



École nationale  
de la statistique  
et de l'administration  
économique

université  
PARIS-SACLAY

---

## Hidden Markov Models and Sequential Monte-Carlo methods

LE PELLETIER DE WOILLEMONT PIERRE, VINCENT-CUAZ CÉDRIC,  
ZAMOLOTCHIKOV PETR

---

## Sequential Monte-Carlo Methods for Approximate Bayesian Computation

Estimate Tuberculosis Transmission Parameters From Genotype Data

Nicolas Chopin



# Table des matières

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Modèle dynamique d'épidémiologie : Présentation des algorithmes</b>	<b>2</b>
2.1	ABC et SMC adaptatif . . . . .	2
2.1.1	(0) : Initialisation . . . . .	2
2.1.2	(1) : ESS bisection . . . . .	3
2.1.3	(2) : Resampling . . . . .	3
2.1.4	(3) : Adaptative MCMC . . . . .	3
2.2	ABC-PRC . . . . .	4
2.2.1	(0) : Initialisation . . . . .	4
2.2.2	Partial Rejection Control . . . . .	4
2.2.3	Resampling . . . . .	4
<b>3</b>	<b>Résultats</b>	<b>4</b>
3.1	Résultats SMC-ABC adaptatif . . . . .	4
3.2	Résultats ABC-PRC . . . . .	6
<b>4</b>	<b>Documentation des codes fournis</b>	<b>6</b>
4.1	Code de l'ABC-SMC adaptatif . . . . .	6
4.2	Code de l'ABC-PRC . . . . .	7
<b>5</b>	<b>Bibliographie</b>	<b>7</b>

# 1 Introduction

L'inférence bayésienne cherche à déterminer la densité à posteriori  $\pi(\theta|y)$ , issue de la densité à priori  $\pi(\theta)$  et de la vraisemblance des observations  $f(y|\theta)$ . Cette dernière parfois insolvable ou trop coûteuse à calculer, peut être approximée par ABC (Approximate Bayesian Computation) qui nécessite uniquement de pouvoir échantillonner des pseudo-observations  $X$  selon  $f(\cdot|\theta)$ . L'inférence est alors faite sur une densité à posteriori modifiée  $\pi_\epsilon(\theta, x|y)$ , où le niveau de tolérance  $\epsilon$  caractérise la distance maximale entre les vraies observations  $y$  et les pseudo-observations. Dans le cas où l'échantillonnage de pseudo-observations est trop complexe, les algorithmes SMC (Sequential Monte-Carlo) apportent une solution par une décomposition en une séquence de problèmes plus simples.

En ces termes, nous présentons deux approches pour estimer les paramètres de transmission de la Tuberculose. La première, développée dans l'article de Del Moral et al. (2011) [1], établit un algorithme SMC avec noyau MCMC de complexité linéaire dans le nombre de particules  $N$ , où les niveaux de tolérance  $\epsilon$  sont déterminés itérativement. La seconde, issue de l'article Sisson et al. (2009) [2], qui apporte une amélioration à l'algorithme ABC-MCMC avec une séquence de  $\epsilon$  fixée, en modifiant la règle d'acceptation des particules.

## 2 Modèle dynamique d'épidémiologie : Présentation des algorithmes

Le modèle de propagation de la Tuberculose, utilisé dans les articles précédemment cités, est dérivé de celui de Tanaka et al. (2006) [3] pour les données de Small et al. (1994) [4]. Ainsi la vraie observation comprend 326 génotypes ( $g$ ) pour 473 individus ( $n = \sum_{i=1}^g n_i$ )

### 2.1 ABC et SMC adaptatif

Avant de détailler le déroulement des étapes de l'algorithme, voici un aperçu global de celui-ci :

