

# Things to remember for 6.042 Final

David Chen

December 11, 2010

## Abstract

The following is a short enumeration of certain topics that are “good to know” for the upcoming 6.042 exam. Readers should take note that this document is in no way meant to be a comprehensive treatment of the topics on the exam. Nor is this document meant to be a representative sample of the question subjects that will be on the midterm. For any lingering doubts, please consult Prof. Leighton’s text or email 6042-staff@

## 1 Things to remind yourself

This test is not unlike all the other tests that you’ve taken before.

- Justify everything to yourself, especially in counting and probability.
- Part of counting is being extremely picky: have I covered all the cases that describe this set? Am I double-counting somewhere?
- When you are counting things, make sure you are not implicitly inserting an ordering where you don’t need it
- In a related note, do not blindly use combinations in situations where you might need an ordering (for example, problem 3 of the 2008 final).
- Expectation is linear. Variance is not. You can only add variances when the variables are independent.
- On the other hand, when two events are independent and you need to calculate variance, you’d best add the variances of the individual events!
- Chernoff bounds are extremely tight - however, you can only use them when your events are independent
- Before using a result from this class, inspect the assumptions of the result. Are they met?
- Believe in yourself. All the *techniques* that are needed for this test, you have learned. It is up to you to figure out which one to use, and to use it correctly.

## 2 Sums and Asymptotics

Students should be familiar with sums of the form  $\sum_{i=1}^n f(i)$ . It simply means to sum up the values of  $f(i)$  as  $i$  goes from 1 to  $n$ :

$$f(1) + f(2) + \dots + f(n)$$

Calculating sums is a mixture of trickery and “standard techniques”. Before we begin, let’s just review some standard sums (**know these by heart!**)

- $\sum_{i=1}^n i = \frac{n(n+1)}{2}$
- $\sum_{i=1}^n i^2 = \frac{n(n+1)(2n+1)}{6}$  (okay don’t need to know this but still wouldn’t hurt)
- $\sum_{i=0}^{\infty} r^i = \frac{1}{1-r}$  if  $|r| < 1$ . NOTE also that the summation index starts at  $i = 0$ : that is, the sum is  $1 + r + r^2 + \dots$
- $\sum_{i=1}^n 1/i = H_n$ . This isn’t an identity, but a definition. This is called the  $n^{\text{th}}$  harmonic number.

**The Perturbation Method** is an algebraic way to find sum identities. For example, consider

$$S = 1 + x + x^2 + \dots + x^{n-1}$$

We “perturb” it by multiplying an  $x$  to both sides of the equation:

$$xS = x + x^2 + \dots + x^n$$

But look! It seems like a lot of the terms in  $S$  and  $xS$  match each other! With the power of subtraction, we can make a difference (ha, ha) and find that

$$S - xS = 1 - x^n$$

Factoring out  $S$  on the left side and dividing out, we find that

$$S = \frac{1 - x^n}{1 - x}$$

It turns out that sums and asymptotics was on the last test. You guys should have a decent enough understanding of it (right?), so I will just forgo describing asymptotic notation and leave it up to you guys to email or come to office hours if you have trouble with that.

Then again, you should always know Stirling’s Approximation:

$$n! \approx \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$$

### 3 Recurrences

Recurrences are when you define a sequence... recursively. We have two main techniques for solving them. They are useful for finding solutions. however, in order to prove that a solution is correct,

#### 3.1 General Linear Recurrences

The way to solve these problems is pretty formulaic. So you should all get a perfect... right?

**Homogeneous Linear Recurrences** are recurrences of the form

$$f(n) = a_1f(n-1) + a_2f(n-2) + \dots + a_df(n-d)$$

Substitute in  $f(n) = x^n$ , we get

$$x^n = a_1x^{n-1} + a_2x^{n-2} + \dots + a_dx^{n-d}$$

Dividing out by  $x^{n-d}$ , we have the *characteristic equation* of the recurrence,

$$x^d = a_1x^{d-1} + a_2x^{d-2} + \dots + a_{d-2}x + a_d$$

- If  $r$  is a nonrepeated root of this equation, then  $r^n$  is a solution
- If  $r$  is a repeated root with multiplicity  $k$ , then  $r^n, nr^n, \dots, n^{k-1}r^n$  are all solutions

The fact that you have to find roots to an algebraic equation should be a strong hint to review how to solve algebraic equations! Cubic and higher-order equations are difficulter to solve, so any problem on the exam would probably involve guessing solutions to factor it. As for quadratic and lower - you should know how to solve those.

