Available online at www.sciencedirect.com
ScienceDirect
C. R. Acad. Sci. Paris, Ser. I 347 (2009) 205–210
COMPTES RENDUS MATHEMATIQUE

Statistics

ABC methods for model choice in Gibbs random fields

Aude Grelaud [a,b,c], Christian P. Robert [b,c], Jean-Michel Marin [c,d]

[a] *INRA, unité MIG, domaine du Vilvert, 78350 Jouy-en-Josas, France*
[b] *CEREMADE, Université Paris Dauphine, place du Maréchal-de-Lattre-de-Tassigny, 75775 Paris cedex 16, France*
[c] *CREST-INSEE, Timbre J340, 3, avenue Pierre-Larousse, 92240 Malakoff, France*
[d] *Institut de mathématiques et de modélisation de Montpellier, Université Montpellier 2, 34095 Montpellier cedex, France*

Received 25 September 2008; accepted after revision 15 December 2008

Available online 30 January 2009

Presented by Paul Deheuvels

Abstract

We consider the problem of model selection within the class of Gibbs random fields. In a Bayesian framework, this choice relies on the evaluation of the posterior probabilities of all models. We define an extended parameter setting, including the model index and show the existence of a corresponding sufficient statistic made of the conjunction of the sufficient statistics of all models. We use this statistic to derive an ABC algorithm. ***To cite this article: A. Grelaud et al., C. R. Acad. Sci. Paris, Ser. I 347 (2009).***

Résumé

Choix bayésien de champs de Gibbs par méthode ABC. On s'intéresse au problème du choix bayésien de modèles de champs de Gibbs. Ce choix repose sur l'évaluation des probabilités a posteriori des modèles. Nous montrons l'existence d'une statistique exhaustive pour l'ensemble des paramètres, incluant l'indice du modèle, constituée de la concaténation de statistiques exhaustives de chacun des modèles. Nous utilisons cette statistique pour construire un algorithme ABC. ***Pour citer cet article : A. Grelaud et al., C. R. Acad. Sci. Paris, Ser. I 347 (2009).***

Version française abrégée

Les champs de Gibbs [3] sont des modèles probabilistes dont l'expression de la densité de probabilité par rapport à la mesure de comptage est donnée par

$$f(\mathbf{x}|\boldsymbol{\theta}) = \frac{1}{Z_{\boldsymbol{\theta}}} \exp\{\boldsymbol{\theta}^T S(\mathbf{x})\},$$

où $\mathbf{x} \in \mathcal{X} \subset \mathbb{R}^n$ ($\mathcal{X}$ étant un ensemble fini), $S(\cdot)$ est le potentiel définissant le champ de Gibbs, $\boldsymbol{\theta} \in \mathbb{R}^p$ est le paramètre associé au modèle et $Z_{\boldsymbol{\theta}} > 0$ est la constante de normalisation. Dans la majorité des cas, $Z_{\boldsymbol{\theta}}$ est impossible à calculer

E-mail addresses: aude.grelaud@jouy.inra.fr (A. Grelaud), xian@ceremade.dauphine.fr (C.P. Robert), jean-michel.marin@math.univ-montp2.fr (J.-M. Marin).

1631-073X/$ – see front matter
doi:10.1016/j.crma.2008.12.009

ce qui complique l'inférence sur le paramètre $\boldsymbol{\theta}$. Nous pouvons dès maintenant noter que $S(\mathbf{x})$ est une statistique exhaustive pour le modèle.

Les champs markoviens sont des cas particuliers de champs de Gibbs utilisés pour modéliser des données présentant des corrélations spatiales. Dans ce cas, le potentiel $S(\cdot)$ est associé à une relation symétrique de voisinage notée $i \sim i'$, au sens où la densité conditionnelle de x_i ne dépend que des $x_{i'}$ tels que $i \sim i'$.

Notre objectif est de choisir au travers d'un échantillon le modèle le plus approprié parmi un ensemble de M champs de Gibbs de densités (2). Le problème peut être vu comme un choix entre M fonctions de potentiel S_m, $1 \leqslant m \leqslant M$.

