This homework is due at the beginning of class on Thursday January 26th, 2011. You are free to talk with each other and get help. However, you should write up your own solutions and understand everything that you write.

**Problem 1.** The 1987 World Series was tied at two games a piece before the St. Louis Cardinals won the fifth game. According to the Associated Press, “The numbers of history support the Cardinals and the momentum they carry. Whenever the series has been tied 2-2 the team that won the fifth game won the series 71% of the time.” If momentum is not a factor and each team has a 50% chance of winning each game (independently of the previous games), what is the probability that the Game 5 winner will win the series? (The World Series is a best of 7 series of games played between the two teams. That is, the first team to win a total of 4 games wins the series.)

SOLUTION: The winner of Game 5 must win one of the last two games. This can happen in one of two ways. Either the Game 5 winner wins Game 6 ($W_6$) or they lose Game 6 ($W_c_6$) and win Game 7 ($W_7$). Since the outcomes of the games (including the first 5) are assumed to be independent:

$$P(\text{Game 5 winner wins series}) = P(W_6) + P(W_c_6W_7)$$

$$= \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = 75\%.$$  

So, if anything, the ‘numbers of history’ suggest that the Game 5 winner is less likely to win the series than by sheer chance.

**Problem 2.** Suppose Alice and Bob play the following game. Alice flips three coins while Bob flips two. Alice wins if she has more Heads showing than Bob.

a) Find the probability that Alice wins.

SOLUTION: Let $X$ be the number of heads that Alice flips and $Y$ be the number of heads that Bob flips. We partition the outcome space according to the events $Y = 0$, $Y = 1$ and $Y = 2$, so

$$P(\text{Alice wins}) = P(X > Y) = P(Y = 0)P(X > Y | Y = 0) + P(Y = 1)P(X > Y | Y = 1) + P(Y = 2)P(X > Y | Y = 2).$$

Observe that $P(X > Y | Y = i) = P(X > i | Y = i)$, and these events are independent (they depend on separate sets of coins), so $P(X > Y | Y = i) = P(X > i)$ and

$$P(\text{Alice wins}) = P(Y = 0)P(X > 0) + P(Y = 1)P(X > 1) + P(Y = 2)P(X > 2)$$

$$= \frac{17}{48} + \frac{11}{22} + \frac{11}{48}$$

$$= \frac{1}{2}$$

b) What is Alice’s probability of winning if she flips $n + 1$ coins and Bob flips $n$ coins? Is this surprising?
SOLUTION 1 (the hard way): Again, let \( X \) be the number of heads that Alice flips and \( Y \) be the number that Bob flips. We can do a similar (but trickier) calculation to the one above,

\[
P(\text{Alice wins}) = P(X > Y)
\]

\[
= \sum_{i=0}^{n} P(Y = i)P(X > i)
\]

\[
= \sum_{i=0}^{n} \binom{n}{i} \left(\frac{1}{2}\right)^n \sum_{j=i+1}^{n+1} \binom{n+1}{j} \left(\frac{1}{2}\right)^{n+1}
\]

\[
= \left(\frac{1}{2}\right)^{2n+1} \sum_{i=0}^{n} \binom{n}{i} \sum_{j=i+1}^{n+1} \binom{n+1}{j}.
\]

At this point we observe that:

\[
\sum_{i=0}^{n} \left(\binom{n}{i} \sum_{j=0}^{n+1} \binom{n+1}{j} \right) = \sum_{i=0}^{n} \left(\binom{n}{n-i} \sum_{j=n+1}^{n+1} \binom{n+1}{n+1-j} \right)
\]

\[
= \sum_{k=0}^{n} \binom{n}{k} \sum_{\ell=0}^{k} \binom{n+1}{\ell}
\]

where in the second line we substituted \( k = n - i \) and \( \ell = n + 1 - j \) and reversed the order of summation (for instance, \( j = i + 1 \) corresponds to \( \ell = n + 1 - (i + 1) = n - i = k \), so the first summand in the second sum is now the last). Together with the observation that

\[
\sum_{i=0}^{n} \binom{n}{i} \sum_{j=0}^{n+1} \binom{n+1}{j}
\]

\[
= 2^n \cdot 2^{n+1} = 2^{2n+1}
\]

we see that

\[
\sum_{i=0}^{n} \binom{n}{i} \sum_{j=i+1}^{n+1} \binom{n+1}{j} = \frac{1}{2}2^{2n+1} = 2^{2n}
\]

so we have our answer:

\[
P(\text{Alice wins}) = \frac{1}{2}
\]

but that was tedious! Instead, we could use more probabilistic reasoning.

