CS 70 Discrete Mathematics and Probability Theory
Fall 2013 Vazirani Induction Practice

1. **Variations**

   Suppose we were trying to prove $P(n)$ is true for all $n \in N$ by induction on n. Instead we succeeded in proving $\forall k \in N$ if $P(k)$ is true then $P(k+2)$ is true. For each of the following assertions below, state whether (A) it must always hold, or (N) it can never hold, or (C) it can hold but need not always. Give a very brief (one or two sentence) justification for your answers. The domain of all quantifiers is the natural numbers.

   (a) $\forall n \geq 0\ P(n)$.

   (b) If $P(0)$ is true then $\forall n\ P(n+2)$ is true.

   (c) If $P(0)$ is true then $\forall n\ P(2n)$ is true.

   (d) $\forall n\ P(n)$ is false.

   (e) If $P(0)$ and $P(1)$ are true then $\forall n\ P(n)$ is true.

   (f) We can conclude that ($\forall n \leq 10$ $P(n)$ is true), and ($\forall n > 10\ P(n)$ is false).

2. **Chocolate!**

   Chocolate often comes in rectangular bars marked off into smaller squares. It is easy to break a larger rectangle into two smaller rectangles along any of the horizontal or vertical lines between the squares. Suppose I have a bar containing k squares and wish to break it down into its individual squares. Prove that *no matter which way I break it*, it will take exactly $k-1$ breaks to do this.

3. **Recursion**

   Let the function g be defined recursively on the natural numbers as follows:
   $g(0) = 0, g(1) = 1$, and $g(n) = 5g(n-1) - 6g(n-2)$, for all $n \geq 2$. Show that $\forall n \in N, g(n) = 3^n - 2^n$.

4. **Some Identities by Induction.**

   (a) For $n \in N$ with $n \geq 2$, define s_n by

   $$s_n = (1 - \frac{1}{2}) \times (1 - \frac{1}{3}) \times \cdots \times (1 - \frac{1}{n}).$$

   Prove that $s_n = 1/n$ for every natural number $n \geq 2$.

   (b) Let $a_n = 3^{n+2} + 4^{2n+1}$. Prove that 13 divides a_n for every $n \in N$. (Hint: What can you say about $a_{n+1} - 3a_n$?)

   (c) Prove that $2^n < n!$ for all integers $n \geq 4$.

   (d) Prove that $1 \cdot 1! + 2 \cdot 2! + \cdots + n \cdot n! = (n+1)! - 1$ for all integers $n \in \mathbb{N}$.

5. **A pizza proof.**

   Working at the local pizza parlor, I have a stack of unbaked pizza doughs. For a most pleasing presentation, I wish to arrange them in order of size, with the largest pizza on the bottom. I know how to place my spatula under one of the pizzas and flip over the whole stack above the spatula (reversing their order). This is the only move I know that can change the order of the stack; however, I am willing to keep repeating this move until I get the stack in order. Is it always possible to get the pizzas in order? Prove your answer.

6. **Grading proofs**

   Assign a grade of A (correct) or F (failure) to each of the following proofs. If you give a F, please explain exactly everything that is wrong with the structure or the reasoning in the *proof*. You should justify all your answers (remember, saying that the claim is false is *not* a justification).

   (a) **Claim**: For every $n \in \mathbf{N}$ with $n \geq 1$, $n^2 + n$ is odd.

   Proof:

   The proof will be by induction.
   Base case: The natural number 1 is odd.
   Inductive step: Suppose $k \in \mathbf{N}$ and $k^2 + k$ is odd. Then,

   $$(k+1)^2 + (k+1) = k^2 + 2k + 1 + k + 1 = (k^2 + k) + (2k + 2)$$

   is the sum of an odd and an even integer. Therefore, $(k+1)^2 + (k+1)$ is odd. By the Principle of Mathematical Induction, the property that $n^2 + n$ is odd is true for all natural numbers n.

   (b) **Claim**: For all $x, y, n \in \mathbf{N}$, if $\max(x, y) = n$, then $x = y$.

   Proof:

   The proof will be by induction.
   Base case: Suppose that $n = 0$. If $\max(x, y) = 0$ and $x, y \in N$, then $x = 0$ and $y = 0$, hence $x = y$.
   Induction step: Assume that, whenever we have $\max(x, y) = k$, then $x = y$ must follow. Next suppose x, y are such that $\max(x, y) = k + 1$. Then it follows that $\max(x - 1, y - 1) = k$, so by the inductive hypothesis, $x - 1 = y - 1$. In this case, we have $x = y$, completing the induction step.

   (c) **Claim**: $\forall n \in \mathbf{N}.\ n^2 \leq n$.

   Proof:

   The proof will be by induction.
   Base case: When $n = 0$, the statement is $0^2 \leq 0$ which is true.
   Induction step: Now suppose that $k \in \mathbf{N}$, and $k^2 \leq k$. We need to show that

   $$(k+1)^2 \leq k + 1$$

   Working backwards we see that:

   $$\begin{aligned} (k+1)^2 &\leq k+1 \\ k^2 + 2k + 1 &\leq k+1 \\ k^2 + 2k &\leq k \\ k^2 &\leq k \end{aligned}$$

   So we get back to our original hypothesis which is assumed to be true. Hence, for every $n \in N$ we know that $n^2 \leq n$.

CS 70 Discrete Mathematics and Probability Theory
Fall 2013 Vazirani Stable Marriage Practice

1. **Stable Marriage True-or-False?**

   For each of the following claims, state whether the claim is true or false. If it is true, give a *short* proof; if it is false, give a *simple* counterexample.

   (a) In a stable marriage instance, if man M and woman W each put each other at the top of their respective preference lists, then M must be paired with W in every stable pairing.

   (b) In a stable marriage instance with at least two men and two women, if man M and woman W each put each other at the bottom of their respective preference lists, then M cannot be paired with W in any stable pairing.

   (c) For every $n > 1$, there is a stable marriage instance with n men and n women which has an unstable pairing in which every unmatched man-woman pair is a rogue couple.

   (d) An instance of the stable marriage problem (i.e., a set of men and women and their associated preference lists) must have at least two different stable pairings, a male-optimal one and a female-optimal one.

   (e) In any instance of the stable marriage problem, if woman w is matched to her first choice by the traditional marriage algorithm, then she must be matched to her first choice in all stable pairings.

   (f) If man m is rejected by woman w at some point in the traditional marriage algorithm, then no stable pairing exists in which m is matched to w.

2. **Stable marriage**

   Consider a set of four boys (a, b, c, d) and four girls (1, 2, 3, 4) with the preferences shown below.

boy	preferences	girl	preferences
a	1>2>3>4	1	d>b>c>a
b	2>1>4>3	2	a>d>b>c
c	1>3>2>4	3	a>b>c>d
d	2>1>3>4	4	d>c>a>b

   (a) Run the traditional marriage algorithm on this instance. Show each stage of the algorithm, and give the resulting matching, expressed as a set of boy-girl pairs. You can do this by hand, or you can try to program it!

   (b) The matching you found above is boy-optimal. Find now a girl-optimal stable matching. Compare the two matchings.

3. **Cold feet and Eager Beavers.**

   Consider a slightly different setting for the stable marriage algorithm: suppose that some of the men are nervous about proposing, and it takes them extra time to work up the courage to ask the women to

marry them. Other men are overly eager, and do not have the courtesy to wait 24 hours between when they were rejected and when they next propose. The result of this is that some men might procrastinate for several days, while others might propose and get rejected several times in a single day. Can the order of the proposals change the resulting pairing? Give an example of such a change or prove that the pairing that results is the same.

4. **Man-Optimal, Woman-Optimal**

   In a particular instance of the stable marriage problem with n men and n women, it turns out that there are exactly three distinct stable pairings, $\mathscr{P}_1, \mathscr{P}_2, \mathscr{P}_3$. Also, each woman W has a different partner in the three pairings. Therefore, each woman has a clear preference ordering of the three pairings (according to the ranking of her partners in her preference list). Now, suppose that for woman W_1 this order is $\mathscr{P}_1 > \mathscr{P}_2 > \mathscr{P}_3$. True or false: every woman has the same preference ordering $\mathscr{P}_1 > \mathscr{P}_2 > \mathscr{P}_3$. Justify your answer carefully, using facts about Stable Marriage that we proved in class.

