 Mylene.Maida@math.univ-lille1.fr

Devoir à la maison n°4
Le processus de Galton-Watson
À rendre le jeudi 11 décembre 2014

Le modèle que nous allons décrire et étudier dans ce problème a été introduit par Francis Galton et Henry Watson à la fin du dix-neuvième siècle pour déterminer des probabilités d'extinction de noms de familles illustres en Grande-Bretagne. Ils formulaient le problème à peu près de la manière suivante : *Si un homme a une probabilité p_1 d'avoir un fils, p_2 d'en avoir deux, etc. ; si chacun de ses fils éventuels est dans le même cas et ainsi de suite, quelle est la probabilité pour qu'à terme, cette branche de la famille s'éteigne ?*
Le modèle proposé est un cas particulier de processus de branchement, qui modélisent l'évolution d'une population dans laquelle les individus se reproduisent indépendamment les uns des autres.

Décrivons maintenant plus précisément le modèle dit de Galton-Watson. La population à la génération n est constituée de Z_n individus asexués qui donnent naissance, tous en même temps, aux individus de la génération $n+1$ et meurent instantanément. La loi du nombre d'enfants est la même pour tous les individus, quelle que soit la génération à laquelle ils appartiennent. Tous les individus se comportent de manière indépendante et indépendamment de la taille de la population. On se donne un germe de probabilité $(p_k)_{k\geq 0}$, où p_k est la probabilité pour un individu d'avoir k enfants. **On suppose que $\mathbf{p_0 \neq 1}$.** Autrement dit, si on se donne une collection $(X_{j,k})_{j,k\in\mathbb{N}}$ de variables aléatoires indépendantes et de même loi (dite de reproduction) $\mu := \sum_{\ell\geq 0} p_\ell \delta_\ell$, alors pour tout $n \in \mathbb{N}$, on peut écrire

$$Z_{n+1} = \sum_{i=1}^{Z_n} X_{n,k}$$

($X_{n,k}$ représente le nombre d'enfants du $k$ième individu de la génération n).
Le but de ce problème est d'étudier le comportement de la suite de variables aléatoires $(Z_n)_{n\in\mathbb{N}}$. Pour cela, on va étudier la suite des fonctions génératrices correspondantes. Nous aurons besoin de quelques résultats généraux sur les fonctions génératrices.

Soit X une variable aléatoire à valeurs dans $\mathbb{N}$. On rappelle que sa fonction génératrice est donnée par $G_X : s \mapsto \sum_{k=0}^{\infty} P(X=k)s^k$ et qu'elle est définie au moins sur $[-1,1]$. Elle a les propriétés suivantes, que l'on ne demande pas de redémontrer : si X possède une espérance finie, alors $E(X) = G'_X(1)$; si X possède une variance finie, alors $V(X) = G''_X(1) - (G'_X(1))^2 + G'_X(1)$; si X et Y deux variables aléatoires indépendantes alors leur somme $X+Y$ a pour fonction génératrice $G_{X+Y} = G_X \cdot G_Y$.

1. Calculer $G_X(0)$ et $G_X(1)$.
2. Montrer que si $P(X=0) \neq 1$, G_X est convexe et strictement croissante sur $[0,1]$.
3. Montrer que si $P(X=0) + P(X=1) < 1$, G_X est strictement convexe.

On revient maintenant à l'étude du processus $(Z_n)_{n\in\mathbb{N}}$.
Pour $s \in [-1,1]$, on pose $F(s) = \sum_{k=0}^{\infty} p_k s^k$.

4. Si $Z_0 = 1$, quelle est la fonction génératrice de la variable aléatoire Z_1 ?
5. Pour $\ell \in \mathbb{N}$, justifier que, si $Z_0 = \ell$, Z_1 a pour fonction génératrice F^ℓ.

On définit la suite $(F_n)_{n\in\mathbb{N}}$ de fonctions sur $[-1,1]$ de la façon suivante :

$$\begin{aligned} F_0(s) &= s \\ F_1(s) &= F(s) \\ F_n(s) &= F(F_{n-1}(s)), \text{pour tout } n \geq 2. \end{aligned}$$

Dans toute la suite du problème, on suppose que l'on part d'un seul individu, donc $\mathbf{Z_0 = 1}$.

