

BIRTHDAY PARADOX

DISCRETE STRUCTURES II

DARRYL HILL

BASED ON THE TEXTBOOK:

DISCRETE STRUCTURES FOR COMPUTER SCIENCE: COUNTING,
RECURSION, AND PROBABILITY

BY MICHIEL SMID

Birthday Paradox

Let d = number of days in a year

(365, 366, 400, it doesn't matter, we can generalize to any number)

$n \geq 2$ is the number of people. Each has a birthday on precisely one day, and each person's birthday is chosen uniformly at random.

The *Birthday Paradox* is the answer to the question, what is the probability that 2 people were born on the same day?

To find the probability we establish the **Sample Space** and define the **Event**.

What is the **Sample Space**? It is all possible combinations of n birthdays.

Let $b_1 \in \{1..d\}$ be a (birth)day.

Let $b_2 \in \{1..d\}$ be a (birth)day.

...

Let $b_n \in \{1..d\}$ be a birthday.

An **Outcome** of the **Sample Space** would be an n -tuple: $(b_1, b_2, b_3, \dots, b_n)$. The **Sample Space** is all possible **Outcomes**.

Formally: $S = \{(b_1, b_2, b_3, \dots, b_n) : \text{for each } b_i \in \{1, \dots, d\}\}$

Birthday Paradox

For example, let $d = 365$ and let $n = 3$. So there are 365 days in a year and 3 people.

The **Sample Space**?

Let $b_1 \in \{1..365\}$ be a birthday.

Let $b_2 \in \{1..365\}$ be a birthday.

Let $b_3 \in \{1..365\}$ be a birthday.

The **Sample Space** is all possible combinations:

(b_1, b_2, b_3)

Or more formally:

$$S = \{(1,1,1), (1,1,2), \dots (55,23,1), (55,23,2), \dots (365, 365, 365)\}$$

Any three people P_1, P_2, P_3 have birthdays (b_1, b_2, b_3) where $(b_1, b_2, b_3) \in S$.

So if $(b_1, b_2, b_3) = (55, 55, 10)$ for example, then P_1 and P_2 have the same birthday.

Birthday Paradox

For example, let $d = 365$ and let $n = 3$. So there are 365 days in a year and 3 people.

The **Sample Space**?

Let $b_1 \in \{1..365\}$ be a birthday.

Let $b_2 \in \{1..365\}$ be a birthday.

Let $b_3 \in \{1..365\}$ be a birthday.

The **Sample Space** is all possible combinations:

(b_1, b_2, b_3)

Or more formally:

$$S = \{(1,1,1), (1,1,2), \dots (55,23,1), (55,23,2), \dots (365, 365, 365)\}$$

How many elements in S ?

Product Rule:

Task 1: choose one of 365 days for the first birthday

Task 2: choose one of 365 days for the second birthday

Task 3: choose one of 365 days for the third birthday

$$|S| = 365 \cdot 365 \cdot 365 = 365^3$$

Birthday Paradox

Let d = the number of days in a year.

Let n = the number of people.

In general, the **Sample Space** consists of all possible n -tuples:

$$S = \{(b_1, b_2, b_3, \dots, b_n) : \text{for each } b_i \in \{1, \dots, d\}\}$$

Or, if you prefer set notation,
let $B = \{1, 2, \dots, d\}$. Then

$$S = B \times B \times B \times \dots \times B$$

What is $|S|$?

We can “build” the elements of S using the procedure:

Task 1 : Choose one of $\{1, \dots, d\}$ for b_1

Task 2 : Choose one of $\{1, \dots, d\}$ for b_2

...

Task n : Choose one of $\{1, \dots, d\}$ for b_n

$$(b_1, b_2, b_3, \dots, b_n)$$

Thus the number of elements in S is

$$|S| = d \cdot d \cdot \dots \cdot d = d^n$$

Birthday Paradox

Let d = the number of days in a year.

Let n = the number of people.

In general, the **Sample Space** consists of all possible n -tuples:

$$S = \{(b_1, b_2, b_3, \dots, b_n) : \text{for each } b_i \in \{1, \dots, d\}\}$$

$$|S| = d^n$$

Now we define the **Event**:

Let **Event** A_n = “out of n people, ≥ 2 people have the same birthday”

We want to find $\Pr(A_n)$.

