Nonparametric Bayesian estimation of level sets

Ghislaine Gayraud
Judith Rousseau

Nonparametric Bayesian estimation of level sets

Ghislaine Gayraud
CREST, Timbre J340, 3 av. P. Larousse, 92241 Malakoff Cedex, and Université de Rouen,
&
Judith Rousseau
CREST, Timbre J340, 3 av. P. Larousse, 92241 Malakoff Cedex, and Université Paris V

Abstract

We propose a Bayesian nonparametric estimation of level sets on $\mathbb{R}^2$, based on the density of n iid observations. The prior on the density function puts mass on piecewise constant functions. The pieces, on which the function is constant, are Voronoi tiles generated by a spatial point process. The values taken by the function on each tile are driven by a random Markov field. We prove that such a prior leads to a strongly consistent posterior distribution. This implies in particular that the Bayes estimate of the level set is consistent in terms of the Lebesgue measure of the symmetric difference.

We also give simulation results to study the numerical performances of such an estimate.

Résumé

Nous considérons une estimation non paramétrique et bayésienne d'ensembles de niveau dans $\mathbb{R}^2$. La loi a priori charge les fonctions constantes par morceaux. Ces fonctions sont constantes sur les cellules d'une tesselation de Voronoi engendrée par un processus ponctuel. Les valeurs prises par la fonction sur les cellules sont issues d'un champ de Markov. Nous démontrons que la loi a posteriori est alors convergente presque sûrement, ce qui implique la convergence des estimateurs bayésiens, lorsque la fonction de perte est la mesure de Lebesgue de la différence symétrique.

Nous effectuons par ailleurs des simulations afin d'étudier les performances de ce type d'estimateurs.

Key words : Bayesian nonparametric estimation, level sets, consistency, point process, Voronoi tesselation.

AMS 2000 SUBJECT CLASSIFICATION, Primary : 62C10, 62G05, 62G20; Secondary : 62M30, 65C40.

1 Introduction

In this paper, we are interested in estimating a λ–level set of a density or a density λ–contour cluster defined as the set where the density is larger than some fixed known positive constant λ. More precisely $X^{(n)} = (X_1, \dots, X_n)$, $X_i \in E$, is distributed with respect to an unknown probability P with density f, where E is a known bounded subset of $\mathbb{R}^2$. This problem has been widely studied due to its large number of statistical applications; in particular, it is used as an auxiliary tool in different statistical fields. It is related to the location of mass concentration and thus to questions concerning the multi-modality of distributions. For instance by considering different values of λ, level set estimation can be used to determine the number and the positions of modes. This can also be used in cluster analysis, since the population λ-clusters are defined as the connected components of the level set, see Hartigan (1975). Confidence regions can also be viewed as level sets and estimated as such. In econometrics, level sets are usually expressed in terms of regression functions and allow to evaluate the performance of an enterprise in terms of technical efficiency.

Hartigan (1987) in the bivariate case and Müller and Sawistzki (1991a, 1991b) in the univariate case, were the first to investigate the estimation level sets using the fact that a λ–level set of a density is the largest set S maximizing the excess mass. The excess mass is defined as $P(S) - \lambda\mu(S)$, where $\mu(S)$ denotes the Lebesgue measure of S. Both of them consider the special case of level sets defined as a finite collection of convex sets. Nolan (1991) considers, in the d-dimensional case, level sets estimates belonging to a collection of ellipsoids. In all the papers cited above, it is assumed that the underlying distribution has a density contour cluster lying in some specified class $\mathcal{C}$. From another point of view, Polonik (1995) gives some asymptotic results pointing out the relation between the behaviour of level sets estimates obtained from the empirical excess mass and the choice of the class $\mathcal{C}$. Tsybakov (1997) introduces new estimates which are based on the maximisation of local empirical excess masses instead of using, as in the previous papers, global empirical excess mass; he proves that these estimates are minimax for some specific classes.

Here, we investigate a completely different way to estimate the λ–level sets of a density since we use a Bayesian nonparametric approach. We thus define a class of prior distributions on the set of all probability distributions on E, which, we think, is appropriate to the problem at hand, i.e. the estimation of λ–level sets of a density. To our knowledge there is hardly any work done on bayesian nonparametric estimation of λ–level sets. The class of prior distributions that we are using is motivated by Arjas and Gasberra (1994), who used similar prior distributions in the setup of nonparametric curve estimation. There, curves are defined on a finite interval and parameterised by function values at a finite number of points. In their paper, the number and the locations of these points are random. This leads to an infinite-dimensional parameter space. Following from Arjas and Gasberra's paper (1994), there has been a number of papers such as Green (1995), Arjas and Heikkinen (1997, 1998, 1999), Denison, Mallick and Smith (1998), Nicholls (1998) and for a more detailed review, see Heikkinen (1998). These papers focus mainly on the definition of the prior distribution and on the implementation of the posterior distribution, via an MCMC algorithm.

Let us briefly state the problem and the framework we focus on : let us denote

$$G_\lambda = \{x \in E : f(x) \geq \lambda\}$$

the quantity of interest, where λ is a given positive number. The main purpose of this paper is to provide some reasonable Bayesian estimate of G_λ. Since G_λ is a subset of $\mathbb{R}^2$, we believe it is appropriate to construct the prior distribution using some spatial process ξ, as will be detailed in Section 2. The idea is basically the same as in Heikkinen (1998) and in Arjas and Gasbarra (1994). The prior density puts mass on functions that are piecewise constants on tiles that are constructed using a Voronoi tesselation generated by ξ. However this does not mean that the Bayesian estimates obtained from the above prior are necessarily piecewise constant. In particular posterior means are usually smoother functions.

One must note that we could use a Bayesian estimate of f, say $\hat{f}$, and then estimate G_λ by $\{x; \hat{f}(x) \geq \lambda\}$. However this would be suboptimal in terms of decision theory. Thus we consider, as an estimate of G_λ, the bayesian estimator associated with a loss function, defined as some specific distance between two sets, and with the prior described in Section 2.

Moreover, since we are considering a nonparametric approach, it would be unrealistic to try to construct a purely subjective prior distribution. Therefore it is necessary to assess the convergence of the posterior distribution as well as the convergence of the estimate we are considering, to make sure that the prior we are choosing is reasonable. There has been a certain number of studies on the convergence of posterior distributions, in a bayesian nonparametric setup, in the last two decades. Diaconis and Freedman (1986) argue that Bayes estimates can be inconsistent if the underlying mechanism allows an infinite number of possible values, more precisely they prove that in the case of Bayesian inference on some infinite-dimensional parameter, even if the prior puts positive mass in weak neighbourhoods of the true parameter, it does not entail that the posterior mass of every weak neighbourhood of the true parameter tends to 1. Barron, Schervish and Wasserman (1999) give general conditions on the prior probability, under which the posterior probability of any Hellinger neighbourhood of the true density converges to 1 almost surely.

We then use the convergence of the posterior distribution to prove the convergence of the estimator of G_λ we are considering.

An outline of the paper is as follows. In Section 2, we define the prior on the parameter and also the loss function, which gives us an explicit expression for the Bayesian estimate of G_λ. The results on the convergence of the posterior distribution and of the estimator are stated in Section 3. The proofs are postponed in Section 5. Section 4 is devoted to additional remarks and simulations.

2 Definition of the Bayesian estimate

2.1 Construction of the prior π_0

As we said in the previous section, the construction of the prior is close to Heikkinen's (1998). Indeed, we use a marked point process to put mass on densities that are piecewise constant. However, as it is usually the case with bayesian estimation, this does not imply that the Bayes estimate of f is piecewise constant; in particular, if one considers the posterior mean of f.

Let us define more precisely the prior we use : we parameterise f by its values at some points, say $\xi = (\xi_1, \dots, \xi_K) \in E^K$, that are realisations of a point process; note that K is random. Typically, the point process ξ would be a Poisson process with intensity function $l : E \to [0, \infty[$. In this case, the number of points K follows the Poisson distribution with expectation $\mathcal{L}(E) = \int_E l(x)\mu(dx)$, where μ is the Lebesgue measure on $\mathbb{R}^2$, and conditional on K the locations $\xi_1, \dots, \xi_K$ are i.i.d. random variables with density $l(\cdot)/\mathcal{L}(E)$. As a special case, homogeneous Poisson processes satisfy $l(x) = c$, $\forall x \in E$, where c is a positive constant. Conditional on ξ, the Voronoi tesselation generated by those points provides a partition of E whose elements are tiles defined as follows :

$$V_j(\xi) = \{z \in E : d(z, \xi_j) \leq d(z, \xi_l), \forall l\}, \quad j = 1, \dots K,$$

where $d(\cdot,\cdot)$ denotes the Euclidean distance in $\mathbb{R}^2$. Note that the $V_j(\xi)$'s are convex polygons. This leads to a natural neighbourhood relation between the points ξ_j's by letting pairs of points having a common boundary be neighbours; denote $\sim$ this relation, therefore $\xi_i \sim \xi_j$ means that ξ_i and ξ_j are neighbours. We then construct f as a piecewise constant function on the $V_j(\xi)$'s :

$$f(z) = \frac{\sum_{j=1}^{K} \eta_j \mathbb{1}_{z \in V_j(\xi)}}{\sum_{j=1}^{K} \eta_j \mu(V_j(\xi))}. \tag{1}$$

However, more sophisticated forms of interpolation could also be considered (see Denison *et al.* 1998).

