Dirk Metzler WS 2013/2014

COMPUTATIONAL POPULATION GENETICS — EXERCISE SHEET 3

1. Use software like Tracer to evaluate the results of the preliminary LAMARC runs from exercises 2 and 3 of exercise sheet 2 and decide which MCMC options may be appropriate for the full analysis. Start LAMARC runs with these parameters.

2. Simulate datasets for a population that consists of four sub-populations, with gene flow between the populations.

   (a) How accurately can you estimate the rates and favored directions of migration between each pair of sub-populations and how does this depend on the number of loci and other properties of the dataset?

   (b) Assume that each of the sub-populations has a substructure. How does this influence the results, especially if the the sampling was limited to some of the sub-subpopulations. What if the sub-substructure is ignored or unknown?

   Design the study and start the first preliminary test runs whose results we can discuss next week to decide about the program option settings for the rest of the study.

3. Assume that two random variables A and B with finite state spaces S_A and S_B and a joint distribution $\Pr(A = a, B = b)$ are given. Further assume that for each $a \in S_A$ there is a Markov chain on S_B with transition law $P^{(a)}_{b\to b'}$ with equilibrium distribution $\Pr(B = b|A = a)$, and for all $b \in S_B$ there is a Markov chain on S_A with transition law $P^{(b)}_{a\to a'}$ with equilibrium distribution $\Pr(A = a|B = b)$. Define a Markov chain transition law as follows: Given the current state (a, b), toss a fair coin, and if it shows "head" go to (a, b') with probability $P^{(a)}_{b\to b'}$. If it shows "tail", got to (a', b) with probability $P^{(a)}_{b\to b'}$. Show that the resulting Markov chain has equilibrium distribution $\Pr(A = a, B = b)$.

4. Explore with simulated datasets how one can increase the efficiency of MCMC Methods like LAMARC and IM/IMa by using MCMCMC and fine-tuning the heating parameters.