Modèle de croissance et de division d'une cellule

Caroline Robet

29 janvier 2016

Table des matières

1 Un modèle markovien d'arbre **2**

1.1 Présentation de la généalogie des cellules 2

1.2 Relations vérifiées par les paramètres des cellules 2

1.3 Comportement de la moyenne 3

2 Estimation statistique du taux de division **9**

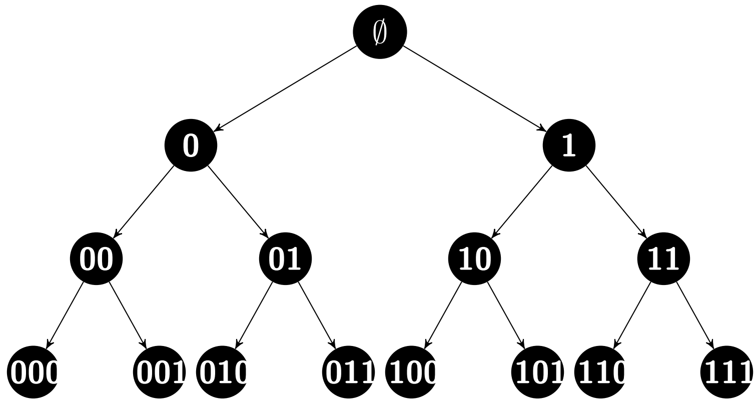
2.1 Deux schémas d'observation 9

2.2 Estimation du taux de division 10

1 Un modèle markovien d'arbre

1.1 Présentation de la généalogie des cellules

On part d'une cellule appelée racine et notée $\emptyset$. Cette cellule croit jusqu'à se diviser en deux cellules de tailles identiques. On définit alors l'arbre généalogique binaire engendré par $\emptyset$ par $\mathcal{U} := \cup_{n=0}^{\infty}\{0,1\}^n$ (où $\{0,1\}^0 = \emptyset$)



Début d'un arbre généalogique d'une cellule

Chaque nœud $u \in \mathcal{U}$ de l'arbre est identifié à une cellule descendante de la cellule $\emptyset$ et est caractérisé par $(\xi_u, b_u, d_u, \tau_u)$ où :
- ξ_u est la taille de la cellule u à sa naissance
- b_u est la date de naissance de la cellule u
- d_u est la durée de vie de la cellule u
- τ_u est la taux de croissance de la cellule u.

On désigne par u^- le parent de u.

1.2 Relations vérifiées par les paramètres des cellules

1. L'évolution $(\xi_t^u, t \in [b_u, b_u + d_u[)$ de la taille de la cellule u pendant sa durée de vie est gouvernée par :
$$\xi_t^u = \xi_u \exp(\tau_u(t - b_u)) \text{ pour } t \in [b_u, b_u + d_u[. \tag{1}$$
On est dans le cas d'une croissance exponentielle de la cellule suivant son taux de croissance.
2. Chaque cellule meurt en donnant naissance à deux cellules de même taille. La date de naissance de u correspond à la date de mort de son parent. On obtient
$$\xi_u = \frac{1}{2}\xi_{u^-} \exp(\tau_{u^-} d_{u^-}) \tag{2}$$
3. La loi de la durée de vie d_u est donnée par
$$\mathbb{P}(d_u \in dt | \xi_u = x, \tau_u = v) = B(x\exp(vt)) \exp\Big(-\int_0^t B(x\exp(vs))ds\Big) dt$$

où $B(x)$ pour $x \in]0,\infty[$ est un taux de division.

4. Finalement, le taux de croissance de u est transmis par le taux τ_{u^-} de son parent à l'aide d'un noyau de transition ρ tel que

$$\mathbb{P}(\tau_u \in dv' | \tau_{u^-} = v) = \rho(v, dv') \tag{3}$$

avec $v > 0$ et où $\rho(v, dv')$ est une probabilité sur $]0,\infty[$ pour tout $v > 0$.

Le modèle dynamique $((\xi_u, \tau_u), u \in \mathcal{U})$ est complètement déterminé par les équations précédentes, comme une chaîne de Markov de conditions initiales $(\xi_\emptyset, \tau_\emptyset)$ sur la racine. Cette chaîne peut être vue comme un processus de Markov continu en posant

$$(\xi_t^u, \tau_t^u) = \begin{cases} (\xi_u \exp(\tau_u(t - b_u)), \tau_u) & \text{si } t \in [b_u, b_u + d_u[\\ (0,0) & \text{sinon} \end{cases} \tag{4}$$

On définit $(X(t), V(t)) = ((X_1(t), V_1(t)), (X_2(t), V_2(t)), \dots)$, le processus des tailles et des taux de croissance des cellules vivantes dans le système à l'instant t. Ces deux quantités sont reliées avec la relation suivante :

$$\sum_{i=1}^{+\infty} 1_{\{X_i(t)>0\}} \delta_{(X_i(t), V_i(t))} = \sum_{u \in \mathcal{U}} 1_{\{b_u \leqslant t < b_u + d_u\}} \delta_{(\xi_t^u, \tau_t^u)} \tag{5}$$