6. Montrer que pour tout $n \in \mathbb{N}$, Z_n a pour fonction génératrice F_n.
7. On étudie maintenant un cas particulier. Supposons que pour tout $k \geq 0$, $p_k = 2^{-(k+1)}$.
   (a) Vérifier que la famille $(p_k)_{k\geq 0}$ est bien un germe de probabilité
   (b) Calculer la fonction génératrice F de la variable aléatoire Z_1.
   (c) Calculer ensuite F_n pour tout $n \in \mathbb{N}$.
   (d) En déduire $P(Z_n = k)$ pour tout $n, k \in \mathbb{N}$.

On s'intéresse maintenant à la probabilité d'extinction, notée π, de la population, c'est-à-dire la probabilité qu'il existe $n \in \mathbb{N}$ tel que $Z_n = 0$. **On suppose dans la suite que $\mathbf{p_0 > 0}$.**

8. Justifier que $\lim_{n\to\infty} F_n(0)$ existe et est égale à π.
9. Calculer π dans l'exemple ci-dessus où pour tout $k \geq 0$, $p_k = 2^{-(k+1)}$.
10. Justifier que dans le cas général, π est le plus petit point fixe de la fonction F dans l'intervalle $[0, 1]$.

La variable aléatoire Z_1 est positive, elle admet donc une espérance $m \in [0, \infty]$. Si $m < 1$, on dit qu'on est dans le cas sous-critique, si $m = 1$ dans le cas critique et si $m > 1$ dans le cas sur-critique. On va étudier le comportement de la suite $(Z_n)_{n\geq 0}$ dans les trois cas.

11. Tracer soigneusement l'allure de la courbe de la fonction F sur $[0, 1]$ dans les trois cas.
12. Montrer que dans les cas critique et sous-critique (soit $m \leq 1$), la probabilité d'extinction π est égale à 1.
13. Montrer que dans le cas sur-critique, F admet exactement un point fixe dans $[0, 1[$, qui est donc égal à π. *Indication : on pourra considérer la fonction $x \mapsto \frac{1-F(x)}{1-x}$ sur $[0, 1[$.*
14. Dans quel cas se trouve-t-on dans l'exemple où pour tout $k \geq 0$, $p_k = 2^{-(k+1)}$?

On examine maintenant plus précisément ce qu'il se passe dans le cas sur-critique lorsque le processus ne s'éteint pas. **On suppose que $\mathbf{m \in]1, \infty[}$ et on suppose toujours $\mathbf{p_0 > 0}$.**

15. Montrer que pour tout $n \geq 1$, $E(Z_n) = m^n$.
16. On étudie maintenant un processus légèrement différent. Soit $k \in \mathbb{N}^*$ fixé. Le processus $(W_n)_{n\in\mathbb{N}}$ est défini de la même façon que $(Z_n)_{n\in\mathbb{N}}$ mais avec $W_0 = k$. On définit alors une suite de réels $(u_n)_{n\in\mathbb{N}}$ par

$$u_n = P(W_n = k \text{ et } W_i \neq k \text{ pour tout } 0 < i < n).$$

   Le nombre u_n est donc la probabilité que le processus prenne la valeur k au rang n pour la première fois après la génération 0. On définit de même, pour tout $r \geq 2$, la probabilité $u_n^{(r)}$ que le processus prenne la valeur k au rang n pour la $r$ième fois après la génération 0.
   (a) Justifier que $\sum_{n=1}^{\infty} u_n < 1$.
   (b) Soit U la fonction génératrice de la suite $(u_n)_{n\in\mathbb{N}}$, donnée par $U(s) = \sum_{n=1}^{\infty} u_n s^n$, pour tout $s \in [0, 1]$. Montrer que pour $r \geq 1$, la fonction génératrice de la suite $(u_n^{(r)})_{n\in\mathbb{N}^*}$ est U^r.
   (c) Montrer que la probabilité pour que la suite $(W_n)_{n\in\mathbb{N}}$ prenne la valeur k une infinité de fois est nulle.
17. En déduire que la probabilité pour que la suite $(Z_n)_{n\in\mathbb{N}}$ prenne la valeur k une infinité de fois est nulle.
18. Conclure que $P(Z_n \xrightarrow[n\to\infty]{} +\infty) = 1 - \pi$.