

# Simulering og Statistikk - Module 1, January 2014

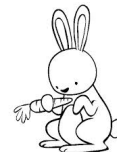
Dr S.B. van Albada and Prof Dr H.G. Schaathun

In this tutorial we focus on different ways to simulate complex systems. As an example, we consider the predator-prey system, which is the topic of this module.

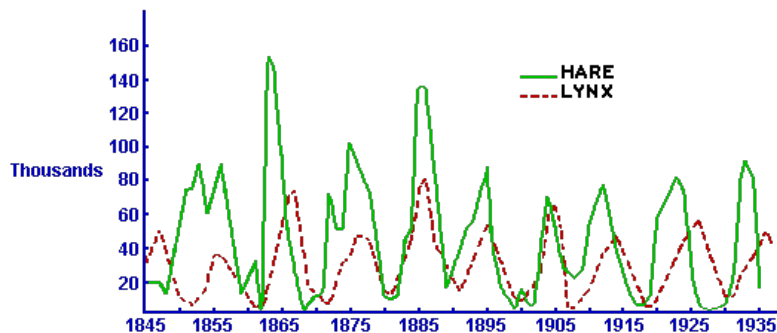
## I. Predator-prey model



In the predator-prey model, we consider the dynamics of populations of a predator species (e.g., foxes) and a prey species (e.g., rabbits). In some areas, foxes are major predators of rabbits. The populations of both species in these areas show large oscillations over time.



An example of such oscillations in a different predator-prey system, consisting of numbers of a predator, the *Canadian Lynx*, and its prey, the *Snowshoe Hare*, are shown in the following figure<sup>1</sup>:



**Figure 1.** This figure shows the numbers of *Canadian Lynx* and *Snowshoe Hares* that were trapped for their pelts over a period of time covering nearly a century. We may assume that these numbers were proportional to the total numbers of animals in the populations.

In order to be able to fully understand a predator-prey system such as that of the lynx and the hare, we have to be sure that other factors than the interaction between the predators and preys do not play a big role. To begin with, if, in addition to the lynx, there would be another predator of the hare, we would need information about (at least) the numbers of that other predator in order to fully understand the system. Secondly, in case the food supply of the hare would not be abundant, we would need to consider that as well. Moreover, if the lynx would have another food source in addition to the hare, we would have to take that into account as well. Lastly, in case of any other external factor such as a substantial change in climate, we would not be able to fully understand the dynamics of the

<sup>1</sup> Source: <https://www.math.duke.edu/education/ccp/materials/diffeq/predprey/pred1.html>

predator-prey system without more information.



1. On average, what was the period of oscillation of the lynx population<sup>2</sup>?



2. On average, what was the period of oscillation of the hare population<sup>3</sup>?



3. On average, do the peaks of the predator population match or slightly precede or slightly lag those of the prey population? If they don't match, by how much do they differ? (Measure the difference, if any, as a fraction of the average period.)<sup>4</sup>



4. How many lynxes and hares were approximately caught in the worst years?

## II. Covariance and correlation

In the previous tutorial, we have studied properties of a single stochastic variable  $X$ , such as the mean  $\mu_X = E[X]$  and the variance  $\sigma_X^2 = E[(X - \mu_X)^2]$ . We have learnt that the variance is the mean squared distance from the mean.

In the data set shown in figure 1, we do not have a single stochastic variable, but two: For each year, we have the number of trapped hares, and the number of trapped lynxes. In such cases where more than one stochastic variable are measured simultaneously, it is interesting to know if these variables change together, since this could mean that there is a dependence between them. If stochastic variables change together, we say that they are *correlated*.



5. A company produced 5 models of boxes of different volume and height. Table 1 shows the different models with their prices, as well as how many items were sold:

Model	Volume (L)	Height (cm)	Price (kr.)	Numbers sold
A	2	10	300	90
B	3	15	400	80
C	4	20	350	60
D	5	25	500	20
E	6	30	450	50

Table 1. *Five different boxes with their volumes, height, price, and numbers sold.*

- Make a plot of the height as a function of the volume, a plot of the price as a function of the volume and a plot of the numbers of boxes sold as a function of volume.
- Is there a correlation between volume and height?
- Is there a correlation between volume and price?
- Which correlation is larger?

