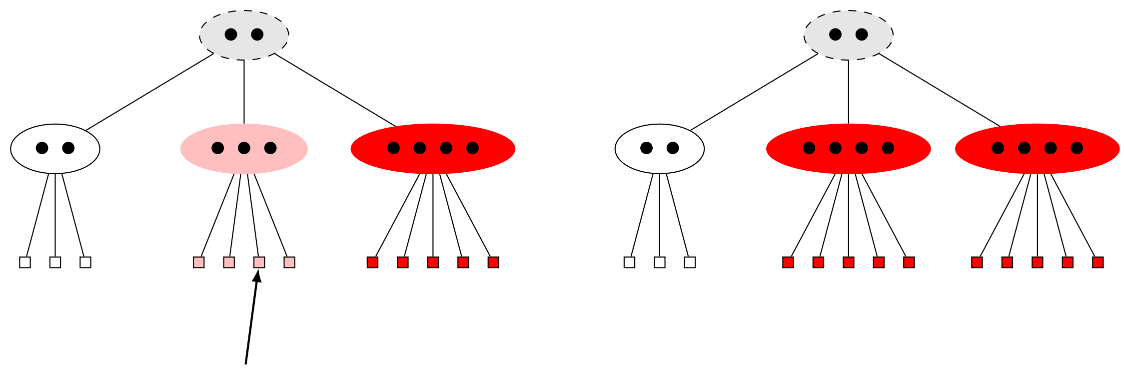


From B-trees to large Pólya urns

Brigitte CHAUVIN

University of Versailles St-Quentin, LMV

Joint work with Danièle GARDY
Nicolas POUYANNE
Dai-Hai TON-THAT

Bath, Branching Structures, June, 9-11, 2014

A continuous-time branching process

$(X(t))_{t\geq 0}$ is a continuous-time multitype branching process, with m types (colors), $m \geq 2$

Any particle of type j is equipped with a $\mathcal{E}xp(m+j-1)$ clock.

When a clock rings,

$\rightarrow$ one particle of type $j+1$, if $j = 1, \dots, m-1$

$\rightarrow$ two particles of type 1, if $j = m$.