It then remains to define the prior distribution on $\eta = (\eta_1, ..., \eta_K)$ conditional on ξ. Because our aim is to estimate level sets associated with mainly smooth functions, we consider that, conditionally on ξ, η is a locally dependent Markov random field, associated with the neighbourhood relation $\sim$ defined previously. In other words, each realisation of f is parameterised by a marked point pattern

$$(\xi_1, \eta_1), ..., (\xi_K, \eta_K).$$

For simplicity's sake, we consider the following parameterisation for the weights : $\eta_j = e^{\theta_j}, \quad j = 1, \ldots, K$ and we denote $\tilde{\theta} = (\theta_1, \ldots, \theta_K)$. Also, the conditional distribution of θ given ξ would be a gaussian Markov random field defined as follows : $\theta_1 = 0$ and

$$\pi_0[\theta_i | \theta_j, j \sim i, \xi] \propto \exp\left\{ -\frac{\gamma}{H_i} \left(\theta_i - \sum_{j \sim i} \frac{\theta_j}{H_i d_{ij}} \right)^2 \right\}, \quad i \geq 2$$

where

$$H_i = \sum_{j \sim i} \frac{1}{d_{ij}},$$

and d_{ij} denotes the Euclidean distance in $\mathbb{R}^2$ between ξ_i and ξ_j and γ is a hyperparameter. Then, the joint distribution of $\theta = (\theta_2, \ldots, \theta_K)$ is given by

$$\pi_0(\theta_2, ..., \theta_K | \xi) \propto \prod_{i=1}^{K} \exp\left\{ -\frac{\gamma}{2} \sum_{j \sim i} \frac{(\theta_i - \theta_j)^2}{d_{ij}} \right\}.$$

One must note that this distribution is proper due to the restriction that $\theta_1 = 0$. The propriety of the prior is an important condition, in Bayesian nonparametric inference. In particular, we believe that without this restriction, i.e. $\theta_1 = 0$, the posterior distribution would not converge almost surely, see Section 3. We could have considered another restriction on $\tilde{\theta}$, however the arbitrariness of this restriction should not be of much effect, at least any restriction in the form $\theta_i = c$, for some i and some c is equivalent to the one we have chosen. We motivate the introduction of the d_{ij}'s in the definition of the above prior by the fact that among the neighbours of ξ_i say, those that are closer should have more influence than those that are further away.

The hyperparameter γ indicates our belief in the smoothness of f. Small γ's allow for less smooth densities.

However this approach can be generalised to other types of marked point patterns, as long as the distribution P of the number K of points satisfies some technical condition that will be given in Section 3, as long as, conditionally on K, the distribution of ξ has a density with respect to the Lebesgue measure on $I\!R^2$ and as long as, conditionally on ξ, the distribution of θ does not allow values, associated with points that are quite closed, be too different. These conditions will be given explicitly in Section 3.

We now define the loss function we consider and we give the expression of the Bayes estimate obtained under this loss function.

2.2 Loss function and Bayes estimate

We recall that our quantity of interest is $G_\lambda = \{x \in E; f(x) \geq \lambda\}$, where λ is fixed. The loss function that we consider is the Lebesgue measure of the symmetric difference, i.e. if $G, \hat{G} \subset E$,

$$L(G,\hat{G}) = \int_E 1\!\!I_{G\cap\hat{G}^c}(x)d\mu(x) + \int_E 1\!\!I_{G^c\cap\hat{G}}(x)d\mu(x). \tag{2}$$

This loss function is commonly used in set estimation. We could also consider an asymmetric loss function in the form :

$$a_1\int_E 1\!\!I_{G\cap\hat{G}^c}(x)d\mu(x) + a_2\int_E 1\!\!I_{G^c\cap\hat{G}}(x)d\mu(x),$$

with $a_1, a_2 > 0$, depending on the problem at hand. The results would not be significantly different.

Let π be any prior on the joint distribution of the observations then the Bayes estimate of G_λ, associated with the loss function defined by (2) and with the prior π has the form :

$$\hat{G}^\pi_\lambda \quad = \quad \left\{y \in E; \pi[y \in G_\lambda|X^{(n)}] \geq 1/2\right\}.$$

Indeed, the Bayes estimate minimizes

$$E^\pi\left[L(G_\lambda,\hat{G}_\lambda)\right] \quad = \quad \int_{\hat{G}^c_\lambda}\pi[y\in G_\lambda|X^{(n)}]d\mu(y) + \int_{\hat{G}_\lambda}\pi[y\in G^c_\lambda|X^{(n)}]d\mu(y).$$

The minimum is then achieves for $\{y \in E; \pi[y \in G|X^{(n)}] \geq \pi[y \in G^c|X^{(n)}]\} = \hat{G}^\pi_\lambda$.

In other words this is the set of points y such that the posterior probability that $f(y) \geq \lambda$ is greater than the posterior probability that $f(y) < \lambda$. This result is valid for any set estimation problem, as soon as the loss function is L.

Under the asymmetric loss function, one would have to compare the posterior probability with $a_2/(a_2 + a_1)$.

3 Main results

In this part, we focus more particularly on the consistency properties of our method. And, first, we present what kind of consistency we consider and the tools we use. Following from the paper of Barron, Schervish and Wasserman (1999), we give a result on the consistency of the posterior distribution in Hellinger distance (which is equivalent to the distance of the total variation).

Let us first introduce some notations :
Let $\mathcal{P}$ be the set of all probability measures that are absolutely continuous with respect to the Lebesgue measure μ on E and denote f_Q the Radon-Nikodym derivative of $Q \in \mathcal{P}$ with respect to μ. Denote $\mathcal{D}(\cdot,\cdot)$ the Hellinger pseudo-metric on $\mathcal{G} = \{f \;:\; f \geq 0 \text{ and } \int f d\mu < +\infty\}$.

Recall that $P \in \mathcal{P}$ is the true distribution of the i.i.d. random variables $X_1, \dots, X_n$ and f the corresponding density probability. Let $\mathcal{J}(\cdot, \cdot)$ be the Kullback-Leibler information

$$\mathcal{J}(f_P, f_Q) = \int \log\left(\frac{f_P(x)}{f_Q(x)}\right) f_P(x) d\mu(x),$$

for $f_P, f_Q \in \mathcal{G}$ and the integrand is stated to zero provided $f_P(x) = 0$. Finally for each $\varepsilon > 0$, define

$$N_\varepsilon = \{Q \in \mathcal{P} : \mathcal{J}(f, f_Q) \leq \varepsilon\} \tag{3}$$
$$A_\varepsilon = \{Q \in \mathcal{P} : \mathcal{D}(f, f_Q) \leq \varepsilon\} \tag{4}$$

Let π be a prior on $\mathcal{P}$.

Definition 1 *For $\varepsilon > 0$ and $\mathcal{C} \subseteq \mathcal{P}$, define $\mathcal{H}(\mathcal{C}, \varepsilon)$ to be the logarithm of the infimum of the set of all k such that there exist nonnegative functions $f_1^U, \dots, f_k^U$ satisfying :*

1. *$\int f_i^U(x) d\mu(x) \leq 1 + \varepsilon$ for all i,*

2. *for each $P \in \mathcal{C}$ there exists i such that $f_P \leq f_i^U$ μ–a.s.*

We now recall Theorem 1 of Barron *et al.* (1999), which enables us to prove the strong consistency of the posterior distribution. To do so, we first state the two conditions that have to be checked in their theorem.

A1 For every $\varepsilon > 0$, $\pi(N_\varepsilon) > 0$.

A2 For every $e > 0$, there exists a sequence $(\mathcal{F}_n)_{n=1}^{\infty}$ of subsets of $\mathcal{P}$, and positive, real numbers c_1, c_2 c_3 and ε such that

$$c_3 < ([e - \sqrt{\varepsilon}]^2 - \varepsilon)/2, \quad \varepsilon < e^2/4,$$

and such that

(i) $\pi(\mathcal{F}_n^c) \leq c_1 \exp(-nc_2)$ for all but finitely many n;

(ii) $\mathcal{H}(\mathcal{F}_n, \varepsilon) \leq nc_3$ for all but finitely many n.

