Staff Solutions to Conflict Final 2

Problem 1 (Expectation) (6 points).
I have twelve cards:

\[1\ 1\ 2\ 2\ 3\ 3\ 4\ 4\ 5\ 5\ 6\ 6\]

I shuffle them and deal them in a row. For example, I might get:

\[1\ 2\ 3\ 3\ 4\ 6\ 1\ 4\ 5\ 5\ 2\ 6\]

What is the expected number of adjacent pairs with the same value? In the example, there are two adjacent pairs with the same value, the 3’s and the 5’s.

We can award partial credit only if you show your work.

Solution. Consider an adjacent pair. The left card matches only one of the other 11 cards, which is equally likely to be in any of the 11 other positions. Therefore, the probability that an adjacent pair matches is \(\frac{1}{11}\). Since there are 11 adjacent pairs, the expected number of matches is \(11 \times \frac{1}{11} = 1\) by linearity of expectation.

Problem 2 (Variance) (8 points).
Let \(R\) be a positive integer valued random variable.

(a) If \(\text{Ex}[R] = 2\), how large can \(\text{Var}[R]\) be?

Solution. The variance of \(R\) is unbounded. For example, suppose \(R_n\) is a random variable that equals \(n + 1\) with probability \(\frac{1}{n}\) and equals 1 otherwise. So \(\text{Ex}[R_n] = \frac{(n - 1)}{n} + n + 1 / n = 2\) as required. However,

\[
\text{Var}[R_n] = \text{Ex}[R_n^2] - \text{Ex}^2[R_n] = \frac{n^2}{n} - 2 > n + 2.
\]

So \(\text{Var}[R_n]\) grows unboundedly.

In fact, there are positive integer valued random variables with expectation 2 and infinite variance (see Problem 19.28).

(b) How large can \(\text{Ex}[1/R]\) be?

Solution. \(\text{Ex}[1/R]\) is maximized at 1 when \(R = 1\) with probability 1. Intuitively, this can be seen by trying to shift some of the probability mass away from 1. No matter how you do it, you will always wind up with an expectation that is less than 1.

Creative Commons 2013, Eric Lehman, F Tom Leighton, Albert R Meyer.
Formally, we can prove this as follows. Suppose $\Pr[R = 1] = 1 - \delta$ for some $\delta > 0$. Then

$$\text{Ex}[R] = 1 \cdot (1 - \delta) + \sum_{i=2}^{\infty} \frac{1}{n} \cdot \Pr[R = n]$$

$$\leq 1 \cdot (1 - \delta) + \sum_{i=2}^{\infty} \frac{1}{2} \cdot \Pr[R = n]$$

$$\leq 1 \cdot (1 - \delta) + \frac{1}{2} \sum_{i=2}^{\infty} \Pr[R = n]$$

$$\leq 1 \cdot (1 - \delta) + \frac{1}{2} \cdot \delta$$

$$\leq 1 - \delta \cdot \frac{3}{2}.$$  

So, for any positive $\delta$, $\text{Ex}[R]$ is less than 1. Therefore, the given distribution maximizes $\text{Ex}[1/R]$.  

(c) If $R \leq 2$, that is $R$ takes only the values 1 and 2, how large can $\text{Var}[R]$ be? 

**Solution.**

$$\frac{1}{4}$$

We know $\text{Var}[R] = \text{Var}[R - 1]$ and $R - 1$ is a Bernoulli variable. If $\Pr[R = 1] = p$, then we know the variance is $p(1 - p)$ which is maximum for $p = 1/2$ (see Problem 19.15). 

**Problem 3 (Modular Sum of Digits) (8 points).**

The sum of the digits of the base 10 representation of an integer is congruent modulo 9 to that integer. For example,

$$763 \equiv 7 + 6 + 3 \pmod{9}.$$ 

This is not always true for the base 11 representation, however. For example,

$$(763)_{11} = 7 \cdot 11^2 + 6 \cdot 11 + 3 \equiv 3 \not\equiv 5 \equiv 7 + 6 + 3 \pmod{11}.$$ 

For exactly what integers $k \in (1, 10]$ is it true that the sum of the digits of the base 11 representation of every nonnegative integer is congruent modulo $k$ to that integer? (No explanation is required, but no part credit without an explanation.)

