Stochastic process

Qualitative paradigm ($\mathcal{QPN}$ **):**
* representation of a bio-molecular process qualitative and minimally described by its topology
* behavior forms a discrete state space
* time not associated to transitions therefore descriptions are time-free
* qualitative analysis considers however all possible behavior of the system under any timing.

Timed information can be added to the qualitative description in two ways - stochastic and continuous.

➢ The deterministic approach: it regards the time evolution as a continuous, wholly predictable process which is governed by a set of coupled, ordinary differential equations (the "reaction-rate equations")

➢ the stochastic approach: it considers the time evolution as a kind of random-walk process which is governed by a single differential-difference equation (the "master equation"), but unfortunately the stochastic master equation is often mathematically intractable.

Stochastic process

Let's take a simple example: we want to follow the evolution of concentration of the protein B according to time.

The number of proteins B in a bacterium at time $t + dt$ is equal to
- the number of proteins B at time t
- plus the number of proteins B synthesized during the time dt
- minus the number of degraded proteins B during the time dt.

The volume of the bacteria being considered as constant, we can easily translate the number of molecules into concentrations and *vice versa*.

If κ and γ are the rate of synthesis and degradation of B, the evolution of the concentration of B per unit time is given, in a deterministic approach, by the following ordinary differential equation:

$$\frac{d[B]}{dt} = \kappa - \gamma[B]$$

This equation reflects the average comportment of a cell in a population.

In a stochastic approach, the variability of the comportment of the different cells from the population will be taken into account. Thus we will think in terms of probability and a specific differential equation will be used named master equation.

Stochastic process

A master equation is a differential equation describing the temporal evolution of a probability distribution. This distribution represents the probability of a system to occupy each of the sets of discrete states according to a continuous time variable.

In our simple case:

➢ At the cell level, we will work with a probability $P(n,t)$ to have n proteins B in a cell at time t. For example, it will have 25% chance of having 5 proteins B in the cell, 50% chance of having 6 proteins B and 25% chance of having 7 proteins.

➢ At the population level, we will have a distribution. If the size of the population is of 1000 cells, that means that approximatively 250 cells will own 5 proteins B, 500 cells will possess 6 proteins and 250 cells will have 7 proteins.

Stochastic process

How to write the master equation ?

If P(n,t) is the probability to have n proteins B in a cell at time t, the goal is to describe the evolution of P(n,t) as a function of time $\dfrac{d[P(n,t)]}{dt}$

<u>Hypothesis:</u> we know the state of the system at time t, *i.e.*, the number of proteins B in the cell.

<u>Question:</u> what is the probability P($n,t+dt$) to have n proteins B at time $t+dt$, with dt very small so that only one synthesis or degradation reaction can occur.

<u>Answer:</u> three cases have to be considered
- at time t, there are n-1 proteins B and that will be the synthesis reaction that will contribute to P($n,t+dt$) for a value: $\kappa\, dt$ P(n-1,t)
- at time t, there are n+1 proteins B and that will be the degradation reaction that will contribute to P($n,t+dt$) for a value: γ (n+1) dt P(n+1,t) (the degradation rate depends on the protein number that is why the term (n+1) appears in the equation)
- at time t, there are n proteins B and this number does not change. For that, we have to consider the probability of having n proteins at t P(n,t) from which is subtracted the probability that a synthesis reaction occurs during the time dt ($\kappa\, dt$ P(n,t) and the probability that a degradation event takes place during the time dt (γ n dt P(n,t).

$$P(n, t + dt) = \kappa\, P(n-1, t)\, dt + \gamma\, (n+1) P(n+1, t)\, dt + P(n, t) - \kappa\, P(n, t)\, dt - \gamma\, n\, P(n, t) dt$$

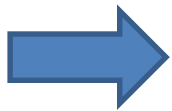# Stochastic process

$$P(n, t + dt) = \kappa\, P(n - 1, t)\, dt + \gamma\, (n + 1)P(n + 1, t)\, dt + P(n, t) - \kappa\, P(n, t)\, dt - \gamma\, n\, P(n, t)dt$$

