Analysis of residuals from segmentation of noisy images

Rasmus Waagepetersen

Abstract

The prior model or penalizing term in Bayesian image analysis is typically a Markov random field parametrized by one or more smoothing parameters. For many commonly applied Markov random field penalizing terms there do not exist both objective and practically applicable methods for choosing the smoothing parameters. In this paper we discuss approaches to analysis of residuals in a simple case of image segmentation, and study to which extent such an analysis can provide information on whether a suitable value of the smoothing parameter has been used.

Keywords: Bayesian image analysis, oversmoothing, penalized likelihood, residuals.

1 Introduction

Segmentation of a digitized image corrupted by noise is an intensively studied problem of statistical inference. The research in this area was in the statistical community initiated by the papers by Geman & Geman (1984) and Besag (1986). A common feature of the methods introduced in these papers is that the segmentation is obtained by maximizing a posterior density where the prior is a Markov random field (MRF) which penalizes undesirable behaviour of the estimate. The prior models are typically not realistic representatives of the variability in the true image, and it therefore often seems more appropiate to consider the object function for the maximization as a penalized likelihood rather than a posterior density. A standard example of a penalizing term is the Ising model (see Section 2.1) which is used to penalize binary image estimates with excessive presence of boundary between pixels of different colours or classes.

Unfortunately there exists no both rigorous and practically applicable method for choosing the smoothing parameter in the Ising model (see e.g. the discussion in Dinten et al., 1991). The importance of this problem is stressed by the results in Greig, Porteous & Seheult (1989) which show, that the penalized likelihood estimate is quite sensitive to the choice of smoothing parameter, when the Ising model is used as the penalizing term. It seems fair to say that lack of applicable methods for choosing smoothing parameters is a general problem when non-Gaussian MRF's are used for penalized likelihood estimation. Various methods for choosing the smoothing parameter in Gaussian MRF penalizing terms are discussed in e.g. Thompson, Brown, Kay & Titterington (1991) and Künsch (1994). Estimation of prior parameters from training data may be a useful approach in a genuine Bayesian framework where a realistic prior modelling is applied, see Møller & Waagepetersen (1996).

The image to be estimated is usually a very complex quantity. In contrast often quite simple models are used for the noise which has corrupted the original image. In Section 3 of this paper we use the simple structure of frequently applied noise models to propose methods for analyzing the residuals obtained from penalized likelihood estimates and we study whether these methods are useful for detection of over or undersmoothing. Our basic example is the case where the Ising model is used for estimation of a binary image. It should however be noted that the methods studied are also applicable in situations where more advanced boundary models (see e.g. Grenander & Miller, 1994) are used. First, in Section 2, some definitions are given, and we discuss some ad hoc guidelines for the choice of the smoothing parameter.

2 Definitions and initial remarks

The true image is represented by a parameter $\theta \in S = \{0, \dots, k-1\}^L$ where k is the number of different colours/classes in the image and $L = \{0, \dots, N-1\} \times \{0, \dots, M-1\}$ is a rectangular lattice. It is for simplicity in the following assumed that $k = 2$ so that binary images are considered, and two different models for the observed image X are discussed:

I $X_i = (\theta_i + \eta_i) \bmod 2$, $i \in L$, where the η_i's are independent with $P(\eta_i = 1) = 1 - P(\eta_i = 0) = q$, $q \in]0, 1[$ (binary white noise).

II $X_i = \theta_i + \epsilon_i$, $i \in L$ where $\epsilon = (\epsilon_i)_{i \in L} \sim N(0, \sigma^2 I)$, $\sigma^2 > 0$ (additive Gaussian white noise).

The penalized likelihood then takes the form

$$pl(\theta; x) \propto p(\theta; \beta) \prod_{i \in L} f(x_i; \theta_i)$$

where $f(x_i; \theta_i)$ is the likelihood of x_i given that the true pixel value is θ_i, $p(\cdot; \beta)$ is the smoothing term, and β is the smoothing parameter. We will concentrate on the case where $p(\cdot; \beta)$ is the second order Ising model on L, i.e.

$$p(\theta; \beta) \propto \exp(-\beta \sum_{\substack{\{i,j\} \subseteq L: \\ i \sim j}} 1(\theta_i \neq \theta_j)), \ \theta \in S, \tag{1}$$

where $\sim$ is the second order neighbour relation on L, so that $i \sim j \Leftrightarrow ||i-j|| < 2$. For smoothing purposes values of $\beta > 0$ are always used. For a pixel $i \in L$, $N_i = \{j \in \mathbf{Z}^2 | i \sim j\}$ is the $\sim$-neighbourhood of i, and for any subset C of $\mathbf{Z}^2$ we let $\dot{C} = \{i \in C | N_i \subseteq C\}$ denote the "interior" of C. Further, for each $\theta \in \{0,1\}^{\mathbf{Z}^2}$ we define the relation $\sim_\theta$ on $\mathbf{Z}^2$ by

$$i \sim_\theta j \Leftrightarrow \exists\, i_1, \dots, i_n \in \mathbf{Z}^2 : i = i_1 \sim i_2 \sim \cdots \sim i_n = j \text{ and } \theta_{i_1} = \cdots \theta_{i_n},$$

and let $\mathcal{K}(\theta)$ be the set of maximal nonempty cliques given by $\sim_\theta$. Hence $\mathcal{K}(\theta)$ becomes the set of maximal connected components of constant colour in the image θ. The set of maximal connected components of colour 1 is denoted $\mathcal{K}^{(1)}(\theta)$. Finally, $\mathcal{K}$ is the set of connected subsets of $\mathbf{Z}^2$, i.e. $K \in \mathcal{K}$ if and only if for each $i, j \in K$ there exists $i_1, \dots, i_n \in K$ such that $i = i_1 \sim i_2 \sim \cdots \sim i_n = j$.

2.1 Geometrical characteristics of connected components and choice of smoothing parameter

In this section we shall consider a certain geometrical characteristic for connected components generated by a binary white noise, and study how this characteristic may be useful for constructing guidelines for the choice of smoothing parameter. Later on (in Section 3.3 and Section 3.4.3) it turns out that the empirical distribution for this characteristic is also of interest as a summary statistic for detecting under or oversmoothing.

Let $\theta^{(1)}$ be the image consisting entirely of zeros, and let $\theta^{(2)}$ be a "noisy" version of $\theta^{(1)}$ given by $\theta_i^{(2)} = 1$, $i \in A$, and $\theta_i^{(2)} = \theta_i^{(1)}$, $i \in L \setminus A$, where $A \in \mathcal{K}$ is a subset of L. Let further $bd(A) = \#\{(i,j) \in A \times (\mathbf{Z}^2 \setminus A) | i \sim j\}$ denote the length of the boundary of the set A measured in terms of the number of $\sim$-adjacent pairs of pixels where the first pixel is in A, and the second pixel is outside A. Assume in order to avoid edge effects that $A \subseteq \dot{L}$. Then in case I,

$$pl(\theta^{(1)}; \theta^{(2)}) > pl(\theta^{(2)}; \theta^{(2)}) \Leftrightarrow \beta^{-1} < -\frac{bd(A)}{|A|} \log(\frac{q}{1-q})^{-1}, \qquad (2)$$

where $|A|$ denotes the number of pixels in A. For given q and β this means, that it is determined by the ratio $bd(A)/|A|$ whether the "true image" $\theta^{(1)}$ has a higher penalized likelihood than $\theta^{(2)}$ given the "data" $\theta^{(2)}$.

