The Tragedy of Conditional Probability

Not Everything is Equally Likely

• Say \( n \) balls are placed in \( m \) urns
  • Each ball is equally likely to be placed in any urn
• Counts of balls in urns are not equally likely!
  • Example: two balls (A and B) placed with equal likelihood in two urns (Urn 1 and Urn 2)

<table>
<thead>
<tr>
<th>Urn 1</th>
<th>Urn 2</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>A, B</td>
<td>-</td>
<td>1/4</td>
</tr>
<tr>
<td>A</td>
<td>B</td>
<td>2/4</td>
</tr>
<tr>
<td>B</td>
<td>A</td>
<td>1/4</td>
</tr>
<tr>
<td>-</td>
<td>A, B</td>
<td>1/4</td>
</tr>
</tbody>
</table>

A Few Useful Formulas

• For any events A and B:
  \[
P(A \cap B) = P(B \cap A) \quad \text{(Commutativity)}
\]
  \[
P(A \cap B) = P(A | B) P(B)
  = P(B | A) P(A) \quad \text{(Chain rule)}
\]
  \[
P(A \cap B^c) = P(A) - P(AB) \quad \text{(Intersection)}
\]
  \[
P(A \cap B) \geq P(A) + P(B) - 1 \quad \text{(Bonferroni)}
\]

Generality of Conditional Probability

• For any events A, B, and E, you can condition consistently on E, and these formulas still hold:
  \[
P(A \cap B | E) = P(B | A \cap E) P(A | E)
\]
  \[
P(A | B \cap E) = \frac{P(B | A \cap E) P(A | E)}{P(B | E)} \quad \text{(Bayes Thm.)}
\]
• Can think of E as "everything you already know"
• Formally, \( P(\cdot | E) \) satisfies 3 axioms of probability

Dissecting Bayes Theorem

• Recall Bayes Theorem (common form):
  \[
P(H | E) = \frac{P(E | H) P(H)}{P(E)}
\]
  • "Posterior" = "Likelihood" \times "Prior"
  • Odds(H | E):
  \[
  \frac{P(H | E)}{P(H^c | E)} = \frac{P(E | H) P(H)}{P(E | H^c) P(H^c)}
  \]
  • How odds of H change when evidence E observed
  • Note that \( P(E) \) cancels out in odds formulation
  • This is a form of probabilistic inference

It Always Comes Back to Dice

• Roll two 6-sided dice, yielding values \( D_1 \) and \( D_2 \)
  • Let E be event: \( D_1 = 1 \)
  • Let F be event: \( D_2 = 1 \)
  • What is \( P(E), P(F), \) and \( P(EF) \)?
    \[
    P(E) = 1/6, \quad P(F) = 1/6, \quad P(EF) = 1/36
    \]
  • \( P(EF) = P(E) P(F) \rightarrow E \) and \( F \) \text{ independent}
  • Let G be event: \( D_1 + D_2 = 5 \) \text{ \{ (1, 4), (2, 3), (3, 2), (4, 1) \}}
  • What is \( P(E), P(G), \) and \( P(EG) \)?
    \[
    P(E) = 1/6, \quad P(G) = 4/36 = 1/9, \quad P(EG) = 1/36
    \]
    \[
    P(EG) \neq P(E) P(G) \rightarrow E \text{ and } G \text{ dependent}
    \]
Independence

- Two events E and F are called independent if:
  \[ P(EF) = P(E) P(F) \]
  Or, equivalently: \[ P(E | F) = P(E) \]
- Otherwise, they are called dependent events
- Three events E, F, and G independent if:
  \[ P(EFG) = P(E) P(F) P(G), \]
  \[ P(EF) = P(E) P(F), \]
  \[ P(EG) = P(E) P(G), \]
  \[ P(FG) = P(F) P(G) \]

Let’s Do a Proof

- Given independent events E and F, prove:
  \[ P(E | F) = P(E | F^c) \]
- Proof:
  \[ P(E | F^c) = P(E) - P(EF) \]
  \[ = P(E) - P(E) P(F) \]
  \[ = P(E) [1 - P(F)] \]
  \[ = P(E) P(F^c) \]
  So, E and \( F^c \) independent, implying that:
  \[ P(E | F^c) = P(E) = P(E | F) \]
- Intuitively, if E and F are independent, knowing whether F holds gives us no information about E

