

Processus auto-régressifs de bifurcation et division cellulaire

Benoîte de Saporta, Anne Gégout-Petit, Laurence Marsalle

Plan de l'exposé

Introduction

BAR avec données manquantes

- Modèle d'observation

- Estimateurs

- Convergence

- Modèle multi-arbres

BAR à coefficients aléatoires

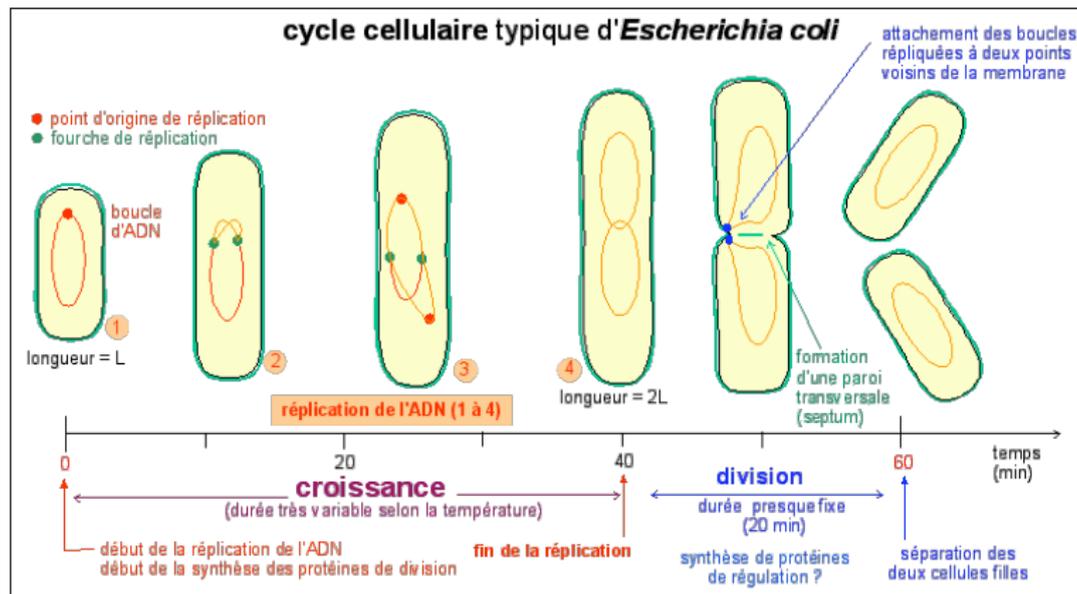
- Modèle

- Lois des grands nombres

Conclusion

Division cellulaire

film



Premier modèle BAR

[Cowan & Staudte 1986] Modèle auto-régressif de bifurcation

$$\begin{cases} X_{2k} &= a + bX_k + \epsilon_{2k} \\ X_{2k+1} &= a + bX_k + \epsilon_{2k+1} \end{cases}$$

$(\epsilon_{2k}, \epsilon_{2k+1})$ gaussiennes iid

$$\mathbb{E}[\epsilon_{2k+i}] = \sigma^2, \mathbb{E}[\epsilon_{2k}\epsilon_{2k+1}] = \rho$$

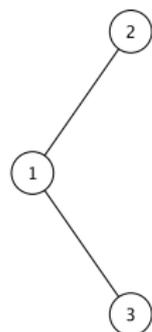
Régime stationnaire si $X_1 \sim \mathcal{N}\left(\frac{a}{1-b}, \frac{\sigma^2}{1-b^2}\right)$

①

Premier modèle BAR

[Cowan & Staudte 1986] Modèle auto-régressif de bifurcation

$$\begin{cases} X_{2k} &= a + bX_k + \epsilon_{2k} \\ X_{2k+1} &= a + bX_k + \epsilon_{2k+1} \end{cases}$$



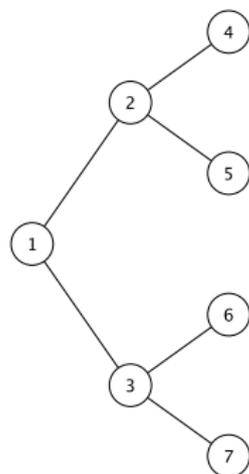
$(\epsilon_{2k}, \epsilon_{2k+1})$ gaussiennes iid

$$\mathbb{E}[\epsilon_{2k+i}] = \sigma^2, \mathbb{E}[\epsilon_{2k}\epsilon_{2k+1}] = \rho$$

Régime stationnaire si $X_1 \sim \mathcal{N}\left(\frac{a}{1-b}, \frac{\sigma^2}{1-b^2}\right)$

Premier modèle BAR

[Cowan & Staudte 1986] Modèle auto-régressif de bifurcation



$$\begin{cases} X_{2k} &= a + bX_k + \epsilon_{2k} \\ X_{2k+1} &= a + bX_k + \epsilon_{2k+1} \end{cases}$$

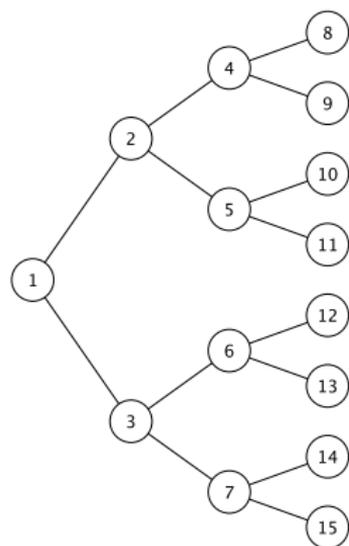
$(\epsilon_{2k}, \epsilon_{2k+1})$ gaussiennes iid

$$\mathbb{E}[\epsilon_{2k+i}] = \sigma^2, \mathbb{E}[\epsilon_{2k}\epsilon_{2k+1}] = \rho$$

Régime stationnaire si $X_1 \sim \mathcal{N}\left(\frac{a}{1-b}, \frac{\sigma^2}{1-b^2}\right)$

Premier modèle BAR

[Cowan & Staudte 1986] Modèle auto-régressif de bifurcation



$$\begin{cases} X_{2k} &= a + bX_k + \epsilon_{2k} \\ X_{2k+1} &= a + bX_k + \epsilon_{2k+1} \end{cases}$$

