PROFESSOR: Our goal for today is to basically analyze this simple model to death. So we're first going to try to understand the deterministic behavior of this model gene expression, where we just get transcription of mRNA, and then translation of protein. And after we think we understand the mean behavior, the deterministic dynamics, then we will try to understand just stochastic behavior in this model.

So we're going to try to understand what's the distribution of mRNA in a cell in this simple situation. What's the distribution of protein? What's going to be the bursting behavior? Everything you can possibly think of to ask about this model, we will hopefully have asked by the end of today's class.

This simple model of gene expression, as was indicated in the review, is perhaps a reasonable description of gene expression in bacteria, when the gene is in some active state. So there's no repressor, for example, bound. Although maybe even in the presence of a repressor, if it's binding and unbinding, maybe you still end up getting some sort of renormalization that looks like this. But this is first order, a reasonable description of gene expression in bacteria.

And it's the model that was basically used in the Sunney Xie paper that we talked about on Tuesday. And hopefully this model will allow us to think a little bit more deeply about the data that they obtained in that paper.

As always, we want to start by understanding the basic aspects of the model. So what we're going to do, is we're going to go through a series of questions of increasing difficulty. And in some of them, we are indeed, the answers will end up being something divided by something. In which case you take advantage of your cards, and illustrate that by putting something on top, something below.

But just first in this model, what is the unit of time? So if I say t is equal to 1, or delta t is equal to 1, what am I referring to? So we're not use the cards. But in particular, the question is, is delta t equal to 1, is that a cell cycle necessarily? Yes or no, ready, three, two, one.

Well I guess now maybe I've complicated things by-- well, this was really going to be relevant for the later ones. All right. Now I've totally confused you. But can somebody offer why it may or may not be-- how do we think about the unit of time in this model?

AUDIENCE: Usually the lifetime of one of the species [INAUDIBLE].

PROFESSOR: Right. OK so indeed, what we often do in these non-dimensionalized models is we set something equal to 1. Have we set anything equal to 1 here? No. So in principle, we've said there's some degradation rate of the mRNA, some degradation rate of the protein. And in general, those will be given in some units involving seconds or minutes or hours. So in general, so at this stage, we have not yet-- we have not actually gotten to this sort of non-dimensionalized version of any model.

So in this case this is going to be something like a seconds, or minutes, or hours, whatever units we use for those degradation rates. So we have not done anything where it's the cell generation time, or the protein lifetime, mRNA lifetime, or anything like that. Here everybody happy with this statement so far?

So we'll go ahead and vote here. So we're going to do some A, B, C, D's. And you can always combine anything you want. So we'll go ahead and say this is the synthesis rate of the mRNA. This is the degradation rate for the mRNA, the synthesis rate for the protein, the degradation rate for the protein. And if you're just confused, you can just do this. But in general, for any of the questions we're going to do, you can do some combination of these guys by putting things in numerator and denominator. Yes?

AUDIENCE: Calculate the population of the cells that were hidden?

PROFESSOR: Yes. Question is, if you just look at the cell population, and you find it's growing

exponentially, the question is what is going to be that rate of exponential growth. Have I done something wrong already?

AUDIENCE: [INAUDIBLE]?

PROFESSOR: OK. But I am going to say that now for this we're going to assume that the protein is stable. So it's not actually degraded. This is to remind you of what we read about in chapter one, maybe of Uri's book, maybe chapter two. I'll give you 10 seconds to think about this.

Do you need more time? All right, Ready, three, two, one. OK. We got a bunch of C's and a bunch of D's and some E's. All right. So the E's are going to argue with me, presumably rather than a neighbor. OK. I think that there are enough people that are disagreeing on this to maybe go ahead, and turn. You should be able to find somebody that disagrees with you. The distribution was a bit patchy, unfortunately. Did you guys-- you guys are worried that you're not going to be able to find somebody. OK. Fine, fine. Yeah?

AUDIENCE: So if the protein is stable, ah, so the mRNA may not be stable?

PROFESSOR: The mRNA may not be stable.

AUDIENCE: Ah, OK. That makes sense.

PROFESSOR: And in general which one typically has a longer lifetime?

AUDIENCE: Proteins.

PROFESSOR: Proteins typically have a longer lifetime. Right. So mRNA are actively degraded, typically. They're also just kind of less stable intrinsically. But what we're going to assume for now is that we're working with stable proteins. In which case the growth rate of the population will just be this effective degradation rate of the protein.

So in this model, even if we say there's no active degradation of the protein, still there's going to be some effective degradation that's due to dilution. So we can say effective, if you like. So the rate of exponential growth of the population will be equal

to this effective degradation rate for the protein, if it's stable.

AUDIENCE: So you're talking about the population of the protein?

PROFESSOR: No. The growth rate of the cell population. So this is if we go in there, and you go into your spectrophotometer. And you measure population-- numbers in function of time is growing exponentially. It'll grow exponentially with this rate. Because this is what's causing the dilution.

In some ways if you stop making the protein, and you double the number of cells, and that means the concentration of the protein in each cell has to go down by a factor of two. So that's the statement. Are there any questions about why I'm making this argument? Yes?

AUDIENCE: What was the relevance of the protein being stable?

PROFESSOR: All right. So the relevance of the protein being stable, because this is in general, this delta, this is the effective rate. This is going to be equal to the growth rate of the population. So you might call it gamma growth plus the actual degradation. I don't want to use the same, but I'll just say plus the degradation rate. And this is a true physical degradation, true degradation rate of the protein.

So if it's stable, then we say that this thing is zero. So when we say stable protein, it means there's no degradation of the protein. So this physical degradation rate is zero. And then the effective degradation rate of the protein is just equal to the growth rate of the population.