Generalized Independence

- General definition of Independence:
  Events \( E_1, E_2, \ldots, E_n \) are independent if for every subset \( E_{i_1}, E_{i_2}, \ldots, E_{i_r} \) (where \( r \leq n \)) it holds that:
  \[ P(E_{i_1} E_{i_2} \ldots E_{i_r}) = P(E_{i_1}) P(E_{i_2}) \ldots P(E_{i_r}) \]
- Example: outcomes of \( n \) separate flips of a coin are all independent of one another
  - Each flip in this case is called a “trial” of the experiment

Two Dice

- Roll two 6-sided dice, yielding values \( D_1 \) and \( D_2 \)
  - Let E be event: \( D_1 = 1 \)
  - Let F be event: \( D_2 = 6 \)
  - Are E and F independent? Yes!
  - Let G be event: \( D_1 + D_2 = 7 \)
  - Are E and G independent? Yes!
  - P(E) = 1/6, P(G) = 1/6, P(E G) = 1/36 [roll (1, 6)]
  - Are F and G independent? Yes!
  - P(F) = 1/6, P(G) = 1/6, P(F G) = 1/36 [roll (1, 6)]
  - Are E, F and G independent? No!
  - P(EFG) = 1/36 ≠ 1/216 = (1/6)(1/6)(1/6)

Generating Random Bits

- A computer produces a series of random bits, with probability \( p \) of producing a 1.
  - Each bit generated is an independent trial
  - E = first \( n \) bits are 1’s, followed by a 0
  - What is \( P(E) \)?
- Solution
  - \[ P(\text{first } n \ 1's) = P(1^{\text{st}} \text{ bit}=1) P(2^{\text{nd}} \text{ bit}=1) \ldots P(n^{\text{th}} \text{ bit}=1) = p^n \]
  - \[ P(n+1 \text{ bit}=0) = (1-p) \]
  - \[ P(E) = P(\text{first } n \ 1's) P(n+1 \text{ bit}=0) = p^n (1-p) \]

Coin Flips

- Say a coin comes up heads with probability \( p \)
  - Each coin flip is an independent trial
- P(\( n \) heads on \( n \) coin flips) = \( p^n \)
- P(\( n \) tails on \( n \) coin flips) = \( (1-p)^n \)
- P(first \( k \) heads, then \( n-k \) tails) \( =p^k (1-p)^{n-k} \)
- P(exactly \( k \) heads on \( n \) coin flips) \( = \binom{n}{k} p^k (1-p)^{n-k} \)
Hash Tables
- $m$ strings are hashed (equally randomly) into a hash table with $n$ buckets
  - Each string hashed is an independent trial
  - $E = \text{at least one string hashed to first bucket}$
  - What is $P(E)$?
- Solution
  - $F_i = \text{string } i \text{ not hashed into first bucket}$ (where $1 \leq i \leq m$)
  - $P(F_i) = 1 - 1/n = (n - 1)/n \text{ (for all } 1 \leq i \leq m)$
  - Event $(F_1,F_2,...,F_m) = \text{no strings hashed to first bucket}$
  - $P(E) = 1 - P(F_1,F_2,...,F_m) = 1 - P(F_1)P(F_2)...P(F_m)$
  
  $$= 1 - (n - 1/n)^m$$
  - Similar to $\geq 1$ of $m$ people having same birthday as you

Yet More Hash Table Fun
- $m$ strings are hashed (unequally) into a hash table with $n$ buckets
  - Each string hashed is an independent trial, with probability $p_i$ of getting hashed to bucket $i$
  - $E = \text{At least one bucket } i \text{ has } \geq 1 \text{ string hashed to it}$
- Solution
  - $F_i = \text{at least one string hashed into } i\text{-th bucket}$
  - $P(E) = P(F_1 \cup F_2 \cup ... \cup F_n) = 1 - P(F_1 \cup F_2 \cup ... \cup F_n)^c$ (DeMorgan’s Law)
  
  $$= 1 - P(F_1^c F_2^c ... F_n^c)$$
  - $P(F_1^c F_2^c ... F_n^c) = P(\text{no strings hashed to buckets 1 to } k)$
    
    $$= (1 - p_1 - p_2 - ... - p_k)^m$$
  - $P(E) = 1 - (1 - p_1 - p_2 - ... - p_n)^m$