**Solution.**

$$2, 5, 10.$$ 

Summing the digits mod $k$ clearly works when $11 \equiv 1 \pmod{k}$. This is equivalent to $k \mid 11 - 1 = 10$. So the three factors of 10 are $k$’s that work.

To see why only these $k$’s work, just look at hex representation of 11, namely, the string 10. The digit-sum requirement means $11 \equiv 1 + 0 = 1 \pmod{k}$. 

■
Problem 4 (Induction) (12 points).\textbf{STAFF NOTE:} (a) 4 pts, (b) 5, (c) 3

Define the\textit{ Triple Fibonacci} numbers $T_0, T_1, \ldots$ recursively by the rules

\begin{align*}
T_0 &= T_1 := 3, \\
T_n &= T_{n-1} + T_{n-2} & \quad \text{(for $n \geq 2$).}
\end{align*}

(a) Prove that all Triple Fibonacci numbers are divisible by 3.

\textbf{Solution. \textit{Proof.}} We use strong induction on $n$ with induction hypothesis

\[ P(n) := 3 \mid T_n. \]

\textbf{Base case:} ($n = 0, 1$). $P(0)$ and $P(1)$ are true since $T_0 = T_1 = 3$ are both divisible by 3.

\textbf{Inductive step:} For $n \geq 2$, $T_n$ is defined by (1). By strong induction, we may assume that $T_{n-1}$ and $T_{n-2}$ are divisible by 3, and hence $T_n$ is the sum of two numbers that are divisible by three, and therefore is itself divisible by 3.

(b) Prove that the GCD of every pair of consecutive Triple Fibonacci numbers is 3.

\textbf{Solution. \textit{Proof.}} By induction on $n$ with induction hypothesis

\[ Q(n) := [\gcd(T_n, T_{n-1}) = 3]. \]

\textbf{Base case:} ($n = 1$). $P(1)$ holds since $\gcd(3, 3) = 3$.

\textbf{Inductive step:} To prove that $\gcd(T_{n+1}, T_n) = 3$, we use the recursive definition 1,

\[
\gcd(T_{n+1}, T_n) = \gcd(T_n, T_{n+1} - T_n) = \gcd(T_n, T_{n-1}) = 3
\]

(by induction hypothesis $Q(n)$).

(c) Express the generating function $T(x)$ for the Triple Fibonacci as a quotient of polynomials. (You do\textit{ not} have to find a formula for $[x^n]T(x)$.)

\textbf{Solution.}

\[
T(x)(1 - x - x^2) = T_0 + T_1x - 2T_0x = 3
\]

so

\[
T(x) = \frac{3}{1 - x - x^2}
\]

Problem 5 (counting poker high cards) (7 points).

In a standard 52-card deck (13 ranks and 4 suits), a hand is a 5-card subset of the set of 52 cards. Express the answer to each part as a formula using factorial, binomial, or multinomial notation.

(a) Let $H$ be the set of all hands.

What is $|H|$?
Solution. \( |H| = \binom{52}{5} \)

(b) Let \( H_{NP} \) be the set of all hands that does not include a pair, that is, no two card in the hand have the same rank.
What is \( |H_{NP}| \)?

Solution. \( |H_{NP}| = \binom{13}{5} \binom{4}{1}^5 \)

(c) Let \( H_S \) be the set of all hands that is a straight, i.e. the rank of the five cards are consecutive. The order of the ranks is \((A, 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, k, A)\), note that \( A \) is appears twice.
What is \( |H_S| \)?

Solution. \( |H_S| = \binom{10}{1} \binom{4}{1}^5 \)

(d) Let \( H_F \) be the set of all hands that is a flush, that is, the suit of the five cards are identical.
What is \( |H_F| \)?

Solution. \( |H_F| = \binom{13}{5} \binom{4}{1} \)

(e) Let \( H_{SF} \) be the set of all straight flush hands that is both a straight and a flush.
What is \( |H_{SF}| \)?

Solution. \( |H_{SF}| = \binom{10}{1} \binom{4}{1} \)

(f) Let \( H_{HC} \) be the set of all high card hands that is hands that do not include a pair, are not straights, and are not flushs.
What is \( |H_{HC}| \)?