Now, what we have are a bunch of things that are solutions to the linear recurrence. We usually also give boundary conditions too, enough so that the solution to the recurrence is unique. Because that previous sentence was confusing, let's try doing an example. Suppose my recurrence were:

$$f(n) = 3f(n-1) - 2f(n-2)$$

with initial conditions  $f(0) = 6$  and  $f(1) = 11$ . Formidable numbers indeed.

We first hypothesize that  $f(n) = x^n$ , and obtain the characteristic equation of  $x^2 = 3x - 2$ . It is easy to see that this factors into

$$(x-1)(x-2) = 0$$

so the roots of the equation are 1 and 2. From this, we know that the solution to this recurrence must be of the form

$$A(1^n) + B(2^n) = f(n)$$

We can now plug in this value of  $f(n)$  into our initial conditions to find the solution.  $f(0) = 6$ , so  $A + B = 6$ . Furthermore,  $f(1) = 11$ , so  $A + 2B = 11$ . Through the quick magic of subtraction, we find  $A = 1$  and  $B = 5$ , and arrive at the happy solution of

$$f(n) = 1 + 5(2^n)$$

**nonhomogeneous Linear Recurrences** are slightly more difficult to solve. They are recurrences of the form

$$f(n) = a_1f(n-1) + a_2f(n-2) + \dots + a_df(n-d) + g(n)$$

with initial conditions. The thing to do here is to find a *particular solution* that satisfies the entire equation, and to find the *homogenous solution*, and then to add them together. We already know how to find the homogeneous solution: just remove  $g$  from the picture to obtain a homogeneous equation, and solve as before. **STRONG NOTE:** we use the initial conditions at the **END** to determine the coefficients in the homogeneous solution.

So the steps are:

- remove  $g(n)$  to obtain a homogeneous equation, then find the homogeneous solution.
- Guess a particular solution. Hints for this are in the FTL book on page 301.
- Add the homogeneous and particular solutions together
- Use the initial conditions to solve for coefficients.

### 3.2 The Akra-Bazzi method

This is our weapon of choice for solving divide-and-conquer recurrences. Given a recurrence of the form

$$T(x) = \begin{cases} \text{is nonnegative and bounded} & \text{for } 0 \leq x \leq x_0 \\ \sum_{i=1}^k a_i T(b_i x + h_i(x)) + g(x) & \text{for } x > x_0 \end{cases}$$

where

1.  $a_1, \dots, a_k$  are positive constants
2.  $b_1, \dots, b_k$  are constants between 0 and 1.
3.  $x_0$  is large enough such that  $T$  is well-defined.
4.  $g(x) \geq 0$  and  $|g'(x)| = O(n^m)$  for some integer  $m$ .
5.  $|h_i(x)| = O(x/\log^2 x)$

Then,

$$T(x) = \Theta \left( x^p \left( 1 + \int_1^x \frac{g(u)}{u^{p+1}} du \right) \right)$$

where  $p$  obeys

$$\sum_{i=1}^k a_i b_i^p = 1$$

There, that wasn't so bad... was it?

## 4 Basic Counting

Being able to count stuff is pretty important for this class, both for pure combinatorics (how many ways can things happen?) and for probability (probability of event A happening = (# ways for event A to happen)/(# of total possibilities)). Thus, it is important to know how to count!

### 4.1 Sums

This is the easiest thing in the world. Say you have two sets,  $A$  and  $B$ , and you want their union. If  $A$  and  $B$  are disjoint (that is, they do not overlap), then their union has size

$$|A \cup B| = |A| + |B|$$

This sounds obvious, and it is. The real challenge is figuring out when to use this. One uses this counting method usually when they partition the set of “stuff they want” into multiple groups - often times, counting is easier on certain subsets. For example, consider counting the number of ways one can get a five-card poker hand higher than an ace-high flush. We would break the counting into different sets of hands: one for full houses, one for four-of-a-kinds, and one for straight flushes, and add those up in the end. Why? Because those sets are disjoint, and we want the union of them. It is cases like this when we use the summing rule.