Since all of S has uniform probability,

$$\Pr(A_n) = \frac{|A_n|}{|S|}.$$

We are left with finding $|A_n|$.

Birthday Paradox

Let d = the number of days in a year.

Let n = the number of people.

In general, the **Sample Space** consists of all possible n -tuples:

$$S = \{(b_1, b_2, b_3, \dots, b_n) : \text{for each } b_i \in \{1, \dots, d\}\}$$

$$|S| = d^n$$

For $n = 2$:

$$A_2 = \{(1,1), (2,2), (3,3), \dots, (d,d)\}$$

$$|A_2| = d$$

$$\Pr(A_2) = \frac{|A_2|}{|S|} = \frac{d}{d^2} = \frac{1}{d}$$

So if there were 2 people and 365 days, the probability that they have the same birthday is

$$\Pr(2 \text{ people have the same birthday}) = \frac{1}{365}$$

Birthday Paradox

Let d = the number of days in a year.

Let n = the number of people.

In general, the **Sample Space** consists of all possible n -tuples:

$$S = \{(b_1, b_2, b_3, \dots, b_n) : \text{for each } b_i \in \{1, \dots, d\}\}$$

$$|S| = d^n$$

For $n = d + 1$ what can we say?

(Each element in S is a tuple of $d + 1$ elements chosen from a set of d elements).

For example, if S were the days of the week, so $|S| = 7$, and there were 8 people, then an element of S might look like:

(1,2,3,4,5,6,7,5)

For $n = d + 1$

$\Pr(2 \text{ people have the same birthday}) = 1$

Birthday Paradox

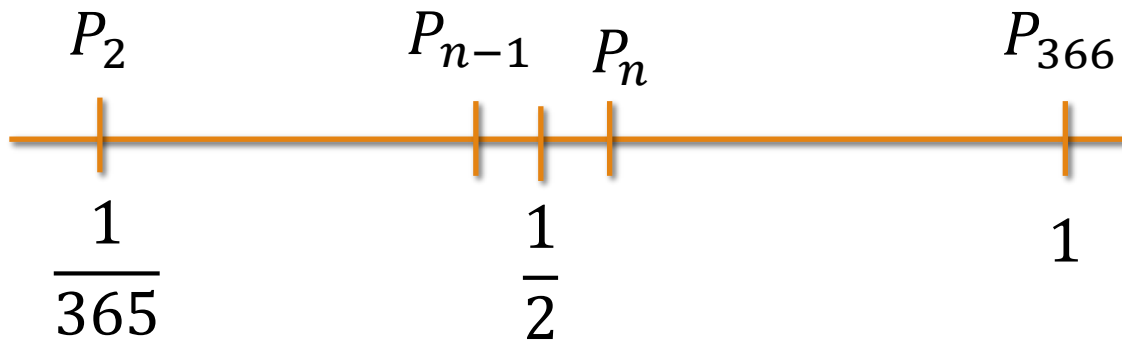
Let d = the number of days in a year.

Let n = the number of people.

$$|S| = d^n$$

Let **Event** A_n = "out of n people, ≥ 2 people have the same birthday"

$$\text{Let } P_n = \Pr(A_n)$$



What is n ?

$$2 \leq n \leq d$$

Difficult to compute size of A_n .

What do we do when $|A_n|$ is difficult to count?

We can try counting the complement

$\overline{A_n}$ = "n people with different birthdays"

