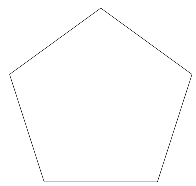
Final exam sample solutions

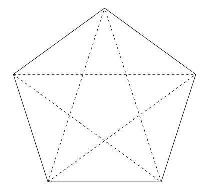
Please— do not read or discuss these solutions in the exam room while others are still taking the exam.

Problem 1. [Proof by Induction] (14 points)

A *diagonal* of a polygon is a line connecting two different, non-adjacent vertices. Let $d(n)$ denote the number of diagonals in a polygon with n vertices.

For instance, shown below on the left is a polygon with five vertices.





On the right each diagonal has been drawn as a dashed line. (The solid lines are not diagonals.) We can see there are five diagonals. Therefore, $d(5) = 5$.

(a) Calculate $d(3)$. You do not need to justify your answer.
$d(3) = \boxed{0}$

(b) Calculate $d(4)$. You do not need to justify your answer.
$d(4) = \boxed{2}$

(c) Prove by induction that $d(n) = \frac{n(n-3)}{2}$ for every $n \geq 3$.

Base case: For $n = 3$, $d(3) = 0$ as calculated in part (a), and indeed $\frac{3\times 0}{2} = 0$.
Induction hypothesis: Suppose $d(n) = \frac{n(n-3)}{2}$ for some n with $n \geq 3$.
Inductive step: We will start by proving that

$$d(n+1) = d(n) + n - 1.$$

To justify this, consider a polygon with $n+1$ vertices. Let v be one of its vertices, with neighbors u and w. Remove v and replace the two edges u-v and v-w with the diagonal connecting u to w. This gives us a smaller polygon with n vertices. Obviously, this smaller polygon has $d(n)$ diagonals.
How many diagonals does the original polygon (with $n+1$ vertices) have? Well, there are three kinds of diagonals in the original polygon:

- Every diagonal of the smaller polygon is a diagonal of the original polygon.
- The line from u to w is a diagonal of the original polygon, but not of the smaller polygon.
- There is a diagonal from v to each of the $n-2$ other non-adjacent vertices in the original polygon. (This diagonal is not present in the smaller polygon, since v is not a vertex of the smaller polygon.)

These are all of the diagonals of the original polygon. There are $d(n)$ diagonals of the first type, 1 diagonal of the second type, and $n-2$ diagonals of the third type. Therefore,

$$d(n+1) = d(n) + 1 + n - 2 = d(n) + n - 1,$$

as claimed earlier.
Now, by the induction hypothesis, $d(n) = \frac{n(n-3)}{2}$, so

$$d(n+1) = d(n) + n - 1 = \frac{n(n-3)}{2} + n - 1 = \frac{n^2 - 3n}{2} + \frac{2n-2}{2} = \frac{n^2 - n - 2}{2} = \frac{(n+1)(n-2)}{2}.$$

Problem 2. [True/False] (10 points)

Do not justify your answers on this problem. You will receive 2 points for each part that you answer correctly, 1 point for each part you leave blank, and 0 points for each part that you answer incorrectly, so if you are not sure, you may want to leave it blank.

For each of the following, determine if the implication is true for all predicates P and Q or if it may be false for some predicates. Circle either YES or NO accordingly.

(a) $(\forall n \in \mathbb{N})(\neg(P(n) \iff Q(n))) \implies (\forall n \in \mathbb{N})(P(n) \vee Q(n))$:
true for all P,Q? [YES] NO

Comment: If $\neg(P(n) \iff Q(n))$ is true, then either $P(n)$ is true or $Q(n)$ is.

(b) $(\forall n \in \mathbb{N})(P(n) \implies Q(n)) \implies (\neg\exists n \in \mathbb{N})(Q(n) \implies \neg P(n))$:
true for all P,Q? YES [NO]

Comment: A counterexample is $P(n) =$ false and $Q(n) =$ false for all n.

