Branching Distributional Equations and their Applications

Mariana Olvera-Cravioto

UNC Chapel Hill

molvera@unc.edu

August 24th, 2018

Analyzing the attracting endogenous solution

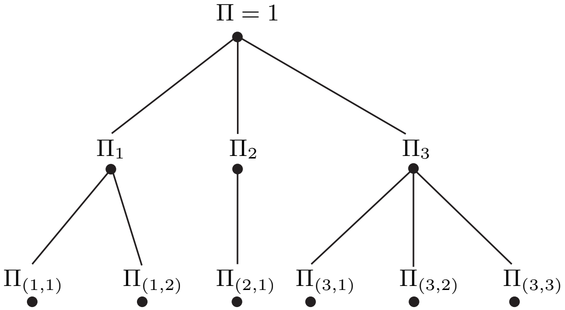
- Consider the non-homogeneous smoothing transform:

$$R \stackrel{\mathcal{D}}{=} \sum_{i=1}^{N} C_i R_i + Q,$$

  where the $\{R_i\}$ are i.i.d. copies of R independent of $(Q, N, \{C_i\})$.
- Let R be the attracting endogenous solution.
- **Goal:** Analyze $P(R > t)$ as $t \to \infty$.
- In the context of scale-free graphs, explain why PageRank and the in-degree follow the same power law.

Weighted branching trees

- Number of offspring N, generic branching vector $(Q, N, C_1, C_2, \dots)$.
- The weighted branching tree $\mathcal{T}$:



- Each node in the tree has a weight $\Pi_{(i_1,\dots,i_n)}$ defined via the recursion

$$\Pi_{i_1} = C_{i_1}, \qquad \Pi_{(i_1,\dots,i_n)} = C_{(i_1,\dots,i_n)} \Pi_{(i_1,\dots,i_{n-1})}, \qquad n \geq 2,$$

and $\Pi = 1$ is the weight of the root node.

The attracting endogenous solution

- **Notation:** Write $\mathbf{i} = (i_1, \ldots, i_k)$ and $(\mathbf{i}, j) = (i_1, \ldots, i_k, j)$.
- Let $A_0 = \{\emptyset\}$ and

$$A_k = \{(\mathbf{i}, i_k) : \mathbf{i} \in A_{k-1},\, 1 \leq i_k \leq N_{\mathbf{i}}\}, \qquad k \geq 1.$$

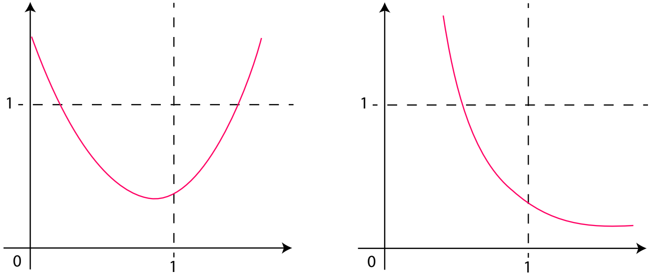
- In terms of the weighted branching tree, the attracting endogenous solution to the non-homogeneous smoothing transform is:

$$R = \sum_{k=0}^{\infty} \sum_{\mathbf{i} \in A_k} \Pi_{\mathbf{i}} Q_{\mathbf{i}}$$

- In general, this may not be the only solution...

A quick note on multiple solutions

- Define $m(\theta) = E\left[\sum_{i=1}^{N} |C_i|^{\theta}\right]$.
- A sufficient condition for $|R| < \infty$ is the existence of $\beta \in (0, 1]$ such that $m(\beta) < 1$.



- **Theorem:** (Alsmeyer-Meiners '10) If there exist $\beta \in (0, 1]$ and $\alpha > \beta$ such that $1 = m(\beta) < m(\theta)$ for $\theta \in [0, \beta)$ and $m(\alpha) = 1$, then there are multiple solutions.

Three heavy-tailed settings

- Linear nonhomogeneous equation: $R \stackrel{\mathcal{D}}{=} \sum_{i=1}^{N} C_i R_i + Q$
- Jelenković, O-C, '12, O-C '12: Solution R can have heavy tails if $(C_1, C_2, \dots)$ is nonnegative and:
  - $E\left[\sum_{i=1}^{N} C_i^{\alpha}\right] = 1$ for some $\alpha > 0$.
  - $P\left(\sum_{i=1}^{N} C_i > t\right) \sim t^{-\alpha} L(t)$ $\quad$ ($P(N > t) \sim t^{-\alpha} L(t)$ and $\{C_i\}$ i.i.d.)
  - $P(Q > t) \sim t^{-\alpha} L(t)$
- Tail of R in each case:
  - $P(R > t) \sim H t^{-\alpha}$
  - $P(R > t) \sim H P\left(\sum_{i=1}^{N} C_i > t\right)$ $\quad$ ($P(R > t) \sim H P(N > t)$)
  - $P(R > t) \sim H P(Q > t)$

