**Outbreaker Likelihoods**

**Genetic pseudo-likelihood**

This describes the probability of observing the genetic data (sequence of i) given our model parameters (the mutation rate) and the proposed transmission tree (described by the ancestor αi, its sequence, and the number of generations between αi and i). It describes the probability of observing the number of genetic differences, given the time between sequences (given by number of generations), the number of bases, and the mutation rate. This can be described by a binomial distribution, with x differences occurring over a potential y bases given a probability of change μ.

**Epidemiological pseudo-likelihood**

This describes the probability of observing the collection given the model parameters (coverage) and the proposed transmission tree (the proposed date of infection, the proposed ancestor αi, the proposed time of infection of αi, and the proposed number of generations). This consists of several parts. The first describes the probability of the proposed delay from infection to collection, simply described with the time-to-collection distribution. The second describes the probability of the given transmission event given the infectiousness of the donor, simply using the generation time distribution. It also incorporates previous generation transmission events, using a convolution function.

The final component describes the probability of having that number of undetected generations (the generation parameter describes the assumed number of undetected generations between detected events) given the sampling coverage. If you have 100% coverage, the only possible number of generations is 1. It is described by a negative binomial distribution (which describes the probability of observing a given number of failures before a specified number of successes occur) – in this case, the probability of achieving one success (sampling a case) before (generation – 1) failures (unobserved cases). In R this is given by dnbinom(#failures, #successes, prob), which describes the probability of missing (generation – 1) cases before sampling a case, given the sampling coverage.