<sup>2</sup> Source: <https://www.math.duke.edu//education/ccp/materials/diffeq/predprey/pred1.html>

<sup>3</sup> Source: <https://www.math.duke.edu//education/ccp/materials/diffeq/predprey/pred1.html>

<sup>4</sup> Source: <https://www.math.duke.edu//education/ccp/materials/diffeq/predprey/pred1.html>

- e. Is there a correlation between the volume of the boxes and how many of them were sold? If so, is it a positive, or a negative correlation?

The *covariance* is an estimator that tells us if two stochastic variables  $X$  and  $Y$  are correlated:

$$\sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)]. \quad (1)$$

In case of a discrete probability distribution:

$$\sigma_{XY} = \frac{1}{n} \sum_i (X_i - \mu_X)(Y_i - \mu_Y) \quad (2)$$



6. In this exercise we compute the population covariance  $\sigma_{VH}$  between volume (V) and height (H), the population covariance  $\sigma_{VP}$  between volume and price (P) and the population covariance  $\sigma_{VN}$  between volume and numbers sold (N) (see table 1).

- Compute the average volume, height, price and number of the boxes sold.
- Use equation (2) to compute the covariances  $\sigma_{VH}$ ,  $\sigma_{VP}$  and  $\sigma_{VN}$ . Do the numbers reflect the correlation between the different quantities? In what sense?

We have seen that the covariance shows if there is a positive, or a negative correlation between two stochastic variables. The result is proportional to the spread in  $X$  and the spread in  $Y$ . Another measure for the correlation between two stochastic variables  $X$  and  $Y$  is the *correlation coefficient*. It is defined as:

$$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}. \quad (3)$$

The correlation coefficient returns a value between -1 (perfectly anti-correlated) and 1 (perfectly correlated). A correlation coefficient of zero means that the variables are uncorrelated.



7. In this exercise we compute the correlation coefficients  $\rho_{VH}$ ,  $\rho_{VP}$  and  $\rho_{VN}$  (see table 1).

- Compute the standard deviations  $\sigma_V$ ,  $\sigma_H$ ,  $\sigma_P$  and  $\sigma_N$ .
- Compute the correlation coefficients  $\rho_{VH}$ ,  $\rho_{VP}$  and  $\rho_{VN}$ .
- Do the correlation coefficients reflect the observed amounts of correlation in the data observed in exercise 5?



8. In this exercise we study the correlation between the numbers of trapped lynxes and hares (see figure 1).

- Do you expect to find a correlation between the numbers of trapped hares and lynxes in figure 1?

function Correlation

```
clear;
years=1845:2:1903;
```

```

lynx_pelts =
1000*[32,50,12,10,13,36,15,12,6,6,65,70,40,9,20,34,45,40,15,15,60,80,26,18,37,
50,35,12,12,25];           %Numbers of lynxes trapped
hare_pelts =
1000*[20,20,52,83,64,68,83,12,36,150,110,60,7,10,70,100,92,70,10,11,137,137,
18,22,52,83,18,10,9,65];   %Numbers of hares trapped
cov(lynx_pelts, hare_pelts) % the covariance matrix
corrcoef(lynx_pelts, hare_pelts) % the matrix of correlation coefficients
figure
plot(lynx_pelts, hare_pelts,'x')
xlabel('lynx pelts')
ylabel('hare pelts')

```

- b. Download and run the above MATLAB script Correlations.m.
- c. Study the scatter plot in which the number of trapped hares is plotted as a function of the number of trapped lynxes. What value for the correlation coefficient would you approximately expect?

The code returns the covariance matrix and the matrix of correlation coefficients for the numbers of trapped hares (H) and lynxes (L). These matrices consist of the covariances  $\begin{bmatrix} \sigma_{LL} & \sigma_{LH} \\ \sigma_{HL} & \sigma_{HH} \end{bmatrix}$  and correlation coefficients  $\begin{bmatrix} \rho_{LL} & \rho_{LH} \\ \rho_{HL} & \rho_{HH} \end{bmatrix}$ . Study the numbers.