Dans la suite, nous nous servirons des hypothèses suivantes :

Hypothèse 1. *Le taux de division B est continu, $B(0) = 0$ et* $\displaystyle\int_0^\infty \frac{B(x)}{x} = +\infty$. *Le noyau de transition ρ est défini sur un compact $\mathcal{E} \subset]0,\infty[$.*

Théorème 2. *Sous l'hypothèse 1, la loi de $((X(t), V(t)), t \geq 0)$ ou $((\xi_u, \tau_u), u \in \mathcal{U})$ ou $((\xi_t^u, \tau_t^u), t \geq 0, u \in \mathcal{U})$ est bien définie sur un espace de probabilité approprié avec presque sûrement aucune accumulation de sauts.*

Si μ est une mesure de probabilité sur $\mathcal{S} =]0,\infty[\times \mathcal{E}$, on notera de manière indifférente par $\mathbb{P}_\mu$ la loi de ces 3 processus lorsque la racine $(\xi_\emptyset, \tau_\emptyset)$ suit la loi μ.

1.3 Comportement de la moyenne

Dans toute cette section, la loi de la racine est $\mu = \mu_x$ avec $x \in]0,+\infty[$ telle que $\mu_x(\{x\} \times \mathcal{E}) = 1$ (ie $\xi_\emptyset = x$ p.s).

Définition 3. *Pour $u \in \mathcal{U}$, on pose $m^i u$ le i-ème parent de u dans son arbre généalogique. On définit alors $\overline{\tau_t^u}$ comme le taux de croissance accumulé de la cellule u et de ses ancêtres jusqu'au temps t par :*

$$\overline{\tau_t^u} = \begin{cases} \sum_{i=1}^{|u|} \tau_{m_u^i} d_{m_u^i} + \tau_t^u (t - b_u) & \textit{pour } t \in [b_u, b_u + d_u[\\ 0 & \textit{sinon} \end{cases} \tag{6}$$

On va ensuite choisir de manière équiprobable une branche de l'arbre de la manière suivante : pour tout $k \geq 1$, si ν_k désigne la cellule marquée à la k-ème génération, on a

$$\mathbb{P}(\nu_k = u) = \begin{cases} 2^{-k} \text{ pour } u \in \mathcal{U} \text{ tel que } |u| = k \\ \quad 0 \text{ sinon} \end{cases}$$

où $|u| = k$ si $u = (u_0, u_1, \ldots, u_k)$. De plus, pour $t \geq 0$, la relation $b_{\nu_{C_t}} \leq t < b_{\nu_{C_t}} + d_{\nu_{C_t}}$ défini de manière unique un processus de comptage $(C_t)_{t \geq 0}$ avec $C_0 = 0$. Le processus C_t permet de définir un processus marqué de taille à la date t, taux de croissance et taux accumulé de croissance définis par

$$(\chi(t), V(t), \overline{V}(t)) = (\xi_t^{\nu_{C_t}}, \tau_t^{\nu_{C_t}}, \overline{\tau_t^{\nu_{C_t}}}) \text{ pour } t \in [b_{\nu_{C_t}}, b_{\nu_{C_t}} + d_{\nu_{C_t}}[.$$

On a $\chi(t)$ est la taille de la cellule ν_{C_t} pendant sa durée de vie, donc

$$\chi(t) = \xi_{C_t} \exp(\tau_{C_t}(t - b_{C_t})) = \frac{1}{2} \xi_{C_t^-} \exp(\tau_{C_t^-} \zeta_{C_t^-}) \exp(\tau_{C_t}(t - b_{C_t}))$$

d'où par récurrence immédiate

$$\chi(t) = \frac{1}{2^{C_t}} \xi_\emptyset \exp(\overline{\tau_t^{\nu_{C_t}}}) = \frac{x \exp(\overline{V}(t))}{2^{C_t}}. \tag{7}$$

De plus, $V(t) \in [e_{min}, e_{max}]$ car on a supposé $\mathcal{E}$ compact. On remarque alors que

$$\overline{V}(t) \in [e_{min} t, e_{max} t]. \tag{8}$$

Le comportement de $(\chi(t), V(t), \overline{V}(t))$ qui correspond à la taille, au taux de croissance et au taux de croissance cumulé de la branche aléatoire de l'arbre peut-être exprimé comme une fonction de l'arbre généalogique complet grâce à la formule "many-to-one".

