Lecture - 32

Thursday, 26 October 2016 (15:10- 16:00)

Mitochondrial Eve

In 1987, Cann, Stoneking and Wilson published a paper in the journal Nature, where they claimed that the mother of all the existing humans is the same. If every person builds his maternal lineage, all the lineages merge at the same lady- Eve, who existed about 20,000 years ago, probably in Africa. By maternal lineage, we mean that a person looks to his mother, she looks to her mother, she looks to her mother and so on. This chain is called a lineage. It has been shown in the Figure 1

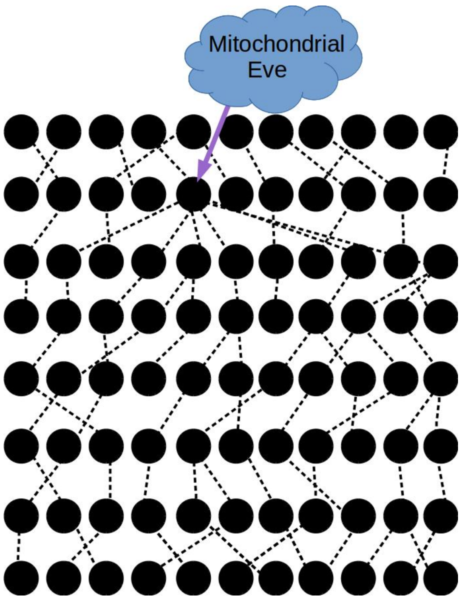

Fig. 1: Mitochondrial Eve

In this lecture, we try to model this phenomenon. Next, we describe the problem statement.

1 Problem Statement

Assume every generation consists on N people each. Every person at level j chooses a parent at level $j-1$ uniformly at random. We take a sample of k people from the lowermost generation and try to find where there maternal lineage collide. This has been shown in the Figure 2

Aim: Find the expected number of generations before which the mitochondrial eve existed.

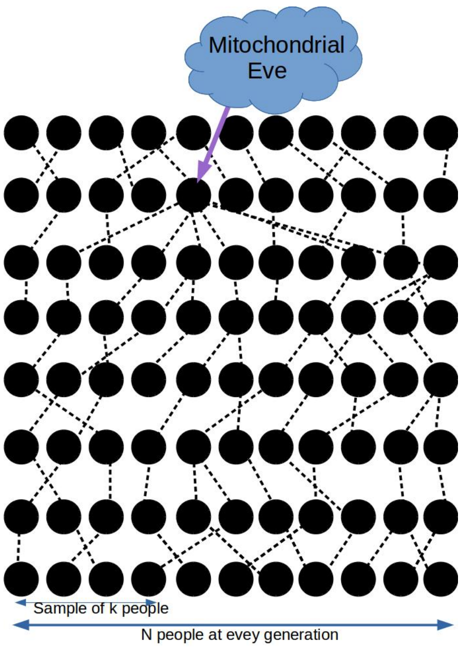

Fig. 2: Problem Statement

We know that at the lowest generation there are k lineages since there are k people. As we move upward, one or more of these lineages merge together and the number of lineages is reduced. In Figure 3, some of the nodes where two or more lineages merge have been marked red. Mitochondrial eve is the node where the number of lineages become 1.

2 Analysis

Consider a particular generation where the number of lineages is j. Probability that the previous generation also has the same number of lineages $j=$ Pr(No two lineages have merged together).

Pr(No 2 lineages merge together)= $1(1-\frac{1}{N})(1-\frac{2}{N})...(1-\frac{j-1}{N})$ (See how one can apply birthday paradox here.)

$= 1-(\frac{1}{N}+\frac{2}{N}+.......+\frac{j-1}{N})+\frac{g(j)}{N^2}+\frac{h(j)}{N^3}+...$, where $g(j), h(j), ...$ are some functions of j

Here we use our first assumption. According to this assumption, N is a very large number as compared to k, hence very large as compared to j as well. Hence all the term which contain N^2 or a greater power of N in the denominator are discarded.

Hence,

Pr(No 2 lineages merge together)= $1-(\frac{1}{N}+\frac{2}{N}+.......+\frac{j-1}{N})$

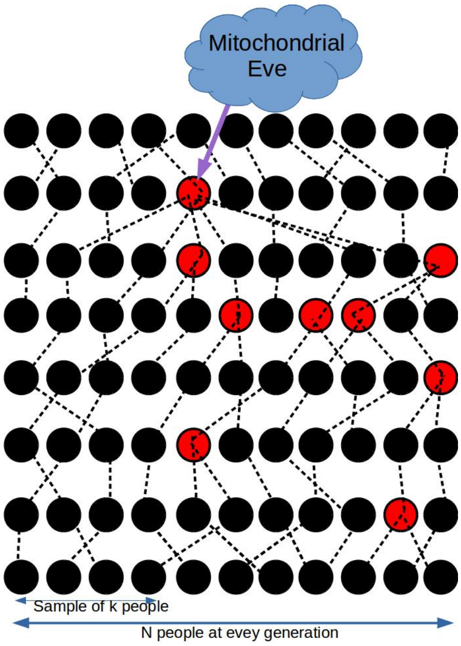

Fig. 3: Merging of lineages

$= 1 - \frac{j(j-1)}{2N}$

Pr (There is at least one pair of lineages that merge)$= \frac{j(j-1)}{2N}$ ....eq(1)

2.1 Ways in which merging can occur

Merging of lineages can occur in many ways. This has been shown in Figure 4.

