The Contour Process of Crump-Mode-Jagers Branching Processes

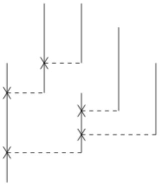
Emmanuel Schertzer (LPMA Paris 6), with Florian Simatos (ISAE Toulouse)

June 24, 2015

Crump-Mode-Jagers trees

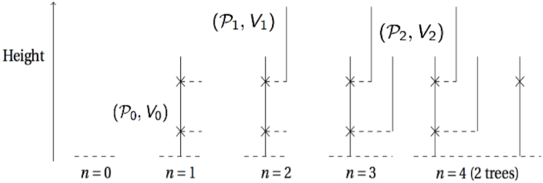
- Crump–Mode–Jagers (CMJ) branching processes generalize Galton–Watson process.
- Individuals live for a random duration and give birth at random times during their life-time.
- Each individual is characterized by a random pair $(\mathcal{P}, V) \in \mathcal{M} \times \mathbb{R}^+$
  1. V is the life-length of the individual
  2. $V \geq \tau^1 \geq \tau^2 \cdots \geq \tau^{|\mathcal{P}|}$

$$\mathcal{P} = \sum_{i=1}^{|\mathcal{P}|} \delta_{\tau^k}$$



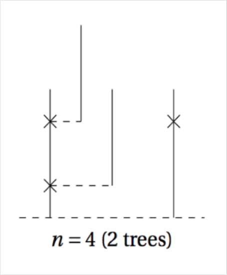
Crump-Mode-Jagers forests

- A CMJ forest is constructed from a sequence of i.i.d. life-descriptors $\{(\mathcal{P}_n, V_n)\}_{n\geq 0} \in \mathcal{M} \times \mathbb{R}^+$
- The root's life length is equal to V_0. We graft new individuals at the atoms of $\mathcal{P}_0$.
- The process is repeated until the population gets extinct, which happens almost surely in the (sub)critical case $\mathbb{E}(|\mathcal{P}|) \leq 1$.
- At extinction time, generate a new tree according to the same procedure.



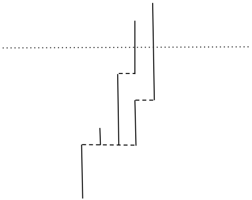
Chronological and genealogical structure

- Genealogy: the Galton Watson forest generated from the sequence $\{|\mathcal{P}_n|\}_{n\geq 0}$. This GW forest encodes the genealogical structure of the CMJ.
- Chronology: the CMJ itself.



Population dynamics

- The population size of a finite (but) large CMJ process is not Markovian.
- Ex: Bellman-Harris processes. individuals beget their children at death, independently of their life-time.
- Need to keep track of the population size and the age structure in the population. This might be represented by a measure-valued process (Sagitov ('97)).
- General case: need to keep track of the age structure and birth events occurring in the past.



The contour process approach

- Different point of view: contour processes.
- The contour process of a CMJ process encodes not only the population dynamics, but also the geometry of the underlying trees (or forests).
- In the particular case of Poissonian birth events, Lambert ('10) showed that the contour process is a Lévy process.
- Using this approach, we will be able to show that CMJ forests fall into three main categories:
  1. If edges are short enough, CMJ forests are obtained by a deterministic stretching of their underlying genealogy.
  2. If the offspring distribution has a finite second finite, a CMJ forest looks asymptotically like a Bellman-Harris forest.
  3. For long edges and large number of offspring, our approach provides an educated guess for a natural scaling limit.

Coding discrete planar trees

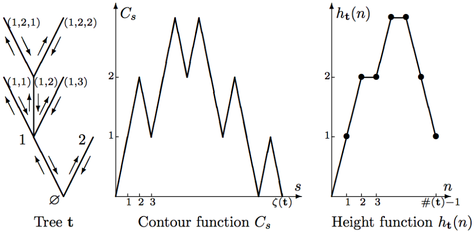
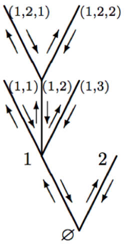


Figure : pictures taken from Le Gall's lecture notes

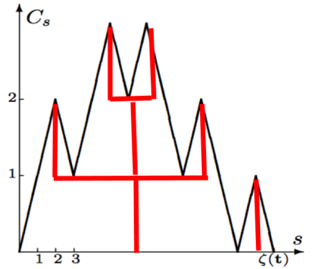
- Contour process: distance from the origin of a particle exploring the tree from left to right by travelling at unit speed along the edges.
- Height process: distance from the origin of the n^{th} vertex visited in lexicographical order.

