

Testing for Spatial Correlation in Nonstationary Binary Data, with Application to Aberrant Crypt Foci in Colon Carcinogenesis

Tatiana V. Apanasovich,¹ Simon Sheather,² Joanne R. Lupton,^{3,*}
Natasa Popovic,^{3,**} Nancy D. Turner,^{3,***} Robert S. Chapkin,^{3,****}
Leslie A. Braby,⁴ and Raymond J. Carroll⁵

¹Department of Statistics, Texas A&M University College Station, Texas, U.S.A.
email: tanya@stat.tamu.edu

²Australian Graduate School of Management, University of New South Wales, Sydney,
NSW 2052, Australia
email: simonsh@agsm.edu.au

³Human Nutrition, Texas A&M University College Station, Texas, U.S.A.

**email:* jlupton@tamu.edu

***email:* natasa-popovic@tamu.edu

****email:* nancy-turner@ansc.tamu.edu

*****email:* r-chapkin@tamu.edu

⁴Department of Nuclear Engineering, Texas A&M University College Station, Texas, U.S.A.
email: labraby@tamu.edu

⁵Department of Statistics, Texas A&M University College Station, Texas, U.S.A.
email: carroll@stat.tamu.edu

SUMMARY. In an experiment to understand colon carcinogenesis, all animals were exposed to a carcinogen, with half the animals also being exposed to radiation. Spatially, we measured the existence of what are referred to as aberrant crypt foci (ACF), namely, morphologically changed colonic crypts that are known to be precursors of colon cancer development. The biological question of interest is whether the locations of these ACFs are spatially correlated: if so, this indicates that damage to the colon due to carcinogens and radiation is localized. Statistically, the data take the form of binary outcomes (corresponding to the existence of an ACF) on a regular grid. We develop score-type methods based upon the Matern and conditionally autoregressive (CAR) correlation models to test for the spatial correlation in such data, while allowing for nonstationarity. Because of a technical peculiarity of the score-type test, we also develop robust versions of the method. The methods are compared to a generalization of Moran's test for continuous outcomes, and are shown via simulation to have the potential for increased power. When applied to our data, the methods indicate the existence of spatial correlation, and hence indicate localization of damage.

KEY WORDS: Aberrant crypt foci; Binary data; Carcinogenesis; Colon cancer; Conditionally autoregressive models; Moran's test; Robustness; Score tests; Spatial statistics.

1. Introduction

This article is concerned with testing for spatial correlation when the outcomes are binary. The problem arises naturally from an important question in colon carcinogenesis. In our experiments, the colon can be thought of as a cylindrical tube cut lengthwise into two pieces. One piece is used for other experiments, while the other is laid out flat onto a slide, see Figure 1. Animals are exposed to a carcinogen, with half of them also exposed to radiation. They are then sacrificed, and images of the colon are obtained by various staining devices. A typical image is given in Figure 2: a color version of this is given at <http://stat.tamu.edu/>

[~carroll/techreports.html](#). Here, we see three types of structures:

1. The grayish region is lymphatic tissue, called Peyer's patches.
2. The small white dots are normal colonic crypts, whose function is to produce cells that line the colon. More details accessible by a statistical audience on the role of colonic crypts are given in Morris et al. (2001, 2002, 2003, to appear).
3. The larger dark and distended regions are *aberrant crypt foci*, or ACF for short, which are crypts that have been

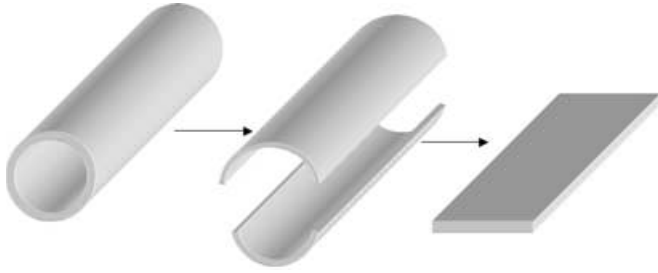


Figure 1. A cartoon showing the process of laying the colon onto a slide.

changed morphologically by the carcinogen and radiation. For technical reasons, it is not possible to determine accurately the existence of an ACF within lymphatic tissue (Peyer's patches). See Bird (1995) and Bird and Good (2000) for the importance of ACF in colon carcinogenesis.

Our interest is in the aberrant crypt foci, which we denote by ACF. These are precursors to colon cancer, and hence almost everything about them are of biological relevance.

The data are clearly naturally spatial. By any measure, they are also nonstationary, as the proximal (front) and distal (back) regions of the colon behave far differently in terms of the likelihood of ACF formation. It is not feasible in practice to measure the locations of ACF, so we formed a rectangular grid of locations and recorded (by hand) the existence of an ACF within each location; see Figure 3 for an illustration. Thus, the data available to us are the grid of locations along with the binary indicator of an ACF.

