Foundations of Computer Science
Lecture 17

Independent Events

Independence is a Powerful Assumption
The Fermi Method
Coincidence and the Birthday Paradox
Application to Hashing
Random Walks and Gambler’s Ruin
New information changes a probability.

Conditional probability.

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.

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

Toss two coins.

\[
\begin{align*}
\Pr[\text{Coin 1}=\text{H}] &= \frac{1}{2} \\
\Pr[\text{Coin 2}=\text{H}] &= \frac{1}{2} \\
\Pr[\text{Coin 1}=\text{H} \ \text{AND} \ \text{Coin 2}=\text{H}] &= \frac{1}{4}
\end{align*}
\]

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

\[
\Pr[\text{Coin 1}=\text{H} \ \text{AND} \ \text{Coin 2}=\text{H}] = \frac{1}{4} = \frac{1}{2} \times \frac{1}{2} = \Pr[\text{Coin 1}=\text{H}] \times \Pr[\text{Coin 2}=\text{H}].
\]

\[
\Pr[\text{rain AND clouds}] = \Pr[\text{rain}] = \frac{1}{7} \gg \frac{1}{35} = \Pr[\text{rain}] \times \Pr[\text{clouds}]. \quad \text{(not independent)}
\]
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
\[ P[A \text{ AND } B] = P[A \cap B] = P[A] \times P[B]. \]
In general, $P[A \cap B] = P[A \mid B] \times P[B]$. Independence means that
\[ P[A \mid B] = P[A]. \]

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?

\[ A = A_1 \cap A_2 \cap A_3 \cap A_4 \cap A_5 \cap A_6 \cap A_7 \cap A_8 \]  
(all criteria must be met)

Independence:

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

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

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

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

\[ \Omega \begin{bmatrix} \text{HHH} & \text{HHT} & \text{HTH} & \text{HTT} & \text{THH} & \text{THT} & \text{TTH} & \text{TTT} \end{bmatrix} \]

\[ P(\omega) \begin{bmatrix} \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} \end{bmatrix} \]

\[ A_1 = \{ \text{coins 1,2 match} \} \]
\[ A_2 = \{ \text{coins 2,3 match} \} \]
\[ A_3 = \{ \text{coins 1,3 match} \} \]

- \[ P[A_1] = P[A_2] = P[A_3] = \frac{1}{2} \].
- \[ P[A_1 \cap A_2] = P[A_2 \cap A_3] = P[A_1 \cap A_3] = \frac{1}{4} \] (independent)
- \[ P[A_1 \cap A_2 \cap A_3] = \frac{1}{4} \].

(1,2) match AND (2,3) match → (1,3) match.

2-way independent, not 3-way independent.

\[ A_1, \ldots, A_n \text{ are independent if the probability of any intersection of distinct events is the product of the event-probabilities of those events,} \]
\[ P[A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}] = P[A_{i_1}] \cdot P[A_{i_2}] \cdots P[A_{i_k}]. \]
Coincidence: Let’s Try to Find a FOCS-Twin

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

- Birthdays are \textit{independent} (no twins, triplets, \ldots) and all birthdays are equally likely.

\[
\begin{array}{cccccccccc}
S_1 & S_2 & S_3 & \cdots & & & & & & B = 366 \\
1 & 2 & 3 & 4 & \cdots & & & & & \\
\end{array}
\]

\[
\begin{align*}
P[s_1 \text{ has no FOCS-twin}] &= \left(\frac{B-1}{B}\right)^{N-1} = \left(\frac{365}{366}\right)^{199} \\
P[s_2 \text{ has no FOCS-twin} \mid s_1 \text{ has no FOCS-twin}] &= \left(\frac{B-2}{B-1}\right)^{N-2} = \left(\frac{364}{365}\right)^{198} \\
P[s_3 \text{ has no FOCS-twin} \mid s_1, s_2 \text{ have no FOCS-twin}] &= \left(\frac{B-3}{B-2}\right)^{N-3} = \left(\frac{363}{364}\right)^{197} \\
&\vdots \\
P[s_k \text{ has no FOCS-twin} \mid s_1, \ldots, s_{k-1} \text{ have no FOCS-twin}] &= \left(\frac{B-k}{B-k+1}\right)^{N-k} = \left(\frac{366-k}{366-k+1}\right)^{N-k} \\
P[s_1, \ldots, s_k \text{ have no FOCS-twin}] &= \left(\frac{365}{366}\right)^{199} \times \left(\frac{364}{365}\right)^{198} \times \left(\frac{363}{364}\right)^{197} \times \cdots \times \left(\frac{366-k}{366-k+1}\right)^{N-k} \approx 0.58
\end{align*}
\]

### Finding a FOCS-twin by the \( k \)th student with class size 200

<table>
<thead>
<tr>
<th>( k )</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>23</th>
<th>25</th>
</tr>
</thead>
<tbody>
<tr>
<td>chances (%)</td>
<td>42.0</td>
<td>66.3</td>
<td>80.4</td>
<td>88.6</td>
<td>93.3</td>
<td>96.1</td>
<td>97.7</td>
<td>98.7</td>
<td>99.2</td>
<td>99.5</td>
<td>99.999</td>
<td>100</td>
</tr>
</tbody>
</table>
The Birthday Paradox

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

Same as asking for

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

Answer:

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

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

Example Queries

\[
\begin{align*}
\text{search(apples)} &= \{\text{page.1, page.2}\} \\
\text{search(hate)} &= \{\text{page.2, page.3}\} \\
\text{search(bananas)} &= \{\text{page.3}\}
\end{align*}
\]

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

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

\text{search}(w): \text{GOTO hash-table row } H(w).

