SCIE1000 Project, Semester 1 2011

Topic 2: Population Models and Genetics

1 Required background:

1.1 Science

To complete this project you will require some background information on three topics: geometric models of unconstrained growth, the logistic map and constrained growth, and Mendelian genetics.

Discrete growth models:

All of the growth models we cover in lectures are smooth, continuous functions, so the population size can be predicted at **any** time, including fractional time values. However, scientists are often primarily concerned with discrete "snap-shot" measurements of a population at certain time steps, such as at the end of each month. They can then use **discrete** models of the smooth functions. One advantage of this is that discrete models can substantially simplify the calculations while still giving sufficiently good results.

When modelling a phenomenon P at discrete steps, the value at step i is often denoted P_i where $i = 0, 1, 2, \ldots$. For example, P_3 may be a population at time 3 months (depending on the time step). In a discrete model it is often fairly easy to write a mathematical equation that defines the value at the **next** time step in terms of the value at the **current** time step. That is, the equation will show how to calculate P_{i+1} from P_i. Usually a specific value is known for the first value P_0; this is called the **initial value**.

As we saw in lectures, any quantity whose rate of change is proportional to its current value follows an exponential function. The discrete form of an exponential function is called a *geometric growth model*. Exponential functions and geometric growth models represent *unconstrained* growth, which is often realistic over short time periods. However, over longer times, growth is *constrained*. Later in lectures we will see the *logistic model* of constrained population growth; the discrete version of this is called the *logistic map*.

Mendelian genetics:

In the 19th Century, Gregor Mendel conducted an elegant series of experiments that revealed patterns governing the inheritance of physical characteristics in garden peas. He showed that some varieties always bred true to form. When he cross-pollinated plants with different characters, all the first-generation (F1) offspring only displayed one trait, the other apparently having disappeared. However, when he cross-pollinated the F1 hybrids, around one quarter of the second generation (F2) offspring displayed the missing trait. Mendel's observations led to the establishment of three rules of classical genetics:

- Physical traits are passed from parents to offspring by units of inheritance (genes);
- Offspring inherit two copies (alleles) of every gene, one from each parent; and
- Some genes are dominant, so are expressed even when the two alleles differ, and some are recessive, so are only expressed when the two alleles are the same.

If an individual has two identical alleles of a gene then it is called *homozygous* for that gene, and if the alleles differ then it is called *heterozygous*.

When considering population genetics, it should be apparent that genes and allele frequencies across the population can change as time passes, resulting in evolutionary change. However, the Hardy-Weinberg theorem states that the frequencies of alleles and genotypes in a population's gene pool remain constant from generation to generation whenever the following five conditions are met (which is uncommon in nature, except over short time periods). Such populations are then said to be in a state of *Hardy-Weinberg equilibrium*.

1. Random mating (which is disrupted when individuals select mates, especially related ones);

2. Large population size (as small populations show chance fluctuations, known as genetic drift);

3. No allele transfer between populations;

4. No mutations (gene insertions, deletions, substitutions); and

5. No natural selection (that is, no differential survival or reproductive success of individuals).

Knowing that gametes (sperm and egg cells) carry only one copy of each chromosome, and therefore only one allele for each gene, allows predictions to be made about mating outcomes and the patterns of inheritance. By coding dominant alleles with capital letters (for example, P) and recessive alleles with lower-case letters (for example, p), it is possible to construct a 2×2 matrix, called a *Punnett square*, to determine all possible combinations of alleles in offspring, and the relative frequencies of these combinations. For example, consider inheritance of a single gene in a population, with the frequency of allele P equal to $x = 0.8$ and that of allele p equal to $y = 0.2$. Then we can calculate the relative frequencies of all three possible genotypes:

- the probability of homozygous PP is $x^2 = 0.8 \times 0.8 = 0.64$.
- the probability of homozygous pp is $y^2 = 0.2 \times 0.2 = 0.04$.
- the probability of heterozygous Pp or pP is $2xy = 2(0.8 \times 0.2) = 0.32$.

