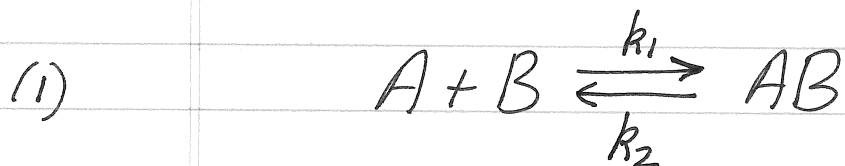


CS Peskin 9/28/2018

Law of mass action

Consider the reversible reaction



in a well-mixed solution. The number of occurrences of the forward reaction per unit time per unit volume is proportional to the concentration of A and also to the concentration of B, so it must be of the form

$$(2) \quad k_1 [A][B]$$

where $[]$ denotes concentration.

Similarly, the number of occurrences of the reverse reaction per unit time per unit volume is given by

$$(3) \quad k_2 [AB]$$

The parameters k_1 and k_2 are called rate constants. Note that they have different units.

$$k_1 \sim \frac{1}{\text{time} \cdot \text{concentration}} = \frac{\text{volume}}{\text{time}}$$

$$k_2 \sim \frac{1}{\text{time}}$$

Here concentration is defined as the number of molecules per unit volume, so it has units of $1/\text{volume}$, since the number of molecules is a pure number, which has no units. Chemists like to define concentration instead as the number of mols per unit volume and treat mole as a unit. For conversion, note that 1 mole = N_A molecules, where N_A is Avogadro's number, roughly 6×10^{23} .

From (1-3), we get the following system of differential equations for $[A]$, $[B]$, $[AB]$:

$$(4) \quad \frac{d}{dt} [A] = -k_1 [A][B] + k_2 [AB]$$

$$(5) \quad \frac{d}{dt} [B] = -k_1 [A][B] + k_2 [AB]$$

$$(6) \quad \frac{d}{dt} [AB] = k_1 [A][B] - k_2 [AB]$$

Note that there are two conserved quantities

$$(7) \quad \frac{d}{dt} ([A] + [AB]) = 0$$

$$(8) \quad \frac{d}{dt} ([B] + [AB]) = 0$$

We therefore introduce

$$(9) \quad [A^*] = [A] + [AB]$$

$$(10) \quad [B^*] = [B] + [AB]$$

So that $[A^*]$ is the total concentration of molecules of type A in the system, regardless of whether they are free or bound to B, and $[B^*]$ is the total concentration of molecules of type B, regardless of whether they are free or bound to A.

According to (7-8), $[A^*]$ and $[B^*]$ are constants. Since concentrations cannot be negative

$$(11) \quad [AB] \leq [A^*]$$

and

$$(12) \quad [AB] \leq [B^*]$$

and we can summarize these inequalities by

$$(13) \quad [AB] \leq \min([A^*], [B^*])$$

Because of (9-10), the system (4-6) can be reduced to one equation in one unknown:

$$(14) \quad \frac{d}{dt} [AB] = k_1 ([A^*] - [AB]) ([B^*] - [AB]) - k_2 [AB]$$

$$= k_1 \left(([A^*] - [AB]) ([B^*] - [AB]) - K [AB] \right)$$

where

$$(15) \quad K = \frac{k_2}{k_1}$$

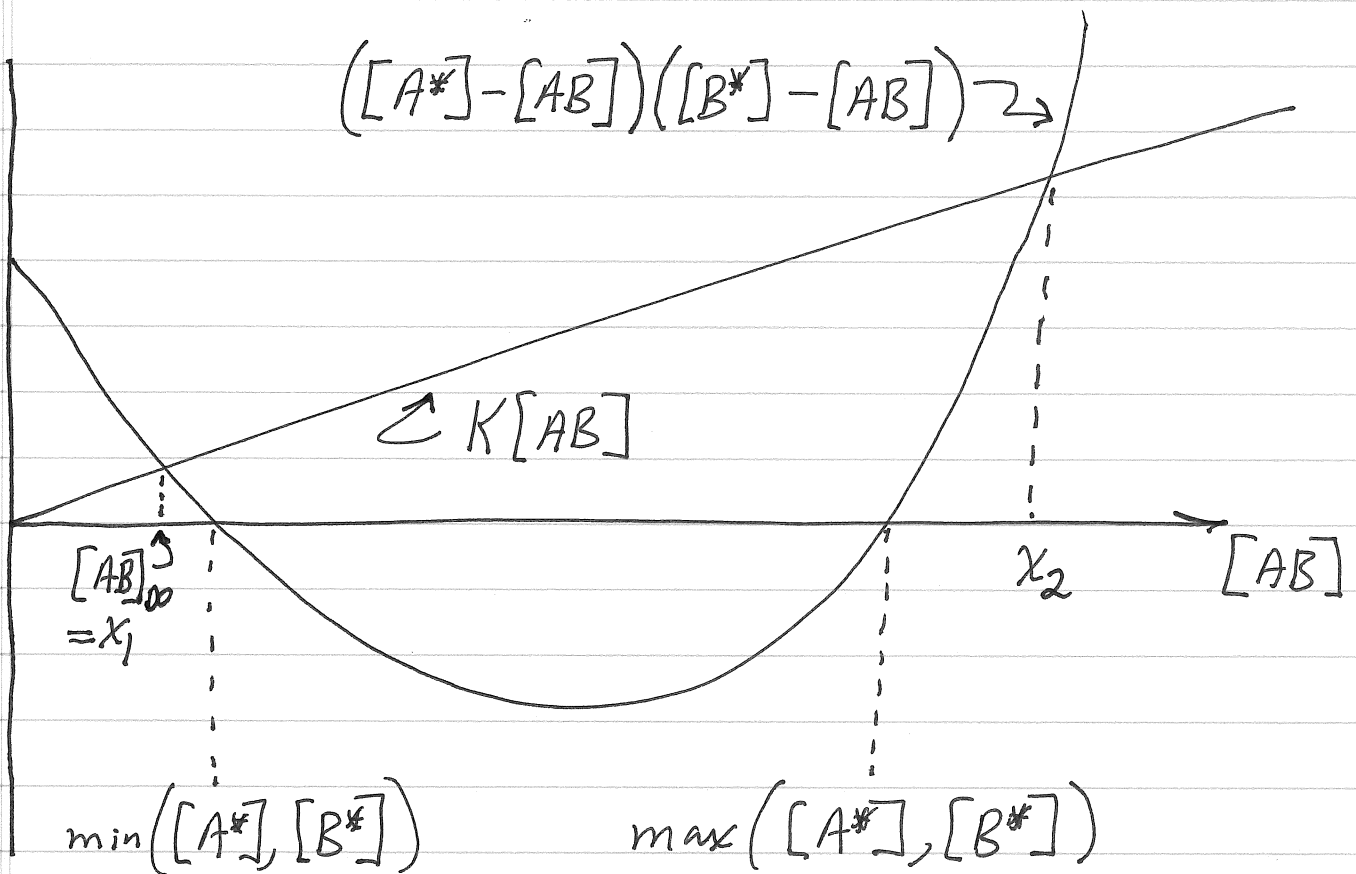
Note that K has units of concentration.

If we look for a steady solution of (14), i.e., a value of $[AB]$ such that $\frac{d}{dt} [AB] = 0$,

we get the equation

$$(16) \quad ([A^*] - [AB])([B^*] - [AB]) = K[AB]$$

The graphical solution of which looks like this:



Clearly, there are two positive solutions but only the smaller one satisfies (13) so there is only one steady state, which we call $[AB]_{\infty}$. It is the

unique solution of (16) that satisfies (13).