Proposition 4. *(Formule many-to-one) Sous l'hypothèse 1, soit $\mathbb{P}_x$ la probabilité sous la loi initiale μ_x de $\emptyset$. Alors $\forall t \geq 0$, on a*

$$\mathbb{E}_x\left[\phi(\chi(t), V(t), \overline{V}(t))\right] = \mathbb{E}_x\Big[\sum_{u\in\mathcal{U}} \xi_t^u \frac{\exp(-\overline{\tau_t^u})}{x}\phi(\xi_t^u, \tau_t^u, \overline{\tau_t^u})\Big]$$

pour tout $\phi : \mathcal{S} \times [0, +\infty[\to [0, +\infty[$

Démonstration. Pour $v \in \mathcal{U}$, on note $I_v = [b_v, b_v + d_v[$. Comme $\chi(t) = \frac{x\exp(\overline{V}(t))}{2^{C_t}}$ par la formule (7), on a

$$\begin{aligned}
\mathbb{E}_x[\phi(\chi(t), V(t), \overline{V}(t))] &= \mathbb{E}_x\left[\phi\left(\frac{x\exp(\overline{V}(t))}{2^{C_t}}, V(t), \overline{V}(t)\right)\right] \\
&= \mathbb{E}_x\left[\sum_{v\in\mathcal{U}}\phi\left(\frac{x\exp(\overline{\tau_t^v})}{2^{|v|}}, \tau_t^v, \overline{\tau_t^v}\right)1_{\{t\in I_v, v_{C_t}=v\})}\right]
\end{aligned}$$

On définit alors $\mathcal{H}_n = \sigma((\xi_u, d_u, \tau_u)\ \textit{avec}\ |u| \leq n)$. On a par conditionnement que

$$\mathbb{P}(v_{C_t} = v|\mathcal{H}_{|v|}) = \frac{1}{2^{|v|}} = \frac{\xi_v\exp(-\overline{\tau_{b_v}^v})}{x}$$

On obtient alors par conditionnement par $\mathcal{H}_{|v|}$,

$$\begin{aligned}
&\mathbb{E}_x\left[\sum_{v\in\mathcal{U}}\phi\left(\frac{x\exp(\overline{\tau_t^v})}{2^{|v|}}, \tau_t^v, \overline{\tau_t^v}\right)1_{\{t\in I_v, v_{C_t}=v\})}\right] \\
&\quad= \mathbb{E}_x\left[\sum_{u\in\mathcal{U}}\phi\left(\frac{x\exp(\overline{\tau_t^v})}{2^{|v|}}, \tau_t^v, \overline{\tau_t^v}\right)1_{\{t\in I_v\}}\frac{\xi_v\exp(-\overline{\tau_{b_v}^v})}{x}\right] \\
&\quad= \mathbb{E}_x\left[\sum_{u\in\mathcal{U}}\xi_t^u\frac{\exp(-\overline{\tau_t^u})}{x}\phi(\xi_t^u, \tau_t^u, \overline{\tau_t^u})\right]
\end{aligned}$$

□

Lemme 5. *On note $\mathcal{F}_t$ la filtration engendrée par $(\chi(s), V(s), s \leqslant t)$ donnés par la cellule marquée. Pour $h > 0$ suffisamment petit, on a*

$$\mathbb{P}_x(C_{t+h} - C_t = 1|\mathcal{F}_t) = B(X(t))h + h\epsilon(h)$$

où $|\epsilon(h)| \to 0$ lorsque $h \to 0$ et

$$\mathbb{P}_x(C_{t+h} - C_t \geq 2) = \mathcal{O}(h^2)$$

Théorème 6. *(Comportement de la mesure de la moyenne empirique) On travaille sous l'hypothèse 1. Soit μ une probabilité sur $\mathcal{S} =]0, +\infty[\times\mathcal{E}$. On définit $n(t, dx, dv)$ par*

$$\langle n(t,.), \varphi\rangle = \mathbb{E}_\mu\left[\sum_{i=1}^{\infty}\varphi(X_i(t), V_i(t))\right] \quad \forall\varphi \in C_0^1(\mathcal{S})$$

Alors $n(t,.)$ est solution (au sens faible) de

$$\begin{cases} \partial_t n(t,x,v) + v\partial_x(xn(t,x,v)) + B(x)n(t,x,v) = 4B(2x)\displaystyle\int_{\varepsilon}\rho(v',v)n(t,2x,dv') \\ n(0,x,v) = n^{(0)}(x,v) \quad x \geq 0 \end{cases}$$

avec pour condition initiale $n^{(0)}(dx,dv) = \mu(dx,dv)$.