### 4.2 Products

You've all seen this problem in elementary school. Joe is in the lunch line at Next Dining. There are 3 choices at the grill, 5 choices of drinks, and 2 desserts. If you get exactly one of each type of food, how many ways can you make your dinner? The answer, unsurprisingly, is  $5 * 3 * 2 = 30$ .

The big idea behind the product rule is that you are taking *products* of sets. A dinner in this case is really a length-3 vector of (g, dr, de), where the first entry is associated to a grill item, the second to a drink, and the third to a dessert. Any time that you are making... a structure that has different elements from different sets, you would want to consider the product rule.

The product rule is stated like so: if  $A_1, A_2, \dots, A_k$  are sets, then the set  $A_1 \prod A_2 \prod \dots \prod A_k$  has size

$$|A_1 \times A_2 \times \dots \times A_k| = |A_1| \cdot |A_2| \cdot \dots \cdot |A_k|$$

### 4.3 Taking things in order

Let's say I have a bag of  $n$  numbers:  $\{1, 2, \dots, n\}$ , and I want to take out  $k$  of them to make a  $k$ -length vector (what I do with my vectors is none of your business). How many ways are there to do this? Well, start with the leftmost entry of the vector. There are  $n$  choices for that entry. Moving to the right, we successively have  $n - 1$  choices, then  $n - 2$ , and so on. In total, we have  $n(n - 1) \dots (n - k + 1)$  ways to make this vector, or  $n!/(n - k)!$  ways to do this. Note also that when  $k = n$ , this comes out to  $n!$ .

This is actually a pretty powerful idea!  $n!$  is the number of ways we can order  $n$  distinct things.  $n!/(n - k)!$  is the number of ways we can order  $k$  things chosen from  $n$  distinct things. What can we do with this?

Well, for one, we can now make assignments of stuff to people. For example: Say we were playing a game show, and there are 8 different prizes. Arthur, Buster, and Charlene each get a prize: how many ways are there to give prizes out? Well, we are simply choosing 3 prizes out of 8 prizes, where order matters. Essentially, we are making a 3-vector  $(P_1, P_2, P_3)$  where each  $P_i$  is a prize, and the first entry corresponds to Arthur's prize, the second corresponds to Buster's prize, and the third one is Charlene's prize. There are  $8!/5!$  ways to do that.

There is also another extremely important idea here: that of taking away ordering (basically, the bookkeeper problems from the recitation). Let's say I've already counted all the orderings of these  $k$  things, but there are always  $m$  things that I want to consider equivalent. Then, I must divide the entire counted sum by  $m!$  to "get rid of" the ordering of these  $m$  identical objects. The reason for this is a symmetry argument, and it is encouraged for students to reason out why this is so for themselves.

The example we did in class was: how many rearrangements of the letters in the word "book-keeper" are there? Well, there are 10 letters in the word, so if we treat them as distinct entities, there are  $10!$  ways to order it. However, there are repeated letters: "o" and "k" show up twice, and "e" shows up three times. So we must divide out by  $2!$  twice and  $3!$  once to get rid of the repeats: the final answer is

$$10!/2!2!3!$$

### 4.4 Choosing items from a set

Often times, we also care only about how many ways there are to select  $k$  items from a set of  $n$  items. Order doesn't matter; all that matters is that we have the ones we want. In that case, there are

$$\binom{n}{k} = \frac{n!}{k!(n - k)!}$$

ways to do choose  $k$  objects from  $n$  things. For example, if I'm a pokemon trainer with 151 distinct pokemon, and I need to bring 6 along for my team (let's pretend there is no sense of "first position pokemon"), there are  $\binom{151}{6}$  ways to do so. If I'm making pizzas and can afford 3 toppings from a selection of 15 toppings, there are  $\binom{15}{3}$  ways for me to do so. Once again, this is used when order doesn't matter. It just matters what gets chosen and what doesn't get chosen.

Say I ask you the question of: how many binary strings of length  $n$  are there with  $k$  ones? If the variable naming and placement of this question in the “choosing” section didn’t tip you off, you should know that the answer is  $\binom{n}{k}$ . But why? We can think of an  $n$ -bit string as a sequence of  $n$  blank slots, and we get to choose the slots where we place 1’s (by “choosing slots,” we mean choosing the index it is associated with). It does not matter if the chosen slots are in any order, only that they are chosen. Because there are only ones and zeroes in this string, choosing all the places where we have ones also uniquely determines where the zeroes are.