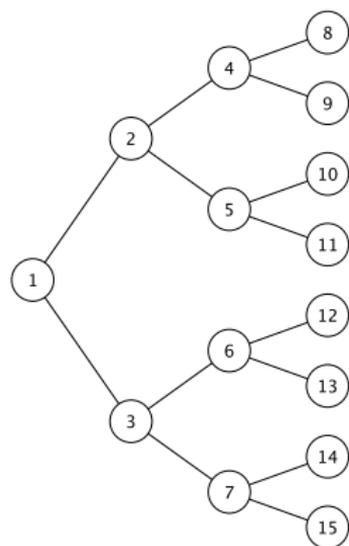
$(\epsilon_{2k}, \epsilon_{2k+1})$ gaussiennes iid

$$\mathbb{E}[\epsilon_{2k+i}] = \sigma^2, \mathbb{E}[\epsilon_{2k}\epsilon_{2k+1}] = \rho$$

Régime stationnaire si $X_1 \sim \mathcal{N}\left(\frac{a}{1-b}, \frac{\sigma^2}{1-b^2}\right)$

Premier modèle BAR

[Cowan & Staudte 1986] Modèle auto-régressif de bifurcation



$$\begin{cases} X_{2k} &= a + bX_k + \epsilon_{2k} \\ X_{2k+1} &= a + bX_k + \epsilon_{2k+1} \end{cases}$$

$(\epsilon_{2k}, \epsilon_{2k+1})$ gaussiennes iid

$$\mathbb{E}[\epsilon_{2k+i}] = \sigma^2, \mathbb{E}[\epsilon_{2k}\epsilon_{2k+1}] = \rho$$

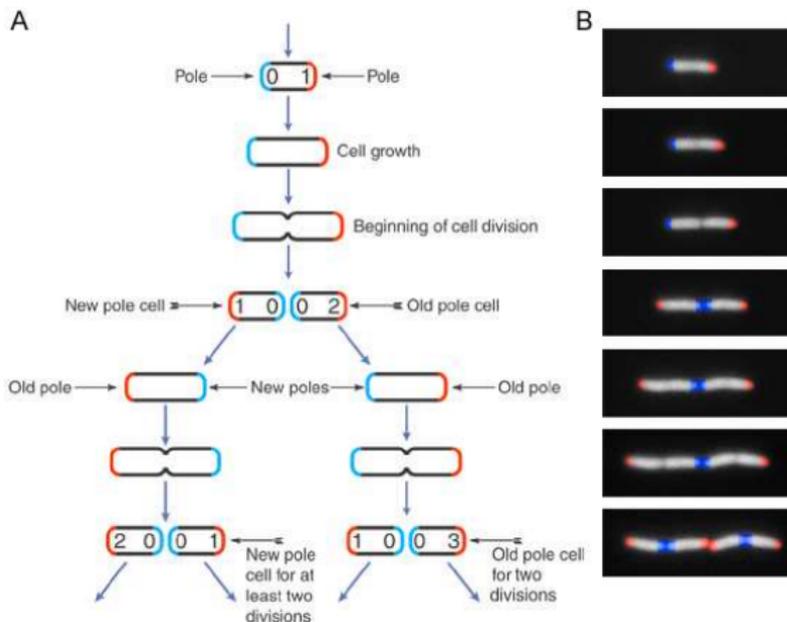
Régime stationnaire si $X_1 \sim \mathcal{N}\left(\frac{a}{1-b}, \frac{\sigma^2}{1-b^2}\right)$

Estimer les paramètres pour mesurer les corrélations

- ▶ b corrélation mère-fille
- ▶ $\phi = b^2 + (1 - b^2)\rho/\sigma^2$ corrélation entre soeurs

Asymétrie de la division

[Stewart & al. 2005]



BAR asymétrique

[Guyon 2007] Modèle asymétrique

$$\begin{cases} X_{2k} &= a + bX_k + \epsilon_{2k} \\ X_{2k+1} &= c + dX_k + \epsilon_{2k+1} \end{cases}$$

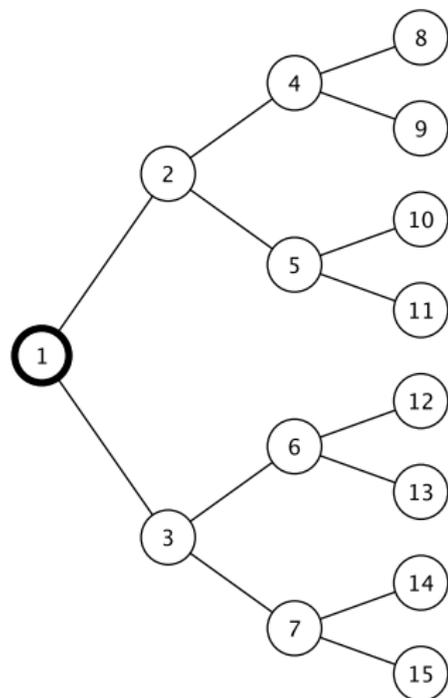
$(\epsilon_{2k}, \epsilon_{2k+1})$ gaussiennes iid, $\mathbb{E}[\epsilon_{2k+i}] = 0$, $\mathbb{E}[\epsilon_{2k}\epsilon_{2k+1}] = \rho$
pas de régime stationnaire

Estimer les paramètres pour tester l'asymétrie

- ▶ $(a, b) = (c, d)$
- ▶ $a/(1 - b) = c/(1 - d)$

Méthode chaînes de Markov bifurquantes en utilisant la structure d'arbre par générations