---

#### Algorithm 1 ABC-SMC Del Moral

---

```

1: procedure ADAPTATIVE_ABC_SMC
2:   (0) : Initialisation  $\rightarrow \{\theta_0^{(i)}\}_{i \in [N]}, \{x_{0,1:M}^{(i)}\}_{i \in [N]}, \epsilon_0, W_0$ 
3:   while  $\epsilon_t > \epsilon_{target}$  do
4:     (1) : ESS_bissection  $\{\theta_t^{(i)}\}_{i \in [N]}, \{x_{t,1:M}^{(i)}\}_{i \in [N]}, \epsilon_t, W_t \rightarrow \epsilon_{t+1}, W_{t+1}$ 
5:     if  $ESS(W_{t+1}) < ESS_{minimal}$  then
6:       (2) : Resampling  $\{\theta_t^{(i)}\}_{i \in [N]}, \{x_{t,1:M}^{(i)}\}_{i \in [N]}, W_{t+1} \rightarrow \text{new } \{\theta_t^{(i)}\}_{i \in [N]}, \{x_{t,1:M}^{(i)}\}_{i \in [N]}, W_{t+1}$ 
7:       (3) : Adaptative_MCMC  $\{\theta_t^{(i)}\}_{i \in [N]}, \{x_{t,1:M}^{(i)}\}_{i \in [N]}, \epsilon_{t+1}, W_{t+1} \rightarrow \{\theta_{t+1}^{(i)}\}_{i \in [N]}, \{x_{t+1,1:M}^{(i)}\}_{i \in [N]}$ 

```

---

#### 2.1.1 (0) : Initialisation

- On génère  $N$  paramètres  $\theta := (\phi, \tau, \xi)$  correspondant respectivement aux taux de naissance, de mort et de mutation, selon les a priori donnés :  $\phi \sim \Gamma(0.1, 1)$ ,  $\tau|\phi \sim U_{[0, \phi]}$ ,  $\xi \sim TN_{\mathbb{R}^+}(0.198, 0.6735^2)$ .
- Pour chaque  $\theta$ , on simule  $M$  pseudo-observations selon le processus suivant : La source est un unique individu (un génotype). Pour 10000 itérations, on suit le processus de propagation selon les probabilités issues de  $\theta$ . Puis, on fait un tirage sans remise de  $n$  individus parmi la population générée.
- On initialise la séquence de niveau de tolérance par  $\epsilon_0 = \infty$  et les poids des particules par  $W_0 = (W_0^{(i)} = \frac{1}{N})_{i \in [N]}$

En comparaison avec l'article de Tanaka [3], Le procédé de génération des pseudo-observations adopté dans l'article ne détaille pas les cas où la simulation aboutit à une population présentant moins de  $n$  individus ( $C1$ ), voire à une population éteinte ( $C2$ ). Ce phénomène est fréquent pour un paramètre  $\theta$  présentant une forte propension à muter. Dans le modèle dynamique proposé par Tanaka, où la simulation est faite jusqu'à obtenir une population de 10000 individus, il associe

une probabilité a posteriori nulle pour ces valeurs de  $\theta$  en cas d'extinction de la population. La condition sur le nombre minimal d'individus n'étant pas pertinente par cette approche, nous la considérons semblable à une extinction. Pour transposer la décision faite dans l'article de Tanaka, nous modifions la statistique  $\eta$  et la distance  $\rho$  pour que de telles pseudo-observations soient placées à une distance infinie de la vraie observation.

### 2.1.2 (1) : ESS bisection

Ainsi, comme introduit précédemment, la pertinence d'une pseudo-observation  $x$  est définie par sa distance  $\rho$  entre sa statistique  $\eta(x)$  et celle de l'observation réelle  $\eta^{true}$  :

$$\eta(x) = \begin{cases} \infty & \text{si } x \text{ vérifie (C1) ou (C2)} \\ (g, 1 - \sum_{i=1}^g (\frac{n_i}{n})^2) & \text{sinon} \end{cases}, \rho(\eta(x), \eta^{true}) = \begin{cases} \infty & \text{si } x \text{ vérifie (C1) ou (C2)} \\ \frac{1}{n} |\eta_1(x) - \eta_1^{true}| + |\eta_2(x) - \eta_2^{true}| & \text{sinon} \end{cases}$$

