

CONTENTS

1	Continuous-Time Monte-Carlo of Reaction Systems	1
1.1	The Exponential Probability Density	1
1.1.1	Fishing	1
1.1.2	Time to the first reaction event	3
1.2	The Doob-Gillespie Method	4
1.2.1	Basic procedure	4
1.3	Stochastic Rate Constants	7
1.3.1	Useful numbers	8
1.3.2	Rescaling a Stochastic System	9
1.3.3	Example: Do-Undo loop and saturation	10

*Lecture Notes [v0.1]. Walter Fontana, Systems Biology, Harvard Medical School.
Please send bug reports to walter@hms.harvard.edu*

CHAPTER 1

CONTINUOUS-TIME MONTE-CARLO OF REACTION SYSTEMS

1.1 THE EXPONENTIAL PROBABILITY DENSITY

1.1.1 Fishing

A process in which the past does not influence the present is called memoryless. Fishing might be an example. Having failed to catch a fish in the past 3 hours does not increase the chances of catching one in the next hour. More formally, if X is the time of an event (catching a fish), then a memoryless process is defined as one for which

$$P(X > t + s | X > t) = P(X > s). \quad (1.1)$$

The point being that $P(X > s)$ is independent of t . This amounts to saying that the probability of an event occurring during the time interval between t and $t + dt$ is $a dt$, where a is a positive constant independent of t . From (1.1) and the relation between conditional and joint probabilities, $p(X | Y)p(Y) = p(X; Y)$, we obtain

$$P(X > t + s | X > t) = \frac{P(X > t + s; X > t)}{P(X > t)} = \frac{P(X > t + s)}{P(X > t)} = P(X > s), \quad (1.2)$$

where the last equation provides us with an equivalent definition of a memoryless process:

$$P(X > t + s) = P(X > t) P(X > s). \quad (1.3)$$

Imagine that at time t_0 we arrived at a lake containing, sadly, a single fish and would like to tell our friends what the chances are of having dinner, that is, catching our only fish,

at time $t_0 + t$. For this, we need to calculate the probability $p(t)dt$ – or the probability density $p(t)$ – of catching the fish between t and $t + dt$, given that we started fishing at time t_0 . For the sake of less clutter, assume that $t_0 = 0$.

The probability of catching *no* fish in the time period up to $t + \Delta t$, which we denote as $P_0(t + \Delta t)$, is given by the probability of having caught no dinner up to t , $P_0(t)$, times the probability of catching no dinner in the subsequent interval from t to $t + \Delta t$, given by $1 - a\Delta t$ for small Δt :

$$P_0(t + \Delta t) = P_0(t)(1 - a\Delta t). \quad (1.4)$$

Notice that this is an instance of the memoryless property (1.3), where X denotes the time of catching a fish: $P(X > t + s)$ corresponds to $P_0(t + \Delta t)$, $P(X > t)$ to $P_0(t)$ and $P(X > s)$ to $(1 - a\Delta t)$. We rearrange (1.4) into

$$\frac{P_0(t + \Delta t) - P_0(t)}{\Delta t} = -P_0(t)a, \quad (1.5)$$

and let $\Delta t \rightarrow 0$, yielding the differential equation

$$\frac{dP_0(t)}{dt} = -P_0(t)a \quad (1.6)$$

with solution

$$P_0(t) = e^{-at}. \quad (1.7)$$

$P_0(0) = 1$, since we started out with no fish.

If $P_0(t)$ is the probability of catching nothing up to time t , then $P(t) = 1 - P_0(t)$ is the probability of catching the fish within time t . Thus,

$$P(t) = 1 - e^{-at}. \quad (1.8)$$

We're not done yet. $P(t)$ is the cumulative distribution that dinner happens at some point before time t . We want to know the chances $p(t)$ for dinner at time t . For this, we must take the derivative of $P(t)$:

$$p(t) = \frac{dP(t)}{dt} = ae^{-at}. \quad (1.9)$$

This is the exponential probability density of a memoryless process with parameter a . If this strikes you like radioactive decay, that's because radioactive decay is a memoryless process; and so is the failure of a light bulb.

