Biconditioned random trees

Vanessa Dan
Supervised by Igor Kortchemski and Cyril Marzouk

CMAP, École polytechnique
DMA, École Normale Supérieur

8 Octobre 2025

Summary

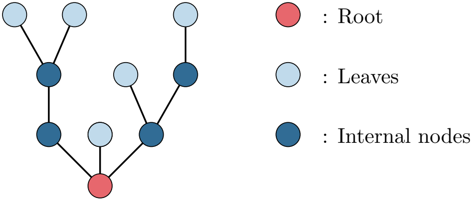
1. Introduction to the objects

2. Limit behavior of a BGW tree conditionned to have n vertices and k leaves

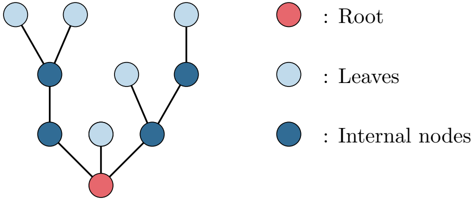
3. Limit behavior of a BGW tree conditionned to have n vertices and k internal nodes

Introduction to the objects

Trees



Trees



Weight of a tree :

Let $\mathbf{w} = (w_k)_{k\geq 0}$ be a sequence of non-negative real numbers. The weight of a tree τ is

$$\omega(\tau) := \prod_{u\in\tau} w_{c_u(\tau)},$$

where $c_u(\tau)$ is the number of children of the vertex u in τ.

Random Trees :

Random Trees :

• A uniform tree with n vertices is a tree chosen uniformly at random among the trees with n vertices.

Random Trees :

• A uniform tree with n vertices is a tree chosen uniformly at random among the trees with n vertices.

• A simply generated tree with n vertices (SG_n), denoted by S_n, is a tree chosen proportionally to its weight among the trees with n vertices, i.e.

$$\mathbb{P}(S_n = \tau) = \frac{\omega(\tau)}{\sum_{t \in \mathcal{T}_n} \omega(t)}, \ \tau \in \mathcal{T}_n.$$

Random Trees :

• A uniform tree with n vertices is a tree chosen uniformly at random among the trees with n vertices.

• A simply generated tree with n vertices (SG_n), denoted by S_n, is a tree chosen proportionally to its weight among the trees with n vertices, i.e.

$$\mathbb{P}(S_n = \tau) = \frac{\omega(\tau)}{\sum_{t \in \mathcal{T}_n} \omega(t)}, \ \tau \in \mathcal{T}_n.$$

• Let μ be a probability distribution on $\mathbb{N}$.
A Bienaymé-Galton-Watson tree with offspring distribution μ (μ-BGW) is a random tree where all its individuals reproduce independently of each other according to the distribution μ.

Random Trees :

• A uniform tree with n vertices is a tree chosen uniformly at random among the trees with n vertices.

• A simply generated tree with n vertices (SG_n), denoted by S_n, is a tree chosen proportionally to its weight among the trees with n vertices, i.e.

$$\mathbb{P}(S_n = \tau) = \frac{\omega(\tau)}{\sum_{t \in \mathcal{T}_n} \omega(t)}, \ \tau \in \mathcal{T}_n.$$

• Let μ be a probability distribution on $\mathbb{N}$.
A Bienaymé-Galton-Watson tree with offspring distribution μ (μ-BGW) is a random tree where all its individuals reproduce independently of each other according to the distribution μ.

If $\sum_{j \geq 0} w_j z^j$ has a positif radius of convergence, then a simply generated tree with n vertices can be seen as a $\tilde{\omega}$-BGW for a certain probability $\tilde{\omega}$.

Conditioned random trees

How to make a random tree 'large' ?

Conditioned random trees

How to make a random tree 'large' ?
$\longrightarrow$ Condition it to have n vertices and let $n \to \infty$.

Asymptotic behavior is well understood – even under conditioning by number of leaves or internal nodes.

[Aldous, Duquesne & Le Gall, Kortchemski, Rizzolo]

Conditioned random trees

How to make a random tree 'large' ?

$\longrightarrow$ Condition it to have n vertices and let $n \to \infty$.

