

The following content is provided under a Creative Commons license. Your support will help MIT OpenCourseWare continue to offer high-quality educational resources for free. To make a donation or to view additional materials from hundreds of MIT courses, visit MIT OpenCourseWare at ocw.mit.edu.

WILLIAM GREEN: All right, so I know some of you have succeeded to do the homework and some of you, I think, have not. Is this correct?

AUDIENCE: Yeah.

WILLIAM GREEN: OK. So I was wondering if someone who has succeeded to do their homework might comment on how small a mesh do you need to converge.

AUDIENCE: [INAUDIBLE]

WILLIAM GREEN: It's about l ? L ? OK, so you need something on the order of l to converge. Is that correct? So if you're trying to do the problem using mesh much bigger than l , you should probably try a tighter mesh. Yes?

AUDIENCE: [INAUDIBLE]

WILLIAM GREEN: All right. Yes?

AUDIENCE: [INAUDIBLE]

WILLIAM GREEN: Yes. Yes. All right. And has anyone managed to get the [INAUDIBLE] solution to actually be consistent with the [INAUDIBLE] solution?

AUDIENCE: Something like 3% or 4% or so.

WILLIAM GREEN: 3% or 4%, OK. And I assume that the [INAUDIBLE] is also using a mesh of similar size? Hard to tell?

AUDIENCE: I used like a triangular system--

WILLIAM GREEN: Yeah, yeah, but I mean, it's really, really tiny ones at the bottom? If you want me to blow it up, I can just take a look and see to be sure. All right, and is backslash able to handle a million by million matrix?

AUDIENCE: Like 10 seconds with [INAUDIBLE].

WILLIAM GREEN: [INAUDIBLE] OK. So you need to-- so do the sparse allocation. And MATLAB is so smart that it just can handle it with a million by million, which is pretty amazing, actually. That's a pretty big matrix. All right, sorry, this is too loud.

All right, so last time, we were doing some elementary things about probability. Actually, any more questions about the homework problem before we get started?

AUDIENCE: What's the answer?

WILLIAM GREEN: What's the answer? You could ask your classmates. Any other questions? All right. So I had you confused a little bit with this formula probability of either A or B. So I asked what the probability of-- I flipped two coins-- that one of them would be a head. And I could see a lot of consternation.

The general formula for this is it's the probability of A plus the probability of B minus the probability of A and B. It can't just be the two of them added together because if you have 50% chance for head of the penny and the dime is 50% chance, this would add up to 100% chance that you'll get a head, but you know sometimes it's true. So this is the formula.

And then the probability of A and B is often written in terms of the conditional probabilities, the probability of A times the probability that B would happen given that A already happened, which is also equal to the other way around. And this has to be read carefully. It means B already happened, and then you want to know the probability of A given that B already happened. So it's sort of like this way-- I don't know-- the way I think about it. Like, this happened first. and now I'm checking the probability that that's going to happen.

Now, a nice little example of this is given in [INAUDIBLE] textbook. And I think it's nice enough that it's worthwhile to spend a few minutes talking about it. So he was-- [INAUDIBLE] who wrote the textbook, was not actually a numerical guy. He was a polymer chemist. And so he gave a nice polymer example.

So if you have a polymer and the monomers have some big molecule, and at one side, they have a sort of acceptor group, and the other side, some kind of donor group-- we'll call it D, I guess. And these are the monomers. And so they can link together. The donor can react to the acceptor. So you can end up with things like this and so on.

So this is the monomer. This is the dimer. Then you could keep on [INAUDIBLE] like this. And many, many, many of the materials you use every day, the fabrics in the seats that you're sitting on, the backs of the seats, your clothing, the binder holding the chalk together, all this stuff is made from polymers like this. So this is a pretty important, actually, practical problem.

And so you start with the monomers, and they react where you have A reacting plus D, over and over again. And we want to understand the statistics of what chain lengths are going to make, maybe what weight percent or what would the average microweight be, something like that would be the kind of things we care about.

So a way to think about it is if I've reacted this to some extent and I just grab a random polymer chain, any molecule in there, and I look and find, let's say, the unreacted D end-- so any oligomer is going to have one unreacted D end. You can see no matter how long I make it, there will still be one unreacted D end. And I'm neglecting the possibility this might circle around and make a loop. So assuming no loops, then any molecule I grab is going to have one unreacted D end.

