Geometric summary statistics for ABC model choice between hidden Gibbs random fields

Julien Stoehr, Pierre Pudlo, Lionel Cucala

Geometric summary statistics for ABC model choice between hidden Gibbs random fields

Julien Stoehr*, Pierre Pudlo* and Lionel Cucala*

February 6, 2014

Abstract

Selecting between different dependence structures of a hidden Markov random field can be very challenging, due to the intractable normalizing constants in the likelihoods and the sum over all possible latent random fields. Approximate Bayesian Computation (ABC) algorithms provide a model choice method in the Bayesian paradigm. The scheme compares the observed data and many numerical simulations through summary statistics. When the Gibbs random field is directly observed, Grelaud et al. (2009) exhibit sufficient summary statistics that immediately guarantee the consistency of the ABC algorithm. But, when the random field is hidden, those statistics are not sufficient anymore. We provide new summary statistics based on the geometry of the image, more precisely a clustering analysis of pixels. To assess their efficiency, we also derive a conditional misclassification rate evaluating the power of ABC algorithms which may be of independent interest.

Keywords: Approximate Bayesian Computation, model choice, hidden Gibbs random fields, summary statistics, misclassification rate

1 Introduction

Gibbs random fields are polymorphous statistical models that are useful to analyse different types of spatially correlated data such as shades of grey on a rectangular grid of pixels for digital images. The autobinomial model (Besag, 1974) and its particular case the Potts one, is used to describe the spatial distribution of discrete random variables (shades of grey or colors) on a lattice (grid of pixels).

Despite the wide range of applications, Gibbs random fields present major difficulties from an inference point of view (Grelaud et al., 2009, Friel, 2012, 2013, Everitt, 2012, Cucala and Marin, 2013). Selecting between two different dependence structures can be very challenging, due to the normalizing constants in the likelihoods which are intractable for all but very small lattices. When the Gibbs random field is observed, this problem can be termed "doubly intractable" since the likelihood, but also, the posterior probability are not available. Here, since the random field is not directly observed like for hidden Markov models, the complexity of the problem grows to become "triply intractable". Currently, there exists very few works on that model choice problem (Forbes and Peyrard, 2003, Friel et al., 2009, Cucala and Marin, 2013).

Reeves and Pettitt (2004) and Friel and Rue (2007) present a recursive algorithm to compute the normalizing constants for bicolor images when the number of pixels is small. But for larger images, the algorithm is not manageable anymore and approximate methods should be used.

*I3M – UMR CNRS 5149, Université Montpellier II, France

Approximate Bayesian Computation (ABC) (Rubin, 1984, Tavaré et al., 1997, Pritchard et al., 1999, Marin et al., 2012, Baragatti and Pudlo, 2014) can provide a model choice method in the Bayesian paradigm (Grelaud et al., 2009, Didelot et al., 2011). The algorithm might be seen as an accept-reject method rejecting simulated datasets which are far from the observed data. Indeed, we simulate jointly a model index m from a prior $\pi(\cdot)$, a parameter value θ_m from the prior $\pi_m(\cdot)$ and a dataset y from the Gibbs distribution $f_m(\cdot|\theta_m)$. The scheme compares the observed data y^{obs} with many such numerical simulations y through summary statistics $S(y)$. In that case, we obtain a Monte Carlo approximation of the posterior probabilities of each model, namely

$$\mathbb{P}(\mathcal{M} = m|S(y^{\text{obs}})) \propto \int f_m(y^{\text{obs}}|\theta_m)\pi_m(\theta_m)\mathrm{d}\theta_m.$$

Variational approches to Bayesian model selection have been presented in the literature. We mention for example the work of Caimo and Friel (2013), based on a reversible jump Markov chain Monte Carlo algorithm (Caimo and Friel, 2011), in the context of exponential random graph models.

It is straightforward to show that the consistency of the ABC algorithm holds if the summary statistics $S(y)$ are sufficient with respect to the model choice problem. Otherwise summary statistics have to be chosen carefully (Robert et al., 2011, Marin et al., 2014). When the Gibbs random field is directly observed, Grelaud et al. (2009) exhibited sufficient summary statistics directly based on the potential function of the Gibbs distributions. Thus,

$$\mathbb{P}(\mathcal{M} = m|y^{\text{obs}}) = \mathbb{P}(\mathcal{M} = m|S(y^{\text{obs}})), \text{ for all } m.$$

However when the random field is hidden, this property does not hold anymore due to lack of sufficiency of the above summary statistics. Robert et al. (2011) showed that the use of ABC with insufficient summary statistics to Bayesian model choice problems can lead to inaccurate results: it may pick the wrong model. Although, the claim of Robert et al. (2011) should be nuanced since the example detailed in the paper is very specific and despite this first warning, the authors developped some further results. Indeed Marin et al. (2014) provide generic conditions – much weaker than sufficiency – implying the consistency of the ABC procedure. Beside those results, Blum (2010a) and Fearnhead and Prangle (2012) have shown that the quality of this approximation decreases with the dimension of $S(y)$. Thus, a balance should be found between low dimensional and informative set of summary statistics. ABC then appears as a major method to manage intractable likelihood. Prangle et al. (2013) proposed a scheme to build automatically statistics $S(y)$ with good properties but based on a first ABC run that can be time consuming. Whereas some other works have been done to select the vector $S(y)$ among a large set of summary statistics (Blum, 2010b, Blum et al., 2013), we aim here to select the most informative statistics within many low dimensional vectors.