Barron *et al.* (1999) prove the consistency of the posterior distribution under these two hypotheses.

Theorem 1 of Barron *et al.* (1999):

Let A_ε be defined by (4). Under conditions **A1** *and* **A2**, *for every $\varepsilon > 0$,*

$$\lim_{n\to\infty} \pi(A_\varepsilon | X^{(n)}) = 1 \text{ a.s. } [P].$$

In our setup conditions **A1** and **A2** will be satisfied under the following assumptions :

C1 The distribution of K (the number of points generated in E), satisfies :

$$P(K = k) > 0, \forall k \geq 3, \quad P(K \leq 2) = 0$$

and

$$\forall \tau > 0, \exists \rho > 0, \quad \text{s.}t. \quad P[K \geq \tau n/\log n] \leq e^{-n\rho}.$$

C2 Conditionally on K, the distribution of ξ is absolutely continuous with respect to the Lebesgue measure on E^K.

C3 Conditionally on ξ, the joint distribution of $(\theta_2, \dots, \theta_K)$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^{K-1}$ and satisfies :

$$\exists \beta, \alpha, r > 0, \quad \text{s.t.} \quad P[|\theta_i - \theta_j| > n^\beta d_{ij}^\alpha |\xi] \leq e^{-nr}, \quad \forall i \sim j,$$

where β, α and r do not depend on ξ.

Condition **C2** is important to ensure the regularity of the Voronoi tesselation.
We then obtain :

Proposition 1 *Under assumptions* **C1**–**C3**, *conditions* **A1** *and* **A2** *are satisfied and*

$$\lim_{n\to\infty} \pi(A_\varepsilon | X^{(n)}) = 1 \ a.s.\ [P] \tag{5}$$

holds.

Note that (5) implies the almost sure convergence of the posterior mean of f, which is a common bayesian estimate of f, in terms of the Hellinger distance $\mathcal{D}$.
Now, we obtain a result concerning the Bayes estimate of the level set G_λ.

Proposition 2 *If (5) is satisfied and if* $\mu(\partial(G_\lambda)) = 0$,

$$\lim_{n\to\infty} L(\hat{G}_\lambda^\pi, G_\lambda) = 0, \quad a.s.[P],$$

where L is defined by (2).

The condition $\mu(\partial G_\lambda) = 0$ is quite classical and it is assumed in all the papers studying consistency properties of level set estimates. Section 5.2. enlightens on the necessity of this hypothesis.

4 Remarks

4.1 Generality of conditions C1-C3.

Conditions **C1-C3** are fairly general, in particular the type of priors that we have considered satisfies these conditions. Let ξ be a Poisson process with intensity function l. Then

$$\begin{aligned} P(K \geq k_n) &= e^{-l(E)} \sum_{k=k_n}^{\infty} \frac{l(E)^k}{k!} \\ &\leq \kappa \exp\left\{-k_n \left(\log k_n + \kappa'\right)\right\}, \end{aligned}$$

where κ, κ' are positive constants. When $k_n = \tau n / \log n$ and n is large enough,

$$P(K \geq k_n) \quad \leq e^{-\tau n/2},$$

and assumption **C1** is proved. By definition of a poisson process **C2** is also satisfied. Now consider

$$\pi_0(\theta_2, ..., \theta_K | \xi) \propto \prod_{i=1}^{K} \exp\left\{ -\frac{\gamma}{2} \sum_{j\sim i} \frac{(\theta_i - \theta_j)^2}{d_{ij}} \right\}.$$

Then, if $i \sim j$ and $(\theta_i - \theta_j)^2/d_{ij} > n^{2\beta}$,

$$\frac{\gamma}{2}\theta^t W \theta = \sum_{l \sim t} \gamma \frac{(\theta_l - \theta_t)^2}{d_{lt}} \geq \gamma n^{2\beta},$$

where W is the inverse of the covariance matrix of θ. Therefore $\theta^t W \theta$ is a χ^2_{K-1} random variable. Let $K \leq \tau n/\log n$, $\alpha = 1$ and $\beta > 1/2$ (independent of K), then there exists $r > 0$ independent of K such that

$$P[\chi^2_{K-1} \geq n^{2\beta}] \leq e^{-nr},$$

and condition **C3** is satisfied.

4.2 Simulations

For the simulations we use the reversible jump Markov chain Monte Carlo algorithm proposed by Heikkinen (1998).

Starting with initial parameters $\xi^{(0)}$ and $\theta^{(0)}$, with $K^{(0)} \geq 6$, we construct an iterative sequence $(\xi^{(t)}, \theta^{(t)})$, $t \in I\!N$ in the following way : let $(\xi^{(t)}, \theta^{(t)})$ be the current state, with K the number of generating points. We choose among the three following moves with probability, h_K, b_K and d_K respectively. Let $\delta, C > 0$ be hyperparameters of the sampler :

1. move 1 : *K fixed move*. We propose $\theta' = (0, \theta'_2, \cdots, \theta'_K)$, where $\theta'_k \sim \mathcal{U}_{[\theta_k^{(t)} - \delta, \theta_k^{(t)} + \delta]}$ and $\theta'_i = \theta_i^{(t)}$, $i \neq k$ and where k is uniformly chosen in $\{2, \cdots, K\}$. $\xi' = \xi^{(t)}$.

2. move 2 : *birth*. We propose $\xi'_{K+1} \sim \mathcal{U}_E$ and $\theta'_{K+1} = \underline{\theta}^{(t)} + \varepsilon$, with

$$\underline{\theta}^{(t)} = \sum_{k \sim K+1} \frac{\mu(V_k(\xi)) - \mu(V'_k(\xi'))}{\mu(V_{K+1}(\xi'))} \theta_k,$$

and $\varepsilon \in I\!R$ is distributed according to the density $g(\varepsilon) = Ce^{C\varepsilon}/(1 + e^{C\varepsilon})^2$.

3. move 3 : *death*. We select uniformly k in $\{1, \cdots, K\}$. If $k = 1$ then $\xi' = (\xi'_1, \cdots, \xi'_{K-1}) = (\xi_2^{(t)}, \cdots, \xi_K^{(t)})$ and $\tilde{\theta}' = (1, \theta_3^{(t)}, \cdots, \theta_K^{(t)})$. If $k > 1$ then, as in Heikkinen (1998), we re-index $\xi^{(t)}$ and $\theta^{(t)}$ by switching k and K, so that $K' = K - 1$ and $\xi' = (\xi_1^{(t)}, \cdots, \xi_{K-1}^{(t)})$.

h_K, b_K, d_K and the exact expressions of the acceptance probabilities in each moves are given in Heikkinen (1998). The program we have run was written in Matlab.

We have considered the following density f defined on $E = [0,1]^2$: using the transformation $g(z) = e^z/(1 + e^z)$, we have considered $Z = (g(x), g(y)) \sim f$, where x and y are independent and are distributed according to :

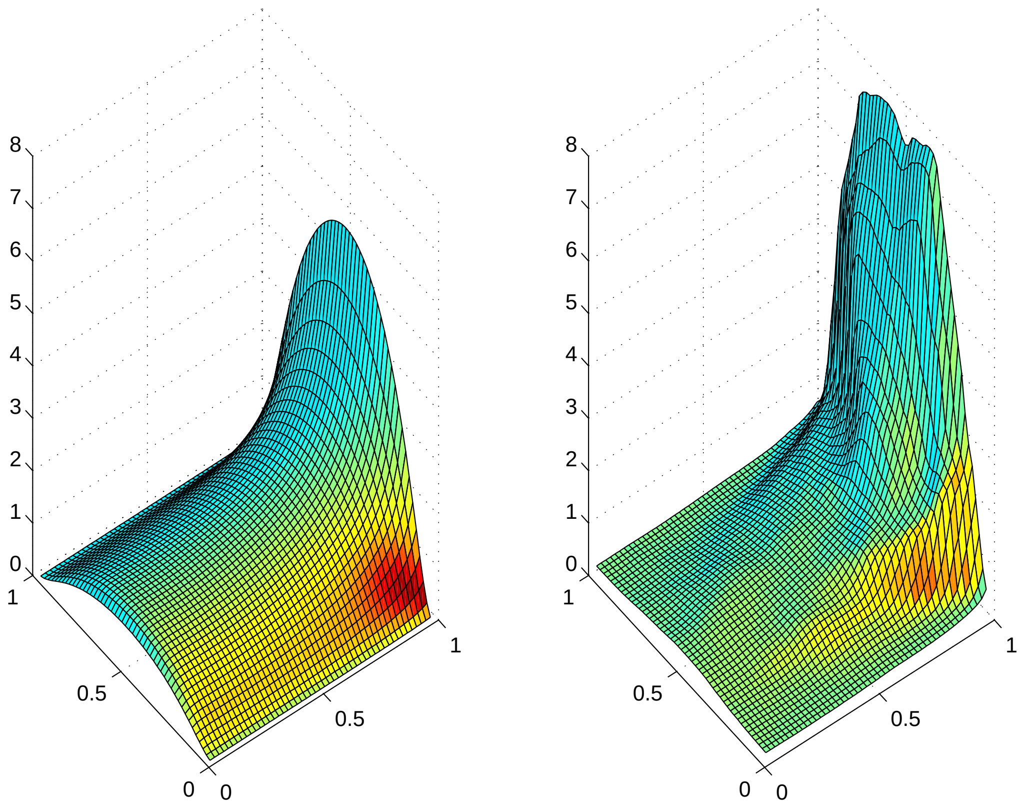
- $x \sim \mathcal{N}(0,1)$ and $y \sim \mathcal{N}(1,4)$. We have estimated the level sets associated with $\lambda = 0.2,\ 0.5,\ 1.5,\ 2$.