In order to be able to clean up homogeneous regions of an image the inequality (2) must hold with high probability when A is substituted by the random set $\tilde{A}$ distributed as the typical noise component which will now be defined. Let η be a binary white noise on $\mathbf{Z}^2$ with error rate q, i.e. η_i, $i \in \mathbf{Z}^2$ *iid* and $P(\eta_i = 1) = 1 - P(\eta_i = 0) = q$, $0 < q < 1$. There then exists a value q_c of q such that all members of $\mathcal{K}^{(1)}(\eta)$ are finite a.s. whenever $q < q_c$, see e.g. Grimmett (1989). The value q_c is the socalled critical value for site percolation on the graph given by $\mathbf{Z}^2$ and $\sim$. Consider any nonempty finite subset K of $\mathbf{Z}^2$ and let

$t_1(K) = \min\{i \in \mathbf{Z} | \exists j \in \mathbf{Z} : (i,j) \in K\}$, $t_2(K) = \min\{j \in \mathbf{Z} | (t_1(K), j) \in K\}$ and $t(K) = (t_1(K), t_2(K))$. Then $t(K)$ is the down and leftmost corner of K. Suppose now that $q < q_c$ and define the random set $\tilde{K}$ to be equal to the component $K \in \mathcal{K}^{(1)}(\eta)$ with $t(K) = (0,0)$, if such a component exists in $\mathcal{K}^{(1)}(\eta)$, and let $\tilde{K} = \emptyset$ otherwise. The typical component of $\mathcal{K}^{(1)}(\eta)$ is then distributed as $\tilde{K}|\tilde{K} \neq \emptyset$. It is known (see Kesten, 1982) that $q_c + p_c = 1$, where p_c is the critical value for site percolation on the graph given by $\mathbf{Z}^2$ and the first order neighbour relation on $\mathbf{Z}^2$. It is furthermore believed that p_c is close to 0.59, but the exact value is not known. The lower bound 0.556 and upper bound 0.680 for p_c is provided by van den Berg & Ermakov (1996) and Wiermann (1995), respectively. All $K \in \mathcal{K}^{(1)}(\eta)$ are thereby finite with probability one whenever $q < 0.320 < q_c$.

The analytic calculation of the distribution of $bd(\tilde{A})/|\tilde{A}|$ is difficult and involves complicated combinatorics. The distribution is however quite easy to estimate unbiasedly by using Monte Carlo and plus sampling. To do this one estimates the empirical distribution of $bd(K)/|K|$, $K \in \{K \in \mathcal{K}^{(1)}(\eta) | t(K) \in L'\}$ for some fixed finite lattice $L' \subseteq \mathbf{Z}^2$. Note that all $K \in \mathcal{K}^{(1)}(\eta)$ with $t(K) \in L'$ will be observable within a sufficiently large finite lattice $\tilde{L}(\eta)$ containing L', so that one only needs to simulate η on the finite lattice $\tilde{L}(\eta)$. The size of $\tilde{L}(\eta)$ of course depends on the actual realization of η. It is required that q is known or that a precise estimate of q is available. If e.g. $q = 0.15$ then $bd(\tilde{A})/|\tilde{A}| < 5.40$ with a probability less than 1%, so unless the boundaries in the true image are very rough, then by (2), the value 0.33 for β should be a suitable lower limit for the values of β to be used. Similarly, if $q = 0.25$, $bd(\tilde{A})/|\tilde{A}| < 4.74$ with a probability less than 1% so that an lower limit for β becomes 0.23.

A similar rule can be deduced in case II. Suppose again that $\theta^{(1)}$ is the image consisting entirely of zeros, and let $\hat{\theta}(x)$ be the maximum likelihood estimate given a noisy version x of the original image $\theta^{(1)}$. Then $\hat{\theta}(x)$ is a realization of a binary white noise with rate $\tilde{q}$, where $\tilde{q} = P(X_i > 0.5)$ is the probability of making a misclassification of a pixel with the maximum likelihood estimate. Suppose that $A \in \mathcal{K}^{(1)}(\hat{\theta}(x)) \neq \emptyset$, and let $\theta_i^{(2)} = 1$, $i \in A$, and $\theta_i^{(2)} = \theta_i^{(1)}$, $i \in L \setminus A$. Then,

$$pl(\theta^{(1)}; x) > pl(\theta^{(2)}; x) \Leftrightarrow \beta^{-1} < -\frac{bd(A)}{\sum_{i \in A} \log(\frac{f(x_i;0)}{f(x_i;1)})} \tag{3}$$

This inequality together with simulations of the Gaussian noise process may in a similar manner as described above be used to obtain a lower bound for choices of β.

Similar guidelines for the choice of smoothing parameter are discussed in Ripley (1988) and Künsch (1994). These authors however do not utilize the characteristics of the particular noise process when determining lower bounds for the smoothing parameter. An upper bound could be provided as in Künsch (1994) by specifiying a set A for which the opposite inequality of (2) should hold.

Finally one may notice that the notion of connected components is also useful when constructing algorithms for maximization of the penalized likelihood. A well known iterative maximization procedure is the ICM-algorithm (Besag, 1986) which tends to get stuck in local minima because it only updates one pixel at a time. Subsets of pixels suitable for simultaneous updating can be identified as small connected components as demonstrated in Waagepetersen (1995).

3 Analysis of residuals

We shall now discuss various approaches to analysis of the residuals from an image segmentation. The residual image will be denoted R and the penalized likelihood estimate $\hat{\theta}(X)$, so that in the case I (binary white noise), $R = (X + \hat{\theta}(X)) \bmod 2$ and in case II (Gaussian noise), $R = X - \hat{\theta}(X)$. The methods considered are based on the assumption that R behaves similar to the noise process when precise estimates are obtained. In the experiments we shall try to use the methods to choose among estimates obtained with varying values of the smoothing parameter. We shall furthermore investigate how robust the methods are towards the fact that very precise restorations can not always be obtained, so that the assumption of similarity between R and the noise process becomes less realistic. In this section we will mainly focus on case I, but by considering the binary random fields $\tilde{R} = (1(R_i > 0))_{i \in L}$ or $R' = (1(|R_i| > c))_{i \in L}$, $c > 0$, the methods developed may also be applicable in case II, see subsection 3.4.4. Note that in practice, oversmoothing will usually be more problematic than undersmoothing, since oversmoothed estimates may be quite appealing visually.

3.1 Test for independence

If $\hat{\theta}(X)$ is an oversmoothed estimate one may expect that clusters of 1's will appear in the residual image on locations where components in the true image have been smoothed away. A test for spatial independence may therefore be useful for identifying oversmoothed images. It is not possible to determine the true distribution of the residuals, and instead we shall apply the external field Ising model given by

$$p(r; \alpha, \gamma) \propto \exp(\alpha \sum_{i \in L} r_i + \gamma \sum_{\substack{\{i,j\} \subseteq L: \\ i \sim j}} 1(r_i \neq r_j)), \; r \in \{0,1\}^L$$

as a simple model for the residuals, where the possible spatial correlation is modelled through the parameter γ. If a perfect estimate is obtained then this Ising model is the true residual distribution when $\alpha = \log(q/(1-q))$ and $\gamma = 0$.