Asymptotic behavior is well understood – even under conditioning by number of leaves or internal nodes.

[Aldous, Duquesne & Le Gall, Kortchemski, Rizzolo]

Biconditioned random trees

How to make a random tree 'large' while controlling its 'shape' (perimeter, volume...) ?

Conditioned random trees

How to make a random tree 'large' ?
$\longrightarrow$ Condition it to have n vertices and let $n \to \infty$.

Asymptotic behavior is well understood – even under conditioning by number of leaves or internal nodes.

[Aldous, Duquesne & Le Gall, Kortchemski, Rizzolo]

Biconditioned random trees

How to make a random tree 'large' while controlling its 'shape' (perimeter, volume...) ?
$\longrightarrow$ Condition it to have n vertices and k_n leaves or internal nodes, with $n \to \infty$.

Asymptotic behavior remains largely unexplored.

[Labarbe & Marckert, Kargin, Kortchemski & Marzouk]

Introduction to the objects

Conditioned random trees

How to make a random tree 'large' ?

$\longrightarrow$ Condition it to have n vertices and let $n \to \infty$.

Asymptotic behavior is well understood – even under conditioning by number of leaves or internal nodes.

[Aldous, Duquesne & Le Gall, Kortchemski, Rizzolo]

Biconditioned random trees

How to make a random tree 'large' while controlling its 'shape' (perimeter, volume...) ?

$\longrightarrow$ Condition it to have n vertices and k_n leaves or internal nodes, with $n \to \infty$.

Asymptotic behavior remains largely unexplored.

[Labarbe & Marckert, Kargin, Kortchemski & Marzouk]

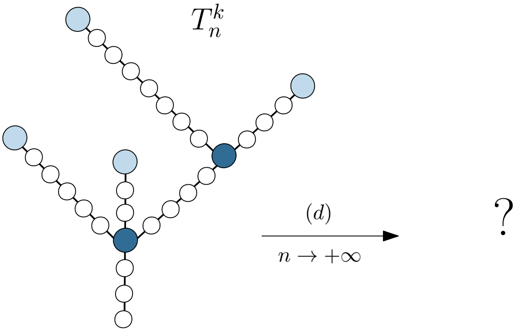
My work : Study BGW trees conditioned on n vertices and k_n leaves/internal nodes.

In this presentation : Focus on the case $k_n := k$.

Limit BGW tree with n vertices and k leaves

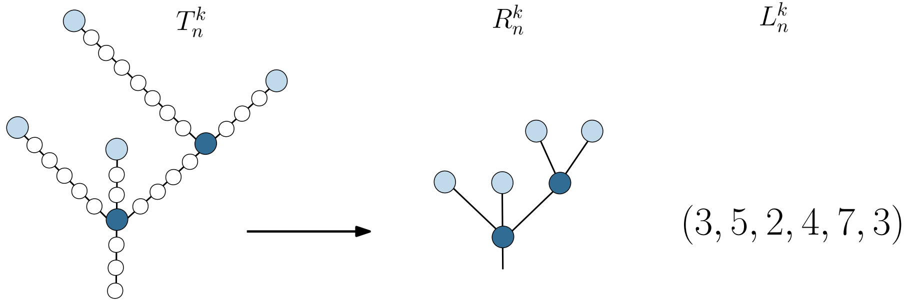
Notations : • k is a fixed integer
• μ a probability distribution such that $\mu(1) > 0$ and $\mu(2) > 0$
• T_n^k is a μ-BGW conditioned to have n vertices and k leaves

Goal : Study the limiting behavior of T_n^k as $n \to \infty$.