Comme suggéré dans [9] (chapitre 7) nous considérons un nouveau vecteur de paramètres

$$(\mathcal{M}, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_M) \in \Theta = \{1, \ldots, M\} \times \bigcup_{m=1}^{M} \Theta_m,$$

qui inclut l'indice du modèle $\mathcal{M}$. Nous montrons dans la Proposition 2.1 que $S(\mathbf{x}) = \{S_1(\mathbf{x}), \ldots, S_M(\mathbf{x})\}$ est une statistique exhaustive pour ce vecteur de paramètres.

Dans un cadre bayésien, nous définissons une distribution a priori pour l'indice du modèle, $\pi(\mathcal{M} = m)$, ainsi que pour le paramètre conditionnellement à la valeur m de l'indice du modèle, $\pi_m(\boldsymbol{\theta}_m)$, définie sur l'espace Θ_m. Le choix du modèle repose alors sur les probabilités a posteriori des modèles

$$\mathbb{P}(\mathcal{M} = m|\mathbf{x}) \propto \int_{\Theta_m} f_m(\mathbf{x}|\boldsymbol{\theta}_m)\pi_m(\boldsymbol{\theta}_m)\,\mathrm{d}\boldsymbol{\theta}_m \pi(\mathcal{M} = m).$$

Les méthodes utilisées habituellement pour estimer ces quantités nécessitent d'être en mesure de calculer exactement la vraisemblance (voir [2] par exemple). Elles ne sont pas applicables ici car la constante de normalisation n'est pas disponible. Des méthodes basées sur l'utilisation de variables instrumentales [7] permettent de s'affranchir du fait que la constante de normalisation est manquante. Cependant, elles sont extrêmement gourmandes en temps de calcul. Récemment, des méthodes permettant de faire de l'inférence sur les paramètres sans utiliser la vraisemblance sont apparues (voir [8,1] et [6]). Dans cette note, étendue dans [4], nous utilisons une méthode de ce type pour estimer les probabilités a posteriori des modèles.

L'Algorithme d'échantillonnage exact de [8] est le suivant : pour un vecteur de données $\mathbf{x}^0$ de distribution (1) et une distribution a priori $\pi(\boldsymbol{\theta})$ pour le paramètre $\boldsymbol{\theta}$, une valeur $\boldsymbol{\theta}^*$ est générée selon la distribution a priori, $\boldsymbol{\theta}^* \sim \pi(\cdot)$, puis une valeur $\mathbf{x}^*$ est générée selon $\mathbf{x}^* \sim f(\cdot|\boldsymbol{\theta}^*)$. Le processus est répété jusqu'à ce que $\mathbf{x}^* = \mathbf{x}^0$. L'Algorithme 1 génère ainsi un paramètre $\boldsymbol{\theta}$ de distribution $\pi(\boldsymbol{\theta}|\mathbf{x}^0)$.

Cet algorithme est difficile, voire impossible, à utiliser quand $\mathbf{x}^* = \mathbf{x}^0$ est un événement de faible probabilité. Pour éviter ce problème, le schéma ABC (Approximate Bayesian Computation) accepte les paramètres $\boldsymbol{\theta}^*$ pour lesquels $\rho(T(\mathbf{x}^0), T(\mathbf{x}^*)) < \epsilon$ où $T(\cdot)$ est une statistique résumée, $\rho(\cdot,\cdot)$ une distance et ϵ une tolérance [1]. L'Algorithme 2 génère des paramètres de distribution $\pi\{\boldsymbol{\theta}|\rho(T(\mathbf{x}^*), T(\mathbf{x}^0)) < \epsilon\}$ qui peut être une bonne approximation de la distribution $\pi(\boldsymbol{\theta}|\mathbf{x}^0)$ lorsque ϵ est petit et $T(\cdot)$ est une statistique suffisamment représentative. (Voir [5] et [10] pour des évaluations de ces approximations.)

Rappelons que l'objectif est d'estimer les probabilités a posteriori des modèles $\mathbb{P}(\mathcal{M} = m|\mathbf{x})$. Nous proposons le schéma ABC-MC (Algorithme 3) qui génère un échantillon $(\boldsymbol{\theta}^{i*}_{m^{i*}}, m^{i*})_{(1 \leqslant i \leqslant N)}$ dont la distribution est approximativement $\pi(m, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_M|\mathbf{x}^0)$. Une estimation fondée sur cet échantillon est donnée par la fréquence empirique des visites du modèle

$$\widehat{\mathbb{P}}\big(\mathcal{M} = m|\mathbf{x}^0\big) = \sharp\big\{m^{i*} = m\big\}/N,\ 1 \leqslant m \leqslant M.$$

Nous avons étudié les performances de cette méthode sur deux cas de champs de Gibbs pour lesquels la constante de normalisation est disponible. La Fig. 1 permet de comparer la vraie valeur de la probabilité a posteriori du modèle à notre estimation dans le cas exact, puis en introduisant un seuil. Dans les deux cas, les valeurs estimées sont très proches de la vraie valeur.

