The Tragedy of Conditional Probability

Thanks xkcd!  http://xkcd.com/795/
A Few Useful Formulas

- For any events $A$ and $B$:
  
  $P(A \cap B) = P(B \cap A)$ \hspace{1cm} (Commutativity)
  
  $P(A \cap B) = P(A | B) P(B)$ \hspace{1cm} (Chain rule)
  
  $= P(B | A) P(A)$
  
  $P(A \cap B^c) = P(A) - P(AB)$ \hspace{1cm} (Intersection)
  
  $P(A \cup B) \geq P(A) + P(B) - 1$ \hspace{1cm} (Bonferroni)
Generality of Conditional Probability

• For any events A, B, and E, you can condition consistently on E, and these formulas still hold:

\[ P(A \cap B \mid E) = P(B \cap A \mid E) \]

\[ P(A \cap B \mid E) = P(A \mid B \cap E) \cdot P(B \mid E) \]

\[ P(A \mid B \cap E) = \frac{P(B \mid A \cap E) \cdot P(A \mid E)}{P(B \mid E)} \quad (\text{Bayes’ Thm.}) \]

• Can think of E as “everything you already know”
• Formally, \( P(\bullet \mid E) \) satisfies 3 axioms of probability
Dissecting Bayes’ Theorem

- Recall Bayes’ Theorem (common form):

\[
P(H \mid E) = \frac{P(E \mid H) P(H)}{P(E)}
\]

- Prior: Probability of H *before* you observe E
- Likelihood: Probability of E given that H holds
- Posterior: Probability of H *after* you observe E
Odds

- Odds of an event defined as:
  \[
  \frac{P(A)}{P(A^c)} = \frac{P(A)}{1 - P(A)}
  \]

- Odds of H given observed evidence E:
  \[
  \frac{P(H | E)}{P(H^c | E)} = \frac{P(H) P(E | H) / P(E)}{P(H^c) P(E | H^c) / P(E)}
  = \frac{P(H) P(E | H)}{P(H^c) P(E | H^c)}
  \]

- After observing E, just update odds by:
  \[
  \frac{P(E | H)}{P(E | H^c)}
  \]
Coins and Urns?!

- An urn contains 2 coins: A and B
  - A comes up heads with probability $\frac{1}{4}$
  - B comes up heads with probability $\frac{3}{4}$
  - Pick coin (equally likely), flip it, and it comes up heads
  - What are odds that A was picked (note: $A^c = B$)?
    
    $$
    \frac{P(A \mid \text{heads})}{P(A^c \mid \text{heads})} = \frac{P(A) \cdot P(\text{heads} \mid A)}{P(A^c) \cdot P(\text{heads} \mid A^c)}
    $$
    
    $$
    = \frac{\frac{1}{2} \cdot \frac{1}{4}}{\frac{1}{2} \cdot \frac{3}{4}} = \frac{1}{3}
    $$

  - Odds are 1/3:1 (or probability $\frac{1}{4}$) that A was picked
  - Note: before observing heads $P(A) / P(A^c) = 1:1$
    - Equally likely to pick A vs. not picking A (1 out of 2 chance)
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 $P(EF)$?
  ▪ $P(E) = 1/6$, $P(F) = 1/6$, $P(EF) = 1/36$
  ▪ $P(EF) = P(E) \cdot P(F) \rightarrow E \text{ and } F \text{ independent}$

• Let $G$ be event: $D_1 + D_2 = 5$ \{(1, 4), (2, 3), (3, 2), (4, 1)\}

• What is $P(E)$, $P(G)$, and $P(EG)$?
  ▪ $P(E) = 1/6$, $P(G) = 4/36 = 1/9$, $P(EG) = 1/36$
  ▪ $P(EG) \neq P(E) \cdot P(G) \rightarrow E \text{ and } G \text{ dependent}$
Independence

• Two events E and F are called **independent** if:
  \[ P(\text{EF}) = P(E) \times P(F) \]
  Or, equivalently: \( P(E \mid F) = P(E) \)
• Otherwise, they are called **dependent** events

• Three events E, F, and G independent if:
  \[ P(\text{EFG}) = P(E) \times P(F) \times P(G), \text{ and} \]
  \[ P(\text{EF}) = P(E) \times P(F), \text{ and} \]
  \[ P(\text{EG}) = P(E) \times P(G), \text{ and} \]
  \[ P(\text{FG}) = P(F) \times P(G) \]
Let’s Do a Proof

• Given independent events E and F, prove:
  \[ P(E \mid F) = P(E \mid F^c) \]

• 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:

\[ P(E \mid F^c) = P(E) = P(E \mid F) \]