(c) $\Big(P(3) \wedge (\forall n \in \mathbb{N})\big(P(n+1) \iff (P(n) \wedge P(n-1) \wedge \cdots \wedge P(0))\big)\Big) \implies (\forall n \in \mathbb{N})(P(n))$:
true for all P,Q? [YES] NO

Comment: Instantiating the universally quantified statement with $n = 2$, we see that $P(2) \wedge P(1) \wedge P(0)$ is true. Now $(\forall n \in \mathbb{N})(P(n))$ follows by induction on n.

For each of the following, circle either TRUE or FALSE. $\mathbb{N}$ denotes the set of natural numbers and $\mathbb{R}$ denotes the set of real numbers.

(d) [TRUE] or FALSE: The set $S = \{x \in \mathbb{R} : (\exists n \in \mathbb{N})(x^2 = n)\}$ is countable.

Comment: We can enumerate all the elements of S: $0, \sqrt{1}, -\sqrt{1}, \sqrt{2}, -\sqrt{2}, \sqrt{3}, -\sqrt{3}, \ldots$.

(e) TRUE or [FALSE]: The function $f : \{0,1\}^* \times \{0,1\}^* \to \{0,1\}$ is computable, where f is defined by

$$f(P,I) = \begin{cases} 1 & \text{if program } P \text{ eventually halts and returns 3 when run on input } I, \\ 0 & \text{if } P \text{ never halts, or eventually halts and returns something other than 3, when run on input } I. \end{cases}$$

Comment: If you could compute f, then you could solve the halting problem as follows. Given a program P and an input I, where we want to know whether P eventually halts on input I, define the program Q (which implicitly includes the code of P as a subroutine) as follows:

$Q(I)$:
1. Call $P(I)$ (as a subroutine).
2. Return 3.

Now $f(Q,I) = 1$ if and only if P eventually halts on input I. But we know the halting problem is not computable, so therefore f must not be computable either.

Problem 3. [Probability and Counting] (14 points)

Suppose X is a number chosen uniformly at random from the set $\{0,1,\ldots,419\}$. Do not justify your answers on this problem.

(a) What is the probability that $X = 4$?

$\boxed{1/420}$

(b) What is the probability that $3X \equiv 4 \pmod{20}$?

$\boxed{1/20}$

Comment: Since $\gcd(3,20) = 1$, 3 has an inverse modulo 20, so the equation has a single solution modulo 20 (in fact, it is $X \equiv 8 \pmod{20}$). Therefore, there is one solution of this equation between $0 \ldots 19$, another solution between $20 \ldots 39$, and so on (in fact, the solutions are 8, 28, 48, and so on). In total, there are 21 solutions in the range $0 \ldots 419$.

(c) What is the probability that $3X \equiv 4 \pmod{21}$?

$\boxed{0}$

Comment: There are no integers X and n such that $3X = 21n + 4$ since two terms in the equation are divisible by 3 and the other term is not. Therefore, there is no X satisfying $3X \equiv 4 \pmod{21}$.

(d) Let q be a large prime. Alice has a message, which she represents as a polynomial $P(x)$ of degree $n-1$, working modulo q. She wants to sends the message over an unreliable communication channel to Bob. Therefore, she uses the error-correcting code described in lecture to encode the message before transmission: in other words, Alice sends the packets $c_1 = P(1)$, $c_2 = P(2)$, $c_3 = P(3)$, $\ldots$, $c_{n+k} = P(n+k)$ (all reduced modulo q). Unfortunately, Bob receives only $n-3$ of these packets; the rest are all lost. Bob decides to enumerate all of the messages that Alice might have sent, i.e., all the messages that are consistent with the $n-3$ packets he received. Given the $n-3$ packets that Bob received, how many possible messages are there?

$\boxed{q^3}$

Comment: If we knew the value of 3 more packets, Alice's message would be uniquely determined. There are q possibilities for each packet, and they can be independently chosen, so there are $q \times q \times q = q^3$ possible values for those 3 packets.

Problem 4. [Probability] (10 points)

A health study tracked a group of people for five years. At the beginning of the study, 20% were classified as heavy smokers, 30% as light smokers, and 50% as nonsmokers. Results of the study showed that light smokers were twice as likely as nonsmokers to die during the five-year study, but only half as likely as heavy smokers.

Suppose we select, uniformly at random, a participant from this study, and it turns out that this participant died at some point during the five-year period. Calculate the probability that this participant was classified as a heavy smoker at the beginning of the study. Show your calculation clearly.

