Stat 110 Strategic Practice 3, Fall 2011

Prof. Joe Blitzstein (Department of Statistics, Harvard University)

1 Continuing with Conditioning

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 $\frac{1}{2} \leq p \leq 1$.

   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 $\frac{1}{2} \leq p \leq 1$).

   (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). (SEE SEPARATE PAGE)

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

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

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

   $\boxed{\text{False.}}$ (a) If X and Y are independent and Y and Z are independent, then X and Z are independent. COUNTEREXAMPLE: consider $Z = 3X$

   $\boxed{\text{False.}}$ (b) If X and Y are independent, then they are conditionally independent given Z. (SEE BELOW.)

   $\boxed{\text{False.}}$ (c) If X and Y are conditionally independent given Z, then they are independent. (SEE SEPARATE PAGE)

$$P(X=x)P(Y=y) = P(X=x, Y=y) \not\Rightarrow$$
$$P(X=x|Z=z)P(Y=y|Z=z) = P(X=x, Y=y|Z=z)$$

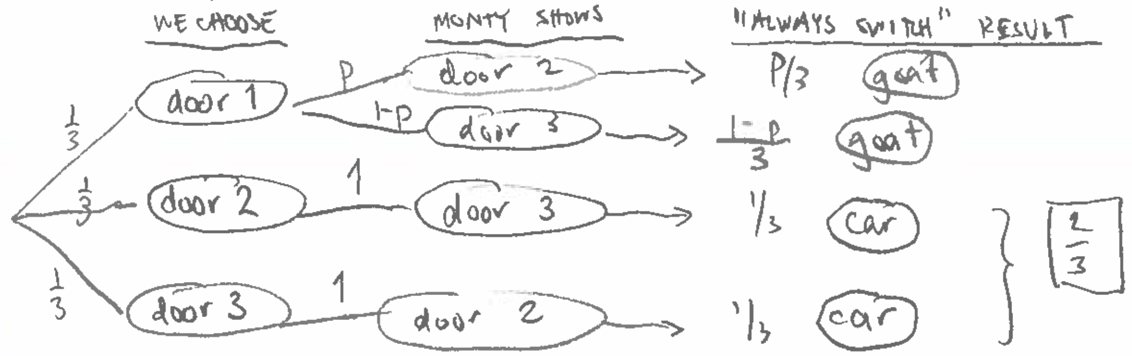
2b. COUNTEREXAMPLE:
Let X be indicator for first coin flip, Y be indicator for second, where indicator is $\{0,1\}$.
Let $X+Y = Z$. If we know Z, these equalities do not hold:

$$\begin{cases} P(Y=y|X=x, Z=z) \stackrel{?}{=} P(Y=y|Z=z) \\ P(X=x|Y=y, Z=z) \stackrel{?}{=} P(X=x|Z=z) \end{cases}$$

(e.g. consider what happens to Y given knowing X and Z vs just Z).

① [Continuing with conditioning, SP3]

✓ A. Unconditional probability of success w/ "always switch" strategy? (Assume C_1 for concreteness.)



✓ B. Conditional probability of success w/ "always switch" strategy given Monty shows door 2 (M_2)? Assume C_1, get car ≡ end up on door 1

$$\frac{\frac{1}{3}}{\frac{1}{3}+\frac{p}{3}} = \boxed{\frac{1}{1+p}} = \frac{P(\text{get car} \mid M_2)}{[P(\text{get car} \mid M_2) + P(\neg \text{get car} \mid M_2)]} \cdot \cancel{\frac{P(M_2)}{P(M_2)}}$$

✓ C. ... given Monty shows door 3 (M_3). Assume C_1, get car ≡ end up on door 1.

$$\frac{\frac{1}{3}}{\frac{1}{3}+\frac{1-p}{3}} = \boxed{\frac{1}{2-p}}$$

following same reasoning as above.

③ [Bernoulli & Binomial, SP3]

1A ✓ "First to 4" requires counting 4 wins within [4, 7] games.

