Algorithme SAEM : Par delà l'étape de simulation

Juliette Chevallier [1]

[1] *Institut de Mathématiques de Toulouse, INSA Toulouse,*
`juliette.chevallier@insa-toulouse.fr`

Résumé. L'algorithme d'espérance-maximisation ou algorithme EM permet de maximiser la vraisemblance de modèles à données manquantes dans des cadres très généraux. Lorsque l'étape d'espérance nous est inaccessible, on peut recourir à une approximation stochastique de l'algorithme EM (SAEM). La convergence du SAEM vers les maxima locaux de la vraisemblance observée ainsi que son efficacité numérique ont été démontrées. En dépit de cette performance, l'algorithme SAEM est très sensible à ses conditions initiales. Afin de palier ce problème, nous avons proposé une nouvelle classe d'algorithmes SAEM dont nous avons démontré la convergence vers des maxima locaux. Cette classe repose sur la simulation par une loi approchée de la vraie loi conditionnelle dans l'étape de simulation.

Après avoir rappelé quelques propriétés de l'algorithme SAEM, nous présenterons des développements récents visant à étendre son applicabilité. En particulier, nous concentrerons notre présentation sur l'amélioration de l'étape de simulation et présenterons une amélioration de l'algorithme SAEM basée sur des techniques de recuit simulé ou "tempering" permettant de favoriser sa convergence vers des maxima globaux : le SAEM tempéré ou tempering-SAEM.

Mots-clés. Algorithmes de type EM, Approximation stochastique, Distributions tempérées

Abstract. The expectation-maximization (EM) algorithm is a powerful computational technique for maximum likelihood estimation in incomplete data models. When the expectation step cannot be performed in closed form, a stochastic approximation of the EM algorithm (SAEM) can be used. The convergence of the SAEM toward critical points of the observed likelihood has been proved, and its numerical efficiency demonstrated. However, sampling from the posterior distribution may be intractable or have a high computational cost. Moreover, despite appealing features, the limit position of this algorithm can strongly depend on its starting one. In this talk, we propose a method to overcome these two limitations : sampling from an approximation of the distribution in the expectation phase of the SAEM.

After recalling some SAEM algorithm properties, we will present recent developments aiming at extending its applicability. In particular, we will concentrate our presentation on improving the simulation step, focusing on the tempering-SAEM. Inspired by the simulated annealing algorithm, the tmp-SAEM proposes to temper the posterior distribution of the SAEM sampling step to favor its convergence towards global maxima.

Keywords. EM-like algorithms, Stochastic approximation, tempering distributions

1 Estimation du maximum de vraisemblance par des algorithmes de type EM

Malheureusement, le problème de l'estimation du maximum de vraisemblance n'a généralement pas de solution sous forme close. Dans ce contexte, l'algorithme d'espérance-maximisation (EM) et ses variantes ont démontré leur efficacité pratique et restent des outils privilégiés.

1.1 L'algorithme EM et ses variantes

L'algorithme EM itère deux étapes jusqu'à convergence : une étape d'espérance, l'étape E, dans laquelle on calcule l'espérance de la vraisemblance en tenant compte des dernières variables observées et une étape de maximisation, l'étape M, dans laquelle on estime le maximum de vraisemblance des paramètres en maximisant la vraisemblance trouvée à l'étape E.

Autrement dit, étant donné l'estimée courante θ_k du paramètre θ que l'on cherche à estimer, l'algorithme EM itère successivement les deux étapes suivantes :

Étape E : Calculer l'espérance de log-vraisemblance conditionnelle

$$Q(\theta|\theta_k) \;=\; \int_{\mathcal{Z}} \log q(y,z;\theta) q(z|y;\theta_k) \, \mathrm{d}z \;=\; \mathbb{E}\left[\log q(Z|y,\theta_k)\right] \,,$$

Étape M : Maximiser $Q(\,\cdot\,|\theta_k)$ dans l'ensemble des paramètres admissibles Θ

$$\theta_{k+1} \;\in\; \operatorname*{argmax}_{\theta\in\Theta} Q(\theta|\theta_k) \,,$$

où $q(y,z;\theta)$ désigne la vraisemblance complète de notre modèle et $q(z|y;\theta)$ la vraisemblance conditionnelle.