AUDIENCE: OK, so no degradation means stable, basically?

PROFESSOR: Yes, sorry, yeah. Any other questions about what I mean by this?

So now what we want to do is ask a few other quantities about this model. So for example, what will be the number of mRNA per cell?

And this is always going to be the mean. I'll give you 20 seconds. In this model what is the mean number of mRNA per cell? All right. Ready? Three, two, one. And we

have let's say a majority of the group is saying it's A over B, which corresponds to the synthesis rate of the mRNA divided by the degradation.

Some people are this one? Yes, so this is indeed, synthesis rate divided by the degradation rate. Now this is saying that what happens later doesn't really matter, for the mean mRNA number. Because it's just that it's going to be made at some rate. Its lifetime is given by 1 over delta m. Now this thing, of course, is again as always, the effective degradation rate. So it's the sum of the sort of physical degradation rate, plus this dilution due to growth.

But in general, the true degradation, the physical degradation is much faster than the cell division rate. So this is very close to actually just the physical degradation rate. But in any case, it's just delta m, regardless. Are there any questions about why this is the way it is? Yes?

AUDIENCE: Does it matter whether it's only physical? Because wouldn't it be the same if it were--

PROFESSOR: It doesn't matter that it's only-- exactly. That's what I was trying to say. So the way that this is written, it doesn't matter whether the-- this is the answer regardless of whether the physical degradation rate is much larger than the growth rate or not. Yeah.

All right. What is this protein molecules per mRNA? How many protein molecules are made from each mRNA? Protein produced-- Do you need more time? Remember. This is again, the mean number of proteins produced from a single mRNA or each mRNA.

Let's go ahead and vote, so I can see where we are. Ready? Three, two, one. OK. So we have, I'd say, so at least a majority are saying it's going to be C over B Now. All right. So this is interesting. So this is saying that really what's happening is that there's a competition once you make an mRNA that the proteins are going to be getting fired off at some rate. But eventually it's going to be degraded.

It's a competition between those two rates that determines basically how many

proteins, how many times do you fire off a protein before you get degraded. Any questions about that logic?

AUDIENCE: Can you please just repeat that one more time?

PROFESSOR: Sure. Right, so what we're assuming is that OK, an mRNA is produced. And that's already happened. So it doesn't matter what Sm is anymore. So now we have an mRNA. Eventually this mRNA will be degraded. But before that happens, we want to know, basically how many proteins do we expect to be made.

Now if Sp and delta m are the same that means you kind of expect one protein to be made on average, before it's degraded. Or if Sp we're twice delta m, then you would get two proteins made before it was degraded. Now this is a mean statement. We're about to start thinking-- in 10 minutes, we'll think about this distribution. And so we have to be careful. But in terms of mean behavior, this thing is true.

AUDIENCE: So is this different than it has been for the number of proteins per mRNA in the cell?

PROFESSOR: Is this different from the number of proteins in the cell?

AUDIENCE: Number of proteins per mRNA in the cell. Because then you will have to do the protein concentration over mRNA concentration.

PROFESSOR: OK. Right. So this is not the same thing as asking about the ratio of the number of proteins. And we can calculate that as well. Yeah these are different. This is the number of protein molecules produced from each mRNA. So this is just talking about production. Because indeed, the degradation rates are going to be different. So then we can see what that ends up being.

Any other questions about why this one is what it is? How about the number of mRNA produced per cell cycle? And for now we're going to ignore factors of log two. Do you need more time? So another 10 seconds.

AUDIENCE: Produced but not degraded?

PROFESSOR: Produced, yes. We're just talking about production. Because we've already

calculated a number of mRNA in the cell. But now we want to know the mean number produced. For example, this is the same as the mean number of protein bursts observed in Sunney Xie's paper. But this is just the number of mRNA produced per cell cycle.

All right. Let's see where we are. Ready? Three, two, one. All right. So we've got lots of A's over D's. That's sounds nice. So this is going to be some synthesis rate. But now the relevant thing is this delta p. Because that's the cell division rate. So it's barring issues of log two, it's approximately the synthesis rate of the mRNA divided by delta p. Because this is the growth rate of population.

Cell generation time is log two off of that. Are there any questions about that statement? All right, so this is the mean number. Now from the paper, we know how this thing is distributed. We should probably-- we're going to use a bunch of distributions over the next couple.

So we can-- we like exponential distributions. We like geometric distributions. We like Poisson. We like Gaussian. And we like gamma. These are various probability distributions. The question is, how is it that now, not the mean, but how is the number of mRNA produced per cell cycle distributed. Ready? Three, two, one.

All right. We've got some-- this side of the rooms a little bit slower, maybe. But that's OK. So maybe some people are not confident of this statement. OK. So this one ends up being Poisson. So this is indeed how the number of-- this is number mRNA per cycle.

Now this is-- so Poisson, in general, that's what you get if there's some probability per unit time that something's going to happen, and you want to know how many of them happen in some finite time period. That's basically the definition of a Poisson. And this is, if you recall, this is what we talked about on Tuesday. The probability observe n, it's given by this mean number. So if lambda is the mean, then we get lambda to the n, over n factorial, e to the minus lambda.

If you go ahead and calculate the mean of this, you indeed get lambda. So lambda

is equal to the mean, which in this case was around, well in the case of Sunney's paper, does anybody remember what that roughly was? It was around one.

Now what about this other one? So we also have another mRNA problem, which is that we calculated the mean number of mRNA per cell. If you look at a cell, the mean number is this. But what's the probability distribution of the number of mRNA per cell? So we probably-- I'm trying think it-- you probably don't yet know this answer.