Limit BGW tree with n vertices and k leaves

A bijection



Study the limit behavior of T_n^k $\Longleftrightarrow$ Study the limit behavior of R_n^k and L_n^k

Limit BGW tree with n vertices and k leaves

Proposition

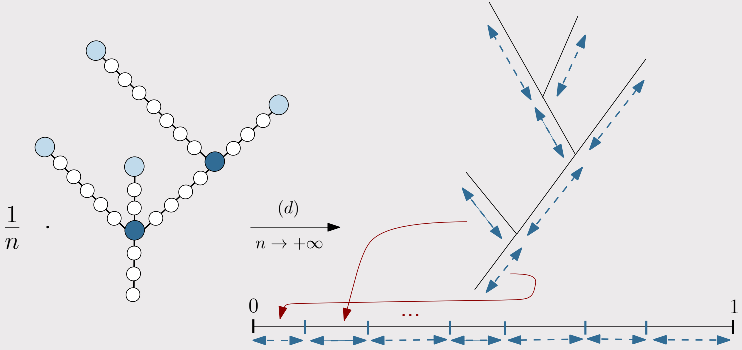
- $R_n^k \xrightarrow[n\to\infty]{(d)}$ Uniform binary tree with k leaves

- $\dfrac{L_n^k}{n} \xrightarrow[n\to\infty]{(d)}$ List of lengths of the intervals in an increasing reordering of $2k-2$ independent uniforms on $[0,1]$

Limit BGW tree with n vertices and k leaves

Theorem

Let μ be a probability distribution on $\mathbb{N}$ such that $\mu(1) > 0$ and $\mu(2) > 0$. Every $\mu-$BGW conditioned to have n vertices and k leaves converges in distribution, as n tends to infinity, to a uniform binary tree with k leaves, where the branch lengths (viewed in lexicographic order) normalized by n correspond to the lengths of the intervals formed by $2k-2$ uniforms on $[0,1]$.



Limit BGW tree with n vertices and k leaves

Theorem

Let μ be an offspring distribution with $\mu(1) > 0$ and fix $k \geq 2$ with $k \in \mathcal{K}(\mu)$. Then, we have the following convergence in distribution :

$$\left(R_n^k, \frac{L_n^k}{n}\right) \xrightarrow[n\to+\infty]{(d)} (R^k, \Delta)$$

where R^k is a uniform random tree in $\mathbb{T}^k(\mu)$ and Δ is a Dirichlet random variable with parameter $(1, \ldots, 1)$ independent of R^k.

Limit BGW tree with n vertices and k leaves

Theorem

Let μ be an offspring distribution with $\mu(1) > 0$ and fix $k \geq 2$ with $k \in \mathcal{K}(\mu)$. Then, we have the following convergence in distribution :

$$\left(R_n^k, \frac{L_n^k}{n}\right) \xrightarrow[n\to+\infty]{(d)} (R^k, \Delta)$$

where R^k is a uniform random tree in $\mathbb{T}^k(\mu)$ and Δ is a Dirichlet random variable with parameter $(1, \ldots, 1)$ independent of R^k.

- $\mathcal{K}(\mu) := \{1 + \sum_{\substack{i\in \mathrm{Supp}(\mu)\\ i>0}} b_i(i-1) : b_i \in \mathbb{N}\}$
- $\mathbb{T}^k(\mu)$ the set of trees having $b_j^{\max}$ vertices with j children, where $(b_j^{\max})_{j\in\mathrm{Supp}(\mu)}$ is the sequence that maximizes $\sum_{j\in\mathrm{Supp}(\mu)} b_j$ under the constraint

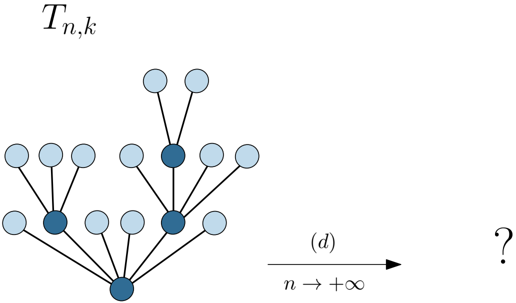
"$b_1 = 0$ and $\sum_{\substack{i\in \mathrm{Supp}(\mu)\\ i>0}} b_i(i-1) = k-1$".