SOLUTION 2 (the ‘easy’ way): Suppose that Alice first flips \( n \) coins, then flips one more for a total of \( n + 1 \) coin flips. Let \( X_n \) be the number of heads that Alice has on her first \( n \) flips. If \( X_n < Y \) then Alice cannot win, because even if her last coin is heads she can at best tie with Bob. If \( X_n > Y \) then Alice has won, regardless of the outcome of her last coin flip. If \( X_n = Y \), then Alice wins if and only if her last coin shows heads. The quantities \( X_n \) and \( Y \)
have the same distribution and are independent of one another (each is the number of heads showing in \( n \) independent coin flips). So by symmetry

\[ P(X_n > Y) = P(X_n < Y). \]

Letting \( W \) be the event that Alice wins, we compute

\[
P(W) = P(X_n < Y)P(W|X_n < Y) + P(X_n > Y)P(W|X_n > Y) + P(X_n = Y)P(W|X_n = Y)
\]

\[ = P(X_n < Y) \cdot 0 + P(X_n > Y) \cdot 1 + P(X_n = Y) \cdot \frac{1}{2}
\]

\[ = \frac{1}{2} [2P(X_n > Y) + P(X_n = Y)]
\]

\[ = \frac{1}{2} [P(X_n < Y) + P(X_n > Y) + P(X_n = Y)]
\]

\[ = \frac{1}{2}
\]

While this may seem surprising at first, the last argument shows that the game is either determined by Alice’s first \( n \) flips (and Bob’s flips) in favor of either player with equal probability, or by Alice’s last flip, which is Heads with probability 1/2.

**Problem 3** (p.31 #5 a, d). Let \( \Omega \) be the sample space corresponding to three tosses of a coin. Give a verbal description of the following events:

a) \( A = \{HHH, HHT, HTH, HTT\}; \)

SOLUTION: Each of the outcomes in \( A \) has a heads as the first toss, and these are all the outcomes in \( \Omega \) in which the first toss is a heads. Hence \( A= \)“the first toss lands heads.”

b) \( B = \{HHH, HHT, HTH, THH\}; \)

SOLUTION: Each of outcomes in \( B \) has at least two heads, and each outcome in \( \Omega \) with at least two heads appears in \( B \). Therefore \( B= \)”At least two out of the three coins land heads.”

c) Show that \( A \) and \( B \) are not independent.

SOLUTION: \( P(A) = P(B) = \frac{1}{2} \), but \( P(A \cap B) = \frac{3}{8} \neq \left( \frac{1}{2} \right)^2 \)

d) Find a third event, \( C \), so that \( P(C) = \frac{1}{2} \) and \( P(A \cap B \cap C) = P(A)P(B)P(C) \). So the events \( A, B, C \) are not independent, but they have this property.

SOLUTION: There are a few answers that work here. Any event that has exactly one of the outcomes \( HHH, HHT, \text{ or } HTH \) and any three other outcomes will suffice. For example, \( C = \{HHT, HTT, THT, TTT\} \), which is the event that “The last toss lands tails”, will work. Clearly \( P(C) = \frac{1}{2} \), and we can easily check that \( P(A \cap B \cap C) = \frac{1}{8} = \left( \frac{1}{2} \right)^3 \).
Problem 4 (p.31 #11). Let $A, B,$ and $C$ be three events. Use the inclusion-exclusion formula for two events to derive the following inclusion-exclusion formula for three events:

