Solutions to the Final Examination

Problem 1 (10 points). For each of the following logical formulas with domain of discourse the natural numbers, $\mathbb{N}$, indicate whether it is a possible formulation of

   I: the **I**nduction Axiom,

  S: the **S**trong Induction Axiom,

  L: the **L**east Number Principle (called the Well-ordering Principle by Rosen), or

  N: **N**one of these.

For example, the Rule for the Least Number Principle in the appendix could be expressed with the following formula, so it gets labelled "L".

$$[\exists n P(n)] \longrightarrow \exists m \, [P(m) \wedge \forall n \, [P(n) \longrightarrow m \le n]] \qquad \underline{\quad \textbf{L} \quad}$$

This is a multiple choice problem: do not explain your answer.

(a) (2 points) $\quad [P(b) \wedge [\forall k \ge b \; P(k) \longrightarrow P(k+1)]] \longrightarrow \forall k \ge b \; P(k)$

Solution. I. This is a formulation of the Induction Axiom with base case b instead of zero. $P(k) \longrightarrow P(k+1)$ for $k \ge b$ describes the induction step. ■

(b) (2 points) $\quad [P(0) \wedge \forall k \, [\forall m \le k \; P(m)] \longrightarrow P(k+1)] \longrightarrow \forall k \; P(k)$

Solution. S. Strong Induction with base case 0. ■

(c) (2 points) $\quad [\forall n \, [\forall m < n \; P(m)] \longrightarrow P(n)] \longrightarrow \forall n \; P(n)$

Solution. S. This is the formulation of Strong Induction with no base case. ■

(d) (2 points) $[\exists n \; P(n)] \longrightarrow \exists n \; [P(n) \wedge \forall k < n \; \overline{P(k)}]$

Solution. L.   ∎

(e) (2 points) $\forall n \; [P(n) \longrightarrow \exists n \; [P(n) \wedge \forall k \; P(k) \longrightarrow n \leq k]]$

Solution. L. This is an equivalent way of expressing the example formula from the appendix.   ∎

Problem 2 (10 points). An integer, m, *divides* an integer, n, in symbols, $m \mid n$, iff there is an integer k such that $km = n$.

Claim. *For any prime, p, and positive integers $x_1, x_2, \ldots, x_n$, if $p \mid x_1 x_2 \cdots x_n$, then $p \mid x_i$ for some i between 1 and n.*

Underline the sentence where the following proof goes wrong and explain.

> *False proof.* [By strong induction on n.]
>
> The induction hypothesis is the Claim itself.
>
> Base case ($n = 1$): When $n = 1$, we have $p \mid x_1$, therefore we can let $i = 1$ and conclude $p \mid x_i$.
>
> Induction step: Now assuming the claim holds for all $k \leq n$, we must prove it for $n + 1$.
>
> So suppose $p \mid x_1 x_2 \cdots x_{n+1}$. Let $y_n = x_n x_{n+1}$, so $x_1 x_2 \cdots x_{n+1} = x_1 x_2 \cdots x_{n-1} y_n$. Since the righthand side of this equality is a product of n terms, we have by induction that p divides one of them. If $p \mid x_i$ for some $i < n$, then we have the desired i. Otherwise $p \mid y_n$. But y_n is a product of the two terms x_n, x_{n+1}. Therefore, we have by strong induction that p divides one of them. So in this case $p \mid x_i$ for $i = n$ or $i = n + 1$.   □

Solution. Notice that the proof above never uses the fact that p is prime, indicating that something is wrong, since the claim is false when p is not prime.

The statement "we have by strong induction that p divides one of them" is the place where the proof breaks down: it appeals to strong induction to justify applying the induction hypothesis for $2 = k \leq n$. But the base case was $n = 1$, so we can't assume $2 \leq n$.