For each graph, we have run 750 000 iterations with $n = 1000$ observations and where the hyperparameters are : $\delta = 0.8$ in move 1, $C = 5$ in move 2, $\gamma = 0.1$ in the definition of the prior and the parameter of the Poisson distribution is 30.

The results are reported in figures 1, 2 and 3. Figure 1 shows the true density and the estimated one using the posterior mean. Figure 2 and 3 show the true level sets and our Bayes estimators for different levels : $\alpha = 0.2,\ 0.5,\ 1.5,\ 2$.

These simulations show an important feature of level set estimation that was already given in Proposition 2, namely that the estimation is not good when the Lebesgue measure of the set $\{f(x) = \lambda\}$ is positive. Indeed, whereas for $\lambda = 0.2, 1.5, 2$ the estimation is rather good, when $\lambda = 0.5$ the estimation is quite poor. By looking at the true density, we realize the the density is very flat around the level 0.5, whereas it is quite steep around the other two levels. It thus shows that it is difficult to estimate the flat parts of f. We had also tested other levels and the results where quite as good as for $\lambda = 0.2, 1.5, 2$.

Figure 1: Left side: true density and Right side : estimated density



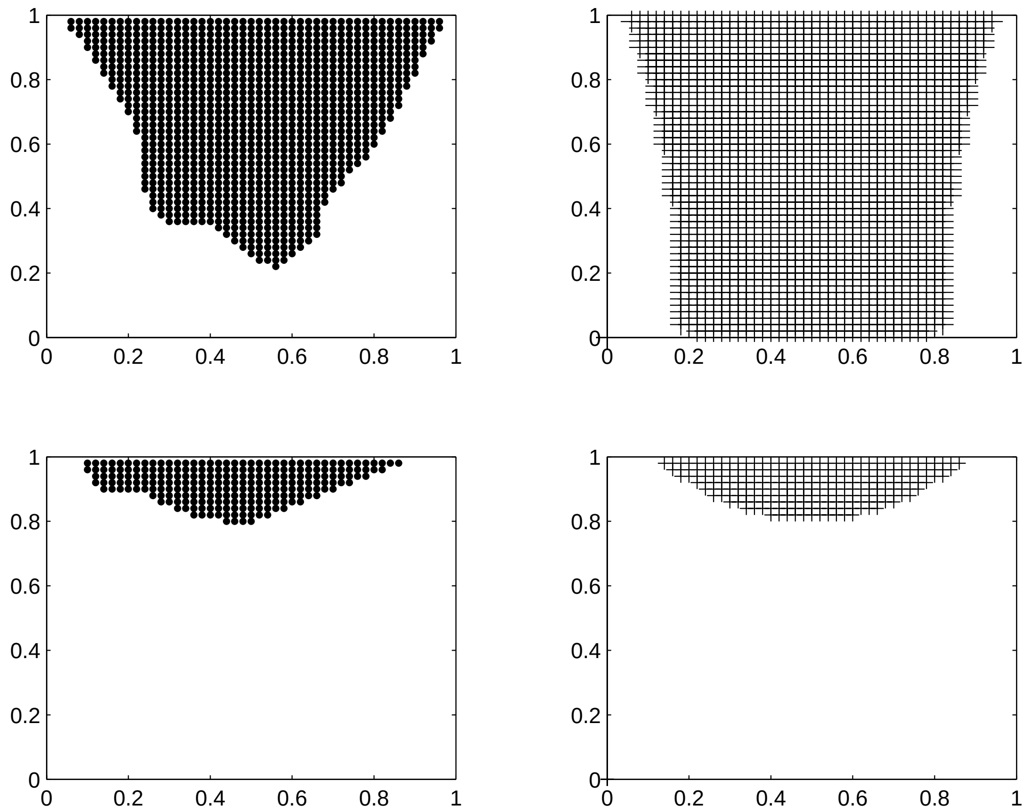


Figure 2: (a) Top left : estimated level set, $\lambda = 0.5$, (b) Top right : true level set, $\lambda = 0.5$, (c) Bottom left : estimated level set, $\lambda = 2$, (d) Bottom right : true level set, $\lambda = 2$

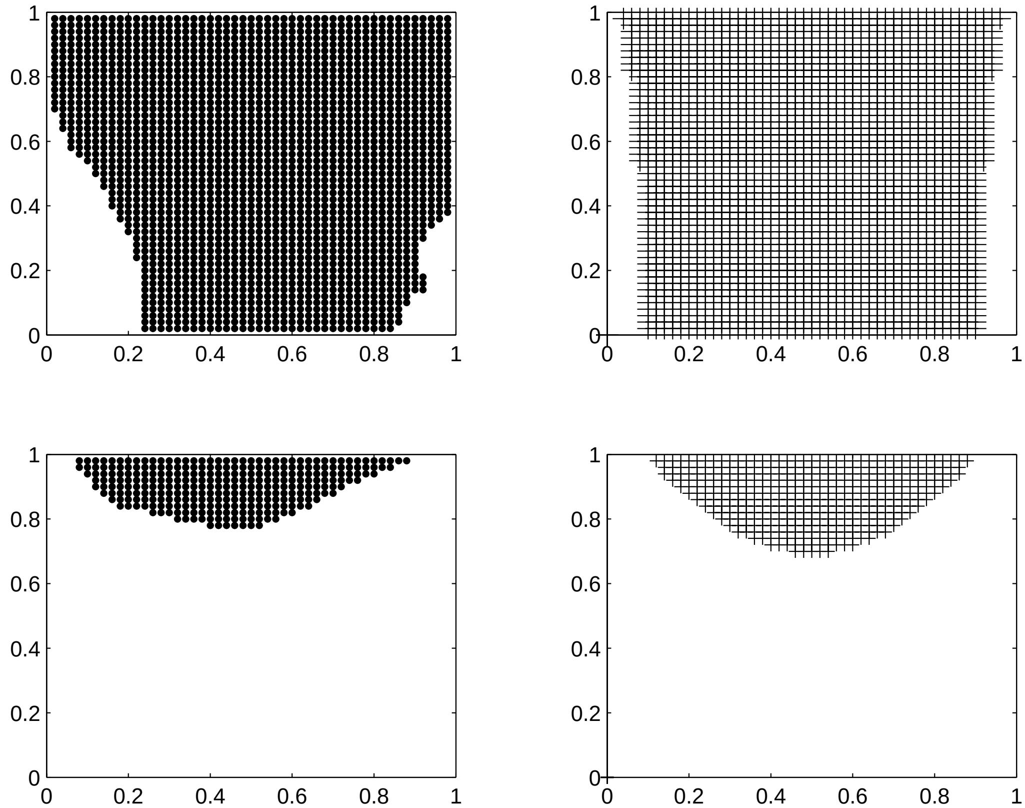


Figure 3: (a) Top left : estimated level set, $\lambda = 0.2$, (b) Top right : true level set, $\lambda = 0.2$, (c) Bottom left : estimated level set, $\lambda = 1.5$, (d) Bottom right : true level set, $\lambda = 1.5$

5 Proofs

5.1 Proof of Proposition 1

This proof is reduced to verify that both assumptions **A1** and **A2** are satisfied. Throughout the proof, C will denote any positive constant and $\|\cdot\|$ will denote the Euclidean norm in the spaces $\mathbb{R}^l$, $l \geq 2$. We will also denote $f_{\theta,\xi}(x)$ densities in the form (1), i.e.

$$f_{\theta,\xi}(x) = \sum_{i=1}^{K} \frac{e^{\theta_i} \mathbb{1}_{x \in V_i}}{\sum_{i=1}^{K} e^{\theta_i} \mu_i},$$

where μ_i denotes the Lebesgue measure of V_i, $K > 2$, and $V_i = V_i(\xi)$ is defined as previously.
1. Assumption **A1** :
Set $\eta_i = \exp(\theta_i)$ so that $\eta_i > 0$ for all i.

