COMSM0302 - WEEK 2 EXERCISES

JAMES MARSHALL
OCTOBER 25, 2006

1. Exercise 1 - Stochastic Universal Sampling

Extend your implementation of the Simple Genetic Algorithm from week 1 to use Stochastic Universal Sampling.

2. Exercise 2 - Approximation of minimum population size for binary chromosomes

Given that for a population of N binary chromosomes of length ℓ, the probability that the population contains at least one of each allele at each locus is

$$P_2^* = \left(1 - \left(\frac{1}{2}\right)^{N-1}\right)^{\ell},$$

prove that for arbitrarily large $P_2^* < 1$ the minimum population size to achieve that probability is given by

$$N \approx \left\lceil 1 + \log_2\left(\frac{-\ell}{\ln P_2^*}\right)\right\rceil.$$

Hint: $e^x = \lim_{y\to\infty}\left(1 + \frac{x}{y}\right)^y$, i.e. as y increases, e^x is an increasingly good approximation of the right hand side of the equation just given.

3. Exercise 3 - Mean selection probability under linear ranking

Prove that the probability of selecting the median ranked individual under linear ranking is the same as the mean selection probability taken over all ranked individuals.

Hint: There is a simple closed-form solution for the sum of an arithmetic sequence. You can find this in the lecture notes.

4. Exercise 4 - Equivalence of soft tournament selection and linear ranked selection

Given that the selection probability for the ith individual under soft tournament selection with 'best-wins' probability p and tournament size $\tau = 2$ is

$$P(i) = p\frac{2(i-1)}{N(N-1)} + (1-p)\frac{2(N-i)}{N(N-1)},$$

derive expressions for α and β in terms of p and N that make this selection operator generate equivalent selection probabilities to linear ranked selection. Calculate the selection intensity ϕ given by both of these expressions. Verify your value of ϕ by calculating it directly from the expression for $P(i)$ above for the fittest string (ranked N) and the median string (ranked $\frac{N+1}{2}$).

5. Exercise 5 - Derivation of the Hardy-Weinberg equilibrium distribution

Given a population of homozygotes AA, heterozygotes Aa and aA, and homozygotes aa, in the frequency distribution $p : 2q : r$, derive the distribution after one generation of random mating without selection. Assume any individual can mate with any other, as in a GA.

Hint: Work out all the different ways in which an individual of each type may be produced by two different parents, and calculate the frequency with which this occurs.