CS 70 Discrete Mathematics and Probability Theory
Spring 2015 Vazirani Final Exam Solution

PRINT your name: _____, _____
(last) (first)

SIGN your name: _____

PRINT your student ID: _____

CIRCLE your exam room: Wheeler Auditorium 220 Hearst Gym

Name of the person sitting to your left: _____

Name of the person sitting to your right: _____

Please write your name and student ID on every page (**after** the exam starts).

Please write your answers in the spaces provided in the test. There are extra pages at the end that you can use as scratch papers. We will not grade them unless you clearly tell us in the problem's main page to look there.

You may consult three single-sided sheets of handwritten notes. Apart from that, you may not look at books, notes, etc. Calculators and computers are not permitted.

You have 180 minutes. There are 11 questions provided, but you need to do only 9 for a total of 170 points:

- Choose three out of questions 4, 5, 6, and 7.
- Choose three out of questions 8, 9, 10, and 11.

Clearly mark which three questions you want us to grade (you might even cross out the question you don't want us to grade). Please note that there is no advantage to attempting all four questions — if you do not indicate which three questions you want us to grade, then we will only grade the first three questions.

You may use the number of points as a rough guide for the amount of time to allocate to that question. Note that many of the points are for proofs and justifications for your answers. Please make sure you spend the time to write clear, correct and concise justifications. Also make sure you read the questions carefully. Good luck!

Do not turn this page until your instructor tells you to do so.

Name: _____ SID: _____

1. **True or False (30 points)**

   For each question below, circle True if the statement is true, and circle False if the statement is false. No justification needed. 2 points per question; no partial credit.

   (a) A proposition and its contrapositive cannot both be true.
   Circle one: **True** **False**
   Answer: False. A proposition is always equivalent to its contrapositive.

   (b) The proposition $(A \wedge B) \vee (\neg A \wedge B) \vee \neg B$ can never be false.
   Circle one: **True** **False**
   Answer: True. The proposition can be simplified to $B \vee \neg B$, which is always true.

   (c) The hypercube graph always has an Eulerian tour.
   Circle one: **True** **False**
   Answer: False. To have an Eulerian tour the degree of each vertex must be even. For the hypercube graph this is only true when the dimension is even.

   (d) If $f : A \to B$ is an injective (1-1) function, then there exists a surjective (onto) function $g : B \to A$.
   Circle one: **True** **False**
   Answer: True. We can construct g by assigning $g(b) = a$ if there exists $a \in A$ with $f(a) = b$, and assigning $g(b)$ arbitrarily otherwise.

   (e) If $\gcd(a,b) = d$, then a has no factor larger than d.
   Circle one: **True** **False**
   Answer: False. $\gcd(a,b) = d$ means a and b have no common factor larger than d, but a itself can still have factors larger than d.

   (f) In RSA with modulus $n = 91$ and encryption power $e = 5$, the decryption power is $d = 73$ because $de = 365 \equiv 1 \bmod 91$.
   Circle one: **True** **False**
   Answer: False. We need $de \equiv 1 \bmod (p-1)(q-1)$. In this case $p = 7$, $q = 13$, so $(p-1)(q-1) = 72$, and we need $d = 29$ so that $de = 145 \equiv 1 \bmod 72$.

   (g) If the multiplicative inverse $a^{-1} \bmod p$ exists for all $a \in \{1, \ldots, p-1\}$, then p is a prime.
   Circle one: **True** **False**
   Answer: True. The condition implies p is relatively prime to all $a \in \{1, \ldots, p-1\}$, which means p is a prime.

   (h) For any $d \in \mathbb{N}$, the set of polynomials of degree d with integer coefficients is countable.
   Circle one: **True** **False**
   Answer: True. The set of polynomials of degree d with integer coefficients is in bijection to $\mathbb{Z}^{d+1}$, which is countable.

(i) If a sample space Ω has n sample points, then there are 2^n possible events.
Circle one: **True** **False**
Answer: True. An event is just a subset of the sample space.

(j) If $P(A \mid B) = 1$, then $P(B) \leq P(A)$.
Circle one: **True** **False**
Answer: True. By Bayes' rule $P(B)/P(A) = P(B \mid A)/P(A \mid B) = P(B \mid A) \leq 1$.

