Coalescent Process With Changing Population Sizes

Biostatistics 666

Lecture by Guest Expert
Sebastian Zoellner

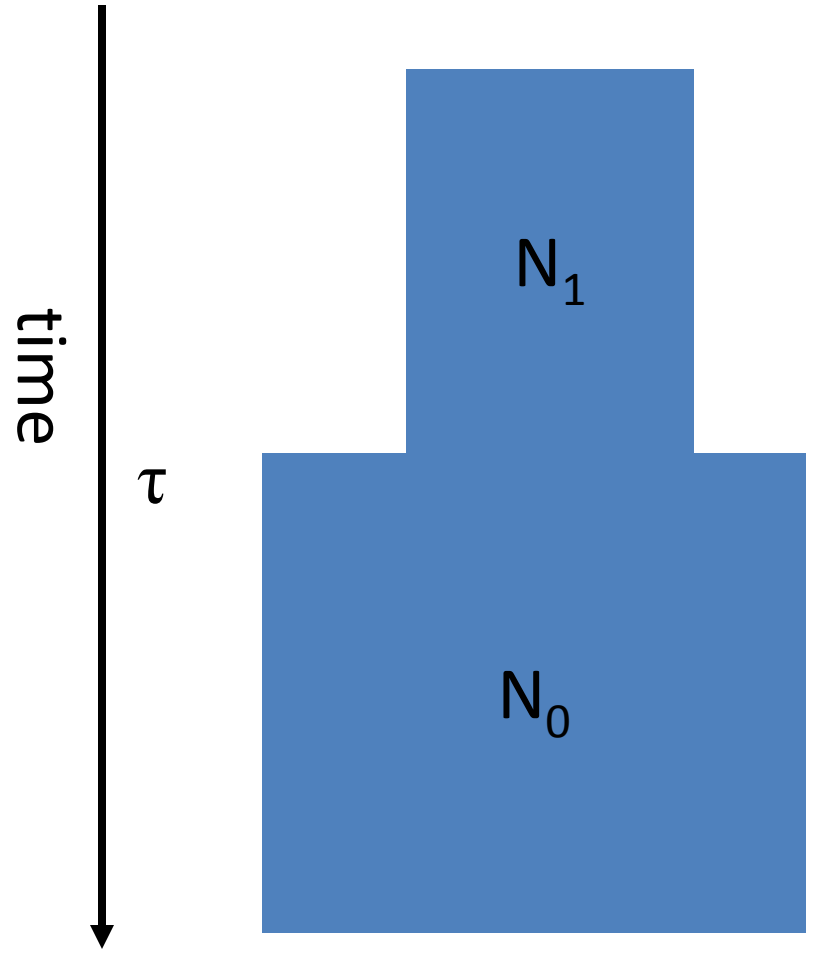
Basic Assumptions of the Coalescent

(and interesting alternatives!)

- Mating is Random
  - ➢ Subdivided Population
- Population Size is Constant
  - ➢ Changing Population Sizes
- Shape of Genealogy is Independent of the Population Size
  - ➢ Natural Selection is equivalent to a subdivided population where population size is changing

