Solutions to the Final Examination

STAFF NOTE: 1.5 pts each, round up at end

Problem 1 (numbers short answer) (9 points).

Circle true or false for the statements below, and provide counterexamples for those that are false. Variables, \(a, b, c, m, n\) range over the integers and \(m, n > 1\).

(a) If \(a \equiv b \pmod{n}\), then \(p(a) \equiv p(b) \pmod{n}\) for any polynomial \(p(x)\) with integer coefficients.

Solution. true

(b) If \(\gcd(a, b) \neq 1\) and \(\gcd(b, c) \neq 1\), then \(\gcd(a, c) \neq 1\).

Solution. false \(a = 2 \cdot 3, b = 3 \cdot 5, c = 5 \cdot 7\)

(c) If no integer linear combination of \(a\) and \(b\) equals 2, then neither does any integer linear combination of \(a^2\) and \(b^2\).

Solution. true

(d) If \(a \equiv b \pmod{\phi(n)}\) for \(a, b > 0\), then \(c^a \equiv c^b \pmod{n}\).

Solution. false. Need \(c\) relatively prime to \(n\). Counterexample: \(n = 4, \phi(n) = 2; a = 1, b = 3, \) so \(a \equiv b \pmod{\phi(n)}, c = 2, \) so \(c^a = 2 \neq 0 = c^b \pmod{4}\).

(e) If \(a, b > 1\), then \([a\) has a multiplicative inverse \(\pmod{b}\) iff \(b\) has a multiplicative inverse \(\pmod{a}\).]

Solution. true

(f) If \(\gcd(a, n) = 1\), then \(a^{n-1} \equiv 1 \pmod{n}\).

Solution. false Let \(a = 5, n = 6\).

STAFF NOTE: part(a) 1pt for circling all but the first, 0pts anything else
part(b) 1.5 pts each, round up at end

Problem 2 (graphs short answer) (8 points).
(a) Circle all the properties below that are preserved under graph isomorphism.

- The vertices can be numbered 1 through 7.
- There is a cycle that includes all the vertices.
- There are exactly two spanning trees.
- The OR of two properties that are preserved under isomorphism.

Solution. All are preserved.

(b) For the following four statements about finite simple graphs, circle those that are true, and provide counterexamples for those that are false.

- Every graph has a spanning tree.  
  true  false
  Solution. false. Any disconnected graph is a counterexample.

- A graph with three vertices cannot have exactly one vertex of degree 1. 
  true  false
  Solution. true. There are only two such connected graphs: a triangle (with no degree 1 vertices) and a length three line graph with two degree 1 vertices. Otherwise, the graph has a degree 0 vertex in which case the other two have the same degree.

- The number of leaves in a tree is not equal to the number of non-leaf vertices. 
  true  false
  Solution. false. A line graph with 4 vertices has 2 non-leaf vertices and 2 leaves.

- The minimum number of edges possible in a nonplanar 2-colorable graph is 10. 
  true  false
  Solution. false. $K_{3,3}$ is the smallest 2-colorable (bipartite) nonplanar graph, and it has 9 edges. $K_5$ is nonplanar with 10 edges, but not bipartite.

STAFF NOTE: ide base 2pts, constant base 1pt, f+g & fg together 3pts, 2f 2 pts

Problem 3 (structural induction) (8 points).
The Limited 18.01 Functions (LF18’s) are defined similarly to the F18 functions from class problem 7.3, but they don’t have function composition or inverse as a constructor. Namely,

Definition. LF18 is the set of functions of one complex variable defined recursively as follows:

Base cases:
- The identity function, id(z) ::= z for z ∈ ℂ, is an LF18,
- any constant function is an LF18.

Constructor cases: If f, g are LF18’s, then so are

1. $f + g$, $fg$, and $2^f$.