(k) For any random variable X, $\mathbb{E}(X^2) \geq \mathbb{E}(X)^2$.
Circle one: **True** **False**
Answer: True. $\mathbb{E}(X^2) - \mathbb{E}(X)^2 = \text{Var}(X) \geq 0$.

(l) For any random variables X and Y, $\mathbb{E}(X - Y) = \mathbb{E}(X) - \mathbb{E}(Y)$.
Circle one: **True** **False**
Answer: True. Linearity of expectation.

(m) If X and Y are independent random variables, then $\mathbb{E}(X/Y) = \mathbb{E}(X)/\mathbb{E}(Y)$.
Circle one: **True** **False**
Answer: False. $\mathbb{E}(X/Y) = \mathbb{E}(X)\,\mathbb{E}(1/Y)$, but $\mathbb{E}(1/Y) \neq 1/\mathbb{E}(Y)$.

(n) If X and Y are independent random variables, then $\text{Var}(X) = \text{Var}(X - Y) - \text{Var}(Y)$.
Circle one: **True** **False**
Answer: True. $\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y)$.

(o) If X is an exponential random variable, then $P(X \geq s + t \mid X \geq s) = P(X \geq s + t)$.
Circle one: **True** **False**
Answer: False. Should be $P(X \geq s + t \mid X \geq s) = P(X \geq t)$.

2. **Short answers (40 points)**

(a) What distribution would best model each of the following scenarios? Choose from binomial, Poisson, geometric, exponential, uniform, or normal distribution. No justification needed.

i. Number of taxis passing the corner of Euclid Ave and Hearst Ave between 5 pm and 6 pm on a weekday.
Answer: Poisson

ii. Number of customers who purchase a lottery ticket before someone hits the jackpot.
Answer: Geometric

iii. Number of balls in the first urn in a Polya urn process, with two urns each starting with one ball.
Answer: Uniform

iv. Times of finishers in the NY Marathon.
Answer: Normal

v. Number of girls in a family with 6 kids.
Answer: Binomial

vi. Number of miles a car can run before the engine fails.
Answer: Exponential

(b) What is the number of poker (5 card) hands with 2 pairs? Explain your calculation.
(A poker hand has 5 cards. The 2 pairs must be of different ranks, and the last card must also be different. So $(2\diamondsuit, 2\clubsuit, 4\heartsuit, 4\spadesuit, 8\diamondsuit)$ is an example of a poker hand with 2 pairs, but $(2\diamondsuit, 2\clubsuit, 2\heartsuit, 2\spadesuit, 8\diamondsuit)$ and $(2\diamondsuit, 2\clubsuit, 4\heartsuit, 4\spadesuit, 4\diamondsuit)$ are not. The ordering of the cards does not matter.)
Answer: There are $\binom{13}{2}$ ways to choose the two ranks to be the pairs. For each chosen rank, there are $\binom{4}{2}$ ways to choose the two suits for the pair with that rank. The last (fifth) card must be of different rank from the two pairs, so there are $52-8=44$ choices. Then the total number of poker hands with two pairs is:

$$\binom{13}{2}\binom{4}{2}\binom{4}{2}44$$

(c) What is the number of ways of placing k labelled balls in n labelled bins such that no two balls are in the same bin? Assume $k \leq n$. Explain your calculation.
Answer: There are n bin choices for the first ball. There are $n-1$ bin choices for the second ball, since it cannot go to the same bin as the first ball. Continuing this way, there are $n-i+1$ bin choices for the i-th ball, so the total number of ways is

$$n \cdot (n-1) \cdots (n-k+1) = \frac{n!}{(n-k)!} = \binom{n}{k} k!$$

(d) X and Y are independent random variables modulo n. You don't know the distribution of X, but you know that Y is uniformly distributed. What can you say about the distribution of $Z = X + Y \mod n$? Justify your answer.

Answer: Z is uniformly distributed modulo n. For any specific value of $X = x$, when we add a uniform random number Y, the resulting value $x + Y$ is still uniformly distributed mod n. Then it is easy to show that no matter what the distribution of X is, the distribution of $Z = X + Y$ is still uniform.

Formally, you can also calculate that for any value i mod n,

$$\begin{aligned}
\Pr[Z = i] &= \sum_{j=0}^{n-1} \Pr[Z = i, Y = j] \\
&= \sum_{j=0}^{n-1} \Pr[X = i - j, Y = j] \\
&= \sum_{j=0}^{n-1} \Pr[X = i - j] \cdot \Pr[Y = j] \\
&= \frac{1}{n} \cdot \sum_{j=0}^{n-1} \Pr[X = i - j] \\
&= \frac{1}{n}.
\end{aligned}$$

Common mistakes: Many students say that the distribution of Z is the same as (or similar to) the distribution of X because Y is uniform so it is "essentially constant", which is incorrect.