1. Introduction

The models considered in this Note are Gibbs random fields whose definition relies on a potential function $S(\cdot)$ [3]:

Definition 1.1. Gibbs random fields are probabilistic models associated with the density function

$$f(\mathbf{x}|\boldsymbol{\theta}) = \frac{1}{Z_{\boldsymbol{\theta}}} \exp\{\boldsymbol{\theta}^T S(\mathbf{x})\}, \tag{1}$$

where $\mathbf{x} \in \mathcal{X} \subset \mathbb{R}^n$, $\mathcal{X}$ is finite, $S(\cdot)$ is the potential function defining the random field, taking values in $\mathbb{R}^p$, $\boldsymbol{\theta} \in \mathbb{R}^p$ is the associated scale parameter and $Z_{\boldsymbol{\theta}}$ is the corresponding normalising constant.

We can already note that $S(\cdot)$ is a sufficient statistic for model (1). In most realistic settings, the summation $Z_{\boldsymbol{\theta}} = \sum_{\mathbf{x}\in\mathcal{X}} \exp\{\boldsymbol{\theta}^T S(\mathbf{x})\}$ involves too many terms to be manageable and this greatly complicates the inference on the scale parameter $\boldsymbol{\theta}$.

An important case of Gibbs random fields are Markov random fields, used to model the dependency within spatially correlated data. In this case, the potential function $S(\cdot)$ is associated with a symmetric neighbourhood structure, meaning that the conditional density of x_i only depends on the $x_{i'}$ such that i and i' are neighbours, denoted by $i \sim i'$. For instance, if we consider Potts models, $S(\mathbf{x}) = \sum_{i'\sim i} \mathbb{I}_{\{x_i = x_{i'}\}}$. In a Bayesian context, the model selection is driven by the posterior probabilities of the models. We propose here and in [4] a likelihood-free algorithm that approximates these quantities.

2. Model choice in Gibbs random fields

In a model choice perspective, we face M Gibbs fields in competition, each one being associated with a potential function S_m $(1 \leqslant m \leqslant M)$, i.e. with corresponding density

$$f_m(\mathbf{x}|\boldsymbol{\theta}_m) = \exp\{\boldsymbol{\theta}_m^T S_m(\mathbf{x})\} / Z_{\boldsymbol{\theta}_m,m}, \tag{2}$$

where $\boldsymbol{\theta}_m \in \Theta_m$ and $Z_{\boldsymbol{\theta}_m,m}$ is the unknown normalising constant. For instance, in the case of Potts models, the choice is between M neighbourhood relations $i \overset{m}{\sim} i'$ $(1 \leqslant m \leqslant M)$ with $S_m(\mathbf{x}) = \sum_{i \overset{m}{\sim} i'} \mathbb{I}_{\{x_i = x_{i'}\}}$.

As suggested in [9] (Chapter 7) we consider an extended parameter that includes the model index $\mathcal{M}$,

$$(\mathcal{M}, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_M) \in \Theta = \{1, \ldots, M\} \times \bigcup_{m=1}^{M} \Theta_m.$$

We define a prior distribution on the model index, $\pi(\mathcal{M} = m)$, as well as a prior distribution on the model parameter conditional on the value m of $\mathcal{M}$, $\pi_m(\boldsymbol{\theta}_m)$. Formally, the choice between those models follows from the posterior probabilities of the models. The computational target is thus

$$\mathbb{P}(\mathcal{M} = m|\mathbf{x}) \propto \int_{\Theta_m} f_m(\mathbf{x}|\boldsymbol{\theta}_m)\pi_m(\boldsymbol{\theta}_m)\,\mathrm{d}\boldsymbol{\theta}_m \pi(\mathcal{M} = m),$$

that is to say, the marginal of the posterior distribution on $\mathcal{M}$ given $\mathbf{x}$.