So I grab a molecule. I start at the unreacted D end, and I look at the A that's next to it. And I say, is that A reacted or not? So if it's a monomer, I grab the D. I look over here. The A is unreacted. So the probability that it's a monomer is going to be equal sort of like $1 - P$ where P is the probability that As react. So it didn't react, so just like that.

This one, the one next to it has reacted. So this is just going to be the probability of a dimer is the probability that my nearest neighbor reacted and next neighbor is unreacted, right? Is that OK?

So I can write this way. I could say, what's the probability that my nearest reacted times a conditional probability, next unreacted if nearest is reacted? So far, so good? You guys are OK with this?

So I grabbed a chain. I'm trying to see if it's a dimer. I'm going to calculate the probability that this next acceptor group has been reacted to a donor group. If it has reacted, then I'm going to check the next one after that. So this is the nearest neighbor. This is the next nearest neighbor. And I want this to be unreacted. If that's both true, then I have a dimer. If either one of those is false, [INAUDIBLE]. Is that OK?

So now I need to have a probability. So what's the probability that the nearest one is reacted? There's some probability that things have reacted. So this is going to be my P , probability that things reacted. And I wanted this to be unreacted.

Now, there's a question. Are these correlated or not? Now, in reality, everything's correlated to everything. So probably, they're correlated. But if we're trying to make a model and think about it, the fact that this thing reacted at this side doesn't really affect this side if this is a big enough [INAUDIBLE] So to a good approximation, this is independent of whether or not it's reacted or not. So this is still going to have the ordinary probability of being unreacted, which would be $1 - P$.

So I could write down that the probability of being a monomer is equal to $1 - P$. The probability of being a dimer is equal to P times $1 - P$. What's the probability of being a trimer? P^2 times $1 - P$. And in general, the probability of being an n -mer is equal to P^n times $1 - P$. So now you guys are statistical polymer chemists.

So this derivation was derived by a guy named Flory. He got the Nobel Prize. He's a pretty important guy. If you want to learn a lot about him, I think both Professor Cohen and Professor Rutledge teach classes that are basically, learn what Mr. Flory figured out. Well, maybe that's a little bit too strong, but pretty much. There's another guy named [INAUDIBLE] that did a bit too, so [INAUDIBLE] and Flory. Basically everything about polymers worked out by at these guys. And all they did was just probability theory, so it was a piece of cake.

And so this is the probability that you have an n -mer. So now we can compute things like, what is the expectation value of the chain length? How many guys link together? And that's defined to be the sum of n times the probability of n . So that, in this case, is going to be sum of n times P^n times $1 - P$.

Now, a lot of these kinds of simple series summations, there's formulas for it. And maybe in high school, you guys might have studied series. I don't know if you remember. And so you can look up. And some of these have analytical formulas that are really simple. But you can just leave it this way too, because you get a value numerically with MATLAB, no trouble.

You can also figure out what is the concentration of oligomers with n units in them. And so that's going to be equal to the total concentration of polymers times the probability that it has n . So this one, we just worked out.

The total concentration, a way to figure that out is to think about there's one monomer or one monomer-- I'll call this a polymer too. This is a polymer with one unit. There's one polymer molecule per unreacted end, unreacted D end. So it's really, I want to know how many are unreacted.

So that's going to be $1 - P$ times the amount of monomer I had to start with. It could be A or D. It doesn't matter. It's like, how many of them-- I started with a certain amount of free ends. What fraction of them have reacted based on $1 - P$. Yeah, it's $1 - P$.

So as P goes-- well, yeah, it goes backwards. Yeah, as P goes to infinity-- I think that's right. Yeah, when P is-- well, I'm totally confused here now. $1 - P$ sound right? Maybe I did the reasoning backwards. This is definitely the right formula. I'm just confusing myself with my language.

This is a, at least for me, endemic problem with probability is you could say things very glibly. You've got to think of exactly what you mean. So the concentration of unreacted ends, so initially, this was equal to A. It was all unreacted ends. And as the process proceeds, as P increases, then at the end, it's going to be very small. So this is right.