From tree to contour and vice-versa.

- To every rooted planar tree corresponds a contour process.
- The genealogy can be recovered from the sub-excursions nested in the contour process.
- Real tree: tree constructed from an arbitrary excursion.

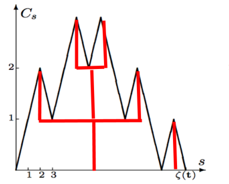


Tree t



Contour function C_s

Real trees (Aldous, Evans, Pitman Duquesne, Le Gall, etc ...)



Contour function C_s

- Let f be a non-negative continuous function made of finite excursions with $f(0) = 0$.
- $x \sim y$ iff $\inf_{[x\wedge y, x\vee y]} f = f(x) = f(y)$
- $\mathcal{T}_f = \mathbb{R}^+ / \sim$.
- $d_f(x, y) \;=\; f(x) + f(y) - 2\inf_{[x\wedge y, x\vee y]} f$.

Real trees

$(\mathcal{T}_f, d_f)$ defines a real forest in the sense that

(i) **(Unique geodesics.)** There is a unique isometric map $\psi^{a,b}$ from $[0, d_f(a,b)]$ into $\mathcal{T}_f$ such that $\psi^{a,b}(0) = a$ and $\psi^{a,b}(d_f(a,b)) = b$.

(ii) **(Loop free.)** If q is a continuous injective map from $[0,1]$ into $\mathcal{T}_f$, such that $q(0) = a$ and $q(1) = b$, we have $q([0,1]) = \psi^{a,b}([0, d_f(a,b)])$.

- Continuous Random Forest (Aldous ('91)): Real forest encoded by a reflected Brownian motion.

Scaling limit of Galton Watson genealogies: finite variance case

- Let $\mathcal{F}$ be a Galton Watson forest such that

$$\sum_{k=0}^{\infty} kp(k) = 1 \text{ (critical case) and } 0 < \sum_{k=0}^{\infty} k^2 p(k) - 1 = \sigma^2 < \infty$$

  and $\mathcal{C}$ its contour process.
- $\frac{1}{\sqrt{n}}(\mathcal{H}(n\cdot), \mathcal{C}(n\cdot))$ converges in the weak topology to $\frac{2}{\sigma}(|w|(\cdot), |w|(\cdot/2))$, where w is a standard Brownian motion.
- This indicates that the (rescaled) Galton Watson forest converges to the real forest encoded by a reflected Brownian motion.

Scaling limit of Galton Watson genealogies: infinite variance case (Le Gall Le Jan ('98), Duquesne Le Gall ('02))

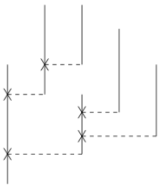
- Let $\mathcal{F}$ be a Galton Watson forest such that

$$\sum_{k=0}^{\infty} kp(k) = 1 \quad \text{(critical case)}$$

- The offspring distribution $\{p(k)\}$ is in the domain of attraction of the stable law with exponent $\alpha \in (1,2)$.
- There exists $\epsilon_n \to 0$ such that $\epsilon_n (\mathcal{H}(n\cdot), \mathcal{C}(n\cdot))$ converges in the weak topology to a continuous random process $(\mathcal{H}_\infty(\cdot), \mathcal{H}_\infty(\cdot/2))$.
- The limiting process is not Markovian but can be expressed as a functional of a spectrally positive Lévy process

Back to the original question

- In this talk, we will consider the height process and the contour process of a CMJ forest.
- Let $\mathbb{H}$ (resp., $\mathbb{C}$) be the height (resp., contour) process of the CMJ.
- Let $\mathcal{H}$ (resp., $\mathcal{C}$) be the height (resp., contour) process of the underlying Galton Watson forest.