From the above equation, we can deduce the derivative form that corresponds to the **master equation**:

$$\frac{dP(n, t)}{dt} = \kappa\, [P(n - 1, t) - P(n, t)] + \gamma[\, (n + 1)P(n + 1, t) - n\, P(n, t)]$$

However, for more complex system, the master equations can no longer be analytically integrated.
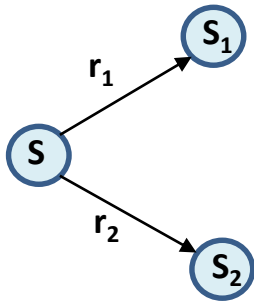
Stochastic simulations

The Gillespie stochastic simulation algorithm

The essential point of this algorithm is to create two random numbers to determine:
- time at which the next reaction occurs
- which next chemical reaction occurs

Stochastic simulations

If two transitions are enabled then the two events are competing :

There is a race between events. The fastest event is executed first. This execution will modify globally the state of the system.

Continuous Time Markov Chains :
It is a discrete set of states connected by transitions. Each transition is associated with a rate of an exponential distribution. In each state transitions compete in a race condition: the fastest one determines the new state and the time elapsed. In the new state, the race condition starts over (memoryless property).

Stochastic simulations

A reaction R_n, for example $A + B \rightarrow C$, can occur when a molecule of species A and a molecule of species B collide with enough energy. The probability $P(R_n, dt)$ that the reaction R_n occurs during the infinitesimal time interval dt **is proportional to the duration of the time interval dt, the number of possible collisions h_n (here the product of the number of molecules of species A and B respectively) and kinetics (rate) c_n specific to this reaction R_n.** We obtain:

$$P\,(R_n,\, dt) = h_n c_n dt$$

The term $a_n = h_n c_n$ **is called the propensity of the reaction R_n.**
Each propensity is the probability per unit time that a specific reaction occurs.

If there are M reactions, the probability per unit time that any reaction occurs is just the sum of the propensity of each reaction:

$$a_0 = \sum_{r=1}^{r=M} a_r$$

Example: two reactions with propensities a_1 and a_2 respectively. The probability that reaction 1 occurs is given by $a_1 = \dfrac{a_1}{a_1 + a_2} = \dfrac{a_1}{a_0}$

Stochastic Petri net

- Each transition gets its own local timer.
   When a transition becomes enabled (enough tokens in its pre-places), the local timer is set to an initial value computed by means of the corresponding probability distribution (in general, this value will be different for each run of simulation). The local timer is then decremented at a constant speed, and when the timer reaches zero, the transition is fired. If many transitions are enabled, a race of the next firing will take place.

- Technically, various probability distributions can be chosen to determine the random values for the local timers. Biochemical systems are prototypes for exponentially distributed reactions
   The firing rates of transitions will follow an exponential definition which could be described by a single parameter λ.
   The firing rate will be described by its own parameter λ to specify its local time behavior.
   The waiting time is an exponential distributed random variable $X_t \in [0,\infty[$ with the probability density function:
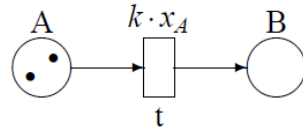
$$f_{X_t}(\tau) = \lambda_t(m)e^{-\lambda_t(m)\tau} \quad , \tau \geq 0$$

Here, as we are working with biochemical (molecular) network, the stochastic hazard $\lambda_t(m)$ for the transition t will correspond to the *stochastic mass-action hazard function. It will depend on the* transition-specific stochastic rate constant and on the number of tokens present in the preplaces of the transition t.

Stochastic Petri net

Example 1:

$k \cdot x_A$

A   B

k is the rate constant of the reaction

t

transition *t is* enabled because its input place *A* is marked. A firing time $\tau 1$ is thus chosen for *t*, drawn from the negative exponential distribution of parameter $k\, x_A = 2k$, and a clock starts to countdown from $\tau 1$ to 0. When the clock reaches 0, transition *t* fires. A new marking is obtained $x_A = 1$, $x_B = 1$.
After the firing, transition *t* is still enabled, but its rate has now become $k\, x_A = k$.
Consequently, its new firing time $\tau 2$ will be selected from an exponential random variable different from the one out of which $\tau 1$ was sampled. Again, a clock is set to countdown until the new firing time is reached. At that time, the marking is changed to $x_A = 0$, $x_B = 2$, where no transitions are enabled anymore and the evolution stops.