And the concentration of unreacted ends is equal to the total concentration of polymers, the number of polymers [INAUDIBLE]. So it's this times P^{n-1} times [INAUDIBLE]. All right, and this is called the Flory redistribution. And that gives the concentrations of all your oligomers after you do a polymerization if they're all uncorrelated and you don't form any loops.

It's often very important to know the width of the distribution. If you make a polymer, you want to make things have as monodisperse as possible. It's because you'd really like to make this pure chemical.

There's some polymer chain length which is optimal for your purpose. You want to try to make sure that the average value, average value, this is going to be equal to the value you want. So you want to keep running P up until you reach the point where the average chain length is the chain length that's optimal for your application.

If you make the polymer too long, then it's going to be hard to dissolve it. It's going to be hard to handle and it's can be solid. If you make it too short, then it may not have the mechanical properties you need for the polymer to have. So there's some optimal choice. So you typically

run the conversion until P reaches a number so that this is your optimal value, but then you care about what's the dispersion about that optimal value.

And particularly, the unreacted monomers that are left might be a problem because they might leach out over time because they might still be liquids, or even gases that come out. So this famous problem, people made baby bottles and they have some leftover small molecules in the baby bottles. And then they can leach out into the milk, and the mothers don't appreciate that. So there's a lot of real practical problems about how to do this.

So anyway, you'd be interested in the width of the distribution. So we define what's called the variance. And the variance of n is written this way. And it's just defined to be the expectation value of n squared minus the expectation value of n squared. These two are always different or almost always different, so it's not 0.

So this is equal to the summation of n squared times the probability of n minus-- all right? And a lot of times in the polymer field, what they'll take is they'll take the square root of this and they'll compare σ_n divided by expectation value of n . This is a dimensionless number because σ_n will have the dimensions.

σ_n squared is dimensions n squared. This is dimensions of n , so it's a dimensionless number. And that's-- I think they call it dispersity of polymer, something like that.

Now, notice that when we use these [INAUDIBLE], when we wrote it this way, it's implicitly that these things are divided by the summation of the probability of n . But because these probabilities sum to 1, I can just leave it out. But sometimes, it may be difficult for you to figure out exactly what the probabilities are and you'll need a scaling factor to force this thing to be equal to 1. So sometimes, people leave these in the denominator.

There's another thing you might care about, which would be like, what's the weight percent of P_n ? So what fraction of the weight of the polymer is my particular oligomer, P_n ? [INAUDIBLE] sorry, some special one, P_m . And I want to know its weight percent.

So that's going to be equal to the weight of P_m in the mix divided by the total weight. So that's equal to the weight of m times the probability of m divided by the total weight, which is going to be the weight of all these guys, times the probability of each of them.

And you can see this is different. This is not the same as-- not equal to, right? It's not the same thing. So just watch out when you do this. And in fact, in the polymer world, they always

have to say, I did weight average. I did number average, because they're different. Is this OK?
Yeah?

So I would-- my general confidence, at least for me, if I skip steps, I always get it wrong when I do probability. So don't skip steps. Do it one by one by one, what you really mean. Then you'll be OK.

All right, now, this is a cute little example. It's discrete variables. It's easy to count everything. Very often, we care about probability distributions of continuous variables. And we have to do those probability density functions that I talked about last time which have units in them.

And so as we mentioned last time, if you want to know the probability that a continuous variable, that x is a member of this interval, the probability this is true is equal to $P(x)$ times dx . And so this quantity has units of 1 over x , whatever the units of x are. And then you have to multiply it by x in order to get the units to be dimensionless, which is what the probability is.

And this is like obvious things, like $\int P(x) dx$. [INAUDIBLE] value as possible of x has got to be equal to 1. It's a probability, which is the same as saying that probability of x is some value anywhere is 1. So there's some [INAUDIBLE] you measure.

And you can also have a probability that x is less than or equal to x' . And that's the integral from negative infinity to x' of $P(x) dx$. And the mean is just the integral of $x P(x)$. And you can compute the x squared. The average of x and x squared, same thing. You average of anything.

You can put these together. You can get σ^2 is equal to $\int x^2 P(x) dx$ minus the average squared. So that's the variance of x .

