A new numerical method for computing the quasi-stationary distribution of subcritical Galton-Watson branching processes

Stefano Massei

Joint work with Sophie Hautphenne (University of Melbourne)

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

stefano.massei@epfl.ch

Hobart, 13 February 2019

Branching processes

Branching processes are stochastic processes describing the dynamics of a population of individuals which reproduce and die independently, according to some specific probability distributions.

Branching processes have numerous applications in population biology and phylogenetics

Galton-Watson branching processes

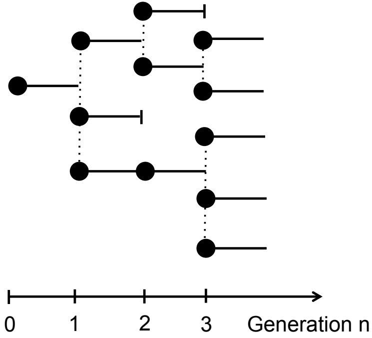
- Time is discrete and represents successive generations
- Each individual has a unit lifetime, at the end of which it might give birth to one or more offsprings simultaneously
- The offspring distribution is described by a random variable θ taking non-negative integer values with corresponding probabilities

$$p_j = \mathbb{P}[\theta = j], \quad j \geq 0.$$

- All individuals behave independently of each other
- The function $P(z) := \sum_j p_j z^j$ is called probability generating function (p.g.f) of the offspring distribution
- If the expected value of the offspring distribution $m = P'(1) = \sum_j j p_j < 1$ then we have extinction with probability 1

Galton-Watson branching processes

A realisation of a GW process through 3 generations starting with a single individual at generation 0:



Quasi stationary distribution

In some models it is of interest to consider populations that are certain to become extinct ($m < 1$), yet appear to be stationary over any reasonable time scale.
In that case, one can consider the so called quasi stationary distribution of the process, i.e., the asymptotic distribution of the population size, conditional on its survival.

Theorem (Yaglom)

For each $j = 1, 2, \ldots,$

$$\lim_{n\to\infty} \mathbb{P}[Z_n = j \mid Z_n \neq 0] = g_j$$

exists, and $\sum_j g_j = 1$.

Moreover, the p.g.f. $G(z) = \sum_j g_j z^j$ *satisfies the equation*

$$G(P(z)) = m\, G(z) + 1 - m.$$

The problem we address in this talk

Given $P(z) = \sum_{j\geq 0} p_j z^j$ such that

$$p_j \geq 0, \quad P(1) = \sum_{j\geq 0} p_j = 1, \quad P^{(1)}(1) = \sum_{j\geq 1} j p_j = m \in (0,1),$$

we want to find $G(z) = \sum_{j\geq 0} g_j z^j$ that solves

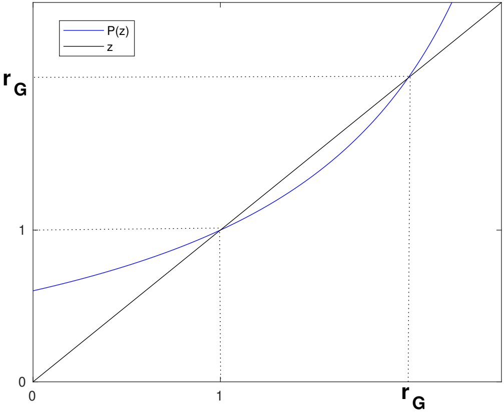
$$\begin{cases} G(P(z)) = mG(z) + 1 - m & z \in [0,1] \\ G(0) = 0 \\ G(1) = 1 \\ g_j \geq 0 \end{cases}.$$

The existence of $G(z)$ is guaranteed by Yaglom's theorem.
Uniqueness is ensured if $G(z)$ is analytic at $z = 1$.

Analyticity of $G(z)$

Theorem (Königs 1884)

If $P(z)$ is analytic on $\mathcal{B}(0, r_P)$ with $r_P > 1$ then $G(z)$ is analytic on $B(0, r_G)$ where r_G is either the solution of $z = P(z)$ on $(1, \infty)$ or ∞.



Computing $G(z)$ in the analytic case

$$\boxed{G(P(z)) = mG(z) + 1 - m} \tag{1}$$

Without imposing boundary conditions, (1) admits infinite solutions of the form

$$G(z) = 1 + s \cdot g(z), \qquad s \in \mathbb{C},$$

where $g(z)$ verifies $g(P(z)) = m \cdot g(z)$. In particular, we have to find an eigenvector of $g \to g \circ P$ associated with the eigenvalue m.