Solution.

\[
|H_{HC}| = |H_{NP}| - |H_S| - |H_F| + |H_{SF}|
\]
\[
= \binom{13}{5} \binom{4}{1}^5 - \binom{10}{1} \binom{4}{1}^5
\]
\[
- \binom{13}{5} \binom{4}{1} + \binom{10}{1} \binom{4}{1}
\]
\[
= \left( \binom{4}{1}^5 - \binom{4}{1} \right) \left( \binom{13}{5} - \binom{10}{1} \right)
\]

Problem 6 (Tournament Probability) (10 points).\textbf{STAFF NOTE:} (a) 2 pts, (b) 3, (c) 3, (d) 2

The results of a round robin tournament in which every two people play each other and one of them wins can be modelled a tournament digraph—a digraph with exactly one edge between each pair of distinct vertices, but we’ll continue to use the language of players beating each other.
An $n$-player tournament is $k$-neutral for some $k \in [0, n)$, when, for every set of $k$ players, there is another player who beats them all. For example, being 1-neutral is the same as not having a “best” player who beats everyone else.

This problem shows that for any fixed $k$, if $n$ is large enough, there will be a $k$-neutral tournament of $n$ players. We will do this by reformulating the question in terms of probabilities. In particular, for any fixed $n$, we assign probabilities to each $n$-vertex tournament digraph by choosing a direction for the edge between any two vertices, independently and with equal probability for each edge.

(a) For any set $S$ of $k$ players, let $B_S$ be the event that no contestant beats everyone in $S$. Express $\Pr[B_S]$ in terms of $n$ and $k$.

**Solution.** The probability that a player outside $S$ beats everyone in $S$ is $\frac{1}{2^k}$. So the probability such a player did not beat everyone in the group is $1 - \frac{1}{2^k}$. There are $n - k$ players outside of the group, so

$$\Pr[B_S] = \left[1 - \left(\frac{1}{2}\right)^k\right]^{n-k}.$$  

(b) Let $Q_k$ be the event equal to the set of $n$-vertex tournament digraphs that are not $k$-neutral. Prove that

$$\Pr[Q_k] \leq \binom{n}{k} \alpha^{n-k},$$

where $\alpha := 1 - (1/2)^k$.

**Hint:** Let $S$ range over the size-$k$ subsets of players, so

$$Q_k = \bigcup_S B_S.$$

Use Boole’s inequality.

**Solution.**

$$\Pr[Q_k] = \Pr[\bigcup_S B_S] \quad \text{(hint)}$$

$$\leq \sum_S \Pr[B_S] \quad \text{(Boole’s Inequality)}$$

$$= |\{S \mid |S| = k\}| \cdot \Pr[B_S]$$

$$= \binom{n}{k} \alpha^{n-k} \quad \text{(part (a))}$$

(c) Conclude that if $n$ is enough larger than $k$, then $\Pr[Q_k] < 1$.

**Solution.**

$$\Pr[Q_k] \leq \binom{n}{k} \alpha^{n-k} \quad \text{(by part (b))}$$

$$= \frac{1}{\alpha^k} \binom{n}{k} \alpha^n$$

$$\leq (1/\alpha)^k n^k \alpha^n.$$
But for any fixed $k$, this last term approaches 0 as $n$ goes to infinity, so $Pr[Q_k]$ will be less than 1 for all large $n$.

To see why it approaches 0, note that $0 < \alpha < 1$, so $1/\alpha > 1$ and hence $n^k = o((1/\alpha)^n)$. Therefore,

$$n^k \alpha^n = \frac{n^k}{(1/\alpha)^n} \rightarrow 0.$$ 

\(\square\)

(d) Explain why the previous result implies that for every integer $k$, there is a $k$-neutral tournament.

**Solution.** Suppose $n$ is large enough that $Pr[Q_k] < 1$. Then $Pr[\bar{Q}_k] > 0$, which implies there must be at least one tournament graph (outcome) in $\bar{Q}_k$, that is, at least one tournament graph that is $k$-neutral. In fact, as $n$ grows, it follows that almost all $n$-player tournaments will be $k$-neutral.