The expected time to dinner is $\langle t \rangle = \int_0^\infty tp(t)dt$, which we integrate by parts:

$$\langle t \rangle = a \int_0^\infty te^{-at}dt = a \left(-\frac{1}{a} [te^{-at}]_0^\infty - \frac{1}{a^2} [e^{-at}]_0^\infty \right) = \frac{1}{a}. \quad (1.10)$$

The parameter a has the dimension of a rate: s^{-1} . The second moment is obtained similarly (integrate by parts twice):

$$\langle t^2 \rangle = a \int_0^\infty t^2 \exp(-at)dt = 2/a^2, \quad (1.11)$$

from which we obtain the variance σ^2 as

$$\sigma^2 = \langle t^2 \rangle - \langle t \rangle^2 = \frac{2}{a^2} - \frac{1}{a^2} = \frac{1}{a^2}. \quad (1.12)$$

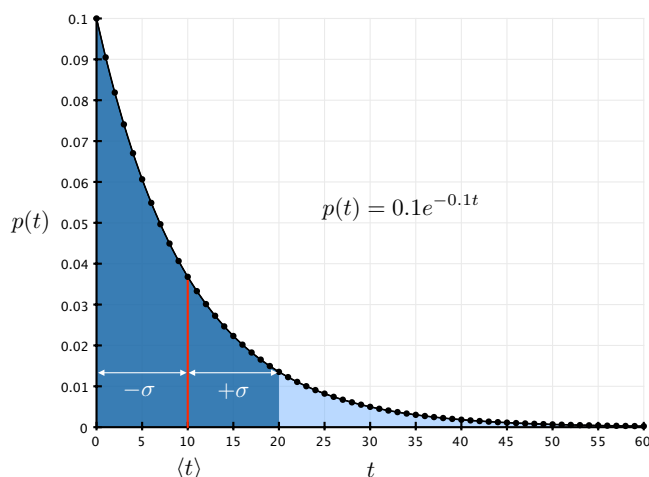


Figure 1.1 An exponential probability density.

1.1.2 Time to the first reaction event

Consider the reaction scheme



with rate constant γ . Imagine further a "system" consisting of only a single molecule of A . We now ask for the probability (density) that the only possible reaction event occurs at time t .

The basic assumption is that the process leading to a reaction event is, for any particular combination of reactants (there is only one in our case), a memoryless process. According to section 1.1.1, this means that our reaction event occurs with probability γdt in the time interval $[t, t + dt]$, independent of t . By the same reasoning that led to equation (1.9) this event will happen at time t with probability density

$$p(t) = \gamma e^{-\gamma t}. \quad (1.14)$$

The expected time $\langle t \rangle$ to a reaction is $1/\gamma$. This makes sense: the reciprocal of a unimolecular rate constant is the average time to a reaction event. There is nothing specifically "chemical" in this; from a purely kinetic point of view, A might as well be a radioactively decaying plutonium atom with half-life $\ln 2/\gamma$.

Now imagine n_A molecules of A in the system. What is the probability distribution $p_{(1)}(t)$ for the time to the next reaction event? A system consisting of n_A molecules of A , each of which reacts according to scheme (1.1.2) with rate parameter γ , is like a lake with n_A fish, all identical in their odds of being caught at any instant dt , and asking for the probability $p(t)$ that our *first* catch occurs at time t . We proceed exactly with the same setup as in (1.4), but now the probability of having no reaction occurring (catching no fish) in the interval $[t, t + \Delta t]$ is $1 - n_A \gamma \Delta t$,

$$P_0(t + \Delta t) = P_0(t)(1 - n_A \gamma \Delta t). \quad (1.15)$$

Following the reasoning in section 1.1.1, we end up with the analog of (1.9):

$$p(t) = n_A \gamma e^{-n_A \gamma t}. \quad (1.16)$$

Equation (1.16) is about the time to the *first* (or "next") reaction¹. After the reaction has occurred, the system has changed (we are left with $n_A - 1$ molecules of A). In fact, we should write $n_A(t)$ rather than just n_A , to make it clear that the parameter $n_A(t)\gamma$ in (1.16) is *time-dependent*. This is why we can use it to only compute the distribution of the first event, after which the parameter needs to be updated.

