APC/MAT 199 – Math Alive – Topics for Final Paper

Here are suggestions for topics for the "Paper in lieu of Final Exam."

There are two options:

1. you can write an essay related to a topic covered in class, using appropriate mathematical analysis within your discussion,
2. you can solve a problem that requires mathematical thinking.

In the list below, the first category is marked with (Essay), the second with (Problem). Both options require you to engage and use mathematics! The difference is that in the second option, the emphasis is more on solving a definite mathematical problem that we pose, whereas in the first option, the emphasis is more on showing how mathematical problems arise in nature or society, and then analyzing other people's attempts to solve them.

For essays, we expect a text of at least 3,500 words. If your essay is shorter, don't start padding it. (Padding *will* lower your grade.) You may have the gift of succinctness, and be able to produce the depth and breadth we expect in fewer words, and then your essay will earn as much credit as a longer one would. In practice, we have found, however, that shorter essays are typically the result of insufficient research or reflection; hence the length guideline. Look up and list at the end of the essay several relevant references. Document your arguments with examples from those references, wherever relevant. Use mathematical analysis wherever relevant. For an example of how to incorporate mathematical terms and equations into the flow of English prose, see the lecture notes for the various units of this course.

For mathematical projects we want to see something that is related to the material covered in class, but was not actually covered in the class or in the labs. Again, we ask that you prepare a substantial report, explaining what the problem is you looked at, how it fits in with a topic we saw in class, and then explaining clearly what you figured out, and how you did it. If there are some conclusions to be drawn, give them! There is no length minimum for a mathematical project. Give the references that you used. You will be graded not only on whether you get the right answer, but also on the quality and lucidity of your presentation and explanation of HOW you got the result you did.

You can also do something that is intermediate, where you need to explain some mathematical concept to make your point, and where you then go on to a more general discussion. If you'd like to propose a topic of this nature, you'll need to clear it with I. Daubechies or S. Hughes.

If you pick a topic not on the list below, you also MUST clear it with I. Daubechies or S. Hughes.

In every case, you MUST mention all the sources you have used for your essay/problem.

Please read the list of topics related to the last three modules, and see if you'd like one of these.

Some topics from the midterm are available for final topics as well. They are listed here again, together with some other topics related to the first three modules and that we did not list earlier.

Make sure that the topic is sufficiently different from the one you did in the midterm; it should also pertain to a different module than your midterm paper. When in doubt, ask I. Daubechies or S. Hughes.

The absolute deadline for handing in the final papers is 5:00 pm on Tuesday May 12 (Dean's Date) at Fine 218 (use the box near the door). Exceptions will not be made if you lack permission of the Dean.

POSSIBLE TOPICS

- Birth, Death, Growth and Chaos
  - Beetlemania (Essay)
- Graph Theory
  - Small World (Essay)
  - The Five-Color Map Theorem (Problem)
  - Coloring Edges (Problem)
- Voting and Social Choice
  - A Survey of Apportionment Methods (Essay)
  - The Quota Method of Apportionment (Problem)
- Cryptography
  - Carmichael Numbers (Problem)
  - Divisibility Tests (Problem)
  - Pseudorandom Number Generation (Problem)
  - Primality Tests (Problem)
  - Secret Sharing Schemes (Essay)
- Error Correction and Compression
  - Check Digits Schemes (Problem)
- Probability and Statistics
  - How Many Restaurants? (Problem)
  - Cowboy Duel (Problem)

1 Birth, Death, Growth and Chaos

1.1 Beetlemania (Essay)

Write a mathematical review of the following article on population modeling of flour beetles: R.F. Constantino et al., Science 275: 389-391, 1997. Use the 'Beetlemania' article from 'What's Happening in the Mathematical Sciences 1998-99', Barry Cipra, AMS Press, 1999, as a guide. Copies of the latter article are available from Valerie Marino at the reception desk on the second floor in Fine Hall (room 205). Describe the rationale for each of the terms in the iterated map model for the three life stages of the beetle, and interpret them, and the regular and chaotic dynamics to which they lead, in the light of our discussions of the logistic equation in Birth, Death, Growth and Chaos unit.