Idea: Rephrase the composition with $P(z)$ as an integral operator:

$$g(P(z)) = \int_{\partial B(0,r)} \frac{g(t)}{t - P(z)} dt = m \cdot g(z).$$

Computing $G(z)$ in the analytic case

$$\boxed{\int_{\partial B(0,r)} \frac{g(t)}{t - P(z)} dt = m \cdot g(z)}$$

Approximating the integral with the trapezoidal rule and evaluating the expression in the scaled roots of the unit yields

$$\sum_{h=1}^{n} g(r\xi_h) \cdot \frac{r\xi_h}{n} \cdot (r\xi_h - P(r\xi_j))^{-1} - m \cdot g(r\xi_j) \approx 0 \quad j = 1, \dots, n, \tag{2}$$

where $\xi_h = e^{\frac{2\pi h}{n}}$.

Expression (2) says that the vector $\mathbf{v_g} := [g(r\xi_1), \dots, g(r\xi_n)]^T$ verifies

$$A\mathbf{v_g} \approx 0, \qquad A = (a_{jh})_{j,h=1,\dots,n}, \qquad a_{jh} = \begin{cases} \frac{r\xi_h}{n(r\xi_h - P(r\xi_j))} & h \neq j \\ \frac{r\xi_h}{n(r\xi_h - P(r\xi_h))} - m & h = j \end{cases}.$$

Computing $G(z)$ in the analytic case

$$A\mathbf{v_g} \approx 0$$

Strategy:

- Approximate $\mathbf{v_g} = [g(r\xi_1), \ldots, g(r\xi_n)]^T$ with an eigenvector of A associated with its smallest eigenvalue.
- Apply the inverse fast Fourier transform to $\mathbf{v_g}$ getting the coefficients of a degree $n-1$ polynomial that interpolates $\widetilde{g}(z) := g(r \cdot z)$ on the roots of the unit.
- Use $G(0) = 0$ to compute $s \in \mathbb{C}$ such that $\widetilde{G}(z) := G(r \cdot z) = 1 + s\widetilde{g}(z)$.
- Retrieve the (approximate) coefficients of $G(z)$ by rescaling those of $\widetilde{G}(z)$.

Evaluation-Interpolation strategy

Algorithm 1 Compute_G$(P(z), n, r)$

$m \leftarrow P'(1)$

$\xi \leftarrow \left(r \cdot e^{\frac{2\pi \mathbf{i} j}{n}}\right)_{j=1,\dots,n}$

$A \leftarrow \left(\frac{\xi_h}{n(\xi_h - P(\xi_j))}\right)_{j,h=1,\dots,n}$

$A \leftarrow A - m \cdot I_n$

$\mathbf{v}_g \leftarrow \texttt{eigs}(A)$

$\mathbf{w} \leftarrow \texttt{ifft}(\mathbf{v}_g)$

$\mathbf{w} \leftarrow -\frac{1}{w_1}\mathbf{w}, \quad w_1 \leftarrow 0$ now $\sum_{j=1}^{n} w_j z^{j-1}$ interpolates $\widetilde{G}(z) = G(r \cdot z)$

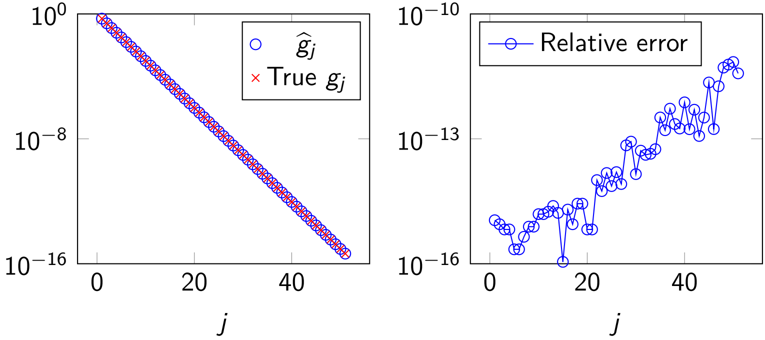
return $\left(\frac{w_j}{r^{j-1}}\right)_{j=1,\dots,n}$

Benchmark example

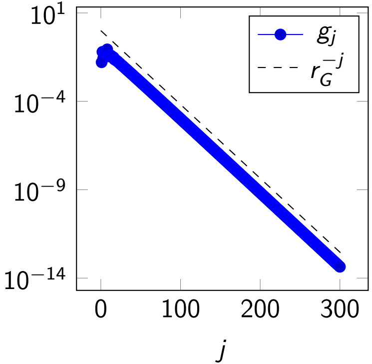
A linear fractional Galton-Watson process is a process in which the offspring distribution is a modified geometric distribution:

$$P(z) = \sum_{j=0}^{\infty} p_j z^j, \qquad p_j = (1-p_0)(1-p)p^{j-1},\ j \geq 1$$

for some parameters $p_0, p \in (0,1)$. In this case, the QSD is geometric with parameter p/p_0. Here we took $p = 0.3$ and $p_0 = 0.6$.



