

Théorèmes limites pour les Chaînes de Markov bifurcantes. Application à la détection du vieillissement cellulaire.

Julien Guyon

28 août 2008

E. Coli : la jeunesse éternelle ?

Un modèle mathématique du vieillissement chez *E. coli*

Chaînes de Markov bifurcantes

Détection du vieillissement chez *E. coli*

E. Coli : la jeunesse éternelle ?

Un modèle mathématique du vieillissement chez *E. coli*

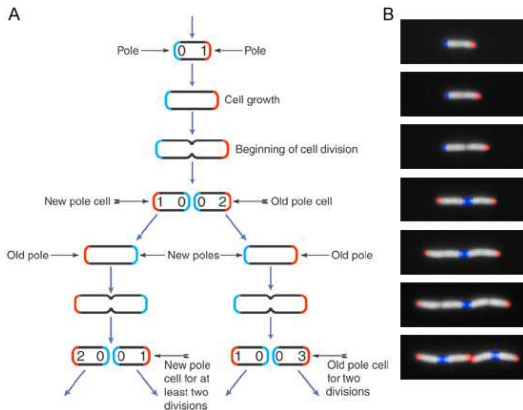
Chaînes de Markov bifurcantes

Détection du vieillissement chez *E. coli*

Introduction

- ▶ **Contexte.** Coopération avec le Laboratoire de génétique moléculaire, évolutive et médicale (INSERM U571, Faculté de Médecine Necker, Paris) : F. TADDÉI, E. J. STEWART, A. LINDNER and G. PAUL.

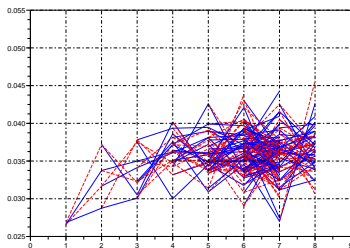
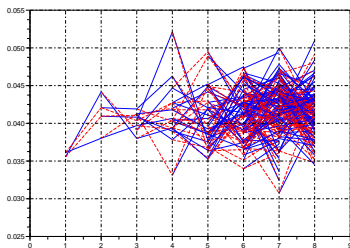
- ▶ **Objectifs.**
 - ▶ Détecter et quantifier le vieillissement cellulaire chez *E. Coli*, sur un seul film.
 - ▶ Détecter et quantifier de la variabilité phénotypique à génotype et environnement constants = Réviser un des dogmes de la biologie!

Le cycle d'*E. coli*, d'après E. J. STEWART *et al.*

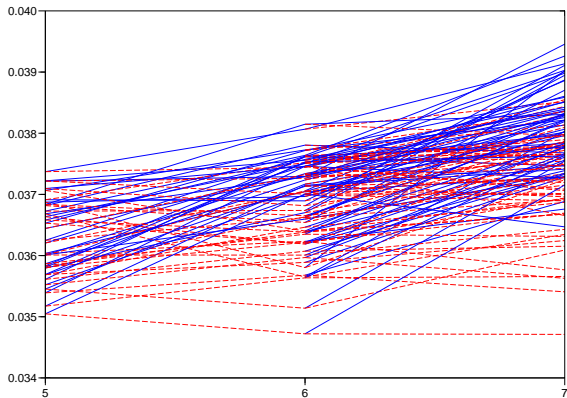
Pourquoi lier âge et vieux pôle ?

- ▶ Des agrégats de protéines mal refermées peuvent se former à un pôle et y restent.
- ▶ Ils sont automatiquement transmis à la fille “vieux pôle”, à partir de la 2e division.
- ▶ Ils sont symptomatiques d'un faible taux de croissance, lui-même lié au vieillissement.
- ▶ Mettre tous les problèmes du même côté : une stratégie avantageée par la sélection naturelle ?

