1 Simple coalescent

We assume the standard coalescent process. From \(n\) sample lineages, coalescent process runs until there are \(k\) sample lineages left. Now we randomly select one of the \(k\) lineages. What is the probability of this lineage have only one descendant leaf (i.e. singleton) in the coalescent tree?

2 Expected SFS

In class, I derived the expected branch length leading to a leaf in a coalescent tree is equal to 2, and this leads to \(E(\xi_1) = \Theta\), where \(\xi_i\) is equal to the number of sites with \(i\) mutant alleles. This is a special case for the expected SFS. In general, \(E(\xi_i) = \frac{2^i}{i}\). Now your task is to derive this general formula.

We take the following approach. Suppose there are \(k\) (\(2 \leq k \leq n\) lineages left, where \(n\) is the number of sample lineages to begin with) lineages at some point of time. We say we are at stage \(k\) if there are \(k\) sample lineages. We say one lineage \(a\) is of size \(i\) if \(a\) has \(i\) leaves below it. Then by linearity of expectation, we have:

\[
E(\xi_i) = \sum_{k=2}^{n} \sum_{j=1}^{k} E(\text{number of mutations on j-th lineage}) \text{Prob(lineage j at stage k is of size i}).
\]

Part 1

Derive \(E(\text{number of mutations on j-th lineage})\): the expectation of number of mutations occurring on a branch at stage \(k\).

Now, we consider \(\text{Prob(lineage j at stage k is of size i)}\), the probability of one lineage at stage \(k\) has size \(i\). Let us consider this situation. We have \(k\) lineages and think of adding \(n-k\) lineages by the following process: “when adding a lineage \(x\), \(x\) is randomly glued to one of the existing lineages (the original \(k\) lineages plus the newly added lineages)”. From genealogical point of view, we are looking forward in time and these additional lineages correspond to those coalesced into the \(k\) remaining lineages. After all \(n-k\) lineages are added, we have \(k\) groups of lineages; each group contains one of the original \(k\) lineages and those glued directly or indirectly to this lineage. So to make the \(j\)-th branch have \(i\) leaves below, the group formed by the \(j\)-th branch must contain additional \(i-1\) out of \(n-k\) lineages.

Part 2

Now suppose we are adding the \(n-k\) lineages one by one according to the above procedure at stage \(k\). What is the probability of having the first \(i-1\) lineages fall to the group of the \(j\)-th lineage and the others do not? What is the probability of having the first lineage is not in the group of the \(j\)-th lineage but the following \(i-1\) lineages are? In general, what is the overall probability that exactly \(i-1\) lineages become members in the group of the \(j\)-th lineage out of the \(n-k\) lineages?

Part 3

Now use your solution for Part 2 to derive \(E(\xi_i)\).