Birthday Paradox

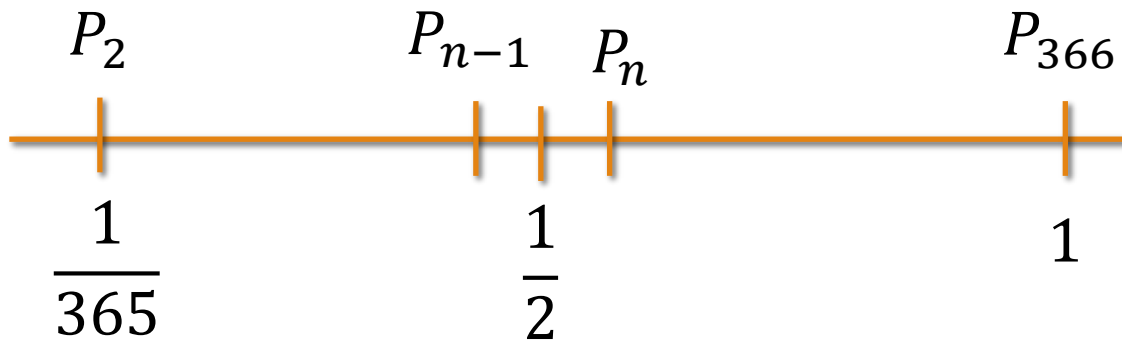
Let d = the number of days in a year.

Let n = the number of people.

$$|S| = d^n$$

Let **Event** A_n = “out of n people, ≥ 2 people have the same birthday”

$$\text{Let } P_n = \Pr(A_n)$$



$\overline{A_n}$ is the set consisting of all sequences (b_1, b_2, \dots, b_n) of distinct numbers chosen from $\{1, \dots, d\}$.

To count $\overline{A_n}$ we can use the Product Rule

Our Procedure is:

Task 1: choose a number for b_1 . d ways

Task 2: choose a number for b_2 . $d - 1$ ways

Task 3: choose a number for b_3 . $d - 2$ ways

...

Task n : choose a number for b_n . $d - n + 1$

Birthday Paradox

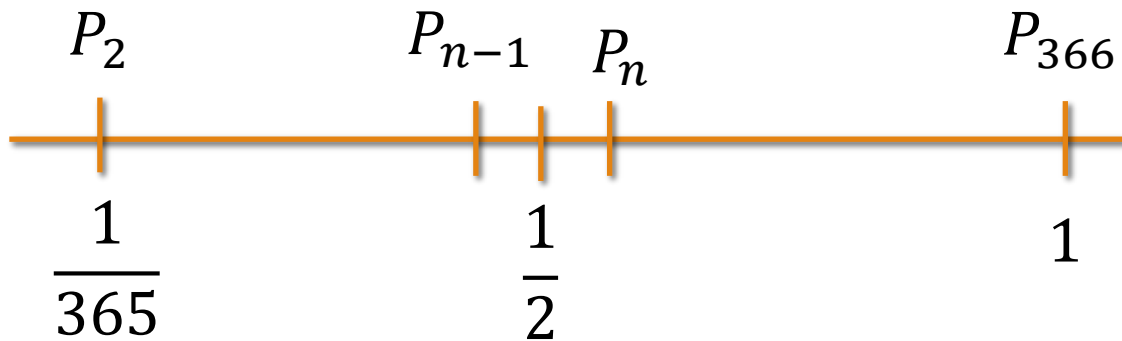
Let d = the number of days in a year.

Let n = the number of people.

$$|S| = d^n$$

Let **Event** A_n = “out of n people, ≥ 2 people have the same birthday”

$$\text{Let } P_n = \Pr(A_n)$$



$\overline{A_n}$ is the set consisting of all sequences (b_1, b_2, \dots, b_n) of distinct numbers chosen from $\{1, \dots, d\}$.

To count \overline{A} we can use the Product Rule

$$|\overline{A}| = d \cdot (d - 1) \cdot \dots \cdot (d - n + 1)$$

$$= \frac{d!}{(d - n)!}$$

Birthday Paradox

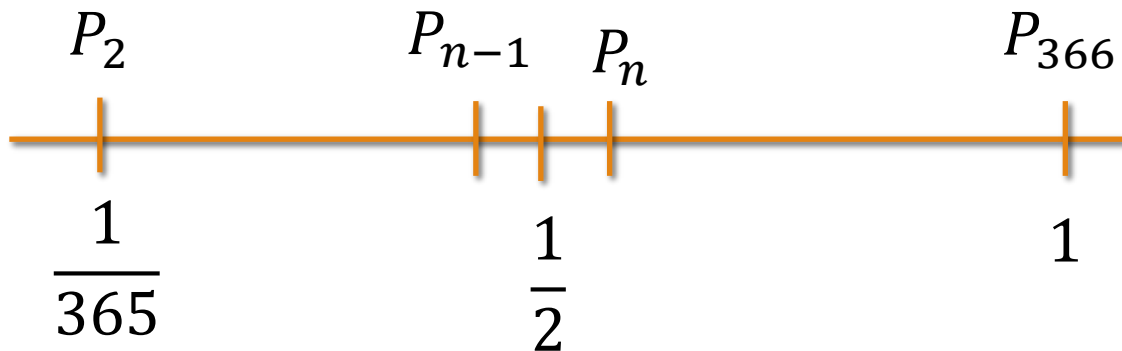
Let d = the number of days in a year.

