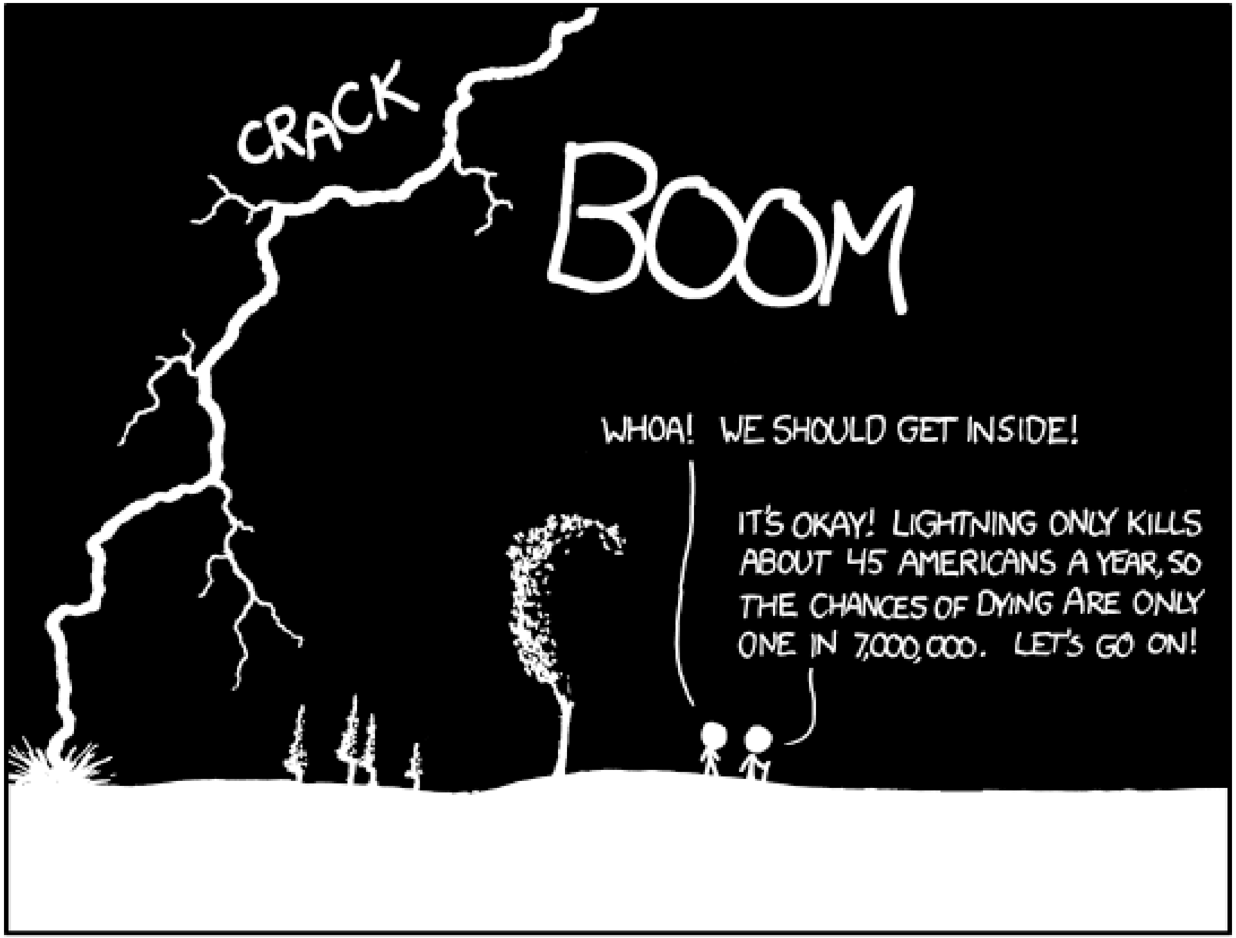
The Tragedy of Conditional Probability



THE ANNUAL DEATH RATE AMONG PEOPLE WHO KNOW THAT *STATISTIC* IS ONE IN SIX.

Thanks xkcd! **`http://xkcd.com/795/`**

A Few Useful Formulas

- For any events A and B:

$P(A\ B) \quad = P(B\ A)$ (Commutativity)

$P(A\ B) \quad = P(A \mid B)\ P(B)$ (Chain rule)

$\qquad\qquad = P(B \mid A)\ P(A)$

$P(A\ B^c) \quad = P(A) - P(AB)$ (Intersection)

$P(A\ B) \quad \geq P(A) + P(B) - 1$ (Bonferroni)

Generality of Conditional Probability

- For any events A, B, and E, you can condition <u>consistently</u> on E, and these formulas still hold:

$$P(A\,B \mid E) = P(B\,A \mid E)$$

$$P(A\,B \mid E) = P(A \mid B\,E)\,P(B \mid E)$$

$$P(A \mid B\,E) = \frac{P(B \mid A\,E)\,P(A \mid E)}{P(B \mid E)} \qquad \text{(Bayes' Thm.)}$$

- 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):

$$\underbrace{P(H \mid E)}_{\text{“Posterior”}} = \frac{\underbrace{P(E \mid H)}_{\text{“Likelihood”}}\ \underbrace{P(H)}_{\text{“Prior”}}}{P(E)}$$

- 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:

$$\frac{P(H \mid E)}{P(H^c \mid E)} = \frac{P(H)\, P(E \mid H) \,/\, P(E)}{P(H^c)\, P(E \mid H^c) \,/\, P(E)}$$

$$= \frac{P(H)\, P(E \mid H)}{P(H^c)\, P(E \mid H^c)}$$

- After observing E, just update odds by: $\dfrac{P(E \mid H)}{P(E \mid H^c)}$

Coins and Urns?!