Many students also attempt to calculate the expectation and variance of X, Y, and Z. But note that to calculate the expectation of a random variable, we need to multiply the possible values of that random variable with the corresponding probabilities. In this case the possible values are integers mod n, and when we multiply them with real numbers, the result doesn't necessarily make sense.

(e) X and Y are independent random variables with normal distribution with mean m_1 and m_2 respectively, and variance σ_1^2 and σ_2^2 respectively. Describe the distribution of $Z = X + Y$ (including mean and variance).

Answer: The sum of two independent normal random variables is normal. So Z is normally distributed with mean $m_1 + m_2$ and variance $\sigma_1^2 + \sigma_2^2$.

3. **Chicken McNugget (10 points)**

McDonald's sells chicken McNuggets only in 6, 9 and 20 piece packages. This means that you cannot purchase exactly 8 pieces, but can purchase 15. The Chicken McNugget Theorem theorem states that the largest number of pieces you *cannot* purchase is 43 (i.e., you cannot purchase exactly 43 pieces, and 43 is the largest number that you cannot purchase).

Formally state the Chicken McNugget Theorem using quantifiers.

Answer: The theorem has two components: (i) you cannot purchase 43 pieces, and (ii) you can purchase any number larger than 43 pieces. The correct translation of the theorem is:

$$\left(\nexists a,b,c \in \mathbb{N},\ 6a+9b+20c=43 \right) \wedge \left(\forall n \geq 44,\ \exists a,b,c \in \mathbb{N},\ 6a+9b+20c=n \right)$$

Common mistakes:

- Using $a,b,c \in \mathbb{Z}$ (integers, can be negative) instead of $a,b,c \in \mathbb{N}$ (natural numbers, always nonnegative). This is incorrect, since you cannot purchase a negative number of pieces.
- Swapping the order of the quantifiers, e.g., $\exists a,b,c \in \mathbb{N},\ \forall n \geq 44,\ 6a+9b+20c=43$. This is incorrect, since it claims there is a choice of a,b,c that works for all values of n (whereas the value of n should determine the choice of a,b,c).
- Using incorrect quantifiers, e.g., $\forall a,b,c \in \mathbb{N},\ \forall n \geq 43,\ 6a+9b+20c=n$.
- Not using quantifiers in some parts, e.g., defining $P(n)$ as the proposition that "you can purchase n nuggets", instead of writing "$\exists a,b,c \in \mathbb{N},\ 6a+9b+20c=n$".

Name: _____ SID: _____

Instruction: Answer any three of the next four questions (questions 4, 5, 6, 7).
Clearly indicate which three questions you want us to grade. If you do all four questions, we will only grade the first three.

4. **Random proposition (15 points)**

Suppose $x_1, x_2, \ldots, x_k$ are chosen independently and uniformly at random from $\{\text{True}, \text{False}\}$.

(a) What is the probability that the proposition $Q_1 = x_2 \wedge x_3 \wedge \cdots \wedge x_k$ is true? (Note that Q_1 does not involve x_1.) Explain your answer.
Answer: $Q_1 = x_2 \wedge x_3 \wedge \cdots \wedge x_k$ is true if and only if $x_2, \ldots, x_k$ are all true. Since the x_i's are independent, this happens with probability $(1/2)^{k-1} = 1/2^{k-1}$.

(b) Let $Q = Q_1 \vee Q_2 \vee \cdots \vee Q_k$ where $Q_i = x_1 \wedge \cdots \wedge x_{i-1} \wedge x_{i+1} \wedge \cdots \wedge x_k$. (Note that Q_i does not involve x_i, but does involve all other $k-1$ variables.) Prove that $\Pr[Q \text{ is true}] \leq k/2^{k-1}$.
Answer: Q is true if and only if at least one of Q_i is true. From part (a) we know that each Q_i is true with probability $1/2^{k-1}$. By union bound,

$$\Pr[Q \text{ is true}] = \Pr[\text{at least one } Q_i \text{ is true}] \leq \sum_{i=1}^{k} \Pr[Q_i \text{ is true}] = \frac{k}{2^{k-1}}.$$

