CS 70 Discrete Mathematics and Probability Theory
Spring 2017 Rao Midterm 2 Solutions

PRINT Your Name: Oski Bear

SIGN Your Name: $\mathscr{OSKI}$

PRINT Your Student ID: ____

CIRCLE your exam room:
Pimentel 1 GPB 100 Hearst Annex A1 Soda 320 Latimer 120 Other

Name of the person sitting to your left: Papa Bear

Name of the person sitting to your right: Mama Bear

- After the exam starts, please *write your student ID (or name) on every odd page* (we will remove the staple when scanning your exam).
- We will not grade anything outside of the space provided for a problem unless we are clearly told in the space provided for the question to look elsewhere.
- On questions 1-2: You need only give the answer in the format requested (e.g., true/false, an expression, a statement.) We note that an expression may simply be a number or an expression with a relevant variable in it. **For short answer questions, correct clearly identified answers will receive full credit with no justification. Incorrect answers may receive partial credit.**
- On question 3-8, do give arguments, proofs or clear descriptions as requested.
- You may consult one sheet of notes. Apart from that, you may not look at books, notes, etc. Calculators, phones, and computers are not permitted.
- There are **14** single sided pages on the exam. Notify a proctor immediately if a page is missing.
- **You may, without proof, use theorems and lemmas that were proven in the notes and/or in lecture.**
- **You have 120 minutes: there are 8 questions on this exam worth a total of 115 points.**

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

1. Modular Arithmetic/RSA (30 points, 3 points each part)

Please write your answer in the provided box, or bubble in the corresponding option.

1. (Short Answer.) Give a number y modulo 35, where $y = 0 \pmod 5$ and $y = 1 \pmod 7$.
   Answer: 15. This is $5 \times (5^{-1} \pmod 7) \pmod{35}$ or one can enumerate the multiples of 5 and check.
2. (Short Answer.) Give a number y modulo 35, where $y = 1 \pmod 5$ and $y = 0 \pmod 7$.
   Answer: 21. Ths is $7 \times (7^{-1} \pmod 5) \pmod{35}$ or one can enumerate the multiples of 7 and check.
3. (Short Answer)Give a number y modulo 35, where $y = 4 \pmod 5$ and $y = 3 \pmod 7$.
   Answer: 24, which is $((4 \times 21) + (3 \times 15)) \pmod{35}$.
4. (True/False) The public key d is relatively prime to $(p-1)(q-1)$.
   Answer: True. Since d has an inverse (e) modulo $(p-1)(q-1)$, it must be that $gcd(d,(p-1)(q-1)) = 1$.
5. Consider an RSA scheme where $p = 23, q = 5$ and $e = 3$. What is d?
   Answer: $(p-1)(q-1) = 88$.

$$
\begin{aligned}
3(0) + (1)(88) &= 88 \\
3(1) + (0)(88) &= 3 \\
3(-29) + (1)(88) &= 1
\end{aligned}
$$

Thus $d = -29 = 59$.

6. CRT/FLT/Euler. Recall that the Chinese Remainder Theorem (CRT) implies for distinct primes $p_1, \dots, p_k$, and $N = p_1 \cdot p_2 \cdots p_k$ that there is a unique $x \in \{0, \dots, N-1\}$ which is the solution for $x = z_i \mod p_i$, for some $z_1, \dots, z_k$.