Note that the reasoning above is fine for every $n \neq 2$, so the whole proof would be fine if we had an argument to prove the claim for $n = 2$. This case is indeed true and can be proved as follows: if p does not divide x_1, then, because p is prime, $\gcd(p, x_1) = 1$ and so

$$kp + mx_1 = 1 \tag{1}$$

for some integers k and m. Multiplying both sides of (1) by x_2 yields

$$kpx_2 + mx_1x_2 = x_2. \tag{2}$$

But when $n = 2$, we are given that $p \mid x_1x_2$. Hence, we have that p divides both terms in the sum on the lefthand side of (2) and therefore divides the righthand side. That is, $p \mid x_2$. ∎

Problem 3 (20 points). Given a simple graph G, we apply the following operation to the graph: pick two vertices $u \neq v$ such that either

1. there is an edge of G between u and v and there is also a path from u to v which does *not* include this edge; in this case, delete the edge $\{u, v\}$.

2. or, there is no path from u to v; in which case, add the edge $\{u, v\}$.

We keep repeating these operations until it is no longer possible to find two vertices $u \neq v$ to which an operation applies.

Assume the vertices of G are the integers $1, 2, \ldots, n$ for some $n \geq 2$. This procedure can be modelled as a state machine whose states are all possible simple graphs with vertices $1, 2, \ldots, n$. The start state is G, and the final states are the graphs on which no operation is possible.

(a) (10 points) For any state, G', let e be the number of edges in G', c be the number of connected components it has, and s be the number of simple cycles. For each of the derived variables below, indicate the *strongest* of the properties that it is guaranteed to satisfy, no matter what the starting graph G is. The choices for properties are: *constant, strictly increasing, strictly decreasing, weakly increasing, weakly decreasing, none of these*. The derived variables are

(i) $e - s$

   Solution. weakly increasing ∎

(ii) $3c/2 + e$

Solution. strictly decreasing ∎

(iii) $c + s$

Solution. strictly decreasing ∎

(iv) (c, e), partially ordered coordinatewise

Solution. weakly decreasing ∎

(v) (c, e), ordered lexicographically

Solution. strictly decreasing ∎

(b) (6 points) Choose a derived variable from above and prove that it is strictly decreasing in some well-founded partial order. Conclude that the procedure terminates.

Solution. To show that the variable (ii) strictly decreases, note that the rule for deleting an edge ensures that the connectedness relation does not change, so neither does the number of connected components c. Meanwhile the number of edges e decreases by one when an edge is deleted. Therefore the variable $3c/2 + e$ decreases by one. The rule for adding an edge ensures that the number of connected components c decreases by one and the number of edges e increases by one. Therefore the variable $3c/2 + e$ decreases by $1/2$.

To show that the variable (iii) strictly decreases, note that the rule for deleting an edge ensures that the number of connected components c does not change and the number of simple cycles s decreases by n, where $n \geq 1$. Therefore the variable $c + s$ decreases by n. The rule for adding an edge ensures that the number of connected components c decreases by one and the number of simple cycles s does not change. Therefore the variable $c + s$ decreases by one.

To show that the lexicographically ordered (c, e) strictly decreases, note that the rule for deleting an edge ensures that the number of connected components c does not change and the number of edges e decreases by one. The rule for adding an edge ensures that the number of connected components c decreases by one. ∎

(c) (4 points) Prove that any final state must be a tree on the vertices.

Solution. We use the characterization of a tree as a cycle-free, connected, simple graph.

A final state must be connected, because otherwise there would be two vertices with no path between them, and then a transition adding the edge between them would be possible, contradicting finality of the state.

A final state can't have a cycle, because deleting any edge on the cycle would be a possible transition. ∎

Problem 4 (10 points). This problem is about game trees for two-player games of perfect information. Player 1 is the player who plays first. Player 1 wins iff Player 2 loses, but the game may end in a draw, or may continue forever.

For each of the parts below, indicate whether the statement is *true* or *false*. If the claim is false, briefly describe a counter-example.

(a) (2 points) If Player 1 has a winning strategy, then the game cannot continue forever.

Solution. False. The game may not end if Player 1 plays stupidly—not using his winning strategy. INSERT COUNTEREXAMPLE. ∎

(b) (2 points) If the game tree is finite-path, then the game cannot continue forever.