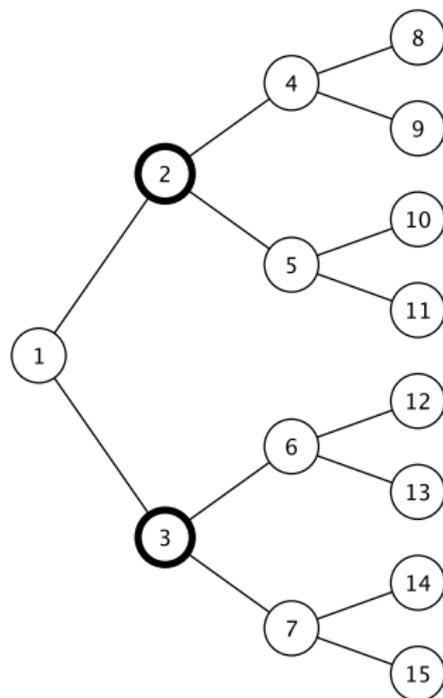
Génération



Génération 0:

$$G_0 = \{1\}$$

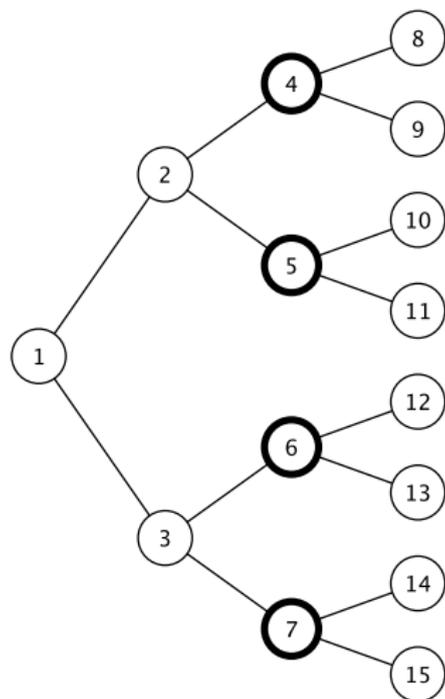
Génération



Génération 1:

$$G_1 = \{2, 3\}$$

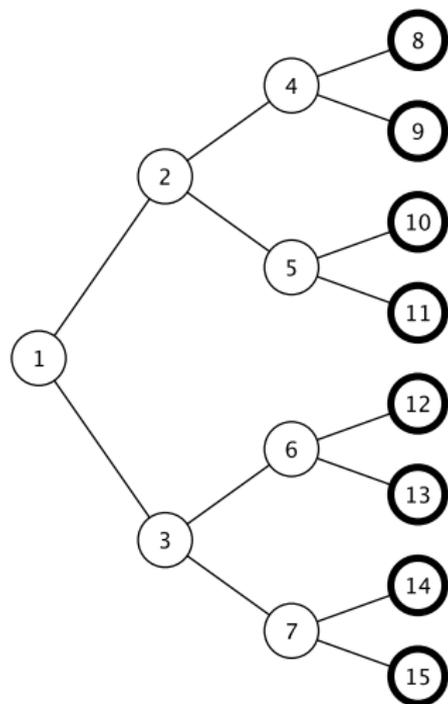
Génération



Génération 2:

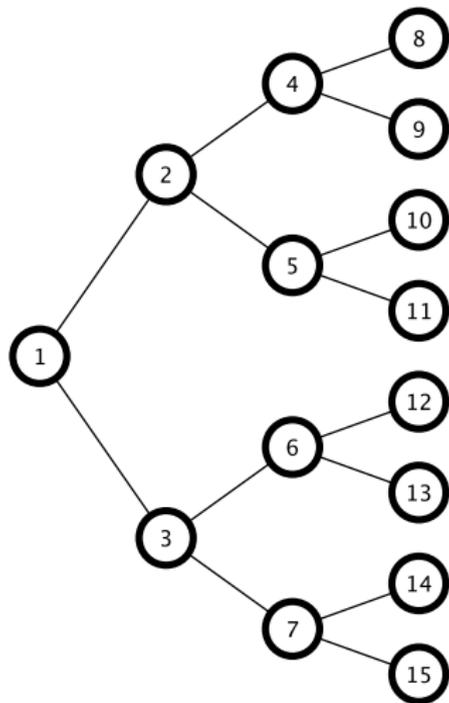
$$G_2 = \{4, 5, 6, 7\}$$

Génération

Génération n :

$$\mathbb{G}_n = \{2^n, 2^n + 1, \dots, 2^{n+1} - 1\}$$

Génération



Arbre jusqu'à la génération n :

$$\mathbb{T}_n = \bigcup_{\ell=0}^n \mathbb{G}_\ell$$

Méthode par chaîne de Markov bifurquante

- ▶ définition d'un modèle de **chaîne de Markov** sur un arbre binaire

$$\mathbb{E} \left[\prod_{k \in \mathbb{G}_n} f_k(X_{2k}, X_{2k+1}) \mid \sigma(X_j, j \in \mathbb{T}_n) \right] = \prod_{k \in \mathbb{G}_n} P f_k(X_k)$$

- ▶ comportement asymptotique de (X_k) donnée par la **chaîne induite**

$$\begin{cases} Y_0 &= X_1, \\ Y_{n+1} &= A_{n+1} + B_{n+1} Y_n \end{cases}$$

(A_n, B_n) iid de loi $(a + \epsilon_2, b) \mathbb{1}_{\{\zeta=1\}} + (c + \epsilon_3, d) \mathbb{1}_{\{\zeta=0\}}$,
 $\zeta \sim \text{Bernoulli}(1/2)$ **lignée aléatoire**

Première contribution

[EJP 2009] Modèle asymétrique

$$\begin{cases} X_{2k} &= a + bX_k + \epsilon_{2k} \\ X_{2k+1} &= c + dX_k + \epsilon_{2k+1} \end{cases}$$

Hypothèses

$\mathcal{F}_n = \sigma\{X_k, k \in \mathbb{T}_n\}$ filtration des générations

- ▶ moments d'ordre 8 pour le bruit
- ▶ différence de martingale $\mathbb{E}[\epsilon_{2k+i} | \mathcal{F}_n] = 0$ pour tout $k \in \mathbb{G}_n$, ϵ_{2k+i} indépendant de ϵ_{2l+j} conditionnellement à \mathcal{F}_n pour tout $k \neq l \in \mathbb{G}_n$
- ▶ $\mathbb{E}[\epsilon_{2k+i}^2 | \mathcal{F}_n] = \sigma^2$, $\mathbb{E}[\epsilon_{2k}\epsilon_{2k+1} | \mathcal{F}_n] = \rho$ pour tout $k \in \mathbb{G}_n$
- ▶ vitesse de convergence des estimateurs
- ▶ méthode martingale

Méthode martingale

Convergence des martingales L^2

(M_n) martingale **scalaire** bornée dans L^2

$$\langle M \rangle_n = \sum_{k=0}^n \mathbb{E}[(M_{k+1} - M_k)^2 \mid \mathcal{F}_k]$$

Si $\lim_{n \rightarrow \infty} \langle M \rangle_n = +\infty$, alors $\frac{M_n}{\langle M \rangle_n} \rightarrow 0$ p.s.