Deux films



L'arbre moyen, sur 95 films

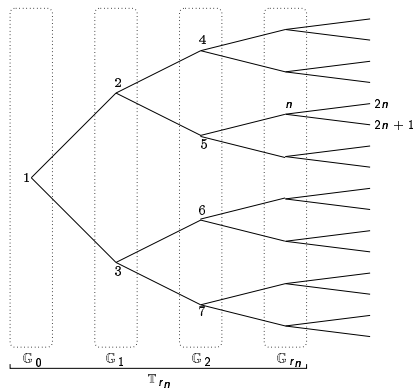


E. Coli : la jeunesse éternelle ?

Un modèle mathématique du vieillissement chez *E. coli*

Chaînes de Markov bifurcantes

Détection du vieillissement chez *E. coli*

L'arbre binaire \mathbb{T} 

Modèle autorégressif bifurcant

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

- ▶ $\alpha_0, \alpha_1 \in (-1, 1)$, $\beta_0, \beta_1 \in \mathbb{R}$
- ▶ $((\varepsilon_{2n}, \varepsilon_{2n+1}), n \geq 1)$ i.i.d., $(\varepsilon_{2n}, \varepsilon_{2n+1}) \sim \mathcal{N}_2(0, \Gamma)$ avec

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

- ▶ Objectifs
 - ▶ estimer le paramètre $\theta = (\alpha_0, \beta_0, \alpha_1, \beta_1)$, ρ and σ^2 ,
 - ▶ tester l'hypothèse nulle $H_0 = \{(\alpha_0, \beta_0) = (\alpha_1, \beta_1)\}$ contre son alternative $H_1 = \{(\alpha_0, \beta_0) \neq (\alpha_1, \beta_1)\}$.

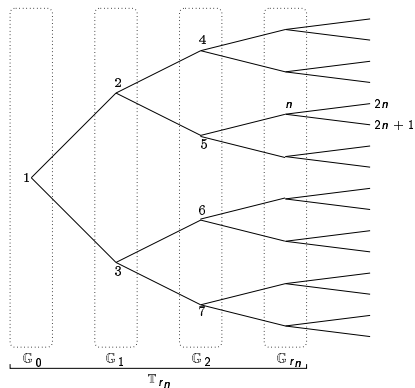
E. Coli : la jeunesse éternelle ?

Un modèle mathématique du vieillissement chez *E. coli*

Chaînes de Markov bifurcantes

Détection du vieillissement chez *E. coli*

Chaînes de Markov bifurcantes



Chaînes de Markov bifurcantes

► Description.

- Une famille de probabilités $P(x, dydz)$ modélise la probabilité que la fille de type 0 (nouveau pôle) soit dans l'état (taux de croissance) y et la fille de type 1 (vieux pôle) soit dans l'état z sachant que la mère est dans l'état x .
- Sachant X_n , on tire (X_{2n}, X_{2n+1}) suivant $P(X_n, dydz)$.
- La loi de X_{2n} sachant X_n peut différer de celle de X_{2n+1} sachant X_n .
- Sachant X_n , X_{2n} et X_{2n+1} peuvent ne pas être indépendants.

► Objectif.

- Démontrer des lois des grands nombres et un théorème central limite.

Quantités considérées

Moyenne de f sur :

▶ la génération r : $\frac{1}{|\mathbb{G}_r|} \sum_{n \in \mathbb{G}_r} f(X_n)$.

▶ les $r + 1$ premières générations : $\frac{1}{|\mathbb{T}_r|} \sum_{n \in \mathbb{T}_r} f(X_n)$.

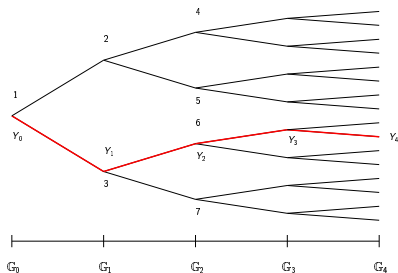
Explosion combinatoire :

▶ $|\mathbb{G}_r| = 2^r$

▶ $|\mathbb{T}_r| = 2^{r+1} - 1 = 2|\mathbb{G}_r| - 1$.

La chaîne de Markov induite ($Y_r, r \in \mathbb{N}$)

On part de la racine et on choisit une des deux filles au hasard :



Loi faible des grands nombres

- ▶ **Théorème.** Si la chaîne de Markov induite est ergodique, de loi stationnaire μ , alors pour toute f continue bornée, dans L^2 ,

$$\lim_{r \rightarrow \infty} \frac{1}{|\mathbb{G}_r|} \sum_{n \in \mathbb{G}_r} f(X_n) = \lim_{r \rightarrow \infty} \frac{1}{|\mathbb{T}_r|} \sum_{n \in \mathbb{T}_r} f(X_n) = (\mu, f).$$