Solution. True. ∎

(c) (2 points) One of the players must have a non-losing strategy.

Solution. False. Consider a game in which the players have only one move at each step, but the game never ends. ∎

For the remaining parts, we only consider games with finite-path game trees.

(d) (2 points) If player 1 has a non-losing strategy, then so must player 2.

Solution. False. Player 1's non-losing strategy might be a winning strategy. ∎

(e) (2 points) If the game can only end after an odd number of moves, then Player 1 has an advantage.

Solution. False. It's perfectly possible for Player 2 to have a winning strategy in an odd-number-of-move-game. ∎

Problem 5 (10 points). Consider the seven letter word *armeyer*.

(a) (3 points) How many different sequences of these seven letters are there?

Solution. $\binom{7}{1,1,1,2,2,1} = 7!/2!2!$ ∎

(b) (7 points) How many such sequences are there that do not contain either of the words *eye* or *ram*? . Explain your answer.

Solution. Inclusion-Exclusion.

$$
\begin{aligned}
\text{Arrangements with eye: use } \{eye, a, r, m, r\} &= 5!/2! \\
\text{Arrangements with ram: use } \{ram, r, e, y, e\} &= 5!/2! \\
\text{Arrangements with ram and eye: use } \{ram, eye, r\} &= 3! \\
\text{Arrangements with neither ram nor eye} &= 7!/2!2! - 5!/2! - 5!/2! + 3! = 1146.
\end{aligned}
$$

∎

Problem 6 (15 points). A pizza house is having a promotional sale. Their commercial reads:

> Buy 2 large pizzas at the regular price, and for each pizza get up to 11 different toppings from 11 possible absolutely free (no double toppings). That's $4,194,304$ different ways to design your order!

The ad writer was a former Harvard student who had figured out that $(2^{11})^2 = 4,194,304$. He came up with this number by reasoning that the number of ways to choose different toppings for one pizza is all the possible subsets of the set of 11 toppings, which is 2^{11}. Since there are two pizzas, the total possible combinations of pizzas are $(2^{11})^2$.

Unfortunately, the number $(2^{11})^2$ is actually wrong.

(a) (5 points) Explain what is wrong with the Harvard student's counting.

Solution. The number of ways to choose different toppings for one pizza is all the possible subsets of the set of 11 toppings, which is 2^{11}.

However, when the student says that there are $(2^{11})^2$ different ways to choose toppings for two pizzas, he is making the mistake of counting the combination "one mushroom and anchovy pizza, and one green-pepper pizza" as different from the combination "one green-pepper pizza and one mushroom and anchovy pizza ." (At least he knew enough not to count a mushroom and anchovy pizza as different from an anchovy and mushroom pizza). ∎

(b) (3 points) In how many ways can you choose toppings for the two pizzas?

Solution. The right solution is

$$2^{11} + \binom{2^{11}}{2} = 2,098,176,$$

namely, the number of ways to choose two pizzas with the same topping, plus the number of ways to choose 2 pizzas with different sets of toppings. ∎

(c) (7 points) In how many ways can you choose toppings for n pizzas?

Solution.

$$\binom{2^{11} - 1 + n}{n}.$$

This formula is an instance of the "stars and bars" formula. There are (2^{11}) urns (different types of pizza) and n balls (pizzas). How many different ways can you place the balls into the urns. ∎

Problem 7 (15 points). We consider a variation of Monty Hall's game. The contestant must pick one of *four* doors, with a prize randomly placed behind one door and goats behind the other three. Then, instead of always opening a door to reveal a goat, Carol *randomly* opens two of the three doors that the contestant hasn't picked. This means she may reveal two goats, or she may reveal the prize and a goat. If she reveals the prize, then the entire game is *restarted*, that is, the prize is again randomly placed behind some door, the contestant again picks a door, and so on until Carol finally reveals two goats. Then the contestant can choose to *stick* with his original choice of door or *switch* to the remaining unopened door. He wins if the prize is behind the door he last chooses.