5. **One side wins, the other loses**

   Mr. and Mrs. Matchmaker are at work in their matchmaking agency. There are n men and n women, each having strict preferences over people of the opposite gender as suitors for marriage. The Matchmakers want to produce a stable pairing.

   Minions who work for the matchmakers have produced two proposals each detailing one set of pairings. Mr. Matchmaker proposes the following scheme to produce the final pairings which are announced to the clients: for each man look at all of the women who are matched with him in at least one of the proposals (there could be one or two women). Then match this man with the best (according to his preferences) of these women.

   (a) Prove that Mr. Matchmaker's scheme actually results in a matching in which no two men are matched with the same woman.

   (b) Prove that the matching produced is stable.

   (c) Mrs. Matchmaker is very suspicious of this scheme and she thinks that it will do a terrible job for women. Help her by proving that Mr. Matchmaker's scheme results in a matching in which each woman is matched to her least favorite suitor among the two proposals produced by the minions.

   (d) In class you learned that the propose and reject algorithm results in a pairing which is optimal for men and pessimal for women. Give another proof using the things you learned in previous parts that such a pairing which is optimal for men and pessimal for women always exists.

6. I'm too good to marry

   In the stable marriage problem, suppose that some men and women have standards and would not just settle for anyone. In other words, in addition to the orderings they have, they prefer being alone to being with some of the lower-ranked individuals (in their own preference list). A pairing would ultimately have to be partial, i.e. some individuals would remain single.

   The notion of stability here should be adjusted a little bit: a pairing is stable if there is no rogue couple, and there is no paired individual who prefers being single over being with his/her pair.

   (a) Prove that the stable marriage algorithm still yields a stable pairing. You can approach this by introducing imaginary mates (one for each person) that people marry if they are single. How should you adjust the preference lists of people, including those of the newly introduced imaginary ones for this to work?

(b) As you saw in the lecture, we may have different stable pairings. But interestingly, if a person remains single in one stable pairing, s/he must remain single in any other stable pairing as well (there really is no hope for some people!). Prove this fact using Mr. Matchmaker's scheme and its properties from the previous question (you don't need to have proved them to use the results here).

CS 70 Discrete Mathematics and Probability Theory
Fall 2013 Vazirani Modular Arithmetic Practice

1. **Modular arithmetic**

   Solve the following equations for x and y modulo the indicated modulus, or show that no solution exists. Show your work.

   (a) $7x \equiv 1 \pmod{15}$.

   (b) $10x + 20 \equiv 11 \pmod{23}$.

   (c) $5x + 15 \equiv 4 \pmod{20}$.

   (d) The system of simultaneous equations $3x + 2y \equiv 0 \pmod{7}$ and $2x + y \equiv 4 \pmod{7}$.

2. **Modular inverse**

   Prove that the equation $ax \equiv ay \bmod n$ implies $x \equiv y \bmod n$ whenever $\gcd(a,n) = 1$. Show that the condition $\gcd(a,n) = 1$ is necessary by supplying a counterexample with $\gcd(a,n) > 1$.

3. **Fibonacci numbers and Euclid**

   Recall that the Fibonacci numbers $F(0), F(1) \ldots$ are given by $F(0) = F(1) = 1$ and the recurrence

   $$F(n+1) = F(n) + F(n-1), \qquad n \geq 2.$$

   (a) Show that for any $n \geq 0$, $\gcd(F(n+1), F(n)) = 1$.

   (b) Show that $aF(n+1) + bF(n) = 1$, where $a = F(n-1)$ and $b = -F(n)$ if n is odd, and $a = -F(n-1)$ and $b = F(n)$ if n is even.

4. **Modular counting**

   What is the size of the the set $\{0a, 1a, 2a, 3a, \ldots, (x-1)a\}$ modulo x, if $gcd(x,a) = 4$ and $a \neq 0$? (Consider ia and ja to be the same if $ia = ja \pmod{x}$.)

5. **Modular arithmetic proof**

   Give a proof to the following theorem. You will likely find the use of modular arithmetic useful.

   Theorem. If $a_1, \ldots, a_n$ is a sequence of n integers (not necessarily distinct), prove that there is some nonempty subsequence whose sum is a multiple of n.

6. **Euclid**

   Let p, q, and r be distinct primes. Prove that there exist integers a, b, and c such that:
   $a \cdot (pq) + b \cdot (qr) + c \cdot (rp) = 1$.

7. **Modular inverse**

   Prove that the equation $ax \equiv ay \bmod n$ implies $x \equiv y \bmod n$ whenever $\gcd(a,n) = 1$. Show that the condition $\gcd(a,n) = 1$ is necessary by supplying a counterexample with $\gcd(a,n) > 1$.

8. **Binary gcd**

(a) Prove that the following statements are true for all $m, n \in \mathbb{N}$.

$$\begin{aligned} \text{If } m \text{ is even and } n \text{ is even,} &\quad \gcd(m,n) = 2\gcd(m/2, n/2). \\ \text{If } m \text{ is even and } n \text{ is odd,} &\quad \gcd(m,n) = \gcd(m/2, n). \\ \text{If } m, n \text{ are both odd and } m \geq n, &\quad \gcd(m,n) = \gcd((m-n)/2, n). \end{aligned}$$

(b) Fill in the missing part of the following template to get an alternative algorithm for computing the gcd.

$\gcd(m,n)$:
1. If $m = 0$, return n. If $n = 0$, return m.
2. If m is even and n is even, return $2 \cdot \gcd(m/2, n/2)$.
3. If m is even and n is odd, return $\gcd(m/2, n)$.
4. If m is odd and n is even, return $\gcd(m, n/2)$.
5. ??????????.

Prove that the resulting algorithm correctly computes the gcd.

CS 70 Discrete Mathematics and Probability Theory
Fall 2013 Vazirani RSA Practice

1. **Euler and Little Fermat**

   (a) What is $7^{3,000,000,000}$ mod 41? Justify your answer.

   (b) What is $2^{3^{2,001}}$ mod 47? Show your work. (Note that a^{b^c} means $a^{(b^c)}$, not $(a^b)^c$.)

   [Hint: Use the theorems alluded to in the title of this problem.]

2. **Fermat's Little Theorem**

   Fermat's Little Theorem states that, if p is prime, then for any $a \in \{1,2,\ldots,p-1\}$, $a^{p-1} = 1 \bmod p$.

   (a) Prove Fermat's Little Theorem. [HINT: Show that the set of $p-1$ numbers $\{a\cdot 1, a\cdot 2, \ldots, a\cdot(p-1)\}$ are all distinct and non-zero mod p. Then multiply them together.]

   (b) Suppose you wish to use a triple prime analog of RSA. Let $N = pqr$, where p,q,r are primes. Suppose that $gcd(e,(p-1)(q-1)(r-1)) = 1$. Show that $E(x) = x^e \bmod N$ is a bijection. What is the decryption key?

3. **RSA**

   Show how to determine p and q given N and $\varphi(N) = (p-1)(q-1)$. (In other words, given N and the value $\varphi(N) = (p-1)(q-1)$, it is possible to factor N efficiently. This shows that determining $\varphi(N)$ is "as hard as factoring.")

4. **RSA**

   In this problem you play the role of Amazon, who wants to use RSA to be able to receive messages securely.

   (a) Amazon first generates two large primes p and q. He picks $p = 13$ and $q = 19$ (in reality these should be 512-bit numbers). He then computes $N = pq$. Amazon chooses e from $e = 37, 38, 39$. Only one of those values is legitimate, which one? (N,e) is then the public key.

   (b) Amazon generates his private key d. He keeps d as a secret. Find d. Explain your calculation.

   (c) Bob wants to send Amazon the message $x = 102$. How does he encrypt his message using the public key, and what is the result?

   (d) Amazon receives an encrypted message $y = 141$ from Charlie. What is the unencrypted message that Charlie sent him?

5. **Squaring**

   In practical implementations of RSA, it is common to use $e = 3$ as the public exponent, because this provides performance enhancements.