   (a) Say $x = 0$, what is the value of each $z_i = x \pmod{p_i}$?
   Answer: $z_i = 0$ for all i. $x = 0$ is $0 \pmod{p_i}$ for all p_i

   (b) Use the CRT to show that the set $S_N = \{x : gcd(x, N) = 1\}$ has size $(p_1 - 1)(p_2 - 1) \cdots (p_k - 1) = \prod_{i=1}^k (p_i - 1)$.
   Answer: For $x \in S_N$, $x = z_i \pmod{p}_i$, $z_i \neq 0$ since $gcd(x, p_i) = 1$.
   Thus, the number of possibilities for each z_i is $p_i - 1$. By the second rule of counting, we get that the total number is $\prod_{i=1}^k (p_i - 1)$.
   *First rule of counting

   (c) Show that for any $a \in S_N$ that $a^{\prod_{i=1}^k (p_i - 1)} \equiv 1 \pmod{N}$.
   Answer: Consider the function $f(x) = ax \pmod{N}$, for $a \in S_N$.
   This is a bijection from S_N to S_N as a has a multiplicative inverse and $ax \in S_N$ since $gcd(a, N) = 1$ and $gcd(x, N) = 1$.
   Thus, the set $T = \{ax | x \in S_N\} = S_N$. Taking the product of the elements of T and setting them equal to those of S_N, yields $a^{|S_N|} \prod_{x \in S_N} x = \prod_{x \in S_N} x \pmod{N}$.
   By observing that each element of S_N has an inverse, we can simplify the expression to $a^{|S_N|} = 1 \pmod{N}$. Replacing $|S_N|$ with $\prod_{i=1}^k (p_i - 1)$ yields the result.

   (d) Assume one chooses x uniformly from $\{0, \dots, N-1\}$. What is the probability that x is in S_N?
   Answer: $\frac{|S_N|}{N}$.

   (e) What is the probability that x is in S_N given that $gcd(x, p_1) = 1$?
   Answer: $\frac{|S_N|}{N \times \frac{p_1 - 1}{p_1}} = \frac{\prod_{i=2}^k (p_i - 1)}{\prod_{i=2}^k (p_i)}$

2. **Countability/Halting** (9 points, 3 for each True/False + 4 points)

Please write your answer in the provided box, or bubble in the corresponding option.

1. (True/False) For a countable set of strings, some of which have possibly infinite size, the union of all the finite sized substrings of each string is countable. (A substring of a string can be formed by choosing the positions in the original string to include in the substring. Here there will be a finite number of positions.)

   Answer: True. For one string, the set is countable as the substrings can be enumerated according by enumerating all 2^k substrings of the first k letters of the string. Now a countable union of countable sets is countable, again using interleaving; enumerate the first k elements of the first k sets, and then enumerate the say $2k$ elements of the first $2k$ elements. This rough enumeration may output the same element many times, so simply cross out an element that appeared previously.

2. (True/False) For any irrational number, there exists a computer program to compute it. (A program computes a number when it outputs any digit at a fixed finite time, even while running perhaps forever. For example, there are programs that compute $\sqrt{2}$ and π.)

   Answer: False. The set of computer programs is countable; the irrational numbers are uncountable.

3. (True/False). Let $f : X \to \mathscr{P}(X)$ be a function from a set X to its power set. True/False: The set $\{x \in X : x \notin f(x)\}$ is not in the range of f.

   Answer: True. Let $D = \{x \in X : x \notin f(x)\}$. Consider some set in the range, $S = f(y)$, if S contains y, then D does not, and if S does not contain y than D does. Thus $D \neq S$ for any S in the range.

4. (Argument/Proof.) Let M be a program that takes in a program P and an input x, and tells if you some variable v in the program P ever becomes the value 42 when you run P on the input x. Prove that such a program M cannot exist. **Answer:** Show that we can reduce *Halt* to M, or in other words, if we can solve M, we can solve the halting problem.

```
def Halt($P$, $x$)
    def $P''(x)$:
     Same code as in $P$, with instances of the variable $v$ are
     replaced by a new, unused variable so that there is no variable
     $v$ in the entire program
    def $P'(x)$:
      Run $P''$ on $x$
       $v = 42$
    return $M(P',x)$
```

   We have then solved the halting problem by using the program M. However, Halt cannot be solved, so neither can M.

3. Polynomial Related Questions. (24 points, 3 for each part)

Please write your answer in the provided box, or bubble in the corresponding option.