$0 < \tau_1 < \dots < \tau_n < \dots$ are the jump times

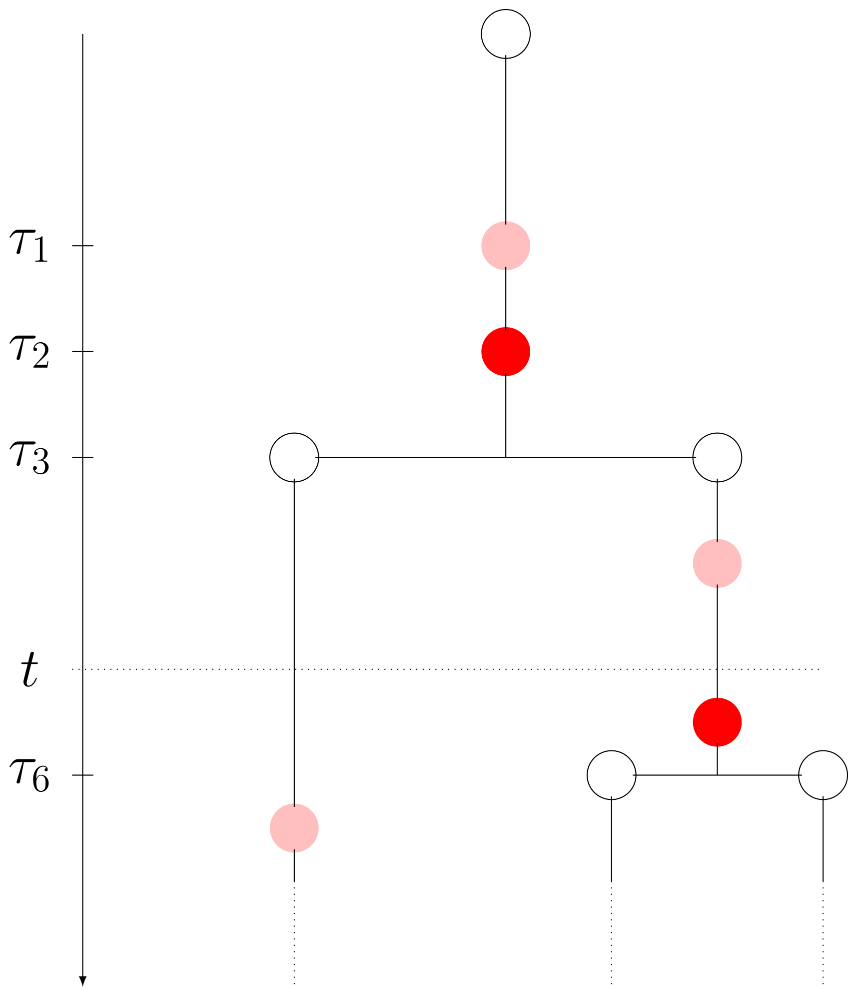
Composition vector

$$X(t) = \begin{pmatrix} X(t)^{(1)} \\ \vdots \\ X(t)^{(m)} \end{pmatrix}$$

$X(t)^{(j)} = \#$ particles of type j alive at time t.

$\boxed{m = 3}$

$type\ \ 1 = \ white$
$type\ \ 2 = \ pink$
$type\ \ 3 = \ red$



$$X(t) = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$$

$$e^{-tA_m}X(t) \text{ is a vector-valued martingale.}$$

where A_m is the transition matrix:

$$A_m = \begin{pmatrix} -m & & & & & 2(2m-1) \\ m & -(m+1) & & & & \\ & m+1 & -(m+2) & & & \\ & & \ddots & \ddots & & \\ & & & \ddots & -(2m-2) & \\ & & & & 2m-2 & -(2m-1) \end{pmatrix}.$$

The eigenvalues are simple, solution of the

Characteristic equation: $\boxed{(\lambda+m)\ldots(\lambda+2m-1) = \dfrac{(2m)!}{m!}}$

$\lambda_1 = 1$

$\lambda_2, \overline{\lambda_2} =$ conjugate eigenvalues with the greatest real part < 1.

$$\sigma_2 := \Re(\lambda_2).$$

2 different asymptotics:

- Gaussian: when $m \le 59$, then $\sigma_2 < \frac{1}{2}$ and

$$X(t) = e^t \xi v_1 (1 + o(1)) + e^{t/2} \sqrt{\xi} G (1 + o(1))$$

where v_1 is an eigenvector for 1;

ξ is Gamma-distributed

G is a Gaussian vector, independant of ξ

the convergence in the first term holds a.s. and in L^p

the convergence in the second term holds in distribution.

- non Gaussian: when $m \ge 60$, then $\frac{1}{2} < \sigma_2 < 1$ and

$$\boxed{X(t) = e^t \xi v_1 (1 + o(1)) + \Re\big(e^{\lambda_2 t} W^{CT} v_2\big)(1 + o(1)) + o(e^{\sigma_2 t})}$$

where

v_1, v_2 are deterministic vectors

W^{CT} is a $\mathbb{C}$-valued martingale's limit

$o(\ )$ means a.s. and in any $L^p, p \ge 1$

Goal: distribution of W^{CT}

Where does the asymptotic expansion come from?

A_m = transition matrix associated with $X(t)$

$$e^{-tA_m}X(t) \text{ is a vector-valued martingale.}$$

Spectral decomposition:

$$X(t) = u_1(X(t))v_1 + u_2(X(t))v_2 + \overline{u_2}(X(t))\overline{v_2} + \dots$$

where

u_1 = coordinate of the first projection along v_1

u_2 = second v_2.

$$(e^{-t}u_1(X(t)))_{t\geq 0} \quad \text{and} \quad (e^{-\lambda_2 t}u_2(X(t)))_{t\geq 0}$$

are continuous-time L^p-bounded martingales.

Notation:

$$\xi := \lim_{t\to+\infty} e^{-t}u_1(X(t)) \qquad W^{CT} := \lim_{t\to+\infty} e^{-\lambda_2 t}u_2(X(t))$$

Summarizing:

$$\xi := \lim_{t \to +\infty} e^{-t} u_1(X(t)) \qquad\qquad W^{CT} := \lim_{t \to +\infty} e^{-\lambda_2 t} u_2(X(t))$$

Theorem ($\sim$ Janson, CLP): for $m \geq 60$,

$$\boxed{X(t) = e^t \xi v_1 (1 + o(1)) + \Re\big(e^{\lambda_2 t} W^{CT} v_2\big)(1 + o(1)) + o(e^{\sigma_2 t})}$$

where

v_1, v_2 are deterministic vectors
W^{CT} is a $\mathbb{C}$-valued martingale's limit
$o(\ )$ means a.s. and in any $L^p, p \geq 1$.