Prove by structural induction that LF18 is closed under composition. That is, using the induction hypothesis,

$$P(f) ::= \forall g ∈ LF18. f ◦ g ∈ LF18,$$

prove that $P(f)$ holds for all $f ∈ LF18$. 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 \( g \circ \text{id}_\mathbb{R} = g \) and the composition of \( g \) with a the constant function is a constant function.

constructor cases: Given \( e, f \in LF_{18} \), we may assume by structural induction that \( P(e) \) and \( P(f) \) both hold, and must prove \( P(h) \) where

- case \( h = e \circ f \) where \( \circ = + \) or \( \cdot \): In this case,
  \[ h \circ g = (e \circ g) \circ (f \circ g) \]
  and since \((g \circ e), (g \circ f) \in LF_{18}\) by hypothesis, so is their \( \circ \) by the constructor rule (1). This proves \( P(h) \) in this case.

- case \( h = 2^f \). This follows similarly since \( 2^f \circ g = 2^f \circ g \).

This completes all the constructor cases, and so \( \forall f \in LF_{18}. P(f) \) follows by structural induction.

STAFF NOTE: 2 pts each

Problem 4 (lining up fall11) (6 points).
There are 10 students \( A, B, \ldots, J \) who will be lined up left to right according to the some rules below.
Translate each of the following rules into predicate formulas with the set of 10 students as the domain of discourse. The only predicates you may use are

- equality and,
- \( F(x, y) \), meaning that “\( x \) is to the left of \( y \).” For example, in the lineup “CDA”, both \( F(C, A) \) and \( F(C, D) \) are true.

(a) Rule I: Student A must not be rightmost.

Solution.
\[ \exists x. F(A, x) \]

(b) Rule II: Student B must be adjacent to C (directly to the left or right of C).

Solution.
\[ \forall x.((x \neq B) \text{ AND } (x \neq C)) \text{ IMPLIES } (F(x, B) \text{ IFF } F(x, C)) \]

(c) Rule III: Student D is always second.
Solution.

\[ \exists x. F(x, D) \text{ AND } (\forall y. ((y \neq D) \text{ AND } (y \neq x)) \text{ IMPLIES } F(D, y)) \]

\[ S T A F F \ \ N O T E: \ 1, \ 3, \ 3 \ p t s \]

\[ a \]

**Problem 5 (infinite binary sequences) (7 points).**

Let \( \{0, 1\}^\omega \) be the uncountable set of infinite binary sequences, and let \( F_n \subset \{0, 1\}^\omega \) be the set of infinite binary sequences whose bits are all 0 after the \( n \)th bit. That is, if \( s := (s_0, s_1, s_2, \ldots) \in \{0, 1\}^\omega \), then

\[ s \in F_n \text{ IFF } \forall i > n. s_i = 0. \]

For example, the sequence \( t \) that starts 001101 with 0’s after that is in \( F_5 \), since by definition \( t_i = 0 \) for all \( i > 5 \). In fact, \( t \) is by definition also in \( F_6, F_7, \ldots. \)

(a) What is the size, \( |F_n| \), of \( F_n \)?

**Solution.**

\[ 2^{n+1} \]

**STAFF NOTE: 1/2 credit for \( 2^n \).**

There are only \( 2^{n+1} \) possible initial sequences of \( n + 1 \) bits for sequences in \( F_n \) and after that all of them are all 0’s.

(b) Explain why the set \( F \subset \{0, 1\}^\omega \) of sequences with only finitely many 1’s, is a countable set. (You may assume without proof any results from class about countability.)

**Solution.** Note that

\[ F = \bigcup_{n=0}^{\infty} F_n. \]

Now \( F_n \) is finite and therefore is countable, and a countable union of countable sets is countable (Problem 5.11), so \( F \) is countable.

An alternative explanation is to notice that the mapping \( f : F \to \mathbb{N} \), where

\[ f(s) := \sum_{n \in \mathbb{N}} s_n 2^n \]

is a bijection.

