The total external length in the evolving Kingman coalescent

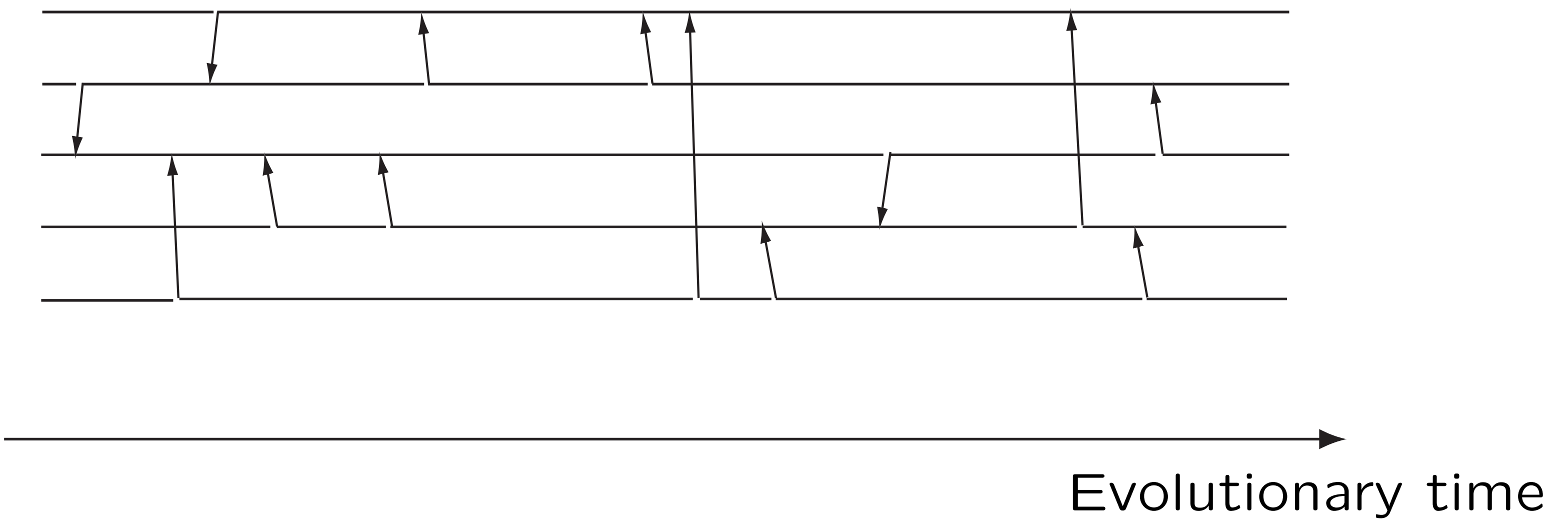
Götz Kersting and Iulia Stanciu

Goethe Universität, Frankfurt am Main

CIRM, Luminy
June 11–15 2012

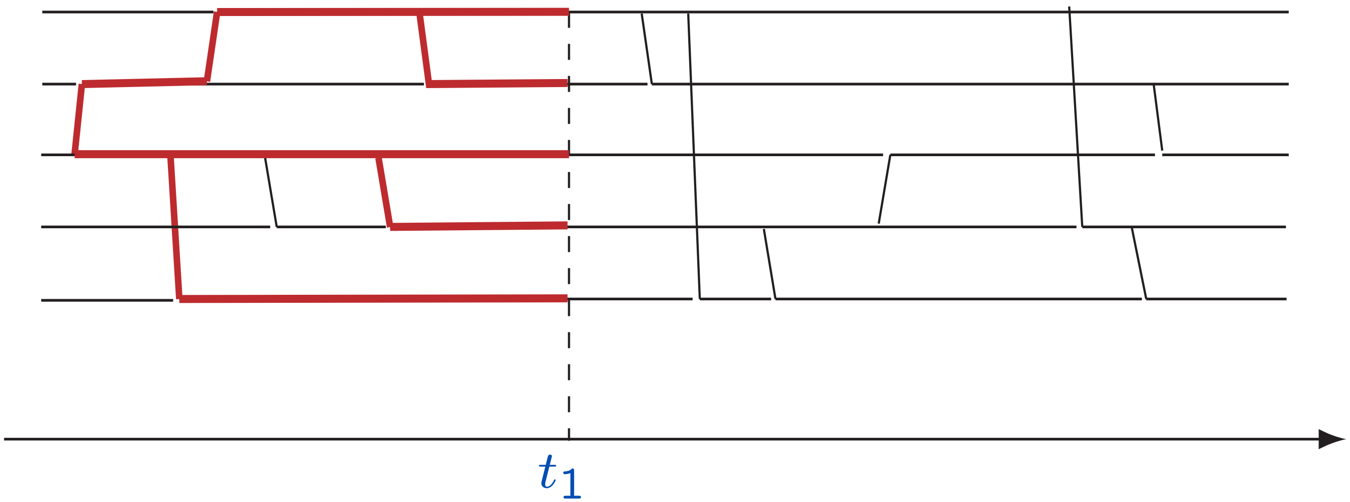
GOETHE UNIVERSITÄT FRANKFURT AM MAIN

The evolving Kingman N-coalescent ($N = 5$):



