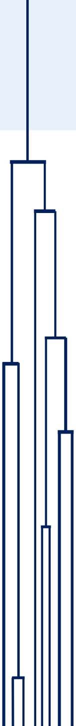
The Kingman Coalescent as a Density on a Space of Trees

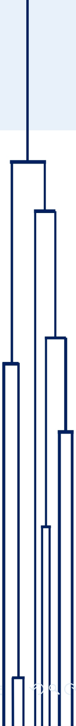
Lena Walter
Freie Universität Berlin
July 31st, 2018
joint work with Christian Haase

Berlin Mathematical School

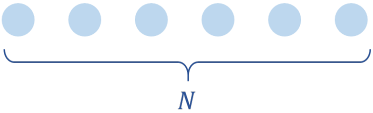


Contents

Contents

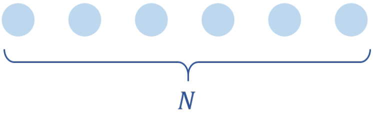


The Kingman n-coalescent



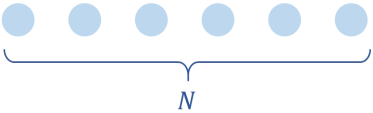
THE KINGMAN n-COALESCENT

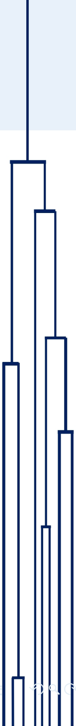
population genetic model:



The Kingman n-coalescent

population genetic model:
Wright-Fisher 30's

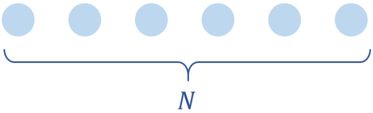


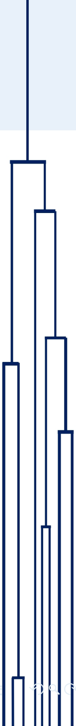


The Kingman n-coalescent

population genetic model:
Wright-Fisher 30's

- fixed population size

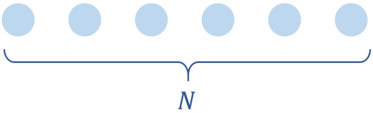




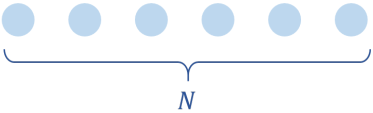
The Kingman n-coalescent

population genetic model:
Wright-Fisher 30's

- fixed population size
- non-overlapping generations



The Kingman n-coalescent



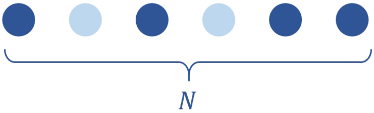
population genetic model:
Wright-Fisher 30's

- fixed population size
- non-overlapping generations
- no recombination

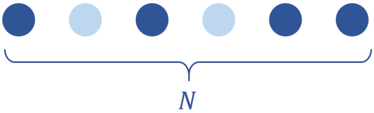
The Kingman n-coalescent

population genetic model:
Wright-Fisher 30's

→ sample n individuals from a population of N



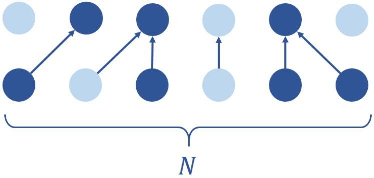
The Kingman *n*-coalescent



population genetic model:
Wright-Fisher 30's

- → sample n individuals from a population of N
- → trace ancestry backwards in time