+ conditions de moment alors $\left(\frac{M_n}{\langle M \rangle_n}\right)^2 = \mathcal{O}\left(\frac{\log(\langle M \rangle_n)}{\langle M \rangle_n}\right)$ p.s.

Mise en œuvre

- ▶ identifier une martingale (vectorielle) pour la filtration des **générations**
- ▶ calculer la limite du crochet
- ▶ **appliquer** le théorème de vitesse de convergence ?

Méthode martingale

Convergence des martingales L^2

(M_n) martingale **scalaire** bornée dans L^2

$$\langle M \rangle_n = \sum_{k=0}^n \mathbb{E}[(M_{k+1} - M_k)^2 \mid \mathcal{F}_k]$$

Si $\lim_{n \rightarrow \infty} \langle M \rangle_n = +\infty$, alors $\frac{M_n}{\langle M \rangle_n} \rightarrow 0$ p.s.

+ conditions de moment alors $\left(\frac{M_n}{\langle M \rangle_n}\right)^2 = \mathcal{O}\left(\frac{\log(\langle M \rangle_n)}{\langle M \rangle_n}\right)$ p.s.

Mise en œuvre

- ▶ identifier une martingale (vectorielle) pour la filtration des **générations**
- ▶ calculer la limite du crochet
- ▶ **redémontrer** le théorème de vitesse de convergence pour une martingale sur un **arbre binaire**

Données réelles

Données de [Stewart & al. 2005]

- ▶ 94 films = 94 généalogies
- ▶ 4 à 9 générations de cellules par généalogie
- ▶ aucune généalogie complète : cellules hors champs ou superposées

Données réelles

Données de [Stewart & al. 2005]

- ▶ 94 films = 94 généalogies
- ▶ 4 à 9 générations de cellules par généalogie
- ▶ aucune généalogie complète : cellules hors champs ou superposées

On ne peut pas appliquer notre procédure d'estimation et de test à ces données

Données réelles

Données de [Stewart & al. 2005]

- ▶ 94 films = 94 généalogies
- ▶ 4 à 9 générations de cellules par généalogie
- ▶ aucune généalogie complète : cellules hors champs ou superposées

On ne peut pas appliquer notre procédure d'estimation et de test à ces données

⇒ Nouvelle procédure d'estimation en tenant compte des données manquantes

Plan de l'exposé

Introduction

BAR avec données manquantes

Modèle d'observation

Estimateurs

Convergence

Modèle multi-arbres

BAR à coefficients aléatoires

Conclusion

Modèle Galton-Watson

[Delmas & Marsalle 2010]

- ▶ chaque cellule a un **type** 0 (pair–nouveau pôle) ou 1 (impair–ancien pôle)
- ▶ probabilité $p(j_0, j_1)$ d'avoir j_0 fille de type 0 et j_1 fille de type 1, tiré **indépendamment** pour chaque cellule
- ▶ Z_n nombre de cellules présentes à la génération n
Galton-Watson
- ▶ une cellule non observée n'a pas de descendante observée
- ▶ inférence sur le BAR partiellement observé par méthode chaîne de Markov bifurquante

Modèle Galton-Watson

[Delmas & Marsalle 2010]

- ▶ chaque cellule a un **type** 0 (pair–nouveau pôle) ou 1 (impair–ancien pôle)
- ▶ probabilité $p(j_0, j_1)$ d'avoir j_0 fille de type 0 et j_1 fille de type 1, tiré **indépendamment** pour chaque cellule
- ▶ Z_n nombre de cellules présentes à la génération n
Galton-Watson
- ▶ une cellule non observée n'a pas de descendante observée
- ▶ inférence sur le BAR partiellement observé par méthode chaîne de Markov bifurquante

Le nombre de cellules filles de chaque type devrait aussi dépendre du type de la mère

Modèle de Galton-Watson à deux types

- ▶ $\delta_k = 1$ si la cellule k est observée, 0 sinon
- ▶ probabilité $p^{(i)}(j_0, j_1)$ pour une cellule mère de type i d'avoir j_0 fille de type 0 et j_1 fille de type 1, tiré indépendamment pour chaque cellule
- ▶ Z_n^i nombre de cellules de type i présentes à la génération n , (Z_n^0, Z_n^1) processus de Galton-Watson à deux types
- ▶ une cellule non observée n'a pas de descendante observée

Extinction

Matrice de descendance

$$P = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix}$$

$p_{i0} = p^{(i)}(1, 0) + p^{(i)}(1, 1)$: nombre moyen de filles de **type 0**

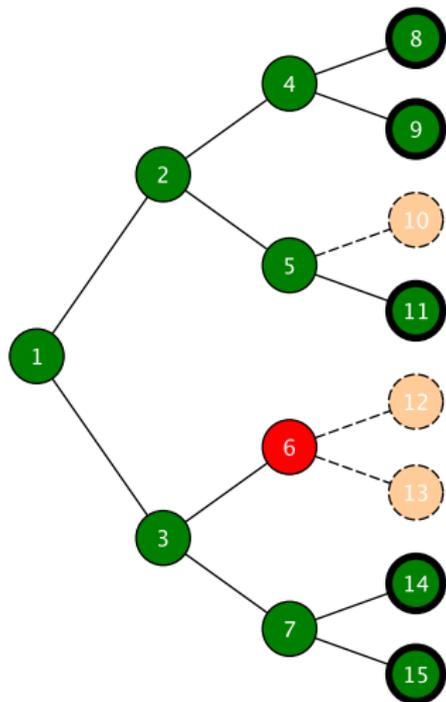
$p_{i1} = p^{(i)}(0, 1) + p^{(i)}(1, 1)$: nombre moyen de filles de **type 1**
d'une mère de type i

Critère d'extinction

π rayon spectral de P

- ▶ si $\pi \leq 1$, extinction presque sure
- ▶ si $\pi > 1$, extinction avec probabilité < 1