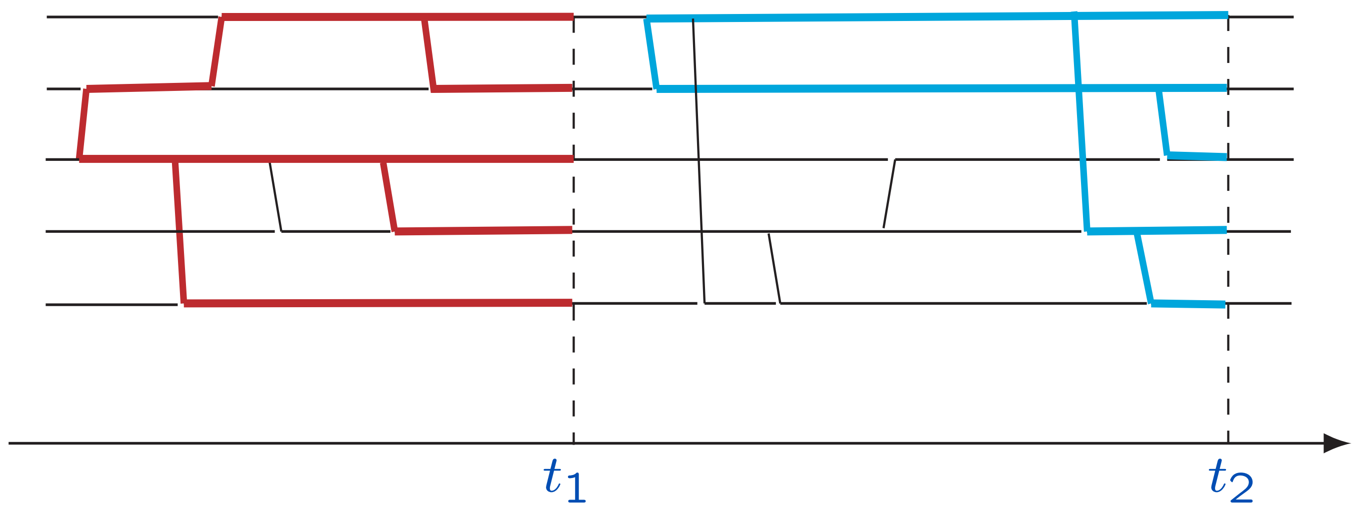
Moran's model with time $-\infty < t < \infty$:
Links between pairs of lines appear at rate 1,
independent between the different pairs.

The evolving Kingman N-coalescent:



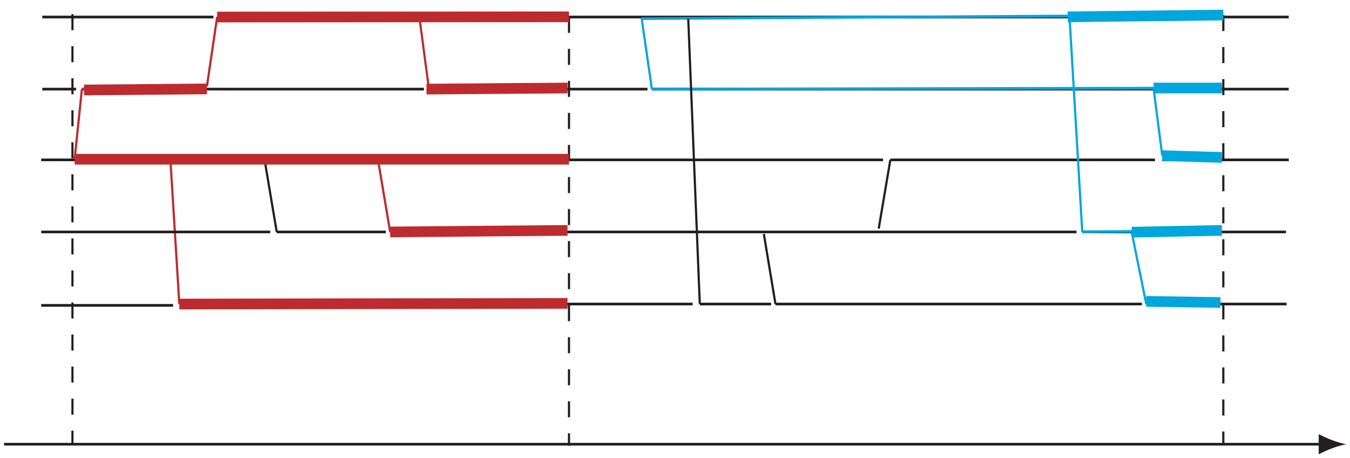
Kingman's coalescent at time t_1

The evolving Kingman N-coalescent:



The coalescent tree evolves in time.

The evolving Kingman N-coalescent:



Evolving time to MRCA

tree topology

total length

total external length

These results are rather different in nature, none covering any other.

Interesting aspects:

– limiting processes with
a.s. continuous paths versus
a.s. (compensated) pure jump paths

– different time scalings

A BASKET OF RESULTS

Theorem: (Pfaffelhuber/Wakolbinger , Donnelly/Kurtz, 2006)

Let $A_N(t)$ be the time to the MRCA of the evolving Kingman N-coalescent at time $t \in \mathbb{R}$. Then, as $N \to \infty$,

$$\Big(A_N(t)\Big)_{t\in\mathbb{R}} \quad \xrightarrow{d} \quad A \ ,$$

where the limiting process $A = (A_t)_{t\in\mathbb{R}}$ is stationary, a.s. pure jump, non-Markovian.

For a related result on the two oldest families in the genealogy see Delmas, Dhersin, and Siri-Jegousse (2010).

Theorem: (Greven, Pfaffelhuber, Winter, 2009, 2010)

Let $T_N(t)$ be the tree, induced by the evolving Kingman N-coalescent at time $t \in \mathbb{R}$ in the space of real trees furnished with the Gromov-weak topology. Then, as $N \to \infty$,

$$\big(T_N(t)\big)_{t\in\mathbb{R}} \quad \xrightarrow{d} \quad T \ ,$$

where the limiting tree-valued process $T = (T_t)_{t\in\mathbb{R}}$ is stationary, a.s. continuous, and unique solution of a martingale problem.