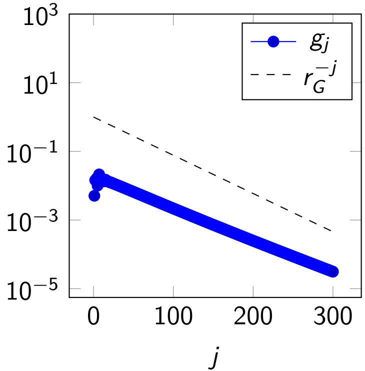
A numerical example



n	Time (s)	Res	$\sum g_j$
256	0.05	$8.67 \cdot 10^{-2}$	0.72
512	0.11	$1.40 \cdot 10^{-2}$	0.96
1,024	0.44	$4.96 \cdot 10^{-4}$	1
2,048	1.62	$8.85 \cdot 10^{-7}$	1
4,096	9.92	$3.80 \cdot 10^{-12}$	1
8,192	78.21	$2.04 \cdot 10^{-15}$	1

$P(z)$ is a degree 8 polynomial, such that $m = 0.776$ and $r_G = 1.101$.
$Res := \max_{j=1,\dots,n} |G(P(\xi_j)) - mG(\xi_j) - 1 + m|$.

Example m "close" to 1



n	Time (s)	Res	$\sum g_j$
256	0.06	$9.94 \cdot 10^{-2}$	0.73
512	0.11	$7.53 \cdot 10^{-2}$	0.74
1,024	0.39	$4.33 \cdot 10^{-2}$	0.97
2,048	1.55	$2.14 \cdot 10^{-2}$	0.67
4,096	9.94	$2.49 \cdot 10^{-2}$	0.65
8,192	79.88	$1.33 \cdot 10^{-2}$	0.86

$P(z)$ is a degree 8 polynomial, such that $m = 0.942$ and $r_G = 1.026$.
$Res := \max_{j=1,\dots,n} |G(P(\xi_j)) - mG(\xi_j) - 1 + m|$.

Structure of the matrix A

The matrix A defined by

$$A = (a_{jh})_{j,h=1,\dots,n}, \qquad a_{jh} = \begin{cases} \frac{r\xi_h}{n(r\xi_h - P(r\xi_j))} & h \neq j \\ \frac{r\xi_h}{n(r\xi_h - P(r\xi_h))} - m & h = j \end{cases}$$

is highly structured, indeed it can be written as

$$A = C_{P,r}^{(n)} \cdot \operatorname{diag}\left(\frac{r\xi_1}{n}, \dots \frac{r\xi_n}{n}\right) - mI_n$$

where $C_{P,r}^{(n)} = (c_{hj}) := \frac{1}{r\xi_h - P(r\xi_j)}$ is a Cauchy matrix.

In particular, Cauchy matrices often exhibit a low numerical rank.
Hence, we would like to approximate A as

$$A \approx UV^* - mI_n$$

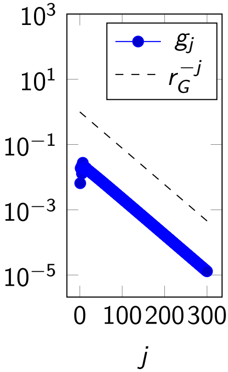
with $U, V \in \mathbb{C}^{n\times k}$ tall and skinny matrices ($k \ll n$).

Exploiting low-rank

We can then modify the evaluation-interpolation procedure as follows:

- Avoid forming explicitly A.
- Compute a low-rank approximation of $C_{P,r}^{(n)} \approx U\widetilde{V}^*$ by means of Adaptive Cross Approximation (ACA).
- Retrieve $C_{P,r}^{(n)} \cdot \operatorname{diag}\left(\frac{r\xi_1}{n}, \dots \frac{r\xi_n}{n}\right) \approx UV^*$ just by rescaling $\widetilde{V}$.
- Solve the eigenvector problem with a $\mathcal{O}(n)$ cost.

The resulting algorithm has $\mathcal{O}(n\log(n))$ cost in time and $\mathcal{O}(n)$ cost in storage.