• 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 $E_1', E_2', \ldots, E_r'$ (where $r \leq n$) it holds that:

  $$P(E_1'E_2'E_3'\ldots'E_r') = P(E_1')P(E_2')P(E_3')\ldots P(E_r')$$

- 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!
  - $P(E) = 1/6$, $P(G) = 1/6$, $P(E \cap G) = 1/36$ [roll (1, 6)]
  - Are $F$ and $G$ independent? Yes!
  - $P(F) = 1/6$, $P(G) = 1/6$, $P(F \cap G) = 1/36$ [roll (1, 6)]
  - Are $E$, $F$ and $G$ independent? No!
  - $P(E \cup F \cup G) = 1/36 \neq 1/216 = (1/6)(1/6)(1/6)$
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 single 0
  - What is $P(E)$?
- Solution
  - $P(\text{first } n \text{ 1’s}) = P(1^{\text{st}} \text{ bit}=1) \cdot P(2^{\text{nd}} \text{ bit}=1) \ldots \cdot P(n^{\text{th}} \text{ bit}=1)$
    - $= p^n$
  - $P(n+1 \text{ bit}=0) = (1 - p)$
  - $P(E) = P(\text{first } n \text{ 1’s}) \cdot P(n+1 \text{ bit}=0) = p^n \cdot (1 - p)$
Coin Flips

- Say a coin comes up heads with probability $p$
  - Each coin flip is an independent trial

- $P(n \text{ heads on } n \text{ coin flips}) = p^n$
- $P(n \text{ tails on } n \text{ coin flips}) = (1 - p)^n$

- $P(\text{first } k \text{ heads, then } n - k \text{ tails}) = p^k (1 - p)^{n-k}$

- $P(\text{exactly } k \text{ heads on } n \text{ coin flips}) = \binom{n}{k} p^k (1 - p)^{n-k}$
Hash Tables

- \( m \) strings are hashed (equally randomly) into a hash table with \( n \) buckets
  - Each string hashed is an independent trial
  - \( E = \) at least one string hashed to first bucket
  - What is \( P(E) \)?

Solution

- \( F_i = \) string \( i \) not hashed into first bucket (where \( 1 \leq i \leq m \))
- \( P(F_i) = 1 – 1/n = (n – 1)/n \) (for all \( 1 \leq i \leq m \))
- Event \( (F_1 F_2 \ldots F_m) = \) no strings hashed to first bucket
- \( P(E) = 1 – P(F_1 F_2 \ldots F_m) = 1 – P(F_1)P(F_2)\ldots P(F_m) = 1 – ((n – 1)/n)^m \)
- 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*ₐ of getting hashed to bucket *i*
  - E = **At least 1 of** buckets 1 to *k* has ≥ 1 string hashed to it

- Solution
  - *F*ᵢ = at least one string hashed into *i*-th bucket
  - \[ P(E) = P(F_1 \cup F_2 \cup \ldots \cup F_k) = 1 - P((F_1 \cup F_2 \cup \ldots \cup F_k)^c) \]
  - \[ = 1 - P(F_1^c F_2^c \ldots F_k^c) \] (DeMorgan’s Law)
  - \[ P(F_1^c F_2^c \ldots F_k^c) = P(\text{no strings hashed to buckets 1 to } k) \]
  - \[ = (1 - p_1 - p_2 - \ldots - p_k)^m \]
  - \[ P(E) = 1 - (1 - p_1 - p_2 - \ldots - p_k)^m \]
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 getting hashed to bucket $i$
  - $E = \text{Each of}$ buckets 1 to $k$ has $\geq 1$ string hashed to it

- Solution
  - $F_i = \text{at least one string hashed into } i\text{-th bucket}$
  - $P(E) = P(F_1 F_2 \ldots F_k) = 1 - P((F_1 F_2 \ldots F_k)^c)$
    
    $= 1 - P(F_1^c \cup F_2^c \cup \ldots \cup F_k^c)$ \hspace{1cm} (DeMorgan’s Law)

    $= 1 - \sum_{i=1}^{k} F_i^c \cup \sum_{r=1}^{k} (-1)^{r+1} \sum_{i_1<i_2<\ldots<i_r} P(F_{i_1}^c F_{i_2}^c \ldots F_{i_r}^c)$

    where $P(F_{i_1}^c F_{i_2}^c \ldots F_{i_r}^c) = (1 - p_{i_1} - p_{i_2} - \ldots - p_{i_r})^m$
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 = $ functional path from A to B exists. What is $P(E)$?

• Solution:
  - $P(E) = 1 - P($all routers fail$)$
  - $= 1 - (1 - p_1)(1 - p_2)\ldots(1 - p_n)$
  - $= 1 - \prod_{i=1}^{n}(1 - p_i)$