Scaling limits of random trees and graphs: exercises

Please send any comments or corrections to goldschm@stats.ox.ac.uk.

For reference: the Dirichlet distribution with parameters $a_1, a_2, \ldots, a_n > 0$ has density

$$\frac{\Gamma(\sum_{i=1}^n a_i)}{\prod_{i=1}^n \Gamma(a_i)} x_1^{a_1-1} \ldots x_n^{a_n-1}$$

with respect to the Lebesgue measure on

$$\left\{ \mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}_+^n : \sum_{i=1}^n x_i = 1 \right\}.$$

1. **(Rémy's algorithm)** Show that after n steps, Rémy's algorithm indeed generates a uniform random element of $\mathbb{T}_n^*$, the set of planted plane binary trees with leaves labelled by $1, 2, \ldots, n$.

2. **(Urns and martingales: the distance between 0 and 1 in Rémy's algorithm)** Consider the urn process which starts with a single black ball, such that when we pick a black ball, we put it back in the urn with one further black and one white ball, and when we pick a white ball, we put it back with a further two white balls. For $n \geq 1$, let B_n be the number of black balls at step 1, so that $B_1 = 1$. Let $b_1 = 1$ and $b_{n+1} = \frac{2^{2n}(n!)^2}{(2n)!}$ for $n \geq 1$.

   (a) Show that $b_{n+1} = \frac{2n}{2n-1} b_n$ for $n \geq 1$.

   (b) Deduce that the process $(M_n)_{n \geq 1}$ defined by $M_n = B_n / b_n$ is a martingale.

   (c) Show that $\mathbb{E}\left[M_{n+1}^2\right] \leq \mathbb{E}\left[M_n^2\right] + \frac{1}{2nb_n}$ and deduce that $(M_n)_{n \geq 1}$ is bounded in L^2.

3. **(Urns and martingales: the proportions of leaves in the different subtrees)** Suppose that an urn initially contains one red ball, one green ball and one blue ball. At each time-step, pick a ball uniformly at random from the urn, look at its colour, and put it back into the urn with two extra balls of the same colour. Let R_n, G_n, B_n denote the number of red, green and blue balls respectively at step $n \geq 0$, so that $(R_0, G_0, B_0) = (1, 1, 1)$. Observe that at step n, there are $2n + 3$ balls in the urn.

   (a) Show that $\left(\frac{1}{2n+3}(R_n, G_n, B_n), n \geq 0\right)$ is a (vector-valued) martingale.

   (b) Deduce that

   $$\frac{1}{2n+3}(R_n, G_n, B_n) \to (\Delta_1, \Delta_2, \Delta_3) \text{ a.s. as } n \to \infty,$$

   for some non-negative random vector $(\Delta_1, \Delta_2, \Delta_3)$ such that $\Delta_1 + \Delta_2 + \Delta_3 = 1$.

(c) For $n \geq 1$, let C_n denote the index of the colour picked at step n (red = 1, green = 2, blue = 3). Show that for $n_1, n_2, n_3 \geq 0$ such that $n_1 + n_2 + n_3 = n$,

$$\begin{aligned}&\mathbb{P}\left(C_1 = \cdots = C_{n_1} = 1, C_{n_1+1} = \cdots = C_{n_1+n_2} = 2, C_{n_1+n_2+1} = \cdots = C_n = 3\right)\\ &\quad = \frac{2(2n_1)!(2n_2)!(2n_3)!(n+1)!}{n_1!n_2!n_3!(2n+2)!}.\end{aligned}$$

(d) Show that for a fixed sequence $\mathbf{c} = (c_1, c_2, \ldots, c_n) \in \{1,2,3\}^n$ and any permutation $\sigma : \{1,2,\ldots,n\} \to \{1,2,\ldots,n\}$,

$$\mathbb{P}\left(C_1 = c_1, C_2 = c_2, \ldots, C_n = c_n\right) = \mathbb{P}\left(C_{\sigma(1)} = c_1, C_{\sigma(2)} = c_2, \ldots, C_{\sigma(n)} = c_n\right).$$

This shows that the random variables $(C_1, \ldots, C_n)$ are exchangeable.