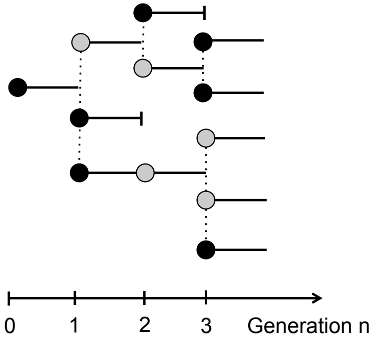
n	Time (s)	Res	$\sum g_j$	$\mathrm{rk}(C_{P,r}^{(n)})$
16,384	2.28	$4.10 \cdot 10^{-4}$	1	264
32,768	6.88	$5.80 \cdot 10^{-7}$	1	366
65,536	21.67	$1.15 \cdot 10^{-10}$	1	465
$1.31 \cdot 10^5$	49.89	$4.33 \cdot 10^{-10}$	1	471
$2.62 \cdot 10^5$	113.19	$5.94 \cdot 10^{-10}$	1	475

$P(z)$ is a degree 8 polynomial, such that $m = 0.942$ and $r_G = 1.026$.
$Res := \max_{j=1,\dots,n} |G(P(\xi_j)) - mG(\xi_j) - 1 + m|$.

Multitype Galton-Watson branching process

Suppose now there are $d > 1$ types of individuals, each type having its own reproduction law.

Example with $d = 2$:



Bivariate version of the problem

Given $P_j(x,y) = \sum_{h,k\geq 0} p_{hk}^{(j)} x^h y^k$ such that

$$p_{hk}^{(j)} \geq 0, \quad P_j(1,1) = 1, \quad j = 1,2$$

and the spectral radius m of the matrix $\begin{bmatrix} \frac{\partial P_1(1,1)}{\partial x} & \frac{\partial P_1(1,1)}{\partial y} \\ \frac{\partial P_2(1,1)}{\partial x} & \frac{\partial P_2(1,1)}{\partial y} \end{bmatrix}$ is less than 1, we want to find $G(z) = \sum_{h,k\geq 0} g_{hk} x^h y^k$ that solves

$$\begin{cases} G(P_1(x,y), P_2(x,y)) = mG(x,y) + 1 - m & x,y \in [0,1] \\ G(0,0) = 0, \quad G(1,1) = 1 \\ g_{h,k} \geq 0 \end{cases}.$$

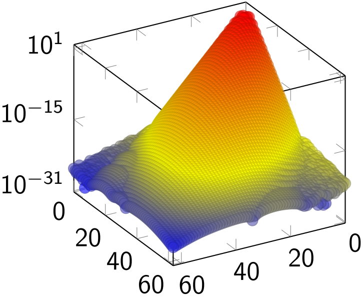
Idea: Once again, exploit:

$$G(P_1(x,y), P_2(x,y)) = \frac{1}{(2\pi\mathbf{i})^2} \int_{\partial\mathcal{B}(0,r_1)\times\partial\mathcal{B}(0,r_2)} \frac{G_j(\tilde{x},\tilde{y})}{(\tilde{x} - P_1(x,y))(\tilde{y} - P_2(x,y))} \, d\tilde{x} \, d\tilde{y}$$

A 2D example

We take $P_1(x, y)$ and $P_2(x, y)$ bivariate polynomials of degree $(2, 2)$.

n	Time (s)	Res	$\sum \widehat{g}_{h,k}$
16	0.12	0.27	1.17
32	0.18	$6.59 \cdot 10^{-2}$	0.9
64	0.86	$2.26 \cdot 10^{-3}$	1
128	5.66	$1.45 \cdot 10^{-5}$	1
256	33.45	$9.53 \cdot 10^{-10}$	1
512	218.49	$5.95 \cdot 10^{-12}$	1



	$p_{0,0}^{(j)}$	$p_{0,1}^{(j)}$	$p_{0,2}^{(j)}$	$p_{1,0}^{(j)}$	$p_{1,1}^{(j)}$	$p_{1,2}^{(j)}$	$p_{2,0}^{(j)}$	$p_{2,1}^{(j)}$	$p_{2,2}^{(j)}$
P_1	0.798	0.029	0.009	0.015	0.010	0.022	0.052	0.020	0.045
P_2	0.694	0.041	0.057	0.035	0.027	0.043	0.024	0.051	0.028

$m = 0.5884$, $r_1 = 1.2462$, $r_2 = 1.4101$.

Conclusion

- Discretizing the functional equation allows the design of fast and accurate solvers for the quasi stationary distribution.

Some open questions:

- Prove the stability of the discretization scheme
- In the bivariate case the matrix that arises from the discretization is the unfolding of a 4-th order tensor. Can we gain in representing it via tensor formats?

Full story:

S. Hautphenne, S. M., A low-rank technique for computing the quasi stationary distribution of subcritical Galton-Watson processes, Arxiv, 2019.