On cherche alors à déterminer le seuil de tolérance  $\epsilon_{t+1}$  qui vérifie l'équation :

$$ESS(W_{t+1}, \epsilon_{t+1}) = \alpha ESS(W_t, \epsilon_t)$$

En considérant un noyau de transition MCMC de densité invariante et l'approximation usuelle du noyau rétroactif associé, avant d'être renormalisés les poids sont actualisés selon la formule suivante :

$$W_{t+1}^{(i)} \propto W_t^{(i)} \frac{\sum_{k=1}^M \mathbf{1}_{A_{\epsilon_{t+1}, y}(X_{k,t}^{(i)})}}{\sum_{k=1}^M \mathbf{1}_{A_{\epsilon_t, y}(X_{k,t}^{(i)})}} \text{ où } \mathbf{1}_{A_{\epsilon, y}(z)} = \{z | \rho(\eta(z), \eta^{true}) \leq \epsilon\}$$

On applique alors un algorithme de bisection qui soulève plusieurs interrogations.

- L'ESS exprimée par  $ESS(W) = (\sum_{i=1}^N W^{(i)})^{-1}$  peut ne pas être définie si l'ensemble des particules sont rejetées i.e si  $\epsilon < \min_x \rho(x, y^{true})$ . On considère dans ce cas qu'elle vaut 0 pour continuer la recherche vers des  $\epsilon$  plus grand.
- L'ESS est une fonction constante par morceaux et croissante selon  $\epsilon$ . En effet, les poids  $W^{(i)}$  sont des probabilités caractérisant l'acceptation de la particule  $x_i$ . La diminution de  $\epsilon$  implique un rejet plus conséquent des particules et une augmentation de la variance des poids. Ainsi L'ESS approximant l'inverse de la variance des poids diminuera.
- Le choix des bornes pour approximer notre ESS cible est crucial à chaque itération. Empiriquement, la recherche d'un  $\epsilon$  pour approximer  $_{cible} n[\alpha ESS(W_t) - \delta, \alpha ESS(W_t) + \delta]$  nécessitait un trade-off entre le risque que l'algorithme n'aboutisse pas et des variations trop dégénérées de l'ESS. On a alors choisi de construire des bornes asymétriques en considérant les distances de l' $ESS_{cible}$  à son prédécesseur et son successeur. On introduit également de manière asymétrique, un terme de sélection asymétrique caractérisé par  $\delta$  pour forcer l'ESS vers des valeurs plus petites :  $ESS_{cible} \in [\alpha ESS(W_t)(1 - \delta(1 - \alpha)), \alpha ESS(W_t)(1 + \frac{\delta}{2}(\frac{1}{\alpha} - 1))]$

### 2.1.3 (2) : Resampling

Dans le cas où l'ESS des poids actualisés est inférieure à  $\frac{N}{2}$ , on construit un nouvel échantillon par tirage avec remise selon les poids  $W_{t+1}$ . Puis les poids sont réinitialisés à  $W_0$ .

### 2.1.4 (3) : Adaptive MCMC

Les particules vivantes sont déplacées selon un noyau de transition MCMC à densité invariante caractérisée par le seuil de tolérance  $\epsilon$ . En notant  $q_n$  la densité de la perturbation issue des marches aléatoires normales données, une nouvelle particule est acceptée selon la probabilité donnée par le ratio de Metropolis-Hastings :

$$\min \left\{ 1, \frac{\sum_{k=1}^M \mathbf{1}_{A_{\epsilon_{t+1}, y}(X_{k,t}^{(i)})} q_n(\theta^*, \theta) \pi(\theta^*)}{\sum_{k=1}^M \mathbf{1}_{A_{\epsilon_{t+1}, y}(X_{k,t+1}^{(i)})} q_n(\theta, \theta^*) \pi(\theta)} \right\}$$

## 2.2 ABC-PRC

L'algorithme PRC (partial rejection control) se présente de la manière suivante :