## 4.5 Stars and Bars

The “stars and bars” problem is a pretty standard problem in combinatorics. Essentially, asks: how many solutions are there to the equation

$$x_1 + x_2 + \dots + x_n = c, \text{ where } x_i \geq 0, \forall i$$

The way to view this is through partitions. Imagine a binary string of length  $d + n - 1$ , with exactly  $n - 1$  ones and the rest zeros. We can create a bijection (one-to-one correspondence) between all such strings and all assignments of  $x_i$  that satisfy the above constraints. The idea is to imagine the 1’s as partitions of the 0’s, and the 0’s as single counters. Thus, let  $x_1$  be the number of 0’s to the left of the first 1, let  $x_2$  be the number of 0’s between the first and second 1, and so on. From every such string, we can give an assignment of variables; and from each assignment of variables, we can generate the corresponding binary string. Thus, this is a bijection, and so the number of solutions to the equation is equal to the number of such strings, which we have already found to be

$$\binom{n + d - 1}{n - 1}$$

## 4.6 Combinatorial Proof

A combinatorial proof is one in which you try to count the same set in two different ways, thus giving you an equality of two seeming not-the-same algebraic things. It’s pretty magical, and the only thing I can say is just open your mind and try to count in different ways.

## 4.7 Pigeonhole Principle

The idea behind the Pigeonhole Principle is really quite simple. It is just that if you have  $n + 1$  things, and  $n$  places to put them in, then there must be one place that contains at least 2 things. If a student is taking 6 exams, and there are only 5 times, they must have a conflict. In pigeonhole problems, it is extremely important to properly identify what are your pigeons and what are your pigeonholes. Also, it is extremely important to identify *when* to use it. “Collision” type problems are a huge hint to use pigeonhole. That is, if we need to show that there are two things that take on the same value, or fit in the same category. For further examples, prove that in a group of 367 people, two people must share the same birthday.

## 5 Probability

To think about the probability of an event, you need to have two things:

- A universe of possible things that can happen. This is usually called the sample space. Call it  $U$ .
- The set of thing of “stuff you care about”. If you want the probability of event  $A$ , this is the set of all instances of  $A$ .

The probability of event  $A$  happening in the universe is simply the ratio of “ $|A|/|S|$ ”. I use quotation marks because the “size” of  $A$  should not exactly be the number of elements it has - each individual outcome can have different outcomes, so it’s more like a weighted sum of the elements of  $A$ . It’s the ratio of the size of the set you care about to the size of the entire universe of possibility. At its core, probability is inherently about areas - that is why venn diagrams are sometimes very helpful in visualizing probabilistic statements.

There is something in the book about how to break up events into a tree of steps happening. Most people seem to understand this, so I won’t go into it. However, if you are unsure of yourself, please consult section 14.2 and beyond in the FTL text.

### 5.1 Rules about Probability

A countable **sample space**  $\mathcal{S}$  is a nonempty countable set. An event  $w \in \mathcal{S}$  is an **outcome**. A subset of  $\mathcal{S}$  is called an **event**. A probability function  $Pr$  is then a function that maps from  $\mathcal{S}$  to  $\mathbb{R}$  that has the properties:

- $Pr(w) \geq 0, \forall w \in \mathcal{S}$  (there are no negative probabilities)
- $\sum_{w \in \mathcal{S}} Pr(w) = 1$  (total probability of “something happening” is 1).

For an event  $E \subseteq \mathcal{S}$ , we define

$$Pr(E) = \sum_{w \in E} Pr(w)$$

That is, the probability of event  $E$  is the sum of the probabilities of its disjoint constituents. Think about this in the area-of-sets way; it makes a lot of sense!

The following are true:

- For disjoint events  $E$  and  $F$ ,

$$Pr[E \cup F] = Pr[E] + Pr[F]$$

(analogously, for sets  $A$  and  $B$ ,  $|A \cup B| = |A| + |B|$ .)