Here, we consider a particular problem, namely, testing whether the existence of an ACF at one location is predic-

tive of an ACF at neighboring locations. Hence, we want to test for spatial dependence, using the binary outcome of the existence of an ACF. Such spatial dependence, if it exists, is interesting. It suggests that damage to the colon is localized regionally. There may be areas in which greater levels of damage in response to an insult could lead to focused areas of inflammatory responses, or an alteration in the release of signaling molecules that could then affect the regulation of homeostatic mechanisms in colonocytes in adjacent crypts. This localization may help explain why tumors develop from particular ACF, but not from all ACF formed in response to a carcinogen insult.

Thus, testing for the spatial dependence is of biological interest in itself and, as such, it is not merely testing for a nuisance parameter.

One way to test for such spatial dependence is to build a spatial regression model that includes independence as a special case, and then to test for this special case. We instead develop simpler methods that avoid the need to fit any particular spatial model, while still allowing for the nonstationarity in the mean that is inherent in our problem. Hence, the method developed has the potential to be widely applicable in practice.

One of the methods we develop is the binary version of Moran's test (Moran, 1948) that allows for spatial nonstationarity. However, our primary focus is on a test motivated by score testing ideas, and robust versions of this test that ensure that one or two neighboring pairs of ACFs will not in themselves lead to a declaration of spatial dependence.

An outline of this article is as follows. In Section 2, we describe Moran's test and derive the score-type test, which is based upon the Matern and conditionally autoregressive (CAR) correlation models. In Section 3, we describe in detail the main robustness issue with the score-type test and derive

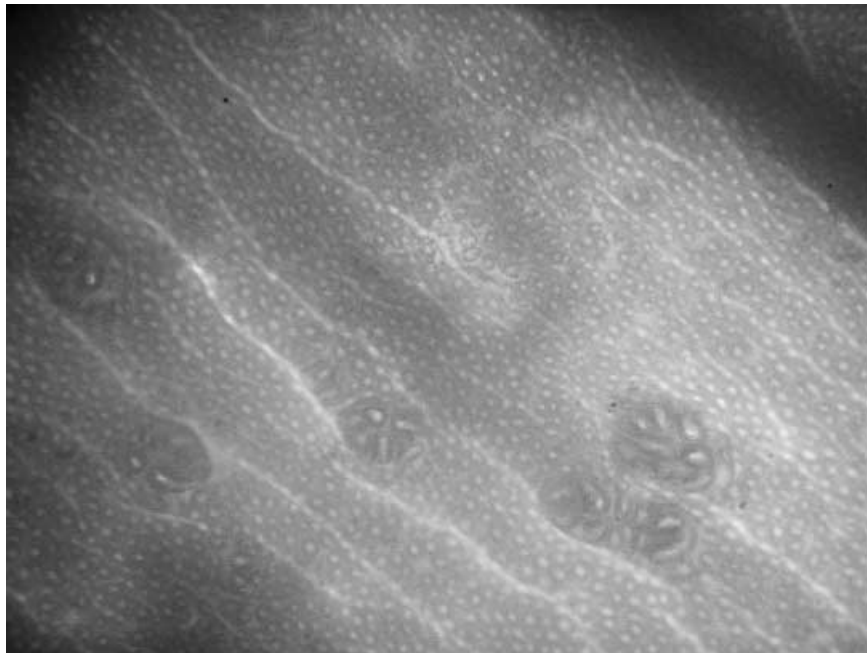


Figure 2. A colon laid lengthwise, showing a Peyer's patch (gray region), normal colon crypts (white dots), and aberrant crypt foci (dark distended shapes).

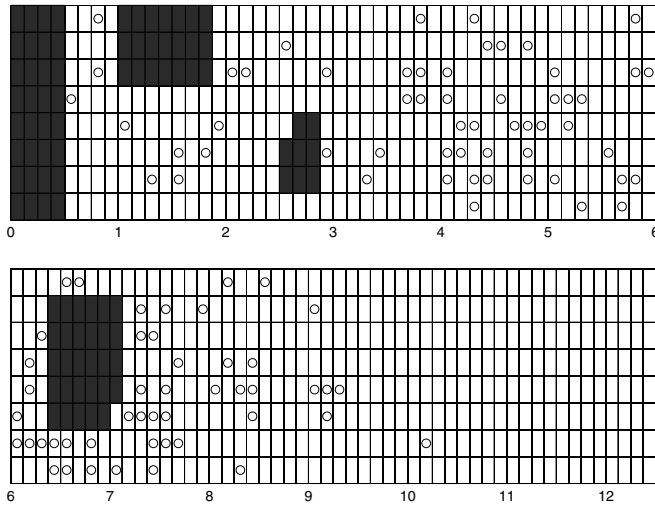


Figure 3. A gridded plot of a rat 263. Dots are coded for ACF indicators and shaded areas are Peyer’s patches.