Alternative solution 1: You can argue that Q is true if and only if either: (i) all x_i's are true, or (ii) exactly one of the x_i's is false. So there are $k+1$ possible configurations of $(x_1, \ldots, x_k)$ that make Q true, out of 2^k possibilities. Then

$$\Pr[Q \text{ is true}] = \frac{k+1}{2^k} \leq \frac{2k}{2^k} = \frac{k}{2^{k-1}}.$$

Alternative solution 2: Let A_i be the indicator random variable that Q_i is true. The proposition Q is true if and only if at least one of the Q_i's is true, which means $A = A_1 + \cdots + A_k \geq 1$. By Markov's inequality,

$$\Pr[Q \text{ is true}] = \Pr[A \geq 1] \leq \frac{\mathbb{E}(A)}{1} = \sum_{i=1}^{k} \mathbb{E}(A_i) = \sum_{i=1}^{k} \frac{1}{2^{k-1}} = \frac{k}{2^{k-1}}$$

where in the calculation above we use the result of part (a) that $\mathbb{E}(A_i) = \Pr[Q_i \text{ is true}] = 1/2^{k-1}$.

5. **Neverloops (15 points)**

The function Neverloops(P) is 0 if program P does not halt on some input x, and 1 if P halts on every input x. Is there a program that computes Neverloops? Justify your answer.

(Note: the standard halting problem, which is uncomputable, asks on input P,x whether program P halts on input x.)

Answer: No. Suppose that on the contrary, there is a program that computes Neverloops. Then we can use it to solve the halting problem through the following function:

```
def TestHalt(P, x):
  def P'(y):
    return P(x)

  if NeverLoops(P') == 1:
    return "halts"
  else:
    return "does not halt"
```

The function above returns "halt" if and only if program P on input x halts. This is the halting problem. So we see that we can easily solve the halting problem if we were able to define the program Neverloops. But we know that the halting problem is unsolvable. Hence, there is a contradiction. There is no program that computes Neverloops.

Common incorrect approaches:

- **Reducing Neverloops to the halting problem instead of reducing the halting problem to Neverloops:** This was a common error. Students tried to show that to find a program for Neverloops, we needed the solution to the halting problem as a subroutine. Since the halting problem is not solvable, Neverloops must not be solvable either.

  Analysis: This approach ignores the possibility that we could find an alternate program for Neverloops which does not involve using the solution of the halting problem. (For instance, it *might* be possible to determine whether a program halts on any input by simply checking if there are any loops in the code, and if so, checking the loop termination clauses, but not enumerating through all possible input.) Without first eliminating the possibility of any such alternate programs, we can **not** conclude that Neverloops is uncomputable.

  On considering the opposite reduction, we observe that by showing how we can solve the halting problem if we had a program for Neverloops, we are claiming to do something which we know is not possible! So, we conclude that we must be mistaken: we can not find a program that computes Neverloops.

- **Making an argument similar to that given in the notes for The Halting Problem:**
  Sample solution: *Suppose there is a program that computes* `Neverloops`*. Then we define a program Turing as follows:*

  ```
  def Turing(P):
    if Neverloops(P) == 0, then loop forever
    else halt
  ```

  Now, we consider `Turing(Turing)`*. If it halts, it should have looped forever, and if it loops forever, it should have halted. Hence, contradiction.*

Analysis: To see why this argument works for the halting problem but not for Neverloops, we need to think about what's different between Neverloops and the halting problem. In the halting problem, we consider P for a specific input x, whereas in Neverloops, we will consider it for all possible inputs. Keeping this in mind, let's think about the different possibilities for `Turing(Turing)`. Suppose `Neverloops(Turing)` returns 0, then Turing loops forever. This means `Neverloops(Turing)` can not return 0. Therefore, it must return 1. But then, `Turing(Turing)` must halt. This is not a contradiction since inputting any halting program to `Turing` will cause it to loop forever (for instance, `Turing(lambda x: x)`). `Neverloops(Turing)` has correctly returned 1.

6. **Coin induction (15 points)**

We have n coins $C_1, \ldots, C_n$. The coins are weighted such that coin C_i comes up Heads with probability $\frac{1}{2i+1}$. Prove by induction that if the n coins are tossed independently, the probability of getting an odd number of Heads is $\frac{n}{2n+1}$.

