Foundations of Computer Science
Lecture 17

Independent Events

Last Time

1. New information changes a probability.
2. Conditional probability.
3. Conditional probability traps.
   - Sampling bias, using $\mathbb{P}[A]$ instead of $\mathbb{P}[A \mid B]$.
   - Transposed conditional, using $\mathbb{P}[B \mid A]$ instead of $\mathbb{P}[A \mid B]$.
   - Medical testing.
4. Law of total probability.
   - Case by case probability analysis.

Today: Independent Events

1. Independence is an assumption
   - Fermi method
   - Multiway independence
2. Coincidence and the birthday paradox
   - Application to hashing
3. Random walk and gambler's ruin

Independence is a Simplifying Assumption

- Sex of first child has nothing to do with sex of second → independent.
- What about eyecolor? (Depends on genes of parent.) → not independent.
- Tosses of different coins have nothing to do with each other →independent.
- Cloudy and rainy days. When it rains, there must be clouds. → not independent.

Toss two coins.

$$\mathbb{P}[\text{Coin 1=H}] = \tfrac{1}{2} \qquad \mathbb{P}[\text{Coin 2=H}] = \tfrac{1}{2} \qquad \mathbb{P}[\text{Coin 1=H AND Coin 2=H}] = \tfrac{1}{4}$$

Toss 100 times: Coin 1 $\approx$ 50H (of these) → Coin 2 $\approx$ 25H (independent)

$$\mathbb{P}[\text{Coin 1=H AND Coin 2=H}] = \tfrac{1}{4} = \tfrac{1}{2} \times \tfrac{1}{2} = \mathbb{P}[\text{Coin 1=H}] \times \mathbb{P}[\text{Coin 2=H}].$$

$$\mathbb{P}[\text{rain AND clouds}] = \mathbb{P}[\text{rain}] = \tfrac{1}{7} \gg \tfrac{1}{35} = \mathbb{P}[\text{rain}] \times \mathbb{P}[\text{clouds}].$$ (not independent)

Definition of Independence

Events A and B are independent if "They have nothing to do with each other."
Knowing the outcome is in B does not change the probability that the outcome is in A.

The events A and B are independent if
$$\mathbb{P}[A \text{ AND } B] \;=\; \mathbb{P}[A \cap B] \;=\; \mathbb{P}[A] \times \mathbb{P}[B].$$
In general, $\mathbb{P}[A \cap B] = \mathbb{P}[A \mid B] \times \mathbb{P}[B]$. Independence means that
$$\mathbb{P}[A \mid B] = \mathbb{P}[A].$$

Independence is a non-trivial assumption, and you can't always assume it.

When you can assume independence

PROBABILITIES MULTIPLY

Fermi-Method: How Many Dateable Girls Are Out There?

A_1 = "Lives nearby"; A_2 = "Right sex"; A_3 = "Right age"; A_4 = "Single";
A_5 = "Educated"; A_6 = "Attractive"; A_7 = "Finds me attractive"; A_8 = "We get along".

$$A = A_1 \cap A_2 \cap A_3 \cap A_4 \cap A_5 \cap A_6 \cap A_7 \cap A_8 \qquad \text{(all criteria must be met)}$$

Independence:
$$\mathbb{P}[A] \;=\; \mathbb{P}[A_1 \cap A_2 \cap A_3 \cap A_4 \cap A_5 \cap A_6 \cap A_7 \cap A_8].$$