   Could we use $e = 2$ for RSA encryption? Why or why not?

6. **Easy RSA**

   In class, we said that RSA uses as its modulus a product of two primes. Let's look at a variation that uses a single prime number as the modulus. In other words, Bob would pick a 1024-bit prime p and a public exponent e satisfying $2 \leq e < p-1$ and $\gcd(e, p-1) = 1$, calculate his private exponent d as the inverse of e modulo $p-1$, publish (e, p) as his public key, and keep d secret. Then Alice could encrypt via the equation $E(x) = \mod(x^e, p)$ and Bob could decrypt via $D(y) = \mod(y^d, p)$.

   Explain why this variation is insecure. In particular, describe a procedure that Eve could use to recover the message x from the encrypted value y that she observes and the parameters (e, p) that are known to her. Make sure to justify why Eve could feasibly carry out this procedure without requiring extravagant computation resources.

7. **OpRSA**

   Anonymous is running a website and needs to be able to securely receive information from people with something interesting to say about a certain potential target. Your job is to help them set up a crypto-system to accomplish this task.

   (a) Anonymous first generates two large primes, p and q. They pick $p = 13$ and $q = 19$ (though in reality these should be 512-bit numbers). They then computes $N = pq$. Anonymous chooses e from $\{37, 38, 39\}$. Only one of these values is legitimate; which one? (N, e) is then the public key.

   (b) Anonymous needs to generate a private key, d, which will be kept as a secret. Help Anonymous find d and explain your calculation.

   (c) Brain wants to send Anonymous the message $x = 102$. How does he encrypt his message using the public key and what is the result?

   (d) Anonymous receives the encrypted message $y = 141$ from Candy, another informant. What is the unencrypted message that Candy sent?

   (e) Try encrypting several other messages (you're free to choose, as long as they're valid). Do you see anything wrong with this crypto-system? (In the real world, Anonymous is a lot smarter than this.)

8. **Because the Moth just doesn't cut it**

   Gandalf the Grey (a good wizard) wanders about on his merry adventures but frequently runs into some troubles with goblins and orcs along the way. Always being the well-prepared wizard that he is, Gandalf has enlisted the service of the Great Eagles to fly him out of sticky situations at a moment's notice. To do this, he broadcasts a short message detailing his dilemma and a nearby eagle will come to his aid.

   While this is all well and good, Saruman the White (an evil wizard) wants in on this eagle concierge service. The eagles can no longer trust just any distress call they receive! Gandalf needs you (a cryptography master) to help him devise a simple scheme that will allow the eagles to verify his identity whenever he broadcasts a message out. Not only that, but the eagles need to know when the message they receive from Gandalf has been tampered with.

   Once you have devised this scheme, Gandalf will tell it to the eagle lord Gwaihir, who will relay it out to the rest of the world (they are loudmouths so they can't keep a secret).

   To summarize:

   (a) Gandalf broadcasts a message m to all of Middle-Earth.

(b) He is able to attach to the message an extra piece of information s that verifies his identity (i.e. cannot be forged) to whomever receives it.

(c) If the message has been modified in transit, recipients of the modified message should be able to detect that it is not original.

(d) Everyone in Middle-Earth knows the scheme (i.e. the algorithm itself is not a secret)

Your job in this problem is to devise an algorithm (like RSA) that meets the above criteria. In your answer, you should formally prove that Gandalf's messages can be successfully verified. You do not need to formally prove (though it should still be the case) that it is difficult to forge/tamper with messages, but you should provide some informal justification.

CS 70 Discrete Mathematics and Probability Theory
Fall 2013 Vazirani Polynomials Practice

1. **Polynomial interpolation**

   (a) Consider the set of four points $\{(0,1),(1,2),(2,4),(4,2)\}$.

   (i) Construct the unique degree-3 polynomial (over the reals) that passes through these four points by writing down and solving a system of linear equations.

   (ii) Repeat part (i) but using the method of Lagrange interpolation. Show your working.

   (b) Find a polynomial $h(x) = ax^2 + bx + c$ of degree at most 2 (over GF(7)) such that $h(0) = 3 \mod 7$, $h(1) = 6 \mod 7$, and $h(2) = 6 \mod 7$.

2. **Roots**

   Let p be prime. Argue that every value z has at most 2 square roots modulo p.

3. **Random Polynomials**

   Let p be a prime. Two polynomials f and g over $GF(p)$ are chosen independently and uniformly at random from all polynomials of degree $d \geq 0$. What is the expected number of intersection points of f and g?

4. **More polynomials!**

   Define the sequence of polynomials by $P_0(x) = x + 12$, $P_1(x) = x^2 - 5x + 5$ and $P_n(x) = xP_{n-2}(x) - P_{n-1}(x)$.
   (For instance, $P_2(x) = 17x - 5$ and $P_3(x) = x^3 - 5x^2 - 12x + 5$.)

   (a) Show that $P_n(7) \equiv 0 \pmod{19}$ for every $n \in \mathbb{N}$.

   (b) Show that, for every prime q, if $P_{2013}(x) \not\equiv 0 \pmod{q}$, then $P_{2013}(x)$ has at most 2013 roots modulo q.

5. **Even more polynomials!**

   Consider two polynomials $p(x), q(x)$ whose product is zero: that is, $p(x) \cdot q(x) = 0$ for all x.

   (a) Show that if $p(x)$ and $q(x)$ are polynomials over the real numbers then in this case, either $p(x) = 0$ for all x or $q(x) = 0$ for all x (or both). (Hint: You may want to first prove this lemma, true in all fields: The set of roots of $p(x) \cdot q(x)$ is the union of the roots of $p(x)$ and $q(x)$.)

   (b) Show that, in contrast, over $GF(p)$ there exist such polynomials whose product is zero but which are both nonzero. (A polynomial $p(x)$ is nonzero if $p(x) \neq 0$ for some, but not necessarily all, x.) (Hint: Fermat's Little Theorem is useful here).

6. **Random Polynomials**

   Recall that a polynomial of degree d has at most d roots. In this problem we will show that most polynomials of degree (at most) d have even fewer roots. We will work modulo p, where p is a prime number such that $d \ll p$.

Let $f(x) = a_d x^d + a_{d-1} x^{d-1} + \cdots + a_0$ be a uniformly random polynomial of degree (at most) d. Recall that $f(x)$ can be picked by either picking each coefficient a_j to be a random number modulo p, or by picking randomly the values of the polynomial at $d+1$ chosen points x and interpolating to get $f(x)$.

Let Z be a random variable whose value is the number of roots of $f(x)$.

(a) What is the probability that $f(x)$ has a root at $x = 1$ (i.e $f(1) = 0$)?

(b) What is E[Z], the expected number of roots of $f(x)$?

(c) What does Markov's bound tell you about the probability that $f(x)$ has at least 10 roots?

(d) What is the probability that $f(x)$ has roots at $x = 1$ and $x = 5$ (i.e. $f(1) = 0$ and $f(5) = 0$)?

(e) Recall that $Var[Z] = E[Z^2] - E[Z]^2$. What is $Var[Z]$?

(f) What does Chebyshev's bound tell you about the probability that $f(x)$ has at least 10 roots?

7. **Secret Sharing**

Consider the following variant of the secret sharing problem. We wish to share a secret among twenty-one people, divided into three groups of seven, so that the following condition is satisfied. A subset of the twenty-one people can recover the secret if and only if it contains majorities (at least four out of seven) of at least two of the groups. How would you modify the standard secret sharing scheme to achieve this condition? (Hint: Try a two-level scheme, one level for groups, the other for people within the group.)

8. **How many secrets?**

A secret sharing scheme is k-secure if and only if any group of k or fewer people has probability at most $\frac{1}{q}$ of recovering the secret, where q is the number of possible choices for the secret (this means that the best strategy such a group has is to guess the secret at random). In the typical secret sharing scheme, the secret is $P(0)$, the value of a certain degree k polynomial (that we construct) at 0. Suppose that, instead, the secret is $P(0), P(1)$ (the values at both 0 and 1). Of course, we also change the algorithm by handing out $P(2), \ldots, P(n+1)$ to the n people instead of handing out $P(1), \ldots, P(n)$. Is this scheme still k-secure? Prover your answer.