Most simulation methods used to generate a sample from the posterior distribution require the likelihood to be available in closed form (see [2] for a review). Here, the unavailability of the normalising constants $Z_{\boldsymbol{\theta}_m,m}$ prevents the use of those methods. [7] propose an auxilary variable method for which the knowledge of the normalising constant is not required, but this approach is very time consuming. As an alternative, we propose an approach based on a likelihood-free algorithm. This method requires to use some summary statistics whose choice has a significant great impact on the quality of the results. The optimal choice is to consider only sufficient statistics and a sufficient statistic for the extended parameter is actually available in this setting.

Proposition 2.1. *$S_m(\cdot)$ being the potential function corresponding to the Gibbs random field m, $S(\mathbf{x}) = \{S_1(\mathbf{x}), \ldots, S_M(\mathbf{x})\}$ is a sufficient statistic for the joint parameter $(\mathcal{M}, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_M)$.*

Proof. As already noted, $S_m(\mathbf{x})$ is a sufficient statistic for model m. Thus, the vector of statistics $S(\mathbf{x}) = \{S_1(\mathbf{x}), \ldots, S_M(\mathbf{x})\}$ is obviously sufficient for each model m as it includes a sufficient statistic for each model. Moreover, the structure of the Gibbs random field allows for a specific factorisation of the distribution $f_m(\mathbf{x}|\boldsymbol{\theta}_m)$. Indeed, the density of $\mathbf{x}$ in model m factorises as

$$f_m(\mathbf{x}|\boldsymbol{\theta}_m) = f_m^1\big(\mathbf{x}|S(\mathbf{x})\big) f_m^2\big(S(\mathbf{x})|\boldsymbol{\theta}_m\big) = \frac{1}{n\{S(\mathbf{x})\}} f_m^2\big(S(\mathbf{x})|\boldsymbol{\theta}_m\big)$$

where $f_m^2(S(\mathbf{x})|\boldsymbol{\theta}_m)$ is the density of $S(\mathbf{x})$ within model m [not to be confused with the density of $S_m(\mathbf{x})$] and $n\{S(\mathbf{x})\} = \sharp\{\tilde{\mathbf{x}} \in \mathcal{X}\colon\ S(\tilde{\mathbf{x}}) = S(\mathbf{x})\}$ is the cardinal of the set of elements of $\mathcal{X}$ with the same value of the sufficient statistic, which does not depend on m (the support of f_m is constant in m). The statistic S is therefore also sufficient for the joint parameter $(\mathcal{M}, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_M)$. □

The fact that the concatenation of the sufficient statistics of each model is sufficient for the joint parameter $(\mathcal{M}, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_M)$ is specific to Gibbs random fields, since $f_m^1(\mathbf{x}|S(\mathbf{x}))$ usually depends on m.

3. ABC methods

Recently, new methods have been developed to evaluate posterior distributions when the likelihood function is analytically or computationally intractable (introduced by [8] and extended in [1] and [6]).

The rejection sampling method introduced by [8] samples exactly from the posterior distribution without computing the likelihood. It can be briefly described as follows: given a dataset $\mathbf{x}^0$ associated with the sampling distribution $f(\cdot|\boldsymbol{\theta})$, and under a prior distribution $\pi(\boldsymbol{\theta})$ on the parameter $\boldsymbol{\theta}$, jointly generate a parameter value $\boldsymbol{\theta}^*$ from the prior, $\boldsymbol{\theta}^* \sim \pi(\cdot)$ and a value $\mathbf{x}^*$ from the sampling distribution $\mathbf{x}^* \sim f(\cdot|\boldsymbol{\theta}^*)$ until $\mathbf{x}^* = \mathbf{x}^0$. The algorithmic rendering is as follows:

Algorithm 1 *(Rejection algorithm).*

(i) Generate $\boldsymbol{\theta}^*$ from the prior $\pi(\cdot)$ and $\mathbf{x}^*$ from the sampling density $f(\cdot|\boldsymbol{\theta}^*)$.
(ii) Accept $\boldsymbol{\theta}^*$ if $\mathbf{x}^* = \mathbf{x}^0$, otherwise, repeat (i).

