

## Elementary modeling of a Markov process

The goal of this project is to model so-called Markov jump processes, widely used in population dynamics, chemical kinetics, theory of radioactive decay, etc. These processes describe systems with discrete states (e.g. given by a number of particles of a given type), with stochastic transitions between these states (see [https://en.wikipedia.org/wiki/Markov\\_chain](https://en.wikipedia.org/wiki/Markov_chain)). The transitions do not depend on the past, but just on the present configuration, and are described through rates. The simplest Markov process is the Poisson process, where at a given rate  $r$  particles of type  $A$  are added to the system. This means that during small time interval  $\Delta t$ , the probability that a particle will be added is  $r\Delta t$ . From the theory of Markov processes it follows that the time intervals between the events of adding a particle  $\tau = t_{i+1} - t_i$  are distributed according to the exponential distribution with density

$$w(\tau) = re^{-r\tau}$$

If the number of particle is  $N$ , then at the transition  $N \rightarrow N + 1$ . If one neglects fluctuations, the mean number of particles grows deterministically via mean-field equations

$$\frac{dN}{dt} = r.$$

More complex processes are the birth and death processes, where each existing particle can produce a new one with rate  $r_b$ , or disappear with rate  $r_d$ . Because the particles are independent, the transition can happen with any of them, and the rates of transitions  $N \rightarrow N + 1$  and  $N \rightarrow N - 1$  are  $r_b N$  and  $r_d N$ , respectively. The mean-field equations are

$$\frac{dN}{dt} = r_b N, \quad \frac{dN}{dt} = -r_d N$$

To describe a saturation of the population growth, one uses a logistic (Verhulst) nonlinear mean-field equation

$$\frac{dN}{dt} = r_b N - r_{nl} N^2$$

which can be considered as a superposition of a linear birth process and of a nonlinear death process.

To model a superposition of elementary Markov processes, one uses **the Gillespie algorithm** [https://en.wikipedia.org/wiki/Gillespie\\_algorithm](https://en.wikipedia.org/wiki/Gillespie_algorithm). This algorithm is based on the fact, that for a superposition of several processes with rates  $r_1, r_2, \dots, r_n$  the time between the events is a random variable distributed exponentially with rate  $r = \sum_{k=1}^n r_k$ . Which of  $n$  events happens, can be modeled by choosing a discrete random number  $p$  out of  $n$  possibilities having probabilities  $\frac{r_1}{r}, \frac{r_2}{r}, \dots, \frac{r_n}{r}$ .

Below you have to model several Markov processes with Gillespie algorithm, and compare the results to mean-field equations.

**Task 1:** Model a death process  $A \xrightarrow{r} \emptyset$  with initial number of particles  $N(0) = 20$  and rate  $r = 0.1$ . Present several sample realization of the Markov process; calculate the average  $\langle N(t) \rangle$  and the standard deviation.

**Task 2:** Model the gene expression for a single cell. The mean field equations read [M. Scott, Tutorial: Genetic circuits and noise, <http://www.math.uwaterloo.ca/~mscott/NoiseTutorial.pdf>.]

$$\begin{aligned}\frac{dM}{dt} &= \lambda_m - d_m M \\ \frac{dP}{dt} &= \lambda_p M - d_p P\end{aligned}$$

for  $\lambda_m = 1$ ,  $\lambda_p = 1$ ,  $d_m = 0.2$ ,  $d_p = 0.02$ .

Find the equilibrium distribution density of  $P$  and compare with to theoretical approximate prediction

$$w(P) \approx \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[ -\frac{(P - \langle P \rangle)^2}{2\sigma^2} \right]$$

where  $\langle P \rangle = \frac{\lambda_m \lambda_p}{d_m d_p}$  and  $\sigma^2 = \langle P \rangle (1 + \frac{\lambda_p}{d_m + d_p})$ . Plot the density in normal and logarithmic scales. Calculate the skewness and the kurtosis of the distribution.

**Task 3:** Model of extinction in the logistic (Verhulst) process.

$$\frac{dN}{dt} = \lambda N - dN^2$$

for  $\lambda = 1$ ,  $d = 0.1$ .

Here the mean level of the population is  $N_0 = \frac{\lambda}{d}$ , but there are stochastic deviations. If at some time  $N = 0$ , the population extincts, but for large  $\hat{N}$  this is rather improbable.

Find a dependence of the mean extinction time starting from  $N(0) = N_0 = \lambda/d$  on the values of  $N_0$  (vary these in the interval  $10 \leq N_0 \lesssim 20$ ).

**Task 4:** SIR model of epidemic spreading. Here there are 3 types of individuals:  $S$  - susceptible,  $I$  - infected,  $R$  - recovered. The rate of infection is proportional to the probability of  $S$  and  $I$  to meet, i.e. to  $SI$ . The rate of recovering is constant. One neglects all other processes like death rate etc. The mean-field equations are

$$\begin{aligned}\frac{dS}{dt} &= -bSI \\ \frac{dI}{dt} &= bSI - cI \\ \frac{dR}{dt} &= cI\end{aligned}$$

Elementary events here is infection, at which  $S \rightarrow S - 1$ ,  $I \rightarrow I + 1$ ; and recovery at which  $I \rightarrow I - 1$ ,  $R \rightarrow R + 1$ . The epidemic ends when  $I = 0$ .

**4.1:** Solve mean-field equations for  $b = 0.02$  and  $c = 0.3$ , starting from  $S = 30$ ,  $I = 1$ ,  $R = 0$ . Find the state at  $t \rightarrow \infty$ , which gives an average number of totally infected persons.

**4.2** Apply the Gillespie method, simulate epidemic spreading as a stochastic process, and find a distribution of overall infected persons during the epidemic for  $S = 30$ ,  $I = 1$ ,  $R = 0$ .