

ABC WITHIN GIBBS

Grégoire Clarté^{1†}, Kerrie Mengersen², Pierre Pudlo³, Christian Robert^{4‡}, Robin Ryder^{5†},
Julien Stoehr^{6†},

¹ *clarte@ceremade.dauphine.fr ;*

² *Department of Statistics, QUT, k.mengersen@qut.edu.au ;*

³ *Aix-Marseille Université, pierre.pudlo@univ-amu.fr ;*

⁴ *xian@ceremade.dauphine.fr ;*

⁵ *ryder@ceremade.dauphine.fr ;*

⁶ *stoehr@ceremade.dauphine.fr ;*

[†] *Université Paris Dauphine ;*

[‡] *Warwick University et Université Paris Dauphine.*

Résumé. Les méthodes ABC pâtissent des grandes dimensions. La grande dimension de l'espace des observations est traité par l'usage de statistiques résumées. Les grands espaces de paramètres, doivent être explorés plus efficacement, nous nous proposons ici d'adapter l'échantillonneur de Gibbs en utilisant des conditionnelles approchées par méthode ABC. La convergence de cet algorithme n'est pas évidente dans le cas général, on démontrera la convergence en variation totale par des méthodes de couplage. L'efficacité de cet échantillonneur est montrée sur plusieurs exemples.

Mots-clés. Méthodes bayésiennes, Algorithmes stochastiques, Statistique computationnelle.

Abstract. ABC methods suffers from high dimensions. Summary statistics can handle high dimension of the observation space. To deal with high dimensioned parameter space we have to explore it more efficiently ; here we propose to adapt Gibbs sampler by plugging ABC approximations of the conditionals. Convergence of this algorithm is not obvious, and we will show the existence of a limiting law with coupling techniques. The efficiency of this sampling scheme will be shown on several examples.

Keywords. Bayesian methods, stochastic algorithms, computational statistics

1 Algorithme

Nous proposons d'adapter l'échantillonneur de Gibbs aux modèles intractables. Supposons un modèle décrit par k paramètres a_1, \dots, a_k , on note x les observations. L'échantillonneur de Gibbs s'écrit :

- initialiser a_1^1, \dots, a_k^1 ;

- pour $t = 1, 2, \dots$:
 - tirer $a_1^{t+1} \sim \pi(a_1|x, a_2^t, \dots, a_k^t)$
 - \vdots
 - tirer $a_k^{t+1} \sim \pi(a_k|x, a_1^{t+1}, \dots, a_{k-1}^{t+1})$.

Dans le cas d'un modèle à vraisemblance intractable, toutes ou certaines conditionnelles peuvent être inaccessibles. On se propose donc de remplacer ces dernières par des approximations ABC de la forme :

$$\pi_\varepsilon(a_i|x, a_1, \dots, a_{i-1}, a_{i+1}, \dots, a_k) \propto \int \pi(a_i)\pi(x^*|a_1, \dots, a_k)\mathbf{1}_{d(x, x^*) < \varepsilon} dx^*.$$

Le nouvel échantillonneur est nommé dans la suite ABC-Gibbs. De manière notable, il est presque aussi simple à implémenter qu'un échantillonneur ABC standard.

Plusieurs avantages sont liés à la décomposition de la loi en conditionnelles approchées. Le cas le plus simple est celui où certaines des conditionnelles sont calculables exactement ou pour lesquelles l'usage d'un algorithme de type Metropolis-Hastings est possible. Il est alors possible de restreindre l'usage de l'approximation ABC pour les parties du modèle qui le nécessitent. De plus, l'exploration de l'espace des paramètres est bien plus efficace. Selon le modèle étudié les gains peuvent être exponentiels (*e.g.* modèles hiérarchiques) ; le choix des statistiques résumées peut aussi être simplifié, non seulement la dimension des données est réduite dans le cas des modèles hiérarchiques mais il est en général plus simple de trouver une statistique intéressante de petite dimension pour un seul paramètre ; et ce d'autant plus que de nouvelles statistiques, faisant appel aux valeurs des autres paramètres, sont disponibles.