\( F \) can also be shown to be countable by defining a \([ \leq 1 \text{ out } \geq 1 \text{ out}] \) mapping (surjective function) from some countable set to \( F \) (Problem 5.8). This is easy using the fact that the set \( \{0, 1\}^* \) of finite binary strings is countable (Problem 5.3). Namely, define the function \( g : \{0, 1\}^* \to F \), by the rule that \( g(z) \) is the finite sequence \( z \) followed by an infinite sequence of 0’s.

(c) Prove that the set of infinite binary sequences with infinitely many 1’s is uncountable. **Hint:** Use parts (a) and (b); a direct proof by diagonalization is tricky.
Solution. The set of sequences with infinitely many 1’s is \( \{0, 1\}^\omega - F \).

Suppose for the sake of contradiction that \( \{0, 1\}^\omega - F \) was countable. Since the union of two countable sets is countable, this implies \( (\{0, 1\}^\omega - F) \cup F = \{0, 1\}^\omega \) is countable, contradicting the fact that \( \{0, 1\}^\omega \) is uncountable.

Using a naive diagonal argument does not work here. Namely, suppose there is a countable list of elements of sequences with infinitely many 1’s and we create a new “diagonal” sequence, \( s \), that differs from the \( n \)th sequence in the list at position \( n \), for all \( n \in \mathbb{N} \). Then \( s \) is not in the list, but there is no guarantee that \( s \) has infinitely many 1’s. So the diagonalization may not yield a “missing” sequence, and the proof breaks down.

By the way, there is a simple fix to the diagonal argument: go down a 30° diagonal instead of 45°. Namely, define a diagonal sequence, \( s' \), that differs from the \( n \)th sequence in the list at position \( 2n \). This still ensures that \( s' \) is not in the list, but it leaves all the odd numbered bits of \( s' \) unspecified. So we can define \( s'_{2n+1} := 1 \) for all \( n \in \mathbb{N} \), guaranteeing that \( s' \) has infinitely many 1’s.

**Problem 6** (congruences) (6 points).

(a) Call a number from 0 to 174 powerful iff some positive power of the number is congruent to 1 modulo 175. What is the probability that a random number from 0 to 174 is powerful?

Solution.

$$24 \quad 35$$

Note that \( x^k \equiv 1 \pmod{n} \) for some \( k \) iff \( x \) has an inverse modulo \( n \) iff \( x \) is relatively prime to \( n \). So being powerful is equivalent to being relatively prime to 175. There are \( \phi(175) = (5^2 - 5)(7 - 1) = 20 \cdot 6 = 120 \) numbers from 0 to 174 that are relatively prime to 175, so

$$\Pr[\text{powerful}] = \frac{120}{175} = \frac{24}{35}.$$  

(b) What is the remainder of \((-12)^{482}\) divided by 175?

Solution. 144.

Since -12 and 175 are relatively prime, we have by Euler’s Theorem that \( (-12)^{\phi(175)} \equiv 1 \pmod{175} \), and so

$$(-12)^{482} = ((-12)^{120})^4 \cdot (-12)^2 \equiv 1^4 \cdot (144) \equiv 144 \pmod{175}.$$  

**STAFF NOTE:** 1,1,1,2(1/2,1/12),2 pts

**Problem 7** (directed graphs and probability) (7 points). (a) For the directed acyclic graph (DAG) \( G_0 \) in Figure 1, a minimum-edge DAG with the same walk relation can be obtained by removing some edges. List these edges (use notation \( u \rightarrow v \) for an edge from \( u \) to \( v \)):

...
Solutions to the Final Examination

Figure 1  The DAG $G_0$

Solution. After removing edges $(1 \to 4)$, $(3 \to 2)$ and $(3 \to 4)$, we get the minimum DAG.

(b) List the vertices in a maximal chain in $G_0$.

Solution. $\{3, 1, 2, 4\}$

Let $G$ be the simple graph shown in Figure 2.
A directed graph $\vec{G}$ can be randomly constructed from $G$ by assigning a direction to each edge independently with equal likelihood.

(e) What is the probability that $\vec{G} = G_0$?