Back to the original question

- Question 1: limiting behavior of $\mathbb{C}$ at large time.
- Question 2: joint convergence of $(\mathbb{C},\ \mathcal{C})$ and $(\mathbb{H}, \mathcal{H})$, i.e., relation between the CMJ forest and its underlying genealogy.

Result 1: Height process in the short edges case

- Define y to be the random variable such that for every test function f:

$$\mathbb{E}\left[f(y)\right] = \sum_{k=1}^{\infty} \left(\frac{1}{k} \sum_{r=1}^{k} \mathbb{E}\left[f\left(A(\mathcal{P}, r)\right) \mid |\mathcal{P}| = k\right] \right) k\mathbb{P}(|\mathcal{P}| = k)$$

  where $A(\mathcal{P}, r)$ is the position of the r^{th} atom on $\mathcal{P}$

- Assume that $\mathbb{E}(y) < \infty$
- Assume that there exists $\epsilon_n \to 0$ and $\mathcal{H}_\infty$ such that

$$\epsilon_n \mathcal{H}(n\cdot) \implies \mathcal{H}_\infty.$$

- Then

$$\epsilon_n \left(\mathcal{H}(n\cdot), \ \mathbb{H}(n\cdot)\right) \to \left(\mathcal{H}_\infty, \mathbb{E}(y)\mathcal{H}_\infty\right).$$

  in the sense of f.d.d.

Result 2: from height to contour processes in the short edges case.

- Assume that $\mathbb{E}(V) < \infty$.
- Assume that there exists $\epsilon_n \to 0$

$$\epsilon_n \mathcal{H}([n\cdot]) \Longrightarrow \mathcal{H}_\infty(\cdot), \quad \epsilon_n \mathcal{C}([n\cdot]) \Longrightarrow \mathcal{C}_\infty(\cdot)$$

- Assume that there exists $\bar{\epsilon}_n \to 0$ such that $\{\bar{\epsilon}_n \mathbb{H}(nt)\}_n$ is tight (for every determinstic t) then

$$\bar{\epsilon}_n \left(\mathbb{H}\left(\frac{[nt]}{2\mathbb{E}(V)} \right) - \mathbb{C}([nt]) \right) \Longrightarrow 0$$

  in the sense of f.d.d..
- Generalization of a result by Duquesne & Le Gall in the discrete setting. Non trivial due to the absence of tightness.

Tightness issue

- X in the domain of attraction of an α-stable law with $\alpha \in (1,2)$.
- $\mathcal{P} = \delta_1(X-1) + \delta_X$ in such a way that $|\mathcal{P}|$ is distributed as X.
- $\mathbb{E}(y) < \infty$.
- $\frac{1}{n^{1-\frac{1}{\alpha}}}\mathcal{H}([n\cdot]) \implies \mathcal{H}_\infty$.
- $\frac{1}{n^{1-\frac{1}{\alpha}}}\mathbb{H}([n\cdot]) \implies \mathbb{E}(y)\mathcal{H}_\infty$.
- $\max_{0,\cdots,n} \mathbb{H} \geq \max_{0,\cdots,n} X_i \sim n^{1/\alpha}$
- Taking $\alpha < \frac{1}{2}(1+\sqrt{5})$, we have $n^{1/\alpha} >> n^{1-\frac{1}{\alpha}}$
- As n goes to ∞, one can find infinitely many edges which do not scale as the height process.

