GAMBLING 101

- You choose a die first, I choose second
- We both throw; higher number wins
- Which die would you choose?

\[
\begin{array}{cccc}
5 & 5 & 5 & 5 \\
7 & 7 & 7 & 7 \\
5 & 5 & 5 & 5 \\
A & B & C & D \\
\end{array}
\]

GAMBLING 101

- Antoine Gombaud (1607-1684) made history for being a loser
  I will roll a die four times; I win if I get a 1
- After a while no one would take the bet
- \[1 - \left(\frac{5}{6}\right)^4 = 0.518\]
GAMBLING 101

• Gombaud invented a new scam: I will roll two dice 24 times; I win if I get a double 1

• Why was he losing money?

• \[ 1 - \left(\frac{35}{36}\right)^{24} = 0.491 \]

• Gombaud wrote to Pascal and Fermat, who subsequently created probability theory

PENNIES AND GOLD

• Three bags contain two gold coins, two pennies, and one of each

• Bag is chosen at random, and one coin from it is selected at random; the coin is gold

• Poll 1: What is the probability that the other coin is gold?

  - \( z = \frac{1}{6} \)
  - \( x = \frac{1}{3} \)
  - \( x = \frac{2}{3} \)
  - \( z = 1 \)

LANGUAGE OF PROBABILITY

Probability can be counterintuitive; we need a formal language!
LANGUAGE OF PROBABILITY

- The sample space is a finite set of elements $S$
- A probability distribution $p$ assigns a non-negative real probability to each element, such that
  $$\sum_{x \in S} p(x) = 1$$

LANGUAGE OF PROBABILITY

- An event is a subset $E \subseteq S$
- $\Pr[E] = \sum_{x \in E} p(x)$
- If each element $x \in S$ has equal probability, the distribution is uniform:
  $$\Pr[E] = \sum_{x \in E} p(x) = \frac{|E|}{|S|}$$

LANGUAGE OF PROBABILITY

- We roll a white die and black die
- $S = \{(1,1), (1,2), (1,3), (1,4), (1,5), (1,6), (2,1), (2,2), (2,3), (2,4), (2,5), (2,6), (3,1), (3,2), (3,3), (3,4), (3,5), (3,6), (4,1), (4,2), (4,3), (4,4), (4,5), (4,6), (5,1), (5,2), (5,3), (5,4), (5,5), (5,6), (6,1), (6,2), (6,3), (6,4), (6,5), (6,6)\}$
- Poll 2: What is the probability that the sum is 7 or 11?
  - $\frac{1}{9}$
  - $\frac{2}{9}$
  - $\frac{3}{9}$
  - $\frac{4}{9}$
**CONDITIONAL PROBABILITY**

- The probability of event $A$ given event $B$ is defined as
  \[ \Pr[A | B] = \frac{\Pr[A \cap B]}{\Pr[B]} \]
- Think of it as the proportion of $A \cap B$ to $B$

---

**PENNIES AND GOLD, REVISITED**

- Three bags contain two gold coins, two pennies, and one of each
- Bag is chosen at random, and one coin from it is selected at random; the coin is gold
- $G_i$: coin in bag $i \in \{1, 2\}$ is gold
- $\Pr[G_1] = \frac{1}{2}$, $\Pr[G_1 \cap G_2] = \frac{1}{3}$
- $\Pr[G_2 | G_1] = \frac{1/3}{1/2} = \frac{2}{3}$

---

**CONDITIONAL PROBABILITY**

- $\Pr[A \cap B] = \Pr[B] \times \Pr[A | B]$
- Interpretation: For $A$ and $B$ to occur, $B$ must occur, and $A$ must occur given that $B$ occurred
- Applying iteratively:
  \[ \Pr[A_1 \cap \cdots \cap A_n] = \Pr[A_1] \times \Pr[A_2 | A_1] \times \cdots \times \Pr[A_n | A_1, \ldots, A_{n-1}] \]

This is called the chain rule
BAYES’ RULE

- \( \Pr[B] \times \Pr[A|B] = \Pr[A \cap B] = \Pr[A] \times \Pr[B|A] \)

Bayes’ rule:

\[
\Pr[A|B] = \frac{\Pr[A] \Pr[B|A]}{\Pr[B]}
\]

MONTY HALL PROBLEM

- Announcer hides prize behind one of three doors at random
- You choose a door
- Announcer opens a door with no prize
- Should you stay with your choice or switch?
**Monty Hall Problem**

