Statistical population genetics

Lecture 9: The fate of Neanderthal

Xavier Didelot

Dept of Statistics, Univ of Oxford
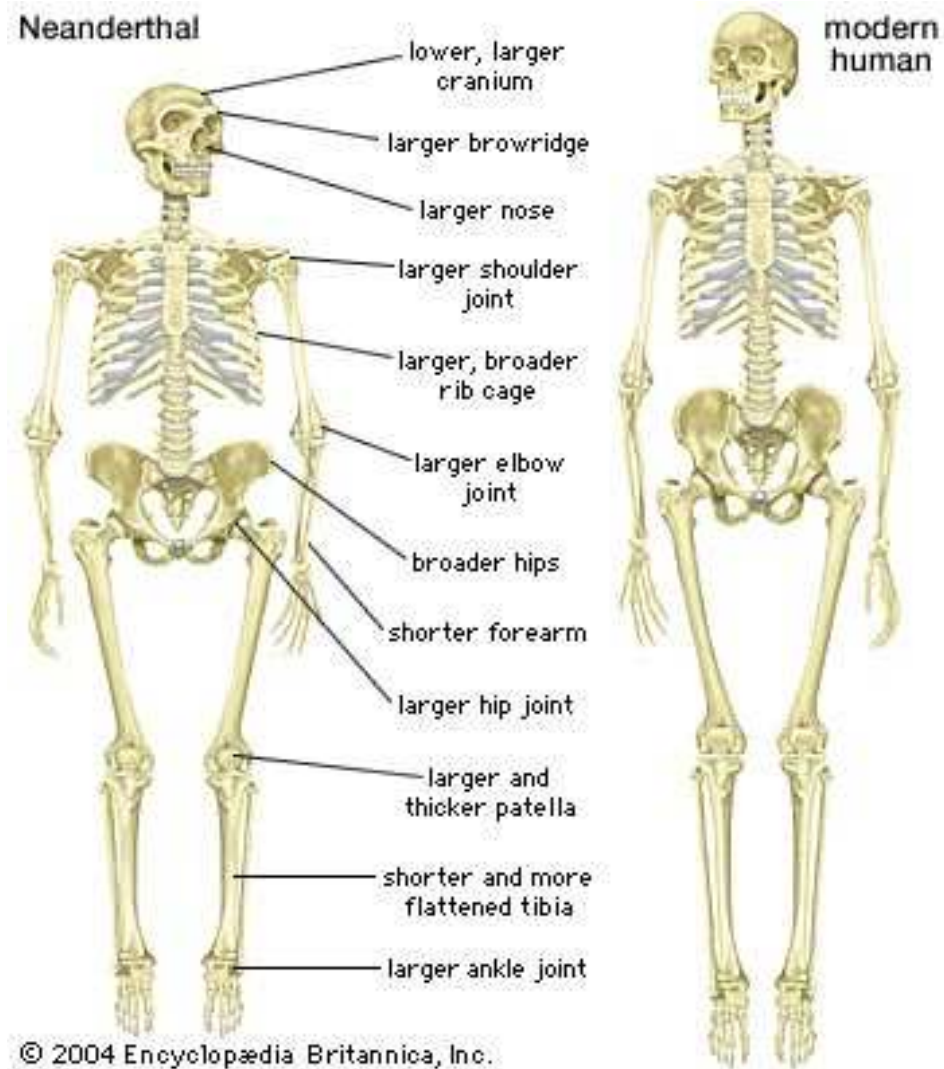
didelot@stats.ox.ac.uk

The Neanderthal

- The Neanderthals were hominids who lived in Europe and western Asia

- Oldest fossils around 100,000 years ago

- Disappear from Asia around 50,000ya

- Disappear from Europe around 30,000ya

The Neanderthal

Morphologically distinct from modern humans

Fate of the Neanderthal

- Neanderthal and Homo sapiens coexisted around 50,000-30,000ya

- Hypothesis 1: Neanderthal was a separate species (ie. unable to breed with Homo sapiens) which became extinct

- Hypothesis 2: Neanderthal was a subspecies, and became integrated into anatomically modern humans

- Homo neanderthalensis or Homo sapiens neanderthalensis?