Limit BGW tree with n vertices and k internal nodes

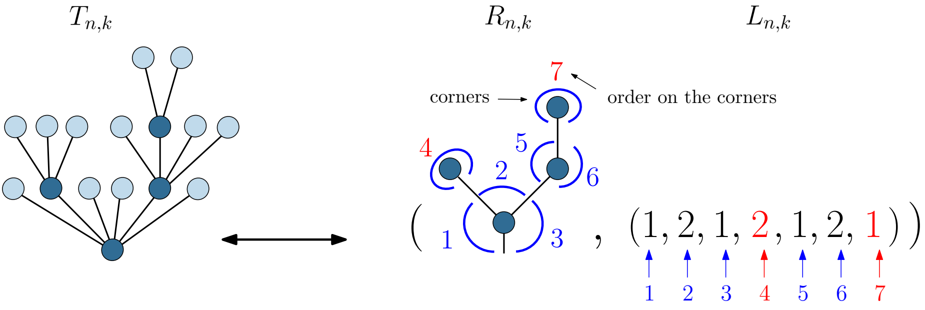
Limit BGW tree with n vertices and k i.n

Notations :
- k is a fixed integer
- $T_{n,k}$ is a μ-BGW conditioned to have n vertices and k internal nodes.

Goal : Study the limiting behavior of $T_{n,k}$ as $n \to \infty$.



A bijection

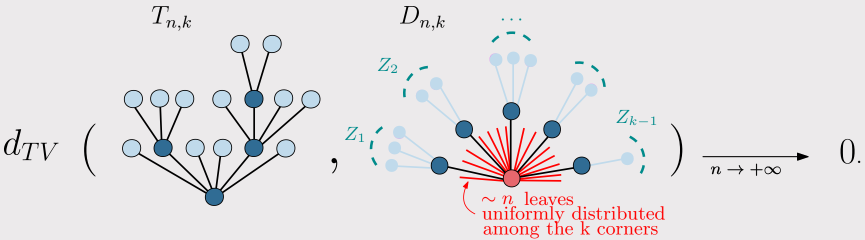


Study the limit behavior of $T_{n,k}$ $\iff$ Study the limit behavior of $R_{n,k}$ and $L_{n,k}$

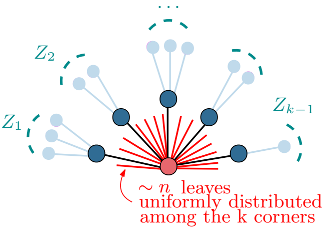
Theorem

If $\mu(i) = l(i)/i^{1+\beta}$ where l is slowly varying and $\beta > 1$, then we have

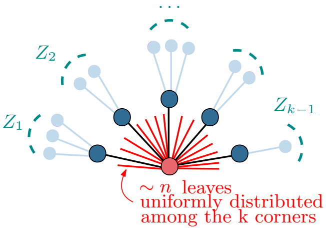
$$d_{TV}(T_{n,k}, D_{n,k}) \xrightarrow[n\to+\infty]{} 0$$



Main steps of the proof :

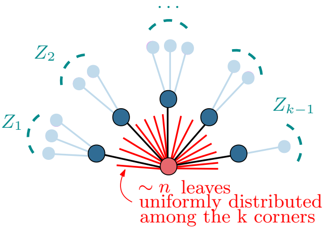


Main steps of the proof :



Step 1 : The outdegrees of the internal nodes of $T_{n,k}$ correspond to the increments of a random walk conditioned to reach a large value in a few steps.

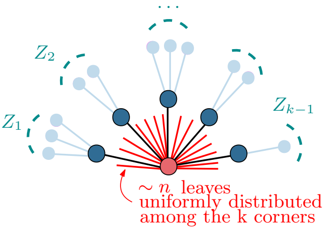
Main steps of the proof :



Step 1 : The outdegrees of the internal nodes of $T_{n,k}$ correspond to the increments of a random walk conditioned to reach a large value in a few steps.

Step 2 : For large n, the vertex with the highest degree will have degree of order n, while the degrees of all other vertices remain bounded.

Main steps of the proof :

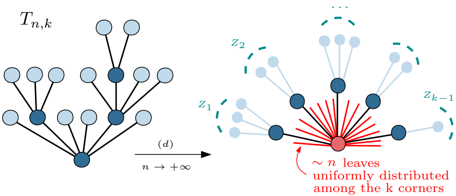


Step 1 : The outdegrees of the internal nodes of $T_{n,k}$ correspond to the increments of a random walk conditioned to reach a large value in a few steps.