For interest, some numbers that work are:

$$k = 1, n = 3; k = 2, n = 21; k = 3, n = 33; k = 4, n = 46;$$
$$k = 5, n = 59; k = 6, n = 72; k = 7, n = 85; k = 8, n = 98.$$ 

\(\square\)

**Problem 7 (coloring complete triangles) (12 points).**

**STAFF NOTE:** (a) 2, (b) 2, (c) 1, (d) 2, (e) 2, (f) 3

Let $K_n$ be the complete graph with $n$ vertices. Each of the edges of the graph will be randomly assigned one of the colors red, green, or blue. The assignments of colors to edges are mutually independent, and the probability of an edge being assigned red is $r$, blue is $b$, and green is $g$ (so $r + b + g = 1$).

A set of three vertices in the graph is called a triangle. A triangle is monochromatic if the three edges connecting the vertices are all the same color.

(a) Let $m$ be the probability that any given triangle, $T$, is monochromatic. Write a simple formula for $m$ in terms of $r, b,$ and $g$.

**Solution.** $m = r^3 + b^3 + g^3$ 

(b) Let $I_T$ be the indicator variable for whether $T$ is monochromatic. Write simple formulas in terms of $m, r, b,$ and $g$ for $Ex[I_T]$ and $Var[I_T]$.

$$Ex[I_T] = \quad Var[I_T] =$$

**Solution.**

$$Ex[I_T] = m, \quad Var[I_T] = m(1 - m).$$ 

\(\square\)
Now assume \( r = b = g = \frac{1}{3} \).

Let \( T \) and \( U \) be distinct triangles.

(c) What is the probability that \( T \) and \( U \) are both monochromatic?

**Solution.**

\[
\frac{1}{3^4}.
\]

If \( T \) and \( U \) do not share an edge, then the three edges of \( T \) match, and independently, the three edges of \( U \) must match, so both match with probability \((1/3)^2 \cdot (1/3)^2\). If they do share an edge, the five edges among them must all match, which happens with probability \((1/3)^4\) as well. \( \square \)

(d) Show that \( I_T \) and \( I_U \) are independent random variables.

**Solution.** Since \( I_T \) and \( I_U \) are indicators for events, it suffices to verify that

\[
\Pr[I_T = 1] \cdot \Pr[I_U = 1] = \Pr[I_T \cdot I_U = 1].
\]

There are two cases depending on whether \( T \) and \( U \) share an edge. In each case,

\[
\Pr[I_T = 1] \cdot \Pr[I_U = 1] = \left(\frac{1}{3}\right)^2 \cdot \left(\frac{1}{3}\right)^2 = \left(\frac{1}{3}\right)^4 = \Pr[I_T \cdot I_U = 1].
\]

\( \square \)

(e) Let \( M \) be the number of monochromatic triangles. Write simple formulas in terms of \( n \) and \( m \) for \( \text{Ex}[M] \) and \( \text{Var}[M] \).

\[
\begin{align*}
\text{Ex}[M] &= m \cdot \binom{n}{3}, \\
\text{Var}[M] &= \text{Var}[I_T] \cdot \# \text{triangles}
\end{align*}
\]

**Solution.**

\[
\begin{align*}
\text{Ex}[M] &= m \cdot \# \text{triangles} \\
&= m \binom{n}{3}, \quad (2) \\
\text{Var}[M] &= \text{Var}[I_T] \cdot \# \text{triangles} \\
&= m(1 - m) \binom{n}{3} = (1 - m) \text{Ex}[M]. \quad (3)
\end{align*}
\]

\( \square \)

(f) Let \( \mu := \text{Ex}[M] \). Prove that

\[
\Pr \left[ |M - \mu| > \sqrt{\mu \log \mu} \right] = O \left( \frac{1}{\log n} \right)
\]
Solution. According to Chebyshev’s Bound:

$$\Pr(|M - \mu| > c\sigma) \leq \frac{1}{c^2}$$

So

\[
\Pr \left[ |M - \mu| > \sqrt{\mu \log \mu} \right] = \Pr \left[ |M - \mu| > \sqrt{\log \mu \sqrt{\mu}} \right] \\
\leq \Pr \left[ |M - \mu| > \sqrt{\log (1-m) \mu} \right] \quad (0 \leq 1 - m \leq 1) \\
= \Pr \left[ |M - \mu| > \sqrt{\log \mu} \sigma \right] \quad \text{(by (3))} \\
\leq \frac{1}{\log \mu} \quad \text{(by Chebyshev)} \\
= O \left( \frac{1}{\log n} \right) \quad \text{(by (2))}.
\]