This paper sets new summary statistics based on the geometry of the image, more precisely connected components of pixels, and assesses their efficiency for an ABC model choice between hidden random fields models. Since the conditions given by Marin et al. (2014) to insure consistency are very difficult to check in practice, we provide a new method based on a conditional misclassification rate to validate the whole procedure.

The paper is organized as follows: Section 2 briefly presents hidden Potts model and the ABC algorithm for model selection. Then, Section 3 introduces our geometric summary statistics and the estimation procedure of the conditional misclassification rate used to assess their efficiency. Finally, Section 4 presents some numerical results.

2 ABC model choice for hidden Gibbs random fields

Let us begin with an introduction on ABC model choice methods. We focus here on the specific case of hidden Potts model.

2.1 Hidden Potts model

Consider a finite set of sites $\mathcal{S} = \{0, \dots, N-1\}$. At each site $i \in \mathcal{S}$ we observe a random variable x_i taking values in $\{0, \dots, K-1\}$, where K is a known integer. Let x denote the random process $(x_0, \dots, x_{N-1})$. When modeling an image, the site i is called a pixel, N is the number of pixels, K is the number of colors and x_i corresponds to the color of the pixel i. Consider an undirected graph $\mathcal{G}$ which defines an adjacency relation on the set of sites $\mathcal{S}$: i and j are adjacent if and only if there is an edge between i and j in the graph $\mathcal{G}$.

The Potts model with parameter β defined on the graph $\mathcal{G}$ has density

$$\pi(x|\mathcal{G},\beta) = \frac{1}{Z(\mathcal{G},\beta)} \exp\left(\beta \sum_{i \overset{\mathcal{G}}{\sim} j} \mathbb{1}\{x_i = x_j\}\right),$$

where the sum $i \overset{\mathcal{G}}{\sim} j$ ranges the edges of $\mathcal{G}$. The normalizing constant $Z(\mathcal{G},\beta)$ – intractable except for small value of N and K –, also called the partition function is

$$Z(\mathcal{G},\beta) = \sum_{x \in \mathcal{X}} \exp\left(\beta \sum_{i \overset{\mathcal{G}}{\sim} j} \mathbb{1}\{x_i = x_j\}\right). \tag{2.1}$$

But in our case, the random field x is not directly observed. We assume that the observations $y = (y_i)_{i \in \mathcal{S}}$ are conditionally independent given x. The conditional distribution of the observation is given by $\pi(y|x) = \prod_i P(y_i|x_i)$, where P is the noise distribution. The one we set is given by

$$P_\alpha(y_i|x_i) = \frac{\exp\{\alpha(2\mathbb{1}\{x_i = y_i\} - 1)\}}{\exp(\alpha) + (K-1)\exp(-\alpha)},$$

and extends the one proposed by Everitt (2012) when $K = 2$. Note that conditionally on x_i and $y_i \neq x_i$, y_i is then uniformly distributed over $\{0, \dots, K-1\} \setminus \{x_i\}$.

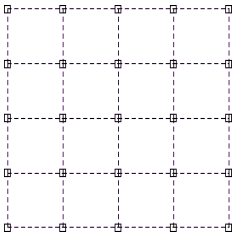
Hence the likelihood of the hidden Potts model with parameter β on the graph $\mathcal{G}$ and noise distribution P_α, denoted HPM$(\mathcal{G}, \alpha, \beta)$, is given by

$$f(y|\alpha,\beta,\mathcal{G}) = \sum_{x \in \mathcal{X}} \pi(x|\mathcal{G},\beta)\, \pi_\alpha(y|x).$$

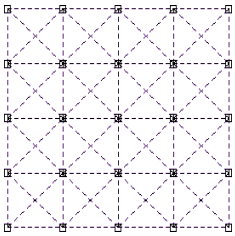
2.2 Bayesian model choice

Our purpose is to select the hidden Gibbs random field that better fits a given picture y^{obs} within two models $\mathcal{M}_4$ and $\mathcal{M}_8$ where $\mathcal{M}_4$ is a HPM$(\mathcal{G}_4, \alpha, \beta)$ with a given prior π_4 on (α, β) and $\mathcal{M}_8$ is a HPM$(\mathcal{G}_8, \alpha, \beta)$ with a given prior π_8 on (α, β) defined more precisely in the following paragraph. The edges of the graphs $\mathcal{G}_4$ and $\mathcal{G}_8$ respectively link the four and the eight directly adjacent neighbours (see Figure 1).