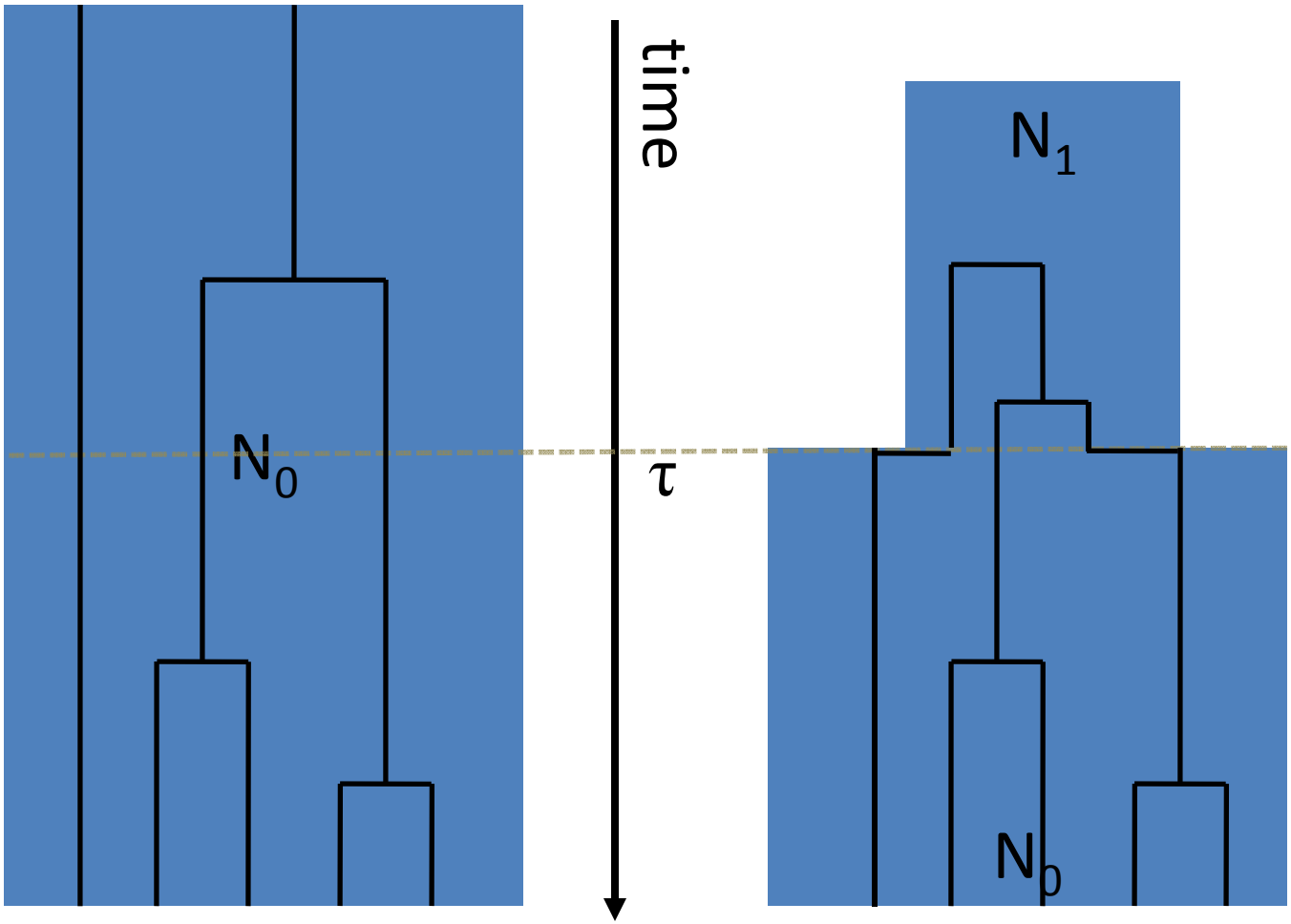
Variable Population Size
Simple Model

- At time τ the population size changes from N_0 to $N_1=\lambda N_0$.
- Hence the coalescence rate before τ is multiplied by λ.
- Conditional on the number of lines j at time τ, the coalescent before τ is independent from the process after τ.

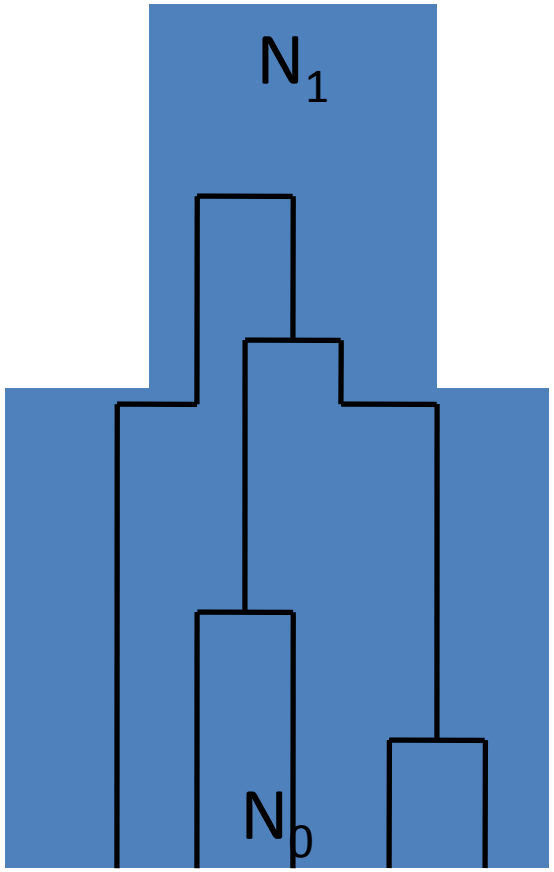


Effect of instant changes of population size

Until time τ, the processes are identical



Effects of changing population size



- What happens to the MPSD*?
- What happens to the number of singletons?
- What happens to the number of variable sites?

*MPSD is the mean pairwise sequence divergence.