7

Also, it is obvious from the foregoing plot and equation (14) that

$$(17) \quad 0 \leq [AB] < [AB]_{\infty}$$

$$\Rightarrow \frac{d}{dt} [AB] > 0$$

and

$$(18) \quad [AB]_{\infty} < [AB] < \min([A^*], [B^*])$$

$$\Rightarrow \frac{d}{dt} [AB] < 0$$

This shows that $[AB]_{\infty}$ is a stable steady state.

We can solve the initial-value problem for $[AB](t)$ in the following way.

Let

$$(19) \quad x = [AB](t)$$

and let x_1, x_2 be the two solutions of (16) that have been found graphically above.*
Then

$$(20) \quad 0 < x_1 < \min([A^*], [B^*]) < \max([A^*], [B^*]) < x_2$$

Note that x_1 and x_2 are solutions of the quadratic equation

$$(21) \quad x^2 - ([A^*] + [B^*] + K)x + [A^*][B^*] = 0$$

and it is an interesting exercise to prove (20) algebraically by writing out the solutions to this equation. (Hint: In the case of x_1 , rationalize the numerators!)

* with $x_1 < x_2$, so that $x_1 = [AB]_{\infty}$

In terms of x , x_1 , x_2 , equation (14) takes the form

$$(22) \quad \frac{dx}{dt} = k_1(x_1 - x)(x_2 - x)$$

Note that

$$(23) \quad \frac{d}{dt} \left(\frac{x_1 - x}{x_2 - x} \right) = - \frac{(x_2 - x) \frac{dx}{dt} - (x_1 - x) \frac{dx}{dt}}{(x_2 - x)^2}$$

$$= - \frac{(x_2 - x_1) k_1 (x_1 - x)(x_2 - x)}{(x_2 - x)^2}$$

$$= - (x_2 - x_1) k_1 \left(\frac{x_1 - x}{x_2 - x} \right)$$

Thus $(x_1 - x)/(x_2 - x)$ satisfies a linear differential equation for which we know the solution:

$$(24) \quad \frac{x_1 - x(t)}{x_2 - x(t)} = \frac{x_1 - x(0)}{x_2 - x(0)} e^{-(x_2 - x_1)k_1 t}$$

Since $x_2 > x_1$, the exponential is decaying.

Since $x_2 - x(t)$ is bounded from below by the positive quantity $x_2 - \min([A^*], [B^*])$, we see immediately from (24) that

$$(25) \quad \lim_{t \rightarrow \infty} x(t) = x_1$$

i.e.,

$$(26) \quad \lim_{t \rightarrow \infty} [AB](t) = [AB]_{\infty}$$

hence the name, $[AB]_{\infty}$.

Equation (24) is easily solved for $x(t)$:

$$(27) \quad x(t) = \frac{x_1 - x_2 \frac{x_1 - x(0)}{x_2 - x(0)} e^{-(x_2 - x_1)k_1 t}}{1 - \frac{x_1 - x(0)}{x_2 - x(0)} e^{-(x_2 - x_1)k_1 t}}$$

and this completes the solution. (The values of x_1 and x_2 are found by solving (21).)

Microscopic mass action

Now we consider the reaction $A + B \rightleftharpoons AB$ from a microscopic point of view. This means that we keep track of the integer number of molecules of each chemical species in our system:

(28) $N_A(t) =$ number of molecules of A at time t

(29) $N_B(t) =$ number of molecules of B at time t

(30) $N_{AB}(t) =$ number of molecules of AB at time t

These are integer-valued random variables. They change in steps of ± 1 whenever an instance of the reaction $A + B \rightarrow AB$ or whenever an instance of the reaction $AB \rightarrow A + B$ occurs.

The specific changes that occur are shown in the following table

<u>reaction</u>	<u>ΔN_A</u>	<u>ΔN_B</u>	<u>ΔN_{AB}</u>
$A+B \rightarrow AB$	-1	-1	+1
$AB \rightarrow A+B$	+1	+1	-1

Of course, the numbers

$$(31) \quad n_A^* = N_A(t) + N_{AB}(t)$$

$$(32) \quad n_B^* = N_B(t) + N_{AB}(t)$$

are constant, and we use lowercase letters for them to emphasize that they are not random. Thus, we really only need to keep track of $N_{AB}(t)$. Its possible values are integers such that

$$(33) \quad 0 \leq N_{AB}(t) \leq \min(n_A^*, n_B^*)$$

Next, we need to discuss the microscopic rate constants r_1 and r_2 for the forward and reverse reactions, and their relationship to the macroscopic rate constants k_1 and k_2 .