The number $\alpha(t) = \gamma n_A(t)$ is the *activity* of the unimolecular reaction (1.13) when there are n_A molecules of A ; it corresponds to the instantaneous reaction rate (flux) of the deterministic kinetics. Since our system consists of just this reaction, it also is the *reactivity* of the system. According to equation (1.10), the average time to the first reaction event is $1/(\gamma n_A)$. This is fairly obvious: the more A molecules in the system, the more "active" the reaction, the more "reactive" the system and the shorter the time interval to the next event.

The case of a bimolecular reaction scheme



is analogous to the development for the unimolecular case. If a system contains n_A and n_B molecules of type A and B , respectively, the number of distinct reaction possibilities (fish) is $n_A n_B$. Each of these possibilities realizes a memoryless process with parameter γ , and the distribution of times to the first reaction becomes

$$p(t) = \alpha e^{-\alpha t} \text{ with } \alpha = \gamma n_A(t) n_B(t). \quad (1.19)$$

The activity of a reaction is always the number of distinct reactant combinations computed from the current number of reactant molecules in the system. If the bimolecular reaction scheme were of the kind $A + A \rightarrow \text{products}$ with rate constant γ (as in a dimerization), the activity would be $\gamma 1/2 n_A(n_A - 1)$.

1.2 THE DOOB-GILLESPIE METHOD

1.2.1 Basic procedure

We now consider a system of n_r reaction schemes labelled r , $r = 1, \dots, n_r$ with activities α_r computed from their rate constants γ_r and the number of distinct reactant combinations.

¹**IYI** — Here's a slightly different derivation. Imagine we repeatedly record samples, each sample consisting of n_A time stamps τ that are generated with the same cumulative distribution $P(t) = p(\tau < t)$, equation (1.8), and probability density $p(t)$, equation (1.9). We then ask for the distribution of the smallest number in each of these samples. To be the smallest time stamp means that $n_A - 1$ time stamps are larger than t , where t has probability $p(t)$ (being *some* time stamp). There are n_A ways of choosing these $n_A - 1$ time stamps whose value is larger than t , and each of these choices has probability $[1 - P(t)]^{n_A - 1}$. Thus, we obtain as the distribution of the value of the smallest time stamp, $p_{(1)}(t)$:

$$p_{(1)}(t) \equiv p(t) = n_A p(t) [1 - P(t)]^{n_A - 1} = n_A \gamma e^{-\gamma t} e^{-(n_A - 1)\gamma t} = n_A \gamma e^{-n_A \gamma t}, \quad (1.17)$$

which is equation (1.16). This is also known as the first *order statistic* of a sample, and the calculation can be generalized to obtain any higher order statistic $p_{(k)}(t)$ (that is, the distribution of the k th smallest number in a sample).

Recall from section 1.1.2 that

$$\alpha_r = \gamma_r n_A \quad \text{for a unimolecular reaction} \quad (1.20)$$

$$\alpha_r = \gamma_r n_A n_B \quad \text{for a bimolecular reaction} \quad (1.21)$$

$$\alpha_r = \gamma_r \frac{1}{2} n_A (n_A - 1) \quad \text{for a reaction between molecules of the same type} \quad (1.22)$$

These are the main reaction schemes that we will encounter in the situations of interest to us.

To simulate the stochastic dynamics of a system consisting of specified molecule numbers, we need to know (1) when the next reaction event happens and (2) which reaction it is. The reasoning is pretty much analogous to section 1.1.2, but we will go a slightly different route to obtain a useful intermediate result.