Theorem: (Pfaffelhuber, Wakolbinger, Weisshaupt, 2011)

Let $L'_N(t)$ be the total length of the evolving Kingman N-coalescent at time $t \in \mathbb{R}$. Then, as $N \to \infty$,

$$\Big(L'_N(t) - 2\log N\Big)_{t\in\mathbb{R}} \quad \stackrel{d}{\to} \quad L' \ ,$$

where the limiting process $L' = (L'_t)_{t\in\mathbb{R}}$ is stationary, a.s. pure jump, non-Markovian.

L' has *infinite quadratic variation* (Knobloch, Stanciu, Wakolbinger, 2011), thus fails to be a semimartingale.

Theorem: (Schweinsberg, 2011)

Let $L''_N(t)$ be the total length of the evolving *Bolthausen-Sznitman* N-coalescent at time $t \in \mathbb{R}$. Then, as $N \to \infty$,

$$\Big(\frac{(\log N)^2}{N} L''_N(\tfrac{t}{\log N}) - \log N - \log\log N\Big)_{t \in \mathbb{R}} \quad \xrightarrow{d} \quad L'' \ ,$$

where the stationary limiting process $L'' = (L''_t)_{t \in \mathbb{R}}$ solves the SDE

$$dL'' = -L''\, dt + dY$$

for a certain Lévy-process Y of index 1.

Theorem: (K., Stanciu, 2012, ongoing work)

Let $L_N(t)$ be the total *external length* of the evolving Kingman N-coalescent. Then

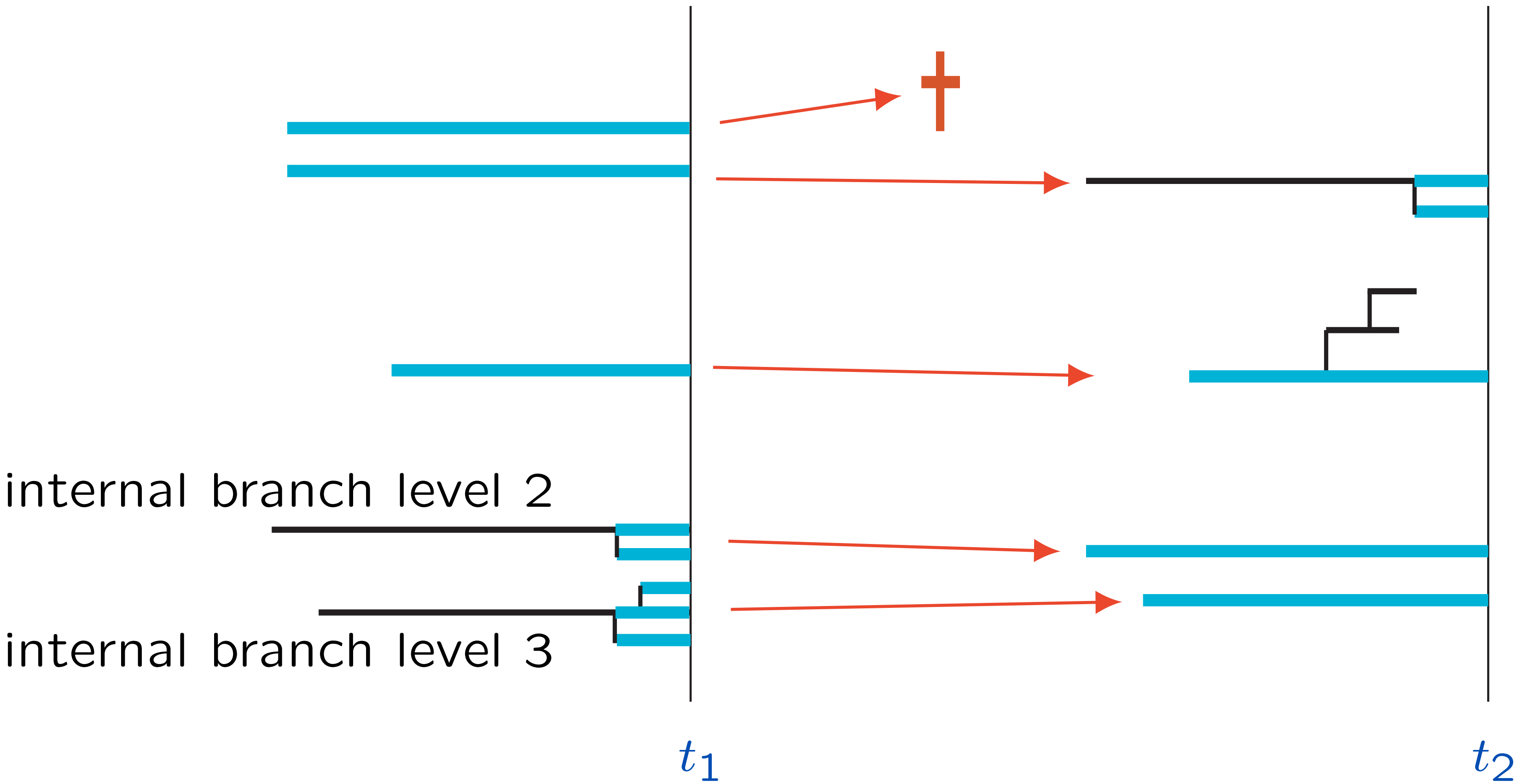
$$\left(\sqrt{\frac{N}{4\log N}}\Big(L_N(\tfrac{t}{N}) - 2\Big)\right)_{t\in\mathbb{R}} \quad \xrightarrow{d} \quad L \ ,$$

where L is a stationary, Gaussian, a.s. continuous, with covariance function

$$\mathbf{Cov}(L_s, L_t) = \Big(\frac{1}{1+|t-s|}\Big)^2 \ .$$

Note the different scaling of time (*real* instead of *evolutionary* periodes).