✓ ⑦ $\sum_{i=4}^{7} \binom{i}{4} p^4 (1-p)^{i-4}$ ⟹ NOT QUITE CORRECT because 5, 6, 7 game series overcount cases.

$$P(A \text{ wins}) = \underbrace{p^4}_{\text{win in } 4} + \underbrace{\binom{4}{3} p^4 (1-p)}_{\text{win in } 5} + \underbrace{\binom{5}{3} p^4 (1-p)^2}_{\text{win in } 6} + \underbrace{\binom{6}{3} p^4 (1-p)^3}_{\text{win in } 7}$$

(binomial coefficients come from the number of ways to arrange wins such that the ways to win series in fewer games are not counted)

e.g. $\binom{5}{3}$ ≡ need to win 3 games within first 5 to win series in game k

✓ 1B. $\boxed{\text{No}}$ because in each scenario, 4 wins (a majority) determines series win, so games played afterward have no bearing on series outcome. Consider simple 3 game series: A can win by [W, W], [L, W, W], [W, L, W] for first case which is same as [W, W, W], [W, W, L], [L, W, W], [W, L, W].

Algebraically,

$$p^4 + \binom{4}{3} p^4 (1-p) + \binom{5}{3} p^4 (1-p)^2 + \binom{6}{3} p^4 (1-p)^3 \overset{\checkmark}{=} \sum_{k=4}^{7} \binom{7}{k} p^k (1-p)^{7-k}$$

for all choices of valid p.

) [Simpson's Paradox, SP3]

A. ~~not~~ possible by LOTP: Given $P(A|E) < P(B|E)$ and $P(A|E^c) < P(B|E^c)$

$\underbrace{P(A|E)P(E) + P(A|E^c)P(E^c)}_{P(A)} \leq \underbrace{P(B|E)P(E) + P(B|E^c)P(E^c)}_{P(B)}$ since probabilities must be $\in [0,1]$.

B. yes, possible if $P(E|B) \geq P(E|B^c)$ or $P(E^c|B) \geq P(E^c|B^c)$. since

$$P(A|B) = P(A|B,E)\boxed{P(E|B)} + P(A|B,E^c)\boxed{P(E^c|B)}$$
$$\vee \qquad \wedge \qquad \vee \qquad \wedge \qquad \vee$$
$$P(A|B^c) = P(A|B^c,E)\boxed{P(E|B^c)} + P(A|B^c,E^c)\boxed{P(E^c|B^c)}$$

(but both pairs cannot be same sign).

STORY: Let A be event of on-time (shipment) departure, B be event shipment leaves very inefficient port, and E be event of a shipment having a good OA. So B^c is event that the port of call is efficient and E^c is event the shipment is moved by a bad OA. An efficient port may look worse, i.e., $P(A|B^c) < P(A|B)$, absent info about what kind of OA is moving the goods if good OAs work using the inefficients ports more frequently!

e.g. $P(A|B) = (.80)\boxed{(.99)} + (.60)(.01) \doteq 0.798$

$$\vee \qquad \wedge \qquad \vee \qquad \wedge \qquad \vee$$

$P(A|B^c) = (.90)\boxed{(.04)} + (.75)(.96) = 0.7545$

) [Continuing with Conditioning, SP3]

2c. $P(Y=y|X=x, Z=z) = P(Y=y|Z=z) \overset{?}{\Rightarrow} P(Y=y|X=x) = P(Y=y)$: $\boxed{\text{No}}$

Consider case when $Z \sim \text{Bern}(\frac{1}{2})$ and z indicates whether a good trucking co. or a bad one was used. Let X be whether demurrage is paid on first shipment, Y be whether demurrage paid on second. Assume that the probability of paying demurrage is p_1 for good co., p_2 for bad co., and $p_2 > p_1$. If $Z=1$ (good co. used) then X and Y are [conditionally] independent and $X, Y \sim \text{Bern}(p_1)$ and i.i.d. However, knowing we paid demurrage on the first shipment ($X=1$) increases likelihood that we are using the bad trucking company. ($P(Y=y|X=x) \neq P(Y=y)$). Same logic holds for $Z=0$ case.

