Convergence of contour processes of random trees

Douglas Rizzolo
Rough Notes

In this lecture, we are interested in studying the asymptotic contour process of a Gatlon-Watson tree with Geometric(1/2) offspring distribution conditioned to be large. Instead of looking at the contour process of a single tree, it will be easier to first consider an infinite forest of trees. We will use Donsker's theorem to find the limiting contour process of the forest and then we will look for large trees that occur naturally in the infinite forest. In order to simplify our lives, we will consider a signed forest. That is, we let $F = ((T_k, U_k))_{k\geq 1}$ be an i.i.d sequence such that T_k and U_k are independent, $\mathbb{P}(U_k = 1) = \mathbb{P}(U_k = -1) = 1/2$ and T_k is a planted Gatlon-Watson tree with Geometric(1/2) offspring distribution, planted meaning the root is conditioned to have degree equal to 1. We will call F a signed forest of planted Gatlon-Watson trees with Geometric(1/2) offspring distribution.

For a tree $\mathbf{t}$, we let $C_{\mathbf{t}} : [0, 2\#\mathbf{t}] \to [0, \infty)$ be the contour function of $\mathbf{t}$. Consider a signed forest $\mathbf{f} = ((\mathbf{t}_k, e_k))_{k\geq 1}$ and let $n_0 = 0$ and $n_t = \min\{j : 2\#\mathbf{t}_1 + \cdots + 2\#\mathbf{t}_j > t\}$. The contour function of the forest is defined by $C_{\mathbf{f}}(t) = e_{n_t} C_{\mathbf{t}_{n_t}}(t - 2[\#\mathbf{t}_1 + \cdots + \#\mathbf{t}_{n_t-1}])$

Theorem 1. $(C_F(k))_{k\geq 0}$ *is a simple random walk.*

Proof. Let $X_1, X_2, \ldots$ be i.i.d with $\mathbb{P}(X_k = 1) = \mathbb{P}(X_k = -1) = 1/2$ and define $S_0 = 0$ and $S_n = X_1 + \cdots + X_n$ for $n \geq 1$. We want to show that $(C_F(k))_{k\geq 0} =_d (S_n)_{n\geq 0}$. There are many ways to do this, but one nice one is to use excursion theory. We decompose the random walk based on its returns to 0. Specifically, let $\tau_0 = 0$ and for $n \geq 1$ define $\tau_n = \min\{n > \tau_{n-1} : S_n = 0\}$. Using the strong Markov property of simple random walk, we see that $((S_{\tau_n+k})_{k=0}^{\tau_{n+1}})_{n\geq 0}$ is an i.i.d. sequence. But $((e_{T_n} C_{T_n}(k))_{k=0}^{2\#T_n})_{n\geq 0}$ is also an i.i.d. sequence and $(S_{\tau_n+k})_{k=0}^{\tau_{n+1}} =_d (e_{T_n} C_{T_n}(k))_{k=0}^{2\#T_n}$ by the formula for $\mathbb{P}(T_k = \mathbf{t})$ and the bijection between rooted ordered trees Dyck paths. □

Corollary 1. *We can construct a sequence* $((C_F^n(k))_{k\geq 0})_{n\geq 0}$ *such that* $(C_F^n(k))_{k\geq 0} =_d (C_F(k))_{k\geq 0}$ *for every* n *and a Brownian motion* $(B_t, 0 \leq t < \infty)$ *on a common probability space such that for every* $T > 0$,

$$\max_{t\in[0,T]} \left| \frac{1}{\sqrt{n}} C_F(nt) - \frac{1}{\sqrt{n}} B_{nt} \right| \xrightarrow[n\to\infty]{p} 0.$$

This is not quite what we need. We need the following corollary of this result, which follows from the Skorokhod representation theorem.