Calculations with Changing Population Size

Processes modeled independently for intervals $[0,\tau]$, $[\tau,\infty)$.

All calculations are done conditional the boundary condition j. Equation (1) gives the probability of j, for a given *n* and τ.

$$h_{n,j}(\tau) = \sum_{k=j}^{n} e^{-\tau\binom{k}{2}} \frac{(2k-1)(-1)^{k-j} j_{(k-1)} n_{[k]}}{j!(k-j)! n_{(k)}} \quad (1)$$

Branching patterns unaffected by changes in population size!

Coalescent times are not independent anymore.

Example: Time to the MRCA

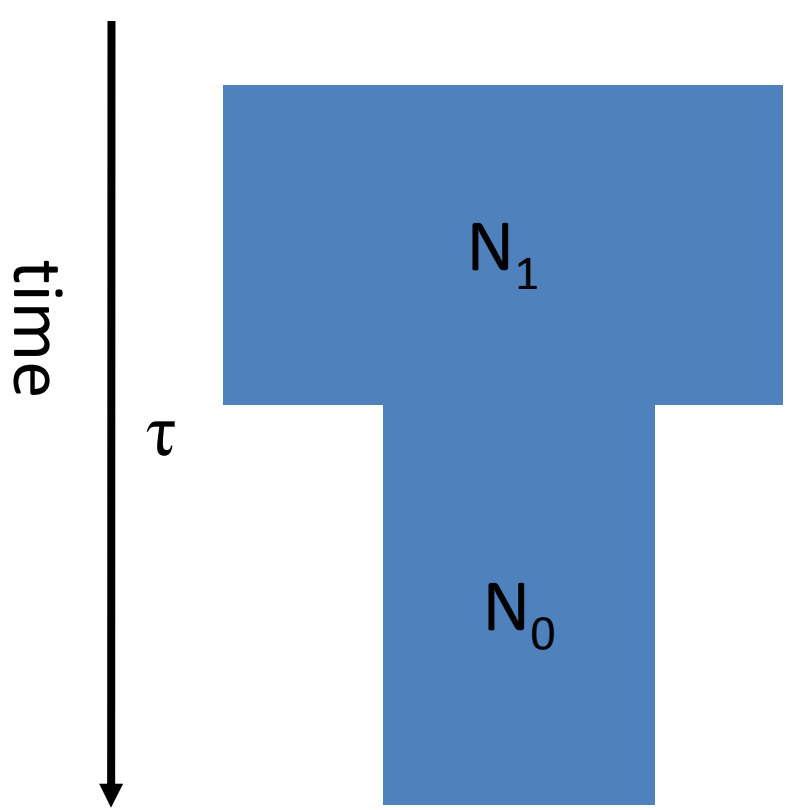
$$E(T_{MRCA}) = E(E(T_{MRCA} \mid j))$$

$$P(j) = h_{n,j}(\tau) = \sum_{k=j}^{n} e^{-\tau\binom{k}{2}} \frac{(2k-1)(-1)^{k-j}\, j_{(k-1)} n_{[k]}}{j!(k-j)!\, n_{(k)}} \quad (1)$$

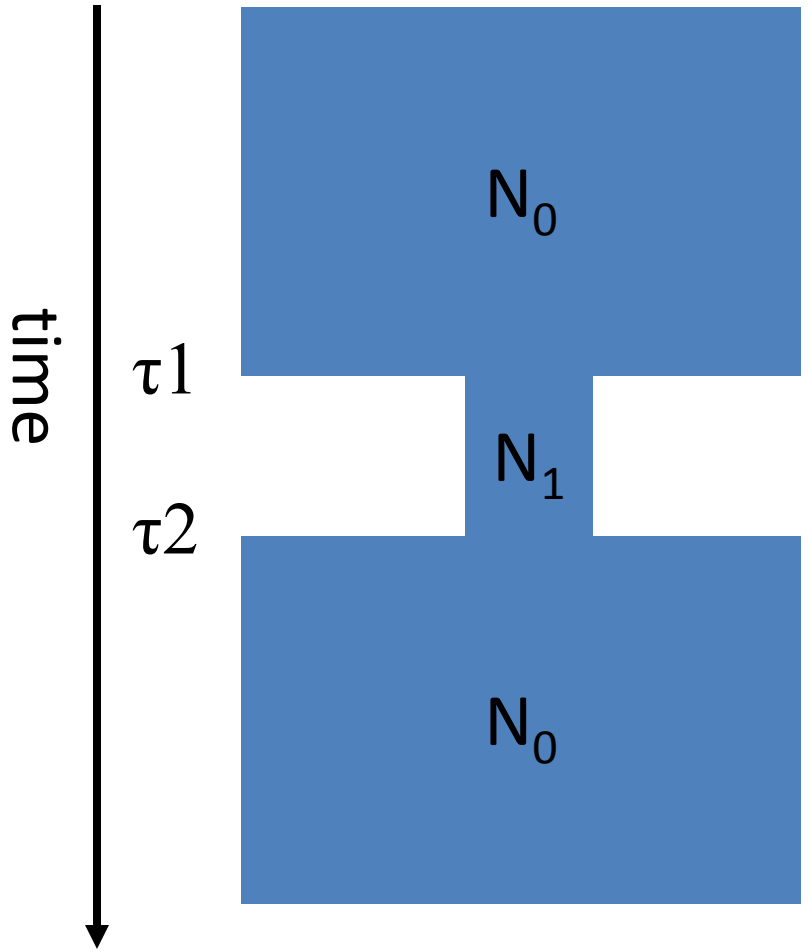
$$E(T_{MRCA} \mid j) = 2\lambda\left(1 - \frac{1}{j}\right) + \tau \text{ for } j > 1$$

