Quantum Techniques for Stochastic Mechanics

John C. Baez[1,2] and Jacob D. Biamonte[2,3]

[1] Department of Mathematics
University of California
Riverside, CA 92521, USA

[2] Centre for Quantum Technologies
National University of Singapore
Singapore 117543

[3] Institute for Scientific Interchange
Via Alassio 11/c
10126 Torino, Italy

email: baez@math.ucr.edu, jacob.biamonte@qubit.org

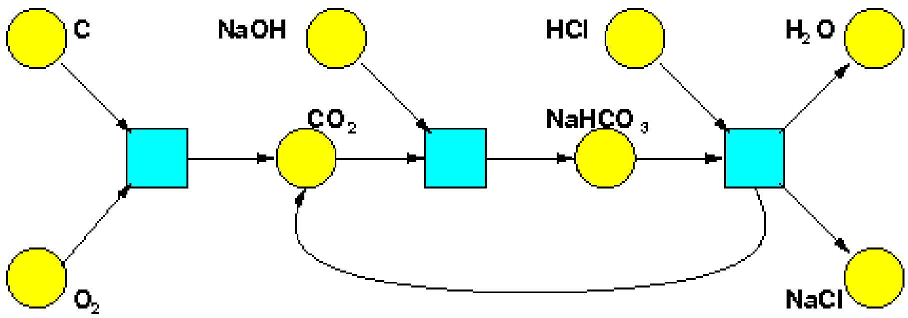
Abstract

Some ideas from quantum theory are just beginning to percolate back to classical probability theory. For example, there is a widely used and successful theory of 'chemical reaction networks', which describes the interactions of molecules in a stochastic rather than quantum way. Computer science and population biology use the same ideas under a different name: 'stochastic Petri nets'. But if we look at these theories from the perspective of quantum theory, they turn out to involve creation and annihilation operators, coherent states and other well-known ideas—but in a context where probabilities replace amplitudes. We explain this connection as part of a detailed analogy between quantum mechanics and stochastic mechanics. We also study the overlap of quantum mechanics and stochastic mechanics, which involves Hamiltonians that can generate either unitary or stochastic time evolution. These Hamiltonians are called 'Dirichlet forms', and they arise naturally from electrical circuits made only of resistors.

Contents

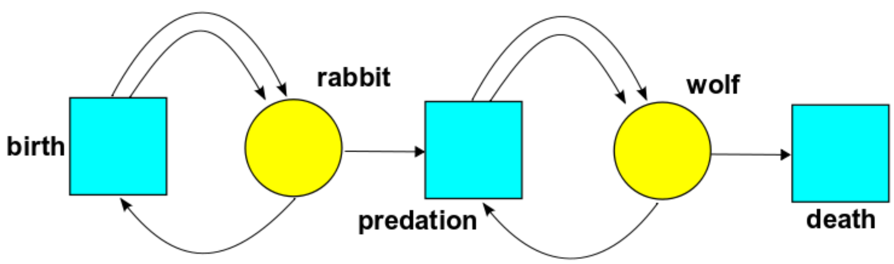
1 Stochastic Petri nets

Stochastic Petri nets are one of many different diagrammatic languages people have evolved to study complex systems. We'll see how they're used in chemistry, molecular biology, population biology and queuing theory, which is roughly the science of waiting in line. Here's an example of a Petri net taken from chemistry:



It shows some chemicals and some reactions involving these chemicals. To make it into a stochastic Petri net, we'd just label each reaction by a nonnegative real number: the reaction rate constant, or **rate constant** for short.

In general, a Petri net will have a set of **states**, which we'll draw as yellow circles, and a set of **transitions**, which we'll draw as blue rectangles. Here's a Petri net from population biology:



Now, instead of different chemicals, the states are different species. And instead of chemical reactions, the transitions are processes involving our species.

This Petri net has two states: **rabbit** and **wolf**. It has three transitions:

- In **birth**, one rabbit comes in and two go out. This is a caricature of reality: these bunnies reproduce asexually, splitting in two like amoebas.
- In **predation**, one wolf and one rabbit come in and two wolves go out. This is a caricature of how predators need to eat prey to reproduce.
- In **death**, one wolf comes in and nothing goes out. Note that we're pretending rabbits don't die unless they're eaten by wolves.

If we labelled each transition with a rate constant, we'd have a stochastic Petri net.

To make this Petri net more realistic, we'd have to make it more complicated. I'm trying to explain general ideas here, not realistic models of specific situations. Nonetheless, this Petri net already leads to an interesting model of population dynamics: a special case of the so-called 'Lotka-Volterra predator-prey model'. We'll see the details soon.

More to the point, this Petri net illustrates some possibilities that our previous example neglected. Every transition has some 'input' states and some 'output' states. But a state can show up more than once as the output (or input) of some transition. And as we see in 'death', we can have a transition with no outputs (or inputs) at all.

But let me stop beating around the bush, and give you the formal definitions. They're simple enough:

Definition 1. *A* **Petri net** *consists of a set S of* **states** *and a set T of* **transitions**, *together with a function*

$$i : S \times T \to \mathbb{N}$$

saying how many copies of each state shows up as **input** *for each transition, and a function*

$$o : S \times T \to \mathbb{N}$$

saying how many times it shows up as **output**.

Definition 2. *A* **stochastic Petri net** *is a Petri net together with a function*

$$r : T \to [0, \infty)$$

giving a **rate constant** *for each transition.*

Starting from any stochastic Petri net, we can get two things. First:

- The master equation. This says how the *probability that we have a given number of things in each state* changes with time.

Since stochastic means 'random', the master equation is what gives stochastic Petri nets their name. The master equation is the main thing I'll be talking about in future blog entries. But not right away!

Why not?

In chemistry, we typically have a huge number of things in each state. For example, a gram of water contains about 3×10^{22} water molecules, and a smaller but still enormous number of hydroxide ions (OH_-), hydronium ions (H_3O_+), and other scarier things. These things blunder around randomly, bump into each other, and sometimes react and turn into other things. There's a stochastic Petri net describing all this, as we'll eventually see. But in this situation, we don't usually want to know the probability that there are, say, exactly $310, 1849, 578, 476, 264$ hydronium ions. That would be too much information!

We'd be quite happy knowing the *expected value* of the number of hydronium ions, so we'd be delighted to have a differential equation that says how this changes with time.

And luckily, such an equation exists; and it's much simpler than the master equation. So, today we'll talk about:

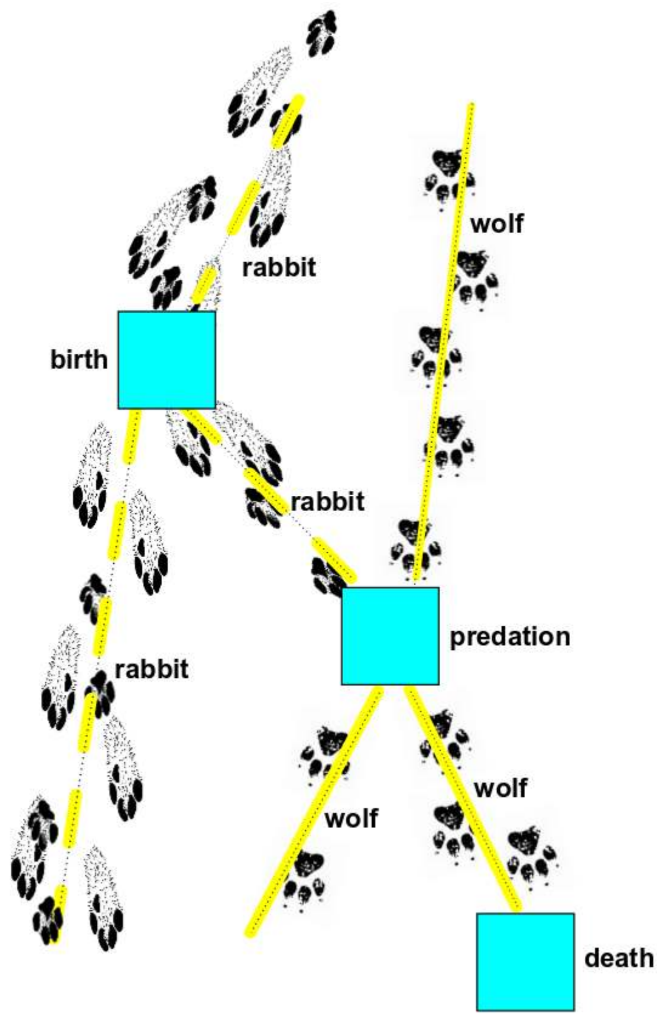
- The rate equation. This says how the *expected number of things in each state* changes with time.

But first, I hope you get the overall idea. The master equation is stochastic: at each time the number of things in each state is a random variable taking values in $\mathbb{N}$, the set of natural numbers. The rate equation is deterministic: at each time the expected number of things in each state is a non-random variable taking values in $[0, \infty)$, the set of nonnegative real numbers. If the master equation is the true story, the rate equation is only approximately true; but the approximation becomes good in some limit where the expected value of the number of things in each state is large, and the standard deviation is comparatively small.

If you've studied physics, this should remind you of other things. The master equation should remind you of the quantum harmonic oscillator, where energy levels are discrete, and probabilities are involved. The rate equation should remind you of the classical harmonic oscillator, where energy levels are continuous, and everything is deterministic.

When we get to the 'original research' part of our story, we'll see this analogy is fairly precise! We'll take a bunch of ideas from quantum mechanics and quantum field theory, and tweak them a bit, and show how we can use them to describe the master equation for a stochastic Petri net.

Indeed, the random processes that the master equation describes can be drawn as pictures:

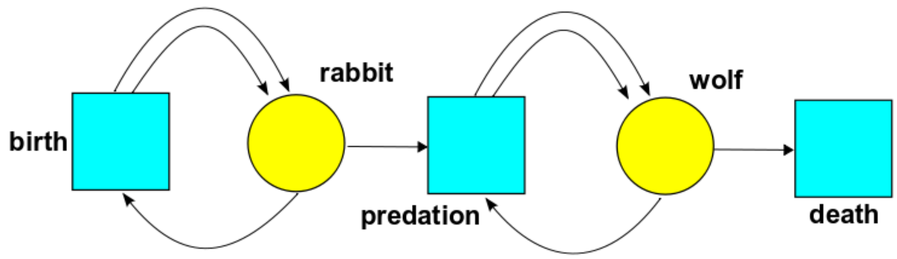


This looks like a Feynman diagram, with animals instead of particles! It's pretty funny, but the resemblance is no joke: the math will back it up.

I'm dying to explain all the details. But just as classical field theory is easier than quantum field theory, the rate equation is simpler than the master equation. So we should start there.

1.1 The rate equation

If you hand me a stochastic Petri net, I can write down its rate equation. Instead of telling you the general rule, which sounds rather complicated at first, let me do an example. Take the Petri net we were just looking at:



We can make it into a stochastic Petri net by choosing a number for each transition:

- the birth rate constant β
- the predation rate constant γ
- the death rate constant δ

Let $x(t)$ be the number of rabbits and let $y(t)$ be the number of wolves at time t. Then the rate equation looks like this:

$$\frac{dx}{dt} = \beta x - \gamma xy$$

$$\frac{dy}{dt} = \gamma xy - \delta y$$

It's really a system of equations, but I'll call the whole thing 'the rate equation' because later we may get smart and write it as a single equation.

See how it works?

- We get a term βx in the equation for rabbits, because rabbits are born at a rate equal to the number of rabbits times the birth rate constant β.
- We get a term $-\delta y$ in the equation for wolves, because wolves die at a rate equal to the number of wolves times the death rate constant δ.
- We get a term $-\gamma xy$ in the equation for rabbits, because rabbits die at a rate equal to the number of rabbits times the number of wolves times the predation rate constant γ.
- We also get a term γxy in the equation for wolves, because wolves are born at a rate equal to the number of rabbits times the number of wolves times γ.

Of course I'm *not* claiming that this rate equation makes any sense biologically! For example, think about predation. The γxy terms in the above equation would make sense if rabbits and wolves roamed around randomly, and whenever a wolf and a rabbit came within a certain distance, the wolf had a certain probability of eating the rabbit and giving birth to another wolf. At least it would be make sense in the limit of large numbers of rabbits and wolves, where we can treat x and y as varying continuously rather than discretely. That's a reasonable approximation to make sometimes. Unfortunately, rabbits and wolves *don't* roam around randomly, and a wolf *doesn't* spit out a new wolf each time it eats a rabbit.

Despite that, the equations

$$\frac{dx}{dt} = \beta x - \gamma xy$$

$$\frac{dy}{dt} = \gamma xy - \delta y$$

are actually studied in population biology. As I said, they're a special case of the **Lotka-Volterra predator-prey model**, which looks like this:

$$\frac{dx}{dt} = \beta x - \gamma xy$$

$$\frac{dy}{dt} = \epsilon xy - \delta y$$

The point is that while these models are hideously oversimplified and thus quantitatively inaccurate, they exhibit interesting *qualititative* behavior that's fairly robust. Depending on the rate constants, these equations can show either a stable equilibrium or stable periodic behavior. And we go from one regime to another, we see a kind of catastrophe called a "Hopf bifurcation". IYou can read about this in week308 and week309 of *This Week's Finds*. Those consider some *other* equations, not the Lotka-Volterra equations. But their qualitative behavior is the same!

If you want stochastic Petri nets that give *quantitatively* accurate models, it's better to retreat to chemistry. Compared to animals, molecules come a lot closer to roaming around randomly and having a chance of reacting when they come within a certain distance. So in chemistry, rate equations can be used to make accurate predictions.

But I'm digressing. I should be explaining the *general recipe for getting a rate equation from a stochastic Petri net!* You might not be able to guess it from just one example. But I sense that you're getting tired. So let's stop now. Next time I'll do more examples, and maybe even write down a general formula. But if you're feeling ambitious, you can try this now:

Problem 1. Can you write down a stochastic Petri net whose rate equation is the Lotka-Volterra predator-prey model:

$$\frac{dx}{dt} = \beta x - \gamma xy$$

$$\frac{dy}{dt} = \epsilon xy - \delta y$$

for arbitrary $\beta, \gamma, \delta, \epsilon \geq 0$? If not, for which values of these rate constants can you do it?

1.2 References

If you want to study a bit on your own, here are some great online references on stochastic Petri nets and their rate equations:

- [2] — Peter J. E. Goss and Jean Peccoud, Quantitative modeling of stochastic systems in molecular biology by using stochastic Petri nets, Proc. Natl. Acad. Sci. USA **95**, 6750-6755 (1998). link
- [3] — Jeremy Gunawardena, Chemical reaction network theory for in-silico biologists, Lecture notes, (2003).
- [1] — Martin Feinberg, Lectures on reaction networks, Lecture notes, (1979).

I should admit that the first two talk about 'chemical reaction networks' instead of Petri nets. That's no big deal: any chemical reaction network gives a Petri net in a pretty obvious way. You can probably figure out how; if you get stuck, just ask.

Also, I should admit that Petri net people say **place** where I'm saying **state**.

Here are some other references, which aren't free unless you have an online subscription or access to a library:

- [4] — Peter J. Haas, Stochastic Petri Nets: Modelling, Stability, Simulation, Springer, Berlin, (2002).
- [5] — F. Horn and R. Jackson, General mass action kinetics, Archive for Rational Mechanics and Analysis, **47**, 81-116, (1972). link
- [7] — Ina Koch, Petri nets - a mathematical formalism to analyze chemical reaction networks, Molecular Informatics **29**, 838-843, (2010). link
- [11] — Darren James Wilkinson, Stochastic Modelling for Systems Biology, Taylor & Francis, New York, (2006).

1.3 Answer to the problem

Here is the answer to the problem:

Problem 1. Can you write down a stochastic Petri net whose rate equation is the Lotka-Volterra predator-prey model:

$$\frac{dx}{dt} = \beta x - \gamma xy$$

$$\frac{dy}{dt} = \epsilon xy - \delta y$$

for arbitrary $\beta, \gamma, \delta, \epsilon \geq 0$? If not, for which values of these rate constants can you do it?

Answer. We can find a stochastic Petri net that does the job for any $\beta, \gamma, \delta, \epsilon \geq 0$. In fact we can find one that does the job for any possible value of $\beta, \gamma, \delta, \epsilon$. But to keep things simple, let's just solve the original problem.

We'll consider a stochastic Petri net with two states, **rabbit** and **wolf** , and four transitions:

- **birth** (1 rabbit in, 2 rabbits out), with rate constant β
- **death** (1 wolf in, 0 wolves out), with rate constant δ
- **jousting** (1 wolf and 1 rabbit in, R rabbits and W wolves out, where R, W are arbitrary natural numbers), with rate constant κ
- **dueling** (1 wolf and 1 rabbit in, R' rabbits and W' wolves out, where R', W' are arbitrary natural numbers) with rate constant κ'.

All these rate constants are nonnegative.

This gives the rate equation:

$$\frac{dx}{dt} = \beta x + (R-1)\kappa xy + (R'-1)\kappa' xy$$

$$\frac{dy}{dt} = (W-1)\kappa xy + (W'-1)\kappa' xy - \delta y$$

This is flexible enough to do the job.

For example, let's assume that when they joust, the massive, powerful wolf always kills the rabbit, and then eats the rabbit and has one offspring ($R = 0$ and $W = 2$). And let's assume that in a duel, the lithe and clever rabbit always kills the wolf, but does not reproduce afterward ($R' = 1$, $W' = 0$).

Then we get

$$\frac{dx}{dt} = \beta x - \kappa xy$$

$$\frac{dy}{dt} = (\kappa - \kappa')xy - \delta y$$

This handles the equations

$$\frac{dx}{dt} = \beta x - \gamma xy$$

$$\frac{dy}{dt} = \epsilon xy - \delta y$$

where $\beta, \gamma, \delta, \epsilon \geq 0$ and $\epsilon \geq \gamma$. In other words, the cases where more rabbits die due to combat than wolves get born!

I'll let you handle the cases where fewer rabbits die than wolves get born.

If we also include a death process for rabbits and birth process for wolves, we can get the fully general Lotka-Volterra equations:

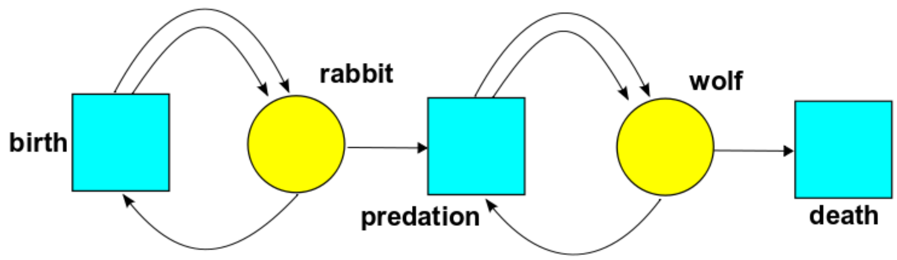
$$\frac{dx}{dt} = \beta x - \gamma xy$$

$$\frac{dy}{dt} = \epsilon xy - \delta y$$

It's worth noting that biologists like to study these equations with different choices of sign for the constants involved: the predator-prey Lotka-Volterra equations and the competitive Lotka-Volterra equations.

2 The rate equation of a stochastic Petri net

As we saw previously in Section 1, a Petri net is a picture that shows different kinds of things and processes that turn bunches of things into other bunches of things, like this:



The kinds of things are called **states** and the processes are called **transitions**. We see such transitions in chemistry:

$H + OH \to H_2O$

and population biology:

amoeba → amoeba + amoeba

and the study of infectious diseases:

infected + susceptible → infected + infected

and many other situations.

A "stochastic" Petri net says the rate at which each transition occurs. We can think of these transitions as occurring randomly at a certain rate—and then we get a stochastic process described by something called the "master equation". But for starters, we've been thinking about the limit where there are very many things in each state. Then the randomness washes out, and the expected number of things in each state changes deterministically in a manner described by the "rate equation".

It's time to explain the general recipe for getting this rate equation! It looks complicated at first glance, so I'll briefly state it, then illustrate it with tons of examples, and then state it again.

One nice thing about stochastic Petri nets is that they let you dabble in many sciences. Last time we got a tiny taste of how they show up in population biology. This time we'll look at chemistry and models of infectious diseases. I won't dig very deep, but take my word for it: you can do a lot with stochastic Petri nets in these subjects! I'll give some references in case you want to learn more.

2.1 Rate equations: the general recipe

Here's the recipe, really quickly:

A stochastic Petri net has a set of **states** and a set of **transitions**. Let's concentrate our attention on a particular transition. Then the ith state will appear m_i times as the **input** to that transition, and n_i times as the **output**. Our transition also has a **reaction rate** $0 < \infty$.

The rate equation answers this question:

$$\frac{dx_i}{dt} = ???$$

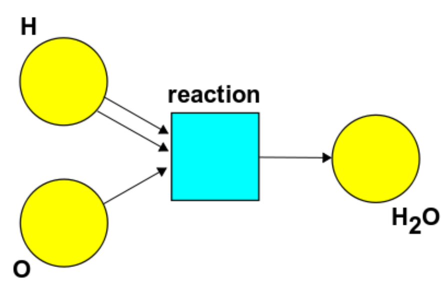
where $x_i(t)$ is the number of **things** in the ith state at time t. The answer is a sum of terms, one for each transition. Each term works the same way. For the transition we're looking at, it's

$$r(n_i - m_i)x_1^{m_1} \cdots x_k^{m_k}$$

The factor of $(n_i - m_i)$ shows up because our transition destroys m_i things in the ith state and creates n_i of them. The big product over all states, $x_1^{m_1} \cdots x_k^{m_k}$, shows up because our transition occurs at a rate proportional to the product of the numbers of things it takes as inputs. The constant of proportionality is the reaction rate r.

2.2 The formation of water (1)

But let's do an example. Here's a naive model for the formation of water from atomic hydrogen and oxygen:



This Petri net has just one transition: two hydrogen atoms and an oxygen atom collide simultaneously and form a molecule of water. That's not really how it goes... but what if it were? Let's use [H] for the number of hydrogen atoms, and so on, and let the reaction rate be α. Then we get this rate equation:

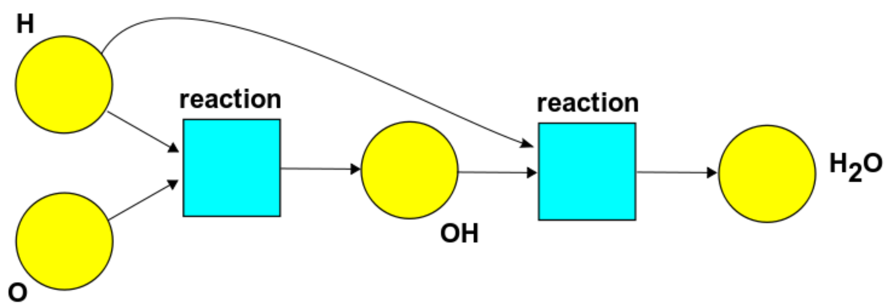
$$\begin{aligned}
\frac{d[\mathrm{H}]}{dt} &= -2\alpha[\mathrm{H}]^2[\mathrm{O}] \\
\frac{d[\mathrm{O}]}{dt} &= -\alpha[\mathrm{H}]^2[\mathrm{O}] \\
\frac{d[\mathrm{H_2O}]}{dt} &= \alpha[\mathrm{H}]^2[\mathrm{O}]
\end{aligned}$$

See how it works? The reaction occurs at a rate proportional to the product of the numbers of things that appear as *i*nputs: two H's and one O. The constant of proportionality is the rate constant α. So, the reaction occurs at a rate equal to $\alpha[\mathrm{H}]^2[\mathrm{O}]$. Then:

- Since two hydrogen atoms get used up in this reaction, we get a factor of -2 in the first equation.
- Since one oxygen atom gets used up, we get a factor of -1 in the second equation.
- Since one water molecule is formed, we get a factor of $+1$ in the third equation.

2.3 The formation of water (2)

Let me do another example, just so chemists don't think I'm an absolute ninny. Chemical reactions rarely proceed by having *three* things collide simultaneously—it's too unlikely. So, for the formation of water from atomic hydrogen and oxygen, there will typically be an intermediate step. Maybe something like this:



Here OH is called a 'hydroxyl radical'. I'm not sure this is the *most* likely pathway, but never mind—it's a good excuse to work out another rate equation. If the first reaction has rate constant α and the second has rate constant β, here's what we get:

$$\begin{aligned} \frac{d[\mathrm{H}]}{dt} &= -\alpha[\mathrm{H}][\mathrm{O}] - \beta[\mathrm{H}][\mathrm{OH}] \\ \frac{d[\mathrm{OH}]}{dt} &= \alpha[\mathrm{H}][\mathrm{O}] - \beta[\mathrm{H}][\mathrm{OH}] \\ \frac{d[\mathrm{O}]}{dt} &= -\alpha[\mathrm{H}][\mathrm{O}] \\ \frac{d[\mathrm{H_2O}]}{dt} &= \beta[\mathrm{H}][\mathrm{OH}] \end{aligned}$$

See how it works? Each reaction occurs at a rate proportional to the product of the numbers of things that appear as inputs. We get minus signs when a reaction destroys one thing of a given kind, and plus signs when it creates one. We don't get factors of 2 as we did last time, because now no reaction creates or destroys *two* of anything.

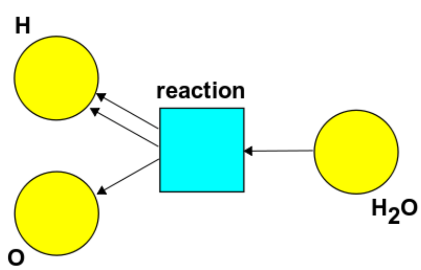
2.4 The dissociation of water (1)

In chemistry every reaction comes with a reverse reaction. So, if hydrogen and oxygen atoms can combine to form water, a water molecule can also 'dissociate' into hydrogen and oxygen atoms. The rate constants for the reverse reaction can be different than for the original reaction... and all these rate constants depend on the temperature. At room temperature, the rate constant for hydrogen and oxygen to form water is a lot higher than the rate constant for the reverse reaction. That's why we see a lot of water, and not many lone hydrogen or oxygen atoms. But at sufficiently high temperatures, the rate constants change, and water molecules become more eager to dissociate.

Calculating these rate constants is a big subject. I'm just starting to read this book, which looked like the easiest one on the library shelf:

- [8] — S. R. Logan, Chemical Reaction Kinetics, Longman, Essex, (1996).

But let's not delve into these mysteries today. Let's just take our naive Petri net for the formation of water and turn around all the arrows, to get the reverse reaction:



If the reaction rate is α, here's the rate equation:

$$\begin{aligned}
\frac{d[\mathrm{H}]}{dt} &= 2\alpha[\mathrm{H}^2\mathrm{O}] \\
\frac{d[\mathrm{O}]}{dt} &= \alpha[\mathrm{H}^2\mathrm{O}] \\
\frac{d[\mathrm{H}_2\mathrm{O}]}{dt} &= -\alpha[\mathrm{H}^2\mathrm{O}]
\end{aligned}$$

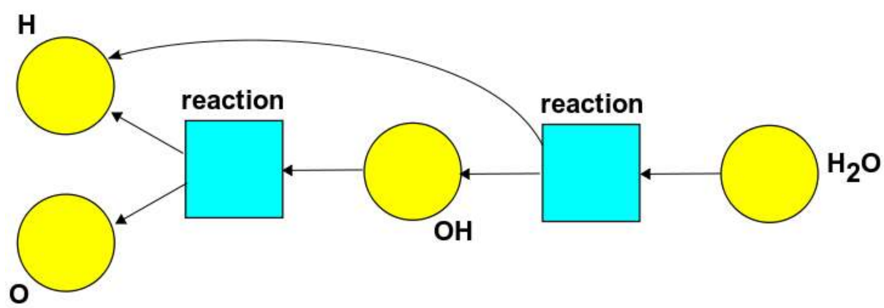
See how it works? The reaction occurs at a rate proportional to $[\mathrm{H}^2\mathrm{O}]$, since it has just a single water molecule as input. That's where the $\alpha[\mathrm{H}^2\mathrm{O}]$ comes from. Then:

- Since two hydrogen atoms get formed in this reaction, we get a factor of $+2$ in the first equation.
- Since one oxygen atom gets formed, we get a factor of $+1$ in the second equation.

- Since one water molecule gets used up, we get a factor of +1 in the third equation.

2.5 The dissociation of water (2)

Of course, we can also look at the reverse of the more realistic reaction involving a hydroxyl radical as an intermediate. Again, we just turn around the arrows in the Petri net we had:



Now the rate equation looks like this:

$$
\begin{aligned}
\frac{d[\mathrm{H}]}{dt} &= +\alpha[\mathrm{OH}] + \beta[\mathrm{H_2O}] \\
\frac{d[\mathrm{OH}]}{dt} &= -\alpha[\mathrm{OH}] + \beta[\mathrm{H_2O}] \\
\frac{d[\mathrm{O}]}{dt} &= +\alpha[\mathrm{OH}] \\
\frac{d[\mathrm{H_2O}]}{dt} &= -\beta[\mathrm{H_2O}]
\end{aligned}
$$

Do you see why? Test your understanding of the general recipe.

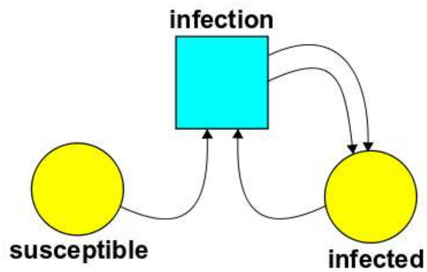
By the way: if you're a category theorist, when I said "turn around all the arrows" you probably thought "opposite category". And you'd be right! A Petri net is just a way of presenting of a strict symmetric monoidal category that's freely generated by some objects (the states) and some morphisms (the transitions). When we turn around all the arrows in our Petri net, we're getting a presentation of the *opposite* symmetric monoidal category. For more details, try:

- [10] — Vladimiro Sassone, On the category of Petri net computations, *6th International Conference on Theory and Practice of Software Development*, Proceedings of TAPSOFT '95, Lecture Notes in Computer Science **915**, Springer, Berlin, pp. 334-348 (1995).

After I explain how stochastic Petri nets are related to quantum field theory, I hope to say more about this category theory business. But if you don't understand it, don't worry about it now—let's do a few more examples.

2.6 The SI model

The **SI model** is an extremely simple model of an infectious disease. We can describe it using this Petri net:



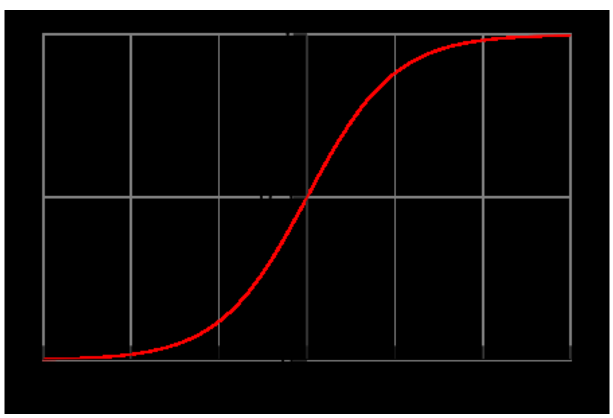
There are two states: **susceptible** and **infected**. And there's a transition called **infection**, where an infected person meets a susceptible person and infects them.

Suppose S is the number of susceptible people and I the number of infected ones. If the rate constant for infection is β, the rate equation is

$$\begin{aligned} \frac{dS}{dt} &= -\beta SI \\ \frac{dI}{dt} &= \beta SI \end{aligned}$$

Do you see why?

By the way, it's easy to solve these equations exactly. The total number of people doesn't change, so $S + I$ is a conserved quantity. Use this to get rid of one of the variables. You'll get a version of the famous logistic equation, so the fraction of people infected must grow sort of like this:



Problem 2. Is there a stochastic Petri net with just one state whose rate equation is the logistic equation:

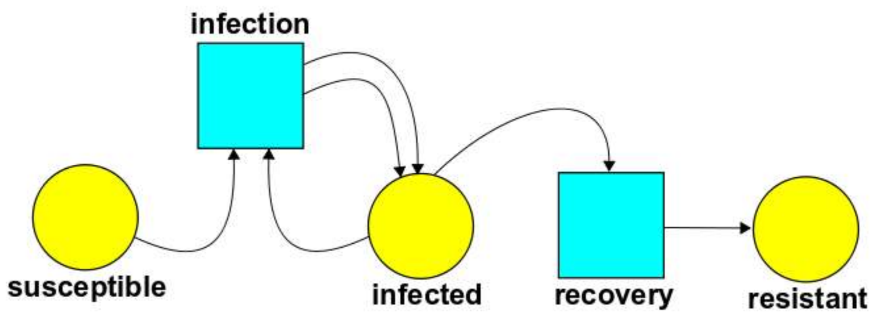
$$\frac{dP}{dt} = \alpha P - \beta P^2?$$

2.7 The SIR model

The SI model is just a warmup for the more interesting **SIR model**, which was invented by Kermack and McKendrick in 1927:

- [6] — W. O. Kermack and A. G. McKendrick, A Contribution to the mathematical theory of epidemics, Proc. Roy. Soc. Lond. A **115**, 700-721, (1927).

The SIR model has an extra state, called **resistant**, and an extra transition, called **recovery**, where an infected person gets better and develops resistance to the disease:



If the rate constant for infection is β and the rate constant for recovery is α, the rate equation for this stochastic Petri net is:

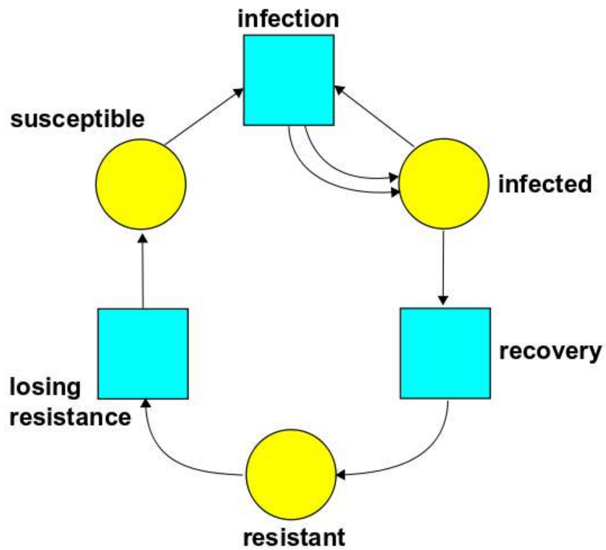
$$
\begin{aligned}
\frac{dS}{dt} &= -\beta SI \\
\frac{dI}{dt} &= \beta SI - \alpha I \\
\frac{dR}{dt} &= \alpha I
\end{aligned}
$$

See why?

I don't know a 'closed-form' solution to these equations. But Kermack and McKendrick found an approximate solution in their original paper. They used this to model the death rate from bubonic plague during an outbreak in Bombay, and got pretty good agreement. Nowadays, of course, we can solve these equations numerically on the computer.

2.8 The SIRS model

There's an even more interesting model of infectious disease called the **SIRS model** . This has one more transition, called **losing resistance** , where a resistant person can go back to being susceptible. Here's the Petri net:



Problem 3. If the rate constants for recovery, infection and loss of resistance are α, β, and γ, write down the rate equations for this stochastic Petri net.

In the SIRS model we see something new: cyclic behavior! Say you start with a few infected people and a lot of susceptible ones. Then lots of people get infected... then lots get resistant... and then, much later, if you set the rate constants right, they lose their resistance and they're ready to get sick all over again! You can sort of see it from the Petri net, which looks like a cycle.

I learned about the SI, SIR and SIRS models from this great book:

- [9] Marc Mangel, The Theoretical Biologist's Toolbox: Quantitative Methods for Ecology and Evolutionary Biology, Cambridge U. Press, Cambridge, (2006).

For more models of this type, see:

- Compartmental models in epidemiology, Wikipedia.

A 'compartmental model' is closely related to a stochastic Petri net, but beware: the pictures in this article are not really Petri nets!

2.9 The general recipe revisited

Now let me remind you of the general recipe and polish it up a bit. So, suppose we have a stochastic Petri net with k states. Let x_i be the number of things in the ith state. Then the rate equation looks like:

$$\frac{dx_i}{dt} = ???$$

It's really a bunch of equations, one for each $1 < i < k$. But what is the right-hand side?