Example 2:

A

$k \cdot x_A \cdot x_B$ C

B

t
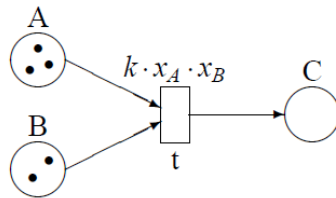
Transition *t* is enabled as both places A and B are not empty.
In the initial marking of the model, there are six several independent ways in which the bimolecular reaction can occur, each one associated to one specific selection of the pair of molecules A and B that react. Thus, the rate associated to transition *t* in the initial marking is: $k\; x_A\, x_B = 6k$.
After the firing, the marking is changed to $x_A = 2$, $x_B = 1$, $x_C = 1$
The subsequent firing of transition *t* will occur at a rate that is: $k\, x_A\, x_B = 2k$.

Stochastic Petri net

The higher the propensity of the reaction, the shorter the waiting time will be.
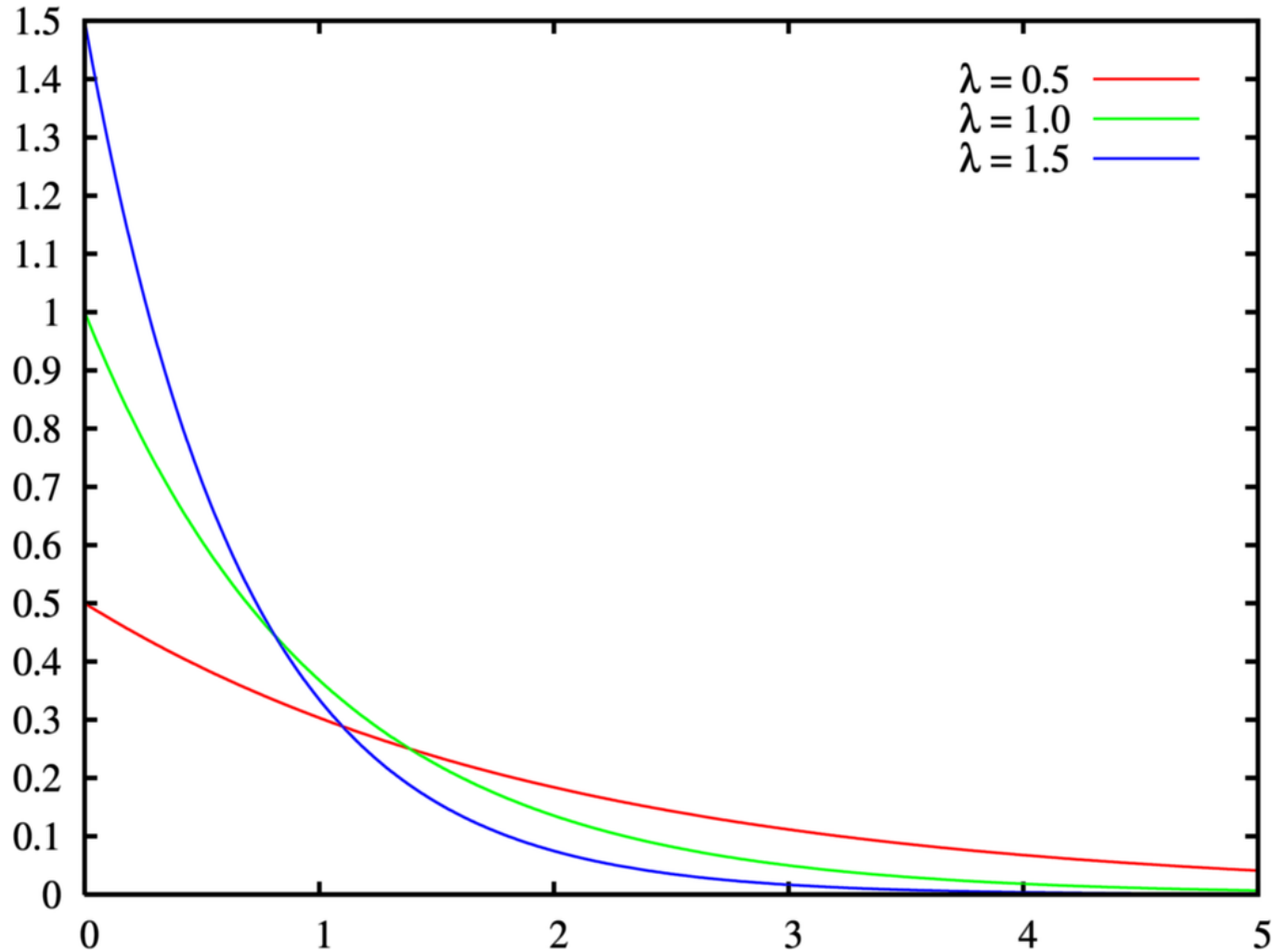
Stochastic simulations: selecting a random reaction time

