Branching Distributional Equations and their Applications

Mariana Olvera-Cravioto

UNC Chapel Hill

molvera@unc.edu

August 21st, 2018

Motivation

- Two problems leading to branching distributional equations:
  - Ranking algorithms on complex networks.
  - Parallel queueing systems with synchronization.

Motivation

- Two problems leading to branching distributional equations:
  - Ranking algorithms on complex networks.
  - Parallel queueing systems with synchronization.
- The object of interest is the **attracting endogenous** solution to a certain max-plus branching stochastic fixed-point equation.

Motivation

- Two problems leading to branching distributional equations:
  - Ranking algorithms on complex networks.
  - Parallel queueing systems with synchronization.
- The object of interest is the **attracting endogenous** solution to a certain max-plus branching stochastic fixed-point equation.
- Naïve Monte Carlo techniques are not a good idea...
- **Goal:** provide an efficient simulation algorithm to study them.

Example: PageRank

- Consider a directed graph on n vertices.
- Generalized PageRank assigns to each vertex i a "rank" r_i satisfying:

$$r_i = q_i + \sum_{j \to i} \frac{\zeta_j}{d_j} \cdot r_j,$$

  where q_i is known as the personalization or teleportation value of vertex i, d_j is the out-degree of vertex j, $|\zeta_j| \leq c < 1$, and the sum is taken over all inbound neighbors of vertex i.
- Multiplying both sides by n, we obtain the "scale-free" ranks $R_i = nr_i$.
- The goal is to compute the limiting distribution of the scale-free rank of a randomly chosen vertex.

Example: PageRank

- Consider random digraphs constructed according to either the directed configuration model or the inhomogeneous random digraph with rank-1 kernel.
- The distribution of the scale-free rank of a randomly chosen vertex, R_ξ, converges, as $n \to \infty$, to

$$R^* = Q_0 + \sum_{j=1}^{N_0} X_j,$$

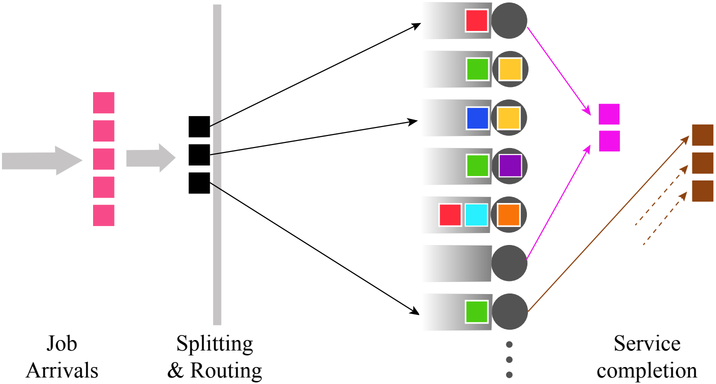
where the $\{X_j\}$ are i.i.d. copies of the attracting endogenous solution to the linear SFPE:

$$X \stackrel{\mathcal{D}}{=} QC + \sum_{j=1}^{N} C X_j$$

Example: Queueing networks

- Consider a network consisting of n computers working in parallel, e.g. a server farm.
- Jobs arrive to the network according to a Poisson process.
- Upon arrival, they are split into several pieces of similar size, with larger jobs having more pieces.
- The pieces are then routed to a random subset of servers to be processed separately.
- **Network rules:**
  - Jobs need to be served on a FCFS basis.
  - All pieces of a job must begin service simultaneously **(Synchronization)**.
- The synchronization causes idleness and blocking.

Example: Queueing networks



Example: Queueing networks

- We study the distribution of the stationary waiting time of a job in the n server network, denoted $W^{(n)}$.
- As the number of servers $n \to \infty$, $W^{(n)}$ converges to the high-order Lindley equation:

$$W \stackrel{\mathcal{D}}{=} \max\left\{0, \max_{1 \le i \le N} (\chi_i - \tau_i + W_i)\right\},$$

where τ_i is exponentially distributed, N denotes the number of fragments of a job, and the $\{\chi_i\}$ are the service requirements of each of the N fragments.

Solutions to branching recursions

- The attracting endogenous solutions to

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

are especially simple, and are given by

$$R = \sum_{\mathbf{i} \in \mathcal{T}} \Pi_{\mathbf{i}} Q_{\mathbf{i}} \qquad \text{and} \qquad R = \bigvee_{\mathbf{i} \in \mathcal{T}} \Pi_{\mathbf{i}} Q_{\mathbf{i}},$$

respectively, where the sum (maximum) is taken over all nodes $\mathbf{i} = (i_1, \ldots, i_n)$ on the tree $\mathcal{T}$.

