

Détection du vieillissement cellulaire au moyen de chaînes de Markov bifurcantes sur un arbre de Galton-Watson

Jean-François Delmas, Laurence Marsalle

► **To cite this version:**

Jean-François Delmas, Laurence Marsalle. Détection du vieillissement cellulaire au moyen de chaînes de Markov bifurcantes sur un arbre de Galton-Watson. 41èmes Journées de Statistique, SFdS, Bordeaux, 2009, Bordeaux, France, France. inria-00386569

HAL Id: inria-00386569

<https://hal.inria.fr/inria-00386569>

Submitted on 22 May 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

DÉTECTION DU VIEILLISSEMENT CELLULAIRE AU MOYEN DE CHÂÎNES DE MARKOV BIFURCANTES SUR UN ARBRE DE GALTON-WATSON

Jean-François Delmas & Laurence Marsalle

*Jean-François Delmas, CERMICS, Univ. Paris-Est, 6-8 av. Blaise Pascal,
Champs-sur-Marne, 77455 Marne La Vallée, France.*

*Laurence Marsalle, Lab. P. Painlevé, CNRS UMR 8524, Bât. M2, Univ. Lille 1, Cité
Scientifique, 59655 Villeneuve d'Ascq Cedex, France.*

Résumé Les processus auto-régressifs bifurcants sont souvent utilisés pour modéliser les divisions cellulaires, voire notamment le travail originel de Cowan et Staudte (1986), et celui de Guyon (2007) concernant la détection du vieillissement cellulaire. Pour prendre en compte la mort possible des cellules, nous utilisons un arbre de Galton-Watson pour modéliser la généalogie incomplète, et introduisons la notion de chaîne de Markov bifurcante sur un arbre de Galton-Watson. Après avoir calculé les estimateurs du maximum de vraisemblance liés à ce modèle, nous établissons leur consistance et leur normalité asymptotique, pour finalement construire un test qui permette de décider du vieillissement des cellules.

Mots clés détection du vieillissement cellulaire, processus de Galton-Watson, auto-régressif bifurcant.

Abstract Bifurcating autoregressive processes are widely used for modelling cell lineages, see e.g. the work of Cowan and Staudte (1986) and that of Guyon (2007) dealing with the detection of cellular aging. To take into account the possible death of cells, we use a Galton-Watson tree to model the maybe incomplete genealogy, and are thus led to define bifurcating Markov chain on Galton-Watson tree. After computing the maximum likelihood estimators for our model, we prove their consistency and asymptotic normality. Finally, we use these results to build a test to decide whether cells age or not.

Key words detection of cellular aging, Galton-Watson process, bifurcating autoregressive process.

1 Introduction

Ce travail a pour origine des expériences effectuées par une équipe de biologistes (Stewart et al. (2005)) sur la bactérie *Escherichia coli*. Cette bactérie est un organisme unicellulaire qui se reproduit en se divisant par son milieu. Les deux cellules filles possèdent donc

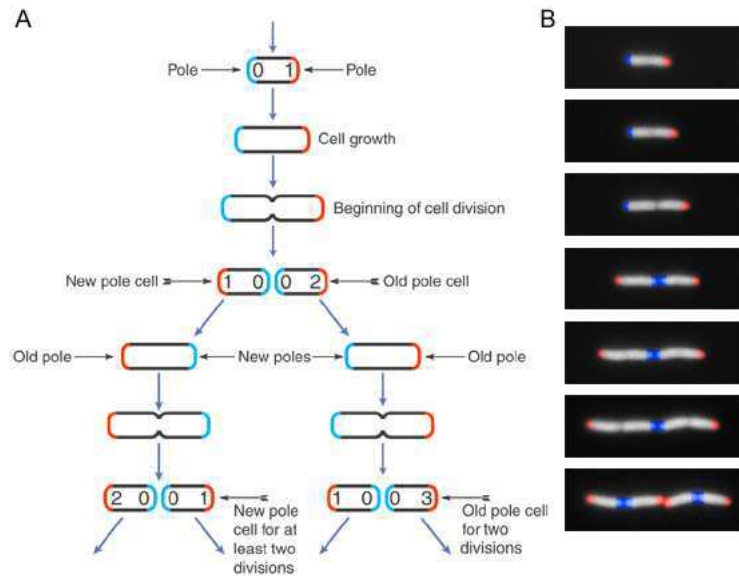


Figure 1: Divisions cellulaires d'*E. coli*, d'après Stewart *et al.* (2005)

