W06: Hashing

CS3230 AY21/22 Sem 2

Table of Contents

Click on the link to jump to the relevant sections!

Analysis using Indicator Random Variables

(Recap)

Indicator Random Variables

Indicator RV is like a form of “counter” for the occurrence of some event.

Indicator Random Variables

Indicator RV is like a form of “counter” for the occurrence of some event.
Let *X* be an indicator RV, where a certain event occurs with probability *p*:

$$X = \begin{cases} 1 & \text{w.p. } p \\ 0 & \text{w.p. } 1 - p \end{cases}$$

You may observe that it being 0 or 1 is the way we “count” whether some event occurs

Indicator Random Variables (Expectation)

Let X be an indicator random variable with probability p of the event happening.

$E[X]$

$= 1\,(p) + 0\,(1 - p)$ [Definition of expectation]

$= p$

$$X = \begin{cases} 1 & \text{w.p. } p \\ 0 & \text{w.p. } 1-p \end{cases}$$

Very useful! Simple (hopefully) to calculate

$$\boxed{E[X] = p}$$

The analysis “pattern” (1 & 2 are “interchangeable”)

Most (but not all) analysis in Randomised Algorithms follow this “pattern”.

1. Identify a Random Variable to “count” what you want (e.g. X. Goal: $E[X]$)

2. Express this RV as a **sum** of random variables (e.g. $X = X_1 + X_2 + \ldots + X_n$)
    a. Calculate the relevant probability for X_1, X_2, ...
    b. Calculate the individual expectation of the “sub”-random variables. ($E[X_1]$, $E[X_2]$, …)

3. Use linearity of expectations on $E[X]$. Then you add up the expectation of the “sub”-random variables (from step 2b)

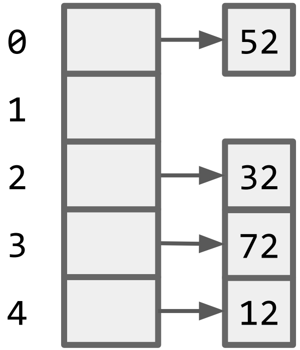
Universal Hashing

Universal Hashing (Motivation)

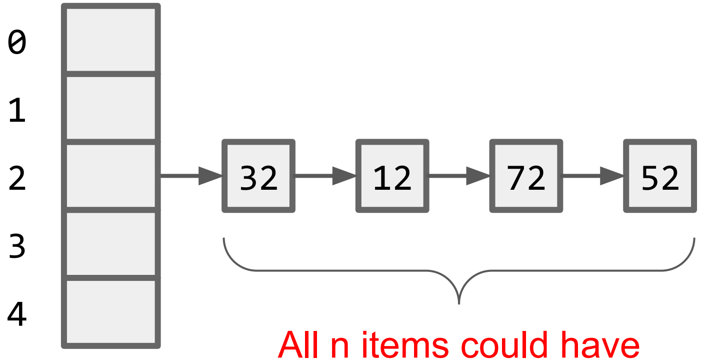
- Suppose your hash table uses a particular hash function *h*

Universal Hashing (Motivation)

- Suppose your hash table uses a particular hash function *h*
- If an adversary has access to your hash function, then it can force a worst-case scenario: give items that all map to the same bucket!



Ideal!



All n items could have hashed to the same bucket!

Universal Hashing (Motivation)

- Suppose your hash table uses a particular hash function *h*
- If an adversary has access to your hash function, then it can force a worst-case scenario: give items that all map to the same bucket!

- We want to “**fight against**” the adversary! Instead of deterministically choosing hash function:

Universal Hashing (Motivation)

- Suppose your hash table uses a particular hash function *h*
- If an adversary has access to your hash function, then it can force a worst-case scenario: give items that all map to the same bucket!

- We want to “**fight against”** the adversary! Instead of deterministically choosing hash function:
  - **Randomly** choose from a **set** of hash functions
  - Use that hash function instead -- adversary can’t possibly know for sure what it is!

Universal Hashing (Motivation)

- Suppose your hash table uses a particular hash function *h*
- If an adversary has access to your hash function, then it can force a worst-case scenario: give items that all map to the same bucket!

- We want to "**fight against"** the adversary! Instead of deterministically choosing hash function:
  - **Randomly** choose from a **set** of hash functions
  - Use that hash function instead -- adversary can't possibly know for sure what it is!
- We want to choose from a **"good set"** of hash functions -- Universal Hash Family is one way to define this "good set"!

Universal Hashing (Definition)

Definition: Suppose $\mathcal{H}$ is a set of hash functions mapping U to $[M]$. We say $\mathcal{H}$ is ***universal*** if for all $x \neq y$:

$$\frac{|h \in \mathcal{H} : h(x) = h(y)|}{|\mathcal{H}|} \leq \frac{1}{M}.$$

H = Our set of hash functions

U = All the items in the universe

M = How many "buckets" in the hash table

Universal Hashing (Definition)

Number of hash functions where x and y collide

Definition: Suppose $\mathcal{H}$ is a set of hash functions mapping U to $[M]$. We say $\mathcal{H}$ is ***universal*** if for all $x \neq y$:

$$\frac{\boxed{|h \in \mathcal{H} : h(x) = h(y)|}}{|\mathcal{H}|} \leq \frac{1}{M}.$$

H = Our set of hash functions

U = All the items in the universe

M = How many “buckets” in the hash table

Universal Hashing (Definition)

Number of hash functions where x and y collide

Definition: Suppose $\mathcal{H}$ is a set of hash functions mapping U to $[M]$. We say $\mathcal{H}$ is ***universal*** if for all $x \neq y$:

$$\frac{\boxed{|h \in \mathcal{H} : h(x) = h(y)|}}{\boxed{|\mathcal{H}|}} \leq \frac{1}{M}.$$

Total number of hash functions

H = Our set of hash functions

U = All the items in the universe

M = How many “buckets” in the hash table

Universal Hashing (Definition)

Number of hash functions where x and y collide

Definition: Suppose $\mathcal{H}$ is a set of hash functions mapping U to $[M]$. We say $\mathcal{H}$ is ***universal*** if for all $x \neq y$:

$$\frac{\boxed{|h \in \mathcal{H} : h(x) = h(y)|}}{\boxed{|\mathcal{H}|}} \leq \frac{1}{M}.$$

Total number of hash functions

H = Our set of hash functions

U = All the items in the universe

M = How many "buckets" in the hash table

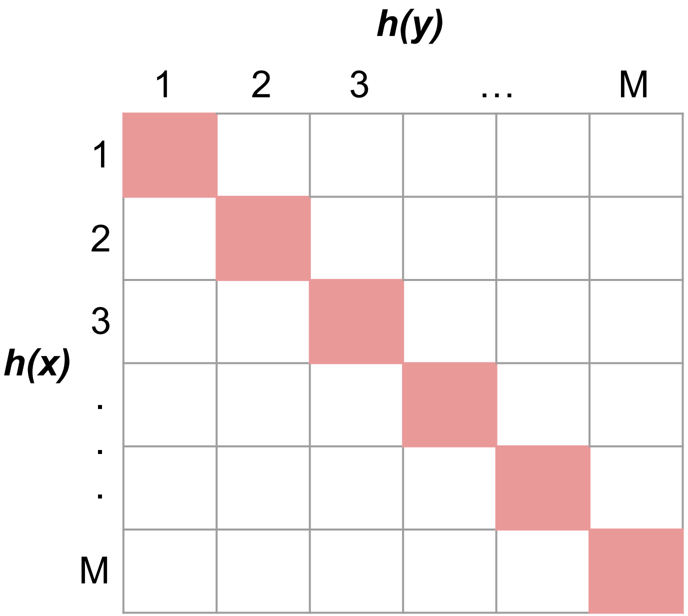
$$\Pr_{h \sim \mathcal{H}}[h(x) = h(y)] \leq \frac{1}{M}.$$

Alternative formulation

Universal Hashing (Illustration)

Universal: $\Pr_{h \sim \mathcal{H}}[h(x) = h(y)] \leq \frac{1}{M}$

Take **any** different *x*, *y* in *U*. Write the number of hash functions *h* in *H* that lie in each box.



If *H* is universal, sum of values in red cells must be at most ***|H|/M***

Example

<u>Definition</u>: Suppose $\mathcal{H}$ is a set of hash functions mapping U to $[M]$. We say $\mathcal{H}$ is ***universal*** if for all $x \neq y$:

$$\frac{|h \in \mathcal{H} : h(x) = h(y)|}{|\mathcal{H}|} \leq \frac{1}{M}.$$

Pairs: (a, b)

M = 2 (result is 0 or 1)

|H| = 3

	a	b
h_1	0	0
h_2	1	0
h_3	0	1

Example

Pairs: (a, b)

M = 2 (result is 0 or 1)

$|H|$ = 3

	a	b
h_1	0	0
h_2	1	0
h_3	0	1

Definition: Suppose $\mathcal{H}$ is a set of hash functions mapping U to $[M]$. We say $\mathcal{H}$ is ***universal*** if for all $x \neq y$:

$$\boxed{\frac{|h \in \mathcal{H} : h(x) = h(y)|}{|\mathcal{H}|}} \leq \frac{1}{M}.$$

LHS:

(a, b) collides for h_1 only

Therefore, since |H| = 3

LHS = ⅓

Example

Definition: Suppose $\mathcal{H}$ is a set of hash functions mapping U to $[M]$. We say $\mathcal{H}$ is ***universal*** if for all $x \neq y$:

$$\frac{|h \in \mathcal{H} : h(x) = h(y)|}{|\mathcal{H}|} \leq \boxed{\frac{1}{M}}$$

Pairs: (a, b)

M = 2 (result is 0 or 1)

|H| = 3

	a	b
h_1	0	0
h_2	1	0
h_3	0	1

LHS:

(a, b) collides for h_1 only

Therefore, since |H| = 3

LHS = ⅓

RHS:

1/M = ½

Example

Definition: Suppose $\mathcal{H}$ is a set of hash functions mapping U to $[M]$. We say $\mathcal{H}$ is ***universal*** if for all $x \neq y$:

$$\boxed{\frac{|h \in \mathcal{H} : h(x) = h(y)|}{|\mathcal{H}|}} \leq \boxed{\frac{1}{M}}$$

Pairs: (a, b)

$M = 2$ (result is 0 or 1)

$|H| = 3$

	a	b
h_1	0	0
h_2	1	0
h_3	0	1

LHS:

(a, b) collides for h_1 only

Therefore, since $|H| = 3$

LHS = ⅓

RHS:

1/M = ½

We have $\boxed{⅓} \leq \boxed{½}$, so this set is universal!

