1. Consider the Monty Hall problem, except that Monty enjoys opening Door 2 more than he enjoys opening Door 3, and if he has a choice between opening these two doors, he opens Door 2 with probability p, where $\frac{1}{2} \le p \le 1$.

   To recap: there are three doors, behind one of which there is a car (which you want), and behind the other two of which there are goats (which you don't want). Initially, all possibilities are equally likely for where the car is. You choose a door, which for concreteness we assume is Door 1. Monty Hall then opens a door to reveal a goat, and offers you the option of switching. Assume that Monty Hall knows which door has the car, will always open a goat door and offer the option of switching, and as above assume that if Monty Hall has a choice between opening Door 2 and Door 3, he chooses Door 2 with probability p (with $\frac{1}{2} \le p \le 1$).

   (a) Find the unconditional probability that the strategy of always switching succeeds (unconditional in the sense that we do not condition on which of Doors 2,3 Monty opens).

   (b) Find the probability that the strategy of always switching succeeds, given that Monty opens Door 2.

   (c) Find the probability that the strategy of always switching succeeds, given that Monty opens Door 3.

$P(\text{open door } 2 \mid \text{car in } 1 \text{ and player picks } 1) = p$

Let event where player picks car b

Let event that switches door and win be S.

Let event that car is in door k be C_k.

$P(S) = P(S \mid \text{player picks car initially}) \cdot P(\text{player picks car initially})$
$+ P(S \mid \text{player picks goat initially}) \cdot P(\text{player picks goat initially})$

$= P(S \mid \text{player picks car initially}) \cdot \frac{1}{3} + \frac{2}{3}$

$\tilde{P}(S) = P(S \mid \text{player picks car initially})$

$\tilde{P}(S) = \tilde{P}(S \mid C_1) \cdot \tilde{P}(C_1) + \tilde{P}(S \mid C_1^c) \cdot \tilde{P}(C_1^c)$
$= P(S \mid C_1 \cap \text{pick car}) \cdot \tilde{P}(C_1) + \tilde{P}(S \mid C_1^c \cap \text{pick car}) \cdot (\ldots)$
$= 0.$

$\Rightarrow P(S) = \frac{2}{3}.$

O_2 : open door 2

I_2 : initial pick is 2

b)

car in	your choice	open	open 2	3	win if switch	probability
1	1	2 (p)	✓		✗	$p/9$
1	1	3 ($1-p$)	✗	✓	✗	$1-p/9$
1	2	3	✗	✓	✓	$1/9$
1	3	2	✓		✓	$1/9$
2	1	3	✗	✓	✓	$1/9$
2	2	1	✗		✗	$1/18$
2	2	3	✗	✓	✗	$1/18$
2	3	1	✗		✓	$1/9$
3	1	2	✓		✓	$1/9$
3	2	1	✗		✓	$1/9$
3	3	1	✗		✗	$1/18$
3	3	2	✓		✗	$1/18$

D_- (first probability branch labelled $\frac{1}{3}$)

b) $\dfrac{1/9 + 1/9}{\left(\frac{p}{9} + \frac{1}{9} + \frac{1}{9} + \frac{1}{18}\right)} = \dfrac{2/9}{\frac{2p+2+2+1}{18}} = \dfrac{2/9}{\frac{2p+5}{18}} = \frac{2}{9} \cdot \frac{18}{2p+5} = \boxed{\frac{4}{2p+5}}$

c) $\dfrac{\frac{1}{9} + \frac{1}{9}}{\left(\frac{1-p}{9} + \frac{1}{9} + \frac{1}{9} + \frac{1}{18}\right)} = \dfrac{2/9}{\frac{2-2p+2+2+1}{18}} = \frac{2}{9} \cdot \frac{18}{7-2p} = \boxed{\frac{4}{7-2p}}$

$\frac{7-2p}{18}$

2. For each statement below, either show that it is true or give a counterexample. Throughout, X, Y, Z are discrete random variables.

(a) If X and Y are independent and Y and Z are independent, then X and Z are independent.

(b) If X and Y are independent, then they are conditionally independent given Z.

(c) If X and Y are conditionally independent given Z, then they are independent.

(d) If X and Y have the same distribution given Z, i.e., for all a and z, we have $P(X = a|Z = z) = P(Y = a|Z = z)$, then X and Y have the same distribution.

a) False. If $X = Z$, then X and Y are independent, Y and Z are independent, but X and Z are dependent.

b) For X and Y to be independent,

$$Pr(X=x) = Pr(X=x \mid Y=y)$$

$$Pr(X=x, Y=y) = Pr(X=x)Pr(Y=y)$$

$$Pr(X=x, Y=y \mid Z=z) \neq Pr(X=x \mid Z=z)Pr(Y=y \mid Z=z)$$

False. Suppose $Z = X - Y$.
Then if $Z=0$, $X=Y \Rightarrow x=y$

Hence $Pr(X=x, Y=y \mid Z=0) = 0$ i.f.f $x \neq y$ and
$Pr(X=x, Y=y \mid Z=0) = 1$ i.f.f $x=y$

However
$$\begin{aligned} Pr(X=x \mid Z=0) &= Pr(X=x \mid Z=0, Y=x)Pr(Y=x \mid Z=0) \\ &\quad + Pr(X=x \mid Z=0, Y\neq x)Pr(Y\neq x \mid Z=0) \\ &= Pr(Y=x \mid Z=0) \end{aligned}$$

Similarly
$$\begin{aligned} Pr(Y=y \mid Z=0) &= Pr(Y=y \mid Z=0, X=y)Pr(X=y \mid Z=0) \\ &\quad + Pr(Y=y \mid Z=0, X\neq y)Pr(X\neq y) \mid Z=0) \\ &= Pr(X=y \mid Z=0) \end{aligned}$$