---

### Algorithm 2 ABC-PRC

---

linenosize=

```

1: procedure ABC_PRC( $(\epsilon_t)_{t \in [T]}$  donnée)
2:   (0) : Initialisation  $\rightarrow \{\theta_i^0\}_{i \in [N]}, W_0$ 
3:   for  $t$  in  $[1, T]$  do
4:     (1) : Partial_Rejection_Control  $\{\theta_i^{t-1}\}_{i \in [N]}, W_{t-1} \rightarrow \{\theta_i^t\}_{i \in [N]}, W_t$ 
5:     if  $ESS(W_t) < ESS_{minimal}$  then
6:       (2) : Resampling  $\{\theta_i^t\}_{i \in [N]}, W_t \rightarrow \text{new } \{\theta_i^t\}_{i \in [N]}, W_t$ 

```

---

Contrairement au premier algorithme où les niveaux de tolérance sont déterminés itérativement, ici la séquence doit être choisie préalablement. Celle-ci doit être suffisamment lisse pour que la distribution initiale des thetas converge graduellement vers la distribution caractérisée par le  $\epsilon$  ciblé.

### 2.2.1 (0) : Initialisation

- Ici la simulation des paramètres  $\{\theta_i^0\}_{i \in [N]}$  est sélective et itérative. Suivant les priors de l'algorithme précédent, on simule des pseudo-observations  $x$  pour un paramètre  $\theta$  donné jusqu'à que la condition  $\rho(x, \eta^{true}) < \epsilon_0$  soit vérifiée. Bien que l'auteur semble utiliser l'algorithme de génération de pseudo-observation de Tanaka[3], nous avons conservé celle de Del Moral [1] pour diminuer les temps de calculs. Les mêmes considérations sont faites quant à la distance et la statistique.

- Les poids des particules sont initialisés à  $W_0 = (W_0^{(i)} = \frac{1}{N})_{i \in [N]}$

### 2.2.2 Partial Rejection Control

- Pour chacune des particules, on commence par échantillonner un  $\theta$  parmi  $\theta_{t-1}^{(i)}$  selon les probabilités  $W_{t-1}^{(i)}$ . On perturbe celle-ci en une nouvelle particule  $\theta^*$  par le noyau de transition  $K_t$  suivant la loi  $N(0, \Sigma)$  où  $\Sigma = \begin{pmatrix} 0.5^2 & 0.225 & 0 \\ 0.225 & 0.5^2 & 1 \\ 0 & 0 & 0.015^2 \end{pmatrix}$ . On génère une pseudo-observation  $x^* \sim f(x|\theta^*)$ . Si celle-ci est à une distance supérieure à  $\epsilon_t$  de  $y^{true}$ , on recommence pour un nouveau  $\theta$ . Le noyau de transition diffère de celui utilisé dans le premier algorithme en considérant une corrélation entre le processus de naissance et de mort, et une perturbation du taux de mutation plus faible.
- Pour les  $N$  nouveaux paramètres  $\theta_t^{(i)}$ , avant d'être renormalisés, les poids sont actualisés suivant la formule suivante :

$$W_t^{(i)} = \frac{\pi(\theta_t^{(i)})}{\sum_{j=1}^N W_{t-1}(\theta_{t-1}^{(j)}) K_t(\theta_t^{(i)} | \theta_{t-1}^{(j)})}$$

### 2.2.3 Resampling

Le ré-échantillonnage est le même que pour l'algorithme précédent.