The right-hand side is a sum of terms, one for each transition in our Petri net. So, let's assume our Petri net has just one transition! (If there are more, consider one at a time, and add up the results.)

Suppose the ith state appears as input to this transition m_i times, and as output n_i times. Then the rate equation is

$$\frac{dx_i}{dt} = r(n_i - m_i)x_1^{m_1} \cdots x_k^{m_k}$$

where r is the rate constant for this transition.

That's really all there is to it! But subscripts make my eyes hurt more and more as I get older—this is the real reason for using index-free notation, despite any sophisticated rationales you may have heard—so let's define a vector

$$x = (x_1, \ldots, x_k)$$

that keeps track of how many things there are in each state. Similarly let's make up an **input vector** :

$$m = (m_1, \ldots, m_k)$$

and an **output vector** :

$$n = (n_1, \ldots, n_k)$$

for our transition. And a bit more unconventionally, let's define

$$x^m = x_1^{m_1} \cdots x_k^{m_k}$$

Then we can write the rate equation for a single transition as

$$\frac{dx}{dt} = r(n - m)x^m$$

This looks a lot nicer! Any equation with more than 20 symbols is too ugly for me to enjoy.

Indeed, this emboldens me to consider a general stochastic Petri net with lots of transitions, each with their own rate constant. Let's write T for the set of transitions and $r(\tau)$ for the rate constant of the transition $\tau \in T$. Let $n(\tau)$ and $m(\tau)$ be the input and output vectors of the transition τ. Then the rate equation for our stochastic Petri net is

$$\frac{dx}{dt} = \sum_{\tau \in T} r(\tau)(n(\tau) - m(\tau))x^{m(\tau)}$$

That's the fully general recipe in a nutshell. I'm not sure yet how helpful this notation will be, but it's here whenever we want it.

Next time we'll get to the really interesting part, where ideas from quantum theory enter the game! We'll see how things in different states randomly transform into each other via the transitions in our Petri net. And someday we'll check that the *expected* number of things in each state evolves according to the rate equation we just wrote down... at least in there limit where there are lots of things in each state.

2.10 Answers to problems

Here are the answers to the problems:

Problem 2. Is there a stochastic Petri net with just one state whose rate equation is the logistic equation:

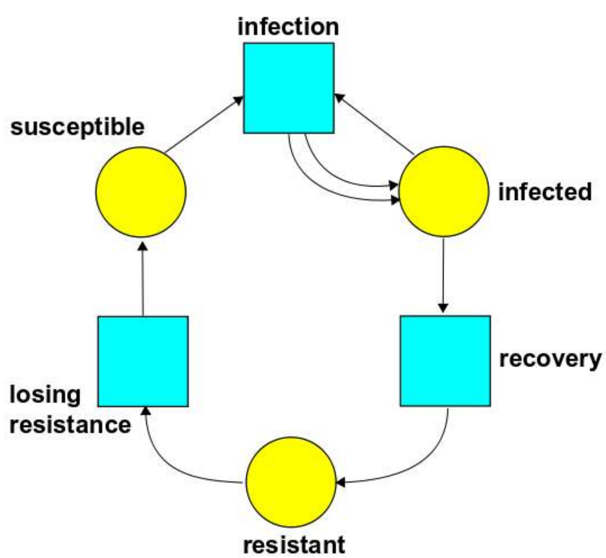
$$\frac{dP}{dt} = \alpha P - \beta P^2; ?$$

Answer. Yes. Use the Petri net with one state, say **amoeba** , and two transitions:

- **fission** , with one amoeba as input and two as output, with rate constant α.
- **competition** , with two amoebas as input and one as output, with rate constant β.

The idea of 'competition' is that when two amoebas are competing for limited resources, one may die.

Problem 3. If the rate constants for recovery, infection and loss of resistance are α, β, and γ, write down the rate equations for this stochastic Petri net:



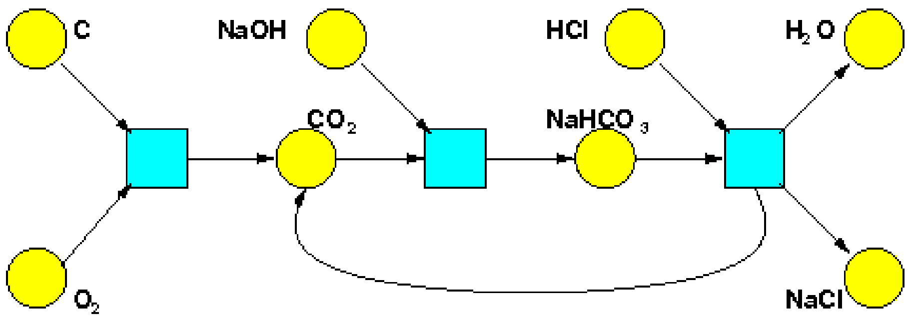
Answer. The rate equation is:

$$\begin{aligned}
\frac{dS}{dt} &= -\beta SI + \gamma R \\
\frac{dI}{dt} &= \beta SI - \alpha I \\
\frac{dR}{dt} &= \alpha I - \gamma R
\end{aligned}$$

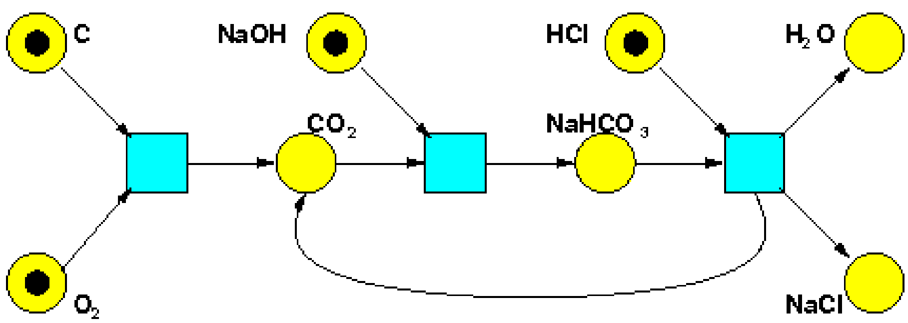
3 The master equation of a stochastic Petri net

Previously, in Section 2 I explained the rate equation of a stochastic Petri net. But now let's get serious: let's see what's *stochastic*—that is, random— about a stochastic Petri net. For this we need to forget the rate equation (temporarily) and learn about the 'master equation'. This is where ideas from quantum field theory start showing up!

A Petri net has a bunch of **states** and a bunch of **transitions**. Here's an example we've already seen, from chemistry:

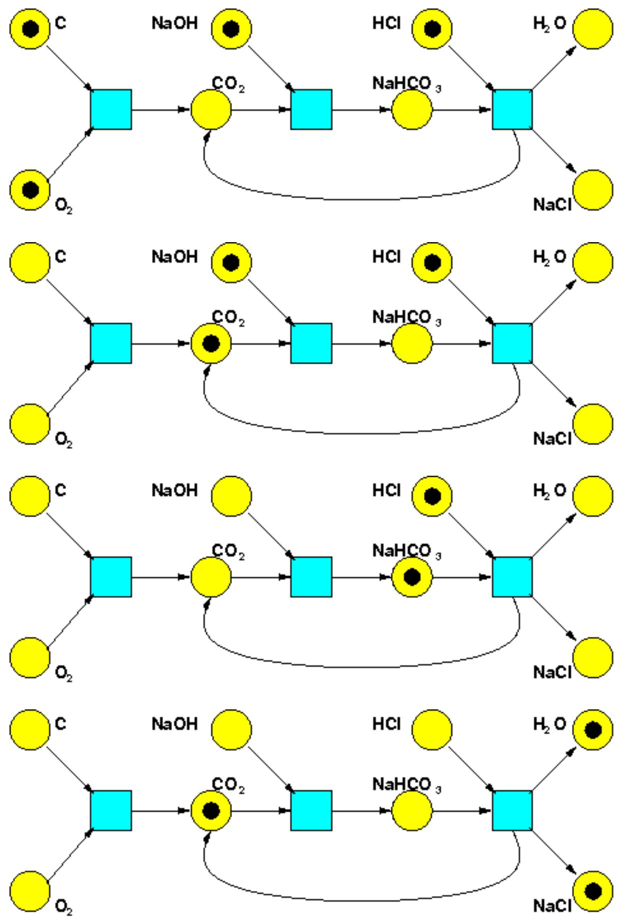


The states are in yellow, the transitions in blue. A **labelling** of our Petri net is a way of putting some number of **things** in each state. We can draw these things as little black dots:



In this example there are only 0 or 1 things in each state: we've got one atom of carbon, one molecule of oxygen, one molecule of sodium hydroxide, one molecule of hydrochloric acid, and nothing else. But in general, we can have any natural number of things in each state.

In a stochastic Petri net, the transitions occur randomly as time passes. For example, as time passes we could see a sequence of transitions like this:



Each time a transition occurs, the number of things in each state changes in an obvious way.

3.1 The master equation

Now, I said the transitions occur 'randomly', but that doesn't mean there's no rhyme or reason to them! The miracle of probability theory is that it lets us state precise laws about random events. The law governing the random behavior of a stochastic Petri net is called the 'master equation'.

In a stochastic Petri net, each transition has a **rate constant**, a nonnegative real number. Roughly speaking, this determines the probability of that transition.

A bit more precisely: suppose we have a Petri net that is labelled in some way at some moment. Then the probability that a given transition occurs in a short time Δt is approximately:

- the rate constant for that transition, times

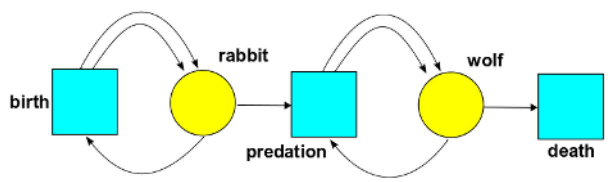
- the time Δt, times
- the number of ways the transition can occur.

More precisely still: this formula is correct up to terms of order $(\Delta t)^2$. So, taking the limit as $\Delta t \to 0$, we get a differential equation describing precisely how the probability of the Petri net having a given labelling changes with time! And this is the **master equation**.

Now, you might be impatient to actually *see* the master equation, but that would be rash. The true master doesn't need to see the master equation. It sounds like a Zen proverb, but it's true. The raw beginner in mathematics wants to see the solutions of an equation. The more advanced student is content to prove that the solution exists. But the master is content to prove that the *equation* exists.

A bit more seriously: what matters is understanding the rules that inevitably lead to some equation: actually *writing it down* is then straightforward.

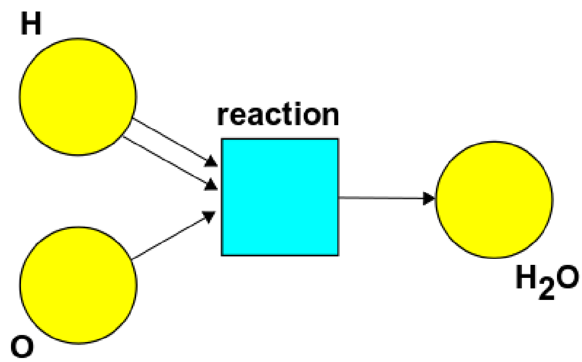
And you see, there's something I haven't explained yet: "the number of ways the transition can occur". This involves a bit of counting. Consider, for example, this Petri net:



Suppose there are 10 rabbits and 5 wolves.

- How many ways can the **birth** transition occur? Since birth takes one rabbit as input, it can occur in 10 ways.
- How many ways can **predation** occur? Since predation takes one rabbit and one wolf as inputs, it can occur in $10 \times 5 = 50$ ways.
- How many ways can **death** occur? Since death takes one wolf as input, it can occur in 5 ways.

Or consider this one:



Suppose there are 10 hydrogen atoms and 5 oxygen atoms. How many ways can they form a water molecule? There are 10 ways to pick the first hydrogen, 9 ways to pick the second hydrogen, and 5 ways to pick the oxygen. So, there are

$$10 \times 9 \times 5 = 450$$

ways.

Note that we're treating the hydrogen atoms as *distinguishable*, so there are 10×9 ways to pick them, not $\frac{10\times 9}{2} = \binom{10}{2}$. In general, the number of ways to choose M distinguishable things from a collection of L is the **falling power**

$$L^{\underline{M}} = L \cdot (L-1) \cdots (L-M+1)$$

where there are M factors in the product, but each is 1 less than the preceding one—hence the term 'falling'.

Okay, now I've given you all the raw ingredients to work out the master equation for any stochastic Petri net. The previous paragraph was a big fat hint. One more nudge and you're on your own:

Problem 4. Suppose we have a stochastic Petri net with k states and one transition with rate constant r. Suppose the ith state appears m_i times as the input of this transition and n_i times as the output. A labelling of this stochastic Petri net is a k-tuple of natural numbers $\ell = (\ell_1, \ldots, \ell_k)$ saying how many things are in each state. Let $\psi_\ell(t)$ be the probability that the labelling is ℓ at time t. Then the **master equation** looks like this:

$$\frac{d}{dt}\psi_{\ell'}(t) = \sum_{\ell} H_{\ell'\ell}\psi_\ell(t)$$

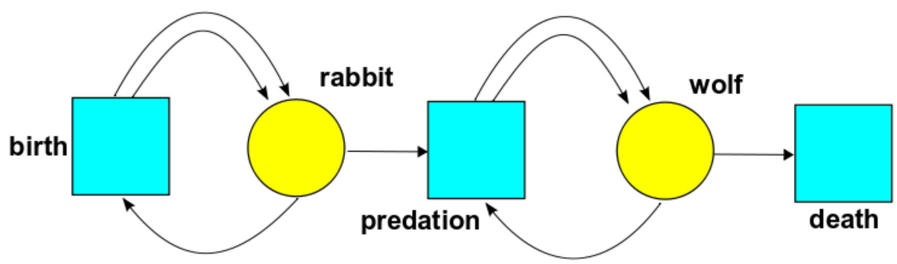
for some matrix of real numbers $H_{\ell'\ell}$. What is this matrix?

You can write down a formula for this matrix using what I've told you. And then, if you have a stochastic Petri net with more transitions, you can just compute the matrix for each transition using this formula, and add them all up.

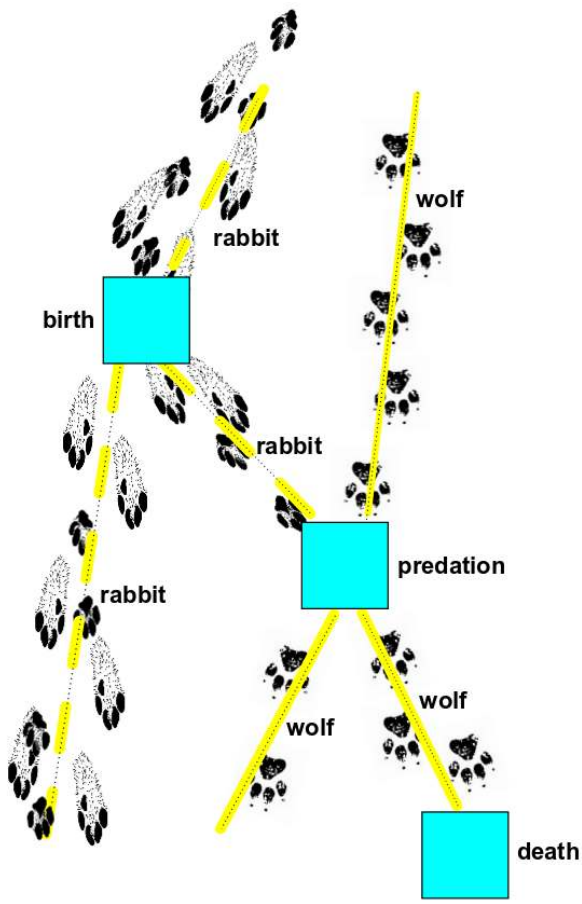
There's a straightforward way to solve this problem, but I want to get the solution by a strange route: I want to *guess* the master equation using ideas from *quantum field theory!*

Why? Well, if we think about a stochastic Petri net whose labelling undergoes random transitions as I've described, you'll see that any possible 'history' for the labelling can be drawn in a way that looks like a Feynman diagram. In quantum field theory, Feynman diagrams show how things interact and turn into other things. But that's what stochastic Petri nets do, too!

For example, if our Petri net looks like this:



then a typical history can be drawn like this:



Some rabbits and wolves come in on top. They undergo some transitions as time passes, and go out on the bottom. The vertical coordinate is time, while the horizontal coordinate doesn't really mean anything: it just makes the diagram easier to draw.

If we ignore all the artistry that makes it cute, this Feynman diagram is just a graph with states as edges and transitions as vertices. Each transition occurs at a specific time.

We can use these Feynman diagrams to compute the probability that if we start it off with some labelling at time t_1, our stochastic Petri net will wind up with some other labelling at time t_2. To do this, we just take a sum over Feynman diagrams that start and end with the given labellings. For each Feynman diagram, we integrate over all possible times at which the transitions occur. And what do we integrate? Just the product of the rate constants for those transitions!

That was a bit of a mouthful, and it doesn't really matter if you followed it in detail. What matters is that it *sounds a lot like stuff you learn when you* *study quantum field theory!*

That's one clue that something cool is going on here. Another is the master equation itself:

$$\frac{d}{dt}\psi_{\ell'}(t) = \sum_{\ell} H_{\ell'\ell}\psi_{\ell}(t)$$

This looks a lot like Schrödinger's equation, the basic equation describing how a quantum system changes with the passage of time.

We can make it look even more like Schrödinger's equation if we create a vector space with the labellings ℓ as a basis. The numbers $\psi_{\ell}(t)$ will be the components of some vector $\psi(t)$ in this vector space. The numbers $H_{\ell'\ell}$ will be the matrix entries of some operator H on that vector space. And the master equation becomes:

$$\frac{d}{dt}\psi(t) = H\psi(t)$$

Compare Schrödinger's equation:

$$i\frac{d}{dt}\psi(t) = H\psi(t)$$

The only visible difference is that factor of i!

But of course this is linked to another big difference: in the master equation ψ describes probabilities, so it's a vector in a *real* vector space. In quantum theory ψ describes probability amplitudes, so it's a vector in a *complex* Hilbert space.

Apart from this huge difference, everything is a lot like quantum field theory. In particular, our vector space is a lot like the Fock space one sees in quantum field theory. Suppose we have a quantum particle that can be in k different states. Then its Fock space is the Hilbert space we use to describe an arbitrary collection of such particles. It has an orthonormal basis denoted

$$|\ell_1 \cdots \ell_k\rangle$$

where $\ell_1, \ldots, \ell_k$ are natural numbers saying how many particles there are in each state. So, any vector in Fock space looks like this:

$$\psi = \sum_{\ell_1,\ldots,\ell_k} \psi_{\ell_1,\ldots,\ell_k} \, |\ell_1 \cdots \ell_k\rangle$$

But if write the whole list $\ell_1, \ldots, \ell_k$ simply as ℓ, this becomes

$$\psi = \sum_{\ell} \psi_{\ell}|\ell\rangle$$

This is almost like what we've been doing with Petri nets!—except I hadn't gotten around to giving names to the basis vectors.

In quantum field theory class, I learned lots of interesting operators on Fock space: annihilation and creation operators, number operators, and so on. So, when I bumped into this master equation

$$\frac{d}{dt}\psi(t) = H\psi(t)$$

it seemed natural to take the operator H and write it in terms of these. There was an obvious first guess, which didn't quite work... but thinking a bit harder eventually led to the right answer. Later, it turned out people had already thought about similar things. So, I want to explain this.

When I first started working on this stuff, I was focused on the difference between collections of *indistinguishable* things, like bosons or fermions, and collections of *distinguishable* things, like rabbits or wolves. But with the benefit of hindsight, it's even more important to think about the difference between quantum theory, which is all about *probability amplitudes*, and the game we're playing now, which is all about *probabilities*. So, next time I'll explain how we need to modify quantum theory so that it's about probabilities. This will make it easier to guess a nice formula for H.

3.2 Answer to the problem

Here is the answer to the problem:

Problem 4. Suppose we have a stochastic Petri net with k states and just one transition, whose rate constant is r. Suppose the ith state appears m_i times as the input of this transition and n_i times as the output. A labelling of this stochastic Petri net is a k-tuple of natural numbers $\ell = (\ell_1, \ldots, \ell_k)$ saying how many things are in each state. Let $\psi_\ell(t)$ be the probability that the labelling is ℓ at time t. Then the master equation looks like this:

$$\frac{d}{dt}\psi_{\ell'}(t) = \sum_\ell H_{\ell'\ell}\psi_\ell(t)$$

for some matrix of real numbers $H_{\ell'\ell}$. What is this matrix?

Answer. To compute $H_{\ell'\ell}$ it's enough to start the Petri net in a definite labelling ℓ and see how fast the probability of being in some labelling ℓ' changes. In other words, if at some time t we have

$$\psi_\ell(t) = 1$$

then

$$\frac{d}{dt}\psi_{\ell'}(t) = H_{\ell'\ell}$$

at this time.

Now, suppose we have a Petri net that is labelled in some way at some moment. Then I said the probability that the transition occurs in a short time Δt is approximately:

- the rate constant r, times
- the time Δt, times
- the number of ways the transition can occur, which is the product of falling powers $\ell_1^{\underline{m_1}} \cdots \ell_k^{\underline{m_k}}$. Let's call this product $\ell^{\underline{m}}$ for short.

Multiplying these 3 things we get

$$r\ell^{\underline{m}}\Delta t$$

So, the *rate* at which the transition occurs is just:

$$r\ell^{\underline{m}}$$

And when the transition occurs, it eats up m_i things in the ith state, and produces n_i things in that state. So, it carries our system from the original labelling ℓ to the new labelling

$$\ell' = \ell + n - m$$

So, *in this case* we have

$$\frac{d}{dt}\psi_{\ell'}(t) = r\ell^{\underline{m}}$$

and thus

$$H_{\ell'\ell} = r\ell^{\underline{m}}$$

However, that's not all: there's another case to consider! Since the probability of the Petri net being in this new labelling ℓ' is going up, the probability of it staying in the original labelling ℓ must be going down by the same amount. So we must also have

$$H_{\ell\ell} = -r\ell^{\underline{m}}$$

We can combine both cases into one formula like this:

$$H_{\ell'\ell} = r\ell^{\underline{m}}\left(\delta_{\ell',\ell+n-m} - \delta_{\ell',\ell}\right)$$

Here the first term tells us how fast the probability of being in the new labelling is going up. The second term tells us how fast the probability of staying in the original labelling is going down.

Note: each column in the matrix $H_{\ell'\ell}$ sums to zero, and all the off-diagonal entries are nonnegative. That's good: in the next section we'll show that this matrix must be 'infinitesimal stochastic', meaning precisely that it has these properties!

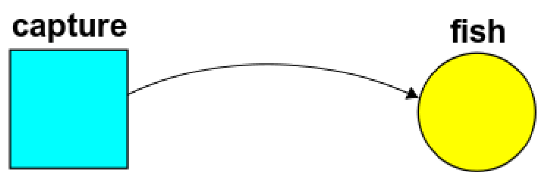
4 Probabilities vs amplitudes

Previously in Section 3 we saw clues that stochastic Petri nets are a lot like quantum field theory, but with probabilities replacing amplitudes. There's a powerful analogy at work here, which can help us a lot. So, this time I want to make that analogy precise.

But first, let me quickly sketch why it could be worthwhile.

4.1 A Poisson process

Consider this stochastic Petri net with rate constant r:



It describes an inexhaustible supply of fish swimming down a river, and getting caught when they run into a fisherman's net. In any short time Δt there's a chance of about $r\Delta t$ of a fish getting caught. There's also a chance of two or more fish getting caught, but this becomes negligible by comparison as $\Delta t \to 0$. Moreover, the chance of a fish getting caught during this interval of time is independent of what happens before or afterwards. This sort of process is called a **Poisson process**.

Problem 5. Suppose we start out knowing for sure there are no fish in the fisherman's net. What's the probability that he has caught n fish at time t?

Answer. At any time there will be some probability of having caught n fish; let's call this probability $\psi(n,t)$. We can summarize all these probabilities in a single power series, called a **generating function**:

$$\Psi(t) = \sum_{n=0}^{\infty} \psi(n,t)\, z^n$$

Here z is a formal variable—don't ask what it means, for now it's just a trick. In quantum theory we use this trick when talking about collections of photons rather than fish, but then the numbers $\psi(n,t)$ are complex 'amplitudes'. Now they are real probabilities, but we can still copy what the physicists do, and use this trick to rewrite the master equation as follows:

$$\frac{d}{dt}\Psi(t) = H\Psi(t)$$

This describes how the probability of having caught any given number of fish changes with time.

What's the operator H? Well, in quantum theory we describe the creation of photons using a certain operator on power series called the **creation operator**:

$$a^\dagger \Psi = z\Psi$$

We can try to apply this to our fish. If at some time we're 100% sure we have n fish, we have

$$\Psi = z^n$$

so applying the creation operator gives

$$a^\dagger \Psi = z^{n+1}$$

One more fish! That's good. So, an obvious wild guess is

$$H = ra^\dagger$$

where r is the rate at which we're catching fish. Let's see how well this guess works.

If you know how to exponentiate operators, you know to solve this equation:

$$\frac{d}{dt}\Psi(t) = H\Psi(t)$$

It's easy:

$$\Psi(t) = \exp(tH)\Psi(0)$$

Since we start out knowing there are no fish in the net, we have

$$\Psi(0) = 1$$

so with our guess for H we get

$$\Psi(t) = \exp(rta^\dagger)1$$

But $a^\dagger$ is the operator of multiplication by z, so $\exp(rta^\dagger)$ is multiplication by e^{rtz}, and

$$\Psi(t) = e^{rtz} = \sum_{n=0}^{\infty} \frac{(rt)^n}{n!} z^n$$

So, if our guess is right, the probability of having caught n fish at time t is

$$\frac{(rt)^n}{n!}$$

Unfortunately, this can't be right, because these probabilities don't sum to 1! Instead their sum is

$$\sum_{n=0}^{\infty} \frac{(rt)^n}{n!} = e^{rt}$$

We can try to wriggle out of the mess we're in by dividing our answer by this fudge factor. It sounds like a desperate measure, but we've got to try something!

This amounts to guessing that the probability of having caught n fish by time t is

$$\frac{(rt)^n}{n!} e^{-rt}$$

And this is *right!* This is called the **Poisson distribution**: it's famous for being precisely the answer to the problem we're facing.

So on the one hand our wild guess about H was wrong, but on the other hand it was not so far off. We can fix it as follows:

$$H = r(a^\dagger - 1)$$

The extra -1 gives us the fudge factor we need.

So, a wild guess corrected by an ad hoc procedure seems to have worked! But what's really going on?

What's really going on is that $a^\dagger$, or any multiple of this, is not a legitimate Hamiltonian for a master equation: if we define a time evolution operator $\exp(tH)$ using a Hamiltonian like this, probabilities won't sum to 1! But $a^\dagger - 1$ is okay. So, we need to think about which Hamiltonians are okay.

In quantum theory, self-adjoint Hamiltonians are okay. But in probability theory, we need some other kind of Hamiltonian. Let's figure it out.

4.2 Probability versus quantum theory

Probability vs quantum theory

Suppose we have a system of any kind: physical, chemical, biological, economic, whatever. The system can be in different states. In the simplest sort of model, we say there's some set X of states, and say that at any moment in time the system is definitely in one of these states. But I want to compare two other options:

- In a **probabilistic** model, we may instead say that the system has a **probability** $\psi(x)$ of being in any state $x \in X$. These probabilities are nonnegative real numbers with

$$\sum_{x \in X} \psi(x) = 1$$

- In a **quantum** model, we may instead say that the system has an **amplitude** $\psi(x)$ of being in any state $x \in X$. These amplitudes are complex numbers with

$$\sum_{x \in X} |\psi(x)|^2 = 1$$

Probabilities and amplitudes are similar yet strangely different. Of course given an amplitude we can get a probability by taking its absolute value and squaring it. This is a vital bridge from quantum theory to probability theory. Today, however, I don't want to focus on the bridges, but rather the *parallels* between these theories.

We often want to replace the sums above by integrals. For that we need to replace our set X by a measure space, which is a set equipped with enough structure that you can integrate real or complex functions defined on it. Well, at least you can integrate so-called 'integrable' functions—but I'll neglect all issues of analytical rigor here. Then:

- In a **probabilistic** model, the system has a **probability distribution** $\psi : X \to \mathbb{R}$, which obeys $\psi \geq 0$ and

$$\int_X \psi(x)\, dx = 1$$

- In a **quantum** model, the system has a **wavefunction** $\psi : X \to \mathbb{C}$, which obeys

$$\int_X |\psi(x)|^2\, dx = 1$$

In probability theory, we integrate ψ over a set $S \subset X$ to find out the probability that our systems state is in this set. In quantum theory we integrate $|\psi|^2$ over the set to answer the same question.

We don't need to think about sums over sets and integrals over measure spaces separately: there's a way to make any set X into a measure space such that by definition,

$$\int_X \psi(x)\, dx = \sum_{x \in X} \psi(x)$$

In short, integrals are more general than sums! So, I'll mainly talk about integrals, until the very end.

In probability theory, we want our probability distributions to be vectors in some vector space. Ditto for wave functions in quantum theory! So, we make up some vector spaces:

- In probability theory, the probability distribution ψ is a vector in the space

$$L^1(X) = \{\psi : X \to \mathbb{C} : \int_X |\psi(x)|\, dx < \infty\}$$

- In quantum theory, the wavefunction ψ is a vector in the space

$$L^2(X) = \{\psi : X \to \mathbb{C} : \int_X |\psi(x)|^2\, dx < \infty\}$$

You may wonder why I defined $L^1(X)$ to consist of *complex* functions when probability distributions are real. I'm just struggling to make the analogy seem as strong as possible. In fact probability distributions are not just real but nonnegative. We need to say this somewhere... but we can, if we like, start by saying they're complex-valued functions, but then whisper that they must in fact be nonnegative (and thus real). It's not the most elegant solution, but that's what I'll do for now.

Now:

- The main thing we can do with elements of $L^1(X)$, besides what we can do with vectors in any vector space, is integrate one. This gives a linear map:
$$\int : L^1(X) \to \mathbb{C}$$
- The main thing we can with elements of $L^2(X)$, besides the besides the things we can do with vectors in any vector space, is take the inner product of two:
$$\langle \psi, \phi \rangle = \int_X \overline{\psi}(x)\phi(x)\, dx$$
This gives a map that's linear in one slot and conjugate-linear in the other:
$$\langle -, - \rangle : L^2(X) \times L^2(X) \to \mathbb{C}$$

First came probability theory with $L^1(X)$; then came quantum theory with $L^2(X)$. Naive extrapolation would say it's about time for someone to invent an even more bizarre theory of reality based on $L^3(X)$. In this, you'd have to integrate the product of *three* wavefunctions to get a number! The math of *L^p spaces* is already well-developed, so give it a try if you want. I'll stick to L^1 and L^2 today.

4.3 Stochastic versus unitary operators

Now let's think about time evolution:

- In probability theory, the passage of time is described by a map sending probability distributions to probability distributions. This is described using a **stochastic operator**
$$U : L^1(X) \to L^1(X)$$
meaning a linear operator such that
$$\int U\psi = \int \psi$$
and
$$\psi \geq 0 \quad \Rightarrow \quad U\psi \geq 0$$

- In quantum theory the passage of time is described by a map sending wavefunction to wavefunctions. This is described using an **isometry**

$$U : L^2(X) \to L^2(X)$$

meaning a linear operator such that

$$\langle U\psi, U\phi\rangle = \langle \psi, \phi\rangle$$

In quantum theory we usually want time evolution to be reversible, so we focus on isometries that have inverses: these are called **unitary** operators. In probability theory we often consider stochastic operators that are *not* invertible.

4.4 Infinitesimal stochastic versus self-adjoint operators

Sometimes it's nice to think of time coming in discrete steps. But in theories where we treat time as continuous, to describe time evolution we usually need to solve a differential equation. This is true in both probability theory and quantum theory:

In probability theory we often describe time evolution using a differential equation called the **master equation**:

$$\frac{d}{dt}\psi(t) = H\psi(t)$$

whose solution is

$$\psi(t) = \exp(tH)\psi(0)$$

In quantum theory we often describe time evolution using a differential equation called Schr0̈dinger's equation:

$$i\frac{d}{dt}\psi(t) = H\psi(t)$$

whose solution is

$$\psi(t) = \exp(-itH)\psi(0)$$

In fact the appearance of i in the quantum case is purely conventional; we could drop it to make the analogy better, but then we'd have to work with 'skew-adjoint' operators instead of self-adjoint ones in what follows.

Let's guess what properties an operator H should have to make $\exp(-itH)$ unitary for all t. We start by assuming it's an isometry:

$$\langle \exp(-itH)\psi, \exp(-itH)\phi\rangle = \langle \psi, \phi\rangle$$

Then we differentiate this with respect to t and set $t = 0$, getting

$$\langle -iH\psi, \phi\rangle + \langle \psi, -iH\phi\rangle = 0$$

or in other words:

$$\langle H\psi, \phi\rangle = \langle \psi, H\phi\rangle$$

Physicists call an operator obeying this condition self-adjoint. Mathematicians know there's more to it, but today is not the day to discuss such subtleties, intriguing though they be. All that matters now is that there is, indeed, a correspondence between self-adjoint operators and well-behaved 'one-parameter unitary groups' $\exp(-itH)$. This is called Stone's Theorem.

But now let's copy this argument to guess what properties an operator H must have to make $\exp(tH)$ stochastic. We start by assuming $\exp(tH)$ is stochastic, so

$$\int \exp(tH)\psi = \int \psi$$

and

$$\psi \geq 0 \quad \Rightarrow \quad \exp(tH)\psi \geq 0$$

We can differentiate the first equation with respect to t and set $t = 0$, getting

$$\int H\psi = 0$$

for all ψ.

But what about the second condition,

$$\psi \geq 0 \quad \Rightarrow \quad \exp(tH)\psi \geq 0?$$

It seems easier to deal with this in the special case when integrals over X reduce to sums. So let's suppose that happens... and let's start by seeing what the *first* condition says in this case.

In this case, $L^1(X)$ has a basis of 'Kronecker delta functions': The Kronecker delta function δ_i vanishes everywhere except at one point $i \in X$, where it equals 1. Using this basis, we can write any operator on $L^1(X)$ as a matrix.

As a warmup, let's see what it means for an operator

$$U : L^1(X) \to L^1(X)$$

to be stochastic in this case. We'll take the conditions

$$\int U\psi = \int \psi$$

and

$$\psi \geq 0 \quad \Rightarrow \quad U\psi \geq 0$$

and rewrite them using matrices. For both, it's enough to consider the case where $\psi$is a Kronecker delta, say δ_j.

In these terms, the first condition says

$$\sum_{i \in X} U_{ij} = 1$$

for each column j. The second says

$$U_{ij} \geq 0$$

for all i, j. So in this case, a stochastic operator is just a square matrix where each column sums to 1 and all the entries are nonnegative. (Such matrices are often called left stochastic.)

Next, let's see what we need for an operator H to have the property that $\exp(tH)$is stochastic for all $t \geq 0$. It's enough to assume t is very small, which lets us use the approximation

$$\exp(tH) = 1 + tH + \cdots$$

and work to first order in t. Saying that each column of this matrix sums to 1 then amounts to

$$\sum_{i \in X} \delta_{ij} + tH_{ij} + \cdots = 1$$

which requires

$$\sum_{i \in X} H_{ij} = 0$$

Saying that each entry is nonnegative amounts to

$$\delta_{ij} + tH_{ij} + \cdots \geq 0$$

When $i = j$ this will be automatic when t is small enough, so the meat of this condition is

$$H_{ij} \geq 0 \mathrm{if} i \neq j$$

So, let's say H is an **infinitesimal stochastic** matrix if its columns sum to zero and its off-diagonal entries are nonnegative.

I don't love this terminology: do you know a better one? There should be some standard term. People here say they've seen such an operator called a 'stochastic Hamiltonian'. The idea behind my term is that any infintesimal stochastic operator should be the infinitesimal generator of a stochastic process.

In other words, when we get the details straightened out, any 1-parameter family of stochastic operators

$$U(t) : L^1(X) \to L^1(X) \qquad t \geq 0$$

obeying

$$U(0) = I$$

$$U(t)U(s) = U(t+s)$$

and continuity:

$$t_i \to t \quad \Rightarrow \quad U(t_i)\psi \to U(t)\psi$$

should be of the form

$$U(t) = \exp(tH)$$

for a unique 'infinitesimal stochastic operator' H.