(e) Deduce that

$$\mathbb{P}\left(R_n = 2n_1 + 1, G_n = 2n_2 + 1, B_n = 2n_3 + 1\right) = \frac{\binom{2n_1}{n_1}\binom{2n_2}{n_2}\binom{2n_3}{n_3}}{(n+1)\binom{2n+2}{n+1}}.$$

(f) Now fix $x_1, x_2, x_3 \geq 0$ such that $x_1 + x_2 + x_3 = 1$. Using Stirling's formula, show that

$$\begin{aligned}&\mathbb{P}\left(R_n = 2\lfloor nx_1\rfloor + 1, G_n = 2\lfloor nx_2\rfloor + 1, B_n = 2\lfloor nx_3\rfloor + 1\right)\\ &\quad \sim Cn^{-2}x_1^{-1/2}x_2^{-1/2}x_3^{-1/2},\end{aligned}$$

where C is a constant.

(g) Deduce that $(\Delta_1, \Delta_2, \Delta_3) \sim \text{Dirichlet}(1/2, 1/2, 1/2)$.

For a fascinating account of exchangeability and its consequences, see the classic paper by J. F. C. Kingman, **Uses of exchangeability**, *Annals of Probability*, **6**(2), pp.183–197 (1978).

4. **(Dirichlet and gamma distributions)**

(a) Show that if

$$(D_1, D_2, \ldots, D_n) \sim \text{Dirichlet}(a_1, a_2, \ldots, a_n)$$

and $G \sim \text{Gamma}(\sum_{i=1}^n a_i, 1)$ are independent then

$$G \times (D_1, D_2, \ldots, D_n) \stackrel{d}{=} (G_1, G_2, \ldots, G_n),$$

where $G_1 \sim \text{Gamma}(a_1, 1), G_2 \sim \text{Gamma}(a_2, 1), \ldots, G_n \sim \text{Gamma}(a_n, 1)$ are independent.

(b) Show, moreover, that

$$\left(\frac{G_1}{\sum_{i=1}^n G_i}, \frac{G_2}{\sum_{i=1}^n G_i}, \ldots, \frac{G_n}{\sum_{i=1}^n G_i}\right) \stackrel{d}{=} (D_1, D_2, \ldots, D_n)$$

and is independent of $\sum_{i=1}^n G_i \sim \text{Gamma}(\sum_{i=1}^n a_i, 1)$.

5. **(Inhomogeneous Poisson processes)**

   (a) Suppose that $(P(t), t \geq 0)$ is an inhomogeneous Poisson process of rate $\lambda = (\lambda(t), t \geq 0)$, and let $(Q(t), t \geq 0)$ be an ordinary Poisson process of rate 1. Show that

   $$(P(t), t \geq 0) \stackrel{d}{=} (Q(\Lambda(t)), t \geq 0),$$

   where $\Lambda(t) := \int_0^t \lambda(x) dx$.

   (b) Suppose that $C_1, C_2, \ldots$ are the points of an inhomogeneous Poisson process of rate $\lambda(t) = t$, $t \geq 0$. Use (a) to show that if $E_1, E_2, \ldots$ are i.i.d. Exponential(1/2) then

   $$(C_k,\ k \geq 1) \stackrel{d}{=} \left(\sqrt{\sum_{i=1}^{k} E_i},\ k \geq 1 \right).$$

   For much more about general Poisson processes, see the classic book by J. F. C. Kingman, **Poisson processes**, Oxford University Press (1993).

6. **(Uniform random labelled trees)** Let T be a BGW tree with Poisson(1) offspring distribution and total progeny N.

   (a) Fix a particular rooted ordered tree t with n vertices having numbers of children $c_v, v \in \mathrm{t}$. What is $\mathbb{P}(T = \mathrm{t})$?

