

Estimation de paramètres avec le Calcul Bayésien Approché (ABC)

Maxime Lenormand

Journées Réseau Mexico, 22 Novembre 2012

NANTES Ifremer

Laboratoire d'Ingénierie pour les Systèmes Complexes
Travail en commun avec **Franck Jabot** et **Guillaume Deffuant**



Prototypical Policy Impacts on Multifunctional Activities in rural municipalities

A collaborative project under the
EU Seventh Framework Programme



Plan

- ▶ Introduction
- ▶ Le modèle SimVillages
- ▶ Approximate Bayesian Computation
- ▶ Adaptive Approximate Bayesian Computation
- ▶ Résultats
- ▶ Conclusion et perspectives

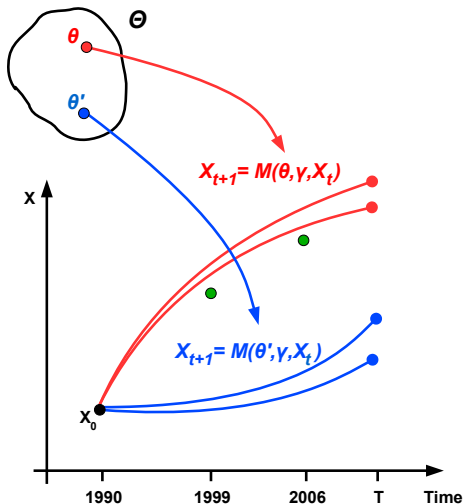
Modèle de microsimulation dynamique

- Modèle individu-centré
- Stochastique
- Espace des paramètres de grande dimension
- Coût computationnel élevé par simulation

Estimer la valeur des paramètres

- Calibrer le modèle
- Comprendre son comportement
- Réaliser une analyse d'incertitude
- Valider le modèle

Introduction



Données D ; On cherche à estimer un paramètre θ

Plan

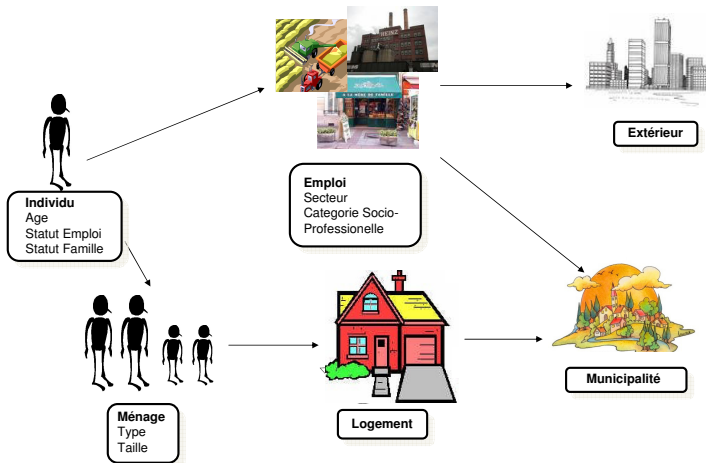
- ▶ Introduction
- ▶ **Le modèle SimVillages**
- ▶ Approximate Bayesian Computation
- ▶ Adaptive Approximate Bayesian Computation
- ▶ Résultats
- ▶ Conclusion et perspectives

Le modèle SimVillages

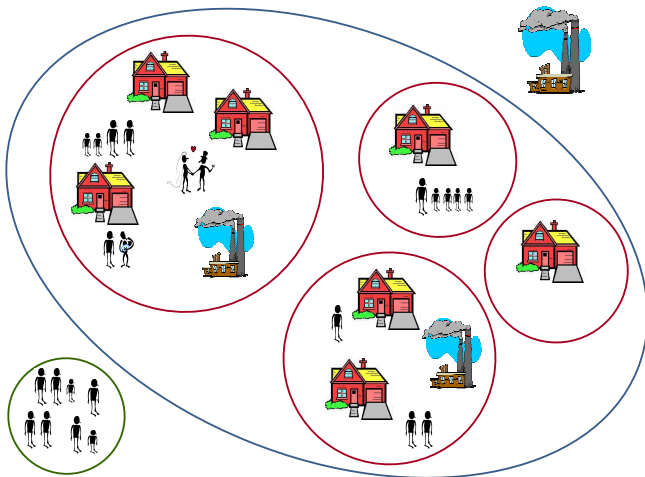
- Modèle d'évolution municipale de :
 - la population (taille et distribution des âges)
 - la distribution des activités et des logements
 - l'offre d'emplois liés à la présence de population
- en interaction avec les autres municipalités d'une région
- sous l'effet de scénario :
 - d'offre municipales d'emplois et de logements
 - donnant les contraintes liées à l'évolution du contexte

⇒ **Pour répondre à la question du devenir de la présence de population dans le rural**