(a) (6 points) Let R be the number of times the game is restarted before Carol picks two goats. What is $\mathrm{E}[R]$?

Solution. Think of *not* having to restart as a failure. So $\Pr\{\text{failure}\} = 1/4 + 3/4 * 1/3 = 1/2$, and $\mathrm{E}[R]$ is mean time to failure minus one—because we are only counting the number of "successes"—namely, $1/\Pr\{\text{failure}\} - 1 = 2 - 1 = 1$. So the answer is 1. ∎

(b) (5 points) When Carol finally reveals two goats, the contestant has the choice of sticking or switching. Let's say that the contestant adopts the strategy of sticking. What is the probability that the contestant wins with this strategy?

Solution. To analyze this setup, we define two events:

GP: The event that the contestant **g**uesses the door with the **p**rize behind it on his first guess.

OP: The event that the game is restarted at least once.

The probability of a win is $1/2$, because

$$\begin{aligned} w &= \Pr\{W \mid GP\}\Pr\{GP\} &&+ \Pr\{W \mid \overline{GP} \cap OP\}\Pr\{\overline{GP} \cap OP\} \\ & && + \Pr\{W \mid \overline{GP} \cap \overline{OP}\}\Pr\{\overline{GP} \cap \overline{OP}\} \\ &= 1 \cdot 1/4 && + w \cdot 3/4 \cdot 2/3 + 0 \cdot 3/4 \cdot 1/3 \\ &= 1/4 + w/2. \end{aligned}$$

So $w(1 - 1/2) = 1/4 \longrightarrow w = 1/2$. ∎

(c) (4 points) For any final outcome where the contestant wins with a "stick" strategy, he would lose if he had used a "switch" strategy, and vice versa. In the original Monty Hall game, we concluded immediately that the probability that he would win with a "switch" strategy was $1 - \Pr\{\text{win with stick}\}$. Why isn't this conclusion quite as obvious for this new, restartable game? Briefly explain why this conclusion still sound.

Solution. Switching strategies turns wins to losses for *terminated* games. So the probability of win with switch is $t - \Pr\{W\}$ where t is the probability of termination. In original Monty Hall, the game obviously terminated after one door-opening, so t was 1.

The extra complication here is that it is possible for the game to run forever, but this event has probability 0, so t is still 1, and the conclusion is still sound. ∎

Problem 8 (10 points). A *simple k-cycle* in a simple graph is an undirected path going through each of k distinct vertices exactly once and ending where it started (where $k > 2$). A simple k-cycle can be represented by the sequence of k distinct vertices $v_1, v_2, \dots, v_k$ along the path.

Note that every simple k-cycle can be represented by many sequences. For example, the 4-cycle represented by 1234 is as the same cycle represented by 2341 or 3412 because a cycle does not have to start at any particular vertex. It is also represented by 4321, because the cycle is undirected.

(a) (3 points) How many different sequences of vertices represent a given simple k-cycle?

Solution. $2k$ ∎

(b) (2 points) In a complete graph on n vertices (i.e. all possible edges are present), how many simple k-cycles are there?

Solution. $P(n, k)/2k$.

The number of ways of choosing a ordered sequence of k vertices is $P(n, k)$, but each simple k-cycle is represented by $2k$ such sequences. ∎

Suppose we construct a simple n-node graph, $G = (V, E)$, randomly as follows: For every set of two distinct vertices $\{v, v'\}$, toss a biased coin whose probability of coming up heads is p. The undirected edge between v and v' is included in E iff the coin comes up heads. Assume that all coin tosses are mutually independent.

(c) (5 points) What is the expected number of simple k-cycles in G? (Note: If you were not able to solve part (b), you may let b denote the answer to part (b) and express your answer in terms of b, k, and p).

Solution. A particular simple k-cycle (call it e) will occur only if all of the edges in that cycle exist in the randomly constructed $G = (V, E)$. Since each edge is independently added with probability p, we have $\Pr\{e \text{ occurs in } G\} = p^k$.