The Kingman *n*-coalescent

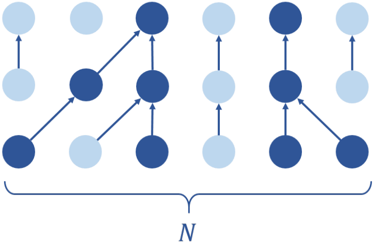


population genetic model:
Wright-Fisher 30's

→ sample n individuals from a population of N

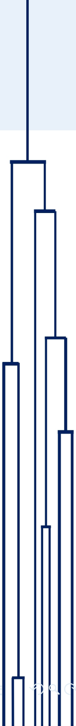
→ trace ancestry backwards in time

The Kingman n-coalescent

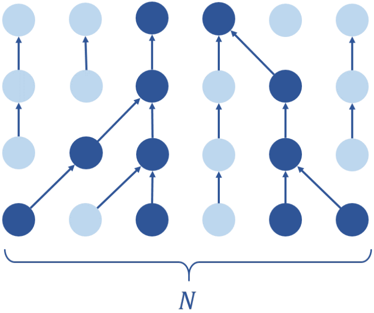


population genetic model:
Wright-Fisher 30's

→ sample n individuals from a population of N
→ trace ancestry backwards in time

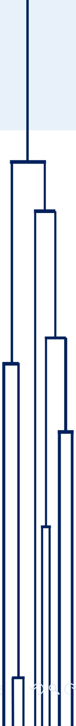


The Kingman n-coalescent

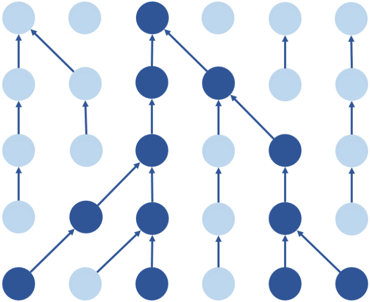


population genetic model:
Wright-Fisher 30's

→ sample n individuals from a population of N
→ trace ancestry backwards in time



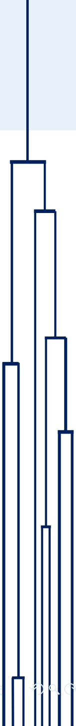
The Kingman n-coalescent



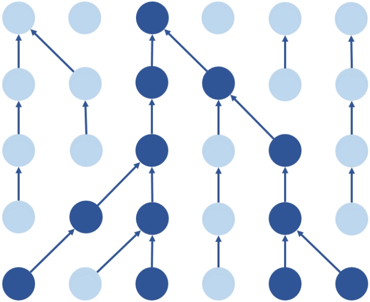
population genetic model:
Wright-Fisher 30's

→ sample n individuals from a population of N
→ trace ancestry backwards in time

MRCA (Most Recent Common Ancestor)?

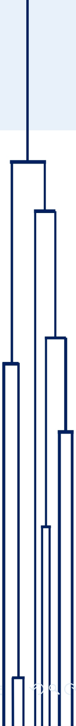


The Kingman n-coalescent

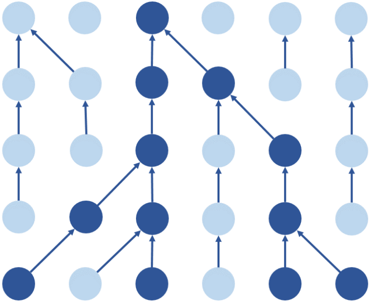


MRCA (Most Recent Common Ancestor)?

Consider two lineages:



The Kingman n-coalescent

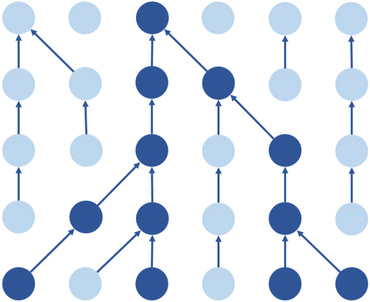


MRCA (Most Recent Common Ancestor)?

Consider two lineages:

$P(\text{c. 1 generation ago}) = \frac{1}{N}$

The Kingman n-coalescent



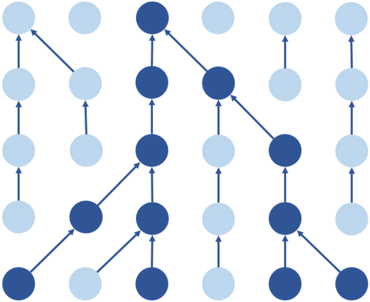
MRCA (Most Recent Common Ancestor)?

Consider two lineages:

$P(\text{c. 1 generation ago}) = \frac{1}{N}$

$P(\text{c. 2 generations ago}) = \left(1 - \frac{1}{N}\right) \frac{1}{N}$

The Kingman n-coalescent



MRCA (Most Recent Common Ancestor)?

Consider two lineages:

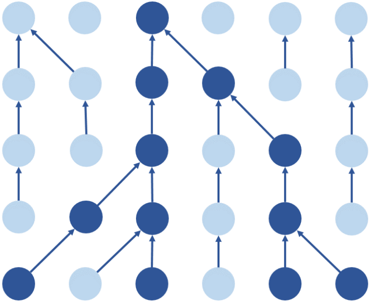
$P(\text{c. 1 generation ago}) = \frac{1}{N}$

$P(\text{c. 2 generations ago}) = \left(1 - \frac{1}{N}\right)\frac{1}{N}$

$\vdots$

$P(\text{c. t generations ago}) = \left(1 - \frac{1}{N}\right)^{t-1}\frac{1}{N}$

The Kingman n-coalescent



MRCA (Most Recent Common Ancestor)?

Consider two lineages:

$P(\text{c. 1 generation ago}) = \frac{1}{N}$

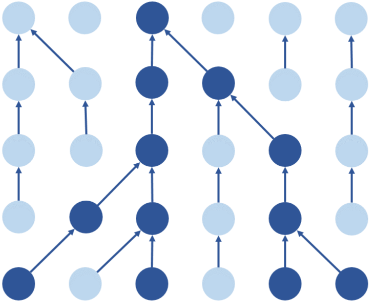
$P(\text{c. 2 generations ago}) = \left(1 - \frac{1}{N}\right)\frac{1}{N}$

$\vdots$

$P(\text{c. t generations ago}) = \left(1 - \frac{1}{N}\right)^{t-1}\frac{1}{N}$

$\rightarrow$ geometric distribution

The Kingman n-coalescent



MRCA (Most Recent Common Ancestor)?