robust alternatives to this test. Section 4 describes a set of simulations that suggest that the score test and its robust modifications can be more powerful than Moran’s test, as well as having test level closer to the nominal. Section 5 returns to the aberrant crypt foci data in detail and shows some evidence of spatial correlation, and hence localization of damage to the colon. Section 6 has concluding remarks. All technical details are given in an Appendix.

2. Methods

Generalized linear mixed models are widely used models for spatially dependent binary data (Breslow and Clayton, 1993; Diggle, Moyeed, and Tawn, 1997; Heagerty and Lele, 1998). These models are convenient for modeling the dependence of a response variable, Y_i , measured at $i = 1, \dots, n$ sites, as well as on measured covariates, X_i . We use a multivariate probit model to model spatial dependence and nonstationarity. Let $I(\cdot)$ be the indicator function. Let the ϵ_i for $i = 1, \dots, n$ be independent and normally distributed, with mean 0 and variance 1. Let λ_i denote random effects responsible for possible spatial dependence. For a parameter ρ and a correlation matrix $\Omega(\rho)$, the λ s are assumed to be normally distributed, with mean 0 and covariance matrix $\sigma_\lambda^2 \Omega(\rho)$. Let μ_i be systematic effects incorporating nonstationarity. Then, the multivariate probit model is defined as $Y_i = I(\mu_i + \lambda_i + \epsilon_i > 0)$, so that

$$\text{pr}(Y_i = 1 \mid \lambda_i, \mu_i) = \Phi(\mu_i + \lambda_i), \tag{1}$$

where $\Phi(\bullet)$ is the univariate standard normal distribution function. We are interested in assessing whether $\rho = 0$, in which case $\Omega(0)$ is the identity matrix.

The rest of this section is taken up with defining the two methods we use.

2.1 Moran’s Test

Moran’s test for spatial dependence (Moran, 1948) was developed for stationary data. For stationary data, $\mu_i = \mu(X_i) \equiv \mu$,

and the test is as follows. Let Z^{vec} be the vector of observations Y minus their sample mean. Let W^{mat} be an $n \times n$ matrix with the (i, j) the element equal to 1 if sites (i, j) , $i \neq j$ are neighbors and equal 0 otherwise. Moran’s test statistic is (up to a constant of proportionality) $(Z^{\text{vec}})^T W^{\text{mat}} Z^{\text{vec}} / (Z^{\text{vec}})^T Z^{\text{vec}}$. Note that Moran’s test statistic takes on the classic form of any autocorrelation coefficient: the numerator term is a measure of covariance and the denominator term is a measure of variance. Its values are compared to a nontrivial expression, see Cliff and Ord (1981, pp. 19–21).

For nonstationary numerical data, Moran’s test is usually modified (Cliff and Ord, 1981) by subtracting predicted values from the observations rather than the mean.

For our case of nonstationary binary data, we modify Moran’s test in the usual way, namely by letting Z^{vec} be the vector of standardized residuals from an ordinary Probit regression of the Y s on the X s: $(Y_i - \hat{Y}_i) / \{\hat{Y}_i(1 - \hat{Y}_i)\}^{1/2}$.

2.2 Score Test

Rao’s score statistic (Rao, 1973) is a standard tool for carrying out hypothesis testing. In many situations, it has the advantage over likelihood ratio and Wald tests, because all calculations are carried out under the hypothesis, except for the derivation of the test statistic itself. In our context of a multivariate probit model, it does not appear possible to derive such an explicit formula for the score test for an arbitrary correlation function.

We instead take a different approach, one that yields a readily computed test statistic. Our idea is to look at *pairs* of observations, and derive a score test statistic for correlation using such pairs when the correlation is of the Matern class (Stein, 1999). We will then combine this test statistic over many pairs. As we show in the Appendix, it turns out that the resulting test is the same as the score test for a particular version of the conditionally autoregressive correlation model (Besag, 1974; Richardson, Guihenneuc, and Lasserre, 1992; Cressie, 1993).

