Nonhomogeneity Analysis Using Borrowed Strength
Author(s): Carey E. Priebe
Source: *Journal of the American Statistical Association*, Vol. 91, No. 436 (Dec., 1996), pp. 1497–1503
Published by: American Statistical Association
Stable URL: http://www.jstor.org/stable/2291575
Accessed: 09/12/2010 13:51

Nonhomogeneity Analysis Using Borrowed Strength

Carey E. PRIEBE

This article develops a "borrowed strength" methodology for estimating local probability densities in a random field. Under a piecewise stationarity condition similarities between the densities in different regions of the field can be exploited by estimating a global mixture density and imposing the parameter-space support of this borrowed strength estimate on the local estimation problems. The local estimates are then used in an analysis of the homogeneity of the field, which is shown to benefit from the method of borrowing strength.

KEY WORDS: Change detection; Maximum likelihood; Mixture model; Multiple comparison; Profile likelihood; Random field.

1. INTRODUCTION

Given a random field $\xi(\mathbf{x})$, it is often reasonable to consider the domain of definition of this field to be a union of disjoint subregions of homogeneity. Under such an assumption, one can test for homogeneity of the field by estimating the regional characteristics and comparing them in a traditional multiple comparisons framework. For instance, in image processing one may consider $\xi(\mathbf{x})$: $R^0 \to \Re$, where R^0 is an $M_1 \times M_2$ lattice of pixel locations and the value of a field observation represents pixel intensity. Assuming for simplicity that the image is made up of two disjoint regions $R^0 = R^1 \cup R^2$ with associated probability density functions $\xi(\mathbf{x}) \sim \alpha^i(\xi)$ for $\mathbf{x} \in R^i$, the question of interest can be posed as

$$H_0 \text{: Homogeneity } (\alpha^1 = \alpha^2)$$

versus $\qquad\qquad$ (1)

$$H_1 \text{: Nonhomogeneity } (\alpha^1 \neq \alpha^2).$$

That is, the image is made up entirely of pixels drawn from a single process (H_0) or composed of two disparate regions (H_1). (The application then entails the inference that nonhomogeneity warrants further investigation; for instance, that regions with statistical characteristics differing from the norm should be the subject of additional attention or more sophisticated, potentially computationally intensive processing.) In this case, given probability density estimates $\hat{\alpha}^1$ and $\hat{\alpha}^2$, large values of a statistic $T = d(\hat{\alpha}^1, \hat{\alpha}^2)$ for some distance d defined on the space of probability densities under consideration will indicate nonhomogeneity. This same scenario can be adapted to almost any image or signal processing application whenever the fundamental question being addressed is one of field homogeneity.

This article seeks to address two issues pertaining to tests of the type described in (1). Although these two issues may at first glance appear unrelated, they in fact come together in the practice of nonhomogeneity analysis, as we demonstrate.

First, can the estimates used in the statistic T be improved on under the assumption that there are fundamental similarities in the regional characteristics, even under the alternative hypothesis? We show the answer to this question to be "yes" when the α^i have certain characteristics in common. In particular, when the α^i are finite mixture models whose underlying mixture components are invariant across the entire field domain R^0 in terms of their location in parameter space, the local mixture probability density functions differ only in their mixing coefficients. A "borrowed strength" estimate that exploits this invariance by using all of the observed data to develop an estimate of the invariant parameters and imposing this estimate as a constraint on the estimation of the local α^i can produce superior local estimates and hence a superior test for homogeneity.

The second aspect of the nonhomogeneity analysis addressed herein concerns a more realistic setup for a test of homogeneity for random fields. Because there is seldom prior knowledge of the location of the local regions of interest R^i, it is necessary to introduce a regional structure on R^0 and test for homogeneity under this structure. Let $R^0 = \cup \tilde{R}^i$ $(i = 1, \ldots, \tilde{r})$, where the $\tilde{R}^i$ are (possibly overlapping) neighborhoods. For instance, in the investigation of scan statistics (Cressie 1993, pp. 312–313), one often considers ε balls, in which case $\tilde{R}^\tau = B(\tau, \varepsilon) = \{\mathbf{x} \in R^0 \colon \|\mathbf{x} - \tau\| < \varepsilon\}$ for $\tau \in R^0$ and $\varepsilon > 0$. The statistic to be used is then identified as $T = \max_{i, j \in \{1, \ldots, \tilde{r}\}} d(\hat{\alpha}^i, \hat{\alpha}^j)$, and the test is defined as

$$H_0 \text{: Homogeneity } (\alpha^i = \alpha^j \quad \forall\, i, j)$$

versus $\qquad\qquad$ (2)

$$H_1 \text{: Nonhomogeneity } (\exists\, i, j \text{ such that } \alpha^i \neq \alpha^j).$$

On observing $\xi_1^i, \ldots, \xi_{n^i}^i$ from each region $\tilde{R}^i$, the goal is to determine whether there are anomalous regions in the field or whether the $\tilde{r}$ populations are identical (the field is homogeneous).