2 Graph Theory

2.1 Small World (Essay)

Read both parts of the following two-part article in American Scientist:

- Hayes, B., Graph Theory in Practice: Part I, American Scientist, Vol. 88, No. 1. (Jan.-Feb. 2000), p. 9.
- Hayes, B., Graph Theory in Practice: Part II, American Scientist, Vol. 88, No. 2. (Mar.-Apr. 2000), p. 104.

In the second part, the author discusses various models of graphs that might approximate (to a greater or lesser extent) small-world network graphs such as the internet or the Hollywood graph.

Your job is to explain some of these models as best you can, being as mathematically precise as possible. You will need to follow up some facts into other articles, preferably in the primary literature. First start with Erdös-Renyi random graphs. Describe how to construct a random graph. How do these graphs succeed and how do they fail in modeling typical small-world network graphs?

Now we consider the more elaborate models of Watts and Strogatz. Watts and Strogatz start with a graph based on a lattice and then modify it. Explain what a lattice is. What is percolation in a lattice? How would this relate to small-world network phenomena?

Now describe the Watts-Strogatz model in mathematical detail. Describe the strengths and weaknesses of their model. Discuss how Kleinberg modified the Watts-Strogatz model for the purposes of investigating how one might look for short paths while navigating in a small world. You can find a lot of good materials linked on Kleinbergs webpage at Cornell. Describe the Kleinberg "greedy algorithm" for finding short paths through his networks. Be as mathematical as possible in explaining what Kleinberg found.

Discuss "power laws." Try to find (on the internet or in the literature) as many estimates of power laws for various small-world graphs (such as the internet, or Hollywood, or call graphs) as you can.

2.2 The Five-Color Map Theorem (Problem)

Prove that any planar graph (and therefore every geographical map) is 5-colorable. Your proof must be logically and mathematically complete, thorough, and correct, and presented with great clarity, and in your own words. While trying to understand this problem, you may read other peoples proofs, but you may not look at them while writing down your own version of the proof. Dont make notes or other records of other peoples proofs. Wait at least an hour after looking at other peoples proofs before working on your own to make sure you have internalized the arguments yourself.

2.3 Coloring Edges

During the class, we were interested in colorings of vertices. Now we shall consider colorings of edges.

Let G be the complete graph on 7 vertices, i.e., G has 7 vertices, and each pair of vertices is connected by an edge (but there are no multiple edges or loops). How many edges are there in G? How many triangles are there in G? (Be careful in counting: For any three vertices, we count only one triangle involving them, since we do not care about what order the vertices and edges are traversed going around the triangle.)

Now we color all the edges of G; each edge is either red or blue. A *one-color triangle* in G is a triangle all of whose edges are colored one color (either all three edges are blue, or all three are red, but no mix of colors is allowed). Prove that there are at least four one-color triangles in G no matter how we chose to color the edges of$G$. Show that there is at least one way of coloring G that makes it have only four one-color triangles.

3 Voting and Social Choice

3.1 A Survey of Apportionment Methods (Essay)

The United States Constitution says the following regarding the House of Representatives:

> Representatives and direct Taxes shall be apportioned among the several States which may be included within this Union, according to their respective Numbers.... The Number of Representatives shall not exceed one for every thirty Thousand, but each State shall have at Least one Representative...

Of course, it is practically impossible that all states have exactly the same proportion of representatives to people. This is the problem of apportionment.

Read about and discuss the various methods suggested to try to come as close as possible to satisfying what the Constitution asks for. Be sure to cover the methods of Hamilton, Jefferson, Webster, and Hill-Huntington. Describe each method as precisely and mathematically as you can. Discuss the strengths and weaknesses of each method, using mathematical analysis. Describe different measures people have used to evaluate how equitable a proposed apportionment is. You will find *For All Practical Purposes* (on reserve in the Fine Hall Library) useful for this.