Selecting a model m is driven by the posterior probability of both models $\mathcal{M}_4$ and $\mathcal{M}_8$. Say that $\Theta_m \subset \mathbb{R}^2$ is the parameters space of model m, $m \in \{\mathcal{M}_4, \mathcal{M}_8\}$. The Bayesian analysis of a dataset y^{obs} requires defining



(a)



(b)

Figure 1: Underlying graph $\mathcal{G}$ of model HPM$(\mathcal{G}, \alpha, \beta)$. (a) The four closest neighbour graph $\mathcal{G}_4$ defining model $\mathcal{M}_4$. (b) The eight closest neighbour graph $\mathcal{G}_8$ defining model $\mathcal{M}_8$

a prior distribution on the index model $\pi(\mathcal{M} = m)$ and prior distributions $\pi_m(\theta)$ on parameter $\theta = (\alpha, \beta) \in \Theta_m$. Thus, the posterior probability of model m is given by

$$\pi(m|y) \propto \pi(\mathcal{M} = m) \int_{\Theta_m} f_m(y|\theta)\pi_m(\theta)\mathrm{d}\theta.$$

In our case,

$$\pi(m|y) \propto \int_{\Theta_m} \frac{1}{Z(\mathcal{G},\beta)} \frac{1}{C(\alpha)} Q(y, \mathcal{G}, \theta)\pi_m(\theta)\mathrm{d}\theta, \tag{2.2}$$

where $Z(\mathcal{G},\beta)$ is the partition function defined by (2.1), $C(\alpha) = \{\exp(\alpha) + (K-1)\exp(-\alpha)\}^N$ and $Q(y, \mathcal{G}, \theta) = \sum_{x \in \mathcal{X}} \exp\left\{\beta \sum_{i \overset{\mathcal{G}}{\sim} j} \mathbb{1}\{x_i = x_j\} + \alpha\left(2\sum_{i=0}^{N-1} \mathbb{1}\{x_i = y_i\} - N\right)\right\}$.

To estimate the posterior probability, we thus face a triple intractable problem. The summation over all the latent fields $x \in \mathcal{X}$, that appears in $Z(\mathcal{G},\beta)$ and $Q(y, \mathcal{G}, \theta)$, generally involves too many terms (K^n) to be calculated explicitly or numerically and the integral on Θ_m might be hard to manage. Thus (2.2) can not be computed for realistic number of sites, instead approximate methods have to be used(see Section 2.3). In the same spirit as the forward-backward algorithm for hidden Markov model (Rabiner, 1989), Friel and Rue (2007) provide a recursive algorithm based on Reeves and Pettitt (2004) to compute (2.2). However, the algorithm is manageable only if the total number of pixels does not exceed 400 pixels with the simplest 2-color model ; in addition, the complexity of the algorithm grows exponentially with the number of colors.

2.3 Approximate Bayesian Computation for model choice

Since its introduction by Tavaré et al. (1997), ABC algorithms have aimed at estimating posterior probabilities in settings where the likelihood is intractable. We refer the reader to published reviews on ABC (Marin et al., 2012, Baragatti and Pudlo, 2014) and focus on the model choice procedure.

ABC simulates data y for many parameter values θ_m under each model m (Algorithm 1) which are then compared to the observed data y^{obs}. Algorithm 1 produces the ABC reference table, that is a large set of particules (m, θ, y) drawn from the Bayesian model. We note it $\mathcal{D} = \left\{\left(m^j, \theta^j, S\left(y^j\right)\right) \mid j = 1, \ldots, n_{\text{REF}}\right\}$.

Algorithm 1: Simulation of ABC reference table

Result: A reference table of size n_{REF}
for $j \leftarrow 1$ **to** n_{REF} **do**
  draw m from the prior π;
  draw $\theta = (\alpha, \beta)$ from the prior π_m;
  draw y from the likelihood $f_m(\cdot|\theta)$;
  compute $S(y)$;
  set $(m^j, \theta^j, S(y^j)) \leftarrow (m, \theta, S(y))$;
end

Among the particles of $\mathcal{D}$, ABC keeps the ones from models under which $S(y)$ is close to $S(y^{\text{obs}})$ in the sense of a distance ρ, where S is a vector of summary statistics, see Algorithm 2.

Algorithm 2: ABC model choice algorithm returning n_{POST} particles

Data: A reference table $\mathcal{D}$ of size n_{REF},
A function $S(\cdot)$,
A number of particles selected n_{POST}.
Result: A sample of size n_{POST} distributed according to the ABC posterior approximation

sort $\mathcal{D}$ according to $\rho(S(y^j), S(y^{\text{obs}}))$;
keep the n_{POST} first particles;

Denore ϵ the quantile of the distance, that is the disance of the $n^{\text{th}}_{\text{POST}}$ closest particle. The accepted particles m^j at the end of the Algorithm 2 are distributed (Biau et al., 2013) according to

$$\pi\left(m \,\middle|\, \rho(S(y^j), S(y^{\text{obs}})) < \epsilon\right),$$

and

$$\widehat{\pi}_{ABC}\left(m \,\middle|\, S\left(y^{\text{obs}}\right)\right) = \frac{\sum_{j=1}^{n_{\text{REF}}} \mathbb{1}\{m^j = m, \rho(S(y^j), S(y^{\text{obs}})) < \epsilon\}}{\sum_{j=1}^{n_{\text{REF}}} \mathbb{1}\{\rho(S(y^j), S(y^{\text{obs}})) < \epsilon\}}$$

is a Monte Carlo estimate of the posterior probability of model $\mathcal{M}_m$.