Implicit Renewal Theory

- When $N \equiv 1$ and $E[|C|^\alpha] = 1$, $E[|C|^\alpha \log^+ |C|] < \infty$, Goldie's Implicit Renewal Theorem (1991) gives the asymptotics of the solutions to

$$R \stackrel{\mathcal{D}}{=} CR + Q \quad \text{and} \quad R \stackrel{\mathcal{D}}{=} CR \vee Q.$$

- (Jelenković, O-C, '12) **Generalization to trees:**
  - Natural condition $E\left[\sum_{i=1}^N |C_i|^\alpha\right] = 1$.
  - For nonnegative weights the appropriate renewal measure is

$$\eta(dt) = e^{\alpha t} E\left[\sum_{i=1}^N 1(\log C_i \in dt)\right].$$

  - Real valued weights lead to a matrix renewal measure.

Conditions for the theorem

- Let $(N, C_1, C_2, \ldots)$ be a random vector, where $N \in \mathbb{N} \cup \{\infty\}$.
- For the applications of the theorem let $(Q, N, C_1, C_2, \ldots)$ be a random vector, where $N \in \mathbb{N} \cup \{\infty\}$, and $P(|Q| > 0) > 0$.
- Suppose the $\{\log C_j\}$ are nonlattice.
- Let R be a real valued random variable independent of $(N, C_1, C_2, \ldots)$.
- The theorem can be applied to many recursions of the form $R \stackrel{\mathcal{D}}{=} \Phi\left(Q, N, \{C_i\}, \{R_i\}\right)$, e.g.

$$R \stackrel{\mathcal{D}}{=} \left(\bigvee_{i=1}^{N} C_i R_i\right) \vee Q \qquad \text{and} \qquad R \stackrel{\mathcal{D}}{=} \left(\bigvee_{i=1}^{N} C_i R_i\right) + Q$$

Implicit Renewal Theorem for Trees

- **Theorem:** Assume $0 < \mu \triangleq E\left[\sum_{j=1}^{N} |C_j|^\alpha \log |C_j|\right] < \infty$ and $E\left[\sum_{j=1}^{N} |C_j|^\alpha\right] = 1$.
  - If $\{C_i\} \geq 0$ a.s., $E[(\pm R)^+)^\beta] < \infty$ for any $0 < \beta < \alpha$, and

$$\int_0^\infty \left| P(\pm R > t) - E\left[\sum_{j=1}^{N} 1(\pm C_j R > t)\right]\right| t^{\alpha-1} dt < \infty, \qquad (1)$$

then

$$P(\pm R > t) \sim H_\pm t^{-\alpha}, \qquad t \to \infty,$$

where $0 \leq H_\pm < \infty$ is given by

$$H_\pm = \frac{1}{\mu} \int_0^\infty v^{\alpha-1} \left(P(\pm R > v) - E\left[\sum_{j=1}^{N} 1(\pm C_j R > v)\right]\right) dv.$$

.

Implicit Renewal Theorem for Trees

- **Theorem:** ... continued.
  - If $P(C_j < 0) > 0$ for some $j \geq 1$, $E[|R|^\beta] < \infty$ for any $0 < \beta < \alpha$ and both conditions given by (1) are satisfied, then

$$P(R > t) \sim P(R < -t) \sim Ht^{-\alpha}, \qquad t \to \infty,$$

    where $0 \leq H = (H_+ + H_-)/2 < \infty$.

- **Remarks:**
  - Lattice version, i.e., $\{\log C_i\}$ lattice valued, also available.
  - If (1) holds both for α and $\alpha + \epsilon$, and some other technical conditions are satisfied,

$$\left| P(\pm R > t) - H_\pm t^{-\alpha} \right| = o(t^{-\epsilon})$$

  - This theorem can be applied to various max-plus recursions.

Application to the linear recursion

Theorem: Suppose for some $\alpha > 0$,

- $E[|Q|^\alpha] < \infty$, $0 < E\left[\sum_{i=1}^N |C_i|^\alpha \log |C_i|\right] < \infty$ and $E\left[\sum_{i=1}^N |C_i|^\alpha\right] = 1$.
- $E\left[\sum_{i=1}^N |C_i|\right] < 1$ and $E\left[\left(\sum_{i=1}^N |C_i|\right)^\alpha\right] < \infty$, if $\alpha > 1$; or,
- $E\left[\left(\sum_{i=1}^N |C_i|^{\alpha/(1+\epsilon)}\right)^{1+\epsilon}\right] < \infty$ for some $0 < \epsilon < 1$, if $0 < \alpha \leq 1$.

Then, if $P(C_j < 0) > 0$ for some $j \geq 1$,

$$P(R > t) \sim P(R < -t) \sim Ht^{-\alpha}, \qquad t \to \infty,$$

where $H \geq 0$ is given by

$$H = \frac{E\left[\left|\sum_{i=1}^N C_i R_i + Q\right|^\alpha - \sum_{i=1}^N |C_i R_i|^\alpha\right]}{2\alpha E\left[\sum_{i=1}^N |C_i|^\alpha \log |C_i|\right]}.$$

The PageRank example

- That's nice, but....