The last step follows because according to (2),

$$\mu = m \left( \frac{n}{3} \right) = \Theta(n^3),$$

so

$$\frac{1}{\log \mu} = \Theta \left( \frac{1}{\log n^3} \right) = \Theta \left( \frac{1}{\log n} \right).$$

\[\blacksquare\]

Problem 8 (Conditional Probability) (6 points). STAFF NOTE: (a) 1 pt, (b) 1, (c) 2, (d) 2

There are two decks of cards, the red deck and the blue deck. They differ slightly in a way that makes drawing the eight of hearts slightly more likely from the red deck than from the blue deck.

One of the decks is randomly chosen and hidden in a box. You reach in the box and randomly pick a card that turns out to be the eight of hearts. You believe intuitively that this makes the red deck more likely to be in the box than the blue deck.

Your intuitive judgment about the red deck can be formalized and verified using some inequalities between probabilities and conditional probabilities involving the events

- \( R := \text{Red deck is in the box}, \)
- \( B := \text{Blue deck is in the box}, \)
- \( E := \text{Eight of hearts is picked from the deck in the box}. \)

(a) State an inequality between probabilities and/or conditional probabilities that formalizes the assertion, “picking the eight of hearts from the red deck is more likely than from the blue deck.”

Solution.

$$\Pr[E | R] > \Pr[E | B].$$ \hspace{1cm} (4)

\[\blacksquare\]

(b) State a similar inequality that formalizes the assertion “picking the eight of hearts from the deck in the box makes the red deck more likely to be in the box than the blue deck.”
Solution. 
\[
\Pr \left[ R \mid E \right] > \Pr \left[ B \mid E \right].
\] (5)

(c) Assuming the each deck is equally likely to be the one in the box, prove that the inequality of part (a) implies the inequality of part (b).

Solution. From (4) and the definition of conditional probability,
\[
\frac{\Pr[\text{E AND } R]}{\Pr[R]} > \frac{\Pr[\text{E AND } B]}{\Pr[B]}.
\]
Also, \(\Pr[R] = \Pr[B] = 1/2\) by assumption. This implies
\[
\Pr[\text{E AND } R] > \Pr[\text{E AND } B].
\]
Dividing both sides of this inequality by \(\Pr[E]\) completes the proof:
\[
\Pr \left[ R \mid E \right] := \frac{\Pr[\text{E AND } R]}{\Pr[E]} > \frac{\Pr[\text{E AND } B]}{\Pr[E]} = \Pr \left[ B \mid E \right].
\]

(d) Suppose you couldn’t be sure that the red deck and blue deck were equally likely to be in the box. Could you still conclude that picking the eight of hearts from the deck in the box makes the red deck more likely to be in the box than the blue deck? Briefly explain.

Solution. No. In the extreme case, suppose the red deck consisted solely of the eight of hearts, but an adversary *almost never* chose to put the red deck in the box. Also, suppose there was a reasonable possibility, say 1 in 10, of drawing the eight of hearts from the blue deck. Then picking the eight of hearts is ten times more likely from the red deck than from the blue deck, but the probability that the blue deck is in the box remains nearly certain.

Problem 9 (Counting paths; Combinatorial Proof) (8 points). **STAFF NOTE**: (a) 1 pt, (b) 4 (c) 3

Solve the following counting problems. You may leave factorials and binomial coefficients in your solutions. No explanation is required.

(a) Theory Hippo wants to travel from point \((0, 0)\) to point \((20, 20)\) using only steps that add one to the \(x\) or the \(y\) coordinate. How many different paths can he choose from?

Solution.
\[
\binom{40}{20} = \frac{40!}{20! 20!}
\]
He must choose 20 of 40 steps to increment \(x\) and the remaining 20 to increment \(y\).