9. **Spies**

The president wants to authorize military generals to launch nuclear weapons without the direct approval of the president if enough of them sanction this launch. As you might remember, secret sharing schemes come in handy here. But now there is a new twist. We have been informed that spies have infiltrated the army, and have even become generals. When it is time for a nuclear launch, the spies can give false information and therefore break the usual secret-sharing scheme.

There are 100 generals in the military. The president knows that out of those 100 at most 5 are spies. He wants a secret sharing mechanism where any group of 9 generals cannot launch the nuclear missiles. However he wants n generals to always be able to launch the nuclear missiles, no matter how many of the 5 spies are there among them. What mechanism should the president use? What is the correct bound n that guarantees any n generals can launch the missiles (even if there are spies among them).

10. **Win at Poker**

A *pseudorandom number generator* is a way of generating a large quantity of random-looking numbers, if all we have is a little bit of randomness (known as the *seed*). One simple scheme is the *linear congruential generator*, where we let x_0 denote the seed and define

$$x_{t+1} = (ax_t + b) \bmod m$$

for some modulus m and some constants a, b. (Notice that $0 \leq x_t < m$ holds for every t.)

You've discovered that a popular web site uses a linear congruential generator to generate poker hands for its players. For instance, it uses x_0 to pseudo-randomly pick the first card to go into your hand, x_1 to pseudo-randomly pick the second card to go into your hand, and so on. For extra security, the poker site has kept the parameters a and b secret, but you do know that the modulus is $m = 2^{31} - 1$ (which is prime).

Suppose that you can observe the values x_0, x_1, x_2, x_3, and x_4 from the information available to you, and that the values $x_5, \ldots, x_9$ will be used to pseudo-randomly pick the cards for the next person's hand. Describe how to efficiently predict the values $x_5, \ldots, x_9$, given the values known to you.

CS 70 Discrete Mathematics and Probability Theory
Fall 2013 Vazirani Error-correction Practice

1. **Error-correcting codes**

   In this question we will go through an example of error-correcting codes. Since we will do this by hand, the message we will send is going to be short, consisting of $n = 3$ numbers, each modulo 5, and the number of errors will be $k = 1$.

   (a) First, construct the message. Let $a_0 = 3$, $a_1 = 4$, and $a_2 = 2$; use the polynomial interpolation formula to construct a polynomial $P(x)$ of degree 2 (remember that all arithmetic is mod 5) so that $P(0) = a_0$, $P(1) = a_1$, and $P(2) = a_2$; then extend the message to length $n + 2k$ by adding $P(3)$ and $P(4)$. What is the polynomial $P(x)$ and what is the message that is sent?

   (b) Suppose the message is corrupted by changing a_0 to 0. Use the Berlekamp-Welsh method to detect the location of the error and to reconstruct the original message $a_0a_1a_2$. Show clearly all your work.

2. **Error-Detecting Codes**

   In the realm of error-correcting codes, we usually want to recover the original message if we detect any errors, and we want to provide a guarantee of being able to do this even if there are k malicious errors. Suppose that instead we are satisfied with detecting whether there is any error at all and do not care about the original message if we detect any errors. In class you saw that for recovering from at most k malicious errors when transmitting a message of length n you need to extend your message by $2k$ symbols and send a message of length $n + 2k$. But since we don't require recovering the original message, it is conceivable that we might need less symbols.

   (a) Formally suppose that we have a message consisting of n symbols that we want to transmit. We want to be able to detect whether there is any error if we are guaranteed that there can be at most k malicious errors. How should we extend our message (i.e. by how many symbols should we extend, and how should we get those symbols) in order to be able to detect whether our message has been corrupted on its way? You may assume that we work in $GF(p)$ for a very large prime number p.

   (b) If you were to detect non-malicious errors (i.e. random perturbations of some symbols), how would you do this using just one additional symbol? Obviously you won't be able to guarantee detection, but provide a reason why you think most cases of non-malicious errors will be detected by your algorithm.

3. **Possible Messages**

   Suppose Alice wants to transmit to Bob a polynomial P of degree ≤ 1 over $GF(5)$. She sends packets indicating the values of $P(0), P(1), P(2)$, and $P(3)$ so that Bob will be able to recover P even if two packets are dropped. However, a disaster happens and three packets are dropped: Bob only receives a packet indicating that $P(2) = 3$. Help Bob find a list of all the polynomials that P could have been given this information. (Note that Bob already knowns that the polynomial has degree ≤ 1 and is over $GF(5)$.)

4. **Magic!**

In this problem we will investigate what happens when in error-correcting codes there are fewer errors than the decoding algorithm is able to handle. For the entire problem we are working in $GF(7)$.

Assume that we wish to transfer a message of length 2 which we denote by (m_1, m_2). Each m_i is a member of $GF(7)$. We also wish to be able to correct up to $k = 2$ errors. Using the error-correcting codes we learned in class, we have to first find a polynomial $P(x)$ of degree at most 1 such that $P(1) = m_1$ and $P(2) = m_2$. Then we have to extend the message we send by $2k$ symbols. i.e. we will send $P(1), P(2), P(3), P(4), P(5), P(6)$ to the recipient.

(a) Consider an example where $(m_1, m_2) = (4, 2)$. What are the six symbols that are transmitted?

(b) Now assume that you have received these numbers: $5, 3, 4, 0, 3, 6$. i.e. if there were no errors then we would have $P(1) = 5, P(2) = 3, P(3) = 4, P(4) = 0, P(5) = 3, P(6) = 6$. Now, write down the linear equations that help decode error-correcting codes.

(c) In this part try to solve the linear equations you got in the previous section. You should observe that there are multiple solutions to these equations. Pick two different solutions and for each one write down the error-locating polynomial $E(x)$ and the polynomial $Q(x)$. In each of the two solutions divide $Q(x)$ by $E(x)$ to get the original polynomial. Do you get the same polynomial in both cases?

(d) Factorize $E(x)$ in each one of the two solutions you got to get its roots. Do they have a common root? What does that tell you about the position of errors in the transmitted message?

5. **Check Digits**

In this problem, we'll look at two real-world applications of check-digits.

In the first part, we'll look at International Standard Book Numbers (ISBNs). These are 10-digit codes $(d_1 d_2 \ldots d_{10})$ which are assigned by the publisher. These 10 digits contain information about the language, the publisher, and the number assigned to the book by the publisher. Additionally, the last digit d_{10} is a "check digit" selected so that $\sum_{i=1}^{10} i \cdot d_i \equiv 0 \mod 11$. (*Note that the letter X is used to represent the number 10 in the check digit.*)

(a) Suppose you have very worn copy of the (recommended) textbook for this class. You want to list it for sale online but you can only read the first nine digits: 0-07-288008-? (the dashes are only there for readability). What is the last digit? Please show your work, even if you actually have a copy of the textbook.

(b) Wikipedia says that you can determine the check digit by computing $\sum_{i=1}^{9} i \cdot d_i \mod 11$. Show that Wikipedia's description is equivalent to the above description.

(c) Prove that changing any single digit of the ISBN will render the ISBN invalid. That is, the check digit allows you to *detect* a single-digit substitution error.

(d) Does the check digit allow you to detect all two-digit errors (i.e., all errors where a pair of digits, not necessarily adjacent, are entered incorrectly)?

(e) Can you *switch* any two digits in an ISBN and still have it be a valid ISBN? For example, could 012345678X and 015342678X both be valid ISBNs?

(f) Now we'll look at another checksum formula: the Luhn formula (also known as the Luhn algorithm). This formula is used to verify the validity of credit card numbers. You can read more about it and see an example at `http://en.wikipedia.org/wiki/Luhn_algorithm` The algorithm is as follows:

i. Double each even-positioned digit, when counting from *right to left*.

ii. Determine the sum of the digits from each of the products in step (a).

iii. Sum the numbers from step (b). Find the sum of the unaffected digits (odd-positioned digits) in the original number. Combine these sums.

iv. Verify the account number by determining if the sum from step (c) is equivalent to 0 mod 10.