• We first assume that there exist (θ, ξ) such that $f(x) = f_{\theta,\xi}(x)$. Set $N_\varepsilon^*(f_{\theta,\xi}) = \{f_{\theta',\xi'} : \mathcal{J}(f_{\theta,\xi}, f_{\theta',\xi'}) \leq \varepsilon\}$ and define $\kappa(A,B) = \{f_{\theta',\xi'} :$ for all $i = 1 \ldots, K$, $|\eta_i - \eta_i'| \leq A\eta_i, \|\xi_i' - \xi_i\| \leq B,\ \xi_i' \in V_i\}$. To prove that $\pi(N_\varepsilon^*) > 0$, we prove first that for any $\varepsilon > 0$, there exist $A, B > 0$ such that $\kappa(A,B) \subset N_\varepsilon^*$. Then, since, conditionally on K, the prior is absolutely continuous with respect to Lebesgue measure, $\pi[\kappa(A,B)|K] > 0$ and assumption **A1** is established.

Let $f_{\theta',\xi'}$ in $\kappa(A,B)$,

$$\mathcal{J}(f_{\theta,\xi}, f_{\theta',\xi'}) = \sum_{i,j} \frac{\eta_i}{\sum_{i=1}^{K} \eta_i \mu_i} \log\left(\frac{\eta_i}{\sum_{l=1}^{K} \eta_l \mu_l} \frac{\sum_{l=1}^{K} \eta_l' \mu_l'}{\eta_j'}\right) \mu(V_i \cap V_j'),$$

where $\mu'_l = \mu(V'_l)$ for $l = 1, ..., K$. Note that

$$x \in V_i \cap V'_j \Leftrightarrow \begin{cases} \|x - \xi_i\| \le \|x - \xi_l\| & l \neq i \\ \|x - \xi'_j\| \le \|x - \xi'_l\| & l \neq j. \end{cases}$$

This implies that

$$\mu(V_i \cap V'_j) \le \mu(\{x \in V_i : \|x - \xi_j\| \le \|x - \xi_l\| + 2B, l \neq j\}). \tag{6}$$

Two cases occur :
- If $j \neq i$, following from (6), there exists $C(\xi) < \infty$, such that for all $B > 0$

$$\mu(V_i \cap V'_j) \le C(\xi)B,$$

since $\|\xi_l - \xi'_l\| \le B$, $\forall l = 1, \dots, K$.
- If $j = i$,

$$\mu(V_i) \ge \mu(V_i \cap V'_i) = \mu(V_i) - \sum_{j \neq i} \mu(V_i \cap V'_j) \ge \mu(V_i) - (K-1)C(\xi)B$$

and $\mu'_i \le \mu(V_i \cap V'_i) + (K-1)C(\xi)B$. So the Kullback-Leibler divergence satisfies

$$\begin{aligned} \mathcal{J}(f_{\theta,\xi}, f_{\theta',\xi'}) &\le \sum_i \frac{\eta_i}{\sum_{i=1}^K \eta_i \mu_i} \log\left(\frac{\eta_i}{\sum_{i=1}^K \eta_i \mu_i} \frac{\sum_{j=1}^K \eta'_j \mu'_j}{\eta'_i}\right) \mu_i + \\ & \quad \sum_{i,j,i \neq j} \frac{\eta_i}{\sum_{i=1}^K \eta_i \mu_i} \log\left(\frac{\eta_i}{\sum_{i=1}^K \eta_i \mu_i} \frac{\sum_{j=1}^K \eta'_j \mu'_j}{\eta'_j}\right) c(\eta, \xi) B \\ &\le C(\eta, \xi)(A + B), \end{aligned}$$

where $C(\eta, \xi)$ is a positive constant; the last inequality holds provided A and B are small enough. So we can choose A and B such that $\mathcal{J}(f_{\theta,\xi}, f_{\theta',\xi'})$ is less than ε.

Therefore for such A and B's, $\kappa(A, B) \subset N^*_\varepsilon$.

• We now consider any function $f > 0$ defined on E and such that $\int_E f(x)d\mu(x) = 1$. Then, for all $\varepsilon > 0$, there exists θ, ξ such that $\mathcal{J}(f, f_{\theta,\xi}) \le \varepsilon/2$. Therefore, $N^*_{\varepsilon/2}(f_{\theta,\xi}) \subset N^*_\varepsilon(f)$

• Finally, $\pi(N^*_\varepsilon(f_{\theta,\xi})) > 0$ and $N^*_{\varepsilon/2} \subset N_\varepsilon(f)$, where $N_\varepsilon(f)$ is defined by (3) and assumption **A1** is satisfied.

2. Assumption **A2** :
Let $\tau > 0$ and $k_n = \tau n / \log n$, set

$$\mathcal{F}_n = \left\{ f_{\theta,\xi}; K \le k_n, \forall i, j \le K, |\theta_i - \theta_j| \le n^\beta d_{ij}^\alpha \right\},$$

where β and α are defined in assumption **C3**.
We now prove that $\pi(\mathcal{F}_n^c) \le e^{-nr_1}$, for some $r_1 > 0$:
C1-C3 imply that

$$\begin{aligned} \pi(\mathcal{F}_n^c) &\le P[K \ge k_n] + \sum_{k=3}^{k_n} P(K = k) \sum_{i<j} \pi \left[|\theta_i - \theta_j| > n^\beta d_{ij}^\alpha \middle| K = k \right] \\ &\le e^{-n\rho} + e^{-nr} \sum_{k=3}^{k_n} k(k-1)/2 \\ &\le e^{-n\rho} + e^{-nr} C k_n^3 \\ &\le e^{-nr_1}, \end{aligned}$$

with $r_1 < \min(\rho, r/2)$, when n is large enough.
It thus remains to prove part **(ii)** of Assumption **A2**.

Let $e > 0$ and $c_3, \varepsilon > 0$ such that $\varepsilon < e^2/4$ and $c_3 < ([e - \sqrt{\varepsilon}]^2 - \varepsilon)/2$. Let $\mathcal{F}_n(k)$ be the set of $f_{\theta,\xi} \in \mathcal{F}_n$ such that the number of points ξ_i is equal to k. Then, $\mathcal{F}_n = \cup_{k=3}^{k_n} \mathcal{F}_n(k)$.

Let $(\xi_1, \theta_1, ..., \xi_k, \theta_k)$ be a parameter in $(E \times I\!R)^k$ satisfying the conditions of $\mathcal{F}_n(k)$ and $\delta, \delta' > 0$. The idea of the proof is to construct an upper bound f^U and to obtain conditions on δ and δ' such that for any $(\xi_1', \theta_1', ..., \xi_k', \theta_k')$ satisfying $\|\xi_i - \xi_i'\| \le \delta$ and $|\theta_i - \theta_i'| < \delta'$,

$$f'(x) = f_{\theta',\xi'}(x) \le f^U(x) \quad \text{and} \int_E f^U(x) d\mu(x) \le 1 + \varepsilon,$$

where f^U depends on θ, ξ. Then, we count the number of balls $\{|\theta' - \theta| \le \delta', \|\xi_i' - \xi_i\| \le \delta, i = 1, ..., k\}$ that is needed to cover $\mathcal{F}_n(k)$ and we prove that it is small enough for **A2** to be satisfied.

We denote $d_1 = n^{-h_0}$, with $h_0 > 0$ some positive constant which we will fix later on. We divide $\xi_1, ..., \xi_k$ into groups, say $(\xi_{i_1}, ..., \xi_{i_{n_i}})$, $i = 1, ..., q$ and $n_1 + ... + n_q = k$. These groups are defined by

- (i) $d_{i_l i_{l+1}} \le d_1$, $l = 1, ..., n_i - 1$, $i = 1, ..., q$.
- (ii) $\forall i \ne j \le q$, $\forall l \le n_i, l' \le n_j$ $d_{i_l j_{l'}} > d_1$.

