

BIOINFORMATICS
Population Dynamics – Lab classes

I. Introduction (no lab class)
II. Exponential growth

Numerical exercises

1. Between 1700 and 1800, the human population grew steadily on a world-wide scale. It increased from approximately 600 millions to 910 millions of persons during that time. Estimate r .
2. The *United Nations Demographic Yearbook* estimates that the human population grew from 4490 millions to 5290 millions between 1980 and 1990.
 - a) Compare the value of r between this past decade and the one estimated for the XVIII century in Exercise 1.
 - b) Assuming that the human population growth remains at the same rate and knowing that there is about 1.49×10^{14} m² of dry land on the planet (including the Antarctic, central Greenland, Sahara, etc.), how long will it take before the number of humans exceeds the number of available square meters on earth ? (why not try to guess before calculating ?).
3. The differences between the instantaneous birth and death rates of human populations in four countries (estimated for 1991) are presented below (units are year⁻¹):

Ghana	Argentina	United Kingdom	Canada
0.031	0.012	0.002	0.006

How long will each of these countries take to duplicate their populations ? (what is the assumption underlying your calculations ?).

III. Regulated growth in continuous time: The logistic equation

Summary

Let us assume that a given population grows according to the logistic equation, but the observations N_1, N_2, N_3, \dots gathered by an ecologist, have errors relative to the baseline model, caused both by environmental variability and by measurement. It will be assumed that these errors are normally distributed around the expected value predicted by the model. The goal of this lab session is to simulate such error-prone observations computationally, using pre-specified values of r and K

Simulations with EXCEL

1. The logistic equation of continuous breeders in a spreadsheet.

- 1) Open a new spreadsheet and write down the symbols of the logistic equation, r, N_0, K , respectively, in cells A1:A3. In cells B1:B3 enter their numeric values, for example: 1.2, 5, 100.

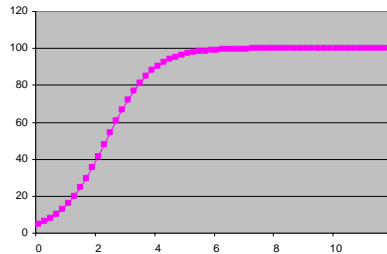
In A6 enter 0 and in A7 enter the formula =A6+0.2. Copy and paste this formula into A8:A65 thus getting a column with the time instants for which N_t will be computed (every time interval $\Delta t=0.2$).

2) In B6 enter the integral solution of the logistic equation $=(B\$3*B\$2)/(B\$2+(B\$3-B\$2)*EXP(-B\$1*A6))$, that is:

$$N_t = \frac{KN_0}{N_0 + (K - N_0)e^{-rt}}$$

The formula gets the parameters in cells B1:B3 and the time from column A. Copy and paste the formula into B7:B65.

3) Do the logistic graph using the cell arrays A6:A65 (abscissa) and B6:B65 (ordinate). To which value does the graph tend asymptotically? Change the value in B1 gradually to 0.5, 1.0, 1.5, 1.7, ... and follow the corresponding changes in the graph.



2. A stochastic logistic equation with a generator of $N(\mu, \sigma)$ prn's

Assume that despite a natural propension for the population to grow according to the logistic graph plotted, environmental variability and measurement errors, cause deviations in the observations of N_t , and these deviations are normally distributed, $N(\mu, \sigma)$ as follows:

- the mean μ , equals the value of the logistic equation at time t ,
- the standard deviation is proportional to μ . That is, $\sigma = c\mu$, where c is a constant.

The normality assumption is plausible, as there is usually many sources of error in field observations of N_t , and the relation $\sigma = c\mu$ is not uncommon in ecological data.

Before proceeding, here's a recipe to generate prn's with distribution $N(\mu, \sigma)$, taking advantage of the so-called Box-Muller transformation (this transformation maps a bivariate uniform distribution into a bivariate Normal).

Consider two prn's, $R_1 \in R_2$, drawn from a standard uniform distribution, for instance generated by function RAND() of Excel. It is possible to show that:

$$X = \mu + \sigma \sqrt{-2 \ln R_1} \cos 2\pi R_2$$

is a prn from a Normal distribution with mean μ and standard deviation σ , that is:

$$X \in N(\mu, \sigma)$$

We may now proceed with our spreadsheet exercise.