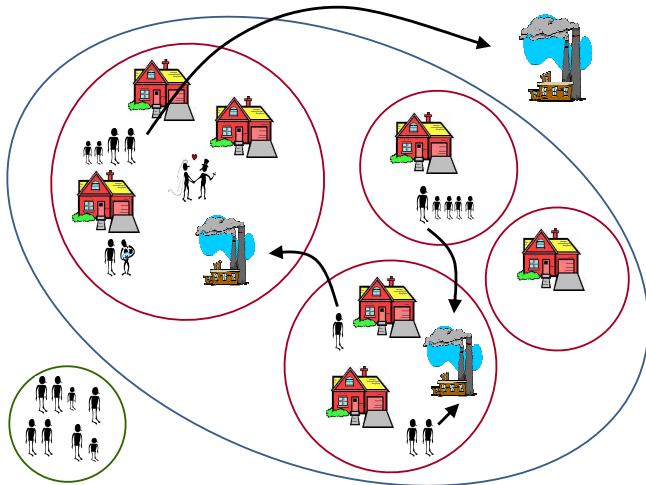
Le modèle SimVillages



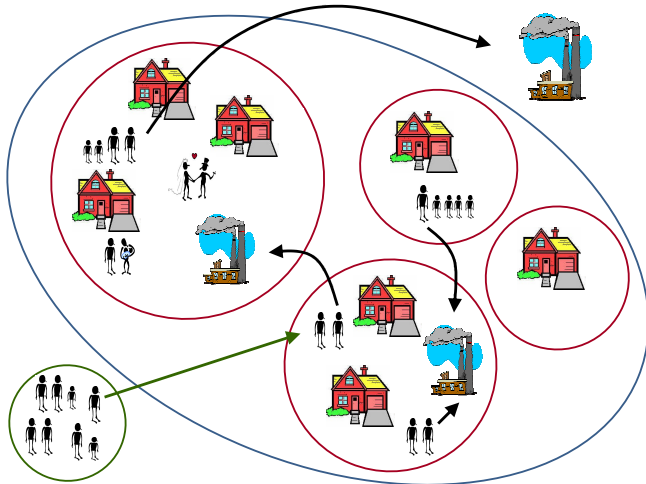
Le modèle SimVillages



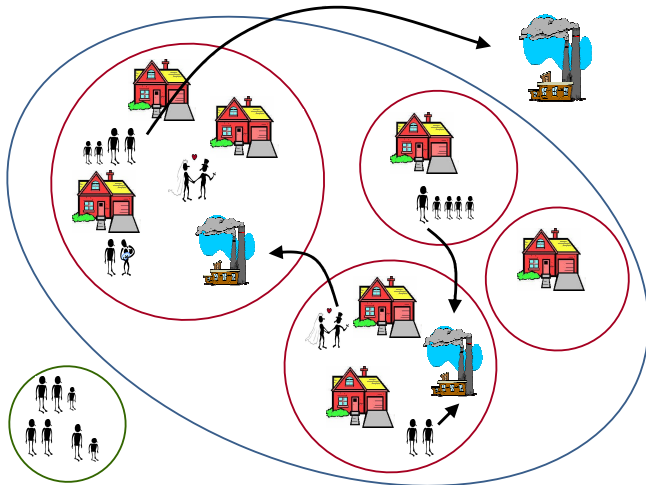
Le modèle SimVillages



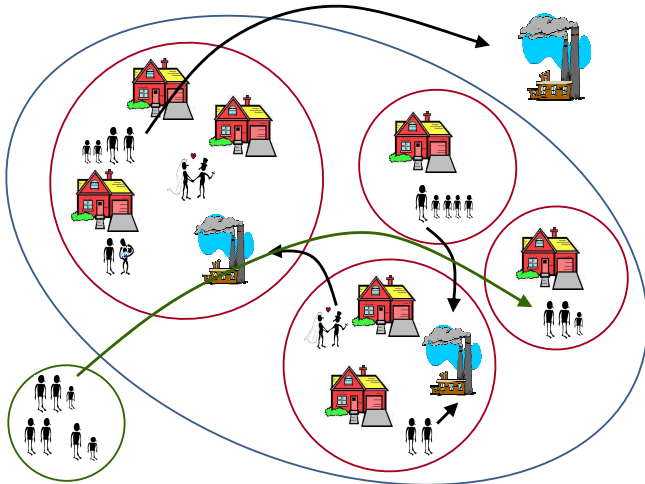
Le modèle SimVillages



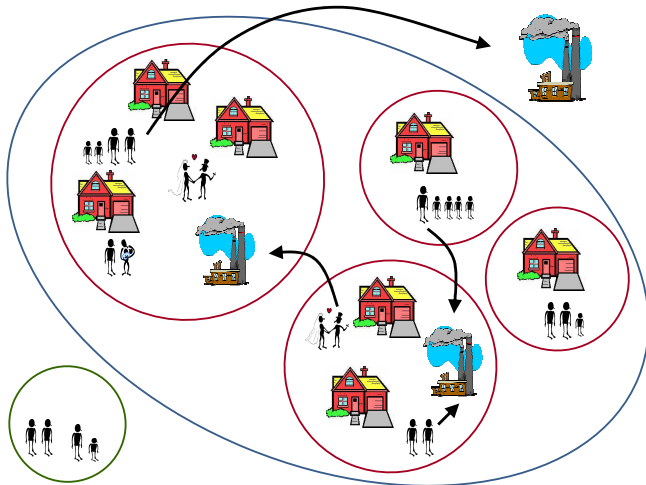
Le modèle SimVillages



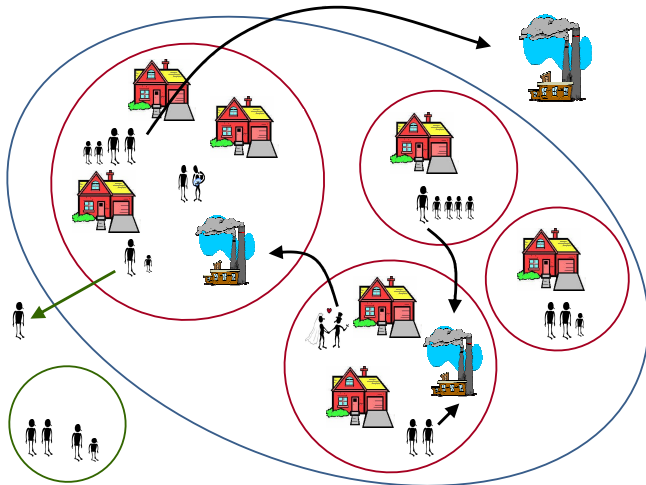
Le modèle SimVillages



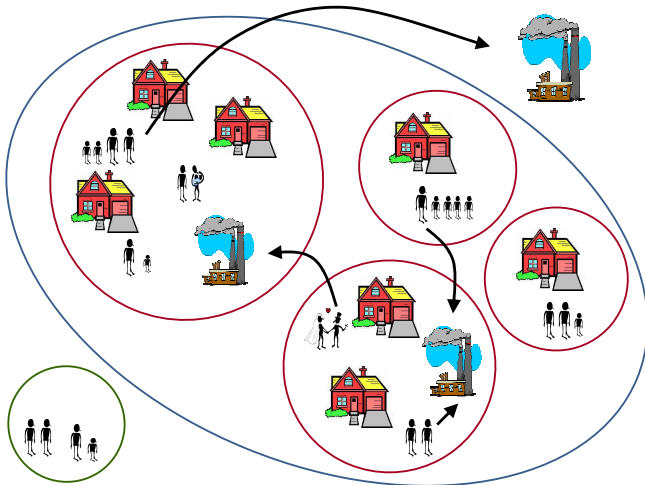