Finite tree approximation

- Often, the map Φ defines a contraction under a suitable metric.
- More precisely, let

$$R_{\mathbf{i}}^{(j)} = \Phi\left(Q_{\mathbf{i}}, N_{\mathbf{i}}, \{C_{(\mathbf{i},r)}\}, \{R_{(\mathbf{i},r)}^{(j-1)}\}\right), \quad 1 \le j \le k,$$

  for $\mathbf{i}$ in generation $k-j$, and $R_{\mathbf{i}}^{(0)} = Q_{\mathbf{i}}$ for $\mathbf{i}$ in generation k.
- Define μ_k to be the probability measure of $R^{(k)}$ and μ that of the attracting endogenous solution R.
- In many cases we can show that there exist a metric d and a constant $0 < c < 1$ such that

$$d(\mu_k, \mu) \le c^k d(\mu_0, \mu).$$

- Hence, it suffices to compute the distribution of $R^{(k)} = R_{\emptyset}^{(k)}$ for large k.

A simulation approach

- In principle, we could simulate a tree for k generations (k large) and use

$$R \approx R^{(k)}.$$

- For the linear and maximum recursion this would imply computing

$$R^{(k)} = \sum_{\mathbf{i} \in \mathcal{T}, |\mathbf{i}| \leq k} \Pi_{\mathbf{i}} Q_{\mathbf{i}} \qquad \text{and} \qquad R^{(k)} = \bigvee_{\mathbf{i} \in \mathcal{T}, |\mathbf{i}| \leq k} \Pi_{\mathbf{i}} Q_{\mathbf{i}},$$

where $|(i_1, \ldots, i_n)| = n$.

A simulation approach

- In principle, we could simulate a tree for k generations (k large) and use
$$R \approx R^{(k)}.$$
- For the linear and maximum recursion this would imply computing
$$R^{(k)} = \sum_{\mathbf{i} \in \mathcal{T}, |\mathbf{i}| \leq k} \Pi_{\mathbf{i}} Q_{\mathbf{i}} \qquad \text{and} \qquad R^{(k)} = \bigvee_{\mathbf{i} \in \mathcal{T}, |\mathbf{i}| \leq k} \Pi_{\mathbf{i}} Q_{\mathbf{i}},$$
where $|(i_1, \ldots, i_n)| = n$.
- **Problem:** This approach requires an average of $(E[N])^k$ i.i.d. copies of $(Q, N, \{C_i\})$!

The Population Dynamics algorithm

1. *Initialize*: Set $j = 0$. Simulate a sequence $\{Q_i\}_{i=1}^m$ of i.i.d. copies of Q and let $\hat{R}_i^{(0,m)} = Q_i$ for $i = 1, \ldots, m$. Output $\mathscr{P}^{(0,m)} = \left(\hat{R}_1^{(0,m)}, \hat{R}_2^{(0,m)}, \ldots, \hat{R}_m^{(0,m)}\right)$ and update $j = 1$.
2. While $j \leq k$:
   2.1 Simulate a sequence $\{(Q_i, N_i, C_{(i,1)}, C_{(i,2)}, \ldots)\}_{i=1}^m$ of i.i.d. copies of the generic branching vector, independent of everything else.
   2.2 Let
   $$\hat{R}_i^{(j,m)} = \Phi\left(Q_i, N_i, \{C_{(i,r)}\}, \{\hat{R}_{(i,r)}^{(j-1,m)}\}\right), \qquad i = 1, \ldots, m, \quad (1)$$
   where the $\hat{R}_{(i,r)}^{(j-1,m)}$ are sampled uniformly with replacement from the pool $\mathscr{P}^{(j-1,m)}$.
   2.3 Output $\mathscr{P}^{(j,m)} = \left(\hat{R}_1^{(j,m)}, \hat{R}_2^{(j,m)}, \ldots, \hat{R}_m^{(j,m)}\right)$ and update $j = j + 1$.