Consider two lineages:

P(c. 1 generation ago) $= \frac{1}{N}$

P(c. 2 generations ago) $= \left(1 - \frac{1}{N}\right)\frac{1}{N}$

$\vdots$

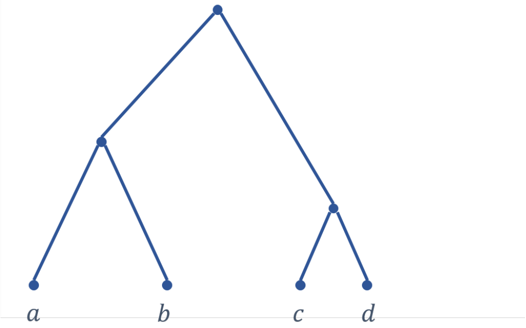
P(c. t generations ago) $=$ $\left(1 - \frac{1}{N}\right)^{t-1}\frac{1}{N}$

$\rightarrow$ geometric distribution

Theorem[Kingman '82]

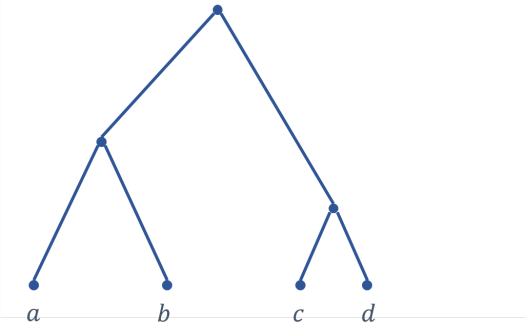
For $N \rightarrow \infty$ this will converge to the Kingman n-coalescent.

The Kingman n-coalescent



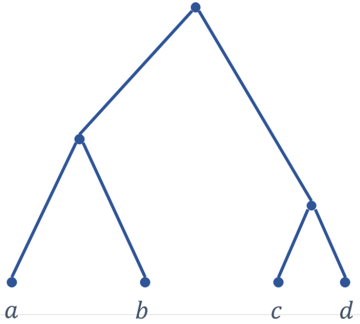
→ consider their genealogy as a binary rooted equidistant n-tree

The Kingman n-coalescent



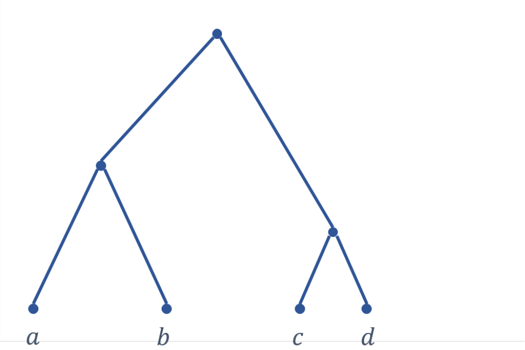
→ consider their genealogy as a binary rooted equidistant n-tree

The Kingman n-coalescent



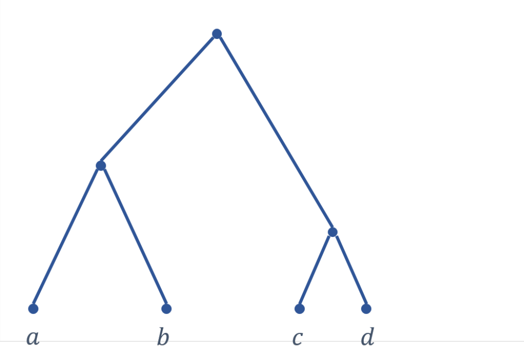
→ consider their genealogy as a binary rooted equidistant n-tree