Le modèle SimVillages



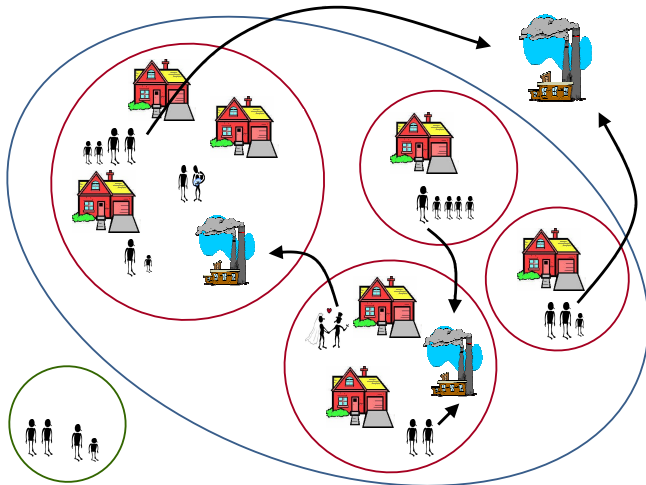
Le modèle SimVillages



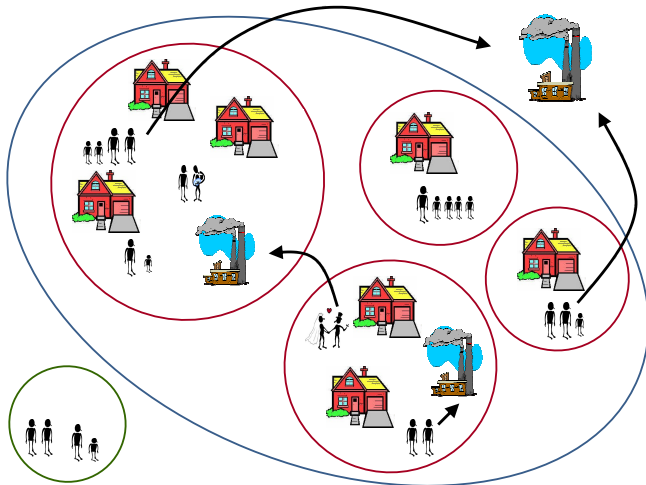
Le modèle SimVillages



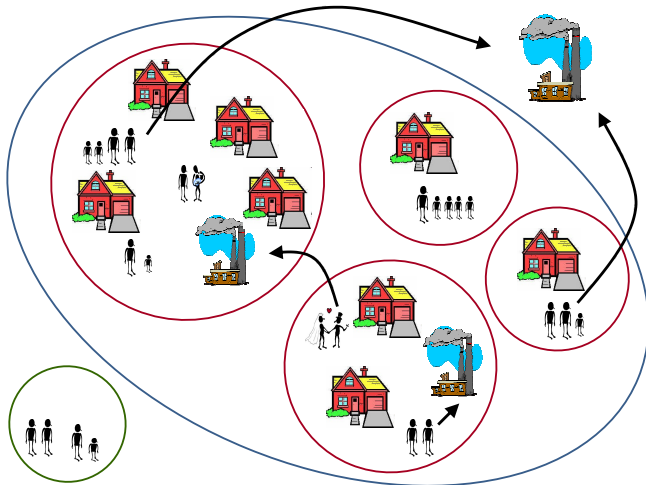
Le modèle SimVillages



Le modèle SimVillages



Le modèle SimVillages



Le modèle SimVillages

Données

Year	Population (thousands)	Variation of population (%)
1962	1 657	
1968	1 592	-4,10
1975	1 481	-7,50