These two issues are considered together, as they can be seen to give rise to a trade-off. The lack of knowledge of the location of potential subregions of nonhomogeneity necessitates that the second version of nonhomogeneity analysis be adopted and that the regions $\tilde{R}^i$ be chosen to be rel-

atively small compared to the anticipated size of the true but unknown R^i. This in turn implies that the number of observations per region is small and hence the estimates $\hat{\alpha}^i$ $(i = 1, \ldots, \tilde{r})$ will have relatively large variance. Therefore, any improvement in these estimates that can be gained by borrowing strength will be potentially significant. Furthermore, because the $\tilde{R}^i$ are necessarily smaller in their spatial extent than the nonhomogeneities anticipated, there likely will be numerous regions $\tilde{R}^i$ completely contained within a given R^i and hence having the same statistical characteristics. This implies that there will be information relevant to the estimation of $\hat{\alpha}^i$ in at least some of the regions $\tilde{R}^j$ for $j \neq i$, and the assumption that borrowing strength can improve the local estimations is reasonable.

2. FORMULATION

2.1 Problem Statement

Let $R^0 \subset \Re^n$ be the domain of definition of a random field $\xi(\mathbf{x})$: $R^0 \to \Re^m$ and consider $R^0 = \cup R^i$ $(i = 1, \ldots, r)$, where the r disjoint, connected regions R^i have each associated with them random variables $\boldsymbol{\xi}(\mathbf{x}), \mathbf{x} \in R^i$. Assume that the observations from region R^i are identically distributed with common marginal density $\alpha^i(\boldsymbol{\xi})$; that is, each R^i is a subregion of homogeneity and $\boldsymbol{\xi}(\mathbf{x}) \sim \alpha^i(\boldsymbol{\xi})$ for $\mathbf{x} \in R^i$. R^0 can be termed *piecewise stationary* in the sense of being r stationary random fields $\boldsymbol{\xi}^i$ embedded via $\boldsymbol{\xi}(\mathbf{x}) = \sum_{i=1}^r \boldsymbol{\xi}^i I_{\{m(\mathbf{x})=i\}}$, where $I_{\{m(\mathbf{x})=i\}}$ is the indicator function for an embedding field $m(\mathbf{x})$ taking on the values $1, \ldots, r$. This approach has been used by, for instance, Carlstein and Lele (1994). In the simplest case, the goal is to perform the test of the hypothesis of homogeneity in (2). If this test is to be performed without choosing a specific type of nonhomogeneity in the alternative, then it is necessary to develop estimates $\hat{\alpha}^i$ for each i and reject the null hypothesis of homogeneity for large values of $T = \max_{i,j \in (1,\ldots,\tilde{r})} d(\hat{\alpha}^i, \hat{\alpha}^j)$.

2.2 Borrowed Strength Methodology

The fundamental assumption made regarding the densities $\alpha^i(\boldsymbol{\xi})$ is that they are finite mixture models. We assume that $\alpha^i(\boldsymbol{\xi}) = \alpha(\boldsymbol{\xi}; \boldsymbol{\theta}^i) = \alpha(\boldsymbol{\xi}; \boldsymbol{\psi}^i, \boldsymbol{\lambda}^i)$ is a mixture of m^i component densities $C(\boldsymbol{\xi}; \boldsymbol{\gamma})$. That is,

$$\alpha^i(\boldsymbol{\xi}) = \sum_{t=1}^{m^i} \pi_t^i C(\boldsymbol{\xi}; \boldsymbol{\gamma}_t^i)$$

where

$$\boldsymbol{\theta}^i = (\boldsymbol{\psi}^i, \boldsymbol{\lambda}^i) = ((\boldsymbol{\gamma}_1^i, \ldots, \boldsymbol{\gamma}_{m^i}^i), (\pi_1^j, \ldots, \pi_{m^i}^j)).$$

Furthermore, we assume that $m^i = m$ and $(\boldsymbol{\gamma}_1^i, \ldots, \boldsymbol{\gamma}_{m^i}^i) = \boldsymbol{\psi}^i = \boldsymbol{\psi}^0 = (\boldsymbol{\gamma}_1^0, \ldots, \boldsymbol{\gamma}_m^0)$ for all i. Thus

$$\alpha^i(\boldsymbol{\xi}) = \alpha(\boldsymbol{\xi}; \boldsymbol{\theta}^i) = \alpha(\boldsymbol{\xi}; \boldsymbol{\psi}^0, \boldsymbol{\lambda}^i) = \sum_{t=1}^m \pi_t^i C(\boldsymbol{\xi}; \boldsymbol{\gamma}_t^0); \quad (3)$$

that is, for univariate normal mixtures, we have $\boldsymbol{\gamma} = (\mu, \nu)$ and

$$\alpha^i(\xi) = \alpha(\xi; (\mu_1^0, \nu_1^0, \ldots, \mu_m^0, \nu_m^0), (\pi_1^i, \ldots, \pi_m^i))$$

$$= \sum_{t=1}^m \pi_t^i \varphi(\xi; \mu_t^0, \nu_t^0). \quad (4)$$

$\boldsymbol{\psi}^0 = (\mu_1^0, \nu_1^0, \ldots, \mu_m^0, \nu_m^0)$ is common to all of the densities α^i, and the difference between the densities is encompassed entirely in the mixing coefficients $\boldsymbol{\lambda}^i = (\pi_1^i, \ldots, \pi_m^i)$. The maximum likelihood performance exhibited in Section 3 for our proposed methodology is valid whenever the component family is an exponential family, the $\boldsymbol{\gamma}_t^0$ are distinct, and $\pi_t^i > 0$ for all i, t.