Note that when $\epsilon \to 0$, ABC cannot recover anything better than $\pi\left(m \,\middle|\, S(y) = S(y^{\text{obs}})\right)$. For sufficient summary statistics S, this has no consequence on the approximation of the posterior distribution since $\pi\left(m \,\middle|\, S(y) = S(y^{\text{obs}})\right) = \pi\left(m \,\middle|\, y = y^{\text{obs}}\right)$. When the Gibbs random field is directly observed, Grelaud et al. (2009) show that the concatenation of sufficients summary statistics R for each model is sufficient for the model choice. They are defined by

$$R(\mathcal{G}_4, y) = \sum_{i \overset{\mathcal{G}_4}{\sim} j} \mathbb{1}\{y_i = y_j\} \text{ and } R(\mathcal{G}_8, y) = \sum_{i \overset{\mathcal{G}_8}{\sim} j} \mathbb{1}\{y_i = y_j\}, \tag{2.3}$$

and represent the number of neighbouring pixels exhibiting the same color in each model. However those summary statistics are not sufficient anymore if the Gibbs random field is hidden. Thus, we introduce new summary statistics based on pixel clusters to reduce the lack of sufficiency of R.

3 Pixels cluster and conditional misclassification rate

The following part sets new summary statistics based on pixel clusters and asses their efficiency thanks to a new validation procedure.

3.1 Some geometrics summary statistics

Consider a picture y and a graph $\mathcal{G}$. We define the graph $\Gamma(\mathcal{G}, y)$ induced by $\mathcal{G}$ on y as follows: there is an edge between pixels i and j in $\Gamma(\mathcal{G}, y)$ if and only if there is an edge between i and j in $\mathcal{G}$ and if both pixels share the same color, that is $y_i = y_j$.

Subsequently, we introduce a set of four summary statistics based on pixel clusters. We remind that the connected components of a graph Γ are all the subgraphs formed by the equivalence classes of the adjacency relation of Γ. In other words, a connected component of a graph Γ is a subgraph of Γ in which any two pixels are connected to each other by a path, and which is connected to no other vertices of Γ. We then define two summary statistics taking values in $\mathbb{N}$: $T(\mathcal{G}, y)$ is the number of connected components of $\Gamma(\mathcal{G}, y)$ and $U(\mathcal{G}, y)$ is the size of the biggest component of $\Gamma(\mathcal{G}, y)$. Thus, given any picture y, the set of geometric summary statistics we consider is $T(\mathcal{G}_4, y)$, $U(\mathcal{G}_4, y)$, $T(\mathcal{G}_8, y)$ and $U(\mathcal{G}_8, y)$. See Figure 2 for an example on a bicolor (red/black) picture y.

$\Gamma(\mathcal{G}_4, y)$ $\Gamma(\mathcal{G}_8, y)$

Figure 2: The induced graph $\Gamma(\mathcal{G}_4, y)$ and $\Gamma(\mathcal{G}_8, y)$ on a given bicolor image y of size 5×5. The four summary statistics on y are thus $T(\mathcal{G}_4, y) = 7$, $U(\mathcal{G}_4, y) = 12$, $T(\mathcal{G}_8, y) = 4$ and $U(\mathcal{G}_8, y) = 16$

This paper compares three nested sets of summary statistics $S_1(y) \subset S_2(y) \subset S_3(y)$ in order to see if we get more information on the hidden random field by adding some geometric summary statistics. The first set is $S_1(y) = \{R(\mathcal{G}_4, y); R(\mathcal{G}_8, y)\}$ where R is defined by (2.3). The second set is $S_2(y) = S_1(y) \cup \{T(\mathcal{G}_4, y); T(\mathcal{G}_8, y)\}$ and the third one is $S_3(y) = S_2(y) \cup \{U(\mathcal{G}_4, y); U(\mathcal{G}_8, y)\}$. Very informally, we believe that the geometry of Potts models with graphs $\mathcal{G}_4$ and $\mathcal{G}_8$ are different in term of such clusters, and that the noise does not destroy this signal.

3.2 Conditional or local misclassification rate

It is straitghtforward looking at Algorithm 2 that each set $S_\ell(y)$ defines a different ABC procedure. The question is then: what is the most relevant set of summary statistics for our model choice purpose? Given a reference table and a set of summary statistics, for any new observed data y, ABC can predict the index of the model with a maximum ABC *a posteriori* rule, say $\widehat{m}_\ell(y)$. Facing some observed data y^{obs}, the user of an ABC algorithm is interested in the error he commits believing $\widehat{m}_\ell(y^{\text{obs}})$ computed with the reference table he has at his disposal, and a given set of summary statistics $S_\ell(y)$. In this paragraph, we define an error,

say $\tau_\ell(y^{\text{obs}})$, or more precisely a misclassification rate given y^{obs}, in order to evaluate the local efficiency of ABC. Then we propose an estimation algorithm of this conditional error rate.