In the following, recall that a polynomial, $P(x)$, contains a point (a,b) when $P(a) = b$. And two polynomials, $P(x)$ and $Q(x)$, intersect at a point (a,b) when $P(a) = Q(a) = b$.

1. (Short Answer.) A degree (at most) five polynomial, $P(x)$, intersects a degree (at most) four polynomial, $Q(x)$ at k points. What is the smallest value of k where one can conclude that $Q(x) = P(x)$? (Here your bound should hold for any $P(x)$,$Q(x)$ that obey the degree bound.) **Answer:** $k = 6$. This is enough as $P(x) - Q(x)$ would have 6 zeros which indicates it is the zero polynomial which indicates $P(x) = Q(x)$. It is necessary since one can take $Q(x)$ to be the zero polynomial and $P(x)$ can have 5 zeros and not be the zero polynomial.

2. (Short Answer.) Consider a set of 11 points, and two degree 4 polynomials, $P(x)$ and $Q(x)$, where $P(x)$ contains k points and $Q(x)$ contains a possibly different set of k points. What is the minimum value of k where we can conclude $P(x) = Q(x)$.

   Answer: 8. The total number of containments (pigeons) is 16, thus at least $16 - 11 = 5$ points (holes) which are contained by both polynomials. Thus, they have 5 points where they intersect which means they are the same polynomial. If each contain fewer than 8 of the 11 points, then we can construct polynomials that only intersect at fewer than 5 points, in which case they could be different.

3. (Short Answer.) Give a degree 2 polynomial modulo 5 that contains $(1,3)$, $(2,0)$, $(3,0)$. **Answer:** $3(x-2)(x-3) \times (1-2)^{-1}(1-3)^{-1} = 4(x-2)(x-3) \pmod 5$.

   Note you do not have to compute the delta polynomials for $x = 2,3$ as those values are zeros.

4. (Short Answer.) If a channel loses f fraction of packets, how many packets does one need to send to recover a message of length n using polynomial encoding.

   Answer: If we lose k packets, we need to send $n+k$ packets. Here, if we send m packets, we could lose $k = mf$ packets. Thus, one gets $m = n + mf$. Solving for m, yields $n/(1-f)$.

5. (Short Answer.) If a channel corrupts f fraction of packets, how many packets does one need to send to recover a message of length n using polynomial encoding.

   Answer: $n/(1-2f)$. Similar to above, except we get $m = n + 2k = n + 2mf$ and solve for m.

6. (Short Answer.) Consider a two variable polynomial $Z(x,y) = P(x)Q(y)$ modulo a prime p where $P(x)$ and $Q(y)$ are nonzero degree d polynomials where $d < p$. What is the maximum number of distinct pairs of (i,j) that satisfy $Z(i,j) = 0 \pmod{p}$.

   Answer: $2pd - d^2$. This is the set $\{(i,j) : P(i) = 0\} \cup \{(i,j) : Q(j) = 0\}$. The first set has size at most pr_1 where r_1 is the number of roots of $P(x)$, the second has size at most pr_2 where r_2 is the number of roots of $Q(x)$, and the intersection size $r_1 r_2$. Use inclusion-exclusion to bound the union and then notice it is maximized when $r_1 = r_2 = d$.

7. (Short Answer.) Given the error polynomial from Berlekamp-Welsh algorithm, $x^2 + 3x + 2 \pmod{11}$, for what 'x' values are the points corrupted?

   Answer: The roots of this polynomial are -1 and -2, or 10 and 9 $\pmod{11}$. That is where the errors are.

8. (True/False.) There exists a bijection between the set of all ordered triples of reals (a,b,c), and the set of polynomials of degree at most 3 that pass through the point (3,3). **Answer:** True. We can map each triple (a,b,c) to the unique polynomial of degree at most three for which $P(0) = a, P(1) = b, P(2) = c, P(3) = 3$.

4. Counting. (9 points, 3 for each short answer + 4 points for combinatorial proof)

Please write your answer in the provided box, or bubble in the corresponding option.