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC).$$

SOLUTION: We have $A \cup B \cup C = (A \cup B) \cup C$. We apply the inclusion-exclusion formula for two events to the events $(A \cup B)$ and $C$:

$$P[(A \cup B) \cup C] = P(A \cup B) + P(C) - P[(A \cup B) \cap C]$$

Again, by the inclusion-exclusion formula applied to $A \cup B$, we have

$$P(A \cup B) = P(A) + P(B) - P(AB)$$

Also, $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$, so we apply inclusion-exclusion to these two sets:

$$P[(A \cap B) \cap C] = P[(A \cap C) \cup (B \cap C)] = P(AC) + P(BC) - P(ABC)$$

(Recall that $(A \cap C) \cap (B \cap C) = A \cap B \cap C = ABC$). Combining these, we get

$$P(A \cup B \cup C) = P[(A \cup B) \cup C]$$

$$= P(A \cup B) + P(C) - P[(A \cup B) \cap C]$$

$$= P(A) + P(B) - P(AB) - P[(A \cup B) \cap C]$$

$$= P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC)$$

Problem 5. How can 5 black and 5 white balls be put into two urns to maximize the probability that a white ball is drawn when we draw from a randomly chosen urn?

SOLUTION: Put one white ball in the first urn and the other nine balls in the second urn. This gives a probability of $(1/2) \cdot 1 + (1/2) \cdot (4/9) = 13/18$ of drawing a white ball. In fact, this solution is best if there are $n$ black balls and $n$ white balls, and gives probability $(1/2) \cdot 1 + (1/2) \cdot \frac{n-1}{2n-1} = \frac{3n-2}{4n-2}$ of drawing a white ball. To prove that we can do no better, let $W = \{\text{draw white ball}\}$, $1 = \{\text{pick urn 1}\}$ and $2 = \{\text{pick urn 2}\}$ so that

$$P(W) = P(1)P(W|1) + P(2)P(W|2) = \frac{1}{2} [P(W|1) + P(W|2)].$$

Let us consider two cases. In the first case, we suppose that each urn has the same number of black balls and white balls (e.g. 3 black and 3 white in urn 1, and 2 black and 2 white in urn 2). In this case, $P(W|1) = P(W|2) = \frac{1}{2}$ so $P(W) = \frac{1}{2}$, but we have already determined that the strategy given above does better.

In the second case, we suppose that one of the urns has (strictly) more white balls than black balls. Without loss of generality, let’s suppose that this is urn 2. This means that the number of black balls in urn 1 (call this $b$) must be strictly greater than the number of white balls in urn 1 (call this $w$). That is, $b \geq w + 1$. Therefore:

$$P(W|1) = \frac{w}{b+w} \leq \frac{w}{2w+1}.$$
Now, observe that the expression on the right side of the above inequality is an increasing function of \( w \) (to see this take a derivative to get \( \frac{1}{(2w+1)^2} > 0 \)), so

\[
P(W|1) \leq \frac{w}{2w + 1} \leq \frac{n - 1}{2n - 1}
\]

since \( w \leq n - 1 \) (we must put at least one ball in urn 2 by our assumption). Note that our strategy of putting exactly one (white) ball in urn 2 satisfies the above inequality with equality. Also, clearly, \( P(W|2) \leq 1 \), and this is also satisfied by equality for our strategy, which implies that no other strategy does better than this one.

**Problem 6** (p.45 #1). In a particular population of men and women, 92% of women are right-handed, and 88% of men are right-handed. Indicate whether each of the statements below is (i) true, (ii) false, or (iii) can’t be decided from the information given.

a) The overall proportion of right-handers in the population is exactly 90%.

   SOLUTION: Cannot be determine from the information given. Observe that we are given \( P(RH|M) \) and \( P(RH|F) \), where \( RH = \) right-handed, \( M = \) male, \( F = \) Female. In order to determine the proportion of right-handers in the population, however, we need to compute \( P(RH) \). To do this, we rely on the fact that males and females completely partition the space:

   \[
P(RH) = P(RH \cap M) + P(RH \cap F) = P(RH|M)P(M) + P(RH|F)P(F) = .88P(M) + .92P(F)
\]

   We see that we can’t determine \( P(RH) \) without knowing the fraction of males and females.

b) The overall proportion of right-handers is between .88 and .92.