This calculation can be completed in a Punnett square, as follows:

		Male parent	
		P $(x = 0.8)$	p $(y = 0.2)$
Female parent	P $(x = 0.8)$	PP $(x^2 = 0.64)$	Pp $(xy = 0.16)$
	p $(y = 0.2)$	pP $(yx = 0.16)$	pp $(y^2 = 0.04)$

(Note that heterozygous Pp is identical to pP; each occurs with a relative frequency of 0.16, so the combined relative frequency is 0.32. Note also that the sum of the probabilities of each possible genoptype equals 1.)

1.2 Mathematics

In a **geometric model** with growth factor equal to a constant r per time period, the value at the next step is $(1 + r)$ times the value at the current step. If $r > 0$ then the value increases as time passes, and if $r < 0$ then the value decreases as time passes. For $i = 0, 1, 2, \ldots$,

$$P_{i+1} = (1 + r)P_i.$$

You might notice that the formula for the geometric model is identical to the formula for compound interest. This makes sense: in each case, the quantity (population and bank balance) is growing by a fixed percentage of the total in each time period.

In the **logistic map**, the growth rate varies, reflecting the negative impact of competition as the population increases. The population can be modelled using the following expression in which r is called the *intrinsic growth rate* and k is a positive constant:

$$P_{i+1} = (1 + g)P_i \quad \text{where} \quad g = (r - kP_i)$$

1.3 Python

To round a number x to n decimal places, the Python command is `y=round(x,n)`. For example, `round(pi,2)` gives 3.14. To plot multiple graphs, use the `figure` command **before** each call to `plot()`. Also, **do not** include the `show()` command until you have drawn the final graph. Here is an example:

```
figure()
plot(x,y1)
figure()
plot(x,y2)
show()
```

2 Questions:

Complete the following questions, showing all working. Unless otherwise stated, each question is independent of the others (so growth rates and populations may change between questions). For simplicity, in all of these calculations and your Python programs, do not round your numbers. Keep all values, including the number of frogs, accurate to at least 4 decimal places.

(0) (0 marks) A number of the following questions vary between students. Let T be the number you get when you add 10000 to the number formed by the **last four digits of your student number**. For example, if your number were 42136712, then $T = 10000 + 6712$, which equals 16712. (It is important that you use the correct value of T for your student number; if you do not, you will lose a large number of marks. If you have any questions, please ask a tutor.)

1. Consider a certain species of Giant Desert Frog[1] that has a life expectancy of one month. In this question, assume that the value of the monthly growth rate r is constant. Biologists conduct population surveys of an isolated habitat, and estimate that the total population is T individuals, where T is the number you calculated in Question 0.

- **(a)** (1 mark) Four months after the first survey, in a follow-up survey, biologists estimate that the population size is now $T + 3000$. Find the value of the **monthly** growth rate r.
- **(b)** (1 mark) Create a table showing the predicted population size at **the end** of each month i from 0 to 12 inclusive; assume that the initial survey occurs at the end of Month 0.
- **(c)** (1 mark) Estimate the population size 24 months **before** the **first** survey.
- **(d)** (1 mark) Find the month in which the population will first equal $(T + 10000)$.
- **(e)** (1 mark) Assume that a population with the same growth rate r initially has T individuals, but that 1% of the population is harvested by the *Natural Medicine Company* (NMC) at the **end of each month**. What is the population size after 4 months? (Hint: assume that the harvest occurs **after** the population size has increased. That is, the frogs breed, then the harvest occurs, then the population is measured. You may assume that the population of T frogs at time 0 is **after** the harvest for that month.)
- **(f)** (1 mark) If NMC maintains its harvest rate, find the month in which the population will first equal $(T + 10000)$.

2. Biologists decide that it is more accurate to model the frog population using a logistic map.

- **(a)** (1 mark) Explain briefly what happens to the growth rate g in a logistic map as the population increases.
- **(b)** Explain the predicted effect on the population (over an extended time period) of:
  - **(i)** (1 mark) larger values of r versus smaller values (if all other factors remain constant)?
  - **(ii)** (1 mark) larger values of k versus smaller values (if all other factors remain constant)?
- **(c)** (1 mark) Biologists propose that for the frog population in Question 1, $r = 0.05$ and $k = 10^{-6}$. Create a table showing the population size P_i at **the end** of each month i from 0 to 12 inclusive. (Assume that no frogs are harvested.)
- **(d)** (1 mark) Repeat Part (c) but now also assume that 1% of the total population is harvested at the end of each month.
- **(e)** (1 mark) On a single set of axes, draw graphs of the populations you calculated in Parts (c) and (d), clearly identifying each population. (You may draw the graph by hand, or using a computer.)