When X is a finite set, this is true—and an infinitesimal stochastic operator is just a square matrix whose columns sum to zero and whose off-diagonal entries are nonnegative. But do you know a theorem characterizing infinitesimal stochastic operators for general measure spaces X? Someone must have worked it out.

Luckily, for our work on stochastic Petri nets, we only need to understand the case where X is a countable set and our integrals are really just sums. This should be almost like the case where X is a finite set—but we'll need to take care that all our sums converge.

4.5 The moral

Now we can see why a Hamiltonian like $a^\dagger$ is no good, while $a^\dagger - 1$ is good. (I'll ignore the rate constant r since it's irrelevant here.) The first one is not infinitesimal stochastic, while the second one is!

In this example, our set of states is the natural numbers:

$$X = \mathbb{N}$$

The probability distribution

$$\psi : \mathbb{N} \to \mathbb{C}$$

tells us the probability of having caught any specific number of fish.

The creation operator is not infinitesimal stochastic: in fact, it's stochastic! Why? Well, when we apply the creation operator, what was the probability of having n fish now becomes the probability of having $n+1$ fish. So, the probabilities remain nonnegative, and their sum over all n is unchanged. Those two conditions are all we need for a stochastic operator.

Using our fancy abstract notation, these conditions say:

$$\int a^\dagger \psi = \int \psi$$

and

$$\psi \geq 0; \Rightarrow; a^\dagger \psi \geq 0$$

So, precisely by virtue of being stochastic, the creation operator fails to be infinitesimal stochastic:

$$\int a^\dagger \psi \neq 0$$

Thus it's a bad Hamiltonian for our stochastic Petri net.

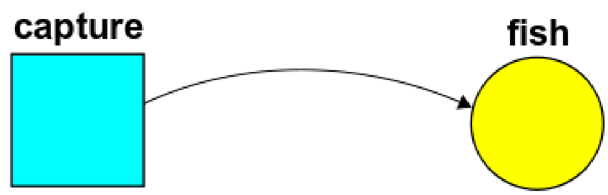
On the other hand, $a^\dagger - 1$ *is* infinitesimal stochastic. Its off-diagonal entries are the same as those of $a^\dagger$, so they're nonnegative. Moreover:

$$\int (a^\dagger - 1)\psi = 0$$

precisely because

$$\int a^\dagger \psi = \int \psi$$

You may be thinking: all this fancy math just to understand a single stochastic Petri net, the simplest one of all!



But next time I'll explain a general recipe which will let you write down the Hamiltonian for *any* stochastic Petri net. The lessons we've learned today will make this much easier. And pondering the analogy between probability theory and quantum theory will also be good for our bigger project of unifying the applications of network diagrams to dozens of different subjects.

5 Operators

Now for the fun part. Let's see how tricks from quantum theory can be used to describe random processes. I'll try to make this post completely self-contained, except at the very end. So, even if you skipped a bunch of the previous ones, this should make sense.

You'll need to know a bit of math: calculus, a tiny bit probability theory, and linear operators on vector spaces. You don't need to know quantum theory, though you'll have more fun if you do. What we're doing here is very similar, but also strangely different—for reasons I explained last time.

5.1 Rabbits and quantum mechanics

Suppose we have a population of rabbits in a cage and we'd like to describe its growth in a stochastic way, using probability theory. Let ψ_n be the probability of having n rabbits. We can borrow a trick from quantum theory, and summarize all these probabilities in a formal power series¿ like this:

$$\Psi = \sum_{n=0}^{\infty} \psi_n z^n$$

The variable z doesn't mean anything in particular, and we don't care if the power series converges. See, in math 'formal' means "it's only symbols on the page, just follow the rules". It's like if someone says a party is 'formal', so need to wear a white tie: you're not supposed to ask what the tie *means*.

However, there's a good reason for this trick. We can define two operators on formal power series, called the **annihilation operator**:

$$a\Psi = \frac{d}{dz}\Psi$$

and the **creation operator**:

$$a^\dagger \Psi = z\Psi$$

They're just differentiation and multiplication by z, respectively. So, for example, suppose we start out being 100% sure we have n rabbits for some particular number n. Then $\psi_n = 1$, while all the other probabilities are 0, so:

$$\Psi = z^n$$

If we then apply the creation operator, we obtain

$$a^\dagger \Psi = z^{n+1}$$

Voilà! One more rabbit!

The annihilation operator is more subtle. If we start out with n rabbits:

$$\Psi = z^n$$

and then apply the annihilation operator, we obtain

$$a\Psi = nz^{n-1}$$

What does this mean? The z^{n-1} means we have one fewer rabbit than before. But what about the factor of n? It means there were n different ways we could pick a rabbit and make it disappear! This should seem a bit mysterious, for various reasons... but we'll see how it works soon enough.

The creation and annihilation operators don't commute:

$$(aa^\dagger - a^\dagger a)\Psi = \frac{d}{dz}(z\Psi) - z\frac{d}{dz}\Psi = \Psi$$

so for short we say:

$$aa^\dagger - a^\dagger a = 1$$

or even shorter:

$$[a, a^\dagger] = 1$$

where the **commutator** of two operators is $[S, T] = ST - TS$.

The noncommutativity of operators is often claimed to be a special feature of *quantum* physics, and the creation and annihilation operators are fundamental to understanding the quantum harmonic oscillator. There, instead of rabbits, we're studying quanta of energy, which are peculiarly abstract entities obeying rather counterintuitive laws. So, it's cool that the same math applies to purely classical entities, like rabbits!

In particular, the equation $[a, a^\dagger] = 1$ just says that there's one more way to put a rabbit in a cage of rabbits, and then take one out, than to take one out and then put one in.

But how do we actually *use* this setup? We want to describe how the probabilities ψ_n change with time, so we write

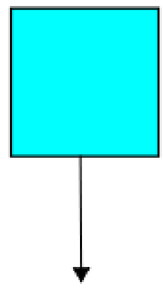
$$\Psi(t) = \sum_{n=0}^{\infty} \psi_n(t) z^n$$

Then, we write down an equation describing the rate of change of Ψ:

$$\frac{d}{dt}\Psi(t) = H\Psi(t)$$

Here H is an operator called the **Hamiltonian**, and the equation is called the **master equation**. The details of the Hamiltonian depend on our problem! But we can often write it down using creation and annihilation operators. Let's do some examples, and then I'll tell you the general rule.

5.2 Catching rabbits



Last time I told you what happens when we stand in a river and catch fish as they randomly swim past. Let me remind you of how that works. But today let's use rabbits.

So, suppose an inexhaustible supply of rabbits are randomly roaming around a huge field, and each time a rabbit enters a certain area, we catch it and add it to our population of caged rabbits. Suppose that on average we catch one rabbit per unit time. Suppose the chance of catching a rabbit during any interval of time is independent of what happens before or afterwards. What is the Hamiltonian describing the probability distribution of caged rabbits, as a function of time?

There's an obvious dumb guess: the creation operator! However, we saw last time that this doesn't work, and we saw how to fix it. The right answer is

$$H = a^\dagger - 1$$

To see why, suppose for example that at some time t we have n rabbits, so:

$$\Psi(t) = z^n$$

Then the master equation says that at this moment,

$$\frac{d}{dt}\Psi(t) = (a^\dagger - 1)\Psi(t) = z^{n+1} - z^n$$

Since $\Psi = \sum_{n=0}^\infty \psi_n(t) z^n$, this implies that the coefficients of our formal power series are changing like this:

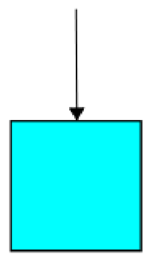
$$\frac{d}{dt}\psi_{n+1}(t) = 1$$

$$\frac{d}{dt}\psi_n(t) = -1$$

while all the rest have zero derivative at this moment. And that's exactly right! See, $\psi_{n+1}(t)$ is the probability of having one more rabbit, and this is going up at rate 1. Meanwhile, $\psi_n(t)$ is the probability of having n rabbits, and this is going down at the same rate.

Problem 6. Show that with this Hamiltonian and any initial conditions, the master equation predicts that the expected number of rabbits grows linearly.

5.3 Dying rabbits



Don't worry: no rabbits are actually injured in the research that Jacob Biamonte is doing here at the Centre for Quantum Technologies. He's keeping them well cared for in a big room on the 6th floor. This is just a thought experiment.

Suppose a mean nasty guy had a population of rabbits in a cage and didn't feed them at all. Suppose that each rabbit has a unit probability of dying per unit time. And as always, suppose the probability of this happening in any interval of time is independent of what happens before or after that time.

What is the Hamiltonian? Again there's a dumb guess: the annihilation operator! And again this guess is wrong, but it's not far off. As before, the right answer includes a 'correction term':

$$H = a - N$$

This time the correction term is famous in its own right. It's called the **number operator**:

$$N = a^\dagger a$$

The reason is that if we start with n rabbits, and apply this operator, it amounts to multiplication by n:

$$N z^n = z \frac{d}{dz} z^n = n z^n$$

Let's see why this guess is right. Again, suppose that at some particular time t we have n rabbits, so

$$\Psi(t) = z^n$$

Then the master equation says that at this time

$$\frac{d}{dt}\Psi(t) = (a - N)\Psi(t) = n z^{n-1} - n z^n$$

So, our probabilities are changing like this:

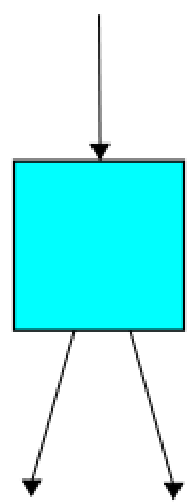
$$\frac{d}{dt}\psi_{n-1}(t) = n$$

$$\frac{d}{dt}\psi_n(t) = -n$$

while the rest have zero derivative. And this is good! We're starting with n rabbits, and each has a unit probability per unit time of dying. So, the chance of having one less should be going up at rate n. And the chance of having the same number we started with should be going *down* at the same rate.

Problem 7. Show that with this Hamiltonian and any initial conditions, the master equation predicts that the expected number of rabbits decays exponentially.

5.4 Breeding rabbits



Suppose we have a strange breed of rabbits that reproduce asexually. Suppose that each rabbit has a unit probability per unit time of having a baby rabbit, thus effectively duplicating itself.

As you can see from the cryptic picture above, this 'duplication' process takes one rabbit as input and has two rabbits as output. So, if you've been paying attention, you should be ready with a dumb guess for the Hamiltonian: $a^\dagger a^\dagger a$. This operator annihilates one rabbit and then creates two!

But you should also suspect that this dumb guess will need a 'correction term'. And you're right! As always, the correction terms makes the probability of things staying the same *go down* at exactly the rate that the probability of things changing *goes up*.

You should guess the correction term... but I'll just tell you:

$$H = a^\dagger a^\dagger a - N$$

We can check this in the usual way, by seeing what it does when we have n rabbits:

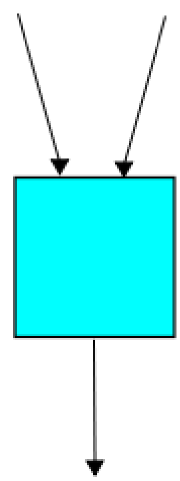
$$Hz^n = z^2 \frac{d}{dz} z^n - nz^n = nz^{n+1} - nz^n$$

That's good: since there are n rabbits, the rate of rabbit duplication is n. This is the rate at which the probability of having one more rabbit goes up... and also the rate at which the probability of having n rabbits goes down.

Problem 8. Show that with this Hamiltonian and any initial conditions, the master equation predicts that the expected number of rabbits grows exponentially.

5.5 Dueling rabbits

Let's do some stranger examples, just so you can see the general pattern.



Here each pair of rabbits has a unit probability per unit time of fighting a duel with only one survivor. You might guess the Hamiltonian $a^\dagger aa$, but in fact:

$$H = a^\dagger aa - N(N-1)$$

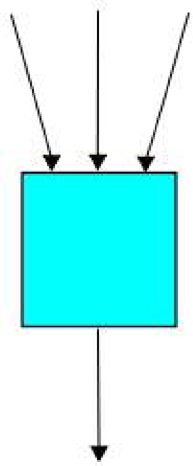
Let's see why this is right! Let's see what it does when we have n rabbits:

$$Hz^n = z\frac{d^2}{dz^2} z^n - n(n-1)z^n = n(n-1)z^{n-1} - n(n-1)z^n$$

That's good: since there are $n(n-1)$ ordered pairs of rabbits, the rate at which duels take place is $n(n-1)$. This is the rate at which the probability of having one less rabbit goes up... and also the rate at which the probability of having n rabbits goes down.

(If you prefer *unordered* pairs of rabbits, just divide the Hamiltonian by 2. We should talk about this more, but not now.)

5.6 Brawling rabbits



Now each *triple* of rabbits has a unit probability per unit time of getting into a fight with only one survivor! I don't know the technical term for a three-way fight, but perhaps it counts as a small 'brawl' or 'melee'. In fact the Wikipedia article for 'melee' shows three rabbits in suits of armor, fighting it out:

[PIC of rabbits fighting]

Now the Hamiltonian is:

$$H = a^\dagger a^3 - N(N-1)(N-2)$$

You can check that:

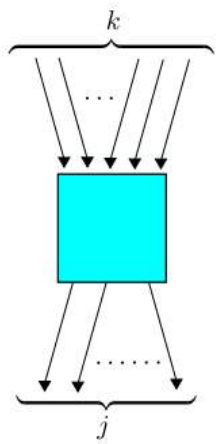
$$Hz^n = n(n-1)(n-2)z^{n-2} - n(n-1)(n-2)z^n$$

and this is good, because $n(n-1)(n-2)$ is the number of ordered triples of rabbits. You can see how this number shows up from the math, too:

$$a^3 z^n = \frac{d^3}{dz^3} z^n = n(n-1)(n-2)z^{n-3}$$

5.7 The general rule

Suppose we have a process taking k rabbits as input and having j rabbits as output:



I hope you can guess the Hamiltonian I'll use for this:

$$H = {a^\dagger}^j a^k - N(N-1)\cdots(N-k+1)$$

This works because

$$a^k z^n = \frac{d^k}{dz^k} z^n = n(n-1)\cdots(n-k+1) z^{n-k}$$

so that if we apply our Hamiltonian to n rabbits, we get

$$H z^n = n(n-1)\cdots(n-k+1)(z^{n+j-k} - z^n)$$

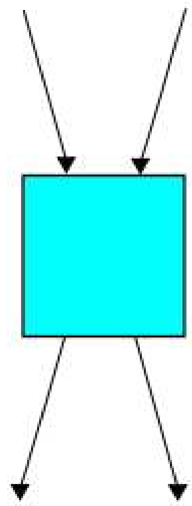
See? As the probability of having $n+j-k$ rabbits goes up, the probability of having n rabbits goes down, at an equal rate. This sort of balance is necessary for H to be a sensible Hamiltonian in this sort of stochastic theory (an 'infinitesimal stochastic operator', to be precise). And the rate is exactly the number of ordered k-tuples taken from a collection of n rabbits. This is called the kth **falling power** of n, and written as follows:

$$n^{\underline{k}} = n(n-1)\cdots(n-k+1)$$

Since we can apply functions to operators as well as numbers, we can write our Hamiltonian as:

$$H = {a^\dagger}^j a^k - N^{\underline{k}}$$

5.8 Kissing rabbits



Let's do one more example just to test our understanding. This time each pair of rabbits has a unit probability per unit time of bumping into each other, exchanging a friendly kiss and walking off. This shouldn't affect the rabbit population at all! But let's follow the rules and see what they say.

According to our rules, the Hamiltonian should be:

$$H = {a^\dagger}^2 a^2 - N(N-1)$$

However,

$${a^\dagger}^2 a^2 z^n = z^2 \frac{d^2}{dz^2} z^n = n(n-1) z^n = N(N-1) z^n$$

and since z^n form a 'basis' for the formal power series, we see that:

$${a^\dagger}^2 a^2 = N(N-1)$$

so in fact:

$$H = 0$$

That's good: if the Hamiltonian is zero, the master equation will say

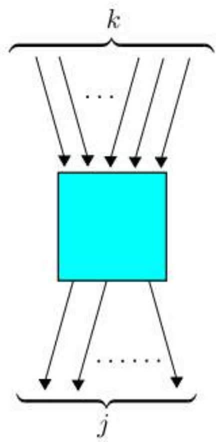
$$\frac{d}{dt} \Psi(t) = 0$$

so the population, or more precisely the probability of having any given number of rabbits, will be constant.

There's another nice little lesson here. Copying the calculation we just did, it's easy to see that:

$${a^\dagger}^k a^k = N^{\underline{k}}$$

This is a cute formula for falling powers of the number operator in terms of annihilation and creation operators. It means that for the general transition we saw before:



we can write the Hamiltonian in two equivalent ways:

$$H = {a^\dagger}^j a^k - N^{\underline{k}} = {a^\dagger}^j a^k - {a^\dagger}^k a^k$$

Okay, that's it for now! We can, and will, generalize all this stuff to stochastic Petri nets where there are things of many different kinds—not just rabbits. And we'll see that the master equation we get matches the answer to the problem in Part 4. That's pretty easy. But first, we'll have a guest post by Jacob Biamonte, who will explain a more realistic example from population biology.

5.9 References

There has been a lot of work on using annihilation and creation operators for stochastic systems. Here are some papers on the foundations:

- M. Doi, Second-quantization representation for classical many-particle systems, *Jour. Phys. A* **9** (1976), 1465–1477.

- M. Doi, Stochastic theory of diffusion-controlled reactions. *Jour. Phys. A* **9** (1976), 1479–1495.

- L. Peliti, Path integral approach to birth-death processes on a lattice, *Jour. de Physique* **46** 1469-83.

- J. Cardy, Renormalization group approach to reaction-diffusion problems. Available at arXiv:cond-mat/9607163.

- D. C. Mattis and M. L. Glasser, The uses of quantum field theory in diffusion-limited reactions, *Rev. of Mod. Phys.* **70** (1998), 979–1001.

Here are some papers on applications:

- U. Taüber, *et al.* Applications of field-theoretic renormalization group methods to reaction-diffusion problems, *Jour. Phys. A* **38** (2005), R79. Also available as arXiv:cond-mat/0501678.
- M. A. Buice, and J. D. Cowan, Field-theoretic approach to fluctuation effects in neural networks, *Phys. Rev. E* **75** (2007), 051919.
- M. A. Buice and J. D. Cowan, Statistical mechanics of the neocortex, *Prog. Biophysics & Mol. Bio.* **99** (2009), 53–86.
- P. J. Dodd and N. M. Ferguson, A many-body field theory approach to stochastic models in population biology, *PLoS ONE* **4** (2009), e6855.

5.10 Answers

Here are the answers to the problems:

Problem 6. Show that with the Hamiltonian

$$H = a^\dagger - 1$$

and any initial conditions, the master equation predicts that the expected number of rabbits grows linearly.

Answer. Here is one answer, thanks to David Corfield on Azimuth. If at some time t we have n rabbits, so that $\Psi(t) = z^n$, we have seen that the probability of having any number of rabbits changes as follows:

$$\frac{d}{dt}\psi_{n+1}(t) = 1, \qquad \frac{d}{dt}\psi_n(t) = -1, \qquad \frac{d}{dt}\psi_m(t) = 0 \mathrm{otherwise}$$

Thus the rate of increase of the expected number of rabbits is $(n+1) - n = 1$. But any probability distribution is a linear combination of these basis vector z^n, so the rate of increase of the expected number of rabbits is always

$$\sum_n \psi_n(t) = 1$$

so the expected number grows linearly.

Here is a second solution, using some more machinery. This machinery is overkill here, but it will be useful for solving the next two problems and also many other problems.

In the general formalism described in Section **??**, I used $\int \psi$ to mean the integral of a function over some measure space, so that probability distributions are the functions obeying

$$\int \psi = 1$$

and

$$\psi \geq 0$$

In the examples today this integral is really a sum over $n = 0, 1, 2, \ldots$, and it might be confusing to use integral notation since we're using derivatives for a completely different purpose. So let me define a sum notation as follows:

$$\sum \Psi = \sum_{n=0}^{\infty} \psi_n$$

This may be annoying, since after all we really have

$$\Psi|_{z=1} = \sum_{n=0}^{\infty} \psi_n$$

but please humor me.

To work with this sum notation, two rules are very handy.

Rule 1: For any formal power series Φ,

$$\sum a^\dagger \Phi = \sum \Phi$$

I mentioned this in the previous section: it's part of the creation operator being a stochastic operator. It's easy to check:

$$\sum a^\dagger \Phi = z\Phi|_{z=1} = \Phi|_{z=1} = \sum \Phi$$

Rule 2: For any formal power series Φ,

$$\sum a\Phi = \sum N\Phi$$

Again this is easy to check:

$$\sum N\Phi = \sum a^\dagger a\Phi = \sum a\Phi$$

These rules can be used together with the commutation relation

$$[a, a^\dagger] = 1$$

and its consequences

$$[a, N] = a, \qquad [a^\dagger, N] = -a^\dagger$$

to do many interesting things.

Let's see how! Suppose we have some observable O that we can write as an operator on formal power series: for example, the number operator, or any power of that. The expected value of this observable in the probability distribution Ψ is

$$\sum O\Psi$$

So, if we're trying to work out the time derivative of the expected value of O, we can start by using the master equation:

$$\frac{d}{dt}\sum O\Psi(t) = \sum O\frac{d}{dt}\Psi(t) = \sum OH\Psi(t)$$

Then we can write O and H using annihilation and creation operators and use our rules.

For example, in the problem at hand, we have

$$H = a^\dagger - 1$$

and the observable we're interested in is the number of rabbits

$$O = N$$

so we want to compute

$$\sum OH\Psi(t) = \sum N(a^\dagger - 1)\Psi(t)$$

There are many ways to use our rules to evaluate this. For example, Rule 1 implies that

$$\sum N(a^\dagger - 1)\Psi(t) = \sum a(a^\dagger - 1)\Psi(t)$$

but the commutation relations say $aa^\dagger = a^\dagger a + 1 = N + 1$, so

$$\sum a(a^\dagger - 1)\Psi(t) = \sum (N + 1 - a)\Psi(t)$$

and using Rule 1 again we see this equals

$$\sum \Psi(t) = 1$$

Thus we have

$$\frac{d}{dt}\sum N\Psi(t) = 1$$

It follows that the expected number of rabbits grows linearly:

$$\sum N\Psi(t) = t + c$$

Problem 7. Show that with the Hamiltonian

$$H = a - N$$

and any initial conditions, the master equation predicts that the expected number of rabbits decays exponentially.

Answer. We use the machinery developed in our answer to Problem 1. We want to compute the time derivative of the expected number of rabbits:

$$\frac{d}{dt}\sum N\Psi = \sum NH\Psi = \sum N(a-N)\Psi$$

The commutation relation $[a,N] = a$ implies that $Na = aN - a$. So:

$$\sum N(a-N)\Psi = \sum (aN - N - N^2)\Psi$$

but now Rule 2 says:

$$\sum (aN - N - N^2)\Psi = \sum (N^2 - N - N^2)\Psi = -\sum N\Psi$$

so we see

$$\frac{d}{dt}\sum N\Psi = -\sum N\Psi$$

It follows that the expected number of rabbits decreases exponentially:

$$\sum N\Psi(t) = ce^{-t}$$

Problem 8. Show that with the Hamiltonian

$$H = {a^\dagger}^2 a - N$$

and any initial conditions, the master equation predicts that the expected number of rabbits grows exponentially.

Answer. We use the same technique to compute

$$\frac{d}{dt}\sum N\Psi(t) = \sum NH\Psi(t) = \sum (N{a^\dagger}^2 a - N)\Psi(t)$$

First use the commutation relations to note:

$$N({a^\dagger}^2 a - N) = Na^\dagger N - N^2 = a^\dagger (N+1)N - N^2$$

Then:

$$\sum (a^\dagger (N+1)N - N^2)\Psi(t) = \sum ((N+1)N - N^2)\Psi(t) = \sum N\Psi(t)$$

So, we have

$$\frac{d}{dt}\sum N\Psi(t) = \sum N\Psi(t)$$

It follows that the expected number of rabbits grows exponentially:

$$\sum N\Psi(t) = ce^{t}$$

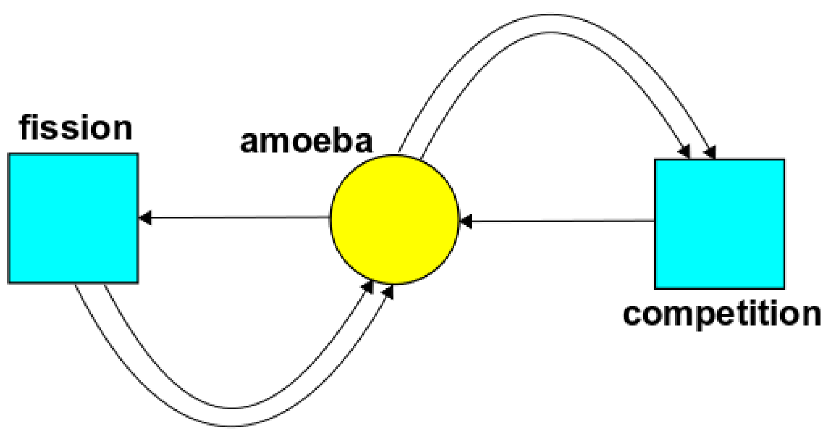
6 A stochastic Petri net from population biology

This post is part of a series on what John and I like to call *Petri net field theory* . Stochastic Petri nets can be used to model everything from vending machines to chemical reactions. Chemists have proven some powerful theorems about when these systems have equilibrium states. We're trying to bind these old ideas into our fancy framework, in hopes that quantum field theory techniques could also be useful in this deep subject. We'll describe the general theory later; today we'll do an example from population biology.

Those of you following this series should know that I'm the calculation bunny for this project, with John playing the role of the wolf. If I don't work quickly, drawing diagrams and trying to keep up with John's space-bending quasar of information, I'll be eaten alive! It's no joke, so please try to respond and pretend to enjoy anything you read here. This will keep me alive for longer. If I did not take notes during our meetings, lots of this stuff would have never made it here, so hope you enjoy.

6.1 Amoeba reproduction and competition

Here's a stochastic Petri net:



It shows a world with one state, **amoeba**, and two transitions:

- **reproduction**, where one amoeba turns into two. Let's call the rate constant for this transition α.
- **competition**, where two amoebas battle for resources and only one survives. Let's call the rate constant for this transition β.

We are going to analyse this example in several ways. First we'll study the *deterministic* dynamics it describes: we'll look at its rate equation, which turns out to be the logistic equation, familiar in population biology. Then we'll study the *stochastic* dynamics, meaning its master equation. That's where the ideas from quantum field theory come in.

6.2 The rate equation

If $P(t)$ is the population of amoebas at time t, we can follow the rules explained in Part 3 and crank out this **rate equation**:

$$\frac{dP}{dt} = \alpha P - \beta P^2$$

We can rewrite this as

$$\frac{dP}{dt} = kP(1 - \frac{P}{Q})$$

where

$$Q = \frac{\alpha}{\beta}, \qquad k = \alpha$$

What's the meaning of Q and k?

- Q is the **carrying capacity**, that is, the maximum sustainable population the environment can support.
- k is the **growth rate** describing the approximately exponential growth of population when $P(t)$ is small.

It's a rare treat to find such an important differential equation that can be solved by analytical methods. Let's enjoy solving it.

We start by separating variables and integrating both sides:

$$\int \frac{dP}{P(1 - P/Q)} = \int k dt$$

We need to use partial fractions on the left side above, resulting in

$$\int \frac{dP}{P} + \int \frac{dP}{Q - P} = \int k dt$$

and so we pick up a constant C, let $A = \pm e^{-C}$, and rearrange things as

$$\frac{Q - P}{P} = Ae^{-kt}$$

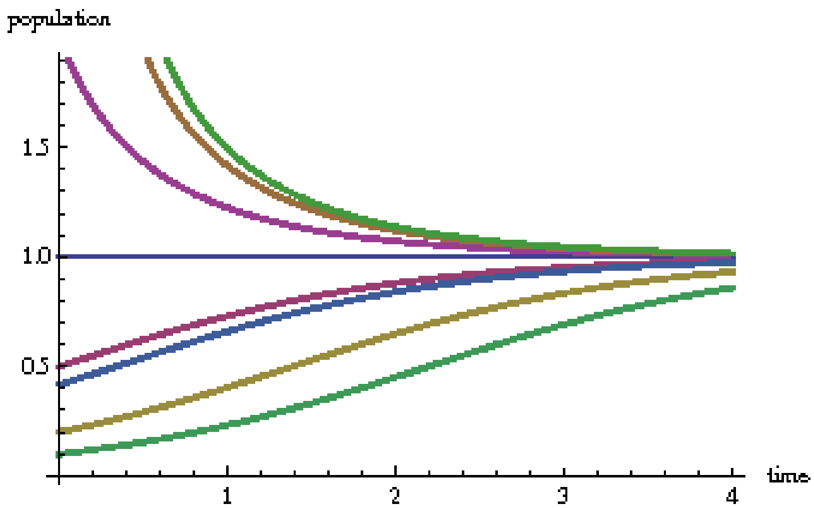
so the population as a function of time becomes

$$P(t) = \frac{Q}{1 + Ae^{-kt}}$$

At $t = 0$ we can determine A uniquely. We write $P_0 := P(0)$ and A becomes

$$A = \frac{Q - P_0}{P_0}$$

The model now becomes very intuitive. Let's set $Q = k = 1$ and make a plot for various values of A:



We arrive at three distinct cases:

- **equilibrium** ($A = 0$). The horizontal blue line corresponds to the case where the initial population P_0 exactly equals the carrying capacity. In this case the population is constant.
- **dieoff** ($A < 0$). The three decaying curves above the horizontal blue line correspond to cases where initial population is higher than the carrying capacity. The population dies off over time and approaches the carrying capacity.
- **growth** ($A > 0$). The four increasing curves below the horizontal blue line represent cases where the initial population is lower than the carrying capacity. Now the population grows over time and approaches the carrying capacity.

6.3 The master equation

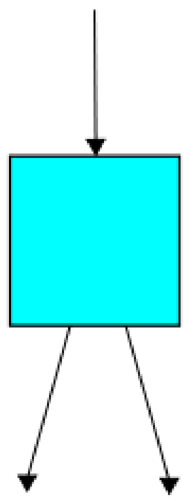
Next, let us follow the rules explained in Part 6 to write down the master equation for our example. Remember, now we write:

$$\Psi(t) = \sum_{n=0}^{\infty} \psi_n(t) z^n$$

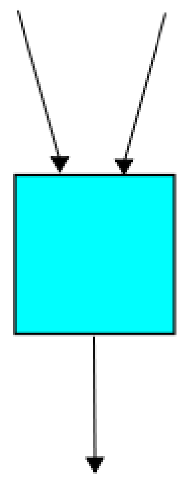
where $\psi_n(t)$ is the probability of having n amoebas at time t, and z is a formal variable. The **master equation** says:

$$\frac{d}{dt}\Psi(t) = H\Psi(t)$$

where H is an operator on formal power series called the **Hamiltonian**. To get the Hamiltonian we take each transition in our Petri net and build an operator built from creation and annihilation operators, as follows. Reproduction works like this:



while competition works like this:



Here a is the annihilation operator, $a^\dagger$ is the creation operator and $N = a^\dagger a$ is the number operator. Last time John explained precisely how the N's arise. So the theory is already in place, and we arrive at this Hamiltonian:

$$H = \alpha(a^\dagger a^\dagger a - N) + \beta(a^\dagger aa - N(N-1))$$

Remember, α is the rate constant for reproduction, while β is the rate constant for competition.

The master equation can be solved: it's equivalent to $\frac{d}{dt}(e^{-tH}\Psi(t)) = 0$ so that $e^{-tH}\Psi(t)$ is constant, and so

$$\Psi(t) = e^{tH}\Psi(0)$$

and that's it! We can calculate the time evolution starting from any initial probability distribution of populations. Maybe everyone is already used to this, but I find it rather remarkable.

Here's how it works. We pick a population, say n amoebas at $t = 0$. This would mean $\Psi(0) = z^n$. We then evolve this state using e^{tH}. We expand this operator as

$$\begin{aligned} e^{tH} &= \textstyle\sum_{n=0}^{\infty} \frac{1}{n!} t^n H^n \\ &= 1 + tH + \frac{1}{2}t^2H^2 + \cdots \end{aligned}$$

This operator contains the full information for the evolution of the system. It contains the *histories* of all possible amoeba populations—an amoeba mosaic if you will. From this, we can construct amoeba Feynman diagrams.

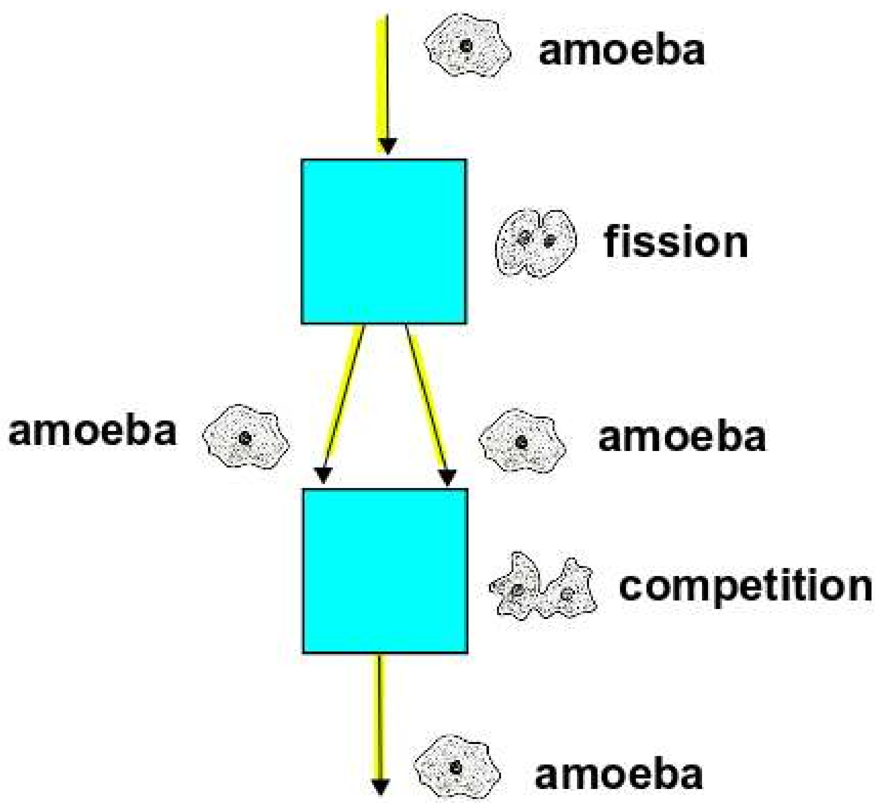
To do this, we work out each of the H^n terms in the expansion above. The first-order terms correspond to the Hamiltonian acting once. These are proportional to either α or β. The second-order terms correspond to the Hamiltonian acting twice. These are proportional to either α^2, $\alpha\beta$ or β^2. And so on.