No, Really, it’s More Hash Table Fun
- $m$ strings are hashed (unequally) into a hash table with $n$ buckets
  - Each string hashed is an independent trial, with probability $p_i$ of functioning (where $1 \leq i \leq n$)
  - $E = \text{Each of buckets 1 to } k \text{ has } \geq 1 \text{ string hashed to it}$
- Solution
  - $F_i = \text{at least one string hashed into } i\text{-th bucket}$
  - $P(E) = P(F_1 \cup F_2 \cup ... \cup F_n) = 1 - P(F_1 \cup F_2 \cup ... \cup F_n)^c$ (DeMorgan’s Law)
  
  $$= 1 - P(F_1^c F_2^c ... F_n^c)$$
  - $P(F_1^c F_2^c ... F_n^c) = P(\text{all routers fail})$
    
    $$= (1 - p_1)(1 - p_2)...(1 - p_n)$$
  - $P(E) = 1 - (1 - p_1 - p_2 - ... - p_n)^m$

Reminder of Geometric Series
- Geometric series: $x^0 + x^1 + x^2 + x^3 + ... + x^n = \sum_{i=0}^{n} x^i$
- From your “Calculation Reference” handout:
  
  $$\sum_{i=0}^{n} x^i = \frac{1-x^{n+1}}{1-x}$$
  - As $n \to \infty$, and $|x| < 1$, then
  
  $$\sum_{i=0}^{n} x^i = \frac{1-x^{n+1}}{1-x} \to \frac{1}{1-x}$$

Sending Messages Through a Network
- Consider the following parallel network:
- $n$ independent routers, each with probability $p_i$ of functioning (where $1 \leq i \leq n$)
  - $E = \text{functional path from A to B exists. What is } P(E)$?
- Solution:
  - $P(E) = 1 - P(\text{all routers fail})$
    
    $$= 1 - (1-p_1)(1-p_2)...(1-p_n)$$
  - $P(E) = 1 - \prod_{i=1}^{n} (1 - p_i)$

Simplified Craps
- Two 6-sided dice repeatedly rolled (roll = ind. trial)
  - $E = 5 \text{ is rolled before a 7 is rolled}$
  - What is $P(E)$?
- Solution
  - $F_0 = \text{no 5 or 7 rolled in first } n - 1 \text{ trials}, 5 \text{ rolled on } n\text{th trial}$
  - $P(E) = P(F_0) = \sum_{i=0}^{n-1} P(F_i)$
    
    $$= P(5 \text{ on any trial}) + (4/36) P(7 \text{ on any trial}) = 6/36$$
    
    $$= (1 - (10/36)^{n-1}) (4/36) + (26/36)^{n-1} (4/36)$$
  - $P(E) = \frac{4}{36} \sum_{i=0}^{n-1} \left(\frac{26}{36}\right)^i = \frac{4}{36} \sum_{i=0}^{n-1} \left(\frac{26}{36}\right)^{i} = \frac{4}{36} \left(1 - \frac{26}{36}\right)^2 = \frac{2}{3}$
DNA Paternity Testing

- Child is born with (A, a) gene pair (event \( B_{A,a} \))
  - Mother has (A, A) gene pair
  - Two possible fathers: \( M_1: (a, a) \) \quad \( M_2: (a, A) \)
  - \( P(M_1) = p \quad P(M_2) = 1 - p \)
  - What is \( P(M_1 | B_{A,a}) \)?

- Solution
  \[
  P(M_1 | B_{A,a}) = \frac{P(M_1 B_{A,a})}{P(B_{A,a})}
  = \frac{P(M_1)P(B_{A,a} | M_1)}{P(B_{A,a})}
  = \frac{1 \cdot p}{1 - p + \frac{1}{2} (1 - p)} = \frac{2p}{1 + p} > p
  \]
  \( M_1 \) more likely to be father than he was before, since
  \( P(M_1 | B_{A,a}) > P(M_1) \)