

abc : un paquetage R pour le calcul bayésien approché

K. Csilléry^a, L. Lemaire^b, O. François^b and M.G.B. Blum^b

^a Ecologie des forêts méditerranéennes
INRA UR 629
Site Agroparc, Domaine Saint Paul
kati.csillery@gmail.com

^b Univ. Grenoble Alpes, TIMC-IMAG
CNRS, TIMC-IMAG
Faculté de médecine, 38000 Grenoble
louisiane.lemaire@orange.fr, olivier.francois@imag.fr, michael.blum@imag.fr

Mots clefs : Statistique bayésienne, Ecologie, Génétique des populations, Simulations.

Le calcul bayésien approché (« Approximate Bayesian Computation » en anglais) est une technique de statistique bayésienne qui permet de faire de l'inférence statistique dans des modèles où le calcul de la vraisemblance est trop coûteux d'un point de vue numérique. Cette technique a fait ses preuves en génétique des populations dans le cadre des modèles de *coalescence* [1]. Dans ces modèles, les paramètres à estimer sont en général de l'ordre de la dizaine tandis que le nombre de variable latentes est beaucoup plus grand ce qui rend difficile voire impossible le calcul de la vraisemblance et l'application des méthodes de Monte-Carlo par chaînes de Markov. Dans le calcul bayésien approché, l'inférence statistique repose sur des simulations numériques du modèle et un algorithme de rejet éventuellement suivi de méthodes dites d'ajustement [2,3].

Dans cet exposé, je présenterai comment effectuer les différentes étapes de l'analyse statistique avec le calcul bayésien approché : inférence des paramètres, comparaison de modèles et évaluation de l'adéquation du modèle aux données. Après une présentation des concepts statistiques, je montrerai comment effectuer ces opérations avec le paquetage R *abc* [4]. J'illustrerai les différentes étapes du calcul bayésien approché avec un exemple issu de la génétique des populations où l'objectif est d'inférer la démographie passée de populations humaines à partir d'indicateurs statistiques de la variabilité génétique.

Références

- [1] Beaumont, M. A., Zhang, W., & Balding, D. J. (2002). Approximate Bayesian computation in population genetics. *Genetics*, 162(4), 2025-2035.
- [2] Blum, M. G. B., & François, O. (2010). Non-linear regression models for Approximate Bayesian Computation. *Statistics and Computing*, 20(1), 63-73.
- [3] Blum, M. G. B. (2010). Approximate Bayesian computation: a nonparametric perspective. *Journal of the American Statistical Association*, 105(491), 1178-1187.
- [4] Csilléry, K., François, O., & Blum, M. G. B. (2012). abc: an R package for approximate Bayesian computation (ABC). *Methods in ecology and evolution*, 3(3), 475-479.