2.2.1. Score statistic for pairs. We first compute the joint probability distribution of any two binary responses Y_i and Y_j . Let $\Phi(\bullet)$ and $\phi(\bullet)$ be the univariate standard normal distribution and density functions, respectively. Let $\Phi_2(\mu_1, \mu_2, \rho)$ be the bivariate standard normal probability of being below μ_1 and μ_2 when the correlation is ρ . Define $\mu_i^* = \mu_i / (1 + \sigma_\lambda^2)^{1/2}$. Then,

$$\begin{aligned} \text{pr}(Y_i = 1 \mid \mu_i) &= \Phi(\mu_i^*); \\ \text{pr}(Y_i = 1, Y_j = 1 \mid \mu_i, \mu_j) &= \Phi_2\{\mu_i^*, \mu_j^*, \sigma_\lambda^2 \Omega_{ij}(\rho) / (1 + \sigma_\lambda^2)\}. \end{aligned} \tag{2}$$

We will calculate the score-type test based on $k = 1, \dots, N$ pairs. Consider the k th pair $(Y_{1k} = i, Y_{2k} = j)$, where we write $\text{pr}(Y_{1k} = 1) = \Phi(\mu_{1k}^*)$ and $\text{pr}(Y_{2k} = 1) = \Phi(\mu_{2k}^*)$. Also define $\text{pr}(Y_{1k} = i, Y_{2k} = j) = \pi_{ijk}(\rho, \sigma_\lambda^2) = \pi_{ijk}$, so that $\text{pr}(Y_{1k} = i) = \pi_{i \cdot k}$ and $\text{pr}(Y_{2k} = j) = \pi_{\cdot j k}$, where the “dots” indicate summation. A useful fact is that

$$\pi_{11k}(\rho = 0, \sigma_\lambda^2) = \Phi(\mu_{1k}^*) \Phi(\mu_{2k}^*). \tag{3}$$

If we define $Z_{ijk} = I(Y_{1k} = i, Y_{2k} = j)$, then the log likelihood is

$$\log L(\rho) = \sum_k \{Z_{00k} \log(\pi_{00k}) + Z_{01k} \log(\pi_{01k}) + Z_{10k} \log(\pi_{10k}) + Z_{11k} \log(\pi_{11k})\}. \tag{4}$$

Formal differentiation of (4) and evaluated at the null hypothesis $\rho = 0$ would yield the essential part of the score statistic. Let d_k be the Euclidean distance between the members of the k th pair. Recent literature (e.g., Stein, 1999, p. 31–33) advocates the use of the Matern family, for which the covariance functions have general form

$$C_M(d_k) = \frac{\sigma_S^2}{2^{\nu-1}\Gamma(\nu)}(d_k/\rho)^\nu K_\nu(d_k/\rho), \quad \sigma_S^2, \rho, \nu > 0, \tag{5}$$

where K_ν is the modified Bessel function of order ν which, in turn, do not have a closed form for general ν . If $\nu = m + (1/2)$ for $m = 0, 1, 2, \dots$, then (5) has a simple form. However, as we show in the Appendix, there is a difficulty with this approach. When we use as the correlation function a member of the Matern class with $\nu = m + \frac{1}{2}$, the score evaluated at $\rho = 0$ is identically 0, so that nothing useful results. As we show in the Appendix, our approach is to focus only on those pairs that are exactly the same distance apart, in which case the score becomes a nontrivial statistic times a common constant that equals 0 when $\rho = 0$. Removing this common constant leads to a score equal to

$$\mathcal{G}_k(\mu_{1k}^*, \mu_{2k}^*) = \frac{(Y_{1k} - \pi_{1k})(Y_{2k} - \pi_{1k})\phi(\mu_{1k}^*)\phi(\mu_{2k}^*)}{\pi_{1k}(1 - \pi_{1k})\pi_{1k}(1 - \pi_{1k})}. \tag{6}$$

In practice, we implement our score-type test as follows. Recall model (1), and let the μ 's depend on covariates X and a parameter β_* , i.e., $\mu(X, \beta_*)$, with the property that for any constant c , $c\mu(X, \beta_*) = \mu(X, \beta_{**})$ for some β_{**} . As seen in (2), under the null hypothesis, the Y s are independent and $\text{pr}(Y = 1|X) = \Phi\{\mu(X, \beta_*)/(1 + \sigma_\lambda^2)^{1/2}\} = \Phi\{\mu(X, \beta)\}$, say. Thus, we can estimate β consistently under both the null and alternative models via a probit regression with probability function $\text{pr}(Y = 1|X) = \Phi\{\mu(X, \beta)\}$. Call the estimate $\hat{\beta}$. Modify (6) appropriately by defining

$$\mathcal{H}_k(\beta) = \frac{[Y_{1k} - \Phi\{\mu(X_{1k}, \beta)\}][Y_{2k} - \Phi\{\mu(X_{2k}, \beta)\}]\phi\{\mu(X_{1k}, \beta)\}\phi\{\mu(X_{2k}, \beta)\}}{\Phi\{\mu(X_{1k}, \beta)\}[1 - \Phi\{\mu(X_{1k}, \beta)\}]\Phi\{\mu(X_{2k}, \beta)\}[1 - \Phi\{\mu(X_{2k}, \beta)\}]}. \tag{7}$$