True. ~~False.~~ → conditionally identical dist $\overset{?}{\Longrightarrow}$ identical dist.

(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. Consider good player outcomes = bad player outcomes on a day that's bad for good player, good for bad player

account for all z!

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).

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.*

   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.

   H ≡ person will hurt Stampy
   D ≡ person is ivory dealer
   L = person has lots of ivory

   (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). (SEE BACK)

   ALMOST (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

needed help w/ notation & formalizing the reasoning below.

LISA

↑ ivory ⇒ ↑ ivory dealer ⇒ ↑ desire to ~~want~~ more ivory ⇒ ↑ need to hurt Stampy

HOMER

↑ ivory ⇒ ↓ desire to ~~want~~ more ivory ⇒ ↓ need to hurt stampy

assume perfect correlation

2b. Homer is arguing that $P(H|L) < P(H) < P(H|L^c)$.

Lisa is arguing that $P(H|L) > P(H) > P(H|L^c)$

by considering $P(D|L) > P(D) > P(D|L^c)$.

2c. Homer can still be wrong even if

$$\begin{array}{ccc} P(H|L,D) & & P(H|L,D^c) \\ \wedge & \& & \wedge \\ P(H|L^c,D) & & P(H|L^c,D^c) \end{array}$$

(Lisa's argument)

$$\begin{array}{ccccccc} P(H|L) & = & P(H|L,D) & P(D|L) & + & P(H|L,D^c) & P(D^c|L) \\ \vee & & \wedge & \vee & & \wedge & \wedge \\ P(H) & & & P(D) & & & \\ \vee & & & \vee & & & \\ P(H|L^c) & = & P(H|L^c,D) & P(D|L^c) & + & P(H|L^c,D^c) & P(D^c|L^c) \end{array}$$

reasoning that the evidence about Blackheart makes it less likely that he will harm Stampy. (SEE BACK OF PREVIOUS PAGE)

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. (SEE ATTACHMENT)

4 Bernoulli and Binomial

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).

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). (SEE BELOW).

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. (SEE BACK)

   (b) Show that $X - Y$ is *not* Binomial. The support of $X-Y$ is $[-m, n]$ but the support of binomial dist. is strictly posi[tive]

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

2. Let $X \sim \text{Bern}(p)$ (indicator). A sequence of independent experiments yields $X_1, X_2, \dots X_n$. Let $Y = X_1 + X_2 + \dots X_n \Rightarrow Y \sim \text{Bin}(n, p)$.

$P(Y=y) = \dfrac{\binom{n}{y} p^y q^{(n-y)}}{\sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k}}$ so unconditionally, each outcome of n independent experiments is equally likely. Conditionally this is also true because the p[robabilities of] each unique $\{X_1, X_2, \dots X_n\}$ such that $Y=y$ all equal $p^y q^{n-y}$ and since they are disjoint events, their union = $\binom{n}{y} p^y q^{n-y}$. So conditionally P(each unique comb of $\{X_1 .. X_n\} | Y=y) = \frac{1}{\binom{n}{y}}$.

4. $X \sim \text{Bin}(n,p)$; $Y \sim \text{Bin}(m,p)$. Show $X+Y \sim B(n+m,p)$.

<u>STORY</u>: Imagine running experiment of flipping n coins that have p. chance of landing heads. Imagine doing same experiment m more times. The total number of heads would be $\text{Bin}(n+m,p)$ given that $X \sim \text{Bin}(n,p)$, $Y \sim \text{Bin}(m,p)$.