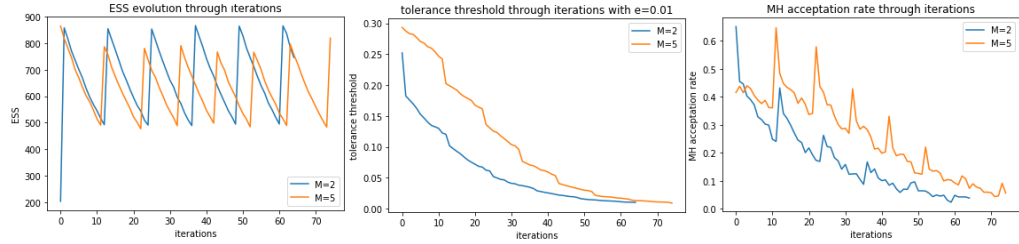
## 3 Résultats

### 3.1 Résultats SMC-ABC adaptatif

- Initialement, l'indication sur l'espérance du prior  $\phi \sim \Gamma(0.1, 1)$  semble être une erreur. Celle-ci menait à une distribution semblable pour le taux de transmission avec un rapport de 100. Cela induisait des taux de mutation supérieurs à ceux de naissance et de décès,

impliquant un grand nombre d'échecs pour la génération de pseudo-observations et un ESS initial autour des 200, nécessitant de commencer par un ré-échantillonnage pour que la bissection fonctionne. Nous avons inversé les coefficients du prior  $\phi \sim \Gamma(1, 0.1)$  pour avoir une espérance de 10 et une variance de 100, en adéquation avec la distribution du taux de transmission donné. Ceci réduisant grandement la propension à muter de la maladie et l'échec des générations. Nos résultats ont ainsi pu être comparables à ceux de Del Moral.

- Nous avons premièrement fait des simulations pour voir l'impact du paramètre  $M$  afin de déterminer l'impact de la discrimination apportée par la génération d'au moins  $n$  individus. Les simulations ont été faites pour  $N = 1000$ ,  $M = 2; 5$ ,  $\alpha = 0.95$ ,  $\epsilon = 0.01$  et  $\delta = 0.1$ . Tout d'abord, on peut observer les différents sauts faits à chaque ré-échantillonnage. D'après l'article lorsque  $M$  augmente en considérant les autres paramètres fixés, la convergence vers  $\epsilon$  est plus rapide. Nous observons ici l'inverse, ceci pouvant être expliqué par une distribution initiale moins précise. En revanche, le taux d'acceptation du noyau de Metropolis-Hastings est bien supérieur pour  $M = 5$



- Pour comparer nos résultats avec ceux de l'article, nous avons effectué 5 simulations pour  $N = 1000$ ,  $M = 15$ ,  $\alpha = 0.9$  et un  $\epsilon = 0.01$  plus faible pour des raisons de temps de calculs. Ci-dessous on peut observer l'évolution pour l'un d'entre eux. Les autres simulations étaient caractérisées par des évolutions similaires, et nécessitaient plus ou moins deux itérations pour converger vers le seuil de tolérance ciblé. Les résultats analogues présentés dans l'article semblent être moins influencés par les décrochages suite au resampling. Pour cet exemple, les queues de distributions du taux de transmission et du temps de doublement sont moins épaisses que pour le résultat présenté dans l'article avec un epsilon plus faible. En revanche, sur les 5 simulations des variances plus importantes sont logiquement observées.