1) In cell D3 enter a numeric value for the constant c . For example 0.1.

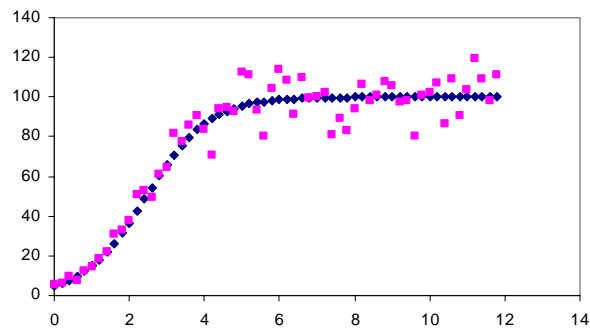
In D6, enter the formula

$$=B6+(B6*\$D\$3*\text{SQRT}(-2*\text{LN}(\text{RAND()}))*\text{COS}(2*\text{PI}()*\text{RAND()}))$$

which takes N_t (in B6) computed from the logistic, adds the standard deviation ($B6*\$D\3) multiplied by the square root factor and the cosen of the Box-Muller transformation. Also notice the uniform prn's generated by $\text{RAND}()$.

Copy and paste the formula into cells D7:D65. We get in column D a succession of values for N_t with an underlying logistic trend, but affect by normally distributed errors – we have a realization of a stochastic logistic curve.

2) Make a graph with cells A6:A6 in abscissas (time steps), B6:B65 (deterministic logistic) and D6:D65 (stochastic logistic) in ordinates. You should get something like this:



Change c gradually in D3 ($c=0.05, 0.1, 0.2, 0.3$) while keeping everything else constant. As c increases, the scatter of points of the stochastic logistic also increases (why ?).

Set $c=0.2$ and suppose you had gathered field observations of N_t , like the ones in the graph after $t=5$ (cells D36:D65) when the ascending part of the curve could not be seen. Would you be able to recognize the existence of a deterministic model underlying these observations ?

Place the cursor on an empty cell of the spreadsheet and click on the <delete> button, as you observe what is going on in the scatter of points. The points change as you press <delete>, because Excel generates new prn's given by function $\text{RAND}()$ and thus new error deviations form the deterministic logistic..

IV. Exploring demographic stochasticity with a spreadsheet

Literature:

Donovan, TM, and JE Goetz. 1999. Exploring demographic and environmental stochasticity with spreadsheets. *Bull. Ecological Society of America* 80(4):235-239.

Simulations with EXCEL

Demographic stochasticity (DS): simulation of a population of size $N=10$ individuals between t and $t+1$.

In a given population, the probability that an average individual survives during 1 year is $S=0.85 \text{ year}^{-1}$ and the average probability of having one child is $b= 0.21 \text{ year}^{-1}$ (it's assumed that it is not possible to have more than 1 child per year). The product bS is thus the per capita likelihood of having a child *and* that he/she survives 1 year. The finite rate of growth is $\lambda=S+bS$).

1) Open a new spreadsheet (herein it is assumed that you are using Excel, but other spreadsheets have equivalent functions) and enter the numerical value of S in cell B1 and the value of bS in D1. In G2 enter the formula $=10*(B1+D1)$ in order to get a deterministic estimate of N_{t+1} , assuming you start with $N_t=10$ individuals at time t . You have just computed $N_{t+1}=10\lambda$.

Can you trust this estimate given such a small initial value of N_t ? Can you understand why DS turns out to be relevant when projecting the future of small populations ?

2) In cells A5:A14 enter the digits 1 to 10, each representing one of the 10 individuals in the population at t . In cells B5:B14 enter the formula $=\text{RAND}()$. This formula is a pre-defined function of Excel that generates prn's (pseudo-random numbers) with a uniform distribution between 0 and 1. The prn's are re-generated each time the spreadsheet does calculations (you can see the prn's changing, for example by placing the cursor in an empty cell and clicking repeatedly in <delete>).

Do you remember what stating that the distribution is uniform between 0 and 1 means ?