chacune une extrémité neuve, appelée nouveau pôle, et une extrémité ancienne, appelée vieux pôle. On attribue un âge à la cellule, l'âge de son vieux pôle, au sens du nombre de divisions depuis lesquelles ce vieux pôle existe. En s'appuyant sur la figure 1, on peut remarquer qu'à chaque génération une cellule donne naissance à deux cellules, qui ont chacune un nouveau pôle, et qu'une des deux cellules présente un vieux pôle d'âge 1 (qui correspond au nouveau pôle de sa mère), alors que l'autre présente un pôle d'âge plus grand que 1 (qui correspond au vieux pôle de sa mère). La première cellule est appelée fille de type nouveau pôle, et la seconde fille de type vieux pôle. Les données expérimentales de Stewart *et al.* (2005) suggèrent que le taux de croissance de la fille de type nouveau pôle est significativement plus grand que le taux de croissance de la fille de type vieux pôle.

Guyon (2007) a modélisé l'évolution du taux de croissance par une chaîne de Markov asymétrique sur un arbre binaire régulier, encore appelée chaîne de Markov bifurcante. Ce modèle permet de prendre en compte une répartition asymétrique du taux de croissance de la cellule entre ses deux filles, selon leur type. A partir de là, Guyon construit un test pour détecter une différence entre les taux de croissance des deux filles selon leur type de pôle. Dans ce modèle, on suppose que les cellules ne meurent jamais (la mort signifie ici que la cellule ne se divise pas). Cette hypothèse est satisfaite dans le cas d'un milieu saturé en nourriture, mais dans des conditions de stress, les cellules mortes peuvent représenter une part plus significative de la population. Nous prenons en compte cet effet aléatoire en utilisant un arbre de Galton-Watson pour modéliser la généalogie, et nous étudions un modèle de chaîne de Markov sur un arbre de Galton-Watson au lieu

d'un arbre régulier. L'estimation des paramètres d'un auto-régressif bifurcant a déjà été étudiée, nous mentionnons, dans le cas symétrique, Hwang, Basawa et Yeo (2008), et le récent travail de Bercu, De Saporta et Gégoud-Petit (2008) dans le cas asymétrique.

2 Le modèle

Nous décrivons la généalogie des cellules par un arbre binaire. La cellule initiale est numérotée 1, elle donne naissance à deux filles, 2 et 3. Plus généralement, la cellule i donne naissance à deux filles numérotées $2i$ et $2i + 1$, et par convention, $2i$ désignera la cellule de type nouveau pôle, $2i + 1$ la cellule de type ancien pôle. Les ensembles

$$\mathbb{G}_r = \{2^r, \dots, 2^{r+1} - 1\}, \quad \mathbb{T}_r = \cup_{q=0}^r \mathbb{G}_q \quad \text{et} \quad \mathbb{T} = \cup_{r \geq 0} \mathbb{G}_r$$

désignent alors respectivement la r -ème génération, le sous-arbre des $r + 1$ premières générations et l'arbre binaire complet. Le taux de croissance de la cellule i est X_i . Si une cellule meurt (*i.e.* elle ne se divise pas), nous considérons qu'elle ne croît pas. Nous travaillons avec le modèle de type auto-régressif suivant :

- avec probabilité $p_{1,0}$ la cellule i se divise en deux cellules vivantes $2i$ et $2i + 1$ et

$$\begin{cases} X_{2i} &= \alpha_0 X_i + \beta_0 + \varepsilon_{2i}, \\ X_{2i+1} &= \alpha_1 X_i + \beta_1 + \varepsilon_{2i+1}, \end{cases}$$

où $((\varepsilon_{2i}, \varepsilon_{2i+1}), i \geq 1)$ est une suite i.i.d. de vecteurs gaussiens centrés, de matrice de covariance

$$\sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \quad \sigma^2 > 0, \quad \rho \in (-1, 1),$$