This algorithm is difficult, or even impossible, to implement when the equality $\mathbf{x}^* = \mathbf{x}^0$ has a low probability. Two levels of approximations are adopted to define the Approximate Bayesian Computation (ABC) algorithm [1]. First, the parameter value $\boldsymbol{\theta}^*$ is accepted when a distance ρ between $\mathbf{x}^*$ and $\mathbf{x}^0$ is within a tolerance ϵ is less time-consuming. Moreover, the data $\mathbf{x}^0$ being usually of high dimension, another level of approximation is enforced by replacing the distance $\rho(\mathbf{x}^*, \mathbf{x}^0)$ with a corresponding distance between summary statistics $\rho(T(\mathbf{x}^*), T(\mathbf{x}^0))$, T being not sufficient. The algorithm reads as:

Algorithm 2 *(ABC algorithm).*

(i) Generate $\boldsymbol{\theta}^*$ from the prior $\pi(\cdot)$ and $\mathbf{x}^*$ from the sampling density $f(\cdot|\boldsymbol{\theta}^*)$.
(ii) Compute the distance $\rho(T(\mathbf{x}^0), T(\mathbf{x}^*))$.
(iii) Accept $\boldsymbol{\theta}^*$ if $\rho(T(\mathbf{x}^0), T(\mathbf{x}^*)) < \epsilon$, otherwise, repeat from (i).

The output of the ABC algorithm is therefore a simulation from the distribution $\pi\{\boldsymbol{\theta}|\rho(T(\mathbf{x}^*), T(\mathbf{x}^0)) < \epsilon\}$ and it can be considered as a good approximation of the distribution $\pi(\boldsymbol{\theta}|\mathbf{x}^0)$ if the tolerance ϵ is small and the summary statistic T is close to being sufficient. [5] consider the selection of the statistic T in terms of likelihood ratio variations, while [10] replaces the approximation with an exact simulation based on a convolution with an arbitrary kernel.

The ABC algorithm does sample exactly from the posterior distribution $\pi(\boldsymbol{\theta}|\mathbf{x}^0)$ when $\epsilon = 0$ and when T is a sufficient statistic for the parameter $\boldsymbol{\theta}$. The proof is obvious since $\mathbb{P}(\mathcal{M} = m|\mathbf{x}) = \mathbb{P}(\mathcal{M} = m|T(\mathbf{x}))$ when T is a sufficient statistic. The tolerance ϵ is usually chosen as an empirical quantile (1% or 10%) of $\rho(T(\mathbf{x}^*), T(\mathbf{x}^0))$ when $\mathbf{x}^*$ is simulated from the marginal distribution, see [1].

Our target is to estimate the posterior probabilities of the models, $\mathbb{P}(\mathcal{M} = m|\mathbf{x})$. We have shown in Section 1 that the vector of statistics $S(\mathbf{x}) = \{S_1(\mathbf{x}), \ldots, S_M(\mathbf{x})\}$ is sufficient for the joint parameter $(\mathcal{M}, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_M)$. Hence, we now propose the ABC-MC algorithm (where MC stands for Model Choice) which allows to sample approximately from the joint posterior distribution $\pi\{(\mathcal{M}, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_M)|\mathbf{x}^0\}$:

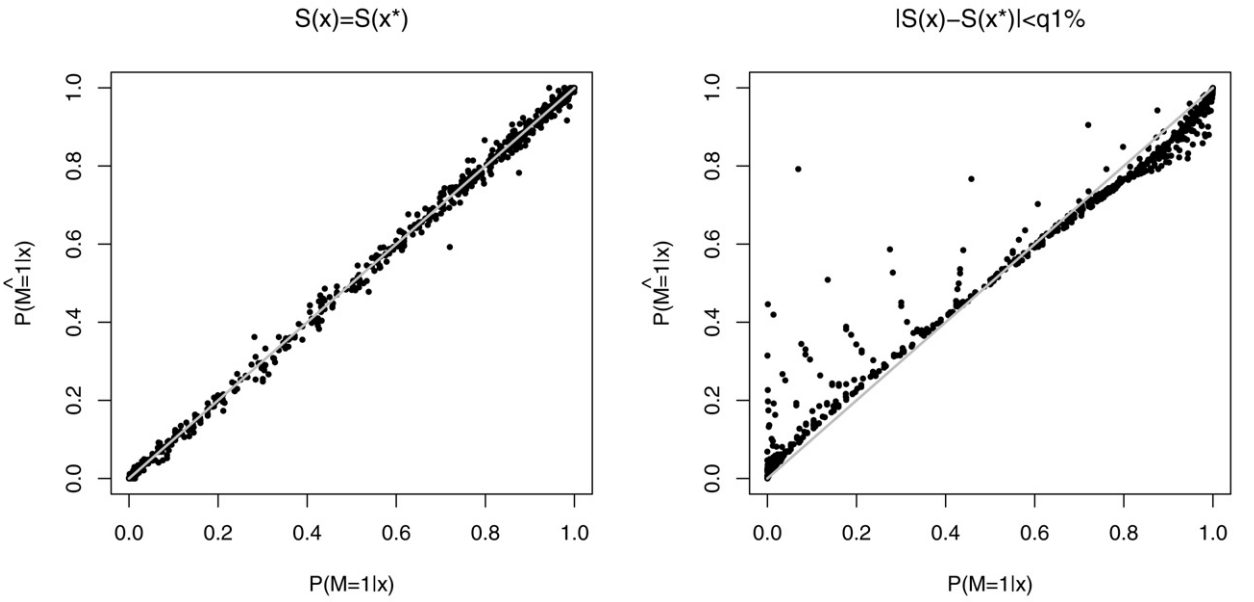

Fig. 1. (Left) Comparison of the true probability $\mathbb{P}(\mathcal{M}=1|\mathbf{x}^0)$ with $\widehat{\mathbb{P}}(\mathcal{M}=1|\mathbf{x}^0)$ over 2000 simulated sequences and 4×10^6 proposals from the prior. The grey line is the diagonal. (Right) Same comparison when using a tolerance ϵ corresponding to the 1% quantile on the distances.

Fig. 1. (Gauche) Comparaison de la vraie valeur $\mathbb{P}(\mathcal{M}=1|\mathbf{x}^0)$ à $\widehat{\mathbb{P}}(\mathcal{M}=1|\mathbf{x}^0)$ sur 2000 séquences simulées et en utilisant un échantillon de 4×10^6 propositions. La ligne grise correspond à la diagonale. (Droite) Même comparaison en utilisant un seuil ϵ correspondant au quantile à 1% des distances.

Algorithm 3 *(ABC algorithm for model choice (ABC-MC)).*

(i) Generate m^* from the prior $\pi(\mathcal{M}=m)$.
(ii) Generate $\boldsymbol{\theta}^*_{m^*}$ from the prior $\pi_{m^*}(\cdot)$.
(iii) Generate $\mathbf{x}^*$ from the sampling distribution $f_{m^*}(\cdot|\boldsymbol{\theta}^*_{m^*})$.
(iv) Compute the distance $\rho(S(\mathbf{x}^0), S(\mathbf{x}^*))$.
(v) Accept $(\boldsymbol{\theta}^*_{m^*}, m^*)$ if $\rho(S(\mathbf{x}^0), S(\mathbf{x}^*)) < \epsilon$, otherwise, repeat from (i).

As noted above, the algorithm is exact when $\epsilon = 0$ since S is sufficient.

A standard Monte Carlo approximation of the posterior probabilities based on a sample of N values $(\boldsymbol{\theta}^{i*}_{m^{i*}}, m^{i*})$ $(1 \leqslant i \leqslant N)$, generated from this algorithm is the empirical frequencies of visits to the models, namely

$$\widehat{\mathbb{P}}\big(\mathcal{M}=m|\mathbf{x}^0\big) = \sharp\big\{m^{i*}=m\big\}/N \quad (1 \leqslant m \leqslant M)$$

where $\sharp\{m^{i*}=m\}$ denotes the number of simulated m^{i*}'s equal to m. Those posterior probability estimators can be plugged-in to approximate the Bayes factor associated with the data $\mathbf{x}^0$ in favour of a model m relative to another model m'.