5. $P(X=k \mid X+Y=j)$ is the probability of getting k tagged elk from a sample of j elk. (Y is r.v. that describes non-tagged elk in sample.)

$$P(X=k \mid X+Y=j) = \frac{P(X=k,\ X+Y=j)}{P(X+Y=j)} \qquad \text{by Bayes Theorem}$$

$$= \frac{P(X=k)\ P(Y=j-k)}{P(X+Y=j)} \qquad \text{by ind. of } X, Y$$

$$= \frac{\binom{n}{k} \cancel{p^{k}} \cancel{q^{n-k}} \binom{m}{j-k} \cancel{p^{j-k}} \cancel{q^{m+k-j}}}{\binom{n+m}{j} \cancel{p^{j}} \cancel{q^{n+m-j}}} \qquad \text{by Bin Theorem}$$

$$= \frac{\binom{n}{k}\binom{m}{j-k}}{\binom{n+m}{j}} \qquad \text{same solution to elk problem.}$$

Stat 110 Homework 3, Fall 2011

Prof. Joe Blitzstein (Department of Statistics, Harvard University)

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? (SEE BACK)

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

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 $\frac{P(D|H)}{P(D|H^c)}$. (SEE ATTACHMENT)

(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.

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"

$\boxed{\text{Yes}}$ if $P(A_1, A_2 | B)$ is sufficiently large (i.e. $P(A_1 \cap A_2 | B) >$ $P(A_1|B) - P(A_1|C) + P(A_2|B) - P(A_2|C) + P(A_1, A_2|C)$).

STORY:

. Two approaches to solving this problem: "UNCONDITIONAL" & "BAYESIAN" ways. For both assume that we choose door #1 (for concreteness) and lets also say Monty shows goats behind doors #2, #5, #6.

"BAYESIAN" approach:

$$P(C_1 \mid M_{2,5,6}) = \frac{P(M_{2,5,6} \mid C_1)\, P(C_1)}{P(M_{2,5,6})} = \frac{\frac{1}{\cancel{\binom{6}{3}}} \cdot \frac{1}{7}}{\frac{1}{\cancel{\binom{6}{3}}}} = \frac{1}{7}$$

which is the probability of winning car if we stay w/ door #1. The complement is the probability of winning if we switch to door #3, #4, OR #7, so the probability of winning switching to any one of these doors is $\frac{(1-\frac{1}{7})}{3} = \boxed{\frac{2}{7}}$

"UNCONDITIONAL" approach:

$P(\text{get car}) = P(\text{get car} \mid C_1)\, \overset{\frac{1}{7}}{\overset{\shortparallel}{P(C_1)}} + \ldots P(\text{get car} \mid C_7)\, \overset{\frac{1}{7}}{\overset{\shortparallel}{P(C_7)}}$. If we switch $P(\text{get car} \mid C_1) = 0$, so $P(\text{get car}) = \frac{1}{7}\left(6 \cdot \frac{1}{3}\right) = \boxed{\frac{2}{7}}$

$= \frac{1}{7}\left[P(\text{get car} \mid C_2) + \ldots + P(\text{get car} \mid C_7)\right]$. $P(\text{get car} \mid C_i)$ for $2 \le i \le 7$ is $\frac{1}{3}$ since we have $\frac{1}{3}$ chance of picking correctly from the 3 doors that are closed (we don't count door #1 b/c we are presupposing switching).

∴ Generalizing the above, we can see by symmetry that if we have n doors, Monty shows m doors after our initial choice, the probability of getting car from switching to one of the remaining $n-m-1$ doors is $\boxed{\frac{(1-\frac{1}{n})}{n-m-1}}$.

2A $\underbrace{\boxed{\dfrac{P(H|D)}{P(H^c|D)}}}_{\text{posterior odds}} = \dfrac{P(D|H)\cdot P(H)}{P(D)\cdot P(H^c|D)} = \dfrac{P(D|H)\,P(H)}{P(D \cap H^c)}$

$= \underbrace{\dfrac{P(D|H)}{P(D|H^c)}}_{\text{likelihood ratio}} \cdot \underbrace{\boxed{\dfrac{P(H)}{P(H^c)}}}_{\text{prior odds}}$ ✓

2B. UPDATE simultaneously:

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

$$= \frac{(0.95)^2(.01)}{(0.05)^2(.99)} = 3.\overline{64}$$

(converted to probability) $\Rightarrow \dfrac{3.\overline{64}}{3.\overline{64}+1} \approx 0.784$ ✓

UPDATE sequentially:

$$\frac{P(D|+_1)}{P(D^c|+_1)} = \frac{P(+_1|D)}{P(+_1|D^c)} \cdot \frac{P(D)}{P(D^c)} = \frac{(.95)(.01)}{(.05)(.99)}$$

$$\frac{P(D|+_2)}{P(D^c|+_2)} = \frac{P(+_2|D)}{P(+_2|D^c)} \cdot \frac{P(D|+_1)}{P(D^c|+_1)} = \frac{(.95)}{(.05)} \cdot \frac{(.95)}{(.05)} \cdot \frac{(.01)}{(.99)}$$

$$= \frac{(0.95)^2(.01)}{(0.05)^2(.99)}$$

interpreting the events, as well as giving specific numbers); otherwise, show that it is impossible for this phenomenon to happen.

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. (SEE BACK)

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

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).

ALMOST

(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. (SEE ATTACHMENT)

(b) Find p_7. (SEE ATTACHMENT)

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

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.

ALMOST

(a) If A answers ~~m questions~~, what is the PMF of the number of questions she gets right? $A \sim \boxed{\text{Bin}(m, p_1)} = \binom{m}{k} p_1^k (1-p_1)^{m-k}$

(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. (SEE BELOW)

r.v. ≠ 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. (SEE BACK).

SEE BIO-HAZARD 3.7.7 (p. 115)

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

6b) $\boxed{\text{Bin}(m, p_1) + \text{Bin}(n, p_2)}$; this is $\text{Bin}(m+n, p)$ iff $p = p_1 = p_2$ and A is independent of B (where $A \sim \text{Bin}(m, p_1)$, $B \sim \text{Bin}(n, p_2)$). Let $T = A + B$, we are looking for $P(T=t) = \sum_{j=0}^{t} P(A+B=t \mid A=j) P(A=j)$ by LOTP. This simplifies to $\sum_{j=0}^{t} P(B=t-j) P(A=j) = \sum_{j=0}^{t} \binom{n}{t-j} p_2^{t-j} (1-p_2)^{n-t+j} \binom{m}{j} p_1^j (1-p_1)^{m-j}$.

. $P(W) = P(W|w_1, w_2)\, P(w_1, w_2) + P(W|w_1, l_2)\, P(w_1, l_2) + P(W|l_1, w_2)\, P(l_1, w_2) + P(W|l_1, l_2)\, P(l_1, l_2)$ $\Big\}$ $1 \cdot p^2 + 2pq\, P(W) + \cancel{0 \cdot q^2}$

$P(W) - 2pq\, P(W) = p^2$

$P(W) = \boxed{\dfrac{p^2}{1-2pq}}$

This is the same as gambler's ruin with $N=4$, $i=2$ and $p_0 = 0$, $p_N = 1 \Rightarrow \begin{cases} \dfrac{1-\left(\frac{q}{p}\right)^2}{1-\left(\frac{q}{p}\right)^4} & \text{if } p \neq q \\ \frac{1}{2} & \text{if } p = q \end{cases}$ (from worked example/problem in SP3)

This is the same as PART A: $\dfrac{\frac{p^2-q^2}{p^2}}{\frac{p^4-q^4}{p^4}} = \boxed{\dfrac{p^2}{p^2+q^2}} \stackrel{?}{=} \dfrac{p^2}{1-2pq}$

because $p+q=1$: $p^2+q^2 \stackrel{?}{=} 1-2pq$

$(p+q)^2 \stackrel{?}{=} 1$

$1 \stackrel{\checkmark}{=} 1$

. $P(A \text{ wins}) = \sum_{k=1}^{\infty} P(A \text{ wins} | \text{win in rd. } k)\, P(\text{win in rd. } k) = p_1 + (1-p_1)(1-p_2)p_1 + (1-p_1)^2(1-p_2)^2 p_1 + \ldots$