- $\mathbb{E}(V) < \infty, \mathbb{E}(y) < \infty$
- Combining (R1) and (R2)

$$\epsilon_n(\mathcal{C}([n\cdot], \mathbb{C}([n\cdot]) \to (\mathcal{C}_\infty,\ \mathbb{E}(y)\mathcal{C}_\infty(\cdot/2\mathbb{E}(V)))$$

- Sagitov ('95)

$$\epsilon_n Z([nt]) \to \frac{\mathbb{E}(V)}{\mathbb{E}(y)} \mathcal{Z}(t/2\mathbb{E}(y))$$

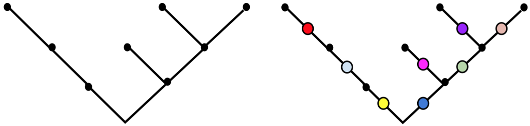
  where $\mathcal{Z}$ is a CSBP starting at 1 and Z the discrete CMJ branching processes BP starting with $[1/\epsilon_n]$ individuals.
- Question: What if $\mathbb{E}(y) = \infty, \mathbb{E}(V) = \infty$? We distinguish between two cases (1) $\mathbb{E}(|\mathcal{P}|^2) < \infty$, and (2) $\mathbb{E}(|\mathcal{P}|^2) = \infty$.

Result 3: Height process for offspring distribution with finite variance case

- $\mathbb{E}(|\mathcal{P}|^2) < \infty$ but y is the domain of attraction of an α-stable law with $\alpha \in (0,1)$.
- There exists $\epsilon_n \to 0$ such that the joint distribution of

$$\left(\frac{1}{n^{1/2}} \mathcal{H}(n\cdot),\ \epsilon_n \mathbb{H}(n\cdot) \right)$$

can be asymptotically described in terms of the Poisson snake.



- Continuum analog of Bellman-Harris forests.
- Start from the genealogical structure.
- Mark every edge with an independent random number according to life-length distribution V.
- The BH forest can be recovered by a simple (random) stretching of the underlying genealogy.
- Informelly, our results indicates that if $|\mathcal{P}|$ has a finite second moment, then CMJ branching processes behave as a Bellman Harris forest.

Poisson snake (Warren ('02), Abraham Delmas ('02) etc.)

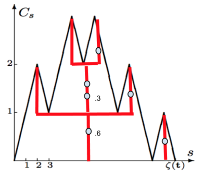
- Let $(\mathcal{F}_{\frac{2}{\sigma}|w|}, d_{\frac{2}{\sigma}|w|})$ be the real forest induced by a reflected Brownian motion.
- Let $\lambda_{\frac{2}{\sigma}|w|}$ be the branch length measure, i.e. $\forall a, b$

$$\lambda_{\frac{2}{\sigma}|w|}([a, b]) \;=\; d_{\frac{2}{\sigma}|w|}(a, b).$$

- Conditioned on $(\mathcal{F}_{\frac{2}{\sigma}|w|}, d_{\frac{2}{\sigma}|w|})$, mark the forest with a Poisson Point process on $\mathcal{F}_{\frac{2}{\sigma}|w|} \times \mathbb{R}^+$ with intensity measure

$$\lambda_{\frac{2}{\sigma}|w|} \times \frac{dl}{l^{\alpha+1}}$$

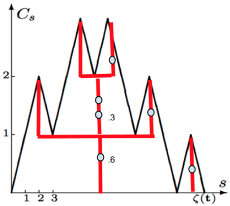
.



Contour function C_t

- $\mathbb{H}_\infty(t)$ is the sum of all the atoms along the branch $[\rho, t]$.
- $$\left(\frac{1}{\sqrt{n}}\mathcal{H}([nt]),\ \epsilon_n \mathbb{H}([nt]);\ t \geq 0\right) \to \left(\frac{2}{\sigma}|w|(t),\ \mathbb{H}_\infty(t);\ t \geq 0\right)$$

  in the sense of f.d.d.
- The chronological ancestral line is obtained by random dilatation of the genealogical ancestral line.