   (b) Condition on the event $\{N = n\}$. Assign the vertices of T a uniformly random labelling by $[n]$, and let $\tilde{T}$ be the labelled tree obtained by forgetting the ordering and the root. Show that $\tilde{T}$ has the same distribution as T_n, a uniform random tree on n vertices.
   Hint: it suffices to show that the probability of obtaining a particular tree t *is a function of* n *only.*

7. **(Other combinatorial trees)** Let T be a BGW tree with offspring distribution $(p(k), k \geq 0)$ and total progeny N.

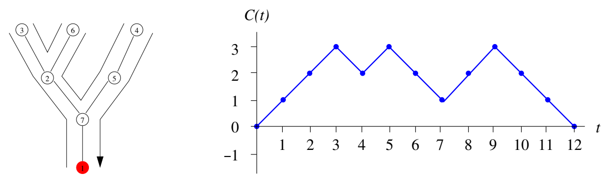
   (a) Show that if $p(0) = 1/2$ and $p(2) = 1/2$ then, conditional on $N = 2n - 1$, T is uniform on the set of rooted (unplanted!) plane binary trees with n leaves.

   (b) Show that if $p(k) = 2^{-k-1}, k \geq 0$ then, conditional on $N = n$, T is uniform on the set of ordered rooted trees with n vertices.

8. **(Height process and depth-first walk)** Let $(X(k), 0 \leq k \leq n)$ be the depth-first walk of a tree T of size n, and let $(H(k), 0 \leq k \leq n-1)$ be its height process. Convince yourself (e.g. by drawing a picture) that for $0 \leq i \leq n-1$,

   $$H(i) = \#\left\{ 0 \leq j \leq i-1 : X(j) = \min_{j \leq k \leq i} X(k) \right\}.$$

9. **(Contour function and convergence to the Brownian CRT)** In Lecture 1, we discussed the depth-first walk and the height function of a tree. A third encoding which is often used is the so-called *contour function* $(C(i), 0 \leq i \leq 2(n-1))$. For a rooted ordered tree $\mathbf{t}$, we imagine a particle tracing the outline of the tree from left to right at speed 1. (The picture below is for a labelled tree, with a planar embedding given by the labels.) Notice that we visit every vertex a number of times given by its degree.



Let T_n be a BGW tree with Geometric offspring distribution $p(k) = \left(\frac{1}{2}\right)^{k+1}, k \geq 0$, conditioned to have total progeny $N = n$, as in Question 7(b). Let $(C^n(i), 0 \leq i \leq 2(n-1))$ be its contour function. It will be convenient to define a somewhat shifted version: let $\tilde{C}^n(0) = 0$, $\tilde{C}^n(2n) = 0$ and, for $1 \leq i \leq 2n-1$, $\tilde{C}^n(i) = 1 + C(i-1)$. As usual, we write d_n for the graph distance in T_n. The *Gromov-Hausdorff distance* between (isometry classes of) compact metric spaces (X, d) and (X', d') is defined to be

$$\mathrm{d}_{\mathrm{GHP}}((X,d),(X',d')) = \frac{1}{2} \inf_R \mathrm{dis}(R),$$

where the infimum is over correspondences R between X and X'.

(a) Show that $(\tilde{C}^n(i), 0 \leq i \leq 2n))$ has the same distribution as a simple symmetric random walk (i.e. a random walk which makes steps of $+1$ with probability $1/2$ and steps of -1 with probability $1/2$) conditioned to return to the origin for the first time at time $2n$.
Hint: first consider the unconditioned BGW tree with this offspring distribution.

(b) It's straightforward to interpolate linearly to get a continuous function $\tilde{C}^n : [0, 2n] \to \mathbb{R}_+$. Let $(\tilde{T}_n, \tilde{d}_n)$ be the $\mathbb{R}$-tree encoded by this linear interpolation. Show that

$$\mathrm{d}_{\mathrm{GH}}((T_n, d_n), (\tilde{T}_n, \tilde{d}_n)) \leq 1.$$

Hint: notice that (T_n, d_n) has only n points, whereas $(\tilde{T}_n, \tilde{d}_n)$ is an $\mathbb{R}$-tree and consists of uncountably many points. Draw a picture and find a correspondence.

