CS109: Independence

Generalized Chain Rule

Chain Rule

Definition of conditional probability:

$$P(E|F) = \frac{P(EF)}{P(F)}$$

normalization const

The Chain Rule:

$$\underbrace{P(EF) = P(E|F)P(F)}$$

$$P(F|E)P(E) = \cancel{P(EF)} = P(E|F)P(F)$$

Generalized Chain Rule

$E_1 E$

$$P(E_1 E_2 E_3 \dots E_n)$$
$$= P(E_1)P(E_2|E_1)P(E_3|E_1 E_2) \dots P(E_n|E_1 E_2 \dots E_{n-1})$$

Pop Quiz

$$P(E_1 E_2 E_3 \ldots E_n) = P(E_1)P(E_2|E_1) \ldots P(E_n|E_1 E_2 \ldots E_{n-1})$$ Chain Rule

You're planning to go to a friend's outdoor Groundhog Day party.

Let C = there is candy
M = there is music
W = you wear a groundhog mask
E = no one wears your mask

When all four events occur, the party is **awesome.**

What is $P(\text{awesome party}) = P(CMWE)$?

A. $P(C)P(M|C)P(W|CM)P(E|CMW)$
B. $P(M)P(C|M)P(W|MC)P(E|MCW)$
C. $P(W)P(E|W)P(CM|EW)$
D. A, B, and C
E. A and B

Pop Quiz

$P(E_1E_2E_3 \dots E_n) = P(E_1)P(E_2|E_1) \dots P(E_n|E_1E_2 \dots E_{n-1})$ Chain Rule

You're planning to go to a friend's outdoor Groundhog Day party.

Let C = there is candy
M = there is music
W = you wear a groundhog mask
E = no one wears your mask

When all four events occur, the party is **awesome.**

What is $P(\text{awesome party}) = P(CMWE)$?

A. $\underbrace{\underbrace{P(C)P(M|C)}_{P(MC)}P(W|CM)}_{P(CMW)}P(E|CMW) \rightarrow P(CMWE)$ ✓

B. $P(M)\underbrace{P(C|M)}_{MC}\underbrace{P(W|MC)}_{PCMC}\underbrace{P(E|\overline{MCW})}_{P(\dots)} \rightarrow P(CMWE)$

C. $\underbrace{P(W)P(E|W)}_{\Rightarrow P(EW)}P(CM|EW) \rightarrow P(CMWE)$

D. A, B, and C

E. A and B

Chain Rule is a way of introducing "order" and "procedure" into probability.

Think of the children

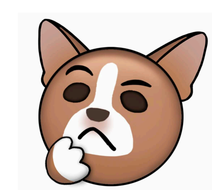
Two parents both have an (A, a) gene pair.

- Each parent will pass along one of their genes to their first child.
- The probability of the first child having curly hair is $P(E_1)$.
- The parents have three children.

What is the probability that all three children have curly hair?

Let $\mathrm{E}_1, \mathrm{E}_2, \mathrm{E}_3$ be the events that child 1, 2, and 3 have curly hair, respectively.

$$P(E_1E_2E_3) = P(E_1)P(E_2|E_1)P(E_3|E_1E_2)$$



Independence I

Independence

Two events E and F are defined as independent if:

$$P(EF) = P(E)P(F)$$

Otherwise E and F are called dependent events.

If E and F are independent, then:

$$P(E|F) = P(E)$$

Intuition through proof

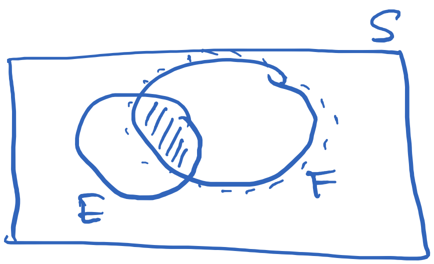
Independent events E and F $\Leftrightarrow$ $P(EF) = P(E)P(F)$

Statement:

If E and F are independent, then $P(E|F) = P(E)$.