Define the events H = heavy smoker, L = light smoker, N = non-smoker, D = dies. Let $p = \Pr[D|N]$, so

$$\Pr[N] = 0.5 \qquad \Pr[D|N] = p$$
$$\Pr[L] = 0.3 \qquad \Pr[D|L] = 2p$$
$$\Pr[H] = 0.2 \qquad \Pr[D|H] = 4p.$$

Using the total probability rule,

$$\Pr[D] = \Pr[D|N]\Pr[N] + \Pr[D|L]\Pr[L] + \Pr[D|H]\Pr[H] = 0.5p + 0.6p + 0.8p = 1.9p.$$

By Bayes' rule,

$$\Pr[H|D] = \frac{\Pr[D|H]\Pr[H]}{\Pr[D]} = \frac{0.8p}{1.9p} = \frac{8}{19}.$$

Problem 5. [Random Coalitions] (15 points)

The country of Elbonia has two political parties, the Republicrats and the Democans. The Elbonian House of Representatives is formed by choosing 100 Elbonians uniformly at random. For part (a), find the exact probability, as an expression with no summations. For parts (b) and (c), give a numerical value that approximates the probability you are asked for (you do not need to calculate the probability exactly; an approximation is sufficient). Show your work on all parts.

(a) If each political party has the support of exactly half the Elbonian population, what is the probability that the Republicrat party ends up with a majority in the House (i.e., there are more than 50 Republicrat members of the House)? Your answer should be an exact expression, with no summations.

Let R = the number of Republicrat members of the House. Now,

$$\Pr[R < 50] + \Pr[R = 50] + \Pr[R > 50] = 1.$$

By symmetry, $\Pr[R < 50] = \Pr[R > 50]$. Also, $R \sim \text{Binomial}(100, \frac{1}{2})$, so $\Pr[R = 50] = \binom{100}{50}2^{-100}$. Plugging in these observations, we find

$$\binom{100}{50}2^{-100} + 2\Pr[R > 50] = 1,$$

and thus,

$$\Pr[R > 50] = \tfrac{1}{2} - \binom{100}{50}2^{-101}.$$

(b) The president happens to be a Democan, so the Republicrat party needs a veto-proof majority of 70% of the House to pass its desired legislation. However, due to a massive scandal involving the Democan president, many Elbonians have shifted their party allegiance such that Elbonia is now 80% Republicrat. If new representatives are chosen for the House, approximately what is the probability that the Republicrats will have a veto-proof majority in the House? Your answer should be a number that approximates this probability, accurate to 2 digits after the decimal point.

Let $R =$ the number of Republicrat members of the House. Then $R \sim \text{Binomial}(100, 0.8)$, so $\mathbb{E}(R) = 80$ and $\text{Var}(R) = 100 \times \frac{4}{5} \times \frac{1}{5} = 16$.
Now we'll use the normal approximation and convert to a standard normal distribution. Defining $Z = \frac{R-80}{4}$, we find that Z has approximately the standard normal distribution. Then

$$\Pr[R \geq 70] \approx \Pr[Z \geq \tfrac{70-80}{4}] = \Pr[Z \geq -2.5] = \Pr[Z \leq 2.5] \approx 0.9938.$$

(c) The Elbonian Senate is formed by choosing 25 Elbonians at random from each of the four provinces of Elbonia, for a total of 100 senators. Suppose that all Elbonians from the president's home province are Democan, all residents of the Republicrat leader's province are Republicrat, and $\frac{2}{3}$ of Elbonians from each of the remaining two provinces are Republicrat. The Senate has different rules than the House: in the Senate, a 60% majority is veto-proof. What is the approximate probability that the Republicrats will achieve a veto-proof majority in the Senate? Your answer should be a number that approximates this probability, accurate to 2 digits after the decimal point.