For this it is necessary to consider the volume V of the system, since it will influence in particular the microscopic rate constant r_1 .

Let r_1 be the probability per unit time (see below) that a particular molecule of type A (not bound to B) and a particular molecule of type B (not bound to A) will combine to form AB. Since we are talking about a particular pair of molecules, this probability per unit time is not affected by the numbers of molecules of the different types. It is, however, affected by the volume V of the system, since it takes longer for a particular molecule of type A and a particular molecule of type B to find each other in a larger volume.

Similarly, let r_2 be the probability per unit time that a particular molecule of the type AB will undergo the reaction $AB \rightarrow A+B$. Again, since we are talking about a particular molecule, r_2 is unaffected by the numbers of molecules of the various species that may be present. Unlike r_1 , however, r_2 is also unaffected by the volume of the system, since there is no need for two molecules to find each other for the reaction to occur.

The probability per unit time of an event is defined as follows. Given an interval $(t, t+\Delta t)$ in which an event may or may not occur, let $P(t, t+\Delta t)$ be the probability that the event does occur in this interval. Then

$$(34) \quad r = \lim_{\Delta t \rightarrow 0} \frac{P(t, t+\Delta t)}{\Delta t}$$

if this limit exists, is called the probability per unit time of the event (at time t).

Note that r is not a probability. In particular, it has units of reciprocal time, so its value will change depending on the unit of time that is chosen. Also, although r cannot be negative, there is no upper limit to its set of allowed values. This point needs to be emphasized, since it is a very common mistake to confuse the concept of probability per unit time with that of probability and therefore to think that a probability per unit time cannot exceed 1.

Now we are ready to consider the relationship between the microscopic and macroscopic rate constants. First, consider the forward reaction $A+B \rightarrow AB$. The expected number of occurrences per unit time in our system will be

(35) $r_1 N_A N_B$

since $N_A N_B$ is the number of pairs of molecules, one of type A and the other of type B, and each such pair has the probability per unit time

r_1 of undergo the reaction $A+B \rightarrow AB$.
 For comparison with the macroscopic system, we divide by the volume V to get the number of occurrences per unit volume per unit time:

$$(36) \quad \frac{r_1 N_A N_B}{V} = \frac{r_1 V N_A N_B}{V^2}$$

$$= (r_1 V) [A][B]$$

In the last step, we have introduced the concentrations by writing $[A] = N_A/V$ and $[B] = N_B/V$. Comparison with equation (2) now shows that $r_1 V = k_1$,
 or

$$(37) \quad r_1 = \frac{k_1}{V}$$

Since the macroscopic rate constant k_1 is independent of volume, this simple relationship quantifies the influence of volume

on the probability per unit time for a given pair of molecules to find each other and react. It is amazing that the result is so simple, and also that we can derive it without worrying about the details of the process of diffusion by which the molecules actually move.

Similarly, for the reverse reaction $AB \rightarrow A+B$, the expected number of occurrences per unit time per unit volume is

$$(38) \quad \frac{r_2 N_{AB}}{V} = r_2 [AB]$$

and comparison with (3) gives the result

$$(39) \quad r_2 = k_2$$

There is no volume dependence in this case because there is no need for molecules to find each other to undergo the reaction.

Since the microscopic rate constants r_1 and r_2 are probabilities per unit time, they both have the same units, $1/\text{time}$. Equations (37) & (39) therefore remind us again that the units of the two macroscopic rate constants k_1 and k_2 are different.

In particular, the rate constant of the forward reaction, which is bimolecular, has units of volume/time. This reflects the physical influence of volume in a bimolecular reaction.