Answer:

Base case: When $n = 1$ we only have 1 coin C_1, which comes up Heads with probability $\frac{1}{2\cdot 1+1} = \frac{1}{3}$, which agrees with the formula $\frac{n}{2n+1} = \frac{1}{3}$.

Inductive hypothesis: Assume the statement is true for n coins, namely,

$$\Pr[\text{odd \# of Heads in } n \text{ tosses}] = \frac{n}{2n+1}.$$

Inductive step: Now suppose we have $n+1$ coins. There are two possible ways of getting an odd number of Heads in $n+1$ tosses:

- get an even number of Heads in the first n tosses and the $(n+1)$-st toss comes up Heads
- get an odd number of Heads in the first n tosses and the $(n+1)$-st toss comes up Tails

Since the coin tosses are independent, the desired probability is

$$\begin{aligned}
\Pr[\text{odd \# of Heads in } n+1 \text{ tosses}] &= \Pr[\text{even \# of Heads in } n \text{ tosses}] \cdot \Pr[C_{n+1} \text{comes up Heads}] \\
&\quad + \Pr[\text{odd \# of Heads in } n \text{ tosses}] \cdot \Pr[C_{n+1} \text{comes up Tails}] \\
&= \left(1 - \frac{n}{2n+1}\right) \cdot \frac{1}{2(n+1)+1} + \frac{n}{2n+1} \cdot \left(1 - \frac{1}{2(n+1)+1}\right) \\
&= \frac{n+1}{(2n+1)(2n+3)} + \frac{n(2n+2)}{(2n+1)(2n+3)} \\
&= \frac{(n+1)(2n+1)}{(2n+1)(2n+3)} \\
&= \frac{n+1}{2n+3},
\end{aligned}$$

as claimed. In the second equality above we have used the inductive hypothesis.

7. **Drawers of socks (15 points)**

A chest of drawers has two drawers. 10 different pairs of socks are randomly placed in the two drawers (each of the 20 socks is equally likely to be placed in either drawer).

(a) Let N be the number of complete pairs of socks in the first drawer. Find the distribution of N. Specify the parameter(s).

Answer: Each sock is placed in the first drawer with probability $\frac{1}{2}$, and the socks are independent, so each pair of socks can be considered as a coin flip that comes up Heads (the pair is in the first drawer) with probability $\frac{1}{4}$. So the distribution of N is binomial with parameter $n = 10$ and success probability $p = \frac{1}{4}$.

(b) What is the probability that at least one drawer has no complete pairs of socks? Explain your calculation.

Answer: Let A be the event that the first drawer has no complete pairs of socks, and B be the event that the second drawer has no complete pairs. The probability that we want is

$$\Pr[A \cup B] = \Pr[A] + \Pr[B] - \Pr[A \cap B].$$

From part (a), we have $\Pr[A] = \Pr[N = 0] = (\frac{3}{4})^{10}$. Similarly, $\Pr[B] = (\frac{3}{4})^{10}$.

$A \cap B$ is the event that both drawers have no complete pairs. Each of the ten pairs of socks has probability $\frac{1}{2}$ of being split between the two drawers, so therefore $A \cap B$ happens with probability $\Pr[A \cap B] = (\frac{1}{2})^{10}$.

Therefore, the probability that either the first or the second drawer has no complete pairs is

$$\Pr[A \cup B] = \left(\frac{3}{4}\right)^{10} + \left(\frac{3}{4}\right)^{10} - \left(\frac{1}{2}\right)^{10} = 2\left(\frac{3}{4}\right)^{10} - \left(\frac{1}{2}\right)^{10}.$$

Grading notes: Many people tried to compute the probability that no drawer has no complete pairs of socks. This is the probability that both drawers have at least one complete pair of socks. This is unfortunately much more complicated than the original problem, as some pairs can still be split between the two drawers. The inclusion-exclusion allows us to avoid this difficulty.

When computing $\Pr[A \cap B]$, note that A and B are not independent, so $\Pr[A \cap B] \neq \Pr[A]\Pr[B]$.

Instruction: Answer any three of the next four questions (questions 8, 9, 10, 11).
Clearly indicate which three questions you want us to grade. If you do all four, we will only grade the first three.

8. **Base disease (15 points)**

Suppose that 1 percent of the population has a certain disease. There is a test for the disease, but it's not always correct.

