

COMSM0302 - WEEK 2 SOLUTIONS

JAMES MARSHALL
OCTOBER 30, 2007

1. EXERCISE 2 - APPROXIMATION OF MINIMUM POPULATION SIZE FOR BINARY CHROMOSOMES

We can rewrite P_2^* as

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

Recall that $e^x = \lim_{y \rightarrow \infty} \left(1 + \frac{x}{y}\right)^y$, so we want to rewrite P_2^* in a similar form.

Given that P_2^* is written as the sum of 1 and a fraction $\frac{-1}{2^{N-1}}$, all raised to the power ℓ , we simply need to rewrite the fraction so that its denominator is ℓ . As the value of a fraction is unchanged by multiplying its numerator and denominator by the same constant, we can rewrite our fraction as $\frac{-\ell/2^{N-1}}{\ell}$. This gives us the exponential approximation

$$P_2^* \approx e^{-\ell/2^{N-1}}.$$

Now taking the natural logarithm of both sides we have

$$\ln P_2^* \approx \frac{-\ell}{2^{N-1}},$$

which can be rearranged to give

$$2^{N-1} \approx \frac{-\ell}{\ln P_2^*}.$$

To find an expression for N we again take the logarithm of both sides, this time to the base 2, rearrange, and take the ceiling of the right-hand side as we are interested in the smallest *integer* population size that will give us our desired probability P_2^* or better, hence we arrive at

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

In fact this approximation is unnecessary. An exact solution can be found as

$$N = \lceil 1 - \log_2(1 - P_2^{*1/\ell}) \rceil.$$

It is unclear why Reeves and Rowe used the approximation.

2. EXERCISE 3 - MEAN SELECTION PROBABILITY UNDER LINEAR RANKING

The median individual is that with rank $\frac{N+1}{2}$, hence the probability of selecting the median individual is

$$\alpha + \beta \frac{N+1}{2}.$$

The closed-form solution for the sum of an arithmetic sequence of form

$$\sum_{k=1}^{k=N} (\alpha + \beta k)$$

is

$$N \left(\alpha + \beta \frac{N+1}{2} \right),$$

hence the average of the sequence is that expression divided through by N , which is

$$\alpha + \beta \frac{N+1}{2}.$$

3. EXERCISE 4 - EQUIVALENCE OF SOFT TOURNAMENT SELECTION AND LINEAR RANKED SELECTION

For linear ranking, $P(i) = \alpha + \beta i$.

For soft tournament selection with $\tau = 2$,

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

Thus to find expressions for α and β that make the two selection schemes equivalent, we need to write the above expression for soft tournament selection as the sum of two expressions, one of which has a coefficient i , thus

$$P(i) = \frac{2N - 2p(N+1) + (4p-2)i}{N(N-1)}.$$

This gives us

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

$$\beta = \frac{2(2p-1)}{N(N-1)},$$

and hence $\phi = 2p$.

4. EXERCISE 5 - DERIVATION OF THE HARDY-WEINBERG EQUILIBRIUM DISTRIBUTION

The frequencies for the AA, Aa, aA, and aa types are p , q , q and r respectively. Consider the frequencies of these types in the next generation, produced by random mating in the absence of selection, and assuming any individual can mate with any other.

In the next generation the homozygous AA type can be produced by the crosses (in which the first allele comes from the first parent, and the second allele from the second parent) AA \times AA, Aa \times AA, AA \times aA, and Aa \times aA. Hence the total frequency of the AA type in the next generation will be $p^2 + pq + pq + q^2 = (p+q)^2$. By the same reasoning the frequency of the aa type will be $(q+r)^2$.

In the next generation the heterozygote type Aa or aA type will be generated by the crosses AA \times Aa, aA \times AA, Aa \times Aa, aA \times aA, aa \times aA, Aa \times aa, AA \times aa and aa \times AA. Hence the total frequency of heterozygotes in the next generation will be $pq + qp + q^2 + q^2 + rq + qr + pr + rp = 2(p+q)(q+r)$.

Thus the distribution in the next generation will be

$$(p+q)^2 : 2(p+q)(q+r) : (q+r)^2.$$