Contour function C_s

Result 4: Contour process for offspring distribution with finite variance case

- $\mathbb{E}(|\mathcal{P}|^2) < \infty$. V is in the domain of attraction of a β-stable law with $\beta \in (0,1)$.
- Assume that there exists $\epsilon_n, \bar{\epsilon}_n \to 0$

$$\epsilon_n \mathcal{H}([n\cdot]) \Longrightarrow \mathcal{H}_\infty(\cdot), \bar{\epsilon}_n \mathbb{H}([n\cdot]) \Longrightarrow \mathbb{H}_\infty(\cdot),$$

  There exists $\tilde{\epsilon}_n \to 0$,

$$(\epsilon_n \mathbb{H}([n\cdot]), \epsilon_n \mathbb{C}([n\cdot])) \ \to \ (\mathbb{H}_\infty, \mathbb{C}_\infty),$$

  in the sense of f.d.d. such that $\mathbb{H}_\infty$ and $\mathbb{C}_\infty$ are independent.
- Let $\bar{\mathbb{C}}(t)$ be the date of birth (height) of the individual visited at time t. Then

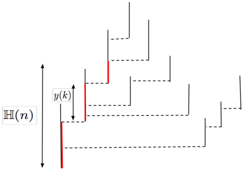
$$(\epsilon_n \mathbb{H}, \tilde{\epsilon}_n \bar{\mathbb{C}}) \ \to \ (\mathbb{H}_\infty, \mathbb{H}_\infty \circ \Gamma^{-1}),$$

  (again in the sense of f.d.d.) where Γ is a β-stable subordinator independent of $\mathbb{H}_\infty$.

Summary

- $\mathbb{E}(V), \mathbb{E}(y) < \infty$: the chronology is recovered from the genealogy by a deterministic space-time change.
- $\mathbb{E}(V), \mathbb{E}(y) = \infty$ by $\mathbb{E}(|\mathcal{P}|^2) < \infty$: the chronology is obtained by a random dilation of the genealogy and a random time change.

Fundamental decomposition of the spine



- $\mathcal{H}(n)$: number of ancestors of n
- $y(k)$ = age of the k^{th} ancestor when it begets the $(k-1)^{th}$ ancestor.
- The number of ancestors and the length can be computed from the Lukasiewicz path. $S(0) = 0$ and

$$S(n+1) - S(n) = |\mathcal{P}_n| - 1.$$

Random walk with $\mathbb{E}(S(n+1) - S(n)) = 0$ and negative increments of size 1.

Fundamental decomposition of the spine

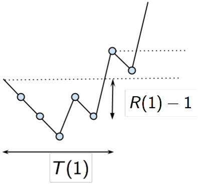
- Define $\{T(k)\}$ the sequence of (weak) ascending ladder times:

$$T(0) = 0; \quad \text{and for } k \geq 1$$

$$T(k+1) = \inf\{k > T(k) \ : \ S(k) = \sup_{\{0,\cdots,k\}} S\}.$$

- Define $R(k)$ as the undershoot upon reaching the running maximum at time $T(k)$ incremented by 1 unit

$$\text{for } k \geq 1, \ R(k) = S(T(k-1)) - S(T(k)-1) + 1$$



The dual walk and the genealogical height process.

- $\omega = (\mathcal{P}_k, V_k)_{k\in\mathbb{Z}}$
- $\vartheta^n(\omega) = (\mathcal{P}_{n-1-k}, V_{n-1-k})_{k\in\mathbb{Z}}$
- Dual walk at n:

$$S \circ \vartheta^n \;=\; (S(n) - S(n-k);\; k \geq 0)\,,$$

- $\mathcal{A}(n)$ indices of the ancestors of n.

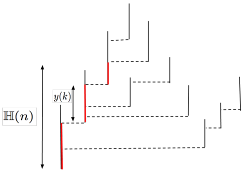
$$\mathcal{A}(n) \;=\; \{n - T(k)\circ\vartheta^n \;:\; T(k)\circ\vartheta^n \leq n\}.$$

- $\mathcal{H}(n) = |\{k \leq n :\; T(k)\circ\vartheta^n \leq n\}|$

The dual walk and the chronological height process

- $A(\mathcal{P}, k)$: location of the rth atom on $\mathcal{P}$.
- $y(k) = A(\mathcal{P}_{T(k)-1}, R(k))$ random functional of the ladder height process.
- $y(k) \circ \vartheta^n$ is the contribution of the k^{th} ancestor: age of the k^{th} when it begets the $(k-1)^{th}$ ancestor.
- The pair $(\mathcal{H}(n)\ ,\ \mathbb{H}(n))$ is equal to

$$\left(|\{k \leq n : \ T(k) \leq n\}|\ ,\ \sum_{k \leq n:\ T(k) \leq n} y(k) \right) \circ \vartheta^n.$$



(R2) and (R3) from the spine decomposition

- From standard excursion theory, the process

$$\left(\left(T(k)\ ,\ \sum_{i=1}^{k} y(i)\right)\ ;\ k \geq 0\right)$$

defines a bivariate renewal process.