Estimators

- For large enough m, we expect that

$$\hat{F}_{k,m}(x) \triangleq \frac{1}{m}\sum_{i=1}^{m} 1(\hat{R}_i^{(k,m)} \leq x) \approx P(R^{(k)} \leq x) \triangleq F_k(x).$$

- Moreover, for functions h "nice enough", we would want to estimate $E[h(R^{(k)})]$ using

$$\frac{1}{m}\sum_{i=1}^{m} h(\hat{R}_i^{(k,m)}).$$

- **Problem:** Although the $\{\hat{R}_i^{(k,m)}\}$ are identically distributed, they are not independent and do not in general have distribution F_k.

Assumptions

- We have two different types of assumptions:
- **Assumption p:** Suppose that for some $p \geq 1$ there exists a constant $0 < H_p < \infty$ such that

$$E\left[\left|\Phi\left(Q, N, \{C_i\}, \{X_i\}\right) - \Phi\left(Q, N, \{C_i\}, \{Y_i\}\right)\right|^p\right] \leq H_p E\left[|X_1 - Y_1|^p\right],$$

  for $\{(X_i, Y_i)\}$ i.i.d., independent of $(Q, N, \{C_i\})$, and such that $\{X_i\} \sim F$ and $\{Y_i\} \sim G$.
- **Assumption φ:** There exists a function $\varphi : \mathbb{R} \to \mathbb{R}_+$ such that

$$|\Phi(q, n, \{c_r\}, \{x_r\}) - \Phi(q, n, \{c_r\}, \{y_r\})| \leq \sum_{r=1}^{n} \varphi(c_r)|x_r - y_r|$$

  for any vectors $(q, n, \{c_r\}, \{x_r\})$ and $(q, n, \{c_r\}, \{y_r\})$ for which Φ is well-defined.

The Wasserstein metric

- **Definition:** Let F and G be distribution functions on $\mathbb{R}$ having finite pth moment, $p \geq 1$. The Wasserstein distance of order p between F and G is

$$d_p(F,G) = \inf_{X \sim F, Y \sim G} \left(E\left[|X-Y|^p\right]\right)^{1/p}$$
$$= \left(\int_0^1 |F^{-1}(u) - G^{-1}(u)|^p du\right)^{1/p},$$

  where $f^{-1}(t) = \inf\{x \in \mathbb{R} : f(x) \geq t\}$.

- Convergence in the Wasserstein metric of order p is equivalent to weak convergence plus convergence of the absolute moments of order up to p.
- Recall that

$$F_k(x) = P(R^{(k)} \leq x) \quad \text{and} \quad \hat{F}_{k,m}(x) = \frac{1}{m}\sum_{i=1}^{m} 1(\hat{R}_i^{(k,m)} \leq x).$$

Convergence in mean

- **Theorem A:** Fix $1 \leq p < \infty$ and suppose that Φ satisfies *Assumption p* for p. Assume further that for any fixed $k \in \mathbb{N}$, $\max_{0 \leq j \leq k} E[(R^{(j)})^p] < \infty$. Then,

$$E\left[d_p(\hat{F}_{k,m}, F_k)^p\right] \leq \left(\sum_{r=0}^{k} (H_p^{1/p})^r\right)^{p-1} \sum_{r=0}^{k} (H_p^{1/p})^{k-r} E\left[d_p(F_{r,m}, F_r)^p\right],$$

where $F_{r,m}$ is the empirical distribution function of an i.i.d. sample of size m from F_r. Moreover, if $\max_{0 \leq j \leq k} E\left[|R^{(j)}|^q\right] < \infty$ for $q > p \geq 1$, $q \neq 2p$, then

$$E\left[d_p(\hat{F}_{k,m}, F_k)^p\right] \leq K m^{-\alpha} \left(\sum_{r=0}^{k} (H_p^{1/p})^r\right)^{p-1} \sum_{r=0}^{k} (H_p^{1/p})^{k-r} (E[|R^{(j)}|^q])^{p/q},$$

where $K = K(p, q)$ and $\alpha = \min\{(q-p)/q, 1/2\}$.

Some remarks

- The theorem does not require $H_p < 1$.
- When $H_p < 1$, the bound provided by the theorem becomes independent of the depth of the recursion k.
- In some cases, failure to find $H_p < 1$ for some $p \geq 1$ implies the instability of the recursion.
- The rate at which $E[d_p(\hat{F}_{k,m}, F_k)^p]$ converges to zero is that of $\max_{0 \leq j \leq k} E[d_p(F_{j,m}, F_j)^p]$ (up to a constant), which corresponds to the "perfect" i.i.d. pool.