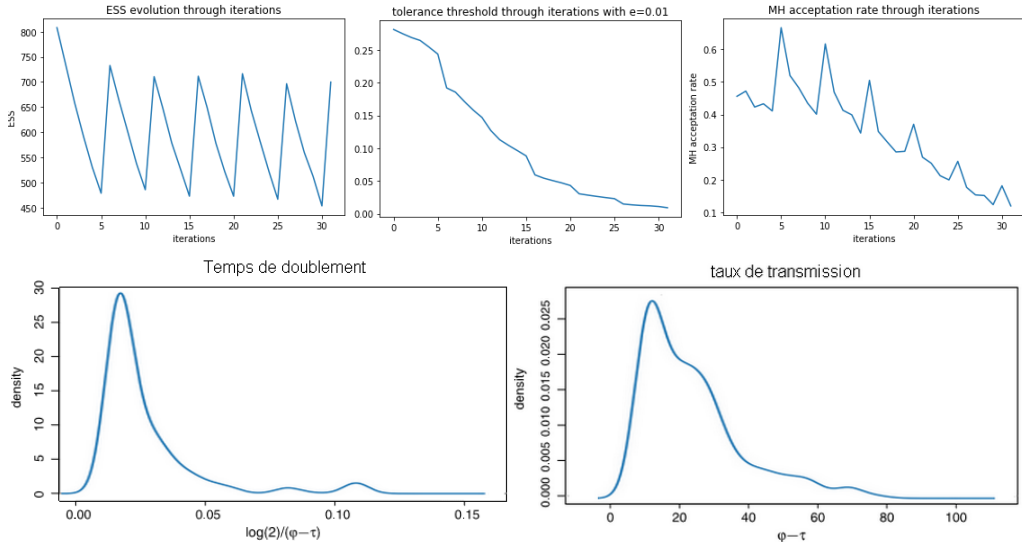
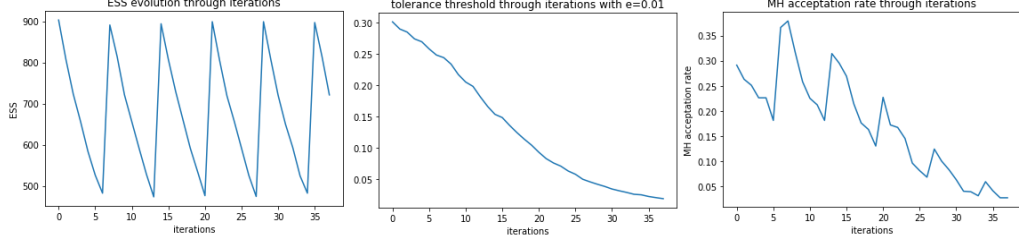


TABLE 1 – moyenne et variance des distributions à posteriori pour 5 expériences

	Moyenne	Variance
Naissance $\phi$	27.58(1.08)	18.60(0.25)
Décès $\tau$	0.92(0.13)	0.61(0.17)
Mutation $\xi$	0.22(0.05)	0.08(0.03)

- Afin de comparer avec l'algorithme de Sisson où  $M = 1$ , nous avons tester plusieurs fois des simulations avec  $N = 1000, M = 1, \alpha = 0.9, \epsilon = 0.01$ . Dans le meilleur cas, celles-ci échouaient vers  $\epsilon = 0.02$ . Nous n'avons pas eu le temps de faire de simulations pour  $N$  ou  $\alpha$  plus grand. Afin de comparer avec l'algorithme de Sisson, on observe que l'algorithme à converger à  $\epsilon \approx 0.025$  avec 22909 générations de pseudo-observation.



### 3.2 Résultats ABC-PRC

Nous avons lancé l'algorithme pour  $N = 1000, (\epsilon) = (2, 0.5, 0.1, 0.05, 0.025, 0.01)$  :

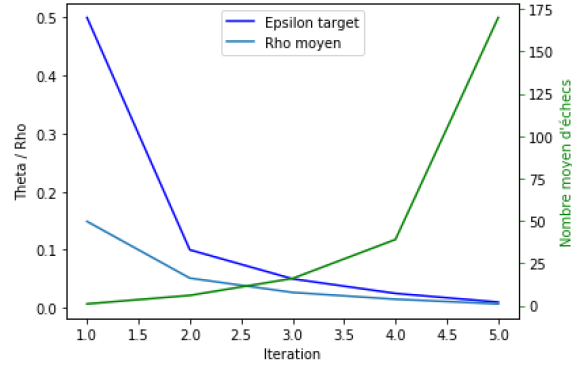


FIGURE 1 – Évolution de la distance moyenne des pseudo-observations pour l'ensemble des  $(\theta)$  sélectionnés et du nombre d'échecs moyen pour déterminer un  $\theta$

On considère à chaque nouveau seuil,  $Rho$  la distance moyenne des pseudo-observations et le nombre moyen d'échecs pour l'ensemble des paramètres  $\theta$ . Bien qu'une contrainte supplémentaire existe pour qu'une pseudo-observation soit acceptée, nos résultats sont cohérents avec ceux de l'article [2]. Pour cette simulation, environ 65000 générations de pseudo-observations sont faites pour arriver à  $\epsilon = 0.025$ .