- In general, the difficulty stems from the fact that those two renewal processes are correlated in a non-trivial way.
- $\mathbb{E}(y) < \infty$: $\sum_{k=1}^{n} y(k) \ \sim \ n\mathbb{E}(y)$
- $\mathbb{E}(|\mathcal{P}^2|) < \infty$: the two subordinators become independent.

Long edges and large number of offsprings

- Start from a P.P.P. $\{(t_i, \mathcal{P}_{t_i})\}_{t_i \in I}$ on $\mathbb{R} \times \mathcal{M}$ with intensity measure $dt \times \lambda(d\mathcal{P})$.
- Let π be the push-forward measure of λ by the map $(t, \mathcal{P}) \to (t, |\mathcal{P}|)$ and assume that $\int_0^\infty x \wedge x^2 \pi(dx) < \infty$.
- $\{(t_i, |\mathcal{P}_{t_i}|\}_{t_i \in I}$ is a Poisson Point Process on $\mathbb{R} \times \mathbb{R}^+$ with intensity measure $dt \times \pi(dx)$.
- From the sequence $\{(t_i, |\mathcal{P}_{t_i}|)\}_{t_i \in I}$, one can construct a Lévy process X_t with Laplace exponent

$$\alpha\lambda \ + \int_0^\infty (e^{-\lambda x} - 1 + \lambda x)\pi(dx)$$

- Choose α, π such that X is of infinite variation ($\int_0^\infty x\pi(dx) = \infty$) and does not drift to $+\infty$.

- $S_t = \max_{[0,t]} X$
- Let L be the local at 0 of $S - X$.
- Construct a discrete sequence from the ascending ladder height process corresponding to X and analogously to the discrete case, we consider

$$\left(L^{-1}_{u_i},\ S^{-}_{L^{-1}_{u_i}} - X^{-}_{L^{-1}_{u_i}},\ \mathcal{P}_{L^{-1}_{u_i}} \right)_{t_i \in I}$$

where the t_i's correspond to the jump times of S.

- This defines a P.P.P. with intensity

$$\mu(dt\ dr\ d\mathcal{P}) \ = \ 1_{r \in [0,|\mathcal{P}|]} \lambda(d\mathcal{P}) \mu_r(dt)$$

where μ_r is the law of the $\inf\{u : X_u = -r\}$

- By mimicking the spine decomposition described above, one can construct a continuum analog of the height process.

- Start P.P.P. $\{(t_i, \mathcal{P}_{t_i})\}_{i \in I}$.
- $X \circ \vartheta^t = X(t) - X(t - \cdot)$: dual Lévy process
- $A(\mathcal{P}, x) = \sup\{u \ : \ \mathcal{P}([u, \infty)) \geq x\}$.
- At jump times of S,

$$y(u) = A(\mathcal{P}_u, S_u^- - X_u^-)$$

- $\mathbb{H}(t) = \left(\sum_{u \leq t \ : \ \Delta S_u > 0} y(u)\right) \circ \vartheta^t$. (Random transformation of the Lévy process $X \circ \vartheta^t$.)
- $\mathcal{H}(t) = \left(\sum_{u \leq t \ : \ \Delta S_u > 0} 1 \right) \circ \vartheta^t$ (Duquesne Le Gall)

Open Questions

- Convergence in the long edges and large number of offspring case.
- Ray-Knight type of theorems.
  1. Local time for generalized height processes.
  2. Can we identify those local times with an underlying continuum branching processes (measure valued process).

Thank you !