You can also do this with any function. So you can say that the average value of a function is equal to the integral of $f(x) P(x) dx$. This is an average value of a function of a random variable described by probability density function with $P(x)$. And then you can get things like $\int f^2(x) P(x) dx$ is equal to the integral of $f(x)$, quantity squared, $P(x)$ minus-- all right?
Everything's OK? Yeah.

All right, so a lot of times, people are going to say, we do sampling from $P(x)$. So sampling from $P(x)$ means that we have some probability distribution function, $P(x)$ of x , and we want to have

one value of x that we draw from that probability distribution. When we say it that way, we mean that we're more likely to find x 's where P_x has a high value and we're less likely to draw an x value that P_x has a low value. So that's what's sampling from.

Now, you can do that mathematically using random number generators in MATLAB, for example, and we'll do that sometimes. But you do it all the time when you do experiments. So the experiment has some probability density function that you're going to observe something, you're going to measure something. And you don't know what that distribution is, but every time you make a measurement, you're sampling from that distribution.

So that's the key conceptual idea is that there is a P_x of x out there for our measurement. So you're trying to measure how tall I am. Every time you measure it, you're drawing from a distribution of experimental measurements of Professor Green's height. And there is some P_x of x that exists even though you don't know what it is.

And each time you make the measurement, you're drawing numbers from that distribution. And if you draw a lot of them, then you can do an average. And it should be an average that's close to this. If you drew an infinite number of values, then you're sampling this. You can make a histogram plot of the heights you measure of me, and it should have some shape that's similar to P_x of x . Does that make sense? All right.

So actually, everyday you're drawing from probability distributions. You just didn't know it. It's like [INAUDIBLE] street. The probability the bus is going to hit me or not and the bus driver is going to stop, I think there's a high probability, but I'm always a little worried, actually. Good. I'm drawing from-- it's a particular instance of that probability distribution about whether the bus driver's really going to stop or not. And if I sample enough times, I might be dead. But anyway, all right.

Often we have multiple variables. So you can write down-- you can define \hat{P}_x . So now I have multiple x 's. It's like more than one variable of x . And I wanted the probability density function of them. I'm going to measure this and this and this and this, all right?

And this is equal to the probability that x_1 is a member of the set, x_1 , x_1 plus dx_1 , and x_2 is a member of x_2 , x_2 plus dx_2 , and that. That's what probability density function means with multiple variables.

So this is very common for us because we often measure and experiment more than one

thing, right? So you measure the flow rate and the temperature. You measure the yield and the absorption at some wavelength that corresponds to an impurity.

Usually when you experiment, you often measure multiple things. And so you're sampling from multiple observable simultaneously. And implicitly, you're sampling from some complicated PDF like this even though you don't know the shape of the PDF usually to start with.

And so then when you have this multiple variable case, you can define a thing called the covariance matrix where the elements of the matrix C_{ij} are equal to $\langle x_i x_j \rangle$, the mean of that product, minus $\langle x_i \rangle \langle x_j \rangle$. And so you can see that, for example, σ_i^2 is equal to C_{ii} , but the diagonal elements are just the variances. But now we have the covariances because we measured, let's say, two things.

All right, so suppose we do n measurements and we compute the average of our repeats. So we'd just repeat the same measurements over and over. So suppose you measure my height and my weight. Every time I go to the medical clinic, they always measure my height, my weight, my blood pressure. You've got three numbers.

And I could go back in there 47 times, and they'll do it 47 times. And if a different technician measured it using different [INAUDIBLE] and a different scale, I might get a different number. Sometimes, I forget to take my shoes off so I'm a little bit taller than I would have been. So the numbers go up and down. They fluctuate, right? You'd expect that, right? If you looked at my medical chart, it's not the same number every time.

But you'd think, if everything's right in the world, that I'm an old guy. I've been going to the medical clinic for a long time. If I look at my chart and average all those numbers, it should be somewhere close to the true value of those numbers. So I should have that the average values experimentally, which I just define to be the averages-- this is the number of experiments.

OK, so I can have these averages. And I would expect that as n goes to infinity, I hope that my experimental values go to the same value of x that I would have gotten from the true probability distribution function. If I knew what $P(x)$ of x is and I evaluated the integral and I got x , I think it should be the same as the experiment as long as I did enough repeats. So this is almost like an article of faith here, yeah? It's what you'd expect.