   SOLUTION: True. From above, we know that

   \[
P(RH) = .88P(M) + .92P(F)\]

   

   \[P(RH) = .88P(M) + .92(1 - P(M))\]

   \[P(RH) = .92 - 0.04P(M)\]

   Since \( P(M) \) is between 0 and 1, we see that the most \( P(RH) \) can be is 0.92, and the least it can be is .88.

c) If the sex ratio in the population is 1 to 1, then the proportion of right-handers is 90%.

   SOLUTION. True, because if \( P(M) = P(F) = .5 \), then

   \[P(RH) = (0.88)(.5) + (.92)(.5) = .9.\]

   d) If the proportion of right-handers is 90%, then the sex ratio must be 1 to 1.
SOLUTION. True, because

\[ P(RH) = 0.88P(M) + 0.92P(F) \]

\[ 0.9 = 0.88P(M) + 0.92(1 - P(M)) \]

\[ 0.9 = 0.92 - 0.04P(M) \]

\[ \Rightarrow 0.02 = 0.04P(M) \]

\[ \Rightarrow 0.5 = P(M) \] \hspace{1cm} (2)

e) If there are at least three times as many women as men in the population, then the overall
population of right handers is at least 91%.

SOLUTION. True. If there are at least three times as many women as men, then the fraction
of women is at least 0.75, and the fraction of men is at most 0.25. We already know that

\[ P(RH) = 0.92 - 0.04P(M) \]

This is a decreasing function of \( P(M) \), and thus it is smallest when \( P(M) \) is largest, i.e. when
\( P(M) \) is at most 0.25; therefore, \( P(RH) \geq 0.92 - 0.01 = 0.91 \).

Problem 7. (p. 46, #11) Assume that identical twins are always of the same sex, equally likely
boys or girls. Assume that for fraternal twins, the firstborn is equally likely to be a boy or a girl, and
so is the secondborn, independently of the first. Assume that proportion \( p \) of twins are identical,
and \( q = 1 - p \) of twins are fraternal. Find formulae in terms of \( p \) for the following probabilities:

a) \( P(\text{both boys}) \)

SOLUTION. Our sample space is the space of twins, which is partitioned by the sets \( I = \{\text{Identical twins}\} \) and \( F = \{\text{Fraternal twins}\} \). Let \( BB = \{\text{both boys}\} \). Then

\[ P(BB) = P(BB \cap I) + P(BB \cap F) \]

\[ = P(BB|I)P(I) + P(BB|F)P(F) \]

\[ = 0.5p + 0.25(1 - p) = \frac{1 + p}{4} \] \hspace{1cm} (3)

b) \( P(\text{firstborn boy and secondborn girl}) \)

SOLUTION. Let \( BG = \text{firstborn boy, secondborn girl} \). Since identical twins are always of
the same gender, \( P(BG \cap I) = 0 \). We have

\[ P(BG) = P(BG \cap I) + P(BG \cap F) \]

\[ = 0 + P(BG|F)P(F) \]

\[ = 0.25(1 - p) = \frac{(1 - p)}{4} \] \hspace{1cm} (4)

c) \( P(\text{secondborn girl}|\text{firstborn boy}) \)
SOLUTION. We need to first calculate $P($firstborn boy$)$. Let $1B=$firstborn boy. We have

$$P(1B) = P(1B \cap I) + P(1B \cap F)$$
$$= P(1B|I)P(I) + P(1B|F)P(F)$$
$$= 0.5p + 0.5(1 - p)$$
$$= 1/2$$

(5)

Let $2G =$ secondborn girl. We have

$$P(2G \cap 1B) = P(2G \cap 1B|F)P(F) + P(2G \cap 1B|I)P(I)$$
$$= 0.25(1 - p) + 0p = 0.25(1 - p)$$

(6)

and therefore

$$P(2G|1B) = \frac{P(2G \cap 1B)}{P(1B)} = \frac{0.25(1-p)}{0.5} = \frac{1}{2}(1-p)$$

