M. Phil. in COMPUTATIONAL BIOLOGY

Friday, 11 May, 2018   2:00 pm to 4:00 pm

COMPUTATIONAL BIOLOGY

*Attempt **ALL** questions.*

*There are **THREE** questions in total.*

The questions carry equal weight.

STATIONERY REQUIREMENTS
Cover sheet
Treasury Tag
Script paper

SPECIAL REQUIREMENTS
Calculator - students are permitted
to bring an approved calculator.

**You may not start to read the questions
printed on the subsequent pages until
instructed to do so by the Invigilator.**

1 Population genetic analyses of genomic data

Complete both parts A and B of the question.

A) In a single-locus, diploid system, the fitnesses of the homozygous and heterozygous genotypes are given by:

$$w_{aa} = 1$$
$$w_{Aa} = 1 + hs \tag{1}$$
$$w_{AA} = 1 + s$$

a) What name is commonly given to the parameter h in the above equation?

b) What value of h would specify an 'additive' model of fitness?

c) Suppose that mutation creates a single variant A in a population that is otherwise fixed for the aa genotype. We here consider the probability that the variant goes on to reach a frequency of at least 1% in the population before dying out. Compared to the additive case, how is this probability affected under each of the following cases?
i) h=1
ii) h=0
iii) h=1.4
iv) h=-0.2
Explain your answers in each case.

d) We now consider the probability that the variant goes on to fix in the population. Compared to the additive case, how is this probability affected under each of the following cases?
i) h=1.4
ii) h=-0.2
Explain your answers in each case.

B) Equation 2 describes the evolution of the allele frequency of a single-locus two-allele system (alleles labelled 0 and 1), the term q_i^1 denoting the fraction of allele 1 at locus i in the genome. The noise term $\chi_t(q_i^1)$ has mean equal to zero, variance equal to $q_i^1(1 - q_i^1)/N$, and is uncorrelated in time.

$$\frac{dq_i^1}{dt} = \sigma q_i^1(1 - q_i^1) + \mu(1 - 2q_i^1) + \chi_t(q_i^1) \tag{2}$$

a) What does μ denote in this equation? What does N denote in the variance of $\chi_t(q_i^1)$?

b) Suppose that the variant allele is deleterious, such that $\sigma < 0$, and that the population is large, such that the noise term can be neglected. What frequency does the allele reach at equilibrium?

Hint: If $ax^2 + bx + c = 0$ then

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \tag{3}$$

You may suppose that μ is small.

2   Scientific Programming

1. Study the following function **s**, and the subsequent calls to **s**. What is printed out by **s** during the execution of the four final expressions, and what are the instantiated values of **a**, **b**, **c** and **d**?   [50%]

   State briefly what the function is doing.

   ```
   s = function(L,k) {
     ## L is a vector of numbers.  k is a positive integer.

     print(c(k, L)) ## show what is printed each time here
     if (length(L) < 2) {
       return (L[k])
     }
     p = 1; x = L[p]
     p1 = p2 = p3 = c()
     for (y in L) {
       if (y < x)  p1 = c(p1, y)
       if (y == x) p2 = c(p2, y)
       if (y > x)  p3 = c(p3, y)
     }
     if ( k <= length(p1) ) {
       return ( s(p1, k) )
     } else {
       z = length(p1) + length(p2)
       if ( k > z ) {
         return ( s(p3, k - z) )
       } else {
         return ( x )
       }
     }
   }


   ## Evaluate the following expressions:
   L = c(6, 4, 19, 2, 17, 18, 2)
   a = s(L, 1)
   b = s(L, 4)
   c = s(L, 7)
   d = s(L, 8)
   ```

2. Study the following three functions: f, digits, k. Determine what each function is doing by evaluating the expressions for x, y and z after the function definitions. In each case, show what is printed out during the evaluation. [50%]

```
f = function(v) {
  ## v is a vector of integers
  c = 1; sum = 0
  for (i in v) {
    sum = sum + (i*c)
    c = c*10
  }
  sum
}


digits = function(n) {
  ## Assume n is a positive integer less than 10,000
  a = rep(0,4)
  for (i in 1:length(a) ) {
    a[i] = n %% 10                   # remainder
    n = n %/% 10                     # division, ignore remainder.
  }
  a
}


k = function(n, debug=TRUE) {
  loop = TRUE; steps = 0
  while (loop) {
    d = digits(n)
    v1 = sort(d, decreasing=TRUE)      # largest number first
    v2 = rev(v1)                       # reverse the list
    n1 = f(v2) - f(v1)
    cat(paste(n, f(v2), f(v1), n1, "\n"))
    loop = (n != n1) && (n1 != 0)      # a && b means a AND b
    n = n1
    steps = steps+1
  }
  c(n1, steps)
}


### Evaluate these expressions.  Show what is printed in each case.
x = k(9024)
y = k(1987)
z = k(5555)
```

3   Focus area within Genome Informatics: Gene regulation

In the era of cheap sequencing, discuss genome-wide approaches to understanding gene regulation in eukaryotes.

END OF PAPER