Proof:

$$P(E|F) = \frac{P(EF)}{P(F)}$$ Definition of conditional probability



$$= \frac{P(E)P(F)}{P(F)}$$ Independence of E and F

$$= P(E)$$ Taking the bus to cancellation city

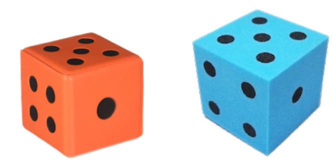
$P(E) = 0.30$
$P(E|F) = 0.30$

Knowing that F happened does not change our belief that E happened.

Dice, our misunderstood friends

Independent events E and F $\Leftrightarrow$ $P(EF) = P(E)P(F)$
$P(E|F) = P(E)$

- Roll two 6-sided dice, yielding values D_1 and D_2.
- Let event E: $D_1 = 1$
  event F: $D_2 = 6$
  event G: $D_1 + D_2 = 5$



$G = \{(1,4), (2,3), (3,2), (4,1)\}$ $|G| = 4$

1. Are E and F independent?

$P(E) = 1/6$
$P(F) = 1/6$
$P(EF) = 1/36$

✅ <u>independent</u>

2. Are E and G independent?

$P(E) = 1/6$
$P(G) = 4/36 = 1/9$
$P(EG) = 1/36 \neq P(E)P(G)$

❌ <u>dependent</u>

Generalizing independence

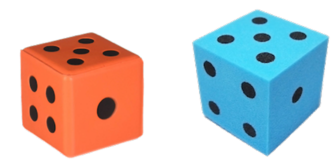
Three events E, F, and G are independent if:

$P(EFG) = P(E)P(F)P(G)$, and
$P(EF) = P(E)P(F)$, and
$P(EG) = P(E)P(G)$, and
$P(FG) = P(F)P(G)$

n events $E_1, E_2, \ldots, E_n$ are independent if:

for $r = 1, \ldots, n$:
for every subset $E_1, E_2, \ldots, E_r$:
$P(E_1, E_2, \ldots, E_r) = P(E_1)P(E_2) \cdots P(E_r)$

Dice, increasingly misunderstood (still our friends)



- Each roll of a 6-sided die is an independent trial.
- Two rolls: D_1 and D_2.
- Let event E: $D_1 = 1$
  event F: $D_2 = 6$
  event G: $D_1 + D_2 = 7$

$\frac{|G|}{36} = \frac{6}{36} = \frac{1}{6}$

$G = \{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\}$

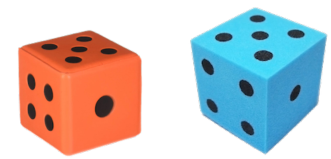
pairwise

triple

1. Are E and F independent? ✅
2. Are E and G independent?
3. Are F and G independent?
4. Are E, F, G independent?

$P(EF) = 1/36$

Dice, increasingly misunderstood (still our friends)



- Each roll of a 6-sided die is an independent trial.
- Two rolls: D_1 and D_2.
- Let event E: $D_1 = 1$

  event F: $D_2 = 6$

  event G: $D_1 + D_2 = 7$

$$G = \overbrace{\{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\}}^{6}$$

1. Are E and F independent? ✅
2. Are E and G independent? ✅
3. Are F and G independent? ✅
4. Are E, F, G independent? ❌

$$P(E) = \frac{1}{6}$$

$$P(G) = \frac{1}{6} = \frac{6}{36}$$

$$P(EG) = \frac{1}{36} = \frac{1}{6} \cdot \frac{1}{6}$$

$$P(G) = \frac{1}{6}$$

$$\neq$$

$$P(G|EF) = 1$$

$$P(GEF) \neq P(E)P(F)P(G)$$

Pairwise independence does not imply universal independence of >2 events!

Independence II

Independent trials

We often are interested in experiments consisting of n **independent trials**.