- This debate has been raging for over a century

- But now we have a new way to find out... DNA

Krings *et al.* (1997)

Neandertal DNA Sequences and the Origin of Modern Humans

Abstract: DNA was extracted from the Neandertal-type specimen found in 1856 in western Germany. [...] Sequence comparisons with human mtDNA sequences show that the Neandertal sequence falls outside the variation of modern humans. [...] This suggests that Neandertals went extinct without contributing mtDNA to modern humans.

Krings *et al.* (1997)

986 modern humans   Neanderthal
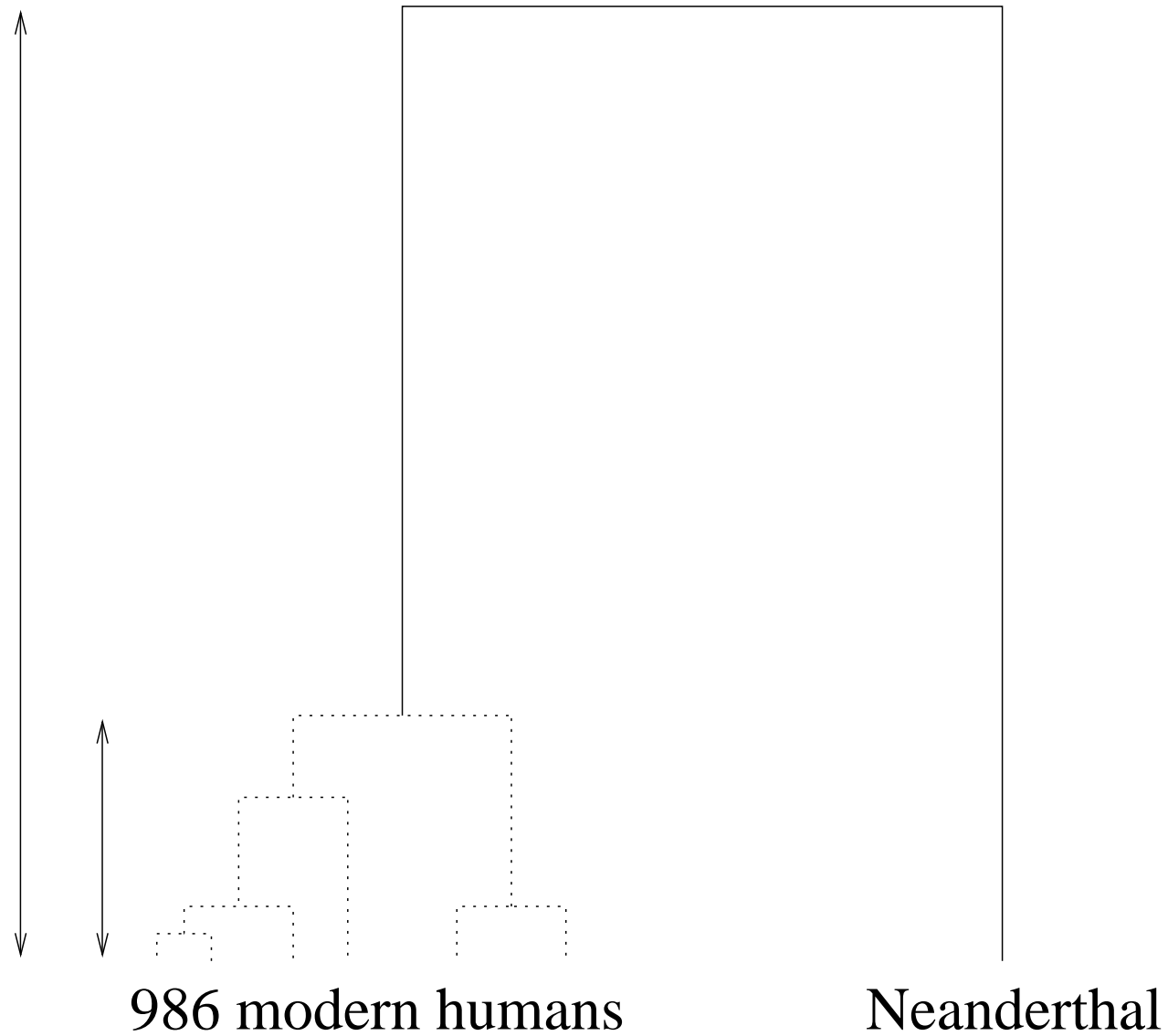
Krings *et al.* (1997)

- Krings *et al.* (1997) sequenced a mitochondrial DNA sequence recovered from a Neanderthal skeleton

- Compared it to a sample of 986 sequences from modern humans

- They reconstructed a genealogy where all modern humans coalesce together before coalescing with Neanderthal

- Based on this, they rejected the hypothesis of inbreeding

- Can we quantify the evidence using statistics?