The dynamics of the external lengths:



So let for the *static* Kingman N-coalescent (at time $t = 0$)

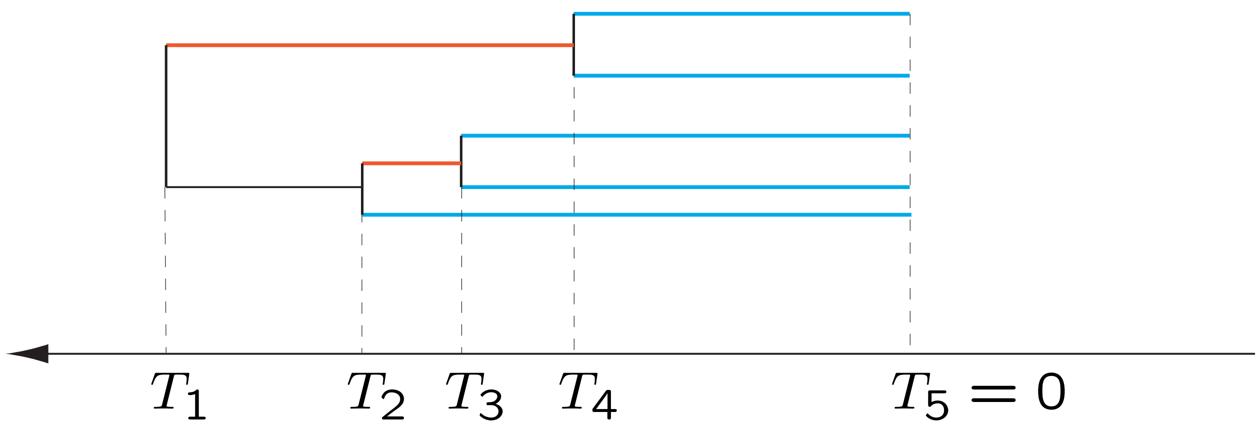
$$L_N^i = \text{ total internal branch length of level } i \ ,$$

in particular for $i = 1$

$$L_N^1 = \text{ total external branch length } .$$

How do we gain access to these quantities?

Branch numbers $V_N, \ldots, V_2$ *and* $W_N, \ldots, W_2$.



$V_2 = 0$ $\qquad$ $V_5 = 5$

$W_2 = 1$ $\qquad$ $W_5 = 0$

$$L_N^1 = \sum_{i=2}^{N} V_i (T_{i-1} - T_i) \ , \qquad L_N^2 = \sum_{i=2}^{N} W_i (T_{i-1} - T_i)$$

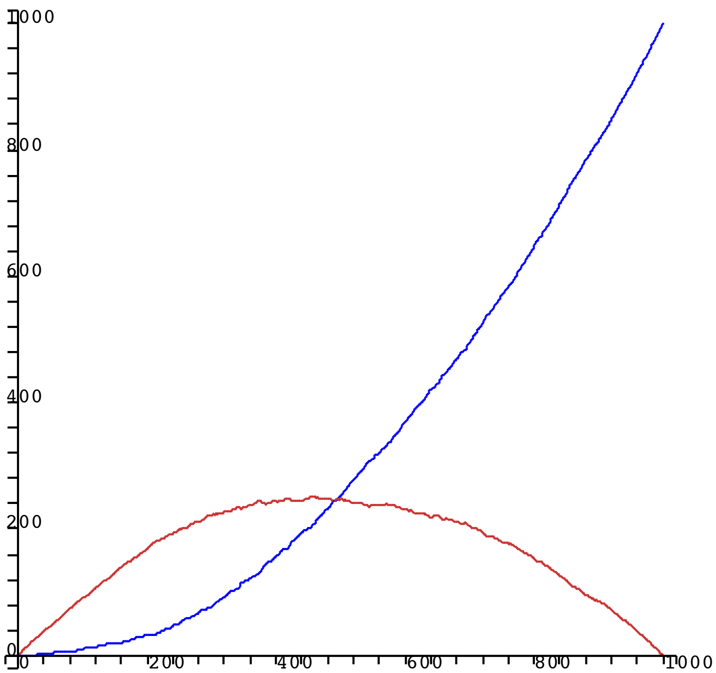
$$L_N^1 = \sum_{i=2}^{N} V_i(T_{i-1} - T_i)$$

$$= \sum_{k=1}^{N-1} T_k(V_{k+1} - V_k) + T_1 V_2$$

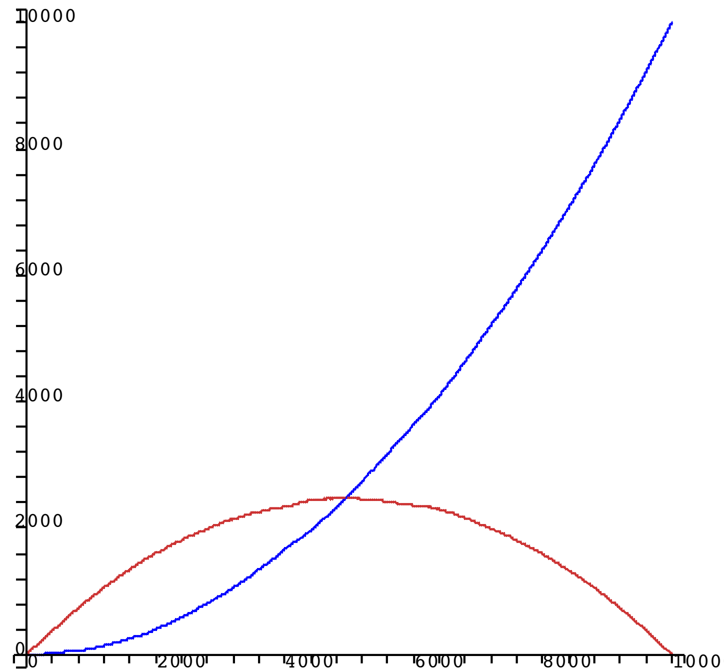
$$\approx \sum_{k=1}^{N-1} \frac{2}{k} \Delta V_k$$

ΔV_k is easy to handle for k close to N.