For clarification, an example from Wikipedia is shown below. In this example, $x = 3$ is the check digit.

Using the Luhn algorithm, determine the check digit x for the following number: $601143871005123x$.

(g) Can this algorithm detect if any two digits are switched? If not, which will it miss and why? *(Hint: you may look on Wikipedia to get started but explain the answer in your own words.)*

CS 70 Discrete Mathematics and Probability Theory
Fall 2013 Vazirani Graphs Practice

1. **Decomposing a non-Eulerian graph into Walks.**

   We proved in class that a connected (undirected) graph has an Eulerian cycle if and only if every vertex has even degree.

   Prove that if a graph G on n vertices has exactly $2d$ vertices of odd degree, then there are d walks (i.e. paths that can go through the same vertex more than once) that *together* cover all the edges of G (i.e., each edge of G occurs in exactly one of the d walks; and each of the walks should not contain any particular edge more than once).

2. **Directing a graph.**

   Suppose we have an (undirected) graph in which all vertices have even degree. Briefly describe how you would give a direction to each edge such that the indegree equals the outdegree of every vertex in the resulting directed graph.

3. **Trees**

   A *tree* is a connected undirected graph that has no cycles.

   (a) Show that every tree contains at least one vertex of degree 1.

   (b) Prove by induction on n that every tree with n vertices has exactly $n-1$ edges. [HINT: Use part (a) for the inductive step.]

4. **Polygons**

   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.

   $d(3) =$

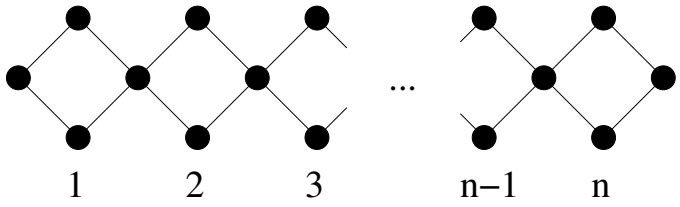
   (b) Calculate $d(4)$. You do not need to justify your answer.

   $d(4) =$

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

5. **Chains**

Consider "chain" graphs, shaped like this (with n "links"):



How many different Eulerian tours, starting and ending at the leftmost vertex, does a chain graph with n links have?

6. **Hypercube routing**

Recall that an n-dimensional hypercube contains 2^n vertices, each labeled with a distinct n bit string, and two vertices are adjacent if and only if their bit strings differ in exactly one position.

(a) The hypercube is a popular architecture for parallel computation. Let each vertex of the hypercube represent a processor and each edge represent a communication link. Suppose we want to send a packet for vertex x to vertex y. Consider the following "bit-fixing" algorithm:

> In each step, the current processor compares its address to the destination address of the packet. Let's say that the two addresses match up to the first k positions. The processor then forwards the packet and the destination address on to its neighboring processor whose address matches the destination address in at least the first $k+1$ positions. This process continues until the packet arrives at its destination.

Consider the following example where $n = 4$: Suppose that the source vertex is (1001) and the destination vertex is (0100). Give the sequence of processors that the packet is forwarded to using the bit-fixing algorithm.

(b) In general, for an arbitrary source vertex and arbitrary destination vertex, how many edges must the packet traverse under this algorithm? Give an exact answer in terms of the n-bit strings labeling source and destination vertices.

(c) Consider the following example where $n = 3$: Suppose that x is (010) and y is (100). What is the length of the shortest path between x and y? What is the set of all vertices and the set of all edges that lie on shortest paths between x and y. Do you see a pattern?

(d) Answer the last question for an arbitrary pair of vertices x and y in the hypercube. Can you describe the set of vertices and the set of edges that lie on shortest paths between x and y?

CS 70 Discrete Mathematics and Probability Theory
Fall 2013 Vazirani Counting + Sample Spaces Practice

1. **Counting, counting and counting**

   The only way to learn counting is to practice, practice, practice, so here is your chance to do so. We encourage you to leave your answer as an expression (rather than trying to evaluate it to get a specific number).

   (a) How many 10-bit strings are there that contain exactly 4 ones?

   (b) How many different 13-card bridge hands are there? (A bridge hand is obtained by selecting 13 cards from a standard 52-card deck. The order of the cards in a bridge hand is irrelevant.)

   (c) How many different 13-card bridge hands are there that contain no aces?

   (d) How many different 13-card bridge hands are there that contain all four aces?

   (e) How many different 13-card bridge hands are there that contain exactly 6 spades?

   (f) How many 99-bit strings are there that contain more ones than zeros?

   (g) If we have a standard 52-card deck, how many ways are there to order these 52 cards?

   (h) Two identical decks of 52 cards are mixed together, yielding a stack of 104 cards. How many different ways are there to order this stack of 104 cards?

   (i) How many different anagrams of FLORIDA are there? (An anagram of FLORIDA is any re-ordering of the letters of FLORIDA, i.e., any string made up of the letters F, L, O, R, I, D, and A, in any order. The anagram does not have to be an English word.)

   (j) How many different anagrams of ALASKA are there?

   (k) How many different anagrams of ALABAMA are there?

   (l) How many anagrams does the word PAPASAN have where the S and N are not adjacent? For example, count APPSANA but do not count APAANSP.

   (m) We have 9 balls, numbered 1 through 9, and 27 bins. How many different ways are there to distribute these 9 balls among the 27 bins?

   (n) We throw 9 identical balls into 7 bins. How many different ways are there to distribute these 9 balls among the 7 bins such that no bin is empty?

   (o) How many different ways are there to throw 9 identical balls into 27 bins?

   (p) How many ways are there to place 50 unlabeled balls in 9 labeled bins where each bin contains at least as many balls as its bin number. (That is, bin 1 contains at least 1 ball, bin 2 contains at least 2, and so on.)

   (q) There are exactly 20 students currently enrolled in a class. How many different ways are there to pair up the 20 students, so that each student is paired with one other student?

2. **More Counting**

   Let A, B be finite sets, with $|A| = m$ and $|B| = n$.

(a) How many subsets does A have? (Recall that the empty set and A are both subsets of A.)

(b) How many distinct functions $f : A \to B$ are there from A to B?

(c) Suppose $m = n$. How many distinct bijections are there from A to B?

3. **And More Counting**

How many non-negative integer solutions $(x_1, \ldots, x_7)$ are there to the following equation?

$$x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 = 2003$$
$$x_1 \geq 0, \ldots, x_7 \geq 0, \qquad x_1, \ldots, x_7 \in \mathbb{Z}$$

Order matters. For instance, $(1, 2002, 0, 0, 0, 0, 0)$ counts as a different solution than $(2002, 1, 0, 0, 0, 0, 0)$.

4. **Sum of digits**

Choose a number uniformly at random between 0 and 999,999, inclusive. What is the probability that the digits sum to 19?

5. **Algebraic vs. combinatorial proofs**

Consider the following identity:

$$\binom{2n}{2} = 2\binom{n}{2} + n^2.$$

(a) Prove the identity by algebraic manipulation (using the formula for the binomial coefficients).

(b) Prove the identity using a combinatorial argument. (Write both sides as the answer to a question of the form "how many ways can you...?")

6. **Red cards**

Consider a deck with just the four aces (red: hearts, diamonds; black: spades, clubs). Melissa shuffles the deck and draws the top two cards.

Given that Melissa has the ace of hearts, what is the probability that Melissa has both red cards?

Given that Melissa has at least one red card, what is the probability that she has both red cards?

7. **Sample Space and Events**

Consider the sample space Ω of all outcomes from flipping a coin 4 times.

(a) List all the outcomes in Ω. How many are there?

(b) Let A be the event that the first flip is a Heads. List all the outcomes in A. How many are there?

(c) Let B be the event that the third flip is a Heads. List all the outcomes in B. How many are there?

(d) Let C be the event that the first flip and the third flip are both Heads. List all the outcomes in C. How many are there?

(e) Let D be the event that the first flip or the third flip is a Heads. List all the outcomes in D. How many are there?

(f) Are the events A and B disjoint? Express the event C in terms of A and B. Express the event D in terms of A and B.