Thus $Pr(X=x \mid Z=0)Pr(Y=y \mid Z=0) = Pr(Y=x \mid Z=0)Pr(X=y \mid Z=0)$

If $x=y$, $Pr(X=x \mid Z=0)Pr(Y=y \mid Z=0) = Pr(Y=x \mid Z=0)Pr(X=x \mid Z=0) \neq 1$

If $x \neq y$, $Pr(X=x \mid Z=0)Pr(Y=y \mid Z=0) = Pr(Y=x \mid Z=0)Pr(X=y \mid Z=0) \neq 0$

Thus $Pr(X=x, Y=y \mid Z=0) \neq Pr(X=x \mid Z=0)Pr(Y=y \mid Z=0)$
and not conditional given Z. $\square$.

c) If $P(X=x \cap Y=y \mid Z=z) = P(X=x \mid Z=z)\, P(Y=y \mid Z=z)$
then $P(X=x \cap Y=y) = P(X=x)\, P(Y=y)$
$\Rightarrow P(Y=y \mid X=x) = P(Y=y)$

False.

Let Z be the skill level of a chess player.
Let X be the indicator if player wins first match against player.
Let Y be the indicator if player wins second match against player.

Suppose $X=1$, then player wins first match. That provides information on the skill level of the opponent hence $P(Y=1 \mid X=1) \neq P(Y=1)$.

But suppose we know the skill level of the player, then X and Y will then be independent.

d) $P(X=a \mid Z=z) = P(Y=a \mid Z=z)\ \forall a, z.$

Suppose $\{z_1, \ldots, z_n\} = Z$.

By LOTP

$P(X=a) = P(X=a \mid Z=z_1) P(Z=z_1) + \ldots + P(X=a \mid Z=z_n) P(Z=z_n)$ for all a
$P(Y=a) = P(Y=a \mid Z=z_1) P(Z=z_1) + \ldots + P(Y=a \mid Z=z_n) P(Z=z_n)$ for all a

Since $P(X=a \mid Z=z_k) = P(Y=a \mid Z=z_k)$ for $k=1, \ldots, n$, then
$P(X=a) = P(Y=a)$ for all a
$\Rightarrow X$ and Y have identical distributions

$\square$.

2 Simpson's Paradox

1. (a) Is it possible to have events A, B, E such that $P(A|E) < P(B|E)$ and $P(A|E^c) < P(B|E^c)$, yet $P(A) > P(B)$? That is, A is less likely under B given that E is true, and also given that E is false, yet A is more likely than B if given no information about E. Show this is impossible (with a short proof) or find a counterexample (with a "story" interpreting A, B, E).

   (b) Is it possible to have events A, B, E such that $P(A|B, E) < P(A|B^c, E)$ and $P(A|B, E^c) < P(A|B^c, E^c)$, yet $P(A|B) > P(A|B^c)$? That is, given that E is true, learning B is evidence against A, and similarly given that E is false; but given no information about E, learning that B is true is evidence in favor of A. Show this is impossible (with a short proof) or find a counterexample (with a "story" interpreting A, B, E).

1) $P(A|E) < P(B|E)$ and $P(A|E^c) < P(B|E^c)$ but $P(A) > P(B)$?

$$P(A|E^c)P(E^c) + P(A|E)P(E) = P(A)$$
$$P(B|E^c)P(E^c) + P(B|E)P(E) = P(B)$$

By contradiction,
Suppose $P(A) > P(B)$

$$P(A|E^c)P(E^c) + P(A|E)P(E) > P(B|E^c)P(E^c) + P(B|E)P(E)$$
$$[P(A|E^c) - P(B|E^c)]P(E^c) + [P(A|E) - P(B|E)]P(E) > 0$$

Since $1 \geq P(E) \geq 0$ by 1st axiom,
and $P(A|E^c) - P(B|E^c) < 0$
$P(A|E) - P(B|E) < 0$ by definition,
then $P(A) - P(B) < 0$ which is a contradiction.

Thus, $P(A) \leq P(B)$

$\square$

b) Lemma 1: If $P(A|B) > P(A|B^c)$, then $P(A|B) > P(A)$

Pf: $P(A|B) > P(A|B^c)$

$$P(A|B^c) = \frac{P(B^c|A)P(A)}{P(B^c)}$$
$$= \frac{(1 - P(B|A))P(A)}{1 - P(B)}$$
$$\Rightarrow P(A|B)(1 - P(B)) > P(A) - P(B|A)P(A)$$
$$P(A|B)(1 - P(B)) > P(A) - \frac{P(A|B)P(B)}{P(A)}P(A)$$
$$P(A|B) - \cancel{P(B)P(A|B)} > P(A) - \cancel{P(A|B)P(B)}$$
$$P(A|B) > P(A)$$

$\square$

Corollary: If $P(A|B,C) < P(A|B^c, C)$, then $P(A|B,C) < P(A|C)$

Pf: $P(A|B,E) < P(A|B^c, E)$
Let $\tilde{P}(A) = \tilde{P}(A|E)$

$$\tilde{P}(A|B) < \tilde{P}(A|B^c)$$
$$\Rightarrow \tilde{P}(A|B)(1 - \tilde{P}(B)) < \tilde{P}(A) - \tilde{P}(B|A)P(A)$$

