The Birthday Paradox

June 10, 2020

Birthday Paradox

Let $m < n$. We figure out m, n later.
We will put m balls into n boxes uniformly at random.
Goal: How big does m have to be before the prob that some box has 2 balls is $\geq \frac{1}{2}$?

Birthday Paradox

Let $m < n$. We figure out m, n later.
We will put m balls into n boxes uniformly at random.
Goal: How big does m have to be before the prob that some box has 2 balls is $\geq \frac{1}{2}$?

We ask opp: What is prob that NO box has ≥ 2 balls?

Birthday Paradox

Let $m < n$. We figure out m, n later.
We will put m balls into n boxes uniformly at random.
Goal: How big does m have to be before the prob that some box has 2 balls is $\geq \frac{1}{2}$?

We ask opp: What is prob that NO box has ≥ 2 balls?

- Number of ways to put balls into boxes: n^m

Birthday Paradox

Let $m < n$. We figure out m, n later.
We will put m balls into n boxes uniformly at random.
Goal: How big does m have to be before the prob that some box has 2 balls is $\geq \frac{1}{2}$?

We ask opp: What is prob that NO box has ≥ 2 balls?

- Number of ways to put balls into boxes: n^m
- Number of ways to put balls into boxes: so that no box has ≥ 2 balls: $n(n-1)\cdots(n-m+1)$

Birthday Paradox

Let $m < n$. We figure out m, n later.
We will put m balls into n boxes uniformly at random.
Goal: How big does m have to be before the prob that some box has 2 balls is $\geq \frac{1}{2}$?

We ask opp: What is prob that NO box has ≥ 2 balls?

- Number of ways to put balls into boxes: n^m
- Number of ways to put balls into boxes: so that no box has ≥ 2 balls: $n(n-1)\cdots(n-m+1)$

The probability is

$$\frac{n(n-1)(n-2)\cdots(n-m+1)}{n^m}$$

Approx

$$\frac{n(n-1)(n-2)\cdots(n-m+1)}{n^m}$$

Approx

$$\frac{n(n-1)(n-2)\cdots(n-m+1)}{n^m}$$

$$= \frac{n}{n} \times \frac{n-1}{n} \times \frac{n-2}{n} \times \cdots \times \frac{n-m+1}{n}$$

Approx

$$\frac{n(n-1)(n-2)\cdots(n-m+1)}{n^m}$$

$$= \frac{n}{n} \times \frac{n-1}{n} \times \frac{n-2}{n} \times \cdots \times \frac{n-m+1}{n}$$

$$= 1 \times \left(1 - \frac{1}{n}\right) \times \left(1 - \frac{2}{n}\right) \times \cdots \times \left(1 - \frac{m-1}{n}\right)$$

Approx

$$\frac{n(n-1)(n-2)\cdots(n-m+1)}{n^m}$$

$$= \frac{n}{n} \times \frac{n-1}{n} \times \frac{n-2}{n} \times \cdots \times \frac{n-m+1}{n}$$

$$= 1 \times \left(1 - \frac{1}{n}\right) \times \left(1 - \frac{2}{n}\right) \times \cdots \times \left(1 - \frac{m-1}{n}\right)$$

Recall $e^{-x} \sim 1 - x$ for x small. So we have

Approx

$$\frac{n(n-1)(n-2)\cdots(n-m+1)}{n^m}$$

$$= \frac{n}{n} \times \frac{n-1}{n} \times \frac{n-2}{n} \times \cdots \times \frac{n-m+1}{n}$$

$$= 1 \times \left(1 - \frac{1}{n}\right) \times \left(1 - \frac{2}{n}\right) \times \cdots \times \left(1 - \frac{m-1}{n}\right)$$

Recall $e^{-x} \sim 1 - x$ for x small. So we have

$$\sim e^{-1/n} \times e^{-2/n} \times \cdots \times e^{-(m-1)/n} = e^{-(1/n)(1+2+\cdots+(m-1))}$$

Approx

$$\frac{n(n-1)(n-2)\cdots(n-m+1)}{n^m}$$

$$= \frac{n}{n} \times \frac{n-1}{n} \times \frac{n-2}{n} \times \cdots \times \frac{n-m+1}{n}$$

$$= 1 \times \left(1 - \frac{1}{n}\right) \times \left(1 - \frac{2}{n}\right) \times \cdots \times \left(1 - \frac{m-1}{n}\right)$$

Recall $e^{-x} \sim 1 - x$ for x small. So we have

$$\sim e^{-1/n} \times e^{-2/n} \times \cdots \times e^{-(m-1)/n} = e^{-(1/n)(1+2+\cdots+(m-1))}$$

$$\sim e^{-m^2/2n}$$

Actual Numbers!

If $m < n$ and you put m balls in n boxes at random then prob that ≥ 2 balls in same box is approx:

$$1 - e^{-m^2/2n}$$

Actual Numbers!

