

Mélange de distributions comme fonction d'importance dans l'échantillonnage préférentiel combiné avec l'algorithme de Monte Carlo par Chaîne de Markov

Dorota Gajda*, Chantal Guihenneuc-Jouyaux**
Judith Rousseau***

*Biostatistiques, CESP Centre de recherche en Epidémiologie et Santé des Populations,
U1018, Inserm, F-94807, Villejuif, France
Université Paris Sud, UMRS1018, Villejuif, F-94807, France
dorota.gajda@inserm.fr,

** EA 4064 (épidémiologie environnementale : impact sanitaire des pollutions),
Faculté des Sciences Pharmaceutiques et Biologiques, Université Paris Descartes,
4, av de l'Observatoire, 75006 Paris, France

*** Université Paris Dauphine, Paris, France

Résumé. Les algorithmes de Monte Carlo par Chaîne de Markov (MCMC) sont très souvent utilisés pour estimer les lois a posteriori ainsi que leurs moments dans le cadre d'un modèle Bayésien. En effet, selon les modèles, les lois a posteriori ou leurs moments peuvent ne pas avoir d'expression analytique et le recours à des méthodes d'approximation est donc indispensable. Lors de l'étude empirique d'estimateurs, différents jeux donnés sont simulés sous le même modèle (et avec les mêmes valeurs de paramètres) et pour chaque jeu de données, les estimations a posteriori des paramètres sont obtenus via MCMC. Cette procédure est répétée pour d'autres valeurs des paramètres. Globalement, les temps de calcul peuvent être très importants. L'échantillonnage préférentiel (Importance Sampling en anglais, IS) combiné avec les algorithmes MCMC est une solution permettant de réduire ce temps de calcul. En effet, l'IS nécessite le choix d'une fonction d'importance que nous proposons construite comme mélange de lois a posteriori présélectionnées sur quelques jeux données simulés, lois a posteriori déjà estimées via MCMC. Les autres calculs ne nécessitent plus le recours aux algorithmes MCMC. Les approches évoquées ici sont illustrées sur deux exemples de modèles de Poisson.

1 Introduction

L'échantillonnage préférentiel (Importance Sampling en anglais, IS) est présenté ici comme une méthode d'optimisation algorithmique dans le cas de l'étude empirique d'un estimateur. En effet, même s'il est possible d'avoir des propriétés asymptotiques des estimateurs, les études empiriques sont nécessaires afin d'évaluer leurs comportements dans un cadre non asymptotique. La démarche consiste alors à définir ces situations caractéristiques (taille d'échantillon, valeurs