Note: If you wanted to generate uniform prn's between a and b , you could use the formula $=\text{rand}()*(b-a)+a$

3) In cell C5 enter formula $=\text{IF}(B5<B\$1;1;0)$. This formula compares the value in B5 with S in B1. If $B5<S$, the individual survives and places 1 in cell C5; otherwise, the individual dies and a 0 is placed in cell C5. Copy and paste the formula into the array of cells C5:C14.

In cell D5 enter the formula $=\text{IF}(B5<D\$1;1;0)$. This formula compares the value in B5 with bS in D1. If $B5<bS$, the individual has a child, the child survives and places 1 in cell D5; otherwise, no child was born and/or did not survive and a 0 is placed in D5. Copy and paste the formula into the array D5:D14.

In cells C15 and D15 enter, respectively, formulas $=\text{SUM}(C5:C14)$ and $=\text{SUM}(D5:D14)$. These formulas sum, respectively, the survivals and the newborns at time $t+1$, projected from the initial 10 individuals.

The sum of survivals and newborns are the population in $t+1$. You have just projected the population, taking into account the random factor known as SD. Can you trust that this single estimate will be the population in $t+1$?

4) In cells A18:A67 input the digits 1 to 50, which will stand for 50 replicates of the population projection from t to $t+1$ (I'm making life easy for you, as normally far more replicates would be done).

Copy and paste the value (use Edit\Paste Special\Values !) of C15 and D15 into, respectively, B18 and C18 (this special way of copy-paste ignores formulas and copies only the numerical values. Repeat to cells B19 and C19, B20 and C20, B21 and C21, etc... Notice that every time you copy-paste, the values in the arrays C5:C15 and D5:D15 change as the functions $\text{RAND}()$ generate new prn's.

Note: this is a bothersome piece of work that could be avoided if you were to make a “macro”. See below how to do that.

In cell D18 enter the formula `=B18+C18`, which gives replicate 1 of the population at time $t+1$. Copy and paste into cells D18:D67, thus getting 50 replicates of the population in $t+1$.

Each replicate is an estimate of N_{t+1} . Are all the possible estimates of N_{t+1} equally likely ?

5) In cells E18:E36 input the digits 0, 1, 2 ... 18; 0 and 18 are, respectively, lower and upper limits for N_{t+1} , and were determined by visual inspection of the 50 replicates in D18:D67. You may set different limits, if you feel these are not appropriate for your replicates.

In cell F18 enter the formula `=COUNTIF(D$18:D$67;E18)`. This formula counts how many times the numerical value in cell E18 shows up in cells D18:D67. It thus determines the *absolute frequency* of replicates where $N_{t+1}=0$. Copy and paste the formula into cells F18:F36, thus getting the absolute frequency distribution of N_{t+1} .

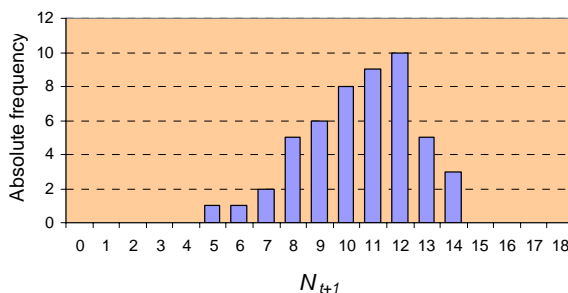
Enter in F39 the formula `=SUM(F18:F36)` simply to get the total number of replicates (50).

6) In G18 enter the formula `=F18/F$39`, which computes the *relative frequency* of replicates where $N_{t+1}=0$. Copy and paste into cells G18:G36, in order to get the relative frequency distribution of N_{t+1} . These are estimates of the probabilities of getting each N_{t+1} .

