Abstract

Wiuf and Posada (2003) introduced a coalescent model considering recombination hotspots. Their model is an extension of Hudson’s coalescent model (1983), the coalescent with uniform recombination rate. They considered heterogeneity in recombination rate along the chromosome. They chose the center of recombination hotspots according to some point process and assumed that recombination happens based on a descending rate from the centers. They used a two-step procedure: using a point process to find the center of the recombination hotspot and then a chosen distribution for the recombination events happening around this hotspot. Here we propose a new model using Hawkes processes which improves on this; our model locates the recombination events and the hotspots in one step.

Hudson’s Coalescent Model

- The algorithm starts with \( k = n \) sequences and looks back in time.
- Time to the next event \( = E(\frac{k(k-1)}{2}) \) with \( \rho = 4Nc \).
- A coalescence with probability \( 1 - \frac{1}{k(k-1)} \).
- A recombination with probability \( \frac{1}{k(k-1)} \).
- If a recombination event it will be placed uniformly along the sequence.
- The algorithm will be repeated until the grand-most recent common ancestor (GMRCA) is reached.

Wiuf and Posada’s Setup

They used Hudson’s algorithm [2] but assumed that recombination events will be placed on the DNA non-uniformly. In what follows, we only focus on the way they evaluated recombination rate and located the hotspots. Here are the key steps they used to do this:
1. The whole chromosome is considered as a line and the region they are interested in as the interval (0,1) on this line (Figure 2).
2. The hotspots centers, \( x_j, j = 1, 2, \ldots \) are chosen according to a point process (Figure 2).
3. Recombination happens around the hotspot, \( x_j \), with rate \( \gamma_j \) per generation.
4. Then the breakpoint is chosen according to a distribution \( x_j \) around \( x_j \) (Figure 3).

New Method

Our method has been setup using a category of point processes called Hawkes processes. The Hawkes processes are a class of self-exciting point processes which can be used to model the events showing the clustering effect. The intensity of this process is as follows:

\[
\lambda(t) = \mu + \sum_{T_i \leq t} g(t - T_i)
\]

in which \( \mu \) is the base intensity, \( g(t) \) is the excitation kernel, \( T_i \) are the occurrence positions, and in our case, \( t \) stands for nucleotide position [1]. The exponential kernel

\[
g(t) = \alpha e^{-\beta t}
\]

is one of the common type of the excitation kernels for Hawkes processes in which \( \alpha \) is the size of self excited jumps and \( \beta \) is the exponential decay rate and \( \alpha, \beta > 0 \).

References