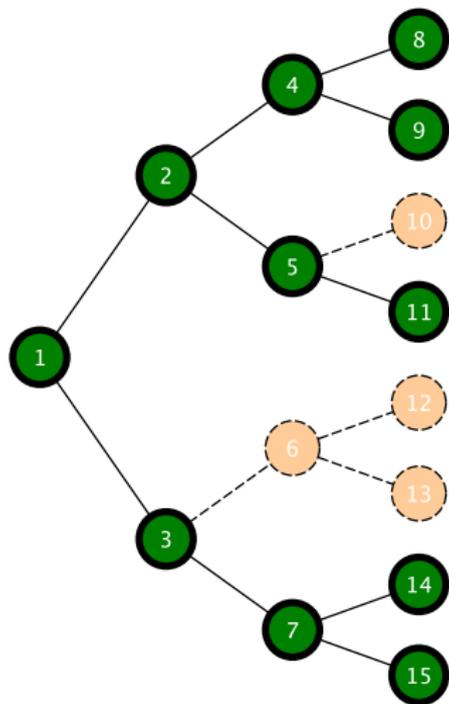
Génération observée



Génération n observée:

$$\mathbb{G}_n^* = \{k \in \mathbb{G}_n ; \delta_k = 1\}$$

Génération observées



Arbre jusqu'à la génération n :

$$\mathbb{T}_n^* = \{k \in \mathbb{T}_n ; \delta_k = 1\} = \cup_{\ell=0}^n \mathbb{G}_\ell^*$$

BAR partiellement observé

$$\begin{cases} X_{2k} &= a + b X_k + \epsilon_{2k} \\ X_{2k+1} &= c + d X_k + \epsilon_{2k+1} \end{cases}$$

Hypothèses

- ▶ indépendance entre (δ_k) et (X_k) et $(\epsilon_{2k}, \epsilon_{2k+1})$
- ▶ bruit différence de martingale et moments d'ordre 8

Estimation de $\theta = (a, b, c, d)^t$: minimiser

$$\Delta_n(\theta) = \frac{1}{2} \sum_{k \in \mathbb{T}_{n-1}} \delta_{2k} (X_{2k} - a - bX_k)^2 + \delta_{2k+1} (X_{2k+1} - c - dX_k)^2.$$

Estimateurs empiriques des moments du bruit

Estimateur de θ Estimateur des moindres carrés pour θ

$$\hat{\theta}_n = \begin{pmatrix} \hat{a}_n \\ \hat{b}_n \\ \hat{c}_n \\ \hat{d}_n \end{pmatrix} = \mathbf{s}_{n-1}^{-1} \sum_{k \in \mathbb{T}_{n-1}} \begin{pmatrix} \delta_{2k} X_{2k} \\ \delta_{2k} X_k X_{2k} \\ \delta_{2k+1} X_{2k+1} \\ \delta_{2k+1} X_k X_{2k+1} \end{pmatrix}$$

avec

$$\mathbf{s}_n = \begin{pmatrix} \mathbf{s}_n^0 & 0 \\ 0 & \mathbf{s}_n^1 \end{pmatrix}$$

$$\mathbf{s}_n^0 = \sum_{k \in \mathbb{T}_n} \delta_{2k} \begin{pmatrix} 1 & X_k \\ X_k & X_k^2 \end{pmatrix} \quad \mathbf{s}_n^1 = \sum_{k \in \mathbb{T}_n} \delta_{2k+1} \begin{pmatrix} 1 & X_k \\ X_k & X_k^2 \end{pmatrix}$$

Convergence avec vitesse

Théorème

$$\mathbb{1}_{\{|G_n^*|>0\}} \|\hat{\theta}_n - \theta\|^2 = \mathbb{1}_{\{|G_n^*|>0\}} \mathcal{O}\left(\frac{\log |\mathbb{T}_{n-1}^*|}{|\mathbb{T}_{n-1}^*|}\right)$$

Preuve: méthode **martingale**

- ▶ identifier une martingale (vectorielle) pour la filtration des **générations** (augmentée de tout le processus d'observation)
- ▶ calculer la limite du crochet
- ▶ théorème de vitesse de convergence pour une martingale sur un arbre binaire de **Galton-Watson**

Martingale principale

$\widehat{\theta}_n - \theta = \mathbf{S}_{n-1}^{-1} \mathbf{M}_n$, avec (\mathbf{M}_n) martingale pour la filtration des générations et observations

$$\mathbf{M}_n = \sum_{k \in \mathbb{T}_{n-1}} \begin{pmatrix} \delta_{2k} \epsilon_{2k} \\ \delta_{2k} X_k \epsilon_{2k} \\ \delta_{2k+1} \epsilon_{2k+1} \\ \delta_{2k+1} X_k \epsilon_{2k+1} \end{pmatrix}$$

$(\mathbf{M}_n)_{n \geq 1}$ de carré intégrable et de crochet $\langle \mathbf{M} \rangle_n = \mathbf{\Gamma}_{n-1}$

$$\mathbf{\Gamma}_n = \begin{pmatrix} \sigma^2 \mathbf{S}_n^0 & \rho \mathbf{S}_n^{0,1} \\ \rho \mathbf{S}_n^{0,1} & \sigma^2 \mathbf{S}_n^1 \end{pmatrix} \quad \text{and} \quad \mathbf{S}_n^{0,1} = \sum_{k \in \mathbb{T}_n} \delta_{2k} \delta_{2k+1} \begin{pmatrix} 1 & X_k \\ X_k & X_k^2 \end{pmatrix}$$

Convergence du crochet

Lois des grands nombres pour les observations (δ_k) , le bruit $(\delta_k \epsilon_k)$,
le BAR $(\delta_{2k+i} X_k^q)$

- ▶ martingales **scalaires** pour différentes filtrations
- ▶ forme spécifique de l'**auto-régression**
- ▶ hypothèse $\max\{|b|, |d|\} < 1$