External branch numbers $0 = V_1, V_2, \ldots, V_N$, and (total) internal branch numbers $U_1 = 1 - V_1, U_2 = 2 - V_2, \ldots, U_N = 1 - V_N$



$N = 1000$



$N = 10000$

Let us look at the randomness within

$$V_N - N \quad (= 0)$$

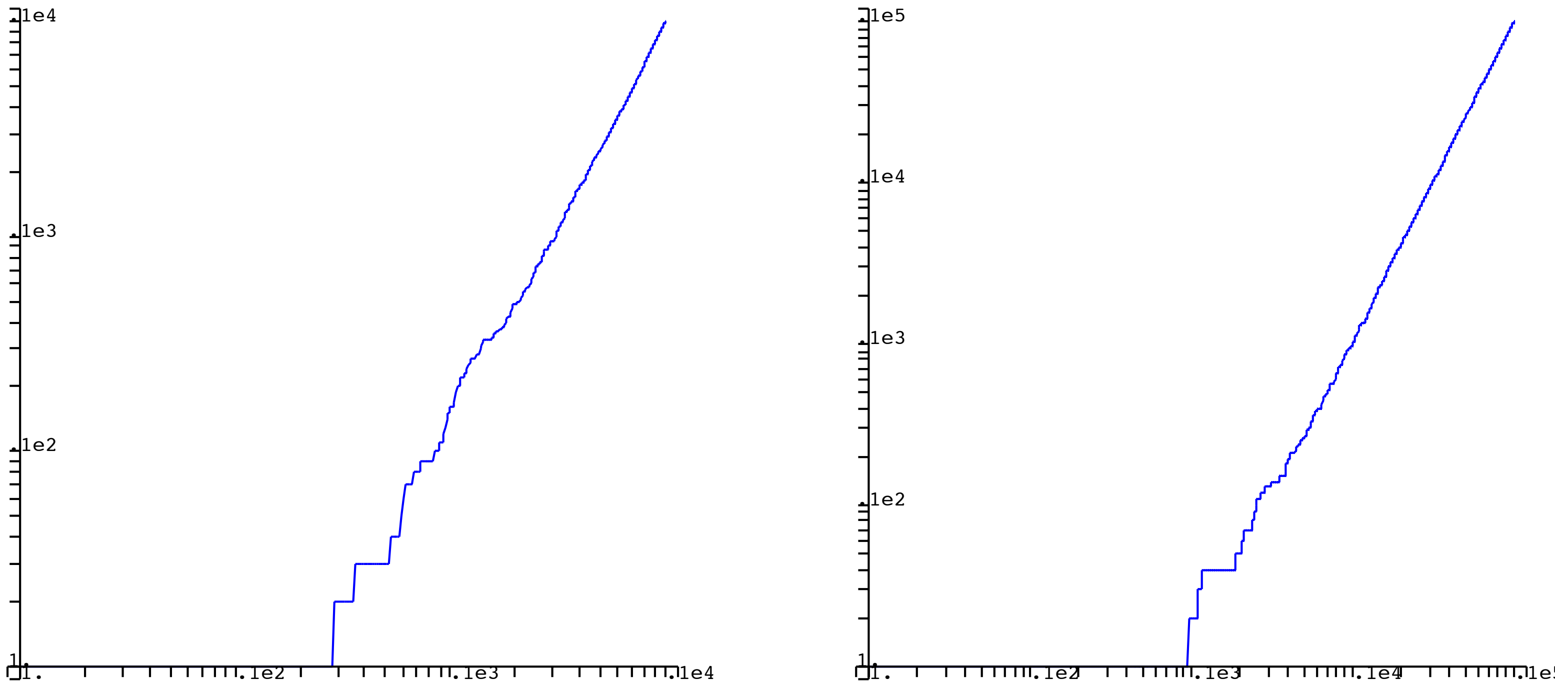
$$V_{N-1} - N + 2$$

$$\vdots$$

$$V_{N-i} - N + 2i$$

$$\vdots$$

Randomness within (V_i) close to N:
$V_N - N + 1, V_{N-1} - N + 3, \ldots$



N^0 $N^{1/2}$ N

Randomness enters (almost) independently.