3.2 The Quota Method of Apportionment (Problem)

For an introduction to this problem, read the above description of the previous topic (A Survey of Apportionment Methods). For this project, it will help to read (in *For All Practical Purposes* or the papers mentioned below) about the apportionment methods (Hamilton, Jefferson, Webster, Hill- Huntington) mentioned in the description of the above project, and to read what is said concerning their strengths and weaknesses, but you do not need to describe or analyze these methods as in the previous project. The focus of this project is different: to understand a new method proposed by Balinski and Young in the following papers:

- Balinski, M. L., and H. P. Young, The Quota Method of Apportionment, The American Mathematical Monthly, Vol. 82, No. 7. (Aug.-Sep., 1975), pp. 701-730.
- Balinski, M. L., and H. P. Young, A New Method for Congressional Apportionment, Proceedings of the National Academy of Sciences, Vol. 71, No. 11. (Nov. 1974), pp. 4602-4606.

Describe precisely the algorithm (procedure) for apportioning votes under this method. Describe precisely what Balinski and Young mean by the following three properties: “housemonotone,” “consistent,” and “satisfies quota.” After looking at their paper and proofs, explain why their quota method satisfies all three properties. This is the difficult part of the project – it involves trying to understand their proof that the method satisfies quota. (Or coming up with your own proof!) You must be mathematically correct, thorough, and clear.

4 Cryptography

4.1 Carmichael Numbers (Problem)

Fermat’s little theorem gave us a way to think about whether a number is prime or not without factoring it. Compute 2^{N-1}(mod N), 3^{N-1}(mod N), 5^{N-1}(mod N), and 7^{N-1}(mod N) for $N = 1729$. What can you conclude? What happens if you try to factor 1729 using a pocket calculator?

Find out about pseudoprimes or Carmichael numbers in a book on elementary Number Theory or by searching the web. Explain what they are and find at least one example of a Carmichael number. What is it about a Carmichael number that makes it “behave like” a prime? What mathematical conditions must be satisfied for a number to be a Carmichael number? Show how these conditions lead to the behavior that mimics true primes.

4.2 Divisibility Tests (Problem)

We know from experience that the numbers that are divisible by 2 are exactly those whose last digit is 0, 2, 4, 6 or 8. Similarly, a number is a multiple of 5 if its last digit is a 0 or a 5. Using modular arithmetic we can explain why these divisibility tests work. Just as we did when we talked about check digits we can expand a number N in powers of 10:

$$54237915 = 5 \times 107 + 4 \times 106 + 2 \times 105 + 3 \times 104 + 7 \times 103 + 9 \times 102 + 1 \times 10 + 5$$

Since 10 is equivalent to 0 (mod 5), this expansion shows that that 54237915 = 5 = 0(mod 5). In other words, N is a multiple of 5. There are many other divisibility tests you may have seen before. A number is divisible by 3 if and only if the sum of its digits is a multiple of 3. For example 1267221 is a multiple of 3 because the sum of its digits is 21 and 21 is a multiple of 3. Use modular arithmetic and the expansion of N into powers of 10 to explain why this test works. A similar test tells you if N is divisible by 11. The test for divisibility by 11 is based on the fact that 10 = -1(mod 11). Figure out what this test is and show why it works. There is also a test for divisibility by 7 and 13 based on the fact that $1001 = 7 \times 11 \times 13$. Can you figure out this test and explain how it works? Try to find a test for divisibility by 17. Can you explain how Fermat's little theorem gives a divisibility test for any prime p?

4.3 Pseudorandom Digits Generation (Problem)

Very often, scientists need random numbers for their computer programs, for example, when they are simulating random processes that occur in nature. Since computers dont actually do anything randomly, it is impossible to actually get a random number from a computer. But for many applications, a set of numbers that seem random are good enough, so we settle for that. A computer typically has a pseudorandom number generator, which is just a program that generates numbers that seem random in many respects, but are not, in fact, random.

Let us use modular arithmetic to try to generate a sequence of random digits. We will use the prime number p = 1069 in our computation. We will use the number 342 as our "seed," or starting point. Now we generate the our first three digits. We calculate seed $\times$ seed (modulo p) = 342 $\times$ 342 (mod 1069) = 443. Then we generate the next three digits by multiplying the result 443 by the seed 342 modulo p, and we will get 777. Then we continue: we multiply each result by the seed modulo 1069 to get the next 3-digit entry of the sequence. In case we obtain a number below 100, we add leading zeros to get three digits total. In case we get a number more than 999, we disregard this number and continue (that is, we do not use it in our sequence, but we still multiply it with 342 to obtain the next 3-digit number).