(7)

d) $P($secondborn girl$|firstborn girl$)

SOLUTION. The methods here are identical to those of the previous part, except that fraternal twins have a positive probability (1/4, to be exact) of being both girls. Let $1G$ and $2G$ be the events corresponding to firstborn girl and secondborn girl, respectively. We have

$$P(2G|1G) = \frac{P(2G \cap 1G)}{P(2G)}$$

$$P(2G) = P(2G|I)P(I) + P(2G|F)P(F) = 0.5p + 0.5(1 - p) = 0.5$$

$$P(2G \cap 1G) = P(2G \cap 1G|I)P(I) + P(2G \cap 1G|F)P(F)$$
$$= 0.5p + 0.25(1 - p) = 0.25 + 0.25p$$

$$\Rightarrow P(2G|1G) = \frac{0.25 + 0.25p}{0.5} = \frac{1}{2}(1 + p)$$

Problem 8 (p.53 #2). An urn contains 4 white and 6 black balls. A ball is chosen at random and its color noted. The ball is then replaced, along with 3 more balls of the same color. Then another ball is drawn at random from the urn.

a) Find the probability that the second ball drawn is white.

SOLUTION. Let $2w$ denote the event that the second ball is white, and $1w$ and $1b$ the events that the first balls are white or black, respectively. We have

$$P(2w) = P(2w \cap 1w) + P(2w \cap 1b) = P(2w|1w)P(1w) + P(2w|1b)P(1b)$$

and $P(2w|1w) = 7/13$, $P(2w|1b) = 4/13$. Further, $P(1w) = 4/10$, $P(1b) = 6/10$.


7
b) Given the second ball is white, what is the probability that the first is black? By definition,

\[ P(1b|2w) = \frac{P(2w \cap 1b)}{P(2w)} \]

and \( P(2w) = \frac{52}{130} \) as computed in the previous part, so we need to compute \( P(2w \cap 1b) = P(2w|1b)P(1b) \). But this is also computed above and is \( \frac{24}{130} \). Therefore \( P(1b|2w) = \frac{24}{52} = \frac{6}{13} \).

c) Suppose the original contents of the urn are \( w \) white balls and \( b \) black balls, and that after a ball is drawn from the urn, it is replaced along with \( d \) more balls of the same color. Show that the probability that the second ball drawn is white is \( \frac{w}{w+b} \), i.e. that this probability does not depend on \( d \).

**SOLUTION.** Here again, we have

\[
P(2w) = P(2w \cap 1w) + P(2w \cap 1b) \\
= P(2w|1w)P(1w) + P(2w|1b)P(1b) \\
= \left( \frac{w + d}{w + b + d} \right) \left( \frac{w}{w + b} \right) + \left( \frac{w}{w + b + d} \right) \left( \frac{b}{w + b} \right) \\
= \frac{w(w + d) + wb}{(w + b)(w + b + d)} \\
= \frac{w^2 + wd + wb}{(w + b)(w + d + b)} \\
= \frac{w(w + d + b)}{(w + b)(w + d + b)} = \frac{w}{w + b}
\]

**Problem 9** (p.54 #5). The fraction of people who have a certain disease is 0.01. In a diagnostic test, a healthy person has a 0.05 chance of being falsely diagnosed as having the disease, and a sick person has a 0.2 chance of being falsely diagnosed as not having the disease.

a) What is the probability a test result is positive?

**SOLUTION.** Let \( Pos=\text{positive}, D=\text{has disease}, D^c=\text{does not have disease}. \)

\[
P(Pos) = P(Pos|D)P(D) + P(Pos|D^c)P(D^c) \\
= (.8)(.01) + (.05)(.99)
\]

b) What is the probability of having the disease but being diagnosed as healthy?

**SOLUTION.** The probability in question is

\[
P(D \cap Neg) = P(D)P(Neg|D) = (.01)(.2) = .002
\]

c) What is the probability that the person is correctly diagnosed and is healthy?

**SOLUTION.** This is \( P(Neg \cap D^c). \) We have

\[
P(Neg \cap D^c) = P(Neg|D^c)P(D^c) \\
= (.95)(.99)
\]
d) Suppose the test result is positive. What is the probability the person actually has the disease?

