Extremal Properties of Random Trees

Eli Ben-Naim
Los Alamos National Lab

I Tree generation model

II Heuristics derivation

III Minimal height distribution

IV Distribution of rank

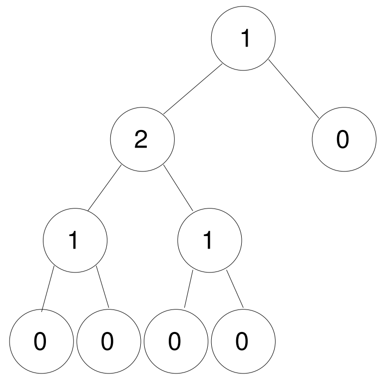
V Conclusions

Paul Krapivsky (Boston)
Satya Majumdar (Toulouse/Tata)

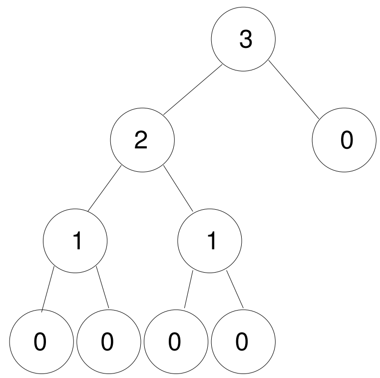
Phys. Rev. E **64**, 035101 (2001)

Extremal Characteristics

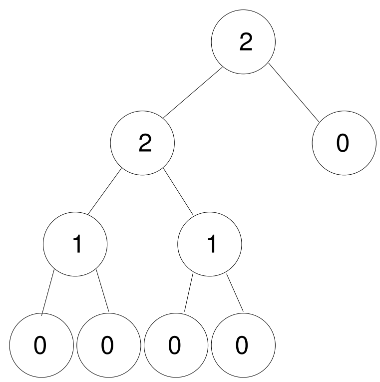
- Minimal height: $i + j \rightarrow \min(i, j) + 1$



- Maximal height: $i + j \rightarrow \max(i, j) + 1$



- Rank: $i + j \rightarrow \max(i, j) + \delta_{i,j}$



Motivation

- Computer Science: Data storage algorithms.
  min/max = best/worst case performance Devroye 86
- Geophysics: Characterization of river networks
  rank=Horton-Strahler index Rothman 98
- Physics: Collision processes in gases van Beijeren 98
  max ≈ largest Lyapunov exponent in gas.