Explain why we want to add leading zeros to numbers with fewer than 3-digits, and explain why we disregard numbers with more than 3 digits. That is, explain why the digits would be less uniformly distributed if we didn't follow these procedures.

Explain why this pseudorandom digit generator will always end up generating digits in cycles. (That is why it is a pseudo-random digit generator.)

Try different seeds. Find seeds that generate cycles of different lengths. Explain how we should try to choose seeds to make the cycle longer. Explain which seeds we need to avoid, so as not to get easy patterns.

4.4 Primality Tests (Problem)

Wilson's Theorem (1770) states that if n is an integer greater than one, then n is prime if and only if $(n-1)! = -1$(mod n). We could use this as a test to see whether a number is prime, but $(n-1)!$ is difficult to calculate as compared to a^{n-1}, so primality tests based on Fermat's little theorem are much more practical.

Explain why Wilson's theorem is true.

If n is an integer greater than one and n is not prime, what do we get when we reduce $(n-1)!$ modulo n?

Make an estimate of how much time proving primality with Wilson's theorem can take.

On the page `http://www.utm.edu/research/primes/prove/index.html` you can find a discussion of primality testing.

This web site contains several primality tests that are good for special cases, that is, when the candidate-prime number has a special form. Pick one of the tests and explain how it works.

The site also gives some probabilistic tests. Pick one of them and explain how it works.

Search the Internet and find the biggest known prime. How many digits does it have? What is special about this number?

Explain how it is possible for the biggest known prime to have many more digits than the largest numbers that present-day factoring algorithms can handle.

4.5 Secret Sharing Schemes (Essay)

Imagine the government trusts you and 15 other people to work on a secret project this Saturday. To enter the extreme-security building you work in, you and your co-workers must enter a secret number, say X, into a computer on the door of the building. You get a secret key (which is different from X), and your co-workers get their own secret keys. Because the government doesn't trust any one of you on your own, or even just two or three of you together, security demands that there must be at least 11 people together, with their 11 different keys, to get access to the building. (Allowance is made for the possibility that up to 4 people cannot make it, because of illness or other reasons.) The idea is that you and those 10 other people must carry out a computation first, together, to figure out the secret key to the door. The computation won't work without all 11 keys. (Note however, that it should work whenever 11 different people among the 15 get together!) What sort of computation should the government set up for you to do?

Here's a quote from the website `http://www.rsasecurity.com`

> Secret sharing schemes were discovered independently by Blakley and Shamir. The motivation for secret sharing is secure key management. In some situations, there is usually one secret key that provides access to many important files. If such a key is lost (for example, the person who knows the key becomes unavailable, or the computer which stores the key is destroyed), then all the important files become inaccessible. The basic idea in secret sharing is to divide the secret key into pieces and distribute the pieces to different persons in a group so that certain subsets of the group can get together to recover the key.

The security scheme of Shamir is based on polynomial interpolation. Please describe in detail how and why this scheme works, and give an example of a set of 4 keys which together encodes a message using Shamir's scheme. If any of these keys are missing, one should not be able to decode the message! For some good descriptions of this scheme, please check:

- `http://www.rsasecurity.com/rsalabs/faq/2-1-9.html`
- `http://www.rsasecurity.com/rsalabs/faq/3-6-12.html`

Note: There are other secret sharing schemes as well. If you would like to use another secret sharing scheme instead, you must ask I. Daubechies or S. Hughes ahead of time.

5 Error Correction and Compression

5.1 Check Digits Schemes (Problem)

In class and in the class notes, we looked at only a few check digit schemes: airline tickets, bank ID numbers on checks, American Express check numbers, Postnet codes. *For All Practical Purpose* mentions more. The notes also give indications of how to find out what mistakes (either single digit errors or inversions of two consecutive digits) can be detected by the check digit pattern. Such schemes usually have the following form (or something equivalent): the sum of the information digits and the check digit (possibly a weighted sum in which different digits are multiplied by different numbers) must come out to zero modulo a modular base b. We assume the information digits are 0 to 9, but we allow the check digit to have higher numerical values (which can be represented with letters or other symbols, as in the ISBN code, where "X" has a value of 10). Show that such schemes that can detect all single digit errors and all inversions MUST have a modular base of at least 11. Explain why the ISBN scheme (described in the notes; it does work mod 11) can detect all these mistakes. Note: Since some of this material was partially covered in class, your explanation should be especially lucid and complete!