[1]For more information on the Giant Desert Frog, ask Dr Robbie Wilson, who lectures into BIOL1030.

(f) (1 mark) In about 50 words, comment on the impact on the population of the harvesting strategy in Part (d), compared with no harvesting in Part (c).

3. **(a)** (0 marks) Write a Python program that uses a logistic map with $r = 0.05$ and $k = 10^{-6}$ (see Question 2) to model a frog population during a given time period. Your program must:

- Prompt the user to enter the following:
  - the initial population;
  - the proportion of the frog population harvested at the end of each month (this proportion may be 0); and
  - the number of months over which the model should run.
- Apply the logistic map to the population for the specified number of months.
- Print the population at the end of each month, and the number of frogs harvested that month.
- On completion, print the final frog population and the total number of frogs harvested.
- Have variables with meaningful names, and be appropriately commented.

Hint(s):

- You may assume that all input values are valid (for example, the initial population and harvest rate will not be negative).

(b) (4 marks) Print and submit a copy of your program. Marks will be awarded for:

* Adherence to the program specifications given above
* Appropriate programming style, structure and logic
* Appropriate print statements, with helpful text explanations of the output
* Use of comments and meaningful variable names

(c) (2 marks) Test your program in the following ways. (You must include a printed copy of the output from your program in your submission; in each case, verify the output against your hand calculations. You may need to add some print commands to your program to do this testing.)

(i) Use your program to repeat Question 2(c).

(ii) Use your program to repeat Question 2(d).

(d) (1 mark) Use your program to model a frog population with initial population size of 10000 and with 1% of the population harvested per month, for 100 months. You do not need to submit a copy of the output; instead, answer the following questions:

(i) What is the final frog population?

(ii) How many frogs are harvested in total?

(iii) What is the population size at the end of Month 50?

(iv) During which month does the harvest size first exceed 250?

All questions above this line must be completed as part of your initial project submission.

Once again, consider populations of Giant Desert Frogs. From now on, assume that the colour of individual frogs is determined by a single **recessive** colour gene, with homozygous cc individuals gold, and other individuals brown. (Thus, all frogs with genotype CC, Cc and cC are brown.)

4. (1 mark) In a certain population, 30% of the individuals are gold and 50% of individuals are heterozygous. Calculate the relative frequency of allele C in this population.

5. All parts of this question refer to a frog population that satisfies the conditions for Hardy-Weinberg equilibrium, and in which the relative frequency of allele C is 0.4. The initial population size is T (where T is the number you calculated in Question 0), and the population is growing according to the logistic map given in Question 2.

(a) (1 mark) Draw a Punnett square for the colour gene, and calculate the relative frequency of each genotype. What is the expected proportion of gold frogs in the population?

(b) (2 marks) Find the number of gold frogs in the population at the ends of each of Months 1, 2 and 3.

(c) (0 marks) Modify your Python program from Question 3 so that it now also includes the genotypes/phenotypes of frogs in the population. In addition to all of the previous input/output, your program must now also:

- Prompt the user to enter the initial relative frequency of allele C in the population.
- Each month, print the number of frogs in the population with each genotype (homozygous CC, heterozygous, and homozygous cc).
- On completion, draw (separate) graphs showing the total frog population at the end of each month, and the number of gold frogs harvested each month. (Section 1.3 shows how to plot multiple graphs in a program.)

(d) (4 marks) Print and submit a copy of your program. Marks will be awarded for:

* Adherence to the program specifications given above
* Appropriate programming style, structure and logic
* Appropriate print statements, with helpful text explanations of the output
* Use of comments and meaningful variable names

(e) (1 mark) Test your program in the following way. (You must include a printed copy of the output from your program in your submission; verify the output against your hand calculations. You do not need to submit a copy of the graphs.)