$\vdots$

$$\tilde{P}(A|B) < \tilde{P}(A)$$
$$\Leftrightarrow P(A|B,E) < P(A|E)$$

$\square$

By contradiction,

suppose $P(A|B) > P(A|B^c)$, and $P(A|B,E) < P(A|B^c,E)$

$P(A|B,E^c) < P(A|B^c,E^c)$

By lemma 1,

$$P(A|B) > P(A)$$

By corollary,

$$P(A|B,E) < P(A|E)$$
$$P(A|B,E^c) < P(A|E^c)$$

Prove that $P(A|B) > P(A) \Rightarrow P(A|B) - P(A) > 0$

$$P(A) = P(A|E)P(E) + P(A|E^c)P(E^c)$$
$$> P(A|B,E)P(E) + P(A|B,E^c)P(E^c)$$

$$P(A|B) = P(A|B,E)P(E|B) + P(A|B,E^c)P(E^c|B)$$

$$\begin{aligned}
P(A|B) - P(A) &= P(A|B,E)[P(E|B) - P(E)] + P(A|B,E^c)[P(E^c|B) - P(E^c)] \\
&= P(A|B,E)\left[\frac{P(B|E)P(E)}{P(B)} - P(E)\right] + P(A|B,E^c)\left[\frac{P(B|E^c)P(E^c)}{P(B)} - P(E^c)\right] \\
&= \frac{P(A|B,E)P(E)}{P(B)}[P(B|E) - 1] + \frac{P(A|B,E^c)P(E^c)}{P(B)}[P(B|E^c) - 1] \\
&\leq 0
\end{aligned}$$

since $\frac{P(A|B,E)P(E)}{P(B)} > 0$ and $P(B|E) - 1 < 0$

and $\frac{P(A|B,E^c)P(E^c)}{P(B)} > 0$ and $P(B|E^c) - 1 < 0$

But we deduced $P(A|B) - P(A) > 0$. Thus, a contradiction is derived.
Hence, this situation is impossible

$\blacksquare$

2. Consider the following conversation from an episode of *The Simpsons*:

> Lisa: *Dad, I think he's an ivory dealer! His boots are ivory, his hat is ivory, and I'm pretty sure that check is ivory.*
> Homer: *Lisa, a guy who's got lots of ivory is less likely to hurt Stampy than a guy whose ivory supplies are low.*

$P(S \mid \text{ivory dealer})$

Here Homer and Lisa are debating the question of whether or not the man (named Blackheart) is likely to hurt Stampy the Elephant if they sell Stampy to him. They clearly disagree about how to use their observations about Blackheart to learn about the probability (conditional on the evidence) that Blackheart will hurt Stampy.

(a) Define clear notation for the various events of interest here.

(b) Express Lisa's and Homer's arguments (Lisa's is partly implicit) as conditional probability statements in terms of your notation from (a).

(c) Assume it is true that someone who has a lot of a commodity will have less desire to acquire more of the commodity. Explain what is wrong with Homer's reasoning that the evidence about Blackheart makes it less likely that he will harm Stampy.

a) Let D be the event that Blackheart is an ivory dealer.
Let I be the event that Blackheart have lots of ivory
Let S be the event that Stampy is hurt.

b) Lisa's argument: $P(S \mid I) > P(S \mid I^c)$ and $P(D \mid I) > P(D)$
Homer's argument: $P(S \mid I) < P(S \mid I^c)$

c) If Blackheart is a dealer then $P(S \mid I) < P(S \mid I^c)$ might not be true.

c)

3 Gambler's Ruin

1. A gambler repeatedly plays a game where in each round, he wins a dollar with probability 1/3 and loses a dollar with probability 2/3. His strategy is "quit when he is ahead by \$2," though some suspect he is a gambling addict anyway. Suppose that he starts with a million dollars. Show that the probability that he'll ever be ahead by \$2 is less than 1/4.

Let event that gambler is ahead by \$2 be G.
Let event that gambler wins be W $\qquad P(W) = 1/3$
Let event that gambler loses be L $\qquad P(L) = 2/3$

$$P(G) = P(G|L)\cdot P(L) + P(G|W)\cdot P(W)$$
$$= P(G|L)\tfrac{2}{3} + \tfrac{1}{3}$$

$$P(G|L) = P(G|LW)\cdot P(W|L) + P(G|LL)\cdot P(L|L)$$
$$= P(G|LW)\cdot\tfrac{1}{3} + P(G|LL)\cdot\tfrac{2}{3}$$

$$\Rightarrow P(G) = \tfrac{1}{3} + \tfrac{2}{3}\left[P(G|LW)\cdot\tfrac{1}{3} + P(G|LL)\cdot\tfrac{2}{3}\right]$$
$$= \tfrac{1}{3} + \tfrac{2}{9}P(G|LW) + \tfrac{4}{9}P(G|LL)$$

Let $p_i = P(G \mid i \text{ more to win})$

$$p_i = P(G \mid W \cap i \text{ more to win})\,P(W) + P(G \mid L \cap i \text{ more to win})\,P(L)$$
$$= p_{i-1}P(W) + p_{i+1}P(L)$$
$$= p_{i-1}\tfrac{1}{3} + p_{i+1}\tfrac{2}{3}$$

Let $p_i = x^i$

$$x^i = x^{i-1}\tfrac{1}{3} + x^{i+1}\tfrac{2}{3}$$
$$x = \tfrac{1}{3} + x^2\tfrac{2}{3}$$

$$\tfrac{2}{3}x^2 - x + \tfrac{1}{3} = 0$$

$$2x^2 - 3x + 1 = 0$$
$$(x-1)(2x-1) = 0$$

$$x = 1 \text{ or } x = \tfrac{1}{2}$$

$$p_i = 1 \text{ or } x = (\tfrac{1}{2})^i$$

$$a + b(\tfrac{1}{2})^i = p_i$$

$$p_0 = 1 \qquad p_{1000000} = a + b(\tfrac{1}{2})^{1000000} = 0$$
$$a = -b(\tfrac{1}{2})^{1000000}$$
$$\Rightarrow a < 0$$

$$b(\tfrac{1}{2})^0 = 1$$
$$b = 1$$

$$p_i < (\tfrac{1}{2})^i \qquad p_2 < (\tfrac{1}{2})^2 = (\tfrac{1}{4})$$

$$\therefore p_2 < \tfrac{1}{4}$$

$\square$.