The likelihood ratio test for $\gamma = 0$ will provide a test for over or undersmoothing. Asymptotic distribution results for the likelihood ratio test are provided within the following framework. Let G be the set of probability measures on $\{0,1\}^{\mathbf{Z}^2}$ whose conditional distributions on nonempty finite subsets of $\mathbf{Z}^2$ agree with the conditional distributions of Ising models defined on finite lattices, and let $\partial C = \{j \in \mathbf{Z}^2 \setminus C \mid \exists\, i \in C : i \sim j\}$ be the set of neighbouring pixels for any $C \subseteq \mathbf{Z}^2$. Then more precisely, $p(\cdot) \in G$ if and only if for every $\Lambda \subseteq \mathbf{Z}^2$ with $0 < |\Lambda| < \infty$ we have that the corresponding conditional probability of $p(\cdot)$ is given by

$$p_\Lambda(r_\Lambda | r_{\mathbf{Z}^2 \setminus \Lambda}; \alpha, \gamma) = p_\Lambda(r_\Lambda | r_{\partial\Lambda}; \alpha, \gamma) =$$
$$c(\alpha, \gamma, r_{\partial\Lambda}) \exp(\alpha \textstyle\sum_{i\in\Lambda} r_i + \gamma \sum_{\substack{\{i,j\}\subseteq \mathbf{Z}^2: \\ i\sim j \text{ and } \{i,j\}\cap\Lambda\neq\emptyset}} 1(r_i \neq r_j)),\ r \in \{0,1\}^{\mathbf{Z}^2},$$

where $c(\alpha, \gamma, r_{\partial\Lambda})$ is a normalizing constant, and for any subset C of $\mathbf{Z}^2$, $r_C = (r_i)_{i\in C}$. The set G is always nonempty. Background material on this question and on when $|G| = 1$ is provided in e.g. Kindermann & Snell (1980) or Guyon (1995).

Let $(L_n)_{n\geq 1}$ be a strictly increasing sequence of finite subsets of $\mathbf{Z}^2$ where $|L_n \setminus \dot{L}_n|/|L_n| \to 0$. Given an observation $r \in \{0,1\}^{\mathbf{Z}^2}$, let further $U_n^1(r)$ be the maximum of the conditional likelihood $p_{L_n}(r_{L_n} | r_{\mathbf{Z}^2\setminus L_n}; \alpha, \gamma)$ and $U_n^0(r)$ be the maximum of the same conditional likelihood under the null hypothesis $\gamma = 0$. Since $|G| = 1$, trivially, under the null hypothesis, we by theorem 5.3.2 in Guyon (1995) have the following asymptotic result:

$$-2\log(U_n^0(R)/U_n^1(R)) \rightrightarrows \chi^2(1).$$

In the case where we have observed r on the finite window L, we can use this result by basing inference on the conditional likelihood $p_{\dot{L}}(r_{\dot{L}} | r_{\partial\dot{L}})$ which is identical to the conditional likelihood with respect to $\dot{L}$ of the Ising model defined on the infinite lattice.

In order to estimate α and γ under the initial hypothesis and to calculate the likelihood ratio we are left with the problem that the normalizing constant $c(\alpha, \gamma, r_{\partial\dot{L}})$ is not available in closed form. This problem can however be resolved by the Markov chain Monte Carlo methods developed in Geyer & Thompson (1992) and Geyer (1991). It is enough to know the normalizing constant up to a constant of proportionality. With p_{imp} an appropiately chosen importance sampling distribution with density say

$$p_{imp}(x) = ch(x),\ x \in \{0,1\}^{\dot{L}},$$

where c is a possibly unknown normalizing constant, we can estimate $(c(\alpha, \gamma; r_{\partial\dot{L}})/c)^{-1}$ by the empirical average

$$\sum_{i=1}^{n} \exp(\alpha \sum_{i\in\dot{L}} Y_i^n + \gamma \sum_{\substack{\{i,j\}\subseteq\dot{L}: \\ i\sim j}} 1(Y_i^n \neq Y_j^n) + \gamma \sum_{\substack{(i,j)\in\dot{L}\times L\setminus\dot{L}: \\ i\sim j \text{ and } i<j}} 1(Y_i^n \neq r_j))/h(Y^n),$$

where $(Y^n)_{n\geq 1}$ is an ergodic sample from the importance sampling distribution. Such a sample can e.g. be generated by using the Metropolis-Hastings algorithm

(Hastings, 1970). The normalizing constant $c(\alpha, 0, r_{\partial \dot{L}})$ is in this case actually known and equal to $1/(1+\exp(\alpha))^{|\dot{L}|}$, but if c is unknown we in order to be able to calculate the likelihood ratio also need to estimate $c(\alpha, 0, r_{\partial \dot{L}})/c$ in the same way as in the case $\gamma \neq 0$.

I found rough approximations $\tilde{\alpha}$ and $\tilde{\gamma}$ to the mle's, and then used an equal mixture of $p_{\dot{L}}(\cdot|r_{L\setminus\dot{L}}; \tilde{\alpha}, \tilde{\gamma})$ and $p_{\dot{L}}(\cdot|r_{L\setminus\dot{L}}; \hat{\alpha}_0, 0)$ as the importance sampling distribution where $\hat{\alpha}_0$ is the mle under the null hypothesis. Then the support of the importance sampling distribution covers the support of the Ising model both under the initial hypothesis and under the null hypothesis, so that the normalizing constant can be estimated accurately in both cases. A solution to the practical and theoretical problems of using this mixture is provided in Geyer (1991). The initial estimates $\tilde{\alpha}$ and $\tilde{\gamma}$ were found by using pseudo maximum likelihood estimation (Besag, 1975) or stochastic approximation (Younes, 1988).

3.2 Tests based on local configurations of the noise process

Let for any $x \in S$, $A(x) = \{i \in \dot{L} \mid x_i = 1\}$, and let for any $x \in \{0,1\}^{\mathbf{Z}^2}$, $B_i(x) = B_i(x_{N_i}) = x_i \sum_{j \in N_i}(1 - x_j)$, $i \in \mathbf{Z}^2$. Then $B_i(x)$ is the number of 0-neighbours of the pixel i when $x_i = 1$, and $B_i(x)$ is zero otherwise. Let further $\chi_i^{(k)}(x) = 1(x_i = 1, B_i(x) = k)$, $k = 0, \ldots, 8$, $i \in L$. We shall in this section consider the statistic $U = \sum_{i \in \dot{L}} B_i(R)/|A(R)| = bd(A(R))/|A(R)|$ which is the boundary to area ratio of the set of 1-pixels given by $R_{\dot{L}}$, and the statistics $U_k = \frac{\sum_{i \in \dot{L}} \chi_i^{(k)}(R)}{|A(R)|}$, $k = 0, \ldots, 8$, which measure the fractions of 1-pixels with k neighbours of opposite colour.

Test for over or undersmoothing may now be deviced by comparing these statistics with their expected values under the assumption that R is a binary white noise. Approximate distribution results for the statistics U and U_k, $k = 0, \ldots, 8$, may be derived as follows: Let as in subsection 2.1 η be a binary white noise on $\mathbf{Z}^2$ with with error rate q, and let $\ddot{C} = \{i \in C \,|\, \forall j \in C : N_i \cap N_j \neq \emptyset \Rightarrow N_j \subseteq C\}$ for any $C \subseteq \mathbf{Z}^2$, so that $\ddot{C}$ is the set of pixels for which the neighbourhood of each neighbour is contained in C. Let further $b(\cdot; n, q)$ denote the probabilities of a binomial distribution with parameters n and q, and let the random fields $B = (B_i)_{i \in \mathbf{Z}^2}$ and $\chi = (\chi_i^{(k)})_{i \in \mathbf{Z}^2}$, $k = 0, \ldots, 8$, on $\mathbf{Z}^2$ be given by $B_i = B_i(\eta) = \eta_i \sum_{j \in N_i}(1 - \eta_j)$ and $\chi_i^{(k)} = \chi_i^{(k)}(\eta) = 1(\eta_i = 1, B_i(\eta) = k)$, $k = 0, \ldots, 8$, $i \in \mathbf{Z}^2$. For these random fields we have the following results:

Theorem 1 *Let $(L_n)_{n \geq 1}$ be a strictly increasing sequence of subsets of $\mathbf{Z}^2$ such that $(|L_n \setminus \ddot{L}_n|)/|L_n| \to 0$. Then*

$$|L_n|^{-1/2} \sum_{i \in L_n} (B_i - \mu(q)) \rightrightarrows N(0, \sigma^2(q))$$

where $\mu(q) = E(B_i) = 8q(1-q)$*, and*

$$\sigma^2(q) = \sum_{\substack{j\in\mathbf{Z}^2: \\ N_0\cap N_j\neq\emptyset}} Cov(B_0, B_j) = 64q - 240q^2 + 352q^3 - 176q^4.$$