- Yes, thanks to the coalescent model!

First analysis

- Assuming that Neanderthal and Homo sapiens are a unique population (ie. with random mating between them), what is the probability p that Neanderthal is last to coalesce in a sample of $n + 1 = 987$ sequences?
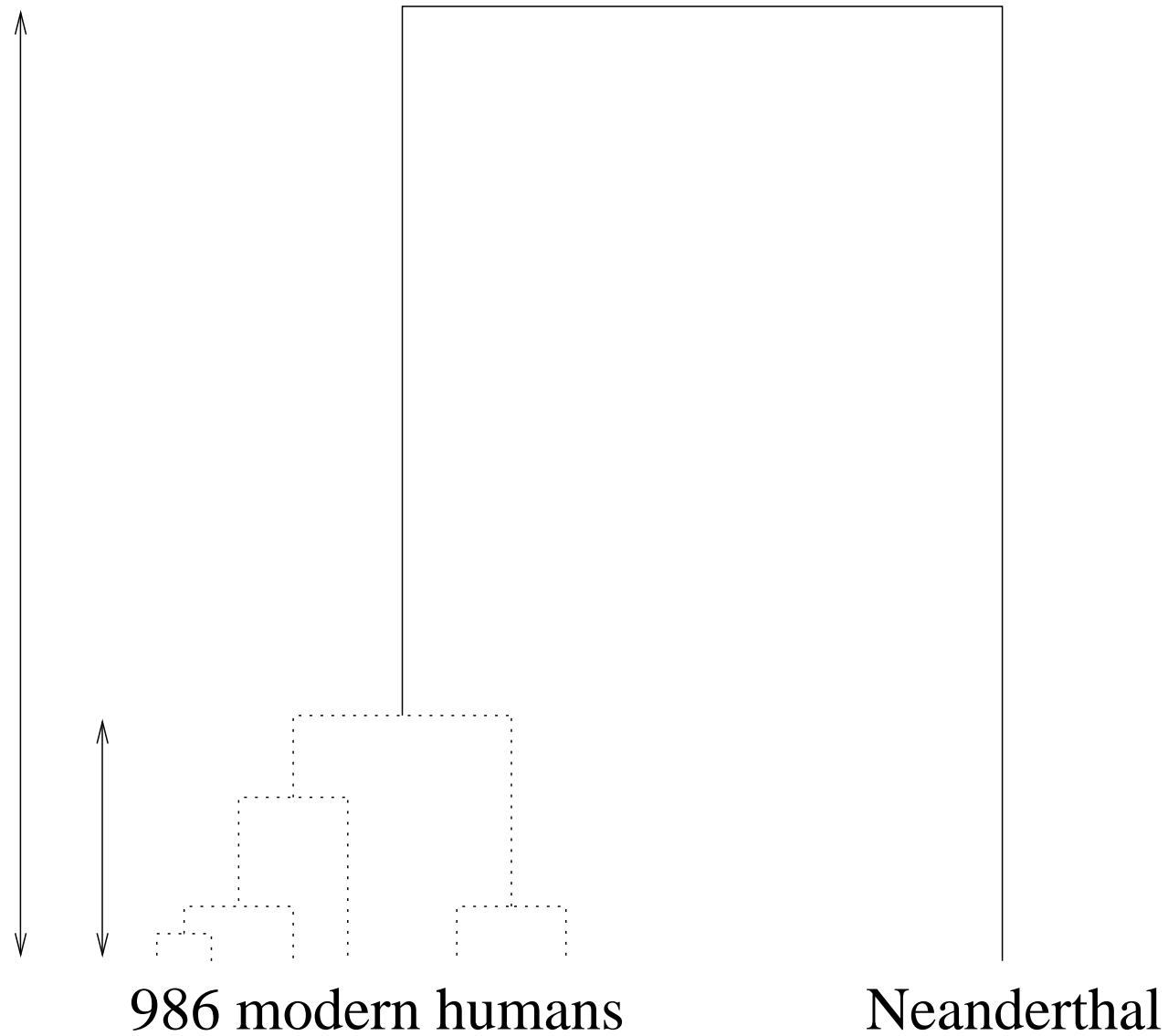
- It is equal to:

$$p = \frac{\binom{n}{2}}{\binom{n+1}{2}} \frac{\binom{n-1}{2}}{\binom{n}{2}} \cdots \frac{\binom{2}{2}}{\binom{3}{2}}$$

$$= \frac{2}{n(n+1)} \approx 2 \cdot 10^{-6}$$

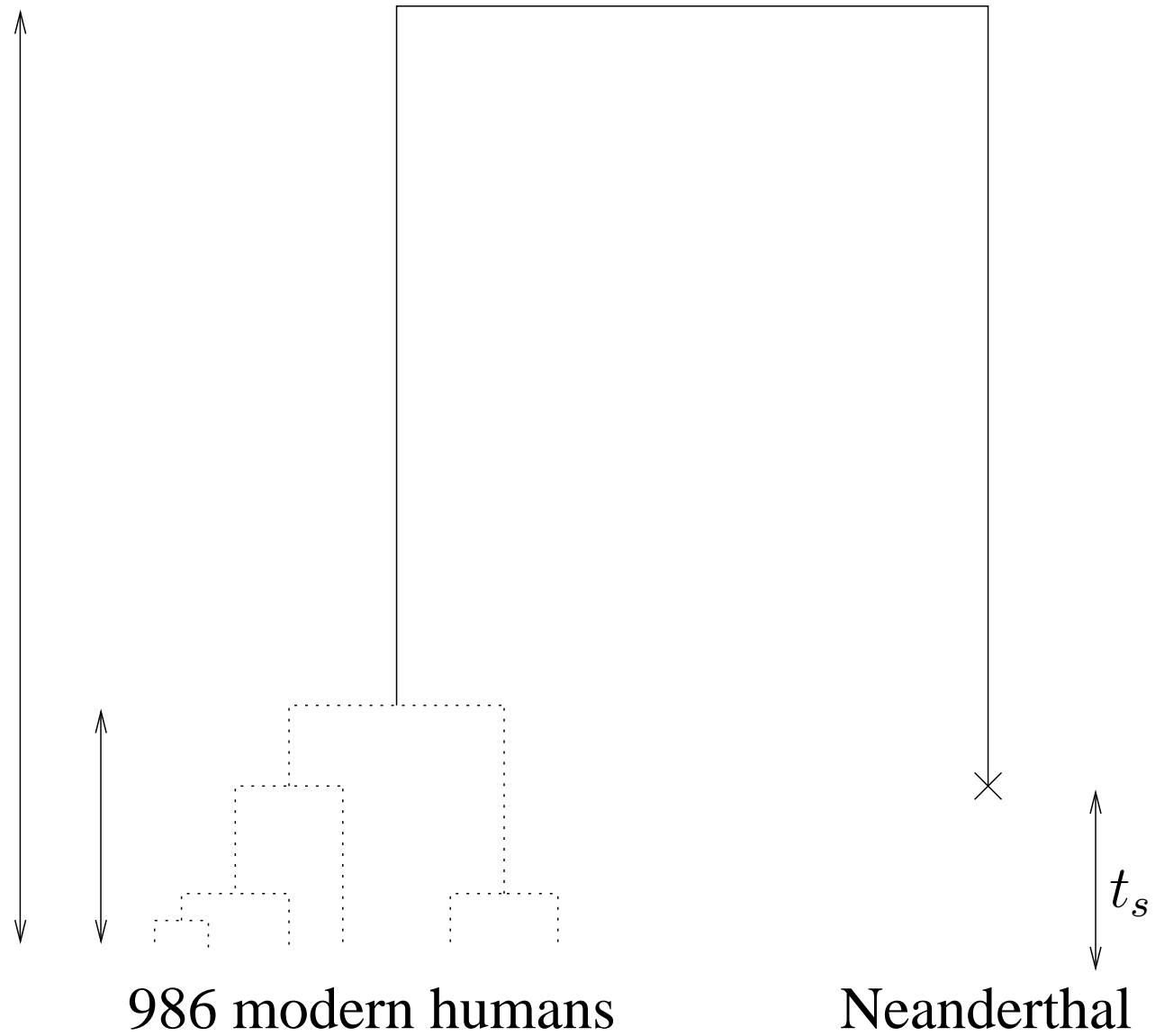- So the probability of this tree branching order is very small indeed...

But...