## 4 Documentation des codes fournis

### 4.1 Code de l'ABC-SMC adaptatif

- **theta()** :simulation d'un paramètre  $\theta$
- **get\_N\_thetas(N)** : simulation de  $N$  paramètres
- **pseudo\_obs( $\theta$ )** : processus de génération des pseudo-observations pour un paramètre  $\theta$
- **M\_pseudo\_obs(M, $\theta$ )** :  $M$  pseudo-observations pour un  $\theta$
- **parallelthetas\_M\_pseudo\_obs(M, ( $\theta$ ))** : parallélisation pour générer  $M$  pseudo-observations pour chaque  $\theta$
- **compute\_eta(pseudo\_obs)** : calcul de la statistique  $\eta$
- **compute\_rho(eta1,eta2)** : calcul de la distance  $\rho$
- **compute\_ESS(W)** : calcul de l'ESS pour un vecteur de poids  $W$ .
- **update\_weights(previous\_W, particles, new\_ε, previous\_ε)** : actualisation des poids.
- **epsilon\_bisection(ESS\_ciblé, ESS\_min,  $\alpha$ ,  $\delta$ , previous\_W, previous\_particles, previous\_ε)** : algorithme de bisection construire la séquence de seuil adaptativement.
- **resampler\_on\_living\_particles(( $\theta$ ), particles,W)** : resampling si  $ESS < ESS_{min}$



- **random\_walk**( $\theta$ ) : perturbation de  $\theta$  selon la marche aléatoire.
- **rw\_density**( $\theta_1, \theta_2$ ) : densité de la perturbation.
- **prior\_density**( $\theta$ ) : densité du prior en  $\theta$
- **MH\_resampling\_theta**( $\theta, \text{particles}, \epsilon$ ) : algorithme de Metropolis-Hastings pour une particule vivante.
- **MH\_resampling**(( $\theta, \text{particles}, \mathbf{W}, \epsilon$ ) : algorithme MH pour l'ensemble des particules
- **delmoral\_SMCABC**( $\mathbf{N}, \mathbf{M}, \alpha, \delta, ESS_{min}, \epsilon$ ) : algorithme SMC-ABC adaptatif utilisant l'ensemble des fonctions précédentes.

## 4.2 Code de l'ABC-PRC

- Les fonctions **theta**, **get\_N\_thetas**, **pseudo\_obs**, **compute\_eta**, **compute\_rho**, **resampler\_on\_living\_particles** sont réutilisées.
- **pseudo\_obs\_sisson\_t1**( $\epsilon_0$ ) : méthode de générations des pseudo-observation particules à l'initialisation.
- **pertub\_theta\_star**( $\theta^*$ ) : perturbation du paramètre  $\theta^*$  selon le noyau de transition
- **Sisson\_ABCPRC**( $(\epsilon), N, ESS_{min}$ ) : Algorithme ABC-PRC

## 5 Bibliographie

- 1 Del Moral, P., Doucet, A., Jasra, A. (2012). An adaptive sequential Monte Carlo method for approximate Bayesian computation. *Statistics and Computing*, 22(5), 1009-1020.
- 2 Sisson, S., Fan, Y., Tanaka, M.M. : Correction : Sequential Monte Carlo without likelihoods. *Proc. Natl. Acad. Sci.* 106, 1760 (2009)
- 3 Tanaka, M. M., Francis, A. R., Luciani, F., Sisson, S. A. (2006). Using approximate Bayesian computation to estimate tuberculosis transmission parameters from genotype data. *Genetics*, 173(3), 1511-1520.
- 4 Small, P. M., Hopewell, P. C., Singh, S. P., Paz, A., Parsonnet, J., Ruston, D. C., ... Schoolnik, G. K. (1994). The Epidemiology of Tuberculosis in San Francisco—A Population Based Study Using Conventional and Molecular Methods. *New England Journal of Medicine*, 330(24), 1703-1709.