The Kingman n-coalescent



→ consider their genealogy as a binary rooted equidistant n-tree

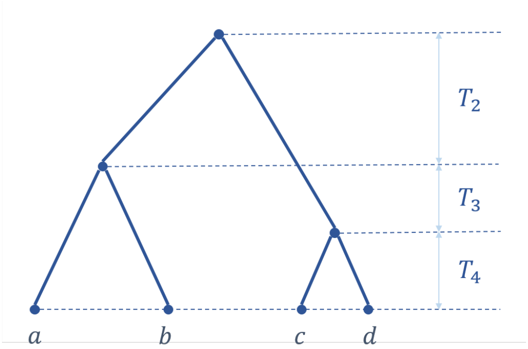
The Kingman n-coalescent



$\rightarrow$ consider their genealogy as a binary rooted equidistant n-tree

Discrete: Choosing pairs to coalesce

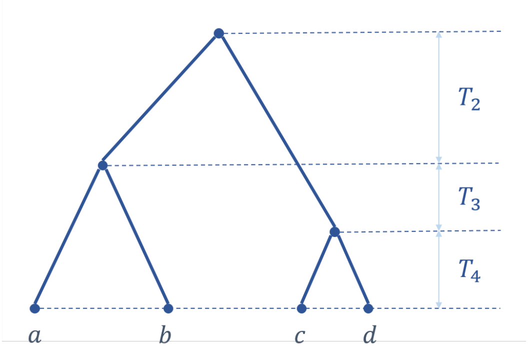
The Kingman n-coalescent



→ consider their genealogy as a binary rooted equidistant n-tree

Discrete: Choosing pairs to coalesce

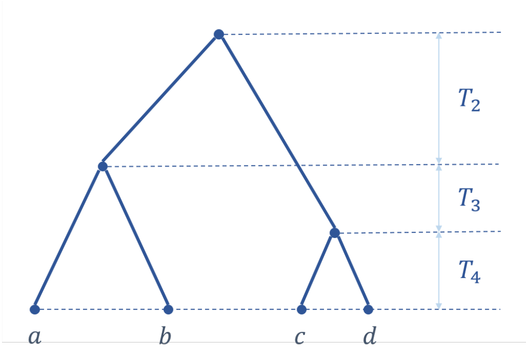
The Kingman n-coalescent



$\rightarrow$ consider their genealogy as a binary rooted equidistant n-tree
$T_j =$ time during which there are exactly j lineages in the sample

Discrete: Choosing pairs to coalesce

The Kingman *n*-coalescent



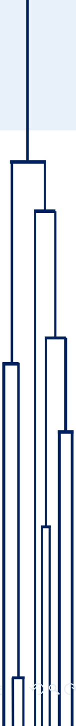
→ consider their genealogy as a binary rooted equidistant *n*-tree

T_j = time during which there are exactly j lineages in the sample

Discrete: Choosing pairs to coalesce

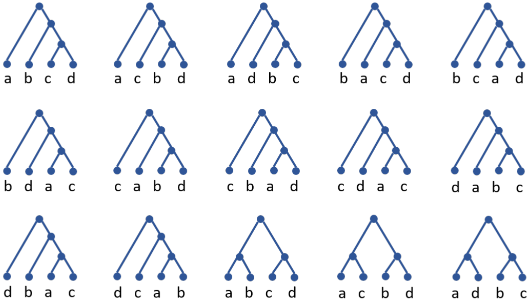
Continuous: Exponentially distributed waiting times $T_j \sim \mathrm{Exp}\binom{j}{2}$

Contents



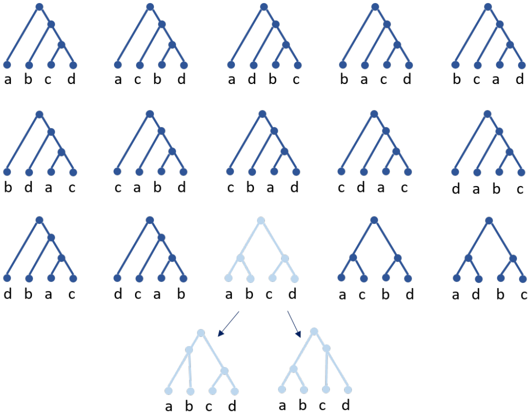
Space of Trees

Coarse combinatorial types of binary rooted equidistant 4-trees



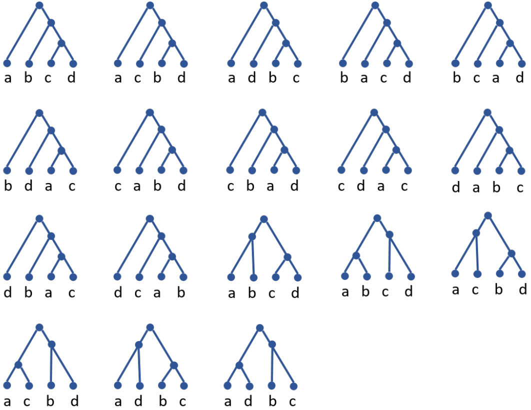
Space of Trees

Coarse combinatorial types of binary rooted equidistant 4-trees

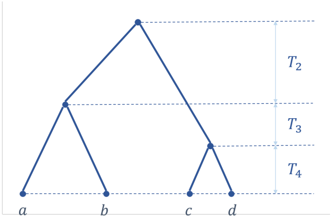


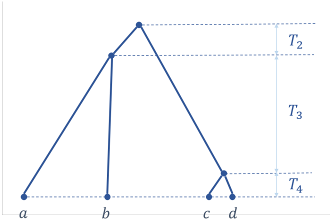
Space of Trees

Fine combinatorial types of binary rooted equidistant 4-trees

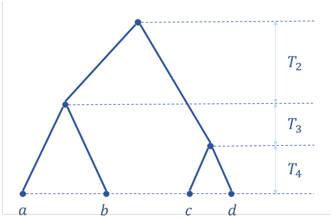


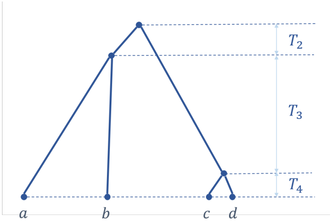
Space of Trees





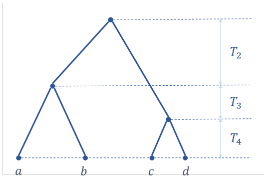
Space of Trees

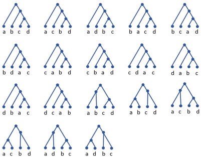




→ Parametrization?