This is where things start to get interesting! To illustrate how it works, we will consider two possibilities for the second-order terms:

1) We start with a lone amoeba, so $\Psi(0) = z$. It reproduces and splits into two. In the battle of the century, the resulting amoebas compete and one dies. At the end we have:

$$\frac{\alpha\beta}{2}(a^\dagger aa)(a^\dagger a^\dagger a)z$$

We can draw this as a Feynman diagram:

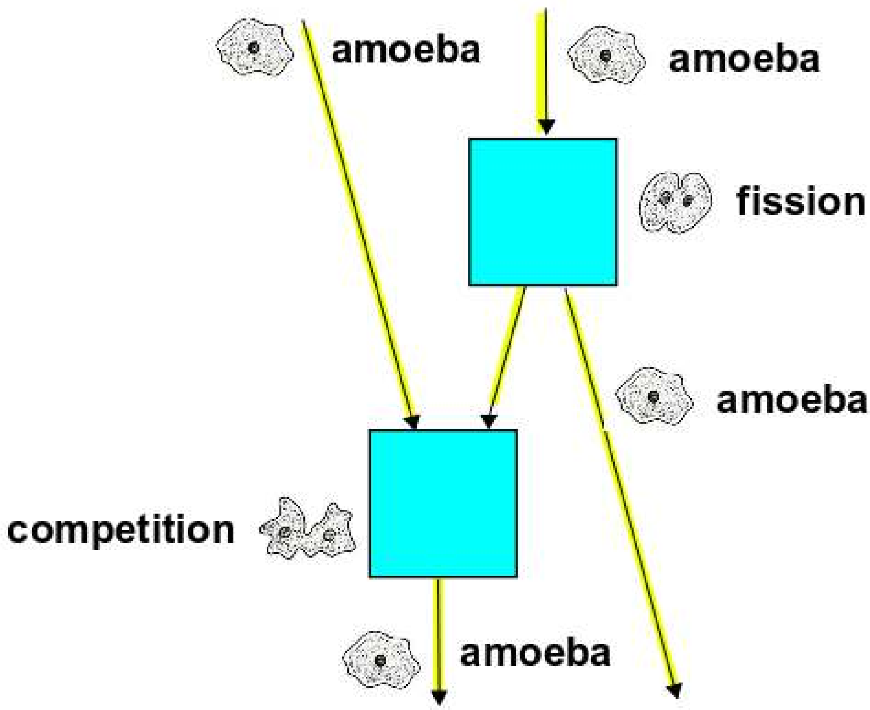


You might find this tale grim, and you may not like the odds either. It's true, the odds could be better, but people are worse off than amoebas! The great Japanese swordsman Miyamoto Musashi quoted the survival odds of fair sword duels as 1/3, seeing that 1/3 of the time both participants die. A remedy is to cheat, but these amoeba are competing *honestly* .

2) We start with two amoebas, so the initial state is $\Psi(0) = z^2$. One of these amoebas splits into two. One of these then gets into an argument with the original amoeba over the Azimuth blog. The amoeba who solved all John's problems survives. At the end we have

$$\frac{\alpha\beta}{2}(a^\dagger aa)(a^\dagger a^\dagger a)z^2$$

with corresponding Feynman diagram:



This should give an idea of how this all works. The exponential of the Hamiltonian gives all possible histories, and each of these can be translated into a Feynman diagram. In a future blog entry, we might explain this theory in detail.

6.4 An equilibrium state

We've seen the equilibrium solution for the rate equation; now let's look for equilibrium solutions of the master equation. This paper:

- D. F. Anderson, G. Craciun and T.G. Kurtz, Product-form stationary distributions for deficiency zero chemical reaction networks, arXiv:0803.3042. [?]

proves that for a large class of stochastic Petri nets, there exists an equilibrium solution of the master equation where the number of things in each state is distributed according to a Poisson distribution. Even more remarkably, these probability distributions are *independent* , so knowing how many things are in one state tells you nothing about how many are in another!

Here's a nice quote from this paper:

- The surprising aspect of the deficiency zero theorem is that the assumptions of the theorem are completely related to the network of the system whereas the conclusions of the theorem are related to the dynamical properties of the system.

The 'deficiency zero theorem' is a result of Feinberg, which says that for a large class of stochastic Petri nets, the rate equation has a unique equilibrium solution. Anderson showed how to use this fact to get equilibrium solutions of the master equation!

We will consider this in future posts. For now, we need to talk a bit about 'coherent states'.

These are all over the place in quantum theory. Legend (or at least Wikipedia) has it that Erwin Schrödinger himself discovered coherent states when he was looking for states of a quantum system that look 'as classical as possible'. Suppose you have a quantum harmonic oscillator. Then the uncertainty principle says that

$$\Delta p \Delta q \geq \hbar/2$$

where Δp is the uncertainty in the momentum and Δq is the uncertainty in position. Suppose we want to make $\Delta p \Delta q$ as small as possible, and suppose we also want $\Delta p = \Delta q$. Then we need our particle to be in a 'coherent state'. That's the definition. For the quantum harmonic oscillator, there's a way to write quantum states as formal power series

$$\Psi = \sum_{n=0}^{\infty} \psi_n z^n$$

where ψ_n is the amplitude for having n quanta of energy. A coherent state then looks like this:

$$\Psi = e^{cz} = \sum_{n=0}^{\infty} \frac{c^n}{n!} z^n$$

where c can be any complex number. Here we have omitted a constant factor necessary to normalize the state.

We can also use coherent states in classical stochastic systems like collections of amoebas! Now the coefficient of z^n tells us the probability of having n amoebas, so c had better be real. And probabilities should sum to 1, so we really should normalize Ψ as follows:

$$\Psi = \frac{e^{cz}}{e^c} = e^{-c} \sum_{n=0}^{\infty} \frac{c^n}{n!} z^n$$

Now, the probability distribution

$$\psi_n = e^{-c} \frac{c^n}{n!}$$

is called a **Poisson distribution**. So, for starters you can think of a 'coherent state' as an over-educated way of talking about a Poisson distribution.

Let's work out the expected number of amoebas in this Poisson distribution. In the answers to the problems in Part 6, we started using this abbreviation:

$$\sum \Psi = \sum_{n=0}^{\infty} \psi_n$$

We also saw that the expected number of amoebas in the probability distribution Ψ is

$$\sum N\Psi$$

What does this equal? Remember that $N = a^\dagger a$. The annihilation operator a is just $\frac{d}{dz}$, so

$$a\Psi = c\Psi$$

and we get

$$\sum N\Psi = \sum a^\dagger a\Psi = c \sum a^\dagger \Psi$$

But we saw in Part 5 that $a^\dagger$ is stochastic, meaning

$$\sum a^\dagger \Psi = \sum \Psi$$

for any Ψ. Furthermore, our Ψ here has

$$\sum \Psi = 1$$

since it's a probability distribution. So:

$$\sum N\Psi = c \sum a^\dagger \Psi = c \sum \Psi = c$$

The expected number of amoebas is just c.

Problem 9. This calculation must be wrong if c is negative: there can't be a negative number of amoebas. What goes wrong then?

Problem 10. Use the same tricks to calculate the standard deviation of the number of amoebas in the Poisson distribution Ψ.

Now let's return to our problem and consider the initial amoeba state

$$\Psi = e^{cz}$$

Here aren't bothering to normalize it, because we're going to look for equilibrium solutions to the master equation, meaning solutions where $\Psi(t)$ doesn't change with time. So, we want to solve

$$H\Psi = 0$$

Since this equation is linear, the normalization of Ψ doesn't really matter.

Remember,

$$H\Psi = \alpha(a^\dagger a^\dagger a - N)\Psi + \beta(a^\dagger aa - N(N-1))\Psi$$

Let's work this out. First consider the two α terms:

$$a^\dagger a^\dagger a\Psi = cz^2\Psi$$

and

$$-N\Psi = -a^\dagger a\Psi = -cz\Psi$$

Likewise for the β terms we find

$$a^\dagger aa\Psi = c^2 z\Psi$$

and

$$-N(N-1)\psi = -a^\dagger a^\dagger aa\Psi = -c^2 z^2\Psi$$

Here I'm using something John showed in Part 6: the product ${a^\dagger}^2 a^2$ equals the 'falling power' $N(N-1)$.

The sum of all four terms must vanish. This happens whenever

$$\alpha(cz^2 - cz) + \beta(c^2 z - c^2 z^2) = 0$$

which is satisfied for

$$c = \frac{\alpha}{\beta}$$

Yipee! We've found an equilibrium solution, since we found a value for c that makes $H\Psi = 0$. Even better, we've seen that the expected number of amoebas in this equilibrium state is

$$\frac{\alpha}{\beta}$$

This is just the same as the equilibrium population we saw in the *rate equation* —that is, the logistic equation! That's pretty cool, but it's no coincidence: in fact, Anderson proved it works like this for lots of stochastic Petri nets.

I'm not sure what's up next or what's in store, since I'm blogging at gun point from inside a rabbit cage:

[Silly pic]

I'd imagine we're going to work out the theory behind this example and prove the existence of equilibrium solutions for master equations more generally. One idea John had was to have me start a night shift—that way you'll get Azimuth posts 24 hours a day.

6.5 Answer

Here are the answers to the problems, provided by David Corfield:

Problem 9. We calculated that the expected number of amoebas in the Poisson distribution with parameter c is equal to c. But this can't be true if c is negative: there can't be a negative number of amoebas. What goes wrong then?

Answer. If the probability of having n amoebas is given by the Poisson distribution

$$\psi_n = e^{-c}\frac{c^n}{n!}$$

then c had better be nonnegative for the probability to be negative when $c = 1$.

Problem 10. Calculate the standard deviation of the number of amoebas in the Poisson distribution.

Answer. The standard deviation is the square root of the variance, which is

$$\sum N^2\Psi - (\sum N\Psi)^2$$

We have seen that for the Poisson distribution,

$$\sum N\Psi = c$$

and using the same tricks we see

$$\begin{array}{rcl} \sum N^2\Psi &=& \sum a^\dagger a a^\dagger a\Psi \\ &=& c\sum a^\dagger a a^\dagger \Psi \\ &=& c\sum a a^\dagger \Psi \\ &=& c\sum (a^\dagger a + 1)\Psi \\ &=& c(c+1) \end{array}$$

So, the variance is $c(c+1) - c^2 = c$ and the standard deviation is $\sqrt{c}$.

7 From stochastic Petri nets to Feynman diagrams

Summer vacation is over. Time to get back to work!

This month, before he goes to Oxford to begin a master's program in Mathematics and the Foundations of Computer Science, Brendan Fong is visiting the Centre for Quantum Technologies and working with me on stochastic Petri nets. He's proved two interesting results, which he wants to explain.

To understand what he's done, you need to know how to get the rate equation and the master equation from a stochastic Petri net. We've *almost* seen how. But it's been a long time since the last article in this series, so today I'll start with some review. And at the end, just for fun, I'll say a bit more about how Feynman diagrams show up in this theory.

Since I'm an experienced teacher, I'll assume you've forgotten *everything I ever said* .

(This has some advantages. I can change some of my earlier terminology—improve it a bit here and there—and you won't even notice.)

7.1 Stochastic Petri nets

Definition. A **Petri net** consists of a set S of **species** and a set T of **transitions**, together with a function

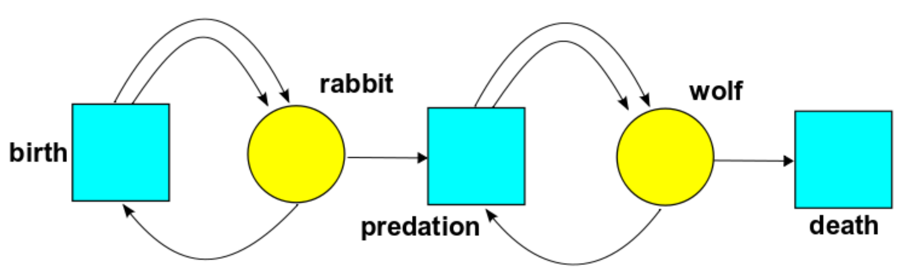
$$i : S \times T \to \mathbb{N}$$

saying how many things of each species appear in the **input** for each transition, and a function

$$o : S \times T \to \mathbb{N}$$

saying how many things of each species appear in the **output**.

We can draw pictures of Petri nets. For example, here's a Petri net with two species and three transitions:



It should be clear that the transition 'predation' has one wolf and one rabbit as input, and two wolves as output.

A 'stochastic' Petri net goes further: it also says the rate at which each transition occurs.

Definition 3. *A* **stochastic Petri net** *is a Petri net together with a function*

$$r : T \to [0, \infty)$$

giving a **rate constant** *for each transition.*

7.2 Master equation versus rate equation

Starting from any stochastic Petri net, we can get two things. First:

- The **master equation**. This says how the *probability that we have a given number of things of each species* changes with time.
- The **rate equation**. This says how the *expected number of things of each species* changes with time.

The master equation is stochastic: it describes how probabilities change with time. The rate equation is deterministic.

The master equation is more fundamental. It's like the equations of quantum electrodynamics, which describe the amplitudes for creating and annihilating individual photons. The rate equation is less fundamental. It's like the classical Maxwell equations, which describe changes in the electromagnetic field in a deterministic way. The classical Maxwell equations are an approximation to quantum electrodynamics. This approximation gets good *in the limit* where there are lots of photons all piling on top of each other to form nice waves.

Similarly, the rate equation can be derived from the master equation *in the limit* where the number of things of each species become large, and the fluctuations in these numbers become negligible.

But I won't do this derivation today! Nor will I probe more deeply into the analogy with quantum field theory, even though that's my ultimate goal. Today I'm content to remind you what the master equation and rate equation *are* .

The rate equation is simpler, so let's do that first.

7.3 The Rate Equation

Suppose we have a stochastic Petri net with k different species. Let x_i be the number of things of the ith species. Then the rate equation looks like this:

$$\frac{dx_i}{dt} = ???$$

It's really a bunch of equations, one for each $1 \leq i \leq k$. But what is the right-hand side?

The right-hand side is a sum of terms, one for each transition in our Petri net. So, let's start by assuming our Petri net has just one transition.

Suppose the ith species appears as input to this transition m_i times, and as output n_i times. Then the rate equation is

$$\frac{dx_i}{dt} = r(n_i - m_i) x_1^{m_1} \cdots x_k^{m_k}$$

where r is the rate constant for this transition.

That's really all there is to it! But we can make it look nicer. Let's make up a vector

$$x = (x_1, \ldots, x_k) \in [0, \infty)^k$$

that says how many things there are of each species. Similarly let's make up an **input vector**

$$m = (m_1, \ldots, m_k) \in \mathbb{N}^k$$

and an **output vector**

$$n = (n_1, \ldots, n_k) \in \mathbb{N}^k$$

for our transition. To be cute, let's also define

$$x^m = x_1^{m_1} \cdots x_k^{m_k}$$

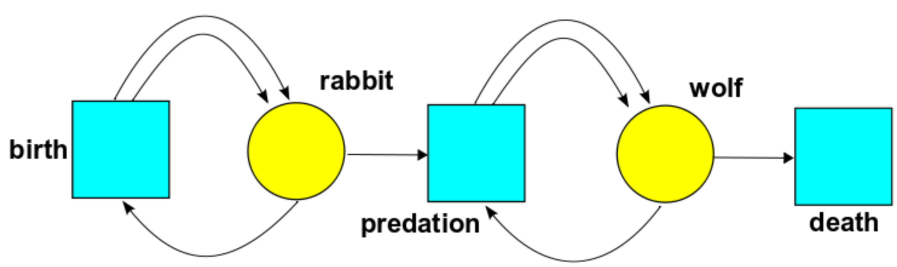
Then we can write the rate equation for a single transition like this:

$$\frac{dx}{dt} = r(n - m)x^m$$

Next let's do a general stochastic Petri net, with lots of transitions. Let's write T for the set of transitions and $r(\tau)$ for the rate constant of the transition $\tau \in T$. Let $n(\tau)$ and $m(\tau)$ be the input and output vectors of the transition τ. Then the rate equation is:

$$\frac{dx}{dt} = \sum_{\tau \in T} r(\tau)\, (n(\tau) - m(\tau))\, x^{m(\tau)}$$

For example, consider our rabbits and wolves:



Suppose

- the rate constant for 'birth' is β,
- the rate constant for 'predation' is γ,
- the rate constant for 'death' is δ.

Let $x_1(t)$ be the number of rabbits and $x_2(t)$ the number of wolves at time t. Then the rate equation looks like this:

$$\frac{dx_1}{dt} = \beta x_1 - \gamma x_1 x_2$$
$$\frac{dx_2}{dt} = \gamma x_1 x_2 - \delta x_2$$

If you stare at this, and think about it, it should make perfect sense. If it doesn't, go back and read Part 2.

7.4 The master equation

Now let's do something new. In Part 6 I explained how to write down the master equation for a stochastic Petri net with just *one* species. Now let's generalize that. Luckily, the ideas are exactly the same.

So, suppose we have a stochastic Petri net with k different species. Let $\psi_{n_1,\dots,n_k}$ be the probability that we have n_1 things of the first species, n_2 of the second species, and so on. The master equation will say how all these probabilities change with time.

To keep the notation clean, let's introduce a vector

$$n = (n_1, \dots, n_k) \in \mathbb{N}^k$$

and let

$$\psi_n = \psi_{n_1,\dots,n_k}$$

Then, let's take all these probabilities and cook up a formal power series that has them as coefficients: as we've seen, this is a powerful trick. To do this, we'll bring in some variables $z_1, \dots, z_k$ and write

$$z^n = z_1^{n_1} \cdots z_k^{n_k}$$

as a convenient abbreviation. Then any formal power series in these variables looks like this:

$$\Psi = \sum_{n \in \mathbb{N}^k} \psi_n z^n$$

We call Ψ a **state** if the probabilities sum to 1 as they should:

$$\sum_n \psi_n = 1$$

The simplest example of a state is a monomial:

$$z^n = z_1^{n_1} \cdots z_k^{n_k}$$

This is a state where we are 100% sure that there are n_1 things of the first species, n_2 of the second species, and so on. We call such a state a **pure state**, since physicists use this term to describe a state where we know for sure exactly what's going on. Sometimes a general state, one that might not be pure, is called **mixed**.

The master equation says how a state evolves in time. It looks like this:

$$\frac{d}{dt}\Psi(t) = H\Psi(t)$$

So, I just need to tell you what H is!

It's called the **Hamiltonian**. It's a linear operator built from special operators that annihilate and create things of various species. Namely, for each state $1 \leq i \leq k$ we have a **annihilation operator**:

$$a_i\Psi = \frac{d}{dz_i}\Psi$$

and a **creation operator**:

$$a_i^\dagger\Psi = z_i\Psi$$

How do we build H from these? Suppose we've got a stochastic Petri net whose set of transitions is T. As before, write $r(\tau)$ for the rate constant of the transition $\tau \in T$, and let $n(\tau)$ and $m(\tau)$ be the input and output vectors of this transition. Then:

$$H = \sum_{\tau \in T} r(\tau)\,({a^\dagger}^{n(\tau)} - {a^\dagger}^{m(\tau)})\,a^{m(\tau)}$$

where as usual we've introduce some shorthand notations to keep from going insane. For example:

$$a^{m(\tau)} = a_1^{m_1(\tau)} \cdots a_k^{m_k(\tau)}$$

and

$${a^\dagger}^{m(\tau)} = {a_1^\dagger}^{m_1(\tau)} \cdots {a_k^\dagger}^{m_k(\tau)}$$

Now, it's not surprising that each transition τ contributes a term to H. It's also not surprising that this term is proportional to the rate constant $r(\tau)$. The only tricky thing is the expression

$$({a^\dagger}^{n(\tau)} - {a^\dagger}^{m(\tau)})a^{m(\tau)}$$

How can we understand it? The basic idea is this. We've got two terms here. The first term:

$${a^\dagger}^{n(\tau)}a^{m(\tau)}$$

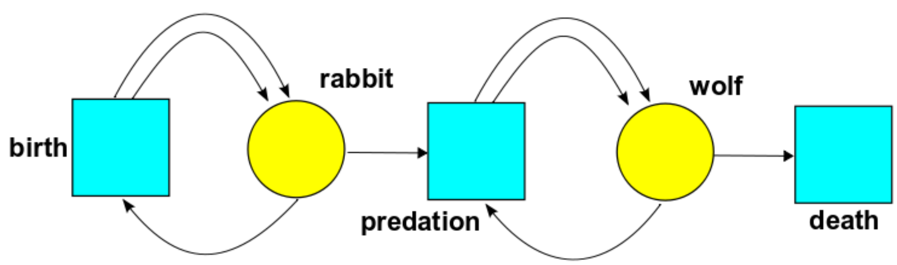
describes how $m_i(\tau)$ things of the ith species get annihilated, and $n_i(\tau)$ things of the ith species get created. Of course this happens thanks to our transition τ. The second term:

$$-a^{\dagger m(\tau)}a^{m(\tau)}$$

is a bit harder to understand, but it says how the probability that *nothing* happens—that we remain in the same pure state— *decreases* as time passes. Again this happens due to our transition τ.

In fact, the second term must take precisely the form it does to ensure 'conservation of total probability'. In other words: if the probabilities ψ_n sum to 1 at time zero, we want these probabilities to still sum to 1 at any later time. And for this, we need that second term to be what it is! In Part 6 we saw this in the special case where there's only one species. The general case works the same way.

Let's look at an example. Consider our rabbits and wolves yet again:



and again suppose the rate constants for birth, predation and death are β, γ and δ, respectively. We have

$$\Psi = \sum_n \psi_n z^n$$

where

$$z^n = z_1^{n_1} z_2^{n_2}$$

and $\psi_n = \psi_{n_1,n_2}$ is the probability of having n_1 rabbits and n_2 wolves. These probabilities evolve according to the equation

$$\frac{d}{dt}\Psi(t) = H\Psi(t)$$

where the Hamiltonian is

$$H = \beta B + \gamma C + \delta D$$

and B, C and D are operators describing birth, predation and death, respectively. (B stands for birth, D stands for death... and you can call predation 'consumption' if you want something that starts with C. Besides, 'consumer' is a nice euphemism for 'predator'.) What are these operators? Just follow the rules I described:

$$B = {a_1^\dagger}^2 a_1 - a_1^\dagger a_1$$

$$C = {a_2^\dagger}^2 a_1 a_2 - a_1^\dagger a_2^\dagger a_1 a_2$$

$$D = a_2 - a_2^\dagger a_2$$

In each case, the first term is easy to understand:

- Birth annihilates one rabbit and creates two rabbits.
- Predation annihilates one rabbit and one wolf and creates two wolves.
- Death annihilates one wolf.

The second term is trickier, but I told you how it works.

7.5 Feynman diagrams

How do we solve the master equation? If we don't worry about mathematical rigor too much, it's easy. The solution of

$$\frac{d}{dt}\Psi(t) = H\Psi(t)$$

should be

$$\Psi(t) = e^{tH}\Psi(0)$$

and we can hope that

$$e^{tH} = 1 + tH + \frac{(tH)^2}{2!} + \cdots$$

so that

$$\Psi(t) = \Psi(0) + tH\Psi(0) + \frac{t^2}{2!}H^2\Psi(0) + \cdots$$

Of course there's always the question of whether this power series converges. In many contexts it doesn't, but that's not necessarily a disaster: the series can still be asymptotic to the right answer, or even better, Borel summable to the right answer.

But let's not worry about these subtleties yet! Let's just imagine our rabbits and wolves, with Hamiltonian

$$H = \beta B + \gamma C + \delta D$$

Now, imagine working out

$$\Psi(t) = \Psi(0) + tH\Psi(0) + \frac{t^2}{2!}H^2\Psi(0) + \frac{t^3}{3!}H^3\Psi(0) + \cdots$$

We'll get lots of terms involving products of B, C and D hitting our original state $\Psi(0)$. And we can draw these as diagrams! For example, suppose we start with one rabbit and one wolf. Then

$$\Psi(0) = z_1 z_2$$

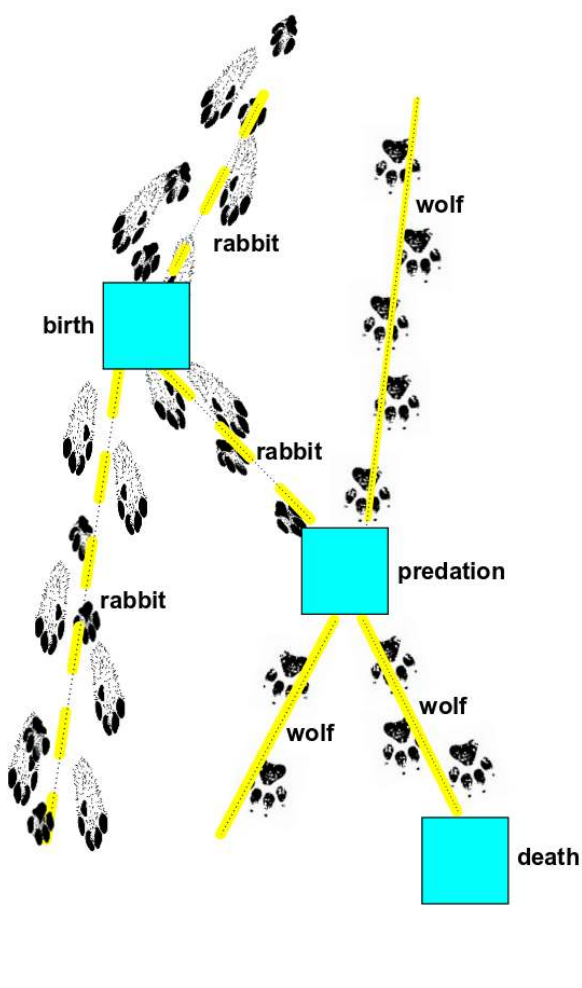
And suppose we want to compute

$$H^3\Psi(0) = (\beta B + \gamma C + \delta D)^3 \Psi(0)$$

as part of the task of computing $\Psi(t)$. Then we'll get lots of terms: 27, in fact, though many will turn out to be zero. Let's take one of these terms, for example the one proportional to:

$$DCB\Psi(0)$$

We can draw this as a sum of Feynman diagrams, including this:



In this diagram, we start with one rabbit and one wolf at top. As we read the diagram from top to bottom, first a rabbit is born (B), then predation occur (C), and finally a wolf dies (D). The end result is again a rabbit and a wolf.

This is just one of four Feynman diagrams we should draw in our sum for $DCB\Psi(0)$, since either of the two rabbits could have been eaten, and either wolf could have died. So, the end result of computing

$$H^3\Psi(0)$$

will involve a lot of Feynman diagrams... and of course computing

$$\Psi(t) = \Psi(0) + tH\Psi(0) + \frac{t^2}{2!}H^2\Psi(0) + \frac{t^3}{3!}H^3\Psi(0) + \cdots$$

will involve even more, even if we get tired and give up after the first few terms. So, this Feynman diagram business may seem quite tedious... and it may not be obvious how it helps.

But it does, sometimes!

Now is not the time for me to describe 'practical' benefits of Feynman diagrams. Instead, I'll just point out one conceptual benefit. We started with what seemed like a purely computational chore, namely computing

$$\Psi(t) = \Psi(0) + tH\Psi(0) + \frac{t^2}{2!}H^2\Psi(0) + \cdots$$

But then we saw—at least roughly—how this series has a clear meaning! It can be written as a sum over diagrams, each of which represents a *possible history* of rabbits and wolves. So, it's what physicists call a 'sum over histories'.

Feynman invented the idea of a sum over histories in the context of quantum field theory. At the time this idea seemed quite mind-blowing, for various reasons. First, it involved elementary particles instead of everyday things like rabbits and wolves. Second, it involved complex 'amplitudes' instead of real probabilities. Third, it actually involved integrals instead of sums. And fourth, a lot of these integrals diverged, giving infinite answers that needed to be 'cured' somehow!

Now we're seeing a sum over histories in a more down-to-earth context without all these complications. A lot of the underlying math is analogous... but now there's nothing mind-blowing about it: it's quite easy to understand. So, we can use this analogy to demystify quantum field theory a bit. On the other hand, thanks to this analogy, all sorts of clever ideas invented by quantum field theorists will turn out to have applications to biology and chemistry! So it's a double win.

8 The Anderson–Craciun–Kurtz theorem

Last time we reviewed the rate equation and the master equation. Both of them describe processes where things of various kinds can react and turn into other things. But:

- In the rate equation, we assume the number of things varies continuously and is known precisely.
- In the master equation, we assume the number of things varies discretely and is known only probabilistically.

This should remind you of the difference between classical mechanics and quantum mechanics. But the master equation is not *quantum* , it's *stochastic* : it involves probabilities, but there's no uncertainty principle going on.

Still, a lot of the math is similar.

Now, given an equilibrium solution to the rate equation—one that doesn't change with time—we'll try to find a solution to the master equation with the same property. We won't *always* succeed—but we often can! The theorem saying how was proved here:

- [?] — D. F. Anderson, G. Craciun and T. G. Kurtz, Product-form stationary distributions for deficiency zero chemical reaction networks.

To emphasize the analogy to quantum mechanics, we'll translate their proof into the language of annihilation and creation operators. In particular, our equilibrium solution of the master equation is just like what people call a 'coherent state' in quantum mechanics.

So, if you know about quantum mechanics and coherent states, you should be happy. But if you don't, fear not!—we're not assuming you do.

8.1 The rate equation

To construct our equilibrium solution of the master equation, we need a special type of solution to our rate equation. We call this type a 'complex balanced solution'. This means that not only is the net rate of production of each species zero, but the net rate of production of each possible *bunch* of species is zero.

Before we make this more precise, let's remind ourselves of the basic setup.

We'll consider a stochastic Petri net with a finite set S of species and a finite set T of transitions. For convenience let's take $S = \{1, \ldots, k\}$, so our species are numbered from 1 to k. Then each transition τ has an input vector $m(\tau) \in \mathbb{N}^k$ and output vector $n(\tau) \in \mathbb{N}^k$. These say how many things of each species go in, and how many go out. Each transition also has rate constant $r(\tau) \in [0, \infty)$, which says how rapidly it happens.

The rate equation concerns a vector $x(t) \in [0, \infty)^k$ whose ith component is the number of things of the ith species at time t. Note: we're assuming this number of things varies continuously and is known precisely! This should remind you of classical mechanics. So, we'll call $x(t)$, or indeed any vector in $[0, \infty)^k$, a **classical state**.

The **rate equation** says how the classical state $x(t)$ changes with time:

$$\frac{dx}{dt} = \sum_{\tau \in T} r(\tau) \left(n(\tau) - m(\tau)\right) x^{m(\tau)}$$

You may wonder what $x^{m(\tau)}$ means: after all, we're taking a vector to a vector power! It's just an abbreviation, which we've seen plenty of times before. If $x \in \mathbb{R}^k$ is a list of numbers and $m \in \mathbb{N}^k$ is a list of natural numbers, we define

$$x^m = x_1^{m_1} \cdots x_k^{m_k}$$

We'll also use this notation when x is a list of *operators* .

8.2 Complex balance

The vectors $m(\tau)$ and $n(\tau)$ are examples of what chemists call **complexes**. A complex is a bunch of things of each species. For example, if the set S consists of three species, the complex $(1, 0, 5)$ is a bunch consisting of one thing of the first species, none of the second species, and five of the third species.

For our Petri net, the set of complexes is the set $\mathbb{N}^k$, and the complexes of particular interest are the **input complex** $m(\tau)$ and the **output complex** $n(\tau)$ of each transition τ.

We say a classical state $c \in [0, \infty)^k$ is **complex balanced** if for all complexes $\kappa \in \mathbb{N}^k$ we have

$$\sum_{\{\tau : m(\tau) = \kappa\}} r(\tau) c^{m(\tau)} = \sum_{\{\tau : n(\tau) = \kappa\}} r(\tau) c^{m(\tau)}$$

The left hand side of this equation, which sums over the transitions with *input complex* κ, gives the rate of consumption of the complex κ. The right hand side, which sums over the transitions with *output complex* κ, gives the rate of production of κ. So, this equation requires that the net rate of production of the complex κ is zero in the classical state c.

Problem 11. Show that if a classical state c is complex balanced, and we set $x(t) = c$ for all t, then $x(t)$ is a solution of the rate equation.

Since $x(t)$ doesn't change with time here, we call it an **equilibrium solution** of the rate equation. Since $x(t) = c$ is complex balanced, we call it **complex balanced** equilibrium solution.

8.3 The master equation

We've seen that any complex balanced classical state gives an equilibrium solution of the *rate* equation. The Anderson–Craciun–Kurtz theorem says that it also gives an equilibrium solution of the *master* equation.

The master equation concerns a formal power series

$$\Psi(t) = \sum_{n \in \mathbb{N}^k} \psi_n(t) z^n$$

where

$$z^n = z_1^{n_1} \cdots z_k^{n_k}$$

and

$$\psi_n(t) = \psi_{n_1,\ldots,n_k}(t)$$

is the probability that at time t we have n_1 things of the first species, n_2 of the second species, and so on.

Note: now we're assuming this number of things varies discretely and is known only probabilistically! So, we'll call $\Psi(t)$, or indeed any formal power series where the coefficients are probabilities summing to 1, a **stochastic state**. Earlier we just called it a 'state', but that would get confusing now: we've got classical states and stochastic states, and we're trying to relate them.

The **master equation** says how the stochastic state $\Psi(t)$ changes with time:

$$\frac{d}{dt}\Psi(t) = H\Psi(t)$$

where the **Hamiltonian** H is:

$$H = \sum_{\tau \in T} r(\tau) \left({a^\dagger}^{n(\tau)} - {a^\dagger}^{m(\tau)} \right) a^{m(\tau)}$$

The notation here is designed to neatly summarize some big products of annihilation and creation operators. For any vector $n \in \mathbb{N}^k$, we have

$$a^n = a_1^{n_1} \cdots a_k^{n_k}$$

and

$${a^\dagger}^n = {a_1^\dagger}^{n_1} \cdots {a_k^\dagger}^{n_k}$$

8.4 Coherent states

Now suppose $c \in [0,\infty)^k$ is a complex balanced equilibrium solution of the rate equation. We want to get an equilibrium solution of the master equation. How do we do it?

For any $c \in [0,\infty)^k$ there is a stochastic state called a **coherent state**, defined by

$$\Psi_c = \frac{e^{cz}}{e^c}$$

Here we are using some very terse abbreviations. Namely, we are defining

$$e^c = e^{c_1} \cdots e^{c_k}$$

and

$$e^{cz} = e^{c_1 z_1} \cdots e^{c_k z_k}$$

Equivalently,

$$e^{cz} = \sum_{n \in \mathbb{N}^k} \frac{c^n}{n!} z^n$$

where c^n and z^n are defined as products in our usual way, and

$$n! = n_1! \cdots n_k!$$

Either way, if you unravel the abbrevations, here's what you get:

$$\Psi_c = e^{-(c_1 + \cdots + c_k)} \sum_{n \in \mathbb{N}^k} \frac{c_1^{n_1} \cdots c_k^{n_k}}{n_1! \cdots n_k!} z_1^{n_1} \cdots z_k^{n_k}$$

Maybe now you see why we like the abbreviations.

The name 'coherent state' comes from quantum mechanics. In quantum mechanics, we think of a coherent state Ψ_c as the 'quantum state' that best approximates the classical state c. But we're not doing quantum mechanics now, we're doing probability theory. Ψ_c isn't a 'quantum state', it's a stochastic state.

In probability theory, people like Poisson distributions. In the state Ψ_c, the probability of having n_i things of the ith species is equal to

$$e^{-c_i} \frac{c_i^{n_i}}{n_i!}$$

This is precisely the definition of a **Poisson distribution** with mean equal to c_i. We can multiply a bunch of factors like this, one for each species, to get

$$e^{-c} \frac{c^n}{n!}$$

This is the probability of having n_1 things of the first species, n_2 things of the second, and so on, in the state Ψ_c. So, the state Ψ_c is a product of independent Poisson distributions. In particular, knowing how many things there are of one species says *nothing all about* how many things there are of any other species!

It is remarkable that such a simple state can give an equilibrium solution of the master equation, even for very complicated stochastic Petri nets. But it's true—at least if c is complex balanced.

8.5 The Anderson–Craciun–Kurtz theorem

Now we're ready to state and prove the big result:

Theorem (Anderson–Craciun–Kurtz). Suppose $c \in [0,\infty)^k$ is a complex balanced equilibrium solution of the rate equation. Then $H\Psi_c = 0$.

It follows that Ψ_c is an equilibrium solution of the master equation. In other words, if we take $\Psi(t) = \Psi_c$ for all times t, the master equation holds:

$$\frac{d}{dt}\Psi(t) = H\Psi(t)$$

since both sides are zero.