1. (a) In the World Series of baseball, two teams (call them A and B) play a sequence of games against each other, and the first team to win four games wins the series. Let p be the probability that A wins an individual game, and assume that the games are independent. What is the probability that team A wins the series?

(b) Give a clear intuitive explanation of whether the answer to (a) depends on whether the teams always play 7 games (and whoever wins the majority wins the series), or the teams stop playing more games as soon as one team has won 4 games (as is actually the case in practice: once the match is decided, the two teams do not keep playing more games).

a) Let $W_k \sim \text{Bern}(p)$ be the random variable representing if A won game k.

Let $A_k = W_1 + \ldots + W_k$

never use k.

$P(A_k = 4) = \binom{k}{4} p^4 (1-p)^{k-4}$

$P(B_k = 4) = \binom{k}{4} (1-p)^4 p^{k-4}$

If WLWWWLL then this sequence should instead fit in $P(A=4 \cap B=1)$ since won at WLWWW.

$P(A_k = 4 \cap B_k = 3) = \binom{7}{4} p^4 (1-p)^3$

$P(A_k = 4 \cap B_k = 2) = \binom{6}{4} p^4 (1-p)^2$

$\vdots$

$P(A_k = 4 \cap B_k = 0) = p^4$

$$P(A_k = 4 \cap B_k < 4) = \binom{7}{4} p^4 (1-p)^3 + \ldots + \binom{5}{4} p^4 (1-p)^1 + p^4$$
$$= p^4 \left(\binom{7}{4}(1-p)^3 + \ldots + \binom{5}{4}(1-p) + 1 \right)$$

Ans: $p^4 \left[\binom{7}{4}(1-p)^3 + \ldots + \binom{5}{4}(1-p) + 1 \right]$

b) 2nd case. 1st case only applies for part (a) when A wins 4 and loses 3 games. However, the cases where A wins 5, A wins 6, and A wins 7 are also included in 1st case but are not required in part (a), since (a) only needs 4 wins.

a) Let A be the variable representing the no. of games won by A.

$P(A = 4 \mid 4 \text{ games}) = p^4$

$P(A = 4 \mid 5 \text{ games}) = p \cdot \binom{4}{3} p^3 (1-p)$

$P(A = 4 \mid 6 \text{ games}) = p \binom{5}{3} p^3 (1-p)^2$

$P(A = 4 \mid 7 \text{ games}) = p \binom{6}{3} p^3 (1-p)^3$

$$\therefore P(A) = p^4 + \binom{4}{3} p^4 (1-p) + p^4 \binom{5}{3} (1-p)^2 + p^4 \binom{6}{3} (1-p)^3$$
$$= p^4 \left[1 + \binom{4}{3}(1-p) + \binom{5}{3}(1-p)^2 + \binom{6}{3}(1-p)^3 \right] \quad \square.$$

b) No. doesn't matter.

In case 1, At some point out of the 7 games, the outcome will already be decided the instant a group wins 4 games, since that determines that the group has majority wins.

2. A sequence of n independent experiments is performed. Each experiment is a success with probability p and a failure with probability $q = 1 - p$. Show that conditional on the number of successes, all possibilities for the list of outcomes of the experiment are equally likely (of course, we only consider lists of outcomes where the number of successes is consistent with the information being conditioned on).

Let S_i be an indicator such that $S_i = 1$ if the ith experiment succeeds, and $S_i = 0$ if the ith experiment fails.

Let $S = S_1 + \dots + S_n$ be a random variable, with $S \sim \text{Bin}(n, p)$.

Let a_k be some outcome with k successes.
Let b_k be another outcome with k successes s.t. $a_k \neq b_k$.

$$P(a_k \mid S = k) = \frac{P(a_k \cap S = k)}{P(S = k)} = \frac{P(a_k)}{P(S=k)} = \frac{p^k q^{n-k}}{\binom{n}{k} p^k q^{n-k}} = \frac{1}{\binom{n}{k}}$$

Similarly, $P(b_k \mid S = k) = \dfrac{P(b_k \cap S = k)}{P(S=k)} = \dfrac{1}{\binom{n}{k}}$

$$P(a_k \mid S = k) = P(b_k \mid S = k)$$
$\Rightarrow$ all outcomes are equally likely.

$\square$.

3. Let $X \sim \text{Bin}(n, p)$ and $Y \sim \text{Bin}(m, p)$, independent of X.

(a) Show that $X + Y \sim \text{Bin}(n + m, p)$, using a story proof.

(b) Show that $X - Y$ is *not* Binomial.

(c) Find $P(X = k|X + Y = j)$. How does this relate to the elk problem from HW 1?

a) Let X be the number of heads given n coin flips, with heads having probability p.
Let Y be the number of heads given m coin flips. with heads having probability p.

Then the number of $m+n$ coin flips is $X + Y = \text{Bin}(n+m, p)$

$\square$.

b) $X \sim \text{Bin}(n, p)$ $Y \sim \text{Bin}(m, p)$.

By contradiction, suppose $X - Y$ is binomial, then

$X - Y \sim \text{Bin}(k, q)$ where $0 \leq q \leq 1$ and $k > 0$.

$\Rightarrow P(X - Y = i) = \binom{k}{i} q^i (1-q)^{k-i}$

$P(X - Y = 0) = P(X = i \cap Y = i)$ for $k = 0, \dots n$ if $n < m$ or $k = 0, \dots, m$ if $m < n$.

