Processus de branchement de Galton-Watson

Pierron Théo

5 janvier 2014

Le but de ce développement est de décrire la propagation d'un nom de famille au cours des générations. Soit p_k la probabilité qu'un homme ait k fils. On suppose que $p_0 \in]0,1[$.

Soit $\mathcal{L}$ la loi $\sum_{k=0}^{\infty} p_k \delta_k$. On note également Z_k le nombre d'hommes à la génération k avec par convention $Z_0 = 1$. On va s'intéresser à la suite $x_k := \mathbb{P}(Z_k = 0)$ des probabilités que la lignée s'éteigne à partir de la génération k.

Soit $G(s) = \sum_{k=0}^{\infty} p_k s^k$ la série génératrice de la loi $\mathcal{L}$.

Lemme 1
Si $T_1, \dots, T_n$ sont iid de série génératrice G, la série génératrice de $S_n := \sum_{i=1}^{n} T_i$ est G^n.

Lemme 2
Pour tout $n \in \mathbb{N}$, on note G_n la série génératrice de Z_n. Alors $G_n(s) = G^{\circ n}(s)$.

Démonstration. Soit n fixé. Pour tout $i \in [\![1, Z_n]\!]$, on note T_i le nombre de fils du i-ème homme de la n-ème génération. Alors T_i suit la loi $\mathcal{L}$ et pour tout $i \neq j$, $T_i \perp\!\!\!\perp T_j$. On a de plus

$$Z_{n+1} = \sum_{i=1}^{Z_n} T_i$$

Le coefficient d'ordre k de G_{n+1} est

$$\begin{aligned}\mathbb{P}(Z_{n+1} = k) &= \sum_{j=0}^{\infty} \mathbb{P}(Z_{n+1} = k, Z_n = j) = \sum_{j=0}^{\infty} \mathbb{P}\left(S_j = k, Z_n = j\right) \\ &= \sum_{j=0}^{\infty} \mathbb{P}\left(S_j = k\right) \mathbb{P}(Z_n = j)\end{aligned}$$

La série G_{n+1} vérifie

$$G_{n+1}(s) = \sum_{k=0}^{\infty} \mathbb{P}(Z_{n+1} = k) s^k = \sum_{k=0}^{\infty}\sum_{j=0}^{\infty} \mathbb{P}\left(S_j = k\right) \mathbb{P}(Z_n = j) s^k$$

Par Fubini, comme tout le monde est positif, on a

$$G_{n+1}(s) = \sum_{j=0}^{\infty} \mathbb{P}(Z_n = j) \underbrace{\sum_{k=0}^{\infty} \mathbb{P}\left(S_j = k\right) s^k}_{\text{série génératrice de } S_j} = \sum_{j=0}^{\infty} \mathbb{P}(Z_n = j) G(s)^j = G_n(G(s))$$

Donc par récurrence, $G_n(s) = G^{\circ n}(s)$. ∎

Lemme 3
G est strictement croissante et convexe sur $[0,1]$. Elle est strictement convexe ssi $p_0 + p_1 \neq 1$.

Démonstration. G est analytique donc C^2 sur $[0,1]$. Par dérivation des séries entières, on a

$$G'(s) = \sum_{k=1}^{\infty} kp_k s^{k-1} \text{ et } G''(s) = \sum_{k=2}^{\infty} k(k-1)p_k s^{k-2}$$

Comme $p_0 < 1$, il existe $k > 0$ tel que $p_k > 0$ donc pour tout $s \geqslant 0$, comme $kp_k s^k > 0$, $G'(s) > 0$ donc G est strictement croissante.

Si $p_0 + p_1 = 1$, $G(s) = p_0 + p_1 s$ est affine donc convexe mais pas strictement. Sinon, il existe $k > 1$ tel que $p_k > 0$ et $G''(s) > 0$ par le même argument donc G est strictement convexe. ■

Lemme 4
Si $G'(1) > 1$, G a un unique point fixe dans $[0,1[$. Sinon, le seul point fixe de G est 1.