The biochemical systems are prototypes for exponentially distributed reactions

As reminder : an exponential distribution models the time of occurrence of a (simple) random event. It is given by a random variable T, with values in $[0, \infty [$, with density : $f(t) = \lambda e^{-\lambda t}$ where λ is the rate of the exponential distribution.

The probability of the event happening within time t is : $P(T \leq t) = 1 - e^{-\lambda t}$ λ is the average density of frequency of events per unit of time.

We have see that the probability per unit of time that any reaction occurs is a_0 (the sum of the propensity of each reaction).
If time t corresponds to time interval $(t\text{-}t_{ref})$ where t_{ref} is a reference time (ex: time when the last system reaction occurred) then :
$$P_{reacted} = 1 - e^{-a_0(t - t_{ref})}$$

The distribution of reaction times is therefore :

$$p(t) = \frac{dP_{reacted}}{dt} = a_0 e^{-a_0(t - t_{ref})}$$

➢ Thus, we need to generate exponentially distributed reaction times. It will allow to determine the time τ (after t_{ref}) corresponding to the time when the next reaction will take place.

Stochastic simulations: selecting a random reaction time

However most random number generators are based on a uniform distribution between 0 and 1. But a uniform random number r_1 could be convert into an exponentially distributed random number as follow:

$$\tau = \frac{1}{a_0} \ln \frac{1}{r_1}$$

Stochastic simulations: Gillepsie algorithm

The system to be simulated involves:
- N molecular species $\{S_1, \ldots, S_N\}$ represented by a vector of dynamic state
$X(t) = (X_1(t), \ldots, X_N(t))$ where $X_i(t)$, is the number of molecules of the species S_i in the system at time t
- M chemical reactions $\{R_1, \ldots R_M\}$. Each reaction R_j is characterized by its propensity a_j and a vector of state change $v_j = \{v_{1j}, \ldots, v_{Nj}\}$, where v_{ij} is the variation of the number of molecules of the species S_i due to the reaction R_j.

Step 1: Determination of the time τ corresponding to the time when the next reaction will occur.
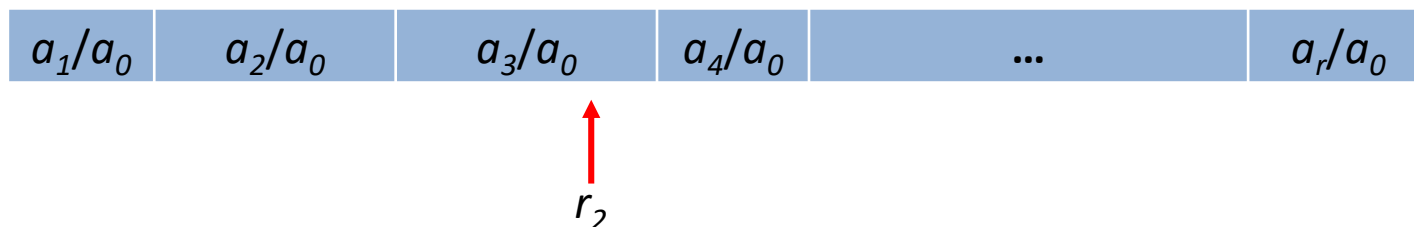Generate an random number, r_1 from an uniform distribution [0,1]. Deduce τ