- n trials, each with the same set of possible outcomes
- n-way independence: an event in one subset of trials is independent of events in other subsets of trials

Examples:

- Flip a coin n times
- Roll a die n times
- Send a multiple-choice survey to n people
- Send n Google Doodle image requests to k different servers

Think of the children as independent trials

Two parents both have an (A, a) gene pair.

- Each parent will pass on one of their genes (each gene equally likely) to a child.
- The probability of **any single child** having curly hair (the recessive trait) is 0.25, independent of other siblings.
- There are three children.

What is the probability that all three children have curly hair?

Let E_1, E_2, E_3 be the events that child 1, 2, and 3 have curly hair, respectively.

$$\begin{aligned} P(E_1 E_2 E_3) &= P(E_1)P(E_2|E_1)P(E_3|E_1E_2) \\ &= P(E_1)P(E_2)P(E_3) \end{aligned}$$

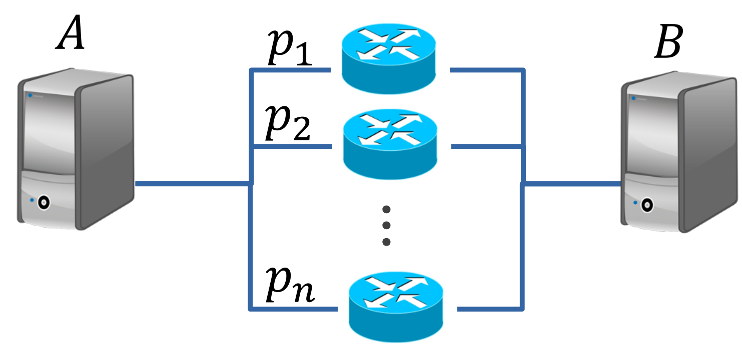
Network reliability

$$\sum_{i=1}^{n} p_i = 1$$

Consider the following parallel network:

- n independent routers, each with probability p_i of functioning (where $1 \le i \le n$)
- E = functional path from A to B exists.

What is $P(E)$?

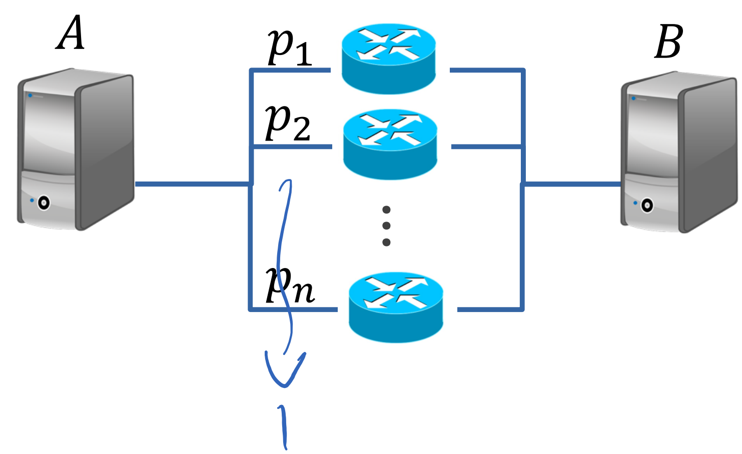


Network reliability

Consider the following parallel network:

- n independent routers, each with probability p_i of functioning (where $1 \leq i \leq n$)
- E = functional path from A to B exists.

What is $P(E)$?



$$
\begin{aligned}
P(E) &= P(\geq 1 \text{ one router works}) \\
&= 1 - P(\text{all routers fail}) \\
&= 1 - (1 - p_1)(1 - p_2) \cdots (1 - p_n) \\
&= 1 - \prod_{i=1}^{n} (1 - p_i)
\end{aligned}
$$

≥ 1 with independent trials: take complement

Ponder

Let's take a two-minute breather.

Slide 21 has two questions to think over by yourself while you do that breathing. We'll go over it together afterwards.