Questions

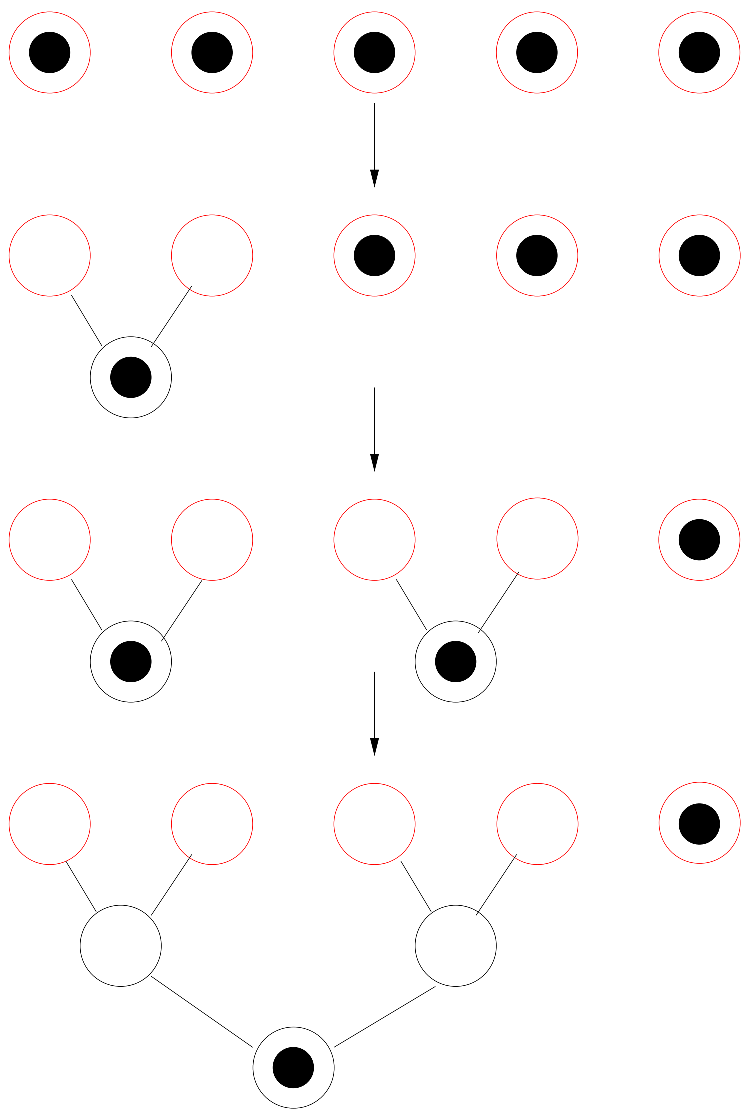
- Growth of average

$$\langle k_{\min} \rangle \simeq v_{\min} \ln N$$

- Distribution of extremal characteristics

$$\begin{aligned} P(k) &= ? \\ \sigma &= ? \end{aligned}$$

Tree generation process



Random binary trees

Recursive tree generation process:

- Initial conditions: System consists of trivial trees.
- Merging process: Pick two roots at random and join them (with fixed rate $= 1$).

Rate equation theory:

- Average over ∞ number of realizations
- Limit of ∞ number of trees at $t = 0$

The number density

- $c(t)$= number at t/number at $t = 0$.
- Rate equation for $c(t)$: $\frac{d}{dt}c(t) = -c^2(t)$
- Initial conditions: $c(0) = 1$

$$c(t) = (1 + t)^{-1} \qquad N(t) = c^{-1} = (1 + t).$$

Heuristic derivation of averages

- Height distribution obeys Poisson statistics:
  $P_n(t) =$ probability that distance from a randomly chosen leaf to root $= n$ at time t.

$$P_n(t) = \frac{[h(t)]^n}{n!} e^{-h(t)}.$$

- Growth of average height: $\frac{d}{dt} h(t) = 2c$

$$h(t) = 2 \ln N.$$

- Assume minimal height: $\langle k_{\min} \rangle \simeq v_{\min} \ln N$
- Estimating $\langle k_{\min} \rangle$ from tail of $P_n(t)$:

$$\sum_{n=0}^{\langle k_{\min} \rangle} P_n(t) \sim N^{-1}$$

- Equation for v

$$v \ln \frac{2e}{v} = 1$$

- Two roots correspond to $v_{\min}$ and $v_{\max}$

$$v_{\min} = 0.373365 \qquad v_{\max} = 4.31107$$

Distribution of minimal height

- Distribution of minimal height:
  $c_k(t)$=fraction of trees with minimal height k

- Evolution equation for $c_k \equiv c_k(t)$

$$\frac{d}{dt}c_k = c_{k-1}^2 + 2c_{k-1}\sum_{j=k}^{\infty} c_j - 2cc_k$$

- Use cumulative distribution A_k & time τ

$$A_k = c^{-1}\sum_{j=k}^{\infty} c_j \qquad \tau = \ln N$$

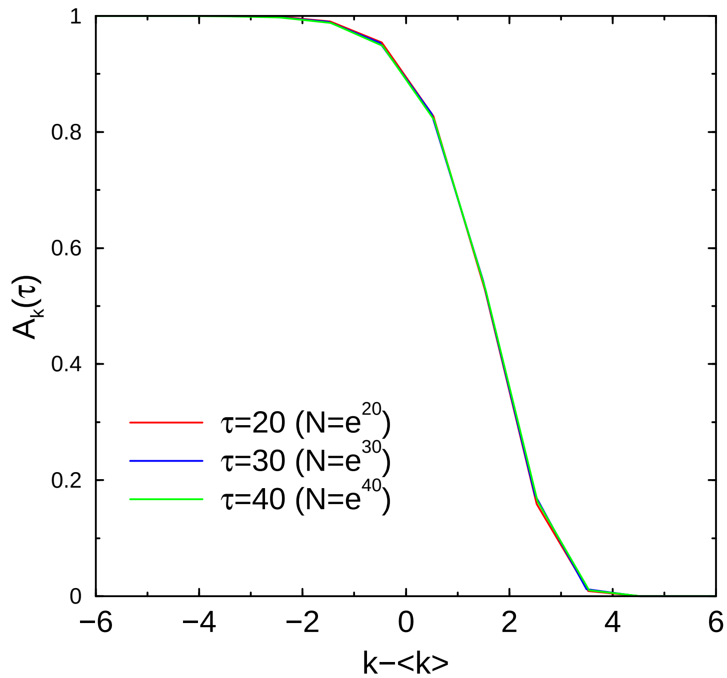
- Nonlinear difference-differential equations

$$\frac{d}{d\tau}A_k = A_{k-1}^2 - A_k$$

- Step function initial conditions

$$A_k(0) = \begin{cases} 1 & k \le 0; \\ 0 & k > 0. \end{cases}$$

Traveling wave solution



- In the asymptotic (large tree) limit

$$\begin{aligned} A_k(\tau) &\rightarrow F(k - v\tau) \qquad\qquad \tau \rightarrow \infty \\ vF'(x) &= F(x) - F^2(x-1). \end{aligned}$$