If $m < n$ and you put m balls in n boxes at random then prob that ≥ 2 balls in same box is approx:

$$1 - e^{-m^2/2n}$$

Recall Our goal is to find m such that prob of 2 in the same box is $\geq \frac{1}{2}$. Hence we need $1 - e^{-m^2/2n} > \frac{1}{2}$:

Actual Numbers!

If $m < n$ and you put m balls in n boxes at random then prob that ≥ 2 balls in same box is approx:

$$1 - e^{-m^2/2n}$$

Recall Our goal is to find m such that prob of 2 in the same box is $\geq \frac{1}{2}$. Hence we need $1 - e^{-m^2/2n} > \frac{1}{2}$:

$$e^{-m^2/2n} < \frac{1}{2}$$

Actual Numbers!

If $m < n$ and you put m balls in n boxes at random then prob that ≥ 2 balls in same box is approx:

$$1 - e^{-m^2/2n}$$

Recall Our goal is to find m such that prob of 2 in the same box is $\geq \frac{1}{2}$. Hence we need $1 - e^{-m^2/2n} > \frac{1}{2}$:

$$e^{-m^2/2n} < \frac{1}{2}$$

$$-\frac{m^2}{2n} < \ln(0.5) \sim -0.7$$

Actual Numbers!

If $m < n$ and you put m balls in n boxes at random then prob that ≥ 2 balls in same box is approx:

$$1 - e^{-m^2/2n}$$

Recall Our goal is to find m such that prob of 2 in the same box is $\geq \frac{1}{2}$. Hence we need $1 - e^{-m^2/2n} > \frac{1}{2}$:

$$e^{-m^2/2n} < \frac{1}{2}$$

$$-\frac{m^2}{2n} < \ln(0.5) \sim -0.7$$

$$\frac{m^2}{2n} > 0.7$$

Actual Numbers!

If $m < n$ and you put m balls in n boxes at random then prob that ≥ 2 balls in same box is approx:

$$1 - e^{-m^2/2n}$$

Recall Our goal is to find m such that prob of 2 in the same box is $\geq \frac{1}{2}$. Hence we need $1 - e^{-m^2/2n} > \frac{1}{2}$:

$$e^{-m^2/2n} < \frac{1}{2}$$

$$-\frac{m^2}{2n} < \ln(0.5) \sim -0.7$$

$$\frac{m^2}{2n} > 0.7$$

$$m > (1.4n)^{1/2}$$

Real Numbers!

If $m > (1.4n)^{1/2}$ and you put m balls in n boxes at random then prob that ≥ 2 balls in same box is over $\frac{1}{2}$.

Real Numbers!

If $m > (1.4n)^{1/2}$ and you put m balls in n boxes at random then prob that ≥ 2 balls in same box is over $\frac{1}{2}$.

$n = 365$.
$m = \lceil (1.4n)^{1/2} \rceil = 23$

Real Numbers!

If $m > (1.4n)^{1/2}$ and you put m balls in n boxes at random then prob that ≥ 2 balls in same box is over $\frac{1}{2}$.

$n = 365$.
$m = \lceil (1.4n)^{1/2} \rceil = 23$

Birthday Paradox: If there are 23 people in a room then prob two have the same birthday is $> \frac{1}{2}$.

Real Numbers!

If $m > (1.4n)^{1/2}$ and you put m balls in n boxes at random then prob that ≥ 2 balls in same box is over $\frac{1}{2}$.

$n = 365$.
$m = \lceil (1.4n)^{1/2} \rceil = 23$

Birthday Paradox: If there are 23 people in a room then prob two have the same birthday is $> \frac{1}{2}$.

How We Use: If $\sim n^{1/2}$ balls put into n boxes then prob 2 in same box is large.

Alternative Proof

Prob balls i, j in same box is $\frac{n}{n^2} = \frac{1}{n}$.
Prob balls i, j NOT in same box is $1 - \frac{1}{n}$.

Alternative Proof

Prob balls i, j in same box is $\frac{n}{n^2} = \frac{1}{n}$.
Prob balls i, j NOT in same box is $1 - \frac{1}{n}$.

Prob NO pair is in same box: Want to say $(1 - \frac{1}{n})^{\binom{m}{2}}$.

Alternative Proof

Prob balls i, j in same box is $\frac{n}{n^2} = \frac{1}{n}$.
Prob balls i, j NOT in same box is $1 - \frac{1}{n}$.

Prob NO pair is in same box: Want to say $(1 - \frac{1}{n})^{\binom{m}{2}}$.

Not quite. Would be true if they are all ind. But good approx.

Alternative Proof

Prob balls i, j in same box is $\frac{n}{n^2} = \frac{1}{n}$.
Prob balls i, j NOT in same box is $1 - \frac{1}{n}$.

Prob NO pair is in same box: Want to say $(1 - \frac{1}{n})^{\binom{m}{2}}$.

Not quite. Would be true if they are all ind. But good approx.

Prob NO pair is in same box $< (1 - \frac{1}{n})^{\binom{m}{2}} \sim e^{-m^2/2n}$.
Prob SOME pair is in same box $> 1 - e^{-m^2/2n}$.
Same as before.