- An urn contains 2 coins: A and B
  - A comes up heads with probability ¼
  - B comes up heads with probability ¾
  - Pick coin (equally likely), flip it, and it comes up heads
  - What are odds that A was picked (note: $A^c = B$)?

$$\frac{P(A \mid \text{heads})}{P(A^c \mid \text{heads})} = \frac{P(A)\ \ P(\text{heads} \mid A)}{P(A^c)\ \ P(\text{heads} \mid A^c)}$$

$$= \frac{\frac{1}{2}\ \ \frac{1}{4}}{\frac{1}{2}\ \ \frac{3}{4}} = 1/3$$

  - Odds are 1/3:1 (or probability ¼) 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)?
  - $P(E) = 1/6$, $P(F) = 1/6$, $P(EF) = 1/36$
  - $P(EF) = P(E)\,P(F)$ $\rightarrow$ E and F ***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)?
  - $P(E) = 1/6$, $P(G) = 4/36 = 1/9$, $P(EG) = 1/36$
  - $P(EG) \neq P(E)\,P(G)$ $\rightarrow$ E and G ***dependent***

Independence

- Two events E and F are called <u>**independent**</u> if:

  $$P(EF) = P(E)\,P(F)$$

  Or, equivalently: $P(E \mid F) = P(E)$
- Otherwise, they are called <u>**dependent**</u> events

- Three events E, F, and G independent if:

  $P(EFG) = P(E)\,P(F)\,P(G)$, and

  $P(EF) = P(E)\,P(F)$, and

  $P(EG) = P(E)\,P(G)$, and

  $P(FG) = P(F)\,P(G)$

Let's Do a Proof

- Given independent events E and F, prove:

  $P(E \mid F) = P(E \mid F^c)$

- Proof:

$P(E F^c)$	$= P(E) - P(EF)$	Intersection
	$= P(E) - P(E)\, P(F)$	Independence
	$= P(E)\, [1 - P(F)]$	Factoring
	$= P(E)\, P(F^c)$	Complement

  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 $E_1, E_2, ..., E_n$ are independent if for every subset $E_{1'}, E_{2'}, ..., E_{r'}$ (where $r \leq n$) it holds that:

$$P(E_{1'}E_{2'}E_{3'}...E_{r'}) = P(E_{1'})P(E_{2'})P(E_{3'})...P(E_{r'})$$

- 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 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!
  - $P(EFG) = 1/36 \neq 1/216 = (1/6)(1/6)(1/6)$

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(first *n* 1's) $= P(1^{st}\text{ bit}=1)\, P(2^{nd}\text{ bit}=1) \ldots P(n^{th}\text{ bit}=1)$
    $= p^n$
  - $P(n+1 \text{ bit}=0) = (1 - p)$
  - $P(E) = P(\text{first } n \text{ 1's})\, P(n+1 \text{ bit}=0) = p^n (1 - p)$

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 = at least one string hashed to first bucket
  - What is P(E)?
- Solution
  - F_i = string i <u>not</u> 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 ≥ 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 = 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(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 = Each of buckets 1 to *k* has ≥ 1 string hashed to it
- Solution
  - F_i = at least one string hashed into *i*-th bucket
  - $P(E) = P(F_1 F_2 \ldots F_k) = 1 - P((F_1 F_2 \ldots F_k)^c)$

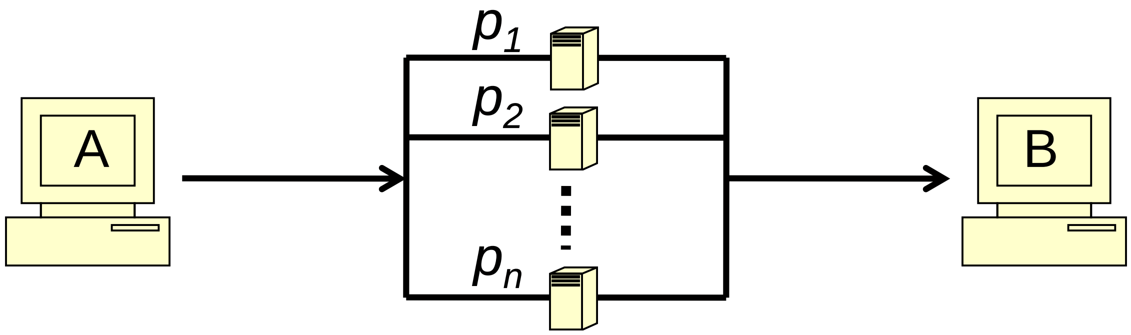
    $= 1 - P(F_1^c \cup F_2^c \cup \ldots \cup F_k^c)$ (DeMorgan's Law)

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

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 $1 \le i \le n$)
  - E = functional path from A to B exists. What is P(E)?
- Solution:
  - P(E) = 1 – P(all routers fail)

$$= 1 - (1 - p_1)(1 - p_2)\ldots(1 - p_n)$$

$$= 1 - \prod_{i=1}^{n} (1 - p_i)$$