Example 2

Definition: Suppose $\mathcal{H}$ is a set of hash functions mapping U to $[M]$. We say $\mathcal{H}$ is ***universal*** if for all $x \neq y$:

$$\frac{|h \in \mathcal{H} : h(x) = h(y)|}{|\mathcal{H}|} \leq \frac{1}{M}.$$

Pairs: (a, b), (a, c), (b, c)

M = 2 (result is 0 or 1)

|H| = 3

RHS = ½

	a	b	c
h_1	0	0	1
h_2	1	1	0
h_3	1	0	1

Example 2

<u>Definition</u>: Suppose $\mathcal{H}$ is a set of hash functions mapping U to $[M]$. We say $\mathcal{H}$ is ***universal*** if for all $x \neq y$:

$$\frac{|h \in \mathcal{H} : h(x) = h(y)|}{|\mathcal{H}|} \leq \frac{1}{M}.$$

Pairs: (a, b), (a, c), (b, c)

M = 2 (result is 0 or 1)

$|H|$ = 3

RHS = ½

	a	b	c
h_1	0	0	1
h_2	1	1	0
h_3	1	0	1

(b, c)

↳ No collision

↳ LHS = 0/3

$0/3 \leq 1/2$

ok!

Example 2

Definition: Suppose $\mathcal{H}$ is a set of hash functions mapping U to $[M]$. We say $\mathcal{H}$ is ***universal*** if for all $x \neq y$:

$$\frac{|h \in \mathcal{H} : h(x) = h(y)|}{|\mathcal{H}|} \leq \frac{1}{M}.$$

Pairs: (a, b), (a, c), (b, c)

M = 2 (result is 0 or 1)

|H| = 3

RHS = ½

	a	b	c
h_1	0	0	1
h_2	1	1	0
h_3	1	0	1

(b, c)

↳ No collision

↳ LHS = $0/3$

$0/3 \leq 1/2$

ok!

(a, c)

↳ Collides for h_3

↳ LHS = $1/3$

$1/3 \leq 1/2$

ok!

Example 2

Definition: Suppose $\mathcal{H}$ is a set of hash functions mapping U to $[M]$. We say $\mathcal{H}$ is ***universal*** if for all $x \neq y$:

$$\frac{|h \in \mathcal{H} : h(x) = h(y)|}{|\mathcal{H}|} \leq \frac{1}{M}.$$

Pairs: (a, b), (a, c), (b, c)

M = 2 (result is 0 or 1)

$|H|$ = 3

RHS = ½

	a	b	c
h_1	0	0	1
h_2	1	1	0
h_3	1	0	1

(b, c)

↳ No collision

↳ LHS = $0/3$

$0/3 \leq 1/2$

ok!

(a, c)

↳ Collides for h_3

↳ LHS = $1/3$

$1/3 \leq 1/2$

ok!

(a, b):

↳ Collides for h_1, h_2

↳ LHS = $2/3$

$\frac{2}{3} > \frac{1}{2}$

Not ok!

Not all pairs satisfy the universality condition!

Therefore not universal

Pairwise Independent Family

Pairwise Independent

H = Our set of hash functions
U = All the items in the universe
M = How many "buckets" in the hash table

Definition: Suppose $\mathcal{H}$ is a set of hash functions mapping U to $[M]$. We say $\mathcal{H}$ is ***pairwise-independent*** if for all $x \neq y$ and any two hash values i_1, i_2:

$$\Pr_{h \sim \mathcal{H}}[h(x) = i_1, h(y) = i_2] = \frac{1}{M^2}.$$

Pairwise Independent

H = Our set of hash functions
U = All the items in the universe
M = How many "buckets" in the hash table

Definition: Suppose $\mathcal{H}$ is a set of hash functions mapping U to $[M]$. We say $\mathcal{H}$ is ***pairwise-independent*** if for all $x \neq y$ and any two hash values i_1, i_2:

$$\Pr_{h \sim \mathcal{H}}[h(x) = i_1, h(y) = i_2] = \frac{1}{M^2}.$$

Intuitively:

- Think of M^2 as all possible pairs of i_1 and i_2: e.g. { (0, 0), (0, 1), (1, 0), (1, 1) }

Pairwise Independent

H = Our set of hash functions
U = All the items in the universe
M = How many "buckets" in the hash table

Definition: Suppose $\mathcal{H}$ is a set of hash functions mapping U to $[M]$. We say $\mathcal{H}$ is ***pairwise-independent*** if for all $x \neq y$ and any two hash values i_1, i_2:

$$\Pr_{h \sim \mathcal{H}}[h(x) = i_1, h(y) = i_2] = \frac{1}{M^2}.$$

Relate it to "independence":
Pr(h(x) = i_1) = 1/m
Pr(h(y) = i_2) = 1/m
Pr both = (1/m)(1/m)

Intuitively:

- Think of M^2 as all possible pairs of i_1 and i_2: e.g. { (0, 0), (0, 1), (1, 0), (1, 1) }

Pairwise Independent

H = Our set of hash functions
U = All the items in the universe
M = How many "buckets" in the hash table

Definition: Suppose $\mathcal{H}$ is a set of hash functions mapping U to $[M]$. We say $\mathcal{H}$ is ***pairwise-independent*** if for all $x \neq y$ and any two hash values i_1, i_2:

$$\Pr_{h \sim \mathcal{H}}[h(x) = i_1, h(y) = i_2] = \frac{1}{M^2}.$$

Relate it to "independence":
$Pr(h(x) = i_1) = 1/m$
$Pr(h(y) = i_2) = 1/m$
Pr both = $(1/m)(1/m)$

Intuitively:

- Think of M^2 as all possible pairs of i_1 and i_2: e.g. { (0, 0), (0, 1), (1, 0), (1, 1) }
- For all distinct x and y, I can choose i_1 and i_2 to be anything I want
  - And the result "should feel uniformly distributed" (remember, M^2 is all possibilities)

Pairwise Independent (Illustration)

Pairwise-independent:

$$\Pr_{h \sim \mathcal{H}}[h(x) = i_1, h(y) = i_2] = \frac{1}{M^2}$$

Take **any** different *x*, *y* in *U*. Write the number of hash functions *h* in *H* that lie in each box.

h(y)

h(x)	1	2	3	…	M
1	$\frac{\lvert H \rvert}{M^2}$	$\frac{\lvert H \rvert}{M^2}$	$\frac{\lvert H \rvert}{M^2}$	…	$\frac{\lvert H \rvert}{M^2}$
2	$\frac{\lvert H \rvert}{M^2}$	$\frac{\lvert H \rvert}{M^2}$	…	…	…
3	$\frac{\lvert H \rvert}{M^2}$	…	…	…	…
⋮	…	…	…	…	$\frac{\lvert H \rvert}{M^2}$
M	$\frac{\lvert H \rvert}{M^2}$	…	…	$\frac{\lvert H \rvert}{M^2}$	$\frac{\lvert H \rvert}{M^2}$

If *H* is pairwise independent, the value in each cell must be the same, i.e. $|H|/M^2$

Pairwise Independent (Intuition)

Definition: Suppose $\mathcal{H}$ is a set of hash functions mapping U to $[M]$. We say $\mathcal{H}$ is ***pairwise-independent*** if for all $x \neq y$ and any two hash values i_1, i_2:

$$\Pr_{h \sim \mathcal{H}}[h(x) = i_1, h(y) = i_2] = \frac{1}{M^2}.$$

If you just have two elements x and y, then their resulting hash output “should feel” random. But **not necessarily true for 3** elements onwards!

Pairwise Independent (Intuition)

Definition: Suppose $\mathcal{H}$ is a set of hash functions mapping U to $[M]$. We say $\mathcal{H}$ is ***pairwise-independent*** if for all $x \neq y$ and any two hash values i_1, i_2:

$$\Pr_{h \sim \mathcal{H}}[h(x) = i_1, h(y) = i_2] = \frac{1}{M^2}.$$

If you just have two elements x and y, then their resulting hash output "should feel" random. But **not necessarily true for 3 elements onwards!**

Let's say we have *h(x) = ax + b* for randomly chosen *a* and *b*. (Note: sloppily defined, just to give an idea)

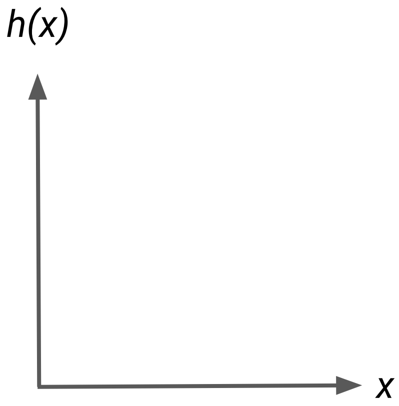
Pairwise Independent (Intuition)

Definition: Suppose $\mathcal{H}$ is a set of hash functions mapping U to $[M]$. We say $\mathcal{H}$ is ***pairwise-independent*** if for all $x \neq y$ and any two hash values i_1, i_2:

$$\Pr_{h \sim \mathcal{H}}[h(x) = i_1, h(y) = i_2] = \frac{1}{M^2}.$$

If you just have two elements x and y, then their resulting hash output "should feel" random. But **not necessarily true for 3 elements onwards!**

Let's say we have *h(x) = ax + b* for randomly chosen *a* and *b*. (Note: sloppily defined, just to give an idea)



For the first point x, it is "random" - you can hash anywhere!