- d. Why do the covariances have such large values?
- e. Why are the diagonal elements of the matrix of correlation coefficients equal to one?
- f. Compare the correlation coefficient to your estimated value in c.

### III. Lotka-Volterra equations

One of the first and basic models that try to explain observed population fluctuations as considered in the previous section, was the *Lotka-Volterra equation*. The Lotka-Volterra equations are a pair of coupled differential equations:

$$\frac{dx}{dt} = x(\alpha - \beta y) \quad (3)$$

$$\frac{dy}{dt} = -y(\gamma - \delta x) \quad (2)$$

The variable  $x$  denotes the number of prey animals,  $y$  the number of predators;  $t$  denotes time and  $\alpha$ ,  $\beta$ ,  $\gamma$  and  $\delta$  are the parameters of the model, all positive numbers. Let us first try to understand the model.



9. In this exercise, we will investigate the meaning of the model parameters.
  - a. What does  $\frac{dx}{dt}$  mean?
  - b. What would  $\frac{dx}{dt} = \frac{10000}{\text{month}}$  mean?
  - c. Explain that reproduction of prey animals can be described by  $\frac{dx}{dt} = \alpha x$ , where  $\alpha$  is the (average) number of offspring of a prey animal per time.

- d. What would  $\frac{dx}{dt} = -\frac{10000}{\text{month}}$  mean?
- e. Explain that the consumption of prey animals by predators can be described by  $\frac{dx}{dt} = -\beta xy$ , where  $\beta$  is the probability per unit time that a given predator animal catches a given prey animal.
- f. Explain that passing away of predator animals can be described by  $\frac{dy}{dt} = -\gamma y$ , where  $\gamma$  is the (average) probability per unit of time that a predator animal passes away.
- g. Explain that reproduction of predator animals can be described by  $\frac{dy}{dt} = \delta xy$ , if we assume that the birth rate of predators increases linearly with the number of prey animals consumed. Hint: Use exercise e.
- h. Explain that the number of predators that are born per number of prey animals that they consume equals  $\delta/\beta$ .

We say that the system of differential equations is *first-order*, as only first derivatives of  $x$  and  $y$  are present. Furthermore, the differential equations are *non-linear*, as products of  $x$  and  $y$  appear in the equations.

#### IV. Numerical integration of differential equations

How can we solve differential equations like the Lotka-Volterra equations? To begin with, non-linear differential equations are often difficult to solve analytically. However, for given parameters and initial conditions, one can often find the solution to great accuracy by numerical methods. An exact solution of the Lotka-Volterra equations actually does exist. However, we will here focus on *numerical integration of the differential equations*, as these methods are generally applicable.

##### IV.1. Euler method

The essence of numerical integration of differential equations is as follows. Consider the definition of the derivative:

$$\frac{dx}{dt} = \lim_{\Delta t \rightarrow 0} \frac{x(t+\Delta t) - x(t)}{\Delta t} \quad (3)$$

If we now, instead of an infinitesimal time step take a finite, but small time step  $\Delta t$ , we can approximate equation 1 as:

$$\frac{x(t+\Delta t) - x(t)}{\Delta t} = x(\alpha - \beta y) \quad (4)$$

or, reordering terms:

$$x(t + \Delta t) = x(t) + \Delta t \cdot x(\alpha - \beta y) \quad (5)$$

For given parameters  $\alpha$  and  $\beta$  and initial conditions (for example,  $x(0)=20\,000$  rabbits) we can approximate the number of rabbits after one small time step (for

example,  $\Delta t = 15$  min). It is that simple! We do the same for the number of predators

$$y(t + \Delta t) = y(t) - \Delta t \cdot y(\gamma - \delta x) \quad (6)$$

and achieve in this way the numbers of predators  $y(t+\Delta t)$  and prey  $x(t+\Delta t)$  after one time step  $\Delta t$ . From there on, we proceed in exactly the same way in order to compute the numbers of animals after two time steps, and so on. This method is called the **Euler method**.