Démonstration. Soit $x \in]0,\infty[$, on va prouver le résultat pour $\mu = \mu_x$. Soit $\varphi \in C_0^1(\mathcal{S})$ positive. On a

$$\langle n(t,.), \varphi\rangle = \mathbb{E}_x\left[\sum_{i=1}^{\infty}\varphi(X_i(t), V_i(t))\right] = \mathbb{E}_x\left[\sum_{u\in\mathcal{U}}\varphi(\xi_t^u, \tau_t^u)\right].$$

Or la formule many-to-one, nous dit que si $g : \mathcal{S}\times[0,+\infty[\to[0,+\infty[$ alors $\forall t \geq 0$, on a

$$\mathbb{E}_x[g(\chi(t), V(t), \overline{V}(t))] = \mathbb{E}_x[\sum_{u\in\mathcal{U}}\xi_t^u\frac{\exp(-\overline{\tau_t^u})}{x}g(\xi_t^u, \tau_t^u, \overline{\tau_t^u})].$$

En particulier, avec $g : (\xi_t^u, \tau_t^u, \overline{\tau_t^u}) \to \frac{x}{\xi_t^u}\exp(\overline{\tau_t^u})\varphi(\xi_t^u, \tau_t^u)$ on obtient

$$\langle n(t,.), \varphi\rangle = \mathbb{E}_x\left[x\varphi(\chi(t), V(t))\frac{\exp(\overline{V}(t))}{\chi(t)}\right]. \tag{9}$$

Pour $h > 0$, on introduit l'opérateur

$$\Delta_h f(t) = \frac{f(t+h)-f(t)}{h}.$$

On va étudier la convergence de $\Delta_h\langle n(t,.), \varphi\rangle$ lorsque $h \to 0$ restreint aux événements $\{C_{t+h} - C_t = 0\}$, $\{C_{t+h} - C_t = 1\}$ et $\{C_{t+h} - C_t \geq 2\}$.

On pose $\mathcal{F}_t$ la filtration générée par la cellule marquée $(\chi(s), V(s), s \leq t)$. Comme $\varphi \in \mathcal{C}_0^1(\mathcal{S})$, il existe $c(\varphi) > 0$ tel que $\varphi(y,v) = 0$ si $y \leq c(\varphi)$ donc

$$\left|\varphi(\chi(t), V(t))\frac{\exp(\overline{V}(t))}{\chi(t)}\right| \leq \sup_{y,v}\left(\left|\frac{\varphi(y,v)}{y}\right|\right)\exp(e_{\max}t) \leq \sup_{y,v}(|\varphi(y,v)|)\frac{\exp(e_{\max}t)}{c(\varphi)} \tag{10}$$

- **Sur l'événement** $\{C_{t+h} - C_t \geq 2\}$:
  Par l'expression précédente et par le lemme, on obtient,

$$\begin{aligned}
&\mathbb{E}_x\left[\left|\Delta_h(\varphi(\chi(t),V(t)))\frac{\exp(\overline{V}(t))}{\chi(t)}1_{\{C_{t+h}-C_t\geq 2\}}\right|\right] \\
&\leq \mathbb{E}_x\left[\frac{2}{h}\sup_{y,v}|\varphi(y,v)|\frac{\exp(e_{\max}t)}{c(\varphi)}1_{\{C_{t+h}-C_t\geq 2\}}\right] \\
&\leq \frac{2}{h}\sup_{y,v}|\varphi(y,v)|\frac{\exp(e_{\max}t)}{c(\varphi)}\mathbb{P}_x(C_{t+h}-C_t\geq 2)\lesssim h
\end{aligned}$$

- **Sur l'événement** $\{C_{t+h} - C_t = 0\}$:
  On a marqué la même cellule entre t et t+h donc le processus V est constant entre t et t+h et donc $\frac{\exp(\overline{V}(s))}{\chi(s)}$ est également constant par la relation (7).

$$\left|\Delta_h\left(\varphi(\chi(t),V(t))\frac{e^{\overline{V}(t)}}{\chi(t)}\right)\right| = \left|\Delta_h\left(\varphi(\chi(.),V(t))\right)\frac{e^{\overline{V}(t)}}{\chi(t)}\right|$$

Or

$$\begin{aligned}
|\Delta_h(\varphi(\chi(.),V(t)))| &= \left|\frac{\varphi(\chi(t+h),V(t))-\varphi(\chi(t),V(t))}{h}\right| \\
&= \left|\frac{\varphi(\chi(t)e^{hV(t)},V(t))-\varphi(\chi(t),V(t))}{h}\right| \\
&= \left|\frac{\chi(t)e^{hV(t)}-\chi(t)}{h}\frac{\varphi\left(\chi(t)e^{hV(t)},V(t)\right)-\varphi(\chi(t),V(t))}{\chi(t)e^{hV(t)}-\chi(t)}\right| \\
&\leqslant \left|\frac{\chi(t)e^{hV(t)}-\chi(t)}{h}\right|\sup_{y,v}|\partial_y\varphi(y,v)|
\end{aligned} \tag{11}$$