However, since we consider

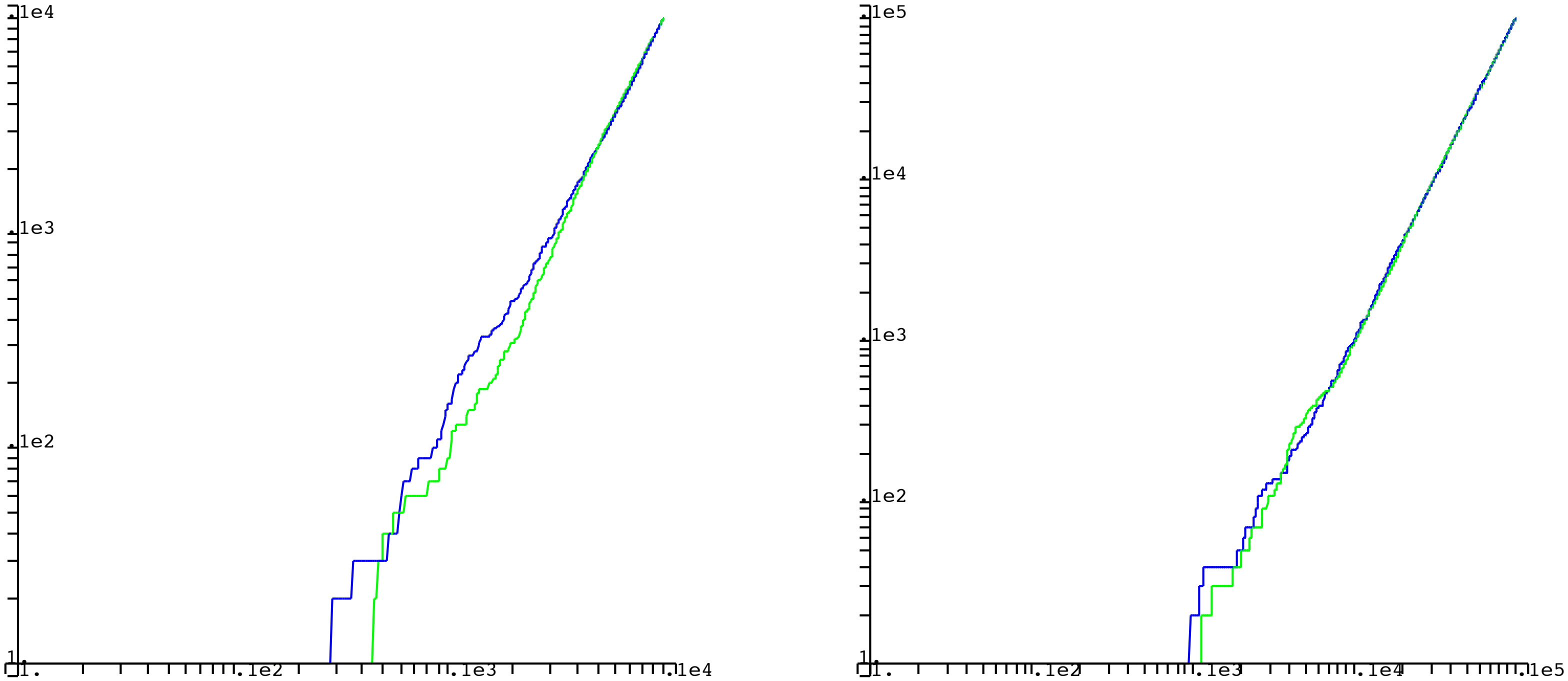
$$L_N^1 \approx \sum_{k=1}^{N-1} \frac{2}{k} \Delta V_k$$

we have to understand randomness at the beginning of $V_2, \ldots, V_N$. So let us compare

$$V_N - N, V_{N-1} - N + 2, \ldots \qquad \text{to} \qquad V_1, V_2, \ldots$$

Randomness within (V_i) close to N and 1:

$V_N - N + 1, V_{N-1} - N + 3, \ldots$ versus $V_1 + 1, V_2 + 1, \ldots$



For the (total) internal numbers $U_1 = 1 - V_1, \ldots, U_N = N - V_N$ we have reversibility:

Theorem: (Janson, K. 2011)

$$(U_1, \ldots, U_{N-1}) \stackrel{d}{=} (U_{N-1}, \ldots, U_1) \ .$$

For another proof see Knobloch, Stanciu, Wakolbinger (2011).

Theorem: (Janson, K. 2011)

$$\sqrt{\frac{N}{4 \log N}} \left(L_N^1 - 2\right) \xrightarrow{d} N(0, 1) \ .$$

The representation of the (total) internal numbers $U_1, \ldots, U_N$ *as diminishing urn:*

– Take urn with *blue* balls, altogether N balls.

– Remove them stepwise:
Successively remove a random pair of balls
and replace it by one *orange* ball.

– If i balls are left,
let U_i the number of orange balls among them
and V_i the number of blue balls.

Note:

$$V_{N-i} - N + 2i = i - U_{N-i} \quad , \quad V_i = i - U_i \ .$$

Now recall

$V_N, \ldots, V_2$ external branch numbers

$W_N, \ldots, W_2$ internal branch numbers level 2

$\vdots$

Note:

$V_N, V_{N-1}, \ldots, V_2$ is a *Markov chain* (inhomogeneous in time).

$(V_N, W_N), (V_{N-1}, W_{N-1}), \cdots, (V_2, W_2)$ is a Markov chain, or

$W_N, \ldots, W_2$ is a Markov chain,
given the *random environment* $V_N, \ldots, V_2$.

The transition probabilities:

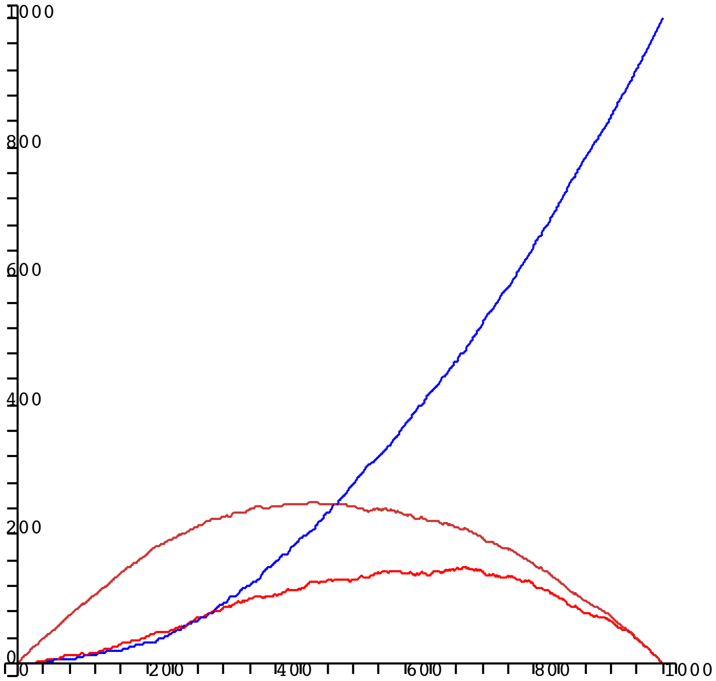
$$P^k_{v,w}(v', w') = \mathbf{P}(V_{k-1} = v', W_{k-1} = w' \mid V_k = v, W_k = w)$$