- Choose door 1, door 2 opens
- $\Pr[P_3|O_2] = \frac{\Pr[P_2]\Pr[O_2|P_2]}{\Pr[O_2]}$
- $\Pr[P_3] = \frac{1}{3}, \Pr[O_2|P_3] = 1,$
  $\Pr[O_2] = 1/2$
- Therefore, $\Pr[P_3|O_2] = 2/3$
- **Poll 3:** Assuming there are five doors, what is the probability of winning when switching?
  - 1. $\frac{3}{15}$
  - 2. $\frac{4}{15}$
  - 3. $\frac{5}{15}$
  - 4. $\frac{6}{15}$

---

**Independence**

- Events $A$ and $B$ are independent if and only if $\Pr[AB] = \Pr[A]\Pr[B]$.
- **Poll 4:** Which of the following events are independent when rolling black die and white die?
  1. Black die is 1, white die is 1
  2. Sum is 2, sum is 3
  3. Black die is 1, product is 2
  4. Black die is 1, sum is 2

---

**The Birthday Paradox**

- $m$ people in a room; suppose all birthdays are equally likely (excluding Feb 29); what is the probability that two people have the same birthday?
- $S = \{1, ..., 365\}^m$, sample $\vec{x} = (x_1, ..., x_m)$
- $E = \{\vec{x} \in S \mid \exists i, j, s.t. x_i = x_j\}$

Apply the chain rule!
THE BIRTHDAY PARADOX

• $E$ is the event that two people share a birthday
• We will compute $\tilde{E}$
• Let $A_i$ be the event that person $i$'s birthday differs from the birthdays of $1, \ldots, i-1$
• $\tilde{E} = A_i \cap \cdots \cap A_n$
• Using the chain rule:
  \[ \Pr(\tilde{E}) = \Pr(A_i) \times \Pr(A_j | A_i) \times \cdots \times \Pr(A_n | A_1, \ldots, A_{n-1}) \]

So what is $\Pr(A_i | A_1, \ldots, A_{i-1})$?

THE BIRTHDAY PARADOX

• $A_1 \cap \cdots \cap A_{i-1}$ means first $i-1$ students had different birthdays
• $i-1$ out of 365 occupied when $i$th birthday is chosen
• $\Pr(A_i | A_1, \ldots, A_{i-1}) = \frac{365 - (i-1)}{365} = 1 - \frac{i-1}{365}$
• $\Pr(\tilde{E}) = 1 \times \left(1 - \frac{1}{365}\right) \times \cdots \times \left(1 - \frac{m-1}{365}\right)$
• $\Pr(E) = 1 - \Pr(\tilde{E})$

THE BIRTHDAY PARADOX

![Graph showing probability distribution]

15251 Fall 2017: Lecture 21
Carnegie Mellon University
THE BIRTHDAY PARADOX

• Poll 5: What is the probability that two people have the same birthday if there are 730 people?
  1. \( \frac{1}{2} \)
  2. 0.75
  3. 0.99999999999997
  4. 1

BIRTHDAY ATTACK*

• A cryptographic hash function ‘scrambles’ a string \( S \) into a \( k \)-bit hash \( f(S) \)
• It should be hard to find a collision: two strings \( S_1, S_2 \) such that \( f(S_1) = f(S_2) \)
• Application: digital signatures
  ◦ Alice wants Bob to sign a message \( m \)
  ◦ They compute \( f(m) \) and it is signed using Bob’s secret key
  ◦ Bad collision: Alice can find a fair contract \( m \) and a fraudulent contract \( m' \) such that \( f(m) = f(m') \)

BIRTHDAY ATTACK*

• The SHA-1 cryptographic hash function uses 160 bits
• To find a collision for SHA-1, take a huge number of strings, hash them all, and hope that two hash to the same string
• If SHA-1 is really safe, each \( f(S) \) should be uniform in \( \{1, \ldots, 2^{160}\} \)
• This is like the birthday problem with \( 2^{160} \) days of the year!
**Birthday Attack***

• To find a collision you would need roughly \( \sqrt{2^{160}} = 2^{80} \) strings
• A crypto hash function is considered broken if you can beat the birthday attack
• SHA-1 collisions can be found using “only” \( 2^{63} \) strings
• On 2/23/2017, Google and CWI announced that they had generated two different PDF files with the same SHA-1 hash

**Summary**

• Terminology:
  • Language of probability
  • Conditional probability
  • Independence
• Principles:
  • Chain rule
  • Bayes’ rule