Further,

$$|L_n|^{-1/2}\sum_{i\in L_n}\left(\frac{\chi_i^{(k)}}{q} - \mu_{(k)}(q)\right) \stackrel{\sim}{\rightarrow} N(0, \sigma^2_{(k)}(q)),\ k = 0,\ldots,8,$$

where $\mu_{(k)}(q) = E(\chi_0^{(k)}/q) = P(B_i = k|\eta_i = 1) = b(k; 8, 1-q)$*, and*

$$\sigma^2_{(k)}(q) = \sum_{\substack{j\in\mathbf{Z}^2: \\ N_0\cap N_j\neq\emptyset}} Cov(\chi_0^{(k)}, \chi_j^{(k)})/q^2 =$$

$$\mu_{(k)}(q)/q + 4(C_1(k,q) + 2C_2(k,q) + C_3(k,q) + C_4(k,q) + C_5(k,q)) - 25\mu_{(k)}(q)^2$$

with $C_n(k,q) = \sum_{m=0}^{n} b(m; n, 1-q)b(k-m; 8-n, 1-q)^2$*,* $n = 1, 2, 3$*, and* $C_n(k,q) = \sum_{m=0}^{n-2} b(m; n, 1-q)b(k-m-1; 8-n+1, 1-q)^2$*,* $n = 4, 6$*.*

Proof: Since B_i and B_j (and $\chi_0^{(k)}$, $\chi_j^{(k)}$) are independent whenever $|N_i\cap N_j| = \emptyset$, $i, j \in \mathbf{Z}^2$, the results follows easily by the central limit theorem for random fields (see e.g. Guyon, 1995). Here $C_n(k,q)$ is $Cov(\chi_0^{(k)}, \chi_j^{(k)})/q^2$ in the case where $|N_i\cap N_j| = n$, $n = 1, 2, 3, 4, 6$. □

Note that $|\dot{L}|^{-1}bd(A(\eta)) = |\dot{L}|^{-1}\sum_{i\in\dot{L}} B_i$. Since the quantity $|\dot{L}|^{-1}\sum_{i\in\dot{L}}\eta_i$ converges a.s. to q as $|L|\to\infty$ and $|A(\eta)| = \sum_{i\in\dot{L}}\eta_i$, the theorem above yields that for large N and M,

$$bd(A(\eta))/|A(\eta)| \dot{\sim} N(\mu(q)/q, \sigma^2(q)/(q^2|\dot{L}|))$$

Similarly,

$$\frac{\sum_{i\in\dot{L}}\chi_i^{(k)}}{|A(\eta)|} \dot{\sim} N(\mu_k(q), \sigma^2_{(k)}(q)/|\dot{L}|),\ k = 0,\ldots,8.$$

When applying tests based on these approximative results one may in case I either insert the value of q used in the likelihood, or alternatively, reestimate q from the residual image. However in case II, when considering $\tilde{R}$, we actually know that we should use $q = 0.5$.

One may finally note that it is also possible to calculate the asymptotic covariance-matrix $\Sigma(q)$ for $(U_0,\ldots,U_8)$ and then consider the statistic $||(U_0 - \mu_0(q),\ldots,U_8 - \mu_8(q))\Sigma(q)^{-1/2}||^2$ which is asymptotically $\chi^2(9)$-distributed.

3.3 Analysis based on geometrical characteristics of connected components

Let for any $x \in S$, the summary statistic $G(x)$ be given by the empirical distribution of the boundary to area ratios $bd(K)/|K|$, $K \in \mathcal{K}^{(1)}(x)$. In order to decide whether $G(R)$ indicates over or undersmoothing one may compare $G(R)$ with envelopes obtained by computing G from a number of simulations of the noise process. This approach is similar to the one proposed in Diggle (1983) for point process summary statistics. An estimate of the error rate q is of course needed in order to perform the simulations.

Note that, because of edge effects, the summary statistic mentioned here is not an unbiased estimator of the distribution of the boundary to area ratio for the typical component of the noise process discussed in section 2.1.

3.4 Experiments

A general problem concerning penalized likelihood estimation of images using non-Gaussian MRF penalizing terms is that statistical properties of the penalized likelihood estimator and the residuals can not be evaluated analytically. The methods proposed must therefore be judged in terms of experiments using simulated data. For this purpose I have considered segmentations of noisy versions of the two (supposingly unobserved) images depicted in Figure 1. The first image is a simple artificial image previously considered in Greig et al. (1989) while the second more complex image is a thresholded version of an microscopic image of a rock sample. In the experiments the images are corrupted by a binary white noise with a 15% error rate, and penalized likelihood estimates are obtained with varying values of the smoothing parameter, and the true value of the error rate in the likelihood. The Ising model (1) is used as the penalizing term, and simulated annealing (Kirkpatrick et al., 1983) is used for the maximization of the penalized likelihood.

Figure 1: Test images.

3.4.1 Test for independence

I have applied the test for independent residuals to residuals from segmentations of both the first and second image. Maximum likelihood estimates of γ, test probabilities based on the asymptotic $\chi^2(1)$-distribution of the likelihood ratio test, and rates of misclassified pixels for five repetitions of each experiment are reported in Table 1 and Table 2.

β	Er.	Rs.	$\hat{\gamma}$	ϵ	β	Er.	Rs.	$\hat{\gamma}$	ϵ
0.30	4.2	11.1	0.037	34.5	0.52	1.8	13.3	0.004	91.6
	5.0	11.4	0.018	64.6		1.7	13.8	-0.004	90.7
	5.0	10.9	0.057	16.1		1.4	13.6	-0.010	76.5
	5.2	10.5	0.020	65.6		2.3	13.2	-0.025	45.7
	5.2	10.5	0.032	1.9		2.3	13.5	0.022	28.5
0.40	2.6	12.3	-0.027	45.0	0.60	2.3	13.9	-0.066	3.3
	2.4	13.1	-0.015	67.1		2.1	14.4	-0.063	4.3
	2.5	12.7	-0.054	11.7		2.1	14.2	-0.076	1.2
	2.9	12.4	-0.046	18.0		2.7	13.6	-0.076	1.3
	2.9	12.3	0.003	90.6		3.8	14.9	-0.118	0.0
0.45	1.7	13.2	0.014	69.2	0.68	2.5	14.2	-0.099	0.0
	1.7	13.8	-0.009	79.6		2.8	15.2	-0.117	0.0
	1.4	13.6	-0.015	64.9		3.1	15.3	-0.157	0.0
	2.4	13.1	-0.007	85.0		2.9	13.6	-0.074	1.5
	2.4	13.5	0.025	11.9		4.6	15.7	-0.163	0.0

Table 1: Results from five repetitions of the experiment using the first test image. Er. is the rate of misclassified pixels and Rs. is the frequency of 1's in the residual image, both in %, and ϵ is the test probability (p-value).