Suppose $n < m$,

$$P(X - Y = 0) = \sum_{i=0}^{n} P(X = i) P(Y = i)$$
$$= \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} \binom{m}{i} p^i (1-p)^{m-i}$$
$$= \sum_{i=0}^{n} \binom{n}{i} \binom{n}{i} p^{2i} (1-p)^{n+m-2i}$$

From the definition of a binomial $X-Y$,

$$P(X-Y=0) = \binom{k}{0} q^0 (1-q)^k$$
$$= (1-q)^k$$

Hence $$(1-q)^k = \sum_{i=0}^{n} \binom{n}{i}^2 p^{2i} (1-p)^{n+m-2i}$$

Binomial cannot be negative, but $X-Y$ can be.

c) $$P(X=k \mid X+Y=j) = \frac{P(X=k \cap X+Y=j)}{P(X+Y=j)}$$

$$= \frac{P(X=k \cap Y=j-k)}{P(X+Y=j)}$$

$$= \frac{\binom{n}{k} p^k (1-p)^{n-k} \binom{m}{j-k} p^{j-k} (1-p)^{m-j+k}}{\binom{n+m}{j} p^j (1-p)^{n+m-j}}$$

$$= \frac{\binom{n}{k}\binom{m}{j-k} p^j (1-p)^{n+m-j}}{\binom{n+m}{j} p^j (1-p)^{n+m-j}}$$

$$= \frac{\binom{n}{k}\binom{m}{j-k}}{\binom{n+m}{j}}$$

Identical to Elk problem. Let total population of elks be $n+m$, with n tagged elks and m untagged elks.

Draw a elk sample of size j. What is the probability that the sample size j has exactly k tagged elks?

Homework.

1. (a) Consider the following 7-door version of the Monty Hall problem. There are 7 doors, behind one of which there is a car (which you want), and behind the rest of which there are goats (which you don't want). Initially, all possibilities are equally likely for where the car is. You choose a door. Monty Hall then opens 3 goat doors, and offers you the option of switching to any of the remaining 3 doors.

Assume that Monty Hall knows which door has the car, will always open 3 goat doors and offer the option of switching, and that Monty chooses with equal probabilities from all his choices of which goat doors to open. Should you switch? What is your probability of success if you switch to one of the remaining 3 doors?

(b) Generalize the above to a Monty Hall problem where there are $n \geq 3$ doors, of which Monty opens m goat doors, with $1 \leq m \leq n-2$.

$7 - 3 = 4$

(a) Let W be the event that you win given that you switch.
Let G be the event that you picked a goat initially.
Let M be the event that 3 goat doors are opened.

$$P(W|M) = P(W|M \cap G)\,P(G) + P(W|M \cap G^c)\,P(G^c)$$
$$= P(W|M \cap G)\,6/7$$

$P(W|M \cap G) = \frac{1}{3}$ (Since after M, there are 3 closed doors unpicked, one of which contains a car)

$$\therefore\ P(W|M) = \frac{1}{3} \cdot \frac{6}{7} = \frac{2}{7}$$

$P(G^c) = \frac{1}{7} < P(W|M)$
$\therefore$ You should switch.

b) $P(W|M) = P(W|M \cap G)\,P(G)$
$= P(W|M \cap G)\,\frac{n-1}{n}$

$$P(W|M \cap G) = \frac{1}{n-m-1}$$

$$P(W|M) = \frac{1}{n-m}\left(\frac{n-1}{n}\right) = \frac{n-1}{n(n-m)} \qquad \frac{n-1}{n(n-m-1)}$$

$\boxed{\text{Ans: } \frac{n-1}{n(n-m)}}$ Ans: $\frac{n-1}{n(n-m-1)}$

2)

2. The *odds* of an event with probability p are defined to be $\frac{p}{1-p}$, e.g., an event with probability 3/4 is said to have odds of 3 to 1 in favor (or 1 to 3 against). We are interested in a hypothesis H (which we think of as a event), and we gather new data as evidence (expressed as an event D) to study the hypothesis. The *prior* probability of H is our probability for H being true before we gather the new data; the *posterior* probability of H is our probability for it after we gather the new data. The *likelihood ratio* is defined as $\frac{P(D|H)}{P(D|H^c)}$.

(a) Show that Bayes' rule can be expressed in terms of odds as follows: *the posterior odds of a hypothesis H are the prior odds of H times the likelihood ratio.*

(b) As in the example from class, suppose that a patient tests positive for a disease afflicting 1% of the population. For a patient who has the disease, there is a 95% chance of testing positive (in medical statistics, this is called the *sensitivity* of the test); for a patient who doesn't have the disease, there is a 95% chance of testing negative test (in medical statistics, this is called the *specificity* of the test).

The patient gets a second, independent test done (with the same sensitivity and specificity), and again tests positive. Use the odds form of Bayes' rule to find the probability that the patient has the disease, given the evidence, *in two ways*: in one step, conditioning on both test results simultaneously, and in two steps, first updating the probabilities based on the first test result, and then updating again based on the second test result.

Prior: $P(H)$
Posterior: $P(H|D)$
likelihood: $\frac{P(D|H)}{P(D|H^c)}$

a) posterior odds: $\frac{P(H|D)}{P(H^c|D)}$

prior odds: $\frac{P(H)}{P(H^c)}$

$$\frac{P(H|D)}{P(H^c|D)} = \frac{P(H)}{P(H^c)} \cdot \frac{P(D|H)}{P(D|H^c)}$$

$$\frac{P(H|D)}{P(H)P(D|H)} = \frac{P(H^c|D)}{P(H^c)P(D|H^c)}$$

$$\frac{P(H)P(D|H)}{P(H|D)} = \frac{P(H^c)P(D|H^c)}{P(H^c|D)}$$

Using LOTP, $P(D) = P(D|H)P(H) + P(D|H^c)P(H^c)$
$P(H^c)P(D|H^c) = P(D) - P(D|H)P(H)$

$$\Rightarrow \frac{P(H)P(D|H)}{P(H|D)} = \frac{P(D) - P(D|H)P(H)}{1 - P(H|D)}$$

$$P(H)P(D|H) - P(H)P(H|D)P(D|H)$$
$$= P(D)P(H|D) - P(H)P(D|H)P(H|D)$$

$$P(H)P(D|H) = P(D)P(H|D)$$

$$P(H|D) = \frac{P(H)P(D|H)}{P(D)}$$

$\square$.

b) Let T_1 be event that patient tests positive in first test
Let T_2 be event that patient tests positive in second test.
Let D be event that patient is infected with disease.