On en déduit que

$$\begin{aligned}
\left|\Delta_h\left(\varphi(\chi(t),V(t))\frac{e^{\overline{V}(t)}}{\chi(t)}\right)\right| &\leqslant \left|\frac{e^{hV(t)}-1}{h}\right|\sup_{y,v}|\partial_y\varphi(y,v)|e^{\overline{V}(t)} \\
&\leqslant V(t)e^{V(t)h}\sup_{y,v}|\partial_y\varphi(y,v)|e^{\overline{V}(t)} \\
&\leqslant e_{max}e^{e_{max}h}\sup_{y,v}|\partial_y\varphi(y,v)|e^{e_{max}t} \\
&\leq \sup_{y,v}|\partial_y\varphi(y,v)|e_{\max}e^{2e_{max}t}
\end{aligned} \tag{12}$$

Comme $\mathbb{P}_x(C_{t+h}-C_t=0)\to 1$, on obtient par convergence dominée que

$$\mathbb{E}_x\left[x\Delta_h(\varphi(\chi(t),V(t)))\frac{\exp(\overline{V}(t))}{\chi(t)}1_{\{C_{t+h}-C_t=0\}}\right]\to_{h\to 0}\mathbb{E}_x\left[x\partial_1\varphi(\chi(t),V(t))V(t)\exp(\overline{V}(t))\right]$$

Par la formule many-to-one, cette quantité vaut $\langle n(t,dx,dv), xv\partial_x\varphi\rangle$.

– **Sur l'événement** $\{C_{t+h} - C_t = 1\}$:
On a $\chi(t+h) = \frac{\chi(t)}{2} + \varepsilon_1(h)$ car la taille à la naissance de la cellule marquée en t+h est la moitié de la taille finale de la cellule marquée au temps t. On obtient donc

$$\varphi(\chi(t+h), V(t+h)) = \varphi\left(\frac{\chi(t)}{2}, V(t+h)\right) + \varepsilon_2(h).$$

Enfin, $\exp(\overline{V}(t+h)) = \exp(\overline{V}(t)) + \varepsilon_3(h)$ car $\overline{V}$ est continue. De plus, $\forall i$, ε_i est aléatoire mais $|\varepsilon_i(h)| \le \epsilon_1(h) \to 0$ lorsque $h \to 0$ où $\epsilon_1(h)$ est déterministe grâce à (7) et (8). Or,

$$V(t+h) = \tau_{v_{C_t+1}} \quad \text{sur l'événement } \{C_{t+h} - C_t = 1\}.$$

Il s'en suit que

$$\begin{aligned}
&\mathbb{E}_x\left[\varphi(\chi(t+h), V(t+h))\frac{\exp(\overline{V}(t+h))}{\chi(t+h)}1_{\{C_{t+h}-C_t=1\}}\right]\\
&= \mathbb{E}_x\left[\varphi(\chi(t)/2, \tau_{v_{C_t+1}})\frac{2\exp(\overline{V}(t))}{\chi(t)}1_{\{C_{t+h}-C_t=1\}}\right] + \epsilon_2(h)\\
&= \mathbb{E}_x\left[\varphi(\chi(t)/2, \tau_{v_{C_t+1}})\frac{2\exp(\overline{V}(t))}{\chi(t)}1_{\{C_{t+h}-C_t\ge 1\}}\right] + \epsilon_3(h)
\end{aligned}$$

où $\epsilon_2(h), \epsilon_3(h) \to 0$ lorsque $h \to 0$.
On conditionne relativement à $\mathcal{F}_t \vee \tau_{v_{C_t}+1}$ et on utilise le fait que $\{C_{t+h} - C_t \ge 1\}$ et $\tau_{v_{C_t}+1}$ sont indépendants pour pouvoir appliquer la première partie du lemme. On en déduit

$$\begin{aligned}
&\mathbb{E}_x\left[\varphi(\chi(t+h), V(t+h))\frac{\exp(\overline{V}(t+h))}{\chi(t+h)}1_{\{C_{t+h}-C_t=1\}}\right]\\
&= \mathbb{E}_x\left[\varphi(\chi(t)/2, \tau_{v_{C_t+1}})\frac{2\exp(\overline{V}(t))}{\chi(t)}B(\chi(t))h\right] + \epsilon_4(h)\\
&= \mathbb{E}_x\left[\int_\varepsilon \varphi(\chi(t)/2, v')\rho(V(t), dv')\frac{2\exp(\overline{V}(t))}{\chi(t)}B(\chi(t))h\right] + \epsilon_4(h)
\end{aligned}$$

avec $\epsilon_3(h) \to 0$ lorsque $h \to 0$.
Finalement, en utilisant de nouveau le lemme en conditionnant on a :

$$\begin{aligned}
&\mathbb{E}_x\left[x\Delta_h(\varphi(\chi(t), V(t)))\frac{e^{\overline{V}(t)}}{\chi(t)}1_{\{C_{t+h}-C_t=1\}}\right]\\
&\qquad \to_{h\to 0} \mathbb{E}_x\left[x\Big(\int_\varepsilon 2\varphi(\chi(t)/2, v')\rho(V(t), dv') - \varphi(\chi(t), V(t))\Big)\frac{e^{\overline{V}(t)}}{\chi(t)}B(\chi(t))\right]
\end{aligned} \tag{13}$$