Corollary 2. *We can construct an ind*$(C_F(k))_{k\geq 0}$ *and a Brownian motion* $(B_t, 0 \leq t < \infty)$ *on a common probability space such that for every* $T > 0$,

$$\max_{t\in[0,T]} \left| \frac{1}{\sqrt{n}} C_F^n(nt) - B_t \right| \xrightarrow[n\to\infty]{a.s.} 0.$$

How can we use this to obtain results about conditioned Galton-Watson trees? We make the following observation: If $Y_1, Y_2, \dots$ is an i.i.d sequence, A is a set such that $\mathbb{P}(Y_1 \in A) > 0$, and $\sigma_A = \inf\{n : Y_n \in A\}$, then $\mathbb{P}(Y_{\sigma_A} \in B) = \mathbb{P}(Y_1 \in B | Y_1 \in A)$. That is, Y_{σ_A} is distributed like Y_1 conditioned to be in A. For example, if $A = \{\mathbf{t} : \mathrm{ht}(\mathbf{t}) > n\}$ then T_{σ_A} is distributed like a Galton-Watson tree conditioned to have height greater than n.

What sort of events can Corollary 2 be used to condition on? Let us give some heuristics. Suppose that $(A_n)_{n\geq 0}$ is a sequence of events such that $\mathbb{P}(T_1 \in A_n) \sim cn^{-1/2}$. Note that σ_{A_n} has a geometric distribution with parameter $\mathbb{P}(T_1 \in A_n)$. The the expected number of trees we need to look at to see a tree in A_n is $\mathbb{E}\sigma_{A_n} = \mathbb{P}(T_1 \in A_n)^{-1} \sim c^{-1}n^{1/2}$. In Corollary 2, we look at the first nT vertices in the forest, how many trees is this? The Otter-Dwass formula and the local central limit imply that $\mathbb{P}(\#T_1 = n) \sim Cn^{-3/2}$. The stable central limit theorem implies that

$$\frac{1}{n^2}\sum_{k=1}^{n} \#T_k \xrightarrow{d} Y,$$

where Y is a totally asymmetric stable distribution. Consequently, $\sum_{k=1}^{\sqrt{n}} \#T_k \approx nY$, so Corollary 2 looks at the first $\sqrt{n}$ trees in the forest. Thus we can expect that with positive probability the first tree in A_n will occur before we have seen the first nT vertices of the forest. Thus if we can pick out the part of the contour process corresponding to this tree, we should be able to prove a functional limit theorem for its contour process.

Example If $A_n = \{\mathbf{t} : \mathrm{ht}(\mathbf{t}) \geq \sqrt{n}\}$ then $\mathbb{P}(T_1 \in A_n) \sim n^{-1/2}$, so we can condition on these events.

Example If $A_n = \{\mathbf{t} : \#\mathbf{t} \geq n\}$ then $\mathbb{P}(T_1 \in A_n) \sim cn^{-1/2}$, so we can condition on these events.

Non-Example If $A_n = \{\mathbf{t} : \#\mathbf{t} = n\}$ then $\mathbb{P}(T_1 \in A_n) \sim cn^{-3/2}$, so we cannot condition on these events.

In order to make this heuristic rigorous, we need the following result about Brownian motion that seems clear but is somewhat tricky to prove.

Proposition 1. *Letting $\gamma = \inf\{t > 0 : B_t = 0\}$ we have $\gamma = 0$ a.s.*

Corollary 1. *The zero set of Brownian motion has no isolated points.*

Proof. For $q \in \mathbb{Q}^+$ define $\tau_q = \inf\{t > q : B_t = 0\}$. By the strong Markov property, $\inf\{t > \tau_q : B_t = 0\} = 0$ a.s. Thus this holds almost surely simultaneously for all $q \in \mathbb{Q}^+$. Suppose that $B_t = 0$ t does not equal τ_q for any $q \in \mathbb{Q}^+$. For every $q < t$, $q \leq \tau_q < t$. Letting $q \uparrow t$ proves the result. □

To illustrate the method, we prove the following theorem.

Theorem 1. *Let T_n be distributed like a Galton-Watson tree conditioned to have height at least $\sqrt{n}$. Then*

$$\left(\frac{1}{\sqrt{n}} C_{T_n}(nt)\right)_{0 \leq t \infty} \xrightarrow{d} (B_t^1, 0 \leq t < \infty).$$

where B^1 is a Brownian excursion conditioned to have height at least 1.

Corollary 2. *If $a < b < x$ then $\mathbb{P}_x(\tau_a < \tau_b) = (b-x)/(b-a)$. $\mathbb{P}(ht(T_n) \geq b) = (b-1)/b$.*

Proof. Proof by martingales. Proof by random walk. (Somewhat redundant). □

Convergence of the corresponding height process. Vertices occur after up-steps. Large deviations concentration about $2t$ of the time change.