(c) Suppose that we have continuous excursions $f : [0,1] \to \mathbb{R}_+$ and $g : [0,1] \to \mathbb{R}_+$ which encode $\mathbb{R}$-trees $(\mathcal{T}_f, d_f)$ and $(\mathcal{T}_g, d_g)$. For $t \in [0,1]$, let $p_f(t)$ be the image of t in the tree $\mathcal{T}_f$ and similarly for $p_g(t)$. Now define a correspondence

$$R = \left\{(x,y) \in \mathcal{T}_f \times \mathcal{T}_g : x = p_f(t),\ y = p_g(t) \text{ for some } t \in [0,1]\right\}.$$

Show that $\mathrm{dis}(R) \leq 4\|f - g\|_\infty$.
Hint: recall how the metric in an $\mathbb{R}$-tree is related to the function encoding it.

(d) Observe that the variance of the step-size in a simple symmetric random walk is 1. Hence, by Kaigh's theorem, we have

$$\frac{1}{\sqrt{2(n-1)}}(C^n(2(n-1)t), 0 \le t \le 1) \xrightarrow{d} (e(t), 0 \le t \le 1) \tag{*}$$

as $n \to \infty$. Use this, (b) and (c) to prove directly that $(T_n, \frac{d_n}{\sqrt{n}})$ converges to a constant multiple of the Brownian CRT in the Gromov-Hausdorff sense.
Hint: you may want to use Skorokhod's representation theorem in order to work on a probability space where the convergence (*) occurs almost surely.

An exposition of this approach is given by Jean-François Le Gall & Grégory Miermont, **Scaling limits of random trees and planar maps**, Lecture notes for the Clay Mathematical Institute Summer School in Buzios, July 11 to August 7, 2010, *available at* `http://perso.ens-lyon.fr/gregory.miermont/Cours_Buzios.pdf`.

10. **(The total population of a BGW process)** Let N be the total population size in a BGW branching process with offspring distribution $(p(k), k \ge 0)$. Recall that $N = \inf\{k \ge 0 : X(k) = -1\}$, where the depth-first walk $(X(k), k \ge 0)$ is a random walk with $X(0) = 0$ and step distribution $\nu(k) = p(k+1)$, $k \ge -1$.

(a) Consider a possible path for X which first hits -1 at time n. There are n different cyclic rearrangements of the n steps

$$X(1) - X(0), X(2) - X(1), \ldots, X(n) - X(n-1)$$

of X i.e.

$$X(i+1) - X(i), X(i+2) - X(i+1), \ldots, X(i+n) - X(i+n-1),$$

for $0 \le i \le n-1$, where the indices are taken mod n. We always have $\sum_{k=0}^{n-1}(X(i+k+1) - X(i+k)) = -1$. Show that only one of these cyclic rearrangements results in the walk hitting -1 for the *first time* at n.

(b) Use this to argue that

$$\mathbb{P}(N = n) = \frac{1}{n}\mathbb{P}(X(n) = -1).$$

(c) Suppose that we have $p(0) = p(2) = 1/2$, so that X is a simple symmetric random walk. What is $\mathbb{P}(N = 2n-1)$ for $n \ge 1$? Deduce, using the bijection between lattice excursions and planted plane binary trees, that $|\mathbb{T}_n| = \frac{1}{n}\binom{2n-2}{n-1}$.

(d) Suppose that we have $p(k) = e^{-1}/k!$, $k \ge 0$. Show that

$$\mathbb{P}(N = n) = \frac{(\lambda n)^{n-1} e^{-\lambda n}}{n!}, \quad n \ge 1.$$

This is known as the *Borel distribution*.