Let n = the number of people.

$$|S| = d^n$$

Let **Event** A_n = “out of n people, ≥ 2 people have the same birthday”

$$\text{Let } P_n = \Pr(A_n)$$



$$\begin{aligned} |\bar{A}| &= d \cdot (d - 1) \cdot \dots \cdot (d - n + 1) \\ &= \frac{d!}{(d-n)!} \end{aligned}$$

$$\begin{aligned} \Pr(A_n) &= 1 - \Pr(\bar{A}) \\ &= 1 - \frac{|\bar{A}|}{|S|} \\ &= 1 - \frac{d!}{d^n(d-n)!} \end{aligned}$$

As a sanity check, we can plug in $n = 2$:

$$\begin{aligned} &= 1 - \frac{365!}{365^2(365 - 2)!} = \frac{365!}{365^2 363!} \\ &= 1 - \frac{364}{365} = \frac{1}{365} \end{aligned}$$

Birthday Paradox

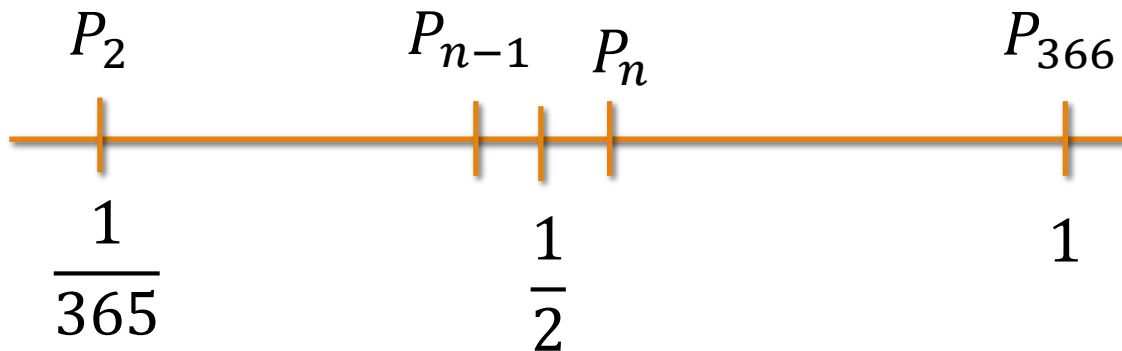
Let d = the number of days in a year.

Let n = the number of people.

$$|S| = d^n$$

Let **Event** A_n = “out of n people, ≥ 2 people have the same birthday”

$$\text{Let } P_n = \Pr(A_n)$$



$$\begin{aligned}\Pr(A_n) &= 1 - \Pr(\bar{A}) \\ &= 1 - \frac{|\bar{A}|}{|S|} \\ &= 1 - \frac{d!}{d^n(d-n)!}\end{aligned}$$

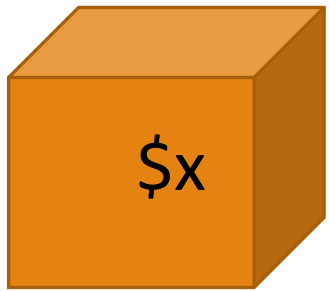
So what is n ? Assuming $d = 365$, we can plug in values for n . If you take $n = 23$ you get

$$\begin{aligned}&1 - \frac{365!}{365^n(365-n)!} \\ &1 - \frac{365!}{365^{23}(365-23)!} \\ &\approx 0.5073\end{aligned}$$

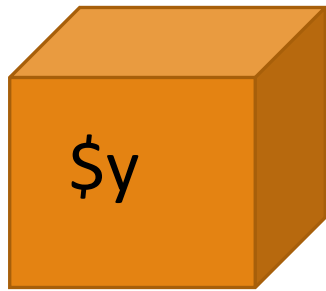
Secret Box

We have two boxes.