Loi forte des grands nombres

- **Théorème.** Si la chaîne de Markov induite est **géométriquement** ergodique, de loi stationnaire μ , alors pour toute f continue bornée, **presque sûrement**

$$\lim_{r \rightarrow \infty} \frac{1}{|\mathbb{G}_r|} \sum_{n \in \mathbb{G}_r} f(X_n) = \lim_{r \rightarrow \infty} \frac{1}{|\mathbb{T}_r|} \sum_{n \in \mathbb{T}_r} f(X_n) = (\mu, f).$$

Théorème central limite

- ▶ Pour $f(x, y, z)$ continue bornée
 - ▶ $Pf(x) = \int_{S^2} f(x, y, z) P(x, dy, dz) = \mathbb{E}[f(X_1, X_2, X_3) | X_1 = x]$.
 - ▶ $(\mu, Pf^2) = \int \mu(dx) \int_{S^2} f(x, y, z)^2 P(x, dydz)$.
- ▶ **Théorème.** Supposons Y ergodique. Alors

$$\lim_{r \rightarrow \infty} \frac{1}{|\mathbb{T}_r|} \sum_{n \in \mathbb{T}_r} f(X_n, X_{2n}, X_{2n+1}) = \lim_{r \rightarrow \infty} \frac{1}{|\mathbb{T}_r|} \sum_{n \in \mathbb{T}_r} Pf(X_n) = (\mu, Pf).$$

De plus, avec $\sigma^2 = (\mu, Pf^2) - (\mu, (Pf)^2)$,

$$\frac{1}{\sqrt{|\mathbb{T}_r|}} \left(\sum_{n \in \mathbb{T}_r} (f(X_n, X_{2n}, X_{2n+1}) - Pf(X_n)) \right) \xrightarrow[n \rightarrow \infty]{\text{(loi)}} \mathcal{N}(0, \sigma^2).$$

E. Coli : la jeunesse éternelle ?

Un modèle mathématique du vieillissement chez *E. coli*

Chaînes de Markov bifurcantes

Détection du vieillissement chez *E. coli*

Détection du vieillissement chez *E. coli*

Modèle autorégressif bifurcant pour le taux de croissance :

$$\begin{cases} X_{2n} = \alpha_0 X_n + \beta_0 + \varepsilon_{2n} & (\text{type 0, nouveau pôle}), \\ X_{2n+1} = \alpha_1 X_n + \beta_1 + \varepsilon_{2n+1} & (\text{type 1, vieux pôle}), \end{cases}$$

- ▶ $\alpha_0, \alpha_1 \in (-1, 1)$, $\beta_0, \beta_1 \in \mathbb{R}$,
- ▶ $(\varepsilon_{2n}, \varepsilon_{2n+1})$ vecteurs gaussiens bidimensionnels i.i.d. avec variance commune $\sigma^2 > 0$ et corrélation $\rho \in (-1, 1)$.
- ▶ Paramètres : $\alpha_0, \alpha_1, \beta_0, \beta_1, \sigma^2, \rho$.

Estimation des paramètres au vu de \mathbb{T}_{r+1}

$$\begin{aligned} \hat{\alpha}_\delta^r &= \text{Cov}^r(X_n, X_{2n+\delta}) / \text{Var}^r(X_n), \\ \hat{\beta}_\delta^r &= \text{Mean}^r(X_{2n+\delta}) - \hat{\alpha}_\delta^r \text{Mean}^r(X_n), \end{aligned} \quad \text{où } \delta \in \{0, 1\},$$

$$\text{Mean}^r(X_n) = \frac{1}{|\mathbb{T}_r|} \sum_{n \in \mathbb{T}_r} X_n, \quad \text{Mean}^r(X_{2n+\delta}) = \frac{1}{|\mathbb{T}_r|} \sum_{n \in \mathbb{T}_r} X_{2n+\delta},$$

$$\text{Cov}^r(X_n, X_{2n+\delta}) = \frac{1}{|\mathbb{T}_r|} \sum_{n \in \mathbb{T}_r} X_n X_{2n+\delta} - \text{Mean}^r(X_n) \text{Mean}^r(X_{2n+\delta}),$$

$$\text{Var}^r(X_n) = \frac{1}{|\mathbb{T}_r|} \sum_{n \in \mathbb{T}_r} X_n^2 - \left(\text{Mean}^r(X_n) \right)^2.$$