Let the collection of all observations in the field be $\Xi_{n^0}^0 = \cup \Xi_{n^i}^i$, where $\Xi_{n^i}^i = \{\boldsymbol{\xi}_j^i\}$ $j = 1, \ldots, n^i$ represents the observations from region $\tilde{R}^i$.

Borrowed Strength Methodology

1. Introduce a regional structure on R^0; $R^0 = \cup \tilde{R}^i$ $(i = 1, \ldots, \tilde{r})$.

2. Obtain the borrowed strength estimate $\tilde{\boldsymbol{\psi}}^0$ of $\boldsymbol{\psi}^0$ for the entire random field R^0 using all the observations $\Xi_{n^0}^0$.

3. Obtain the estimate $\tilde{\boldsymbol{\lambda}}^i$ of $\boldsymbol{\lambda}^i$ for each region $\tilde{R}^i$, using the local observations $\Xi_{n^i}^i$ and the borrowed strength estimate $\tilde{\boldsymbol{\psi}}^0$ of $\boldsymbol{\psi}^0$.

4. Large values of

$$T = \max_{i,j \in (1,\ldots,\tilde{r})} d(\tilde{\alpha}^i(\tilde{\boldsymbol{\psi}}^0, \tilde{\boldsymbol{\lambda}}^i), \tilde{\alpha}^j(\tilde{\boldsymbol{\psi}}^0, \tilde{\boldsymbol{\lambda}}^j))$$

indicate nonhomogeneity.

We make the claim that using the statistic T thus obtained in the test for nonhomogeneity (2) yields an improvement over the analogous conventionally estimated statistic, particularly when the estimates are obtained via maximum likelihood.

For the examples presented in Section 3, the integrated squared error (ISE) $\text{ISE}(\tilde{\alpha}^i, \tilde{\alpha}^j) = \int_{-\infty}^{\infty} (\tilde{\alpha}^i(\boldsymbol{\xi}) - \tilde{\alpha}^j(\boldsymbol{\xi}))^2 d\boldsymbol{\xi}$ is used as the measure of distance, although there are numerous other possible choices.

2.3 Investigation Under Independence

Consider the case of independent observations. When each region R^i has the same density, given by $\alpha(\boldsymbol{\xi}; \boldsymbol{\theta})$, the joint likelihood is given by

$$L^J(\boldsymbol{\theta}) = \prod_{i=1}^r \prod_{j=1}^{n^i} \alpha(\boldsymbol{\xi}_j^i; \boldsymbol{\theta}) = \prod_{j=1}^{n^0} \alpha(\boldsymbol{\xi}_j; \boldsymbol{\theta}), \quad (5)$$

and when the possibly unique parameter vector for each region R^i is given by $\boldsymbol{\theta}^i$, this joint likelihood is given by

$$L^J(\boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^r) = \prod_{i=1}^r \prod_{j=1}^{n^i} \alpha(\boldsymbol{\xi}_j^i; \boldsymbol{\theta}^i). \quad (6)$$

In this case an estimate for $\boldsymbol{\theta}^i$ is obtained by maximizing the regional likelihood

$$L^i(\boldsymbol{\theta}^i) = \prod_{j=1}^{n^i} \alpha(\boldsymbol{\xi}_j^i; \boldsymbol{\theta}^i). \quad (7)$$

When it can be assumed that there are common elements of the parameter vectors θ^i, so that $\theta^i = (\psi^0, \lambda^i)$, the joint likelihood (6) becomes

$$L^J(\psi^0, \lambda^1, \ldots, \lambda^r) = \prod_{i=1}^{r} \prod_{j=1}^{n^i} \alpha(\xi_j^i; \psi^0, \lambda^i). \quad (8)$$

Under the likelihood (8), it is inefficient to estimate θ^i with $\bar{\theta}^i = (\bar{\psi}^i, \bar{\lambda}^i)$ obtained by maximizing the regional likelihood

$$L^i(\psi^0, \lambda^i) = \prod_{j=1}^{n^i} \alpha(\xi_j^i; \psi^0, \lambda^i), \quad (9)$$

as this ignores information regarding ψ_0 available in the other regional samples Ξ^j $(j \neq i)$. It does make sense, and in fact it is common practice, to estimate θ^i with $\hat{\theta}^i = (\hat{\psi}^0, \hat{\lambda}^i)$ obtained by maximizing the joint likelihood (8).

However, superior estimates can be obtained, under the condition of nonorthogonality between ψ^0 and λ^i (Cox and Reid 1987) and for specific values of n^0, n^i with respect to the similarity between λ^i and λ^j, by combining the joint likelihood estimate $\tilde{\psi}^0$ obtained by maximizing

$$L^0(\psi^0, \lambda^0) = \prod_{j=1}^{n^0} \alpha(\xi_j; \psi^0, \lambda^0), \quad (10)$$

with the regional profile likelihood estimate $\tilde{\lambda}^i$ obtained by maximizing

$$L^i(\lambda^i | \tilde{\psi}^0) = \prod_{j=1}^{n^i} \alpha(\xi_j^i; \lambda^i | \tilde{\psi}^0). \quad (11)$$

Estimating $\theta^i = (\psi^0, \lambda^i)$ via $\theta^i = (\tilde{\psi}^0, \tilde{\lambda}^i)$ obtained using (10) and (11) is termed *borrowed strength maximum likelihood*.