(g) Suppose now the coin is flipped $n \geq 3$ times instead of 4 flips. Compute $|\Omega|, |A|, |B|, |C|, |D|$.

8. **Probability Models**

Suppose you have two coins, one is biased with a probability of p coming up Heads, and one is biased with a probability of q coming up Heads. Answer the questions below, but you don't need to provide justifications.

(a) Suppose $p = 1$ and $q = 0$.

i. You pick one of the two coins randomly and flip it. You repeat this process n times, each time randomly picking one of the two coins and then flipping it. Consider the sample space Ω of all possible length n sequences of Heads and Tails so generated. Give a reasonable probability assignment (i.e. assign probabilities to all the outcomes) to model the situation.

ii. Now you pick one of the two coins randomly, but flip the *same* coin n times. Identify the sample space for this experiment together with a reasonable probability assignment to model the situation. Is your answer the same as in the previous part?

(b) Repeat the above two questions for arbitrary values of p and q. Express your answers in terms of p and q.

CS 70 Discrete Mathematics and Probability Theory
Fall 2013 Vazirani Conditional Probability Practice

1. **Independence**

   We flip two unbiased coins: a nickel and a dime, and consider the following events:

   (a) The nickel comes up heads.
   (b) The dime comes up heads.
   (c) The nickel and dime both come up heads.
   (d) Exactly one of the nickel and dime comes up heads.
   (e) The nickel and dime both come up the same way.

   State without proof whether each of the following pairs of events are independent:

   - (a) and (b):
   - (a) and (c):
   - (a) and (d):
   - (a) and (e):
   - (c) and (b):
   - (d) and (b):
   - (e) and (b):
   - (c) and (d):
   - (c) and (e):
   - (d) and (e):

   State without proof whether each of the following triples of events are independent:

   - (a), (b), (c):
   - (a), (b), (d):
   - (a), (b), (e):
   - (b), (d), (e):
   - (a), (c), (e):

2. **Correlation**

   It was suggested in class that, when $\Pr[A|B] > \Pr[A]$, then A and B may be viewed intuitively as being positively correlated. One might wonder whether "being positively correlated" is a symmetric relation. Prove or disprove: If $\Pr[A|B] > \Pr[A]$ holds, then $\Pr[B|A] > \Pr[B]$ must necessarily hold, too. (You may assume that both $\Pr[A|B]$ and $\Pr[B|A]$ are well-defined, i.e., neither $\Pr[A]$ nor $\Pr[B]$ are zero.)

3. **Monty Hall Again**

   In the three-door Monty Hall problem, there are two stages to the decision, the initial pick followed by the decision to stick with it or switch to the only other remaining alternative after the host has shown an incorrect door. An extension of the basic problem to multiple stages goes as follow.

   Suppose there are four doors, one of which is a winner. The host says: "You point to one of the doors, and then I will open one of the other non-winners. Then you decide whether to stick with your original pick or switch to one of the remaining doors. Then I will open another (other than the current pick) non-winner. You will then make your final decision by sticking with the door picked on the previous decision or by switching to the only other remaining door.

   (a) How many possible strategies are there?

   (b) For each of the possible strategies, calculate the probability of winning. What is the best strategy?

4. **Smokers**

   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.

5. **The myth of fingerprints**

   A crime has been committed. The police discover that the criminal has left DNA behind, and they compare the DNA fingerprint against a police database containing DNA fingerprints for 20 million people. Assume that the probability that two DNA fingerprints (falsely) match by chance is 1 in 10 million. Assume that, if the crime was committed by someone whose DNA fingerprint is on file in the police database, then it's certain that this will turn up as a match when the police compare the crime-scene evidence to their database; the only question is whether there will be any false matches.

   Let D denote the event that the criminal's DNA is in the database; $\neg D$ denotes the event that the criminal's DNA is not in the database. Assume that it is well-documented that half of all such crimes are committed by criminals in the database, i.e., assume that $\Pr[D] = \Pr[\neg D] = 1/2$. Let the random variable X denote the number of matches that are found when the police run the crime-scene sample against the DNA database.

   (a) Calculate $\Pr[X = 1|D]$.

   (b) Calculate $\Pr[X = 1|\neg D]$.

   (c) Calculate $\Pr[\neg D|X = 1]$. Evaluate the expression you get and compute this probability to at least two digits of precision.

   As it happens, the police find exactly one match, and promptly prosecute the corresponding individual. You are appointed a member of the jury, and the DNA match is the only evidence that the police present. During the trial, an expert witness testifies that the probability that two DNA fingerprints (falsely) match by chance is 1 in 10 million. In his summary statement, the prosecutor tells the jury that this means that the probability that the defendant is innocent is 1 in 10 million.

(d) What is wrong with the prosecutor's reasoning in the summary statement?

(e) Do you think the defendant should be convicted? Why or why not?

6. **Poisoned pancakes**

You have been hired as an actuary by IHOP corporate headquarters, and have been handed a report from Corporate Intelligence that indicates that a covert team of ninjas hired by Denny's will sneak into some IHOP, and will have time to poison five of the pancakes being prepared (they can't stay any longer to avoid being discovered by Pancake Security). Given that an IHOP kitchen has 50 pancakes being prepared, and there are ten patrons, each ordering five pancakes (which are chosen uniformly at random from the pancakes in the kitchen), calculate the probabilities that the first patron:

(a) will not receive any poisoned pancakes;

(b) will receive exactly one poisoned pancake;

(c) will receive at least one poisoned pancake;

(d) will receive at least one poisoned pancake given that the second patron received at least one poisoned pancake;

(e) Calculate the probability that any of the first three receive at least one poisoned pancake.

7. **Colorful coins**

We are given three coins. The first coin is a fair coin painted blue on the heads side and white on the tails side. The other two coins are biased so that the probability of heads is p. They are painted blue on the tails side and red on the heads side. One coin is randomly chosen and flipped twice.

(a) Describe the outcomes in the sample space, and give their probabilities. [NOTE: You may want to draw a tree to illustrate the sample space.]

(b) Now suppose two coins are chosen randomly *with replacement* and each flipped once. Describe the outcomes in the sample space in this new experiment, and give their probabilities. Are they the same as in part (a)? [NOTE: You may want to draw a tree to illustrate the sample space.]

(c) Now suppose two coins are chosen randomly *without replacement* and each flipped once. Describe the outcomes in the sample space in this new experiment, and give their probabilities. Are they the same as in parts (a) or (b)? [NOTE: You may want to draw a tree to illustrate the sample space.]

(d) Suppose the probability that the two sides that land face up are the same color is $\frac{29}{96}$ in the experiment in part (c). What does this tell you about the possible values of p?

(e) Let A be the event that you get a head on the first flip and B is the event that you get a head on the second flip. In each of the experiments in (a), (b) and (c), determine whether A and B are independent events.

8. **A paradox in conditional probability?**

Here is some on-time arrival data for two airlines, A and B, into the airports of Los Angeles and Chicago. (Predictably, both airlines perform better in LA, which is subject to less flight congestion and less bad weather.)

	Airline A		Airline B	
	# flights	# on time	#flights	# on time
Los Angeles	600	534	200	188
Chicago	250	176	900	685

(a) Which of the two airlines has a better chance of arriving on time into Los Angeles? What about Chicago?

(b) Which of the two airlines has a better chance of arriving on time overall?

(c) Do the results of parts (a) and (b) surprise you? Explain the apparent paradox, and interpret it in terms of conditional probabilities.

9. **A flippant choice**

We have noted that if a fair coin is flipped three times, there are eight equally probable outcomes: HHH, HHT, HTH, HTT, THH, THT, TTH, and TTT. Two CS 70 students play a game based on coin flipping. Player A selects one of the triplets just listed; player B selects a different one. The coin is then repeatedly flipped until one of the chosen triplets appears as a run and wins the game. For example, if player A chooses HHT and player B chooses THT and the flips are THHHT, player A wins.

Fill in the table below to show player B's best choice of triplet for each possible choice that player A makes, and the probability of player B winning with a best choice. Then explain why the odds for one player winning are so lopsided.