Let $T =$ the number of Republicrat members of the Senate. Let $U =$ the number of Republicat Senators from the last two provinces. Then $U \sim \text{Binomial}(50, \frac{2}{3})$, so $\mathbb{E}(U) = \frac{100}{3}$ and $\text{Var}(U) = 100 \times \frac{2}{3} \times \frac{1}{3} = \frac{100}{9}$.
Now we wish to compute $\Pr[T \geq 60]$. Note that $T = U + 25$, so $\Pr[T \geq 60] = \Pr[U \geq 35]$. Again, we'll use the normal approximation. Define $Z = \frac{U - 100/3}{10/3}$, so that Z has approximately a standard normal distribution. Then

$$\Pr[U \geq 35] \approx \Pr[Z \geq \tfrac{35-100/3}{10/3}] = \Pr[Z \geq 0.5] = 1 - \Pr[Z \leq 0.5] \approx 1 - 0.6915 = 0.3085.$$

Problem 6. [Law of Large Numbers] (8 points)

The *Law of Large Numbers* is said to hold for a sequence of random variables $S_1, S_2, S_3, S_4, \ldots$ if for every $\varepsilon > 0$,

$$\lim_{n\to\infty} \Pr[\,\left|\tfrac{1}{n}S_n - \mathbb{E}(\tfrac{1}{n}S_n)\right| > \varepsilon] = 0.$$

In the course we have shown that the Law of Large Numbers holds if $S_n = X_1 + \cdots + X_n$, where the X_i's are i.i.d. random variables. This problem explores if the Law of Large Numbers hold under other circumstances.

Packets are sent from a source to a destination node over the Internet. Each packet is sent on a certain route. Each route has a failure probability of p and different routes fail independently. If a route fails, all packets sent along that route are lost. You can assume that the routing protocol has no knowledge of which route fails.

For each of the following routing protocols, determine whether the Law of Large Numbers holds when S_n is defined as the total number of received packets out of n packets sent. Circle YES if the Law of Large Number holds, or NO if not. Do not justify your answer. You will receive 2 points for each part that you answer correctly, 1 point for each part you leave blank, and 0 points for each part that you answer incorrectly, so if you are not sure, you may want to leave it blank. (Whenever convenient, you can assume below that n is even.)

(a) [YES] or NO: Each packet is sent on a completely different route.

Comment: Here $S_n \sim \text{Binomial}(n, 1-p)$, and in particular, $S_n = X_1 + \cdots + X_n$ where X_i is 0 with probability p and 1 otherwise, so this is the classic situation for the Law of Large Numbers.

(b) $\boxed{\text{YES}}$ or NO: The packets are split into $n/2$ pairs of packets. Each pair is sent together on its own route (i.e., different pairs are sent on different routes).

Comment: Here $S_n/2 \sim \text{Binomial}(n/2, 1-p)$, so a Law of Large Numbers holds here too.

(c) YES or $\boxed{\text{NO}}$: The packets are split into 2 groups of $n/2$ packets. All the packets in each group are sent on the same route, and the two groups are sent on different routes.

Comment: Here

$$\tfrac{1}{n}S_n = \begin{cases} 0 & \text{with probability } p^2 \\ 1/2 & \text{with probability } 2p(1-p) \\ 1 & \text{with probability } (1-p)^2, \end{cases}$$

so the Law of Large Numbers does not hold: $\left|\frac{1}{n}S_n - (1-p)\right|$ is either $1-p$, $|p-\frac{1}{2}|$, or p, no matter how large n is. In particular, $\Pr[\left|\frac{1}{n}S_n - (1-p)\right| > \varepsilon]$ is strictly positive when ε is smaller than all three of these quantities.

(d) YES or $\boxed{\text{NO}}$: All the packets are sent on one route.

Comment: Here $\frac{1}{n}S_n$ is either 0 (with probability p) or 1 (with probability $1-p$). Therefore, $\left|\frac{1}{n}S_n - (1-p)\right|$ is either p or $1-p$, no matter how large n is. It follows that $\Pr[\left|\frac{1}{n}S_n - (1-p)\right| > \varepsilon]$ is strictly positive for small values of ε, no matter how large n is.