Cependant la nature et l'existence d'une loi limite à cet échantillonneur ne sont pas évidentes. On démontrera l'existence d'une loi limite à la chaîne de Markov décrite par la procédure précédente en montrant que l'application associée est contractante dans l'espace des mesures de probabilité. Pour ce faire nous utiliserons des techniques inspirées des travaux de Monmarché (2017). La loi limite étant inaccessible par la méthode précédente, nous devons ensuite calculer la distance entre cette loi et la loi *a posteriori* issue de l'échantillonneur de Gibbs exact.

En tant que fusion de l'échantillonneur de Gibbs avec les méthodes ABC, ABC-Gibbs bénéficie des avantages et des inconvénients qui leurs sont propres. Nominément, il est possible que, pour des statistiques σ_1 et σ_2 associées à a_1 et a_2 , les conditionnelles $\pi_\varepsilon(a_1|\sigma_1(x), a_2, \dots, a_k)$ et $\pi_\varepsilon(a_2|\sigma_2(x), a_1, \dots, a_k)$ ne soient pas compatibles — c'est à dire qu'il n'existe pas de loi jointe qui ait ces deux conditionnelles pour marginales —, auquel cas l'algorithme ne pourra pas converger vers une loi pertinente. En tant que méthode ABC, la majeure partie du temps de calcul de cet algorithme est due à la simulation de pseudo-observations et au calcul des statistiques résumées, aussi nous utiliseront le nombre de pseudo données simulées comme point de comparaison.

On se contentera ici de donner un résultat de convergence dans un cas particulier, un modèle hiérarchique de la forme $a \rightarrow b \rightarrow x$ avec $a \in \mathcal{A}$ et $b \in \mathcal{B}$ les paramètres et $x \in \mathcal{X}$ les observations. On suppose que la vraisemblance $\pi(b|a)$ est intractable tandis que $\pi(x|b)$ est accessible exactement.

1.1 Résultats théoriques

Existence d'une loi limite Si les deux conditions suivantes sont vérifiées,

$$(H1) \quad K_1 = \inf_{b \in \mathcal{B}} \pi(B_{b, \eta/4}) > 0 ;$$

$$(H2) \quad K_2 = \inf_{a \in \mathcal{A}} \inf_{b \in \mathcal{B}} \pi(B_{b, 3\eta/2} | a, x) > 0,$$

où $B_{x,r}$ désigne la boule centrée en x de rayon r . Alors l'échantillonneur ABC-Gibbs a une mesure invariante et converge vers celle-ci à vitesse géométrique de taux $1 - K_1 K_2$ en variation totale.

Remarquons que les précédentes hypothèses impliquent que \mathcal{A} et \mathcal{B} soient de mesure finie. Ces conditions sont vérifiées si \mathcal{A} et \mathcal{B} sont compacts et $b \mapsto \int \pi(b|a)\pi(a)da$ et $b \mapsto \pi(x|b)$ sont continues sur \mathcal{B} .

On peut aussi alléger les hypothèses au prix d'une vitesse de convergence moindre. Il est également possible de calculer une borne supérieure sur la distance en variation totale entre la loi résultant de cet échantillonneur et la loi issue d'un Gibbs exact, laquelle tend vers zéro lorsque la précision ABC tend vers 0.

2 Exemples jouets

Modèle hiérarchique normal Le résultat précédent s'étend simplement au modèle hiérarchique constitué d'un hyperparamètre a , de paramètres b_i chacun réalisation indépendante d'une loi $\pi(\cdot|a)$, et d'observations x_i chacune réalisation indépendante d'une loi $\pi(\cdot|b_i)$. On note \mathcal{X} l'ensemble des observations et \mathcal{B} l'ensemble des paramètres, on retrouve alors un modèle similaire à celui étudié dans la partie précédente.