Proof. To prove the Anderson–Craciun–Kurtz theorem, we just need to show that $H\Psi_c = 0$. Since Ψ_c is a constant times e^{cz}, it suffices to show $He^{cz} = 0$. Remember that

$$He^{cz} = \sum_{\tau \in T} r(\tau) \left({a^\dagger}^{n(\tau)} - {a^\dagger}^{m(\tau)} \right) a^{m(\tau)} \, e^{cz}$$

Since the annihilation operator a_i is given by differentiation with respect to z_i, while the creation operator $a_i^\dagger$ is just multiplying by z_i, we have:

$$He^{cz} = \sum_{\tau \in T} r(\tau) \, c^{m(\tau)} \left(z^{n(\tau)} - z^{m(\tau)} \right) e^{cz}$$

Expanding out e^{cz} we get:

$$He^{cz} = \sum_{i \in \mathbb{N}^k} \sum_{\tau \in T} r(\tau) c^{m(\tau)} \left(z^{n(\tau)} \frac{c^i}{i!} z^i - z^{m(\tau)} \frac{c^i}{i!} z^i \right)$$

Shifting indices and defining negative powers to be zero:

$$He^{cz} = \sum_{i \in \mathbb{N}^k} \sum_{\tau \in T} r(\tau) c^{m(\tau)} \left(\frac{c^{i-n(\tau)}}{(i-n(\tau))!} z^i - \frac{c^{i-m(\tau)}}{(i-m(\tau))!} z^i \right)$$

So, to show $He^{cz} = 0$, we need to show this:

$$\sum_{i \in \mathbb{N}^k} \sum_{\tau \in T} r(\tau) c^{m(\tau)} \frac{c^{i-n(\tau)}}{(i-n(\tau))!} \, z^i = \sum_{i \in \mathbb{N}^k} \sum_{\tau \in T} r(\tau) c^{m(\tau)} \frac{c^{i-m(\tau)}}{(i-m(\tau))!} \, z^i$$

We do this by splitting the sum over T according to output and then input complexes, making use of the complex balanced condition:

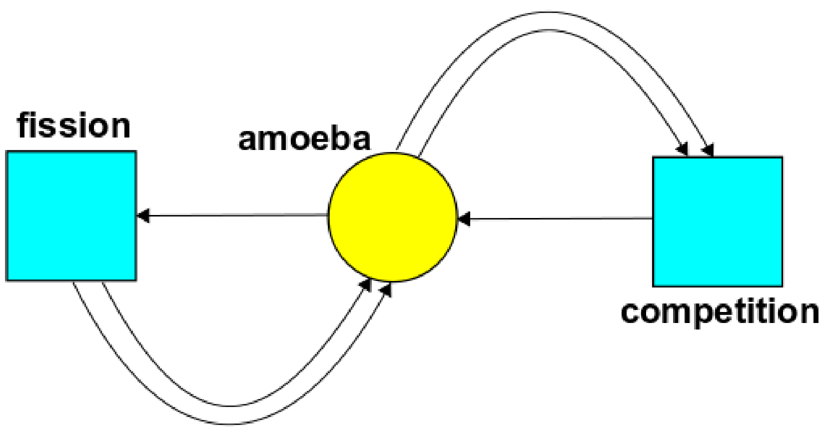
$$\begin{aligned}
\sum_{i \in \mathbb{N}^k} \sum_{\kappa \in \mathbb{N}^k} \sum_{\{\tau : n(\tau) = \kappa\}} r(\tau) c^{m(\tau)} \frac{c^{i-n(\tau)}}{(i-n(\tau))!} \, z^i &= \textstyle\sum_{i \in \mathbb{N}^k} \sum_{\kappa \in \mathbb{N}^k} \frac{c^{i-\kappa}}{(i-\kappa)!} \, z^i \sum_{\{\tau : n(\tau) = \kappa\}} r(\tau) c^{m(\tau)} \\
&= \textstyle\sum_{i \in \mathbb{N}^k} \sum_{\kappa \in \mathbb{N}^k} \frac{c^{i-\kappa}}{(i-\kappa)!} \, z^i \sum_{\{\tau : m(\tau) = \kappa\}} r(\tau) c^{m(\tau)} \\
&= \textstyle\sum_{i \in \mathbb{N}^k} \sum_{\kappa \in \mathbb{N}^k} \sum_{\{\tau : m(\tau) = \kappa\}} r(\tau) c^{m(\tau)} \frac{c^{i-m(\tau)}}{(i-m(\tau))!} \, z^i
\end{aligned}$$

This completes the proof! It's just algebra, but it seems a bit magical, so we're trying to understand it better.

I hope you see how amazing this result is. If you know quantum mechanics and coherent states you'll understand what I mean. A coherent state is the "best quantum approximation"; to a classical state, but we don't expect this quantum state to be *exactly* time-independent when the corresponding classical state is, *except* in very special cases, like when the Hamiltonian is quadratic in the creation and annihilation operators. Here we are getting a result like that much more generally... but only given the "complex balanced" condition.

8.6 An example

We've already seen one example of the Anderson–Craciun–Kurtz theorem back in Part 7. We had this stochastic Petri net:



We saw that the rate equation is just the logistic equation, familiar from population biology. The equilibrium solution is complex balanced, because pairs of amoebas are getting created at the same rate as they're getting destroyed, and *single* amoebas are getting created at the same rate as *they're* getting destroyed.

So, the Anderson–Craciun–Kurtz theorem guarantees that there's an equilibrium solution of the master equation where the number of amoebas is distributed according to a Poisson distribution. And, we actually checked that this was true!

Next time we'll look at another example.

You can also read comments on Azimuth, and make your own comments or ask questions there!

8.7 Answers to problems

Here is the answer to the problem, provided by David Corfield:

Problem 11. Show that if a classical state c is complex balanced, and we set $x(t) = c$ for all t, then $x(t)$ is a solution of the rate equation.

Answer. Assuming c is complex balanced, we have:

$$\begin{aligned}
\sum_{\tau \in T} r(\tau) m(\tau) c^{m(\tau)} &= \textstyle\sum_{\kappa} \sum_{\{\tau : m(\tau) = \kappa\}} r(\tau) m(\tau) c^{m(\tau)} \\
&= \sum_{\kappa} \sum_{\{\tau : m(\tau) = \kappa\}} r(\tau) \kappa c^{m(\tau)} \\
&= \sum_{\kappa} \sum_{\{\tau : n(\tau) = \kappa\}} r(\tau) \kappa c^{m(\tau)} \\
&= \sum_{\kappa} \sum_{\{\tau : n(\tau) = \kappa\}} r(\tau) n(\tau) c^{m(\tau)} \\
&= \sum_{\tau \in T} r(\tau) n(\tau) c^{m(\tau)}
\end{aligned}$$

So, we have

$$\sum_{\tau \in T} r(\tau) \left(n(\tau) - m(\tau) \right) c^{m(\tau)} = 0$$

and thus if $x(t) = c$ for all t then $x(t)$ is a solution of the rate equation:

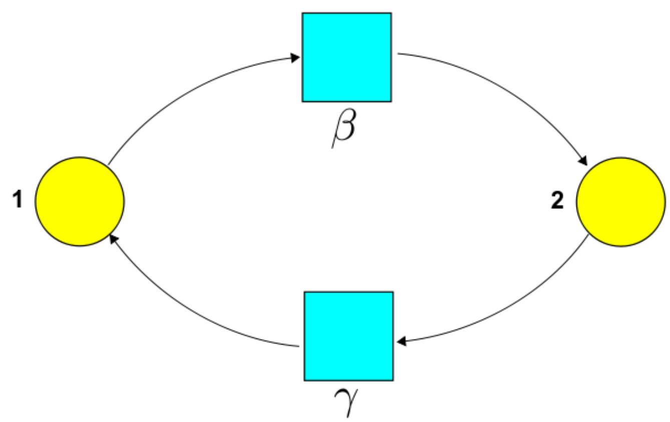
$$\frac{dx}{dt} = 0 = \sum_{\tau \in T} r(\tau) \left(n(\tau) - m(\tau) \right) x^{m(\tau)}$$

9 An example of the Anderson-Craciun-Kurtz theorem

Last time Brendan showed us a proof of the Anderson–Craciun–Kurtz theorem on equilibrium states. Today we'll look at an example that illustrates this theorem. This example brings up an interesting 'paradox'—or at least a problem. Resolving this will get us ready to think about a version of *Noether's theorem* relating conserved quantities and symmetries. Next time Brendan will state and prove a version of Noether's theorem that applies to stochastic Petri nets.

9.1 A reversible reaction

In chemistry a type of atom, molecule, or ion is called a **chemical species**, or **species** for short. Since we're applying our ideas to both chemistry and biology, it's nice that 'species' is also used for a type of organism in biology. This stochastic Petri net describes the simplest reversible reaction of all, involving two species:



We have species 1 turning into species 2 with rate constant β, and species 2 turning back into species 1 with rate constant γ. So, the rate equation is:

$$\begin{aligned} \frac{dx_1}{dt} &= -\beta x_1 + \gamma x_2 \\ \frac{dx_2}{dt} &= \beta x_1 - \gamma x_2 \end{aligned}$$

where x_1 and x_2 are the amounts of species 1 and 2, respectively.

9.2 Equilibrium solutions of the rate equation

Let's look for *equilibrium* solutions of the rate equation, meaning solutions where the amount of each species doesn't change with time. Equilibrium occurs when each species is getting created at the same rate at which it's getting destroyed.

So, let's see when

$$\frac{dx_1}{dt} = \frac{dx_2}{dt} = 0$$

Clearly this happens precisely when

$$\beta x_1 = \gamma x_2$$

This says the rate at which 1's are turning into 2's equals the rate at which 2's are turning back into 1's. That makes perfect sense.

9.3 Complex balanced equilibrium

In general, a chemical reaction involves a *bunch* of species turning into a *bunch* of species. Since 'bunch' is not a very dignified term, a bunch of species is usually called a **complex**. We saw last time that it's very interesting to study a strong version of equilibrium: **complex balanced equilibrium**, in which each *complex* is being created at the same rate at which it's getting destroyed.

However, in the Petri net we're studying today, all the complexes being produced or destroyed consist of a single species. In this situation, any equilibrium solution is automatically complex balanced. This is great, because it means we can apply the Anderson–Craciun–Kurtz theorem from last time! This says how to get from a complex balanced equilibrium solution of the *rate equation* to an equilibrium solution of the *master equation* .

First remember what the master equation says. Let $\psi_{n_1,n_2}(t)$ be the probability that we have n_1 things of species 1 and n_2 things of species 2 at time t. We summarize all this information in a formal power series:

$$\Psi(t) = \sum_{n_1,n_2=0}^{\infty} \psi_{n_1,n_2}(t) z_1^{n_1} z_2^{n_2}$$

Then the master equation says

$$\frac{d}{dt}\Psi(t) = H\Psi(t)$$

where following the general rules laid down in Part 8,

$$H = \beta(a_2^\dagger - a_1^\dagger)a_1 + \gamma(a_1^\dagger - a_2^\dagger)a_2$$

This may look scary, but the **annihilation operator** a_i and the **creation operator** $a_i^\dagger$ are just funny ways of writing the partial derivative $\partial/\partial z_i$ and multiplication by z_i, so

$$H = \beta(z_2 - z_1)\frac{\partial}{\partial z_1} + \gamma(z_1 - z_2)\frac{\partial}{\partial z_2}$$

or if you prefer,

$$H = (z_2 - z_1)\,(\beta\frac{\partial}{\partial z_1} - \gamma\frac{\partial}{\partial z_2})$$

The first term describes species 1 turning into species 2. The second describes species 2 turning back into species 1.

Now, the Anderson–Craciun–Kurtz theorem says that whenever (x_1, x_2) is a complex balanced solution of the rate equation, this recipe gives an equilibrium solution of the master equation:

$$\Psi = \frac{e^{x_1 z_1 + x_2 z_2}}{e^{x_1 + x_2}}$$

In other words: whenever $\beta x_1 = \gamma x_2$, we have
we have

$$H\Psi = 0$$

Let's check this! For starters, the constant in the denominator of Ψ doesn't matter here, since H is linear. It's just a normalizing constant, put in to make sure that our probabilities ψ_{n_1,n_2} sum to 1. So, we just need to check that

$$(z_2 - z_1)(\beta\frac{\partial}{\partial z_1} - \gamma\frac{\partial}{\partial z_2})e^{x_1 z_1 + x_2 z_2} = 0$$

If we do the derivatives on the left hand side, it's clear we want

$$(z_2 - z_1)(\beta x_1 - \gamma x_2)e^{x_1 z_1 + x_2 z_2} = 0$$

And this is indeed true when $\beta x_1 = \gamma x_2$.

So, the theorem works as advertised. And now we can work out the probability ψ_{n_1,n_2} of having n_1 things of species 1 and n_2 of species 2 in our equilibrium state Ψ. To do this, we just expand the function Ψ as a power series and look at the coefficient of $z_1^{n_1} z_2^{n_2}$. We have

$$\Psi = \frac{e^{x_1 z_1 + x_2 z_2}}{e^{x_1 + x_2}} = \frac{1}{e^{x_1} e^{x_2}} \sum_{n_1,n_2}^{\infty} \frac{(x_1 z_1)^{n_1}}{n_1!} \frac{(x_2 z_2)^{n_2}}{n_2!}$$

so we get

$$\psi_{n_1,n_2} = \frac{1}{e^{x_1}} \frac{x_1^{n_1}}{n_1!} \cdot \frac{1}{e^{x_2}} \frac{x_1^{n_2}}{n_2!}$$

This is just a product of two independent Poisson distributions!

In case you forget, a Poisson distribution says the probability of k events occurring in some interval of time if they occur with a fixed average rate and independently of the time since the last event. If the expected number of events is λ, the Poisson distribution is

$$\frac{1}{e^{\lambda}} \frac{\lambda^k}{k!}$$

and it looks like this for various values of λ: [FIG off of wiki]

It looks almost like a Gaussian when λ is large, but when λ is small it becomes very lopsided.

Anyway: we've seen that in our equilibrium state, the number of things of species $i = 1, 2$ is given by a Poisson distribution with mean x_i. That's very nice and simple... but the amazing thing is that these distributions are *independent* .

Mathematically, this means we just *multiply* them to get the probability of finding n_1 things of species 1 and n_2 of species 2. But it also means that knowing how many things there are of one species says nothing about the number of the other.

But something seems odd here. One transition in our Petri net consumes a 1 and produces a 2, while the other consumes a 2 and produces a 1. The total number of particles in the system never changes. The more 1's there are, the fewer 2's there should be. But we just said knowing how many 1's we have tells us nothing about how many 2's we have!

At first this seems like a paradox. Have we made a mistake? Not exactly. But we're neglecting something.

9.4 Conserved quantities

Namely: the equilibrium solutions of the master equation we've found so far are not the only ones! There are other solutions that fit our intuitions better.

Suppose we take any of our equilibrium solutions Ψ and change it like this: set the probability ψ_{n_1,n_2} equal to 0 unless

$$n_1 + n_2 = n$$

but otherwise leave it unchanged. Of course the probabilities no longer sum to 1, but we can rescale them so they do.

The result is a new equilibrium solution, say Ψ_n. Why? Because, as we've already seen, no transitions will carry us from one value of $n_1 + n_2$ to another. And in this new solution, the number of 1's is clearly *not* independent from the number of 2's. The bigger one is, the smaller the other is.

Problem 12. Show that this new solution Ψ_n depends only on n and the ratio $x_1/x_2 = \gamma/\beta$, not on anything more about the values of x_1 and x_2 in the original solution

$$\Psi = \frac{e^{x_1 z_1 + x_2 z_2}}{e^{x_1 + x_2}}$$

Problem 13. What is this new solution like when $\beta = \gamma$? (This particular choice makes the problem symmetrical when we interchange species 1 and 2.)

What's happening here is that this particular stochastic Petri net has a 'conserved quantity': the total number of things never changes with time. So, we can take any equilibrium solution of the master equation and—in the language of quantum mechanics—'project down to the subspace' where this conserved quantity takes a definite value, and get a new equilibrium solution. In the language of probability theory, we say it a bit differently: we're 'conditioning on' the conserved quantity taking a definite value. But the idea is the same.

This important feature of conserved quantities suggests that we should try to invent a new version of Noether's theorem. This theorem links conserved quantities and *symmetries* of the Hamiltonian.

There are already a couple versions of Noether's theorem for classical mechanics, and for quantum mechanics... but now we want a version for *stochastic* mechanics. And indeed one exists, and it's relevant to what we're doing here. We'll show it to you next time.

You can also read comments on Azimuth, and make your own comments or ask questions there!

10 A stochastic version of Noether's theorem

Noether proved lots of theorems, but when people talk about Noether's theorem, they always seem to mean her result linking *symmetries* to *conserved quantities* . Her original result applied to classical mechanics, but today we'd like to present a version that applies to 'stochastic mechanics'—or in other words, Markov processes.

What's a Markov process? We'll say more in a minute—but in plain English, it's a physical system where something hops around randomly from state to state, where its probability of hopping anywhere depends only on where it is *now* , not its past history. Markov processes include, as a special case, the stochastic Petri nets we've been talking about.

Our stochastic version of Noether's theorem is copied after a well-known *quantum* version. It's yet another example of how we can exploit the analogy between stochastic mechanics and quantum mechanics. But for now we'll just present the stochastic version. Next time we'll compare it to the quantum one.

10.1 Markov processes

We should and probably *will* be more general, but let's start by considering a finite set of states, say X. To describe a Markov process we then need a matrix of real numbers $H = (H_{ij})_{i,j \in X}$. The idea is this: suppose right now our system is in the state i. Then the probability of being in some state j changes as time goes by—and H_{ij} is defined to be the *time derivative* of this probability right now.

So, if $\psi_i(t)$ is the probability of being in the state i at time t, we want the **master equation** to hold:

$$\frac{d}{dt}\psi_i(t) = \sum_{j \in X} H_{ij}\psi_j(t)$$

This motivates the definition of 'infinitesimal stochastic', which we recall from Part 5:

Definition. Given a finite set X, a matrix of real numbers $H = (H_{ij})_{i,j \in X}$ is **infinitesimal stochastic** if

$$i \neq j \Rightarrow H_{ij} \geq 0$$

and

$$\sum_{i \in X} H_{ij} = 0$$

for all $j \in X$.

The inequality says that if we start in the state i, the probability of being in some other state, which starts at 0, can't go down, at least initially. The equation says that the probability of being *somewhere or other* doesn't change. Together, these facts imply that:

$$H_{ii} \leq 0$$

That makes sense: the probability of being in the state i, which starts at 1, can't go up, at least initially.

Using the magic of matrix multiplication, we can rewrite the master equation as follows:

$$\frac{d}{dt}\psi(t) = H\psi(t)$$

and we can solve it like this:

$$\psi(t) = \exp(tH)\psi(0)$$

If H is an infinitesimal stochastic operator, we will call $\exp(tH)$ a **Markov process**, and H its **Hamiltonian**.

(Actually, most people call $\exp(tH)$ a **Markov semigroup**, and reserve the term Markov process for another way of looking at the same idea. So, be careful.)

Noether's theorem is about 'conserved quantities', that is, observables whose expected values don't change with time. To understand this theorem, you need to know a bit about observables. In stochastic mechanics an **observable** is simply a function assigning a number O_i to each state $i \in X$.

However, in quantum mechanics we often think of observables as matrices, so it's nice to do that here, too. It's easy: we just create a matrix whose diagonal entries are the values of the function O. And just to confuse you, we'll also call this matrix O. So:

$$O_{ij} = \begin{cases} O_i & \text{if} \quad i = j \\ 0 & \text{if} \quad i \neq j \end{cases}$$

One advantage of this trick is that it lets us ask whether an observable commutes with the Hamiltonian. Remember, the **commutator** of matrices is defined by

$$[O, H] = OH - HO$$

Noether's theorem will say that $[O, H] = 0$ if and only if O is 'conserved' in some sense. What sense? First, recall that a **stochastic state** is just our fancy name for a probability distribution ψ on the set X. Second, the **expected value** of an observable O in the stochastic state ψ is defined to be

$$\sum_{i \in X} O_i \psi_i$$

In Part 5 we introduced the notation

$$\int \phi = \sum_{i \in X} \phi_i$$

for any function ϕ on X. The reason is that later, when we generalize X from a finite set to a measure space, the sum at right will become an integral over X. Indeed, a sum is just a special sort of integral!

Using this notation and the magic of matrix multiplication, we can write the expected value of O in the stochastic state ψ as

$$\int O\psi$$

We can calculate how this changes in time if ψ obeys the master equation... and we can write the answer using the commutator $[O, H]$:

Lemma. Suppose H is an infinitesimal stochastic operator and O is an observable. If $\psi(t)$ obeys the master equation, then

$$\frac{d}{dt}\int O\psi(t) = \int [O, H]\psi(t)$$

Proof. Using the master equation we have

$$\frac{d}{dt}\int O\psi(t) = \int O\frac{d}{dt}\psi(t) = \int OH\psi(t)$$

But since H is infinitesimal stochastic,

$$\sum_{i \in X} H_{ij} = 0$$

so for any function ϕ on X we have

$$\int H\phi = \sum_{i,j \in X} H_{ij}\phi_j = 0$$

and in particular

$$\int HO\psi(t) = 0$$

Since $[O, H] = OH - HO$, we conclude from (1) and (2) that

$$\frac{d}{dt}\int O\psi(t) = \int [O, H]\psi(t)$$

as desired.

The commutator doesn't look like it's doing much here, since we also have

$$\frac{d}{dt}\int O\psi(t) = \int OH\psi(t)$$

which is even simpler. But the commutator will become useful when we get to Noether's theorem!

10.2 Noether's theorem

Here's a version of Noether's theorem for Markov processes. It says an observable commutes with the Hamiltonian iff the expected values of that observable *and its square* don't change as time passes:

Theorem 4. *Suppose H is an infinitesimal stochastic operator and O is an observable. Then*

$$[O, H] = 0$$

if and only if

$$\frac{d}{dt} \int O\psi(t) = 0$$

and

$$\frac{d}{dt} \int O^2\psi(t) = 0$$

for all $\psi(t)$ obeying the master equation.

If you know Noether's theorem from quantum mechanics, you might be surprised that in this version we need not only the observable *but also its square* to have an unchanging expected value! We'll explain this, but first let's prove the theorem.

Proof. The easy part is showing that if $[O, H] = 0$ then $\frac{d}{dt} \int O\psi(t) = 0$ and $\frac{d}{dt} \int O^2\psi(t) = 0$. In fact there's nothing special about these two powers of t; we'll show that

$$\frac{d}{dt} \int O^n\psi(t) = 0$$

for all n. The point is that since H commutes with O, it commutes with all powers of O:

$$[O^n, H] = 0$$

So, applying the Lemma to the observable O^n, we see

$$\frac{d}{dt} \int O^n\psi(t) = \int [O^n, H]\psi(t) = 0$$

The backward direction is a bit trickier. We now assume that

$$\frac{d}{dt} \int O\psi(t) = \frac{d}{dt} \int O^2\psi(t) = 0$$

for all solutions $\psi(t)$ of the master equation. This implies

$$\int OH\psi(t) = \int O^2H\psi(t) = 0$$

or since this holds for *all* solutions,

$$\sum_{i \in X} O_i H_{ij} = \sum_{i \in X} O_i^2 H_{ij} = 0$$

We wish to show that $[O, H] = 0$.

First, recall that we can think of O is a diagonal matrix with:

$$O_{ij} = \begin{cases} O_i & \text{if} \quad i = j \\ 0 & \text{if} \quad i \neq j \end{cases}$$

So, we have

$$[O, H]_{ij} = \sum_{k \in X} (O_{ik} H_{kj} - H_{ik} O_{kj}) = O_i H_{ij} - H_{ij} O_j = (O_i - O_j) H_{ij}$$

To show this is zero for each pair of elements $i, j \in X$, it suffices to show that when $H_{ij} \neq 0$, then $O_j = O_i$. That is, we need to show that if the system can move from state j to state i, then the observable takes the same value on these two states.

In fact, it's enough to show that this sum is zero for any $j \in X$:

$$\sum_{i \in X} (O_j - O_i)^2 H_{ij}$$

Why? When $i = j$, $O_j - O_i = 0$, so that term in the sum vanishes. But when $i \neq j$, $(O_j - O_i)^2$ and H_{ij} are both non-negative—the latter because H is infinitesimal stochastic. So if they sum to zero, they must each be individually zero. Thus for all $i \neq j$, we have $(O_j - O_i)^2 H_{ij} = 0$. But this means that either $O_i = O_j$ or $H_{ij} = 0$, which is what we need to show.

So, let's take that sum and expand it:

$$\begin{aligned} \sum_{i \in X} (O_j - O_i)^2 H_{ij} &= \sum_i (O_j^2 H_{ij} - 2 O_j O_i H_{ij} + O_i^2 H_{ij}) \\ &= O_j^2 \sum_i H_{ij} - 2 O_j \sum_i O_i H_{ij} + \sum_i O_i^2 H_{ij} \end{aligned}$$

The three terms here are each zero: the first because H is infinitesimal stochastic, and the latter two by equation (3). So, we're done!

10.3 Markov chains

So that's the proof... but why do we need both O *and its square* to have an expected value that doesn't change with time to conclude $[O, H] = 0$? There's an easy counterexample if we leave out the condition involving O^2. However, the underlying idea is clearer if we work with Markov chains instead of Markov processes.

In a Markov process, time passes by continuously. In a Markov chain, time comes in discrete steps! We get a Markov process by forming $\exp(tH)$ where H is an infinitesimal stochastic operator. We get a Markov chain by forming the operator $U, U^2, U^3, \ldots$ where U is a 'stochastic operator'. Remember:

Definition 5. *Given a finite set X, a matrix of real numbers $U = (U_{ij})_{i,j \in X}$ is* **stochastic** *if*

$$U_{ij} \geq 0$$

for all $i, j \in X$ and

$$\sum_{i \in X} U_{ij} = 1$$

for all $j \in X$.

The idea is that U describes a random hop, with U_{ij} being the probability of hopping to the state i if you start at the state j. These probabilities are nonnegative and sum to 1.

Any stochastic operator gives rise to a **Markov chain** $U, U^2, U^3, \ldots$. And in case it's not clear, that's how we're *defining* a Markov chain: the sequence of powers of a stochastic operator. There are other definitions, but they're equivalent.

We can draw a Markov chain by drawing a bunch of states and arrows labelled by transition probabilities, which are the matrix elements U_{ij}:

[PIC off of wikipedia, markov chain]

Here is Noether's theorem for Markov chains:

Theorem 6. *Suppose U is a stochastic operator and O is an observable. Then*

$$[O, U] = 0$$

if and only if

$$\int OU\psi = \int O\psi$$

and

$$\int O^2 U\psi = \int O^2 \psi$$

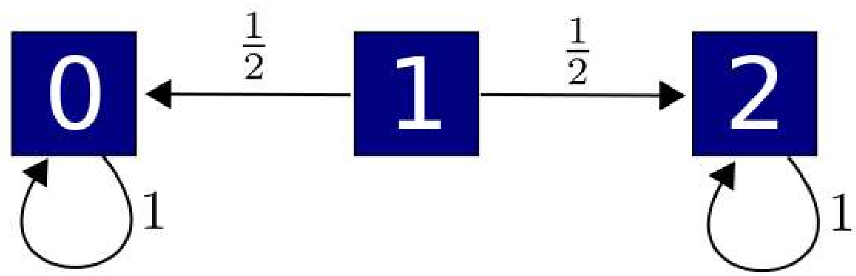
for all stochastic states ψ.

In other words, an observable commutes with U iff the expected values of that observable *and its square* don't change when we evolve our state one time step using U.

You can probably prove this theorem by copying the proof for Markov processes:

Problem 14. Prove Noether's theorem for Markov chains.

But let's see why we need the condition on the square of observable! That's the intriguing part. Here's a nice little Markov chain:



where we haven't drawn arrows labelled by 0. So, state 1 has a 50% chance of hopping to state 0 and a 50% chance of hopping to state 2; the other two states just sit there. Now, consider the observable O with

$$O_i = i$$

It's easy to check that the expected value of this observable doesn't change with time:

$$\int OU\psi = \int O\psi$$

for all ψ. The reason, in plain English, is this. Nothing at all happens if you start at states 0 or 2: you just sit there, so the expected value of O doesn't change. If you start at state 1, the observable equals 1. You then have a 50% chance of going to a state where the observable equals 0 and a 50% chance of going to a state where it equals 2, so its *expected* value doesn't change: it still equals 1.

On the other hand, we do *not* have $[O, U] = 0$ in this example, because we *can* hop between states where O takes different values. Furthermore,

$$\int O^2U\psi \neq \int O^2\psi$$

After all, if you start at state 1, O^2 equals 1 there. You then have a 50% chance of going to a state where O^2 equals 0 and a 50% chance of going to a state where it equals 4, so its expected value changes!

So, that's why $\int OU\psi = \int O\psi$ for all ψ is not enough to guarantee $[O, U] = 0$. The same sort of counterexample works for Markov processes, too.

Finally, we should add that there's nothing terribly sacred about the *square* of the observable. For example, we have:

Theorem 7. *Suppose H is an infinitesimal stochastic operator and O is an observable. Then*

$$[O, H] = 0$$

if and only if

$$\frac{d}{dt} \int f(O)\psi(t) = 0$$

for all smooth $f : \mathbb{R} \to \mathbb{R}$ and all $\psi(t)$ obeying the master equation.

Theorem 8. *Suppose U is a stochastic operator and O is an observable. Then*

$$[O, U] = 0$$

if and only if

$$\int f(O)U\psi = \int f(O)\psi$$

for all smooth $f : \mathbb{R} \to \mathbb{R}$ and all stochastic states ψ

These make the 'forward direction' of Noether's theorem stronger... and in fact, the forward direction, while easier, is probably more useful! However, if we ever use Noether's theorem in the 'reverse direction', it might be easier to check a condition involving only O and its square.

Here's the answer to the problem:

Problem 15. Suppose U is a stochastic operator and O is an observable. Show that O commutes with U iff the expected values of O *and its square* don't change when we evolve our state one time step using U. In other words, show that

$$[O, U] = 0$$

if and only if

$$\int OU\psi = \int O\psi$$

and

$$\int O^2 U\psi = \int O^2\psi$$

for all stochastic states ψ.

Answer. One direction is easy: if $[O, U] = 0$ then $[O^n, U] = 0$ for all n, so

$$\int O^n U\psi = \int UO^n\psi = \int O^n\psi$$

where in the last step we use the fact that U is stochastic.

For the converse we can use the same tricks that worked for Markov processes. Assume that

$$\int OU\psi = \int O\psi$$

and

$$\int O^2 U\psi = \int O^2\psi$$

for all stochastic states ψ. These imply that

$$\sum_{i\in X} O_i U_{ij} = O_j \qquad (1)$$

and

$$\sum_{i\in X} O_i^2 U_{ij} = O_j^2 \qquad (2)$$

We wish to show that $[O,U]=0$. Note that

$$\begin{aligned} [O,U]_{ij} &= \sum_{k\in X}(O_{ik}U_{kj} - U_{ik}O_{kj}) \\ &= (O_i - O_j)U_{ij} \end{aligned}$$

To show this is always zero, we'll show that when $U_{ij}\neq 0$, then $O_j = O_i$. This says that when our system can hop from one state to another, the observable O must take the same value on these two states.

For this, in turn, it's enough to show that the following sum vanishes for any $j\in X$:

$$\sum_{i\in X}(O_j - O_i)^2 U_{ij}$$

Why? The matrix elements U_{ij} are nonnegative since U is stochastic. Thus the sum can only vanish if each term vanishes, meaning that $O_j = O_i$ whenever $U_{ij}\neq 0$.

To show the sum vanishes, let's expand it:

$$\begin{aligned} \sum_{i\in X}(O_j - O_i)^2 U_{ij} &= \sum_i (O_j^2 U_{ij} - 2O_jO_iU_{ij} + O_i^2 U_{ij}) \\ &= O_j^2\sum_i U_{ij} - 2O_j\sum_i O_iU_{ij} + \sum_i O_i^2 U_{ij} \end{aligned}$$

Now, since (1) and (2) hold for all stochastic states ψ, this equals

$$O_j^2 \sum_i U_{ij} - 2O_j^2 + O_j^2$$

But this is zero because U is stochastic, which implies

$$\sum_i U_{ij} = 1$$

So, we're done!

11 Comparing quantum mechanics and stochastic mechanics

Last time we proved a version of Noether's theorem for stochastic mechanics. Now I want to compare that to the more familiar quantum version.

But to do this, I need to say more about the analogy between stochastic mechanics and quantum mechanics. And whenever I try, I get pulled toward explaining some technical issues involving analysis: whether sums converge, whether derivatives exist, and so on. I've been trying to avoid such stuff—not because I dislike it, but because I'm afraid *you* might. But the more I put off discussing these issues, the more they fester and make me unhappy. In fact, that's why it's taken so long for me to write this post!

So, this time I will gently explore some of these technical issues. But don't be scared: I'll *mainly* talk about some simple big ideas. Next time I'll discuss Noether's theorem. I hope that by getting the technicalities out of my system, I'll feel okay about hand-waving whenever I want.

And if you're an expert on analysis, maybe you can help me with a question.

First, we need to recall the analogy we began sketching in Part 5, and push it a bit further. The idea is that stochastic mechanics differs from quantum mechanics in two big ways:

- First, instead of complex amplitudes, stochastic mechanics uses nonnegative real probabilities. The complex numbers form a ring; the nonnegative real numbers form a mere rig, which is a 'ri**n**g without **n**egatives'. Rigs are much neglected in the typical math curriculum, but unjustly so: they're almost as good as rings in many ways, and there are lots of important examples, like the natural numbers $\mathbb{N}$ and the nonnegative real numbers, $[0, \infty)$. For probability theory, we should learn to love rigs.

  But there are, alas, situations where we need to subtract probabilities, even when the answer comes out negative: namely when we're taking the *time derivative* of a probability. So sometimes we need $\mathbb{R}$ instead of just $[0, \infty)$.

- Second, while in quantum mechanics a state is described using a 'wave-function', meaning a complex-valued function obeying

  $$\int |\psi|^2 = 1$$

  in stochastic mechanics it's described using a 'probability distribution', meaning a nonnegative real function obeying

  $$\int \psi = 1$$

So, let's try our best to present the theories in close analogy, while respecting these two differences.

11.1 States

We'll start with a set X whose points are **states** that a system can be in. Last time I assumed X was a finite set, but this post is so mathematical I might as well let my hair down and assume it's a measure space. A measure space lets you do integrals, but a finite set is a special case, and then these integrals are just sums. So, I'll write things like

$$\int f$$

and mean the integral of the function f over the measure space X, but if X is a finite set this just means

$$\sum_{x \in X} f(x)$$

Now, I've already defined the word 'state', but both quantum and stochastic mechanics need a more general concept of state. Let's call these 'quantum states' and 'stochastic states':

- In **quantum mechanics**, the system has an **amplitude** $\psi(x)$ of being in any state $x \in X$. These amplitudes are complex numbers with

$$\int |\psi|^2 = 1$$

  We call $\psi : X \to \mathbb{C}$ obeying this equation a **quantum state**.

- In **stochastic mechanics**, the system has a **probability** $\psi(x)$ of being in any state $x \in X$. These probabilities are nonnegative real numbers with

$$\int \psi = 1$$

  We call $\psi : X \to [0, \infty)$ obeying this equation a **stochastic state**.

In quantum mechanics we often use this abbreviation:

$$\langle \phi, \psi \rangle = \int \overline{\phi} \psi$$

so that a quantum state has

$$\langle \psi, \psi \rangle = 1$$

Similarly, we could introduce this notation in stochastic mechanics:

$$\langle \psi \rangle = \int \psi$$

so that a stochastic state has

$$\langle \psi \rangle = 1$$

But this notation is a bit risky, since angle brackets of this sort often stand for expectation values of observables. So, I've been writing $\int \psi$, and I'll keep on doing this.

In quantum mechanics, $\langle \phi, \psi \rangle$ is well-defined whenever both ϕ and ψ live in the vector space

$$L^2(X) = \{\psi : X \to \mathbb{C} : \int |\psi|^2 < \infty\}$$

In stochastic mechanics, $\langle \psi \rangle$ is well-defined whenever ψ lives in the vector space

$$L^1(X) = \{\psi : X \to \mathbb{R} : \int |\psi| < \infty\}$$

You'll notice I wrote $\mathbb{R}$ rather than $[0, \infty)$ here. That's because in some calculations we'll need functions that take negative values, even though our stochastic states are nonnegative.

11.2 Observables

A state is a way our system can be. An observable is something we can measure about our system. They fit together: we can measure an observable when our system is in some state. If we repeat this we may get different answers, but there's a nice formula for average or 'expected' answer.

In quantum mechanics, an **observable** is a self-adjoint operator A on $L^2(X)$. The **expected value** of A in the state ψ is

$$\langle \psi, A\psi \rangle$$

Here I'm assuming that we can apply A to ψ and get a new vector $A\psi \in L^2(X)$. This is automatically true when X is a finite set, but in general we need to be more careful.