$$\tau = \frac{1}{a_0} \ln \frac{1}{r_1}$$

Step 2: Random choice of the reaction that occurs at time τ
Generate a second random number, r_2 from an uniform distribution [0,1].
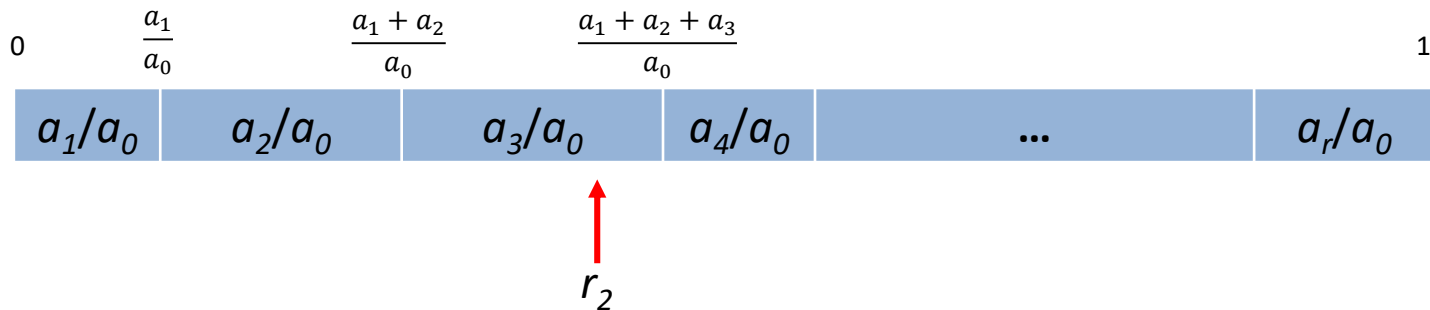The probability that the next reaction to occur is R_r is a_r / a_0 (a_0 probability per unit time that any reaction occurs).
Rank the probability of each reaction. Figure out in which reaction interval r_2 falls.

a_1/a_0	a_2/a_0	a_3/a_0	a_4/a_0	...	a_r/a_0

r_2

Stochastic simulations: Gillepsie algorithm

Step 2:

$$0 \qquad \frac{a_1}{a_0} \qquad \frac{a_1 + a_2}{a_0} \qquad \frac{a_1 + a_2 + a_3}{a_0} \qquad\qquad\qquad 1$$

a_1/a_0	a_2/a_0	a_3/a_0	a_4/a_0	...	a_r/a_0

r_2

Using r_2 as a pointer to determine which reaction will occur means to find the value of k that satisfied the following inequality:

$$\sum_{r=1}^{k-1} a_r < r_2 \, a_0 \leq \sum_{r=1}^{k} a_r$$

Alternatively, we are looking for the index k that is the smallest integer in the interval [1, M] such that:

$$\sum_{r=1}^{k} a_r > r_2 \, a_0 \qquad \text{In our example, the chosen reaction will be } R_3$$

Step 3: the choice of the reaction will change the number of molecules at time (t+τ) for the molecular species that are concerned by the reaction. The vector $X(t) = (X_1(t), ..., X_N(t))$ will be updated to represent the new number of molecules of each species.
New reaction propensities must be calculated.

Gillepsie algorithm

1. Setup: Store initial populations and rate constants, set $t = 0$,
2. Calculate reaction propensities.
3. Generate two uniform random numbers, r_1 and r_2.
4. Calculate τ, the time to next reaction, using r_1.
5. Determine the next reaction using r_2.
6. Add τ to t.
7. Update the populations based on the reaction chosen.
8. Go to step 2 until some chosen stopping criterion is reached (exhaustion of a chemical, target simulation time reached, . . .)

If we want the kind of information we can get from the master equations, many independent runs of simulations must be performed   and then average across realizations.