Dans ce modèle, la conditionnelles $\pi(b|x, a)$ se simplifie en $\prod_i \pi(b_i|x_i, a)$, ce qui assure un gain important. La symétrie du modèle implique qu'il n'y a pas de problème d'incompatibilité des conditionnelles approchées. Dans cet exemple, on utilise des données synthétiques issues de lois normales $x_i \sim \mathcal{N}(b_i, \sigma I_{10})$; tous les priors et toutes les conditionnelles sont normales, les variances sont connues.

On compare les résultats de l'échantillonneur ABC-Gibbs et de l'échantillonneur ABC simple en mesurant la variance des estimateurs *a posteriori* approchées des paramètres et hyperparamètres. Ici, on a fait varier le nombre de points en fixant le nombre total de simulations, c'est à dire que plus N_{eps} augmente moins il y a de points dans l'échantillon final. On observe alors une compétition entre l'erreur Monte-Carlo qui explique la croissance linéaire finale et le biais dû à la non convergence de Gibbs qui explique la décroissance

de la variance initiale. Dans tous les cas on constate qu'ABC-Gibbs est bien plus efficace qu'ABC simple, ce qui était attendu car le modèle est pour ainsi dire construit pour être traité efficacement par Gibbs. Remarquons qu'il est impossible à ABC simple de reconstituer les valeurs des paramètres, la hiérarchie est presque réduite à $a \rightarrow (x_1, \dots, x_k)$, tandis qu'ABC-Gibbs y parvient.