	1990		1999		2006	
	Number	%	Number	%	Number	%
Farmers	60	10,8	72	15,2	44	8,3
Craftmen, storekeepers, business owners	124	22,3	50	10,5	65	12,3
Top executive managers, upper intellectual profession	24	4,3	32	6,8	15	2,8
Intermediary professions	104	18,7	58	12,2	98	18,6
Employees	144	25,9	169	35,7	186	35,2
Workers	100	18,0	93	19,6	120	22,7
Total	556	100	474	100	528	100

	Effectif	%
Number of households which moved there since less than 10 years	200	100
Less than 2 years	46	23
from 2 to 4 years	83	41,5
from 5 to 9 years	71	35,5
Number of households which moved there since 10 years or more	270	100
from 10 to 19 years	82	30,4
from 20 to 29 years	71	26,3
30 years and more	117	43,3
Total of households	470	

	2006	
	Number	%
Number of active people working in their municipality of residence	337	77,6
Number of active people working out of their municipality of residence	97	22,4
Total active employed people	434	100

Modèle

Génération



Calibration

Le modèle SimVillages

Caractéristiques du modèle

- 260 communes (Cantal)
- 8 paramètres (Ex: Nombre moyen d'enfants par femme)
- 6 indicateurs (Ex: Pyramide des âges)
- Année de départ : 1990
- Années de calibration : 1999 et 2006
- Temps moyen par simulation : 1 minute

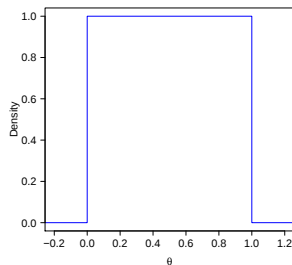
Comment estimer les paramètres ?

Plan

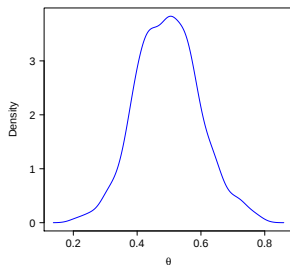
- ▶ Introduction
- ▶ Le modèle SimVillages
- ▶ **Approximate Bayesian Computation**
- ▶ Adaptive Approximate Bayesian Computation
- ▶ Résultats
- ▶ Conclusion et perspectives

Approximate Bayesian Computation

$$\mathbb{P}(\theta|D) = \frac{\mathbb{P}(D|\theta)\pi(\theta)}{\mathbb{P}(D)}$$



$\pi(\theta)$



$\mathbb{P}(\theta|D)$

Approximate Bayesian Computation

Pourquoi l'inférence Bayésienne ?

- Estimation d'une densité
- Théorie mathématique

Approximate Bayesian Computation

Pourquoi l'inférence Bayésienne ?

- Estimation d'une densité
- Théorie mathématique

Problèmes

- Calcul de la vraisemblance $L(\theta|D) = \mathbb{P}(D|\theta)$?
- Comment faire lorsque ce calcul est trop difficile ou trop coûteux ?

Approximate Bayesian Computation

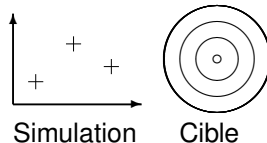
- 1 Simuler $\theta^* \sim \pi(\theta)$
- 2 Simuler $D' \sim f(\cdot|\theta^*)$
- 3 Si $D' = D$, accepter θ^* , sinon le rejeter
- 4 Répéter jusqu'à ce qu'un échantillon de la taille désirée soit obtenu



Distribution a priori
 $\pi(\theta)$



Distribution a posteriori
 $\mathbb{P}(D' = D|\theta)\pi(\theta)$



(Pritchard et al., 1999)

Approximate Bayesian Computation

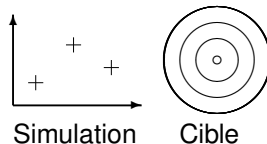
- 1 Simuler $\theta^* \sim \pi(\theta)$
- 2 Simuler $D' \sim f(\cdot|\theta^*)$
- 3 Si $D' = D$, accepter θ^* , sinon le rejeter
- 4 Répéter jusqu'à ce qu'un échantillon de la taille désirée soit obtenu



Distribution a priori
 $\pi(\theta)$



Distribution a posteriori
 $\mathbb{P}(D' = D|\theta)\pi(\theta)$



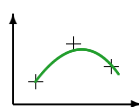
(Pritchard et al., 1999)

Approximate Bayesian Computation

- 1 Simuler $\theta^* \sim \pi(\theta)$
- 2 Simuler $D' \sim f(\cdot|\theta^*)$
- 3 Si $D' = D$, accepter θ^* , sinon le rejeter
- 4 Répéter jusqu'à ce qu'un échantillon de la taille désirée soit obtenu