Problem 7. [Actuarial Calculations] (15 points)

A city has just added 100 new female recruits to its police force. The city will provide a pension to each new hire who remains with the force until retirement. In addition, if the new hire is married at the time of her retirement, a second pension will be provided for her husband. A consulting actuary makes the following assumptions:

- Each new recruit has a 0.4 probability of remaining with the police force until retirement.
- Given that a new recruit reaches retirement with the police force, the probability that she is not married at the time of retirement is 0.25.
- The number of pensions that the city will provide on behalf of each new hire is independent of the number of pensions it will provide on behalf of any other new hire.

Let the random variable P denote the total number of pensions that the city will provide to the 100 new hires and their husbands. Show your calculations for each part.

(a) Compute $\mathbb{E}(P)$.

Let P_i = the number of pensions provided for the ith hire and her husband. Let R_i be the event that the ith hire remains with the department until retirement, and M_i the event that she is married at the time of retirement. Then

$$\Pr[R_i] = 0.4 \qquad\qquad \Pr[\overline{M_i}|R_i] = 0.25.$$

We now compute the distribution of P_i.

$$\begin{aligned}
\Pr[P_i = 0] &= \Pr[\overline{R_i}] = 0.6 \\
\Pr[P_i = 1] &= \Pr[R_i \cap \overline{M_i}] = \Pr[\overline{M_i}|R_i]\Pr[R_i] = 0.25 \times 0.4 = 0.1 \\
\Pr[P_i = 2] &= \Pr[R_i \cap M_i] = \Pr[M_i|R_i]\Pr[R_i] = 0.75 \times 0.4 = 0.3.
\end{aligned}$$

Then $\mathbb{E}(P_i) = 0.1 \times 1 + 0.3 \times 2 = 0.7$.
Finally, since $P = \sum_{i=1}^{n} P_i$, $\mathbb{E}(P) = \sum_{i=1}^{n} \mathbb{E}(P_i) = 70$.

(b) Compute $\mathrm{Var}(P)$.

First, we compute $\mathbb{E}(P_i^2) = 0.1 \times 1 + 0.3 \times 2^2 = 1.3$.
Then $\mathrm{Var}(P_i) = \mathbb{E}(P_i^2) - \mathbb{E}(P_i)^2 = 1.3 - 0.49 = 0.81$. Since the P_i's are mutually independent,

$$\mathrm{Var}(P) = \sum_{i=1}^{n} \mathrm{Var}(P_i) = 81.$$

(c) Determine a non-trivial upper bound on $\Pr[P \geq 90]$.
(An upper bound is non-trivial if it is strictly less than 1.)

By Markov's inequality, we have $\Pr[P \geq 90] \leq \frac{\mathbb{E}(P)}{90} = \frac{7}{9}$.
Alternatively, by Chebyshev's inequality, we have

$$\Pr[P \geq 90] = \Pr[P - 70 \geq 20] \leq \Pr[|P - 70| \geq 20] \leq \frac{\mathrm{Var}(P)}{20^2} = \frac{81}{400}.$$

Problem 8. [Car Insurance] (14 points)

The Elbonian currency is the elbomark. An Elbonian company sells car insurance to Elbonian individuals. When an individual's car is damaged, the individual reports the amount of money they lost to the insurance company. Assume that each automobile loss reported to the insurance company is uniformly distributed between 0 and 2 elbomarks. The loss can take on any real number in this range.

The insurance company covers each such loss as follows: if the loss is less than 1 elbomarks, then the insurance covers half of it; if the loss is more than 1 elbomarks, then the insurance covers half of the first 1 elbomark, and any amount in excess of that will be covered in full. For example, if the loss is 1.8 elbomarks, the amount paid to the individual is $0.5 + (1.8 - 1) = 1.3$ elbomarks.

Show your work on all parts.

(a) Compute and plot the cdf (cumulative distribution function) of the amount paid to the individual.

