A Few Useful Formulas

- For any events A and B:

$P(A\,B)$	$= P(B\,A)$	(Commutativity)
$P(A\,B)$	$= P(A \mid B)\,P(B)$ $= P(B \mid A)\,P(A)$	(Chain rule)
$P(A\,B^c)$	$= P(A) - P(AB)$	(Intersection)
$P(A\,B)$	$\geq P(A) + P(B) - 1$	(Bonferroni)

Generality of Conditional Probability

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

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

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

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

- 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):

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

- Odds(H | E):

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

- 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)?
  - $P(E) = 1/6$, $P(F) = 1/6$, $P(EF) = 1/36$
  - $P(EF) = P(E)\,P(F)$ → 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)?
  - $P(E) = 1/6$, $P(G) = 4/36 = 1/9$, $P(EG) = 1/36$
  - $P(EG) \neq P(E)\,P(G)$ → E and G *dependent*

Independence

- Two events E and F are called **independent** if:

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

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

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

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

  $P(E\,F^c) = P(E) - P(EF)$

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

  $= P(E)\,[1 - P(F)]$

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

  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 $r \le n$) 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 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 0
  - What is P(E)?
- Solution
  - P(first n 1's) = P(1^{st} bit=1) P(2^{nd} bit=1) ... P(n^{th} bit=1) = p^n
  - P($n+1$ bit=0) = (1 – p)
  - P(E) = P(first n 1's) P($n+1$ 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 not hashed into first bucket (where $1 \le i \le m$)
  - P(F_i) = 1 – 1/n = (n – 1)/n (for all $1 \le i \le m$)
  - Event ($F_1F_2\ldots F_m$) = no strings hashed to first bucket
  - P(E) = 1 – P($F_1F_2\ldots F_m$) = 1 – P(F_1)P(F_2)…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) $= P(F_1 F_2 \dots F_k) = 1 - P((F_1 F_2 \dots F_k)^c)$

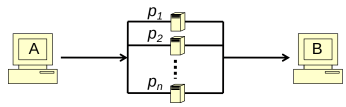
    $= 1 - P(F_1^c \cup F_2^c \cup \dots \cup F_k^c)$ (DeMorgan's Law)

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

  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)

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

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

Reminder of Geometric Series

- Geometric series: $x^0 + x^1 + x^2 + x^3 + \ldots + x^n = \sum_{i=0}^{n} x^i$
- 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

$$\sum_{i=0}^{n} x^i = \frac{1 - x^{n+1}}{1 - x} \to \frac{1}{1 - x}$$

Simplified Craps

- Two 6-sided dice repeatedly rolled (roll = ind. trial)
  - E = 5 is rolled before a 7 is rolled
  - What is P(E)?
- Solution
  - F_n = no 5 or 7 rolled in first n – 1 trials, 5 rolled on n^{th} trial
  - P(E) $= P\left(\bigcup_{n=1}^{\infty} F_n\right) = \sum_{n=1}^{\infty} P(F_n)$
  - P(5 on any trial) = 4/36 P(7 on any trial) = 6/36
  - $P(F_n) = (1 - (10/36))^{n-1} (4/36) = (26/36)^{n-1} (4/36)$
  - P(E) $= \frac{4}{36} \sum_{n=1}^{\infty} \left(\frac{26}{36}\right)^{n-1} = \frac{4}{36} \sum_{n=0}^{\infty} \left(\frac{26}{36}\right)^{n} = \frac{4}{36} \frac{1}{\left(1 - \frac{26}{36}\right)} = \frac{2}{5}$

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) M_2: (a, A)
  - $P(M_1) = p$ $P(M_2) = 1 - p$
  - What is $P(M_1 \mid B_{A,a})$?
- Solution
  - $P(M_1 \mid B_{A,a}) = P(M_1 B_{A,a}) / P(B_{A,a})$

$$= \frac{P(B_{A,a} \mid M_1) P(M_1)}{P(B_{A,a} \mid M_1) P(M_1) + P(B_{A,a} \mid M_2) P(M_2)}$$

$$= \frac{1 \cdot p}{1 \cdot p + \frac{1}{2}(1 - p)} = \frac{2p}{1 + p} > p$$

M_1 more likely to be father than he was before, since $P(M_1 \mid B_{A,a}) > P(M_1)$