Now we are ready to undertake the stochastic simulation of the reversible reaction $A+B \rightleftharpoons AB$. We will use event-driven simulation, also called the Gillespie method, especially in the context of stochastic chemical kinetics. This means that the simulation jumps from event to event, where each event is the occurrence of one of the two reactions $A+B \rightarrow AB$ or $AB \rightarrow A+B$.

When an event can occur at any time $t > 0$ with a probability per unit time r that is constant, then the time T at which it first occurs has an exponential distribution with mean $1/r$. This can be simulated in Matlab by the line

$$(40) \quad T = (-\log(\text{rand})) / r$$

Here "rand" in Matlab generates a random number with uniform distribution on the interval $(0, 1)$, and "log" is the natural logarithm.

Then $-\log(\text{rand})$ is a positive number that turns out to have an exponential distribution on $(0, \infty)$ with mean 1. After multiplication by $1/r$, we then have an exponentially distributed random number with mean $1/r$, as desired.

In our case, however, we typically have two competing processes that might occur, and whichever occurs first is the next event. Suppose the system happens to be in a state with N_A molecules of type A, N_B molecules of type B, and N_{AB} molecules of type AB immediately after a reaction has occurred. Until the next reaction actually happens (thereby changing the state) the probability per unit time of the forward reaction is

(41) $r_1 N_A N_B$

and the probability per unit time of the reverse reaction is

(42) $r_2 N_{AB}$

Thus we choose two random times

(43) $T_1 = (-\log(\text{rand})) / (r_1 * N_A * N_B)$

(44) $T_2 = (-\log(\text{rand})) / (r_2 * N_{AB})$

and whichever is smaller determines both the time of the next reaction and also which reaction it is.

If either (but not both) of the denominators in (43-44) is equal to zero, this does not cause any difficulty. For example, if $N_{AB} = 0$, then the reverse reaction cannot happen, and Matlab will evaluate T_2 as "Inf". This will not stop the program. When T_1 and T_2 are compared, T_1 will be smaller, and everything will proceed as it should.

If both denominators are zero, we are in a degenerate case in which nothing can happen because one species or the other is completely absent, so there is no point in doing the simulation.

Here is Matlab code for the main loop:

```

% initialize r1, r2, NA, NB, NAB, tmax
t = 0
while t < tmax
    T1 = (-log(rand)) / (r1 * NA * NB)
    T2 = (-log(rand)) / (r2 * NAB)
    if (T1 < T2)
        NA = NA - 1; NB = NB - 1; NAB = NAB + 1
        t = t + T1
    else
        NA = NA + 1; NB = NB + 1; NAB = NAB - 1
        t = t + T2
    end
end
end

```

When plotting or processing output from such a simulation, keep in mind that the event times are not equally spaced and that the state variables are constant between events. When computing time averages of any quantity, keep in mind that each value has to be weighted by the amount of time that it has that value.

It is easy to generalize the foregoing to a completely arbitrary network of reactions. We just need some notation.

Let j be an index designating a species, with $j = 1 \dots n_s$, and let k be an index generating a reaction, with $k = 1 \dots n_r$. Forward and reverse reactions are considered separately. Let

$$(45) \quad N(j) = \text{number of molecules of species } j.$$

These are the state variables of our problem. Let

$$(46) \quad dN(j, k) = \text{the change in } N(j) \text{ that happens when reaction } k \text{ occurs}$$

These are small integers that can be positive, negative, or zero. They do not change during the simulation. They encode the effect of reaction k .

Finally, we need a function

(47) $\text{rates}(N, r)$

where N is the current state of the system, see (45), and r is a vector of rate constants.

The output of rates should be a row vector giving the probability per unit time that each reaction will occur.

The details of the function rates will be application-specific. A general-purpose version could be written if we wanted to restrict consideration to mass-action kinetics, but in practice more complicated rate functions are often used, so we will not do that here.