In stochastic mechanics, an **observable** is a real-valued function A on X. The **expected value** of A in the state ψ is

$$\int A\psi$$

Here we're using the fact that we can multiply A and ψ and get a new vector $A\psi \in L^1(X)$, at least if A is bounded. Again, this is automatic if X is a finite set, but not otherwise.

11.3 Symmetries

Besides states and observables, we need 'symmetries', which are transformations that map states to states. We use these to describe how our system changes when we wait a while, for example.

In quantum mechanics, an **isometry** is a linear map $U : L^2(X) \to L^2(X)$ such that

$$\langle U\phi, U\psi\rangle = \langle \phi, \psi\rangle$$

for all $\psi, \phi \in L^2(X)$. If U is an isometry and ψ is a quantum state, then $U\psi$ is again a quantum state.

In stochastic mechanics, a **stochastic operator** is a linear map $U : L^1(X) \to L^1(X)$ such that

$$\int U\psi = \int \psi$$

and

$$\psi \geq 0 \Rightarrow U\psi \geq 0$$

for all $\psi \in L^1(X)$. If U is stochastic and ψ is a stochastic state, then $U\psi$ is again a stochastic state.

In quantum mechanics we are mainly interested in invertible isometries, which are called **unitary** operators. There are lots of these, and their inverses are always isometries. There are, however, very few stochastic operators whose inverses are stochastic:

Problem 16. Suppose X is a finite set. Show that any isometry $U : L^2(X) \to L^2(X)$ is invertible, and its inverse is again an isometry.

Problem 17. Suppose X is a finite set. Which stochastic operators $U : L^1(X) \to L^1(X)$ have stochastic inverses?

This is why we usually think of time evolution as being reversible quantum mechanics, but not in stochastic mechanics! In quantum mechanics we often describe time evolution using a '1-parameter group', while in stochastic mechanics we describe it using a 1-parameter *semi* group... meaning that we can run time forwards, but not backwards.

But let's see how this works in detail!

11.4 Time evolution in quantum mechanics

In quantum mechanics there's a beautiful relation between observables and symmetries, which goes like this. Suppose that for each time t we want a unitary operator $U(t) : L^2(X) \to L^2(X)$ that describes time evolution. Then it makes a lot of sense to demand that these operators form a 1-parameter group:

Definition 9. *A collection of linear operators U(t) (t $\in \mathbb{R}$) on some vector space forms a* **1-parameter group** *if*

$$U(0) = 1$$

and

$$U(s+t) = U(s)U(t)$$

for all $s, t \in \mathbb{R}$.

Note that these conditions force all the operators $U(t)$ to be invertible.

Now suppose our vector space is a Hilbert space, like $L^2(X)$. Then we call a 1-parameter group a **1-parameter unitary group** if the operators involved are all unitary.

It turns out that 1-parameter unitary groups are either continuous in a certain way, or so pathological that you can't even prove they exist without the axiom of choice! So, we always focus on the continuous case:

Definition 10. *A 1-parameter unitary group is* **strongly continuous** *if* $U(t)\psi$ *depends continuously on* t *for all* ψ*, in this sense:*

$$t_i \to t \Rightarrow \|U(t_i)\psi - U(t)\psi\| \to 0$$

Then we get a classic result proved by Marshall Stone back in the early 1930s. You may not know him, but he was so influential at the University of Chicago during this period that it's often called the "Stone Age". And here's one reason why:

Stone's Theorem. There is a one-to-one correspondence between strongly continuous 1-parameter unitary groups on a Hilbert space and self-adjoint operators on that Hilbert space, given as follows. Given a strongly continuous 1-parameter unitary group $U(t)$ we can always write

$$U(t) = \exp(-itH)$$

for a unique self-adjoint operator H. Conversely, any self-adjoint operator determines a strongly continuous 1-parameter group this way. For all vectors ψ for which $H\psi$ is well-defined, we have

$$\left.\frac{d}{dt}U(t)\psi\right|_{t=0} = -iH\psi$$

Moreover, for any of these vectors, if we set

$$\psi(t) = \exp(-itH)\psi$$

we have

$$\frac{d}{dt}\psi(t) = -iH\psi(t)$$

When $U(t) = \exp(-itH)$ describes the evolution of a system in time, H is is called the **Hamiltonian**, and it has the physical meaning of 'energy'. The equation I just wrote down is then called **Schrödinger's equation**.

So, simply put, in quantum mechanics we have a correspondence between observables and nice one-parameter groups of symmetries. Not surprisingly, our favorite observable, energy, corresponds to our favorite symmetry: time evolution!

However, if you were paying attention, you noticed that I carefully avoided explaining how we define $\exp(-itH)$. I didn't even say what a self-adjoint operator is. This is where the technicalities come in: they arise when H is unbounded, and not defined on all vectors in our Hilbert space.

Luckily, these technicalities evaporate for *finite-dimensional* Hilbert spaces, such as $L^2(X)$ for a finite set X. Then we get:

Stone's Theorem (Baby Version). Suppose we are given a *finite-dimensional* Hilbert space. In this case, a linear operator H on this space is self-adjoint iff it's defined on the whole space and

$$\langle \phi, H\psi \rangle = \langle H\phi, \psi \rangle$$

for all vectors ϕ, ψ. Given a strongly continuous 1-parameter unitary group $U(t)$ we can always write

$$U(t) = \exp(-itH)$$

for a unique self-adjoint operator H, where

$$\exp(-itH)\psi = \sum_{n=0}^{\infty} \frac{(-itH)^n}{n!}\psi$$

with the sum converging for all ψ. Conversely, any self-adjoint operator on our space determines a strongly continuous 1-parameter group this way. For all vectors ψ in our space we then have

$$\left.\frac{d}{dt}U(t)\psi\right|_{t=0} = -iH\psi$$

and if we set

$$\psi(t) = \exp(-itH)\psi$$

we have

$$\frac{d}{dt}\psi(t) = -iH\psi(t)$$

11.5 Time evolution in stochastic mechanics

We've seen that in quantum mechanics, time evolution is usually described by a 1-parameter group of operators that comes from an observable: the Hamiltonian. Stochastic mechanics is different!

First, since stochastic operators aren't usually invertible, we typically describe time evolution by a mere 'semigroup':

Definition 11. *A collection of linear operators* $U(t)$ *(*$t \in [0, \infty)$*) on some vector space forms a* **1-parameter semigroup** *if*

$$U(0) = 1$$

and

$$U(s + t) = U(s)U(t)$$

for all $s, t \geq 0$.

Now suppose this vector space is $L^1(X)$ for some measure space X. We want to focus on the case where the operators $U(t)$ are stochastic and depend continuously on t in the same sense we discussed earlier.

Definition 12. *A 1-parameter strongly continuous semigroup of stochastic operators* $U(t) : L^1(X) \to L^1(X)$ *is called a* **Markov semigroup**.

What's the analogue of Stone's theorem for Markov semigroups? I don't know a fully satisfactory answer! If you know, please tell me.

Later I'll say what I *do* know—I'm not *completely* clueless—but for now let's look at the 'baby' case where X is a finite set. Then the story is neat and complete:

Theorem 13. *Suppose we are given a* finite *set* X. *In this case, a linear operator* H *on* $L^1(X)$ *is* **infinitesimal stochastic** *iff it's defined on the whole space,*

$$\int H\psi = 0$$

for all $\psi \in L^1(X)$, *and the matrix of* H *in terms of the obvious basis obeys*

$$H_{ij} \geq 0$$

for all $j \neq i$. *Given a Markov semigroup* $U(t)$ *on* $L^1(X)$, *we can always write*

$$U(t) = \exp(tH)$$

for a unique infinitesimal stochastic operator H, *where*

$$\exp(tH)\psi = \sum_{n=0}^{\infty} \frac{(tH)^n}{n!}\psi$$

with the sum converging for all ψ. *Conversely, any infinitesimal stochastic operator on our space determines a Markov semigroup this way. For all* $\psi \in L^1(X)$ *we then have*

$$\left.\frac{d}{dt}U(t)\psi\right|_{t=0} = H\psi$$

and if we set

$$\psi(t) = \exp(tH)\psi$$

we have the **master equation***:*

$$\frac{d}{dt}\psi(t) = H\psi(t)$$

In short, time evolution in stochastic mechanics is a lot like time evolution in quantum mechanics, except it's typically not invertible, and the Hamiltonian is typically not an observable .

Why not? Because we defined an observable to be a function $A : X \to \mathbb{R}$. We can think of this as giving an operator on $L^1(X)$, namely the operator of multiplication by A. That's a nice trick, which we used to good effect last time. However, at least when X is a finite set, this operator will be diagonal in the obvious basis consisting of functions that equal 1 one point of X and zero elsewhere. So, it can only be infinitesimal stochastic if it's zero!

Problem 18. If X is a finite set, show that any operator on $L^1(X)$ that's both diagonal and infinitesimal stochastic must be zero.

11.6 The Hille–Yosida theorem

I've now told you everything you really need to know... but not everything I want to say. What happens when X is not a finite set? What are Markov semigroups like then? I can't abide letting this question go unresolved! Unfortunately I only know a partial answer.

We can get a certain distance using the Hille–Yosida theorem, which is much more general.

Definition. A Banach space is vector space with a norm such that any Cauchy sequence converges.

Examples include Hilbert spaces like $L^2(X)$ for any measure space, but also other spaces like $L^1(X)$ for any measure space!

Definition 14. *If V is a Banach space, a 1-parameter semigroup of operators $U(t) : V \to V$ is called a* **contraction semigroup** *if it's strongly continuous and*

$$\|U(t)\psi\| \leq \|\psi\|$$

for all $t \geq 0$ and all $\psi \in V$.

Examples include strongly continuous 1-parameter unitary groups, but also Markov semigroups!

Problem 19. Show any Markov semigroup is a contraction semigroup.

The Hille–Yosida theorem generalizes Stone's theorem to contraction semigroups. In my misspent youth, I spent a lot of time carrying around Yosida's book *Functional Analysis* . Furthermore, Einar Hille was the advisor of my thesis advisor, Irving Segal. Segal generalized the Hille–Yosida theorem to *nonlinear* operators, and I used this generalization a lot back when I studied nonlinear partial differential equations. So, I feel compelled to tell you this theorem:

Hille–Yosida Theorem. Given a contraction semigroup $U(t)$ we can always write

$$U(t) = \exp(tH)$$

for some densely defined operator H such that $H - \lambda I$ has an inverse and

$$\|(H - \lambda I)^{-1}\psi\| \leq \frac{1}{\lambda}\|\psi\|$$

for all $\lambda > 0$ and $\psi \in V$. Conversely, any such operator determines a strongly continuous 1-parameter group. For all vectors ψ for which $H\psi$ is well-defined, we have

$$\left.\frac{d}{dt}U(t)\psi\right|_{t=0} = H\psi$$

Moreover, for any of these vectors, if we set

$$\psi(t) = U(t)\psi$$

we have

$$\frac{d}{dt}\psi(t) = H\psi(t)$$

If you like, you can take the stuff at the end of this theorem to be what we mean by saying $U(t) = \exp(tH)$. When $U(t) = \exp(tH)$, we say that H **generates** the semigroup $U(t)$.

But now suppose $V = L^1(X)$. Besides the conditions in the Hille–Yosida theorem, what *extra* conditions on H are necessary and sufficient for H to generate a Markov semigroup? In other words, what's a definition of 'infinitesimal stochastic operator' that's suitable not only when X is a finite set, but an arbitrary measure space?

I asked this question on Mathoverflow a few months ago, and so far the answers have not been completely satisfactory.

Some people mentioned the Hille–Yosida theorem, which is surely a step in the right direction, but not the full answer.

Others discussed the special case when $\exp(tH)$ extends to a bounded self-adjoint operator on $L^2(X)$. When X is a finite set, this special case happens precisely when the matrix H_{ij} is *symmetric* : the probability of hopping from j to i equals the probability of hopping from i to j. This is a fascinating special case, not least because when H is both infinitesimal stochastic *and* self-adjoint, we can use it as a Hamiltonian for both stochastic mechanics *and* quantum mechanics! Someday I want to discuss this. However, it's just a special case.

After grabbing people by the collar and insisting that I wanted to know the answer to the question I actually asked—not some vaguely similar question—the best answer seems to be Martin Gisser's reference to this book:

- Zhi-Ming Ma and Michael Röckner, *Introduction to the Theory of (Non-Symmetric) Dirichlet Forms* , Springer, Berlin, 1992.

This book provides a very nice self-contained proof of the Hille–Yosida theorem. On the other hand, it does *not* answer my question in general, but only when the skew-symmetric part of H is dominated (in a certain sense) by the symmetric part.

So, I'm stuck on this front, but that needn't bring the whole project to a halt. We'll just sidestep this question.

For a good well-rounded introduction to Markov semigroups and what they're good for, try:

- Ryszard Rudnicki, Katarzyna Pichòr and Marta Tyran-Kamìnska, Markov semigroups and their applications.

You can also read comments on Azimuth, and make your own comments or ask questions there!

Here are the answers to the problems. The answer to Problem 2 is an expanded version of one given by Graham Jones.

Problem 20. Suppose X is a finite set. Show that any isometry $U : L^2(X) \to L^2(X)$ is invertible, and its inverse is again an isometry.

Answer. Remember that U being an isometry means that it preserves the inner product:

$$\langle U\psi, U\phi\rangle = \langle \psi, \phi\rangle$$

and thus it preserves the L^2 norm

$$\|U\psi\| = \|\psi\|$$

given by $\|\psi\| = \langle \psi, \psi\rangle^{1/2}$. It follows that if $U\psi = 0$, then $\psi = 0$, so U is one-to-one. Since U is a linear operator from a *finite-dimensional* vector space to itself, U must therefore also be onto. Thus U is invertible, and because U preserves the inner product, so does its inverse: given $\psi, \phi \in L^2(X)$ we have

$$\langle U^{-1}\phi, U^{-1}\psi\rangle = \langle \phi, \psi\rangle$$

since we can write $\phi' = U^{-1}\phi$, $\psi' = U^{-1}\psi$ and then the above equation says

$$\langle \phi', \psi'\rangle = \langle U\phi', U\psi'\rangle$$

Problem 21. Suppose X is a finite set. Which stochastic operators $U : L^1(X) \to L^1(X)$ have stochastic inverses?

Answer. Suppose the set X has n points. Then the set of stochastic states

$$S = \{\psi : X \to \mathbb{R} : \psi \geq 0, \quad \int \psi = 1\}$$

is a **simplex**. It's an equilateral triangle when $n = 3$, a regular tetrahedron when $n = 4$, and so on.

[PIC off wikipedia]

In general, S has n corners, which are the functions ψ that equal 1 at one point of S and zero elsewhere. Mathematically speaking, S is a **convex set**, and its corners are its **extreme points**: the points that can't be written as **convex combinations** of other points of S in a nontrivial way.

Any stochastic operator U must map S into itself, so if U has an inverse that's also a stochastic operator, it must give a bijection $U : S \to S$. Any linear transformation acting as a bijection between convex sets must map extreme points to extreme points (this is easy to check), so U must map corners to corners in a bijective way. This implies that it comes from a permutation of the points in X.

In other words, any stochastic matrix with an inverse that's also stochastic is a **permutation matrix**: a square matrix with every entry 0 except for a single 1 in each row and each column.

It is worth adding that there are lots of stochastic operators whose inverses are not, in general, stochastic. We can see this in at least two ways.

First, for any measure space X, every stochastic operator $U : L^1(X) \to L^1(X)$ that's 'close to the identity' in this sense:

$$\|U - I\| < 1$$

(where the norm is the operator norm) will be invertible, simply because *every* operator obeying this inequality is invertible! After all, if this inequality holds, we have a convergent geometric series:

$$U^{-1} = \frac{1}{I - (I - U)} = \sum_{n=0}^{\infty} (I - U)^n$$

Second, suppose X is a finite set and H is infinitesimal stochastic operator on $L^1(X)$. Then H is bounded, so the stochastic operator $\exp(tH)$ where $t > 0$ will always have an inverse, namely $\exp(-tH)$. But for t sufficiently small, this inverse $\exp(-tH)$ will only be stochastic if $-H$ is infinitesimal stochastic, and that's only true if $H = 0$.

In something more like plain English: when you've got a finite set of states, you can formally run any Markov process backwards in time, but a lot of those 'backwards-in-time' operators will involve negative probabilities for the system to hop from one state to another!

Problem 22. If X is a finite set, show that any operator on $L^1(X)$ that's both diagonal and infinitesimal stochastic must be zero.

Answer. We are thinking of operators on $L^1(X)$ as matrices with respect to the obvious basis of functions that equal 1 at one point and 0 elsewhere. If H_{ij} is an infinitesimal stochastic matrix, the sum of the entries in each column is zero. If it's diagonal, there's at most one nonzero entry in each column. So, we must have $H = 0$.

Problem 23. Show any Markov semigroup $U(t) : L^1(X) \to L^1(X)$ is a contraction semigroup.

Answer. We need to show

$$\|U(t)\psi\| \leq \|\psi\|$$

for all $t \geq 0$ and $\psi \in L^1(X)$. Here the norm is the L^1 norm, so more explicitly we need to show

$$\int |U(t)\psi| \leq \int |\psi|$$

We can split ψ into its positive and negative parts:

$$\psi = \psi_+ - \psi_-$$

where

$$\psi_\pm \geq 0$$

Since $U(t)$ is stochastic we have

$$U(t)\psi_\pm \geq 0$$

and

$$\int U(t)\psi_\pm = \int \psi_\pm$$

so

$$\begin{array}{rcl} \int |U(t)\psi| &=& \int |U(t)\psi_+ - U(t)\psi_-| \\ &\leq& \int |U(t)\psi_+| + |U(t)\psi_-| \\ &=& \int U(t)\psi_+ + U(t)\psi_- \\ &=& \int \psi_+ + \psi_- \\ &=& \int |\psi| \end{array}$$

12 Comparing the quantum and stochastic versions of Noether's theorem

Unlike some recent posts, this will be very short. I merely want to show you the quantum and stochastic versions of Noether's theorem, side by side.

Having made my sacrificial offering to the math gods last time by explaining how everything generalizes when we replace our finite set X of states by an infinite set or an even more general measure space, I'll now relax and state Noether's theorem only for a finite set. If you're the sort of person who finds that unsatisfactory, you can do the generalization yourself.

12.1 Two versions of Noether's theorem

Let me write the quantum and stochastic Noether's theorem so they look almost the same:

Definition 15. *Let X be a finite set. Suppose H is a self-adjoint operator on $L^2(X)$, and let O be an observable. Then*

$$[O, H] = 0$$

if and only if for all states $\psi(t)$ obeying Schrödinger's equation

$$\frac{d}{dt}\psi(t) = -iH\psi(t)$$

the expected value of O in the state $\psi(t)$ does not change with t.

Theorem. *Let X be a finite set. Suppose H is an infinitesimal stochastic operator on $L^1(X)$, and let O be an observable. Then*

$$[O, H] = 0$$

if and only if for all states $\psi(t)$ obeying the master equation

$$\frac{d}{dt}\psi(t) = H\psi(t)$$

the expected values of O and O^2 in the state $\psi(t)$ do not change with t.

This makes the big difference stick out like a sore thumb: in the quantum version we only need the expected value of O, while in the stochastic version we need the expected values of O and O^2!

Brendan Fong proved the stochastic version of Noether's theorem in Part 11. Now let's do the quantum version.

12.2 Proof of the quantum version

My statement of the quantum version was silly in a couple of ways. First, I spoke of the Hilbert space $L^2(X)$ for a finite set X, but any finite-dimensional Hilbert space will do equally well. Second, I spoke of the "self-adjoint operator" H and the "observable" O, but in quantum mechanics an observable is the *same thing* as a self-adjoint operator!

Why did I talk in such a silly way? Because I was attempting to emphasize the similarity between quantum mechanics and stochastic mechanics. But they're somewhat different. For example, in stochastic mechanics we have two very different concepts: infinitesimal stochastic operators, which *generate symmetries* , and functions on our set X, which *are observables* . But in quantum mechanics something wonderful happens: self-adjoint operators both *generate symmetries* and *are observables!* So, my attempt was a bit strained.

Let me state and prove a less silly quantum version of Noether's theorem, which implies the one above:

Theorem 16. *Suppose H and O are self-adjoint operators on a finite-dimensional Hilbert space. Then*

$$[O, H] = 0$$

if and only if for all states $\psi(t)$ obeying Schrödinger's equation

$$\frac{d}{dt}\psi(t) = -iH\psi(t)$$

the expected value of O in the state $\psi(t)$ does not change with t:

$$\frac{d}{dt}\langle\psi(t), O\psi(t)\rangle = 0$$

Proof. The trick is to compute the time derivative I just wrote down. Using Schrödinger's equation, the product rule, and the fact that H is self-adjoint we get:

$$\begin{aligned}
\frac{d}{dt}\langle\psi(t), O\psi(t)\rangle &= \langle -iH\psi(t), O\psi(t)\rangle + \langle\psi(t), O(-iH\psi(t))\rangle \\
&= i\langle\psi(t), HO\psi(t)\rangle - i\langle\psi(t), OH\psi(t))\rangle \\
&= -i\langle\psi(t), [O, H]\psi(t)\rangle
\end{aligned}$$

So, if $[O, H] = 0$, clearly the above time derivative vanishes. Conversely, if this time derivative vanishes for all states $\psi(t)$ obeying Schrödinger's equation, we know

$$\langle\psi, [O, H]\psi\rangle = 0$$

for all states ψ and thus all vectors in our Hilbert space. Does this imply $[O,H]=0$? Yes, because i times a commutator of a self-adjoint operators is self-adjoint, and for any self-adjoint operator A we have

$$\forall\psi \langle\psi, A\psi\rangle = 0 \qquad \Rightarrow \qquad A = 0$$

This is a well-known fact whose proof goes like this. Assume $\langle\psi, A\psi\rangle = 0$ for all ψ. Then to show $A=0$, it is enough to show $\langle\phi, A\psi\rangle = 0$ for all ϕ and ψ. But we have a marvelous identity:

$$\begin{aligned}\langle\phi, A\psi\rangle \quad = \quad & \tfrac{1}{4}\left(\langle\phi+\psi,\, A(\phi+\psi)\rangle - \langle\psi-\phi,\, A(\psi-\phi)\rangle\right. \\ & \left. +i\langle\psi+i\phi,\, A(\psi+i\phi)\rangle - i\langle\psi-i\phi,\, A(\psi-i\phi)\rangle\right)\end{aligned}$$

and all four terms on the right vanish by our assumption. □

The marvelous identity up there is called the **polarization identity**. In plain English, it says: if you know the diagonal entries of a self-adjoint matrix in every basis, you can figure out *all* the entries of that matrix in every basis.

Why is it called the 'polarization identity'? I think because it shows up in optics, in the study of polarized light.

12.3 Comparison

In both the quantum and stochastic cases, the time derivative of the expected value of an observable O is expressed in terms of its commutator with the Hamiltonian. In the quantum case we have

$$\frac{d}{dt}\langle\psi(t), O\psi(t)\rangle = -i\langle\psi(t), [O,H]\psi(t)\rangle$$

and for the right side to *always* vanish, we need $[O,H]=0$, thanks to the polarization identity. In the stochastic case, a perfectly analogous equation holds:

$$\frac{d}{dt}\int O\psi(t) = \int [O,H]\psi(t)$$

but now the right side can always vanish even without $[O,H]=0$. We saw a counterexample in Part 11. There is nothing like the polarization identity to save us! To get $[O,H]=0$ we need a supplementary hypothesis, for example the vanishing of

$$\frac{d}{dt}\int O^2\psi(t)$$

Okay! Starting next time we'll change gears and look at some more examples of stochastic Petri nets and Markov processes, including some from chemistry. After some more of that, I'll move on to networks of other sorts. There's a really big picture here, and I'm afraid I've been getting caught up in the details of a tiny corner.

You can also read comments on Azimuth, and make your own comments or ask questions there!

13 An example: chemistry and the Desargues graph

We've been doing a lot of hard work lately. Let's take a break and think about a fun example from chemistry!

13.1 The ethyl cation

Suppose you start with a molecule of ethane, which has 2 carbons and 6 hydrogens arranged like this:

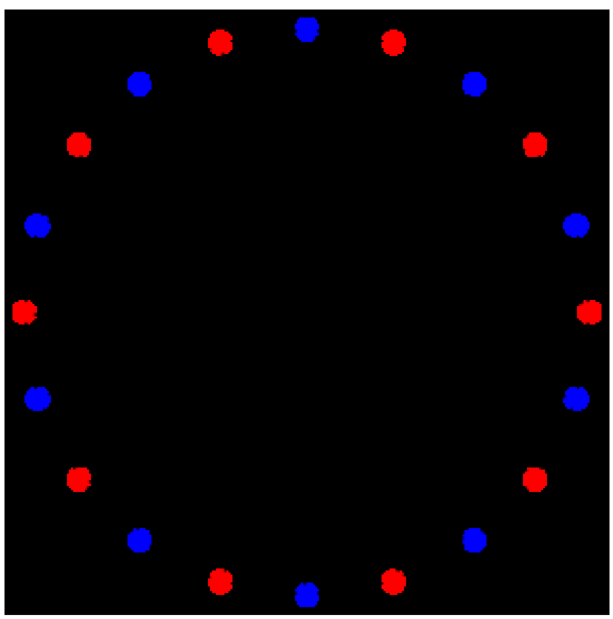
[PIC jpg]

Then suppose you remove one hydrogen. The result is a positively charged ion, or 'cation'. When I was a kid, I thought the opposite of a cation should be called a 'dogion'. Alas, it's not.

This particular cation, formed from removing one hydrogen from an ethane molecule, is called an 'ethyl cation'. People used to think it looked like this:

[PIC jpg]

They also thought a hydrogen could hop from the carbon with 3 hydrogens attached to it to the carbon with 2. So, they drew a graph with a vertex for each way the hydrogens could be arranged, and an edge for each hop. It looks really cool:



The red vertices come from arrangements where the *first* carbon has 2 hydrogens attached to it, and the blue vertices come from those where the *second* carbon has 2 hydrogens attached to it. So, each edge goes between a red vertex and a blue vertex.

This graph has 20 vertices, which are arrangements or 'states' of the ethyl cation. It has 30 edges, which are hops or 'transitions'. Let's see why those numbers are right.

First I need to explain the rules of the game. The rules say that the 2 carbon atoms are distinguishable: there's a 'first' one and a 'second' one. The 5 hydrogen atoms are also distinguishable. But, all we care about is which carbon atom each hydrogen is bonded to: we don't care about further details of its location. And we require that 2 of the hydrogens are bonded to one carbon, and 3 to the other.

If you're a physicist, you may wonder why the rules work this way: after all, at a fundamental level, identical particles *aren't* really distinguishable. I'm afraid I can't give a fully convincing explanation right now: I'm just reporting the rules as they were told to me!

Given these rules, there are 2 choices of which carbon has two hydrogens attached to it. Then there are

$$\binom{5}{2} = \frac{5 \times 4}{2 \times 1} = 10$$

choices of which two hydrogens are attached to it. This gives a total of $2 \times 10 = 20$ states. These are the vertices of our graph: 10 red and 10 blue.

The edges of the graph are transitions between states. Any hydrogen in the group of 3 can hop over to the group of 2. There are 3 choices for which hydrogen atom makes the jump. So, starting from any vertex in the graph there are 3 edges. This means there are $3 \times 20/2 = 30$ edges.

Why divide by 2? Because each edge touches 2 vertices. We have to avoid double-counting them.

13.2 The Desargues graph

The idea of using this graph in chemistry goes back to this paper:

- A. T. Balaban, D. F?rca?iu and R. B?nic?, Graphs of multiple 1,2-shifts in carbonium ions and related systems, *Rev. Roum. Chim.* **11** (1966), 1205.

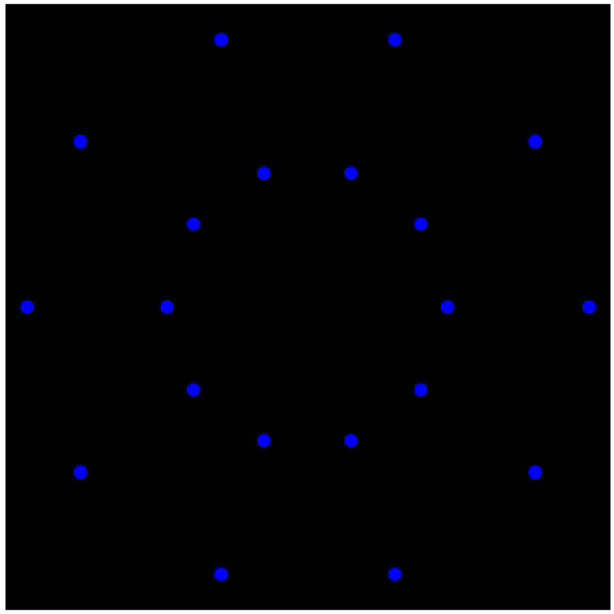
This paper is famous because it was the first to use graphs in chemistry to describe molecular transitions, as opposed to using them as pictures of molecules!

But this particular graph was already famous for other reasons. It's called the **Desargues-Levi graph**, or **Desargues graph** for short:

- Desargues graph, Wikipedia.

Later I'll say why it's called this.

There are lots of nice ways to draw the Desargues graph. For example:



The reason why we can draw such pretty pictures is that the Desargues graph is very symmetrical. Clearly any permutation of the 5 hydrogens acts as a symmetry of the graph, and so does any permutation of the 2 carbons. This gives a symmetry group $S_5 \times S_2$, which has $5! \times 2! = 240$ elements. And in fact this turns out to be the full symmetry group of the Desargues graph.

The Desargues graph, its symmetry group, and its applications to chemistry are discussed here:

- Milan Randic, Symmetry properties of graphs of interest in chemistry: II: Desargues-Levi graph, *Int. Jour. Quantum Chem.* **15** (1997), 663-682.

13.3 The ethyl cation, revisited

We can try to describe the ethyl cation using probability theory. If at any moment its state corresponds to some vertex of the Desargues graph, and it hops randomly along edges as time goes by, it will trace out a random walk on the Desargues graph. This is a nice example of a Markov process!

We could also try to describe the ethyl cation using quantum mechanics. Then, instead of having a *probability* of hopping along an edge, it has an *amplitude* of doing so. But as we've seen, a lot of similar math will still apply.

It should be fun to compare the two approaches. But I bet you're wondering which approach is correct. This is a somewhat tricky question, at least for me. The answer would seem to depend on how much the ethyl cation is interacting with its environment—for example, bouncing off other molecules. When a system is interacting a lot with its environment, a probabilistic approach seems to be more appropriate. The relevant buzzword is 'environmentally induced decoherence'.

However, there's something much more basic I have tell you about.

After the paper by Balaban, F.rca.iu and B.nic. came out, people gradually realized that the ethyl cation *doesn't really look like the drawing I showed you!* It's what chemists call 'nonclassical' ion. What they mean is this: its actual structure is not what you get by taking the traditional ball-and-stick model of an ethane molecule and ripping off a hydrogen. The ethyl cation really looks like this:

[PIC jpg]

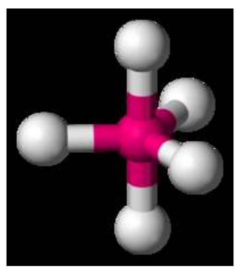
For more details, and pictures that you can actually rotate, see:

- Stephen Bacharach, Ethyl cation, Computational Organic Chemistry.

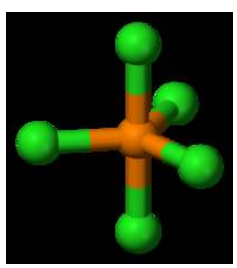
So, if we stubbornly insist on applying the Desargues graph to realistic chemistry, we need to find some other molecule to apply it to.

13.4 Trigonal bipyramidal molecules

Luckily, there are lots of options! They're called trigonal bipyramidal molecules. They look like this:

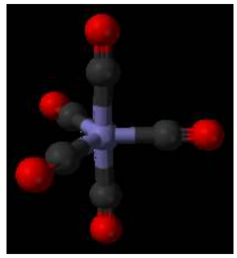


The 5 balls on the outside are called 'ligands': they could be atoms or bunches of atoms. In chemistry, 'ligand' just means something that's stuck onto a central thing. For example, in phosphorus pentachloride the ligands are chlorine atoms, all attached to a central phosphorus atom:



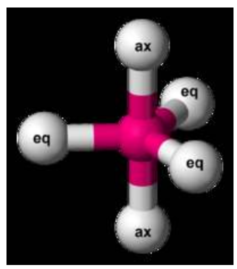
It's a colorless solid, but as you might expect, it's pretty nasty stuff: it's not flammable, but it reacts with water or heat to produce toxic chemicals like hydrogen chloride.

Another example is iron pentacarbonyl, where 5 carbon-oxygen ligands are attached to a central iron atom:



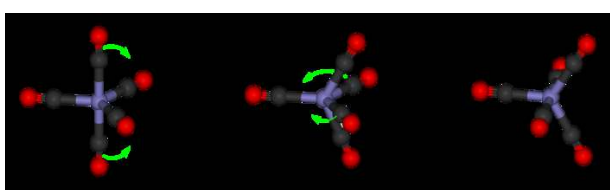
You can make this stuff by letting powdered iron react with carbon monoxide. It's a straw-colored liquid with a pungent smell!

Whenever you've got a molecule of this shape, the ligands come in two kinds. There are the 2 'axial' ones, and the 3 'equatorial' ones:



And the molecule has 20 states... but only if count the states a certain way. We have to treat all 5 ligands as distinguishable, but think of two arrangements of them as the same if we can rotate one to get the other. The trigonal bipyramid has a rotational symmetry group with 6 elements. So, there are 5! / 6 = 20 states.

The transitions between states are devilishly tricky. They're called **pseudorotations**, and they look like this:



If you look very carefully, you'll see what's going on. First the 2 axial ligands move towards each other to become equatorial. Now the equatorial ones are no longer in the horizontal plane: they're in the plane facing us! Then 2 of the 3 equatorial ones swing out to become axial. This fancy dance is called the Berry pseudorotation mechanism.

To get from one state to another this way, we have to pick 2 of the 3 equatorial ligands to swing out and become axial. There are 3 choices here. So, if we draw a graph with states as vertices and transitions as edges, it will have 20 vertices and $20 \times 3/2 = 30$ edges. That sounds suspiciously like the Desargues graph!

Problem 24. Show that the graph with states of a trigonal bipyramidal molecule as vertices and pseudorotations as edges is indeed the Desargues graph.

I think this fact was first noticed here:

- Paul C. Lauterbur and Fausto Ramirez, Pseudorotation in trigonal-bipyramidal molecules, *J. Am. Chem. Soc.* **90** (1968), 6722.6726.

Okay, enough for now! Next time I'll say more about the Markov process or quantum process corresponding to a random walk on the Desargues graph. But since the Berry pseudorotation mechanism is so hard to visualize, I'll *pretend* that the ethyl cation looks like this:

[PIC jpeg]

and I'll use this picture to help us think about the Desargues graph.

That's okay: everything we'll figure out can easily be translated to apply to the real-world situation of a trigonal bipyramidal molecule. The virtue of math is that when two situations are 'mathematically the same', or 'isomorphic', we can talk about either one, and the results automatically apply to the other. This is true even if the one we talk about doesn't actually exist in the real world!

13.5 Drawing the Desargues graph

Before we quit for today, let's think a bit more about how we draw the Desargues graph. For this it's probably easiest to go back to our naive model of an ethyl cation:

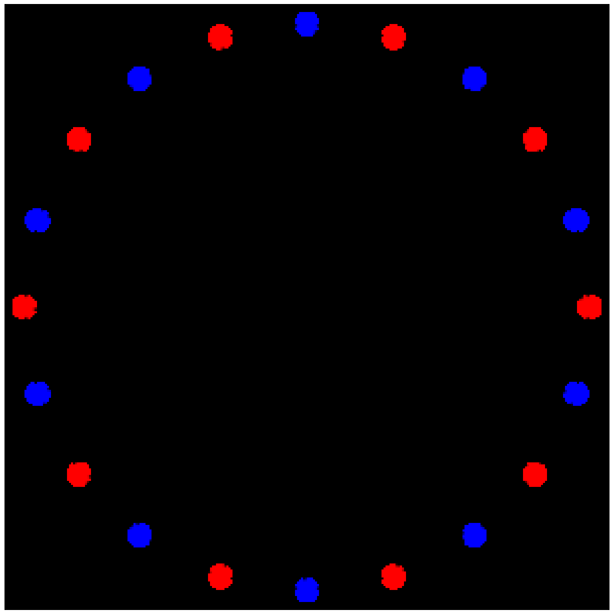
[PIC jpeg]

Even though ethyl cations don't really look like this, and we should be talking about some trigonal bipyramidal molecule instead, it won't affect the math to come. Mathematically, the two problems are isomorphic! So let's stick with this nice simple picture.