Let Y = the amount of damage to the individual's car, X = the insurance payout. Then

$$X = \begin{cases} \frac{1}{2}Y & \text{if } 0 \leq Y \leq 1, \\ Y - \frac{1}{2} & \text{if } 1 \leq Y \leq 2. \end{cases}$$

We first compute the cdf of Y:

$$\Pr[Y \leq y] = \begin{cases} 0 & \text{if } y < 0, \\ \frac{y}{2} & \text{if } 0 \leq y \leq 2, \\ 1 & \text{if } y > 2. \end{cases}$$

Now we compute the cdf of X. If $0 \leq x \leq \frac{1}{2}$,

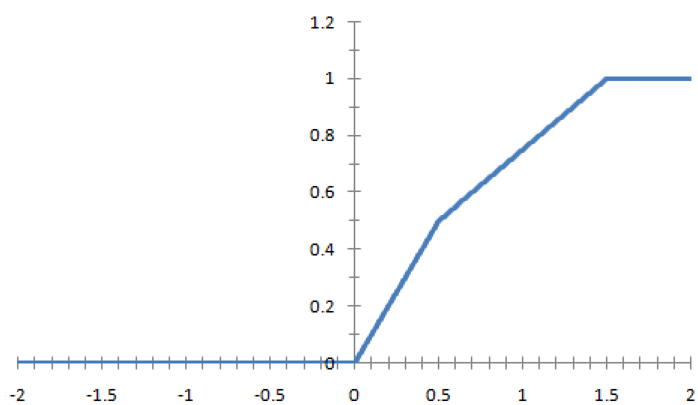
$$\Pr[X \leq x] = \Pr[\tfrac{1}{2}Y \leq x] = \Pr[Y \leq 2x] = x.$$

If $\frac{1}{2} \leq x \leq \frac{3}{2}$,

$$\Pr[X \leq x] = \Pr[Y - \tfrac{1}{2} \leq x] = \Pr[Y \leq x + \tfrac{1}{2}] = \tfrac{1}{2}x + \tfrac{1}{4}.$$

Therefore the cdf of X is

$$\Pr[X \leq x] = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } 0 \leq x \leq \frac{1}{2} \\ \frac{1}{2}x + \frac{1}{4} & \text{if } \frac{1}{2} \leq x \leq \frac{3}{2} \\ 1 & \text{if } x > \frac{3}{2}. \end{cases}$$



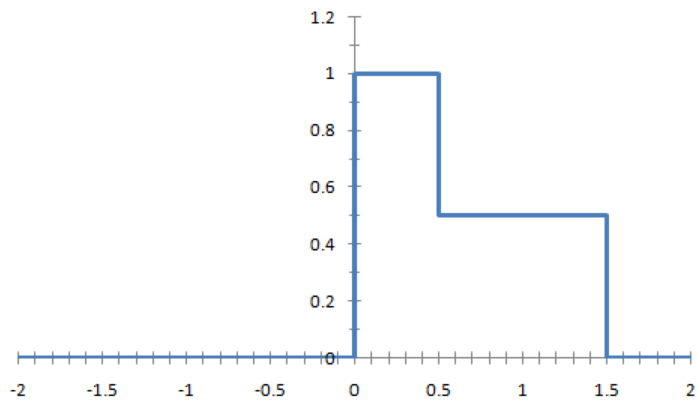
(b) Compute and plot the pdf (probability density function) of the amount paid to the individual.

The pdf of X can be calculated by differentiating the cdf of X:

$$f(x) = \frac{d}{dx}\Pr[X \le x].$$

In this way, we obtain the following pdf:

$$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } 0 \le x \le \frac{1}{2} \\ \frac{1}{2} & \text{if } \frac{1}{2} \le x \le \frac{3}{2} \\ 0 & \text{if } x > \frac{3}{2}. \end{cases}$$



(c) Compute the expectation of the amount paid to the individual.

$$\begin{aligned} \mathbb{E}(X) &= \int_{-\infty}^{\infty} x f(x)\,dx \\ &= \int_0^{\frac{1}{2}} x\,dx + \int_{\frac{1}{2}}^{\frac{3}{2}} \frac{1}{2}x\,dx \\ &= \left[\frac{1}{2}x^2\right]_0^{\frac{1}{2}} + \left[\frac{1}{4}x^2\right]_{\frac{1}{2}}^{\frac{3}{2}} \\ &= \frac{1}{8} + \frac{9}{16} - \frac{1}{16} \\ &= \frac{5}{8}. \end{aligned}$$

Congratulations on finishing the semester. Hope you did great. Here's a last joke for you:

Q: Why couldn't Pythagoras get a car loan?
A: He couldn't find anyone to cosine!