- The Neanderthal individual died at time $t_s \in [30,000; 100,000]$ years

- By that time, the number of human lineages that gave rise to the n sequences observed today was potentially much smaller.

Krings *et al.* (1997)

986 modern humans   Neanderthal

But...

986 modern humans   Neanderthal

t_s

But...

- The Neanderthal individual died at time $t_s \in [30,000; 100,000]$ years

- By that time, the number of human ancestors that gave rise to the n sequences observed today was potentially much smaller.

- Taking $M_e = 3,400$ and 20 years for the average length of a generation, we have $t_s \in [0.44; 1.47]$ coalescent units.

- Let $A_n(t_s)$ denote the number of ancestors of the 986 humans at time t_s in the past, what is its distribution?

- We compute the probability that Neanderthal is last to coalesce:

$$\mathbb{P}(\text{last}) = \sum_{k=1}^{986} \mathbb{P}(\text{last}|A_n(t_s) = k)\mathbb{P}(A_n(t_s) = k)$$

Number of ancestors in the past
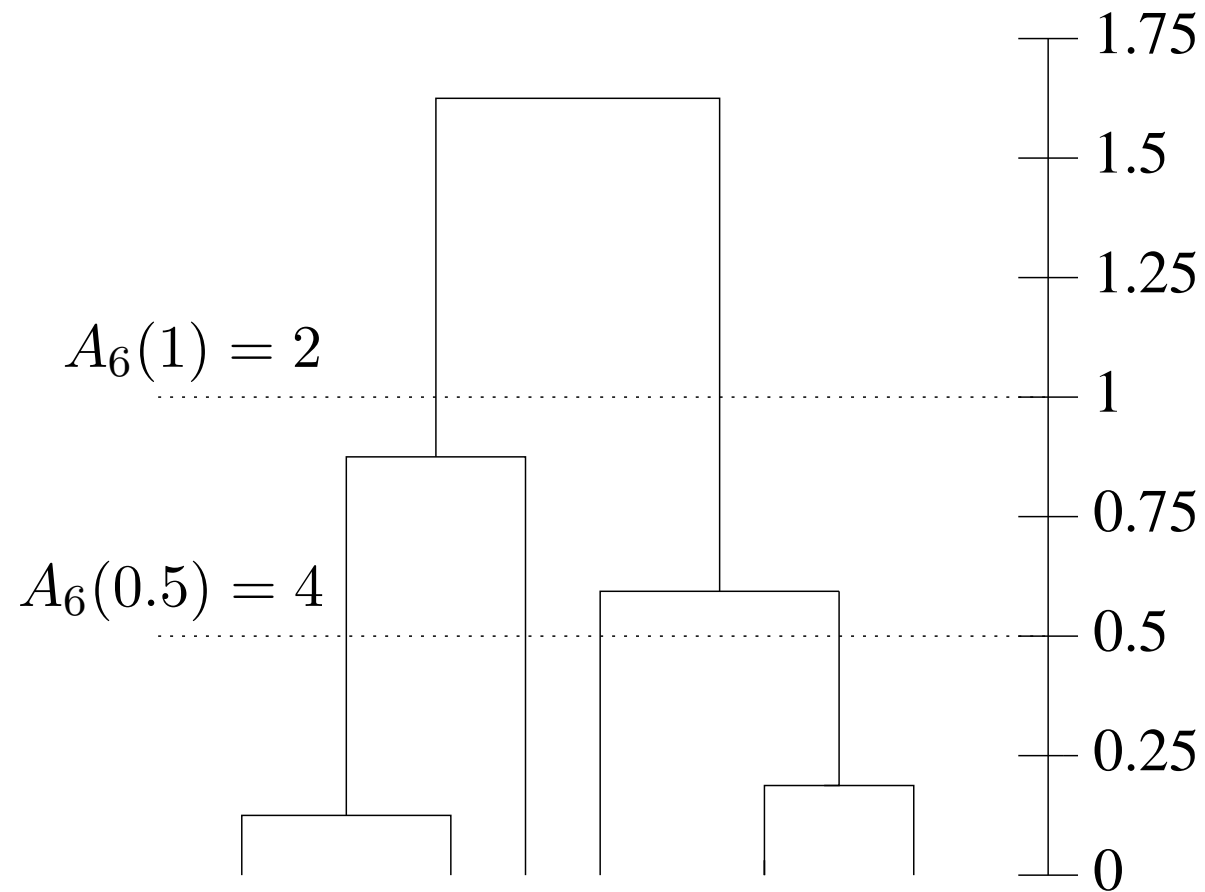
Number of ancestors in the past

- Despite the simplicity of the formulation of the coalescent model, even simple questions can have complicated answers.