Space of Trees

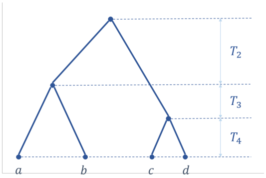


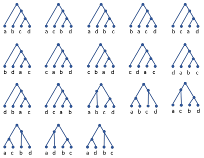


→ Parametrization

- global

Space of Trees





→ Parametrization

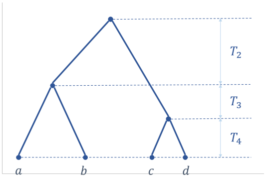
- global

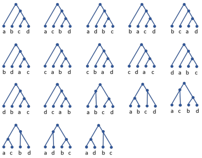
Definition

$$\mathrm{MUM}_n := \{\, d \in \mathbb{R}_{\geq 0}^{\binom{n}{2}} : \forall i, j, k : \max\{d_{ij}, d_{ik}, d_{jk}\} \text{ is attained at least twice}\}$$

Metrics that are UltraMetrics

Space of Trees





→ Parametrization

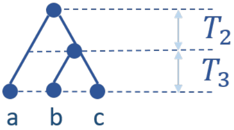
- global
- local

Definition

$$\mathrm{MUM}_n := \{\, d \in \mathbb{R}_{\geq 0}^{\binom{n}{2}} : \forall i,j,k : \max\{d_{ij}, d_{ik}, d_{jk}\}$$
$$\text{is attained at least twice}\}$$

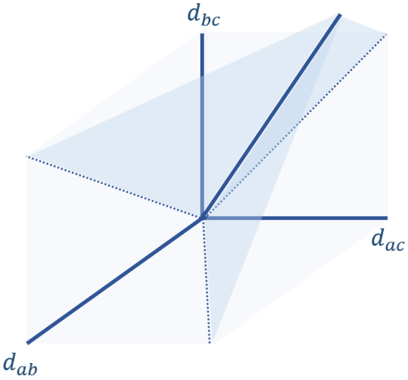
Metrics that are UltraMetrics

Space of Trees

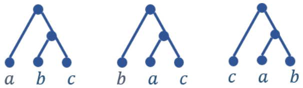


→ Parametrization

- global: coordinates
  d_{ab}, d_{ac}, d_{bc}

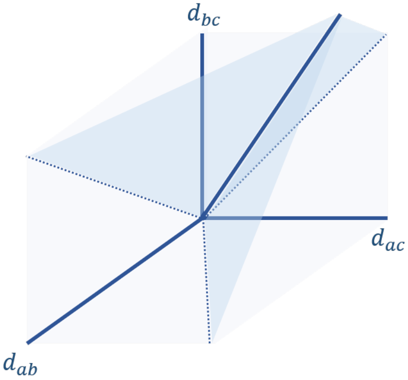


Space of Trees

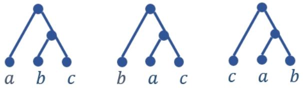


→ Parametrization

- global: coordinates d_{ab}, d_{ac}, d_{bc}

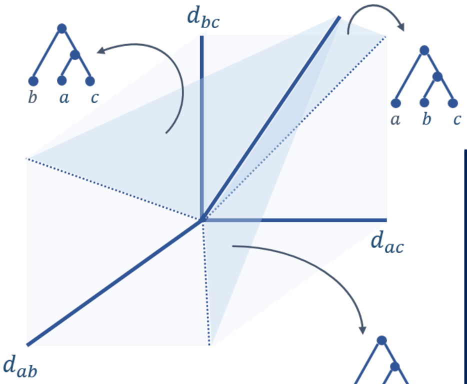


Space of Trees

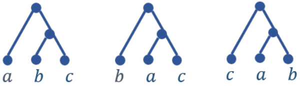


→ Parametrization

- global: coordinates d_{ab}, d_{ac}, d_{bc}

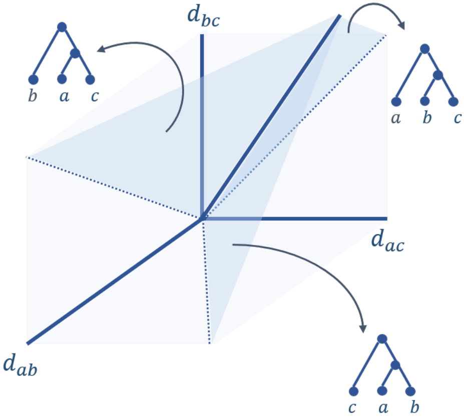


Space of Trees

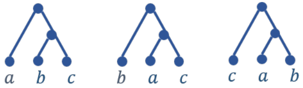


→ Parametrization

- global: coordinates d_{ab}, d_{ac}, d_{bc}
  → one cone for each combinatorial type

