Dirk Metzler WS 2017/2018

COMPUTATIONAL POPULATION GENETICS — EXERCISE SHEET 1

---

1. Simulate the following with paper and pencil and a dice (or the R command `sample(6,1)`) as source of random.
   (a) A haploid Wright-Fisher population of 6 haploid individuals
   (b) Three individuals are sampled from the present generation. Identify their most recent common ancestors.
2. A number $k \in \{2, 3, 4, 5, 10, 20, 50, 100, 200, 1000\}$ of haploid individuals are sampled from a large Hardy-Weinberg population. Calculate for their genealogy…
   (a) …the expected time τ to their most recent common ancestor and …
   (b) …the expected total length ℓ of all branches and …
   (c) …the expected number of mutations on these trees for $\theta = 10$.
   (d) Visualize how τ and ℓ depend on k.
   (e) Find a function that can easily be expressed without sum signs and approximates how ℓ depends on k.
3. Use the R package scrm to simulate trees under three different scenarios:
   (a) The standard case of a panmictic population of constant size
   (b) An exponentially growing population
   (c) A population that consists of two subpopulations with migration.

   Visualize the genealogies. For which growth rates and migration rates do the trees begin to look different than the standard coalescent?
4. FOR BIOINFORMATICIANS AND OTHER PROGRAMMING ENTHUSIASTS: Develop a program that reads a sample size and outputs a random standard coalescent for this sample size in Newick format.
5. Assume a haploid Wright-Fisher population of N individuals. Thus, we assume discrete generations and each individual in generation 0 selects purely randomly and independent of all other individuals one parent in generation -1.
   (a) Let a be an individual in generation -1 and let X be the number of his or her kids in generation 0. Compute $\Pr(X = 2)$.
   (b) Compute $\Pr(X = k)$ for general $k \in \mathbb{N}$.
   (c) Which fraction the individuals of a generation do not have offspring?
   (d) Assume that the population size N is very large. Then, $\Pr(X = k)$ can be approximated by a term of the form
   $$b^c \cdot e^d / f!.$$
   Which are the appropriate values for b, c, d and f?
   (e) Make a rough calculation: Which fraction of the individuals of a generation will have offspring

i. after 10 generations?
ii. after 100 generations?
iii. after $10 \cdot N$ generations?

(f) in a generation $-g$, that is g generations before the present generation 0, let 0.2 be the fraction of individuals who still have offspring in the present population 0. Which fraction of individuals from generation $-(g+1)$ do still have offspring in generation 0?

(g) Does your answer to question 5f shed new light on part 5e?