In what follows, we work given an ABC reference table $\mathcal{D}$. Then, with any new data y^* in $\mathscr{Y}$, the ABC algorithm with a set of summary statistics $S_\ell(y)$ can predict an index model using the value of m that maximises the ABC estimates of the posterior probability of model $\mathscr{M}_m$. The predicted index might be seen as the following, perfectly deterministic function of y^*, namely

$$\widehat{m}_\ell(y^*) = \arg\max_m \widehat{\pi}_{\text{ABC}}(m|S_\ell(y^*)). \tag{3.1}$$

First, we introduce the following function of m^* and $y^* \in \mathscr{Y}$ as

$$\tau(m^*, S_\ell(y^*)) = \mathbb{P}\Big(\widehat{m}_\ell(Y) \neq m^* \Big| \rho\left(S_\ell(Y), S_\ell(y^*)\right) \leq \eta\Big) \tag{3.2}$$

where $\mathbb{P}$ denotes an integral when Y is distributed according to the marginal distribution (in y) of the Bayesian model. We shall remark here that the radius η of the neighborhood of $S_\ell(y^*)$ can be fixed independently of the quantile ϵ used in Algorithm 2.

Actually, we can see the prediction of m as a machine learning problem. Assume $\eta = 0$, the well known *misclassification rate* used in machine learning to estimate the accuracy of $\widehat{m}_\ell$ is then

$$\tau_\ell = \sum_m \iint \tau(m, S_\ell(y))\, \pi(\mathcal{M} = m)\pi_m(\theta) f_m(y|\theta)\, \mathrm{d}\theta \mathrm{d}y, \tag{3.3}$$

that is to say the expected value of $\tau(m, S_\ell(y))$ with respect to the Bayesian model. We advocate here in favor of a conditional version of this error given $S_\ell(y^*)$, namely,

$$\tau(S_\ell(y^*)) = \sum_m \pi(m|S_\ell(y^*))\tau(m, S_\ell(y^*)). \tag{3.4}$$

Explicitly we integrate $\tau(m, S_\ell(y))$ over the distribution of the Bayesian model given $S_\ell(y) = S_\ell(y^*)$. The marginal in m of this conditional distribution is clearly $\pi(m|S_\ell(y^*))$.

Algorithm 3: Estimation of $\tau(m^*, S_\ell(y^*))$ given an ABC reference table $\mathcal{D}$

Data: an ABC reference table $\mathcal{D}$,
a function $S(\cdot)$,
a number of particles selected n_{NEI}
Result: $\widehat{\tau}(m^*, S(y^*))$ defined in (3.2)

choose randomly a particule (m^i, θ^i, y^i) in $\mathcal{D}$;
set $(m^*, \theta^*, y^*) \leftarrow (m^i, \theta^i, y^i)$;
remove (m^i, y^i) from $\mathcal{D}$;
foreach *particle* y^j *among the* n_{NEI} *nearest particles from* y^* **do**
  compute $\widehat{m^j}$ with Algorithm 2 using the summary statisctics S_ℓ and $\mathcal{D}$;
end
compute $\widehat{\tau}(m^*, S(y^*)) \leftarrow \dfrac{1}{n_{\text{NEI}}} \displaystyle\sum_{n_{\text{NEI}} \text{ nearest particles}} \mathbb{1}\left\{\widehat{m^j} \neq m^*\right\}$;

Algorithm 3 estimates the function $\tau(m^*, S_\ell(y^*))$ with the ABC reference table at hand and provides realizations of $\widehat{\tau}(m^*, S_\ell(y^*))$ when (m^*, y^*) is distributed with respect to the Bayesian model. The global

accuracy of the ABC model choice procedure based on the ℓ^{th} set of summary statistics might be evaluated with $\widehat{\tau}_\ell$, the average of the realizations provided by Algorithm 3. The estimate $\widehat{\tau}_\ell$ is in the same spirit as a cross-validation error rate (Hastie et al., 2009) using a non exhaustive leave-one-out procedure. However the estimate $\widehat{\tau}(S_\ell(y^*))$ of the conditional error might be a better evaluation of the error when we are interested in conducting a Bayesian analysis of only one data set y^* since it is local. Comparing those error rates given y^* (or various projection of y^* with summary statistics) provides a procedure to select the relevant summary statistics that are efficient locally around y^*.

The goal we strive here is to construct a predictor of the index model given y based on a collection of predictors (3.1) for various summary sets and also to estimate its error with (3.4). Selecting the best set $\widehat{\ell}(y^\star)$ leads to a procedure based on the Bayesian machinery, as empirical Bayesian estimators for instance. Indeed, $\widehat{\pi}_{\text{ABC}}(m|S_{\widehat{\ell}(y^\star)}(y^\star))$ uses the data twice: a first one to calibrate the set of summary statistics, and a second one to compute the ABC posterior. But as mentioned above, no ABC procedure comparing data sets through non sufficient statistics can claim approximating the true posterior $\pi(m|y)$ (Robert et al., 2011). The sole guarantee that remains is that ABC will pick the correct model if we provide enough data (Marin et al., 2014) and the proposed predictor is reliable.