- This result is due to Tavaré (1984)

Theorem (Number of ancestors in the past)**.**

Let $A_n(t)$ denote the number of ancestors at time t back in time

$\mathbb{P}(A_n(t) = k)$ *is:*

$$
\begin{cases}
1 - \left(\sum_{i=2}^{n} e^{-i(i-1)t/2} \prod_{j=2, j \neq i}^{n} \frac{j(j-1)}{j(j-1) - i(i-1)} \right) & k = 1 \\
\mathbb{P}(A_n(t) = k) = \frac{2}{k(k-1)} \sum_{i=k}^{n} \frac{i(i-1)}{2} e^{-i(i-1)t/2} \prod_{j=k, j \neq i}^{n} \frac{j(j-1)}{j(j-1) - i(i-1)} & k > 1
\end{cases}
$$

Notations

- Time before first coalescence for i individuals:

$$T_i \sim \text{Exponential}\left(\frac{i(i-1)}{2}\right)$$

- Time before n individuals find k ancestors:

$$T_{n,k} = \sum_{i=k+1}^{n} T_i$$

- Example: $T_{n,1}$ is the time to the most recent common ancestor of n individuals (ie. the height of the whole tree)

- The distribution of $T_{n,k}$ can be computed using a convolution

Number of ancestors in the past

Proof.

For there to be k ancestors at time t, exactly $n - k$ coalescent events must have occur ed:

$$\begin{cases} T_{n,k} & = T_n + T_{n-1} + \ldots + T_{k+1} & < t \\ T_{n,k-1} & = T_n + T_{n-1} + \ldots + T_{k+1} + T_k & > t \end{cases}$$

In the case $k = 1$, the second condition is always satisfied since a single ancestor can not coalesce any further. Therefore:

$$\mathbb{P}(A_n(t) = 1) = \mathbb{P}(T_{n,1} < t) = \int_{x=0}^{t} f_{T_{n,1}}(x)\mathrm{d}x$$

Number of ancestors in the past

Using a convolution of exponentials, we can show that the distribution of $T_{n,1}$ is:

$$f_{T_{n,1}}(x) = \sum_{i=2}^{n} \frac{i(i-1)}{2} e^{-i(i-1)x/2} \prod_{j=2,j\neq i}^{n} \frac{j(j-1)}{j(j-1) - i(i-1)}$$

Therefore:

$$\mathbb{P}(A_n(t) = 1) = \sum_{i=2}^{n} \frac{i(i-1)}{2} \int_{x=0}^{t} e^{-i(i-1)x/2} dx \prod_{j=2,j\neq i}^{n} \frac{j(j-1)}{j(j-1) - i(i-1)}$$

$$= \sum_{i=2}^{n} \left(1 - e^{-i(i-1)t/2}\right) \prod_{j=2,j\neq i}^{n} \frac{j(j-1)}{j(j-1) - i(i-1)}$$

$$= 1 - \left(\sum_{i=2}^{n} e^{-i(i-1)t/2} \prod_{j=2,j\neq i}^{n} \frac{j(j-1)}{j(j-1) - i(i-1)} \right)$$

The last equality comes from the fact that the distribution of $T_{n,1}$ integrates to 1.