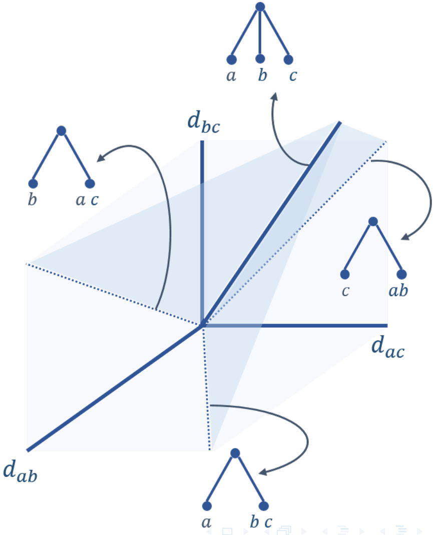


Space of Trees

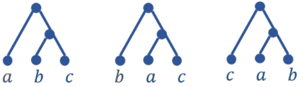


→ Parametrization

- global: coordinates d_{ab}, d_{ac}, d_{bc}
  → one cone for each combinatorial type

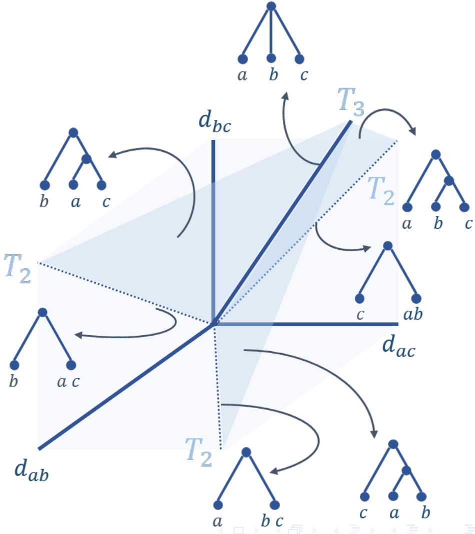


Space of Trees

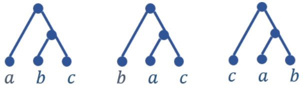


→ Parametrization

- global: coordinates d_{ab}, d_{ac}, d_{bc}
  → one cone for each combinatorial type

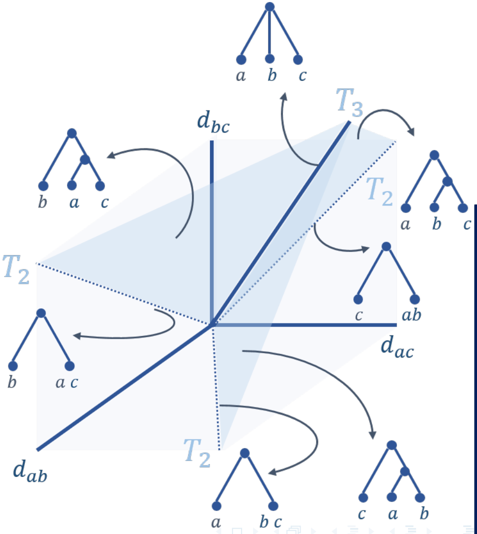


Space of Trees

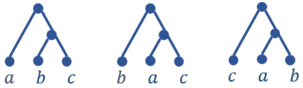


→ Parametrization

- global: coordinates d_{ab}, d_{ac}, d_{bc}
  → one cone for each combinatorial type

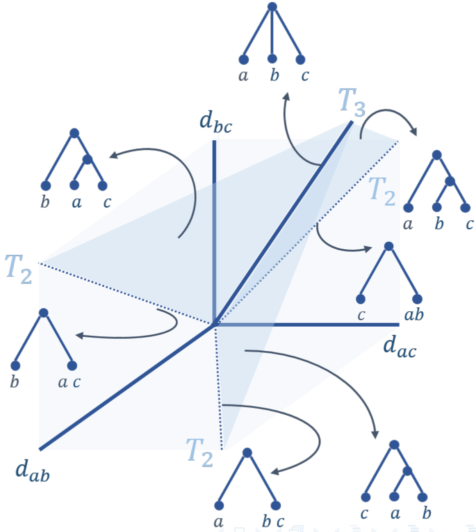


Space of Trees

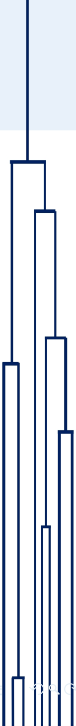


→ Parametrization

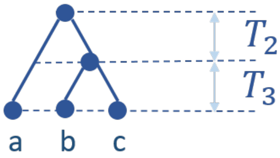
- global: coordinates d_{ab}, d_{ac}, d_{bc}
  → one cone for each combinatorial type
- local: coordinates T_2, T_3

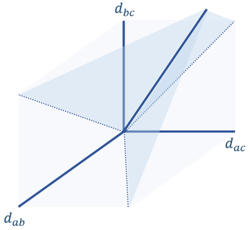


Contents



A Density





Proposition [Haase, W.]