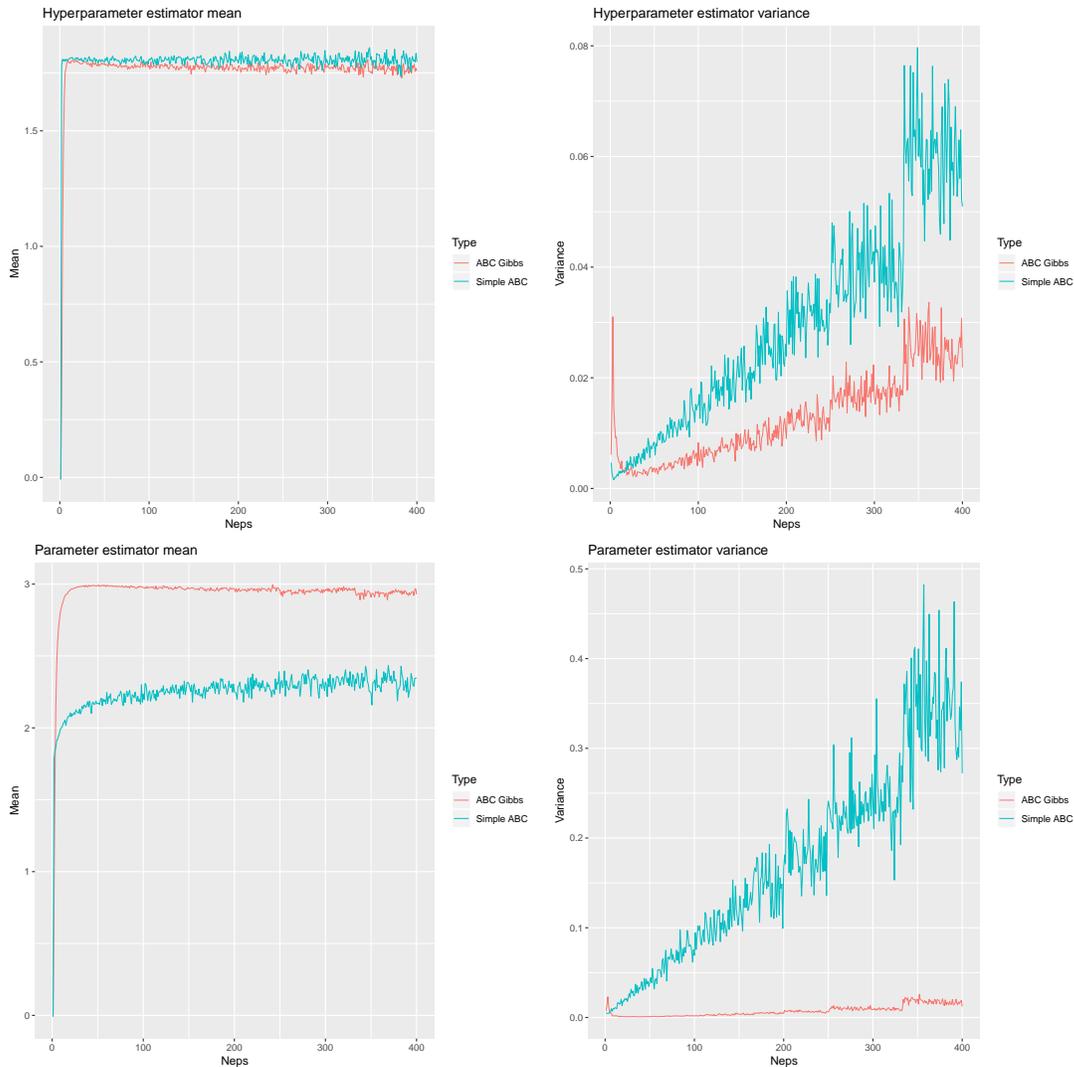


Figure 1: Valeur et variance de l'estimateur de l'hyperparamètre et de l'un des paramètres.

Problème inverse Une application intéressante des méthodes ABC est la résolution des problèmes inverses en EDP ou EDPS. Le point de vue bayésien sur ces problèmes

est communément accepté, Kaipio (2011), mais les vraisemblance sont généralement inaccessibles. Cependant la simulation de solutions est possible (avec néanmoins de sérieux problèmes de temps de calcul), ce qui incite à utiliser des méthodes ABC.

On propose ici un exemple jouet sur données simulées, les données sont issues d’une simulation par différences finies de l’équation :

$$\partial_t f = a(x)\partial_x^2 f ; x \in [0, 1], f(1, t) = f(0, t), \forall t.$$

On dispose donc des valeurs $f(x_i, t_j)$ pour i et j des indices de discrétisation. On cherche à reconstruire a comme fonction constante par morceaux $a(x) = \sum_{i=1}^k a_i \mathbf{1}_{[x_{i-1}, x_i]}$.

On a ici choisi des statistiques résumées adaptées à chaque paramètre : pour a_i c’est la projection sur x_{i-1}, x_i, x_{i+1} . Les simulations sont menées avec $k = 10$. On constate encore une fois que l’estimateur *a posteriori* approché issu de notre échantillonneur a une variance inférieure. On constate aussi que la loi reconstruite par notre échantillonneur est bien plus piquée que celle issue d’ABC simple.

Conclusion

Gibbs-ABC constitue un élément de réponse pour traiter des modèles complexes alliant des paramètres de grande dimension et une vraisemblance intractable. Cependant de nombreux problèmes découlent de son usage, parmi les plus importants on peut citer le choix des statistiques résumées — il est possible d’imaginer un modèle où la qualité de la statistique dépendrait de la valeur des autres paramètres —, leur possible incompatibilité, et le choix du seuil de précision de l’algorithme. Cependant, les résultats théoriques montrent qu’il existe une base à l’usage de ces méthodes.

Parmi les améliorations possibles, il est envisageable de développer une version adaptative du précédent algorithme, pour résoudre le problème du choix du seuil. Nous nous attacherons aussi à appliquer ces méthodes à des modèles réels.

Bibliographie

Jari P. Kaipio, Colin Fox (2011), The Bayesian Framework for Inverse Problems in Heat Transfer, *Heat Transfer Engineering*, 32, 9, 718–753.

Pierre Monmarché (2018), Elementary coupling approach for non-linear perturbation of markov processes with mean-field jump mechanisms and related problems.