Prior: $P(D) = 1/100$
Posterior: $P(D|T_1, T_2) = ?$
Likelihood: $\frac{P(T_1 \cap T_2|D)}{P(T_1 \cap T_2|D^c)} = \frac{P(T_1|D)P(T_2|D)}{P(T_1|D^c)P(T_2|D^c)} = \frac{95/100 \cdot 95/100}{5/100 \cdot 5/100}$

$$= \frac{95 \cdot 95}{5 \cdot 5} = 19 \cdot 19$$

$$\frac{P(D|T_1,T_2)}{P(D^c|T_1,T_2)} = \frac{P(D)}{P(D^c)} \cdot \frac{P(T_1,T_2|D)}{P(T_1,T_2|D^c)}$$

$$= \frac{(1/100)}{99/100} \cdot 19 \cdot 19 = \frac{19^2}{99}$$

$$\frac{P(D|T_1,T_2)}{1-P(D|T_1,T_2)} = \frac{19^2}{99}$$

Let $p = P(D|T_1,T_2)$

$$\frac{p}{1-p} = \frac{19^2}{99}$$

$$99p = 19^2 - 19^2 p$$

$$(99+19^2)p = 19^2$$

$$p = \frac{19^2}{99+19^2} \approx 0.785 \text{ (3 s.f.)}$$

$P(D) = 1/100 \quad P(D|T_1) = ? \quad \frac{P(T_1|D)}{P(T_1|D^c)} = \frac{95}{5} = 19$

$$\frac{P(D|T_1)}{1-P(D|T_1)} = \frac{95}{5} \cdot \frac{1}{99} = \frac{19}{99}$$

$$99p = 19 - 19p$$

$$118p = 19$$

$$p = \frac{19}{118}$$

$$P(D|T_1) = \frac{19}{118}$$

Let $\tilde{P}(D) = P(D|T_1)$

$\tilde{P}(D) = \frac{19}{118} \quad \tilde{P}(D|T_2) = ? \quad \frac{\tilde{P}(T_2|D)}{\tilde{P}(T_2|D^c)} = \frac{95}{5} = 19$

$$\Rightarrow \frac{\tilde{P}(D|T_2)}{1-\tilde{P}(D|T_2)} = 19 \cdot \left(\frac{19/118}{1-19/118}\right) = 19 \cdot \frac{19}{99} = \frac{19^2}{99}$$

$$p = \frac{19^2}{99} - \frac{19^2}{99}p$$

$$\frac{460}{99}p = \frac{19^2}{99}$$

$$p \approx 0.785 \quad //$$

3. Is it possible to have events A_1, A_2, B, C with $P(A_1|B) > P(A_1|C)$ and $P(A_2|B) > P(A_2|C)$, yet $P(A_1 \cup A_2|B) < P(A_1 \cup A_2|C)$? If so, find an example (with a "story" interpreting the events, as well as giving specific numbers); otherwise, show that it is impossible for this phenomenon to happen.

Yes. Let B be event that doctor B is doing surgery.
Let C be event that doctor C is doing surgery.

Let A_1 be event that surgery is heart surgery.
Let A_2 be event that surgery is stitching.

		B	C
A_1	success	1	87
	fail	0	13
A_2	success	39	0
	fail	11	1

100

B overall = 0.784
C overall = 0.861

Simpson's paradox.

4. Calvin and Hobbes play a match consisting of a series of games, where Calvin has probability p of winning each game (independently). They play with a "win by two" rule: the first player to win two games more than his opponent wins the match. Find the probability that Calvin wins the match (in terms of p), in two different ways:

(a) by conditioning, using the law of total probability.

(b) by interpreting the problem as a gambler's ruin problem.

a) $P(\text{Calvin wins}) = P(\text{Calvin wins} \mid WW) \cdot P(WW) + 2P(\text{Calvin wins} \mid WL)\, P(WL) + P(\text{Calvin wins} \mid LL) \cdot P(LL)$

$= p^2 + 2P(\text{Calvin wins} \mid WL)\, P(WL)$

$= p^2 + 2P(\text{Calvin wins})\, p(1-p)$

Let $P(\text{Calvin wins}) = h$

$h = p^2 + 2hp(1-p)$

$h - 2hp(1-p) = p^2$

$h(1-2p(1-p)) = p^2$

$h = \dfrac{p^2}{1-2p+2p^2}$ //

4b) Let C be event where Calvin wins.

Let G_k be event that Calvin must win k games to win the match.

Let W be event that Calvin wins the latest match.

$P(C \mid G_k) = P(C \mid G_k \cap W)\, P(W) + P(C \mid G_k \cap W^c)\, P(W^c)$

$= P(C \mid G_{k-1})\, p + P(C \mid G_{k+1})(1-p)$

Let $f_k = P(C \mid G_k)$

$f_k = f_{k-1}\, p + f_{k+1}(1-p)$

Let $f_k = x^k$

$x^k = x^{k-1} p + x^{k+1}(1-p)$

$x = p + x^2(1-p)$

$x^2(1-p) - x + p = 0$

$x = \dfrac{1 \pm \sqrt{1-4(1-p)p}}{2(1-p)} = \dfrac{1 \pm \sqrt{1-4p+4p^2}}{2(1-p)}$

$= \dfrac{1 \pm \sqrt{(2p-1)(2p-1)}}{2(1-p)}$

$= \dfrac{1 \pm (2p-1)}{2(1-p)} = \dfrac{1+2p-1}{2(1-p)}$ or $\dfrac{1-2p+1}{2(1-p)}$