Stochastic Petri net

One simulation run describes at least one path in the state space graph (Gillepsie algorithm).
It is also possible to perform multiple simulation runs and average the results of all runs.
Thus, an averaged time course will be computed. The more simulation runs are performed, the more precise is the averaged time course. All single simulation runs will fluctuate around the averaged time course.
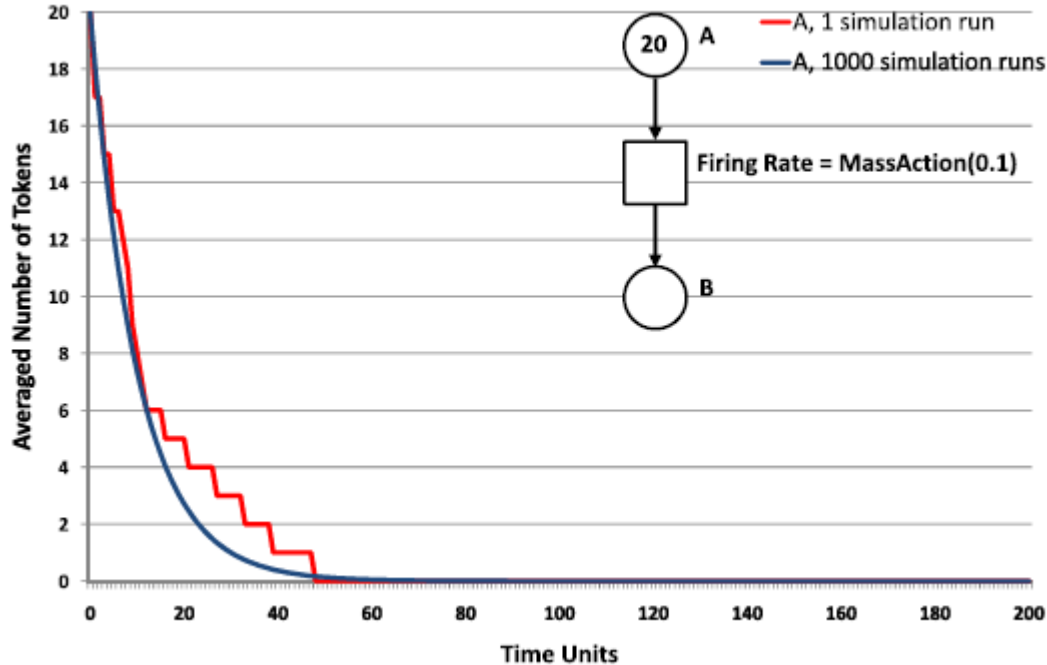
Stochastic Petri net

Stochastic paradigm ($\mathcal{SPN}$):

- preserve a discrete number of tokens on its place, but a firing rate is associated with each transition (random variables defined by probability distributions)
- the firing rates are state dependent and specified by rate functions
- all reactions, which occur in the $\mathcal{QPN}$, can still occur in the $\mathcal{SPN}$, but their likelihood depends on the probability distribution of the associated firing rates
- thus, the system behavior is described by the same discrete space as in the $\mathcal{QPN}$
- all qualitative properties valid in the $\mathcal{QPN}$ are also valid in the $\mathcal{SPN}$, and *vice versa*

- before firing of an enabled transition $t \in T$, a waiting time has to elapse

Stochastic Petri net

In stochastic Petri net, new type of transitions (timed transitions) are available:

➢ **Deterministic transitions** : they have contrary to stochastic transitions a deterministic firing delay which is specified by an integer constant. The delay is always relative to the time point where the transition gets enabled.

➢ **Immediate transitions**: they are a very special kind of deterministic transitions with zero firing delay, *i.e.* they fire immediately after getting enabled, and always prior to (general) deterministic and stochastic transitions. Consequently, getting enabled and the firing itself coincide for immediate transitions

➢ **Scheduled transitions**: they are another special case of deterministic transitions. The deterministic firing occurs according to a schedule specifying absolute points of the simulation time. A schedule can specify just a single time point, or equidistant time points within a given interval, triggering the firing once or periodically. However, transitions only fire at their scheduled time points if they are enabled. Scheduled transitions can dramatically restrict the (qualitative) net behavior. Scheduled transitions can be replaced by net components consisting of immediate and deterministic transitions only; Thus, they do not extend the modelling power.

Scheduled transitions are described by [Start, Repetition, End]. They fire as soon as the fixed time interval elapsed during the given time-points.

Stochastic Petri net

Summary