For the first image the estimates obtained with $\beta = 0.30$ are clearly undersmoothed while the estimates for $\beta = 0.60$ and $\beta = 0.68$ are oversmoothed. The best estimates (in terms of visual appearance and rates of misclassified pixels) are obtained with $\beta = 0.52$, so in this case the lower bound 0.33 for β provided by the guideline described in subsection 2.1 was quite conservative. A sample of estimated images with different choices of the smoothing parameter is shown in Figure 2. Also the corresponding residual images and simulations of the fitted Ising models are shown. The loglikelihood ratio test clearly shows that the oversmoothed estimates obtained with $\beta = 0.60$ and $\beta = 0.68$ are unsatisfactory. Consider the upper right image in Figure 2. That this is an oversmoothed estimate is visually not obvious if the true image is unknown, but the likelihood ratio test shows that a suboptimal estimate has been achieved. Unfortunately the test for independence is not able to identify the undersmoothed estimates obtained with $\beta = 0.30$. If one assumes that the error rate is known, and tests

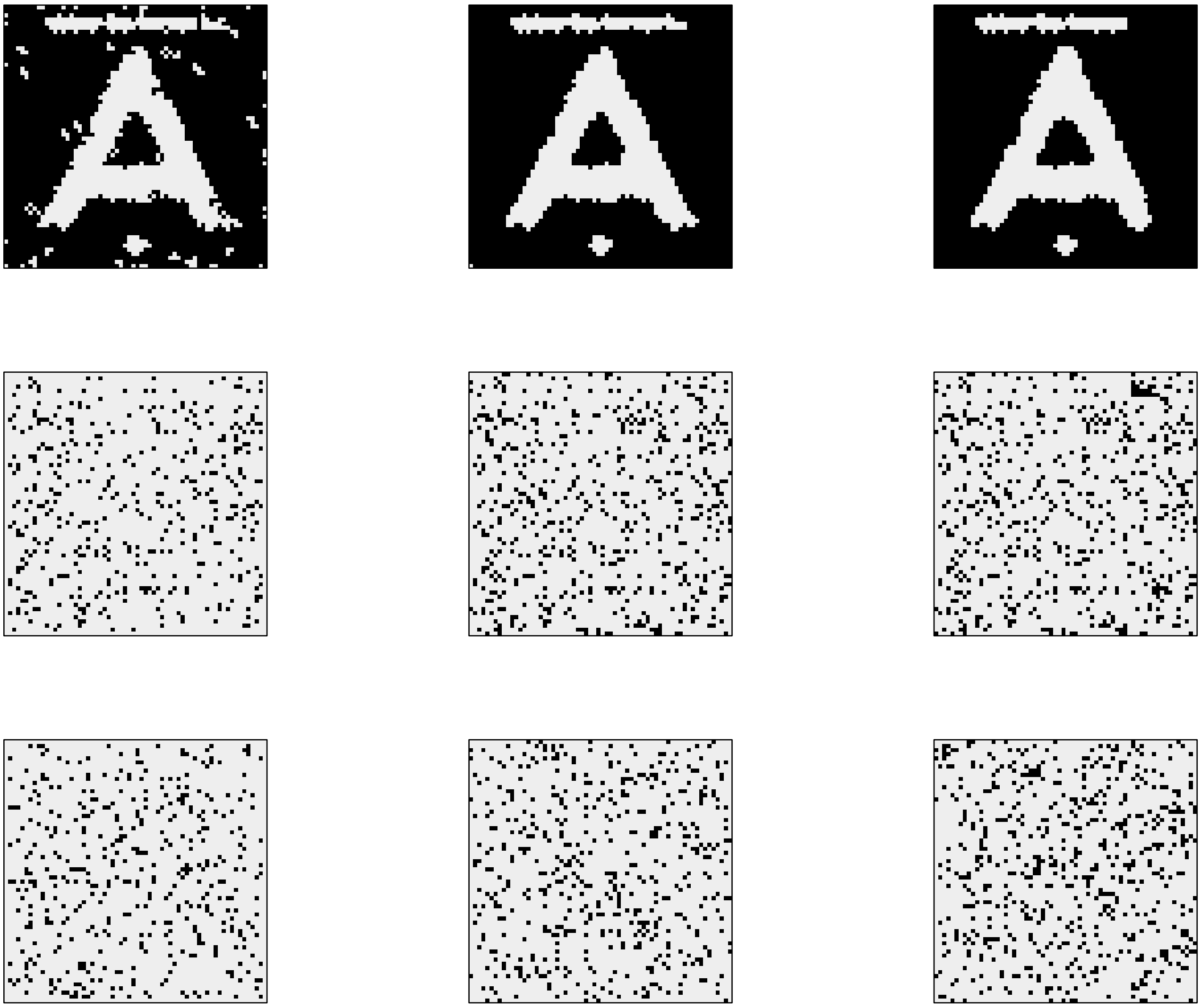

Figure 2: Upper row: estimates from the last repetition of the experiment, and β equal to 0.30, 0.52 and 0.60. Middle row: corresponding residuals. Lower row: simulations of the Ising model fitted to the residuals.

β	Er.	Rs.	$\hat{\gamma}$	ϵ	β	Er.	Rs.	$\hat{\gamma}$	ϵ
0.27	8.6	6.1	-0.17	0.3	0.45	5.0	13.7	-0.08	0.3
	8.8	6.2	-0.11	6.3		4.6	13.8	-0.03	42.0
	8.8	6.4	-0.15	1.3		4.6	13.2	-0.02	52.7
	9.0	6.2	-0.11	7.1		4.9	13.5	-0.06	4.7
	9.2	5.4	-0.14	0.0		4.4	12.6	-0.07	2.6
0.30	5.9	9.3	-0.04	36.2	0.52	5.2	14.3	-0.12	0.0
	5.7	9.8	-0.02	63.0		4.7	14.2	-0.06	3.3
	6.1	10.1	-0.00	94.1		4.8	13.6	-0.06	8.1
	6.7	9.3	-0.05	30.4		5.3	13.9	-0.09	0.4
	6.2	8.9	-0.05	0.8		4.2	12.8	-0.09	0.3
0.36	4.6	10.8	-0.07	6.3	0.60	5.9	15.0	-0.16	0.0
	4.7	11.2	-0.07	4.9		5.4	14.8	-0.11	0.0
	5.1	10.8	-0.05	22.9		5.8	14.4	-0.12	0.0
	4.8	11.4	-0.10	0.3		5.9	14.6	-0.12	0.0
	4.7	10.9	-0.12	0.0		5.3	14.0	-0.15	0.0

Table 2: Results from five repetitions of the experiment using the second test image. Er. is the rate of misclassified pixels and Rs. is the frequency of 1's in the residual image, both in %, and ϵ is the test probability (p-value).

the alternative null hypothesis $\gamma = 0$ and $\alpha = \ln(0.15/0.85)$, then very small test probabilities are obtained for all values of β.

Very precise estimates could be obtained for the first test image with error rates as low as 1.4% and the premises for application of the test for independent residuals were thus fulfilled. Things get more difficult for the second more complex image where it is hard to obtain such precise estimates. For the second image I have considered values of β ranging from 0.27 to 0.60, and the best estimates in terms of visual appearance and rates of misclassified pixels were in general obtained with $\beta = 0.45$. The test for independence often senses that even the best estimates are not perfect, and the undersmoothed estimates for $\beta = 0.30$ are in fact most often accepted. The test for independence may however still be useful for detecting oversmoothing since the estimates for $\beta = 0.60$ are always rejected with very small test probabilities.