We can be a bit more abstract, though. A state of the ethyl cation is like having 5 balls, with 3 in one pile and 2 in the other. And we can focus on the first pile and forget the second, because whatever isn't in the first pile must be in the second.

Of course a mathematician calls a pile of things a 'set', and calls those things 'elements'. So let's say we've got a set with 5 elements. Draw a red dot for each 2-element subset, and a blue dot for each 3-element subset. Draw an edge between a red dot and a blue dot whenever the 2-element subset is *contained* in the 3-element subset. We get the Desargues graph.

That's true by definition. But I never proved that this graph looks like this:



I won't now, either. I'll just leave it as a problem:

Problem 25. If we define the Desargues graph to have vertices corresponding to 2- and 3-element subsets of a 5-element set, with an edge between vertices when one subset is contained in another, why does it look like the picture above?

To draw a picture we know is correct, it's actually easier to start with a big graph that has vertices for *all* the subsets of our 5-element set. If we draw an edge whenever an n-element subset is contained in an $(n+1)$-element subset, the Desargues graph will be sitting inside this big graph.

Here's what the big graph looks like: [PIC from wikipedia]

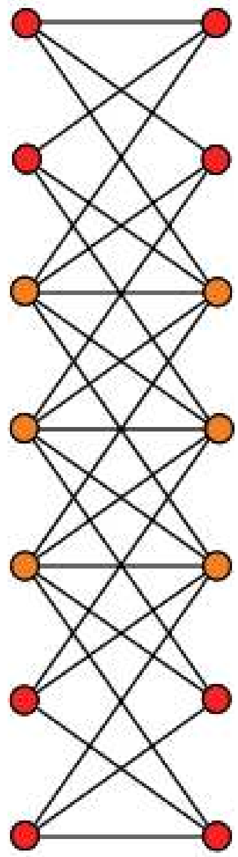
This graph has 2^5 vertices. It's actually a picture of a 5-dimensional hypercube! The vertices are arranged in columns. There's

- one 0-element subset,
- five 1-element subsets,
- ten 2-element subsets,
- ten 3-element subsets,
- five 4-element subsets,
- one 5-element subset.

So, the numbers of vertices in each column go like this:

$$1 \quad 5 \quad 10 \quad 10 \quad 5 \quad 1$$

which is a row in Pascal's triangle. We get the Desargues graph if we keep only the vertices corresponding to 2- and 3-element subsets, like this:



It's less pretty than our earlier picture, but at least there's no mystery to it. Also, it shows that the Desargues graph can be generalized in various ways. For example, there's a theory of bipartite Kneser graphs $H(n, k)$. The Desargues graph is $H(5, 2)$.

13.6 Desargues' theorem

Finally, I can't resist answering this question: why is this graph called the 'Desargues graph'? This name comes from Desargues' theorem, a famous result in projective geometry. Suppose you have two triangles ABC and abc, like this:

[PIC of wikipedia]

Suppose the lines Aa, Bb, and Cc all meet at a single point, the 'center of perspectivity'. Then the point of intersection of ab and AB, the point of intersection of ac and AC, and the point of intersection of bc and BC all lie on a single line, the 'axis of perspectivity'. The converse is true too. Quite amazing!

The **Desargues configuration** consists of all the actors in this drama:

- 10 points: A, B, C, a, b, c, the center of perspectivity, and the three points on the axis of perspectivity and
- 10 lines: Aa, Bb, Cc, AB, AC, BC, ab, ac, bc and the axis of perspectivity

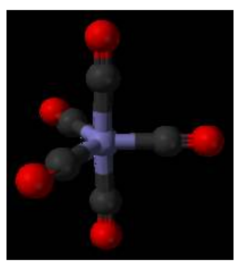
Given any configuration of points and lines, we can form a graph called its **Levi graph** by drawing a vertex for each point or line, and drawing edges to indicate which points lie on which lines. And now for the punchline: Levi graph of the Desargues configuration is the 'Desargues-Levi graph'!—or Desargues graph, for short.

Alas, I don't know how this is relevant to anything I've discussed. For now it's just a tantalizing curiosity.

I didn't get any solutions of Problem 1 the first time I posted this, so I reposed it on Azimuth and the n -Category Café. That brought out lots of great solutions, which I've used to compose the answer here. I'm especially indebted to J.M. Allen, Twan van Laarhoven, Peter Morgan and Johan Swanljung.

Problem 26. Show that the graph with states of a trigonal bipyramidal molecule as vertices and pseudorotations as edges is indeed the Desargues graph.

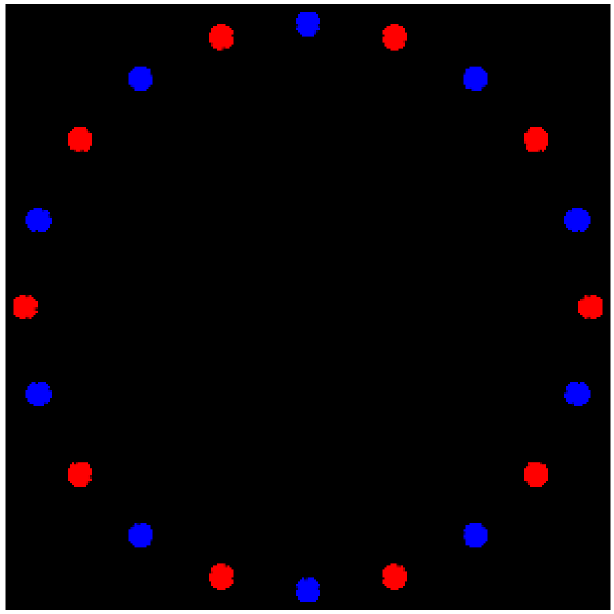
Answer. To be specific, let's use iron pentacarbonyl as our example of a trigonal bipyramidal molecule:



It suffices to construct a 1-1 correspondence between the states of this molecule and those of the ethyl cation, such that two states of this molecule are connected by a transition if and only if the same holds for the corresponding states of the ethyl cation.

Here's the key idea: the ethyl cation has 5 hydrogens, with 2 attached to one carbon and 3 attached to the other. Similarly, the trigonal bipyramidal molecule has 5 carbonyl grops, with 2 axial and 3 equatorial. We'll use this resemblance to set up our correspondence.

There are various ways to describe states of the ethyl cation, but this is the best for us. Number the hydrogens 1,2,3,4,5. Then a state of the ethyl cation consists of a partition of the set $\{1,2,3,4,5\}$ into a 2-element set and a 3-element set, together with one extra bit of information, saying which carbon has 2 hydrogens attached to it. This extra bit is the color here:

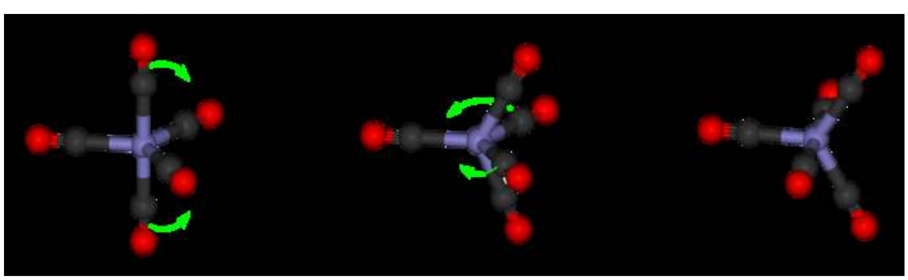


What do transitions look like in this description? When a transition occurs, two hydrogens that belonged to the 3-element set now become part of the 2-element set. Meanwhile, both hydrogens that belonged to the 2-element set now become part of the 3-element set. (Ironically, the one hydrogen that hops is the one that stays in the 3-element set.) Moreover, the extra bit of information changes. That's why every edge goes from a red dot to a blue one, or vice versa.

So, to solve the problem, we need to show that the same description also works for the states and transitions of iron pentacarbonyl!

In other words, we need to describe its states as ways of partitioning the set 1,2,3,4,5 into a 2-element set and a 3-element set, together with one extra bit of information. And we need its transitions to switch two elements of the 2-element set with two of the 3-element set, while changing that extra bit.

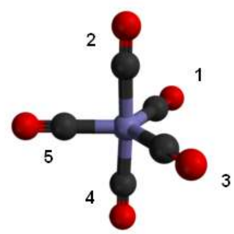
To do this, number the carbonyl groups 1,2,3,4,5. The 2-element set consists of the axial ones, while the 3-element set consists of the equatorial ones. When a transition occurs, two of axial ones trade places with two of the equatorial ones, like this:



So, now we just need to figure out what that extra bit of information is, and why it always *changes* when a transition occurs.

Here's how we calculate that extra bit. Hold the iron pentacarbonyl molecule vertically with one of the equatorial carbonyl groups pointing to your left. Remember, the carbonyl groups are numbered. So, write a list of these numbers, say (a,b,c,d,e), where a is the top axial one, b,c,d are the equatorial ones listed in counterclockwise order starting from the one pointing left, and e is the bottom axial one. This list is some permutation of the list (1,2,3,4,5). Take the sign of this permutation to be our bit!

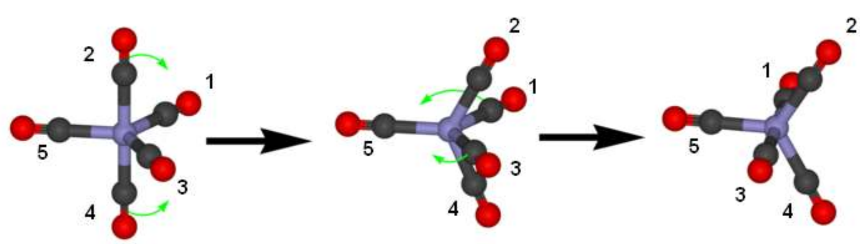
Let's do an example:



Here we get the list (2,5,3,1,4) since 2 is on top, 4 is on bottom, and 5,3,1 are the equatorial guys listed counterclockwise starting from the one at left. The list (2,5,3,1,4) is an odd permutation of (1,2,3,4,5), so our bit of information is *odd* .

Of course we must check that this bit is well-defined: namely, that it doesn't change if we *rotate* the molecule. Rotating it a third of a turn gives an even permutation of the equatorial guys and leaves the axial ones alone, so this is an even permutation. Flipping it over gives an odd permutation of the equatorial guys, but it also gives an odd permutation of the axial ones, so this too is an even permutation. So, rotating the molecule doesn't change the sign of the permutation we compute from it. The sign is thus a well-defined function of the state of the molecule.

Next we must to check that this sign changes whenever our molecule undergoes a transition. For this we need to check that any transition changes our list of numbers by an odd permutation. Since all transitions are conjugate in the permutation group, it suffices to consider one example:



Here we started with a state giving the list (2,5,3,1,4). The transition ttakes us to a state that gives the list (3,5,4,2,1) if we hold the molecule so that 3 is pointing up and 5 to the left. The reader can check that going from one list to another requires an *odd* permutation. So we're done.

Tracy Hall's discussion of the above problem was very interesting and erudite. He also addressed an interesting subsidiary problem, namely: what graph do we get if we discard that extra bit of information that says which carbon in the ethyl cation has 3 hydrogens attached to it and which has 2? The answer is the Petersen graph:

[PIC of wikipedia]

Tracy wrote:

"As some comments have pointed out over on Azimuth, in both cases there are ten underlying states which simply pick out two of the five pendant atoms as special, together with an extra parity bit (which can take either value for any of the ten states), giving twenty states in total. The correspondence of the ten states is clear: an edge exists between state A and state B, in either case, if and only if the two special atoms of state A are disjoint from the two special atoms of state B. This is precisely one definition of the Petersen graph (a famous 3-valent graph on 10 vertices that shows up as a small counterexample to lots of naove conjectures). Thus the graph in either case is a double cover of the Petersen graph—but that does not uniquely specify it, since, for example, both the Desargues graph and the dodecahedron graph are double covers of the Petersen graph.

For a labeled graph, each double cover corresponds uniquely to an element of the $\mathbb{Z}/2\mathbb{Z}$ cohomology of the graph (for an unlabeled graph, some of the double covers defined in this way may turn out to be isomorphic). Cohomology over $\mathbb{Z}/2\mathbb{Z}$ takes any cycle as input and returns either 0 or 1, in a consistent way (the output of a $\mathbb{Z}/2\mathbb{Z}$ sum of cycles is the sum of the outputs on each cycle). The double cover has two copies of everything in the base (Petersen) graph, and as you follow all the way around a cycle in the base, the element of cohomology tells you whether you come back to the same copy (for 0) or the other copy (for 1) in the double cover, compared to where you started.

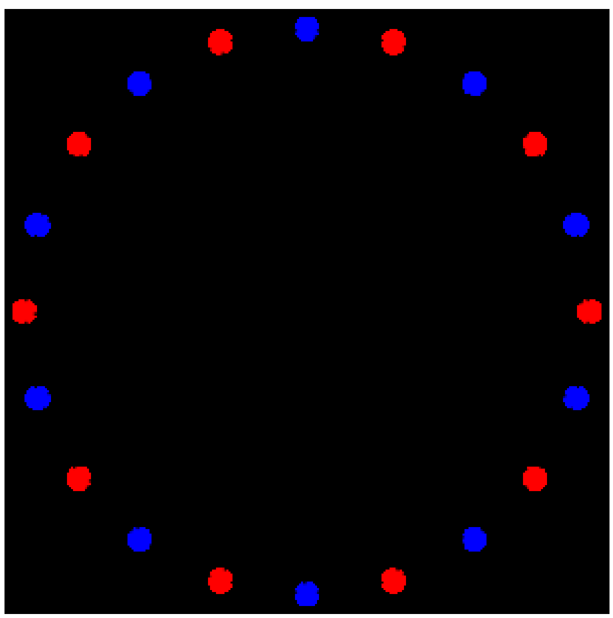
One well-defined double cover for any graph is the one which simply switches copies for every single edge (this corresponding to the element of cohomology which is 1 on all odd cycles and 0 on all even cycles). This always gives a double cover which is a bipartite graph, and which is connected if and only if the base graph is connected and not bipartite. So if we can show that in both cases (the fictitious ethyl cation and phosphorus pentachloride) the extra parity bit can be defined in such a way that it switches on every transition, that will show that we get the Desargues graph in both cases.

The fictitious ethyl cation is easy: the parity bit records which carbon is which, so we can define it as saying which carbon has three neighbors. This switches on every transition, so we are done. Phosphorus pentachloride is a bit trickier; the parity bit distinguishes a labeled molecule from its mirror image, or enantiomer. As has already been pointed out on both sites, we can use the parity of a permutation to distinguish this, since it happens that the orientationpreserving rotations of the molecule, generated by a three-fold rotation acting as a three-cycle and by a two-fold rotation acting as a pair of two-cycles, are all even permutations, while the mirror image that switches only the two special atoms is an odd permutation. The pseudorotation can be followed by a quarter turn to return the five chlorine atoms to the five places previously occupied by chlorine atoms, which makes it act as a four-cycle, an odd permutation. Since the parity bit in this case also can be defined in such a way that it switches on every transition, the particular double cover in each case is the Desargues graph—a graph I was surprised to come across here, since just this past week I have been working out some combinatorial matrix theory for the same graph!

The five chlorine atoms in phosphorus pentachloride lie in six triangles which give a triangulation of the 2-sphere, and another way of thinking of the pseudorotation is that it corresponds to a Pachner move or bistellar flip on this triangulation—in particular, any bistellar flip on this triangulation that preserves the number of triangles and the property that all vertices in the triangulation have degree at least three corresponds to a pseudorotation as described."

The next problem was both posed and answered by Greg Egan on Azimuth.

Problem 2. If we define the Desargues graph to have vertices corresponding to 2- and 3-element subsets of a 5-element set, with an edge between vertices when one subset is contained in another, why does it look like this picture?



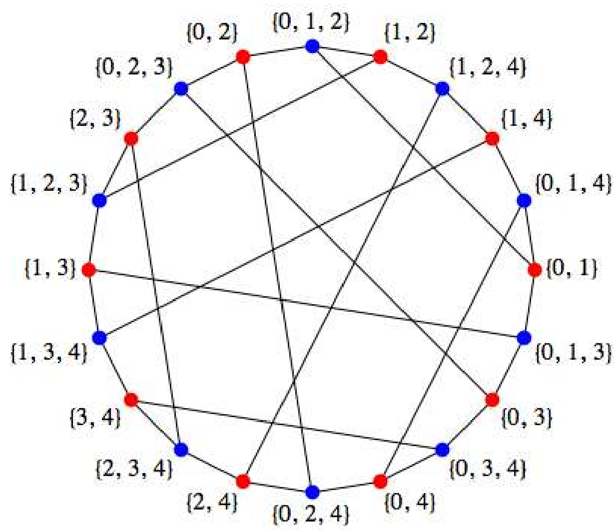
Answer. For $i = 0, ..., 4$, and with all the additions below modulo 5, define five pairs of red dots as:

$$\{i, i+1\}, \{i+1, i+4\}$$

and five pairs of blue dots as:

$\{i, i+1, i+4\}, \{i+1, i+2, i+4\}$

The union of the ith pair of red dots is the first of the ith pair of blue dots, and the union of the second of the ith pair of red dots and the first of the $(i+1)$th pair of red dots is the second of the ith pair of blue dots. So if we form a 20-sided polygon whose vertices are alternating red and blue dots generated in this order, all the edges of the polygon will join red dots to blue dots of which they are subsets:



The first red dot of the ith pair is also a subset of the first blue dot of the $(i+1)$th pair:

$$\{i+1, i+2, i\}$$

which gives the five short chords in the picture, while the second red dot of the ith pair is a subset of the second blue dot of the $(i+2)$ pair:

$$\{i+3, i+4, i+1\}$$

which gives the five long chords in the picture.

14 Markov and quantum processes coming from graph Laplacians

Last time we saw how to get a graph whose vertices are states of a molecule and whose edges are transitions between states. We focused on two beautiful but not completely realistic examples that both give rise to the same highly symmetrical graph: the 'Desargues graph'.

Today I'll show you how a molecule can carry out a random walk on this graph. Then I'll get to work showing how *any* graph gives:

- A Markov process, namely a random walk on the graph.
- A quantum process, where instead of having a *probability* to hop from vertex to vertex as time passes, we have an *amplitude* .

The trick is to use an operator called the 'graph Laplacian', a discretized version of the Laplacian which happens to be both infinitesimal stochastic and self-adjoint. As we saw in Part 12, such an operator will give rise *both* to a Markov process *and* a quantum process (that is, a one-parameter unitary group).

The most famous operator that's both infinitesimal stochastic and self-adjoint is the Laplacian, ∇^2. Because it's both, the Laplacian shows up in *two* important equations: one in stochastic mechanics, the other in quantum mechanics.

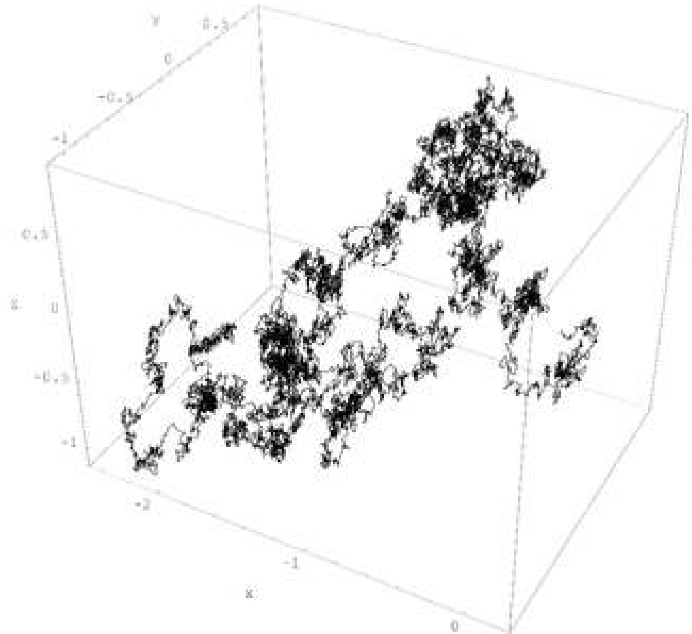
- The **heat equation**:

$$\frac{d}{dt}\psi = \nabla^2\psi$$

describes how the probability $\psi(x)$ of a particle being at the point x smears out as the particle randomly walks around:

[PIC as gif however]

The corresponding Markov process is called 'Brownian motion'.



The **Schrödinger equation**:

$$\frac{d}{dt}\psi = -i\nabla^2\psi$$

describes how the amplitude $\psi(x)$ of a particle being at the point x wiggles about as the particle 'quantumly' walks around.

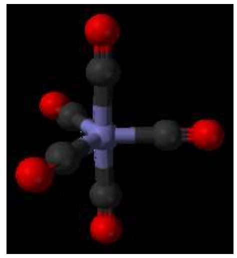
Both these equations have analogues where we replace space by a graph, and today I'll describe them.

14.1 A random walk on the Desargues grap

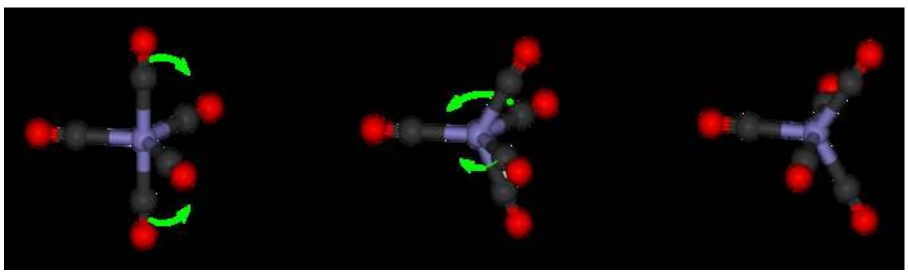
Back to business! I've been telling you about the analogy between quantum mechanics and stochastic mechanics. This analogy becomes especially interesting in chemistry, which lies on the uneasy borderline between quantum and stochastic mechanics.

Fundamentally, of course, atoms and molecules are described by quantum mechanics. But sometimes chemists describe chemical reactions using stochastic mechanics instead. When can they get away with this? Apparently whenever the molecules involved are big enough and interacting with their environment enough for 'decoherence' to kick in. I won't attempt to explain this now.

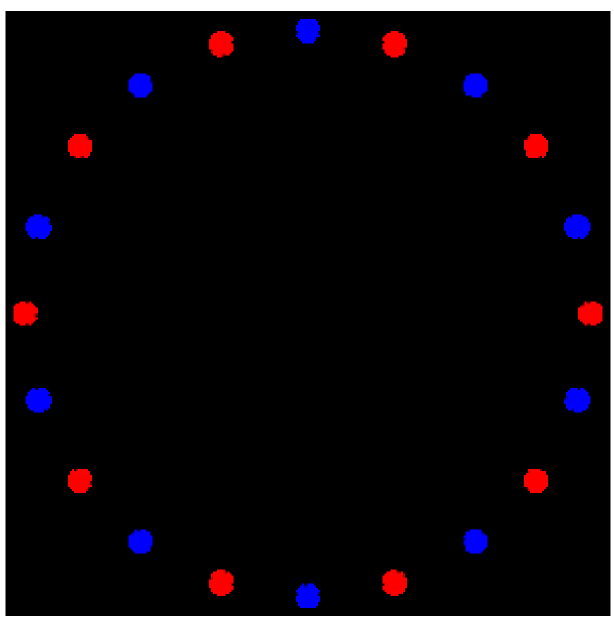
Let's imagine we have a molecule of iron pentacarbonyl with—here's the unrealistic part, but it's not really too bad—distinguishable carbonyl groups:



Iron pentacarbonyl is liquid at room temperatures, so as time passes, each molecule will bounce around and occasionally do a maneuver called a 'pseudorotation':



We can approximately describe this process as a random walk on a graph whose *vertices* are *states* of our molecule, and whose *edges* are *transitions* between states—namely, pseudorotations. And as we saw last time, this graph is the Desargues graph:



Note: all the transitions are reversible here. And thanks to the enormous amount of symmetry, the rates of all these transitions must be equal.

Let's write V for the set of vertices of the Desargues graph. A probability distribution of states of our molecule is a function

$$\psi : V \to [0, \infty)$$

with

$$\sum_{x \in V} \psi(x) = 1$$

We can think of these probability distributions as living in this vector space:

$$L^1(V) = \{\psi : V \to \mathbb{R}\}$$

I'm calling this space L^1 because of the general abstract nonsense explained in Part 12: probability distributions on any measure space live in a vector space called L^1. Today that notation is overkill, since *every* function on V lies in L^1. But please humor me.

The point is that we've got a general setup that applies here. There's a Hamiltonian:

$$H : L^1(V) \to L^1(V)$$

describing the rate at which the molecule randomly hops from one state to another... and the probability distribution $\psi \in L^1(X)$ evolves in time according to the equation:

$$\frac{d}{dt}\psi(t) = H\psi(t)$$

But what's the Hamiltonian H? It's very simple, because it's equally likely for the state to hop from any vertex to any other vertex that's connected to that one by an edge. Why? Because the problem has so much symmetry that nothing else makes sense.

So, let's write E for the set of edges of the Desargues graph. We can think of this as a subset of $V \times V$ by saying $(x, y) \in E$ when x is connected to y by an edge. Then

$$(H\psi)(x) = \sum_{y \mathrm{such that} (x,y) \in E} \psi(y) \quad - \quad 3\psi(x)$$

We're subtracting $3\psi(x)$ because there are 3 edges coming out of each vertex x, so this is the rate at which the probability of staying at x decreases. We could multiply this Hamiltonian by a constant if we wanted the random walk to happen faster or slower... but let's not.

The next step is to solve this discretized version of the heat equation:

$$\frac{d}{dt}\psi(t) = H\psi(t)$$

Abstractly, the solution is easy:

$$\psi(t) = \exp(tH)\psi(0)$$

But to actually compute $\exp(tH)$, we might want to diagonalize the operator H. In this particular example, we can do this by taking advantage of the enormous symmetry of the Desargues graph. But let's not do this just yet. First let's draw some general lessons from this example.

14.2 Graph Laplacians

The Hamiltonian we just saw is an example of a 'graph Laplacian'. We can write down such a Hamiltonian for any graph, but it gets a tiny bit more complicated when different vertices have different numbers of edges coming out of them.

The word 'graph' means lots of things, but right now I'm talking about **simple graphs**. Such a graph has a set of vertices V and a set of edges $E \subseteq V \times V$, such that

$$(x, y) \in E \Rightarrow (y, x) \in E$$

which says the edges are undirected, and

$$(x, x) \notin E$$

which says there are no loops. Let $d(x)$ be the **degree** of the vertex x, meaning the number of edges coming out of it.

Then the **graph Laplacian** is this operator on $L^1(V)$:

$$(H\psi)(x) = \sum_{y \mathrm{such that} (x,y)\in E} \psi(y) \quad - \quad d(x)\psi(x)$$

There is a huge amount to say about graph Laplacians! If you want, you can get started here:

- Michael William Newman, *The Laplacian Spectrum of Graphs* , Masters Thesis, Department of Mathematics, University of Manitoba, 2000.

But for now, let's just say that $\exp(tH)$ is a Markov process describing a random walk on the graph, where hopping from one vertex to any neighboring vertex has unit probability per unit time. We can make the hopping faster or slower by multiplying H by a constant. And here is a good time to admit that most people use a graph Laplacian that's the negative of ours, and write time evolution as $\exp(-tH)$. The advantage is that then the eigenvalues of the Laplacian are ≥ 0.

But what matters most is this. We can write the operator H as a matrix whose entry H_{xy} is 1 when there's an edge from x to y and 0 otherwise, except when $x = y$, in which case the entry is $-d(x)$. And then:

Problem 27. Show that for any finite graph, the graph Laplacian H is **infinitesimal stochastic**, meaning that:

$$\sum_{x \in V} H_{xy} = 0$$

and

$$x \neq y \Rightarrow H_{xy} \geq 0$$

This fact implies that for any $t \geq 0$, the operator $\exp(tH)$ is stochastic—just what we need for a Markov process.

But we could also use H as a Hamiltonian for a *quantum* system, if we wanted. Now we think of $\psi(x)$ as the *amplitude* for being in the state $x \in V$. But now ψ is a function

$$\psi : V \to \mathbb{C}$$

with

$$\sum_{x \in V} |\psi(x)|^2 = 1$$

We can think of this function as living in the Hilbert space

$$L^2(V) = \{\psi : V \to \mathbb{C}\}$$

where the inner product is

$$\langle \phi, \psi \rangle = \sum_{x \in V} \overline{\phi(x)} \psi(x)$$

Problem 28. Show that for any finite graph, the graph Laplacian $H : L^2(V) \to L^2(V)$ is **self-adjoint**, meaning that:

$$H_{xy} = \overline{H}_{yx}$$

This implies that for any $t \in \mathbb{R}$, the operator $\exp(-itH)$ is unitary—just what we need for one-parameter unitary group. So, we can take this version of Schrödinger's equation:

$$\frac{d}{dt}\psi = -iH\psi$$

and solve it:

$$\psi(t) = \exp(-itH)\psi(0)$$

and we'll know that time evolution is unitary!

So, we're in a dream world where we can do stochastic mechanics *and* quantum mechanics with the same Hamiltonian. I'd like to exploit this somehow, but

I'm not quite sure how. Of course physicists like to use a trick called Wick rotation where they turn quantum mechanics into stochastic mechanics by replacing time by imaginary time. We can do that here. But I'd like to do something new, special to this context.

Maybe I should learn more about chemistry and graph theory. Of course, graphs show up in at least two ways: first for drawing molecules, and second for drawing states and transitions, as I've been doing. These books are supposed to be good:

- Danail Bonchev and D.H. Rouvray, eds., *Chemical Graph Theory: Introduction and Fundamentals* , Taylor and Francis, 1991.
- Nenad Trinajstic, *Chemical Graph Theory* , CRC Press, 1992.
- R. Bruce King, *Applications of Graph Theory and Topology in Inorganic Cluster Coordination Chemistry* , CRC Press, 1993.

The second is apparently the magisterial tome of the subject. The prices of these books are absurd: for example, Amazon sells the first for 300 dollars, and the second for 222. Luckily the university here should have them...

14.3 The Laplacian of the Desargues graph

Greg Egan figured out a nice way to explicitly describe the eigenvectors of the Laplacian H for the Desargues graph. This lets us explicitly solve the heat equation

$$\frac{d}{dt}\psi = H\psi$$

and also the Schrödinger equation on this graph.

First there is an obvious eigenvector: any constant function. Indeed, for any finite graph, any constant function ψ is an eigenvector of the graph Laplacian with eigenvalue zero:

$$H\psi = 0$$

so that

$$\exp(tH)\psi = \psi$$

This reflects the fact that the evenly smeared probability distribution is an equilibrium state for the random walk described by heat equation. For a connected graph this will be the only equilibrium state. For a graph with several connected components, there will be an equilibrum state for each connected component, which equals some positive constant on that component and zero elsewhere.

For the Desargues graph, or indeed any connected graph, all the other eigenvectors will be functions that undergo exponential decay at various rates:

$$H\psi = \lambda\psi$$

with $\lambda < 0$, so that

$$\exp(tH)\psi = \exp(\lambda t)\psi$$

decays as t increases. But since probability is conserved, any vector that undergoes exponential decay must have terms that sum to zero. So a vector like this won't be a stochastic state, but rather a deviation from equilibrium. We can write any stochastic state as the equilibrium state plus a sum of terms that decay with different rate constants.

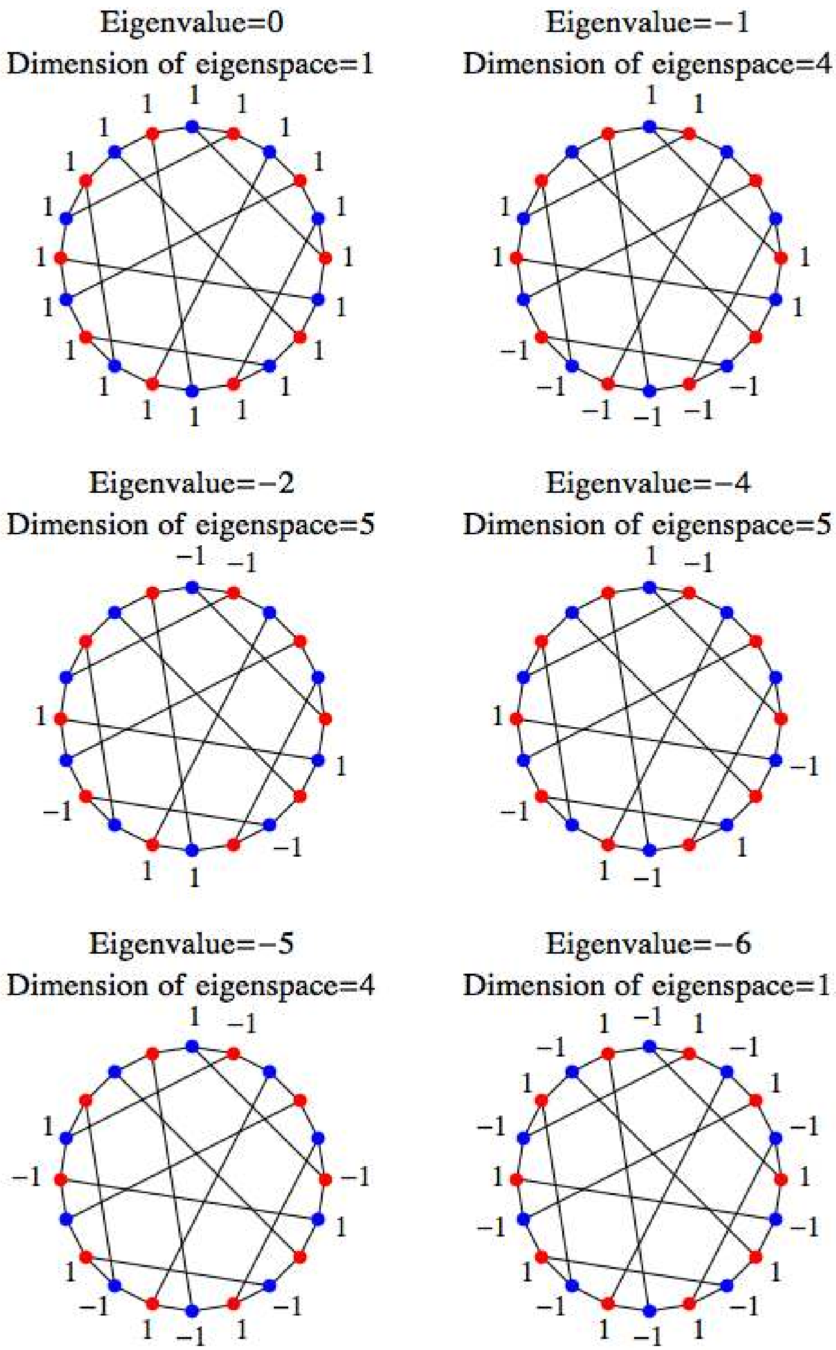
If you put a value of 1 on every red dot and -1 on every blue dot, you get an eigenvector of the Hamiltonian with eigenvalue -6: -3 for the diagonal entry for each vertex, and -3 for the sum over the neighbors.

By trial and error, it's not too hard to find examples of variations on this where some vertices have a value of zero. Every vertex with zero either has its neighbors all zero, or two neighbors of opposite signs:

- Eigenvalue -5: every non-zero vertex has two neighbors with the opposite value to it, and one neighbor of zero.
- Eigenvalue -4: every non-zero vertex has one neighbor with the opposite value to it, and two neighbors of zero.
- Eigenvalue -2: every non-zero vertex has one neighbor with the same value, and two neighbors of zero.
- Eigenvalue -1: every non-zero vertex has two neighbors with the same value, and one neighbor of zero.

So the general pattern is what you'd expect: the more neighbors of equal value each non-zero vertex has, the more slowly that term will decay.