Théorème central limite

Théorème

Conditionnellement à la non extinction

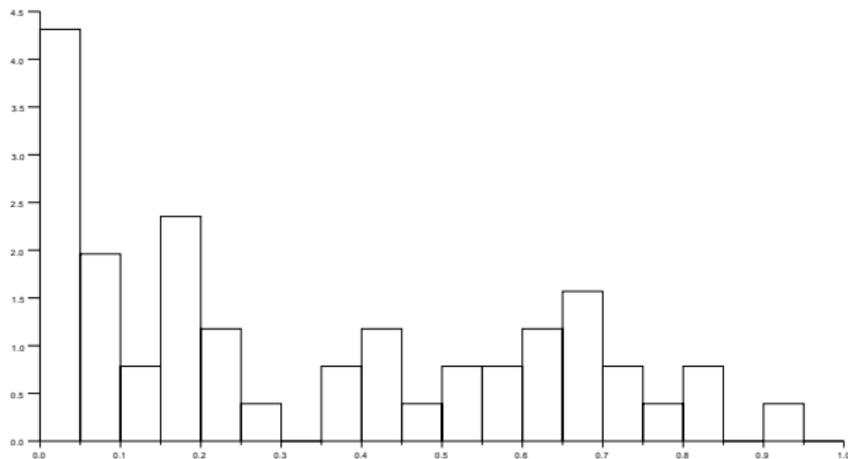
$$\sqrt{|\mathbb{T}_{n-1}^*|}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathbf{S}^{-1} \mathbf{\Gamma} \mathbf{S}^{-1})$$

Deux difficultés

- ▶ normalisation $|\mathbb{T}_{n-1}^*|$ aléatoire
- ▶ résultat conditionné à la non extinction : valable sur l'ensemble de non extinction $\bar{\mathcal{E}} = \cap \{|\mathbb{G}_n^*| > 0\}$ muni de la probabilité $\mathbb{P}_{\bar{\mathcal{E}}}(\cdot) = \mathbb{P}(\cdot \cap \bar{\mathcal{E}}) / \mathbb{P}(\bar{\mathcal{E}})$

Tests de symétrie : données Escherichia coli

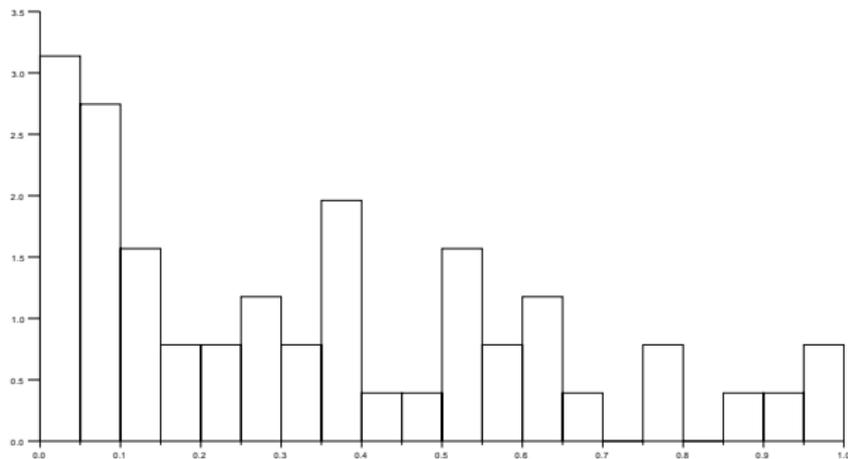
p-valeurs pour les 51 généalogies comportant 8 ou 9 générations



Test $(a, b) = (c, d)$

Tests de symétrie : données Escherichia coli

p-valeurs pour les 51 généalogies comportant 8 ou 9 générations



$$\text{Test } a/(1 - b) = c/(1 - d)$$

Nouveau modèle

Simulations \implies faible puissance des test pour 8 ou 9 générations

Modèle d'estimation multi-arbres

- ▶ utiliser **plusieurs** généalogies (en nombre **fixé**) pour l'inférence
- ▶ les généalogies sont des tirages **iid** du modèle BAR partiellement observé avec les **mêmes paramètres**
- ▶ **union** des ensembles de non-extinction
- ▶ **nouvel** estimateur
- ▶ nouvelles **preuves** de convergence avec les mêmes idées
- ▶ inférence et tests de symétrie sur le **Galton Watson**

Estimateur multi-arbres

Estimateur des moindres carrés pour θ

$$\hat{\theta}_n = \mathbf{S}_{n-1}^{-1} \sum_{j=1}^m \sum_{k \in \mathbb{T}_{n-1}} \begin{pmatrix} \delta_{j,2k} X_{j,2k} \\ \delta_{j,2k} X_{j,k} X_{j,2k} \\ \delta_{j,2k+1} X_{j,2k+1} \\ \delta_{j,2k+1} X_{j,k} X_{j,2k+1} \end{pmatrix}$$

avec

$$\mathbf{S}_n = \begin{pmatrix} \mathbf{S}_n^0 & 0 \\ 0 & \mathbf{S}_n^1 \end{pmatrix}$$

$$\mathbf{S}_n^i = \sum_{j=1}^m \sum_{k \in \mathbb{T}_n} \delta_{j,2k+i} \begin{pmatrix} 1 & X_{j,k} \\ X_{j,k} & X_{j,k}^2 \end{pmatrix}$$

Analyse multi-arbres des données E. coli : BAR

Estimation de $\theta \implies$ hypothèse $\max\{|b|, |d|\} < 1$ vraie

a	0.0203 [0.0197; 0.0210]	c	0.0195 [0.0188; 0.0201]
b	0.4615 [0.4437; 0.4792]	d	0.4782 [0.4605; 0.4959]

Estimation des moments du bruit

σ^2	$1.81 \cdot 10^{-5}$ [$1.12 \cdot 10^{-5}$; $2.50 \cdot 10^{-5}$]
ρ	$0.48 \cdot 10^{-5}$ [$0.44 \cdot 10^{-5}$; $0.52 \cdot 10^{-5}$]

Tests : hypothèse $(a, b) = (c, d)$ rejetée (p-valeur = 10^{-5}),
 hypothèse $a/(1 - b) = c/(1 - d)$ rejetée (p-valeur = $2 \cdot 10^{-3}$)