Player A's choice	Player B's best choice	Player B's probability of winning
HHH		
HHT		
HTH		
HTT		
THH		
THT		
TTH		
TTT		

10. **Stakes well done**

Two players, Alice and Bob, each stake 32 pistoles on a three-point, winner-take-all game of chance. The game is played in rounds; at each round, one of the two players gains a point and the other gains none. Normally the first player to reach 3 points would win the 64 pistoles. However, it starts to rain during the game, and play is suspended at a point where Alice has 2 points and Bob has 1 point. Alice and Bob have to figure out how to split the money.

You should assume that Alice and Bob are evenly matched, so that in each round Alice and Bob each have a 50% chance of winning the round. Assume also that Alice's share should be proportional to the conditional expected valueof her winnings (specifically, her winnings if the game were continued to the end from this point). The same goes for Bob.

Calculate a fair way to distribute the 64 pistoles using this notion of fairness. How many pistoles does Alice receive? Bob?

CS 70 Discrete Mathematics and Probability Theory
Fall 2013 Vazirani Random Variables Practice

1. **Waiting**

   A pair of dice is rolled until either a 4 is rolled (the numbers on the two dice add up to 4) or a 7 is rolled. What is the expected number of rolls needed?

2. **Poisson**

   The number of accidents (per month) at a certain factory has a Poisson distribution. If the probability that there is at least one accident is 1/2, what is the probability that there are exactly two accidents?

3. **Savings**

   Alice buys a piggy bank. Every day she picks a number X uniformly at random from $\{0,1,2\}$. If X is nonzero, she puts X dollars into her piggy bank for that day. If X is zero, she breaks the piggy bank and takes away all the money saved. What is the expected amount of money Alice gets when she breaks the bank?

4. **Quadruply-repeated ones**

   We say that a string of bits has k *quadruply-repeated ones* if there are k positions where four consecutive 1's appear in a row. For example, the string 0100111110 has two quadruply-repeated ones.

   What is the expected number of quadruply-repeated ones in a random n-bit string, when $n \geq 3$ and all n-bit strings are equally likely? Justify your answer.

5. **A flawed shuffle**

   Consider the following bad method for "shuffling" (i.e. randomly permuting the elements of) a 52-element array A.

   (a) Initialize an array *newA* to contain 52 "empty" indicators.

   (b) For $k = 0$ to 51, do the following:

      i. Repeatedly generate a random integer j between 0 and 51 until $A[j]$ isn't empty.

      ii. Copy $A[j]$ to $newA[k]$, and set $A[j]$ to "empty".

   (c) Copy *newA*, which now contains the shuffled elements, back to A.

   Determine the expected number of random integers j that will be generated to produce $newA[k]$.

6. **Find the joker**

   (a) Suppose you take an ordinary deck of 52 playing cards, and add a single joker. If you shuffle it and turn up the cards one at a time until the joker appears, on average how many cards are required until you see the joker?

   (b) Now take a deck of 52 playing cards, add two jokers, shuffle, and turn up cards one at a time until the first time that a joker appears. On average, how many cards are required until you see the first joker?

7. **Games**

   Consider the following game: Alice and Bob will each roll a fair, six-sided die. If Alice's die comes up with a number higher than Bob's, Alice wins $3 from Bob. If Bob's number comes up higher, or if they tie, Bob wins $2 from Alice. Is this game a good deal for Alice? Explain.

8. **Variance**

   You have a die which has one side with a 0, one side with a 2, and four sides with 1s. (So the six sides are 0,1,1,1,1,2.) You roll the die twice.
   Let X be the **product** of the two rolls.

   a. Compute $E[X]$.

   b. Compute $Var[X]$.

9. **St. Petersburg Paradox**
   Toss a fair coin repeatedly until it comes up heads; then stop. If it first comes up heads on the i-th toss, you win $\$2^i$. Let X denote how many dollars you win after playing this game once. Calculate $\mathbf{E}[X]$.

10. **Chopping up DNA**

    (a) In a certain biological experiment, a piece of DNA consisting of a linear sequence (or string) of 4000 nucleotides is subjected to bombardment by various enzymes. The effect of the bombardment is to randomly cut the string between pairs of adjacent nucleotides: each of the 3999 possible cuts occurs independently and with probability $\frac{1}{500}$. What is the expected number of pieces into which the string is cut? What is the variance? Justify your calculation.

    (b) Suppose that the cuts are no longer independent, but highly correlated: when a cut occurs in a particular location, nearby locations are much more likely to be cut as well. The probability of each individual cut remains $\frac{1}{500}$. Does the expected number of pieces increase, decrease, or stay the same? Justify your answer with a precise explanation. Can you say the same about variance?

11. **Random variables modulo p**

    Let the random variables X and Y be distributed independently and uniformly at random in the set $\{0,1,\ldots,p-1\}$, where $p > 2$ is a prime.

    (a) What is the expectation $\mathbf{E}[X]$?

    (b) Let $S = (X+Y) \bmod p$ and $T = XY \bmod p$. What are the distributions of S and T?

    (c) What are the expectations $\mathbf{E}[S]$ and $\mathbf{E}[T]$?

    (d) By linearity of expectation, we might expect that $\mathbf{E}[S] \equiv (\mathbf{E}[X] + \mathbf{E}[Y]) \pmod p$. Explain why this does not hold in the present context; i.e., why does the value for $\mathbf{E}[S]$ obtained in part (b) not contradict linearity of expectation?

    (e) Since X and Y are independent, we might expect that $\mathbf{E}[T] \equiv \mathbf{E}[X]\mathbf{E}[Y] \pmod p$. Does this hold in this case? Explain why/why not?

12. **Random Graph**

    You create a graph on n nodes by adding edge $[i, j]$, for any two of nodes i, j, with probability p.

    (a) What is the probability that the graph has no edge? That it is the complete graph? (In this and the following two questions, give your answer as an expression, however complicated, involving n and/or p.)

(b) What is the probability that the nodes $5, 7$, and 9 will form a triangle (i.e., that all three edges are present)?

(c) What is the expected number of triangles in the graph?

(d) Now you are told that $n = 10$, that $p \leq .5$, and that the variance of the number of edges in the graph is 10.8. What is p? (In this and the next question your answer can be an expression involving real numbers.)

(e) Use Chebyshev's inequality to bound the probability that the graph has at least 40 edges.

13. **The martingale**

Consider a *fair game* in a casino: on each play, you may stake any amount \$$S$; you win or lose with probability $\frac{1}{2}$ each (all plays being independent); if you win you get your stake back plus \$$S$; if you lose you lose your stake.

(a) What is the expected number of plays before your first win (including the play on which you win)?

(b) The following gambling strategy, known as the "martingale," was popular in European casinos in the 18th century: on the first play, stake \$1; on the second play \$2; on the third play \$4; on the kth play $\$2^{k-1}$. Stop (and leave the casino!) when you first win.

Show that, if you follow the martingale strategy, and assuming you have unlimited funds available, you will leave the casino \$1 richer with probability 1.

(c) To discover the catch in this seemingly infallible strategy, let X be the random variable that measures your maximum loss before winning (i.e., the amount of money you have lost *before* the play on which you win). Show that $\mathbf{E}[X] = \infty$. What does this imply about your ability to play the martingale strategy in practice?

14. **The evolution of a social network**

(We give a simplified analysis of the connectivity of a social network.)

Say one person in a class of n people knows a secret, perhaps where the midterm is. Occasionally a randomly chosen person A **who doesn't know the secret** calls a randomly chosen person B ($B \neq A$) and learns the secret if B knows it.

Let X_2 be a random variable that represents the number of calls (no two calls are simultaneous) until two people know the secret.

(a) What is the distribution of X_2?

(b) What is $E[X_2]$?

(c) Let X_i be the number of calls needed to go from $i-1$ people knowing the secret to i people. What is $E[X_i]$?

(d) What is the expected time for everyone to know the secret?

(e) Bound your expression to within a constant factor for large n. Your expression should not have a summation. (You may use $\Theta(\cdot)$ notation, recall that $2n^2 - 5n + 2 = \Theta(n^2)$.)