Let us look at all of the cases one by one:

1. Case (a): The probability that there is exactly one pair in j lineages which merge at the previous generation.
   Given a pair of two lineages, Pr(They Merge)$= 1 \times \frac{1}{N} = \frac{1}{N}$.
   There are a maximum of $\binom{j}{2}$ such pairs. Hence, the probability of this case happening $= \frac{O(j^2)}{N}$

2. Case (b): The probability that there is a set of exactly $\lambda(\lambda > 2)$ lineages which merge together at the previous level.
   Given a set of λ lineages, Pr(They Merge)$= 1\frac{1}{N}\frac{1}{N}$.......$(\lambda - 1\ times)\frac{1}{N} = (\frac{1}{N})^{\lambda-1}$.
   There are a maximum of $\binom{j}{\lambda}$ such pairs. Hence, the probability of this case happening $= \frac{O(j^2)}{N^{\lambda-1}}$.
   The degree of N in the denominator is greater than 1. Since, N is a very large number, we can assume this probability to be almost equal to zero. Hence, in our analysis, we assume that lineages do not merge this way.

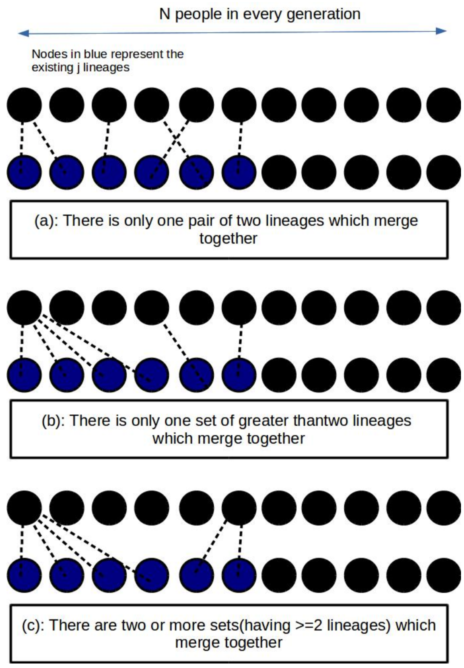

Fig. 4: Different Ways of Merging

3. Case (c): There are more than one sets which merge together to different people in the above generation. Let us take the smallest of such example. Here, number of sets is 2 and every set has 2 elements. This has been shown in Figure 5. Pr(This happening)= $\binom{j}{4} \times \frac{1}{N}\frac{1}{N}$

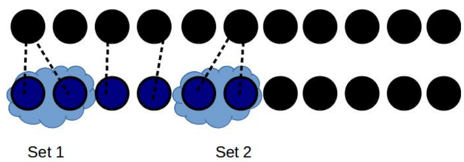

Fig. 5: Smallest possible example of case (c)

Since, N is large, this probability also approaches zero. All the other cases possible where the size of the sets is greater than 2 will comprise of even more higher power of N in the denominator. So, those events can also be considered rare.

Because of the above points, we reach to our second conclusion: **All the merging which happens in a particular generation is only of type (a). Hence, whenever a merging happens, we assume that there is exactly one pair of lineages which have merged.**

2.2 Back to the Analysis

Keeping in mind the above assumption, we proceed further.

From equation 1 Pr (There is at least one pair of lineages that merge)= $\frac{j(j-1)}{2N}$

Because of the assumption that merging always happens because of exactly one pair of lineages, we can write

Pr (Merging)= $\frac{j(j-1)}{2N}$

Our aim is to find the expected number of generations after which the number of lineages become two. Once there are 2 lineages, we know that they merge with a probability of $\frac{1}{N}$.

Number of generations to reach the mitochondrial eve= Number of generations where there were k lineages + Number of generations where there were $k-1$ lineages + ... + Number of generations where there were 2 lineages.

Let the number of generations where there were j lineages be represented with a random variable X_j.

$X = X_k + X_{k-1} + X_{k-2} + ... + X_2$
$E[X] = E[X_k] + E[X_{k-1}] + E[X_{k-2}] + ... + E[X_2]$ Equation2

Finding $E[X_j]$ X_j represents the number of generations where the system remains with j lineages. As soon as there is a merging, this phase will end.

We know Pr(Merging)= $\frac{j(j-1)}{2N} = p$, say

Pr(Not Merging)= $1 - p$

$E[X_j]$= $Pr(X_j \geq 1) + Pr(X_j \geq 2) + ...$
$= 1 + (1-p) + (1-p)^2 + ...$
$= \frac{1}{p}$
$= \frac{2N}{j(j-1)}$

Put in Equation 2
$E[X] = \frac{2N}{k(k-1)} + \frac{2N}{(k-1)(k-2)} + ...\frac{2N}{2}$
$= 2N(\frac{1}{k(k-1)} + \frac{1}{(k-1)(k-2)}) + ...\frac{1}{2\times 1}$

This equation can be solved by writing $\frac{1}{k(k-1)} = \frac{1}{k} - \frac{1}{k-1}$
. Once you write it this way, all the terms except $\frac{1}{k}$ and $\frac{1}{1}$ gets canceled out.

Hence, $E[X] = 2N(1 - \frac{1}{k})$