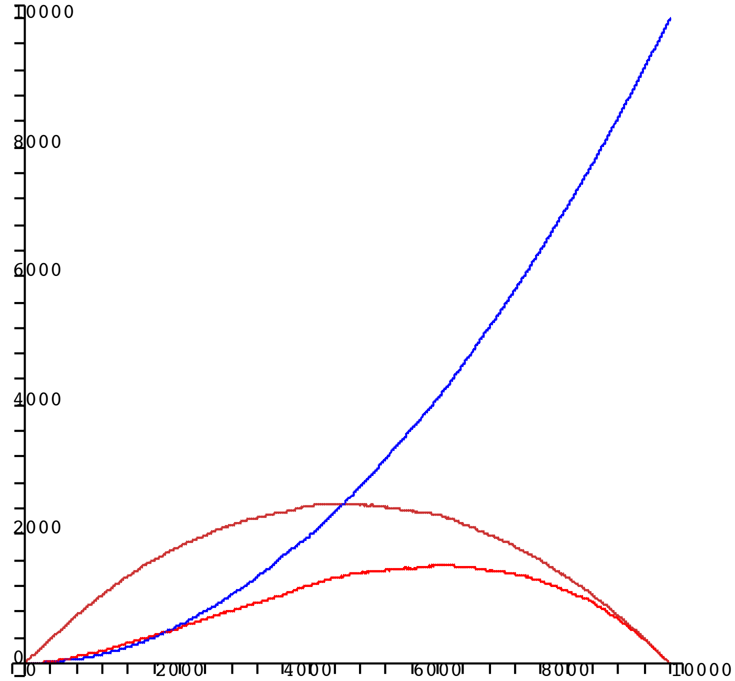
$$P^k_{v,w}(v-2, w+1) = \frac{\binom{v}{2}}{\binom{k}{2}} \qquad P^k_{v,w}(v, w-2) = \frac{\binom{w}{2}}{\binom{k}{2}}$$

$$P^k_{v,w}(v, w) = \frac{\binom{k-v-w}{2}}{\binom{k}{2}} \qquad P^k_{v,w}(v-1, w-1) = \frac{vw}{\binom{k}{2}}$$

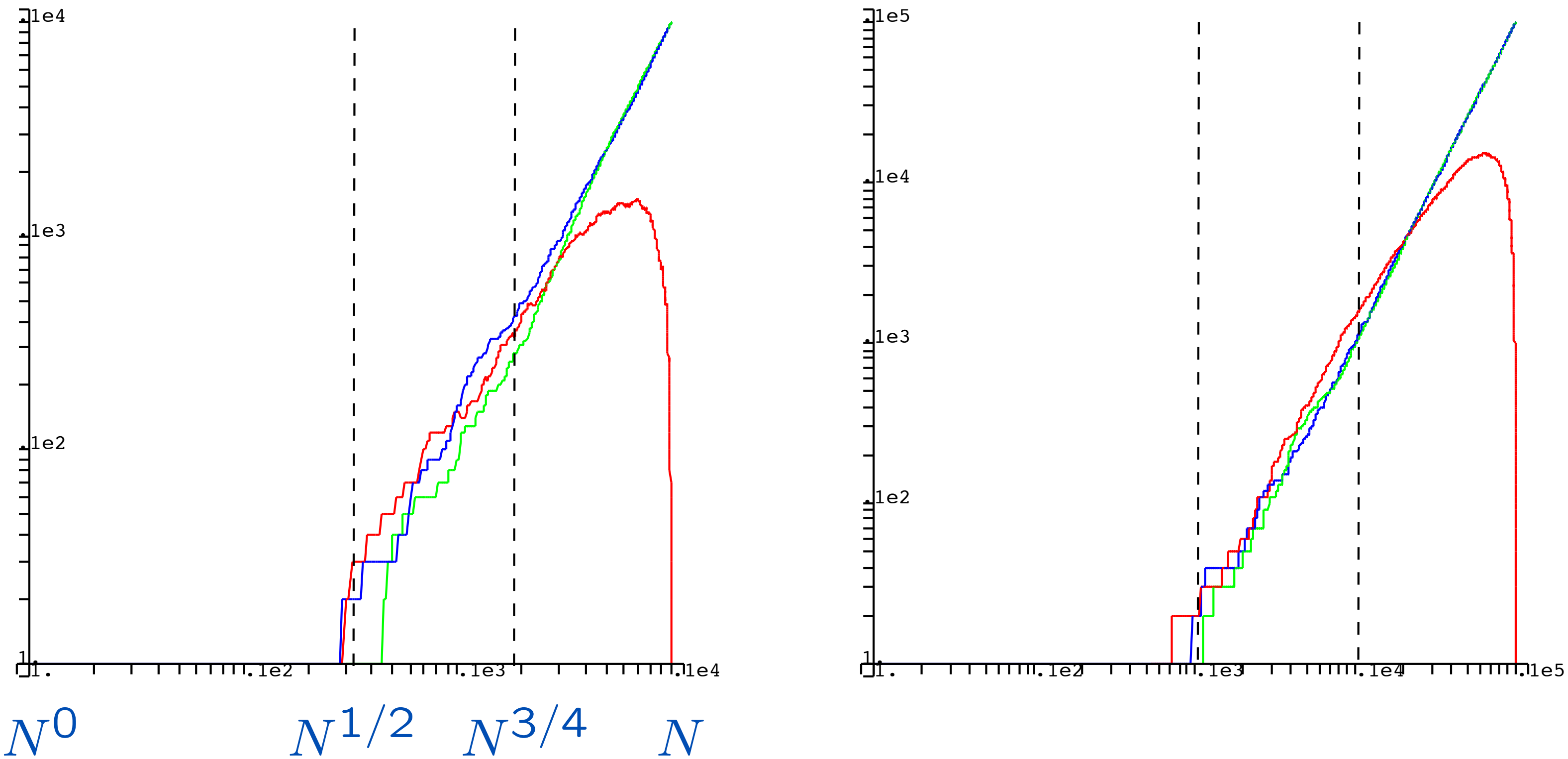
$$P^k_{v,w}(v-1, w) = \frac{v(k-v-w)}{\binom{k}{2}} \qquad P^k_{v,w}(v, w-1) = \frac{w(k-v-w)}{\binom{k}{2}}$$