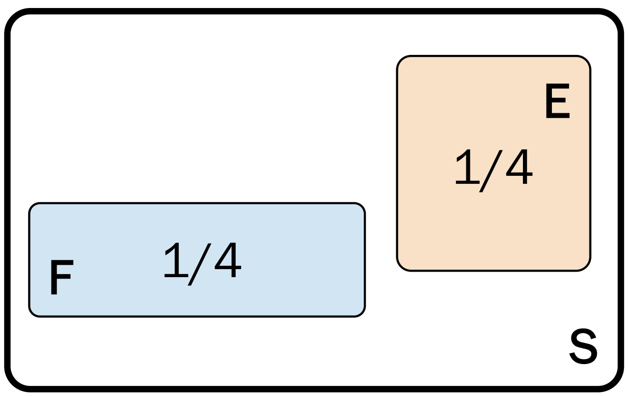
Independence?

Independent events E and F ⟺ $P(EF) = P(E)P(F)$, $P(E|F) = P(E)$

1. True or False? Two events E and F are independent if:
   A. Knowing that F happens means that E can't happen. False
   B. Knowing that F happens doesn't change probability that E happened. True ☺

2. Are E and F independent in the following pictures?

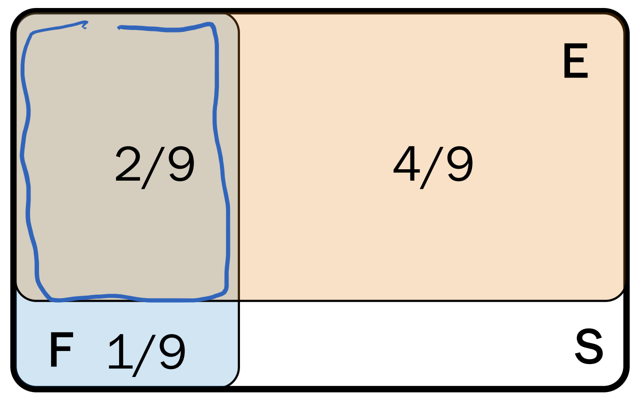
A.



$\underbrace{P(E) = \frac{1}{4}, \; P(F) = \frac{1}{4}}$ $\quad$ $\cancel{P(EF) = \frac{1}{16}}$

$P(EF) = 0$

B.

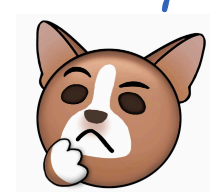


$P(E) = \frac{6}{9}$

$P(F) = \frac{3}{9}$

$P(EF) = \frac{2}{9}$

$\frac{2}{3} \cdot \frac{1}{3} = \frac{2}{9}$



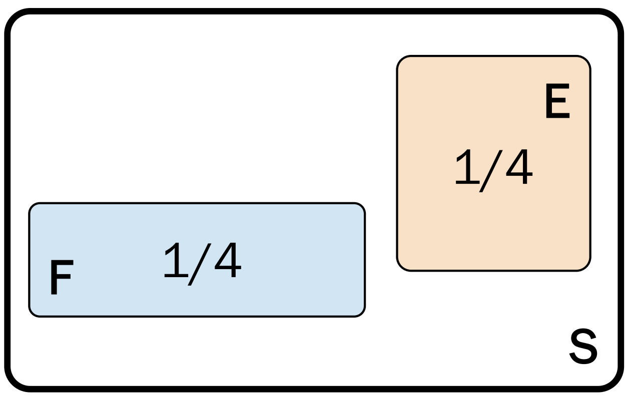
Independence?

Independent events E and F ⟺ $P(EF) = P(E)P(F)$
$P(E|F) = P(E)$

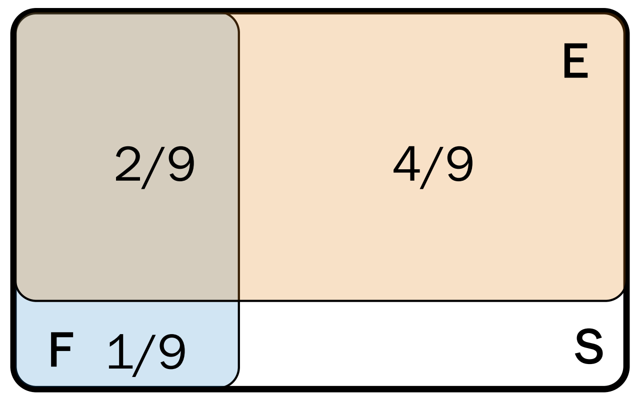
1. True or False? Two events E and F are independent if:
   A. Knowing that F happens means that E can't happen.
   B. Knowing that F happens doesn't change probability that E happened.