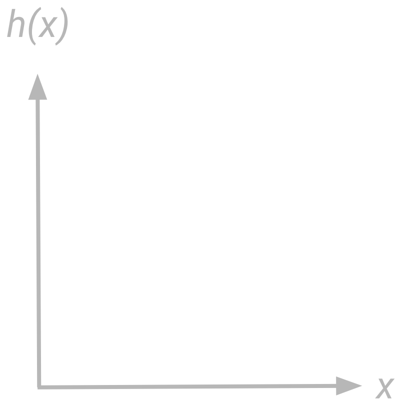
Pairwise Independent (Intuition)

Definition: Suppose $\mathcal{H}$ is a set of hash functions mapping U to $[M]$. We say $\mathcal{H}$ is ***pairwise-independent*** if for all $x \neq y$ and any two hash values i_1, i_2:

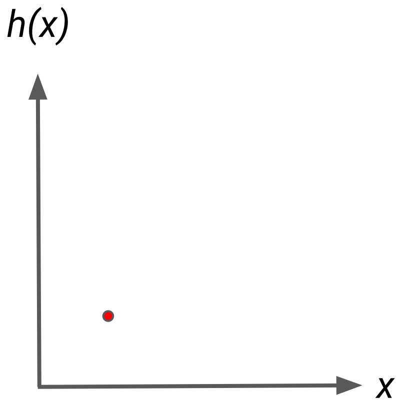
$$\Pr_{h \sim \mathcal{H}}[h(x) = i_1, h(y) = i_2] = \frac{1}{M^2}.$$

If you just have two elements x and y, then their resulting hash output “should feel” random. But **not necessarily true for 3 elements onwards!**

Let’s say we have *h(x) = ax + b* for randomly chosen *a* and *b*. (Note: sloppily defined, just to give an idea)



For the first point x, it is “random” - you can hash anywhere!



Even after the first point is placed, the second point is still “random”!

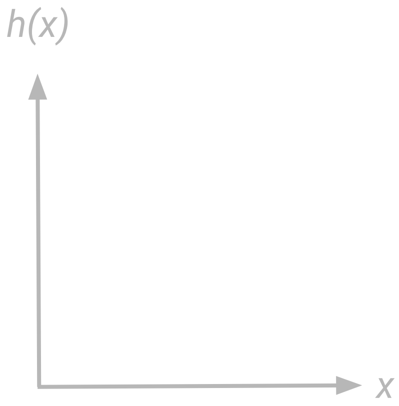
Pairwise Independent (Intuition)

Definition: Suppose $\mathcal{H}$ is a set of hash functions mapping U to $[M]$. We say $\mathcal{H}$ is ***pairwise-independent*** if for all $x \neq y$ and any two hash values i_1, i_2:

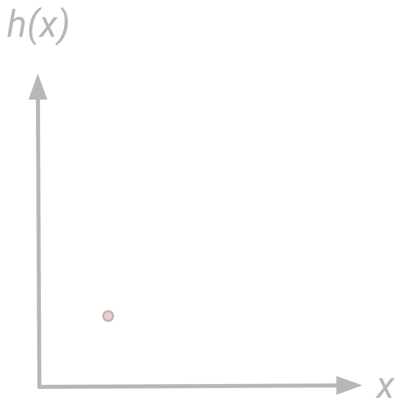
$$\Pr_{h \sim \mathcal{H}}[h(x) = i_1, h(y) = i_2] = \frac{1}{M^2}.$$

If you just have two elements x and y, then their resulting hash output “should feel” random. But **not necessarily true for 3 elements onwards!**

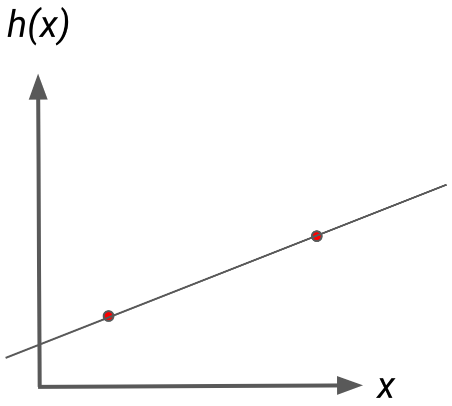
Let’s say we have *h(x) = ax + b* for randomly chosen *a* and *b*. (Note: sloppily defined, just to give an idea)



For the first point x, it is “random” - you can hash anywhere!



Even after the first point is placed, the second point is still “random”!



But after the second point, the third point can be determined! Not independent anymore

Pairwise Independent (Example)

Definition: Suppose $\mathcal{H}$ is a set of hash functions mapping U to $[M]$. We say $\mathcal{H}$ is ***pairwise-independent*** if for all $x \neq y$ and any two hash values i_1, i_2:

$$\Pr_{h \sim \mathcal{H}}[h(x) = i_1, h(y) = i_2] = \frac{1}{M^2}.$$

$|\mathcal{H}| = 4$

$M = 2$ (result is 0 or 1)

	a	b
h_1	1	0
h_2	0	0
h_3	1	1
h_4	0	1

Pairwise Independent (Example)

Definition: Suppose $\mathcal{H}$ is a set of hash functions mapping U to $[M]$. We say $\mathcal{H}$ is ***pairwise-independent*** if for all $x \neq y$ and any two hash values i_1, i_2:

$$\boxed{\Pr_{h \sim \mathcal{H}}[h(x) = i_1, h(y) = i_2]} = \boxed{\frac{1}{M^2}}.$$

$|\mathcal{H}| = 4$

$M = 2$ (result is 0 or 1)

	a	b
h_1	1	0
h_2	0	0
h_3	1	1
h_4	0	1

For all i_1, i_2

$$\boxed{\text{LHS: } \Pr[h(a) = i_1, h(b) = i_2] = \frac{1}{|\mathcal{H}|} = \frac{1}{4}}$$

$$\boxed{\text{RHS: } \frac{1}{M^2} = \frac{1}{2^2} = \frac{1}{4}}$$

$\therefore \text{LHS} = \text{RHS} \quad \forall x, y, i_1, i_2$

Note: Definition with Equality

Pairwise-independent:

$$\Pr_{h\sim\mathcal{H}}[h(x)=i_1,h(y)=i_2] \boxed{=} \frac{1}{M^2}$$

$\Longleftrightarrow$

Pairwise-independent:

$$\Pr_{h\sim\mathcal{H}}[h(x)=i_1,h(y)=i_2] \boxed{\leq} \frac{1}{M^2}$$

Tutorial question, earlier version of slides + recording uses ≤ to define pairwise-independent, whereas this version defines it with equality. The two definitions are **equivalent**. However, equality seems to be more common.

Proof Sketch: Consider all M^2 combinations of i_1 and i_2.

1 = Sum of combinations of Probability of hashing to i_1 and i_2 respectively ≤ $M^2 * (1/M^2) = 1$

First equality is due to probability axioms after considering all cases.
For each probability of hashing to i_1 and i_2, it must be exactly $1/M^2$

Question 1: Pairwise Independent → Universal?

Q1

Does Pairwise-Independent family imply Universal family?

Definition: Suppose $\mathcal{H}$ is a set of hash functions mapping U to $[M]$. We say $\mathcal{H}$ is ***pairwise-independent*** if for all $x \neq y$ and any two hash values i_1, i_2:

$$\Pr_{h \sim \mathcal{H}}[h(x) = i_1, h(y) = i_2] = \frac{1}{M^2}$$

?

Definition: Suppose $\mathcal{H}$ is a set of hash functions mapping U to $[M]$. We say $\mathcal{H}$ is ***universal*** if for all $x \neq y$:

$$\Pr_{h \sim \mathcal{H}}[h(x) = h(y)] \leq \frac{1}{M}$$

Question 1 (Answer)

Yes! Pairwise Independent Family is always Universal Hash family

Question 1 (Answer)

Yes! Pairwise Independent Family is always Universal Hash family

Proof strategy:

1. Use definition of Universal Hash family
2. Express it similar to Pairwise Independent family
3. Since we assume Pairwise Independent, use its property

Question 1 (Answer)

Pairwise-independent:

$$\Pr_{h\sim\mathcal{H}}[h(x)=i_1, h(y)=i_2] = \frac{1}{M^2}$$

Universal: $\Pr_{h\sim\mathcal{H}}[h(x)=h(y)] \leq \frac{1}{M}$

$$\Pr[h(x)=h(y)]$$

Question 2

One side of Universal Family

Question 1 (Answer)

Pairwise-independent:

$$\Pr_{h \sim \mathcal{H}}[h(x) = i_1, h(y) = i_2] = \frac{1}{M^2}$$

Universal: $\Pr_{h \sim \mathcal{H}}[h(x) = h(y)] \leq \frac{1}{M}$

Question 2

Make it look like pairwise independent

$$\Pr[h(x) = h(y)]$$

$$= \sum_i \Pr[h(x) = i, h(y) = i]$$

e.g. M = 2, {0, 1}

Then

Pr[h(x) = h(y)]

= Pr[h(x) = 0, h(y) = 0]

+ Pr[h(x) = 1, h(y) = 1]

Question 1 (Answer)

Pairwise-independent:

$$\Pr_{h\sim\mathcal{H}}[h(x)=i_1, h(y)=i_2] = \boxed{\frac{1}{M^2}}$$

Universal: $\Pr_{h\sim\mathcal{H}}[h(x)=h(y)] \le \frac{1}{M}$

Question 2

Use property of pairwise independent!

$$\Pr[h(x)=h(y)]$$

$$=\sum_i \Pr[h(x)=i, h(y)=i]$$

$$=\boxed{M}\cdot\boxed{\frac{1}{M^2}}$$

There are M possible hash values

Question 1 (Answer)

Pairwise-independent:

$$\Pr_{h \sim \mathcal{H}}[h(x) = i_1, h(y) = i_2] = \frac{1}{M^2}$$

Universal: $\Pr_{h \sim \mathcal{H}}[h(x) = h(y)] \leq \frac{1}{M}$

Question 2

Proven!

$$\boxed{\Pr[h(x) = h(y)]}$$

$$= \sum_i \Pr[h(x) = i, h(y) = i]$$

$$= M \cdot \frac{1}{M^2} = \boxed{\frac{1}{M}}$$

Question 2: Universal $\to$ Pairwise Independent?