Dislocation equations - branching property

Adopt the notations:

$X_k(t) := X(t)$ starting from $(0, \dots, 1, \dots 0)$, one particle of type k.
$W_k := W^{CT}$ starting from $(0, \dots, 1, \dots 0)$, one particle of type k.

$\forall t > \tau$ (the first splitting time),

$$\begin{cases} X_1(t) \overset{\mathcal{L}}{=} X_2(t - \tau^{(1)}) \\ X_2(t) \overset{\mathcal{L}}{=} X_3(t - \tau^{(2)}) \\ \dots \\ X_{m-1}(t) \overset{\mathcal{L}}{=} X_m(t - \tau^{(m-1)}) \\ X_m(t) \overset{\mathcal{L}}{=} X_1^{(1)}(t - \tau^{(m)}) + X_1^{(2)}(t - \tau^{(m)}) \end{cases}$$

where $\tau^{(k)}$ is $\mathcal{E}xp(m + k - 1)$ distributed; $X_1^{(1)}$ and $X_1^{(2)}$ are independent copies of X_1, independent of $\tau^{(m)}$.

Take projections, renormalize, take the limit, $t \to +\infty$:

$$\tau^{(k)} \overset{\mathcal{L}}{=} \mathcal{E}xp(m+k-1)$$

$$\begin{cases} W_1 \overset{\mathcal{L}}{=} e^{-\lambda_2 \tau^{(1)}} W_2 \\ W_2 \overset{\mathcal{L}}{=} e^{-\lambda_2 \tau^{(2)}} W_3 \\ \dots \\ W_{m-1} \overset{\mathcal{L}}{=} e^{-\lambda_2 \tau^{(m-1)}} W_m \\ W_m \overset{\mathcal{L}}{=} e^{-\lambda_2 \tau^{(m)}} \left(W_1^{(1)} + W_1^{(2)} \right) \end{cases}$$

which gives a fixed point equation on W_1 only:

$$W_1 \overset{\mathcal{L}}{=} e^{-\lambda_2 (\tau^{(1)} + \dots + \tau^{(m)})} \left(W_1^{(1)} + W_1^{(2)} \right).$$

$$\boxed{W_1 \overset{\mathcal{L}}{=} V^{\lambda_2} \left(W_1^{(1)} + W_1^{(2)} \right)}$$ where $V = e^{-(\tau^{(1)} + \dots + \tau^{(m)})} \overset{\mathcal{L}}{=} Beta(m, m)$

More on the smoothing equation:

$$\boxed{W \overset{\mathcal{L}}{=} V^{\lambda_2} \left(W^{(1)} + W^{(2)}\right)} \qquad \text{where } V \overset{\mathcal{L}}{=} Beta(m,m)$$

- Th 1: existence and unicity (contraction method)
- Th 2: existence and exponential moments (cascade)
- Th 3: support and density % Lebesgue measure on $\mathbb{C}$ (Liu's method)

$\leftarrow$ discrete analogous process?

B-tree

$X_n = \begin{pmatrix} X_n^{(1)} \\ \vdots \\ X_n^{(m)} \end{pmatrix}$ counts the number of final internal nodes in a B-tree of size n.