La convergence de l'algorithme EM vers un maximum *local* de la vraisemblance observée a été démontrée par [8] puis corrigée par [17]. Cependant les hypothèses requises en première instance étaient difficiles à vérifier et le cadre introduit par [7] fournit des hypothèses plus raisonnables. En particulier, ils supposent que la vraisemblance complète des appartient à une famille exponentielle. En pratique, cette hypothèse n'est pas si restrictive et des modèles complexes satisfont à cette hypothèse.

Outre des techniques pour accélérer sa convergence [12], plusieurs améliorations de l'algorithme EM ont été proposées. Globalement, on peut distinguer deux types d'améliorations : celles concernant l'étape d'espérance et celles concernant l'étape de maximisation. En ce qui concerne cette dernière étape, l'*EM généralisé* [7] ne suppose plus une maximisation de l'espérance à chaque étape mais seulement un accroissement de celle-ci. Ainsi, on peut appliquer l'algorithme EM y compris en l'absence de solution analytique. Dans la version décrite par [10], l'étape de maximisation est réalisée par une méthode de Newton-Raphson.

Les solutions alternatives au calcul de l'espérance consistent en l'introduction d'une part de stochasticité au sein de la procédure d'estimation. Dans l'*EM stochastique* (SEM), [6] proposent de remplacer le calcul de l'espérance par une évaluation numérique de celle-ci *via* une simulation des données manquantes. [16] généralisent cette idée et remplacent le calcul de l'espérance par une approximation de cette dernière par une méthode de Monte-Carlo, donnant lieu au *Monte-Carlo EM* ou MCEM. En modulant le nombre de tirages aléatoires dans la somme de Monte-Carlo [5], on est en particulier à même de mimer un algorithme de type recuit simulé.

1.2 L'algorithme SAEM

Une approche alternative, développée par [7], consiste à remplacer le calcul de l'espérance par une approximation de type Robins-Monro [14], que l'on sait converger vers l'espérance sous des hypothèses *ad hoc*. On parle alors d'*approximation stochastique de l'algorithme EM* ou d'algorithme SAEM. Ainsi, la mise à jour du paramètre courant bénéficie des simulations précédentes. Plus précisément, étant donnée une approximation initiale de l'espérance de la log-vraisemblance Q_0 et une suite de pas positifs $(\gamma_k)_{k\in\mathbb{N}}$, la $k^{\text{ème}}$ itération de l'algorithme SAEM consiste en les trois étapes suivantes :

Étape S : Échantillonner la variable latente z_k selon la densité conditionnelle $q(\,\cdot\,|y;\theta_k)$;

Étape SA : Mettre à jour $Q_k(\theta)$ selon $Q_{k+1}(\theta) \;=\; Q_k(\theta) + \gamma_k\big(\log q(y, z_k;\theta) - Q_k(\theta)\big)$;

Étape M : Maximiser Q_{k+1} dans l'ensemble des paramètres admissibles Θ, *i.e.* trouver $\theta_{k+1} \in \operatorname*{argmax}_{\theta\in\Theta} Q_{k+1}(\theta)$.

En pratique, nous définissons Q_0 comme étant la fonction nulle sur Θ.

1.2.1 Algorithme SAEM et simulation approchée

Lorsque la simulation exacte des variables latentes n'est pas possible, recourir à une simulation approchée par une méthode de Monte-Carlo par chaînes de Markov, ou méthode MCMC pour Markov chain Monte Carlo en anglais, [3, 4, 15] s'avère fructueux. L'idée des méthodes MCMC est de générer une chaîne de Markov convergeant vers la loi cible contre laquelle on veut réaliser la simulation. En particulier, on remplace la génération d'un échantillon selon cette loi compliquée par un nombre, potentiellement élevé, de simulations contre des distributions que l'on espère plus simples.

En se basant sur les travaux de [9] qui établissent la convergence du MCMC-SAEM dans le cas où les variables générées durant la procédure d'estimation restent bornées, [2] démontrent la convergence du MCMC-SAEM en toute généralité. Signalons également que la convergence de cet algorithme ne nécessite qu'une seule étape de MCMC, ce qui le rend très compétitif sur le plan numérique.