Let $V_1, ..., V_k$ be the Voronoi tesselation associated with $(\xi_1, ..., \xi_k)$. We denote $\bar{V}_i = \cup_{l=1}^{n_i} V_{i_l}$, the union of the tiles corresponding to the group i. Then $\bar{V}_i$ is connected, i.e. $\xi_{i_1}, ..., \xi_{i_{n_i}}$ is a sequence of neighbours. We denote $\bar{B}_i$ an h neighbourhood of the boundary of $\bar{V}_i$ denoted by $\partial \bar{V}_i$:

$$\bar{B}_i = \{z; d(z, \partial \bar{V}_i) \le h\},$$

with $h > 0$. Denote $\bar{\theta}_i = max\{\theta_{i_l}, l = 1, ..., n_i\}$ and $\underline{\theta}_i = min\{\theta_{i_l}, l = 1, ..., n_i\}$.

We have :

$$f(x) = f_{\theta,\xi}(x) = \sum_{i=1}^k \frac{e^{\theta_i} 1\!\!I_{x \in V_i}}{\sum_{j=1}^k e^{\theta_j} \mu_j}.$$

We then build the following upper bound for f :

$$f^U(x) = \frac{\sum_{i=1}^k e^{\bar{\theta}_i + \delta'} \left[1\!\!I_{\bar{V}_i}(x) + 1\!\!I_{\bar{B}_i}(x)\right]}{\sum_{i=1}^k e^{\underline{\theta}_i - \delta'} \left[\mu(\bar{V}_i) - \mu(\bar{B}_i)\right]}. \tag{7}$$

Now let $(\xi_1', \theta_1', ..., \xi_k', \theta_k')$ be such that $\|\xi_j - \xi_j'\| \le \delta$ and $|\theta_j - \theta_j'| < \delta'$. We determine the relation between δ and h such that

$$f'(x) \quad = \quad \sum_{i=1}^k \frac{e^{\theta_i'} 1\!\!I_{x \in V_i'}}{\sum_{j=1}^k e^{\theta_j'} \mu_j'} \le f^U(x). \tag{8}$$

We have,

$$\begin{aligned} f'(x) &= \frac{\sum_{i=1}^q \sum_{l=1}^{n_i} e^{\theta'_{i_l}} \left(1\!\!I_{V'_{i_l} \cap \bar{V}_i}(x) + 1\!\!I_{V'_{i_l} \cap \bar{V}_i^c}(x)\right)}{\sum_{i=1}^q \sum_{l=1}^{n_i} e^{\theta'_{i_l}} \left(\mu(V'_{i_l} \cap \bar{V}_i) + \mu(V'_{i_l} \cap \bar{V}_i^c)\right)} \\ &\le \sum_{i=1}^q \frac{e^{\bar{\theta}_i + \delta'} \left(1\!\!I_{\bar{V}_i}(x) + \sum_{j \ne i}^q 1\!\!I_{\bar{V}_i' \cap \bar{V}_j}(x)\right)}{\sum_{i=1}^q e^{\underline{\theta}_i - \delta'} \left(\mu(\bar{V}_i) - \sum_{j \ne i}^q [\mu(\bar{V}_j \cap \bar{V}_i')]\right)} \\ &= e^{2\delta'} \sum_{i=1}^q \frac{e^{\bar{\theta}_i} \left(1\!\!I_{\bar{V}_i}(x) + \sum_{j \ne i}^q 1\!\!I_{\bar{V}_i' \cap \bar{V}_j}(x)\right)}{\sum_{i=1}^q e^{\underline{\theta}_i} \left(\mu(\bar{V}_i) - \sum_{j \ne i}^q [\mu(\bar{V}_j \cap \bar{V}_i')]\right)}. \end{aligned}$$

To prove (8), we will determine in the following lemma, h so that for all $i = 1, .., q$:
- (1) : $\sum_{j \neq i}^{q} 1\!\!I_{\bar{V}_i' \cap \bar{V}_j}(x) \leq 1\!\!I_{\bar{B}_i}(x)$
- (2) : $\sum_{j \neq i}^{q} [\mu(\bar{V}_j \cap \bar{V}_i')] \leq \mu(\bar{B}_i)$.

Let $M = \sup_{x,y \in E} \|x - y\| < \infty$, we have the following result :

Lemma 1 *Choose $h_0 > (\beta + 1)/\alpha$ and $h \leq 2\delta M/d_1$, with $\delta \leq \varepsilon d_1^2/8M$ then $f'(x) \leq f^U(x)$ whenever $\|\xi_i - \xi_i'\| \leq \delta$ and $|\theta_i - \theta_i'| \leq \delta'$, $i = 1, \cdots k$.*

If $\delta \leq \varepsilon d_1^2/192M$,

$$\int f^U(x) d\mu(x) \leq 1 + \varepsilon.$$

The proof of Lemma 2 is given in the appendix.

When k is fixed, the number of such balls needed to cover $\mathcal{F}_n(k)$ (i.e. the number of such upper bounds f^U), is equal to :

$$N_k = N_k^\theta \times N_k^\xi,$$

where $N_k^\theta \leq \left(\frac{3n^{\beta+1}M^\alpha}{\varepsilon}\right)^{k-1}$ is an upper bounds for the number of possible moves of length $\varepsilon/3$ for the θ_i's, $i = 2, ..., k$, $N_k^\xi \leq [C/\delta^2]^k$, and

$$N_k \leq C^k \varepsilon^{-3k} n^{4h_0 k + k(\beta+1)}.$$

Since the number of upper bounds can be bounded by $k_n N_{k_n}$,

$$\begin{aligned}
\mathcal{H}(\mathcal{F}_n, \varepsilon) &\leq k_n \log C - 3k_n \log \varepsilon + (4h_0 k_n + k_n(\beta + 1)) \log n + \log k_n \\
&= \frac{\tau n}{\log n} (\log C - 3 \log \varepsilon + (4h_0 + \beta + 1) \log n) + \log (\tau n / \log n) \\
&\leq 2\tau n (4h_0 + \beta + 1),
\end{aligned}$$

when n is large enough. Choose $\tau < c_3/[2(4h_0 + \beta + 1)]$, then $\mathcal{H}(\mathcal{F}_n, \varepsilon) \leq nc_3$ and assumption **A2** is satisfied.

5.2 Proof of Proposition 2

Recall that,

$$\hat{G}_\lambda^\pi = \{y; \pi[y \in G_\lambda | X^{(n)}] \geq 1/2\}.$$

For clarity's sake, we denote f_0 the true density and G_0 the true level set, throughout the proof. Then,

$$\begin{aligned}
L(\hat{G}_\lambda^\pi, G_0) &= \int_{G_0 \cap (\hat{G}_\lambda^\pi)^c} d\mu(y) + \int_{\hat{G}_\lambda^\pi \cap G_0^c} d\mu(y) \\
&= \int 1\!\!I_{f_0(y)=\lambda} 1\!\!I_{(\hat{G}_\lambda^\pi)^c}(y) d\mu(y) + \int 1\!\!I_{f_0(y)>\lambda} 1\!\!I_{(\hat{G}_\lambda^\pi)^c}(y) d\mu(y) + \int 1\!\!I_{f_0(y)<\lambda} 1\!\!I_{\hat{G}_\lambda^\pi}(y) d\mu(y).
\end{aligned}$$

If $\mu(\{f_0 = \lambda\}) = 0$, the first term of the right hand side of the above equality is equal to zero and

$$\lim_{\delta \to 0} \int_{\lambda < f_0 < \lambda + \delta} d\mu = \lim_{\delta \to 0} \int_{\lambda > f_0 > \lambda - \delta} d\mu = 0.$$

Set $G_0^\delta = \{y; f_0(y) \geq \lambda + \delta\}$ and $G_0^{-\delta} = \{y; f_0(y) \leq \lambda - \delta\}$ for $\delta > 0$. Let $\varepsilon > 0$,then there exists $\delta > 0$ such that

$$L(\hat{G}_\lambda^\pi, G_0) \leq \varepsilon + \int_{G_0^\delta \cap (\hat{G}_\lambda^\pi)^c} d\mu(y) + \int_{G_0^{-\delta} \cap \hat{G}_\lambda^\pi} d\mu(y).$$

Since on $(\hat{G}_\lambda^\pi)^c$, $\pi[y \in G_\lambda^c | X^{(n)}] \geq 1/2$,

$$\begin{aligned} \int_{G_0^\delta \cap (\hat{G}_\lambda^\pi)^c} d\mu(y) &\leq 2 \int_{G_0^\delta} \pi[y \in G_\lambda^c | X^{(n)}] d\mu(y) \\ &= 2E^\pi \left[\int_{G_0^\delta} 1\!\!I_{f(y)<\lambda} d\mu(y) \middle| X^{(n)} \right]. \end{aligned}$$