2. Are E and F independent in the following pictures?

A.



B.



Be careful:
- Independence is NOT mutual exclusion.
- Independence is difficult to visualize graphically.

Independence

Two events E and F are defined as independent if:

$$P(EF) = P(E)P(F)$$

For independent events E and F,

- $P(E|F) = P(E)$
- E and F^C are independent.

new

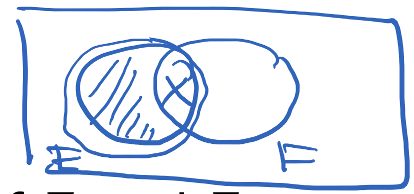
Independence of complements

Statement:

If E and F are independent, then E and F^C are independent.

Proof:

$P(EF^C) = P(E) - P(EF)$	Intersection
$= P(E) - P(E)P(F)$	Independence of E and F
$= P(E)[1 - P(F)]$	Factoring
$= P(E)P(F^C)$	Complement
E and F^C are independent	Definition of independence



Knowing whether F **does or doesn't happen** doesn't change our belief about E happening.

Biased Coin Flips

Suppose we flip a coin n times. Each coin flip is an **independent trial** with probability p of coming up heads and probability q $= 1 - p$ of coming up tails. Write an expression for the following:

$$\overbrace{HHHH \dots H}^{k} \; \overbrace{TTTTTT}^{n-k}$$

1. $P(n$ heads on n coin flips$)$ $= p^n$
2. $P(n$ tails on n coin flips$)$ $\rightarrow q^n = (1-p)^n$
3. $P($first k heads, then $n - k$ tails$)$ $\rightarrow p^k q^{n-k} = p^k (1-p)^{n-k}$
4. $P($exactly k heads on n coin flips$)$

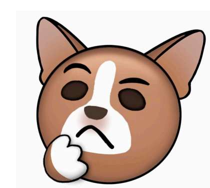
$$H\,T\,\overbrace{HHHH}^{k-1} \cdots H\,T\,T\,T$$

$$p^k q^{n-k}$$

$$\overbrace{HH \cdots H}^{k} \; \overbrace{TTTTT}^{n-k}$$

$$\rightarrow \binom{n}{k} p^k q^{n-k}$$

$$HT\,HT\,HT \dots$$



Biased Coin Flips

Suppose we flip a coin n times. Each coin flip is an **independent trial** with probability p of coming up heads and probability $q = 1 - p$ of coming up tails. Write an expression for the following:

1. $P(n$ heads on n coin flips$)$
2. $P(n$ tails on n coin flips$)$
3. $P($first k heads, then $n - k$ tails$)$
4. $P($exactly k heads on n coin flips$)$

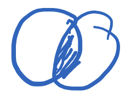
$$\binom{n}{k} p^k (1-p)^{n-k}$$

# of mutually exclusive outcomes

$P($a particular outcome's k heads on n coin flips$)$

Make sure you understand #4! It will come up again.

Probability of events



E or F
$P(E \cup F)$

Mutually exclusive?

Just add!

$P(E) + P(F)$

Inclusion-Exclusion Principle

$P(E) + P(F) - P(E \cap F)$

E and F
$P(EF)$

Independent?

Just multiply!

Chain Rule

Probability of events

E or F

$P(E \cup F)$

Mutually exclusive?

Just add!

$P(E) + P(F)$

Inclusion-Exclusion Principle

$P(E) + P(F) - P(E \cap F)$

E and F

$P(EF)$

Independent?

Just multiply!