10. Given are the following parameters:  $\alpha = 1.1$ ;  $\beta = 0.07$ ;  $\gamma = 1.3$  and  $\delta = 0.02$ ; and as initial conditions: 20.0 foxes and 200.0 rabbits at  $t = 0$ .

- Compute the numbers of foxes and rabbits after a time step of size 0.01.
- Compute the numbers of foxes and rabbits after one and two time steps of size 0.005.
- Compare the results in a. and b.

Numerical integration is a way of simulating a system described by a set of differential equations. It is one of the most-used methods for any engineer or scientist, and is applied, to mention some examples, in flight simulators, chemical process modelling, and simulations of electrical circuits.

However, there exist much faster (and more stable) methods than the Euler method. The most famous (but not the best) of these is the **Runge-Kutta** method. Whereas the Euler method only makes use of the *first derivative* of the right-hand side of differential equations like (1) and (2), Runge-Kutta approximates the functions on the right-hand side to a higher order of choice by cleverly combining several Euler steps. We will not go into the details of the Runge-Kutta method here, but we will use a MATLAB implementation of it in the next section.

## IV.2. Simulation



11. Input and run the following script in MATLAB:

```
clear

% Define boundary conditions:
t0 = 0;
tfinal = 15;
y0 = [200 20]';

% Simulate the differential equation:
[t,y] = ode23('lotka',[t0 tfinal],y0);

plot(t,y)
```

where the function `lotka.m` should be saved as a separate file:

```
function yp = lotka(t,y)
%LOTKA Lotka-Volterra predator-prey model.
```

```

% Copyright 1984–2002 The MathWorks, Inc.
% $Revision: 5.7 $ $Date: 2002/04/15 03:33:21 $

yp = diag([1.1 - .07*y(2), -1.3 + .02*y(1)])*y;

end

```



12. Modify the initial values and parameters of the Lotka-Volterra model in order to best fit the empirical data from figure 1. Consider:

- a. (Average) peak heights of numbers of predators and prey.
- b. (Average) peak widths.
- c. Period of the oscillations.
- d. Phase shift.



13. Compare the output of your best fit quantitatively to the empirical data of figure one. Consider:

- a. The phase shift between predator and prey population peaks.
- b. The values of maxima and minima, including stochastic variations and their possible causes.
- c. Period of the oscillation, including stochastic variations and their possible causes.

As we see very clearly by comparing the observed data to our deterministic model, stochastic fluctuations in the numbers of lynxes and hares play a big role, on top of the big oscillations predicted by the deterministic model. In the next section, we will study a stochastic model that takes into account an important part of these stochastic fluctuations.

### V.1. Gillespie algorithm - Introduction

In the previous section, we have modeled the predator-prey system by a set of deterministic differential equations, without stochastic fluctuations. Changes in the populations were calculated by using average rates for the whole population, such as the average birth rate at which foxes catch rabbits and the average birth rate of rabbits.

However, we did not take into account that the populations of foxes and rabbits actually consist of individuals: The foxes do not continuously catch small portions of rabbit – instead, every now and then a fox catches a rabbit, and sometimes they do not manage to catch a rabbit over a long time. Similarly, every now and then a number of rabbits are born.

These important stochastic effects are taken into account in discrete stochastic algorithms such as the *Gillespie algorithm*. In this algorithm, all events (such as birth of a rabbit) have a given probability per unit time. For example, the *probability* that a rabbit is born during a short time period  $\Delta t$  equals  $ax \cdot \Delta t$ . N.B.

There is a big difference from the *deterministic* simulation given by eq. 5, where the *number* of rabbits born during a time period  $\Delta t$  equals  $\alpha x \cdot \Delta t$ . For example, instead of a continuous decrease of 0.25 foxes per hour, in the Gillespie algorithm, the probability that a fox dies during the next hour is one in four.