Precision in detecting small regions of nonhomogeneity requires that the subfields on which densities are estimated be small, thus there are insufficient data to develop acceptable estimates without borrowing strength. Often, however, there will be sufficient data to obtain suitably accurate constrained local estimates $\tilde{\lambda}^i$ given $\tilde{\psi}^0$. Thus the borrowed strength methodology of constraining the problem of estimating the local densities to that of estimating the mixing coefficients for the subfields can yield superior results.

2.4 An Issue in the Estimation of α^0

It is necessary to note that in practice, care must be taken in estimating the overall field density α^0 (Step 2 in the methodology described earlier). Samples from $\tilde{r}$ different distributions can have an overall distribution indistinguishable from normality, even if each individual distribution is a decidedly nonnormal mixture of normals. Thus it is necessary to perform this initial estimation in such a way as to allow for the "overdetermining" of the overall density. As an example of such a method, a block-recursive adaptive mixtures procedure (Priebe 1994) is utilized in the examples herein. We test local blocks of data, in turn, for the appropriateness of the current model. If we determine that

the current mixture model is insufficient in complexity, we add additional terms to the mixture. This technique ensures that even if the overall density is nearly normal, the estimate of this density will more appropriately reflect any potential underlying mixture structure.

It must be noted that the adaptive mixtures algorithm is recursive and order dependent. It is possible for different orderings to produce significantly different estimates. However, in practice the underlying mixture structure for multiple randomized orderings is quite similar, and the effect of these differences on the performance of the borrowed strength procedure is negligible. For the simulation examples presented in Section 3, we process the observations by the adaptive mixtures algorithm in randomized order without replacement, and we have empirically verified that different random orderings produce nearly identical results.

3. EXAMPLES

3.1 Nile Example

For illustrative purposes, let us consider an example in which the goal is the detection of a single changepoint. The Nile River has been carefully monitored for many years; 100 yearly observations of the annual volume of discharge from 1871 to 1970 were reported by Cobb (1978). There have been many parametric approaches, and various nonparametric approaches as well. Carlstein (1988) summarized these and presented a solution of his own. In general, it is well agreed that there is a changepoint in 1898, with independent meteorological evidence supporting this conclusion.

If only a single changepoint is being sought, general regional analysis is not necessary, but borrowing strength may still be valuable. Let us assume a single changepoint and $R^0 = R^1 \cup R^2$. Here $\{\xi^i(x)\}$ represents the observations available for α^i, and borrowing strength from the other region may be of value if the statistical structure of the two regions is sufficiently similar.

We use the adaptive mixture modeling procedure to obtain $\tilde{\psi}^0$. We develop a normal mixture estimate for $\{\xi^i(x)\}$, the overall data set of 100 observations, in which the number of terms in the mixture model is data driven and the estimate is a local maximum of the likelihood surface determined by the parameter vector ultimately selected. In this case we use seven terms. Why seven terms? Adaptive mixtures produces a finite mixture that is sufficiently complex to model the data well and at the same time impose more structure than a kernel estimator or empirical distribution function. Robustness to the particular model produced by adaptive mixtures has also been considered.

Once $\tilde{\psi}^0$ is obtained, the changepoint is estimated as the x in [1871, 1970] that yields the largest ISE between R^1 and R^2; $\tau = \text{argmax}_x \|\tilde{\alpha}^{x+}, \tilde{\alpha}^{x-}\|_{L^2}$, where R^{x-} is [1871, x] and R^{x+} is [$x + 1$, 1970]. That is, $\tilde{\lambda}^i$ is estimated for each R^{x-} and R^{x+}. The result is $\tau = 1898$ (see Fig. 1). The advantage of borrowing strength can be seen if one assumes for the sake of argument that the seven-term model correctly represents the complexity for this data. In this
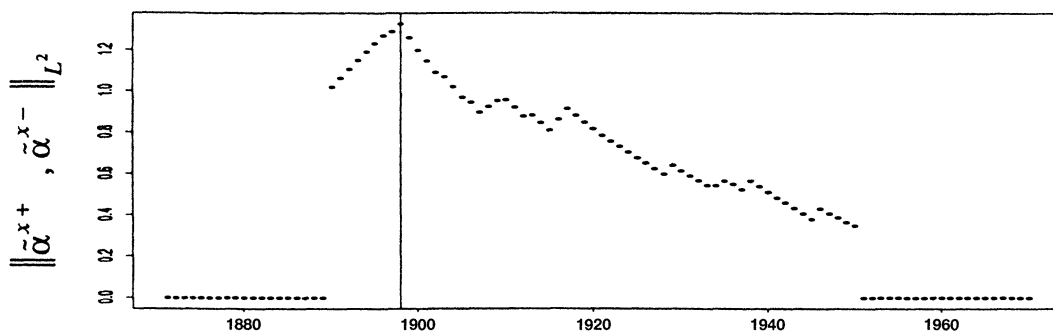
Figure 1. Nile Changepoint Example (cf. Carlstein 1988, Fig. 1).

case borrowed strength requires estimation of the means and variances for a seven-term mixture model, 14 parameters, based on 100 observations. The remaining 6 parameters (the mixing coefficients) are then estimated on the 28 observations [1871, 1898]. Utilizing the same model but without imposing $\tilde{\psi}^0$ on the local estimation problems requires estimating all 20 parameters using only 28 observations, an estimate that will obviously be inferior.