A general-purpose Matlab code for stochastic chemical kinetics is as follows:

The code is so short, if you blink, you'll miss it.

```
% initialize ns, nr, r, dN, N, tmax
t = 0
while t < tmax
    T = (-log(rand(1, nr))) ./ rates(N, r)
    [Tmin, kmin] = min(T)
    N = N + dN(:, kmin)
    t = t + Tmin
end
```

In the foregoing, the line

(48) $[Tmin, kmin] = \min(T)$

takes advantage of a feature of the min function in Matlab. If asked, it produces two outputs, the minimum value (here denoted Tmin) and the index in the input array where that value occurred (here denoted kmin). For us, kmin is the index of the reaction that actually happened. Thus, in the next two ~~and~~ lines, N is updated by $dN(:, kmin)$, and t is updated by Tmin.

Besides doing stochastic simulations of microscopic chemical systems, we can also write differential equations for the probabilities of finding the system in its different possible states as functions of time.

For the reaction that we have been considering,
 $A + B \rightleftharpoons AB$, let

$$(49) \quad P_n(t) = \text{probability that } N_{AB}(t) = n$$

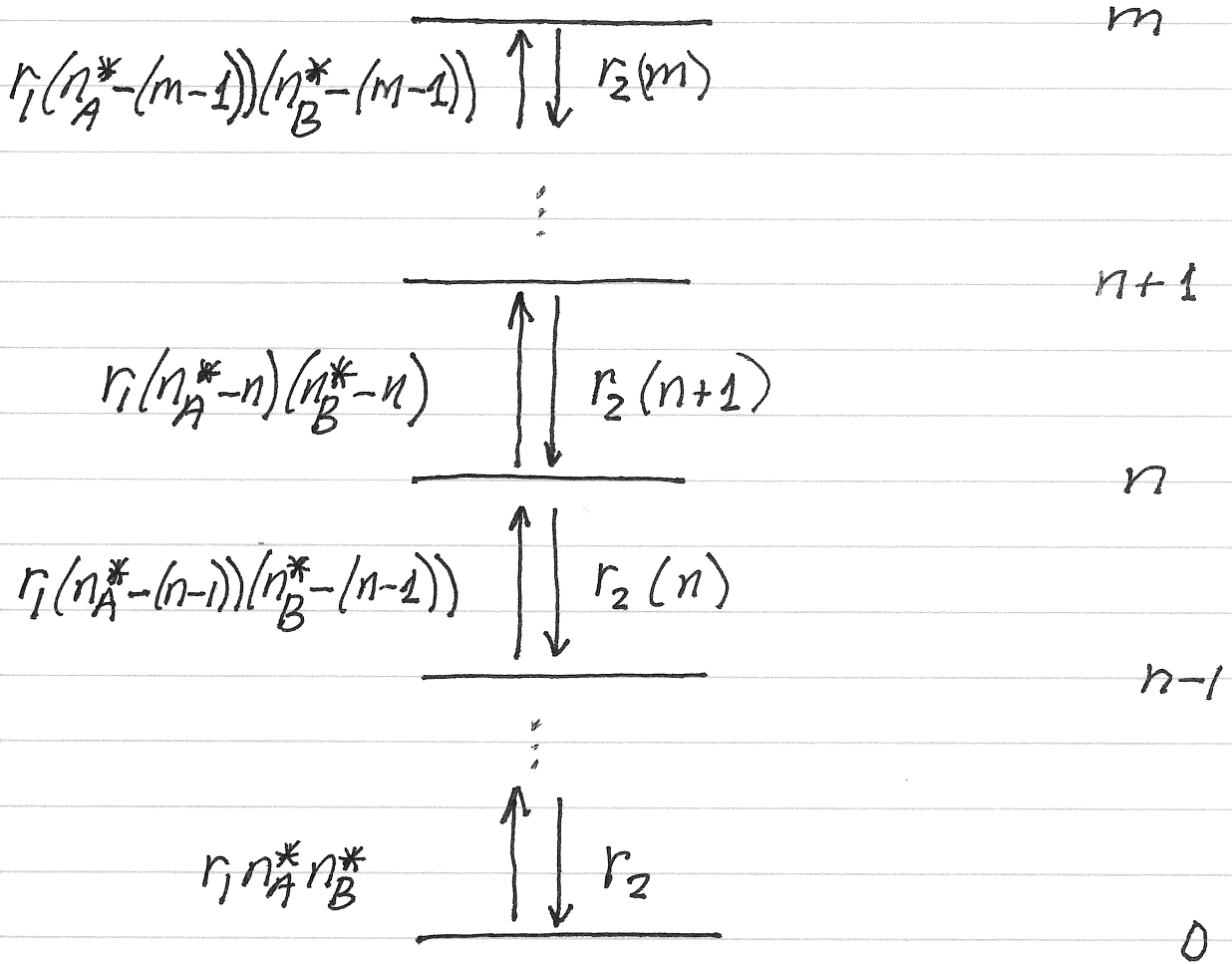
The possible values of n are $0 \dots m$, where