This stochastic effect makes a huge difference, especially if the population size is small. Consider the extreme case that only one fox is present. According to the Gillespie algorithm, the fox will either die, survive or get offspring during the next time step. However, in the differential equation model, a small fraction of the fox will die during the next time step whereas another fraction of the fox will get offspring, such that we will end up with for example 1.03 or 0.98 foxes, which clearly is incorrect.

## V.2. Intermezzo: Poisson process

A process that has a constant probability per unit time (or constant “rate”) is called a *Poisson process*. Some examples:

- The *probability* that a fox passes away from the population during a *short* time interval  $\Delta t$  (i.e. much shorter than the average time between two deaths) is proportional to the duration of the time interval. If we wait twice as long, the probability will be twice as large.
- Radioactive decay: in a *short* time interval (i.e., much shorter than the decay time), the probability that a nucleus decays is proportional to the duration of the time interval.
- In a service center, the probability to get an incoming phone call in a *short* waiting time (i.e., much shorter than the average time between two calls) is proportional to the waiting time.

For a Poisson process with a rate  $\lambda$ , the probability  $p(k)$  for a number of  $k$  events happening in a time  $t$  is given by:

$$p(k) = P(X = k) = \frac{e^{-\lambda t} (\lambda t)^k}{k!} \quad (7)$$

It is a good exercise to prove this on a rainy day! But for the Gillespie algorithm, we actually only need the probability that *no* event has happened in a time  $t$ .



14. Show that the probability that no event has happened during a time  $t$  is given by:  $p(0) = P(X = 0) = e^{-\lambda t}$ .



15. Explain that the probability  $P(X = 0)$  that no event has happened during a time  $t$  is equal to the probability  $P(T > t)$  that the *first* event happens at time  $t$  or later.



16. Show that the cumulative distribution function for the time  $t$  of the first event is given by  $F(t) = P(T \leq t) = 1 - e^{-\lambda t}$ .

A very elegant method (and the most direct) for drawing a random number from a given distribution  $f(x)$  is as follows:



Take a uniformly distributed random number from the interval  $[0, 1]$ , and input it into the inverse of the *cumulative* distribution function  $F^{-1}(x)$ .



17. Show that the inverse cumulative distribution function of  $f(t)$  is given by  $F^{-1}(t) = \frac{1}{\lambda} \ln\left(\frac{1}{1-r}\right)$  starting from exercise 16.

Okay, we have our result: In order to sample a time duration until of the next event of a Poisson process with a rate constant  $\lambda$ , we simply feed a uniformly distributed random number  $r$  from the interval  $[0, 1]$  into  $\frac{1}{\lambda} \ln\left(\frac{1}{r}\right)$ .

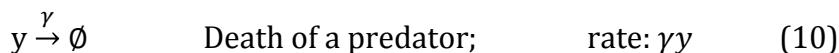
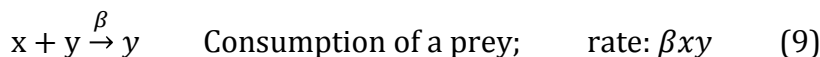
N.B. It is easy to understand that we just as well can use  $r$  instead of  $1-r$  for a uniformly distributed random number in  $[0,1]$  – compare to exercise 17.



18. Draw 5 next event times for a Poisson process of rate constant  $\lambda = 1 \text{ s}^{-1}$  and 5 next event times for a Poisson process of rate constant  $\lambda = 1 \text{ minute}^{-1}$ .

### V.3. The Gillespie algorithm

Now that we know how to handle Poisson processes, we return to the Gillespie algorithm. We will, w.l.o.g., consider the following set of events:



The *total* rate at which an event takes place is simply given by the *sum of the rates*  $r_{\text{tot}} = \alpha x + \beta xy + \gamma y + \delta xy$ . If all four events would on average occur once a month, the total rate for *any* event would be four per month. This is what we use in the Gillespie algorithm: We use the total rate  $r_{\text{tot}}$  in order to sample the next event time.