The variance of (7), under the hypothesis of no spatial correlation, is clearly

$$\mathcal{V}_k(\beta) = \frac{[\phi\{\mu(X_{1k}, \beta)\}\phi\{\mu(X_{2k}, \beta)\}]^2}{\Phi\{\mu(X_{1k}, \beta)\}[1 - \Phi\{\mu(X_{1k}, \beta)\}]\Phi\{\mu(X_{2k}, \beta)\}[1 - \Phi\{\mu(X_{2k}, \beta)\}]}. \tag{8}$$

Our test statistic then is

$$\frac{\sum_k \mathcal{H}_k(\hat{\beta})}{\left\{ \sum_k \mathcal{V}_k(\hat{\beta}) \right\}^{1/2}}. \tag{9}$$

We show in the Appendix that under the hypothesis of no spatial correlation, (9) is asymptotically standard normal; hence, the hypothesis of no spatial correlation can be tested by referring (9) to standard normal quantiles. Notice that the terms in the test statistic's numerator are not independent, though they are uncorrelated. The method of Commenges and Jacqmin-Gadda (1997) can be used to prove asymptotic normality, under the assumption that $\mu(X, \beta)$ and its first two derivatives in β are bounded. The result is asymptotic in the number of pairs, with the same scale of grid and the same spatial scale of autocorrelation.

Terms similar to (6) and the numerator of (9) were derived in a different context by le Cessie and van Houwelingen (1994). They considered the case of classical clustered and not spatial data; this context is crucial, because in our problem, the number of pairs within each animal/colon is large. Even taking this difference of context into account, there still remain important differences with our work. They considered the case that, in effect, the correlation matrix $\Omega(\rho)$ has common correlation for all elements—something not likely to hold for spatial data. Their sum in (9) would thus be over all pairs, and not just pairs of neighbors. In addition, they required multiple clusters (animals). Also, because of their different context, they were not led to notice the problem raised above with straightforward use of the Matern class. They were also not led to show that, under the null hypothesis of no spatial correlation, then for a single animal and a large number of locations, the denominator of (9) is a consistent estimator of the standard deviation of the numerator under the hypothesis of no spatial correlation, taking into account the estimation of β .

Also somewhat similar to our test is work of Jacqmin-Gadda et al. (1997), with their version of (7) having elements of the form $[Y_{1k} - \Phi\{\mu(X_{1k}, \beta)\}][Y_{2k} - \Phi\{\mu(X_{2k}, \beta)\}]w(X_{1k}, X_{2k})$ across all pairs (not just necessarily neighbors) and for an arbitrary function $w(\bullet)$. Their motivation and actual test statistics are however very different: in place of our (1), they start from a logistic family, with a correlation model for the λ -terms that is fixed in advance. In contrast, our work, and that of le Cessie and van Houwelingen, is based on somewhat more standard spatial correlation models involving a free parameter: Matern and CAR in our case, and equicorrelated for le Cessie and van Houwelingen.

2.3 Selecting the Pairs

There are many ways to organize all observations into pairs, bearing in mind that our score-type test is based on the idea that observations in all pairs should be the same distance apart. What we do is the following. Each observation is paired with its closest neighbors in any vertical and horizontal direction, so that each observation will make as many pairs as it has neighbors. Hence, interior observations will have four pairs, ones on edges will have three pairs and ones in corners will have two pairs.

2.4 Conditionally Autoregressive Models

A simple version of the conditionally autoregressive correlation model (Besag, 1974; Richardson et al., 1992) is that the λ s have covariance matrix $\sigma_\lambda^2(I - \rho C)^{-1}$, where C is chosen to be a neighborhood matrix whose (i, j) th

element is equal to 1 if region i and region j ($i \neq j$) are neighbors and I is an identity matrix of appropriate dimension. As we show in the Appendix, Section A.2, our test (9) is the same as the score test for this model and, in this regard, is more general than simply the Matern class.