Suppose we were to use fewer than ten information digits (e.g., we use only 0 to 8 to encode our information) in our modular arithmetic error-detection schemes mentioned above. Is it now possible to design a scheme that detects all single-digit replacements and all transpositions? For each modular base b from 2 to 9, suppose we use only the digits from 0 to $b-1$ to encode information. For each b from 2 to 9, show whether it is possible to design a error-detection scheme that corrects all single-digit replacements and all transpositions. Recall that a system need not use the simple sum of the digits, but can give different weight to different digits. For each b, either design a scheme that detects all single-digit replacements and all transpositions, or show why this is impossible. In each case, you may assume that your information is a string of twenty digits, to which you add one check digit.

6 Probability and Statistics

6.1 How Many Restaurants? (Problem)

Your job is to try to estimate how many pizzerias there are in the Joe's Pizzeria chain. Fortunately for you, each restaurant has a placard on the wall that tells you the number of that particular restaurant, and we are certain that Joe numbers his restaurants consecutively. Unfortunately for you, you are only allowed to visit one randomly selected restaurant. Say you walk into restaurant nr. 543. Then what is your best guess about how many pizzerias there are?

Well, it depends on what you think constitutes a good guess. One method is the Maximum Likelihood Estimator. The general framework: we observe a fact, and we have various explanations of that fact (models of what is going on that would explain the outcome we saw), and we want to decide which model fits best what actually happened. In our example, we have a lot of different models for what we saw. Maybe there are 1000 Joe's pizzerias, and we walked into restaurant number nr. 543. Maybe there are 2000 restaurants. Maybe 7712.

- If our model is that there are 1000 restaurants, then the probability that someone walks randomly into restaurant nr. 543 is 1/1000.
- If our model is that there are 2000 restaurants, then the probability that someone walks randomly into restaurant nr. 543 is 1/2000.
- If our model is that there are 7712 restaurants, then the probability that someone walks randomly into restaurant nr. 543 is 1/7712.

The Maximum Likelihood Estimator says the following: for each model, take the outcome that ACTUALLY

HAPPENED to you, and compute how probable that would be in the model. The "best" model, according to this method, is the model which has the greatest probability of producing the outcome you saw.

In the Joe's pizzeria problem, a model with fewer restaurants (e.g., 1000 instead of 2000) will produce a higher probability of walking at random into nr. 543. Unless the model has fewer than 543 restaurants! (If there are 400 restaurants, the probability of walking into nr. 543 is zero.) So we want to reduce the number of restaurants as much as possible without having fewer than 543. So the Maximum Likelihood Estimator tells us that the best guess for the total number of restaurants is 543.

And, in general, if we walk into restaurant M, then the Maximum Likelihood estimator tells us that we should guess that there are M restaurants! It is clear that you never overestimate the number of restaurants this way, but rather, you always underestimate unless you are lucky enough to walk into the last pizzeria in the chain.

Now your estimator becomes a statistic, that is, a random variable based on what you experience (which we consider to be a random process-which restaurant you go to).

- If there are really N pizzerias, what is the expected value of your maximum likelihood estimator (guess of the number of pizzerias) in terms of N? You walk into restaurant M, and declare M as your guess for the total number of Joe's pizzerias. So what we are asking here is: What is $E[M]$?
- What is the variance of the maximum likelihood estimator in terms of the total number of restaurants? That is, compute $Var(M)$.

You decide that the maximum likelihood estimator is very silly, and want an estimator whose expected value is N when there are N restaurants. Of course you dont know N; you only know the number M of the randomly-chosen restaurant you visit. The maximum likelihood estimator says that M is the guess of the total number, but $E[M]$ is not N.

- Find a simple formula involving M that will have expected value N. We will call this formula (which is a random variable based on M) our "one-visit estimator."
- What is the variance of the one-visit estimator that you just defined?

Now suppose that you are allowed two randomly-selected visits. Because they are totally random and independent, with no preference given to any restaurant, it is even possible that you are sent to the same restaurant twice. Let L and M be the numbers of the restaurants you visit.