Now, the interesting thing about this-- I mean, probably you've done this a lot. You probably

did experiments and you've averaged some things before, right? If everybody in the class tried to measure how tall I was, you guys all wouldn't get the same number. But you'd think that if you took the average of the whole classroom, it might be pretty close to my true height, right?

So the key idea here is that the sigma squared of the x measurement experimental, which we define to be this-- maybe we should do this one at a time. [INAUDIBLE]. Then I can have a vector of these guys for all the different measurements. So there's some error in my height. There's some error in my weight. There's some different error in my blood pressure measurement, but each should have their own variances. I can have the covariances.

OK, so these are all the experimental quantities. You guys maybe even computed all these before in your life. And we expect that this should go like this as n goes to infinity. Now what's going to happen to these guys as n goes to infinity? That's the really important question.

So there's an amazing theory called the central limit theorem of statistics. And what this theorem says, that as n gets large and if trials are uncorrelated and the x 's aren't correlated, which is the same as saying that C_{ij} is equal to 0 off the diagonal, then the probability of making the measurement x is proportional to the Gaussian, the bell curve. All right?

So this is only true as n gets very large. It doesn't specify exactly how large has to be, but it's true for any P_x , any distribution function, probability distribution function. So everything becomes a bell curve if you look at the averages. And σ_i^2 in that limit goes to 1 over $n \sigma_{xi}^2$ experimental.

And this is really important. So what this says is that the width of this Gaussian distribution gets narrower and narrower as you increase the number of repeated experiments or increase the number of samples. So this is really saying that the uncertainty in the mean is scaling as 1 over root n where n is the number of samples or number of experiments that's repeated.

Now, σ , the variance, is not like that at all. So this quantity, actually as you increase n , just goes to a constant. It goes to whatever the real variance is, which if you're measuring me, it might be how good your ruler or something. It'll tell you roughly what the real variance is.

And that number does not go to 0 as the number of repeats happens. I mean, I could get the whole student body to measure how tall I am at MIT, and they're still not going to have 0 variance. It's going to still be some variance, right? So this quantity stays constant as n increases or goes to a constant value once it sort of stabilizes. You have to have enough

samples. But this quantity, the uncertainty in the mean value, gets smaller and smaller and smaller as the square of n .

Now, this is only true in the limit as n is large. Now, this is a huge problem because experimentalists are lazy, and you don't want to do that many measurements. And it's hard to do a measurement.

So for example, the Higgs boson was discovered, what, a year and a half ago, two years ago? And I think altogether, they had like nine observations or something when they reported it, OK? So nine is not infinity. And so they don't have infinitely small error bars on that measurement. And in fact, who knows if it really looks like a Gaussian distribution from such a small sample, but they still reported 90% confidence interval using the Gaussian distribution formula to figure out confidence intervals.

So everybody does this. If n is big, it should be right. And you could prove mathematically it's right, but the formula doesn't really tell you how big is big. So this is like a general problem. And it leads to us oftentimes misestimating how accurate our results are because we're going to use formulas that are based on-- assuming that we've averaged enough repeats that we're in this limit where we can use the Gaussian formulas and get this nice limit formula. But in fact, we haven't really reach that because we haven't done enough repeats.

So anyway, this is just the way life is. That's the way life is. And I think there's even discussions in statistics journals and stuff about how to make corrections and use slightly better forms that get the fact that your distribution of the mean doesn't narrow down to a beautiful Gaussian so fast. It has some stuff in the tails. People talk about that, like low probability events out in the tails of distributions, stuff like that.

So that's a big field of statistics. I don't know too much about it, but it's like-- I mean, it's very practical because-- now unfortunately, oftentimes in chemical engineering, we make so few repeats that we have no chance to figure out what the tails are doing, maybe [INAUDIBLE] our tails. And so this is a big problem for trying to make sure you really have things right.

So I would say in general, this is an optimistic estimate of what the uncertainty in the mean is. Uncertainties are usually bigger. So you shouldn't be surprised if your data doesn't match a model brilliantly well as predicted by this formula. Now, if it's off by some orders of magnitude, you might be a little alarmed. And that might be the normal situation, too. But anyway, if it's just off by a little bit, I wouldn't sweat it because you probably haven't done enough repeats to

be entitled to such a beautiful result as this.