- If  $\{E_0, E_1, \dots\}$  is a collection of disjoint events, then

$$Pr[\cup_{n \in \mathbb{N}} E_n] = \sum_{n \in \mathbb{N}} Pr[E_n]$$



- $Pr[\bar{A}] = 1 - Pr[A]$

(The probability of something not happening is 1 minus the probability it will happen)

There are a few more notes on uniform probability spaces and infinite probability spaces. Uniform probability spaces are just when each outcome is equally likely (think being drawn a hand of cards), and you can use straight combinatorics to find the probability of things happening. For example: say I am dealt a five-card poker hand. What is the probability that I get four-of-a-kind?

The answer is that I have 13 choices for the value of the four-of-a-kind, and 48 choices for the last card. Thus, there are  $13 * 48$  hands that can make this happen. However, there are  $\binom{52}{5}$  ways to get a poker hand, so sadly, my chances of making a nasty hand are  $\frac{13*48}{\binom{52}{5}}$ .

## 5.2 Conditional Probability

Suppose I have two events  $A$  and  $B$ . The probability of  $A$  given  $B$  is:

$$Pr[A|B] = \frac{Pr[A, B]}{Pr[B]}$$

But what does the value on the left mean? It means the probability of event  $A$  occurring, given that event  $B$  has happened. Suppose  $A$  were the event that a student gets a good score on the final exam, and  $B$  is the event that the student studied for the exam. Then we can talk about  $Pr[A|B]$  as: given that some student has studied for the exam, what is the probability that they get an  $A$  on it?

Another way to look at this is in thinking about the sets that these events are. When you condition on  $B$ , you are really just reducing your sample space to only the set of  $B$ . And then, asking about  $Pr[A|B]$  is really finding the area of  $A$  that lies in  $B$ , and dividing out by the area of  $B$  to normalize. It's easy to visualize in venn diagram form, so if you want help with this, show up during office hours.

We have a few more things to remember:

- **Baye's Rule** states that

$$Pr[A|B] = \frac{Pr[B|A]Pr[A]}{Pr[B]}$$

This follows directly from the definition of conditional probability

- **Law of Total Probability, single event:** if  $Pr[E]$  and  $Pr[\bar{E}]$  are nonzero, then

$$Pr[A] = Pr[A|E] \cdot Pr[E] + Pr[A|\bar{E}] \cdot Pr[\bar{E}]$$

Basically,  $Pr[A] = Pr[A, E] + Pr[A, \bar{E}]$ . It makes sense, no? The two terms are disjoint sets, and because  $E \cup \bar{E} =$  the sample space, we are really adding up all instances of  $A$  on the right side.

**Independence** is a pretty important idea in conditional probability. Events  $A$  and  $B$  are independent if  $Pr[B] = 0$  or

$$Pr[A|B] = Pr[A]$$

That is, even if we know that  $B$  happened, no change is made to the probability distribution of  $A$ . If we let  $A$  be the event that a student does well on the exam, and  $B$  be that they studied, then  $A$  and  $B$  are not independent: one affects the other! However, if we flip a coin twice and let  $A$  be that you get a heads on the first try and  $B$  be that you get a heads on the second, then they are independent.

Knowing that events are independent is very powerful, if the events  $E_1, E_2, \dots, E_n$  are mutually independent, then we have

$$Pr[\cap_i E_i] = \prod_i Pr[E_i]$$

## 6 Random Variables, Distributions

A random variable is a variable that can take on different values with certain probabilities. For example, say we flip a coin 30 times, and let  $X$  be the number of heads that appear. Then,  $X = 0$  with a certain probability,  $X = 1$  with another probability, and so on. Note that this means that any function of  $X$  is also a random variable:  $X^2 = 0$  with a certain probability,  $X^2 = 1$  with another probability,  $X^2 = 4$  with another probability, and so on.

A special type of random variable is the **indicator random variable**. It is just a random variable associated to an event that is 1 when the event happens, and 0 when it does not happen. It is useful for finding the expected number of events that happen (when we consider a whole collection of trials). For example, say we flip 10 coins and want to find the number of heads that we get. You use indicator random variables for this!

### 6.1 Expectation

if  $R$  is a random variable in sample space  $S$ , the expectation of  $R$  is:

$$Ex[R] = \sum_{w \in S} R(w)Pr[w]$$

It is a weighted sum of the values that  $R$  can take on, where the weights are the probabilities of the respective values. For example, let  $X$  be the outcome of a dice roll. The expectation is  $E[X] = 1/6(1 + 2 + 3 + 4 + 5 + 6) = 3.5$ .