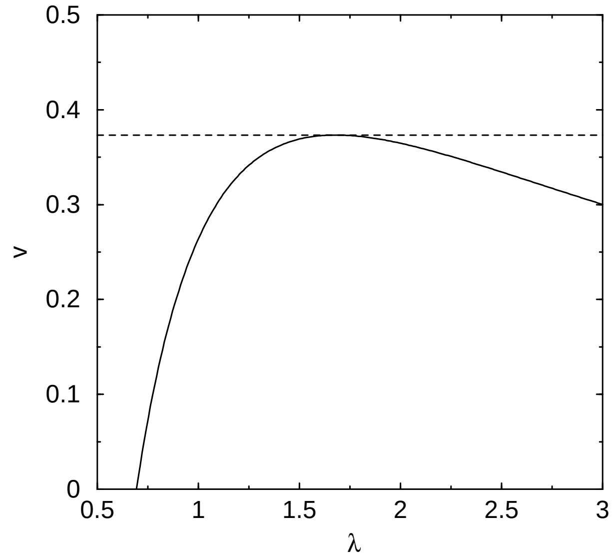
- Linear equations for the tails give:

$$F(x) \cong \begin{cases} 1 - \text{const} \times \exp(\lambda x) & x \rightarrow -\infty; \\ \text{const} \times \exp(-x/v) & x \rightarrow \infty. \end{cases}$$

Exponentially suppressed fluctuations/tails

Velocity selection

KPP/Fisher 37, Bramson 83



- Small height tail gives

$$v = \frac{1 - 2e^{-\lambda}}{\lambda}$$

- Extremum satisfies ($v_{\min} = 0.373365$)

$$v \ln \frac{2e}{v} = 1$$

- v agrees to within 0.1% with numerics

Small height tail dictates velocity

Distribution of maximal height

- Rate equation for distribution

$$\frac{d}{dt}c_k = c_{k-1}^2 + 2c_{k-1}\sum_{j=0}^{k-2} c_j - 2cc_k$$

- Equations for cumulative distribution

$$\frac{d}{d\tau}A_k = -A_k + 2A_{k-1} - A_{k-1}^2$$

- Wave form $A_k(\tau) \to F(k - v\tau)$ as $\tau \to \infty$

$$vF'(x) = F(x) - 2F(x-1) + F^2(x-1)$$

- Exponential tails

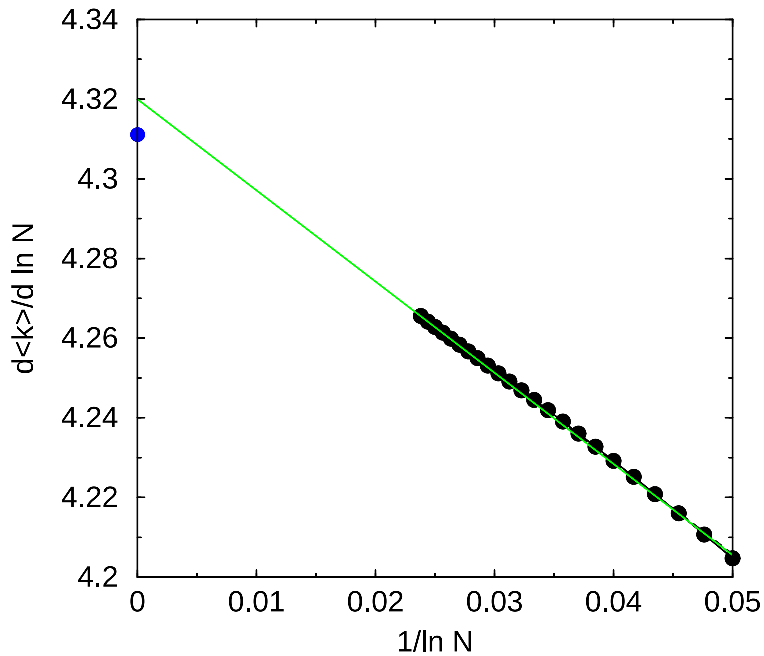
$$F(x) \cong \begin{cases} 1 - \text{const} \times \exp(x/v) & x \to -\infty; \\ \text{const} \times \exp(-\mu x) & x \to \infty. \end{cases}$$