4 Experiments results

Using two numerical experiments, one for $K = 2$ colors and one for $K = 16$ colors, we evaluate the efficiency of the geometric summary statistics from a local and global point of view thanks to the previous misclassification rates.

tab:example

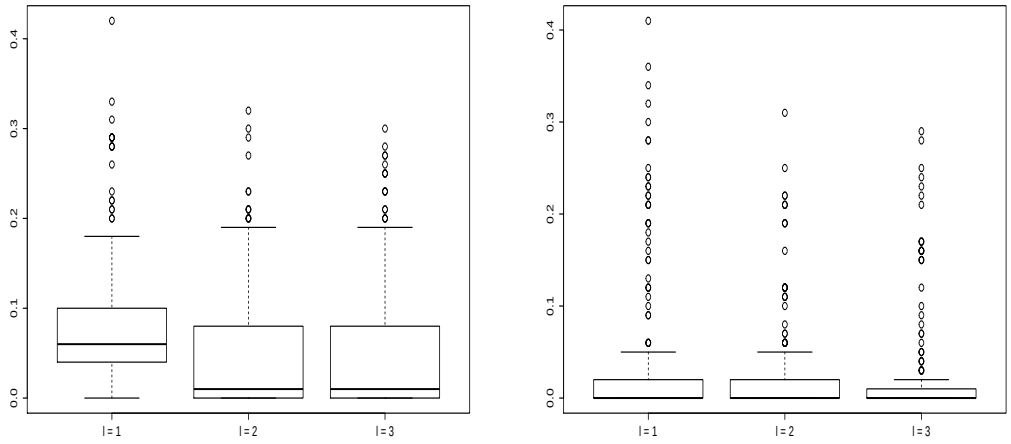
Table 1: Prior distribution and experiment settings

	First experiment with $K = 2$ colors		**Second experiment with $K = 16$ colors**	
Model m (prior probability)	$\mathcal{M}_4$ (0.5)	$\mathcal{M}_8$ (0.5)	$\mathcal{M}_4$ (0.5)	$\mathcal{M}_8$ (0.5)
Parameter priors $\pi_m(\alpha,\beta)$	$\alpha \sim \mathcal{U}(0.42; 2.3)$ $\beta \sim \mathcal{U}(0; 1)$	$\alpha \sim \mathcal{U}(0.42; 2.3)$ $\beta \sim \mathcal{U}(0; 0.35)$	$\alpha \sim \mathcal{U}(1.78; 4.8)$ $\beta \sim \mathcal{U}(0; 2.4)$	$\alpha \sim \mathcal{U}(1.78; 4.8)$ $\beta \sim \mathcal{U}(0; 1)$
Picture size	100×100			
MCMC iterations to generate Potts pictures	2×10^4		4×10^4	
Simulations in the ABC reference table	4×10^5		4×10^5	

We simulated the latent fields with independent runs of the clever MCMC algorithm of Swendsen and Wang (1987). Dependence between pixels is growing with β, such that for large enough β, pictures becomes unicolor (a phase transition occurs). We tuned the upper bound on the prior of β in order to avoir this phenomenon. Moreover, the noise model changes the color of a random amount of pixels, and the prior bounds on α were tuned to change less than 30% of pixels on average.

Given an ABC reference table $\mathcal{D}$, simulated following Table 1, we study the behaviour of the different procedures. Algorithm 3 was run one hundred times for each set S_ℓ with $n_{\text{POST}} = n_{\text{NEI}} = 100$. We shall remark that the following results present two different aspects but both show that the geometric summary statistics improve mostly the prediction given by the ABC algorithm. The integrated misclassification rates given in the Table 2 indicate the general behaviour of the procedure whereas Figure 4 illustrates a local comparison.

Before giving further details, let us point out the suitability of such comparisons in this framework. Considering the observation of Robert et al. (2011) and the inability to check conditions of Marin et al. (2014), one might object that, if the new added summary statistics are not sufficient, we are wasting time since they can give inacurrate results anyway. We would agree if the procedures were mostly picking the wrong model, that is if the error rates $\widehat{\tau}(S_\ell(y))$ defined in (3.4) were high. However, Figure 3 representing the distribution of $\widehat{\tau}(m, S_\ell(y))$ for each procedure $\ell = \{1; 2; 3\}$ shows that the probability of selecting the wrong model is small and therefore we are not trying to compare useless procedures. We shall note also that when we add the geometric summary statistics the error rate $\widehat{\tau}(m, S_\ell(y))$ is indeed decreasing and may even go to zero.



(a) (b)

Figure 3: Comparison of the distributions of $\widehat{\tau}(m, S_\ell(y))$ defined in (3.2) for (a) K =2 colors and (b) K =16 colors.

Let us begin with the generally adopted point of view. We are interested in the averages of the boxplots of Figure 3 which estimate the integrated misclassification rates τ_ℓ defined in (3.3). Results are presented in Table 2. The main point to notice is that the integrated error is divided by $3/2$ for both cases when we add geometric summary statistics.