Distribution a priori
 $\pi(\theta)$



Simulation



Cible

Distribution a posteriori
 $\mathbb{P}(D' = D|\theta)\pi(\theta)$

(Pritchard et al., 1999)

Approximate Bayesian Computation

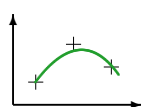
- 1 Simuler $\theta^* \sim \pi(\theta)$
- 2 Simuler $D' \sim f(\cdot|\theta^*)$
- 3 Si $D' = D$, accepter θ^* , sinon le rejeter
- 4 Répéter jusqu'à ce qu'un échantillon de la taille désirée soit obtenu



Distribution a priori
 $\pi(\theta)$



Distribution a posteriori
 $\mathbb{P}(D' = D|\theta)\pi(\theta)$



Simulation



Cible



(Pritchard et al., 1999)

Approximate Bayesian Computation

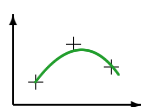
- 1 Simuler $\theta^* \sim \pi(\theta)$
- 2 Simuler $D' \sim f(\cdot|\theta^*)$
- 3 Si $D' = D$, accepter θ^* , sinon le rejeter
- 4 Répéter jusqu'à ce qu'un échantillon de la taille désirée soit obtenu



Distribution a priori
 $\pi(\theta)$



Distribution a posteriori
 $\mathbb{P}(D' = D|\theta)\pi(\theta)$



Simulation



Cible



(Pritchard et al., 1999)

Approximate Bayesian Computation

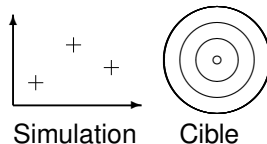
- 1 Simuler $\theta^* \sim \pi(\theta)$
- 2 Simuler $D' \sim f(\cdot|\theta^*)$
- 3 Si $D' = D$, accepter θ^* , sinon le rejeter
- 4 Répéter jusqu'à ce qu'un échantillon de la taille désirée soit obtenu



Distribution a priori
 $\pi(\theta)$



Distribution a posteriori
 $\mathbb{P}(D' = D|\theta)\pi(\theta)$



(Pritchard et al., 1999)

Approximate Bayesian Computation

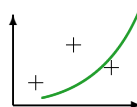
- 1 Simuler $\theta^* \sim \pi(\theta)$
- 2 Simuler $D' \sim f(\cdot|\theta^*)$
- 3 Si $D' = D$, accepter θ^* , sinon le rejeter
- 4 Répéter jusqu'à ce qu'un échantillon de la taille désirée soit obtenu



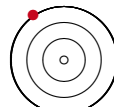
Distribution a priori
 $\pi(\theta)$



Distribution a posteriori
 $\mathbb{P}(D' = D|\theta)\pi(\theta)$



Simulation



Cible

(Pritchard et al., 1999)

Approximate Bayesian Computation

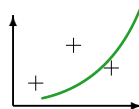
- 1 Simuler $\theta^* \sim \pi(\theta)$
- 2 Simuler $D' \sim f(\cdot|\theta^*)$
- 3 Si $D' = D$, accepter θ^* , sinon le rejeter
- 4 Répéter jusqu'à ce qu'un échantillon de la taille désirée soit obtenu



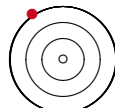
Distribution a priori
 $\pi(\theta)$



Distribution a posteriori
 $\mathbb{P}(D' = D|\theta)\pi(\theta)$



Simulation



Cible



(Pritchard et al., 1999)

Approximate Bayesian Computation

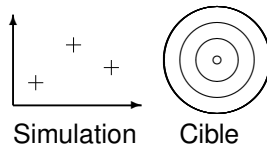
- 1 Simuler $\theta^* \sim \pi(\theta)$
- 2 Simuler $D' \sim f(\cdot|\theta^*)$
- 3 Si $D' = D$, accepter θ^* , sinon le rejeter
- 4 Répéter jusqu'à ce qu'un échantillon de la taille désirée soit obtenu



Distribution a priori
 $\pi(\theta)$



Distribution a posteriori
 $\mathbb{P}(D' = D|\theta)\pi(\theta)$



(Pritchard et al., 1999)

Approximate Bayesian Computation

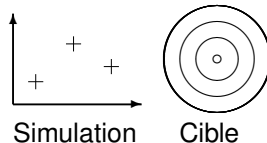
- 1 Simuler $\theta^* \sim \pi(\theta)$
- 2 Simuler $D' \sim f(\cdot|\theta^*)$
- 3 Si $D' = D$, accepter θ^* , sinon le rejeter
- 4 Répéter jusqu'à ce qu'un échantillon de la taille désirée soit obtenu



Distribution a priori
 $\pi(\theta)$



Distribution a posteriori
 $\mathbb{P}(D' = D|\theta)\pi(\theta)$



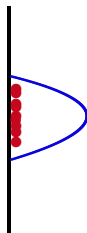
(Pritchard et al., 1999)

Approximate Bayesian Computation

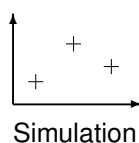
- 1 Simuler $\theta^* \sim \pi(\theta)$
- 2 Simuler $D' \sim f(\cdot|\theta^*)$
- 3 Si $D' = D$, accepter θ^* , sinon le rejeter
- 4 Répéter jusqu'à ce qu'un échantillon de la taille désirée soit obtenu