$\dfrac{2p}{2(1-p)}$ or $\dfrac{2-2p}{2(1-p)}$

$\dfrac{p}{1-p}$ or 1

$$\Rightarrow x = \frac{p}{1-p} \quad \text{or} \quad 1$$

Let $f_k = A\left(\frac{p}{1-p}\right)^k + B$

$f_0 = 1$

$$A\left(\frac{p}{1-p}\right)^0 + B = 1$$

$$A + B = 1 \qquad A = 1 - B$$

$f_4 = 0$

$$A\left(\frac{p}{1-p}\right)^4 + B = 0$$

$$(1-B)\left(\frac{p}{1-p}\right)^4 + B = 0$$

$$\left(\frac{p}{1-p}\right)^4 + B\left(1 - \left(\frac{p}{1-p}\right)^4\right) = 0$$

$$B = \frac{\left(\frac{p}{1-p}\right)^4}{\left(\frac{p}{1-p}\right)^4 - 1} \qquad A = -\frac{1}{\left(\frac{p}{1-p}\right)^4 - 1}$$

$$\Rightarrow f_k = -\frac{\left(\frac{p}{1-p}\right)^k}{\left(\frac{p}{1-p}\right)^4 - 1} + \frac{\left(\frac{p}{1-p}\right)^4}{\left(\frac{p}{1-p}\right)^4 - 1}$$

$$f_2 = \frac{-\left(\frac{p}{1-p}\right)^2 + \left(\frac{p}{1-p}\right)^4}{\left(\frac{p}{1-p}\right)^4 - 1} = \frac{\left(\frac{p}{1-p}\right)^2 \cancel{\left(\left(\frac{p}{1-p}\right)^2 - 1\right)}}{\cancel{\left(\left(\frac{p}{1-p}\right)^2 - 1\right)}\left(\left(\frac{p}{1-p}\right)^2 + 1\right)}$$

$$= \frac{\left(\frac{p}{1-p}\right)^2}{\left(\frac{p}{1-p}\right)^2 + 1} = \frac{\frac{p^2}{(1-p)^2}}{\frac{p^2 + (1-p)^2}{(1-p)^2}}$$

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

$$= \frac{p^2}{p^2 + p^2 - 2p + 1}$$

$$= \frac{p^2}{2p^2 - 2p + 1} \quad //$$

5. A fair die is rolled repeatedly, and a running total is kept (which is, at each time, the total of all the rolls up until that time). Let p_n be the probability that the running total is ever *exactly* n (assume the die will always be rolled enough times so that the running total will eventually exceed n, but it may or may not ever equal n).

(a) Write down a recursive equation for p_n (relating p_n to earlier terms p_k in a simple way). Your equation should be true for all positive integers n, so give a definition of p_0 and p_k for $k < 0$ so that the recursive equation is true for small values of n.

(b) Find p_7.

(c) Give an intuitive explanation for the fact that $p_n \to 1/3.5 = 2/7$ as $n \to \infty$.

a) $p_n = p_{n-1} \cdot \frac{1}{6} + p_{n-2}\frac{1}{6} + \dots + p_{n-6}\frac{1}{6} = \frac{1}{6}(p_{n-1} + \dots + p_{n-6})$

$p_0 = 1$, $p_k = 0$ ~~for~~ $k < 0$.

$p_2 = \frac{1}{6}$

b) $p_0 = 1$

$p_1 = \frac{1}{6}$

$p_2 = \frac{1}{6}(\frac{1}{6} + 1) = \frac{1}{6} \cdot \frac{7}{6} = \frac{7}{6^2}$

$p_3 = \frac{1}{6}(\frac{7}{6^2} + \frac{1}{6} + 1) = \frac{49}{216} = \frac{7^2}{6^3}$

$p_4 = \frac{7^3}{6^4}$ $p_5 = \frac{7^4}{6^5}$ $p_6 = \frac{7^5}{6^6}$

$p_7 = \frac{1}{6}[\frac{7^5}{6^6} + \dots + \frac{1}{6}] = (\frac{1}{6})^2[\frac{7^5}{6^5} + \dots + 1]$

$= (\frac{1}{6})^2 [\frac{1-(\frac{7}{6})^6}{1-\frac{7}{6}}] = (\frac{1}{6})^2 [\frac{1-(\frac{7}{6})^6}{-\frac{1}{6}}]$

$= ((\frac{7}{6})^6 - 1)\frac{1}{6} = \frac{7^6}{6^7} - \frac{1}{6}$.

≈ 0.2536

c) $\lim_{n\to\infty} p_n = \frac{1}{6}(p_{n-1} + \dots + p_{n-6})$

6. Players A and B take turns in answering trivia questions, starting with player A answering the first question. Each time A answers a question, she has probability p_1 of getting it right. Each time B plays, he has probability p_2 of getting it right.

(a) If A answers m questions, what is the PMF of the number of questions she gets right?

(b) If A answers m times and B answers n times, what is the PMF of the total number of questions they get right (you can leave your answer as a sum)? Describe exactly when/whether this is a Binomial distribution.

(c) Suppose that the first player to answer correctly wins the game (with no predetermined maximum number of questions that can be asked). Find the probability that A wins the game.