Three Balls in a Box

Prob balls i, j, k in same box is $\frac{n}{n^3} = \frac{1}{n^2}$.
Prob balls i, j, k NOT in same box is $1 - \frac{1}{n^2}$.

Three Balls in a Box

Prob balls i, j, k in same box is $\frac{n}{n^3} = \frac{1}{n^2}$.
Prob balls i, j, k NOT in same box is $1 - \frac{1}{n^2}$.

Prob NO triple is in same box: $\sim (1 - \frac{1}{n^2})^{\binom{m}{3}} \sim e^{-m^3/6n^2}$
Prob SOME triple is in same box: $\sim 1 - e^{-m^3/6n^2}$

Real Numbers!

If $m < n$ and you put m balls in n boxes at random then prob that ≥ 3 balls in same box is approx:

$$1 - e^{-m^3/6n^2}$$

Real Numbers!

If $m < n$ and you put m balls in n boxes at random then prob that ≥ 3 balls in same box is approx:

$$1 - e^{-m^3/6n^2}$$

To get this $> \frac{1}{2}$ need $1 - e^{-m^3/6n^2} > \frac{1}{2}$

$$e^{-m^3/6n^2} < \frac{1}{2}$$

Real Numbers!

If $m < n$ and you put m balls in n boxes at random then prob that ≥ 3 balls in same box is approx:

$$1 - e^{-m^3/6n^2}$$

To get this $> \frac{1}{2}$ need $1 - e^{-m^3/6n^2} > \frac{1}{2}$

$$e^{-m^3/6n^2} < \frac{1}{2}$$

$$-\frac{m^3}{6n^2} < \ln(0.5) \sim -0.7$$

Real Numbers!

If $m < n$ and you put m balls in n boxes at random then prob that ≥ 3 balls in same box is approx:

$$1 - e^{-m^3/6n^2}$$

To get this $> \frac{1}{2}$ need $1 - e^{-m^3/6n^2} > \frac{1}{2}$

$$e^{-m^3/6n^2} < \frac{1}{2}$$

$$-\frac{m^3}{6n^2} < \ln(0.5) \sim -0.7$$

$$m > (4.2n)^{2/3}$$

Real Numbers!

If $m > (4.2n)^{2/3}$ and you put m balls in n boxes at random then prob that ≥ 3 balls in same box is over $\frac{1}{2}$.

Real Numbers!

If $m > (4.2n)^{2/3}$ and you put m balls in n boxes at random then prob that ≥ 3 balls in same box is over $\frac{1}{2}$.

$n = 365$.
$m = \lceil (4.2n)^{2/3} \rceil = 82$

Real Numbers!

If $m > (4.2n)^{2/3}$ and you put m balls in n boxes at random then prob that ≥ 3 balls in same box is over $\frac{1}{2}$.

$n = 365$.
$m = \lceil (4.2n)^{2/3} \rceil = 82$

Birthday Paradox: $n = 365$ then need $m \geq 82$. SO if 82 people in a room prob is $> \frac{1}{2}$ that three have same bday!

Real Numbers!

If $m > (4.2n)^{2/3}$ and you put m balls in n boxes at random then prob that ≥ 3 balls in same box is over $\frac{1}{2}$.

$n = 365$.
$m = \lceil (4.2n)^{2/3} \rceil = 82$

Birthday Paradox: $n = 365$ then need $m \geq 82$. SO if 82 people in a room prob is $> \frac{1}{2}$ that three have same bday!

How We Use: If $\sim n^{2/3}$ balls put into n boxes then prob 3 in same box is large.

Recap and Generalize

1. $\sim n^{1/2}$ balls put into n boxes, prob 2 in same box.
2. $\sim n^{2/3}$ balls put into n boxes, prob 3 in same box.
3. $\sim n^{3/4}$ balls put into n boxes, prob 4 in same box.
4. $\sim n^{(k-1)/k}$ balls put into n boxes, prob k in same box.

Recap and Generalize

1. $\sim n^{1/2}$ balls put into n boxes, prob 2 in same box.
2. $\sim n^{2/3}$ balls put into n boxes, prob 3 in same box.
3. $\sim n^{3/4}$ balls put into n boxes, prob 4 in same box.
4. $\sim n^{(k-1)/k}$ balls put into n boxes, prob k in same box.

Caveat: The approx we used only works when $k \ll n$.

Recap and Generalize

1. $\sim n^{1/2}$ balls put into n boxes, prob 2 in same box.
2. $\sim n^{2/3}$ balls put into n boxes, prob 3 in same box.
3. $\sim n^{3/4}$ balls put into n boxes, prob 4 in same box.
4. $\sim n^{(k-1)/k}$ balls put into n boxes, prob k in same box.

Caveat: The approx we used only works when $k \ll n$.
Intent: The above is intended for use when the number of balls is small. What happens when the number of balls is large? Do many boxes get many elements in them?

Recap and Generalize

We state the following informally:

Theorem: Let $n \ll N$. There will be n boxes. There are N balls. The balls are put into the boxes randomly. Then, with high probability, MANY boxes will have MANY balls in them.