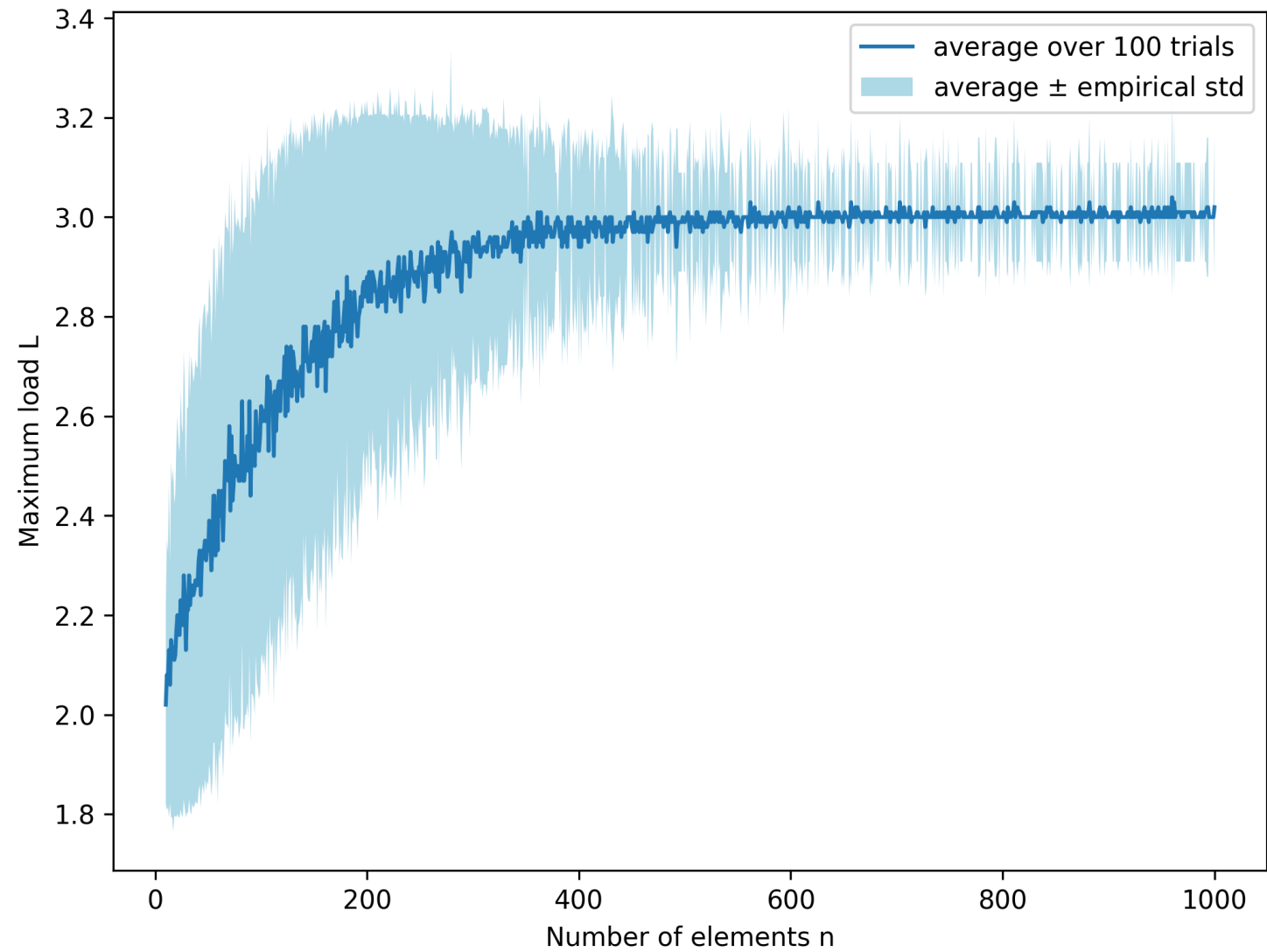
"Now, every time you throw a ball, it selects **two** bins at random, and goes to the least full of the two. What is the maximum expected load?"

Load balancing (a twist)

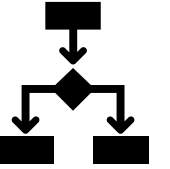


"Now, every time you throw a ball, it selects **two** bins at random, and goes to the least full of the two. What is the maximum expected load?"

- $\Theta(1)$?
- $\Theta(\log \log \mathbf{n})$?
- $\Theta(\sqrt{\log \mathbf{n}})$?
- Something else?



Load balancing (a twist)



"Now, every time you throw a ball, it selects **two** bins at random, and goes to the least full of the two. What is the maximum expected load?"

Theorem. The expected maximum load now $\Theta(\log \log n)$.

