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.
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

   (b) The proposition \((A \land B) \lor (\neg A \land B) \lor \neg B\) can never be false.
       Circle one: True    False

   (c) The hypercube graph always has an Eulerian tour.
       Circle one: True    False

   (d) If \(f: A \rightarrow B\) is an injective (1-1) function, then there exists a surjective (onto) function \(g: B \rightarrow A\).
       Circle one: True    False

   (e) If \(\gcd(a, b) = d\), then \(a\) has no factor larger than \(d\).
       Circle one: True    False

   (f) In RSA with modulus \(n = 91\) and encryption power \(e = 5\), the decryption power is \(d = 73\) because \(de = 365 \equiv 1 \mod 91\).
       Circle one: True    False

   (g) If the multiplicative inverse \(a^{-1} \mod p\) exists for all \(a \in \{1, \ldots, p-1\}\), then \(p\) is a prime.
       Circle one: True    False

   (h) For any \(d \in \mathbb{N}\), the set of polynomials of degree \(d\) with integer coefficients is countable.
       Circle one: True    False
(i) If a sample space \( \Omega \) has \( n \) sample points, then there are \( 2^n \) possible events.
Circle one: True False

(j) If \( P(A \mid B) = 1 \), then \( P(B) \leq P(A) \).
Circle one: True False

(k) For any random variable \( X \), \( \mathbb{E}(X^2) \geq \mathbb{E}(X)^2 \).
Circle one: True False

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

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

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

(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
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.

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

iv. Times of finishers in the NY Marathon.

v. Number of girls in a family with 6 kids.

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

(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 \((2\ Diamond, 2\ Club, 4\ Diamond, 4\ Spade, 8\ Diamond)\) is an example of a poker hand with 2 pairs, but \((2\ Diamond, 2\ Club, 2\ Club, 2\ Spade, 8\ Diamond)\) and \((2\ Diamond, 2\ Club, 4\ Diamond, 4\ Spade, 4\ Diamond)\) are not. The ordering of the cards does not matter.)

(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.
(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 $i \mod n$,

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

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 $\sigma_1^2$ and $\sigma_2^2$ respectively. Describe the distribution of $Z = X + Y$ (including mean and variance).
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 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.
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 \{True, False\}.

(a) What is the probability that the proposition $Q_1 = x_2 \land x_3 \land \cdots \land x_k$ is true? (Note that $Q_1$ does not involve $x_1$.) Explain your answer.

(b) Let $Q = Q_1 \lor Q_2 \lor \cdots \lor Q_k$ where $Q_i = x_1 \land \cdots \land x_{i-1} \land x_{i+1} \land \cdots \land x_k$. (Note that $Q_i$ does not involve $x_i$, but does involve all other $k-1$ variables.) Prove that $\Pr[Q \text{ is true}] \leq k/2^{k-1}$.
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.)
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 $\text{Turing}(\text{Turing})$. Suppose $\text{Neverloops}(\text{Turing})$ returns 0, then Turing loops forever. This means $\text{Neverloops}(\text{Turing})$ cannot return 0. Therefore, it must return 1. But then, $\text{Turing}(\text{Turing})$ must halt. This is not a contradiction since inputting any halting program to $\text{Turing}$ will cause it to loop forever (for instance, $\text{Turing}(\lambda x: x)$). $\text{Neverloops}(\text{Turing})$ 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 \( \frac{n}{2n+1} \).
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).

(b) What is the probability that at least one drawer has no complete pairs of socks? Explain your calculation.
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?
(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).
9. Secret sharing (15 points)

Recall that in a secret sharing scheme the secret $p(0) \mod q$ 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 $a \mod q$.

Now suppose Alice wishes to share a secret that consists of two numbers $a$ and $b$, each mod $q$. She picks a random degree $d$ polynomial $p(x) \mod q$ 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 $a, b$.

(b) Is Alice’s claim correct? If so prove it, and if not give a precise reason why not.
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 $\frac{1}{3}$. 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 $Y$.

(b) What is the mean and standard deviation of $Z$?

(c) Use Chebyshev’s inequality to bound the probability that Bob sells at least 25 umbrellas tomorrow.
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.

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. 
\[
\text{Volume} = \frac{1}{1000} \times 10^5 = 100 \text{ km}^3.
\]

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 $1/10^6$ we have 
\[
E(X) = 100 \times \frac{1}{10^6} = 10^{-4}.
\]

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