Analyse multi-arbres des données E. coli : Galton-Watson

Estimation des lois de reproduction

$p^{(0)}(0, 0)$	0.35579 [0.35574; 0.35583]	$p^{(1)}(0, 0)$	0.35611 [0.35606; 0.35616]
$p^{(0)}(1, 0)$	0.03621 [0.03620; 0.03622]	$p^{(1)}(1, 0)$	0.04707 [0.04706; 0.04708]
$p^{(0)}(0, 1)$	0.04740 [0.04739; 0.04741]	$p^{(1)}(0, 1)$	0.03755 [0.03754; 0.03756]
$p^{(0)}(1, 1)$	0.56060 [0.56055; 0.56065]	$p^{(1)}(1, 1)$	0.55928 [0.55923; 0.55933]

Estimation de π : 1.204 [1.191; 1.217] \implies hypothèse $\pi > 1$ vraie

Tests : hypothèse d'égalité des moyennes des deux lois non rejetée
 (p-valeur= 0.9), hypothèse d'égalité des deux vecteurs rejetée
 (p-valeur= $2 \cdot 10^{-5}$)

Plan de l'exposé

Introduction

BAR avec données manquantes

BAR à coefficients aléatoires

Modèle

Lois des grands nombres

Conclusion

Modèle à coefficients aléatoires

$$\begin{cases} X_{2k} &= (a + \varepsilon_{2k}) + (b + \eta_{2k}) X_k \\ X_{2k+1} &= (c + \varepsilon_{2k+1}) + (d + \eta_{2k+1}) X_k \end{cases}$$

Hypothèses

- ▶ $(\varepsilon_{2k}, \eta_{2k}, \varepsilon_{2k+1}, \eta_{2k+1})$ iid
- ▶ moments d'ordre 32
- ▶ données manquantes Galton Watson simple surcritique

Estimateurs

- ▶ Estimateur des moindres carrés pour θ : même formule
- ▶ Estimateurs des moindres carrés modifiés pour les moments du bruit minimisent

$$\frac{1}{2} \sum_{\ell=1}^{n-1} \sum_{k \in \mathbb{G}_\ell} (\hat{\epsilon}_{2k}^2 - \mathbb{E}[\epsilon_{2k}^2 | \mathcal{F}_\ell^O])^2 + (\hat{\epsilon}_{2k+1}^2 - \mathbb{E}[\epsilon_{2k+1}^2 | \mathcal{F}_\ell^O])^2$$

$$\frac{1}{2} \sum_{\ell=1}^{n-1} \sum_{k \in \mathbb{G}_\ell} (\hat{\epsilon}_{2k} \hat{\epsilon}_{2k+1} - \mathbb{E}[\epsilon_{2k} \epsilon_{2k+1} | \mathcal{F}_\ell^O])^2$$

avec (\mathcal{F}_n^O) filtration des générations et observations et

$$\begin{cases} \epsilon_{2k} &= \delta_{2k}(\varepsilon_{2k} + \eta_{2k} X_k), & \begin{cases} \hat{\epsilon}_{2k} &= \delta_{2k}(X_{2k} - \hat{a}_n - \hat{b}_n X_k), \\ \hat{\epsilon}_{2k+1} &= \delta_{2k}(X_{2k+1} - \hat{c}_n - \hat{d}_n X_k). \end{cases} \\ \epsilon_{2k+1} &= \delta_{2k+1}(\varepsilon_{2k+1} + \eta_{2k+1} X_k), \end{cases}$$

Convergence

Vitesse de convergence

$$\mathbb{1}_{\{|G_n^*|>0\}} \|\hat{\theta}_n - \theta\|^2 = \mathbb{1}_{\{|G_n^*|>0\}} \mathcal{O}\left(\frac{\log |\mathbb{T}_{n-1}^*|}{|\mathbb{T}_{n-1}^*|}\right)$$

Théorème central limite

Conditionnellement à la non extinction

$$\sqrt{|\mathbb{T}_{n-1}^*|}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathbf{S}^{-1} \mathbf{\Gamma} \mathbf{S}^{-1})$$

- ▶ identifier une martingale (vectorielle) pour la filtration des **générations** et observations
- ▶ **calculer la limite du crochet**
- ▶ théorème de vitesse de convergence pour une martingale sur un arbre binaire de **Galton-Watson**

Martingale principale

$\widehat{\theta}_n - \theta = \mathbf{S}_{n-1}^{-1} \mathbf{M}_n$, avec (\mathbf{M}_n) martingale pour la filtration des générations et observations

$$\mathbf{M}_n = \sum_{k \in \mathbb{T}_{n-1}} \begin{pmatrix} \delta_{2k} \epsilon_{2k} \\ \delta_{2k} X_k \epsilon_{2k} \\ \delta_{2k+1} \epsilon_{2k+1} \\ \delta_{2k+1} X_k \epsilon_{2k+1} \end{pmatrix}$$

$$\begin{cases} \epsilon_{2k} &= \delta_{2k} (\epsilon_{2k} + \eta_{2k} X_k), \\ \epsilon_{2k+1} &= \delta_{2k+1} (\epsilon_{2k+1} + \eta_{2k+1} X_k), \end{cases}$$

$(\mathbf{M}_n)_{n \geq 1}$ de carré intégrable et de crochet $\langle \mathbf{M} \rangle_n = \Gamma_{n-1}$ faisant intervenir des termes en $\sum_{k \in \mathbb{T}_{n-1}} \delta_{2k+i} X_k^q$, $0 \leq q \leq 4$

Convergence du crochet

On ne veut pas imposer

$$\max\{|b + \eta_2|, |d + \eta_3|\} < 1$$

⇒ plus de majoration qui gomme l'asymétrie
impossibilité d'utiliser la méthode martingale directe

Convergence du crochet

On ne veut pas imposer

$$\max\{|b + \eta_2|, |d + \eta_3|\} < 1$$

⇒ plus de majoration qui gomme l'asymétrie
impossibilité d'utiliser la méthode martingale directe

⇒ lois des grands nombres par la méthode chaîne de Markov
bifurquante

Chaîne de Markov bifurquante

Chaîne bifurquante sur $\mathbb{R} \cup \partial$

$$X_k^* = X_k \mathbb{1}_{\{\delta_k=1\}} + \partial \mathbb{1}_{\{\delta_k=0\}}$$