A special case is the expectation of an indicator random variable. Consider an indicator random variable  $X$  of an event  $A$ . Then,

$$E[X] = Pr[A](1) + Pr[\bar{A}](0) = Pr[A]$$

Everyone should know that **Expectation is linear**. This means, for random variables  $A$  and  $B$  and for a constant  $C$ :

$$E[A + B] = E[A] + E[B]$$

$$E[cA] = cE[A]$$

This extends into more than 2 variables as well.

Note that this linearity of expectation is perfect for summing indicator random variables. For example, say I roll 60 dice. What is the expected number of 6's that turn up? Well, let  $X_i$  be an indicator random variable for the event that dice  $i$  turns up a 6. Let  $X = \sum_i X_i$ . Then,

$$Ex[X] = Ex\left[\sum_i X_i\right] = \sum_i Ex[X_i] = 60 * 1/6 = 10$$

Easy, no?

Note that while the expectation of a sum is the sum of the expectations, the same is not true of the product of random variables. Given random variables  $A$  and  $B$ ,

$$Ex[AB] = Ex[A]Ex[B]$$

ONLY WHEN  $A$  and  $B$  are independent.

## 6.2 Mean time to failure

The **mean-time-to-failure** problem is a pretty common idea. Say I am flipping a coin where  $Pr[H] = p$  and  $Pr[T] = 1 - p$ . I flip until I get a heads. What is the expected number of flips that I do?

Well, we can simply calculate the expectation directly! Let  $X$  be a random variable describing exactly how many flips I do before hitting heads. Then,  $Pr[X = 1] = p$ ,  $Pr[X = 2] = p(1 - p)$ ,  $Pr[X = 3] = p(1 - p)^2$ , and so on. So

$$Ex[X] = 1 * p + 2p(1 - p) + 3p(1 - p)^2 + \dots$$

After some manipulation, you find that it is simply  $1/p$ . If you can't be bothered to do the summation yourself, remember this result!

## 6.3 Conditional Expectation

Conditional expectation  $Ex[R|A]$  is defined as:

$$Ex[R|A] = \sum_{r \in \text{range}(R)} r \cdot Pr[R = r|A]$$

The law of total expectation states that if  $R$  is a random variable on sample space  $S$ , and  $A_1, A_2, \dots$  is a partition of  $S$ , then

$$Ex[R] = \sum_i Ex[R|A_i]Pr[A_i]$$

## 7 statistics

The variance  $Var[R]$  of a random variable is defined as

$$Var[R] = Ex[(R - Ex[R])^2]$$

We know how to calculate expectations... so this should be easy! Variance also obeys these properties (REMEMBER THEM)

- $Var[R] = Ex[R^2] - Ex^2[R]$ . Remember how squares of random variables are random variables? Remember how to calculate expectations of random variables?
- For any constants  $a, b$ ,  $Var[aR + b] = a^2 Var[R]$ .
- **If  $R_1$  and  $R_2$  are independent**,  $Var[R_1 + R_2] = Var[R_1] + Var[R_2]$ . This is super useful. Try to use it whenever you have to calculate the variance of something that is the sum of variables on independent events.

Finally, we have the following bounds that were discussed in the final recitation:

- **Markov's Inequality**: if  $X$  is a nonnegative random variable and  $c > 0$ , then

$$Pr(X \geq c) \leq \frac{Ex[X]}{c}$$

- **Chebyshev's Inequality**: For all  $x > 0$  and any random variable  $R$ ,

$$Pr(|R - Ex[R]| \geq x) \leq Var[R]/x^2$$

- **Murphy's Law**: If events  $E_1, \dots, E_n$  are mutually independent and  $X$  is the number of these events that occur, then:

$$Pr\{E_1 \cup \dots \cup E_n\} \geq 1 - e^{-Ex(X)}$$

- **Chernoff Bound**: Let  $E_1, \dots, E_n$  be a collection of mutually independent events, and let  $X$  be the number of these events that occur. Then:

$$Pr\{X \geq c Ex(X)\} \leq e^{-(c \ln c - c + 1) Ex(X)} \quad \text{when } c > 1$$