(b) Theory Hippo wants to travel from point \((0, 0)\) to point \((20, 20)\) using only steps that add one to the \(x\) or the \(y\) coordinate. There are Bottomless Pits of Utter Annihilation at points \((5, 5)\) and \((10, 10)\) through which paths cannot pass. How many paths are there for Theory Hippo to follow? **Hint**: Ignoring the pits, let \(N_5\) be the set of paths from \((0, 0)\) to \((20, 20)\) that go through \((5, 5)\); likewise \(N_{10}\) through \((10, 10)\).
Solution.

\[
\binom{40}{20} - \binom{10}{5} \cdot \binom{30}{15} - \binom{20}{10} \cdot \binom{20}{10} + \binom{10}{5} \cdot \binom{10}{5} \cdot \binom{20}{10}.
\]

The set of paths that are blocked by some pit is \(N_5 \cup N_{10}\). By Inclusion-Exclusion,

\[
|N_5 \cup N_{10}| = |N_5| + |N_{10}| - |N_5 \cap N_{10}|
\]

\[
\begin{align*}
&= \binom{10}{5} \cdot \binom{30}{15} + \binom{20}{10} \cdot \binom{20}{10} - \binom{10}{5} \cdot \binom{10}{5} \cdot \binom{20}{10}.
\end{align*}
\]

(6)

Consequently, the number of paths that do not cross a Bottomless Pit of Utter Annihilation is the number, \(\binom{40}{20}\), from part (a) of possible paths, minus (6).

(c) Below is a combinatorial proof of an equation. Fill in the empty boxes in the Theorem statement with the proper expressions.

**Theorem.**

\[
\begin{array}{|c|c|}
\hline
\text{Stinky Peterson owns } n \text{ newts, } t \text{ toads, and } s \text{ slugs. Conveniently, he lives in a dorm with } n + t + s \text{ other students. (The students are distinguishable, but creatures of the same variety are not distinguishable.)} \\
\text{Stinky wants to put one creature in each neighbor's bed. Let } W \text{ be the set of all ways in which this can be done.} \\
\text{On one hand, he could first determine who gets the slugs. Then, he could decide who among his remaining neighbors has earned a toad. Therefore, } |W| \text{ is equal to the expression on the left.} \\
\text{On the other hand, Stinky could first decide which people deserve newts and slugs and then, from among those, determine who truly merits a newt. This shows that } |W| \text{ is equal to the expression on the right.} \\
\text{Since both expressions are equal to } |W|, \text{ they must be equal to each other.} \\
\hline
\end{array}
\]

**Proof.** Stinky Peterson owns \(n\) newts, \(t\) toads, and \(s\) slugs. Conveniently, he lives in a dorm with \(n + t + s\) other students. (The students are distinguishable, but creatures of the same variety are not distinguishable.) Stinky wants to put one creature in each neighbor’s bed. Let \(W\) be the set of all ways in which this can be done.

On one hand, he could first determine who gets the slugs. Then, he could decide who among his remaining neighbors has earned a toad. Therefore, \(|W|\) is equal to the expression on the left.

On the other hand, Stinky could first decide which people deserve newts and slugs and then, from among those, determine who truly merits a newt. This shows that \(|W|\) is equal to the expression on the right.

Since both expressions are equal to \(|W|\), they must be equal to each other.

**Solution.**

\[
\binom{n + t + s}{s} \cdot \binom{n + t}{t} = \binom{n + t + s}{n + s} \cdot \binom{n + s}{n}.
\]

Problem 10 (Stable Matching) (7 points).
Four unfortunate children want to be adopted by four foster families of ill repute. A child can only be adopted by one family, and a family can only adopt one child. Here are their preference rankings (most-favored to least-favored):

<table>
<thead>
<tr>
<th>Child</th>
<th>Families</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bottlecap:</td>
<td>Hatfields, McCoys, Grinches, Scrooges</td>
</tr>
<tr>
<td>Lucy:</td>
<td>Grinches, Scrooges, McCoys, Hatfields</td>
</tr>
<tr>
<td>Dingdong:</td>
<td>Hatfields, Scrooges, Grinches, McCoys</td>
</tr>
<tr>
<td>Zippy:</td>
<td>McCoys, Grinches, Scrooges, Hatfields</td>
</tr>
</tbody>
</table>
(a) Exhibit two different stable matching of Children and Families.