Solution.

$$2^{-|E(G)|} = 2^{-7} = \frac{1}{128}$$

Define the following events with respect to the random graph $\vec{G}$:

$T_1 ::= \text{vertices } 2, 3, 4 \text{ are on a length-3 directed cycle}$,

$T_2 ::= \text{vertices } 1, 3, 4 \text{ are on a length-3 directed cycle}$,

$T_3 ::= \text{vertices } 1, 2, 4 \text{ are on a length-3 directed cycle}$,

$T_4 ::= \text{vertices } 1, 2, 3 \text{ are on a length-3 directed cycle}$.
Figure 2  Simple graph $G$

(d) What is

$\Pr[T_1]$?

$\Pr[T_1 \cap T_2]$?

$\Pr[T_1 \cap T_2 \cap T_3]$?

Solution.

$$\Pr[T_1] = \frac{2}{2^3} = \frac{1}{4}$$

$$\Pr[T_1 \cap T_2] = \frac{1}{16}$$

$$\Pr[T_1 \cap T_2 \cap T_3] = 0.$$

(e) $\vec{G}$ has the property that if it has a directed cycle, then it has a length-3 directed cycle. Use this fact to find the probability that $\vec{G}$ is a DAG.

Solution. The only possible length-3 directed cycles are the ones described by $T_1, \ldots, T_4$. So the given property implies that $\vec{G}$ is a DAG iff $\bigcup_{i \in [1,4]} T_i$ does not occur.

Now using the Inclusion-Exclusion principle,
\[ \Pr[G \text{ is a DAG}] = 1 - \Pr[T_1 \cup T_2 \cup T_3 \cup T_4] \]
\[ = 1 - \sum_{i \in [1,4]} \Pr[T_i] + \sum_{i \neq j} \Pr[T_i \cap T_j] \]
\[ - \sum_{i \neq j \neq k} \Pr[T_i \cap T_j \cap T_k] + \Pr[T_1 \cap T_2 \cap T_3 \cap T_4] \]
\[ = 1 - 4 \cdot \frac{1}{4} + 6 \cdot \frac{1}{16} - 0 + 0 \]
\[ = \frac{3}{8} \]

Here we’re using the fact that the results of part (d) for \( T_1, T_2, T_3 \) hold by symmetry for \( T_i, T_j, T_k \) for all distinct values of \( i, j, k \in [1,4] \).

\[ \text{STAFF NOTE: (a) 1 each, (b) 2pts} \]

\textbf{Problem 8 (asymptotics) (7 points).}

(a) Indicate which of the following asymptotic relations below on the set of nonnegative real-valued functions are \textit{equivalence relations}, (E), strict partial orders (S), weak partial orders (W), or \textit{none} of the above (N).

- \( f \sim g \), the “asymptotically Equal” relation.
  
  \textbf{Solution. E} \hspace{1cm} \star

- \( f = o(g) \), the “little Oh” relation.
  
  \textbf{Solution. S} \hspace{1cm} \star

- \( f = O(g) \), the “big Oh” relation.
  
  \textbf{Solution. N} because it is not antisymmetric, \hspace{1cm} \star

- \( f = \Theta(g) \), the “Theta” relation.
  
  \textbf{Solution. E} \hspace{1cm} \star

- \( f = O(g) \text{ AND NOT}(g = O(f)) \).
  
  \textbf{Solution. S.} \hspace{1cm} \star

(b) Define two functions \( f, g \) that are incomparable under big Oh:

\[ f \neq O(g) \text{ AND } g \neq O(f). \]
Solution. One example is,

\[ f(n) := \begin{cases} n & \text{if } n \text{ is odd}, \\ 0 & \text{if } n \text{ is even} \end{cases}, \quad g(n) := \begin{cases} 0 & \text{if } n \text{ is odd}, \\ n & \text{if } n \text{ is even}. \end{cases} \]

which can also be described by the formulas

\[ f(n) := n \sin \left( \frac{n\pi}{2} \right), \quad g(n) := n \cos \left( \frac{n\pi}{2} \right). \]

**STAFF NOTE:** (a) 2 (b) 2, 2

Problem 9 (counting) (6 points). (a) If the letters in the word FINESSED are randomly permuted, what is the probability that all the vowels are adjacent?