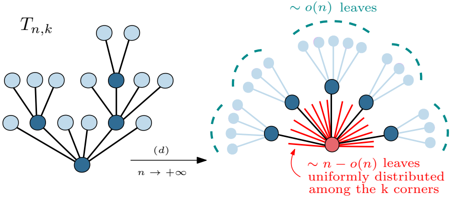
Step 2 : For large n, the vertex with the highest degree will have degree of order n, while the degrees of all other vertices remain bounded.

Step 3 : With high probability, $R_{n,k}$ converges in distribution to a star and the root is the vertex with the largest degree.

$\mu(i) = \frac{l(i)}{i^{1+\beta}}$, l slowly varying, $\beta > 1$

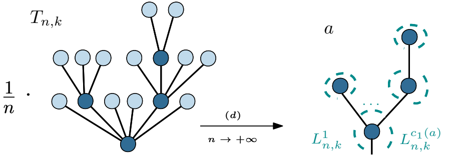


$\mu([i,\infty[) = \frac{l(i)}{i^{\alpha}}$, $l \in \mathcal{L}$, $1 < \alpha < 2$



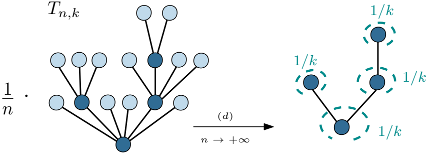
F_μ Δ-analytic and $F_\mu(z) \underset{z \to \rho}{\sim} \frac{c}{(1-z/\rho)^{\alpha}}$

- $R_{n,k} \xrightarrow[n \to \infty]{(d)} SG_k$ where $\omega(k) = \dfrac{\Gamma(k+\alpha)}{\Gamma(k+1)}$
- $\dfrac{L_{n,k}}{n} \xrightarrow[n \to \infty]{(d)} \mathbf{Dir}\left(\left(\frac{c_1(a)+\alpha}{c_1(a)+1}\right)_{i \le c_1(a)+1}, \dots, \left(\frac{c_k(a)+\alpha}{c_k(a)+1}\right)_{i \le c_k(a)+1}\right)$



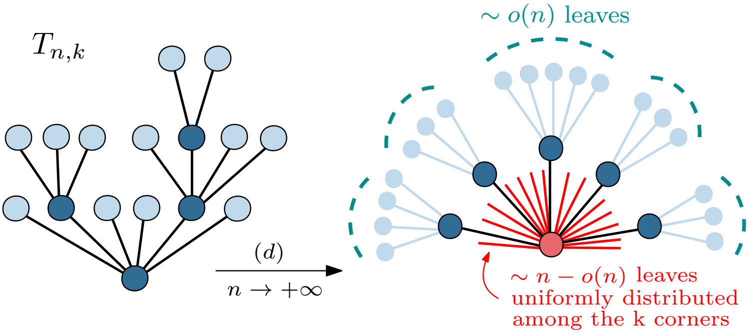
$F_\mu(z) = ce^{P(z)}$, P with non-negative and aperiodic coefficients, $c > 0$

- $R_{n,k} \xrightarrow[n \to \infty]{(d)} \mathcal{P}ois(1)$ - BGW$_k$



Limit BGW tree with n vertices and k i.n

$$\mu([i,\infty[) = \frac{l(i)}{i^\alpha}, \qquad \begin{array}{l} l \in \mathcal{L} \\ 1 < \alpha < 2 \end{array}$$



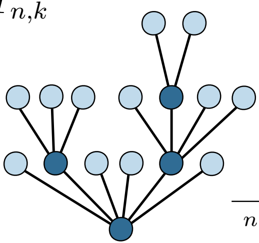
Limit BGW tree with n vertices and k i.n

$$F_\mu(z) = ce^{P(z)}, \quad \begin{array}{l} P \text{ with non-negative and aperiodic coefficients} \\ c > 0 \end{array}$$

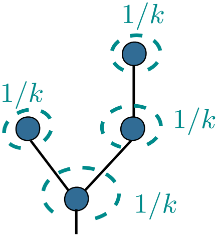
- $R_{n,k} \xrightarrow[n \to \infty]{(d)} \mathcal{P}ois(1)\text{- BGW}_k$