<table>
<thead>
<tr>
<th>Family</th>
<th>Child in 1st match</th>
<th>Child in 2nd match</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grinches:</td>
<td>Zippy, Dingdong, Bottlecap, Lucy</td>
<td></td>
</tr>
<tr>
<td>Hatfields:</td>
<td>Zippy, Bottlecap, Dingdong, Lucy</td>
<td></td>
</tr>
<tr>
<td>Scrooges:</td>
<td>Bottlecap, Lucy, Dingdong, Zippy</td>
<td></td>
</tr>
<tr>
<td>McCoys:</td>
<td>Lucy, Zippy, Bottlecap, Dingdong</td>
<td></td>
</tr>
</tbody>
</table>

Solution. Treat Families as Girls and the result is the following assignment:

<table>
<thead>
<tr>
<th>Family</th>
<th>Children</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grinches:</td>
<td>Lucy</td>
</tr>
<tr>
<td>Hatfields:</td>
<td>Bottlecap</td>
</tr>
<tr>
<td>Scrooges:</td>
<td>Dingdong</td>
</tr>
<tr>
<td>McCoys:</td>
<td>Zippy</td>
</tr>
</tbody>
</table>

Treat Families as Boys and the result is the following assignment:

<table>
<thead>
<tr>
<th>Family</th>
<th>Children</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grinches:</td>
<td>Dingdong</td>
</tr>
<tr>
<td>Hatfields:</td>
<td>Bottlecap</td>
</tr>
<tr>
<td>Scrooges:</td>
<td>Lucy</td>
</tr>
<tr>
<td>McCoys:</td>
<td>Zippy</td>
</tr>
</tbody>
</table>

(b) Examine the matchings from part (a), and explain why these matchings are the only two possible stable matchings between Children and Families.

Hint: In general, there may be many more than two stable matchings for the same set of preferences.

Solution. The two stable matchings of part (a) are respectively Child optimal and Child pessimal. Since the matchings agree that Hatfields adopt Bottlecap, this means that the Hatfields are both the best and worst, and therefore the only, possible adoption for Bottlecap in any stable matching. Likewise for the McCoys adopting Zippy. This leaves only two ways to match up the remaining two children and families, so the two matchings from part (a) are the only ones possible.

Problem 11 (Sampling & Confidence) (10 points).

Yesterday, the programmers at a local company wrote a large program. To estimate the fraction, $b$, of lines of code in this program that are buggy, the QA team will take a small sample of lines chosen randomly and independently (so it is possible, though unlikely, that the same line of code might be chosen more than once). For each line chosen, they can run tests that determine whether that line of code is buggy, after which they will use the fraction of buggy lines in their sample as their estimate of the fraction $b$. 
The company statistician can use estimates of a binomial distribution to calculate a value, $s$, for a number of lines of code to sample which ensures that with 97% confidence, the fraction of buggy lines in the sample will be within 0.006 of the actual fraction, $b$, of buggy lines in the program.

Mathematically, the program is an actual outcome that already happened. The random sample is a random variable defined by the process for randomly choosing $s$ lines from the program. The justification for the statistician’s confidence depends on some properties of the program and how the random sample of $s$ lines of code from the program are chosen. These properties are described in some of the statements below. Indicate which of these statements are true, and explain your answers.

1. The probability that the ninth line of code in the program is buggy is $b$.

Solution. False.
The program has already been written, so there’s nothing probabilistic about the buggyness of the ninth (or any other) line of the program: either it is or it isn’t buggy, though we don’t know which. You could argue that this means it is buggy with probability zero or one, but in any case, it certainly isn’t $b$.

2. The probability that the ninth line of code chosen for the random sample is defective, is $b$.

Solution. True.
The ninth line sampled is equally likely to be any line of the program, so the probability it is buggy is the same as the fraction, $b$, of buggy lines in the program.

3. All lines of code in the program are equally likely to be the third line chosen in the random sample.

Solution. True.
The meaning of “random choices of lines from the program” is precisely that at each of the $s$ choices in the sample, in particular at the third choice, each line in the program is equally likely to be chosen.

4. Given that the first line chosen for the random sample is buggy, the probability that the second line chosen will also be buggy is greater than $b$.

Solution. False.
The meaning of “independent random choices of lines from the program” is precisely that at each of the $s$ choices in the sample, in particular at the second choice, each line in the program is equally likely to be chosen, independent of what the first or any other choice happened to be.

5. Given that the last line in the program is buggy, the probability that the next-to-last line in the program will also be buggy is greater than $b$.

Solution. False.
As noted above, it’s zero or one.