This ends up also being Poisson. We're going to calculate this in a bit. But this is very confusing somehow. That both this thing and this thing, are Poisson. But they're not the same Poisson, in the sense they have different lambdas. Which one is going to be larger? This one or this one? The bottom one, right? And that's because delta m is much larger than delta p, typically.

So indeed, if you ask, in Sunney's paper, for example, there was just over one mRNA produced per cell cycle. But the mean number of mRNA might have been 1/30th of that. Because the degradation rate was just 1 and 1/2 minutes. What that's saying is that in a typical situation you would not see an mRNA in a cell in that condition.

We're going to calculate this in a moment. So don't worry if you don't see why it's a Poisson. But don't get confused. There are two different distributions that arise from the mRNA in the cell or in the cell cycle. And they're different Poissons. And I think that-- I mean I'm sure that in some deep sense there's a reason that they're the same. But it's somehow not immediately obvious.

So there was another one that we might have wanted to do, which is the mean number of proteins in each cell. Now this one is a bit harder. And this one is going to take full advantage of the cards that you have in front of you. So be prepared. I'm going to give you 30 seconds. Because this one you might-- well you might need a little bit more time.

AUDIENCE: This is hard.

PROFESSOR: Yeah. Although I think that it's useful to see that it can be a bit tricky. Because this really is the simplest possible model. We're going to talk about some models that get to be horribly complicated. And so it's useful to just make sure you can nail down the intuition on this model.

All right. Do you need more time? It's OK if this is escaping you at this moment. Why don't we go and see where we are? Ready? Three, two, one. All right. So you know all the naysayers on the cards, now that you've done this, you feel like it's an amazing system.

So it's AC over DB. So the two, the product of the synthesis rates divided by the product of degradation rates. So what we have is the synthesis rate for the mRNA divided by the degradation rate for the mRNA. Synthesis rate for the proteins divided by the degradation rate for the proteins.

Can somebody give us a verbal explanation for why this might have been, or why this is? Yes?

AUDIENCE: It's the same reasoning as the number of mRNA per cell. But instead of just a basel-- like a synthesis rate doesn't depend on the concentration. You're just multiplying the synthesis rate by the number of mRNA.

PROFESSOR: Yeah, that's great. OK. So what you're saying is that this thing here was indeed, we calculate that was the mean number of mRNA in the cell. If you just start with something, and you have a production and degradation rate. OK. Well that means that if you had one mRNA, then indeed that's what the concentration would be, is this Sp divided by delta p.

But now we-- well we multiply that by the number of mRNA, and then we are set. All right. Now another question. We have a distribution, or a mean protein-- wait sorry, mean number. We have the mean number of protein produced from each mRNA, is something.

And the question is, is this the most likely number of proteins to observe. Is the distribution here, now this is a mean, but now we want to start thinking about the

probabilistic stochastic elements. Is this the most likely, is it like the number of proteins observed from an mRNA?

The question is, is this most likely. By which I mean is the probability distribution peaked here. So we're going to do an A as a yes, and B is a no. Does everybody understand the question I'm trying to ask? So an mRNA is here. There's going to be some proteins made from it. This is the mean. What I want to know is, is that we should somehow expect? In a sense, is the distribution peaked, the probability distribution peaked around this value?

And C is-- do I want to do a depends? Well you can always argue after. Do you need more time?

AUDIENCE: Is this, you're saying, is this the most likely number of proteins?

PROFESSOR: Yeah. What I'm wondering is the mode there?

AUDIENCE: Right. But only for this quantity?

PROFESSOR: Only yeah. So now we're not doing means anymore. We want to know if the probability distribution of the protein produced from each mRNA is the mode around this. Ready? Three, two, one. All right. We got a lot of no's, but some yeses.

So this is actually going to be a no. And this was because the probability distribution. The question is, what is the probability distribution for the number of proteins produced from each mRNA. It's going to be one of these. Ready? Three, two, one. All right. So we've got some difference. But I'd say that most of the group is saying it's going to be A or B. And indeed these are almost the same distributions. What's the difference between them?

AUDIENCE: One's discrete--

PROFESSOR: Right. So this guy's discrete. This guy is continuous. Right. Indeed when we're taking about the numbers, then we should get-- it's a geometric. But often we're kind of a little bit loose about these things. So it's not a disaster if you said

exponential. But the key thing is that the distribution looks something like-- so now I've certainly drawn it as an exponential. This is the probability of n as a function of n. Of course, the geometric thing it looks--

AUDIENCE: Sorry. So why should we expect that distribution?

PROFESSOR: Why should we expect the distribution? So one answer is that because that's what you read on Tuesday. But let's go ahead and-- yes, but let's go ahead and calculate it. That's useful.

The way to think about this, in some ways, there's another way to write this perhaps. Which is that imagine you have an mRNA. Now at some rate it's going to be degraded. And maybe we'll keep the degradation rate down, just for-- so there's a degradation rate. But then if you'd like, we could draw it like this. Where this is the synthesis rate for a protein. And out pops a protein.

And so the idea is that is here we're in some state where, OK, here we have an mRNA. Here's the state where we don't have an mRNA. Now this is the competition between those two rates that I was telling you about. There is some degradation rate for the mRNA. Or there's a synthesis rate where we go around this loop. If we come around this loop, we come back to the state with an mRNA. There still is an mRNA intact. Just out pops a protein.

So then what we want to do, is we want to think about what's the number of proteins that we expect, not just the mean, but the actual distribution. So it's useful in these situations to define some probability rho, which is the probability you actually, if you're here, it's the probability that you produce one protein at least. The question is, which path do you take initially.

Well that's just given by the ratios. So there's the rate that we take this circular path divided by the sum of these two other rates. And then what we can do is we can ask, well what is the probability that 0 proteins are produced, probability that we get 0. Well if we take this path initially, will we have 0 proteins?