- avec probabilité p_0 la cellule i se divise en deux cellules dont une seule est vivante, de type nouveau pôle, et

$$X_{2i} = \alpha'_0 X_i + \beta'_0 + \varepsilon'_{2i},$$

où $(\varepsilon'_{2i}, i \geq 1)$ est une suite i.i.d. de gaussiennes $\mathcal{N}(0, \sigma_0'^2)$,

- avec probabilité p_1 la cellule i se divise en deux cellules dont une seule est vivante, de type ancien pôle, et

$$X_{2i+1} = \alpha'_1 X_i + \beta'_1 + \varepsilon'_{2i+1},$$

où $(\varepsilon'_{2i+1}, i \geq 1)$ est une suite i.i.d. de gaussiennes $\mathcal{N}(0, \sigma_1'^2)$.

Pour achever la mise en place du modèle, nous précisons que les coefficients $\alpha_0, \alpha_1, \alpha'_0$ et α'_1 sont dans $] - 1, 1[$, les coefficients $\beta_0, \beta_1, \beta'_0$ et β'_1 sont dans \mathbb{R} , et les trois suites $((\varepsilon_{2i}, \varepsilon_{2i+1}), i \geq 1)$, $(\varepsilon'_{2i}, i \geq 1)$ et $(\varepsilon'_{2i+1}, i \geq 1)$ sont indépendantes.

A partir de ce modèle, détecter le vieillissement revient à tester $(\alpha_0, \beta_0) = (\alpha_1, \beta_1)$ contre $(\alpha_0, \beta_0) \neq (\alpha_1, \beta_1)$.

Nous introduisons de nouvelles notations : \mathbb{G}_r^* , \mathbb{T}_r^* et \mathbb{T}^* désignent les ensembles de cellules vivantes parmi \mathbb{G}_r , \mathbb{T}_r et \mathbb{T} respectivement. Le processus $(|\mathbb{G}_r^*|, r \geq 0)$ est alors un processus de Galton-Watson, dont la loi de reproduction a pour moyenne $m := 2p_{1,0} + p_1 + p_0$. Nous nous plaçons sous l'hypothèse $\mathbf{m} > \mathbf{1}$ (cas surcritique).

3 Résultats

Pour commencer, nous calculons $\hat{\theta}^r = (\hat{\alpha}_0^r, \hat{\beta}_0^r, \hat{\alpha}_1^r, \hat{\beta}_1^r)$, l'estimateur du maximum de vraisemblance de $\theta = (\alpha_0, \beta_0, \alpha_1, \beta_1)$ basé sur \mathbb{T}_r . Pour $\delta \in \{0, 1\}$

$$\hat{\alpha}_\delta^r = \frac{Cov^r(X_i, X_{2i+\delta})}{Var^r(X_i)} \quad \text{et} \quad \hat{\beta}_\delta^r = Moy^r(X_{2i+\delta}) - \hat{\alpha}_\delta^r Moy^r(X_i),$$

où

$$Moy^r(X_i) = \frac{1}{|\tilde{\mathbb{T}}_r|} \sum_{i \in \tilde{\mathbb{T}}_r} X_i, \quad Var(X_i)^r = \frac{1}{|\tilde{\mathbb{T}}_r|} \sum_{i \in \tilde{\mathbb{T}}_r} X_i^2 - (Moy^r(X_i))^2,$$

$$Cov^r(X_i, X_{2i+\delta}) = \frac{1}{|\tilde{\mathbb{T}}_r|} \sum_{i \in \tilde{\mathbb{T}}_r} X_i X_{2i+\delta} - Moy^r(X_i) Moy^r(X_{2i+\delta}),$$

et $\tilde{\mathbb{T}}_r = \{i \in \mathbb{T}_r^*, 2i \in \mathbb{T}^* \text{ et } 2i + 1 \in \mathbb{T}^*\}$. Cet ensemble est naturellement apparu lors du calcul des estimateurs, et plus généralement, ce sont bien évidemment les cellules vivantes qui vont apporter de l'information. Ceci nous amène à étudier le comportement asymptotique du processus $(|\mathbb{G}_k^*|, k \geq 0)$.