3. Robust Score Tests

If the event rates are rare, then one would not expect to have two neighboring pairs of observations for which both Y 's equal 1, unless the correlations are reasonably high. Because of this, one would expect that any score-type test would have the property that when the event rates are small, a pair of neighboring Y 's equal 1 would lead to the rejection of the hypothesis of no spatial correlation. Our test does indeed have this property. In fact, if $Y_{1k} = Y_{2k} = 1$, then (7) becomes $\phi\{\mu(X_{1k}, \beta)\}\phi\{\mu(X_{2k}, \beta)\}/[\Phi\{\mu(X_{1k}, \beta)\}\Phi\{\mu(X_{2k}, \beta)\}]$, which is unbounded as $\mu(X_{1k}, \beta) \rightarrow -\infty$ and $\mu(X_{2k}, \beta) \rightarrow -\infty$. On the other hand, the variance contribution in (8) is bounded in such a circumstance, in fact, it converges to 0. This means that with such a single pair, the test statistic (9) converges to ∞ , as intuition suggests.

The above fact may be looked upon as a strength of the score-type test, but it may also be a flaw. In a particular data set, there may be little evidence of spatial correlation, except for a single pair, and this would make one wary of claiming such a correlation.

In this section, we propose a simple modification of our score-type test that limits the influence of any one pair on the value of the score-type test, while still allowing for considerable power. The method is based on ideas from robustness theory.

To develop this method, rewrite (7) as follows:

$$\begin{aligned} \mathcal{H}_k(\beta) &= [Y_{1k} - \Phi\{\mu(X_{1k}, \beta)\}][Y_{2k} - \Phi\{\mu(X_{2k}, \beta)\}]R(X_{1k}, X_{2k}, \beta); \\ R(X_{1k}, X_{2k}, \beta) &= \frac{\phi\{\mu(X_{1k}, \beta)\}\phi\{\mu(X_{2k}, \beta)\}}{\Phi\{\mu(X_{1k}, \beta)\}[1 - \Phi\{\mu(X_{1k}, \beta)\}]\Phi\{\mu(X_{2k}, \beta)\}[1 - \Phi\{\mu(X_{2k}, \beta)\}]}. \end{aligned} \tag{10}$$

It is the function $R(\bullet)$ that is unbounded, and hence, in the terminology of robustness, (10) is a statistic with unbounded influence. Note that $R(\bullet)$ does not depend on the responses, and in this respect, acts in a fashion similar to that of a design matrix in linear regression.

Methods to bound the influence of “design” points in logistic and linear regression have been investigated by Carroll and Pederson (1993) and by Simpson, Ruppert, and Carroll (1992), respectively. The idea is to redefine the test statistic (10) as

$$\begin{aligned} \mathcal{H}_{k, \text{robust}}(\beta) &= [Y_{1k} - \Phi\{\mu(X_{1k}, \beta)\}][Y_{2k} - \Phi\{\mu(X_{2k}, \beta)\}] \\ &\quad \times H\{R(X_{1k}, X_{2k}, \beta)\}, \end{aligned}$$

for an arbitrary function $H(\bullet)$, and to redefine its variance as

$$\begin{aligned} \mathcal{V}_{k, \text{robust}}(\beta) &= \Phi\{\mu(X_{1k}, \beta)\}[1 - \Phi\{\mu(X_{1k}, \beta)\}] \\ &\quad \times \Phi\{\mu(X_{2k}, \beta)\}[1 - \Phi\{\mu(X_{2k}, \beta)\}] \\ &\quad \times H^2\{R(X_{1k}, X_{2k}, \beta)\}. \end{aligned}$$

We now define two classes of weight functions. The first follows Carroll and Pederson (1993). Let σ_R be the median of the terms $R(X_{1k}, X_{2k}, \beta)$. Define $L_k = R(X_{1k}, X_{2k}, \beta)/\sigma_R$ and for a constant b_{cp} , define $H_{cp}\{R(X_{1k}, X_{2k}, \beta)\} = R(X_{1k}, X_{2k}, \beta)\{1 - (L_k/b_{cp})^2\}^3 I(L_k \leq b_{cp})$. If $b_{cp} = \infty$, we of course get the score-type test, while smaller values of b_{cp} bound the influence. We experimented with some simulated data, and finally choose $b_{cp} = 3$.

The method of Simpson et al. (1992) is similar, namely, $H_{\text{simpson}}\{R(X_{1k}, X_{2k}, \beta)\} = R(X_{1k}, X_{2k}, \beta) \min\{1, (b_{\text{simpson}}/L_k)^\alpha\}$. We took ($b_{\text{simpson}} = 1, \alpha = 1$) and ($b_{\text{simpson}} = 2, \alpha = 2$).

4. Simulations

We performed simulations under two scenarios, with the number of replications equal to 1000. In both cases, we took the test level to be 0.05.