15. **Independent Poisson Variables**

Suppose you see two kinds of blips, red and blue. The blue blips are Poisson with intensity α and the red blips are independent and Poisson with intensity β. Suppose you are color blind and see all blips as being of a single kind. Show that the blips you see are Poisson distributed with intensity $\alpha + \beta$.

16. **Memorylessness**

We begin by proving two very useful properties of the exponential distribution. We then use them to solve a problem about the expected life of a package of batteries.

(a) Let r.v. X have geometric distribution with parameter p. Show that, for any positive integers m, n, we have

$$\Pr[X > m+n \mid X > m] = \Pr[X > n].$$

NOTE: This is called the "memoryless" property of the geometric distribution, because it says that conditioning on the past does not change the future distribution.

(b) Let r.v. X have exponential distribution with parameter λ. Show that, for any positive s, t, we have

$$\Pr[X > s+t \mid X > t] = \Pr[X > s].$$

NOTE: This is the memoryless property of the exponential distribution.

(c) Let r.v.'s X_1, X_2 be independent and exponentially distributed with parameters λ_1, λ_2. Show that the r.v. $Y = \min\{X_1, X_2\}$ is exponentially distributed with parameter $\lambda_1 + \lambda_2$. [Hint: work with cdf 's.]

(d) You have a digital camera that requires two batteries to operate. You purchase n batteries, labeled $1, 2, \ldots, n$, each of which has a lifetime that is exponentially distributed with parameter λ and is independent of all the other batteries. Initially you install batteries 1 and 2. Each time a battery fails, you replace it with the lowest-numbered unused battery. At the end of this process you will be left with just one working battery. What is the expected total time until the end of the process? Justify your answer.

(e) In the scenario of part (d), what is the probability that battery i is the last remaining working battery, as a function of i?

CS 70 Discrete Mathematics and Probability Theory
Fall 2013 Vazirani Continuous RVs Practice

1. **Square**

   Given a point chosen uniformly at random over a unit square (of sidelength 1), what is the density function of the random variable, X, corresponding to the distance to its border?

   Find the expected value of X.

2. **Random arrivals**

   Consider two people arriving uniformly at random in a time interval of one hour. We wish to bound how long the first has to wait for the second.

   Consider choosing two points X and Y independently and uniformly from the unit interval.

   (a) What is the joint density function $f(x,y)$? (Hint: it's not complicated.)

   (b) What is the expected value of $|x-y|$? (This is the expected time between arrivals.)

3. **Linearity of Expectation Again**

   Linearity of expectation holds for discrete r.v. Show that linearity of expectation holds for continuous random variables as well.

4. **Lunch Date**

   Alice and Bob agree to try to meet for lunch between 12 and 1pm at their favorite sushi restaurant. Being extremely busy they are unable to specify their arrival times exactly, and can say only that each of them will arrive (independently) at a time that is uniformly distributed within the hour. In order to avoid wasting precious time, if the other person is not there when they arrive they agree to wait exactly fifteen minutes before leaving. What is the probability that they will actually meet for lunch? Phrase your solution using the language of continuous random variables.

5. **Cumulative Distribution Function**

   In class, the statistics of a random variable are specified by the distribution in the discrete case and specified by the probability density function (pdf) in the continuous case. To unify the two cases, we can define the *cumulative distribution function*(cdf) F for a r.v., which is valid for both discrete and continuous r.v.'s:

   $$F(a) := \Pr[X \leq a], \qquad a \in \Re.$$

   (a) In the discrete case, show that the cdf of a r.v. contains exactly the same information as its distribution, by expressing F in terms of the distribution and expressing the distribution in terms of F.

   (b) In the continuous case, show that the cdf of a r.v. contains exactly the same information as its pdf, by expressing F in terms of the pdf and expressing the pdf in terms of F.

(c) Compute and plot the cdf for (i) $X \sim \text{Geom}(p)$, (ii) $X \sim \text{Exp}(\lambda)$.

(d) Identify two key properties that a cdf of any r.v. has to satisfy.

6. **Memorylessness**

We begin by proving two very useful properties of the exponential distribution. We then use them to solve a problem about the expected life of a package of batteries.

(a) Let r.v. X have geometric distribution with parameter p. Show that, for any positive m,n, we have

$$\Pr[X > m+n \mid X > m] = \Pr[X > n].$$

This is the "memoryless" property of the geometric distribution. Why do you think this property is called memoryless?

(b) Let r.v. X have exponential distribution with parameter λ. Show that, for any positive s,t, we have

$$\Pr[X > s+t \mid X > t] = \Pr[X > s].$$

[This is the "memoryless" property of the exponential distribution.]

(c) Let r.v.'s X_1, X_2 be independent and exponentially distributed with parameters λ_1, λ_2. Show that the r.v. $Y = \min\{X_1, X_2\}$ is exponentially distributed with parameter $\lambda_1 + \lambda_2$. [Hint: work with cdf's.]

(d) You have a digital camera that requires two batteries to operate. You purchase n batteries, labeled $1, 2, \ldots, n$, each of which has a lifetime that is exponentially distributed with parameter λ and is independent of all the other batteries. Initially you install batteries 1 and 2. Each time a battery fails, you replace it with the lowest-numbered unused battery. At the end of this process you will be left with just one working battery. What is the expected total time until the end of the process? Justify your answer.

(e) In the scenario of part (c), what is the probability that battery i is the last remaining working battery, as a function of i?

7. **A difference between discrete and continuous r.v.'s**

Discrete and continuous r.v.'s have a lot of similarities but some differences too.

(a) Suppose X is a discrete r.v. Let the r.v. $Y = cX$ for some constant c. Express the distribution of Y in terms of that of X.

(b) Suppose X is a continuous r.v. Let the r.v. $Y = cX$ for some constant c. Express the pdf of Y in terms of that of X. Is there any difference with the discrete case? [Hint: work with cdf's introduced in Question 5 on this handout.]

(c) If $X \sim N(\mu, \sigma^2)$, what is the density of $Y = cX$?

8. **Normal Distribution**

If a set of grades on a Discrete Math examination in an inferior school (not UC!) are approximately normally distributed with a mean of 64 and a standard deviation of 7.1, find:

(a) the lowest passing grade if the bottom 5% of the students fail the class;

(b) the highest B if the top 10% of the students are given A's.

NOTE: You may assume that if X is normal with mean 0 and variance 1, then $\Pr[X \leq 1.3] \approx 0.9$ and $\Pr[X \leq 1.65] \approx 0.95$.

CS 70 Discrete Mathematics and Probability Theory
Fall 2013 Vazirani Infinity Practice

1. **Cardinalities**
   Answer the following two questions, giving a proof in each case. You may use without proof any result covered in class provided you state it clearly.

   (a) Is the set of *pairs* of natural numbers countable or uncountable?

   (b) Is the set of irrational numbers countable or uncountable?

2. **Countability**

   (a) You are given an array of n bit strings $a_1, a_2, \ldots, a_n$, each of length n. Show how to construct another bit string b of length n such that $b \neq a_i$ for all $1 \leq i \leq n$ by only looking only at n bits in the array. (Looking at the i-th bit of a_j counts as one.)

   (b) Let X be the set of reals from 0.001 to 0.002. Show by a diagonalization argument that X is uncountable.

   (c) Show that the set of perfect powers is countable. (A perfect power is an integer x which equals d^2 where d is an integer.)

3. **Countable or uncountable?**
   Determine whether the following sets are countable or uncountable. For each countable set, display a one-to-one correspondence between the set of natural numbers and that set, or an enumeration of the set. For each uncountable set, explain why it is uncountable.

   (a) The set of binary strings which are palindromes. A string s is a palindrome if it can be written as the concatenation of some string t followed by the reversal of t.

   (b) The set of real numbers in the interval [0,1] whose decimal representation contains a single "1" digit, and all other digits are "0".

   (c) The set of real numbers in the interval [0,1] whose decimal representation contains only "0" and "1" digits (mixed in any order or combination).

   (d) The set of rooted, finite binary trees, in which trees are distinguished only by their shape (that is, you should ignore node values).

4. **Uncomputable**

   We say that two programs are equivalent if they give the same output on every input. Prove that it is impossible to write a computer program that takes as input two pieces of code, code1 and code2, and tests whether the two programs are equivalent.