Estimation des paramètres au vu de \mathbb{T}_{r+1}

$$\hat{\sigma}_r^2 = \frac{1}{2|\mathbb{T}_r|} \sum_{n \in \mathbb{T}_r} (\hat{\varepsilon}_{2n}^2 + \hat{\varepsilon}_{2n+1}^2),$$

$$\hat{\rho}_r = \frac{1}{\hat{\sigma}_r^2 |\mathbb{T}_r|} \sum_{n \in \mathbb{T}_r} \hat{\varepsilon}_{2n} \hat{\varepsilon}_{2n+1},$$

où

$$\hat{\varepsilon}_{2n} = X_{2n} - \hat{\alpha}_0^r X_n - \hat{\beta}_0^r \quad (\text{type 0, nouveau pôle}),$$

$$\hat{\varepsilon}_{2n+1} = X_{2n+1} - \hat{\alpha}_1^r X_n - \hat{\beta}_1^r \quad (\text{type 1, vieux pôle}).$$

La chaîne de Markov induite Y

- ▶ Y est un processus autorégressif généralisé :

$$Y_{r+1} = a_{r+1} Y_r + b_{r+1} + \tilde{\epsilon}_{r+1}, \quad (a_r, b_r, \tilde{\epsilon}_r) \text{ i.i.d.},$$

$$\tilde{\epsilon}_r \sim \mathcal{N}(0, \sigma^2),$$

$$\mathbb{P}((a_r, b_r) = (\alpha_0, \beta_0)) = \mathbb{P}((a_r, b_r) = (\alpha_1, \beta_1)) = 1/2.$$
- ▶ Y est géométriquement ergodique, μ n'est pas gaussienne.
- ▶ $\bar{\alpha} = \frac{\alpha_0 + \alpha_1}{2}$, $\bar{\beta} = \frac{\beta_0 + \beta_1}{2}$, etc.

$$\mu_1 \equiv \int x \mu(dx) = \frac{\bar{\beta}}{1 - \bar{\alpha}}$$

$$\mu_2 \equiv \int x^2 \mu(dx) = \frac{2\bar{\alpha}\bar{\beta} \mu_1 + \bar{\beta}^2 + \sigma^2}{1 - \bar{\alpha}^2}.$$

Loi des grands nombres : exemples

- $f(x, y, z) = x$, $Pf(x) = x$:

$$\text{Mean}^r(X_n) = \frac{1}{|\mathbb{T}_r|} \sum_{n \in \mathbb{T}_r} X_n \xrightarrow[r \rightarrow \infty]{\text{p.s.}} \mu_1.$$

- $f(x, y, z) = xy$, $Pf(x) = x(\alpha_0 x + \beta_0)$:

$$\frac{1}{|\mathbb{T}_r|} \sum_{n \in \mathbb{T}_r} X_n X_{2n} \xrightarrow[r \rightarrow \infty]{\text{p.s.}} (\mu, Pf) = \alpha_0 \mu_2 + \beta_0 \mu_1.$$

Convergence et normalité asymptotique des estimateurs

- ▶ Loi forte des grands nombres :

$$\hat{\theta}^r := (\hat{\alpha}_0^r, \hat{\beta}_0^r, \hat{\alpha}_1^r, \hat{\beta}_1^r) \xrightarrow[r \rightarrow \infty]{\text{p.s.}} (\alpha_0, \beta_0, \alpha_1, \beta_1) =: \theta.$$

$\hat{\sigma}_r^2$ et $\hat{\rho}_r$ convergent p.s. vers σ^2 et ρ .

- ▶ Fluctuations gaussiennes :

$$\sqrt{|\mathbb{T}_r|} (\hat{\theta}^r - \theta) \xrightarrow[r \rightarrow \infty]{(\text{loi})} \mathcal{N}(0, \Sigma),$$

où Σ s'exprime simplement en fonction de σ^2 , ρ , μ_1 and μ_2 .