(1) When does the next reaction happen? Each reaction scheme r is a memoryless process with parameter α_r ending in *its* first reaction. The distribution of times to the first reaction event, $p(t_r)$, is analogous to equation (1.16)

$$p(t_r) = \alpha_r e^{-\alpha_r t_r}, r = 1, \dots, n_r \quad (1.23)$$

Assume that reaction r occurs at time $t_r = t$. What is the probability that this reaction occurred before all others, given that it occurred at t ? We refer to this probability as $\text{prob}(r \text{ is first} \mid t_r = t)$. It is the probability that t is earlier than t_2 and earlier than t_3 and earlier than t_4 , and so on:

$$\begin{aligned} \text{prob}(r \text{ is first} \mid t_r = t) &= p(t < t_1)p(t < t_2) \cdots p(t < t_{r-1})p(t < t_{r+1}) \cdots p(t < t_{n_r}) \\ &= \prod_{j \neq r} p(t < t_j) = \prod_{j \neq r} p(t_j > t). \end{aligned}$$

Using equation (1.23):

$$\begin{aligned} \text{prob}(r \text{ is first} \mid t_r = t) &= \prod_{j \neq r} p(t_j > t) = \prod_{j \neq r} \int_t^\infty \alpha_j e^{-\alpha_j t_j} dt_j \\ &= \prod_{j \neq r} e^{-\alpha_j t} = e^{\alpha_r t} \prod_j e^{-\alpha_j t} = e^{\alpha_r t} e^{-\sum_j \alpha_j t} = e^{\alpha_r t} e^{-\lambda t}, \end{aligned} \quad (1.24)$$

where we defined the *reactivity of the system* as the sum of all reaction activities:

$$\lambda = \sum_{j=1}^{n_r} \alpha_j. \quad (1.25)$$

From this we compute the joint probability density that reaction r is the first reaction and that it occurs at time $t_r = t$:

$$\begin{aligned} \text{prob}(r \text{ is first}; t_r = t) &= \text{prob}(r \text{ is first} \mid t_r = t) \text{prob}(t_r = t) \\ &= e^{\alpha_r t} e^{-\lambda t} \alpha_r e^{-\alpha_r t} = \alpha_r e^{-\lambda t}. \end{aligned} \quad (1.26)$$

However, there are n_r possible reactions r , each of which could be the first one to occur, and we must add up their probabilities. The probability distribution $p(t)$ that the first reaction

event occurs at time t , given n_r possible events r , then becomes

$$\begin{aligned} p(t) &= \text{prob}(\text{first event occurs at } t) \\ &= \sum_r^{n_r} \text{prob}(r \text{ is first}; t_r = t) = \sum_r^{n_r} \alpha_r e^{-\lambda t} = \lambda e^{-\lambda t}. \end{aligned} \quad (1.27)$$

Compare result (1.27) to equation (1.23). Again, we see that the outcome (1.27) is completely analogous to (1.23). Instead of having one (equation 1.9) or n_A (equation 1.16) parallel memoryless processes, we now have as many as there are reactive opportunities in the system of n_r reactions. Many of these potential reaction combinations have the same constant; gathering them together and following the reasoning of section 1.1.2 leads to an exponential probability density with parameter $\lambda = \sum_j \alpha_j$.