Q2

Does Universal family imply Pairwise-independent family?

Definition: Suppose $\mathcal{H}$ is a set of hash functions mapping U to $[M]$. We say $\mathcal{H}$ is ***pairwise-independent*** if for all $x \neq y$ and any two hash values i_1, i_2:

$$\Pr_{h \sim \mathcal{H}}[h(x) = i_1, h(y) = i_2] = \frac{1}{M^2}$$

$$\Uparrow ?$$

Definition: Suppose $\mathcal{H}$ is a set of hash functions mapping U to $[M]$. We say $\mathcal{H}$ is ***universal*** if for all $x \neq y$:

$$\Pr_{h \sim \mathcal{H}}[h(x) = h(y)] \leq \frac{1}{M}$$

Question 2 (Answer)

Pairwise-independent:
$$\Pr_{h\sim\mathcal{H}}[h(x)=i_1, h(y)=i_2] = \frac{1}{M^2}$$

Universal: $\Pr_{h\sim\mathcal{H}}[h(x)=h(y)] \leq \frac{1}{M}$

No! Come up with counterexample:

	a	b
h_1	0	0
h_2	0	1

Universal

LHS: collision for h_1

: $\frac{1}{2}$

RHS: $\frac{1}{M}$: $\frac{1}{2}$

$\therefore$ LHS $\leq$ RHS

Question 2 (Answer)

Pairwise-independent:

$$\Pr_{h \sim \mathcal{H}}[h(x) = i_1, h(y) = i_2] = \frac{1}{M^2}$$

Universal: $\Pr_{h \sim \mathcal{H}}[h(x) = h(y)] \leq \frac{1}{M}$

No! Come up with counterexample:

	a	b
h_1	0	0
h_2	0	1

Universal

LHS: collision for h_1

$: \frac{1}{2}$

RHS: $\frac{1}{M} : \frac{1}{2}$

$\therefore$ LHS $\leq$ RHS

Not Pair-wise Independent

LHS: $\Pr[h(a) = 0, h(b) = 0] = \frac{1}{2}$

RHS: $\frac{1}{M^2} = \frac{1}{2^2} = \frac{1}{4}$

$\therefore$ LHS $\neq$ RHS

$\Rightarrow$ Not pairwise-independent

Pairwise Independent and Universal

- This means that pairwise independent family is a **stronger** notion of hash family!

Pairwise Independent and Universal

- This means that pairwise independent family is a **stronger** notion of hash family!
- Intuitively:
  - In Pairwise Independent Family, you can ***freely*** choose any pair of hash values i_1 *and* i_2
  - In Universal Family, you are ***limited*** to hashes of x and y that collide (equal i_1 *and* i_2)

Pairwise-independent:

$$\Pr_{h \sim \mathcal{H}}[h(x) = i_1, h(y) = i_2] = \frac{1}{M^2}$$

Universal: $\Pr_{h \sim \mathcal{H}}[h(x) = h(y)] \leq \frac{1}{M}$

Question 3:
Pairwise Independent -- in a particular slot

Question 3

- Now you have a hash function from pairwise independent family
- Hash **N** distinct elements
  - At most, what is the expected number of elements which hashes to a particular **slot *j***?

Definition: Suppose $\mathcal{H}$ is a set of hash functions mapping U to $[M]$. We say $\mathcal{H}$ is ***pairwise-independent*** if for all $x \neq y$ and any two hash values i_1, i_2:

$$\Pr_{h \sim \mathcal{H}}[h(x) = i_1, h(y) = i_2] = \frac{1}{M^2}.$$

Question 3

A **pairwise independent** family $\mathcal{H}$ of hash functions mapping $\mathcal{U}$ to $\{1, \dots, M\}$ has the property that for any two distinct universe elements x, y and for any two hash values i_1, i_2:

$$\Pr_{h \sim \mathcal{H}}[h(x) = i_1, h(y) = i_2] \leq \frac{1}{M^2}.$$

Suppose you hash N distinct elements using h randomly drawn from a pairwise independent family. The expected number of elements which hash to slot 1 is at most?

(A) N/M^2
(B) N/M
(C) $N/2M$
(D) $N/4M$

Question 3 (Answer)

Strategy: Indicator Random Variables

1. Identify a Random Variable to “count” what you want (e.g. X. Goal: $E[X]$)
2. Express this RV as a **sum** of random variables (e.g. $X = X_1 + X_2 + \ldots + X_n$)
   a. Calculate the relevant probability for X_1, X_2, ...
   b. Calculate the individual expectation of the “sub”-random variables. ($E[X_1]$, $E[X_2]$, …)
3. Use linearity of expectations on $E[X]$. Then you add up the expectation of the “sub”-random variables (from step 2b)

Question 3 (Answer)

Goal: Expected number items hashing to slot j

Let X be the random variable representing the number of items hashing to slot j

Let X_i be the **indicator random variable** that item i (x_i) hashes to slot j

$X = X_1 + X_2 + \ldots X_N$

Question 3 (Answer)

1. Identify a Random Variable to "count" what you want (e.g. X. Goal: $E[X]$)
2. Express this RV as a **sum** of random variables (e.g. $X = X_1 + X_2 + \ldots + X_n$)
    a. Calculate the relevant probability for $X_1, X_2, \ldots$
    b. Calculate the individual expectation of the "sub"-random variables. ($E[X_1], E[X_2], \ldots$)
3. Use linearity of expectations on $E[X]$. Then you add up the expectation of the "sub"-random variables (from step 2b)

Goal: Expected number items hashing to slot j

Let X be the random variable representing the number of items hashing to slot j

Let X_i be the **indicator random variable** that item i (x_i) hashes to slot j

$X = X_1 + X_2 + \ldots X_N$

Next things to do:

- Calculate $Pr(X_i = 1)$. This is enough to get $E[X_i]$!
- Then we can calculate $E[X]$ easily by linearity of expectations

Question 3 (Computing $Pr(X_i = 1)$)

Pairwise-independent:

$$\Pr_{h \sim \mathcal{H}}[h(x) = i_1, h(y) = i_2] = \frac{1}{M^2}$$

Let X_i be the **indicator random variable** that item i (x_i) hashes to slot j

$$\Pr(X_i = 1)$$

Question 3

Getting ready!

Question 3 (Computing $Pr(X_i = 1)$)

Pairwise-independent:

$$\Pr_{h\sim\mathcal{H}}[h(x) = i_1, h(y) = i_2] = \frac{1}{M^2}$$

Let X_i be the **indicator random variable** that item i (x_i) hashes to slot j

$$\Pr(X_i = 1) = \Pr(h(x_i) = j)$$

Question 3

This is how we defined the indicator!

Question 3 (Computing $Pr(X_i = 1)$)

Pairwise-independent:

$$\Pr_{h\sim\mathcal{H}}[h(x) = i_1, h(y) = i_2] = \frac{1}{M^2}$$

Let X_i be the **indicator random variable** that item i (x_i) hashes to slot j

$$\Pr(X_i = 1) = \Pr(\boxed{h(x_i) = j})$$

$$= \boxed{\sum_{k=0}^{M-1}} \Pr(\boxed{h(x_i) = j}, h(\boxed{y}) = \boxed{k})$$

Question 3

"Make it look like pairwise indep":
k = goes through all M items.
y = item distinct from x_i

Question 3 (Computing $Pr(X_i = 1)$)

Pairwise-independent:

$$\boxed{\Pr_{h\sim\mathcal{H}}[h(x) = i_1, h(y) = i_2] = \frac{1}{M^2}}$$

Let X_i be the **indicator random variable** that item i (x_i) hashes to slot j

$$\Pr(X_i = 1) = \Pr(h(x_i) = j)$$

$$= \sum_{k=0}^{M-1} \boxed{\Pr(h(x_i) = j, h(y) = k)}$$

$$= \sum_{k=0}^{M-1} \boxed{\frac{1}{M^2}}$$

Question 3

Pairwise Independence!

Question 3 (Computing $Pr(X_i = 1)$)

Pairwise-independent:

$$\Pr_{h \sim \mathcal{H}}[h(x) = i_1, h(y) = i_2] = \frac{1}{M^2}$$

Let X_i be the **indicator random variable** that item i (x_i) hashes to slot j

$$\begin{aligned}
\Pr(X_i = 1) &= \Pr(h(x_i) = j) \\
&= \sum_{k=0}^{M-1} \Pr(h(x_i) = j, h(y) = k) \\
&= \boxed{\sum_{k=0}^{M-1}} \frac{1}{M^2} \\
&= \frac{1}{M}
\end{aligned}$$

Question 3

M hash values → M * (1/M^2)

Question 3: Apply linearity!

Goal: Expected number items hashing to slot *j*

Let *X* be the random variable representing the number of items hashing to slot *j*
Let X_i be the **indicator random variable** that item *i* (x_i) hashes to slot *j*

$X = X_1 + X_2 + \dots X_N$

$$E[X] = E[\sum_{i=1}^{\boxed{N}} X_i]$$

Recall: hashing N distinct elements!

$$= \sum_{i=1}^{N} \boxed{E[X_i]}$$

From earlier slide

$$= \sum_{i=1}^{N} \boxed{\frac{1}{M}}$$

$$= \frac{N}{M}$$

Question 4: Same bound for Universal Family?

Question 4

- Now you have a hash function from **universal family instead**
- Hash **N** distinct elements
  - Is the expected number of elements which hashes to a particular **slot *j*** still the same as before? i.e. Still ≤ N/M?

Definition: Suppose $\mathcal{H}$ is a set of hash functions mapping U to $[M]$. We say $\mathcal{H}$ is ***universal*** if for all $x \neq y$:

$$\Pr_{h \sim \mathcal{H}}[h(x) = h(y)] \leq \frac{1}{M}$$

Question 4

The same upper bound as the previous question holds for a hash function drawn from a universal (instead of pairwise independent) family.