1. (Short Answer.) What is the number of boy, girl, dog triples from n boys, m girls, and k dogs?
   Answer: nmk. First rule of counting.

2. (Short Answer.) What is the number of different permutations of the letters "SINHOCHEWHI"?
   Answer: $\frac{11!}{3!2!}$. Second rule of counting.

3. (Short Answer.) What are the number of ways to divide m dollar bills among z people?
   Answer: $\binom{m+z-1}{z-1}$. Stars are m dollars, $z-1$ bars to divide the dollars among z people.

4. Give a combinatorial proof that $\binom{k+n-1}{n-1} = \displaystyle\sum_{i=0}^{k} \binom{k-i+n-2}{n-2}$.

   Answer: The LHS is stars and bars, so it represents the number of ways to throw k indistinguishable balls into n distinguishable bins. The ith term of the RHS summation is again stars and bars, this time counting the number of ways to throw $k-i$ balls into $n-1$ bins. Thus, each term is actually the number of ways to throw i balls into the first bin, then the rest of the k balls into the other bins. So when we sum over all values of i, we get the total number of ways to throw k indistinguishable balls into n bins, the same as the LHS.

5. **Probability** (24 points, 3 for each part 1-8 + 35 points)

Please write your answer in the provided box, or bubble in the corresponding option.

In this question $P(\lambda)$ denotes the Poisson distribution with parameter lambda, $B(n,p)$ denotes the Binomial Distribution with parameters n and p, $G(p)$ denotes the Geometric Distribution with parameter p and $U\{1,\dots,n\}$ denotes the uniform distribution over the range $[1,n]$.

1. (True/False) If $Pr[A] > 0$ and $Pr[B] > 0$, and A and B are disjoint then A and B are not independent. **Answer:** True. $Pr[A|B] = 0 \neq Pr[A] > 0$.
2. (True/False) If $Pr[A|B] = .9$ and $Pr[C|B] = .8$, then $Pr[A] > Pr[C]$? **Answer:** False. Intuititvely, let A='Ebola Lite' and C = 'Flu' similar from class and B='Fever'. We let, $Pr[B] = .1$, $Pr[A] = .1$ with $Pr[A \cap B] = .09$, and $Pr[C] = .5$, and $Pr[C \cap B] = .08$
3. (True/False) For independent $X,Y \sim G(p)$, we have $X+Y \sim G(p/2)$. **Answer:** False. Consider $X,Y \sim G(.5)$. The $Pr[X+Y=2] = \frac{1}{4}$, but for $Z \sim G(.5/2) = G(\frac{1}{4})$, we have $Pr[Z=2] = (\frac{1}{4})(1-1/4) = 3/16$.
4. (True/False) For independent $X,Y \sim P(\lambda)$, we have $X+Y \sim P(2\lambda)$. **Answer:** True. Recall that the Poission can be found by dividing up time into small intervals and considering $B(n,\lambda/n)$ since one thinks of an event as a success in a trial in the limit where there is only one success in a trial. Thus, adding two random variables is really just thinking of $B(n,2\lambda/n)$ since an event now has probability $\lambda/n + \lambda/n - \lambda^2/n^2$ and the third term goes to zero much faster than the first two.
5. (True/False) For independent $X,Y \sim B(n,p)$, we have $X+Y \sim B(2n,p)$. **Answer:** True. X is the number of success in n trials with success probability p, Y is the number of successes in n trials with success probability p, $X+Y$ is the number of successes in $2n$ with probability p and thus is $B(2n,p)$.
6. (True/False) For independent $X,Y \sim U\{1,\dots,n\}$ we have $X+Y \sim U\{1,\dots,2n\}$. **Answer:** False. There are more ways to get n, $(X,Y) = (i,j)$ where $i+j=n$, than there are to get $2n$, $(X,Y) = (n,n)$.
7. (Short Answer) For independent $X \sim B(n,p), Y \sim G(p)$, what is $E[X+Y]$?. **Answer:** $pn + \frac{1}{p}$. Linearity of Expectation.