$\mathbb{P}$["Lives nearby"]	$\frac{\text{number(nearby)}}{\text{number(world)}} \approx \frac{20\text{ million}}{7\text{ billion}} \approx \frac{3}{1000}$
$\mathbb{P}$["Right sex"]	$\frac{1}{2}$ (there are about 50% male and 50% female in the world)
$\mathbb{P}$["Right age"]	$\frac{15}{100}$ (about 15% of people between 20 and 30)
$\mathbb{P}$["Single"]	$\frac{1}{2}$ (about 50% of people are single)
$\mathbb{P}$["Educated"]	$\frac{1}{4}$ (about 25% in the US have a college degree)
$\mathbb{P}$["Attractive"]	$\frac{1}{5}$ (you find 1 in 5 people attractive)
$\mathbb{P}$["Finds me attractive"]	$\frac{1}{10}$ (you are modest)
$\mathbb{P}$["We get along"]	$\frac{1}{16}$ (you get along with 1 in 4 people and assume so for her)

$$\mathbb{P}[\text{"Dateable"}] \;=\; \tfrac{3}{1000} \times \tfrac{1}{2} \times \tfrac{15}{100} \times \tfrac{1}{2} \times \tfrac{1}{4} \times \tfrac{1}{5} \times \tfrac{1}{10} \times \tfrac{1}{16} \times \approx 3.5 \times 10^{-8},$$

1-in-30 million (or 250) dateable girls.

Multiway Independence

Ω	HHH	HHT	HTH	HTT	THH	THT	TTH	TTT
$P(\omega)$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$

A_1 = {coins 1,2 match}
A_2 = {coins 2,3 match}
A_3 = {coins 1,3 match}

- $\mathbb{P}[A_1] = \mathbb{P}[A_2] = \mathbb{P}[A_3] = \frac{1}{2}$.
- $\mathbb{P}[A_1 \cap A_2] = \mathbb{P}[A_2 \cap A_3] = \mathbb{P}[A_1 \cap A_3] = \frac{1}{4}$. (independent)
- $\mathbb{P}[A_1 \cap A_2 \cap A_3] = \frac{1}{4}$. (1,2) match AND (2,3) match → (1,3) match.

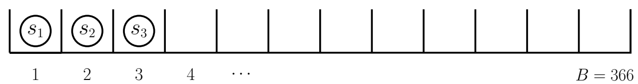
2-way independent, not 3-way independent.

$A_1, \ldots, A_n$ are **independent** if the probability of *any intersection* of distinct events is the *product* of the event-probabilities of those events,
$$\mathbb{P}[A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}] = \mathbb{P}[A_{i_1}] \cdot \mathbb{P}[A_{i_2}] \cdots \mathbb{P}[A_{i_k}].$$

Coincidence: Let's Try to Find a FOCS-Twin

Two hundred students $S = \{s_1, \ldots, s_{200}\}$,

- Birthdays are *independent* (no twins, triplets, ...) and all birthdays are equally likely.



$$\mathbb{P}[s_1 \text{ has no FOCS-twin}] = \left(\tfrac{B-1}{B}\right)^{N-1} = \left(\tfrac{365}{366}\right)^{199}$$
$$\mathbb{P}[s_2 \text{ has no FOCS-twin} \mid s_1 \text{ has no FOCS-twin}] = \left(\tfrac{B-2}{B-1}\right)^{N-2} = \left(\tfrac{364}{365}\right)^{198}$$
$$\mathbb{P}[s_3 \text{ has no FOCS-twin} \mid s_1, s_2 \text{ have no FOCS-twin}] = \left(\tfrac{B-3}{B-2}\right)^{N-3} = \left(\tfrac{363}{364}\right)^{197}$$
$$\vdots$$
$$\mathbb{P}[s_k \text{ has no FOCS-twin} \mid s_1, \ldots, s_{k-1} \text{ have no FOCS-twin}] = \left(\tfrac{B-k}{B-k+1}\right)^{N-k} = \left(\tfrac{366-k}{366-k+1}\right)^{N-k}$$

$$\mathbb{P}[s_1, \ldots, s_k \text{ have no FOCS-twin}] = \left(\tfrac{365}{366}\right)^{199} \times \left(\tfrac{364}{365}\right)^{198} \times \left(\tfrac{363}{364}\right)^{197} \times \cdots \times \left(\tfrac{366-k}{366-k+1}\right)^{N-k} \approx 0.58$$

Finding a FOCS-twin by the kth student with class size 200

k	1	2	3	4	5	6	7	8	9	10	23	25
chances (%)	42.0	66.3	80.4	88.6	93.3	96.1	97.7	98.7	99.2	99.5	99.999	100

The Birthday Paradox

In a party of 50 people, what are the chances that two have the same birthday?