External branch numbers $V_1, \ldots, V_N$, and
(total) internal branch numbers $U_1, \ldots, U_N$ and
internal branch numbers level 2 $W_1, \ldots, W_N$.





Randomness within (V_i) and (W_i):
$V_N - N + 1, V_{N-1} - N + 3, \ldots$ and $V_1 + 1, V_2 + 1, \ldots$
and $W_1 + 1, W_2 + 1, \ldots$



Theorem: (K., Stanciu 2012)

For every $k \in \mathbb{N}$

$$\sqrt{\frac{N}{4\log N}}\,\left(L_N^1 - \mu_1, \ldots, L_N^k - \mu_k\right) \xrightarrow{d} N(0, I_k) \; .$$

with the $k \times k$ identity matrix I_k and

$$\mu_i = \frac{2}{i} \; .$$

Idea of proof:

Reversing time seems no longer feasible.

We couple the Markov chain

$$(V_N, W_N), \dots, (V_2, W_2)$$

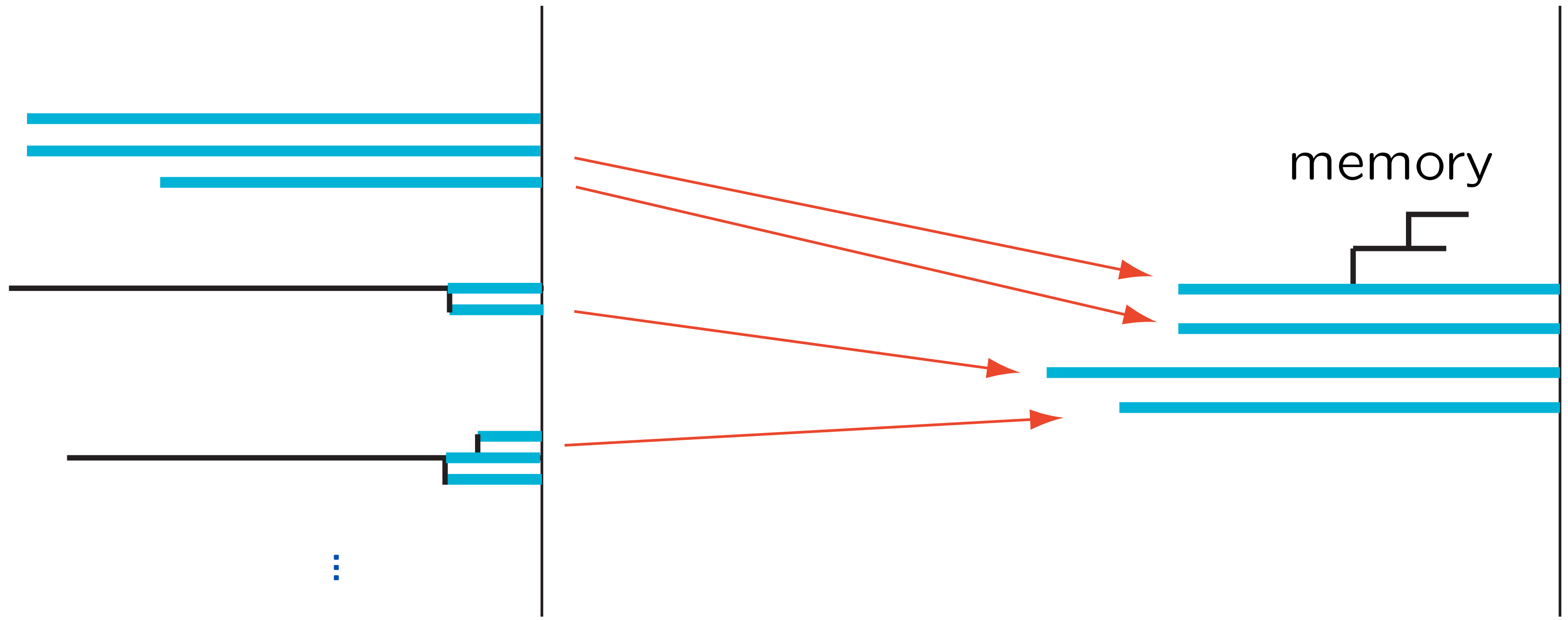
with two independent urns, i.e. with

$$(V_N, \tilde{V}_N), \dots, (V_2, \tilde{V}_2)\ ,$$

where $(\tilde{V}_N, \dots, \tilde{V}_2)$ is an independent copy of $(V_N, \dots, V_2)$.

Now the urns can be reversed.

Back to the evolving coalescent:



$\mathbf{Cov}(L_s, L_t) =$
probability that a critical binary branching process consists of exactly 1 individual at time $|t - s|$.


BARRY
8
22