- $T_{n,k}$

$$\frac{1}{n} \cdot$$

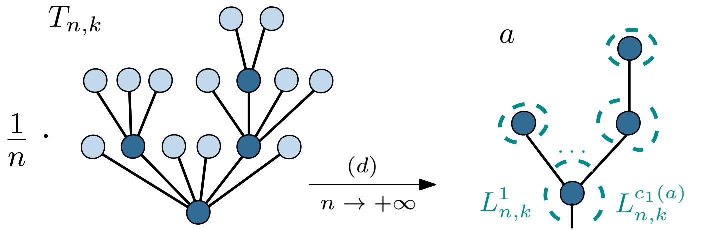


$$\xrightarrow[n \to +\infty]{(d)}$$



$$F_\mu \ \Delta\text{-analytic} \quad \text{and} \quad F_\mu(z) \underset{z \to \rho}{\sim} \frac{c}{(1-z/\rho)^\alpha}$$

- $R_{n,k} \xrightarrow[n \to \infty]{(d)} SG_k$ where $\omega(k) = \dfrac{\Gamma(k+\alpha)}{\Gamma(k+1)}$
- $\dfrac{L_{n,k}}{n} \xrightarrow[n \to \infty]{(d)} \mathbf{Dir}\left(\left(\dfrac{c_1(a)+\alpha}{c_1(a)+1}\right)_{i \leq c_1(a)+1}, \dots, \left(\dfrac{c_k(a)+\alpha}{c_k(a)+1}\right)_{i \leq c_k(a)+1}\right)$



Transfer case : Main steps of the proof

Step 1 : Description of $R_{n,k}$ via generating functions : For all $a \in \mathbb{T}_k$,

$$\mathbb{P}(R_{n,k} = a) = \frac{\zeta_k(a)[z^{n-k-\phi_0(a)}]G_\mu^a(z)}{\sum_{b\in\mathbb{T}_k} \zeta_k(b)[z^{n-k-\phi_0(b)}]G_\mu^b(z)},$$

Transfer case : Main steps of the proof

Step 1 : Description of $R_{n,k}$ via generating functions : For all $a \in \mathbb{T}_k$,

$$\mathbb{P}(R_{n,k} = a) = \frac{\zeta_k(a)[z^{n-k-\phi_0(a)}]G_\mu^a(z)}{\sum_{b\in\mathbb{T}_k} \zeta_k(b)[z^{n-k-\phi_0(b)}]G_\mu^b(z)},$$

where

- $G_\mu^a(z) := \prod_{u\notin I(a)} \widetilde{F}_\mu(z) \prod_{u\in I(a)} F_\mu^{(c_u(a))}(z)$
- $\widetilde{F}_\mu$ the generating function of $\mu(\cdot + 1)$
- $\zeta_k(a) := (\prod_{u\in I(a)} c_u(a)!)^{-1}$

Transfer case : Main steps of the proof

Step 1 : Description of $R_{n,k}$ via generating functions : For all $a \in \mathbb{T}_k$,

$$\mathbb{P}(R_{n,k} = a) = \frac{\zeta_k(a)[z^{n-k-\phi_0(a)}]G_\mu^a(z)}{\sum_{b\in\mathbb{T}_k} \zeta_k(b)[z^{n-k-\phi_0(b)}]G_\mu^b(z)},$$

where • $G_\mu^a(z) := \prod_{u\notin I(a)} \widetilde{F}_\mu(z) \prod_{u\in I(a)} F_\mu^{(c_u(a))}(z)$

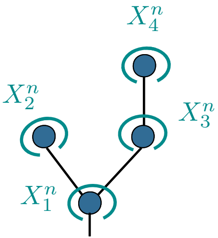
• $\widetilde{F}_\mu$ the generating function of $\mu(\cdot + 1)$

• $\zeta_k(a) := (\prod_{u\in I(a)} c_u(a)!)^{-1}$

Step 2 : $\mathbb{P}(R_{n,k} = a) \xrightarrow[n\to\infty]{} \frac{\prod_{u=1}^k w_u(a)}{\sum_{b\in\mathbb{T}_k} \prod_{u=1}^k w_u(b)}, \quad w_u(a) = \frac{\Gamma(\alpha + c_u(a))}{\Gamma(1 + c_u(a))}$