Collisions: \((\text{hate}, \text{freaks}), (\text{survey}, \text{apples})\)

Problem: What if you search for \text{hate} or \text{survey}?

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

Web-address Directory

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>bananas ( \rightarrow {\text{page.3}} )</td>
</tr>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>hurt ( \rightarrow {\text{page.1}} )</td>
</tr>
<tr>
<td>3</td>
<td>people ( \rightarrow {\text{page.3}} )</td>
</tr>
<tr>
<td>4</td>
<td>dirty ( \rightarrow {\text{page.1, page.2}} )</td>
</tr>
<tr>
<td>5</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>freaks ( \rightarrow {\text{page.2}} ), hate ( \rightarrow {\text{page.2, page.3}} )</td>
</tr>
<tr>
<td>8</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>apples ( \rightarrow {\text{page.1, page.2}} ), survey ( \rightarrow {\text{page.3}} )</td>
</tr>
<tr>
<td>10</td>
<td>health ( \rightarrow {\text{page.1, page.2}} )</td>
</tr>
</tbody>
</table>
Hashing and FOCS-twins

Words \(w_1, w_2 \ldots, w_N\) and Hashing \(\leftrightarrow\) Students \(s_1, s_2, \ldots, s_N\) and Birthdays
\(w_1, \ldots, w_N\) hashed to rows \(0, 1, \ldots, B - 1\) \(\leftrightarrow\) \(s_1, \ldots, s_N\) born on days \(0, 1, \ldots, B - 1\)

No collisions, or HASH-twins \(\leftrightarrow\) 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).

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

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

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

\[
\begin{array}{cccccccccccc}
B & 10 & 20 & 30 & 40 & 50 & 60 & 70 & 80 & 90 & 100 & 500 & 1000 \\
\Pr[\text{no collisions}] & 0.0004 & 0.07 & 0.18 & 0.29 & 0.38 & 0.45 & 0.51 & 0.56 & 0.60 & 0.63 & 0.91 & 0.96 \\
\end{array}
\]

\(B\) large enough \(\rightarrow\) chances of no collisions are high (that’s good). How large should \(B\) be?

**Theorem.** If \(B \in \omega(N^2)\), then \(\Pr[\text{no collisions}] \rightarrow 1\)
### Infinite Outcome Tree

Sequences leading to home:

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td>RLL</td>
<td>RLRLL</td>
</tr>
<tr>
<td>( \frac{1}{2} )</td>
<td>( \left( \frac{1}{2} \right)^3 )</td>
<td>( \left( \frac{1}{2} \right)^5 )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RLRRL</td>
<td>RLRRLRLL</td>
<td>RLRRLRLL</td>
</tr>
<tr>
<td>( \frac{1}{2} )</td>
<td>( \left( \frac{1}{2} \right)^7 )</td>
<td>( \left( \frac{1}{2} \right)^9 )</td>
</tr>
</tbody>
</table>

\[
P((RL)^i L) = \left( \frac{1}{2} \right)^{2i+1}
\]

\[
P[\text{home}] = \frac{1}{2} + \left( \frac{1}{2} \right)^3 + \left( \frac{1}{2} \right)^5 + \left( \frac{1}{2} \right)^7 + \left( \frac{1}{2} \right)^9 + \cdots
\]

\[
= \frac{\frac{1}{2}}{1 - \frac{1}{4}}
\]

\[
= \frac{2}{3}.
\]

### Total Probability

\[
\begin{align*}
P[\text{home}] &= P[L] \cdot P[\text{home} \mid L] \leftarrow \frac{1}{2} \times 1 \\
&\quad + P[RR] \cdot P[\text{home} \mid RR] \leftarrow \frac{1}{4} \times 0 \\
&\quad + P[RL] \cdot P[\text{home} \mid RL] \leftarrow \frac{1}{4} \times P[\text{home}] \\
&= \frac{1}{2} + \frac{1}{4} P[\text{home}].
\end{align*}
\]

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

\[
P[\text{home}] = \frac{\frac{1}{2}}{1 - \frac{1}{4}}
\]

\[
= \frac{2}{3}.
\]
Doubling Up: A Random Walk at the Casino

$P_0 = 0 \quad P_1 =? \quad P_2 =? \quad P_3 =? \quad P_4 = 1$

$\begin{array}{c|c|c|c|c}
\$0 & \$1 & \$2 & \$3 & \$4 \\
\hline
q = 0.6 & p = 0.4 & & & \\
\end{array}$

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

\[ P_1 = qP_0 + pP_2 = pP_2. \]

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

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

Conclusion:

\[ P_2 = \frac{p^2}{1 - 2pq} \approx 0.31 \quad (69\% \text{ chances of RUIN}) \]

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!