(LOTP)

$= p_1 \left(1 + (1-p_1)(1-p_2) + (1-p_1)^2(1-p_2)^2 + (1-p_1)^3(1-p_2)^3 + \ldots\right)$

inf. geo ies for $|<1$)

Let $x = (1-p_1)(1-p_2) \Rightarrow P(A \text{ wins}) = p_1 \left(x^0 + x^1 + x^2 + \ldots\right)$

$= p_1 \cdot \dfrac{1}{1-x}$

(Reintroduce RHS) $= \boxed{p_1 \cdot \dfrac{1}{1-(1-p_1)(1-p_2)}}$

*A-

[HW3]

1
1 1
1 2 1
1 3 3 1
1 4 6 4 1
1 5 10 10 5 1
1 6 15 20 15 6 1

$$p_0 = 1$$

$$p_1 = \frac{1}{6} = \left(\frac{1}{6}\right)\overbrace{(1)}^{p_0}$$

$$p_2 = \left(\frac{1}{6}\right)^2 + \frac{1}{6} = \left(\frac{1}{6}\right)\left(\overbrace{\frac{1}{6}}^{p_1} + \overbrace{1}^{p_0}\right)$$

$$p_3 = \left(\frac{1}{6}\right)^3 + 2\left(\frac{1}{6}\right)^2 + \frac{1}{6} = \left(\frac{1}{6}\right)\left[\overbrace{\left(\frac{1}{6}\right)^2 + \frac{1}{6}}^{p_2} + \overbrace{\frac{1}{6}}^{p_1} + \overbrace{1}^{p_0}\right]$$

$$p_4 = \left(\frac{1}{6}\right)^4 + 3\left(\frac{1}{6}\right)^3 + 3\left(\frac{1}{6}\right)^2 + \frac{1}{6} = \frac{1}{6}\left[\underbrace{\left(\frac{1}{6}\right)^3 + 2\left(\frac{1}{6}\right)^2 + \frac{1}{6}}_{p_3} + \underbrace{\left(\frac{1}{6}\right)^2 + \frac{1}{6}}_{p_2} + \underbrace{\frac{1}{6}}_{p_1} + \underbrace{1}_{p_0}\right]$$

We see a pretty obvious pattern emerge. We need $p_0 = 1$ and the equation to be true $\forall n \in \mathbb{Z}^+$. This can't continue forever since some n are so large a minimum number of rolls are needed.

Taking ~~5~~ 7. $p_7 \neq \frac{1}{6}(\underbrace{p_6 + p_5 + p_4 + p_3 + p_2 + p_1}_{\frac{1}{6}(p_5 + p_4 + p_3 + p_2 + p_1 + p_0)} + p_0)$

will always need at least 2 rolls, at most 7 rolls

when multiplied by $\frac{1}{6}$, yields $p_1 = \frac{1}{6}$, or probability of getting 7 in one roll $\Rightarrow$ invalid!

this is fine since there is a way to get a 6 from one roll; $\frac{1}{6} \cdot p_0 = \frac{1}{6}$.

so keeping w/ pattern does

$$p_7 \stackrel{?}{=} \frac{1}{6}(p_6 + p_5 + p_4 + p_3 + p_2 + p_1)$$

$$\boxed{= \left(\frac{1}{6}\right)^7 + 6\left(\frac{1}{6}\right)^6 + 15\left(\frac{1}{6}\right)^5 + 20\left(\frac{1}{6}\right)^4 + 15\left(\frac{1}{6}\right)^3 + 6\left(\frac{1}{6}\right)^2 + \frac{1}{6}}$$

$$= \frac{1}{6}\left(\cancel{\left(\frac{1}{6}\right)^6} + \cancel{6\left(\frac{1}{6}\right)^5} + \cancel{15\left(\frac{1}{6}\right)^4} + \cancel{20\left(\frac{1}{6}\right)^3} + \cancel{15\left(\frac{1}{6}\right)^2} + \cancel{6\left(\frac{1}{6}\right)}\right)$$