Transfer case : Main steps of the proof

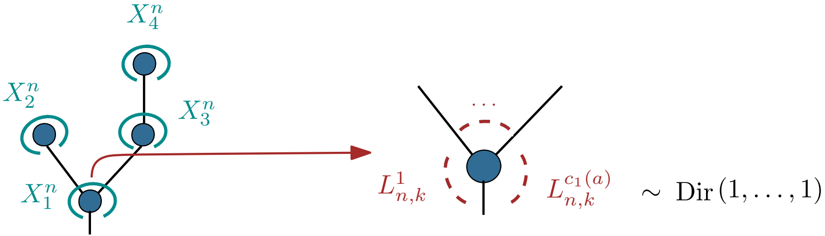
Step 3 : $X_i^n :=$ number of leaves attached to the i-th internal node of $T_{n,k}$



$$\frac{\left(X_1^n, \ldots, X_k^n\right)}{n} \xrightarrow[n\to+\infty]{(d)} \mathbf{Dir}\left(\alpha + c_1(R_{n,k}), \ldots, \alpha + c_k(R_{n,k})\right)$$

Transfer case : Main steps of the proof

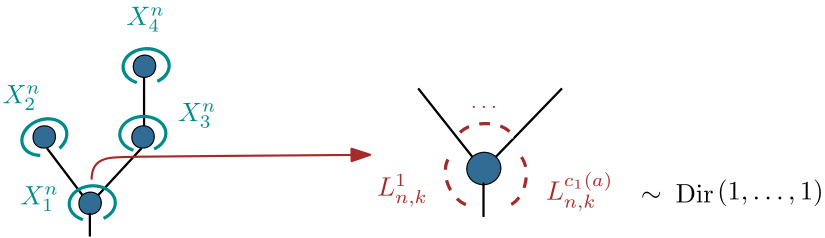
Step 3 : X_i^n := number of leaves attached to the i-th internal node of $T_{n,k}$



$$\frac{(X_1^n,\ldots,X_k^n)}{n} \xrightarrow[n\to+\infty]{(d)} \mathbf{Dir}\big(\alpha + c_1(R_{n,k}),\ldots,\alpha + c_k(R_{n,k})\big)$$

Transfer case : Main steps of the proof

Step 3 : X_i^n := number of leaves attached to the i-th internal node of $T_{n,k}$



$$\frac{(X_1^n,\dots,X_k^n)}{n} \xrightarrow[n\to+\infty]{(d)} \mathbf{Dir}\big(\alpha + c_1(R_{n,k}),\dots,\alpha + c_k(R_{n,k})\big)$$

Step 4 : Conditionally given $R_{n,k} = a$,

$$\frac{L(T_{n,k})}{n} \xrightarrow[n\to+\infty]{(d)} \mathbf{Dir}\left(\left(\frac{c_1(a)+\alpha}{c_1(a)+1}\right)_{1\le i\le c_1(a)+1},\dots,\left(\frac{c_k(a)+\alpha}{c_k(a)+1}\right)_{1\le i\le c_k(a)+1}\right)$$

Thank you for your attention !

References

Armendáriz, I. and Loulakis, M. (2011).
Conditional distribution of heavy tailed random variables on large deviations of their sum.
Stochastic processes and their applications, 121(5) :1138–1147.

Bingham, Goldie, and Teugels (1987).
Regular variation.
Encyclopedia of Mathematics and its Applications.

Björnberg, J. and Stefánsson, S. Ö. (2015).
Random walk on random infinite looptrees.
Journal of Statistical Physics, 158 :1234–1261.

Flajolet, P. and Odlyzko, A. (1990).
Singularity analysis of generating functions.
SIAM Journal on discrete mathematics, 3(2) :216–240.

Lamperti, J. (1958).
An occupation time theorem for a class of stochastic processes.
Transactions of the American Mathematical Society, 88(2) :380–387.

Thévenin, P. (2020).
Vertices with fixed outdegrees in large galton-watson trees.