AUDIENCE: No.

PROFESSOR: No. Right, so this probability is indeed simply equal to the probability that we do this first, which is 1 minus rho. Now what's the probably that we get 1 protein? Well, only 1? That's equal to the probability that we first take this path, and then we take this path. Well we can multiply those probabilities. Because we first take the circular path to make a protein. And then we take the degradation path.

Well what's the probability we get 2? Well that's just that we come around here once, twice, and then degrade. Now if you're not seeing a pattern here, then we're in trouble. So this says the probability of n will then just be equal to rho to the n, 1 minus rho.

And indeed, it's always useful in order to warm up your probability muscles, to check to make sure that this is a normalized probability distribution. So sum over all possible ends indeed goes to 1. And that's just because the sum over a bunch of rho to the n's is equal to 1 divided by 1 minus rho, which is the term there. And that goes to 1.

AUDIENCE: So this is making a pretty strong assumption that they're all independent?

PROFESSOR: Yep, yep. Yep. This is assuming that if you've gone around once, you return. But I've come back to the original state.

AUDIENCE: But do mRNAs like actually get caught in ribosomes--

PROFESSOR: There are a lot of things that can be true. And I would say that in biology and in life, what you do is you first write down the simplest possible model. And then you go and you make measurements. And you ask whether the simplest possible model can adequately explain the data.

And if the answer is no, then you're allowed to start thinking about other things. Because everything's is in principle true. In that mRNA, maybe it's this or that. The question is whether it's significant. And at least from the data from Sunney's group would say that in that condition, in those cells, that those things are not significant, in the sense that you still get a geometric distribution.

Of course it could also be that those other things actually are true and are significant. But then you end up with some new parameters that describe how things look as a result of all the complexity. That's also OK in the sense that I'd say that you can get a quantitative description of the process by describing it as a geometric with just a single free parameter. And they found that the mean was four, or four or five. The mean number of proteins produced from each mRNA. But they got this geometric distribution in that paper. Yes?

And I'll just mention here that the mean of this is rho divided by 1 minus rho. So what you see is that as rho goes to 1, then this thing is going to diverge. And that makes sense. because as rho goes to 1, it's saying that you essentially always synthesize another protein rather than degrading.

And before I move on, I just want to say one more thing, which is that there are many different definitions of the geometric distribution, depending upon whether the probability of rho is the probability of terminating, or the probability of going around, and also depending on whether you're asking what is the-- here we're talking about the probability distribution for the number of proteins produced. Whereas we could have talked about the probability distribution for the number of times we go around this loop before, no, no sorry. That is for the number of proteins produced.

So the other way you could have defined this is the number of times where it's-- the number of cycles that you had to go before you went here, in the sense that if you first go here, you can either call that a 0 or a 1. Do you see what I'm saying? And reasonable people can disagree. But you end up getting distributions that are just a little bit different. So watch out. If you just memorize something, you might have memorized the equation for a different definition of this distribution. Does everyone understand what I tried to say there? Maybe? Yeah?

AUDIENCE: When there's no degradation is it still a Poisson?

PROFESSOR: Ah, if there's no degradation then would this be a Poisson? I mean, this would be infinity, right?

AUDIENCE: Right, [INAUDIBLE] protein, this is done independently, like there's an mRNA. [INAUDIBLE] proteins independently.

PROFESSOR: OK. So I want to be clear. This is, p of n is, this is the probability distribution for number of proteins n, produced from a single mRNA.

Now if there's no degradation of the mRNA, then this thing is not even, I think, defined in that the number of proteins produced from that mRNA just really goes to infinity. If you wanted to ask about the probability distribution for the number of proteins produced in some unit, some period of time that would indeed be a Poisson distribution, assuming that there's no degradation. Do you understand what I'm trying to say?

AUDIENCE: So Sp is like 0?

PROFESSOR: If Sp, I'm sorry. If Sp were 0?

AUDIENCE: Yeah. [INAUDIBLE].

PROFESSOR: OK. And you're saying that what would be Poisson distributed?

AUDIENCE: The number of proteins.

PROFESSOR: Yeah, I think that actually-- no I think that-- I think that you're probably right. That as Sp goes to 0-- I'm a little bit worried that--

AUDIENCE: No, no, no. Zero-th order, sorry.

PROFESSOR: Oh.

AUDIENCE: Not 0. [INAUDIBLE].

PROFESSOR: OK. Right, so the mRNA distribution we're about to find is indeed going to be a Poisson at steady state. And so if there's some process by which the protein distribution is really just mirroring the mRNA distribution, then it will also be Poisson. Although I think you have to be careful about how you actually implement that. Because even in the absence of this geometric bursting, different things, I think, can

happen.

Because for example, if there were exactly 10 proteins produced from each mRNA, then that probability distribution is a shift. But then it's no longer actually going to be Poisson, because the mean and variance are going to scale differently if you do that.

Let's maybe do the Poisson distribution for the mRNA first. And then we can try to touch back on this. So this is a plot of kind of geometric distribution with a mean of 3-4-ish. Is everybody happy with where we are now? OK.

Now from this, what we've said so far is it obvious what the distribution of proteins will be in a cell? We can say obvious, yes. Or not obvious, no. Ready, just verbal, yes? Ready or no? All right. Ready? Three, two, one.

AUDIENCE: No.

PROFESSOR: No. Right. So we've said that the distribution of size of protein bursts from single mRNA is geometric. But that doesn't mean that that's going to be the distribution of proteins in the cell. And indeed after, we're going to find that the distribution of mRNA is going to be Poisson. But even then it's not obvious what the distribution of proteins is.