Some remarks

- The theorem does not require $H_p < 1$.
- When $H_p < 1$, the bound provided by the theorem becomes independent of the depth of the recursion k.
- In some cases, failure to find $H_p < 1$ for some $p \geq 1$ implies the instability of the recursion.
- The rate at which $E[d_p(\hat{F}_{k,m}, F_k)^p]$ converges to zero is that of $\max_{0\leq j\leq k} E[d_p(F_{j,m}, F_j)^p]$ (up to a constant), which corresponds to the "perfect" i.i.d. pool.

  There is no loss in the rate of convergence due to the bootsrapping!

Almost sure convergence

- **Theorem B:** Fix $1 \leq p < \infty$ and suppose that Φ satisfies *Assumption p* for both p and $2p$. Assume further that for any fixed $k \in \mathbb{N}$, $\max_{0 \leq j \leq k} E[(R^{(j)})^{2p}(\log |R^{(j)}|)^+] < \infty$. Then,

$$\lim_{m\to\infty} d_p(\hat{F}_{k,m}, F_k) = 0 \qquad \text{a.s.}$$

- **Theorem C:** Fix $1 \leq p < \infty$ and suppose that Φ satisfies *Assumption φ*. Assume further that $[|R^{(0)}|^{p+\delta} + Z^{p+\delta}] < \infty$ for some $\delta > 0$, where $Z = \sum_{i=1}^{N} \varphi(C_i)$. Then, for any fixed $k \in \mathbb{N}$,

$$\lim_{m\to\infty} d_p(\hat{F}_{k,m}, F_k) = 0 \qquad \text{a.s.}$$

Consistency of estimators

- **Definition:** We say that Θ_n is a *weakly consistent* estimator for θ if, as $n \to \infty$, $\Theta_n \xrightarrow{P} \theta$, and we say that it is *strongly consistent* if $\Theta_n \to \theta$ a.s.

- **Proposition:** Fix $1 \leq p < \infty$ and suppose that $h : \mathbb{R} \to \mathbb{R}$ satisfies $|h(x)| \leq C(1 + |x|^p)$ for all $x \in \mathbb{R}$ and some constant $C > 0$. Then,
  - i.) If $E[d_p(\hat{F}_{k,m}, F_k)^p] \to 0$ as $m \to \infty$, then

$$\frac{1}{m} \sum_{i=1}^{m} h(\hat{R}_i^{(k,m)}) = \int_{-\infty}^{\infty} h(x) d\hat{F}_{k,m}(x) \tag{2}$$

    is a weakly consistent estimator for $E[h(R^{(k)})]$ for each fixed $k \in \mathbb{N}$.

  - ii.) If $d_p(\hat{F}_{k,m}, F_k) \to 0$ a.s. as $m \to \infty$, then (2) is a strongly consistent estimator for $E[h(R^{(k)})]$ for each fixed $k \in \mathbb{N}$.

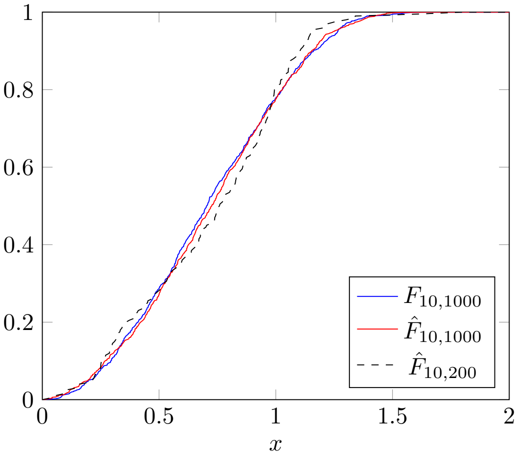
Open problem

- Derive a CLT for the algorithm.
- *Hint:* Suppose that Assumption p holds for some $1 \leq p < \infty$, and that $\max_{0 \leq r \leq k} E[|R^{(r)}|^q] < \infty$ for some $q > 2p$.
- Under these conditions **Theorem B** holds and it is should be easy to verify that

$$\frac{1}{m} \sum_{i=1}^{m} h(\hat{R}_i^{(k,m)}) - E\left[h(\hat{R}_1^{(k,m)}) \middle| \mathscr{P}^{(k-1,m)} \right],$$

properly scaled, satisfies a CLT.

A numerical example (Linear SFPE)



$F_{10,1000}$ is the empirical CDF of $R^{(10)}$ with a sample of 1000 (computation time 883 sec.); $\hat{F}_{10,1000}$ required only 2.1 sec.