SOLUTION. This is \( P(D|Pos) \). We have

\[
P(D|Pos) = \frac{P(Pos \cap D)}{P(Pos)} = \frac{(0.8)(0.01)}{(0.8)(0.01) + (0.05)(0.99)}
\]

e) Do these probabilities admit a long-run frequency interpretation?

SOLUTION. Yes, they do, for instance as the long-run frequency that those who test positive actually have the disease in many repeated testings of people.

Problem 10 (p.54–55 #7). Consider a game in which balls are divided among three boxes. First, I choose a box from among the three boxes, and then I choose ball at random from that particular box. Box 1 contains 1 white and 1 black ball; Box 2 contains 2 white and 1 black ball; Box 3 contains 3 white and 1 black ball. After you see what color ball I pick, you guess which Box I chose from, and the game is played over and over. Your objective is to guess the box correctly as often as possible.

a) Suppose you know that I pick a box at random with probability \((1/3, 1/3, 1/3)\). Your strategy is to guess the box with highest posterior probability given the observed color. What fraction of the time are you likely to be correct?

SOLUTION. Let \( B_1, B_2, B_3 \) represent the events that Box 1, Box 2, or Box 3 were chosen as the box from which to draw. We can compute:

\[
P(\text{Box 1}|W) = \frac{P(W|B_1)P(B_1)}{P(W)} = \frac{\frac{1}{2}}{\frac{1}{3} \left[ \frac{1}{2} + \frac{2}{3} + \frac{3}{4} \right]} = \frac{6}{23}
\]

\[
P(\text{Box 2}|W) = \frac{P(W|B_2)P(B_2)}{P(W)} = \frac{\frac{2}{3}}{\frac{1}{3} \left[ \frac{1}{2} + \frac{2}{3} + \frac{3}{4} \right]} = \frac{8}{23}
\]

\[
P(\text{Box 3}|W) = \frac{P(W|B_3)P(B_3)}{P(W)} = \frac{\frac{3}{4}}{\frac{1}{3} \left[ \frac{1}{2} + \frac{2}{3} + \frac{3}{4} \right]} = \frac{9}{23}
\]

Hence the box with the highest posterior probability given a white ball is Box 3.

Similarly, we can compute the analogous probabilities when the observed ball is black, and we get

\[
P(\text{Box 1}|B) = \frac{P(B|B_1)P(B_1)}{P(B)} = \frac{\frac{1}{2}}{\frac{1}{3} \left[ \frac{1}{2} + \frac{2}{3} + \frac{3}{4} \right]} = \frac{6}{13}
\]

\[
P(\text{Box 2}|B) = \frac{P(B|B_2)P(B_2)}{P(B)} = \frac{\frac{2}{3}}{\frac{1}{3} \left[ \frac{1}{2} + \frac{2}{3} + \frac{3}{4} \right]} = \frac{4}{13}
\]

\[
P(\text{Box 3}|B) = \frac{P(B|B_3)P(B_3)}{P(B)} = \frac{\frac{3}{4}}{\frac{1}{3} \left[ \frac{1}{2} + \frac{2}{3} + \frac{3}{4} \right]} = \frac{3}{13}
\]

So the box with the highest posterior probability given a black ball is Box 1. If you adopt this
strategy, the fraction of time you will be correct is

\[
P(\text{Correct}) = P(\text{Correct} \cap \text{Guess Box 3}) + P(\text{Correct} \cap \text{Guess Box 1})
\]

\[
= P(\text{Correct}|\text{Guess Box 3})P(\text{Guess Box 3}) + P(\text{Correct}|\text{Guess Box 1})P(\text{Guess Box 1})
\]

You guess Box 3 if and only if the observed ball is white, and you guess Box 1 if and only if the observed ball is black. You will be correct only if, when the ball is white, the box is actually Box 3, and similarly when the ball is black, the box is actually Box 1. Hence the above probabilities become:

\[
P(\text{Correct}) = P(B3|W)P(W) + P(B1|B)P(B)
\]

\[
= \frac{9}{23} \cdot \frac{23}{36} + \frac{6}{13} \cdot \frac{13}{36} = \frac{5}{12}
\]

b) Can you do better given any other strategy?