All right. So what we want to do now is we want to introduce kind of a simple version of what's known as the Master Equation. Now you guys are going to do more reading on this for the lecture on Tuesday. Where we're going to talk about the Master Equation, as well as the Fokker-Planck approximation. Maybe the Gillespie algorithm, and so forth.

But I want to start by thinking about the this notion in the simplest possible context. So what we're going to do is we're going to think about the world. So we want to know the steady state, or the equilibrium distribution of mRNA numbers in the cell, given this process. So that's great. We can-- so mRNA distribution, question mark.

Now in this case we don't care about Sp, delta p, because the only things that are

relevant are going to be these. Now what we're going to do is we're going to think about the world in which we just defined states corresponding to the different numbers of these mRNAs.

So there's a state where there's 0. We can't go to the left, less than 0, but we can go to the state where there's 1, or the state where there's 2, and so forth.

Now the description here is supposed to be the analog of this over there. So this is trying to understand the situation where the deterministic equations would be described by m dot is equal to this some synthesis rate, minus a degradation rate that's proportional to the number, so minus delta m times m.

So what you can see is that the deterministic equations are very simple. We already calculated the equilibrium. So when this thing is equal to 0, then we get that m equilibrium is just going to equal to the synthesis rate divided by the degradation rate.

If we're away from the equilibrium in this deterministic approximation, how long is it going to take us to kind of approach our equilibrium? Verbal answer, ready? Three, two, one.

[STUDENTS RESPOND]

PROFESSOR: Right. So it's going to be 1 over delta m. So this tells us the characteristic timescale to come back. So if we're-- this is the equilibrium Sm over delta m. This is m as a function of time. If we're below, we come here. If we're above, we come here. And this time is 1 over delta m. Are the any questions about?

Now this-- so I want to highlight that this is like the world's simplest dynamical equation, almost the world's simplest. Yet what we're going to find is that once we go over and we try to understand the full probability distribution of the stochastic system then it's a little bit more complicated. In particular, we end up with an infinite set of differential equations.

So in general the Master Equation format, where we're going to write differential

equations for how these probabilities change over time. Now what we've done is we've traded a single differential equation for an infinite number of differential equations. So that's a bummer. But on the other hand, it will allow us to do the full stochastic treatment. And it's also a nice, to me, the master equation is useful in kind of two ways.

One is that it's going to be a tool for us to do analytic calculations. But it's also kind of a principled way of organizing your thoughts so that you can go and do stochastic simulations, if that's what you want to do. So it's also just kind of like a weigh station to kind of help you set up your simulation.

So what we're going to do is we're going to ask about the general way that this thing is going to move between different states. In particular, we are going to have some general state in here. Mn, which can go forward or back, Mn plus 1.

Now what we want to do is think about how those probabilities are going to change over time. So we typically have fn. So this is often written as an fn and fn minus 1. And then this is a g. I want to make sure I get the n's and n minus ones correct here.

Typically we write gm plus 1, gn. So these are telling us about the rates of being in this state, with say, n mRNAs, as compared to going here. We're going here. So then what we can do is we can write the change in the probability of mn with respect to time.

Well there are just a few different ways that the probability can change. So we can leave the state in two different ways. fn, gn. So the way that we lose the probability is that we have fn plus gn times the probability that we are in mn. That's an n there.

And then there are going to be two ways that we gain probability. We can gain probably from the mn minus 1. So this is fn minus 1, plus we can get probability from the upper state. That's a gn plus 1 mn plus 1. So this is just saying that the change in the probability of being in this state is going to be given by the probability that we leave the state. Sorry. The probability that we enter the state, minus the

probability that we're leaving the state, kind of the rates.

Now this is going to be true for all n, except for n equal to zero, we don't have the terms over on the left. So this is kind of for all n. So this is, in particular this is for n, basically 0 on up to infinity. So this is a differential equation for the probability for having an mRNA. But this is, we have to have a different equation for each n, 0, 1, 2, 3, 4, 5 on up.

So this is what I mean by converting single differential equation, which is actually an exceedingly simple one, for one that is for an infinite set. And each one is even a little bit more complicated. In general, these f's and n's can be pretty complicated. In this situation they're not so bad.

But let's make sure. Can somebody say what fn and gn are equal to? Any volunteers?

AUDIENCE: [INAUDIBLE]

PROFESSOR: Right so fn, this is rate that we add a new mRNA. Well that's just synthesis rate for mRNA. And this guy is what?

AUDIENCE: Delta.

PROFESSOR: So this is degradation rate. And we actually do have to multiply still by the number n. And that's because as we go further out here to the right, then it is true. The rate at which we come back to the left is increasing. Because there's just more mRNA that can be degraded.

Now it's worth saying that you can, for example, use this to simulate the probability distribution if you start from any distribution you like. So for example, you could start M0 equal to 1. And then just simulate how the probability recalibrates and comes over here. Similarly, you could do it over here. You could start with any probability distribution you want. And you could use this as a framework to calculate what the probability distribution will be at any time later.

But you can also use this just as a way of figuring out what the equilibrium

distribution is going to be. Because at equilibrium, we can just ask, for each one of these arrows, the probability of moving to the right has to be equal to the probability of moving to the left, otherwise we wouldn't be at equilibrium. And that's true for every one of these kinds of pairs of arrows.

And in particular, what we can get, and I want to make sure that-- so but it's not that fn is equal to gn-- so it's really going to end up being that if you see what fn and gn, so that fn is going to have to be equal to g of n plus 1 for all n. Yes?

AUDIENCE: Do you also need to multiply like natural probabilities--

PROFESSOR: Ah, Yes, yes, indeed. So that sorry, times mn times m of n plus 1. So it's the kind of probably flux so we have to equalize.