- Dispersion relation $v = (2e^{\mu} - 1)/\mu$

- Minimum selected $v \ln \frac{2e}{v} = 1$ $v = 4.31107$

Large height tail dictates velocity

Leading asymptotic correction



Keeping next leading term in front location and balancing leading terms gives Derrida 97

$$\langle k_{\min} \rangle = v_{\min} \ln N + \frac{3}{2\lambda} \ln \ln N$$

$$\langle k_{\max} \rangle = v_{\max} \ln N - \frac{3}{2\mu} \ln \ln N$$

Correction to averages is $\mathcal{O}(\ln \ln N)$

Distribution of rank

- Rate equation for distribution

$$\frac{d}{dt}c_k = c_{k-1}^2 - 2c_k \sum_{j=k}^{\infty} c_j$$

- Wave form satisfies

$$vF'(x) = 2F(x)F(x-1) - F(x) - F^2(x-1).$$

- Nonlinear plays role even at the tail!
$F(x) \propto F^2(x-1)$

$$F(x) \cong \begin{cases} 1 - \text{const} \times \exp(x/v) & x \to -\infty; \\ [4\beta v \ln 2] 2^x \exp(-\beta 2^x) & x \to \infty. \end{cases}$$

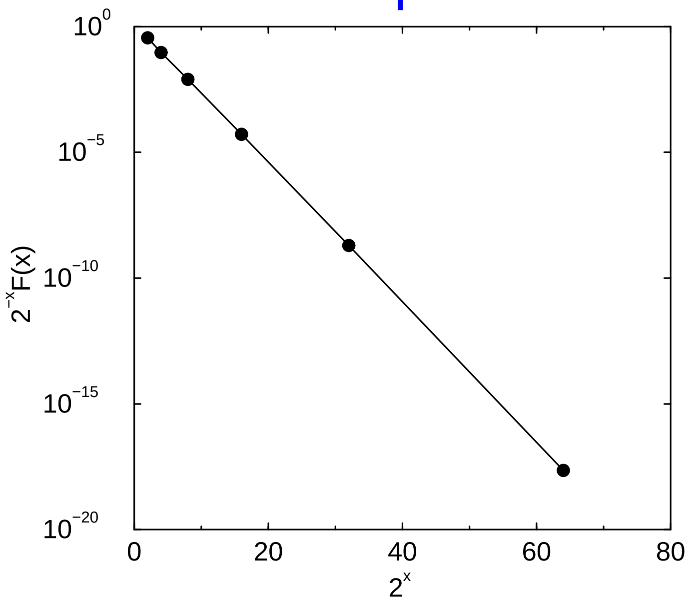
- Velocity is elusive

$$v = 0.89837$$

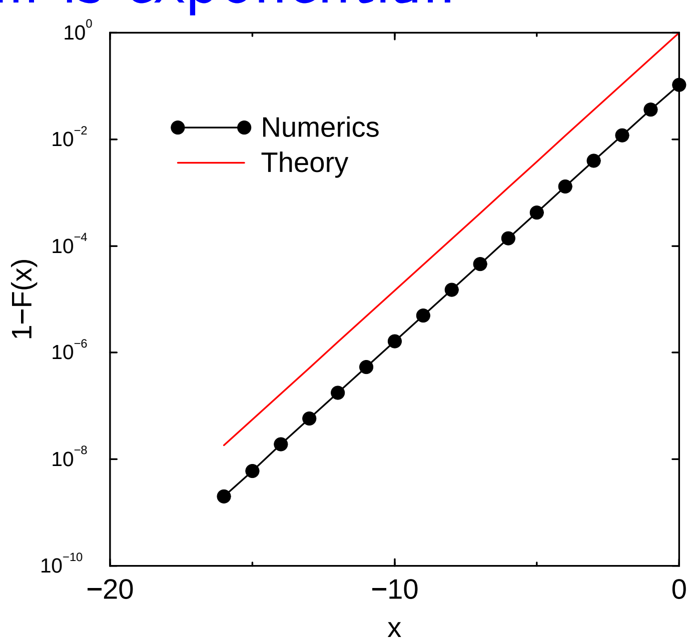
Sharp double exponential large-k tail

Numerical Evidence

Leading tail is double-exponential:



Trailing tail is exponential:



Input

- Initially: system is empty
- Trivial trees (leafs) added at constant rate
- Add input term $\delta_{k,1}$ to rate equations and solve for steady state

- Minimal height – double exponential decay

$$c_k \sim 2^{-2^k}$$

- Maximal height – power law decay

$$c_k \sim k^{-2}$$

- Rank – purely exponential distribution

$$c_k = 2^{-k}$$

Very different behaviors

Conclusions

- Distribution of extremal characteristics approaches traveling wave form asymptotically
- Fluctuations (σ) are finite
- Extremal velocity selection criteria determines growth of averages
- Extreme statistics are exponential

Outlook

- Complete distribution functions
- Selection criteria for rank velocity