(i) Use your program to repeat Question 5(b).

(f) (2 marks) Use your program to model a frog population with initial population size of 10000, relative frequency of allele C equal to 0.4, and with 1% of the population harvested per month, for 100 months. You do not need to submit a copy of the output; instead, answer the following questions:

(i) What is the final population of frogs of each genotype?

(ii) How many frogs are harvested in total?

(iii) What is the number of gold frogs in the population at the end of Month 50?

6. All parts of this question refer to a frog population that satisfies the conditions for Hardy-Weinberg equilibrium, and in which the relative frequency of allele C is 0.4.

(a) Assume that gold frogs contain a special protein that makes them valuable to NMC but that brown frogs do not. As a strategy to increase the commercial value of the population, NMC proposes catching and killing a number of brown frogs each month. If this strategy is implemented:

(i) (1 mark) Explain (in words) what will happen to the relative frequencies of the alleles for the colour gene as time passes. What will happen to the proportion of gold frogs in the population after a large number of generations?

(ii) (1 mark) Does the population remain in Hardy-Weinberg equilibrium? If not, which of the five conditions for Hardy-Weinberg equilibrium is/are no longer met?

(b) (1 mark) If a phenotype-specific disease suddenly kills **all** of the **brown** frogs in the population, find the relative frequency of each allele of the colour gene as soon as they die, and the relative frequency of each genotype in the next generation. What proportion of individuals in the next generation will be gold?

(c) (2 marks) Repeat Part (b), but instead assume that the disease kills all of the **gold** frogs.

7. The relative frequency of allele C in a certain frog population that satisfies the conditions for Hardy-Weinberg equilibrium is 0.5. The initial population size is T (where T is the number you calculated in Question 0), and the population is growing according to the logistic map given in Question 2.

(a) (3 marks) Consider the following scenario:

> A company (NMC) catches 1% of the total frog population at the end of each month, removes any gold frogs they catch from the population and sells them, and returns all brown frogs to the population. Frogs are caught completely at random, so among the captured frogs, the relative frequencies of the genotypes equals the relative frequencies across the entire population.

Estimate the number of gold frogs sold in Month 1, and the total frog population at the end of Month 1, directly after NMC conducts its harvest for that month.

(**Hint:** the number of frogs *caught* in this and subsequent questions is **not** the same as the number of frogs *sold*. The number of frogs sold is the expected number of gold frogs amongst those that are caught.)

(b) (2 marks) Find the relative frequency of each allele of the colour gene and of each genotype at the end of Month 1 (after the harvest), and find the number of gold frogs remaining in the population.

(c) (4 marks) If NMC also catches 1% of the frog population at the end of Month 2:

(i) How many of the frogs caught in Month 2 are gold?

(ii) At the end of Month 2 (after the harvest), what is the total frog population?

(iii) Find the relative frequencies of the colour alleles in the population.

(iv) Find the relative genotype frequencies in the population.

(v) How many gold frogs are in the final population?

8. **(a)** (0 marks) Modify your Python program from Question 5 so that rather than removing frogs of both phenotypes from the population, now only gold frogs are removed. In addition to the previous input/output, your program must now also:

- Prompt the user to enter the **proportion of the total population caught** (at random) by NMC each month.
- Allow NMC to remove any caught gold frogs from the population, with all caught brown frogs returned to the population.
- On completion:
  - · plot a graph of the number of gold frogs harvested each month (with a meaningful title and labels on the axes);
  - · print the total number of frogs in the final population, the number of gold frogs in the final population and the total number of gold frogs harvested.
  - · if the final frog population is less than the initial population, print a warning message saying this.
- Have variables with meaningful names, and be appropriately commented.

(Hint: this question is not easy. It is probably easiest to maintain a count of the number of frogs with each genotype, and update these counts each month. Because only gold frogs are harvested, this will change the relative frequencies of the alleles of the colour gene. This will change the relative frequencies of the genoptypes in the population **next month**.)