Détection du vieillissement cellulaire

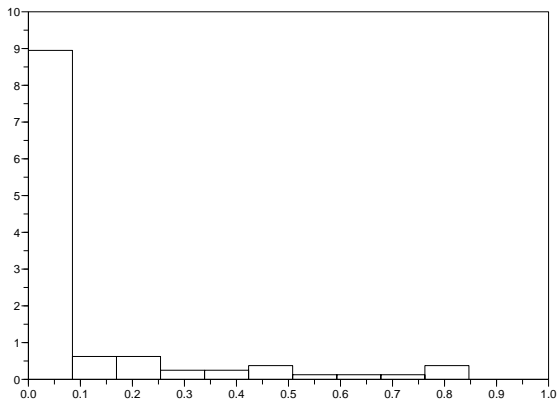
- ▶ But : rejeter $H_0 = \{(\alpha_0, \beta_0) = (\alpha_1, \beta_1)\}$.
- ▶ Test de Wald :

$$\zeta_r = \frac{|\mathbb{T}_r|}{2\hat{\sigma}_r^2 (1 - \hat{\rho}_r)} \left[(\hat{\alpha}_0^r - \hat{\alpha}_1^r)^2 (\hat{\mu}_2^r - \hat{\mu}_1^r) + ((\hat{\alpha}_0^r - \hat{\alpha}_1^r) \hat{\mu}_1^r + \hat{\beta}_0^r - \hat{\beta}_1^r)^2 \right].$$

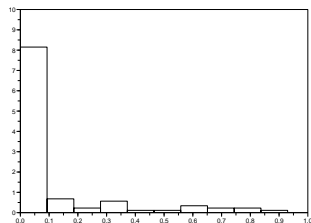
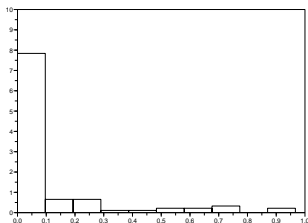
$$\text{Sous } H_0, \quad \zeta_r \xrightarrow[r \rightarrow \infty]{\text{(loi)}} \chi^2(2),$$

$$\text{Sinon,} \quad \zeta_r \xrightarrow[r \rightarrow \infty]{\text{p.s.}} +\infty.$$

- ▶ p -valeur = $\mathbb{P}(\chi^2(2) \geq \zeta_r^{\text{obs}})$. Asymptotiquement uniforme sur $[0, 1]$ sous H_0 . Calculée pour chaque film.

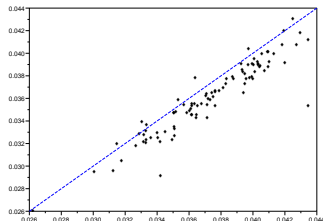
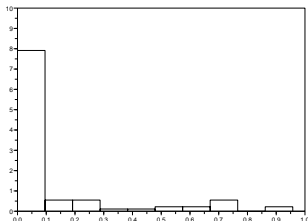


$H_0 = \{(\alpha_0, \beta_0) = (\alpha_1, \beta_1)\}$. Histogramme des 95 p -valeurs



Gauche : $H_0 = \{\alpha_0 = 0\}$. Histogramme des 95 p -valeurs.

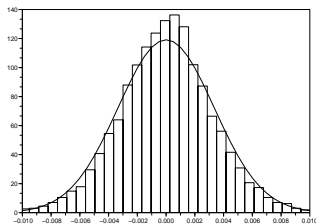
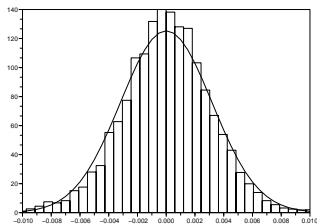
Droite : $H_0 = \{\alpha_1 = 0\}$. Histogramme des 95 p -valeurs.



$$H'_0 = \{\beta_0 / (1 - \alpha_0) = \beta_1 / (1 - \alpha_1)\};$$

gauche : histogramme des 95 p -valeurs ;

droite : $\hat{\beta}_0 / (1 - \hat{\alpha}_0)$ en abscisses, $\hat{\beta}_1 / (1 - \hat{\alpha}_1)$ en ordonnées.



Histogramme des résidus $\hat{\epsilon}_{2n}$ (nouveaux pôles, gauche) et $\hat{\epsilon}_{2n+1}$ (vieux pôles, droite)

Ces résultats ont été publiés dans

- ▶ ESAIM Proceedings (2005)
- ▶ Annals of Applied Probability (2007), volume 17, number 5/6, 1538-1569.