3.4.2 Tests based on local configurations of the noise process

Here we consider only residuals obtained from segmentations of the second image. Assuming that the error rate is known, the statistic U described in Section 3.2 always rejects the undersmoothed estimate obtained with $\beta = 0.27$ on a 5% level of significance and yields small test probabilities for the undersmoothed estimate obtained with $\beta = 0.30$. The test probabilities for the best estimates obtained

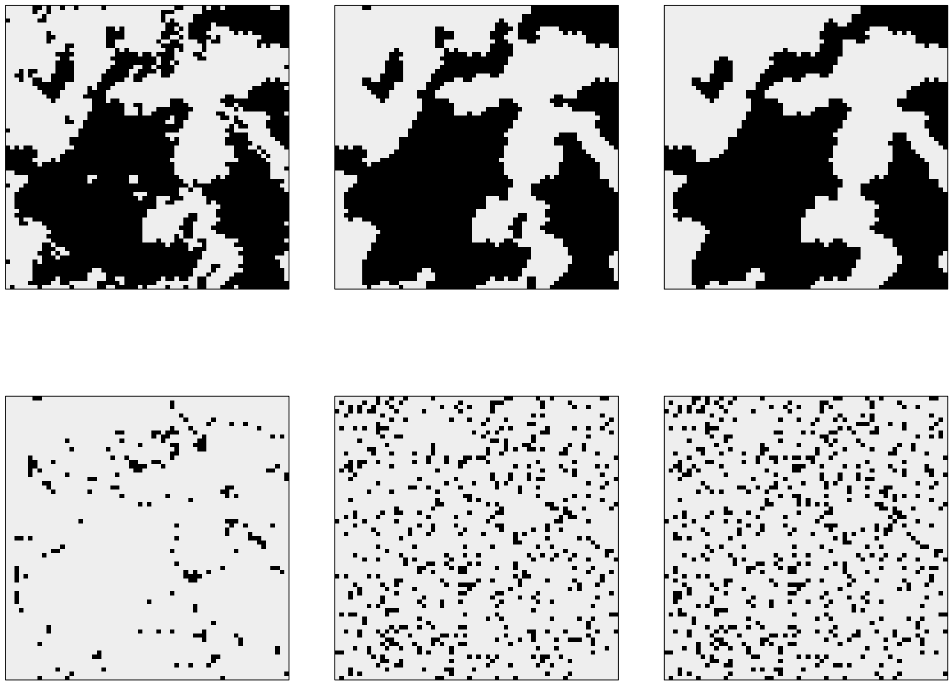

Figure 3: Upper row: estimates obtained with $\beta = 0.30$, $\beta = 0.45$ and $\beta = 0.60$ (last repetition). Lower left: misclassified pixels ($\beta = 0.45$). Lower middle: residual image ($\beta = 0.45$). Lower right: simulation of the noise process.

with β equal to 0.36, 0.45 or 0.52 are high, but unfortunately this is also the case for the oversmoothed estimate. The problem is presumably that the error due to the oversmoothing is averaged away in the statistic U (I have omitted the results obtained with this statistic).

Better results are obtained by considering the statistics U_k, $k = 0, \ldots, 8$. In Table 3 test probabilities for each k and for five repetitions of the experiment are shown. One sees that the tests based on U_k rejects the undersmoothed estimates for k large with very small test probabilities, while the oversmoothed estimates are rejected for k small. Small test probabilities are occasionally encountered also for the suitably smoothed estimates ($\beta = 0.36$ and $\beta = 0.45$), but the general impression is that the tests based on the U_k's are useful for detecting under and oversmoothing.

The results are less promising (see Table 4) when q is estimated by the frequency of 1's in the residual image, and this estimate of q is used instead of the true value of q. The oversmoothed images are still rejected, but it is now not possible to reject the undersmoothed estimates.

γ	0	1	2	3	4	5	6	7	8
-0.27	99.02	93.43	71.54	23.07	0.23	0.00	0.00	0.01	0.00
	99.02	93.43	71.54	23.07	0.23	0.00	0.00	9.24	0.00
	99.02	93.43	71.54	23.07	0.23	0.00	0.00	32.35	0.00
	99.02	93.43	71.54	23.07	0.23	0.00	0.00	1.35	0.00
	99.02	93.43	71.54	23.07	0.23	0.00	0.00	77.22	0.00
-0.30	99.02	93.43	71.54	23.07	0.23	0.00	0.00	4.86	0.00
	99.02	93.43	71.54	23.07	0.23	0.00	0.14	47.33	0.00
	99.02	93.43	71.54	23.07	0.23	0.00	0.05	34.47	0.00
	99.02	93.43	71.54	23.07	0.23	0.00	0.01	13.46	0.00
	99.02	93.43	71.54	23.07	0.99	0.00	0.00	98.90	0.00
-0.36	99.02	93.43	71.54	23.07	13.66	0.34	0.98	13.53	1.62
	99.02	93.43	71.54	88.59	12.47	1.22	7.81	62.26	1.41
	99.02	93.43	71.54	23.07	0.23	2.77	0.75	52.67	0.10
	99.02	93.43	71.54	23.07	5.27	96.83	15.44	78.53	16.51
	99.02	93.43	71.54	23.07	96.31	17.19	36.61	38.03	2.21
-0.45	99.02	93.43	71.54	2.51	0.06	75.43	8.11	97.59	76.07
	99.02	93.43	1.02	17.51	75.27	11.74	33.60	37.46	68.05
	99.02	93.43	0.00	14.16	14.61	13.37	33.57	58.23	3.91
	99.02	93.43	71.54	74.43	93.26	32.37	47.72	59.71	55.05
	99.02	93.43	71.54	23.07	51.94	52.17	72.67	14.56	5.32
-0.52	99.02	0.00	71.54	0.00	0.00	13.03	10.02	75.21	47.32
	99.02	93.43	0.00	0.00	25.74	75.45	19.83	85.59	91.72
	99.02	93.43	0.00	0.00	79.82	35.42	12.99	46.42	9.73
	99.02	93.43	1.10	62.52	3.54	11.50	37.71	56.51	93.45
	99.02	93.43	0.00	78.04	11.48	78.04	66.83	6.93	8.11
-0.60	99.02	0.00	0.00	0.00	0.00	15.49	5.66	18.07	26.06
	99.02	93.43	0.00	0.00	0.00	66.81	8.35	38.08	16.14
	99.02	0.00	0.00	0.00	0.45	59.40	15.36	11.35	54.29
	99.02	93.43	0.00	0.00	0.03	4.45	65.51	28.003	29.97
	99.02	0.00	0.00	0.00	0.10	15.46	46.53	0.89	76.52

Table 3: Test probabilities obtained with the U_k statistics and using the true error rate for five repetitions of the experiment.

γ	0	1	2	3	4	5	6	7	8
-0.27	99.98	99.81	98.61	92.32	67.55	15.22	36.63	0.00	0.09
	99.98	99.80	98.57	92.13	66.95	41.47	89.04	0.28	4.10
	99.98	99.80	98.55	92.04	66.64	40.56	15.26	1.27	3.24
	99.98	99.80	98.57	92.13	66.95	14.62	93.14	0.04	2.29
	99.99	99.89	99.11	94.73	75.79	25.69	42.94	1.03	6.64
-0.30	99.88	98.98	94.07	74.50	26.75	13.09	48.52	6.55	49.11
	99.85	98.74	92.9	70.53	21.18	16.47	26.58	48.81	48.27
	99.84	98.62	92.32	68.65	18.87	11.15	55.51	40.86	64.32
	99.89	98.99	94.13	74.69	27.04	23.2	35.66	14.21	19.69
	99.91	99.15	94.91	77.45	95.30	39.42	42.82	67.72	59.64
-0.36	99.78	98.18	90.25	62.24	38.82	63.54	41.24	24.66	9.41
	99.74	97.94	89.20	20.03	55.25	48.44	18.21	79.89	17.87
	99.78	98.19	90.32	62.47	12.65	15.90	45.18	66.90	30.51
	99.72	97.77	88.44	57.07	90.7	0.05	14.36	97.16	4.81
	99.77	98.11	89.93	61.32	0.18	1.97	1.65	37.45	9.44
-0.45	99.35	95.40	78.75	0.18	0.00	12.60	44.11	80.39	36.94
	99.32	95.23	0.10	4.52	24.66	56.84	92.51	51.73	49.41
	99.44	95.98	0.00	1.25	63.2	98.96	71.53	43.82	79.28
	99.39	95.67	79.77	87.48	25.35	1.43	68.02	45.88	41.55
	99.56	96.69	83.85	45.47	1.20	12.10	13.63	10.69	64.44
-0.52	99.21	0.00	75.51	0.00	0.00	2.90	25.48	65.66	18.74
	99.25	94.77	0.00	0.00	7.25	65.74	50.39	98.71	50.17
	99.37	95.50	0.00	0.00	50.67	77.48	62.88	35.49	78.68
	99.3	95.06	0.16	34.79	0.25	0.91	90.96	45.68	30.35
	99.52	96.48	0.00	65.40	0.05	7.91	21.37	5.05	66.52
-0.60	99.03	0.00	0.00	0.00	0.00	14.80	5.91	17.92	25.15
	98.88	92.61	0.00	0.00	0.00	98.96	3.98	43.84	31.27
	99.20	0.00	0.00	0.00	0.06	97.48	33.65	9.36	99.76
	99.13	94.07	0.00	0.00	0.01	1.70	86.29	25.12	17.21
	99.29	0.00	0.00	0.00	0.00	1.54	98.49	0.67	51.54

Table 4: Test probabilities obtained with the U_k statistics with q estimated from the residuals for five repetitions of the experiment.