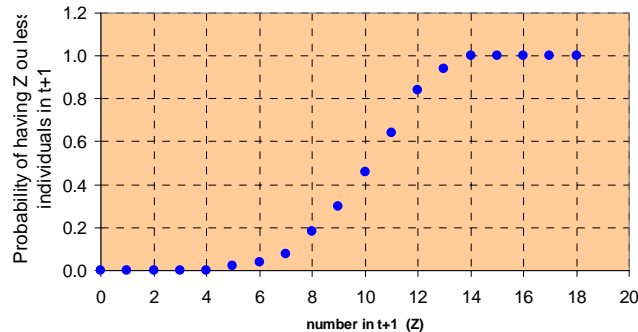
In cell H18 input `=G18` and in cell H19 input `=G19+H18`. Copy and paste cell H19 into cells H19:H36 thus getting the *cumulated relative frequencies* of N_{t+1} . For example, the value in cell H30 is the probability that the population in $t+1$ is *equal or less than* 12 individuals (notice that 12 is the value in E30).

Make sure you understand the probabilistic meaning of the cumulative distribution function just built. If you do, you'll have no trouble to understand what a risk curve is.

7) Build a column graph (or bar chart) using cells E18:E36 for abscissa and F18:F36 for ordinate. This graphs the absolute frequencies of N_{t+1} and should look more less like this,



Built a scatter graph, using cells E18:E36 (abscissa) and H18:H36 (ordinate). This S-shaped graph presents the cumulated probabilities of N_{t+1} , that is the probabilities of N_{t+1} being less or equal to values in abscissas. You have just made the risk curve for the population and it looks like this,



What does risk mean ? How likely are you to get only 8 or less individuals in t+1 ?

8) In cell B70 enter function `=AVERAGE(B18:B67)`, which computes the mean number of survivors in t+1 predicted by the 50 replicates.
 In cell B71 enter function `=STDEV(B18:B67)`, which computes the standard deviation (SD) of survivors around the mean.
 In cell B72 enter function `=B71/B70`, which computes the corresponding coefficient of variation (CV).

Copy cells B70:B72 and paste into C70:C72, getting the mean, SD and CV of the number of newborns in t+1
 Copy cells B70:B72 and paste into D70:D72, getting the mean, SD and CV of the number of individuals in t+1

Which value did you *expect* to get in B70 ? compare with what you got.
 Which value did you *expect* to get in D70 ? compare with what you got.

Suppose instead of an initial population of $N_t=10$ you had $N_t=100$. If everything else (S , b) remained unchanged you should have higher, equal, or lower values in:

- B70 and D70 ? (means)
- B71 and D71 ? (SD's)
- B72 and D72 ? (CV's)

MACRO for step 4)

“Macro” is Excel jargon to designate a sequence of operations that a spreadsheet memorizes and later carries out, whenever the user tells it to. A Macro is computer programme, written down in Visual Basic, which is created and compiled in the Excel environment. Unlike most computer programmes, however, the user does not actually have to write down the programme code or even see it. One just have to tell Excel to start recording a sequence of operations as these are being done (as oposed to being written). Once the sequence is finished, Excel is told to stop recording and to initiate the sequence whenever a certain key-pad is stroke. Next I show how to build a Macro that does step 4) above.

I assume you have just finished step 3) above.

1. In the main menu, select Tools\Macro\Record New Macro. In the dialog box give a name to the Macro (e.g.: population). Also select a key-pad that will launch the Macro in the future (e.g.: Ctrl t). The Macro may now start being recorded. If you make a mistake while recording, corrections to that mistake will also be recorded.
2. In the main menu select Edit\Go To. In the dialog box write the cells id where the cursor is to be positioned, B18:C18. Click OK and watch these cells being highlighted.
3. In the main menu select Insert\Cells. In the dialog box select “move cells down”.
4. In the main menu select Edit\Go To. In the dialog box, write cells C15:D15, where are the recently calculated numbers of survivors and surviving births.
5. In the main menu, select Edit\Copy. In the main menu, select Edit\Go To. In the dialog box enter: B18:C18.
6. In the main menu select Edit\Paste Special. In the dialog box, select “Paste values” (or just “Values” in some Excel versions) in order to paste the numerical values copied from C15:D15.
7. In order to stop recording, select Tools\Macro\Stop Recording in the main menu.

To run the Macro, press **Ctrl t**. If the recording was ok, each replicate (survivals and surviving births) is inserted into columns B and C of row 18 and, every time you press **Ctrl t**, a new replicate is inserted after having pushed the previous one down.