Consider now a situation in which there is no preconceived notion of a single changepoint. To obtain borrowed strength units, a regional structure must first be imposed on R^0, yielding $\tilde{R}^i$. Nine overlapping regions of 20 observations each were chosen: $\tilde{R}^1 = [1871, 1890], \tilde{R}^2 = [1881, 1900], \ldots, \tilde{R}^9 = [1951, 1970]$. $\tilde{\psi}^0$ is estimated as before, and the mixing coefficients $\tilde{\lambda}^i$ are estimated for each region $\tilde{R}^i$ of 20 observations. Table 1 shows these values for each of the seven terms and each of the nine regions. To illustrate, consider the values $\tilde{\lambda}^i$ for terms 1 and 3. The mixing coefficients for these terms indicate the ability to detect nonhomogeneity. For regions $\tilde{R}^1$ and $\tilde{R}^2$, term 1 has a small coefficient (.08 and .05). For regions $\tilde{R}^i$ ($i = 4, \ldots, 9$), term 1 has a coefficient greater than or equal to .57. For region $\tilde{R}^3$, however, term 1 has an intermediate coefficient of .33. Term 3 exhibits the opposite behavior. Again, the coefficient of term 3 for region $\tilde{R}^3$ is dissimilar to that for all other regions.

To investigate nonhomogeneity, we obtain the values of $S_j^i = \mathrm{ISE}(\tilde{\alpha}^i, \tilde{\alpha}^j)$ and investigate the set $\{S_{(2)}^i\}$, the second smallest of these for each given i. The most likely candidate for nonhomogeneity is $\tilde{R}^3 = [1891, 1910]$, as $3 = \mathrm{argmax}_i S_{(2)}^i$. This is consistent with the changepoint occurring in 1898. Table 2 gives the values of S_j^i for each

pair i, j, with $S_{(2)}^i$ in bold. This table shows that regions $\tilde{R}^1$ and $\tilde{R}^2$ are similar to one another and different from all of the other regions. Region $\tilde{R}^3$ differs from all of the other regions, as $S_{(2)}^3 = .33$. Regions $\tilde{R}^i$ ($i = 4, \ldots, 9$) are comparatively similar to one another, with values of $S_{(2)}^i$ less than or equal to .19. Note, however, that if one had no preconceived notion of a single changepoint, then an investigation of region $\tilde{R}^5$, the years [1911, 1930], would be in order, because the value $S_{(2)}^5 = .19$ is far and away the second largest value among the $\{S_{(2)}^i\}$.

3.2 Simulation Example

We now consider a simulation example, depicted in Figure 2, wherein borrowed strength is applied to detect nonhomogeneities in an embedded dependent random field. Consider two stationary, ergodic random fields f^1 and f^2, each with a known dependency structure and known marginals. For this example, we consider a simple dependency. We generate two independent and identically distributed fields ε^i with $\varepsilon^i(\mathbf{x}) \sim \pi^i \varphi_1 + (1 - \pi^i)\varphi_2$ for $i = 1, 2$, with $\pi^1 = .4, \pi^2 = .2, \varphi_1 = N(-2, 1)$, and $\varphi_2 = N(2, \frac{9}{16})$. From these ε^i, we generate $f^i(\mathbf{x}) = |N_\mathbf{x}|^{-1} \sum_{\mathbf{y} \in N_\mathbf{x}} \varepsilon^i(\mathbf{y})$, where $N_\mathbf{x} = \{\mathbf{y} : \|\mathbf{x} - \mathbf{y}\| \leq K\}$ and a value of $K = \sqrt{2}$ is used, yielding $|N_\mathbf{x}| = 5$. Then the marginal density for $f^i(\mathbf{x})$ is known, is identical for all $\mathbf{x}$, and is a mixture of normals. Furthermore, because φ_1 and φ_2 are the same for both ε^1 and ε^2, the individual terms in the mixtures corresponding to the marginals for f^1 and f^2 are the same; that is, there is a matching of (mean, variance) pairs between f^1 and f^2. The marginal densities of these two fields differ only in terms of their mixing coefficients. Figure 2, a and b show realizations of fields f^1 and f^2.

Table 1. $\tilde{\lambda}^i$ for the Nile Example

	Term 1	Term 2	Term 3	Term 4	Term 5	Term 6	Term 7
Region R^1	.05	.26	.58	0	0	.07	.04
Region R^2	.08	.31	.55	0	0	.05	.01
Region R^3	.33	.19	.37	0	.03	.06	.02
Region R^4	.57	.28	.05	.05	.05	0	0
Region R^5	.68	.05	.10	.05	.11	0	0
Region R^6	.78	.16	0	0	.06	0	0
Region R^7	.77	.19	0	0	.04	0	0
Region R^8	.70	.25	0	0	.04	0	0
Region R^9	.57	.37	.03	0	.03	0	0

Table 2. Similarity Matrix (With $S^j_{(2)}$ in Bold) for the Nile Example

	R^1	R^2	R^3	R^4	R^5	R^6	R^7	R^8	R^9
R^1	0	.07	.36	.75	.83	.94	.93	.87	.77
R^2	.07	0	.33	.71	.80	.91	.89	.84	.72
R^3	.36	.33	0	.42	.47	.59	.58	.53	.46
R^4	.75	.71	.42	0	.27	.25	.23	.15	.10
R^5	.83	.80	.47	.27	0	.19	.21	.24	.36
R^6	.94	.91	.59	.25	.19	0	.03	.12	.30
R^7	.93	.89	.58	.23	.21	.03	0	.09	.27
R^8	.87	.84	.53	.15	.24	.12	.09	0	.18
R^9	.77	.72	.46	.10	.36	.30	.27	.18	0