$$(50) \quad m = \min(n_A^*, n_B^*)$$

Of course $P_n(t) \geq 0$ and

$$(51) \quad \sum_{n=0}^m P_n(t) = 1$$

In order to derive the equations for $P_n(t)$, we make use of the following diagram



The label on each arrow is the probability per unit time that the system will make the corresponding transition, given that it is in the state at the tail of the arrow.

Then for $n = 1 \dots (m-1)$ we have

$$(52) \quad \frac{dP_n}{dt} = r_1 \left((n_A^* - (n-1))(n_B^* - (n-1)) P_{n-1} \right. \\ \left. - (n_A^* - n)(n_B^* - n) P_n \right) \\ + r_2 \left((n+1) P_{n+1} - n P_n \right)$$

and the special cases $n=0$ and $n=m$ give

$$(53) \quad \frac{dP_0}{dt} = -r_1 n_A^* n_B^* P_0 + r_2 P_1$$

$$(54) \quad \frac{dP_m}{dt} = r_1 \left(n_A^* - (m-1) \right) \left(n_B^* - (m-1) \right) P_{m-1} \\ - r_2 m P_m$$

It is a good exercise to check that equations (52-54) imply

$$(58) \quad \sum_{n=0}^m \frac{dP_n}{dt} = 0$$

Hint: shift indices in some of the sums.

The steady-state probabilities can be found by setting $dP_n/dt = 0$, but it is easier (and equivalent) to set each of the net fluxes of probability between adjacent levels equal to zero. This gives

$$(56) \quad r_1 (n_A^* - (n-1)) (n_B^* - (n-1)) P_{n-1} \\ = r_2 n P_n$$

or

$$(57) \quad \frac{P_n}{P_{n-1}} = \frac{r_1}{r_2} \frac{(n_A^* - (n-1))(n_B^* - (n-1))}{n}$$

It follows that

$$(58) \quad \frac{P_n}{P_0} = \left(\frac{r_1}{r_2}\right)^n \frac{n_A^*! n_B^*!}{(n_A^*-n)! (n_B^*-n)! n!}$$

The factor that multiplies $\left(\frac{r_1}{r_2}\right)^n$ in the foregoing has an interesting combinatorial interpretation. To see this, we multiply and divide by $n!$ to obtain

$$(59) \quad \frac{P_n}{P_0} = \left(\frac{r_1}{r_2}\right)^n \binom{n_A^*}{n} \binom{n_B^*}{n} n!$$

where

$$(60) \quad \binom{n^*}{n} = \text{" } n^* \text{ choose } n \text{"} = \text{the number of distinct ways of choosing } n \text{ items when there are } n^* \text{ items to choose from}$$

$$= \frac{n^*!}{(n^*-n)! n!}$$

Therefore, if we are given n_A^* items of type A and n_B^* items of type B, and we want to form n pairs of type AB, with no single item participating in more than one pair, then we can accomplish this task as follows:

- i) Choose n items of type A from the n_A^* given items of that type.
- ii) Choose n items of type B from the n_B^* given items of that type.
- iii) Assign each of the chosen items of type A to exactly one of the chosen items of type B.