Solution.

\[ \frac{3}{28} \]

There are \( \binom{8}{2,2,1,1,1,1} = 8!/4 \) equally likely permutations. There are \( \binom{6}{2,1,1,1,1} = 6!/2 \) permutations of FNSSD along with the block of vowels (EEI), and \( \binom{3}{2,1} = 3 \) permutations of the vowel block, so the probability that all vowels are adjacent is

\[ \frac{(6!/2)3}{(8!/4)} = \frac{3}{28}. \]

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

**Theorem.**

\[
\begin{array}{c}
\text{ } \\
\text{ } \\
\end{array}
\]

\[
\begin{array}{c}
\text{ } \\
\text{ } \\
\end{array}
\]

**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.
Solutions to the Final Examination

Solution.

\[
\binom{n+t+s}{s} \cdot \binom{n+t}{t} \cdot \binom{n+s}{n+1} \cdot \binom{n+s}{n}
\]

\[\square\]

**STAFF NOTE:** (a)2 (4 x 1/2) (b)1.5 (c)2.5 pts

**Problem 10 (generating functions) (6 points).**

T-Pain is planning an epic boat trip and he needs to decide what to bring with him.

- He can bring burgers, but they only come in packs of 6.
- He and his two friends can’t decide whether they want to dress formally or casually. He’ll either bring 0 pairs of flip flops or 3 pairs.
- He doesn’t have very much room in his suitcase for towels, so he can bring at most 2.
- In order for the boat trip to be truly epic, he has to bring at least 1 nautical-themed pashmina afghan.

(a) Let \(B(x)\) be the generating function for the number of ways to bring \(n\) burgers, \(F(x)\) for the number of ways to bring \(n\) pairs of flip flops, \(T(x)\) for towels, and \(A(x)\) for Afghans. Write simple formulas for each of these.

\[B(x) : \quad F(x) : \]
\[T(x) : \quad A(x) : \]

(b) Let \(g_n\) be the the number of different ways for T-Pain to bring \(n\) items (burgers, pairs of flip flops, towels, and/or afghans) on his boat trip. Let \(G(x)\) be the generating function \(\sum_{n=0}^{\infty} g_n x^n\). Verify that

\[G(x) = \frac{x}{(1-x)^2}.\]

Solution. The successive generating functions for beer, flip flops, towels, and afghans, are

\[\frac{1}{1-x^6}, \quad (1 + x^3), \quad (1 + x + x^2), \quad x \frac{x}{1-x}.\]

So \(G(x)\) is their product:

\[
\frac{1}{1-x^6} (1 + x^3)(1 + x + x^2) \frac{x}{1-x} = \frac{(1 + x^3)(1 + x + x^2)x}{(1-x^3)(1+x^3)(1-x)} = \frac{(1 + x + x^2)x}{(1-x)(1+x+x^2)(1-x)} = \frac{x}{(1-x)^2}
\]

\[\square\]

(c) Find a simple formula for \(g_n\).
Solutions to the Final Examination

11

Solution. n.
Let

\[ G(x) := \frac{1}{(1 - x)^2}, \]

so the generating function for T-Pain is \( xG(x) \). We know that the coefficient of \( x^n \) in the series for \( (1 - x)^2 \) is, by the Convolution Rule, the number of ways to select \( n \) items of two different kinds, namely, \( \binom{n+1}{1} = n + 1 \), so we conclude that the \( n \)th coefficient in the series for \( G(x) \) is \( n + 1 \).