3.4.3 Geometrical characteristics of connected components

By considering Figure 4 it is clear that the summary statistic G is useful for detecting over and undersmoothing. For the undersmoothed estimates of the rock sample image obtained with $\beta = 0.27$, $G(R)$ lies under the envelopes for high values of the boundary to area ratio. This indicates that the fraction of components with a high boundary to area ratio is higher in the residuals than in the noise process. For the oversmoothed estimates, $G(R)$ lies above the envelopes for small values of the boundary to area ratio where the envelopes almost coincide.

3.4.4 Case II: Gaussian noise

Several options are available for analysis of the residuals when the noise process is a Gaussian white noise. One should however first note that it is not sufficient to consider univariate characteristics of the noise process. If one considers e.g. the sum of squares of the residuals then this statistic is $\sigma^2\chi^2(|L|)$ distributed if a perfect estimate has been obtained. In practice the residuals are correlated, and this means that the observed values of the sum of squares becomes extreme for the $\sigma^2\chi^2(|L|)$ distribution even for the best estimates. In the case where the image is represented by a parameter in $\mathbf{R}^{|L|}$, and a Gaussian penalizing term is used, it is possible to replace $|L|$ by the more appropiate socalled equivalent degrees of freedom (see e.g. Hall & Titterington, 1986). This option is not available in our case. It therefore seems necessary to include spatial information when analyzing the residuals.

One possibility is to fit a CAR (see e.g. Besag, 1974) to the residuals and then test for spatial independence. Also tests based on Morans index of independence (see Guyon, 1995) might be tried. We will however here restrict ourselves to consider what may be achieved by applying the methods proposed in Section 3 to the residual indicator fields $\tilde{R} = (1(R_i > 0))_{i\in L}$ or $\tilde{R}' = (1(|R_i| > c))_{i\in L}$, $c > 0$. A potential advantage of considering $\tilde{R}$ is that $P(\tilde{R}_i = 1)$ is actually known and equal to 0.5 when a perfect estimate is obtained, so that knowledge of σ^2 is not required.

In the experiment a Gaussian white noise with variance 0.5625 was added to the rock sample image in Figure 1. Then the expected rate of misclassified pixels for the maximum likelihood estimate is 25%. Penalized likelihood estimates were obtained with values of β ranging from 0.10 to 1.10, and best estimates were obtained with β around 0.50. The rates of misclassified pixels varied from 11% for the undersmoothed estimates and 7% for the oversmoothed estimates to 4% for the best estimates.

When the Ising model was fitted to the "binarized" residuals $\tilde{R}$ and the test for independence applied, it was possible to reject the most extreme estimates obtained with $\beta = 0.10$ and $\beta = 1.10$. The same approach was applied to R'

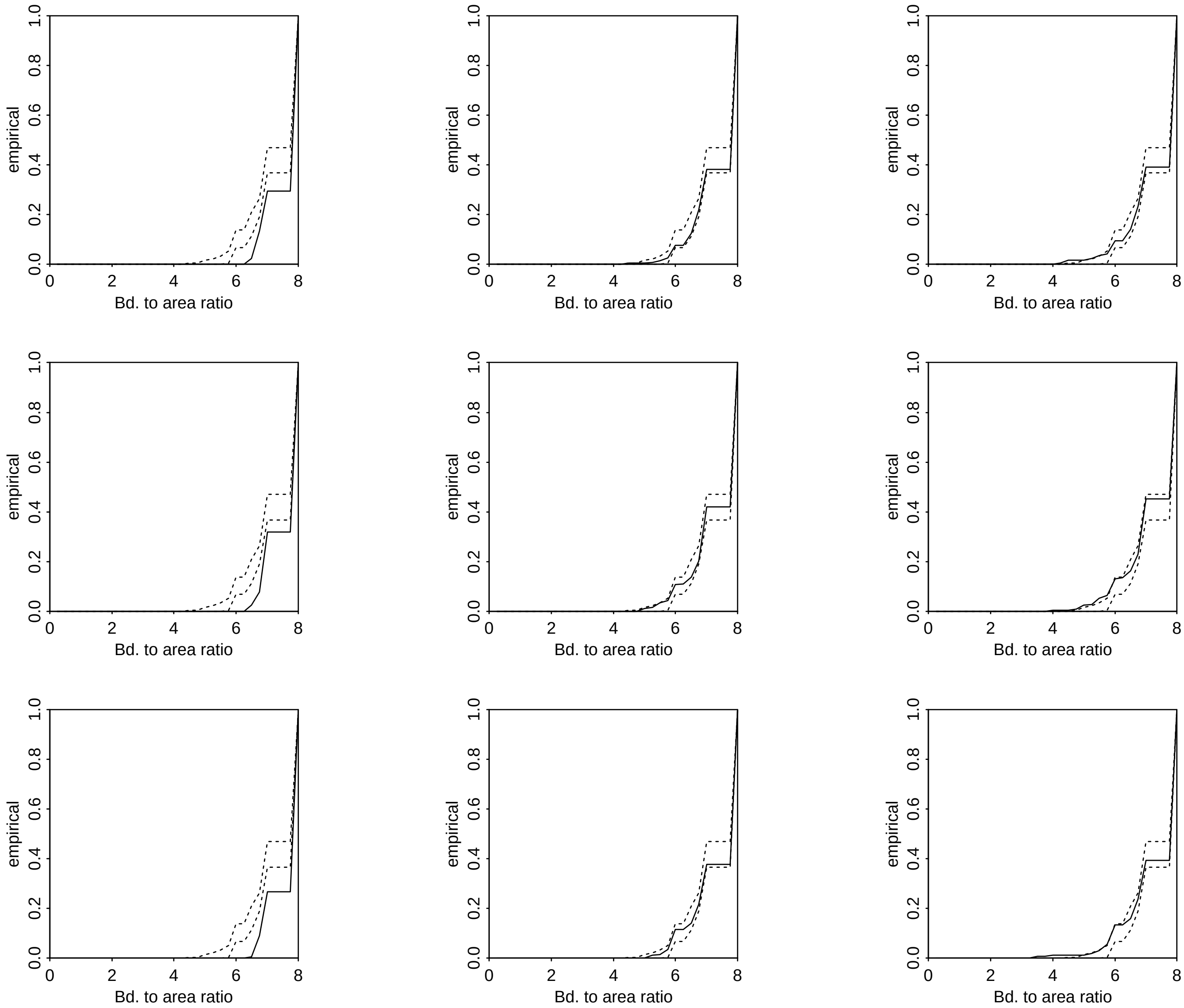

Figure 4: Empirical distribution functions of the boundary to area ratios of the connected components in the residual images for values of β equal to 0.27 (first column), 0.45 (middle column), and 0.60 (last column), and for the last three repetitions of the experiment. The envelopes are calculated from 39 simulations of the noise process.

with $c = 0.72$ so that the expected rate of 1's in R' is 15% for perfect estimates. When testing the hypothesis $\gamma = 0$ and $\alpha = \ln(0.15/0.85)$ it was often possible to identify the best estimates by rejecting the over and undersmoothed estimates. When testing the hypothesis $\gamma = 0$ and $\alpha \in \mathbf{R}$ it was only possible to reject the oversmoothed estimates obtained with $\beta = 1.10$. Similar results were obtained when the tests described in section 3.2 was applied to R'.