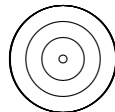
Distribution a priori
 $\pi(\theta)$



Distribution a posteriori
 $\mathbb{P}(D' = D|\theta)\pi(\theta)$



Simulation



Cible

(Pritchard et al., 1999)

Approximate Bayesian Computation

Deux approximations de $D' = D$

- $\rho(D', D) \leq \epsilon$
- $\rho(S(D'), S(D)) \leq \epsilon$

Approximate Bayesian Computation

Deux approximations de $D' = D$

- $\rho(D', D) \leq \epsilon$
- $\rho(S(D'), S(D)) \leq \epsilon$

$$\mathbb{P}(\rho(S(D'), S(D)) \leq \epsilon | \theta) \pi(\theta)$$

Approximate Bayesian Computation

Deux approximations de $D' = D$

- $\rho(D', D) \leq \epsilon$
- $\rho(S(D'), S(D)) \leq \epsilon$

$$\mathbb{P}(\rho(S(D'), S(D)) \leq \epsilon | \theta) \pi(\theta)$$

Les choix de l'ABC

- ϵ ?
- ρ ?
- S ?

Approximate Bayesian Computation

L'ABC dans la littérature (non exhaustif)

- Ajustement des paramètres (Beaumont et al. (2002))
- ABC-MCMC (Marjoram et al. (2003))
- ABC-PRC (Sisson et al. (2007))
- ABC-MCMC (Wegmann et al. (2009))
- Choix de S (Joyce et Marjoram (2008))
- ABC-PMC (Beaumont et al. (2009); Toni et al. (2009))
- Ajustement des paramètres (Blum et François (2010))
- ABC-RSMC (Drovandi et Pettitt (2011))
- ABC-SMC (DelMoral et al. (2012))
- Choix de S (Fearnhead et Prangle (2012))

Plan

- ▶ Introduction
- ▶ Le modèle SimVillages
- ▶ Approximate Bayesian Computation
- ▶ **Adaptive Approximate Bayesian Computation**
- ▶ Résultats
- ▶ Conclusion et perspectives

Adaptive Approximate Bayesian Computation



Prior

(Lenormand et al. (2012))

Adaptive Approximate Bayesian Computation

N particules
(LHS)

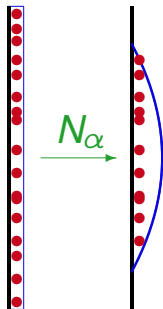


Prior

(Lenormand et al. (2012))

Adaptive Approximate Bayesian Computation

N particules
(LHS) ϵ_1



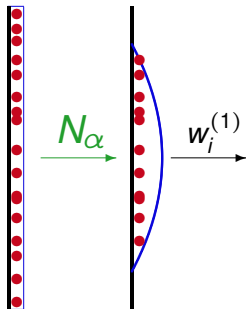
Prior

(Lenormand et al. (2012))

Adaptive Approximate Bayesian Computation

N particles
(LHS)

ϵ_1



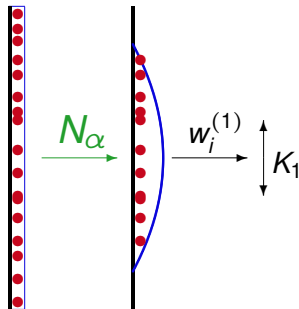
Prior

(Lenormand et al. (2012))

Adaptive Approximate Bayesian Computation

N particles
(LHS)

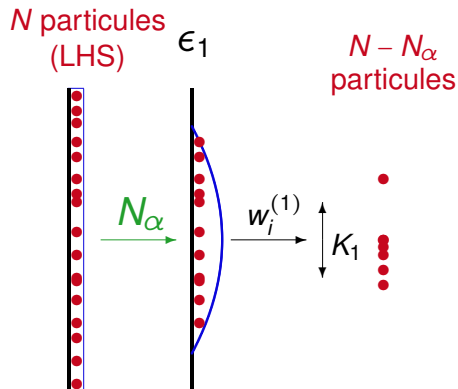
ϵ_1



Prior

(Lenormand et al. (2012))

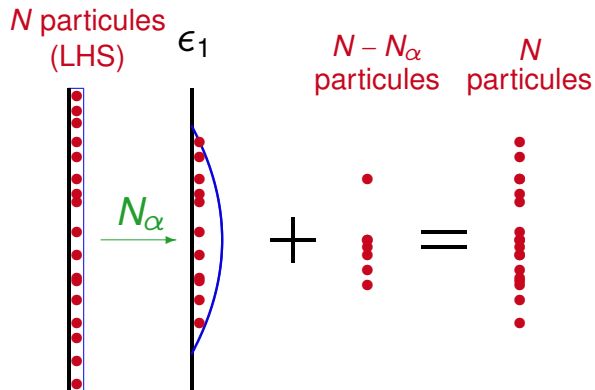
Adaptive Approximate Bayesian Computation



$$p_{acc} = \frac{\sum_{k=N_\alpha+1}^N \mathbb{1}_{\rho(S(D'), S(D)) \leq \epsilon_1}}{N - N_\alpha}$$

(Lenormand et al. (2012))