Lorsque la distribution conditionnelle n'est pas disponible analytiquement, on peut tenter d'approcher cette distribution, par exemple en utilisant des méthodes de calcul bayésien approximatif (ABC, voir [11] pour une revue), également connues sous le nom de techniques sans vraisemblance. Ainsi, pour surmonter le problème d'échantillonnage, [13] ont proposé de combiner l'algorithme SAEM avec une étape ABC, conduisant à l'algorithme ABC-SAEM. Les simulations montrent que cet algorithme peut être calibré pour fournir une inférence précise.

2 Une nouvelle classe d'approximations stochastiques de l'algorithme EM

Malgré sa grande performance numérique, l'algorithme SAEM est très sensible à ses conditions initiales et, malgré la stochasticité de la procédure induite par l'approximation stochastique, peut rester piégé dans des maxima locaux.

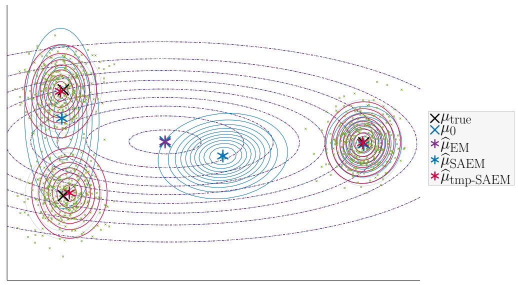
2.1 L'algorithme SAEM approché

Dans [1], nous avons proposé une nouvelle classe d'algorithmes SAEM : les algorithmes SAEM approchés, ou approximated-SAEM en anglais, dont on a démontré la convergence vers des maxima locaux sous des hypothèses standards. Cette classe repose sur la simulation par une loi approchée, en un sens à définir, de la vraie loi conditionnelle dans l'étape de simulation. En particulier, on englobe des algorithmes pré-existants tel que le MCMC-SAEM et l'ABC-SAEM.

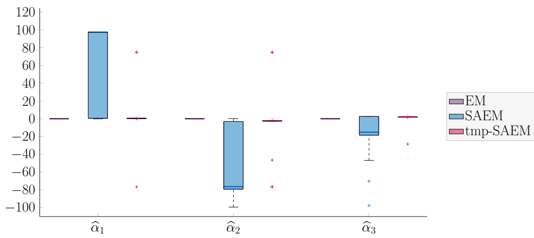
En fait d'approximation de la loi conditionnelle, on se donne une suite de distributions $\tilde{q}_k$ convergeant en moyenne sur tout compact de l'ensemble des paramètres admissibles Θ vers la vraie loi conditionnelle $q(z|y;\theta)$. Ainsi, étant donnée une suite de pas positifs $(\gamma_k)_{k\in\mathbb{N}}$ et notre suite $(q_k)_{k\in\mathbb{N}}$ comme décrite ci-dessus, l'algorithme SAEM approché itère les trois étapes suivantes :

Étape S : Échantillonner une "approximation" de la variable latente $\tilde{z}_k$ selon la loi approchée $\tilde{q}_k(\cdot;\theta_k)$;

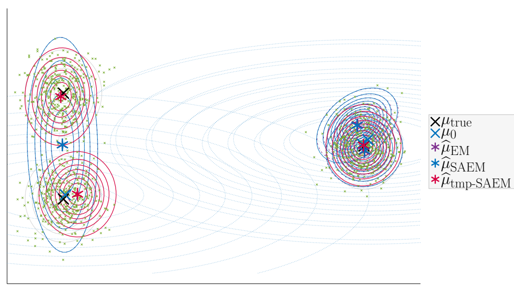
Étape SA : Mettre à jour $Q_k(\theta)$ selon $Q_{k+1}(\theta) \ = \ Q_k(\theta) + \gamma_k\big(\log q(y,\tilde{z}_k;\theta) - Q_k(\theta)\big)$;



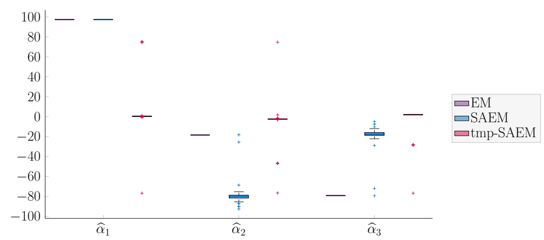
(a) Init. 1 – Performances qualitatives



(b) Init. 1 – Erreurs relatives pour α (en %)



(c) Init. 2 – Performances qualitatives



(d) Init. 2 – Erreurs relatives pour α (en %)

FIGURE 1 – Modèle de mélange gaussien.