So this is nice because this gives us a ratio of things. In particular, this tells us that the probability of being in the n plus 1 divided by the probability of being n, and this is at equilibrium. Is going to be fn divided by gn plus 1. Which is this synthesis rate. And then down here is going to be this degradation rate times, in this case, n plus 1.

So this is useful. Because for example, if we start at m, we could say that m1 over m0-- well maybe we'll even put the m0 over on the right. So then m1, what is that equal to? That's going to be synthesis rate divided degradation rate, times m0.

But then we also know that m2, well that's going to be again, synthesis rate divided by degradation rate. And we're going to get a squared. But then now we have to divide by 1/2 times m0.

Continuing on, m3 we get Sm over delta m cubed, divided by 1 over 3 times 2 times m0. So in general, we get the probability of being in the nth state, is going to be this thing. We'll call it lambda for now. Lambda to the n, divided by n factorial, times m0.

Now what's the-- and I'll-- remember lambda here we've defined it to be the ratio Sm over delta m. Now if we sum over all these probabilities, what should we get?

AUDIENCE: 1.

PROFESSOR: 1. Right, if we sum over this thing, what does that equal to? It's what?

AUDIENCE: Eta lambda.

PROFESSOR: Eta lambda, right? So just remember in this world-- the sum over lambda to the n, n factorial, from n equal to 0 to infinity, this is indeed the definition of e to the lambda. So what that means is that the normalization condition is that m0 has to be equal to e to the minus lambda, which is indeed a Poisson distribution. I'll raise it up a little bit.

So this is saying, OK, to back up. If we just have constant rate of creation of something, constant rate of degradation of that thing, on a per item basis, per unit basis, then you end up getting a Poisson distribution, at equilibrium for the number of that thing, in this case, the number of mRNA in the cell.

Questions about why that is? What happened? How we calculate it?

AUDIENCE: Could you explain why [INAUDIBLE]?

PROFESSOR: Sure. So this is basically f of n. And this is basically this g of n. But remember here n is the number of proteins or the number of mRNA. So then that's in the context of the master equation, then m and n are there. You get n by the current number of m. Does that make sense? Yes?

AUDIENCE: I'm confused how you changed m0 to the e to the minus lambda.

PROFESSOR: OK. Well let's just do it. So mn, this is the probability that we observe n mRNA. And we know that the sum over mn, so all these probabilities from n equal to 0 to infinity, has to be equal to 1. Something has to happen.

Well let's do this sum. This is equal to the sum of lambda to the n, over n factorial m0. But m0, is this a function of n? No. m0 is just, this is just the probability at equilibrium that you have 0 mRNA. So we can just pull this thing out. This is just some number, some probability.

Now the statement is that while this thing, this is the definition of e to the lambda. So

in general, so e to the x we often write is equal to 1, plus x, plus x squared over 2, plus dot, dot, dot. So this thing is indeed just equal e to the lambda.

So what we know is that this is still 1. So m0 times e to the lambda, is equal to 1, so m0 is to the minus lambda.

Any other questions about how we got here? What's going on? Yes?

AUDIENCE: The plot of the solution to the adjoining equation, that would be like the mean value, that would be the behavior of the mean values?

PROFESSOR: That Is the expected behavior of the mean value over time. In this case, fn and gn are both linear functions of the number of the mRNA. Which means that in the context of the master equation, if you ask about the expectation of mn, this quantity is indeed equal to-- it has the same behavior as, over time, as the deterministic equations.

So if f and g are nonlinear, then actually you get a deviation. But in this case, it is indeed the same. What it means that if you compare the stochastic and the deterministic trajectories, what you would see is that this thing is going to be a little bit jagged, or whatnot. And then even at equilibrium it's going to come up and down a little bit. I'm trying to add a little bit of jaggedness because it's discrete.

But the deterministic equation here is what you would get if you average together an infinite number of these stochastic trajectories. Because another one might have come down here. Does that answer?

AUDIENCE: Is m playing a double role? Like in that deterministic equation, m is the concentration of mRNA?

PROFESSOR: I think that I'm-- yeah I think that I should-- my nomenclature I think was not very good. I've used two different things. And now that I'm doing this, I think that I should have-- I should have just called it p of n, or maybe I should've used n here. I think I was trying to be consistent with some of the previous, but I think it was a mistake. Yes?

AUDIENCE: Are you plotting stochastic?

PROFESSOR: I'm plotting-- OK, so no, I'm not. So this is if you run an actual stochastic trajectory. Then at any moment in time, you just have one-- there's some number of mRNA. Whereas the sum over the mn's, this is talking about the probability distribution of the entire thing. So really if you started here, the master equation would give you some distribution for the n's, some distribution for m's. And so if you looked at these over time, than the mean of these distributions is indeed equal to the deterministic behavior. Yes?

AUDIENCE: Is it possible to recover, like how would we recover the differential equation from the master equation? Is that possible? Maybe that would help.

PROFESSOR: Yeah. I think that in the end, there's going to be a one-to-one relationship from, I guess, this differential equation to the master equation. I'm trying to think of any weird case or something funny's going to happen. Is something funny going to happen?

AUDIENCE: No. But like the easy way is just to write them all in terms of the distribution. And you can just differentiate the whole sum. And in that sum, we express the [INAUDIBLE] with your last equation. [INAUDIBLE].

PROFESSOR: Right. But I think this is the much more mathematical way. I mean because I think that actually, I mean, from the differential equation, you actually from the terms here, you can actually construct the master equation. And I think by the same way, you can go from the master equation, and I think that there's going to be a unique differential equation that would have gotten you to that master equation. So I think just from the terms you can do it.

You could also do like moment generating functions to get to how things change. But I mean I think that it's really from this, for example, I think it tells you that that was the differential equation. Does that--

I mean it's sort of-- the way that we typically do things these things, is that we have

a differential equation, and then we construct the master equation. So then we already knew what the differential equation was. But I think just from the terms in your master equation, you can say, all right. This was the differential equation that it started with.