We use a binary (0, 1) Markov random field to model the presence of local nonhomogeneities. Figure 2c shows a realization of such a field, generated using a Gibbs sampler (Geman and Geman 1984) with an initial iid Bernoulli ($p = .46$) field and a 24-pixel square neighborhood. This binary field $m(\mathbf{x})$ is used to embed f^1 and f^2 into field f, shown in Figure 2d, via $f = I_{\{m(\mathbf{x})=1\}}f^1 + I_{\{m(\mathbf{x})=0\}}f^2$. Thus the random field f depicted in Figure 2d is the union of $r = 5$ disjoint regions R^i, and $f(\mathbf{x})$ is identically distributed as $f^2(\mathbf{x})$ for $\mathbf{x}$ in the "background" region associated with $I_{\{m(\mathbf{x})=0\}}$, and $f(\mathbf{x})$ is identically distributed as $f^1(\mathbf{x})$ for $\mathbf{x}$ in the four "anomaly" regions associated with $I_{\{m(\mathbf{x})=1\}}$. Figure 2e shows the marginal densities for f^1, f^2, and f. Note that the marginal for f is nearly identical to that of f^2, due to the sparseness of $m(\mathbf{x})$.

Field f meets the criteria for the application of the borrowed strength scan analysis, and Figure 2, f and g compare the performance of borrowed strength versus local likelihood analysis. These figures depict the results of using the borrowed strength scan methodology described earlier. An 11×11-pixel moving window is scanned throughout the region. At each location the density is estimated, using borrowed strength maximum likelihood on the 121 observations with a $\hat{\psi}^0$ parameter constraint in the one case and standard maximum likelihood on the 121 observations in the other case. Each locality statistic—the estimated
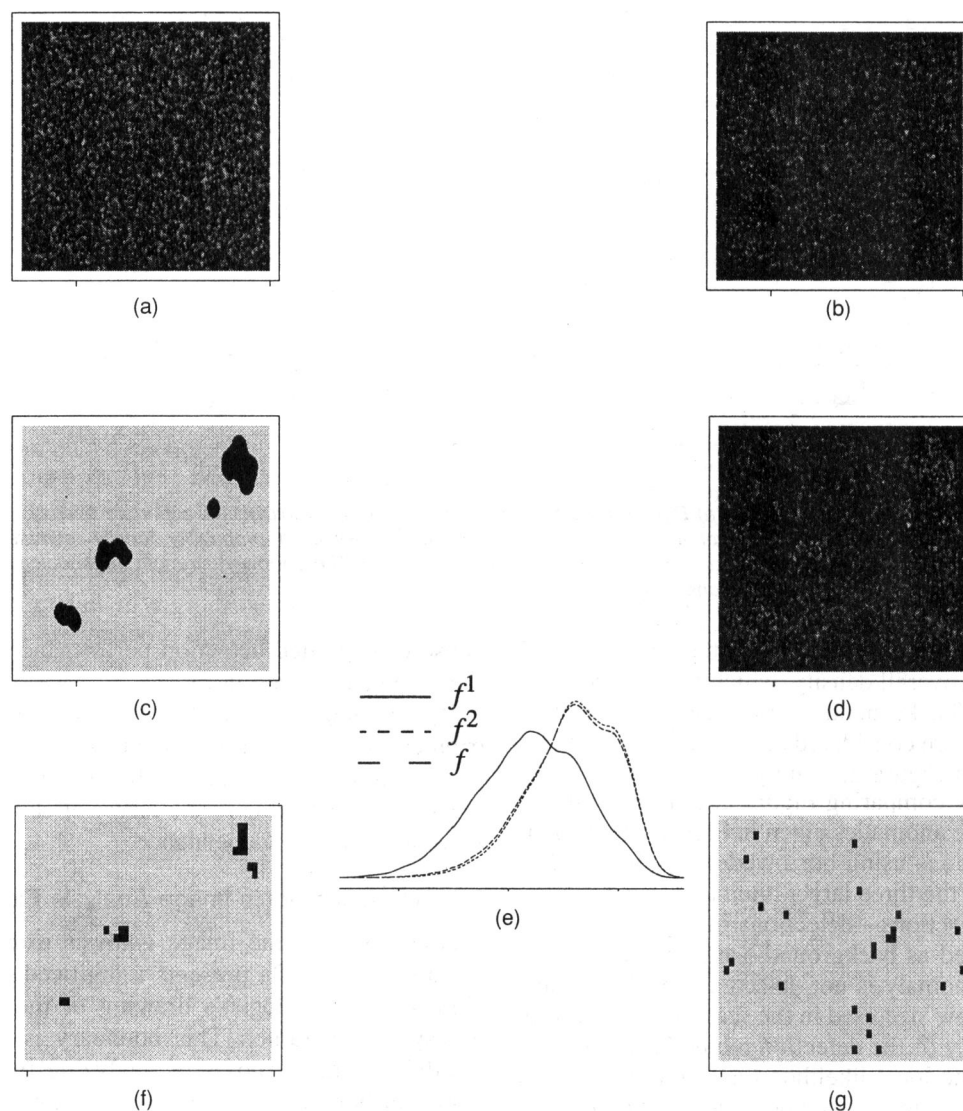