SOLUTION. No. For any other strategy, you must choose a box to guess if the ball is white (call this \(B(W)\)) and a box to guess if the ball is black (call this \(B(B)\)). The long-run fraction of time you are correct will be

\[
P(\text{Correct in any strategy}) = P(B(W)|W)P(W) + P(B(B)|B)P(B)
\]  \hspace{1cm} (8)

Since the probabilities \(P(B)\) and \(P(W)\) are fixed, the best strategy comes from maximizing \(P(B(W)|W)\) and \(P(B(B)|B)\), i.e. choosing the boxes with the maximal posterior probability given each particular observed color.

c) Suppose you use the guessing strategy from (a), but I was in fact randomizing with probabilities \((1/2, 1/4, 1/4)\) instead. Now how would your strategy perform?

SOLUTION. There are different ways to interpret what this question is asking. First, suppose by we keep the strategy of guessing Box 3 if the ball is white and Box 1 if the ball is black. Now, however, the posterior probabilities \(P(B3|W)\) and \(P(B1|B)\) are different. The probability we’ll be right in this case is then

\[
P(B3|W)P(W) + P(B1|B)P(B)
\]

\[
= \frac{9}{48} + \frac{12}{48} = \frac{7}{16}
\]

Alternatively, you could use the strategy of choosing the box which had the highest posterior probability, which is Box 1 if the ball is white and Box 1 if the ball is black (that is, Box 1 no matter what the color of the ball). Here, too, we can compute the probability of being right in the long run:

\[
P(B1|W)P(W) + P(B1|B)P(B)
\]

\[
= \frac{1}{4} \cdot \frac{1}{4} + \frac{1}{4} \cdot \frac{1}{4} = \frac{1}{2}
\]
d) Suppose you knew I was randomizing with either set of prior probabilities. How would you guess which one, and what would you do?

SOLUTION. Over the long run, if the randomizing prior was \((1/3, 1/3, 1/3)\), then the strategy in (a) would be correct about \(5/12\) of the time; if, however, the strategy of always choosing Box 1 were used, it would be correct about \(1/3\) of the time. Hence, one possibility is to always choose Box 1 and see how often you are correct—if the long-run fraction of correct guesses is close to \(1/3\), then the underlying randomization is likely to be \((1/3, 1/3, 1/3)\). If you always choose Box 1 and the long run fraction of correct guesses is close to \(1/2\), then the underlying randomization is likely to be \((1/2, 1/4, 1/4)\). Once you learn which randomization was being used, you should switch to the appropriate strategy for that randomization (i.e. you should choose whichever box has the highest posterior probability given the observed color of the ball and the prior probabilities in the randomization).

Problem 11 (p.70 #2). Suppose a batter’s average (number of hits per at bat) is 0.300 over the season to date. What is the probability that the batter gets at least one hit in the next: (a) two at bats; (b) three at bats; (c) \(n\) at bats? What assumptions are you making?

SOLUTION. If we assume that the outcomes of each time the batter goes to bat are independent and identical, and if we assume that the probability of getting a hit on each trial is 0.3 (the long-run frequency of hits over many trials in the season to date), then the probability of at least one hit in a certain number of times at bat is binomial.

So part (a) is
\[
P(\text{at least one hit in 2 at bats}) = 2(0.3)(0.7) + (0.3)^2.
\]
Alternatively, this is
\[
1 - P(\text{no hits in 2 at bats}) = 1 - (0.7)^2.
\]
Next, part (b) is
\[
P(\text{at least one hit in 3 at bats}) = 3(0.3)(0.7)^2 + 3(0.7)(0.3)^2 + (0.3)^3.
\]
Alternatively, this is
\[
1 - P(\text{no hits in 3 at bats}) = 1 - (0.7)^3.
\]
Finally part (c) is
\[
P(\text{at least one hit in } n \text{ at bats}) = 1 - P(\text{no hits}) = 1 - (0.7)^n.
\]

Problem 12 (p.71 #5). Suppose you are one of \(n\) students in the class, and the birthday of each person is equally likely to be any one of the 365 days of the year.

a) What is the chance that at least one other student has the same birthday as yours?