$$E(T_{MRCA} \mid j = 1) = tP(T_{MRCA} = t \mid t \le \tau)$$

Other Important Scenarios

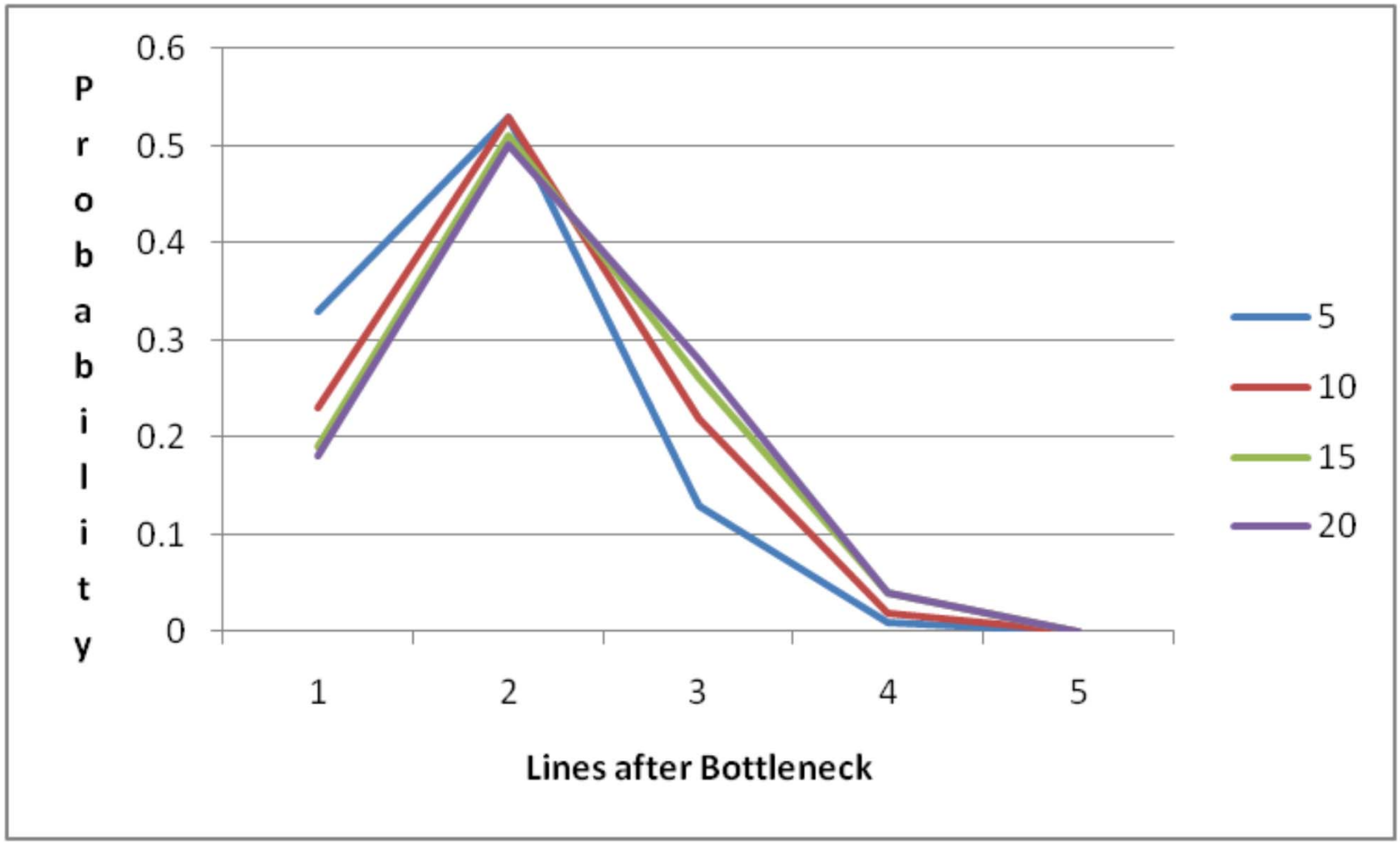


Shrinking Population



Population Bottleneck

Lines after a Bottleneck

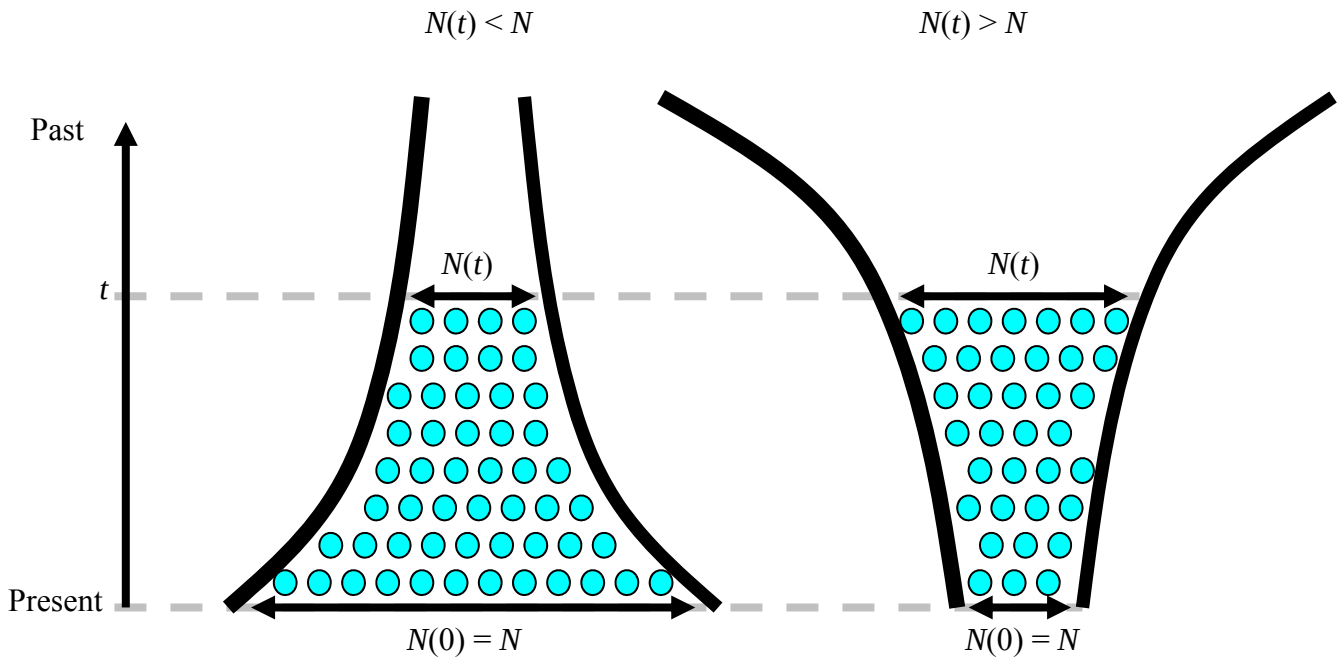


Population Bottlenecks

- Bottlenecks generate rapid loss of ancestral lines.
- Bottleneck's impact summarized by loss of lines.
- With effective population size of N, we expect the same number of coalescent events in time τ as …
- … with effective population size $N\lambda$, but time $\tau\lambda$.

Continuous Population Changes

- In practice, populations do not grow instantly. Instead, exponential growth is a more typical, realistic model.
- To allow this, generate a tree under constant population size and then correct the coalescence times.



LD in the Coalescent

Four Gamete Test for Recombination

Under the infinite sites assumption, a simple test for the presence of recombination can be performed:

- Consider 2 biallelic markers A/a, B/b. If all 4 possible haplotypes AB, Ab, aB, ab exist, a recombination event occurred in the history of the sample.
- This test is conservative, many recombinations will not result in a violation of the 4 gamete test.
- The test is also affected by other violations of the infinite site assumption such as genotyping error.

r^2 and D’

- r^2=1 indicates an ancestry where both mutations occur on the same branch.

- D’=1 indicates that both variants occur on the same tree.