Next, we only need to find out *which* of the four possible events will take place. This is straightforward: The probability for each event to happen is proportional to its rate. We can for example draw a uniformly distributed random number from  $[0,1]$ . Next, we subtract  $\frac{\alpha x}{r_{\text{tot}}}$ . If the result is less than zero, we pick the corresponding event, “birth of prey”. If not, we next subtract  $\frac{\beta xy}{r_{\text{tot}}}$ . If the result is less than zero, we pick the corresponding event, “consumption of a prey”, and so on.

We are now prepared for the full Gillespie scheme:

- 1. Initialization:** Initialize the number of predator and prey animals in the system as well as the rate constants.

2. **Monte Carlo step:** Generate a random number  $r$  and determine the time interval to the next reaction with  $t_{\text{next}} = \frac{1}{\lambda} \ln\left(\frac{1}{r}\right)$ . Next, determine which event will take place by the method described above.
3. **Update:** Increase system time by the randomly generated time  $t_{\text{next}}$  in step 2. Update the numbers of predator and prey animals based on the event that occurred.
4. **Iterate:** Go back to Step 2 unless there are no more predator and prey animals, or until the simulation time has been exceeded.



19. Download the MATLAB directory ssa from the server. Modify the file example.m as follows:

```
function lotka_volterra

tspan = [0, 200]; %seconds
x0 = [20, 200]; %foxes, rabbits
stoich_matrix = [ 0  1 ; %rabbit birth
                 0 -1 ; %rabbit eaten by fox
                 -1 0 ; %fox death
                 1  0 ]; %fox birth

% Rate constants
p.kRBirth = 1.1;
p.kFEatsR = 0.01;
p.kFDeath = 1.3;
p.kFBirth = 0.06;

% Run simulation
%[t,x] = directMethod(stoich_matrix, @propensities_2state, tspan, x0,
p);
[t,x] = firstReactionMethod(stoich_matrix, @propensities_2state,
tspan, x0, p);

% Plot time course
figure(gcf);
stairs(t,x);
set(gca, 'XLim', tspan);
xlim([0 30])
ylim([0 400])
xlabel('time (s)');
ylabel('numbers');
legend({'foxes', 'rabbits'});

end

function a = propensities_2state(x, p)
foxes = x(1);
rabbits = x(2);

a = [p.kRBirth*rabbits;
     p.kFEatsR*foxes*rabbits;
     p.kFDeath*foxes;
     p.kFEatsR*foxes*rabbits];
end
```



20. Run the system and compare the result to the deterministic Lotka-Volterra simulation.



21. Modify the rate constants in order to best fit the experimentally observed data of figure 1.



22. a. How can you explain that:

- i. the number of prey animals sometimes goes to zero?
  - ii. the number of prey animals sometimes keeps growing exponentially?
- b. Why can the number of prey animals not go to zero when the Lotka-Volterra differential equations are applied?

As we have seen, the Gillespie algorithm only deals with integer numbers of animals, in contrast to the Lotka-Volterra system of differential equations. In addition, we are able to simulate stochastic fluctuations in the populations of predator and prey animals by applying the Gillespie algorithm.

However, one important aspect that is neither taken into account by the Lotka-Volterra equations, nor by the Gillespie algorithms, is the effect of spatial variations in the population density of predator and prey animals. In reality, spatial fluctuations play a major role in the predator-prey system. You will learn more about this by running your own agent-based simulation of a predator-prey system.

## VI. Classification of simulation methods

Model	Dimensions	Discrete / continuous	Stochastic or deterministic
Differential equations	0D (no space)	- Continuous time - Continuous numbers of animals	Deterministic
Gillespie	0D (no space)	- Continuous time (but event-driven) - Discrete numbers of animals	Stochastic
Agent-based lattice model	2D lattice	- Discrete time - Discrete space	Stochastic

## VII. Further reading

- [Chapter 7.7](#)
- [Chapter 8.1, 8.2, 8.3, 8.5, 8.6, 8.8.](#)
- [Chapter 9.1, 9.2, 9.3](#)
- [Chapter 12.5](#)

## VIII. Some important exercises:

- 8.12-13      ([Student's t distribution](#))
- 9.2          ([confidence interval Student's t distribution](#))
- 12.8         ([confidence interval for a correlation coefficient](#))