In addition, the integrated misclassification rate is not the most interesting error to consider here. Any user of an ABC algorithm is mainly interested in the best choice of $S_\ell(y)$ he can achieve. In other words, he seeks the set of summary statistics that has the best local behaviour around y^{obs}. The (conditional) misclassification rates are represented in Figure 4. The main drawback is to find a space where we can represent the local error rates for the different observations we have at hand. $\mathcal{Y}$ is clearly not good due to its high dimension. We thus project all the pictures in a common space with $R(\mathcal{G}_4, y)$. This is the x-axis of Figure 4. Replacing $R(\mathcal{G}_4, y)$ by $R(\mathcal{G}_8, y)$ does not change anything since those two summary statistics are

Table 2: Estimation of the integrated misclassification rates $\widehat{\tau}_\ell$

Summary statistics	$\ell = 1$	$\ell = 2$	$\ell = 3$
Error for $K = 2$ colors	0.082	0.049	0.053
Error for $K = 16$ colors	0.038	0.027	0.026

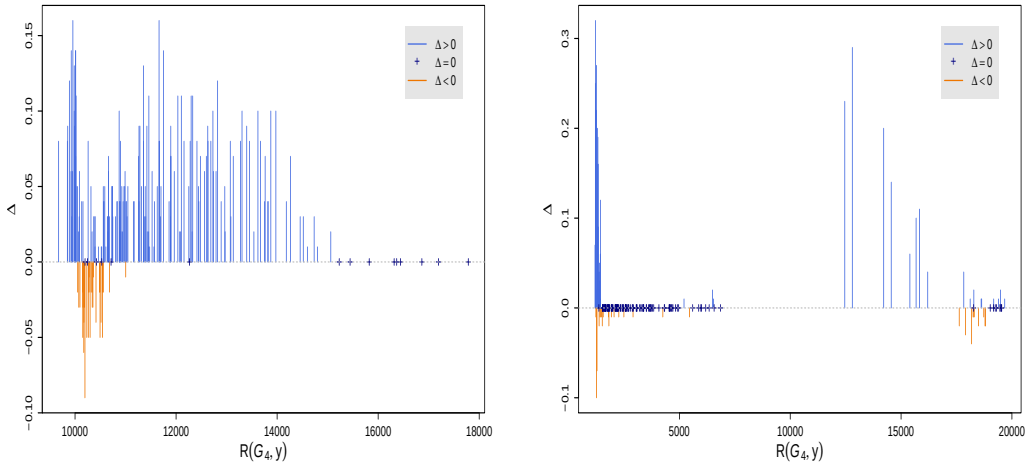
The table contains the estimation of the integrated misclassification rates τ_ℓ defined by (3.3). The estimator $\widehat{\tau}_\ell$ is the average of the distributions of $\widehat{\tau}(m, S_\ell(y))$ presented in Figure 3

strongly correlated. The y-axis of Figure 4 denoted Δ is defined as the relative benefit when adding some geometric summary statistics. It compares by using the conditional misclassification rate $\widehat{\tau}(S_1(y))$ with the $\widehat{\tau}(S_\ell(y))$ for $\ell = 2, 3$

$$\Delta = \max\left\{\widehat{\tau}(S_1(y)) - \widehat{\tau}(S_2(y)), \widehat{\tau}(S_1(y)) - \widehat{\tau}(S_3(y))\right\}.$$

The quantity Δ is thus positive when the geometric statistics improve the ABC procedure and negative otherwise.

Figure 4 shows how much the efficiency is locally dependent. Even if the procedures with $S_2(y)$ and $S_3(y)$ are globally better there exists pictures in $\mathscr{Y}$ for which $\Delta < 0$. In those cases, the geometric statistics do not manage to provide relevant information regarding the model choice issue. But in the others, most common cases, they actually do.



(a) (b)

Figure 4: Comparison of ABC procedures for the three sets of summary statistics S_ℓ with (a) K =2 colors and (b) K =16 colors. $\Delta > 0$ when the geometric statistics improve the ABC procedure and $\Delta < 0$ otherwise

5 Conclusion

The model choice issue between hidden Potts models is one of the interesting statistical setting facing a triple intractable problem. In the paper, we have shown the pertinence of ABC in this context. Moreover, we have presented a new class of statistics that summarize images using pixel clusters. And, we have exhibited their efficiency in ABC model choice procedures to answer the intractable problem. To this aim, we derived a new local error rate, and should like to highlight here that the ability to evaluate the local behaviour of the procedure presents a wider interest than just comparing the performances of algorithms. Indeed, it allows us to select the most relevant set of summary statistics.

Geometric summary statistics based on pixel clusters are relevant for the problem of model selection between hidden Gibbs random fields. Despite the gain being most significant when the summary statistics $T(\mathcal{G}, y)$ and $U(\mathcal{G}, y)$ distinguish the models, there are some given parameters for which they bring little or no information to select the model (Figure 4). We expect that happens for geometries that are too complex to be described only by the number of connected components and the size of the biggest one. Put in another way, the correlation between all the summary statistics is such that we add little or no extra information or maybe even noise in adding the summary based on pixel clusters. Our future work, then, aims to study more precisely the distribution of the size of the connected components in order to find some other geometric summary statistics which might improve the information of the procedure.

A natural extension of this work would be to consider a more realistic noise distribution such as the Gaussian one (Cucala and Marin, 2013). The geometric summary statistics we introduced in our paper can be extended to such a continuous framework if we consider that two neighbouring pixels share the same color if the difference of their grey levels is less than a given threshold. This might also be the subject of our future work.

Acknowledgments

We are grateful to Jean-Michel Marin for his feedback and support.

References

Meïli Baragatti and Pierre Pudlo. An overview on Approximate Bayesian Computation. *ESAIM: Proc.*, 44: 291–299, 2014.