True or False?

False

Universal, from Q2 $\left\{ \begin{array}{c|c|c} & a & b \\ \hline h_1 & 0 & 0 \\ \hline h_2 & 0 & 1 \\ \hline & & \end{array} \right.$

$N=2$ (a and b)

$M=2$ (0 and 1)

$\therefore \frac{N}{M}=1$

False

Universal, from Q2 $\left\{ \begin{array}{c|c|c} & a & b \\ \hline h_1 & 0 & 0 \\ \hline h_2 & 0 & 1 \\ \hline \end{array} \right.$

$N = 2$ (a and b)

$M = 2$ (0 and 1)

$\therefore \frac{N}{M} = 1$

Define IRV X_i:

$$X_i \begin{cases} 1 & \text{if } h(i) = 0 \\ 0 & \text{otherwise} \end{cases}$$

Count it if it hashes to slot 0

$Pr[X_a = 1] = 1$ (both hash gives 0)

$Pr[X_b = 1] = \frac{1}{2}$ (only h_1

False

Universal, from Q2

	a	b
h_1	0	0
h_2	0	1

$N=2$ (a and b)

$M=2$ (0 and 1)

$\therefore \frac{N}{M}=1$

Define IRV X_i:

$$X_i \begin{cases} 1 & \text{if } h(i)=0 \\ 0 & \text{otherwise} \end{cases}$$

Count it if it hashes to slot 0

$Pr[X_a=1]=1$ (both hash gives 0)

$Pr[X_b=1]=\frac{1}{2}$ (only h_1

X: total hashing to slot 0

$$\mathbb{E}[X] = \mathbb{E}[X_a] + \mathbb{E}[X_b]$$

$$= 1 + \frac{1}{2}$$

$$= 1.5$$

False

Universal, from Q2

	a	b
h_1	0	0
h_2	0	1

$N=2$ (a and b)

$M=2$ (0 and 1)

$\therefore \frac{N}{M}=1$

Define IRV X_i:

$$X_i \begin{cases} 1 & \text{if } h(i)=0 \\ 0 & \text{otherwise} \end{cases}$$

$Pr[X_a=1]=1$ (both hash gives 0)

$Pr[X_b=1]=\frac{1}{2}$ (only h_1

X: total hashing to slot 0

$$\boxed{\begin{aligned} \mathbb{E}[X] &= \mathbb{E}[X_a]+\mathbb{E}[X_b] \\ &= 1 + \frac{1}{2} \\ &= 1.5 \end{aligned}}$$

BUT: $1.5 > 1$

$\mathbb{E}[X] > \frac{N}{M}$

Count it if it hashes to slot 0

- Now you have a hash function from **universal family instead**
- Is the expected number of elements which hashes to a particular **slot *j*** still the same as before? i.e. **Still ≤ N/M**?

The following slides will show another example based on the hash family in lecture. If there is no time, it can be skipped

Lecture Example of Universal Hashing

Suppose U is indexed by u-bit strings, and $M = 2^m$.
For any binary matrix A with m rows and u columns:

$$h_A(x) = Ax \pmod 2$$

Claim: $\{h_A : A \in \{0,1\}^{m \times u}\}$ is universal.

Lecture Example of Universal Hashing

Suppose $\boxed{U}$ is indexed by $\boxed{u}$-bit strings, and $\boxed{M} = 2^{\boxed{m}}$.
For any binary matrix A with m rows and u columns:

$$h_A(x) = Ax \pmod 2$$

Claim: $\{h_A : A \in \{0,1\}^{m \times u}\}$ is universal.

What is U? What is u?
What is M? What is m?

Lecture Example

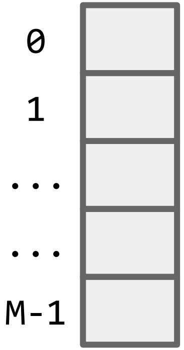
Suppose U is indexed by u-bit strings, and $M = 2^m$.
For any binary matrix A with m rows and u columns:

$$h_A(x) = Ax \pmod 2$$

<u>Claim</u>: $\{h_A : A \in \{0,1\}^{m \times u}\}$ is universal.

U = number of elements in the universe

M = the buckets it is mapping to



Hash table

Lecture Example

Suppose U is indexed by u-bit strings, and $M = 2^m$.
For any binary matrix A with m rows and u columns:

$$h_A(x) = Ax \pmod 2$$

<u>Claim</u>: $\{h_A : A \in \{0,1\}^{m \times u}\}$ is universal.

U = number of elements in the universe

e.g. {0, 1, 2, … 7}

Universe *U*:
{0, 1, 2, 3, 4, 5, 6, 7}

Suppose U is indexed by u-bit strings, and $M = 2^m$.
For any binary matrix A with m rows and u columns:
$$h_A(x) = Ax \pmod 2$$

Lecture Example

Claim: $\{h_A : A \in \{0,1\}^{m \times u}\}$ is universal.

U = number of elements in the universe

e.g. {0, 1, 2, … 7}. The little *u* is the len of binary representation to “index” these *U*

Universe *U*:
{0, 1, 2, 3, 4, 5, 6, 7}

Index	*element in U*
000	*0*
001	*1*
010	*2*
011	*3*
100	*4*
101	*5*
110	*6*
111	*7*

Lecture Example

Suppose U is indexed by u-bit strings, and $M = 2^m$.
For any binary matrix A with m rows and u columns:

$$h_A(x) = Ax \pmod 2$$

Claim: $\{h_A : A \in \{0,1\}^{m \times u}\}$ is universal.

U = number of elements in the universe

e.g. {0, 1, 2, … 7}. The little *u* is the len of binary representation to "index" these *U*

u = θ(log U)

Universe *U*:
{0, 1, 2, 3, 4, 5, 6, 7}

Index	*element in U*
000	*0*
001	*1*
010	*2*
011	*3*
100	*4*
101	*5*
110	*6*
111	*7*

Taken from CS3230 Prerequisites Revision slides

Decimals to Binary

$O(log(x))$ bits to represent!

An integer $x >= 1$, needs $n = \lfloor log_2(x) \rfloor + 1$ bits to represent it

x	Binary Repr.	n
1	1	1
2	10	2
3	11	2
4	100	3
5	101	3
10	1010	4
23	10111	5
63	111111	6
64	1000000	7

Lecture Example

Suppose U is indexed by u-bit strings, and $M = 2^m$.
For any binary matrix A with m rows and u columns:

$$h_A(x) = Ax \pmod 2$$

Claim: $\{h_A : A \in \{0,1\}^{m \times u}\}$ is universal.

M = the number of buckets in the hash table

m = the len of binary representation of M

m=2

idx	M
00	0
01	1
10	2
11	3

Q4: Another Example

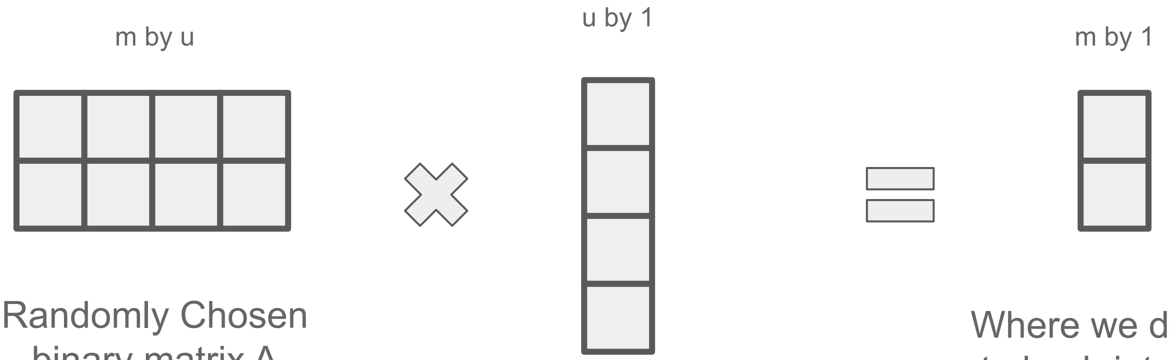
U = size of universe, u = len of binary representation of idx to U
M = number of buckets, m = len of binary representation of idx to M

In lecture, there is an example where we took an element in the universe x, and then we use the hash function $h(x) = Ax \pmod 2$, where A is a m *by* u binary matrix. Over the randomly chosen A, this is a Universal Family

Q4: Another Example

U = size of universe, u = len of binary representation of idx to U
M = number of buckets, m = len of binary representation of idx to M

In lecture, there is an example where we took an element in the universe x, and then we use the hash function $h(x) = Ax \pmod 2$, where A is a m *by* u binary matrix. Over the randomly chosen A, this is a Universal Family



Randomly Chosen binary matrix A

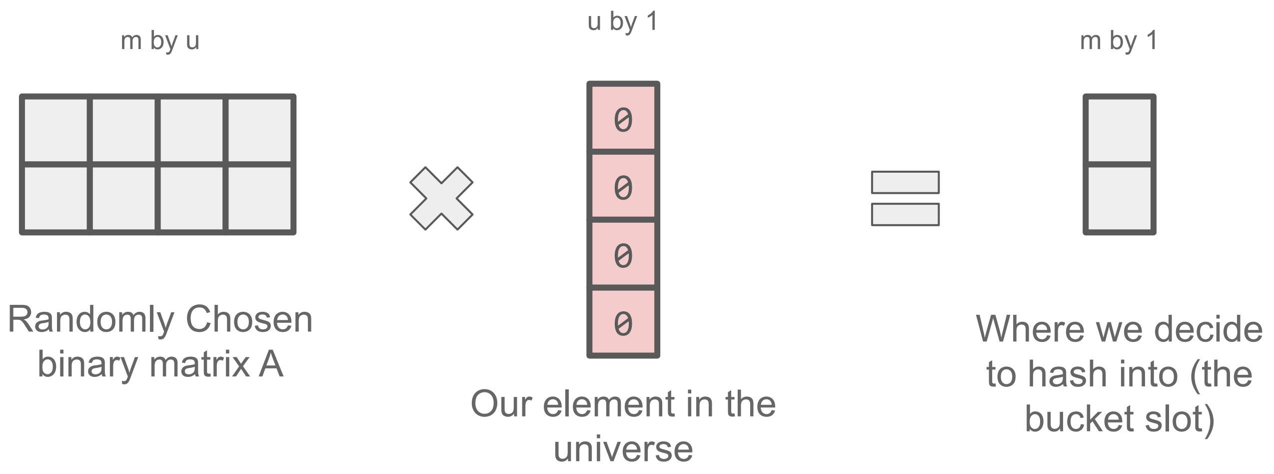
Our element in the universe

Where we decide to hash into (the bucket slot)