Similarly,

$$\begin{aligned} \int_{G_0^{-\delta} \cap \hat{G}_\lambda^\pi} d\mu(y) &\leq 2 \int_{G_0^{-\delta}} \pi[y \in G_\lambda | X^{(n)}] d\mu(y) \\ &= 2E^\pi \left[\int_{G_0^{-\delta}} 1\!\!I_{f(y)\geq\lambda} d\mu(y) \middle| X^{(n)} \right]. \end{aligned}$$

Moreover, when $f_0(y) > \lambda + \delta$ and $f(y) < \lambda$,

$$(\sqrt{f_0(y)} - \sqrt{f(y)})^2 \geq \lambda \left(\sqrt{1 + \frac{\delta}{\lambda}} - 1 \right)^2 = \tau_1$$

and when $f_0(y) \leq \lambda - \delta$ and $f(y) \geq \lambda$,

$$(\sqrt{f(y)} - \sqrt{f_0(y)})^2 \geq \lambda \left(1 - \sqrt{1 - \frac{\delta}{\lambda}} \right)^2 = \tau_2.$$

Thus

$$\begin{aligned} \int_{G_0^\delta \cap (\hat{G}_\lambda^\pi)^c} d\mu(y) &\leq \frac{2}{\tau_1} E^\pi \left[\int_{G_0^\delta} (\sqrt{f_0(y)} - \sqrt{f(y)})^2 d\mu(y) \middle| X^{(n)} \right] \\ &\leq \frac{2}{\tau_1} E^\pi [\mathcal{D}(f, f_0) | X^{(n)}], \end{aligned}$$

and

$$\int_{G_0^{-\delta} \cap \hat{G}_\lambda^\pi} d\mu(y) \leq \frac{2}{\tau_2} E^\pi [\mathcal{D}(f, f_0) | X^{(n)}].$$

Following from Proposition 1, the above quantities go to zero as n goes to infinity, P_0 a.s. and Proposition 2 is proved.

When $\mu(\partial G_\lambda) > 0$, the term

$$\int 1\!\!I_{f_0(y)=\lambda} 1\!\!I_{(\hat{G}_\lambda^\pi)^c} d\mu(y)$$

is not necessarily equal to zero, and might not converge to zero. Indeed even if $\hat{f}$, the posterior mean of f, converges to f_0, on the set of points y such that $f_0(y) = \lambda$, $\hat{f}(y)$ can be quite close to λ but it can be still strictly less than λ. In this case $y \notin \tilde{G} = \{x; \hat{f}(x) \geq \lambda\}$. This is only a heuristic argument and $\tilde{G}$ probably does not behave as well as $\hat{G}_\lambda^\pi$, however it enlightens on the fact that when $\mu(\partial G_\lambda) > 0$ the above term cannot be controlled easily and might not converge to zero. However, note that only this part, i.e. $\{x, f_0(x) = \lambda\}$ might be badly estimated.

6 Appendix : Proof of Lemma 2

For simplicity's sake, we first assume that $q = k$, in other words that $\forall i \neq j \leq k$, $d_{ij} > d_1$.

To begin with, consider ξ_i and ξ_j, neighbours, that are changed into ξ_i', ξ_j' respectively, such that $\|\xi_i - \xi_i'\| \leq \delta$ and $\|\xi_j - \xi_j'\| \leq \delta$. Let L_{ij} be the bisector (see Boots *et al.* p 47-48) of ξ_i, ξ_j and L'_{ij} the bisector of ξ_i', ξ_j'. We bound $(V_i' \cap V_j) \cup (V_i \cap V_j')$ by the union of the symmetric difference between V_i' and V_j and V_j' and V_i. To do so, we consider $h(y)$ the length of $[y, y']$, when $y \in L_{ij}$ and $y' \in L'_{ij}$ is such that y is the projection of y' on L_{ij}. By obtaining an upper bound on $h(y)$, say h, we will be able to construct an h neighbourhood of the boundaries of V_i such that :

$$(V_j' \cap V_i) \cup (V_i' \cap V_j) \subset B_i.$$

We now calculate $h(y)$:

We consider the orthonormal coordinate system defined as follows : ξ_j is the origine, the x coordinate is on the axe defined by the line parallel to L_{ij} going through ξ_j and the y coordinate is on the line (ξ_j, ξ_i). Note that by definition of L_{ij}, the axes are orthogonal. The coordinates of ξ_i are then $(0, d_{ij})$. Denote τ_1 and τ_2, τ_1' and τ_2' real numbers such that the coordinates of ξ_i' are : $(\tau_1, d_{ij} + \tau_2)$, and those of ξ_j' are (τ_1', τ_2'). Then $\delta^2 \geq \tau_1^2 + \tau_2^2$ and $\delta^2 \geq (\tau_1')^2 + (\tau_2')^2$.

Let O be the middle point of $[\xi_i, \xi_j]$ and O' be the middle point of $[\xi_i', \xi_j']$. Let O" be the projection of O' on L_{ij}. Then

$$O = \frac{\xi_i + \xi_j}{2} = (0, d_{ij}/2)$$

and

$$O' = \frac{\xi_i' + \xi_j'}{2} = (\frac{\tau_1 + \tau_1'}{2}, \frac{d_{ij} + \tau_2 + \tau_2'}{2}).$$

Let Ω be the intersection between L_{ij} and L'_{ij}. Then, from Thales, we obtain :

$$\frac{h(y)}{|O'O"|} = \frac{|\Omega y|}{|\Omega O"|}. \tag{9}$$

Let A be the intersection between $\xi_i\xi_j$ and $\xi_i'\xi_j'$. Suppose, without lack of generality, that $|A\xi_j| > |A\xi_i|$ and denote ρ the angle between $[\xi_j, \xi_i)$ and $[\xi_j', \xi_i')$. Then ρ is also the angle between $[\Omega 0')$ and $[\Omega O")$, so that

$$|\tan(\rho)| \;=\; \frac{|O'O"|}{|\Omega O"|} = \frac{|\tau_1'|}{|A\xi_j + \tau_2'|} \leq \frac{|\tau_1'|}{d_{ij}/2}. \tag{10}$$

From (9) and(10), we obtain an upper bound for $h(y)$

$$h(y) \;\leq\; 2\delta M/d_{ij}.$$

Since $d_{ij} > d_1$, $h(y) \leq 2\delta M/d_1 = h$. We can apply this argument, to any neighbours of ξ_i, for all $i = 1, ..., k$. Let B_i be the set :

$$\{z : d(z, \partial V_i) \leq h\},$$

the convex structure of V_i and the above calculations imply that B_i contains

$$\cup_{j \sim i} \left[(V_j \cap V_i') \cup (V_j' \cap V_i) \right].$$

To obtain an upper bound for f' in the form (7), we must make sure that for all i, $\mu_i - \mu(B_i) > 0$. We have,

$$\mu_i = \sum_{j \sim i} \frac{L_{ij} d_{ij}}{2}, \quad \mu(B_i) = 2h \sum_{j \sim i} L_{ij},$$

therefore

$$\mu_i - \mu(B_i) \geq \sum_{j \sim i} L_{ij} \left(\frac{d_1}{2} - 2h \right).$$

Since $2\delta M/d_1 = h$, if we choose $\delta \leq \frac{d_1^2}{8M}$, for all $i = 1, \cdots k$, $\mu_i - \mu(B_i) > 0$.

Moreover, since the V_i's are polygons, when ξ_j and ξ_i are not neighbours, $V_j' \cap V_i = V_i' \cap V_j = \emptyset$ under the above condition on δ. Thus, when $\delta \leq \frac{d_1^2}{8M}$,

$$\begin{aligned} f'(x) &\leq \sum_{i=1}^{k} \frac{e^{\theta_i'} \left(1\!\!I_{x \in V_i} + 1\!\!I_{x \in B_i} \right)}{\sum_{j=1}^{k} e^{\theta_j'} [\mu_j - \mu(B_j)]} \\ &\leq e^{2\delta'} \sum_{i=1}^{k} \frac{e^{\theta_i} \left(1\!\!I_{x \in V_i} + 1\!\!I_{x \in B_i} \right)}{\sum_{j=1}^{k} e^{\theta_j} [\mu_j - \mu(B_j)]} \\ &= f^U(x). \end{aligned}$$

Now, let determine the condition on δ, δ' needed to obtain

$$\begin{aligned} \int f^U(x) d\mu(x) &\leq 1 + \varepsilon \qquad (11) \\ &\Leftrightarrow \\ \sum_i e^{\theta_i} \mu(B_i) [1 + e^{-2\delta'}(1+\varepsilon)] &\leq [e^{-2\delta'}(1+\varepsilon) - 1] \sum_i e^{\theta_i} \mu_i \end{aligned}$$

thus if we choose $\delta' \leq \varepsilon/3$ and $\delta \leq \varepsilon \frac{d_1^2}{48M}$, we have

$$\mu(B_i) \leq \frac{\varepsilon}{6} \mu_i$$

and (11) is satisfied.