$P(E)P(F)$

Chain Rule

$P(E)P(F|E)$

or

$P(F)P(E|F)$

Probability of events

De Morgan's

E or F
$P(E \cup F)$

E and F
$P(EF)$

Mutually exclusive?

Just add!

Inclusion-Exclusion Principle

Independent?

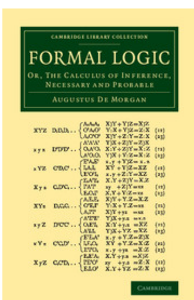
Just multiply!

Chain Rule

Augustus De Morgan

Augustus De Morgan (1806–1871):

British mathematician who wrote the book *Formal Logic* (1847).

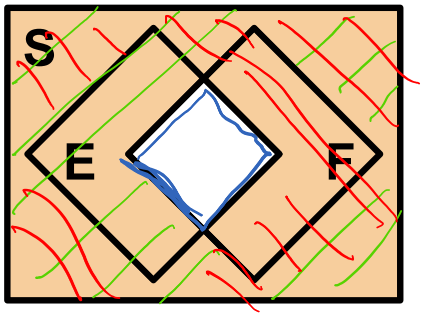


He looks like Jason Alexander (George from Seinfeld)

(but that's not important right now)

De Morgan's Laws

De Morgan's Law lets you alternate between AND and OR.



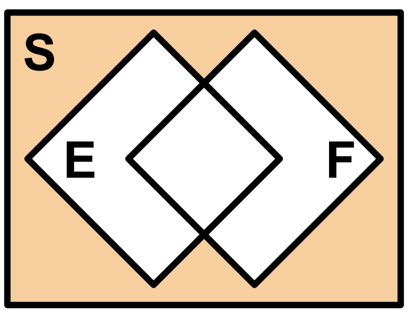
$(E \cap F)^C = E^C \cup F^C$

$$\left(\bigcap_{i=1}^{n} E_i\right)^C = \bigcup_{i=1}^{n} E_i^C$$

In probability:

$$P(E_1 E_2 \cdots E_n)$$
$$= 1 - P\left((E_1 E_2 \cdots E_n)^C\right)$$
$$= 1 - P\left(E_1^C \cup E_2^c \cup \cdots \cup E_n^c\right)$$

Great if E_i^C mutually exclusive!



$(E \cup F)^C = E^C \cap F^C$

$$\left(\bigcup_{i=1}^{n} E_i\right)^C = \bigcap_{i=1}^{n} E_i^C$$

In probability:

$$P(E_1 \cup E_2 \cup \cdots \cup E_n)$$
$$= 1 - P\left((E_1 \cup E_2 \cup \cdots \cup E_n)^C\right)$$
$$= 1 - P(E_1^c E_2^c \cdots E_n^c)$$

Great if E_i independent!

Challenge

Slide 33 presents a few questions exercising your understanding of hash tables, independence, and De Morgan's Law.

With your smarts and Doris's love to guide you, we'll get through it.

Hash table fun

p_1 p_2

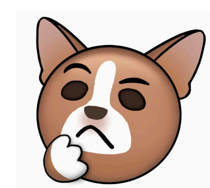
B_1 B_2

- m strings are hashed (not uniformly) into a hash table with n buckets.
- Each string hash is an independent trial w.p. p_i of getting hashed into bucket i.

What is $P(E)$ if

1. E = bucket 1 has $\geq$ 1 string hashed into it?

2. E = at least 1 of buckets 1 to k has $\geq$ 1 string hashed into it?



Hash table fun

- m strings are hashed (not uniformly) into a hash table with n buckets.
- Each string hash is an independent trial w.p. p_i of getting hashed into bucket i.

What is $P(E)$ if

1. E = bucket 1 has ≥ 1 string hashed into it?

Define: S_i = string i hashes to bucket 1
S_i^C = string i doesn't hash to bucket 1

$$P(S_i) = p_1$$
$$P(S_i^C) = 1 - p_1$$

Hash table fun

- m strings are hashed (not uniformly) into a hash table with n buckets.
- Each string hash is an independent trial w.p. p_i of getting hashed into bucket i.