So the \( n \)th coefficient in the series for the generating function \( xG(x) \) is zero for \( n = 0 \), and, for \( n \geq 1 \), is the \( (n - 1) \)th coefficient of \( G \), namely,

\[ (n - 1) + 1 = n. \]

STAFF NOTE: wrong strategy or wrong probability -4pts

Problem 11 (probability) (6 points).
Suppose that Let’s Make a Deal is played according to slightly different rules and with a red goat and a blue goat. There are three doors, with a prize hidden behind one of them and the goats behind the others. No doors are opened until the contestant makes a final choice to stick or switch. The contestant is allowed to pick a door and ask a certain question that the host then answers honestly. The contestant may then stick with their chosen door, or switch to either of the other doors.

If the contestant asks “is the red goat behind one of the unchosen doors?” and the host answers “yes,” is the contestant more likely to win the prize if they stick, switch, or does it not matter? Clearly identify the probability space of outcomes and their probabilities you use to model this situation. What is the contestant’s probability of winning if he uses the best strategy?

Solution. They are more likely to win if they stick.

To model the stick strategy, we can use three equally likely outcomes corresponding to the contestant picking the door with the prize, the red goat or the blue goat. The contestant wins in one outcome —when he picks the prize. Being given that the red goat is behind an unchosen door excludes the outcome of picking the red goat, leaving only a winning outcome and an equally likely losing outcome, so sticking leads to a probability of winning equal to \( 1/2 \).

For the switch strategy, we can use six equally likely outcomes corresponding to what’s behind the picked door and then what’s behind the switched-to door. Given that the red goat is behind an unchosen door rules out two of the outcomes, leaving four equally likely outcomes for which the red goat is not behind the door first picked by the contestant. In two of these outcomes, the contestant first picks the door with the prize, so these are both losing outcomes under the switch strategy. In the other two outcomes, the contestant first picks the door with the blue goat and then switches either to the red goat or the prize. The contestant wins only in the outcome where he switches to the prize. So the contestant wins in only one of the four outcomes, and therefore switching wins with probability only \( 1/4 \).

Problem 12 (conditional probability) (7 points).
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.”

\[ \Pr[E | R] > \Pr[E | B]. \quad (1) \]

(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.”

\[ \Pr[R | E] > \Pr[B | E]. \quad (2) \]

(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 (1) and the definition of conditional probability,

\[ \frac{\Pr[E \text{ AND } R]}{\Pr[R]} > \frac{\Pr[E \text{ AND } B]}{\Pr[B]} . \]

Also, \( \Pr[R] = \Pr[B] = 1/2 \) by assumption. This implies

\[ \Pr[E \text{ AND } R] > \Pr[E \text{ AND } B]. \]

Dividing both sides of this inequality by \( \Pr[E] \) completes the proof:

\[ \Pr[R | E] := \frac{\Pr[E \text{ AND } R]}{\Pr[E]} > \frac{\Pr[E \text{ AND } B]}{\Pr[E]} = \Pr[B | E] . \]

Problem 13 (deviation) (8 points).

Tom has a gambling problem. He plays 240 hands of draw poker, 120 hands of black jack, and 40 hands of stud poker per day. He wins a hand of draw poker with probability 1/6, a hand of black jack with probability 1/2, and a hand of stud poker with probability 1/5.

(a) What is the expected number of hands that Tom wins in a day?
Solution. $240(1/6) + 120(1/2) + 40(1/5) = 108.$

(b) What would the Markov bound be on the probability that Tom will win at least 216 hands on a given day?

Solution. The expected number of games won is 108, so by Markov, $\Pr[R \geq 216] \leq 108/216 = 1/2.$

(c) Assume the outcomes of the card games are pairwise independent. What is the variance in the number of hands won per day? You may answer with a numerical expression that is not completely evaluated.

Solution. The variance can also be calculated using linearity of variance. For an individual hand the variance is $p(1 - p)$ where $p$ is the probability of winning. Therefore the variance is

$$240(1/6)(5/6) + 120(1/2)(1/2) + 40(1/5)(4/5) = 1046/15 = 69 \frac{11}{15}.$$ 

(d) What would the Chebyshev bound be on the probability that Tom will win at least 216 hands on a given day? You may answer with a numerical expression that is not completely evaluated.