Adaptive Approximate Bayesian Computation

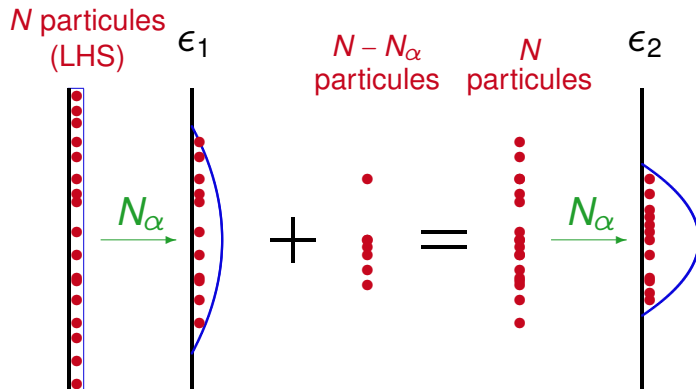


Prior

$$p_{acc} = \frac{\sum_{k=N_\alpha+1}^N \mathbb{1}_{\rho(S(D'), S(D)) \leq \epsilon_1}}{N - N_\alpha}$$

(Lenormand et al. (2012))

Adaptive Approximate Bayesian Computation



Prior

$$p_{acc} = \frac{\sum_{k=N_\alpha+1}^N \mathbb{1}_{\rho(S(D'), S(D)) \leq \epsilon_1}}{N - N_\alpha}$$

(Lenormand et al. (2012))

Plan

- ▶ Introduction
- ▶ Le modèle SimVillages
- ▶ Approximate Bayesian Computation
- ▶ Adaptive Approximate Bayesian Computation
- ▶ **Résultats**
- ▶ Conclusion et perspectives

Résultats

Paramètres

Paramètres	Description	Etendue
θ_1	Nombre d'enfants (pente)	$[0, 0.15]$
θ_2	Nombre d'enfants (intersection)	$[0.5, 3]$
θ_3	Indice de satisfaction du logement	$[0, 0.2]$
θ_4	Probabilité de se mettre en couple	$[0, 1]$
θ_5	Nombre d'essais pour se mettre en couple	$[0, 20]$
θ_6	Probabilité de divorcer	$[0, 0.05]$
θ_7	Seuil de proximité	$[0, 50]$
θ_8	Satisfaction de la demande de logement	$[0, 1]$

Résultats

Statistiques Résumantes

- 6 statistiques (3 en 1999 et 3 en 2006)
- Cible : $\sum_{k=1}^6 \frac{\rho_k(S_k, S'_k)}{\mu_k}$
- μ_k : distance moyenne obtenue avec les N premiers points

Statistique	Description	Distance
S_1	Pyramide des âges	$\mathbb{L}_1$
S_2	Nombre d'habitants par commune	$\mathbb{L}_1$
S_3	Solde Migratoire	$\mathbb{L}_1$

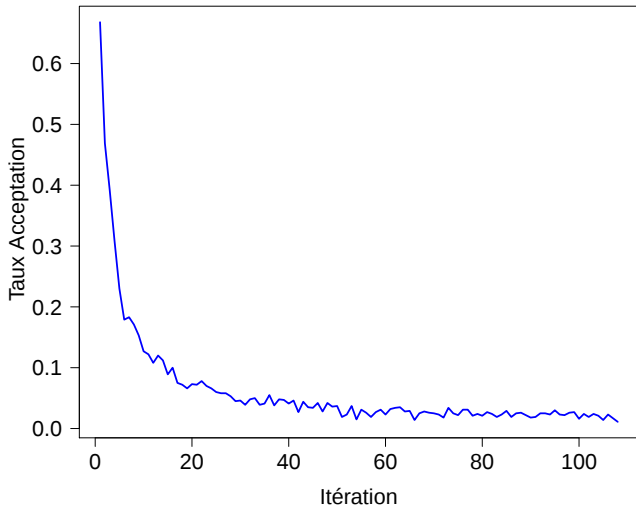
Résultats

Implémentation et exécution

- $N = 2000$
- $\alpha = 0.5$
- Language : JAVA
- Logiciel : OpenMOLE
- Cluster de 24 noeuds
- Durée : 4 jours