Julian Besag. Spatial interaction and the statistical analysis of lattice systems (with Discussion). *Journal of the Royal Statistical Society. Series B (Methodological)*, 36(2):192–236, 1974.

Gérard Biau, Frédéric Cérou, and Arnaud Guyader. New insights into Approximate Bayesian Computation. *Annales de l'Institut Henri Poincaré (B) Probabilits et Statistiques, in press*, 2013.

Michael G. B. Blum. Approximate Bayesian Computation: A Nonparametric Perspective. *Journal of the American Statistical Association*, 105(491):1178–1187, 2010a.

Michael G. B. Blum, Maria Antonieta Nunes, Dennis Prangle, and Scott A. Sisson. A Comparative Review of Dimension Reduction Methods in Approximate Bayesian Computation. *Statistical Science*, 28(2): 189–208, 2013.

Michael G.B. Blum. Choosing the Summary Statistics and the Acceptance Rate in Approximate Bayesian Computation. In *Proceedings of COMPSTAT'2010*, pages 47–56. Springer, Physica-Verlag HD, 2010b. ISBN 978-3-7908-2603-6.

Alberto Caimo and Nial Friel. Bayesian inference for exponential random graph models. *Social Networks*, 33(1):41–55, 2011.

Alberto Caimo and Nial Friel. Bayesian model selection for exponential random graph models. *Social Networks*, 35(1):11 – 24, 2013.

Lionel Cucala and Jean-Michel Marin. Bayesian Inference on a Mixture Model With Spatial Dependence. *Journal of Computational and Graphical Statistics*, 22(3):584–597, 2013.

Xavier Didelot, Richard G. Everitt, Adam M. Johansen, and Daniel J. Lawson. Likelihood-free estimation of model evidence. *Bayesian Analysis*, 6(1):49–76, 2011.

Richard G. Everitt. Bayesian Parameter Estimation for Latent Markov Random Fields and Social Networks. *Journal of Computational and Graphical Statistics*, 21(4):940–960, 2012.

Paul Fearnhead and Dennis Prangle. Constructing summary statistics for approximate Bayesian computation: semi-automatic approximate Bayesian computation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(3):419–474, 2012.

Florence Forbes and Nathalie Peyrard. Hidden Markov random field model selection criteria based on mean field-like approximations. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25 (9):1089–1101, 2003.

N. Friel, A. N. Pettitt, R. Reeves, and E. Wit. Bayesian Inference in Hidden Markov Random Fields for Binary Data Defined on Large Lattices. *Journal of Computational and Graphical Statistics*, 18(2):243–261, 2009.

Nial Friel. Bayesian inference for Gibbs random fields using composite likelihoods. In *Simulation Conference (WSC), Proceedings of the 2012 Winter*, pages 1–8, 2012.

Nial Friel. Evidence and Bayes Factor Estimation for Gibbs Random Fields. *Journal of Computational and Graphical Statistics*, 22(3):518–532, 2013.

Nial Friel and Håvard Rue. Recursive computing and simulation-free inference for general factorizable models. *Biometrika*, 94(3):661–672, 2007.

Aude Grelaud, Christian P. Robert, Jean-Michel Marin, François Rodolphe, and Jean-François Taly. ABC likelihood-free methods for model choice in Gibbs random fields. *Bayesian Analysis*, 4(2):317–336, 2009.

Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The elements of statistical learning: data mining, inference and prediction*, chapter 7. Springer Series in Statistics. Springer, 2nd edition, 2009. ISBN 9780387848587.

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

Jean-Michel Marin, Natesh S. Pillai, Christian P. Robert, and Judith Rousseau. Relevant statistics for Bayesian model choice. *To appear in the Journal of the Royal Statistical Society, Series B*, 2014.

Dennis Prangle, Paul Fearnhead, Murray P. Cox, Patrick J. Biggs, and Niguel P. French. Semi-automatic selection of summary statistics for ABC model choice. *Statistical Applications in Genetics and Molecular Biology*, pages 1–16, 2013.

Jonathan K. Pritchard, Mark T. Seielstad, Anna Perez-Lezaun, and Marcus W. Feldman. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Molecular Biology and Evolution*, 16(12):1791–1798, 1999.

Lawrence R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

Robert Reeves and Anthony N. Pettitt. Efficient recursions for general factorisable models. *Biometrika*, 91 (3):751–757, 2004.

Christian P. Robert, Jean-Marie Cornuet, Jean-Michel Marin, and Natesh S. Pillai. Lack of confidence in approximate Bayesian computation model choice. *Proceedings of the National Academy of Sciences*, 108 (37):15112–15117, 2011.

Donald B. Rubin. Bayesianly justifiable and Relevant Frequency Calculations for the Applied Statistician. *The Annals of Statistics*, 12(4):1151–1172, 1984.

Robert H. Swendsen and Jian-Sheng Wang. Nonuniversal critical dynamics in Monte Carlo simulations. *Physical Review Letters*, 58(2):86–88, 1987.

Simon Tavaré, David J. Balding, Robert C. Griffiths, and Peter Donnelly. Inferring Coalescence Times From DNA Sequence Data. *Genetics*, 145(2):505–518, 1997.