(2) Which reaction happens next? Now that we know the distribution of times to the next reaction in the system, we need to know the probability p_r of it being reaction r . We get the answer immediately from the joint probability (1.26). $p(t)$, equation (1.27), was obtained by summing (1.26) over all reactions r . Now we integrate (1.26) over all times:

$$\begin{aligned} p_r &= \text{prob}(\text{the first event is } r) \\ &= \int_0^\infty \text{prob}(r \text{ is first}; t_r = t) dt = \int_0^\infty \alpha_r e^{-\lambda t} dt = \frac{\alpha_r}{\lambda}. \end{aligned} \quad (1.28)$$

Together, results (1.27) and (1.28) provide us with an easy simulation procedure, given a system with $r = 1, \dots, n_r$ reactions:

```

Continuous Time Monte-Carlo Procedure {
  for each reaction  $r = 1, \dots, n_r$ 
  {
    compute the appropriate activity  $\alpha_r$  using (1.20-1.22)
  }
  compute the system reactivity  $\lambda$  using (1.25)
  reset the simulated wall-clock time  $T$ 

  forever do
  {
    draw a random time  $t$  distributed according to (1.27)
    advance the simulated time,  $T \leftarrow T + t$ 
    draw a random reaction  $r$  according to (1.28)
    execute reaction scheme  $r$ 
    update the molecule numbers in the system
    update the affected activities  $\alpha_j$ 
    update the system reactivity  $\lambda$ 
  }
}

```

Many versions and implementations of this algorithm exist. This is not the place to describe them in detail, although implementation is important because it determines the limits of stochastic network simulations. We shall be a little more concerned with aspects of high-level implementation when we get to rule-based Monte-Carlo.

We have referred to rate constants in the context of stochastic reaction kinetics as parameters of an exponential probability density, equation (1.9). The next section examines the relationship between stochastic and deterministic rate constants.

1.3 STOCHASTIC RATE CONSTANTS

In a chemical reaction scheme r of the kind



the rate constant k fixes the proportionality between the velocity of the reaction v_r (or reaction rate) and the product of the reactant concentrations:

$$v_r = k[A][B] \quad (1.30)$$

The form of (1.30) factors the rate of r into (i) the concentration dependence of encounters between A and B (kinetic law of mass action) and (ii) the "rest", labelled k . This separation is based on the idea that the "rest" is constant over time, unlike the concentrations of reactants. Underlying the constancy of k is the idea of a *single reaction mechanism* – a particular causal path from reactants to products on the free energy surface of the system. Indeed, a time-dependent k would suggest that the mechanism of the reaction is concentration dependent, a telltale sign pointing at multiple mechanisms whose relative significance shifts with the concentration of reactants.

The rate constant k is an average property characterizing a reactive encounter between one molecule of A and one molecule of B . Reaction rate theory from first principles derives a bimolecular rate constant in terms of a collision cross-section between molecules, their relative velocity and internal quantum states. Such theory is a domain of microscopic reaction dynamics, which rests on quantum mechanics, statistical physics, and kinetics.

For our purposes it is not necessary to understand the fundamental theory of reaction rates. However, it is pedagogically useful to have a look at the crude cartoon of a collision between hard spheres. Hard spheres with no activation barrier represent a fictitious situation in which every collision is reactive. In that case, k is straightforward to compute.

Let spheres A and B have radius r_A and r_B , respectively. If A is headed towards B with relative velocity v , there will be an impact if the distance between their centers is less or equal than $d = r_A + r_B$. This condition corresponds to the area of a circle with diameter d , πd^2 , the so-called collision cross-section. The collision cross-section is a general concept referring to an effective area that reactants present to each other in a scattering process. Within time Δt , sphere A will sweep out a cylinder of volume $\pi d^2 v \Delta t$. If the number of moles of B in the system is $[B]$, there will be $\pi d^2 v \Delta t [B]$ opportunities for collision during time Δt for the particular A we are considering. Multiplying by the number of moles of A -spheres present in the volume V in which collisions occur, $[A]V$, and averaging over their velocities yields

$$\pi d^2 \langle v \rangle \Delta t V [A][B]. \quad (1.31)$$

This is the number of collision possibilities, expressed in moles, in the system volume V within Δt time units. Dividing by $\Delta t V$ yields the potential collisions per time and volume unit as $\pi d^2 \langle v \rangle [A][B]$, which equals the instantaneous rate of product formation. The deterministic rate constant in the case of hard spheres is therefore

$$k = \pi d^2 \langle v \rangle. \quad (1.32)$$

In thermal equilibrium, the sphere velocities follow a Maxwell-Boltzmann distribution with $\langle v \rangle = \sqrt{8k_B T / \pi \mu}$, where k_B is the Boltzmann constant, T the absolute temperature and

μ the reduced mass of the two spheres. This at least conveys an intuition of why the rate constant k is temperature and mass dependent. Not surprisingly, the dependence in this cartoon is far from the observed Arrhenius relation.