We can write a similar-- actually, so here I assumed that the x 's are uncorrelated. That's almost never true. If you actually numerically evaluate the C 's, usually they have off-diagonal elements. For example, my weight and my blood pressure are probably correlated. And so you wouldn't expect them to be totally uncorrelated.

And so there's another formula like this. It's given in the notes by Joe Scott that includes the covariance. And you just get a different form of what you'd expect, OK? And the covariance should also converge roughly as 1 over n if you have enough samples. So you should eventually get some covariance.

You can write very similar formulas like this for functions. So if I have a function f of x and that's really what I care about-- remember, I said that I have the average value of f is equal to f of x P_x of x dx . And I could make this vectors if I want.

And I could repeat my function, and I'd get some number. And I could repeat the variance. I have a σ_f . And this is something I like to do a lot of times. Then if we do experimental Δf -- so we don't know what the probability distribution function is usually, or often. So we'll try to evaluate this experimentally.

This is going to be 1 over n , the values f of x little n , the n -th trial. And we could write a similar thing for σ_f , which I just did right there. You can do the same thing. Just make these experimental values now. The σ_f squared experimental should go to 1 over n times the variance. And this was the σ in the mean of f , 1 over n times the variance of the σ .

All right, so this is the same beautiful thing, that the uncertainty in the mean value of f narrows with the number of trials. So you have some original variance that you computed here, either experimentally or from the PDF. Experimentally is fine. And then now you want to know the uncertainty in the mean value, and that drops down with the number of trials in the number of things you average.

So this all leads in two directions. What we're going to talk about first is about comparing models versus experiments where we're sampling by doing the experiment. So that's one really important direction, maybe the most important one. But it also suggests ways you could do numerical integration.

So if I wanted to evaluate an integral that looks like this, $\int f(x) P(x) dx$, and if I had some way to sample from $P(x)$, then one way to evaluate this numerical integral would be to-- sorry, I made this vector [INAUDIBLE] a lot of species there, a lot of directions. If I want to evaluate this multiple integral-- it's a lot of integrals for every dimension of x -- that would be very hard to do, right? We talked about in [INAUDIBLE], if you get more than about three or four of these integral signs, usually you're in big trouble to evaluate the integral.

But you can do it by what's called Monte Carlo sampling where you sample from $P(x)$ and just evaluate the value of f at some particular x points you pull as a sample and just repeat their average. And the average of those things should converge, according to this formula, as you increase the number of samples. And so that's the whole principle of Monte Carlo methods, and we'll come back to that a little bit later.

And you can apply that to a lot of problems. Basically, any problem you have in numerics, you have a choice. You can use deterministic methods or stochastic methods. Deterministic methods, if you can do them, usually are the fastest and more accurate, but stochastic ones are often very easy to program and sometimes are actually the fastest way to do it.

In particular, in this kind of case, we have lots of dimensions, many, many x 's. It turns out that stochastic ones are pretty good way to do it. But we're going to talk mostly about [INAUDIBLE] data because that's going to be important to all of you in your research. So let's talk about that for a minute.

I'll just comment, there's really good notes posted on the [INAUDIBLE] website for all this material, so you should definitely read it. And the textbook has a lot of material. It's maybe not so easy to read as the notes are, but plenty to learn, for sure.

So we generally have a situation where we have an experiment. And what do we have in the experiment? We have some knobs. These are things that we can change. So we can change some valve positions. We can change how much electricity goes into our heaters. We can change the setting on our back pressure regulator. We can change the chemicals we pour into the system. So there's a lot of knobs that we control. And I'm going to call the knobs x .

And then we have parameters. And these are other things that affect the result of the experiment that we don't have control over. And I'm going to call those θ .

So for example, if I do a kinetics experiment, it depends on the rate coefficients. I have no

control of the rate coefficients. They're going to [INAUDIBLE] by God, as far as I know. So they're some numbers, but they definitely affect the result. And if the rate coefficient had a different value, I would get a different result in the kinetic experiment.