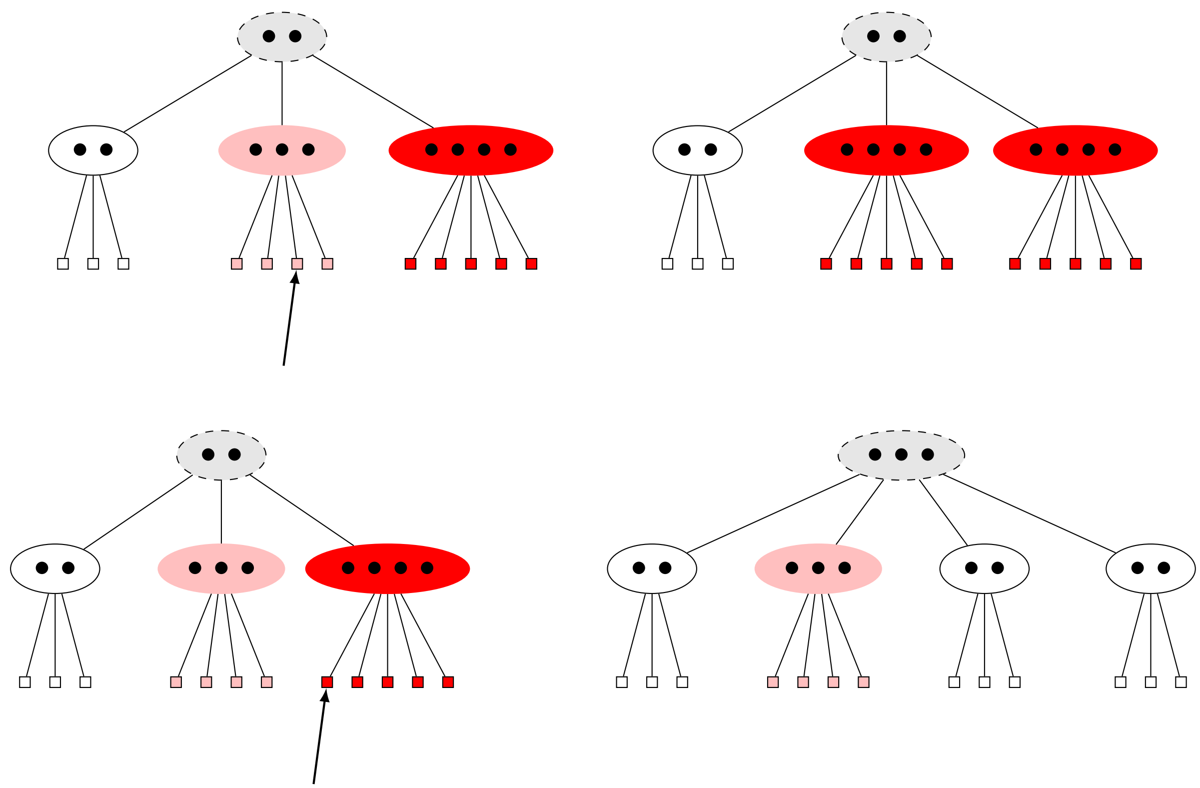
A B-tree with parameter m is a search tree where

- all the leaves are at the same level
- in a node: between $m-1$ and $2m-2$ keys
- insertion of a new key occurs at a leaf

(and the leaves are equally likely)

$\boxed{m = 3}$ $\mathcal{T}_{11} \longrightarrow \mathcal{T}_{12}$

$2 \leq \#$ keys ≤ 4
3 colors: white, pink, red



$$R_3 = \begin{pmatrix} -3 & 4 & \\ & -4 & 5 \\ 6 & & -5 \end{pmatrix}.$$

B-tree

$X_n = \begin{pmatrix} X_n^{(1)} \\ \vdots \\ X_n^{(m)} \end{pmatrix}$ counts the number of final internal nodes in a B-tree of size n.

$G_n = \begin{pmatrix} G_n^{(1)} \\ \vdots \\ G_n^{(m)} \end{pmatrix}$ counts the number of different gaps in a B-tree of size n. $\boxed{G_n = DX_n}$

$$D = \begin{pmatrix} m & & & \\ & m+1 & & \\ & & \ddots & \\ & & & 2m-1 \end{pmatrix}$$

(G_n) is a Pólya urn process, with replacement matrix

$$R_m = \begin{pmatrix} -m & m+1 & & & \\ & -(m+1) & m+2 & & \\ & & \ddots & & \\ & & & -(2m-2) & 2m-1 \\ 2m & & & & -(2m-1) \end{pmatrix}.$$

Asymptotics of the gap process (G_n)

Asymptotics of the node process (X_n)

<u>Theorem</u> ($\sim$ Janson):

- Gaussian if $m \leq 59$,

$$\frac{G_n - nv_1}{\sqrt{n}} \xrightarrow[n\to\infty]{\mathcal{D}} G$$

- non Gaussian: when $m \geq 60$

$$\boxed{G_n = nv_1 + 2\Re\left(n^{\lambda_2} W^{DT} v_2\right) + o\left(n^{\sigma_2}\right)} \qquad \boxed{G_n = DX_n}$$

where

$\rightarrow$ v_1, v_2 are deterministic vectors

$\rightarrow$ W^{DT} is a $\mathbb{C}$-valued martingale's limit $\qquad W^{DT} = \lim_{n\to\infty} \frac{1}{n^{\lambda_2}} u_2\left(G_n\right).$

$\rightarrow$ $o(\ )$ means a.s. and in any $L^p, p \geq 1$.

Goal: distribution of W^{DT}

Connexion DT-CT

$$\boxed{W^{CT} \overset{\mathcal{L}}{=} \xi^{\lambda_2} W^{DT}}$$

where ξ is Gamma-distributed.