Each box has an amount of money,
\$ x or \$ y



or



\$ y

\$ x

The game is this:

Step 1: choose box, count \$'s.

Step 2: decide to keep the box or switch.

Prize money is chosen from a set $A = \{0, 1, 2, 3, \dots, 100\}$ representing dollar amounts.

Thus $x \in A$ and $y \in A$, and $x < y$.

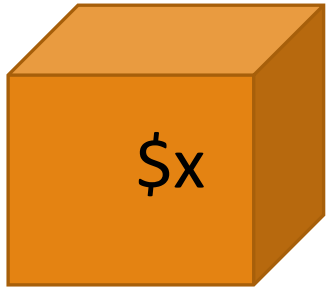
These are not random amounts. They are chosen.

Open box 1, find \$47. Keep or switch?

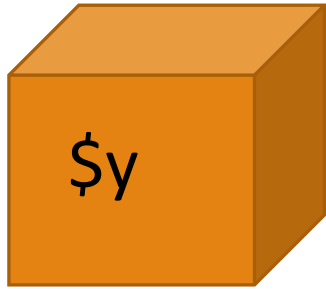
Secret Box

Two boxes, one with $\$x$ one with $\$y$,
 $x < y$ and $x, y \in A$, where

$$A = \{0, 1, 2, 3, \dots, 100\}$$



or



$\$y$

$\$x$

Step 1: choose box, count $\$$'s.

Step 2: decide to keep the box or
switch

Open box 1, find \$47. Keep or switch?

What is a good strategy to find the box with the most money? What can we take advantage of?

Flip a coin? Then our chance of success is $\frac{1}{2}$.

Call the box with $\$y$ the Big Box.

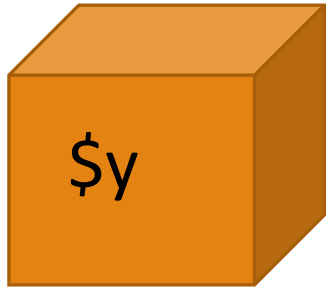
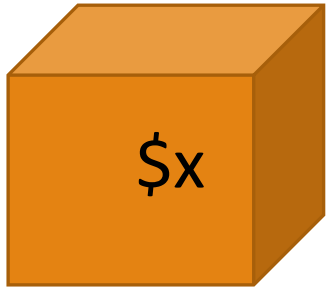
Claim: \exists Algorithm $\Pr(\text{find Big Box}) \geq 0.505$

Depending on $\$x$ and $\$y$, perhaps we do even better.

Secret Box

Two boxes, one with $\$x$ one with $\$y$,
 $x < y$ and $x, y \in A$, where

$$A = \{0, 1, 2, 3, \dots, 100\}$$



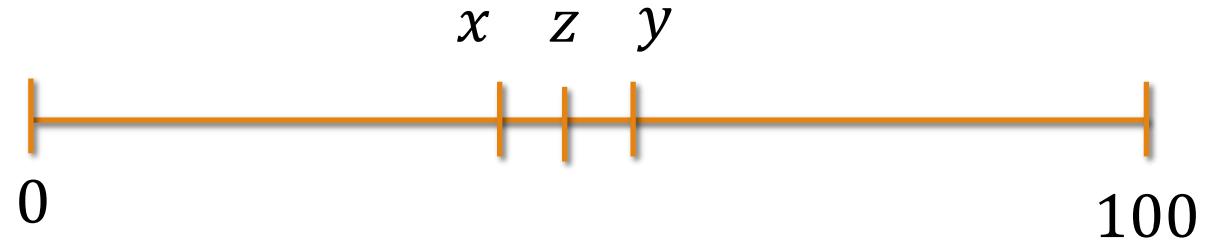
or

$\$y$

$\$x$

Step 1: choose box, count $\$$'s.

Step 2: decide to keep the box or
switch



Let's make an unrealistic assumption.

Assume we know an integer z where $x < z < y$

If we know z then we have a viable strategy.