The number of ways of making each of the above choices is

$$(61) \quad \binom{n_A^*}{n}, \quad \binom{n_B^*}{n}, \quad n!$$

respectively, so the overall number of ways in which the n pairs can be chosen is the product of these factors.

Equation (59) gives each of the P_n in terms of P_0 , and P_0 can then be found from the condition that the sum of the probabilities is equal to 1

$$(62) \quad 1 = \sum_{n=0}^m P_n = P_0 \sum_{n=0}^m \left(\frac{r_1}{r_2}\right)^n \binom{n_A^*}{n} \binom{n_B^*}{n} n!$$

This determines P_0 and then

$$(63) \quad P_n = \frac{\left(\frac{r_1}{r_2}\right)^n \binom{n_A^*}{n} \binom{n_B^*}{n} n!}{\sum_{k=0}^m \left(\frac{r_1}{r_2}\right)^k \binom{n_A^*}{k} \binom{n_B^*}{k} k!}$$

for $n=0 \dots m = \min(n_A^*, n_B^*)$

To evaluate all of the P_n , it is much more efficient to use the recursion relation (57) than the explicit formula (63).

% P_n will be stored in $P(1+(n))$ to avoid 0 index.

$m = \min([nA_{star}, nB_{star}])$

$P = \text{zeros}(m+1, 1)$

$P(1+(0)) = 1$ % temporary value

for $n = 1:m$

$P(1+(n)) = P(1+(n-1)) * (r1/r2) \dots$

$* (nA_{star} - (n-1)) * (nB_{star} - (n-1)) / n$

end

$P = P / \text{sum}(P)$ % normalization

% now evaluate the expected value of NAB

% and the expected value of NAB²

$NAB_{bar} = (0:m) * P$

$NAB2_{bar} = ((0:m).^2) * P$

% and from these the variance and

% standard deviation of NAB

$NAB_{var} = NAB2_{bar} - NAB_{bar}^2$

$NAB_{sig} = \text{sqrt}(NAB_{var})$

Construction of an example in which the numbers of molecules in the system can be varied while keeping concentrations constant for comparison with macroscopic theory

- Choose k_1 and k_2
- Let $V_0 = k_1/k_2$. In all simulations make V a multiple of V_0 , e.g.

$$V = V_0, 4V_0, 16V_0, 64V_0 \dots$$

- Choose two small integers i_A, i_B
In a simulation with volume V ,
let

$$n_A^* = i_A \frac{V}{V_0}, \quad n_B^* = i_B \frac{V}{V_0}$$

Thus n_A^* and n_B^* are integers, and in all cases

$$[A^*] = \frac{n_A^*}{V} = \frac{i_A}{V_0}, \quad [B^*] = \frac{n_B^*}{V} = \frac{i_B}{V_0}$$

so $[A^*], [B^*]$ are independent of V

Compare event-driven simulation results, steady-state probabilities, and macroscopic theory.

Remarks on the above example

The choices of k_1 and k_2 do not affect anything of importance. The constant k_2 , with units of $1/\text{time}$, sets the time scale. Therefore, in event-driven simulation, t_{max} should be set equal to some large multiple of $1/k_2$.

The ratio k_1/k_2 sets the volume scale, and we have made use of this by setting $V_0 = k_1/k_2$. Since $r_1 = k_1/V$ and $r_2 = k_2$, V_0 is the volume at which $r_1 = r_2$. In the special case that $i_A = 1$ and $i_B = 1$ with $V = V_0$, then the one pair of molecules in the system will spend (on average) half its time bound and half its time unbound.

As V increases, the number of molecules of each type increases proportionally, so the relative amount of noise should decrease. The typical scaling of relative noise with n in this kind of situation is $1/\sqrt{n}$, so every multiple of V by 4 should cut the relative noise in half.