6. The expectation of the indicator variable for the last line in the random sample being buggy is $b$.

Solution. True.
The expectation of the indicator variable is the same as the probability that it is 1, namely, it is the probability that the $s$th line chosen is buggy, which is $b$, by the reasoning above.
7. Given that the first two lines of code selected in the random sample are the same kind of statement—they might both be assignment statements, or both be conditional statements, or both loop statements...—the probability that the first line is buggy may be greater than $b$.

Solution. True.

We don’t know how prone to bugginess different kinds of statements may be. It could be for example, that conditionals are more prone to bugginess than other kinds of statements, and that there are more conditional lines than any other kind of line in the program. Then given that two randomly chosen lines in the sample are the same kind, they are more likely to be conditionals, which makes them more prone to bugginess. That is, the conditional probability that they will be buggy would be greater than $b$.

8. There is zero probability that all the lines in the random sample will be different.

Solution. False.

There are $\frac{r!}{(r-s)!}$ ways to choose a sequence of $s$ distinct lines, and so the probability that all lines are distinct is

$$\frac{r!}{(r-s)!} = \frac{r \cdot (r-1) \cdot \ldots \cdot (r-s+1)}{r^s}$$

which is positive as long as the “small” sample size, $s$, is actually less than the length, $r$, of the program. Of course, the probability does approach zero as $s$ approaches $r$ and actually is zero once $s > r$, by the Pigeonhole Principle.

Problem 12 (Structural Induction) (6 points).

Definition. The set RAF of rational functions of one real variable is the set of functions defined recursively as follows:

Base cases:

- The identity function, $\text{id}(r) := r$ for $r \in \mathbb{R}$ (the real numbers), is an RAF,
- any constant function on $\mathbb{R}$ is an RAF.

Constructor cases: If $f, g$ are RAF’s, then so are

1. $f + g$, $fg$, and $f/g$.

(a) Prove by structural induction that RAF is closed under composition. That is, using the induction hypothesis,

$$P(h) := \forall g \in \text{RAF}. h \circ g \in \text{RAF},$$

prove that $P(h)$ holds for all $h \in \text{RAF}$. Make sure to indicate explicitly

- each of the base cases, and
- each of the constructor cases.

Solution. Proof. base cases: We must show $P(\text{id}_\mathbb{R})$ and $P(\text{constant-function})$. But this follows immediately from the fact that $\text{id}_\mathbb{R} \circ g = g$ and the composition of a constant function with any function, $g$, is constant.
**Constructor cases:** Given $e, f \in \text{RAF}$, we may assume by structural induction that $P(e)$ and $P(f)$ both hold, and must prove $P(h)$ for $h = e \circ f$ in the three cases where $\circ = +, \cdot, \text{or } \div$. But the same proof works for all three cases:

$$(e \circ f) \circ g = (e \circ g) \circ (f \circ g)$$

(by definition of composition), and since $(e \circ g), (f \circ g) \in \text{RAF}$ for all $g \in \text{RAF}$ by hypothesis, so is their combination using the operator $\circ$ – that is, their sum/product/quotient – by the constructor rule (1). This proves $P(h)$ for all three of the constructor cases $\circ = +, \cdot, \text{or } \div$.

This completes all the constructor cases, and so $\forall h \in \text{RAF}. P(h)$ follows by structural induction.

(b) Briefly explain why a similar proof using the induction hypothesis

$$Q(g) ::= \forall h \in \text{RAF}. h \circ g \in \text{RAF},$$

would break down.

**Solution.** The problem is that the identity used in the induction step of the previous part, namely that composition distributes from the left over each constructor operation $\circ$, breaks down for composition on the right. For example,

$$h \circ (g_1 + g_2) = (h \circ g_1) + (h \circ g_2)$$

fails when $h$ is nonlinear. For example, if $h(x) ::= x^2$, and $g_1 = g_2 = \text{id}_R$, then

$$[f \circ (g_1 + g_2)](r) = (2r)^2 = 4r^2$$

$$\neq 2r^2 + r^2$$

$$= [(f \circ g_1) + (f \circ g_2)](r).$$

So even though the conclusion $\forall g.Q(g)$ is true, a direct proof using $Q(g)$ as the induction hypothesis quickly gets stuck.