Figure 2. Embedded Random Fields Simulation Example. (a) Random field f^1; (b) random field f^2; (c) map of nonhomogeneities, a binary Markov random field m used to embed f^1 and f^2 into f; (d) random field f; (e) probability densities for f^1, f^2, and f; (f) nonhomogeneity detections using borrowed strength; (g) nonhomogeneity detections using conventional local likelihood. The borrowed strength methodology produces a superior detection rate with significantly fewer false detections.
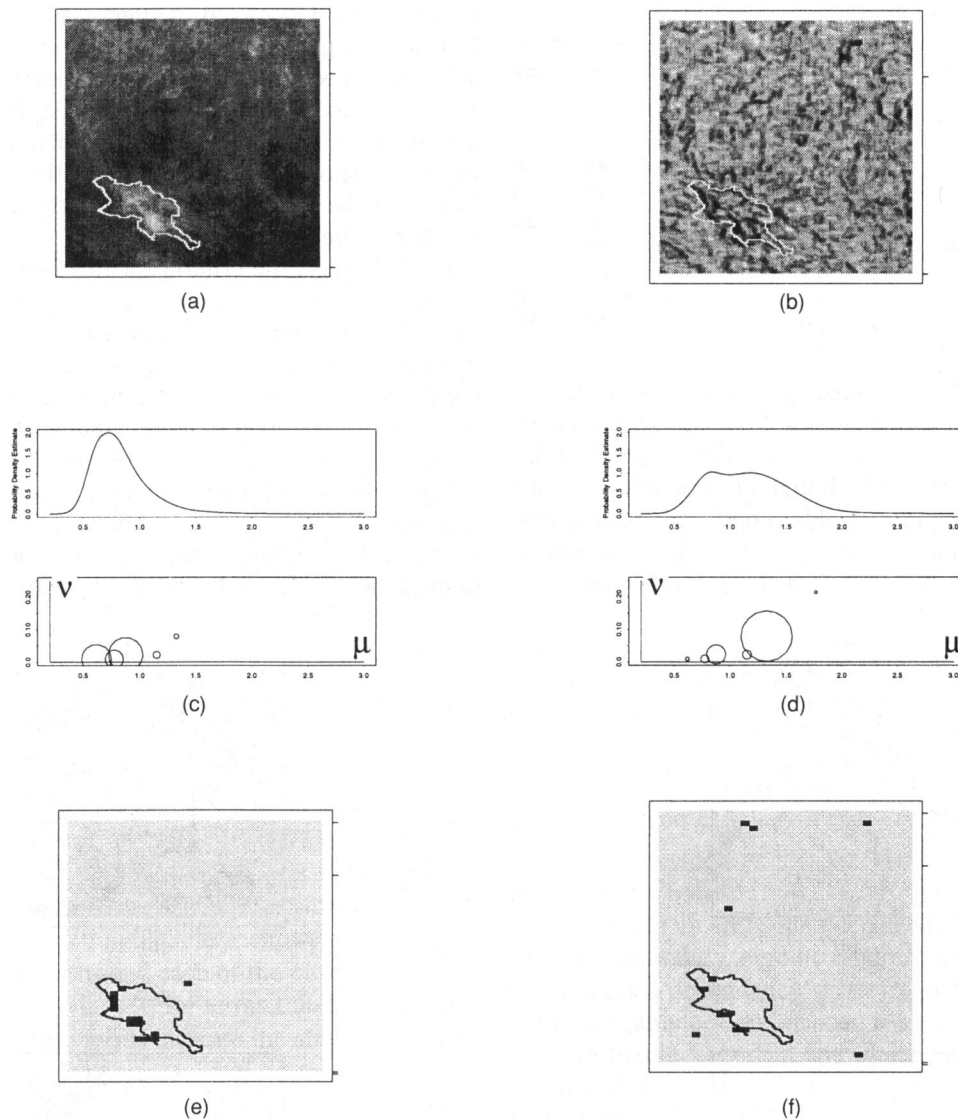
Figure 3. Mammographic Analysis Example. (a) Digitized mammogram (enhanced) and radiologist's boundary overlay; (b) local coefficient of variation field for (a) with boundary overlay; (c) probability density estimate for healthy tissue; (d) probability density estimate for tumorous tissue; (e) detections using borrowed strength; (f) detections using conventional local likelihood. The borrowed strength methodology produces a superior detection rate with significantly fewer false detections.

marginal density for a given window—is then compared, in terms of ISE, to the overall density, which is assumed to be made up mostly of "background." Those scan locations that have the largest ISE are considered anomalies, and these locations are shown in Figure 2, f and g.

As can be seen by comparing Figure 2, f and g, with the true locations for the anomalies given in Figure 2c, the nonhomogeneity detections using borrowed strength are quite impressive. Each of the three larger anomalies has been detected. No false detections—detections in regions that are identically distributed as background—are obtained. However, the smallest anomaly is not detected. This is not surprising, as the window size used in the scan process directly determines the acuity of the detection method for small regions. Results for the local likelihood are quite poor, with numerous false alarms and missed detections. This result is despite the fact that the marginal for background, f^2, is not at all close, as a probability density, to that of the anomalies, f^1. This being the case, one might naively expect superior performance in terms of false detections to come at the ex-

pense of degraded detection performance. That this is not the case can be seen in Figure 2.