Number of ancestors in the past

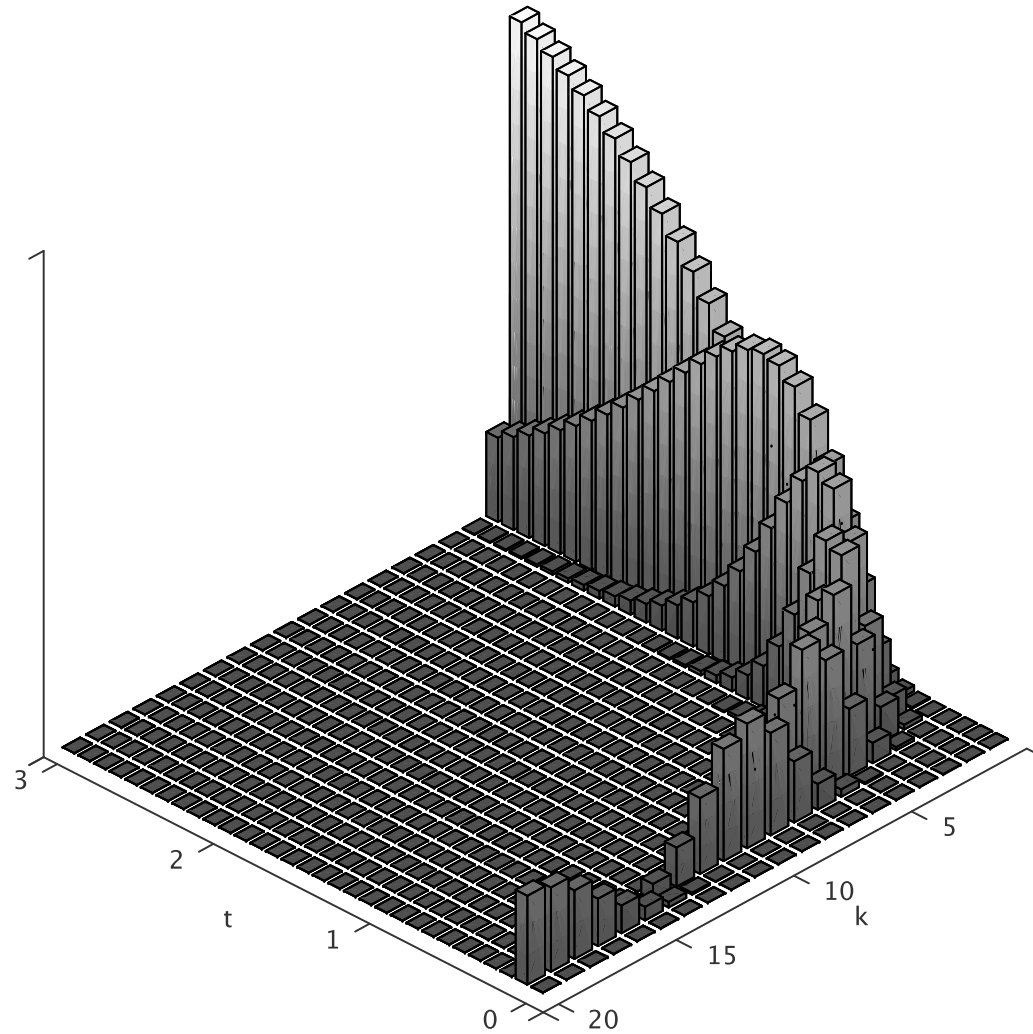For $k > 1$, we need to make sure that the second condition is satisfied too, which leads to:

$$\mathbb{P}(A_n(t) = k) = \mathbb{P}(T_{n,k} < t, T_{n,k-1} > t)$$

$$= \int_{x=0}^{t} f_{T_{n,k}}(x)\mathbb{P}(T_k > t - x)dx$$

Using again a convolution of exponentials, we can show that:

$$f_{T_{n,k}}(x) = \sum_{i=k+1}^{n} \frac{i(i-1)}{2} e^{-i(i-1)x/2} \prod_{j=k+1, j \neq i}^{n} \frac{j(j-1)}{j(j-1) - i(i-1)}$$

and since $\mathbb{P}(T_k > t - x) = \exp(-(t-x)k(k-1)/2)$, we can inject these two expressions in the derivation of $\mathbb{P}(A_n(t) = k)$ to get the result. $\square$

Number of ancestors in the past

The fate of Neanderthal

$$\mathbb{P}(\text{last}) = \sum_{k=1}^{986} \mathbb{P}(\text{last}|A_n(t_s) = k)\mathbb{P}(A_n(t_s) = k)$$

t_s (years)	t_s (cu)	$\mathbb{E}(A_n(t_s))$	$\mathbb{P}(\text{last})$
0	0	986	$2 \cdot 10^{-6}$
30,000	0.44	4.86	0.085
100,000	1.47	1.75	0.56

Summary

- Mistake of Krings *et al.*: the 986 humans **only had a few ancestors** when the Neanderthal died

- We have calculated the distribution for the **number of ancestors**

- Once we take this into account, the evidence against inbreeding is **not very strong**, especially since panmixia is an extreme scenario

- Project to sequence whole Neanderthal genome started in 2006

- **Recombination** means that nDNA is more informative than mtDNA

- We did **not** study recombination in this introductory course, but it is a very important evolutionary force