W^{DT} is a solution of the fixed point distributional equation

$$\boxed{W \overset{\mathcal{L}}{=} V_1^{\lambda_2} W^{(1)} + V_2^{\lambda_2} W^{(2)}} \qquad \text{where } (V_1, V_2) \overset{\mathcal{L}}{=} Dirichlet(m, m)$$

Two proofs:

- connexion and moments
- tree structure and original Pólya urn argument

(ask to Cécile Mailler)

****

Now the last part (C)

Properties of W solution of

$\boxed{W \overset{\mathcal{L}}{=} V_1^{\lambda_2} W^{(1)} + V_2^{\lambda_2} W^{(2)}}$ or $\boxed{W \overset{\mathcal{L}}{=} V^{\lambda_2} \left(W^{(1)} + W^{(2)}\right)}$

- Th 1: existence and unicity (contraction method)
- Th 2: existence and exponential moments (cascade)
- Th 3: support and density % Lebesgue measure on $\mathbb{C}$ (Liu's method)

Theorem 1 *The distributional equation*

$$\boxed{W \overset{\mathcal{L}}{=} V^{\lambda_2}\left(W^{(1)} + W^{(2)}\right)}$$

admits a unique solution W *in the space* $\mathcal{M}_2(C)$ *of probability measures with a second moment and a given expectation* C.

Hints of proof:

$\rightarrow$ Banach fixed point theorem on $\mathcal{M}_2(C)$ endowed with the Wasserstein distance.

$\rightarrow$ simple computations:

$$2\mathbb{E}\left(V^s\right) = \frac{(2m)\dots(m+1)}{(2m-1+s)\dots(m+s)}$$

- λ_2 solution of $\boxed{(\lambda+m)\dots(\lambda+2m-1) = \frac{(2m)!}{m!}} \Longrightarrow \mathbb{E}\left(V^{\lambda_2}\right) = \frac{1}{2}$

- The smoothing transform is $\dfrac{1}{\frac{1}{2}+\Re(\lambda_2)}$-Lipschitz and

$\Re(\lambda_2) > \dfrac{1}{2} \Longrightarrow$ contraction

Theorem 2 *Let W be a solution of*

$$\boxed{W \overset{\mathcal{L}}{=} V^{\lambda_2} \left(W^{(1)} + W^{(2)}\right)}$$

There exist some constants $C > 0$ and $\varepsilon > 0$ such that for all $t \in \mathbb{C}$ with $|t| \leq \varepsilon$,

$$\mathbb{E}e^{\langle t,W\rangle} \leq e^{m\Re(t)+C|t|^2} \quad \textit{and} \quad \mathbb{E}e^{|tW|} \leq 4e^{m|t|+2C|t|^2}. \tag{1}$$

<u>Hints of proof</u>:

$\rightarrow$ introduce the cascade

$(A_u)_{u\in U}$ iid, $A_u \overset{\mathcal{L}}{=} V^{\lambda_2}$

$u = u_1 u_2 \ldots u_{n-1}$

$$Y_n = 2mV^{\lambda_2} \sum_{|u|=n-1} A_{u_1} A_{u_1 u_2} \ldots A_{u_1 \ldots u_{n-1}}$$

so that

$$Y_{n+1} = V^{\lambda_2} \left(Y_n^{(1)} + Y_n^{(2)}\right),$$

(Y_n) is a martingale, bounded in L^2 (computation) since $\Re(\lambda_2) > \dfrac{1}{2}$.

$$Y_n \rightarrow Y_\infty \text{ a.s. and in } L^2,$$

$\rightarrow$ prove the theorem for Y_∞, by recursivity on Y_n.

Theorem 3 *Let W be a solution of*

$$\boxed{W \overset{\mathcal{L}}{=} V^{\lambda_2}\left(W^{(1)} + W^{(2)}\right)} \qquad V \overset{\mathcal{L}}{=} Beta(m,m)$$

(i) The support of W is the whole complex plane $\mathbb{C}$;
(ii) W is absolutely continuous relatively to Lebesgue's measure on $\mathbb{C}$.

W is $\mathbb{C}$-valued
$\lambda_2 \in \mathbb{C}$, the support of V^{λ_2} is a spiral
the sum is mixing the supports. So ??

Hints of proof (à la Liu):

$\rightarrow$ the support is $\mathbb{C} \Longleftarrow z \in Supp(W) \Longrightarrow 2^n v^\lambda z \in Supp(W)$
$+$ arithmetical argument.

$\rightarrow$ equation written on the Fourier transform, and Fourier inversion. Relies on the fact that W is non lattice.

Open questions = More on W

- Is it a known distribution?
- Exact size of the moments
- Shape of the density
- Exploit W^{CT} infinitely divisible, self-decomposable.