The Kingman n-coalescent is given by the continuous density

$$\rho_n(T_2, \ldots, T_n) = \prod_{j=2}^{n} \binom{j}{2} \exp\left(-\binom{j}{2} T_j\right).$$

Contents

Connection to Tropical Geometry

Definition

$$\mathrm{MUM}_n := \{\, d \in \mathbb{R}_{\geq 0}^{\binom{n}{2}} : \forall i, j, k : \max\{d_{ij}, d_{ik}, d_{jk}\} \text{ is attained at least twice}\}$$

Metrics that are Ultrametrics

Connection to Tropical Geometry

Definition

$$\mathrm{DUM}_n := \{ d \in \mathbb{R}^{\binom{n}{2}} : \forall i, j, k : \max \{d_{ij}, d_{ik}, d_{jk}\} \text{ is attained at least twice}\}$$

Dissimilarity maps that are UltraMetrics

Connection to Tropical Geometry

Definition

$$\mathrm{DUM}_n := \{ d \in \mathbb{R}^{\binom{n}{2}} : \forall i, j, k : \max\{d_{ij}, d_{ik}, d_{jk}\} \text{ is attained at least twice}\}$$

Dissimilarity maps that are UltraMetrics

Definition

A *dissimilarity map* on $[n]$ is a map $\delta : [n] \times [n] \to \mathbb{R}$ such that $\delta(i,i) = 0$ for all $i \in [n]$, and $\delta(i,j) = \delta(j,i)$ for all $i, j \in [n]$.

Connection to Tropical Geometry

Definition

$$\mathrm{DUM}_n := \{\, d \in \mathbb{R}^{\binom{n}{2}} : \forall i,j,k : \max\{d_{ij}, d_{ik}, d_{jk}\} \text{ is attained at least twice} \}$$

Dissimilarity maps that are UltraMetrics

Definition

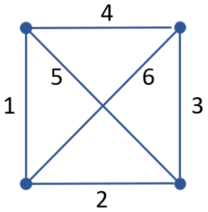
A *dissimilarity map* on $[n]$ is a map $\delta : [n] \times [n] \to \mathbb{R}$ such that $\delta(i,i) = 0$ for all $i \in [n]$, and $\delta(i,j) = \delta(j,i)$ for all $i,j \in [n]$.

Theorem [Ardila Klivans '06]

$$\mathrm{DUM}_n/\mathbb{R}1 = \mathrm{trop}(\mathrm{matroid}(K_n)) = \mathcal{B}(K_n)$$

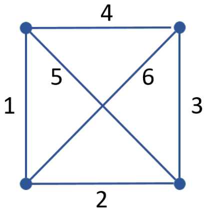
The Bergman fan of a graphical matroid

complete graph K_4



The Bergman fan of a graphical matroid

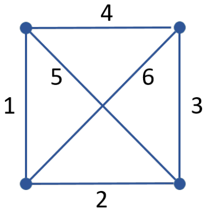
complete graph K_4

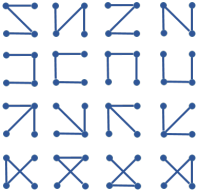


$$E = \{1, 2, 3, 4, 5, 6\}$$

The Bergman fan of a graphical matroid

complete graph K_4

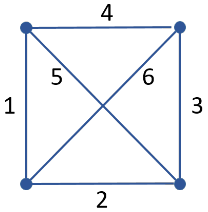




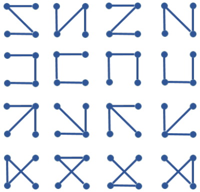
$E = \{1, 2, 3, 4, 5, 6\}$

The Bergman fan of a graphical matroid

complete graph K_4



$E = \{1, 2, 3, 4, 5, 6\}$

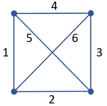


Bases of the graphical matroid
$M(K_4) = (E, B)$

$$B = \{\{2,4,5\}, \{1,3,6\}, \{2,4,6\}, \{1,3,5\}, \{2,3,4\}, \{1,2,4\}, \{1,3,4\}, \{1,2,3\}, \{3,4,6\}, \{2,3,5\}, \{1,4,5\}, \{1,5,6\}, \{4,5,6\}, \{2,5,6\}, \{3,5,6\}\}$$

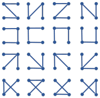
The Bergman fan of a graphical matroid

complete graph K_4 ⇝



$E = \{ 1,2,3,4,5,6\}$

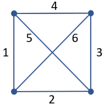
Graphical matroid
$M(K_4) = (E, B)$

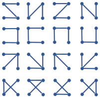


B = {{2,4,5}, {1,3,6}, {2,4,6}, {1,3,5}, {2,3,4}, {1,2,4}, {1,3,4}, {1,2,3}, {3,4,6}, {2,3,5}, {1,4,5}, {1,2,6}, {1,5,6}, {4,5,6}, {2,5,6}, {3,5,6}}

The Bergman fan of a graphical matroid

complete graph K_4 ⇝ Graphical matroid $M(K_4) = (E, B)$



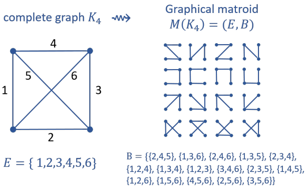


$E = \{1,2,3,4,5,6\}$

B = {{2,4,5}, {1,3,6}, {2,4,6}, {1,3,5}, {2,3,4}, {1,2,4}, {1,3,4}, {1,2,3}, {3,4,6}, {2,3,5}, {1,4,5}, {1,2,6}, {1,5,6}, {4,5,6}, {2,5,6}, {3,5,6}}