Par la formule many-to-one, cette quantité vaut

$$\langle n(t,dx,dv), (\int_\varepsilon 2\varphi(x/2,v')\rho(v,dv') - \varphi(x,v))B(x)\rangle \tag{14}$$

Et par un simple changement de variable, la quantité précédente vaut

$$\langle n(t,2dx,dv), \int_\varepsilon 4\varphi(x,v')\rho(v,dv')B(2x)\rangle - \langle n(t,dx,dv), \varphi(x,v)B(x)\rangle \tag{15}$$

Pour conclure, comme

$$\begin{aligned}
\partial_t\langle n(t,dx,dv),\varphi\rangle &= \lim_{h\to 0}\Delta_h\langle n(t,dx,dv),\varphi\rangle = \lim_{h\to 0}\Delta_h\mathbb{E}_x\left[x\varphi(\chi(t),V(t))\frac{e^{\overline{V}(t)}}{\chi(t)}\right] \\
&= \lim_{h\to 0}\mathbb{E}_x\left[x\Delta_h(\varphi(\chi(t),V(t)))\frac{e^{\overline{V}(t)}}{\chi(t)}1_{\{C_{t+h}-C_t=0\}}\right] + \lim_{h\to 0}\mathbb{E}_x\left[x\Delta_h(\varphi(\chi(t),V(t)))\frac{e^{\overline{V}(t)}}{\chi(t)}1_{\{C_{t+h}-C_t=1\}}\right] \\
&\quad + \lim_{h\to 0}\mathbb{E}_x\left[x\Delta_h(\varphi(\chi(t),V(t)))\frac{e^{\overline{V}(t)}}{\chi(t)}1_{\{C_{t+h}-C_t\geq 2\}}\right] \\
&= \langle n(t,dx,dv), xv\partial_x\varphi\rangle + \langle n(t,2dx,dv), \int_\varepsilon 4\varphi(x,v')\rho(v,dv')B(2x)\rangle \\
&\quad - \langle n(t,dx,dv),\varphi(x,v)B(x)\rangle + 0.
\end{aligned} \tag{16}$$

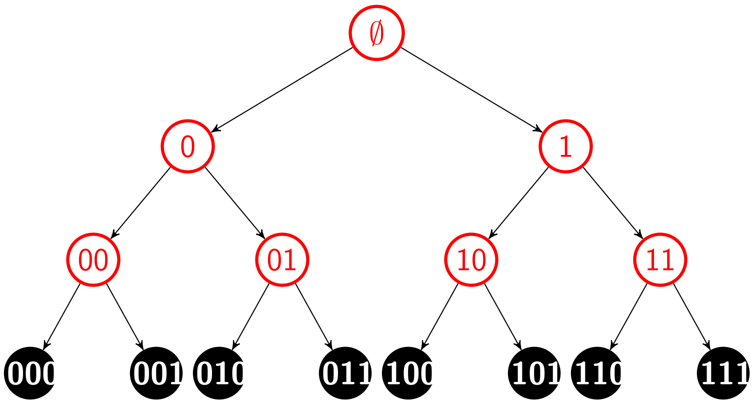
D'où le résultat □

2 Estimation statistique du taux de division

2.1 Deux schémas d'observation

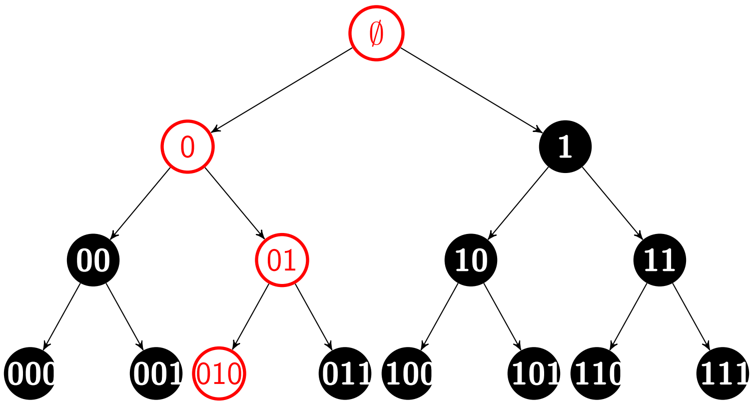
Soit $\mathcal{U}_n \subset \mathcal{U}$ un sous-ensemble de taille n de nœuds connectés tel que si u appartient à $\mathcal{U}_n$ alors son parent u^- aussi. On cherche un estimateur du taux de division B qui serait une fonction $y \to \widehat{B}_n(y) := \widehat{B}_n(y, (\xi_u,\tau_u), u \in \mathcal{U}_n)$ où $y \in]0,+\infty[$. La conclusion statistique est basée sur un schéma d'observation $((\xi_u,\tau_u), u \in \mathcal{U}_n)$. On s'intéresse particulièrement à deux schémas d'observation.