Step 1: if the box we open has $> \$z$: keep box
if the box we open has $< \$z$: switch

if we know z : $\Pr(\text{find BB}) = 1$

Trick: choose random z

Secret Box

$$A = \{0, 1, 2, 3, \dots, 100\}$$

I choose secret $x, y \in A, x < y$

We want to make sure z is between x and y

$$\text{Let } B = \{0.5, 1.5, 2.5, \dots, 99.5\}$$

$$|B| = 100$$

We will choose z uniformly at random from B .

Now our algorithm is as follows:

1.1 Choose a random box to open

Let a = amount of money found.

1.2 Choose random $z \in B$,

$$B = \{0.5, 1.5, 2.5, \dots, 99.5\}$$

2. If $a > z$: keep box

if $a < z$: switch boxes

We know the problem and we know our strategy, so let's analyze this algorithm.

If BB is the event that we select the Big Box, we want to determine $\Pr(BB)$.

Secret Box

$$A = \{0, 1, 2, 3, \dots, 100\}$$

Choose secret $x, y \in A, x < y$

Let $B = \{0.5, 1.5, 2.5, \dots, 99.5\}$

$$|B| = 100$$

1.1 Choose random box, a = amount

1.2 Choose random $z \in B$

2 If $a > z$: keep box

if $a < z$: switch

What is the sample space?

Can think of the sample space as input into our algorithm. We have two inputs:

a : is the money found in the random box.

Either $a = x$ or $a = y$.

z : the random number chosen.

$$z \in B = \{0.5, 1.5, \dots, 99.5\}$$

$$S = \{(a, z): a \in \{x, y\}, z \in B\}$$

$$|S| = 2 \cdot 100 = 200$$

Secret Box

$$A = \{0, 1, 2, 3, \dots, 100\}$$

Choose secret $x, y \in A, x < y$

Let $B = \{0.5, 1.5, 2.5, \dots, 99.5\}$

$$|B| = 100$$

1.1 Choose random box, a = amount

1.2 Choose random $z \in B$

2 If $a > z$: keep box

if $a < z$: switch

$$S = \{(a, z): a \in \{x, y\}, z \in B\}$$

$$|S| = 2 \cdot 100 = 200$$

Step 1.1 and 1.2: choose random element in S

We find Big Box if

$a = x$ and $z > a$ or

$a = y$ and $z < a$

$$BB = \{(a, z): a = x \text{ and } z > a \mid \mid \\ a = y \text{ and } z < a\}$$

$$\Pr(BB) = \frac{|BB|}{|S|}$$

Secret Box

$$A = \{0, 1, 2, 3, \dots, 100\}$$

Choose secret $x, y \in A, x < y$

Let $B = \{0.5, 1.5, 2.5, \dots, 99.5\}$

$$|B| = 100$$

1.1 Choose random box, a = amount

1.2 Choose random $z \in B$

2 If $a > z$: keep box

if $a < z$: switch

$$S = \{(a, z): a \in \{x, y\}, z \in B\}$$

$$|S| = 2 \cdot 100 = 200$$

Step 1.1 and 1.2: choose random element in S

We find Big Box if

$a = x$ and $z > a$ or BB_1

$a = y$ and $z < a$ BB_2

Since these are disjoint sets, we can use the sum rule.