Performance of the techniques for an independently distributed version of this simulation are quite similar, with the difference being that smaller windows can be used in the scan analysis due to the larger effective number of observations under independence.

3.3 Mammographic Image Analysis Example

Now consider the image analysis example depicted in Figure 3. Figure 3a presents a digitized mammogram. Included is a radiologist's drawing of the boundary of the biopsy-proven tumor. This boundary is considered to be "truth" for the purposes of analyzing the relative performance of borrowed strength versus conventional local likelihood scan analysis. Thus R^0 (the breast tissue) is divided into R^H (healthy tissue) and R^T (tumorous tissue). For this image, there are approximately 2,000 tumorous pixels and 100,000 healthy pixels. However, dependency considera-

tions imply that the effective number of observations, as compared to an independent sample, is significantly less.

The local coefficient of variation (σ/μ) field depicted in Figure 3b, rather than intensity, is used throughout this example, as it is well known that pixel brightness cannot be used to detect mammographic anomalies, but there is both physiological and empirical evidence that "texture," or a local roughness measure, is relevant (Hsiao and Sawchuck 1989; Miller and Astley 1992; Priebe et al. 1994). Local coefficient of variation is perhaps the simplest such measure.

Figure 3, c and d, give normal mixture model probability density estimates for the marginals in the two regions R^H and R^T. Included is a representation of the location in (mean, variance) space of the six terms. (The radius of the circles represents the mixing coefficient.) This representation indicates that indeed the mixture model estimates have been constrained to be identical in these individual term locations, differing only in the mixing coefficients. This fact is not at all obvious from examination of the densities themselves. Thus $\alpha^0 = (n^H/n^0)\alpha^H + (n^T/n^0)\alpha^T$, where $n^0 = n^H + n^T$. The estimate of α^0 is obtained using the adaptive mixtures procedure, and the estimates for α^0, α^H, and α^T are extremely close in terms of both visual analysis and ISE to kernel estimates, lognormal approximations, and so on, which were obtained for comparison. This suggests that—at least for this example—mixture estimates with common means and variances can well represent the true marginals. This conclusion appears to be borne out in general through an ongoing analysis of a large set of similar images (Priebe and Marchette 1996). The advantage of the mixture estimates is, of course, that they lend themselves to a borrowed strength methodology.

Figure 3, e and f, present the actual performance of the scan methodologies, in direct analogy to the analysis performed in the simulation example (Fig. 2). Again, it is clear that the nonhomogeneity detections obtained through borrowed strength are significantly more accurate than those for local likelihood, despite the obvious dissimilarity between α^H and α^T.

It would not be argued that this procedure would be used alone to detect anomalies in digital mammography. However, it is clear that any analysis of the utility of a texture-based scan methodology for aiding in an overall detection system will yield unnecessarily pessimistic results unless the applicability of borrowing strength is investigated fully.

4. COMMENTS

This article has presented the basic idea that borrowing strength can allow for better estimation of the characteristics of nonhomogeneous random fields. The idea is motivated through image analysis considerations and a parametric methodology is presented that uses profile likelihood estimation under finite mixture model assumptions. Simulation and experimental examples are included that suggest that borrowing strength will outperform conventional local-area maximum likelihood and provide superior detection of subregions of nonhomogeneity.

REFERENCES

Carlstein, E. (1988), "Nonparametric Change-Point Estimation," *The Annals of Statistics*, 16, 188–197.

Carlstein, E., and Lele, S. (1994), "Nonparametric Change-Point Estimation for Data from an Ergodic Sequence," *Theory of Probability and Its Applications*, 38, 726–733.

Cobb, G. W. (1978), "The Problem of the Nile: Conditional Solution to a Change-Point Problem," *Biometrika*, 65, 243–251.

Cox, D. R., and Reid, N. (1987), "Parameter Orthogonality and Approximate Conditional Inference" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 49, 1–39.

Cressie, N. A. C. (1993), *Statistics for Spatial Data*, New York: Wiley.

Geman, S., and Geman, D. (1984), "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721–741.

Hsiao, J. Y., and Sawchuck, A. A. (1989), "Supervised Texture Image Segmentation Using Feature Smoothing and Probabilistic Relaxation Techniques," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11, 1279–1292.

Miller, P., and Astley, S. (1992), "Classification of Breast Tissue by Texture Analysis," *Image and Vision Computing*, 10, 277–282.

Priebe, C. E. (1994), "Adaptive Mixtures," *Journal of the American Statistical Association*, 89, 796–806.

Priebe, C. E. and Marchette, D. J. (1996), "Characterizing Mammographic Tissue via Borrowed Strength Density Estimation" in *Proceedings of the 3rd International Workshop on Digital Mammography, Chicago, June 1996*, ed. K. Doi, Excerpta Medica International Congress Series (ICS), Amsterdam: Elsevier Science B.V..

Priebe, C. E., Solka, J. L., Lorey, R. A., Rogers, G., Poston, W. Kallergi, M. Qian, W., Clarke, L. P., and Clark, R. A. (1994), "The Application of Fractal Analysis to Mammographic Tissue Classification," *Cancer Letters*, 77, 183–189.