8. (3 points) Find a joint distribution for X and Y such that X and Y are each uniform on the set $\{1,2,3,4,5,6\}$, but (X,Y) is not uniform on $\{1,2,3,4,5,6\} \times \{1,2,3,4,5,6\}$.

   Answer: Take, for example, $\Pr(X=i, Y=i) = 1/6$ for $i = 1, \ldots, 6$.

9. (Quick Calculation) Let $Z \leq 4$ be a random variable with $\mathbf{E}[Z] = 2$. Give an upper bound for $\Pr(Z \leq 0)$.

   Answer: Apply Markov's inequality to $4 - Z \geq 0$, so

$$\Pr(Z \leq 0) = \Pr(4 - Z \geq 4) \leq \frac{4 - \mathbf{E}[Z]}{4} = \frac{1}{2}.$$

10. **Confidence Intervals.** (3 points)

    Consider two shuffled 52-card decks. Let X be the number of cards which are in the same position in both of the decks. Using Chebyshev's Inequality, your friend tells you that X is at least α with probability at most $1/2$. What is α?

    Answer: Note that X is the number of fixed points, so $\mathbf{E}(X) = 1 = \mathrm{Var}(X)$. By Chebyshev,

$$\Pr(|X-1| \geq \alpha - 1) \leq \frac{1}{(\alpha-1)^2} \leq \frac{1}{2},$$

so $\alpha = \sqrt{2} + 1$.

11. (13 points) Let X be geometric with parameter p, Y be Poisson with parameter λ, and $Z = \max(X,Y)$. For full credit, your final answers should not have summations. (Please try to do all parts using answers from previous ones as necessary.)

    (a) (5 points) Compute $P(Y < X)$.

    (b) (4 points) Compute $P(Z \geq X)$.

    (c) (4 points) Compute $P(Z \leq Y)$.

    Answer:

    (a) Condition on Y so you can use the nice property of geometric random variables that $P(X > k) = (1-p)^k$:

$$\begin{aligned}
P(Y < X) &= \sum_{y=0}^{\infty} P(X > Y | Y = y) P(Y = y) \\
&= \sum_{y=0}^{\infty} (1-p)^y \frac{e^{-\lambda}\lambda^y}{y!} \\
&= e^{-\lambda p} e^{\lambda p} \sum_{y=0}^{\infty} \frac{e^{-\lambda}(\lambda(1-p))^y}{y!} \\
&= e^{-\lambda p} \sum_{y=0}^{\infty} \frac{e^{-\lambda(1-p)}(\lambda(1-p))^y}{y!} \\
&= e^{-\lambda p}
\end{aligned}$$

    (to simplify the last summation we observed that the sum could be interpreted as the sum of the probabilities for a Poisson($\lambda(1-p)$) random variable, which is equal to 1)

    (b) 1, the max of X,Y is always at least X.

    (c) $P(Z \leq Y) = P(\max(X,Y) \leq Y) = P(X \leq Y) = 1 - P(X > Y) = 1 - e^{-\lambda p}$

12. (10 points) You are dealt 13 cards without replacement from a standard 52 card deck. Let X be the number of distinct values in your hand (The 13 possible values are Ace, 2, 3, 4, ..., Jack, Queen, King) ignoring suit. For instance, the hand, again dropping suits, (A, A, A, 2, 3, 4, 4, 5, 7, 9, 10, J, J) has 9 distinct values. (We expect expressions here, no need to simplify your answers or multiply out or anything.)

(a) (4 points) Calculate $E[X]$.