Q4: Another Example

U = size of universe, *u* = len of binary representation of idx to U
M = number of buckets, *m* = len of binary representation of idx to M

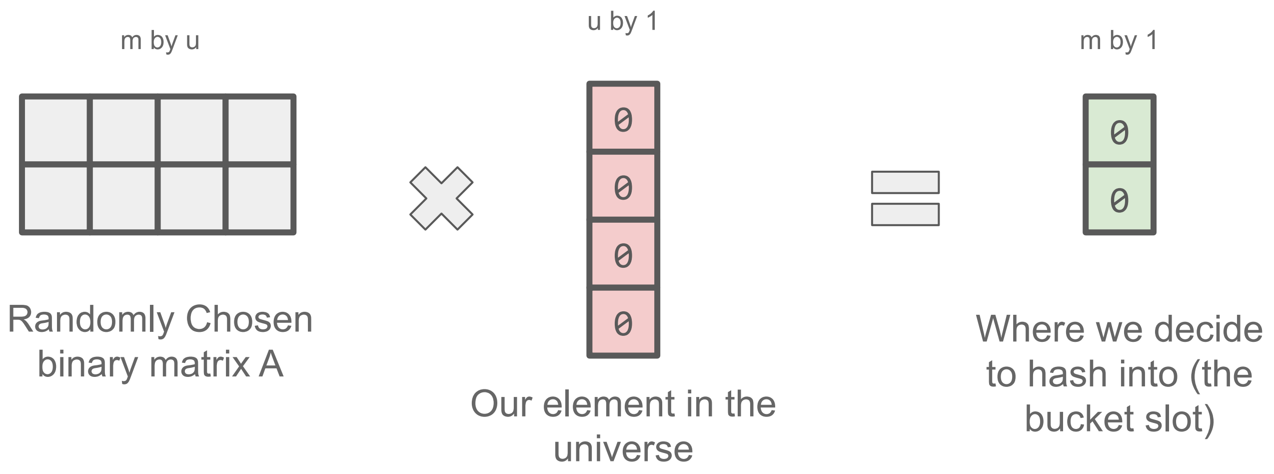
Let's say x = 0



Q4: Another Example

U = size of universe, *u* = len of binary representation of idx to U
M = number of buckets, *m* = len of binary representation of idx to M

Let's say x = 0. Then the resultant h(x) must be 0 as well, regardless of what A is!

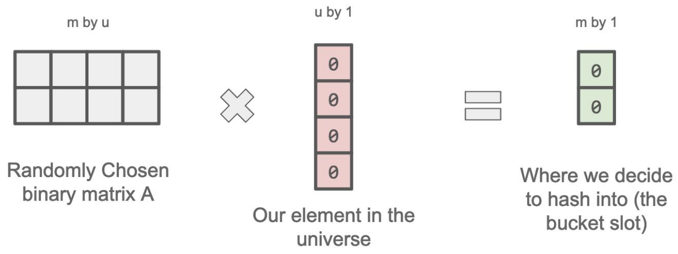


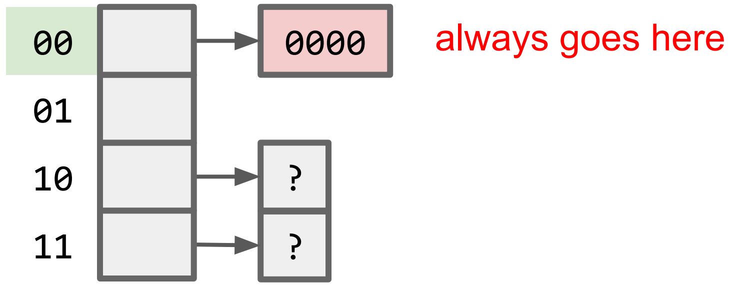
Q4: Another Example

U = size of universe, u = len of binary representation of idx to U
M = number of buckets, m = len of binary representation of idx to M

Let's say x = 0. Then the resultant h(x) must be 0 as well, regardless of what A is!

This means *E[mapping to slot 0]* $\geq$ *1* as long as there is an $x = 0$.





Our hash table

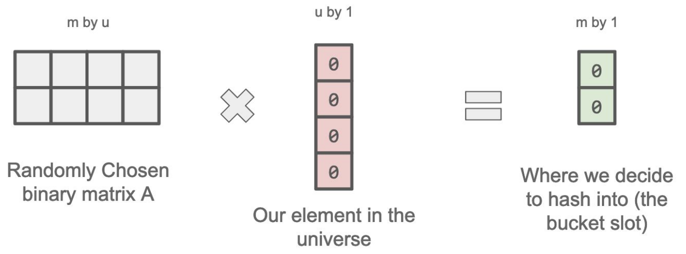
Q4: Another Example

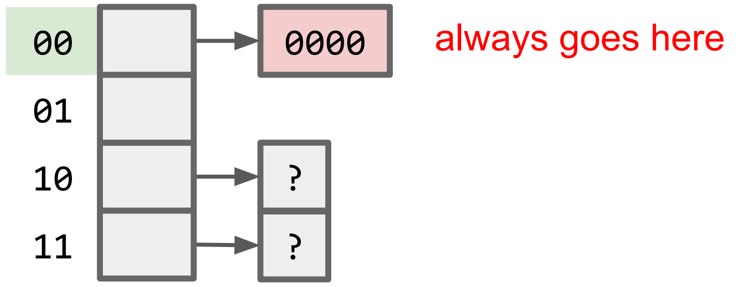
U = size of universe, *u* = len of binary representation of idx to U
M = number of buckets, *m* = len of binary representation of idx to M

Let's say x = 0. Then the resultant h(x) must be 0 as well, regardless of what A is!

This means *E[mapping to slot 0]* ≥ *1* as long as there is an *x = 0*.
Our goal: Is the expectation ≤ N/M?

Ans to q3





Our hash table

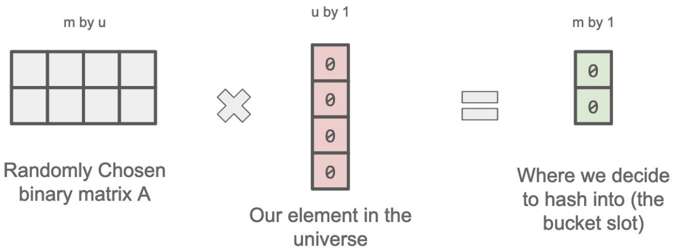
Q4: Another Example

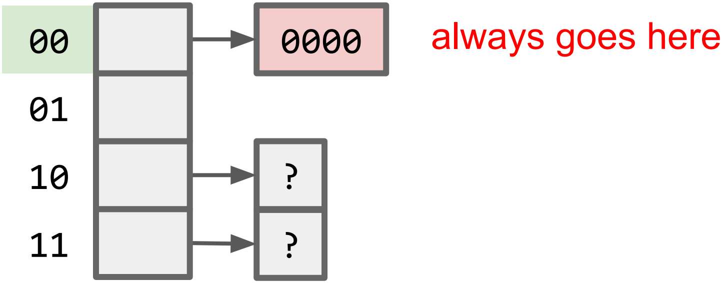
U = size of universe, *u* = len of binary representation of idx to U
M = number of buckets, *m* = len of binary representation of idx to M

Let's say x = 0. Then the resultant h(x) must be 0 as well, regardless of what A is!

This means *E[mapping to slot 0]* ≥ *1* as long as there is an *x = 0*.
Our goal: Is the expectation ≤ N/M? But if M > N, then N/M < 1

Ans to q3





Our hash table

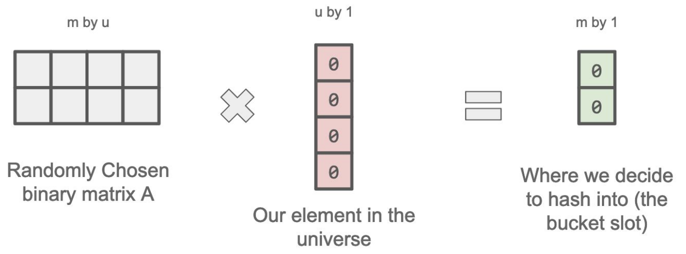
Q4: Another Example

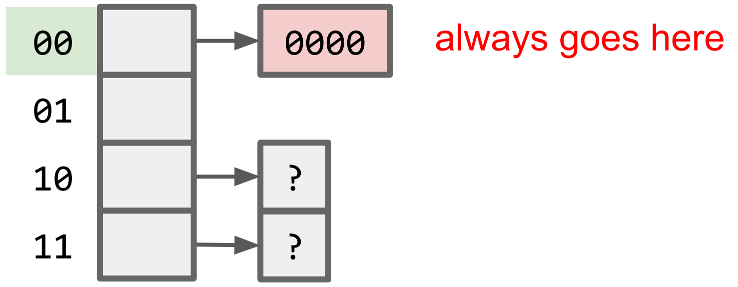
U = size of universe, *u* = len of binary representation of idx to U
M = number of buckets, *m* = len of binary representation of idx to M

Let's say x = 0. Then the resultant h(x) must be 0 as well, regardless of what A is!

This means *E[mapping to slot 0] ≥ 1* as long as there is an *x = 0*.
Our goal: Is the expectation ≤ N/M? But if M > N, then N/M < 1. Cannot satisfy!