Same as asking for

$$\mathbb{P}[s_1, \ldots, s_{50} \text{ have no FOCS-twin}].$$

Answer:

$$\mathbb{P}[\text{no social twins}] = \left(\tfrac{365}{366}\right)^{49} \times \left(\tfrac{364}{365}\right)^{48} \times \left(\tfrac{363}{364}\right)^{47} \times \cdots \times \left(\tfrac{315}{316}\right)^{0} \approx 0.03.$$

Chances are about 97% that two people share a birthday!

Moral: when *searching* for something among many options (1225 pairs of people), *do not be surprised* when you find it.

Search and Hashing

```
http://page.1
dirty apples
hurt health
```

```
http://page.2
health freaks
hate dirty
apples
```

```
http://page.3
survey: people
hate bananas
```

Example Queries

search(`apples`) = {page.1, page.2}
search(`hate`) = {page.2, page.3}
search(`bananas`) = {page.3}

Web-address Directory

`apples` → {page.1, page.2}
`bananas` → {page.3}
`dirty` → {page.1, page.2}
`freaks` → {page.2}
`hate` → {page.2, page.3}
`health` → {page.1, page.2}
`hurt` → {page.1}
`people` → {page.3}
`survey` → {page.3}

$O(\log N)$ search

Hash words into a table (array) using a hash function H(w), e.g:

$$\text{H}(\texttt{hate}) = 8^{17} + 1^{17} + 20^{17} + 5^{17} \pmod{11} = 7$$

search(w): GOTO hash-table row H(w).

Collisions: (hate,freaks), (survey,apples)
Problem: What if you search for hate or survey?

Good hash function maps words independently and randomly.
No collisions → $O(1)$ search (constant time, not $\log N$).

0	`bananas` → {page.3}
1	
2	`hurt` → {page.1}
3	`people` → {page.3}
4	`dirty` → {page.1, page.2}
5	
6	
7	`freaks` → {page.2} `hate` → {page.2, page.3}
8	
9	`apples` → {page.1, page.2} `survey` → {page.3}
10	`health` → {page.1, page.2}

Hashing and FOCS-twins

Words $w_1, w_2 \ldots, w_N$ and Hashing	↔	Students $s_1, s_2, \ldots, s_N$ and Birthdays
$w_1, \ldots, w_N$ HASHED to rows $0, 1, \ldots, B-1$	↔	$s_1, \ldots, s_N$ BORN on days $0, 1, \ldots, B-1$
No collisions, or HASH-twins	↔	No FOCS-twins

Example: Suppose you have $N = 10$ words $w_1, w_2, \ldots, w_{10}$.

$B = 10$ (hash table has as many rows as words).

$$\mathbb{P}[\text{no collisions}] = \left(\tfrac{9}{10}\right)^9 \times \left(\tfrac{8}{9}\right)^8 \times \left(\tfrac{7}{8}\right)^7 \times \left(\tfrac{6}{7}\right)^6 \times \left(\tfrac{5}{6}\right)^5 \times \left(\tfrac{4}{5}\right)^4 \times \left(\tfrac{3}{4}\right)^3 \times \left(\tfrac{2}{3}\right)^2 \times \left(\tfrac{1}{2}\right)^1 \times \left(\tfrac{0}{1}\right)^0 \approx 0.0004.$$

$B = 20$ (hash table has as twice many rows as words).