Démonstration. Si G est affine, 1 est le seul point fixe de G et $G'(1) = p_1 \leqslant 1$.

Si $G'(1) \leqslant 1$, alors par stricte croissance, pour tout $s < 1$, $G'(s) < G'(1) \leqslant 1$ donc en intégrant

$$1 - G(s) = \int_s^1 G'(t)\,\mathrm{d}t < \int_s^1 \mathrm{d}t = 1 - s$$

Donc $G(s) > s$ et G n'a pas d'autre point fixe que 1.

Si $G'(1) > 1$, il existe $\varepsilon > 0$ tel que $G'(1) > 1 + \varepsilon$. Alors par la formule de Taylor, il existe η qui tend vers 0 en 0 tel que

$$G(1-h) = 1 - hG'(1) + h^2\eta(h) < 1 - h + h(h\eta(h) - \varepsilon)$$

pour h assez proche de 0, $h\eta(h) - \varepsilon < 0$ donc $1 - h + h(h\eta(h) - \varepsilon) < 1 - h$. Donc $G(1-h) < 1 - h$. Ainsi, $G - \mathrm{Id}$ est négatif en $1 - h$ et positif en 0 donc par le théorème des valeurs intermédiaires, G a un point fixe p dans $[0, 1-h] \subset [0,1[$.

De plus, ce point fixe est unique : soit $q \neq p$ un autre point fixe de G. Alors G est strictement convexe donc $G|_{]\min(p,q),1[} < \mathrm{Id}$ donc il n'y a pas d'autre point fixe dans $]\min(p,q),1[$, ce qui contredit le fait que $\max(p,q)$ soit fixe. Donc p est unique. ■

<u>Théorème</u> *La probabilité que la lignée s'éteigne à la n-ème génération tend vers 1 si $\mathbb{E}(\mathcal{L}) \leqslant 1$ et converge vers le point fixe p de G dans $]0,1[$ sinon.*

Démonstration. Par définition de Z_n, si $Z_n = 0$ alors $Z_{n+1} = 0$ donc (x_n) est une suite croissante positive. Elle est de plus majorée par 1 donc elle converge vers $x \in [0,1]$. On a de plus

$$x_{n+1} = G_{n+1}(0) = G(G_n(0)) = G(x_n)$$

donc comme G est continue, $x = G(x)$ donc x est un point fixe de G. Comme $\mathbb{E}(\mathcal{L}) = G'(1)$, si $\mathbb{E}(\mathcal{L}) \leqslant 1$, G n'a pas de point fixe dans $[0,1[$ donc $x = 1$.

Si y est un point fixe de G, alors $x_1 = p_0 = G(0) < G(y) = y$ donc pour tout n, $x_n < G^n(y) = y$ donc $x \leqslant y$. x est donc le plus petit point fixe de G qui est donc p. ■

Remarque *On peut aussi s'intéresser au nombre moyen de descendants : on a*

$$\mathbb{E}(Z_n) = G_n'(1) = (G_{n-1} \circ G)'(1) = G'(1)G_{n-1}'(G(1)) = G'(1)G_{n-1}'(1) = G'(1)\mathbb{E}(Z_{n-1})$$

Donc par récurrence $\mathbb{E}(Z_n) = (G'(1))^n$.

Alors, si $\mathbb{E}(\mathcal{L}) > 1$, l'espérance du nombre d'hommes à la génération n suit une progression exponentielle.

Si $\mathbb{E}(\mathcal{L}) < 1$, l'espérance tend vers 0, ce qui est bien cohérent avec le résultat précédent : la lignée s'éteint (en espérance).

Enfin, si $\mathbb{E}(\mathcal{L}) = 1$, l'espérance est constante égale à 1 et la lignée est stable mais (en moyenne) limitée à un seul homme.