(b) (14 marks) Print and submit a copy of your program. Marks will be awarded for:

∗ Adherence to the program specifications given above

- $*$ Appropriate programming style, structure and logic
- $*$ Appropriate print statements, with helpful text explanations of the output
- $*$ Use of comments and meaningful variable names

(c) Test your program in the following ways. (You must include a printed copy of the output from your program in your submission; in each case, verify the output against your hand calculations. You may need to add some print commands to your program to do this testing.)

(i) (1 mark) Use your program to repeat Question 7(a).

(ii) (1 mark) Use your program to repeat Question 7(b).

(iii) (2 marks) Use your program to repeat Question 7(c).

Before answering the rest of the questions on this project, comment out the line(s) of your program that print information about the population each month. All answers only require the graphs, and information about the final population.

(d) (5 marks) Write a brief user guide which explains how to use the program. (Your user guide should **not** assume that the user has read this assignment question sheet.) The guide should contain all necessary information about:

- What the program does.
- What input is requested by the program, and what valid input it can take.
- What output the program gives.
- Any assumptions you have made, or any special cases.

This is a user guide, not a programmer's manual. Do not describe the algorithm you have used or internal details of the program. Instead, if someone with a basic understanding of computers (and a good understanding of the science relevant to your project topic) wanted to run your program, what would they need to know?

(e) (4 marks) Use your program to model a frog population with initial population size of 10000, relative frequency of allele C equal to 0.4, with 1% of the population harvested per month and all harvested gold frogs removed, for 100 months. You must submit a copy of the output.

9. Use your program to model a frog population with initial population 10000, relative frequency of allele C equal to 0.3 and zero frogs caught per month, for a period of 100 months.

(a) (1 mark) Find the final frog population, the number of gold frogs in that population, and the total number of gold frogs harvested.

(b) (2 marks) You should have found that as a proportion of the total frog population, the number of gold frogs in the final population is equal to 0.3^2. Verify that this is true, and explain how this relates to Hardy-Weinberg equilibrium.

10. (25 marks) When answering this question, if relevant you may include some graphs, diagrams, tables of values, equations or mathematical calculations (which will not be included in the word limit), but your response should be predominantly text-based. The report must be typed, written in a professional style, and be within 10% of 2000 words in length. Marks will be deducted for any report with length outside this range.

It is **not** appropriate to use other sources or references in your report; all of the discussion and recommendations should arise solely from your own work, including output from your program. The course Blackboard site contains a Criteria Sheet for this report, showing how marks will be allocated. **(The answer to this question must be submitted in hardcopy and also via Turnitin; see the project overview document.)**

NMC employs you to investigate potential harvesting strategies for a particular frog population. They provide you with the following information:

– The initial population size is T (you calculated this value in Question 0).
– The relative frequency of allele C is 0.3.
– There are three suggested harvesting rates: 0.5% of the total population per month, or 3%, or 10%.
– Their aim is to maximise the number of gold frogs harvested in the next 100 months, although they also want their business to be sustainable after this time. They are willing for you to propose a different harvesting rate to their three suggested ones, if you can demonstrate the advantages of your proposed rate.

Write a report for NMC. The report must meet the following criteria:

– It must be understandable by NMC's management, who have some knowledge of science, ecology and biology, but are not experts.
– It should **not** describe any of the mathematics or programming you used; instead, NMC wants you to provide them with a summary of the results obtained from different harvesting strategies, a discussion of some options, and recommendations about what they should do.
– It must include four sections:
  1. an "Executive Summary" of around 150 words, giving a brief overview of your findings, and listing the recommendations. The rest of the document will expand on the content of the executive summary.
  2. a "Results" section, which **does not** include any discussion, but **does** include:
     · the results and graphs relating to their three proposed harvesting strategies (no discussion should be included here)
     · the optimal harvesting strategy (which you need to identify), which may well be different to their three proposed rates. You must include graphs and results relating to this harvesting rate.
  3. a "Discussion" section, which includes a discussion and analysis of the results, and a description of how you identified the optimal harvesting rate.
  4. a "Conclusions/recommendations" section containing one or more recommendations about what strategy they should follow, with a brief justification of your recommendation(s), and any risks or other relevant information. Your answer should include an explanation of why higher harvest rates are less effective than your identified optimal rate.

The end