Étape M : Maximiser Q_{k+1} : $\theta_{k+1} \in \underset{\theta \in \Theta}{\operatorname{argmax}}\, Q_{k+1}(\theta)\,.$

2.2 Échapper au maxima locaux

Enfin, en se basant sur des techniques de recuit simulé, nous avons proposé une version tempérée de l'algorithme SAEM afin de favoriser sa convergence vers des maxima globaux. Dans cette version, on approche la loi conditionnelle en la tempérant suivant un schéma de températures sinusoïdal amorti.

À la figure 1, nous appliquons cette méthode à l'estimation des paramètres dans les modèles de mélange gaussien et en illustrons ainsi la supériorité numérique sur l'algorithme SAEM classique.

Références

[1] Stéphanie Allassonnière and Juliette Chevallier. A new class of stochastic em algorithms. escaping local maxima and handling intractable sampling. *Computational Statistics & Data Analysis*, 159 :107159, 2021.

[2] Stéphanie Allassonnière, Estelle Kuhn, and Alain Trouvé. Construction of Bayesian deformable models via a stochastic approximation algorithm : A convergence study. *Bernoulli*, 16(3) :641–678, 2010.

[3] Christophe Andrieu, Nando De Freitas, Arnaud Doucet, and Michael I Jordan. An introduction to MCMC for machine learning. *Machine learning*, 50(1-2) :5–43, 2003.

[4] Steve Brooks, Andrew Gelman, Galin Jones, and Xiao-Li Meng. *Handbook of Markov chain Monte Carlo*. CRC press, 2011.

[5] Gilles Celeux, Didier Chauveau, and Jean Diebolt. On stochastic versions of the EM algorithm. Technical report, INRIA, 1995.

[6] Gilles Celeux and Jean Diebolt. The SEM algorithm : a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem. *Computational statistics quarterly*, 2 :73–82, 1985.

[7] Bernard Delyon, Marc Lavielle, and Eric Moulines. Convergence of a stochastic approximation version of the EM algorithm. *The Annals of Statistics*, 27(1) :94–128, 1999.

[8] Arthur Dempster, Nan M. Laird, and Donald B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 39(1) :1–38, 1977.

[9] Estelle Kuhn and Marc Lavielle. Coupling a stochastic approximation version of EM with an MCMC procedure. *ESAIM : Probability and Statistics*, 8 :115–131, 2004.

[10] Kenneth Lange. A gradient algorithm locally equivalent to the EM algorithm. *Journal of the Royal Statistical Society : Series B*, 57(2) :425–437, 1995.

[11] Jean-Michel Marin, Pierre Pudlo, Christian P Robert, and Robin J Ryder. Approximate Bayesian computational methods. *Statistics and Computing*, 22(6) :1167–1180, 2012.

[12] Geoffrey McLachlan and Thriyambakam Krishnan. *The EM algorithm and extensions*, volume 382. John Wiley & Sons, 2007.

[13] Umberto Picchini and Adeline Samson. Coupling stochastic EM and approximate Bayesian computation for parameter inference in state-space models. *Computational Statistics*, 33(1) :179–212, 2018.

[14] Herbert Robbins and Sutton Monro. A stochastic approximation method. *The annals of mathematical statistics*, pages 400–407, 1951.

[15] Christian P. Robert and George Casella. *Monte Carlo Statistical Methods*. Springer Texts in Statistics. Springer-Verlag New York, 1999.

[16] Greg C. Wei and Martin A. Tanner. A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *Journal of the American statistical Association*, 85(411) :699–704, 1990.

[17] Jeff CF. Wu. On the convergence properties of the EM algorithm. *The Annals of statistics*, 11(1) :95–103, 1983.