Answer: Let X_i be the indicator of the ith value appearing in your hand. Then, $X = X_1 + X_2 + \ldots + X_{13}$ (Let 13 correspond to K, 12 correspond to Q, 11 correspond to J). By linearity of expectation then, $E[X] = \sum_{i=1}^{13} E[X_i]$. We can calculate $\Pr[X_i = 1]$ by taking the complement, $1 - Pr[X_i = 0]$, or 1 minus the probability that the card does not appear in your hand. This is $1 - \frac{\binom{48}{13}}{\binom{52}{13}}$. Then, $E[X] = 13\Pr[X_1 = 1] = 13(1 - \frac{\binom{48}{13}}{\binom{52}{13}})$.

(b) (6 points) Calculate $Var[X]$.

Answer:

To calculate variance, since the indicators are not independent, we have to use the formula $E[X^2] = \sum_{i=j} E[X_i^2] + \sum_{i \neq j} E[X_i X_j]$.

$\sum_{i=j} E[X_i^2] = \sum_{i=j} E[X_i] = 13(1 - \frac{\binom{48}{13}}{\binom{52}{13}})$

To calculate $\Pr[X_i X_j = 1]$, we note that $\Pr[X_i X_j = 1] = 1 - \Pr[X_i = 0] - \Pr[X_j = 0] + \Pr[X_i X_j = 0]$.

$\sum_{i \neq j} E[X_i X_j] = 13 \cdot 12 \Pr[X_i X_j = 1] = 13 \cdot 12(1 - \Pr[X_i = 0] - \Pr[X_j = 0] + \Pr[X_i X_j = 0]) = 156(1 - 2\frac{\binom{48}{13}}{\binom{52}{13}} + \frac{\binom{44}{13}}{\binom{52}{13}})$

Putting it all together, we have $Var[X] = E[X^2] - E[X]^2 = 13(1 - \frac{\binom{48}{13}}{\binom{52}{13}}) + 156(1 - 2\frac{\binom{48}{13}}{\binom{52}{13}} + \frac{\binom{44}{13}}{\binom{52}{13}}) - (13(1 - \frac{\binom{48}{13}}{\binom{52}{13}}))^2$

13. (6 points) **Testing for cancer.** Let A be the event that a random male has prostate cancer, B be the event that the person tests positive for cancer. Let $Pr[B|A] = .9$, $Pr[B|\overline{A}] = .1$, and $Pr[A] = .1$.

    (a) (3 points) Given a positive test for a random male, what is the probability he has that cancer? (Can leave as an expression with numbers, no need to do the long division.)
    Answer: This is asking for $Pr[A|B] = \frac{Pr[A \cap B]}{Pr[A \cap B] + Pr[\overline{A} \cap B]}$.
    $Pr[A \cap B] = Pr[B|A] \times Pr[A] = .09$ $Pr[\overline{A} \cap B] = Pr[B|\overline{A}]Pr[\overline{A}] = .09$
    Plugging in, we obtain $.09/.18 = 1/2$.

    (b) (3 points) If one has prostate cancer, one may or may not die from it. We call death from cancer the event C, and say $Pr[C|A] = Pr[C|A \cap B] = .02$. (That is, C is independent of B conditioned on A or in other words the test result doesn't matter so much as the cancer.) If one has surgery, say the probability is p that one dies.
    Say one got a positive test, at what value of p should one not have surgery? In other words, at what value of p is the chance of dying from surgery at least as high as the chance of dying from prostate cancer for someone who tested positive. (Notice $Pr[C|\overline{A}] = 0$ as one can't die from this cancer if one doesn't have it.)
    Answer: We are asking for $Pr[C|B]$; the chance of dying from this cancer given a positive test.
    $Pr[C|B] = Pr[C|A \cap B]Pr[A|B] + Pr[C|\overline{A} \cap B]Pr[\overline{A}|B]$
    Now, $Pr[C|\overline{A} \cap B] = 0$ since $Pr[C|\overline{A}] = 0$, and $Pr[C|A \cap B] = .02$, and $Pr[A|B] = 1/2$ yields a result of $Pr[C|B] = (.02)(1/2) = 0.01$. So, do assess the risks/benefits where possible.