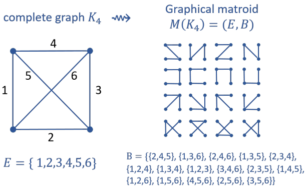
- assign weights $\omega \in \mathbb{R}^{\binom{n}{2}}$ to the edges

The Bergman fan of a graphical matroid



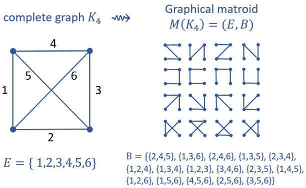
- assign weights $\omega \in \mathbb{R}^{\binom{n}{2}}$ to the edges
- for a basis $B = \{b_1, \dots, b_r\}$ the weight of the basis is given by $\omega_B = \omega_{b_1} + \dots + \omega_{b_r}$

The Bergman fan of a graphical matroid



- assign weights $\omega \in \mathbb{R}^{\binom{n}{2}}$ to the edges
- for a basis $B = \{b_1, \ldots, b_r\}$ the weight of the basis is given by $\omega_B = \omega_{b_1} + \ldots + \omega_{b_r}$
- $\rightarrow$ M_ω = collection of bases of M having minimum $\omega-$weight

The Bergman fan of a graphical matroid



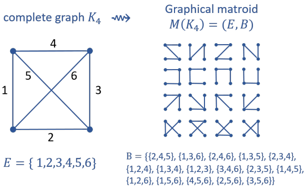
Definition

Bergman fan

$$\mathcal{B}(K_n) := \left\{ \omega \in \mathbb{R}^{\binom{n}{2}} \mid M_\omega \text{ has no loops} \right\}$$

- assign weights $\omega \in \mathbb{R}^{\binom{n}{2}}$ to the edges
- for a basis $B = \{b_1, \ldots, b_r\}$ the weight of the basis is given by $\omega_B = \omega_{b_1} + \ldots + \omega_{b_r}$
- $\rightarrow$ M_ω = collection of bases of M having minimum $\omega-$weight

THE BERGMAN FAN OF A GRAPHICAL MATROID



DEFINITION

Bergman fan

$$\mathcal{B}(K_n) := \left\{ \omega \in \mathbb{R}^{\binom{n}{2}} \mid M_\omega \text{ has no loops} \right\}$$

- assign weights $\omega \in \mathbb{R}^{\binom{n}{2}}$ to the edges
- for a basis $B = \{b_1, \ldots, b_r\}$ the weight of the basis is given by $\omega_B = \omega_{b_1} + \ldots + \omega_{b_r}$
- $\rightarrow$ $M_\omega =$ collection of bases of M having minimum $\omega-$weight

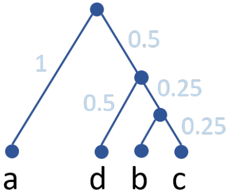
THEOREM [ARDILA KLIVANS '06]

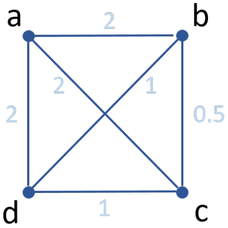
A dissimilarity map is an ultrametric if and only if the corresponding weight vector on the edges of K_n is in the Bergman fan $\mathcal{B}(K_n)$.

The Bergman fan of a graphical matroid

Theorem [Ardila Klivans '06]

A dissimilarity map is an ultrametric if and only if the corresponding weight vector on the edges of K_n is in the Bergman fan $\mathcal{B}(K_n)$.





Space of Trees

Definition

$$\mathrm{DTM}_n := \{ d \in \mathbb{R}^{\binom{n}{2}} : \forall i, j, k, l : \max\{d_{ij} + d_{kl}, d_{ik} + d_{jl}, d_{il} + d_{jk}\} \text{ is attained at least twice} \}$$

Dissimilarity maps that are TreeMetrics

Space of Trees

Definition

$$\mathrm{DTM}_n := \{ d \in \mathbb{R}^{\binom{n}{2}} : \forall i,j,k,l : \max\{d_{ij}+d_{kl}, d_{ik}+d_{jl}, d_{il}+d_{jk}\} \text{ is attained at least twice}\}$$

Dissimilarity maps that are TreeMetrics

Space of Trees

Definition

$$\mathrm{DTM}_n := \{ d \in \mathbb{R}^{\binom{n}{2}} : \forall i, j, k, l : \max\{d_{ij} + d_{kl}, d_{ik} + d_{jl}, d_{il} + d_{jk}\} \text{ is attained at least twice} \}$$

Dissimilarity maps that are TreeMetrics

Theorem [Speyer Sturmfels '03]

$$\mathrm{DTM}_n / \mathbb{R}1 = \mathrm{trop}(\mathrm{Gr}(2, n))$$

Contents

Further Directions

- fan with density in $\mathbb{R}^{\binom{\mathbb{N}}{2}}$?
- fan for species trees?
- include mutations?
- consider non-binary trees?

Thank you for your attention!