- If I choose three constants, a, b, and c, and make my estimator equal to $aL + bM + c$, then what is the expected value of this estimator?
- What constraints should be put on a, b, and c to ensure that the estimator above has expected value equal to N?
- Pick values of a, b, and c which make $E[aL + bM + c] = N$ while making $Var(aL + bM + c)$ as small as possible. You want the variance to be small to be sure that your estimator is likely to be close to N. We call the estimator with these optimal choices of a, b, c the "two-visit estimator."

Now we can imagine going on three independent random visits, seeing restaurants numbered K, L, and M, and making an estimator $aK + bL + cM + d$.

- What constraints on a, b, c, and d are necessary to make sure that our estimator has $E[aK + bL + cM + d] = N$?
- How should we set a, b, c, and d to minimize the variance?
- Show that the estimator produced by your last choices is nonetheless goofy, by giving a scenario where the value it predicts for the total number of restaurants is actually LESS than one of the three numbers K, L, or M, that you ACTUALLY OBSERVED. Show that the two-visit and one-visit estimators developed above do not have this weakness.

Now suppose we are trying to estimate the number of total pizzerias in a county (not just those belonging to the Joe's chain). Now assume that no pizzeria has any information posted in it like Joe's numbers. We are sent by our dispatcher Louie to restaurants at random (repeats possible), and we continue to visit until the first time when we have a repeat visit. Call the total number of restaurants visited V (this is one less than the number of visits, since the last visit is to one of the previously-visited restaurants).

- If there is only one restaurant, what is $E[V]$?
- If there are two restaurants, what is $E[V]$?
- If there are three restaurants, what is $E[V]$?
- If there are four restaurants, what is $E[V]$?
- Can you come up with a formula for $E[V]$ when there are N restaurants? (Your formula might not be simple, and it can be written as a sum of terms ending in $\ldots$, as long as you make clear the generic form of the terms.)

Some facts that may be useful in this problem are that the sum of the first N numbers (i.e., $1+2+3+\ +N)$ is equal to $N(N+1)/2$, and that the sum of the first N squares (i.e., $1+4+9+\ +N^2$) is equal to $N(N+1)(2N+1)/6$.

6.2 Cowboy Duel (Problem)

Three cowboys Clint, Wayne, and Jesse are ready to fight a paint-gun duel. The cowboys take turns shooting and, if someone is hit, he is out. They have agreed on the sequence of shooting, which will be that Jesse will shoot first, Wayne second, and Clint third. They keep shooting in that order until one is paint-splattered and eliminated. Then, the other two continue shooting at each other in turns, starting with whichever of the two remaining cowboys is entitled to the next turn (i.e., the one who didn't hit the first cowboy). The two remaining cowboys go on like this until one of them is hit. The remaining one, unblemished, is then the winner.

It turns out that Clint and Wayne dont like each other, and thus, given the opportunity, Clint will always shoot at Wayne and Wayne will always shoot at Clint. Jesse is not so particular and is mainly interested in winning. Thus, he will choose to shoot at Clint or Wayne or pass (i.e., miss on purpose – we assume he always misses if he wants to) depending on which will maximize his chances of winning.

1. Just to warm up, assume that Clint, Wayne, and Jesse are all perfect shots (i.e., each one always hits his target). What is the optimal strategy for Jesse (i.e., shoot at Clint, shoot at Wayne, or pass) and what are Jesse's chances of winning if he uses that strategy?
2. Now assume that Clint hits his target 90% of the time, Wayne hits his target 80% of the time, and Jesse hits his target 50% of the time. What is the optimal strategy for Jesse? What are Jesse's chances of winning if he uses that strategy? 3
3. Answer the same questions for Clint 90%, Wayne 30%, Jesse 50%.
4. Answer the same questions for Clint 90%, Wayne 20%, Jesse 50%.
5. Answer the same questions for Clint 90%, Wayne 20%, Jesse 10%.
6. Answer the same questions for Clint 30%, Wayne 20%, Jesse 10%.
7. For each of the cases 1 through 6, assuming Jesse uses the optimal strategy, what are the chances that Clint wins?