Let us now connect the deterministic rate constant with its stochastic counterpart. Recall that the stochastic treatment in section 1.1.2 shows the stochastic rate constant γ in conjunction with reactant combinations expressed as products of particle numbers in the system (as opposed to densities or concentrations), e.g. $\gamma n_A n_B$ as in (1.19). In addition, γ (or the parameter a in section 1.1.1) is a probability density, which becomes a probability in the form $\gamma \Delta t$. We therefore need to express the number of possible collisions between hard spheres, e.g. (1.31), in terms of n_A and n_B (expressed here in units of moles, to stay in sync with the derivation of (1.31)). Clearly, $[A] = n_A/V$ and $[B] = n_B/V$, yielding:

$$\pi d^2 \langle v \rangle \Delta t V \frac{n_A}{V} \frac{n_B}{V} = \frac{\pi d^2 \langle v \rangle}{V} \Delta t n_A n_B. \quad (1.33)$$

Thus, for a bimolecular reaction scheme,

$$\gamma = \frac{\pi d^2 \langle v \rangle}{V} = \frac{k}{V} \quad [mol^{-1} s^{-1}]. \quad (1.34)$$

The volume dependence of a bimolecular stochastic rate constant is crucial when converting numerical values of deterministic rate constants for use in stochastic simulations. It requires knowing the reaction volume – whether it is a cell or a subcellular locale – which is not always available. The volume dependence exists for any reaction scheme that involves encounters, hence for all reactions of molecularity $n > 1$: $\gamma = k/V^{(n-1)}$. A unimolecular reaction has no volume dependence (or collision-cross section), since it does not require a collision. The stochastic rate constant of a unimolecular scheme is the same as its deterministic rate constant.

We often want to express γ in "per molecule" rather than "per mol". This only changes the numerical value, not the dimension (both are numbers of molecules):

$$\gamma = \frac{k}{\mathcal{A} V} \quad [molecule^{-1} s^{-1}], \quad (1.35)$$

where $\mathcal{A} = 6.022 \cdot 10^{23}$ is Avogadro's number and k in molar units $M^{-1} s^{-1}$

The Michaelis constant of an enzymatic reaction is $K_m = (k_{-1} + k_2)/k_1$, with k_{-1} and k_1 denoting the off and on constants, respectively, of the enzyme-substrate interaction, and k_2 denoting the catalytic rate constant. The deterministic K_m has unimolecular rate constants in the numerator and a bimolecular rate constant in the denominator from which the stochastic Michaelis constant \mathcal{K}_m inherits the dependence on volume:

$$\mathcal{K}_m = K_m \mathcal{A} V \quad [molecules]. \quad (1.36)$$

1.3.1 Useful numbers

It is useful to have a few approximate cell sizes at hand for stochastic simulations.

- Mammalian cell: $V = 2.25 \cdot 10^{-12} l$ ($1 l = 10^{-3} m^3$), and $\mathcal{A} V = 1.35 \cdot 10^{12}$.

A concentration of $1M$ in a mammalian cell volume corresponds to $1.35 \cdot 10^{12}$ molecules; $1nM \approx 1350$ molecules per cell.