Let I_e be the indicator random variable of the event that a particular k-cycle, e, occurs. So the number, N, of k-cycles is $\sum_e I_e$ where e ranges over all possible k-cycles, *i.e.*, all k-cycles in the complete graph on V.

We know from a previous part that $\Pr\{I_e = 1\} = p^k$, which means that $\mathrm{E}[I_e] = p^k$. Also, the total number of distinct possible k-cycles is $X ::= P(n, k)/2k$.

Therefore

$$\mathrm{E}[N] = \sum_e \mathrm{E}[I_e] = X p^k.$$

∎

Problem 9 (10 points). Let R be a positive integer valued random variable such that

$$f_R(n) = \frac{1}{cn^3},$$

where

$$c ::= \sum_{n=1}^{\infty} \frac{1}{n^3}.$$

(a) (5 points) Prove that $\mathrm{E}\,[R]$ is finite.

Solution.

$$\mathrm{E}\,[R] ::= \sum_{n\in\mathbb{N}^+} n \cdot \frac{1}{cn^3}$$

$$= \sum_{n\in\mathbb{N}^+} \frac{1}{cn^2}$$

$$< 1 + \int_1^\infty \frac{1}{cn^2}$$

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

$$< \infty.$$

Note that $\sum_{n\in\mathbb{N}^+} 1/cn^2 < \infty$ was also shown in Class Problem 1 of Week 7, Friday. ∎

(b) (5 points) Prove that $\mathrm{Var}\,[R]$ is infinite.

Solution. Since

$$\mathrm{Var}\,[R] = \mathrm{E}\,[R^2] - \mathrm{E}^2\,[R],$$

and $\mathrm{E}^2\,[R] < \infty$ by part (a), we need only show that $\mathrm{E}\,[R^2] = \infty$. But

$$\mathrm{E}\,[R^2] ::= \sum_{n\in\mathbb{N}^+} n^2 \frac{1}{cn^3}$$

$$= \sum_{n\in\mathbb{N}^+} \frac{1}{cn}$$

$$= \lim_{n\to\infty} \frac{1}{c} \cdot H_n$$

$$= \infty.$$

Note that $\sum_{n\in\mathbb{N}^+} 1/cn = \infty$ was also shown in Class Problem 1 of Week 7, Friday. ∎

Problem 10 (20 points). Suppose that you are given two biased coins C_p and C_r. Coin C_p flips a head with probability p and coin C_r flips a head with probability r.

The game is to flip C_p and then C_r and keep alternating between coins until you get $2l$ heads in a row, where l is a given positive integer. However, as soon as you get a tail, then

you start over again by first flipping C_p and then alternating between coins[1]. Assume that coin flips are mutually independent of each other.

(a) (10 points) For $l = 3$, write an expression in p and r for the the expected number of coin flips to see $2l$ heads in a row.

Hint: It may be more convenient to use the special expectation formula for natural number valued variables (in the Appendix).

Solution. Solve using Wald's theorem

$$\text{E [number of flips]} = \text{E [number of trials]} \cdot \text{E [length of trial]}$$

$$\text{Pr \{Success of a trial\}} = p \cdot r \cdot p \cdot r \cdot p \cdot r = (pr)^3$$

because we need 6 heads in a row.

$$\text{E [number of trials]} = \frac{1}{\text{Pr \{Success\}}} = \frac{1}{(pr)^3}.$$

Let Q be the length of the try.

$$\begin{aligned}
\text{E}[Q] &= \sum_{i=0}^{5} \text{Pr}\{Q > i\} \\
&= 1 + p + p \cdot r + p \cdot r \cdot p + p \cdot r \cdot p \cdot r + p \cdot r \cdot p \cdot r \cdot p \\
&= 1 + p + pr + p^2 r + p^2 r^2 + p^3 r^2
\end{aligned} \tag{3}$$

Therefore, the **ANSWER** is:

$$\text{E [number of flips]} = \frac{1 + p + pr + p^2 r + p^2 r^2 + p^3 r^2}{(pr)^3}$$

Alternate solution:

You can also compute $\text{E}[Q]$ directly from the normal expectation formula

$$\begin{aligned}
\text{E}[Q] ::= &1 \cdot (1 - p) \\
&+ 2 \cdot p \cdot (1 - r) \\
&+ 3 \cdot p \cdot r \cdot (1 - p) \\
&+ 4 \cdot p \cdot r \cdot p \cdot (1 - r) \\
&+ 5 \cdot p \cdot r \cdot p \cdot r \cdot (1 - p) \\
&+ 6 \cdot p \cdot r \cdot p \cdot r \cdot p \cdot (1 - r) \\
&+ 6 \cdot p \cdot r \cdot p \cdot r \cdot p \cdot r \\
= &(1 - p) + 2p(1 - r) + 3pr(1 - p) + 4p^2 r(1 - r) + 5p^2 r^2(1 - r) + 6p^3 r^2 \tag{4}
\end{aligned}$$

---

[1]For example, if $l = 2$ and the sequence of flips was $HTTHHTHHHH$, then the game took 10 flips and the sequence of coins used was $C_p C_r C_p C_p C_r C_p C_p C_r C_p C_r$

When you simplify (4) you get (3). ∎

(b) (10 points)   What is the expected number of coin flips to see $2L$ heads in a row?

Note: You may express your answer using summations determined by L, but if you do, you should briefly indicate how results from the course imply that there are closed forms for your summations.

Solution.

$$\Pr\{\text{Success of a trial}\} = (pr)^L$$
$$\mathbb{E}\,[\text{number of trials}] = \frac{1}{\Pr\{\text{Success}\}} = \frac{1}{(pr)^L}$$

Let Q be the length of the try.

$$\mathbb{E}\,[Q] = \sum_{i=0}^{2L-1} \Pr\{Q > i\}.$$

If i is even ($i = 2k$), then $\Pr\{Q > i\} = (pr)^k$; if i is odd ($i = 2k+1$) then $\Pr\{Q > i\} = (pr)^k p$. Therefore

$$\mathbb{E}\,[Q] = \sum_{k=0}^{L-1} (pr)^k + \sum_{k=0}^{L-1} (pr)^k p.$$

Both of these sums are simple geometric series so it is possible to find closed forms. ∎

Problem 11 (10 points). (a) (5 points)   Let R be an indicator variable for getting a head on a flip of a fair coin. Calculate the bound on $\Pr\{R \geq 1\}$ using

(i) Markov Bound

(ii) Chebyshev Bound

(iii) Chernoff Bound

Note: $e^{\ln 2} \approx 2$

Solution. In this case $\mu_R = 1/2$ and $\sigma_R^2 = 1/4$, so

- Markov gives $\mu_R/y = 1/2$ which is exactly right,

- The bound from the Chebyshev is

$$\frac{\sigma_R^2}{(y - \mu_R)^2} = \frac{1/4}{(1 - (1/2))^2} = 1$$

However if we assume that the distribution is symmetric about the mean, then Chebyshev bound gives us $1/2$ and so is exactly tight.

- the bound from the Chernoff result is

$$\exp(-(2\ln 2 - 2 + 1)/2) = e^{1/2 - \ln 2} = \sqrt{e}/2 > 0.83,$$

and so is a large overestimate.

∎

(b) (5 points) Prove that $\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y]$, if X, Y are independent. Be sure to justify each step in your proof.

Solution. We will transform the left side into the right side. We begin by applying the alternate definition of variance:

$$\text{Var}[X + Y] = \text{E}\left[(X + Y)^2\right] - \text{E}^2[X + Y]. \tag{5}$$

We will work on the first term and second terms on the righthand side of (5) separately. For the first term,

$$
\begin{aligned}
\text{E}\left[(X + Y)^2\right] &= \text{E}\left[X^2 + 2XY + Y^2\right] \\
&= \text{E}\left[X^2\right] + 2\,\text{E}[XY] + \text{E}\left[Y^2\right] && \text{(linearity of E [])} \\
&= \text{E}\left[X^2\right] + 2\,\text{E}[X]\,\text{E}[Y] + \text{E}\left[Y^2\right]. && \text{(independent multiplicativity of E [])}
\end{aligned}
\tag{6}
$$