In scenario 1, we took data from a rat labeled as 263 in our experiment. Let X be the horizontal distance from the distal part of the colon, let Z be the vertical distance, and let D be the Euclidean distance to the nearest Peyer’s patch. We fit a probit model to these data with probability function

$$\Phi(\beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 Z + \beta_4 D). \tag{11}$$

The original grid was of size 100×8 . For the simulations, we used half of that grid, 50×8 . Data were generated from the ordinary probit fit to this model, with $\beta_0 = -1.83, \beta_1 = 6.96, \beta_2 = -7.34, \beta_3 = -3.12$, and $\beta_4 = -3.46$. We generated correlated data with $\sigma_\lambda^2 = 1$ via the Matern correlation function, with index $3/2$: $\Omega_{ij}(\rho) = \text{corr}(\lambda_i, \lambda_j) = \exp(-d_{ij}/\rho) \times (1 + d_{ij}/\rho)$.

In Scenario 2, data were generated from the model with the following probability function $\Phi(\beta_0 + \beta_1 X + \beta_2 X^2)$, with $\beta_0 = -4.5, \beta_1 = 12.03$, and $\beta_2 = -8, X \in [0.0186, 1]$. The grid was taken to be of size 54×8 . We generated correlated data with $\sigma_\lambda^2 = 1$ via the Matern correlation function, with index $3/2$.

Since we select pairs that are of the same distance apart, all Ω_{ij} are equal to, say, Ω . Define $\psi = \sigma_\lambda^2/(1 + \sigma_\lambda^2)\Omega$; see the Appendix. For power comparison purposes, we vary ψ .

In addition to the test statistics described previously, we also computed the level and power for the test statistic based on (10), but with the function $R(\bullet) \equiv 1$. The motivation for this comes from logistic regression. If in (10), the normal distribution function $\Phi(\bullet)$ and its density $\phi(\bullet)$ were replaced by the logistic distribution and density functions, then $R(\bullet) \equiv 1$. Given estimates of β , the test would formally have bounded influence, and it is one of the tests evaluated in the simulation study done by Jacqmin-Gadda et al. (1997).

The results are given in Table 1. We observed, in a variety of simulations, that Moran’s test had a tendency to be anti-conservative and fail to maintain the level of 0.05. Because of this, we also display results when Moran’s test was adjusted to have exact level 0.05. This was done by running the null case of no spatial correlation many times, computing Moran’s test statistic, and then choosing as the rejection point that value that gave exact null level 0.05. For both scenarios, all the score tests came reasonably close to maintaining the nominal level. In addition, the power of the robust score tests and the original score test are nearly the same. In both scenarios

Table 1

Results of the simulation. The number of replications is 1000. The parameter $\psi = \sigma_\lambda^2 / (1 + \sigma_\lambda^2)\Omega$ indicates the strength of the spatial correlation, with $\psi = 0$ being the case of no spatial dependence. Hence, the column $\psi = 0.00$ is the level of the test, while the other columns are the powers. Scenario 1 and scenario 2 are described in the text. The tests are as follows: “score” refers to the usual score test, “moran” to Moran’s test, “scorecp” to the robust score test with Carroll and Pederson’s weight function, “scores1” to Simpson’s weight function with $b = 3$, $\alpha = 1$, “scores2” to Simpson’s weight function with $b = 4$ and $\alpha = 2$, and “scoreJG” to score test similar to Jacqmin-Gadda et al. (1997). “adj. moran” refers to Moran’s test adjusted to have level exactly 0.05.

Simulation Scenario 1							
$\psi =$	0.00	0.10	0.15	0.20	0.25	0.30	0.40
score	0.05	0.13	0.28	0.46	0.63	0.73	0.82
moran	0.06	0.14	0.26	0.42	0.60	0.66	0.75
adj. moran	0.05	0.11	0.20	0.36	0.55	0.61	0.72
scorecp	0.05	0.13	0.28	0.46	0.63	0.73	0.82
scores1	0.05	0.13	0.28	0.46	0.63	0.73	0.82
scores2	0.05	0.13	0.28	0.46	0.63	0.73	0.82
scoreJG	0.05	0.13	0.28	0.46	0.63	0.73	0.82
Simulation Scenario 2							
$\psi =$	0.00	0.10	0.15	0.20	0.25	0.30	0.40
score	0.05	0.22	0.48	0.73	0.89	0.95	0.98
moran	0.07	0.24	0.45	0.68	0.83	0.90	0.91
adj. moran	0.05	0.21	0.42	0.65	0.81	0.88	0.90
scorecp	0.05	0.22	0.48	0.73	0.89	0.95	0.98
scores1	0.05	0.22	0.48	0.73	0.89	0.95	0.98
scores2	0.05	0.22	0.48	0.73	0.89	0.95	0.98
scoresJG	0.05	0.22	0.48	0.73	0.89	0.95	0.98

score tests have greater power than Moran’s test, when the latter was adjusted to have correct level. The test discussed above, in which $R(\bullet) \equiv 1$, does quite well in terms of power in these simulations. However, we have done other simulations where we do see some loss of power with this choice.