6) Let N_A be the numbers of questions A got correct

$$N_A \sim \text{Bin}(m, p_1).$$

$\text{Bin}(m, p_1)$ is the no. of questions she gets right.

b) Let N_B be the number of questions B got correct

$$N_B \sim \text{Bin}(n, p_2)$$

$$N_B + N_A = \text{Bin}(m, p_1) + \text{Bin}(n, p_2)$$

This is binomial when $p_1 = p_2$. It is not binomial if $p_1 \neq p_2$.

c) $P(A \text{ wins the game}) = P(W_A) + P(L_A L_B W_A) + P(L_A L_B L_A L_B W_A) + \ldots$

$$= p_1 + (1-p_1)(1-p_2)p_1 + (1-p_1)^2(1-p_2)^2 p_1 + \ldots + (1-p_1)^k(1-p_2)^k p_1$$

$$= p_1 \left[1 + (1-p_1)(1-p_2) + \ldots + (1-p_1)^k(1-p_2)^k + \ldots \right]$$

$$= \lim_{k \to \infty} p_1 \left[\frac{1-(1-p_1)^{k+1}(1-p_2)^{k+1}}{1-(1-p_1)(1-p_2)} \right]$$

$$= \frac{p_1}{1-(1-p_1)(1-p_2)}$$

//

7) 7. A message is sent over a noisy channel. The message is a sequence $x_1, x_2, \ldots, x_n$ of n bits ($x_i \in \{0,1\}$). Since the channel is noisy, there is a chance that any bit might be corrupted, resulting in an error (a 0 becomes a 1 or vice versa). Assume that the error events are independent. Let p be the probability that an individual bit has an error ($0 < p < 1/2$). Let $y_1, y_2, \ldots, y_n$ be the received message (so $y_i = x_i$ if there is no error in that bit, but $y_i = 1 - x_i$ if there is an error there).

To help detect errors, the nth bit is reserved for a parity check: x_n is defined to be 0 if $x_1 + x_2 + \cdots + x_{n-1}$ is even, and 1 if $x_1 + x_2 + \cdots + x_{n-1}$ is odd. When the message is received, the recipient checks whether y_n has the same parity as $y_1 + y_2 + \cdots + y_{n-1}$. If the parity is wrong, the recipient knows that at least one error occurred; otherwise, the recipient assumes that there were no errors.

(a) For $n = 5, p = 0.1$, what is the probability that the received message has errors which go undetected?

(b) For general n and p, write down an expression (as a sum) for the probability that the received message has errors which go undetected.

(c) Give a simplified expression, not involving a sum of a large number of terms, for the probability that the received message has errors which go undetected.

Hint for (c): Letting

$$a = \sum_{k \text{ even},\, k \geq 0} \binom{n}{k} p^k (1-p)^{n-k} \text{ and } b = \sum_{k \text{ odd},\, k \geq 1} \binom{n}{k} p^k (1-p)^{n-k},$$

the binomial theorem makes it possible to find simple expressions for $a + b$ and $a - b$, which then makes it possible to obtain a and b.

Let D be the event that an error is detected.
Let E be the event that an error exists.
Let C be the event that the parity check is corrupted.
Let S be the event that the sum has the same parity as original.

$$P(D^c \mid E) = P(D^c \mid E \cap S) P(S \mid E) + P(D^c \mid E \cap S^c) P(S^c \mid E)$$

$$P(S \mid E) = \sum_{k \text{ even } k \geq 2} \binom{n}{k} p^k (1-p)^{n-k} = a$$

$$P(D^c \mid E \cap S) = P(D^c \mid E \cap S \cap C) P(C \mid E \cap S) + P(D^c \mid E \cap S \cap C^c) P(C^c \mid E \cap S)$$
$$= P(C^c \mid E \cap S)$$
$$= 1 - p$$

$$P(S^c \mid E) = \sum_{k \text{ odd}, k \geq 1} \binom{n}{k} p^k (1-p)^{n-k} = b$$

$$P(D^c \mid E \cap S^c) = P(D^c \mid E \cap S^c \cap C) P(C \mid E \cap S^c) + P(D^c \mid E \cap S^c \cap C^c) P(C^c \mid E \cap S^c)$$
$$= p$$

$$P(D^c \mid E) = (1-p)a + bp$$

Let $n = 5$, $p = 0.1$.

$$P(D^c \mid E) = 0.9\left(\binom{5}{2} 0.1^2\, 0.9^3 + \binom{5}{4} 0.1^4\, 0.9\right) + 0.1\left(\binom{5}{1} 0.1\, 0.9^4 + \binom{5}{3} 0.1^3\, 0.9^2 + \binom{5}{5} 0.1^5\right)$$
$$= 0.9[0.0729 + 0.00045] + 0.1[0.32805 + 0.0081 + 0.00001]$$
$$= 0.099631$$

b) $P(D^c \mid E) = (1-p)a + bp$ where

$$a = \sum_{k \text{ even}, k \geq 2} \binom{n}{k} p^k (1-p)^{n-k}$$

$$b = \sum_{k \text{ odd}, k \geq 1} \binom{n}{k} p^k (1-p)^{n-k}$$

c) If parity bit cannot be corrupted, then

$$P(D^c|E) = \sum_{k \text{ even}, k \geq 2} \binom{n}{k} p^k (1-p)^{n-k}$$

$$a+b = \binom{n}{0} p^0 (1-p)^n + \binom{n}{1} p(1-p)^{n-1} + \ldots + \binom{n}{n} p^n$$

$$= [p + (1-p)]^n = 1 \quad — (1)$$

$$a-b = \binom{n}{0} p^0 (1-p)^n - \binom{n}{1} p(1-p)^{n-1} + \ldots + \binom{n}{n} p^n$$

$$= [p - (1-p)]^n = [p-1+p]^n$$

$$= (2p-1)^n \quad — (2)$$

(1) $a+b=1$

$a = 1-b$

(2) $1-b-b = (2p-1)^n$

$1-2b = (2p-1)^n$

$2b = 1-(2p-1)^n$

$b = \frac{1-(2p-1)^n}{2}$

$a = \frac{2-1+(2p-1)^n}{2}$

$= \frac{1+(2p-1)^n}{2}$

$$P(D^c|E) = a - \binom{n}{0}(1-p)^n$$

$$= \frac{1+(2p-1)^n}{2} - (1-p)^n$$

//