1. **Le cas de l'arbre total**. On observe chaque paire (ξ_u,τ_u) jusqu'à la N_n ième génération de l'arbre. On prend dans ce cas $\mathcal{U}_n = \{u \in \mathcal{U} |\ |u| \leq N_n\}$ où $|u| = k$ si $u = (u_0,u_1,\ldots,u_k) \in \mathcal{U}$ et N_n est choisi de telle sorte que 2^{N_n} soit de l'ordre de n.



Arbre jusqu'à la 2ème génération

2. **Le cas d'une branche de l'arbre**. On s'intéresse aux n ancêtres d'une cellule le long d'une ligne fixée de descendants (ie une branche de l'arbre). Cela signifie que pour un $u \in \mathcal{U}$ tel que $|u| = n$, on observe les tailles à la naissance et les taux de croissance des cellules (u_0), (u_0, u_1) jusqu'à $u = (u_0, \ldots u_n)$.



Branche choisie aléatoirement

2.2 Estimation du taux de division

Soit $y = (x, v)$ un élément de l'espace d'état $\mathcal{S} =]0, +\infty[\times\mathcal{E}$ (taille à la naissance, taux de croissance). On introduit le noyau de transition de la taille et du taux de croissance (ξ_u, τ_u) de la cellule u à sa naissance, étant donné la taille à la naissance et le taux de croissance de son parent (ξ_{u^-}, τ_{u^-}) par

$$\mathcal{P}_B(y, dy') := \mathbb{P}((\xi_u, \tau_u) \in dy' | (\xi_{u^-}, \tau_{u^-}) = y).$$

Théorème 7. *On a* $\mathcal{P}_B(y, dy') = \mathcal{P}_B((x, v), x', dv')dx'$ *où*

$$\mathcal{P}_B((x, v), x', dv') = \frac{B(2x')}{vx'} 1_{\{x' \geq x/2\}} \exp\Big(-\int_{x/2}^{x'} \frac{B(2s)}{vs} ds\Big) \rho(v, dv') \qquad (17)$$

Démonstration. On a

$$\begin{aligned}\mathbb{P}(d_{u^-} \in dt|\xi_{u^-} = x, \tau_{u^-} = v) &= B(\xi_t^{u^-}) \exp\Big(- \int_0^t B(\xi_s^{u^-})ds\Big)dt \\ &= B(x\exp(vt))\exp\Big(- \int_0^t B(x\exp(vs))ds\Big)dt\end{aligned} \tag{18}$$

On a alors

$$\mathbb{P}(d_{u^-} > t|\xi_{u^-} = x, \tau_{u^-} = v) = 1 - \exp\Big(- \int_0^t B(x\exp(vs))ds\Big)$$

On utilise alors la relation (2) qui donne la taille à la naissance d'une cellule par rapport aux paramètres de son parent.

Comme la taille à la naissance de u est au moins la moitié de la taille de la naissance de u^-, on se place sur l'événement $\{x' \geq x/2\}$:

$$\begin{aligned}\mathbb{P}(\xi_u \geq x'|\xi_{u^-} = x, \tau_{u^-} = v) &= \mathbb{P}\left(\frac{1}{2}\xi_{u^-}\exp(\tau_{u^-} d_{u^-}) \geq x'|\xi_{u^-} = x, \tau_{u^-} = v\right) \\ &= \mathbb{P}\left(d_{u^-} \geq \frac{1}{\tau_{u^-}}\ln\left(\frac{2x'}{\xi_{u^-}}\right)|\xi_{u^-} = x, \tau_{u^-} = v\right) \\ &= 1 - \exp\Big(- \int_0^{\frac{1}{v}\ln\left(\frac{2x'}{x}\right)} B(x\exp(vs))ds\Big)\end{aligned}$$

On fait alors le changement de variable $z = \frac{x\exp(vs)}{2}$ pour obtenir

$$\mathbb{P}(\xi_u \geq x'|\xi_{u^-} = x, \tau_{u^-} = v) = 1 - \exp\Big(- \int_{x/2}^{x'} \frac{B(2z)}{vz}dz\Big)$$

d'où

$$\mathbb{P}(\xi_u \in dx'|\xi_{u^-} = x, \tau_{u^-} = v) = \frac{B(2x')}{vx'}1_{\{x' \geq x/2\}}\exp\Big(- \int_{x/2}^{x'} \frac{B(2s)}{vs}ds\Big)dx' \tag{19}$$

On en déduit alors que $\mathcal{P}_B(y, dy') = \mathcal{P}_B((x, v), x', dv')dx'$ où

$$\mathcal{P}_B((x, v), x', dv') = \frac{B(2x')}{vx'}1_{\{x' \geq x/2\}}\exp\Big(- \int_{x/2}^{x'} \frac{B(2s)}{vs}ds\Big)\rho(v, dv') \tag{20}$$

□

Proposition 8. *Sous l'hypothèse 1, si $\mathcal{P}_B$ admet une probabilité invariante ν_B de la forme $\nu_B(dy) = \nu_B(x, dv)dx$ alors on a*

$$\nu_B(x) = \frac{B(2x)}{x}E_{\nu_B}\left[\frac{1}{\tau_{u^-}}1_{\{\xi_{u^-} \leq 2x, \xi_u \geq x\}}\right] \tag{21}$$

où E_{ν_B} désigne l'espérance sous la condition que $(\xi_\emptyset, \tau_\emptyset)$ soit de loi ν_B et où on a posé $\nu_B(x) = \int_{\mathcal{E}} \nu_B(y, dv')$.