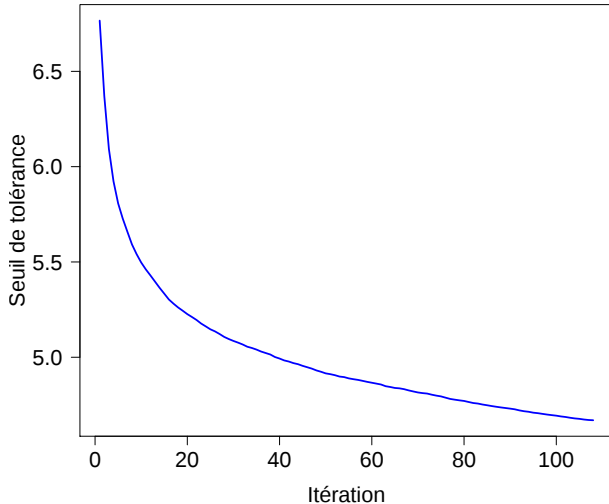
Résultats

Taux d'acceptation



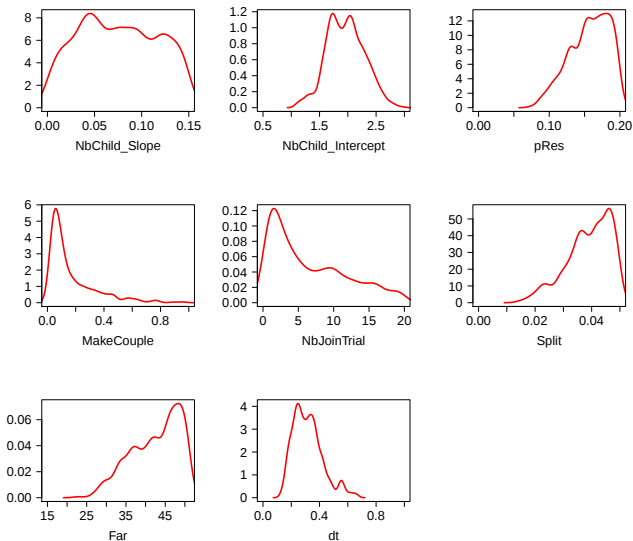
Résultats

Seuil de tolérance



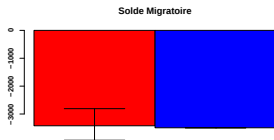
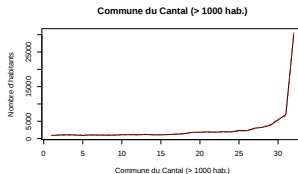
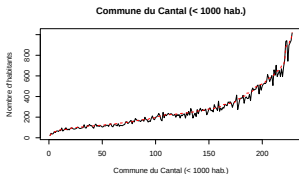
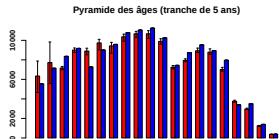
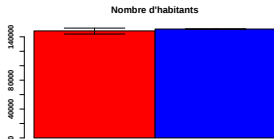
Résultats

Densité a posteriori



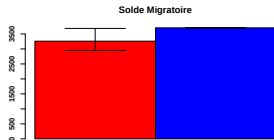
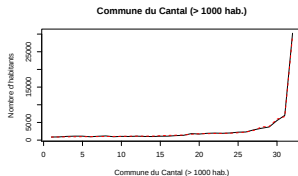
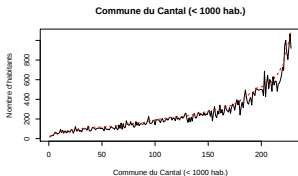
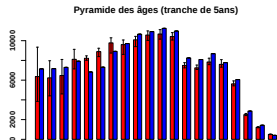
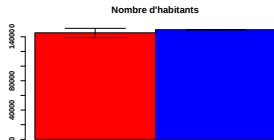
Résultats

Statistiques Résumantes 1999



Résultats

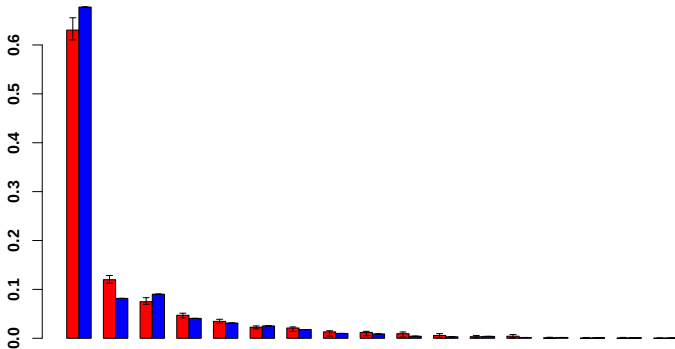
Statistiques Résumantes 2006



Résultats

Validation ?

Distribution de la distance de navettage



Plan

- ▶ Introduction
- ▶ Le modèle SimVillages
- ▶ Approximate Bayesian Computation
- ▶ Adaptive Approximate Bayesian Computation
- ▶ Résultats
- ▶ **Conclusion et perspectives**

Conclusion et perspectives

Conclusion

- Minimisation du nombre de simulations
- Détermination "en ligne" de ϵ
- Calibration de SimVillages avec l'ABC

Perspectives

- Choix des statistiques guidé par une analyse de sensibilité
- ABC versus Optimisation
- Validation/Prediction



Try it now!

```
> install.packages("EasyABC")
```

About EasyABC

The package EasyABC enables to perform efficient approximate bayesian computation (ABC) sampling schemes by launching a series of simulations of a computer code from the R platform, and to retrieve the simulation outputs in an appropriate format for post-processing treatments.

Package maintainer: [Nicolas Dumoulin](#)

Developers: [Franck Jabot](#), [Thierry Faure](#), [Nicolas Dumoulin](#)

News

- 11/20/2012 EasyABC 1.0 is now [available on CRAN](#).
- 11/13/2012 EasyABC is now hosted on [R-Forge](#).
- 11/06/2012 First commit of EasyABC

EasyABC

Algorithme

- Algorithme de réjection
 - Pritchard et al. (1999)
- Algorithme séquentiel
 - Beaumont et al. (2009)
 - Drovandi & Pettitt (2011)
 - Del Moral et al. (2012)
 - Lenormand et al. (2012)
- Algorithme MCMC
 - Marjoram et al. (2003)
 - Wegmann et al. (2009)