- For a randomly chosen person who has the disease, the test comes back positive with probability 0.9 and negative with probability 0.1.
- For a randomly chosen person who doesn't have the disease, the test comes back positive with probability 0.01 and negative with probability 0.99.

(a) The test on a random person comes back positive. What is the probability that the person has the disease?

Answer: Let D be the event that the person has the disease, and T be the event that the test comes back positive. The problem tells us that

$$\Pr[D] = 0.01, \qquad \Pr[T \mid D] = 0.9 \qquad \text{and} \qquad \Pr[T \mid D^c] = 0.01.$$

By the total probability rule,

$$\Pr[T] = \Pr[D] \cdot \Pr[T \mid D] + \Pr[D^c] \cdot \Pr[T \mid D^c] = 0.01 \times 0.9 + 0.99 \times 0.01 = 0.0189.$$

Then by Bayes' rule,

$$\Pr[D \mid T] = \frac{\Pr[T \mid D] \cdot \Pr[D]}{\Pr[T]} = \frac{0.9 \times 0.01}{0.0189} = \frac{10}{21}.$$

(b) Suppose that each test is itself probabilistic — if you perform the test twice on the same person with the disease, each time it comes up positive *independently* with probability 0.9. Similarly, if you perform the test twice on the same person who doesn't have the disease, each time it comes up positive independently with probability 0.01.

You choose a person at random and run the test twice on that person. Suppose the first test comes back positive. What is the probability that the second one comes back positive too?

Hint: Use your answer from part (a).

Answer: From part (a), we know that given the first test comes back positive, the person has the disease with probability $10/21$. Therefore, the probability that the second test comes back positive is

$$\frac{10}{21} \times 0.9 + \frac{11}{21} \times 0.01 = \frac{9.11}{21}.$$

9. Secret sharing (15 points)

Recall that in a secret sharing scheme the secret $p(0) \bmod q$ can be reconstructed from the values of the polynomial $p(x)$ of degree d at any $d+1$ points. However, the values of the polynomial $p(x)$ at any d points reveal absolutely no information about the secret $p(0)$. As we saw in lecture, this condition can be formally stated using conditional probability as follows: $\Pr[p(0) = a \mid p(1), p(2), \ldots, p(d)] = 1/q$ for every $a \bmod q$.

Now suppose Alice wishes to share a secret that consists of two numbers a and b, each $\bmod q$. She picks a random degree d polynomial $p(x) \bmod q$ such that $p(0) = a$ and $p(1) = b$. She distributes shares $p(2), \ldots, p(k)$ as with standard secret sharing (where $k \geq d+2$), and claims that any $d+1$ people can reconstruct the secret, but any d people have absolutely no information about the secret.

(a) Formally state (using conditional probability) Alice's claim that the values $p(2), p(3), \ldots, p(d+1)$ reveal absolutely no information about the secret a, b.

Answer:

$$\Pr[p(0) = a, p(1) = b \mid p(2), p(3), \ldots, p(d+1)] = \frac{1}{q^2}.$$

Common Mistakes:

- Set the above equation to $\frac{1}{q}$. Note that there are q possible values for a and b independently, for a total of q^2 values for the secret.
- Wrote separate equations for $\Pr[p(0) = a]$ and $\Pr[p(1) = b]$, setting them both to $\frac{1}{q}$. While this is a correct equation, it doesn't claim anything about the independence of a and b, which is required for all possible values of the secret to be equally likely. (This statement is in fact true.)

(b) Is Alice's claim correct? If so prove it, and if not give a precise reason why not.

Answer: No, because

$$\Pr[p(0) = a, p(1) = b \mid p(2), p(3), \ldots, p(d+1)] = \frac{1}{q}$$

if the $d+2$ points $p(0) = a, p(1) = b, p(2), \ldots, p(d+1)$ lie on a degree d polynomial, and equal to 0 otherwise. This is because $d+2$ points do not necessarily lie on a degree d polynomial, so some (a,b) pairs are invalid. Alternatively, you could say that knowing one of the values a, b uniquely determines the other, or that there are only q possible polynomials with the points given, but there should be q^2 possible values for (a,b).