Démonstration. Soit $x' \in]0,+\infty[$,
Comme ν_B est une probabilité invariante pour $\mathcal{P}_B$, on a

$$\nu_B(x',dv') = \int_{\mathcal{S}} \mathcal{P}_B((x,v),x',dv')\nu_B(x,dv)dx.$$

On utilise alors l'expression (17) , on obtient

$$\begin{aligned}
\nu_B(x',dv') &= \frac{B(2x')}{x'} \int_{\mathcal{S}} 1_{\{x' \geq x/2\}} \exp\Big(- \int_{x/2}^{x'} \frac{B(2s)}{vs} ds \Big) \frac{\rho(v,dv')}{v} \nu_B(x,dv)dx \\
&= \frac{B(2x')}{x'} \int_0^{2x'} \int_{\mathcal{E}} \exp\Big(- \int_{x/2}^{x'} \frac{B(2s)}{vs} ds \Big) \frac{\rho(v,dv')}{v} \nu_B(x,dv)dx
\end{aligned}$$

Or par l'hypothèse 1, comme $\int_{x/2}^{\infty} \frac{B(2s)}{s} ds = \infty$, on a

$$\int_{x'}^{\infty} \frac{B(2s)}{vs} \exp\Big(-\int_{x/2}^{s} \frac{B(2s')}{vs'} ds'\Big) ds = \Big[- \exp\Big(- \int_{x/2}^{s} \frac{B(2s')}{vs'} ds'\Big)\Big]_{x'}^{\infty} = \exp\Big(-\int_{x/2}^{x'} \frac{B(2s)}{vs} ds\Big)$$

On a alors

$$\begin{aligned}
\nu_B(x',dv') &= \frac{B(2x')}{x'} \int_0^{2x'} \int_{\mathcal{E}} \int_{x'}^{\infty} \frac{B(2s)}{vs} \exp\Big(- \int_{x/2}^{s} \frac{B(2s')}{vs'} ds' \Big) ds \frac{\rho(v,dv')}{v} \nu_B(x,dv)dx \\
&= \frac{B(2x')}{x'} \int_{\mathcal{S}} \int_0^{\infty} 1_{\{x \leq 2x', s \geq x'\}} \frac{B(2s)}{vs} \exp\Big(- \int_{x/2}^{s} \frac{B(2s')}{vs'} ds' \Big) ds \frac{\rho(v,dv')}{v} \nu_B(x,dv)dx \\
&= \frac{B(2x')}{x'} \int_{\mathcal{S}} \int_0^{\infty} 1_{\{x \leq 2x', s \geq x'\}} \mathcal{P}_B((x,v)s,dv') ds \frac{1}{v} \nu_B(x,dv)dx
\end{aligned}$$

On obtient le résultat en intégrant par rapport à dv'. □

Grâce à la relation (21) et par un changement de variable, on obtient

$$B(x) = \frac{x}{2} \frac{\nu_B(x/2)}{E_{\nu_B}\Big[\frac{1}{\tau_{u^-}} 1_{\{\xi_{u^-} \leq x, \xi_u \geq x/2\}}\Big]} \tag{22}$$

sous la condition que le dénominateur soit positif. On peut en déduire un estimateur du taux de division en remplaçant $\nu_B(y/2)$ et l'espérance au dénominateur par une estimation empirique. Pour cela, on prend un noyau $K : [0,+\infty] \to \mathbb{R}^+$ tel que $\int_0^\infty K(y)dy = 1$ et on pose $K_h(y) = \frac{K(y/h)}{h}$ pour $y \in [0,+\infty[$ et $h > 0$. Notre estimateur est défini par

$$\widehat{B}_n(x) = \frac{x}{2} \frac{\frac{1}{n}\sum_{u \in \mathcal{U}_n} K_h(\xi_u - x/2)}{\frac{1}{n}\sum_{u \in \mathcal{U}_n} \frac{1}{\tau_{u^-}} (1_{\{\xi_{u^-} \leq x, \xi_u \geq x/2\}}) \vee C} \tag{23}$$

où $C > 0$ est un seuil qui assure que l'estimateur est bien défini dans tous les cas. L'estimateur dépend donc du choix du noyau K, de la largeur de bande h et du seuil $C > 0$.