Any other questions about what happened here? So we have, I think, a fair number, a fair knowledge of what's going on here now. We know that the equilibrium distribution of mRNA in the cell is going to be Poisson. We also know that the distribution of the number of mRNA produced per sell cycle is also Poisson. But it's a different Poisson from the first one.

We know that the number of proteins produced per mRNA is going to be geometrically distributed. The one thing that we have not yet done is to ask about the distribution of protein in the cell. So let's say something about that.

I'm not going to do the whole derivation. Because it's harder. But I encourage you to-- even the continuous version of the derivation is definitely harder than this. But then the discrete derivation is even worse.

So what we're going to talk about, and the way we'll typically maybe think about this from the standpoint of this class is the continuous approximation to-- oh, that might have ended up being useful. Well it's OK. Is the continuous approximation to the real answer.

And in particular, just the way that the exponential is the continuous approximation of the geometric distribution, in the same way you can think about the equilibrium distribution of protein in the cell. In this model is going to be gamma distributed. But gamma is a continuous distribution. But it's a continuous analog of the negative binomial.

So let me just make sure I'm-- and Sunney Xie actually has a nice *PRL* paper where he derives the gamma distribution. But even earlier actually Paulson had derived this negative binomial distribution, the discrete version of the solution.

So this is the number of protein per cell. We already know the mean. So this is

approximately distributed as a gamma. A gamma is a distribution that requires two parameters to describe. So a Poisson can be described by single parameter. Gamma is typically described by two.

And b is going to be the burst size, whereas a is the mean number of bursts per cell cycle, which is the same as the mean number of mRNA produced, so mean number of bursts.

So the gamma of this a, b. All right. So the gamma of a is the gamma function. It's equal to-- now is it a minus 1 factorial? I always get the-- is it a minus 1 or a plus 1 factorial. Anybody remember this? Yeah, a minus 1.

I mean it's like a lot of things. You look at this equation. It doesn't really mean a whole lot. But I think that a reasonable way to think about this is the gamma is approximately what you get when you add together a different exponentials with length scale, given by b.

When you add probability distributions, you have to do a convolution. So in some ways, the way to think about it, and this kind of makes sense. Because what is happening is that it takes something of order cell division time for these proteins to go away. Because they're stable.

Now each-- and so then what you want to know is how many proteins are kind of produced over the course of a cell cycle. Well that actually you can get at by asking how many bursts are there going to be. And then how big are the bursts?

So indeed, the mean here is equal to a times b. And the variance is equal to a times b squared. So for example, if you have a single exponential distribution, with burst size b, then this is what you get. So this is the probability that you get n proteins. And this is this function of n. So for a single burst, this is exponentially distributed. So this is the continuous version.

Now if we add together multiple of these bursts, this is really saying that we sample from this distribution, say twice. And then we add the resulting value. So this is a convolution. You guys will have an opportunity to practice this on your problem sets.

But what happens is that you end up getting something that looks like-- it's going to go. So it increases linearly. If you added three of them together this increase is quadratic. And it kind of goes like that. So this thing becomes kind of-- it goes from a distribution where it's peaked at 0, to something that's peaked at a nonzero value.

Now you can ask, for example, what happens as for a large a, if you have many bursts, what does this thing look like? Oh, I wish I hadn't erased my probability distributions. So what are the gamma converged to for large a? A normal distribution. Right? So that's the central limit theorem.

If you take any well-behaved probability distribution, you add it. You sample from it many times. Then you end up getting a Gaussian. If you don't remember that very well, then this is something to read about over the weekend. Just like the Poisson is also going to go to-- for large lambda, the Poisson also looks like a Gaussian.

Can somebody give an explanation, an intuitive explanation for why that should be? Why it-- yes?

AUDIENCE: Because in a Poisson distribution, you can't have anything negative.

PROFESSOR: OK. So a Poisson distribution can't have anything-- but now I feel like you're arguing against me. Because a Gaussian has negative values, right?

AUDIENCE: Right. So when the mean is really small, only have [INAUDIBLE].

PROFESSOR: OK. Yeah. All right. So what you're saying is that Poisson for small lambda it can't go negative. OK. No I think that that's true. Yeah, and so somehow the probability distribution is somehow piling up, as you say. What are some other ways of thinking about this?

AUDIENCE: [INAUDIBLE]. Because if you have a low lambda that means it's a Poisson. And then I'm just imagining stretching out. [INAUDIBLE].

PROFESSOR: OK. So I think that's fair. Another way we can think about this, is let's say that we have some process that's occurring randomly over some period of time. And this

could be say, mRNA production. And here this is just the number that we observe here, this is going to be a Poisson, with some mean lambda.

Now let's just say that I take another one, same process, same period of time. How is this guy going to be distributed? So this also Poisson of lambda. Now let's say I take this probability distribution, and I take this probability distribution. And I convolve them. I'm going to do the calculation of my head. I did it.

So for those of you who haven't done convolutions-- whatever. Yes, what's the new distribution going to be? Poisson 2 lambda. And why does that have to be?

AUDIENCE: That line was sort of-- you put it by n.

PROFESSOR: Yeah. That's right. This line, I just kind of like I just made it up. I could have just said, oh. Well it's the same process occurring over here. So we have to have the mean. It's still is going to be a Poisson process. And the mean has to be the-- well we just had twice the length. And indeed, for independent probability distributions, means always add. So this all consistent will all the things we know. So this has to be a Poisson of 2 lambda. If I add another segment on here, it has to Poisson of 3 lambda.

