

There are several overview about the coalescence theory. I like these two the most:

- Felsenstein, J. , M. K. Kuhner, J. Yamato, and P. Beerli. 1999. Likelihoods on coalescents: a Monte Carlo sampling approach to inferring parameters from population samples of molecular data. **Lecture Notes - Monograph Series, Institute of Mathematical Statistics** 33: 163-185.
- Rosenberg NA, Nordborg M. 2002. Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. **Nature Review Genetics** 3(5):380-90.

1 Coalescent

1. When we record all relationships among individuals in a Wright-Fisher population we know exactly who is the ancestor of the current individuals in every generation. Sewall Wright showed that when we pick two individual chromosomes at random, the chance that they have the same ancestor in the last generation is $1/(2N_e)$, where N_e is the effective population size of diploid organisms. In large population this chance is small, but one can calculate how long it takes on average until the two individuals have a common ancestor because the distribution of the waiting time follows a geometric distribution. This expected waiting time is $2N_e$ generations.
2. JFC Kingman realized that when we look backwards in time and changes the discrete time scale to a continuous time scale (using exponential distributions instead of geometric distributions) one can find the probability that out of a sample of size k two random individuals coalesce (have the same ancestor) with probability $k(k-1)/(4N_e)$. The expected waiting time until coalescence is $4N_e/(k(k-1))$ for diploid organisms. The probability allows to calculate the probability $P(G|N_e)$ of a specific genealogy G , one where we know the topology and the branch length. The coalescent has distribution with a very large variance, so that different realization using the same parameter can have very different topologies and branch length.
3. The basic coalescent can be "easily" extended to allow for multiple populations, or other population genetic forces.

2 Maximum likelihood

1. We can construct a maximum likelihood estimator (ML) to infer the population genetic parameters. Unfortunately, we only have genetic data from the most recent generation and we do not know the genealogy. The population parameters population size N_e and migration rate m are confounded with the mutation rate μ and we are not able to disentangle them with only one sample in time, therefore we estimate the compound parameters $\Theta = 4N_e\mu$ and $M = m/\mu$. Because we do not know the genealogy we try to summarize over all of them weighted by their probability.

3 Metropolis-Hastings algorithm

1. The calculation of the MLE would be simple if we could calculate the probabilities for all genealogies, but this is only possible for very small data sets (3-4 individuals). With a small

data sets with 10 individuals we would need to sum over the 2.5×10^9 topologies **and** for each of them we would also need to calculate a 9-dimensional integral (for all possible branch lengths), a task that is very difficult. Typical data sets contain 100 or more individuals.

2. We can use a Markov chain Monte Carlo method (MCMC) to approximate the MLE. We use a technique that was developed by Nicolas Metropolis (1951).
3. Metropolis' recipe:
 - (a) first state
 - (b) perturb old state and calculate probability of new state
 - (c) test if new state is better than old state: accept if ratio of new and old is larger than a random number between 0 and 1
 - (d) move to new state if accepted otherwise stay at old state
 - (e) go to (a).

4 Metropolis coupled Markov chain Monte Carlo

A single chain is often not very successful to explore all possibilities and can miss important areas of the search space because the walk to these regions might lead through regions that have very low probabilities. Charles Geyer and Elisabeth Thompson came up with a scheme that allows a better exploration: use multiple chains that exchange information about their location and probability with each other, and that have different capabilities to move around the space. They coined the term heated chains because the hot chains accept more often and therefore walk around faster than the chain of interest (the cold chain). The analogy comes from physics where gas molecules bounce around much faster when hot than when cold. To achieve this in the MCMC framework one is powering the acceptance-rejection formula: the cold chains accepts with $r^{1/1}$, a twice as hot chain with $r^{1/2}$, a very very hot chain with $r^{1/1000}$. if r is bigger than one each chain will always accept, but if it's smaller than 1 the hot chain will accept more than the cold chain (for example: $r=0.25$, the cold chain will accept with probability 0.25, the twice as hot chain will accept with 0.5 and the very hot chain will accept with 0.999). Every step a pair of chains compares their probabilities and uses an acceptance rejection scheme, when accepted they swap their place. This allows the cold chain to reach regions that are difficult to reach for it alone.

5 Extensions of the coalescent

- Migration rate estimation
- Population growth
- Recombination rate
- many others