If one applies the methods described in section 2.1 for providing lower bounds for β, the lower bound 0.32 is obtained (with this value of β the inequality (3) holds with probability $\approx 1\%$). One can therefore avoid the application of very low values of β.

3.4.5 Discussion

For segmentations of the rock sample image corrupted by a binary noise it was not possible to get error rates much lower than 5%. Figure 3 among other things shows the misclassified pixels due to the estimate obtained with $\beta = 0.45$ in the last repetition of the experiment. This image of course also shows the difference between the residual image and the noise image. Not surprisingly one sees that the errors are made along the boundaries of the true image, and the error mainly originates from noise components (or parts of noise components) being incorporated in the components in the true image and to a smaller extent from fine details in the true image being smoothed away. The error of the estimates implies that the frequency of 1's in the residuals from the suitably smoothed estimates is always $2 - 4\%$ lower than the true error rate. Recall that Figure 4 shows the empirical distribution of the boundary to area ratios (see subsection 3.3) of the connected components in various residual images together with envelopes calculated from 39 simulations of the noise process on L. From the middle column in Figure 4 and the two last plots in the lower row of Figure 3 it seems fair to say, that the clustering given by the noise components is not much different from the clustering exhibited by the greater part of the residual components obtained with $\beta = 0.45$. This suggests that realizations of the residuals resembles realizations of the noise process where a number of noise components have been removed. The frequency of 1's is thereby reduced, but the degree of clustering exhibited by the remaining components still corresponds to the original noise process. This is probably the reason why the test for independence sometimes rejects suitably smoothed estimates.

The tests based on the $U_k's$ are more robust towards the fact that only imperfect estimates can be obtained because the statistic U_k measure the *fraction* of 1-pixels with k neighbours of opposite colour, and is therefore less sensitive towards the fact that the actual frequency of 1's is not right. One may also say that U_k only summarizes information about local configurations in the image. A similar remark holds for the summary statistic G which also turned out to be useful for rejecting both over and undersmoothed estimates.

In practice q will not be known but the conclusions may still be valid if a precise estimate of q is available. In any case it seems necessary to use such a precise estimate of q in the likelihood in order to obtain satisfactory segmentations.

4 Concluding remarks

In this paper various approaches to detection of under or oversmoothing in penalized likelihood segmentation of noisy images have been studied. The method based on testing for spatial independence may be useful for detecting oversmoothing but is not useful for detecting undersmoothing, and even suitably smoothed estimates are sometimes rejected.

Alternatively, we have considered a collection of statistics which are very easy to compute and which are more robust towards the unavoidable perturbations of the residual process. Experiments under the assumption that the error rate is known indicate that these statistics may be useful in practice for detecting under or oversmoothing when a precise estimate of the error rate is available. If the error rate is estimated from the residual image it is still possible to detect oversmoothing but as for the test for independence it is not possible to distinguish undersmoothed estimates from suitably smoothed estimates. Finally it is demonstrated that it is also helpful to consider the empirical distribution of the boundary to area ratios of the connected components in the residual image.

We have focused on segmentation of binary images observed without blur. The methods can in principle also be applied in situations with blur, and are also extendable to cases with multicoloured images. In the case of correlated noise things of course get more complicated. It may e.g. be much more difficult to identify oversmoothing because correlation in the residual process due to oversmoothing may be indistinguishable from the correlation of the noise process.

References

Besag, J. E. (1974), 'Spatial interaction and the statistical analysis of lattice systems (with discussion)', *J. Roy. Statist. Soc. B* **56**, 192–236.

Besag, J. E. (1975), 'Statistical analysis of non-lattice data', *The Statistician* **24**, 179–195.

Besag, J. E. (1986), 'On the statistical analysis of dirty pictures', *J. Roy. Statist. Soc. B* **48**, 259–302.

Diggle, P. J. (1983), *Statistical analysis of spatial point patterns*, Academic Press.

Dinten, J. M., Guyon, X. & Yao, J. F. (1991), On the choice of the regularization parameter: The case of binary images in the Bayesian restauration framework, *in* ‘Spatial statistics and imaging’, Lect. Notes., Inst. Math. Stat., AMS-IMS-SIAM joint summer conf., Hayward, pp. 55–77.

Geman, S. & Geman, D. (1984), ‘Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images’, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**, 721–741.

Geyer, C. J. (1991), Estimating normalizing constants and reweighting mixtures in Markov chain Monte Carlo mixtures, Technical Report 568r, School of Statistics, University of Minnesota.

Geyer, C. J. & Thompson, E. A. (1992), ‘Constrained Monte Carlo maximum likelihood for dependent data (with discussion)’, *J. Roy. Statist. Soc. B* **54**, 657–699.

Greig, D. M., Porteous, B. T. & Seheult, A. H. (1989), ‘Exact maximum a posteriori estimation for binary images’, *J. Roy. Statist. Soc. B* **51**, 271–279.

Grenander, U. & Miller, M. (1994), ‘Representations of knowledge in complex systems’, *J. Roy. Statist. Soc. B* **56**, 549–603.

Grimmett, G. R. (1989), *Percolation*, Springer-Verlag, Berlin.

Guyon, X. (1995), *Random fields on a network*, Springer-Verlag.

Hall, P. & Titterington, D. M. (1986), ‘On some smoothing techniques used in image restoration’, *J. Roy. Statist. Soc. B* **48**, 330–343.

Hastings, W. (1970), ‘Monte Carlo sampling methods using Markov chains and their applications’, *Biometrika* **57**, 97–109.

Kesten, H. (1982), *Percolation theory for mathematicians*, Birkhäuser.

Kindermann, R. & Snell, J. L. (1980), *Markov random fields and their applications*, American Mathematical Society.

Kirkpatrick, S., Gelett, C. D. & Vecchi, M. P. (1983), ‘Optimization by simulated annealing’, *Science* **220**, 621–630.

Künsch, H. R. (1994), ‘Robust priors for smoothing and image restoration’, *Ann. Inst. Statist. Math.* **46**, 1–19.

Møller, J. & Waagepetersen, R. (1996), Markov connected component fields, Research Report 341, Dept. of Theoretical Statistics, University of Aarhus. To appear in Adv. Appl. Prob.

Ripley, B. D. (1988), *Statistical inference for spatial processes*, Cambridge University Press, Cambridge.

Thompson, A. M., Brown, J. C., Kay, J. W. & Titterington, D. M. (1991), 'A study of methods of choosing the smoothing parameter in image restoration by regularization', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **13**, 326–338.

van den Berg, J. & Ermakov, A. (1996), 'A new lower bound for the critical probability of site percolation on the square lattice', *Random structures and algorithms*. To appear.

Waagepetersen, R. (1995), Statistical image analysis and Markov chain Monte Carlo, Unpublished manuscript.

Wiermann, J. C. (1995), 'Substitution method critical probability bounds for the square lattice site percolation model', *Combinatorics, Probability and Computing* **4**, 181–188.

Younes, L. (1988), 'Estimation and annealing for Gibbsian fields', *Annales de l'Institut Henri Poincaré* **24**, 269–294.