But what you see is that we see that Poisson of n lambda, which is the sum over many Poissons. Poissons are well-behaved probability distributions. You add them together, you're going to have to get a Gaussian. So you can see that the Poisson has to become Gaussian for large lambda. And indeed it does.

So there's a comment about this in the--

AUDIENCE: It's a little bit more complicated than this because obviously you always just divide from lambda [INAUDIBLE]. Like you would have to say that Poisson lambda is just like a combination of S--

PROFESSOR: OK. You're saying that if I do this calculation backwards, I'm going to get into trouble. Because if I try to break them--

AUDIENCE: So if you require lambda to be-- you have to have like a significant probability of

getting at least one candidate, right?

PROFESSOR: So I'd say lambda has to be much, much larger than 1. So once you're at lambda of 100, it looks like a Gaussian. And in Sunney's paper, he had a comment about this. Does anybody remember what it was?

Was mRNA production really well described as-- they mention that actually there is some violation of this model in the data.

AUDIENCE: Does it go into eukaryotes?

PROFESSOR: Oh, as soon as you go into eukaryotes, this is why I stay away from them. But even in their data, in E. coli, they actually observed a deviation.

So what they found is that there was a cell cycle dependence to this bursting rate, i.e. the mRNA production over the course of the cell cycle. And presumably their conclusion of this was that you have this guy. And then he turns into, gets longer. And then eventually he septates, and then you get two cells.

What he found is that these longer cells had actually a larger rate of mRNA synthesis than the smaller cells. And actually this makes sense. Because here you maybe have just one copy of the genome. Whereas here you might have-- you're making a second copy. So you might have two copies of that gene. So it may make sense that this bursting rate should grow.

But does that mean that you should not expect it to be a Poisson distribution for the number of bursts per cell cycle? No. It actually is still-- it still is described by a Poisson. Because you can just say, this is the cell cycle. And here this is Poisson of sum lambda 1. Here is a Poisson of sum lambda 2. So there could be a different rate over the course of the thing. But you still have just two Poissons. You still get another Poisson. So adding Poissons, gives you backup Poisson. They don't have to have the same mean lambda.

I just want to make one comment about what you have to do once you start thinking about eukaryotes. And the basic-- so you can see the gamma distribution can either

be peaked at 0, or it can be peaked at nonzero value. So most, for like highly expressed proteins, you'll see that it looks something like this.

Now for eukaryotes, you also have to consider there's some rate that you go between an active and inactive promoter. And this actually makes things much more complicated. So there's a rate going to inactive, a rate going to active. And so now if you look at, for example, the mRNA number per cell, you'll see that it is no longer a Poisson.

And I encourage you, if you're curious about such things, to come up and look at this. The solution for the steady state distribution has been solved analytically. For example, Arjun Raj, who is the author of the review that you guys just read, derived this equation here, which I don't know if you can see. But even from a distance, you can see that this is the solution. And this is just for the mRNA distribution. This is not even getting to the level of the protein.

And it involves many gamma functions, as well as a confluent hypergeometric function of the first kind, which is a disaster. But he went to Courant. He was an applied mathematician. So this is, I guess, this is what you can do after doing a PhD in applied mathematics.

The point though is that it ends up being very complicated. And you can get hugely varying distributions for the mRNA. And indeed this is seen in individual cells. If you look at mammalian cells, just at the mRNA level, you can have some cells that have hardly any mRNA, some that have a huge number. The protein distributions actually end up being more regular than the mRNA distributions. Because of this difference in lifetime.

So the mRNA numbers may fluctuate wildly. But the protein numbers will fluctuate less, because they last longer. So then you do some averaging over this crazy mRNA business.

Now in the last-- yeah, go ahead.

AUDIENCE: In terms of timescale, like all this is switching to the active and inactive promoter,

like to the other--

PROFESSOR: Ah, yes. That's a good question. I think that people argue very much about this. This is kind of minutes. This can be hours. And this is maybe in between those timescales would be typical. And when I say hours, especially like in mammalian cells, they might only divide once a day or so. So then this gets to be many hours. And then I'd say minutes is kind of the--

So there were many biological examples that were discussed in that review. And I'm not going to talk about all them. But I think that it's a nice review. Because it goes over some of the papers that you've read, or that we've talked about over the course of the semester. It also illustrates some different biological context in which noise may play a role.

But I want to mention one study that was done by actually again, Arjun together with Hedia Maamar, in collaboration with Dave Dubnau, where they were studying this process of competence. So in B. subtilis, during sometimes particularly of starvation, or other forms of unhappiness, they kind of pick up DNA from outside. So they'll import DNA. Some of it may just be consumed. But some of it could actually be incorporated into the genome.

Now what they found is that this competence process is mediated by this protein comK. And there was a positive feedback loop, where this guy ends up positively activating itself. And this helps lead to bistability in this network. Only a small fraction of the cells kind of get into this high feedback state. Only a small fraction of them activate competence and then uptake DNA.

And what they were able to show in that study was that it was sort of noise-induced. That they were able to vary both the transcription rate and the translation rate, in a way so as to reduce the noise. The mean is the same. So if you in the context of this model, what they did is they varied transcription rate, and they varied translation rate, each say, by a factor of 2. So they got the same mean, but then different noise.

And we're out of time. I would have you vote. But can anybody remember? If you want to decrease the noise in the number of the proteins, which of these do you want to go up, and which do you want to go down? And which one is going up? Which one's going up let's say?

AUDIENCE: Sm.

PROFESSOR: Sm. Right, so if you want to reduce the noise, but keep the mean constant, you increase the rate of transcription, and you decrease the rate of translation. Because the noise is really driven by this protein bursting behavior here. And that's precisely what they did. They changed those two quantities. They got the same mean, lower noise, and then they reduced the amount of competence in that sur--