$$\stackrel{?}{=} \frac{1}{6}\Big(\underbrace{\cancel{\tfrac{1}{6}}}_{p_1} + \underbrace{\cancel{\tfrac{1}{6}} + \cancel{\left(\tfrac{1}{6}\right)^2}}_{p_2} + \underbrace{\cancel{\left(\tfrac{1}{6}\right)^3} + \cancel{2\left(\tfrac{1}{6}\right)^2} + \cancel{\tfrac{1}{6}}}_{p_3} + \underbrace{\cancel{\left(\tfrac{1}{6}\right)^4} + \cancel{3\left(\tfrac{1}{6}\right)^3} + \cancel{3\left(\tfrac{1}{6}\right)^2} + \cancel{\tfrac{1}{6}}}_{p_4} +$$

$$\underbrace{\cancel{\left(\tfrac{1}{6}\right)^5} + \cancel{4\left(\tfrac{1}{6}\right)^4} + \cancel{6\left(\tfrac{1}{6}\right)^3} + \cancel{4\left(\tfrac{1}{6}\right)^2} + \cancel{\left(\tfrac{1}{6}\right)}}_{p_5} +$$

$$\underbrace{\cancel{\left(\tfrac{1}{6}\right)^6} + \cancel{5\left(\tfrac{1}{6}\right)^5} + \cancel{10\left(\tfrac{1}{6}\right)^4} + \cancel{10\left(\tfrac{1}{6}\right)^3} + \cancel{5\left(\tfrac{1}{6}\right)^2} + \cancel{\left(\tfrac{1}{6}\right)}}_{p_6}\Big)$$

Yes.

$$\boxed{\begin{array}{l} p_n = \frac{1}{6}(p_{n-1} + p_{n-2} + p_{n-3} + p_{n-4} + p_{n-5} + p_{n-6}), \ \forall n \in \mathbb{Z}^+ \\ p_0 = 1 \\ p_k = 0 \text{ for } k < 0 \end{array}}$$

(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.

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).

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.*

   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 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.

ALMOST (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.

DUMB MISTAKE.

Hint for (c): Letting

$$a = \underbrace{\sum_{k \text{ even},\, k \geq 0} \binom{n}{k} p^k (1-p)^{n-k}}_{\text{undetected errors + correct msg.}} \text{ and } b = \underbrace{\sum_{k \text{ odd},\, k \geq 1} \binom{n}{k} p^k (1-p)^{n-k}}_{\text{detected errors}},$$

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.

7A. Undetected errors occur when 2, 4, 6, ... bits flip in transmission, including the parity check bit. e.g. $\underbrace{1110\boxed{1}}_{\text{sent}}$ ✓ $\xrightarrow{\times}$ $\underbrace{1111\boxed{2}}_{\text{received}}$ ✓ no error detected but last two bits flipped! Another e.g., $0000\boxed{0}$ ✓ $\xrightarrow{\times}$ $1111\boxed{0}$ ✓.

$$\boxed{\binom{5}{2}(0.1)^2(0.9)^3 + \binom{5}{4}(0.1)^4(0.9)}$$

7B. $\boxed{\sum_{i \geq 1} \binom{n}{2i} p^{2i} (1-p)^{n-2i}}$ or $\sum_{k \geq 2} \binom{n}{k} p^k (1-p)^{n-k}$ where k is even $\in \mathbb{Z}^+$

7C. $a+b = 1$ since $a+b = (p+q)^n$ where $q = 1-p$. $a-b = (q-p)^n$, also from Binomial Theorem. simplified $a-b = (1-p-p)^n = (1-2p)^n$. So we have $\begin{cases} a+b=1 \\ a-b=(1-2p)^n \end{cases}$, solving for a we get $a = \frac{1+(1-2p)^n}{2}$. Since a includes the case where the message received is exactly the message sent, we are left with $\boxed{\frac{1+(1-2p)^n}{2} - (1-p)^n}$