The molecular weight of sulfur, I have no control over that. That's just a parameter. But if I weigh something and it has a certain number of atoms of sulfur, it's going to be a very important parameter in determining the result.

So we have these two things. And then we're going to have some measurables, things that we can measure. Let's call them y . And in general, we think that if we set the x value and we know the theta values, we should get some measurable values. And so there's a y that the model says that's a function of the x 's and the thetas.

Now, I write this as a simple function like this. This might be really complicated. It might have partial differential equations embedded inside it. It might have all kinds of horrible stuff inside it. But you guys already know how to solve all these problems already because you've done it. You've been in this class through seven homeworks already. And so no problem, right? So if I give you something-- I give you some knobs. I give you some parameters-- you can compute it, all right?

And so then the question is-- that's what the model says. So we could predict the forward prediction of what the model should say if I knew what the parameter values were, if I knew what the knob values were. And I want to-- oftentimes what I measure, y data, which is a function of the knobs, it's implicitly a function of the parameters. I have no control of them, so I'm not going to even put them in here.

So I set the knobs I want. I get some data. I want these two to match each other. I think they should be the same thing if my model is true, yeah? So this is my model, really. But I don't think they should be exactly the same. I mean, just like when you try to measure my height, you don't get exactly the same numbers.

So these y data are not going to be exactly the same numbers as my model would say. So now I have to cope with the fact that I have deviations between the data and the model. And how am I going to handle that, all right?

And also, we have a set of these guys, typically do some repeats. So we have like several numbers for each setting in the x 's, and they don't even agree with each other because they're

all different. Every time I repeated the experiment, I got some different result-- that's my y's-- for each x.

And then I change the x a few times at different knob settings. Then I make some more measurements. And I have a whole bunch of y values that are all scattered numbers that maybe scatter around this model possibly, if I'm lucky, if the model's right.

Often, usually I also don't know if the model's correct. So that's another thing to hold in the back of your mind is like, we're going to this whole comparison assuming the model's correct. And then we might, at the end, decide, hmm, maybe the model's not really right. I may have to go make a new model.

So that's just a thing to keep in the back your mind. But we'll be optimistic to start with, and we'll assume that the model is good. And our only challenge is we just don't have the right values of the thetas, maybe, in my model.

And this is another thing, too. So the thetas are things like rate coefficients and molecular weights and viscosities and stuff that are like properties of the universe, and they're real numbers, maybe. They're also things like the length of my apparatus and stuff like that.

But I don't know those numbers to perfect precision, right? The best number I can find, if I look in the database is, you know, you could find like the speed of light to like 11 significant figures, but I don't know it to the 12th significant figure. So I don't know any of the numbers perfectly.

And a lot of numbers I don't even know at all. So like there's some rate coefficients that no one has ever measured or calculated in the history of the world. And my students have to deal with that a lot in the Green group. So a lot of these are quite uncertain. But there are some that are pretty certain. You have quite a big variance, actually, of how certainly you know the parameter values.

So one idea, a very popular idea, is to say, you know, I have this deviation between the model and the experiment. So I want to sort of do a minimization by varying, say, parameter values of some measure of the error between the model and the data. Somehow, I want to minimize that.

And I have to think about, well, what should I really minimize? And the popular thing to minimize is these guys squared and actually to weight them by some kind of sigma for each one of these guys. So this is-- we should change the notation, make this clearer.

These guys-- one model, and it's the i -th measurement that corresponds to that n -th experiment. So I think that the difference between what I measured and what the model calculated should be sort of scaled by the variance, right? So I would expect that this sum has a bunch of numbers that are sort of order of one because I expect the deviation to be approximately scaled of the variance of my measurements.

And if these deviations are much larger than the variance, then I think my model's not right and what I'm going to try to do right here is I'm going to try to adjust the thetas, the parameters, to try to force the model to agree better to my experiment. And this form looks a lot like this. Do you see this? You see I have a sum of the deviations between the experiment and a theoretical sort of thing divided by some variance?

And so this is the motivation of where this comes from, is that I want to make the probability that I would make this observation experimentally would be maximum if this quantity in the exponent is as small as possible. So I'm going to try to minimize that quantity, and that's exactly what I'm doing over here. Is that all right? OK, so next time when we come back, I'll talk more about how we actually do it.