Example. Let us consider $\mathbf{x} = (x_1, \ldots, x_n)$, with $n = 100$, a vector of binary variables. We compare two Potts models with tractable normalising constants, namely a model with no neighbourhood ($\mathcal{M}=1$)

$$f_1(\mathbf{x}|\boldsymbol{\theta}_1) = \exp\left(\boldsymbol{\theta}_0 \sum_{i=1}^{n} \mathbb{I}_{\{x_i=1\}}\right) \Big/ \big\{1+\exp(\boldsymbol{\theta}_1)\big\}^n,$$

with a model with a nearest neighbour structure ($\mathcal{M}=2$)

$$f_2(\mathbf{x}|\boldsymbol{\theta}_2) = \frac{1}{2}\exp\left(\boldsymbol{\theta}_2 \sum_{i=2}^{n} \mathbb{I}_{\{x_i=x_{i-1}\}}\right) \Big/ \big\{1+\exp(\boldsymbol{\theta}_2)\big\}^{n-1}.$$

If $S_1(\mathbf{x}) = \sum_{i=1}^{n} \mathbb{I}_{\{x_i=1\}}$ and $S_2(\mathbf{x}) = \sum_{i=2}^{n} \mathbb{I}_{\{x_i=x_{i-1}\}}$, the vector $S = \{S_1, S_2\}$ is a sufficient statistic for the joint parameter $(\mathcal{M}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$. We choose $\rho(.)$ as the Euclidean distance.

Under uniform priors for the scale parameters, $\boldsymbol{\theta}_1 \sim \mathcal{U}(-5, 5)$ and $\boldsymbol{\theta}_2 \sim \mathcal{U}(0, 6)$, the bounds ± 5 and 6 being the phase transition values for the corresponding Potts models, the posterior probabilities can be calculated by a straightforward rational fraction integration.

We simulated 1000 datasets under each model with parameters simulated from the priors. For each dataset, we simulated 4×10^6 sets $(m^*, \boldsymbol{\theta}^*_{m^*}, \mathbf{x}^*)$. Fig. 1 shows that the *ABC* approximations of the model posterior probabilities are very close to the true values for all values of $\mathbb{P}(\mathcal{M} = 0|\mathbf{x}^0)$ when we accept the simulations for which $S(\mathbf{x}^*) = S(\mathbf{x}^0)$ (left graph) and also when we introduce a tolerance ϵ equal to the 1% quantile of the distances (right graph), even thought the discrepancy is larger in that case. The number of acceptances can be very low when we accept only the simulations for which $S(\mathbf{x}^*) = S(\mathbf{x}^0)$ (12 in the worst case). Given that using the tolerance version allows for more simulations to be used in the posterior probabilities approximation, we thus recommend using this approach.

Acknowledgements

The authors' research is partly supported by the Agence Nationale de la Recherche (ANR, 212, rue de Bercy, 75012 Paris) through the 2005 project ANR-05-BLAN-0196-01 Misgepop and by a grant from Région Ile-de-France. We thanks F. Rodolphe and J.-F. Taly for their comments.

References

[1] M. Beaumont, W. Zhang, D. Balding, Approximate Bayesian computation in population genetics, Genetics 162 (2002) 2025–2035.
[2] B.P. Carlin, S. Chib, Bayesian model choice via Markov chain Monte Carlo methods, J. Roy. Statist. Soc. Ser. B 57 (1995) 473–484.
[3] N. Cressie, Statistics for Spatial Data, revised ed., Wiley Series in Probability and Statistics, Wiley, New York, 1993.
[4] A. Grelaud, C.P. Robert, J.-M. Marin, F. Rodolphe, J.-F. Taly, ABC methods for model choice in Gibbs random fields, arXiv:0807.2767v1 [stat.CO], 2008.
[5] P. Joyce, P. Marjoram, Approximately sufficient statistics and Bayesian computation, Stat. Appl. Genet. Mol. Biol. 7 (1) (2008) (Article 26).
[6] P. Marjoram, J. Molitor, V. Plagnol, S. Tavaré, Markov chain Monte Carlo without likelihoods, Proc. Natl. Acad. Sci. USA 100 (26) (2003) 15324–15328.
[7] J. Moeller, A.N. Pettitt, R. Reeves, K.K. Berthelsen, An efficient Markov chain Monte Carlo method for distributions with intractable normalising constants, Biometrika 93 (2006) 451–458.
[8] J.K. Pritchard, M.T. Seielstad, A. Perez-Lezaun, M.W. Feldman, Population growth of human Y chromosomes: a study of Y chromosome microsatellites, Mol. Biol. Evol. 16 (1999) 1791–1798.
[9] C.P. Robert, The Bayesian Choice, second ed., Springer-Verlag, New York, 2001.
[10] R. Wilkinson, Approximate Bayesian computation (ABC) gives exact results under the assumption of model error, arXiv:0811.3355v1 [stat.CO], 2008.