What is $P(E)$ if

1. E = bucket 1 has $\geq$ 1 string hashed into it?

Define: S_i = string i hashes to bucket 1
S_i^C = string i doesn't hash to bucket 1

$P(S_i) = p_1$
$P(S_i^C) = 1 - p_1$

WTF (not-real acronym for Want To Find):

$$P(E) = P(S_1 \cup S_2 \cup \cdots \cup S_m)$$

$$= 1 - P\left((S_1 \cup S_2 \cup \cdots \cup S_m)^C\right) \quad \text{Complement}$$

$$= 1 - P\left(S_1^C S_2^C \cdots S_m^C\right) \quad \text{De Morgan's Law}$$

m

$$= 1 - P(S_1^C)P(S_2^C) \cdots P(S_m^C) = 1 - \left(P(S_1^C)\right)^m \quad S_i \text{ independent trials}$$

$$= 1 - (1 - p_1)^m$$

More hash table fun: Possible approach?

- m strings are hashed (not uniformly) into a hash table with n buckets.
- Each string hash is an independent trial w.p. p_i of getting hashed into bucket i.

What is $P(E)$ if

1. E = bucket 1 has $\geq$ 1 string hashed into it?
2. E = at least 1 of buckets 1 to k has $\geq$ 1 string hashed into it?

Define F_i = bucket i has at least one string in it

$$
\begin{aligned}
P(E) &= P(F_1 \cup F_2 \cup \cdots \cup F_k) \\
&= 1 - P\left((F_1 \cup F_2 \cup \cdots \cup F_k)^C\right) \\
&= 1 - P\left(F_1^C F_2^C \cdots F_k^C\right) \\
&?= 1 - \cancel{P\left(F_1^C\right) P\left(F_2^C\right)} \cdots P(F_k^C)
\end{aligned}
$$

⚠️ F_i bucket events are *dependent*!

So we cannot approach with complement.

More hash table fun

- m strings are hashed (not uniformly) into a hash table with n buckets.
- Each string hash is an independent trial w.p. p_i of getting hashed into bucket i.

What is $P(E)$ if

1. E = bucket 1 has $\geq$ 1 string hashed into it?
2. E = at least 1 of buckets 1 to k has $\geq$ 1 string hashed into it?

Define F_i = bucket i has at least one string in it

$$P(E) = P(F_1 \cup F_2 \cup \cdots \cup F_k)$$
$$= 1 - P\left((F_1 \cup F_2 \cup \cdots \cup F_k)^C\right)$$
$$= 1 - P\left(F_1^C F_2^C \cdots F_k^C\right)$$

$= P$(buckets 1 to k **all denied strings**)
$= \left(P(\text{each string hashes to } k + 1 \text{ or higher})\right)^m$
$= (1 - p_1 - p_2 ... - p_k)^m$

$$= 1 - (1 - p_1 - p_2 ... - p_k)^m$$

The fun never stops with hash tables

- m strings are hashed (not uniformly) into a hash table with n buckets.
- Each string hash is an independent trial w.p. p_i of getting hashed into bucket i.

What is $P(E)$ if

1. E = bucket 1 has ≥ 1 string hashed into it? ✅
2. E = at least 1 of buckets 1 to k has ≥ 1 string hashed into it? ✅

Looking for a challenge? ☺

The fun never stops with hash tables

- m strings are hashed (unequally) into a hash table with n buckets.
- Each string hash is an independent trial w.p. p_i of getting hashed into bucket i.

What is $P(E)$ if

1. E = bucket 1 has $\geq$ 1 string hashed into it?
2. E = at least 1 of buckets 1 to k has $\geq$ 1 string hashed into it?
3. E = each of buckets 1 to k has $\geq$ 1 string hashed into it?

Hint: Use Part 2's event definition:

Define F_i = bucket i has at least one string in it

Hint: Try $k = 2$, then $k = 3$, then generalize.