Lemme (Harris (1963)) : *Il existe une variable aléatoire positive W telle que*

$$\frac{|\mathbb{G}_k^*|}{m^k} \xrightarrow[k \rightarrow \infty]{\text{p.s., L}^2} W,$$

avec $\{W = 0\} = \cup_{k \geq 0} \{|\mathbb{G}_k^*| = 0\}$ et $\mathbb{P}(W = 0) < 1$.

Nous pouvons alors énoncer les deux résultats suivants concernant l'estimateur $\hat{\theta}^r$.

Proposition 1 : *Supposons que la loi de X_0 admette des moments de tous ordres. Alors*

$$\mathbf{1}_{\{|\mathbb{G}_r^*| > 0\}} \hat{\theta}^r \xrightarrow[r \rightarrow \infty]{\text{Pr}} \mathbf{1}_{\{W \neq 0\}} \theta.$$

Proposition 2 : *Supposons que la loi de X_0 admette des moments de tous ordres. Alors*

$$\mathbf{1}_{\{|\mathbb{G}_r^*| > 0\}} |\tilde{\mathbb{T}}_r|^{1/2} (\hat{\theta}^r - \theta) \xrightarrow[r \rightarrow \infty]{\text{loi}} \mathbf{1}_{\{W \neq 0\}} G_4,$$

où G_4 est un vecteur gaussien de dimension 4, indépendant de W , centré, et de matrice de covariance

$$\Sigma = \sigma^2 \begin{pmatrix} K & \rho K \\ \rho K & K \end{pmatrix} \quad \text{avec} \quad K = (\mu_2 - \mu_1^2)^{-1} \begin{pmatrix} 1 & -\mu_1 \\ -\mu_1 & \mu_2 \end{pmatrix},$$

où μ_1 et μ_2 sont les moments d'ordre 1 et 2 respectivement d'une certaine loi de probabilité μ .

Nous donnons le schéma de la preuve : le processus $(X_i, i \in \mathbb{T}^*)$ défini ci-dessus est une chaîne de Markov bifurcante sur un arbre de Galton-Watson. L'étude du comportement asymptotique de tels processus fait apparaître une chaîne de Markov standard, Y . Sous l'hypothèse fondamentale d'ergodicité de cette chaîne de Markov auxiliaire, nous prouvons la loi des grands nombres et le TCL fonctionnel pour les chaînes de Markov bifurcantes sur un arbre de Galton-Watson. Les objets limites font intervenir la mesure invariante de Y , que nous appelons μ . Dans le cas particulier du modèle de type auto-régressif décrit ci-dessus, l'ergodicité de la chaîne de Markov auxiliaire Y se montre aisément, et les valeurs de μ_1 et μ_2 sont explicitées en fonction des paramètres du modèle.

Références

- [1] Bercu B., De Saporta B. et Gégoud-Petit A. (2008) Asymptotic analysis for bifurcating autoregressive processes via a martingale approach. *arXiv*, 0807.0528v1.
- [2] Cowan R. et Staudte R.G. (1986) The bifurcating autoregression model in cell lineage. *Biometrics*, 42, 769–783.
- [3] Guyon J. (2007) Limit theorems for bifurcating Markov chains. Application to the detection of cellular aging. *Ann. Appl. Probab.*, 17, 1538–1569.
- [4] Harris T. (1963) *The theory of branching processes*, Springer-Verlag, Berlin.
- [5] Hwang S.Y., Basawa I.V. et Yeo I.K. (2008) Local asymptotic normality for bifurcating autoregressive processes and related inference. *Statistical Methodology*, 6, 61–69.
- [6] Stewart E.J., Madden, R., Paul, G. et Taddéi F. (2005) Aging and death in an organism that reproduces by morphologically symmetric division. *PLoS Biol.*, 3(2), 295–300.