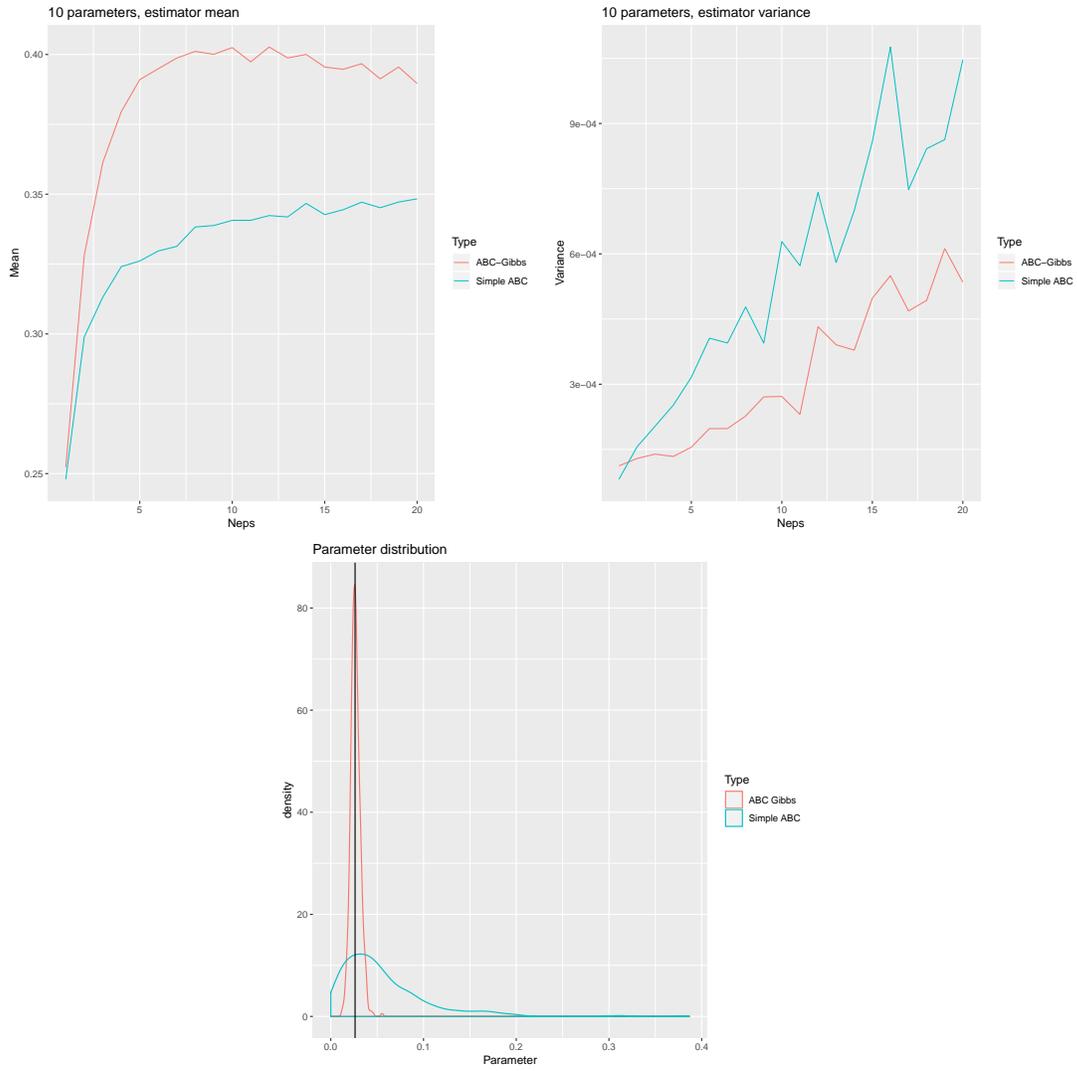


Figure 2: Valeur et variance de l'estimateur de l'un des paramètres, comparaison des distributions résultantes, la vraie valeur est indiquée par une ligne verticale.