- Yeast cell (haploid): $V = 4 \cdot 10^{-14}l$, and $\mathcal{AV} = 2.4 \cdot 10^{10}$.
A concentration of $1M$ in a yeast cell volume corresponds to $2.4 \cdot 10^{10}$ molecules; $1nM \approx 24$ molecules per cell. The volume is doubled in a diploid cell.
- E.coli cell: $V = 10^{-15}l$, and $\mathcal{AV} = 10^8$.
A concentration of $1M$ in a yeast cell volume corresponds to 10^8 molecules; $10nM \approx 1$ molecule per cell.

We occasionally employ for illustrative purposes bimolecular rate constants of 1. A bimolecular $\gamma = 1$ in a mammalian cell volume corresponds to a deterministic $k \approx 10^{12}$. This is two orders of magnitudes faster than a diffusion-controlled k and hence nonsense, if the time scale is interpreted physically. Yet, we often don't care about a factor in the overall time scale.

The following table lists a few ballpark figures for deterministic rate constants and their stochastic counterparts in a mammalian cell volume

process	k	γ	stoch. dimension
general binding	$10^7 - 10^9$	$10^{-5} - 10^{-3}$	$molecule^{-1}s^{-1}$
general unbinding	$10^{-3} - 10^{-1}$	$10^{-3} - 10^{-1}$	s^{-1}
dephosphorylation	1	1	s^{-1}
phosphorylation	0.1	0.1	s^{-1}
receptor dimerization	$2 \cdot 10^6$	$1.6 \cdot 10^{-6}$	$molecule^{-1}s^{-1}$
receptor dissociation	$1.6 \cdot 10^{-1}$	$1.6 \cdot 10^{-1}$	s^{-1}

1.3.2 Rescaling a Stochastic System

Consider reaction (1.29) and imagine two systems with volumes $V_2 = \sigma V_1$, but with the same bimolecular k . While both systems have the same deterministic kinetics, they differ in their stochastic rate constants:

$$\gamma_1 = \frac{k}{A V_1} \quad \gamma_2 = \frac{k}{A \sigma V_1}, \quad (1.37)$$

or

$$\frac{\gamma_1}{\gamma_2} = \frac{V_2}{V_1} = \frac{\lambda V_1}{V_1} = \sigma. \quad (1.38)$$

Let a model M consist of n_s molecular species $i = 1, \dots, n_s$ and n_r reactions $r = 1, \dots, n_r$ occurring in a volume V . Rescaling M with a factor σ means that:

1. All molecule numbers are multiplied by σ (and presumably rounded to some integer):

$$n'_i = \sigma n_i \quad \forall i \quad (1.39)$$

2. All bimolecular stochastic rate constants are divided by σ :

$$\gamma'_r = \frac{1}{\sigma} \gamma_r \quad \forall r \text{ that are bimolecular} \quad (1.40)$$

3. The volume is multiplied by σ

$$V' = \sigma V. \quad (1.41)$$

This ensures that the rescaled system has its stochastic rate constants pegged to the same deterministic rate constants as the original. If $\sigma < 1$ (> 1), the rescaled system is smaller (larger) than the original.

Caution: The average behavior of the system is not changed by rescaling, but the fluctuations are. A DU loop (say) that is not saturated at V will not become saturated at σV , as the particle numbers are scaled accordingly. To change the operating regime of a reaction from unsaturated to saturated, requires either increasing the particle number while keeping the volume constant or decreasing the volume while keeping the particle number constant. This is *not* a rescaling.

1.3.3 Example: Do-Undo loop and saturation

Consider an enzymatic reaction with $k_1 = 10^{-5}$ (binding), $k_{-1} = 0.1$ (unbinding) and $k_2 = 0.1$ (catalytic rate constant), its $\mathcal{K}_m = 2 \cdot 10^4$ molecules in a mammalian cell volume, which corresponds to a $K_m = 0.02 \mu M$. A more reasonable $K_m = 1 \mu M$ comes in at a $\mathcal{K}_m = 10^6$. An enzymatic reaction with such a \mathcal{K}_m requires 10^7 or so molecules to start saturating.