The PageRank example

- That's nice, but.... the weights in PageRank do not satisfy the conditions in the theorem.

The PageRank example

- That's nice, but.... the weights in PageRank do not satisfy the conditions in the theorem.
- In particular, $C_i < 1$ for all i, no root $\alpha > 0$ to $E\left[\sum_{i=1}^{N} C_i^{\alpha}\right] = 1$ satisfying $0 < E\left[\sum_{i=1}^{N} C_i^{\alpha} \log C_i\right] < \infty$.

The PageRank example

- That's nice, but.... the weights in PageRank do not satisfy the conditions in the theorem.
- In particular, $C_i < 1$ for all i, no root $\alpha > 0$ to $E\left[\sum_{i=1}^{N} C_i^{\alpha}\right] = 1$ satisfying $0 < E\left[\sum_{i=1}^{N} C_i^{\alpha} \log C_i\right] < \infty$.
- What then?

The PageRank example

- That's nice, but.... the weights in PageRank do not satisfy the conditions in the theorem.
- In particular, $C_i < 1$ for all i, no root $\alpha > 0$ to $E\left[\sum_{i=1}^{N} C_i^{\alpha}\right] = 1$ satisfying $0 < E\left[\sum_{i=1}^{N} C_i^{\alpha} \log C_i\right] < \infty$.
- What then?
  Answer: traditional heavy-tailed analysis.

Finitely many iterations and exchange of limits

- Generic vector $(Q, N, C_1, C_2, \dots)$, weights $\{C_i\} \geq 0$.
- Define $R^{(k)} = \sum_{i=0}^{k} W_i$, then

$$R^{(k)} \stackrel{\mathcal{D}}{=} \sum_{i=1}^{N} C_i R_i^{(k-1)} + Q, \qquad R^{(0)} = Q,$$

  $\{R^{(k-1)}\}$ i.i.d. copies of $R^{(k-1)}$, independent of $(Q, N, C_1, C_2, \dots)$.
- It can be shown that $R^{(k)} \to R$ a.s. as $k \to \infty$.
- Suppose that either

$$P\left(\sum_{i=1}^{N} C_i > x\right) \sim x^{-\alpha} L(x) \qquad \text{or} \qquad P\left(Q > x\right) \sim x^{-\alpha} L(x)$$

Exchange of limits

- Under these conditions we can show that

$$P\left(R^{(k)} > x\right) \sim x^{-\alpha}\hat{L}_k(x)$$

  for any fixed k.
- Moreover, the limits can be exchanged:

$$\begin{array}{ccc}
R^{(k)} & \xrightarrow{k \to \infty} & R \\
\Big\downarrow {\scriptstyle x \to \infty} & & \Big\downarrow {\scriptstyle x \to \infty} \\
P(R^{(k)} > x) \sim x^{-\alpha}\hat{L}_k(x) & \xrightarrow{k \to \infty} & P(R > x) \sim x^{-\alpha}\hat{L}(x)
\end{array}$$

The asymptotics of $P(R > x)$

- **Theorem:** Let $\rho_\beta = E\left[\sum_{i=1}^N C_i^\beta\right]$. If $P\left(\sum_{i=1}^N C_i > x\right) \sim x^{-\alpha} L(x)$, $\alpha > 1$, $E[|Q|^{\alpha+\epsilon}] < \infty$, $\rho_{\alpha+\epsilon} < \infty$, $E[Q] > 0$, and $\rho \vee \rho_\alpha < 1$, then

$$P(R > x) \sim \frac{(E[Q])^\alpha}{(1-\rho)^\alpha (1-\rho_\alpha)} P\left(\sum_{i=1}^N C_i > x\right).$$

- **Theorem:** If $P(Q > x) \sim x^{-\alpha} L(x)$, $\alpha > 1$, $E[|Q|^\beta] < \infty$ for all $0 < \beta < \alpha$, $E\left[\left(\sum_{i=1}^N C_i\right)^{\alpha+\epsilon}\right] < \infty$, and $\rho \vee \rho_\alpha < 1$, then

$$P(R > x) \sim (1-\rho_\alpha)^{-1} P(Q > x).$$

Back to PageRank

- For PageRank: $P(N > x) \sim x^{-\alpha} L(x)$, $\{C_i\} \in [0, 1)$ i.i.d.
- From large deviations we know

$$P\left(\sum_{i=1}^{N} C_i > x\right) \sim (E[C])^{\alpha} P(N > x) \sim (E[C])^{\alpha} L(x) x^{-\alpha}.$$

- Moreover, the proof of the theorem also reveals that

$$\begin{aligned} P(R > x) &\sim P\left(\sum_{i=1}^{N} C_i R_i > x\right) \\ &\sim P\left(\max_{1 \leq i \leq N} C_i R_i > x\right) + P(N > x/E[CR]), \end{aligned}$$

i.e., either one neighbor has very high rank, or, the page has extremely many neighbors!