$$|BB| = |BB_1| + |BB_2|$$

$$|BB_1| = ?$$

Secret Box

$$A = \{0, 1, 2, 3, \dots, 100\}$$

Choose secret $x, y \in A, x < y$

Let $B = \{0.5, 1.5, 2.5, \dots, 99.5\}$

$$|B| = 100$$

1.1 Choose random box, $a = \text{amount}$

1.2 Choose random $z \in B$

2 If $a > z$: keep box

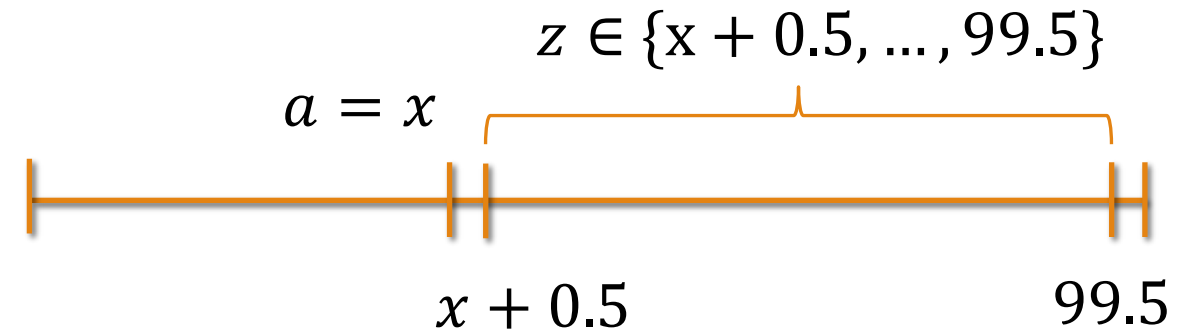
if $a < z$: switch

$$S = \{(a, z) : a \in \{x, y\}, z \in B\}$$

$$|S| = 2 \cdot 100 = 200$$

Step 1.1 and 1.2: choose random element in S

$BB_1 =$ "The event that $x = a$ and $z > a$ " :



z is the range of
 $\{x + 0.5, x + 1.5, \dots, 98.5, 99.5\}$.

If we add 0.5 to each, z is in the range of:
 $\{x + 1, \dots, 100\}$

$$\text{Therefore } |BB_1| = 100 - x$$

Secret Box

$$A = \{0, 1, 2, 3, \dots, 100\}$$

Choose secret $x, y \in A, x < y$

Let $B = \{0.5, 1.5, 2.5, \dots, 99.5\}$

$$|B| = 100$$

1.1 Choose random box, a = amount

1.2 Choose random $z \in B$

2 If $a > z$: keep box

if $a < z$: switch

$$S = \{(a, z): a \in \{x, y\}, z \in B\}$$

$$|S| = 2 \cdot 100 = 200$$

Step 1.1 and 1.2: choose random element in S

We find Big Box if

$$a = x \text{ and } z > a \text{ or } BB_1$$

$$a = y \text{ and } z < a \quad BB_2$$

Since these are disjoint sets, we can use the sum rule.

$$|BB| = |BB_1| + |BB_2|$$

$$|BB_1| = 100 - x$$

$$|BB_2| = ?$$

Secret Box

$$A = \{0, 1, 2, 3, \dots, 100\}$$

Choose secret $x, y \in A, x < y$

Let $B = \{0.5, 1.5, 2.5, \dots, 99.5\}$

$$|B| = 100$$

1.1 Choose random box, $a = \text{amount}$

1.2 Choose random $z \in B$

2 If $a > z$: keep box

if $a < z$: switch

$$S = \{(a, z) : a \in \{x, y\}, z \in B\}$$

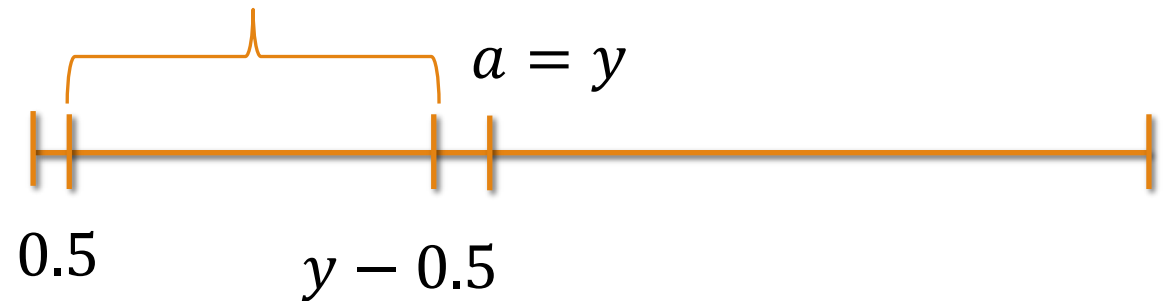
$$|S| = 2 \cdot 100 = 200$$

Step 1.1 and 1.2: choose random element in S

We find Big Box if

$BB_2 = \text{"the event that } a = y \text{ and } z < a\text{"}$

$$z \in \{0.5, \dots, y - 0.5\}$$



If we add 0.5 to $\{0.5, \dots, y - 0.5\}$ we have $\{1, \dots, y\}$, so there are y possible elements to choose z from and $|BB_2| = y$