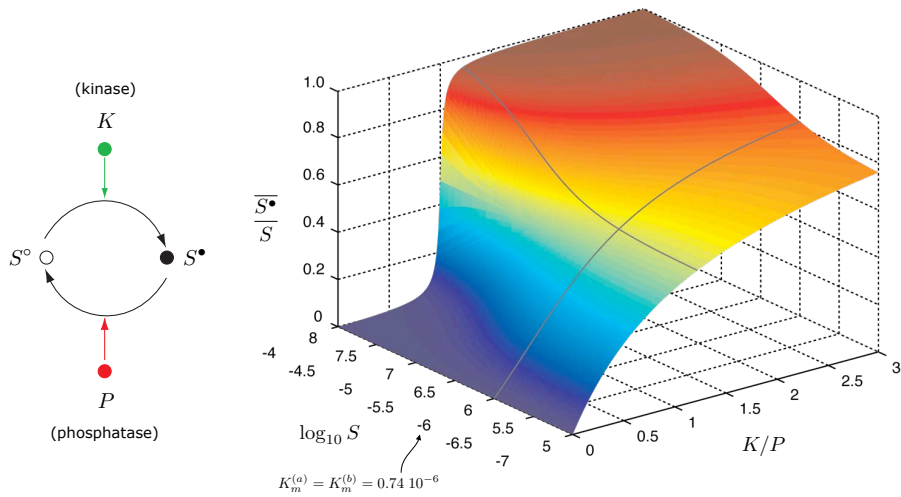


Figure 1.2 Do-Undo loop control surface.

We call a loop of two enzymatic reactions in which one reaction leg undoes what the other achieves a "Do-Undo loop" (DU loop). A frequent example is the phosphorylation of a substrate protein by a kinase and its dephosphorylation by a phosphatase. The (deterministic) control surface of a DU loop with a $\mathcal{K}_m = 10^6$ ($K_m = 1 \mu M$) is shown in Figure 1.2 (note the logarithmic substrate axis). At saturation – above, say, 10^7 molecules – the steady-state of the loop shows a markedly ultrasensitive (threshold-like) response to the kinase-to-phosphatase ratio, in contrast to the hyperbolic response in the unsaturated regime. In order to saturate a DU loop with a $\mathcal{K}_m = 10^6$, a cell would have to contain at least 10 million molecules of substrate protein – too many for a typical substrate protein. Does this mean that such loops can never become saturated? No. The loop could become saturated with far fewer substrate molecules, if we were to confine it to a smaller volume.

What is the volume in which 1000 substrate molecules suffice to saturate a loop with $K_m = 1\mu M$? The answer lies in equation (1.36). We need to determine the volume V that yields a \mathcal{K}_m of, say, 10^2 – well below the 10^3 substrate molecules available – while keeping the deterministic K_m the same:

$$V = \frac{10^2}{6 \cdot 10^{23} 10^{-6}} \sim 10^{-16} l, \quad (1.42)$$

which is roughly 10^{-4} of a mammalian cell volume or about 1/10 of an E.coli volume. The DU-loop with $K_m = 1\mu M$ can operate at saturation with 1000 substrate molecules, provided it occurs in a subvolume much smaller than the mammalian cell.

We can adopt two stances in this case. From a biological stance we might consider the loop as being, for example, confined to a subcellular compartment or a patch of surface. Whether this is realistic depends on what is known about the specific loop. If we know (or want to assume that) the loop is not confined to a subvolume, we would need to simulate with 10^7 substrate molecules to be in the saturated regime. The problem then is that we may not have the computational resources. From a pragmatic stance, however, we can use the same trick as above: *scale* down our reaction system by changing the volume *and* particle numbers simultaneously, allowing us to "play" the loop dynamics in an affordable time (but at the expense of larger fluctuations), while keeping its deterministic characteristics the same. In this case, we would deliberately scale the system by a factor of 10^{-4} to achieve a smaller volume and operate with 10^3 substrate molecules.