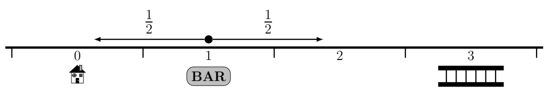
$$\mathbb{P}[\text{no collisions}] = \left(\tfrac{19}{20}\right)^9 \times \left(\tfrac{18}{19}\right)^8 \times \left(\tfrac{17}{18}\right)^7 \times \left(\tfrac{16}{17}\right)^6 \times \left(\tfrac{15}{16}\right)^5 \times \left(\tfrac{14}{15}\right)^4 \times \left(\tfrac{13}{14}\right)^3 \times \left(\tfrac{12}{13}\right)^2 \times \left(\tfrac{11}{12}\right)^1 \times \left(\tfrac{10}{11}\right)^0 \approx 0.07.$$

B	10	20	30	40	50	60	70	80	90	100	500	1000
$\mathbb{P}$[no collisions]	0.0004	0.07	0.18	0.29	0.38	0.45	0.51	0.56	0.60	0.63	0.91	0.96

B large enough → chances of no collisions are high (that's good). How large should B be?

Theorem. If $B \in \omega(N^2)$, then $\mathbb{P}[\text{no collisions}] \to 1$

Random Walk: What are the Chances the Drunk Gets Home?



Infinite Outcome Tree

Sequences leading to home:

L	RLL	RLRLL	RLRLRLL	RLRLRLRLL	⋯
$\frac{1}{2}$	$(\frac{1}{2})^3$	$(\frac{1}{2})^5$	$(\frac{1}{2})^7$	$(\frac{1}{2})^9$	⋯

$$P((\text{RL})^{\bullet i}\text{L}) = (\tfrac{1}{2})^{2i+1}$$

$$\begin{aligned}\mathbb{P}[\text{home}] &= \tfrac{1}{2} + (\tfrac{1}{2})^3 + (\tfrac{1}{2})^5 + (\tfrac{1}{2})^7 + (\tfrac{1}{2})^9 + \cdots \\ &= \frac{\frac{1}{2}}{1-\frac{1}{4}} \\ &= \tfrac{2}{3}.\end{aligned}$$

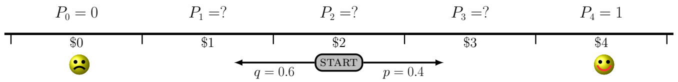
Total Probability

$$\begin{aligned}\mathbb{P}\,[\text{home}] &= \mathbb{P}[\text{L}] \cdot \mathbb{P}[\text{home} \mid \text{L}] && \leftarrow \tfrac{1}{2} \times 1 \\ &\quad + \mathbb{P}[\text{RR}] \cdot \mathbb{P}[\text{home} \mid \text{RR}] && \leftarrow \tfrac{1}{4} \times 0 \\ &\quad + \mathbb{P}[\text{RL}] \cdot \mathbb{P}[\text{home} \mid \text{RL}] && \leftarrow \tfrac{1}{4} \times \mathbb{P}[\text{home}] \\ &= \tfrac{1}{2} + \tfrac{1}{4}\,\mathbb{P}\,[\text{home}].\end{aligned}$$

That is, $(1 - \frac{1}{4})\,\mathbb{P}\,[\text{home}] = \frac{1}{2}$. Solve for $\mathbb{P}[\text{home}]$:

$$\begin{aligned}\mathbb{P}[\text{home}] &= \frac{\frac{1}{2}}{1-\frac{1}{4}} \\ &= \tfrac{2}{3}.\end{aligned}$$

Doubling Up: A Random Walk at the Casino



P_i is the probability to win in the game if you have \$$i$.

$$P_1 = qP_0 + pP_2 = pP_2. \qquad \leftarrow \text{total expectation}$$

$$P_2 = qP_1 + pP_3 = pqP_2 + pP_3 \rightarrow P_2 = \frac{pP_3}{1 - pq}.$$

$$P_3 = qP_2 + pP_4 = \frac{pqP_3}{1 - pq} + p \rightarrow P_3 = \frac{p(1 - pq)}{1 - 2pq}.$$

Conclusion:

$$P_2 = \frac{p^2}{1 - 2pq} \approx 0.31$$

(69% chances of RUIN)

Exercise.

- What if you are trying to double up from \$3? (Answer: 77% chance of RUIN).
- What if you are trying to double up from \$10? (Answer: 98% chance of RUIN).

The *richer* the Gambler, the greater the chances of RUIN!