Common Mistakes:

- Simply quoted the proof from secret sharing that you need $d+1$ points to uniquely determine a degree d polynomial. While it is true that you can't determine the secret exactly with the d points, you can narrow down the options, so some information is revealed.
- Stated that a and b uniquely determine each other, but that this isn't a problem since Alice doesn't reveal either of them. You can look at this from an attacker's perspective - if the attacker guesses a correctly (with probability $\frac{1}{q}$) then they have also found b.
- Stated that the probability of guessing a and the probability of guessing b are both $\frac{1}{q}$, so no information is revealed. Again, this is incorrect because while this statement is true, it doesn't say anything about their independence.

- Several students had the right idea, but attempted to prove by giving an example. There is no such thing as proof by example.

Rubric: Partial credit was given if it was indicated that a and b are not independent, or that the probability of guessing them **both** right was $\frac{1}{q}$. Note that this does not include saying that there are q possible polynomials with the d points given, as this just quotes a conclusion about secret sharing. No points were given for statements/proofs about why secret sharing works.

10. **Umbrella store (15 points)**

Bob has a store that sells umbrellas. The number of umbrellas that Bob sells on a rainy day is a random variable Y with mean 25 and standard deviation $\sqrt{105}$. But if it is a clear day, Bob doesn't sell any umbrellas at all. The weather forecast for tomorrow says it will rain with probability $\frac{1}{5}$. Let Z be the number of umbrellas that Bob sells tomorrow.

(a) Let X be an indicator random variable that it will rain tomorrow. Write Z in terms of X and Y.
Answer: $Z = XY$.

(b) What is the mean and standard deviation of Z?
Answer: We have

$$\mathbb{E}(Z) = \Pr[X=1]\cdot\mathbb{E}(Y) + \Pr[X=0]\cdot 0 = \frac{1}{5}\mathbb{E}(Y) = \frac{1}{5}\cdot 25 = 5$$

and

$$\mathbb{E}(Z^2) = \Pr[X=1]\cdot\mathbb{E}(Y^2) + \Pr[X=0]\cdot 0 = \frac{1}{5}\mathbb{E}(Y^2) = \frac{1}{5}\cdot 730 = 146$$

since $\mathbb{E}(Y^2) = \mathrm{Var}(Y) + \mathbb{E}(Y)^2 = 105 + 625 = 730$. So

$$\mathrm{Var}(Z) = \mathbb{E}(Z^2) - \mathbb{E}(Z)^2 = 146 - 25 = 121.$$

Therefore, the mean of Z is 5 and the standard deviation is $\sqrt{121} = 11$.

(c) Use Chebyshev's inequality to bound the probability that Bob sells at least 25 umbrellas tomorrow.
Answer: Since $\mathbb{E}(Z) = 5$ and $\mathrm{Var}(Z) = 121$,

$$\Pr[Z \geq 25] \;=\; \Pr[Z - \mathbb{E}(Z) \geq 20] \;\leq\; \Pr[|Z - \mathbb{E}(Z)| \geq 20] \;\leq\; \frac{\mathrm{Var}(Z)}{400} \;=\; \frac{121}{400}.$$

11. To infinity and beyond (15 points)

You are the captain of the Bimillennial Eagle, a spaceship that has just returned from hyperspace to ordinary space, only to encounter the debris of a recently destroyed planet. Your maneuvering jets are temporarily out of order. The expected number of pieces of debris in any km^3 of space is $1/10^6$. You reckon that your spaceship has a cross section of area $1/1000$ km^2, and you must travel 10^5 km before you are clear of the debris.

Model the debris field by a Poisson distribution and calculate your chances of getting all the way through the debris field without a collision.

Answer:

The volume of the space that the spaceship travels through is simply the product of its cross section area and the distance it travels, i.e. $\frac{1}{1000}$ $\text{km}^2 \times 10^5$ km $= 100$ km^3.

Now let X be the number of pieces of debris in this space. Given that the expected number of pieces of debris per km^3 is $\frac{1}{10^6}$ we have

$$\mathbb{E}(X) = 100 \text{ km}^3 \times \frac{1}{10^6 \text{ km}^3} = 10^{-4}.$$

The random variable X is a Poisson random variable and is therefore completely characterized by its mean $\lambda = 10^{-4}$. For any $i \geq 0$ we have

$$\Pr[X = i] = \frac{\lambda^i e^{-\lambda}}{i!}.$$

We are interested in the event of having no collisions, i.e. $X = 0$. We have

$$\Pr[X = 0] = \frac{\lambda^0 e^{-\lambda}}{0!} = e^{-\lambda} = e^{-10^{-4}}.$$