SOLUTION. Let \(A\) = at least one other student has the same birthday as you. Then
\[
P(A) = 1 - P(A^c),
\]
and \(A^c\) is the event that no other student has the same birthday as yours. This means all \(n - 1\) other students must be born on some day other than your birthday. Assuming the birthdays of students are independent,
\[
P(A^c) = \left[\frac{364}{365}\right]^{(n-1)}
\]

and hence \( P(A) = 1 - \left[ \frac{364}{365} \right]^{(n-1)} \).

b) How large does the class have to be for this probability to be at least 0.5?

SOLUTION. We want to solve for \( n \) in the equation

\[
\left[ \frac{364}{365} \right]^{(n-1)} \leq 0.5
\]

Taking logs (remember that the logarithm of a number between 0 and 1 is negative, and that multiplying or dividing through an inequality by a negative number reverses the direction of the inequality!) and solving for \( n \), we get

\[
n > 1 + \frac{\ln(0.5)}{\ln(364/365)} \approx 253.65
\]

c) How does this problem differ from the birthday problem discussed in class?

SOLUTION. In the birthday problem, the event of interest was the event that no two students shared the same birthday—here it’s just the event that no one shares a birthday with you. It may seem surprising that we needed so few people to find a pair of people with matching birthdays, but we need so many people to find a birthday that matches yours. In fact, these two findings are consistent: Since some people among your \( n-1 \) classmates likely have the same birthdays, this leaves fewer unique birthdates to match with yours.

We can also generalize this to the case where there are \( N \) possible birthdays rather than 365. We obtain an asymptotic relationship between \( n \) and \( N \) such that \( P(A) \approx 1/2 \) by using the approximation \( \ln(1 + x) \approx x \), so

\[
n \approx 1 + \frac{\ln(0.5)}{\ln(1 - 1/N)} \sim N \ln 2.
\]

Therefore, the class size required to have a 50% chance of another birthday being the same as yours scales linearly with \( N \), the number of possible birthdays, while the class size required to have a 50% chance of some pair of students with the same birthday scales like the square root of \( N \).

**Problem 13.** Suppose that the birthday of each of three people is equally likely to be any one of the 365 days of the year, independently of the others. Let \( B_{ij} \) denote the event that person \( i \) has the same birthday as person \( j \).

a) Are the events \( B_{12} \) and \( B_{23} \) independent?

SOLUTION. Yes. Note that

\[
P(B_{12} \cap B_{23}) = P(1, 2, 3 \text{ all have the same birthday})
\]

12
Now, order the days in the year from 1 to 365.

\[
P(1, 2, 3 \text{ all have the same birthday})
\]

\[
= \sum_{k=1}^{365} P(1, 2, 3 \text{ each have birthdate } k)
\]

\[
= \sum_{k=1}^{365} \left[ \frac{1}{365} \right]^3 = \left[ \frac{1}{365} \right]^2
\]

By the same reasoning, \( P(B_{12}) = P(B_{23}) = \frac{1}{365} \), so indeed

\[
P(B_{12} \cap B_{23}) = \left[ \frac{1}{365} \right]^2 = P(B_{12})P(B_{23}).
\]

b) Are the events \( B_{12}, B_{13}, \) and \( B_{23} \) independent?

SOLUTION. No. Note that

\[
P(B_{12} \cap B_{23} \cap B_{13}) = P(1, 2, 3 \text{ all have the same birthday})
\]

which we already computed above as being equal to \( \left[ \frac{1}{365} \right]^2 \). This is not the same as \( P(B_{12})P(B_{13})P(B_{23}) = \left[ \frac{1}{365} \right]^3 \).

c) Are the events \( B_{12}, B_{23}, B_{13} \) pairwise independent?

SOLUTION. Yes, and you can verify this by repeating the calculation in part (a) for any pair of the three events.