Ans to q3





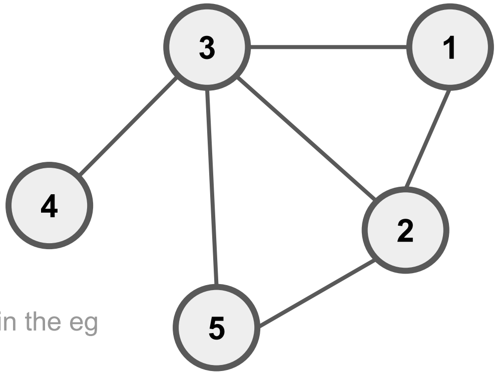
Our hash table

Question 5: Edges across the cut

Question 5

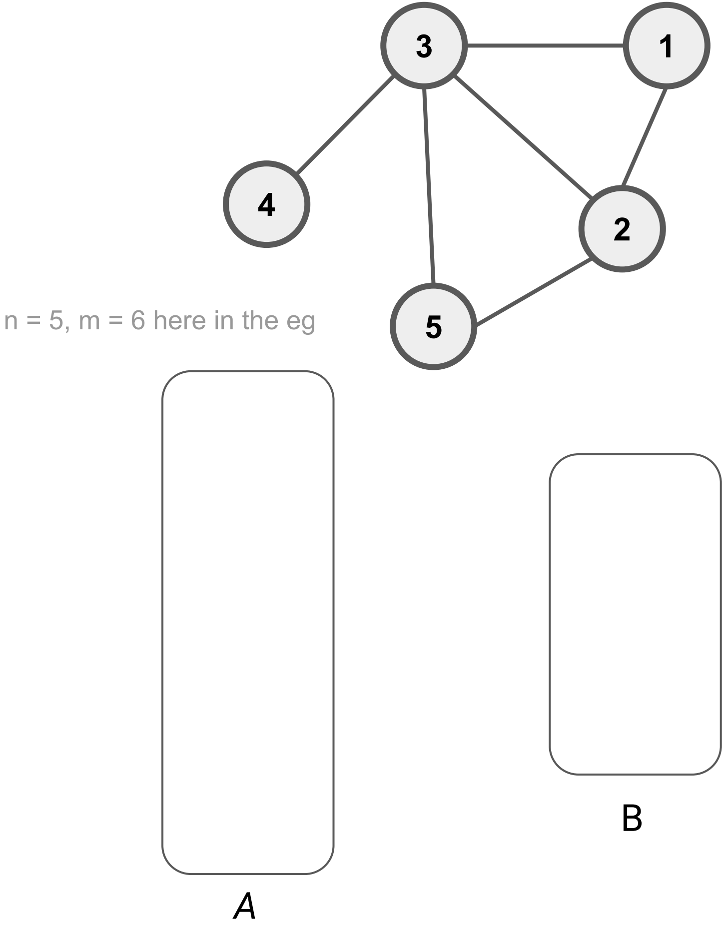
Let *G* be an undirected graph with *n* nodes and *m* edges.

n = 5, m = 6 here in the eg



Question 5

Let G be an undirected graph with n nodes and m edges. Partition the graph into two parts A and B randomly as follows:



Question 5

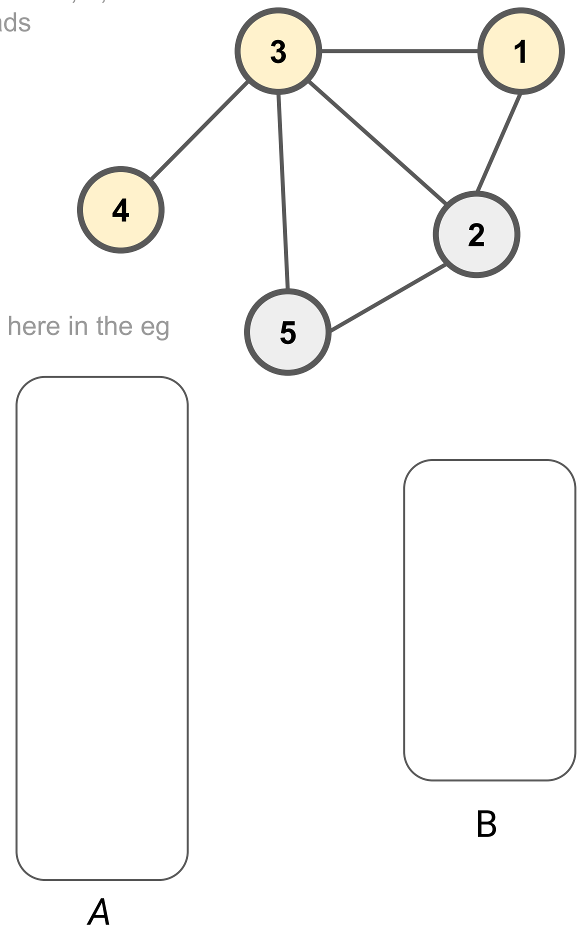
Let G be an undirected graph with n nodes and m edges. Partition the graph into two parts A and B randomly as follows:

For each node v, toss an independent fair coin:

- Heads: Put v in A

Let's say, nodes 1, 3, and 4 gave us heads

n = 5, m = 6 here in the eg

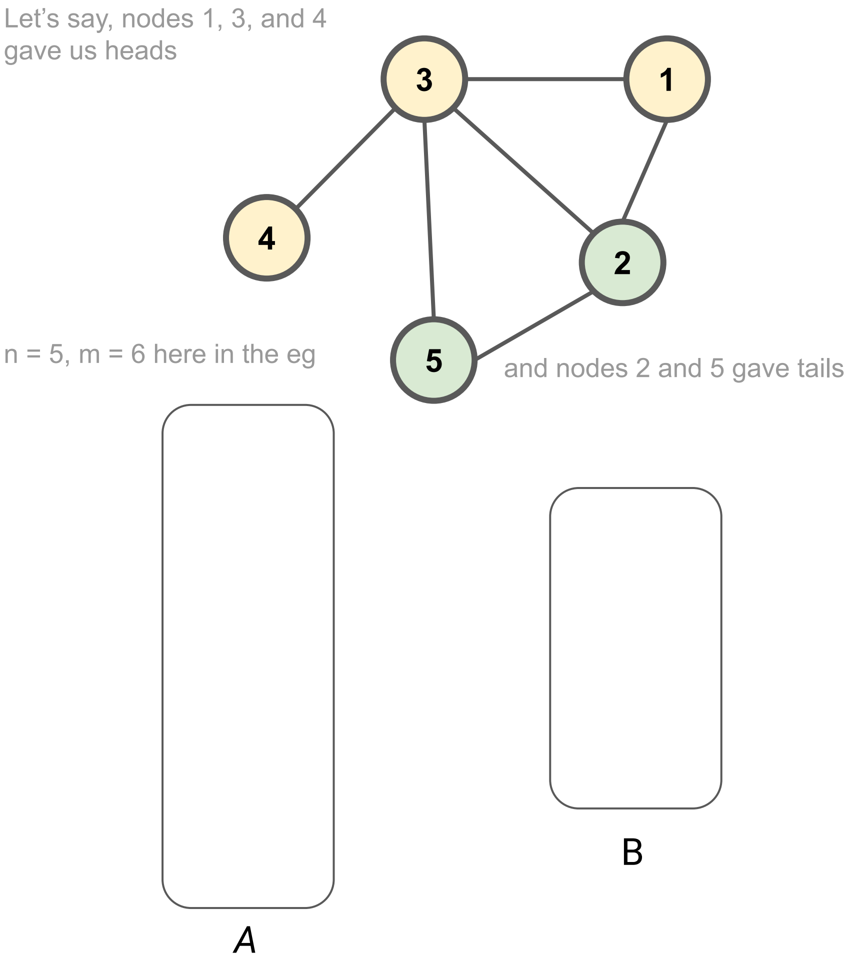


Question 5

Let G be an undirected graph with n nodes and m edges. Partition the graph into two parts A and B randomly as follows:

For each node v, toss an independent fair coin:

- Heads: Put v in A
- Tails: Put v in B

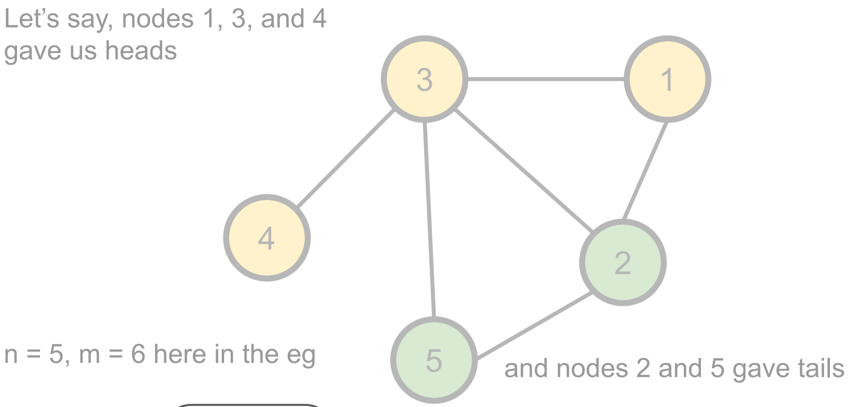


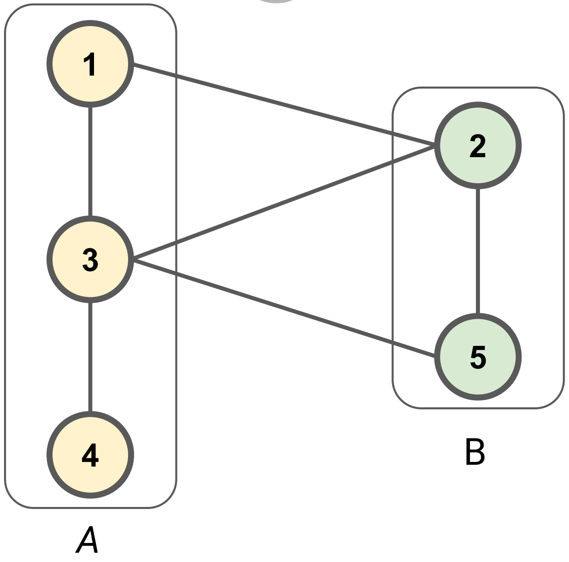
Question 5

Let G be an undirected graph with n nodes and m edges. Partition the graph into two parts A and B randomly as follows:

For each node v, toss an independent fair coin:

- Heads: Put v in A
- Tails: Put v in B





Question 5

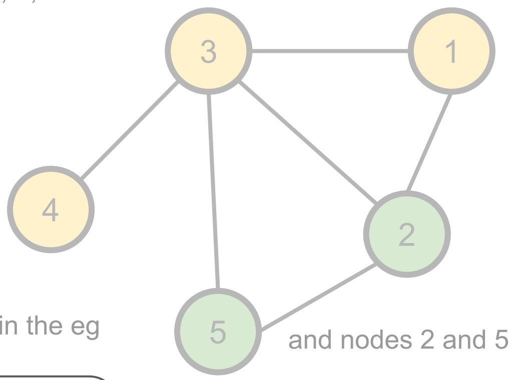
Let G be an undirected graph with n nodes and m edges. Partition the graph into two parts A and B randomly as follows:

For each node v, toss an independent fair coin:

- Heads: Put v in A
- Tails: Put v in B

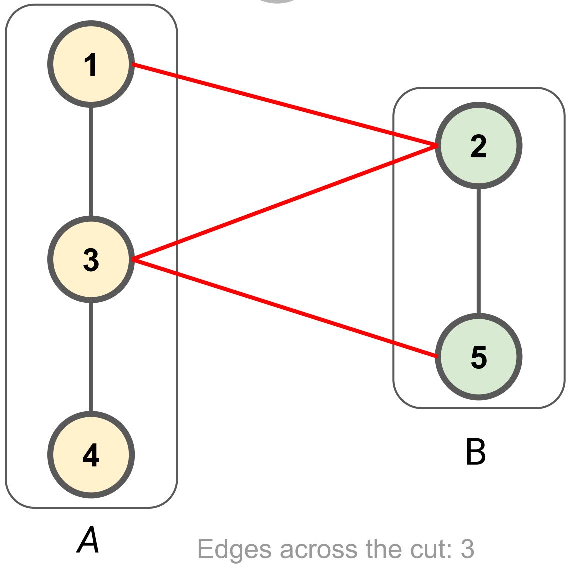
What is the expected number of edges which cross the cut? (One endpoint in A & other in B)

Let's say, nodes 1, 3, and 4 gave us heads



n = 5, m = 6 here in the eg

and nodes 2 and 5 gave tails



Edges across the cut: 3

Question 5 (Answer)

1. Identify a Random Variable to "count" what you want (e.g. X. Goal: $E[X]$)
2. Express this RV as a **sum** of random variables (e.g. $X = X_1 + X_2 + \ldots + X_n$)
   a. Calculate the relevant probability for X_1, X_2, ...
   b. Calculate the individual expectation of the "sub"-random variables. ($E[X_1]$, $E[X_2]$, …)
3. Use linearity of expectations on $E[X]$. Then you add up the expectation of the "sub"-random variables (from step 2b)

Goal: Expected number of edges crossing the cut

1. Identify a Random Variable to “count” what you want (e.g. X. Goal: $E[X]$)
2. Express this RV as a **sum** of random variables (e.g. $X = X_1 + X_2 + \dots + X_n$)
   a. Calculate the relevant probability for $X_1, X_2, \dots$
   b. Calculate the individual expectation of the “sub”-random variables. ($E[X_1], E[X_2], \dots$)
3. Use linearity of expectations on $E[X]$. Then you add up the expectation of the “sub”-random variables (from step 2b)

Question 5 (Answer)

Goal: Expected number of edges crossing the cut

Let X be the random variable representing the number of edges crossing the cut

Question 5 (Answer)

1. Identify a Random Variable to “count” what you want (e.g. X. Goal: $E[X]$)
2. Express this RV as a **sum** of random variables (e.g. $X = X_1 + X_2 + \dots + X_n$)
   a. Calculate the relevant probability for X_1, X_2, ...
   b. Calculate the individual expectation of the “sub”-random variables. ($E[X_1]$, $E[X_2]$, …)
3. Use linearity of expectations on $E[X]$. Then you add up the expectation of the “sub”-random variables (from step 2b)

Goal: Expected number of edges crossing the cut

Let X be the random variable representing the number of edges crossing the cut

For purpose of analysis: label each edge from 1 to m.

Let X_i be the **indicator random variable** that edge i crosses the cut

Question 5 (Answer)

1. Identify a Random Variable to "count" what you want (e.g. X. Goal: $E[X]$)
2. Express this RV as a **sum** of random variables (e.g. $X = X_1 + X_2 + \dots + X_n$)
   a. Calculate the relevant probability for X_1, X_2, ...
   b. Calculate the individual expectation of the "sub"-random variables. ($E[X_1]$, $E[X_2]$, …)
3. Use linearity of expectations on $E[X]$. Then you add up the expectation of the "sub"-random variables (from step 2b)

Goal: Expected number of edges crossing the cut

Let X be the random variable representing the number of edges crossing the cut

For purpose of analysis: label each edge from 1 to m.
Let X_i be the **indicator random variable** that edge i crosses the cut

$X = X_1 + X_2 + \dots X_m$

Question 5 (Answer)

1. Identify a Random Variable to "count" what you want (e.g. X. Goal: $E[X]$)
2. Express this RV as a **sum** of random variables (e.g. $X = X_1 + X_2 + \ldots + X_n$)
   a. Calculate the relevant probability for X_1, X_2, ...
   b. Calculate the individual expectation of the "sub"-random variables. ($E[X_1]$, $E[X_2]$, …)
3. Use linearity of expectations on $E[X]$. Then you add up the expectation of the "sub"-random variables (from step 2b)

Goal: Expected number of edges crossing the cut

Let X be the random variable representing the number of edges crossing the cut

For purpose of analysis: label each edge from 1 to m.
Let X_i be the **indicator random variable** that edge i crosses the cut

$X = X_1 + X_2 + \ldots X_m$

Next things to do:

- Calculate $Pr(X_i = 1)$. This is enough to get $E[X_i]$!
- Then we can calculate $E[X]$ easily by linearity of expectations

Question 5 (Answer)

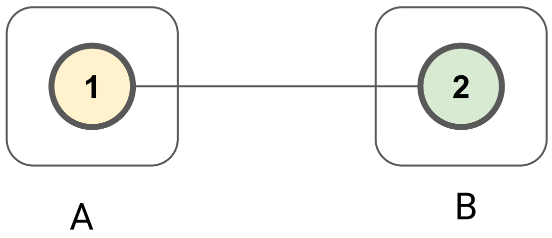
For purpose of analysis: label each edge from 1 to m.
Let X_i be the **indicator random variable** that edge i crosses the cut

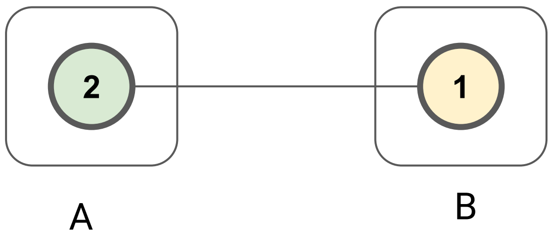
When is $X_i = 1$?

For purpose of analysis: label each edge from *1* to *m*. Let X_i be the **indicator random variable** that edge *i* crosses the cut

Question 5 (Answer)

When is $X_i = 1$? When the two endpoints are in different partitions. 2 cases:





Question 5 (Answer)

For purpose of analysis: label each edge from *1* to *m*.
Let X_i be the **indicator random variable** that edge *i* crosses the cut

When is $X_i = 1$? When the two endpoints are in different partitions. 2 cases:

$Pr(X_i = 1)$
$= (½)(½) + (½)(½)$
$= ¼ + ¼$
$= ½$

First case: 1 goes to A, and 2 goes to B
Second case: 2 goes to A, and 1 goes to B

Implies $E[X_i] = ½$

Let *X* be the random variable representing the number of edges crossing the cut

For purpose of analysis: label each edge from *1* to *m*.
Let X_i be the **indicator random variable** that edge *i* crosses the cut

Question 5 (Answer)

$$E[X] = E[\sum_{j=1}^{m} X_j]$$

Question 5

How we defined X

Question 5 (Answer)

Let *X* be the random variable representing the number of edges crossing the cut

For purpose of analysis: label each edge from *1* to *m*.
Let X_i be the **indicator random variable** that edge *i* crosses the cut

$$E[X] = E[\sum_{j=1}^{m} X_j]$$
$$= \sum_{j=1}^{m} E[X_j]$$

Question 5

Linearity of expectations

Question 5 (Answer)

Let X be the random variable representing the number of edges crossing the cut

For purpose of analysis: label each edge from *1* to *m*.
Let X_i be the **indicator random variable** that edge *i* crosses the cut

$$E[X] = E[\sum_{j=1}^{m} X_j]$$

$$= \sum_{j=1}^{m} E[X_j]$$

$$= \sum_{j=1}^{m} \frac{1}{2}$$

Question 5

Our earlier calculation!

$Pr(X_i = 1)$
$= (½)(½) + (½)(½)$
$= ¼ + ¼$
$= ½$

Implies $E[X_i] = ½$

Question 5 (Answer)

Let *X* be the random variable representing the number of edges crossing the cut

For purpose of analysis: label each edge from *1* to *m*.
Let X_i be the **indicator random variable** that edge *i* crosses the cut

$$
\begin{aligned}
E[X] &= E[\sum_{j=1}^{m} X_j] \\
&= \sum_{j=1}^{m} E[X_j] \\
&= \sum_{j=1}^{m} \frac{1}{2} \\
&= \frac{m}{2}
\end{aligned}
$$

Question 5

Quick mafs