Solution.

$$\Pr[R - 108 \geq 108] \leq \Pr[|R - 108| \geq 108] \leq \frac{V}{108^2} = \frac{1046}{15(108)^2} \approx 0.0059785.$$

(A very slightly better bound of 0.0059430 comes from using the one-sided Chebyshev bound from Problem 18.7.)

STAFF NOTE: (a)6pts (6 x 1) (b) 3pts (choice 3 & 4 correct; choice 3 or 4 alone–2pts, 3 & 4 & another–1pt)

Problem 14 (random sampling) (9 points).

You work for the president and you want to estimate the fraction $p$ of voters in the entire nation that will prefer him in the upcoming elections. You do this by random sampling. Specifically, you select a random voter and ask them who they are going to vote for. You do this $n$ times, with each voter selected with uniform probability and independently of other selections. Finally, you use the fraction $P$ of voters who said they will vote for the President as an estimate for $p$.

(a) Our theorems about sampling and distributions allow us to calculate how confident we can be that the random variable, $P$, takes a value near the constant, $p$. This calculation uses some facts about voters and the way they are chosen. Circle the true facts among the following:

1. Given a particular voter, the probability of that voter preferring the President is $p$.
2. Given a particular voter, the probability of that voter preferring the President is 1 or 0.
3. The probability that some voter is chosen more than once in the sequence goes to zero as $n$ increases.
4. All voters are equally likely to be selected as the third in our sequence of \( n \) choices of voters (assuming \( n \geq 3 \)).

5. The probability that the second voter chosen will favor the President, given that the first voter chosen prefers the President, is greater than \( p \).

6. The probability that the second voter chosen will favor the President, given that the second voter chosen is from the same state as the first, may not equal \( p \).

**Solution.** The preference of any particular voter is a constant: either "the President" or "not the President", so (1) is false and (2) is true. (3) is false; in fact, the Birthday "paradox" implies the probability of some voter being chosen more than once rapidly approaches 1 as \( n \) grows beyond 100. (4) holds by definition of randomly choosing an item from a set. (5) is false because successive voters in the sequence are chosen independently. (6) is true because, for example, the fraction of voters who prefer the President in the largest states may all be \(< p \).

(b) Suppose that according to your calculations, the following is true about your polling:

\[ \Pr[|P - p| \leq 0.04] \geq 0.95. \]

You do the asking, you count how many said they will vote for the President, you divide by \( n \), and find the fraction is 0.53. Among the following, circle the legitimate things you might say in a call to the President:

1. Mr. President, \( p = 0.53! \)
2. Mr. President, with probability at least 95 percent, \( p \) is within 0.04 of 0.53.
3. Mr. President, either \( p \) is within 0.04 of 0.53 or something very strange (5-in-100) has happened.
4. Mr. President, we can be 95\% confident that \( p \) is within 0.04 of 0.53.

**Solution.** You cannot say (1): the only way to know the exact value of the constant \( p \) is to ask all 250,000,000 voters.

You cannot say (2) either: \( p \) is a constant which can either be or not be within 0.04 of 0.53. If it is, then the probability that it is 1, and thus at least 0.95, and therefore (2) will be true. If it is not, then the probability that it is 0, and thus smaller than 0.95, and therefore (2) will be false.

You can say (3): To see why, start with the statement either \(|0.53 - p| \leq 0.04\) or \(|0.53 - p| > 0.04\) which is obviously true. Now read it as follows: Either \( p \) is within 0.04 of 0.53 or it is not and therefore my random variable \( P \) took a value from a set that is hit only 5 times in 100. So, clearly, either \( p \) is within 0.04 of 0.53 or something strange has happened.

It’s probably best to say (4): it means the same thing as (3), but is simpler to say and will be easily understandable by anyone with basic knowledge of sampling theory. And for people without such knowledge, (4) will make intuitive sense and be less likely to cause confusion than the fuller explanation in (3).