In more detail, the eigenvectors are given by the the following functions, together with functions obtained by rotating these pictures. Unlabelled dots are labelled by zero:



In fact you can span the eigenspaces by rotating the particular eigenvectors I drew by successive multiples of 1/5 of a rotation, i.e. 4 dots.

If you do this for the examples with eigenvalues -2 and -4, it's not too hard to see that all five rotated diagrams are linearly independent. If you do it for the examples with eigenvalues -1 and -5, four rotated diagrams are linearly independent. So along with the two 1-dimensional spaces, that's enough to prove that you've exhausted the whole 20-dimensional space.

15 Dirichlet operators and electrical circuits

We've been comparing two theories: stochastic mechanics and quantum mechanics. Last time we saw that any graph gives us an example of *both* theories! It's a bit peculiar, but today we'll explore the intersection of these theories a little further, and see that it has another interpretation. It's also the theory of *electrical circuits made of resistors!*

That's nice, because I'm supposed to be talking about 'network theory', and electrical circuits are perhaps the most practical networks of all:

[PIC gif here]

I plan to talk a lot about electrical circuits. I'm not quite ready to dive in, but I can't resist dipping my toe in the water today. Why don't you join me? It's not too cold!

15.1 Dirichlet operators

Last time we saw that any graph gives us an operator called the 'graph Laplacian' that's both infinitesimal stochastic and self-adjoint. That means we get both:

- a Markov process describing the random walk of a classical particle on the graph. and
- a 1-parameter unitary group describing the motion of a quantum particle on the graph.

That's sort of neat, so it's natural to wonder what are *all* the operators that are both infinitesimal stochastic and self-adjoint. They're called 'Dirichlet operators', and at least in the finite-dimensional case we're considering, they're easy to completely understand. Even better, it turns out they describe electrical circuits made of resistors!

Today let's take a lowbrow attitude and think of a linear operator $H : \mathbb{C}^n \to \mathbb{C}^n$ as an $n \times n$ matrix with entries H_{ij}. Then:

¡li¿ H is **self-adjoint** if it equals the conjugate of its transpose:

$$H_{ij} = \overline{H}_{ji}$$

¡li¿H is **infinitesimal stochastic** if its columns sum to zero and its off-diagonal entries are nonnegative:

$$\sum_i H_{ij} = 0$$

$$i \neq j \Rightarrow H_{ij} \geq 0$$

¡li¿ H is a **Dirichlet operator** if it's both self-adjoint and infinitesimal stochastic.

What are Dirichlet operators like? Suppose H is a Dirichlet operator. Then its off-diagonal entries are ≥ 0, and since

$$\sum_i H_{ij} = 0$$

its diagonal entries obey

$$H_{ii} = -\sum_{i \neq j} H_{ij} \leq 0$$

So all the entries of the matrix H are real, which in turn implies it's symmetric:

$$H_{ij} = \overline{H}_{ji} = H_{ji}$$

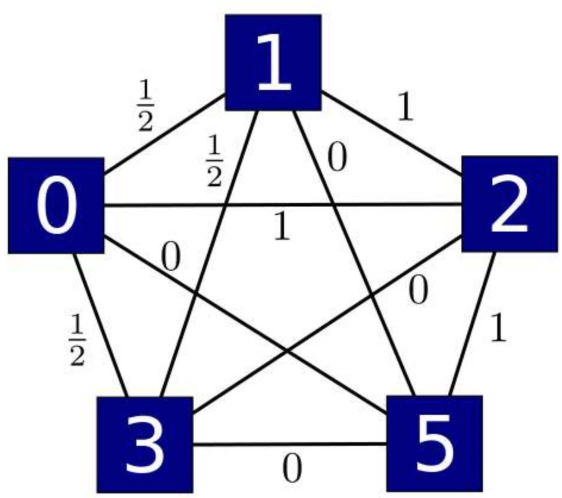
So, we can build any Dirichlet operator H as follows:

- Choose the entries above the diagonal, H_{ij} with $i < j$, to be arbitrary nonnegative real numbers.
- The entries below the diagonal, H_{ij} with $i > j$, are then forced on us by the requirement that H be symmetric: $H_{ij} = H_{ji}$.
- The diagonal entries are then forced on us by the requirement that the columns sum to zero: $H_{ii} = -\sum_{i \neq j} H_{ij}$.

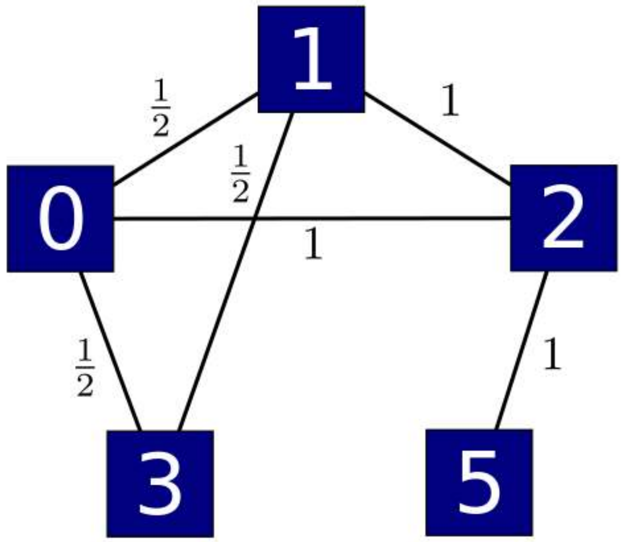
Note that because the entries are real, we can think of a Dirichlet operator as a linear operator $H : \mathbb{R}^n \to \mathbb{R}^n$. We'll do that for the rest of today.

15.2 Circuits made of resistors

Now for the fun part. We can easily *draw* any Dirichlet operator! To this we draw n dots, connect each pair of distinct dots with an edge, and label the edge connecting the ith dot to the jth with any number $H_{ij} \geq 0$.



This contains all the information we need to build our Dirichlet operator. To make the picture prettier, we can leave out the edges labelled by 0:



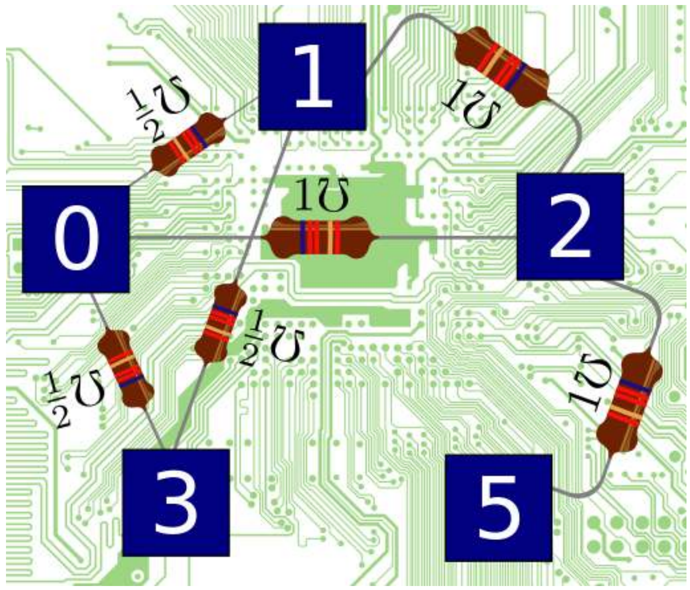
Like last time, the graphs I'm talking about are **simple**: undirected, with no edges from a vertex to itself, and at most one edge from one vertex to another. So:

Theorem 17. *Any finite simple graph with edges labelled by positive numbers gives a Dirichlet operator, and conversely.*

We already talked about a special case last time: if we label all the edges by the number 1, our operator H is called the **graph Laplacian**. So, now we're generalizing that idea by letting the edges have more interesting labels.

What's the meaning of this trick? Well, we can think of our graph as an *electrical circuit* where the edges are *wires*. What do the numbers labelling these wires mean? One obvious possibility is to put a resistor on each wire, and let that number be its resistance. But that doesn't make sense, since we're leaving out wires labelled by 0. If we leave out a wire, that's not like having a wire of zero resistance: it's like having a wire of *infinite* resistance! No current can go through when there's no wire. So the number labelling an edge should be the *conductance* of the resistor on that wire. Conductance is the reciprocal of resistance.

So, our Dirichlet operator above gives a circuit like this:



Here Omega is the symbol for an 'ohm', a unit of resistance... but the upside-down version, namely MHO, is the symbol for a 'mho', a unit of conductance that's the reciprocal of an ohm.

Let's see if this cute idea leads anywhere. Think of a Dirichlet operator $H : \mathbb{R}^n \to \mathbb{R}^n$ as a circuit made of resistors. What could a vector $\psi \in \mathbb{R}^n$ mean? It assigns a real number to each vertex of our graph. The only sensible option is for this number to be the electric potential at that point in our circuit. So let's try that.

Now, what's

$$\langle \psi, H\psi \rangle ?$$

In quantum mechanics this would be a very sensible thing to look at: it would be gives us the expected value of the Hamiltonian H in a state ψ. But what does it mean in the land of electrical circuits?

Up to a constant fudge factor, it turns out to be the *power* consumed by the electrical circuit!

Let's see why. First, remember that when a current flows along a wire, power gets consumed. In other words, electrostatic potential energy gets turned into heat. The power consumed is

$$P = VI$$

where V is the voltage across the wire and I is the current flowing along the wire. If we assume our wire has resistance R we also have Ohm's law:

$$I = V/R$$

so

$$P = \frac{V^2}{R}$$

If we write this using the conductance $C = 1/R$ instead of the resistance R, we get

$$P = CV^2$$

But our electrical circuit has *lots* of wires, so the power it consumes will be a sum of terms like this. We're assuming H_{ij} is the conductance of the wire from the ith vertex to the jth, or zero if there's no wire connecting them. And by definition, the voltage across this wire is the difference in electrostatic potentials at the two ends: $\psi_i - \psi_j$. So, the total power consumed is

$$P = \sum_{i \neq j} H_{ij}(\psi_i - \psi_j)^2$$

This is nice, but what does it have to do with $\langle \psi, H\psi \rangle$?

The answer is here:

Theorem. If $H : \mathbb{R}^n \to \mathbb{R}^n$ is any Dirichlet operator, and $\psi \in \mathbb{R}^n$ is any vector, then

$$\langle \psi, H\psi \rangle = -\frac{1}{2} \sum_{i \neq j} H_{ij}(\psi_i - \psi_j)^2$$

Proof. Let's start with the formula for power:

$$P = \sum_{i \neq j} H_{ij}(\psi_i - \psi_j)^2$$

Note that this sum includes the condition $i \neq j$, since we only have wires going between distinct vertices. But the summand is zero if $i = j$, so we also have

$$P = \sum_{i,j} H_{ij}(\psi_i - \psi_j)^2$$

Expanding the square, we get

$$P = \sum_{i,j} H_{ij}\psi_i^2 - 2H_{ij}\psi_i\psi_j + H_{ij}\psi_j^2$$

The middle term looks promisingly similar to $\langle \psi, H\psi \rangle$, but what about the other two terms? Because $H_{ij} = H_{ji}$, they're equal:

$$P = \sum_{i,j} -2H_{ij}\psi_i\psi_j + 2H_{ij}\psi_j^2$$

And in fact they're zero! Since H is infinitesimal stochastic, we have

$$\sum_i H_{ij} = 0$$

so

$$\sum_i H_{ij}\psi_j^2 = 0$$

and it's still zero when we sum over j. We thus have

$$P = -2\sum_{i,j} H_{ij}\psi_i\psi_j$$

But since ψ_i is real, this is -2 times

$$\langle\psi, H\psi\rangle = \sum_{i,j} H_{ij}\overline{\psi}_i\psi_j$$

So, we're done.

An instant consequence of this theorem is that a Dirichlet operator has

$$\langle\psi, H\psi\rangle \le 0$$

for all ψ. Actually most people use the opposite sign convention in defining infinitesimal stochastic operators. This makes $H_{ij} \le 0$, which is mildly annoying, but it gives

$$\langle\psi, H\psi\rangle \ge 0$$

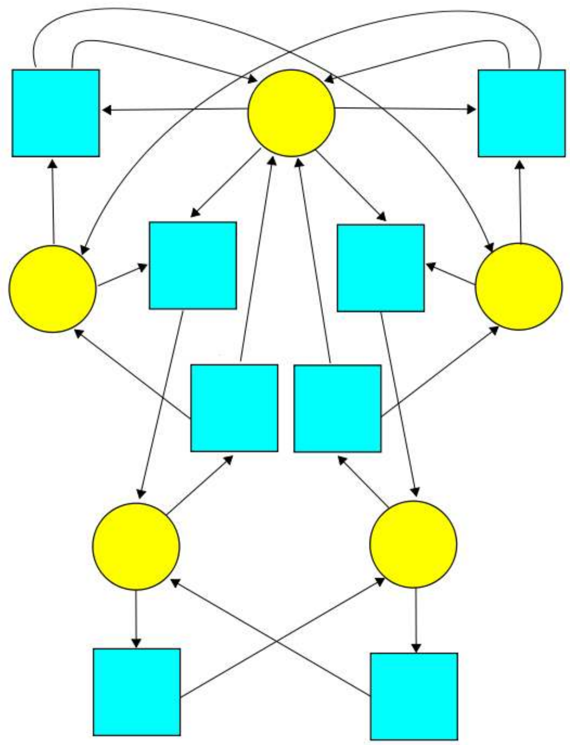
which is nice. When H is a Dirichlet operator, defined with this opposite sign convention, $\langle\psi, H\psi\rangle$ is called a **Dirichlet form**.

15.3 The big picture

Maybe it's a good time to step back and see where we are.

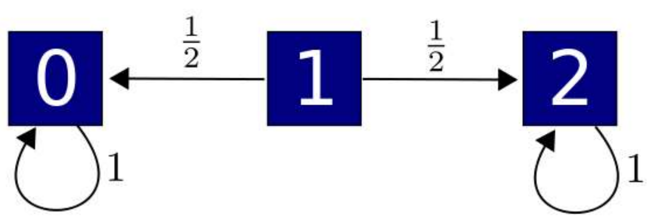
So far we've been exploring the analogy between stochastic mechanics and quantum mechanics. Where do networks come in? Well, they've actually come in twice so far:

1. First we saw that Petri nets can be used to describe stochastic or quantum processes where things of different kinds randomly react and turn into other things. A Petri net is a kind of network like this:



The different kinds of things are the yellow circles; we called them **states**, because sometimes we think of them as different states of a single kind of thing. The reactions where things turn into other things are the blue squares: we called them **transitions**. We label the transitions by numbers to say the rates at which they occur.

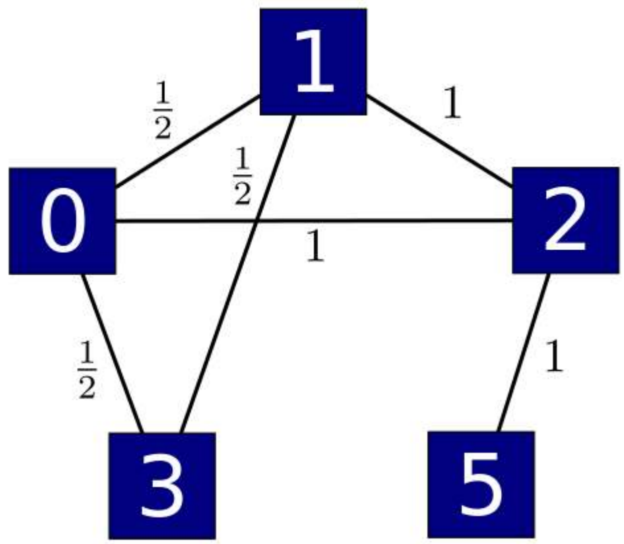
2. Then we looked at stochastic or quantum processes where in each transition a single thing turns into a single thing. We can draw these as Petri nets where each transition has just one state as input and one state as output. But we can also draw them as directed graphs with edges labelled by numbers:



Now the dark blue boxes are states and the *edges* are transitions!

Today we looked at a special case of the second kind of network: the Dirichlet operators. For these the 'forward' transition rate H_{ij} equals the 'reverse' rate

H_{ji}, so our graph can be undirected: no arrows on the edges. And for these the rates H_{ii} are determined by the rest, so we can omit the edges from vertices to themselves:



The result can be seen as an electrical circuit made of resistors! So we're building up a little dictionary:

- Stochastic mechanics: ψ_i is a probability and H_{ij} is a transition rate (probability per time).
- Quantum mechanics: ψ_i is an amplitude and H_{ij} is a transition rate (amplitude per time).
- Circuits made of resistors: ψ_i is a voltage and H_{ij} is a conductance.

This dictionary may seem rather odd—especially the third item, which looks completely different than the first two! But that's good: when things aren't odd, we don't get many new ideas. The whole point of this 'network theory' business is to think about networks from many different viewpoints and let the sparks fly!

Actually, this particular oddity is well-known in certain circles. We've been looking at the *discrete* version, where we have a finite set of states. But in the *continuum* , the classic example of a Dirichlet operator is the Laplacian $H = \nabla^2$. And then:

- The **heat equation**:

$$\frac{d}{dt}\psi = \nabla^2\psi$$

  is fundamental to stochastic mechanics.

- The **Schrödinger equation**:

$$\frac{d}{dt}\psi = -i\nabla^2\psi$$

is fundamental to quantum mechanics.

- The **Poisson equation**:

$$\nabla^2\psi = -\rho$$

is fundamental to electrostatics.

Briefly speaking, electrostatics is the study of how the electric potential ψ depends on the charge density ρ. The theory of electrical circuits made of resistors can be seen as a special case, at least when the current isn't changing with time.

I'll say a lot more about this... but not today! If you want to learn more, this is a great place to start:

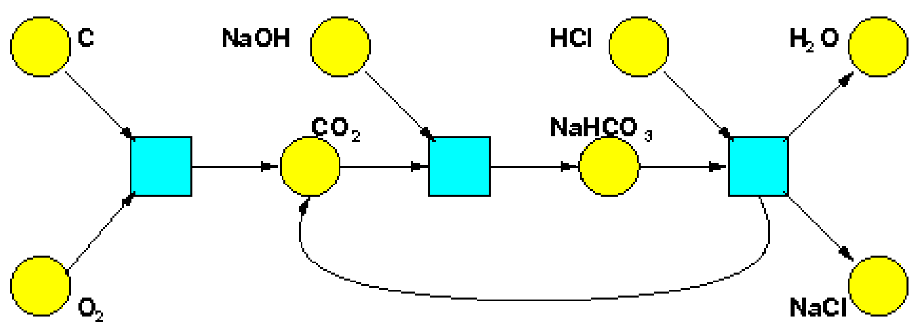
- P. G. Doyle and J. L. Snell, *Random Walks and Electrical Circuits* , Mathematical Association of America, Washington DC, 1984.

This free online book explains, in a really fun informal way, how random walks on graphs, are related to electrical circuits made of resistors. To dig deeper into the continuum case, try:

- M. Fukushima, *Dirichlet Forms and Markov Processes* , North-Holland, Amsterdam, 1980.

16 Feinberg's deficiency zero theorem

We've seen how Petri nets can be used to describe chemical reactions. Indeed our very first example came from chemistry:

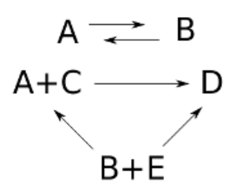


However, chemists rarely use the formalism of Petri nets. They use a different but entirely equivalent formalism, called 'reaction networks'. So now we'd like to tell you about those.

You may wonder: why bother with another formalism, if it's equivalent to the one we've already seen? Well, one goal of this network theory program is to get people from different subjects to talk to each other—or at least be *able* to. This requires setting up some dictionaries to translate between formalisms. Furthermore, lots of deep results on stochastic Petri nets are being proved by chemists—but phrased in terms of reaction networks. So you need to learn this other formalism to read their papers. Finally, this other formalism is actually better in some ways!

Reaction networks

Here's a reaction network:



This network involves 5 **species**: that is, different kinds of things. They could be atoms, molecules, ions or whatever: chemists call all of these species, and there's no need to limit the applications to chemistry: in population biology, they could even be biological species! We're calling them A, B, C, D, and E, but in applications, we'd call them by specific names like CO_2 and HCO_3, or 'rabbit' and 'wolf', or whatever.

This network also involves 5 **reactions**, which are shown as arrows. Each reaction turns one bunch of species into another. So, written out more long-windedly, we've got these reactions:

$$A \to B$$

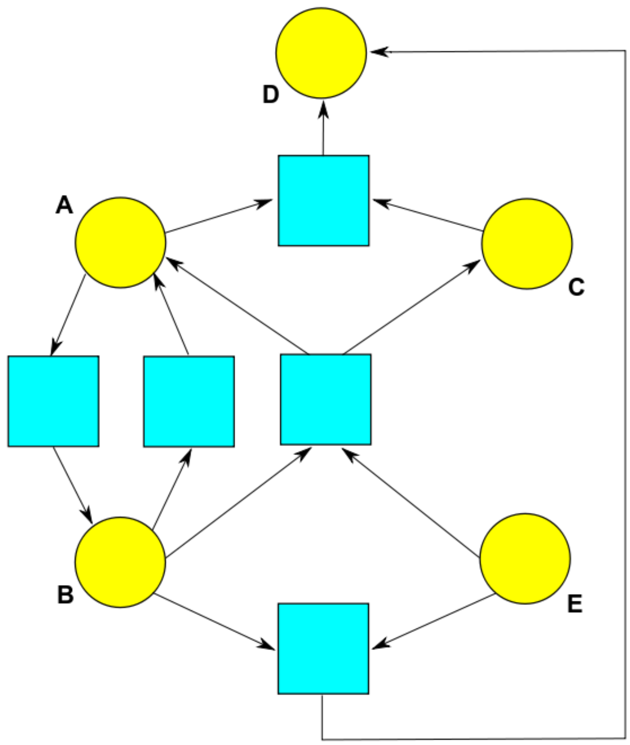
$$B \to A$$

$$A + C \to D$$

$$B + E \to A + C$$

$$B + E \to D$$

If you remember how Petri nets work, you can see how to translate any reaction network into a Petri net, or vice versa. For example, the reaction network we've just seen gives this Petri net:



Each species corresponds to a **state** of this Petri net, drawn as a yellow circle. And each reaction corresponds to **transition** of this Petri net, drawn as a blue square. The arrows say how many things of each species appear as input or output to each transition. There's less explicit emphasis on complexes in the Petri net notation, but you can read them off if you want them.

In chemistry, a bunch of species is called a 'complex'. But what do I mean by 'bunch', exactly? Well, I mean that in a given complex, each species can show up 0,1,2,3... or any natural number of times. So, we can formalize things like this:

Definition. Given a set S of **species**, a **complex** of those species is a function $C : S \to \mathbb{N}$.

Roughly speaking, a reaction network is a graph whose vertices are labelled by complexes. Unfortunately, the word 'graph' means different things in mathematics—appallingly many things! Everyone agrees that a graph has vertices and edges, but there are lots of choices about the details. Most notably:

- We can either put arrows on the edges, or not.
- We can either allow more than one edge between vertices, or not.
- We can either allow edges from a vertex to itself, or not.

If we say 'no' in every case we get the concept of 'simple graph', which we discussed last time. At the other extreme, if we say 'yes' in every case we get the concept of 'directed multigraph', which is what we want now. A bit more formally:

Definition 18. *A* **directed multigraph** *consists of a set V of* **vertices***, a set E of* **edges***, and functions $s, t : E \to V$ saying the* **source** *and* **target** *of each edge.*

Given this, we can say:

Definition 19. *A* **reaction network** *is a set of species together with a directed multigraph whose vertices are labelled by complexes of those species.*

You can now prove that reaction networks are equivalent to Petri nets:

Problem 29. Show that any reaction network gives a Petri net, and vice versa.

In a stochastic Petri net each transition is labelled by a **rate constant**: that is, a numbers in $[0, \infty)$. This lets us write down some differential equations saying how species turn into each other. So, let's make this definition (which is not standard, but will clarify things for us):

Definition 20. *A* **stochastic reaction network** *is a reaction network where each reaction is labelled by a rate constant.*

Now you can do this:

Problem 30. Show that any stochastic reaction network gives a stochastic Petri net, and vice versa.

For extra credit, show that in each of these problems we actually get an equivalence of categories! For this you need to define morphisms between Petri nets, morphisms between reaction networks, and similarly for stochastic Petric nets and stochastic reaction networks. If you get stuck, ask Eugene Lerman for advice. There are different ways to define morphisms, but he knows a cool one.

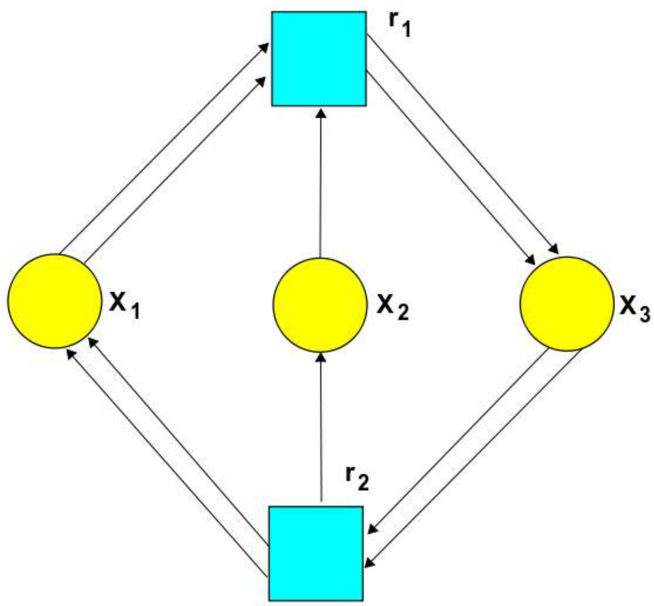
We've been downplaying category theory so far, but it's been lurking beneath everything we do, and someday it may rise to the surface.

16.1 The deficiency zero theorem

You may have already noticed one advantage of reaction networks over Petri nets: they're quicker to draw. This is true even for tiny examples. For example, this reaction network:

$$2X_1 + X_2 \leftrightarrow 2X_3$$

corresponds to this Petri net:



But there's also a deeper advantage. As we saw in Part 8, any stochastic Petri net gives two equations:

- The **master equation**, which says how the *probability that we have a given number of things of each species* changes with time.
- The **rate equation**, which says how the *expected number of things in each state* changes with time.

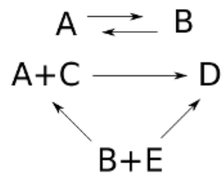
The simplest solutions of these equations are the equilibrium solutions, where nothing depends on time. Back in Part 9, we explained when an equilibrium solution of the rate equation gives an equilibrium solution of the master equation. But when is there an equilibrium solution of the rate equation in the first place?

Feinberg's 'deficiency zero theorem' gives a handy sufficient condition. And this condition is best stated using reaction networks! But to understand it, we need to understand the 'deficiency' of a reaction network. So let's define that, and they say what all the words in the definition mean:

Definition 21. *The* **deficiency** *of a reaction complex is:*

- *the number of vertices minus*
- *the number of connected components minus*
- *the dimension of the stoichiometric subspace.*

The first two concepts here are easy. A reaction network is a graph (okay, a directed multigraph). So, it has some number of vertices, and also some number of connected components. Two vertices lie in the same **connected component** iff you can get from one to the other by a path *where you don't care which way the arrows point* . For example, this reaction network:



has 5 vertices and 2 connected components.

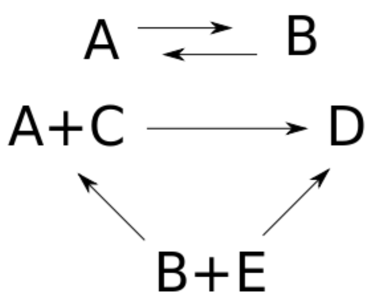
So, what's the 'stoichiometric subspace'? 'Stoichiometry' is a scary-sounding word. According to the Wikipedia article:

'Stoichiometry' is derived from the Greek words ??????? (*stoicheion* , meaning element) and ?????? (*metron* , meaning measure.) In patristic Greek, the word *Stoichiometria* was used by Nicephorus to refer to the number of line counts of the canonical New Testament and some of the Apocrypha.

But for us, stoichiometry is just the art of counting species. To do this, we can form a vector space $\mathbb{R}^S$ where S is the set of species. A vector in $\mathbb{R}^S$ is a function from species to real numbers, saying how much of each species is present. Any complex gives a vector in $\mathbb{R}^S$, because it's actually a function from species to *natural* numbers.

Definition 22. *The* **stoichiometric subspace** *of a reaction network is the subspace* Stoch $\subseteq \mathbb{R}^S$ *spanned by vectors of the form* $x - y$ *where* x *and* y *are complexes connected by a reaction.*

'Complexes connected by a reaction' makes sense because vertices in the reaction network are complexes, and edges are reactions. Let's see how it works in our example:



Each complex here can be seen as a vector in $\mathbb{R}^S$, which is a vector space whose basis we can call A, B, C, D, E. Each reaction gives a difference of two vectors coming from complexes:

- The reaction $A \to B$ gives the vector $B - A$.
- The reaction $B \to A$ gives the vector $A - B$.
- The reaction $A + C \to D$ gives the vector $D - A - C$.
- The reaction $B + E \to A + C$ gives the vector $A + C - B - E$.
- The reaction $B + E \to D$ gives the vector $D - B - E$.

The pattern is obvious, I hope.

These 5 vectors span the stoichiometric subspace. But this subspace isn't 5-dimensional, because these vectors are linearly dependent! The first vector is the negative of the second one. The last is the sum of the previous two. And those are all the linear dependencies, so the stoichiometric subspace is 3 dimensional. For example, it's spanned by these 3 linearly independent vectors: $A - B, D - A - C$, and $D - B - E$.

I hope you see the moral of this example: the stoichiometric subspace is the space of *ways to move in* $\mathbb{R}^S$ *that are allowed by the reactions in our reaction network!* And this is important because the rate equation describes how the amount of each species changes as time passes. So, it describes a point moving around in $\mathbb{R}^S$.

Thus, if Stoch $\subseteq \mathbb{R}^S$ is the stoichiometric subspace, and $x(t) \in \mathbb{R}^S$ is a solution of the rate equation, then $x(t)$ always stays within the set

$$x(0) + \mathrm{Stoch} = \{x(0) + y \colon y \in \mathrm{Stoch}\}$$

Mathematicians would call this set the coset of $x(0)$, but chemists call it the **stoichiometric compatibility class** of $x(0)$.

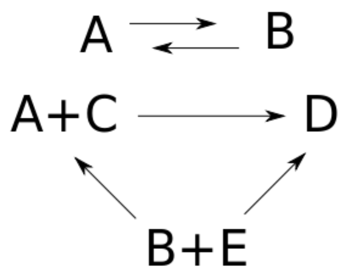
Anyway: what's the deficiency of the reaction complex in our example? It's

$$5 - 2 - 3 = 0$$

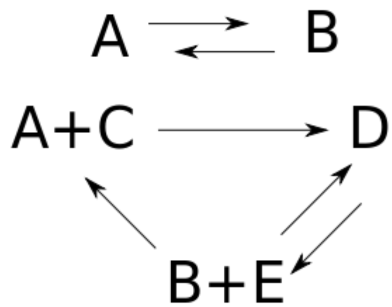
since there are 5 complexes, 2 connected components and the dimension of the stoichiometric subspace is 3.

But what's the deficiency zero theorem? You're almost ready for it. You just need to know *one* more piece of jargon! A reaction network is **weakly reversible** if whenever there's a reaction going from a complex x to a complex y, there's a path of reactions going back from y to x. Here the paths need to follow the arrows.

So, this reaction network is *not* weakly reversible:



since we can get from $A+C$ to D but not back from D to $A+C$, and from $B+E$ to D but not back, and so on. However, the network becomes weakly reversible if we add a reaction going back from D to $B+E$:



If a reaction network isn't weakly reversible, one complex can turn into another, but not vice versa. In this situation, what typically happens is that as time goes on we have less and less of one species. We could have an equilibrium where there's *none* of this species. But we have little right to expect an equilibrium solution of the rate equation that's **positive**, meaning that it sits at a point $x \in (0,\infty)^S$, where there's a nonzero amount of every species.

My argument here is not watertight: you'll note that I fudged the difference between species and complexes. But it can be made so when our reaction network has deficiency zero:

Theorem 23 (Deficiency Zero Theorem (Feinberg))**.** *Suppose we are given a reaction network with a finite set of species S, and suppose its deficiency is zero. Then:*

(i) If the network is not weakly reversible and the rate constants are positive, the rate equation does not have a positive equilibrium solution.

(ii) If the network is not weakly reversible and the rate constants are positive, the rate equation does not have a positive periodic solution, that is, a periodic solution lying in $(0,\infty)^S$.

(iii) If the network is weakly reversible and the rate constants are positive, the rate equation has exactly one equilibrium solution in each positive stoichiometric

compatibility class. Any sufficiently nearby solution that starts in the same stoichiometric compatibility class will approach this equilibrium as $t \to +\infty$. Furthermore, there are no other positive periodic solutions.

This is quite an impressive result. We'll look at an easy example next time.

16.2 References and remarks

The deficiency zero theorem was published here:

- Martin Feinberg, Chemical reaction network structure and the stability of complex isothermal reactors: I. The deficiency zero and deficiency one theorems, *Chemical Engineering Science* **42** (1987), 2229.2268.

These other explanations are also very helpful:

- Martin Feinberg, Lectures on reaction networks, 1979.
- Jeremy Gunawardena, Chemical reaction network theory for in-silico biologists, 2003.

At first glance the deficiency zero theorem might seem to settle all the basic questions about the dynamics of reaction networks, or stochastic Petri nets... but actually, it just means that deficiency zero reaction networks don't display very interesting dynamics in the limit as $t \to +\infty$. So, to get more interesting behavior, we need to look at reaction networks that don't have deficiency zero.

For example, in biology it's interesting to think about 'bistable' chemical reactions: reactions that have two stable equilibria. An electrical switch of the usual sort is a bistable system: it has stable 'on' and 'off' positions. A bistable chemical reaction can serve as a kind of biological switch:

- Gheorghe Craciun, Yangzhong Tang and Martin Feinberg, Understanding bistability in complex enzyme-driven reaction networks, *PNAS* **103** (2006), 8697-8702.

It's also interesting to think about chemical reactions with stable periodic solutions. Such a reaction can serve as a biological clock:

- Daniel B. Forger, Signal processing in cellular clocks, *PNAS* **108** (2011), 4281-4285.

17 Conclusion and Acknowledgement

The network theory series started off as a series of articles on the Azimuth blog. You can read discussion of these articles there, and also make your own comments or ask questions. We thank the readers of Azimuth, especially David Corfield, Greg Egan and Blake Stacey, for many helpful discussions.

References

[1] Martin Feinberg. Lectures on reaction networks. *Lecture notes*, 1979.

[2] Peter J. E. Goss and Jean Peccoud. Quantitative modeling of stochastic systems in molecular biology by using stochastic Petri nets. *Proc. Natl. Acad. Sci. USA*, 95:6750–6755, 1998.

[3] Jeremy Gunawardena. Chemical reaction network theory for in-silico biologists. *Lecture notes*, 2003.

[4] Peter J. Haas. Stochastic Petri Nets: Modelling, Stability, Simulatio. *Springer, Berlin*, 2002.

[5] F. Horn and R. Jackson. General mass action kinetics. *Archive for Rational Mechanics and Analysis, Springer*, 47:81–116.

[6] W. O. Kermack and A. G. McKendrick. A Contribution to the mathematical theory of epidemics. *Proc. Roy. Soc. Lond. A*, pages 700–721.

[7] Ina Koch. Petri nets - a mathematical formalism to analyze chemical reaction networks. *Molecular Informatics*, 29:838–843, 2010.

[8] S. R. Logan. Chemical Reaction Kinetics. *Longman, Essex*, 1996.

[9] Marc Mangel. The Theoretical Biologist's Toolbox: Quantitative Methods for Ecology and Evolutionary Biology. *Cambridge U. Press, Cambridge*, 2006.

[10] Vladimiro Sassone. On the category of Petri net computations. *6th International Conference on Theory and Practice of Software Development, Proceedings of TAPSOFT '95, Lecture Notes in Computer Science*, 915:334–348, 1995.

[11] Darren James Wilkinson. Stochastic Modelling for Systems Biology. *Taylor & Francis, New York*, 2006.