Noyau markovien sur $(\mathbb{R} \cup \partial)^3$: $Pf(\partial) = f(\partial, \partial, \partial)$ et

$$\begin{aligned} Pf(x) &= p(1, 1) \mathbb{E} [f(x, (b + \eta_2)x + a + \varepsilon_2, (d + \eta_3)x + c + \varepsilon_3)] \\ &\quad + p(1, 0) \mathbb{E} [f(x, (b + \eta_2)x + a + \varepsilon_2, \partial)] \\ &\quad + p(0, 1) \mathbb{E} [f(x, \partial, (d + \eta_3)x + c + \varepsilon_3)] \\ &\quad + p(0, 0) f(x, \partial, \partial) \end{aligned}$$

Noyaux sous-markoviens sur \mathbb{R}

$$P_0(x, A) = (p(1, 1) + p(1, 0)) \mathbb{E} [\mathbb{1}_A((b + \eta_2)x + (a + \varepsilon_2))]$$

Chaîne de Markov bifurquante

Chaîne bifurquante sur $\mathbb{R} \cup \partial$

$$X_k^* = X_k \mathbb{1}_{\{\delta_k=1\}} + \partial \mathbb{1}_{\{\delta_k=0\}}$$

Noyau markovien sur $(\mathbb{R} \cup \partial)^3$: $Pf(\partial) = f(\partial, \partial, \partial)$ et

$$\begin{aligned} Pf(x) &= p(1, 1) \mathbb{E} [f(x, (b + \eta_2)x + a + \varepsilon_2, (d + \eta_3)x + c + \varepsilon_3)] \\ &\quad + p(1, 0) \mathbb{E} [f(x, (b + \eta_2)x + a + \varepsilon_2, \partial)] \\ &\quad + p(0, 1) \mathbb{E} [f(x, \partial, (d + \eta_3)x + c + \varepsilon_3)] \\ &\quad + p(0, 0) f(x, \partial, \partial) \end{aligned}$$

Noyaux sous-markoviens sur \mathbb{R}

$$P_1(x, A) = (p(1, 1) + p(0, 1)) \mathbb{E} [\mathbb{1}_A((d + \eta_3)x + (c + \varepsilon_3))]$$

Chaîne induite

(A_n, B_n) iid de loi $(a + \epsilon_2, b + \eta_2)\mathbb{1}_{\{\zeta=1\}} + (c + \epsilon_3, d + \eta_3)\mathbb{1}_{\{\zeta=0\}}$,
 $\zeta \sim \text{Bernoulli}(1/2)$

$$\begin{cases} Y_0 &= X_1, \\ Y_{n+1} &= A_{n+1} + B_{n+1} Y_n \end{cases}$$

- Noyau $Q = (P_0 + P_1)/m$ avec m moyenne de la loi de reproduction
- Tirage aléatoire dans une génération : U uniforme sur \mathbb{G}_n

$$\mathbb{E}[f(Y_n)] = \mathbb{E}[f(X_U) | U \in \mathbb{T}^*]$$

- Lois des grands nombres : ν loi de X_1

$$\left\| \frac{1}{m^n} \sum_{k \in \mathbb{G}_n} f(X_k) \right\|_{L^2}^2 = \frac{\nu Q^n f^2}{m^n} + \frac{2}{m^2} \sum_{\ell=0}^{n-1} \frac{1}{m^\ell} \nu Q^\ell P(Q^{n-\ell-1} f \otimes Q^{n-\ell-1} f)$$

Ergodicité de la chaîne induite

- ▶ loi invariante $\mu \sim \sum B_1 \cdots B_{n-1} A_n$
- ▶ ergodicité **géométrique** pour x^q dès que

$$\mathbb{E}[|B_1|^q] = \frac{\rho(1,0) + \rho(1,1)}{m} \mathbb{E}[|b + \eta_2|^q] + \frac{\rho(0,1) + \rho(1,1)}{m} \mathbb{E}[|d + \eta_3|^q] < 1$$

remplace l'hypothèse $\max\{|b|, |d|\} < 1$

- ▶ loi des grands nombres pour X_k^q requiert moments d'ordre $4q$
- ▶ convergence du crochet
- ▶ vitesse de convergence via méthode martingale

Plan de l'exposé

Introduction

BAR avec données manquantes

BAR à coefficients aléatoires

Conclusion

Comparatif méthode chaîne de Markov et méthode martingale

	Martingale	Chaîne de Markov
bruit	différence de martingale moments d'ordre q	iid moments d'ordre $4q$
b et d	$\max < 1$	moyenne pondérée < 1
observations	Galton-Watson à deux types	Galton-Watson simple deux types?

Références

- [Cowan & Staudte 1986] COWAN AND STAUDTE The bifurcating autoregressive model in cell lineage studies. *Biometrics* (1986).
- [Guyon 2007] GUYON Limit theorems for bifurcating Markov chains. Application to the detection of cellular aging. *Ann. Appl. Probab.* (2007)
- [Delmas & Marsalle 2010] DELMAS AND MARSALLE Detection of cellular aging in a Galton-Watson process. *Stoch. Process. and Appl.* (2010)
- [Stewart & al. 2005] STEWART, MADDEN, PAUL, AND TADDEI Aging and death in an organism that reproduces by morphologically symmetric division. *PLoS Biol.* (2005)
- [EJP 2009] BERCU, DE SAPORTA AND GÉGOUT-PETIT Asymptotic analysis for bifurcating autoregressive processes via a martingale approach. *Electron. J. Probab.* (2009)
- [EJS 2011] DE SAPORTA, GÉGOUT-PETIT AND MARSALLE Parameters estimation for asymmetric bifurcating autoregressive processes with missing data. *Electron. J. Statist.* (2011)
- [SPL 2012] DE SAPORTA, GÉGOUT-PETIT AND MARSALLE Symmetry tests for bifurcating autoregressive processes with missing data. *Statistics & Probability Letters* (2012)
- coefficients aléatoires DE SAPORTA, GÉGOUT-PETIT AND MARSALLE Random coefficients bifurcating autoregressive processes. Arxiv 1205.4840
- multi-arbres DE SAPORTA, GÉGOUT-PETIT AND MARSALLE Statistical study of asymmetry in cell lineage data. Arxiv 1205.3658

MERCI