We now consider the general case. In each group i, $i = 1, ..., q$, by definition of $\mathcal{F}_n$,

$$|\theta_{i_l} - \theta_{i_t}| \leq n^\beta d_{i_l i_t}^\alpha \leq n^\beta d_1^\alpha, \quad \text{when} \quad i_l \sim i_t, \quad l, t \leq n_i.$$

In each group, two points can be linked by a sequence of neighbours, thus, in each group :

$$|\bar{\theta}_i - \underline{\theta}_i| \leq n_i n^\beta d_1^\alpha \leq \frac{\tau n^{\beta + 1 - \alpha h_0}}{\log n}.$$

Choosing $h_0 > (\beta + 1)/\alpha$, we obtain that for all $i = 1, ..., q$:

$$|\bar{\theta}_i - \underline{\theta}_i| \leq n^{-r_0},$$

for some $r_0 > 0$.

We first obtain an upper bound for $f(x)$ as follows :

$$f(x) \leq f^{(1)}(x) = \frac{\sum_{i=1}^{q} e^{\bar{\theta}_i} 1\!\!I_{\bar{V}_i}(x)}{\sum_{i=1}^{q} e^{\underline{\theta}_i} \mu(\bar{V}_i)}.$$

Let now consider moves of length δ for all the points. Let $f'(x)$ be the density corresponding to the (ξ_l', θ_l'), $l = 1, ..., k$. We have :

$$f'(x) \le e^{2\delta'} \frac{\sum_{i=1}^q e^{\bar{\theta}_i} 1\!\!I_{\bar{V}_i'}(x)}{\sum_{i=1}^q e^{\underline{\theta}_i} \mu(\bar{V}_i')}.$$

Recall that the $\bar{V}_i$'s are connected collections of tiles. Thus we can reason exactly as in the previous case : $\partial \bar{V}_i$ is a collection of segments, which are bissectors of points such as $\xi_{i_l}, \xi_{j_{l'}}$, where ξ_{i_l} is in the group i and $\xi_{j_{l'}}$ is in an other group. Therefore $d_{i_l, j_{l'}} > d_1$ and the upper bound of $h(y)$ is still valid. We thus obtain that

$$f'(x) \le f^U(x).$$

Besides,

$$\begin{aligned}
\int_E f^U(x) d\mu(x) \quad &= \quad e^{2\delta'} \frac{\sum_{i=1}^q e^{\bar{\theta}_i} [\mu(\bar{V}_i) + \mu(\bar{B}_i)]}{\sum_{i=1}^q e^{\underline{\theta}_i} [\mu(\bar{V}_i) - \mu(\bar{B}_i)]} \le 1 + \varepsilon \\
&\Leftrightarrow \\
\sum_{i=1}^q \mu(\bar{B}_i) e^{\underline{\theta}_i} \left[e^{\bar{\theta}_i - \underline{\theta}_i} + (1+\varepsilon) e^{-2\delta'} \right] \quad &\le \quad \sum_{i=1}^q \mu(\bar{V}_i) e^{\underline{\theta}_i} \left[(1+\varepsilon) e^{-2\delta'} - e^{\bar{\theta}_i - \underline{\theta}_i} \right].
\end{aligned}$$

The above inequality is satisfied in particular if

$$\mu(\bar{B}_i) \le \mu(\bar{V}_i) \frac{(1+\varepsilon) e^{-2\delta'} - e^{\bar{\theta}_i - \underline{\theta}_i}}{e^{\bar{\theta}_i - \underline{\theta}_i} + (1+\epsilon) e^{-2\delta'}}, \tag{12}$$

for all $i = 1, ..., q$. when $\delta' = \varepsilon/3$,

$$1 + 2\varepsilon/3 \ge (1+\varepsilon) e^{-2\delta'} \ge 1 + \varepsilon/6.$$

We also have that

$$e^{\bar{\theta}_i - \underline{\theta}_i} \le \exp n^{-r_0} \le 1 + \varepsilon/12,$$

when n is large enough. Therefore, if $\delta' = \varepsilon/3$,

$$\mu(\bar{B}_i) \le \mu(\bar{V}_i) \frac{\epsilon}{24}$$

when n is large enough and ε small enough. $\bar{V}_i$ is a collection of tiles, say V_{i_l}, $i_l \in \mathcal{I}_i \subset \{i_1, ..., i_{n_i}\}$ that are bordering on ∂V_i plus possibly other tiles. Therefore

$$\mu(\bar{V}_i) \ge \sum_{i_l \in \mathcal{I}_i} \mu(V_{i_l}) = \sum_{j \sim \mathcal{I}_i} \frac{L_{i_j j} d_{i_j j}}{2} \ge d_1 \sum_{j \sim \mathcal{I}_i} \frac{L_{i_j j}}{2},$$

where $j \sim \mathcal{I}_i$ means that ξ_j does not belong to $\bar{V}_i$, that $\xi_{i_j} \in \bar{V}_i$ and that $\xi_j \sim \xi_{i_j}$. Similarly

$$\mu(\bar{B}_i) = 2h \sum_{j \sim \mathcal{I}_i} L_{i_j j}.$$

Thus, by choosing $h = d_1 \varepsilon / 96$, i.e.

$$\delta = \frac{\varepsilon d_1^2}{192 L} = c \varepsilon d_1^2,$$

(12) is satisfied and $\int f^U(x)d\mu(x) \leq 1 + \varepsilon$.

References

Arjas, E., Gasberra, D. (1994) Nonparametric Bayesian inference from right censored survival data, using the Gibbs sampler. *Statistica Sinica*, **4**, 505–524.

Arjas, E., Heikkinen, J. (1997) An algorithm for nonparametric Bayesian estimation of a Poisson intensity. *Computational Statistics*, **12**, 385–402.

Arjas, E., Heikkinen, J. (1998) Non–parametric Bayesian estimation of a spatial Poisson intensity. *Scandinavian Journal of Statistics*, **25**, 435–450.

Arjas, E., Heikkinen, J. (1999) Modeling a Poisson forest in variable elevations : a nonparametric Bayesian approach. *Biometrics*, **55**, 738–745.

Barron, A. Schervish, M.J. and Wasserman, L. (1999) The Consistency of posterior distribution in nonparametric problems. *Annals of Statist.*, **27**, 2, 536–561.

Boots, A., Chiu, S.N., Okabe, A. and Sugihara, K. (1999) *Spatial Tesselations*, Wiley series in probability and statistics.

Denison, D.G.T., Mallick, B.K. and Smith, A.F.M. (1998) Automatic Bayesian curve fitting. *J.R. Statist. Soc. B*, **60**, 2, 333–350.

Diaconis, P., Freedman, D. (1986) On the consistency of Bayes estimates. *Annals of Statist.*, **14**, 1, 1–26.

Green, P. (1995) Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, **82**, 4, 711–732.

Hartigan, J.A. (1975) *Clustering algorithms.* New York: Wiley.

Hartigan, J.A. (1987) Estimation of a convex density contour in two dimensions. *J. Am. Statist. Ass.*, **82**, 267–270.

Heikkinen, J. (1998) Curve and surface estimation using dynamic step functions. In *Practical Nonparametric and semiparametric Bayesian statistics* (eds D. Dey, P. Müller and D. Sinha), pp. 255–272, N.Y : Springer–Verlag.

Müller, D.W., Sawistzki, G. (1991a) Using excess mass estimates to investigate the modality of a distribution. In *The Frontiers of Statistical Scientific Theory & Industrial Applications*, **26** of the American Series in Mathematical and Management Sciences, pp. 355–382, American Sciences Press, N.Y.

Müller, D.W., Sawistzki, G. (1991b) Excess mass estimates and tests for multimodality. *J. Am. Statist. Ass.*, **86**, 415, 738–746.

Nicholls, G. (1998) Bayesian image analysis with Markov chain Monte Carlo and colored continuum triangulation models. *J.R. Statist. Soc. B*, **60**, 3, 643–659.

Nolan, D. (1991) The excess mass ellipsoid. *J. Multiv. Analysis*, **39**, 348–371.

Polonik, W. (1995) Measuring mass concentration and estimating density contours clusters-an excess mass approach. *Ann. of Statist.*, **23**, 3, 855–881.

Tsybakov, A.B. (1997) On nonparametric estimation of density level sets. *Ann. of Statist.*, **25**, 3, 948–969.