Now we work on the second term on the righthand side of (5).

$$
\begin{aligned}
\text{E}^2[X + Y] &= (\text{E}[X] + \text{E}[Y])^2 && \text{(linearity of E [])} \\
&= \text{E}^2[X] + 2\,\text{E}[X]\,\text{E}[Y] + \text{E}^2[Y].
\end{aligned}
\tag{7}
$$

Subtracting (7) from (6), we have

$$
\begin{aligned}
\text{Var}[X + Y] &= (\text{E}\left[X^2\right] + 2\,\text{E}[X]\,\text{E}[Y] + \text{E}\left[Y^2\right]) - (\text{E}^2[X] + 2\,\text{E}[X]\,\text{E}[Y] + \text{E}^2[Y]) \\
&= (\text{E}\left[X^2\right] - \text{E}^2[X]) + (\text{E}\left[Y^2\right]) - \text{E}^2[Y]) \\
&= \text{Var}[X] + \text{Var}[Y].
\end{aligned}
$$

∎

Problem 12 (10 points). A test tube with a bacterial culture is delivered to a Lab technician at a hospital. Her task is to estimate the percentage of bacteria in the culture that are resistant to penicillin. She will do this by sampling a certain number of bacteria and testing them. Assume her sampling technique is good, *i.e.*, the bacteria are sampled randomly and independently with replacement. Let the percentage of antibiotic-resistant bacteria be b.

(a) (6 points) Write a formula for a number, n, of samples she could take in order to be C% confident that her estimation was no more than d away from the actual value b. The formula for n must not, of course, involve the unknown quantity, b.

Solution. Let p be the probability that the estimation is within d of the actual value b. The Pairwise Sampling Theorem tells us that

$$p ::= \Pr\left\{\left|\frac{S_n}{n} - b\right| \geq d\right\} \leq \frac{\sigma^2}{nd^2}$$

In this case, we are modelling sampling bacteria as n coin flips, and the variance of a flip is $b(1-b)$. This is maximized when $b = 1/2$. This gives us the bound $p \leq 1/4nd^2$. Our percent confidence level is $100(1-p) \geq 100(1 - 1/(4nd^2))$, so in order to be at least 95% confident, we need

$$100(1 - \frac{1}{4nd^2}) \geq 95.$$

Solving for n, we get

$$0.05 \geq \frac{1}{4nd^2}$$
$$n \geq \frac{1}{4 \cdot 0.05c^2} = \frac{5}{c^2}.$$

∎

(b) (4 points) The lab technician is supposed to perform her task daily. She is instructed to warn the medical staff whenever her tests show with 99% confidence that the percentage of penicillin-resistant bacteria in that day's tube exceeds a specified danger threshold.

In early April, the staff receives its first warning of the year from her, and a new resident on the staff orders the nurses to review the medication of all patients taking penicillin, because he realizes there is a very high probability of a dangerous level of penicillin-resistance bacteria in the hospital. When a more senior physician arrives, he overrules the residents' order.

Briefly indicate how you would justify the senior physician's decision to the resident.

Solution. Based on the technician's warning, the resident thought it was a good bet that there were dangerous levels of penicillin-resistance in the hospital. He was exhibiting the common confusion between probability and confidence level.

With 99% confidence level of reporting, we *expect* the technician to get mistaken results about 1% of the time. So after the 100 days to mid-April of reporting no danger, we could expect her to report a dangerous threshold when there isn't any. The resident shouldn't order extra work for the nursing staff when in this situation, he can expect that the danger report will usually be a "false positive.". So the senior physician's decision not to order a full scale review of medication is a sound one.

On the other hand, if the physician is never going to pay attention to the technician's warning, it's silly to waste her time doing estimates. In the face of the technician's warning, a sensible action to take would be to ask her to repeat her estimation. There is only a 1% chance she would report danger again if there was none, so if she does report it after the repeat estimate, it would be prudent to order the medication review. ∎