Secret Box

$$A = \{0, 1, 2, 3, \dots, 100\}$$

Choose secret $x, y \in A, x < y$

Let $B = \{0.5, 1.5, 2.5, \dots, 99.5\}$

$$|B| = 100$$

1.1 Choose random box, a = amount

1.2 Choose random $z \in B$

2 If $a > z$: keep box

if $a < z$: switch

$$S = \{(a, z): a \in \{x, y\}, z \in B\}$$

$$|S| = 2 \cdot 100 = 200$$

Step 1.1 and 1.2: choose random element in S

We find Big Box if

$$a = x \text{ and } z > a \text{ or } BB_1$$

$$a = y \text{ and } z < a \quad BB_2$$

Since these are disjoint sets, we can use the sum rule.

$$|BB| = |BB_1| + |BB_2|$$

$$|BB_1| = 100 - x$$

$$|BB_2| = y$$

Secret Box

$$A = \{0, 1, 2, 3, \dots, 100\}$$

Choose secret $x, y \in A, x < y$

Let $B = \{0.5, 1.5, 2.5, \dots, 99.5\}$

$$|B| = 100$$

1.1 Choose random box, a = amount

1.2 Choose random $z \in B$

2 If $a > z$: keep box

if $a < z$: switch

$$S = \{(a, z): a \in \{x, y\}, z \in B\}$$

$$|S| = 2 \cdot 100 = 200$$

$$BB = BB_1 \cup BB_2$$

$$|BB| = |BB_1| + |BB_2|$$

$$|BB| = 100 - x + y$$

$$\begin{aligned} \text{Then } \Pr(BB) &= \frac{|BB|}{|S|} \\ &= \frac{100 - x + y}{200} \end{aligned}$$

$$= \frac{1}{2} + \frac{y - x}{200}$$

$$\geq \frac{1}{2} + \frac{1}{200} = 0.505$$

But notice the larger $y - x$, the higher the probability that we win.

Secret Box

$$A = \{0, 1, 2, 3, \dots, 100\}$$

Choose secret $x, y \in A, x < y$

Let $B = \{0.5, 1.5, 2.5, \dots, 99.5\}$

$$|B| = 100$$

1.1 Choose random box, a = amount

1.2 Choose random $z \in B$

2 If $a > z$: keep box
if $a < z$: switch

$$S = \{(a, z): a \in \{x, y\}, z \in B\}$$

$$|S| = 2 \cdot 100 = 200$$

$$\Pr(\text{find BB}) = \frac{(100-x)+y}{200}$$

$$= \frac{1}{2} + \frac{y-x}{200}$$

$$\geq \frac{1}{2} + \frac{1}{200} = 0.505$$

This is a lower bound on the probability.

We always find the big box if z is between x and y . That means the exact probability depends on the value $y - x$.

Secret Box

$$A = \{0, 1, 2, 3, \dots, 100\}$$

Choose secret $x, y \in A, x < y$

Let $B = \{0.5, 1.5, 2.5, \dots, 99.5\}$

$$|B| = 100$$

1.1 Choose random box, a = amount

1.2 Choose random $z \in B$

2 If $a > z$: keep box

if $a < z$: switch

$$S = \{(a, z): a \in \{x, y\}, z \in B\}$$

$$|S| = 2 \cdot 100 = 200$$

If $y = \$60$ and $x = \$40$, then

$$\begin{aligned}\Pr(\text{find BB}) &= \frac{(100-x)+y}{200} \\ &= \frac{120}{200} = 0.6\end{aligned}$$

If $y = \$80$ and $x = \$20$, then

$$\begin{aligned}\Pr(\text{find BB}) &= \frac{(100-x)+y}{200} \\ &= \frac{160}{200} = 0.80\end{aligned}$$

If $y = \$100$ and $x = \$0$, then

$$\begin{aligned}\Pr(\text{find BB}) &= \frac{(100-x)+y}{200} \\ &= \frac{200}{200} = 1.00\end{aligned}$$