5. Aberrant Crypt Foci Experiment

The introduction describes the aberrant crypt foci (ACF) experiment, but here we make a few additional remarks about the data collection. The typical rat colon was approximately 10 cm–12 cm long when laid out on a slide. This is far larger than can be read in one go from a microscope. Instead, what was done was to first start with a piece of paper, somewhat like that given in Figure 2, but without the grid lines superimposed. We then simply physically moved the slide horizontally and vertically through the microscope, starting from the proximal part of the colon, noting approximately how far along we were in physical (slide) distance. As we observed ACFs and Peyer’s patches, we made small notations in pencil on the piece of paper.

The image in Figure 1 is approximately 1/5–1/6 square centimeters, and thus takes up approximately 9 grid boxes. This, by the way, is the only image that was recorded, and then primarily for the purpose of illustrating ACFs in this article. One can see multiple (7) ACFs, along with a small Peyer’s patch; in other sections everything seen in the micro-

scope was a Peyer’s patch, as illustrated in Figure 2. What was recorded then was the approximate location of this square image, as well as the existence of ACFs and Peyer’s patches.

Since the work was done manually, the locations of the ACFs and Peyer’s patches, as marked on the paper, are not exact. The first and last authors observed the process numerous times; on this basis, and in collaboration with our colleagues, they decided to use the grid as displayed in Figure 2. We felt that any finer grid would have led to far too much misclassification of location.

All rats were exposed to a chemical carcinogen. One half of the rats were also exposed to radiation. Rats were sacrificed at 4, 6, and 8 weeks, and their colons removed and assayed. There were thus 6 rat groups, and 7 were in each group.

Using model (11), we first computed our score tests on an animal-by-animal basis. We then combined the results as follows. For rat $r = 1, \dots, 7$ in rat group $g = 1, \dots, 6$, each test statistic can be written as T_{rg}/S_{rg} , where T is the numerator of the test statistic and S is its denominator. Since the rats are independent, a simple way to combine the data in a group is to compute the “combined” test statistic

$$\sum_{r=1}^7 T_{rg} / \left(\sum_{r=1}^7 S_{rg}^2 \right)^{1/2} \tag{12}$$

The results are given in Tables 2–3. At 4 weeks after administration of the carcinogen, there is little evidence of strong spatial correlation, with only 1 animal in the irradiated and nonirradiated groups having evidence of correlation, so the combined test was not statistically significant.

However, at 6 and 8 weeks, there are many more ACFs. This is perhaps not unexpected, since it takes some time after an insult via carcinogen or radiation before ACFs form. As seen in Tables 2–3, the combined tests are highly statistically significant at these time points: 10 of the 28 individual rats show individual evidence of spatial correlation. It seems to us, then, that the evidence is fairly strong for spatial correlation within the 6–8 week time period. See McLellan, Medline, and Bird (1991) for discussion of the role of time in developing ACF.

There is an interesting feature in Table 2. Although no p-value for the “18” group (irradiated rats at 8 weeks) was less than 0.05, the overall p-value using the combined test (12) was 0.02. This phenomenon can be explained as follows. It turns out that for these animals, the denominators S_{rg}^2 of (12), for rat r in group g , were on average approximately equal to 14 over the rats. Suppose we know that the numerators, of (12), T_{rg} , for $r = 1, \dots, 7$, are $\text{Normal}\{\mu, (14.76)^2\}$. Then, each individual test has little power for testing $H_0: \mu = 0$, but the mean of the T ’s has much more power. In other words, we might expect the situation for the “18” rats to be the rule, rather than the exception. Indeed, if we set μ equal to 13.14, then the mean of the numerators of the combined test (12), has power 66%, while the univariate tests have power 14%.

The choice of weights $R(\bullet) \equiv 1$ mentioned by Jacqmin-Gadda et al. (1997) sometimes has quite different (higher) p-values from the score tests, with changes in statistical significance at the rat level; note, for example, rat 164. One might expect this to happen if the data were actually generated by a CAR model, for example.

Table 2

Significance levels for irradiated rats. The tests are defined in Table 1. "Combined" is the combined test statistic. The first leading digit in the rat number indicates the animals were irradiated, the second number is the time of sacrifice and the third is the rat number in its group.

rat	score	moran	scorecp	scores1	scores2	scoreJG
141	0.00	0.01	0.00	0.00	0.00	0.00
142	0.67	0.73	0.66	0.66	0.66	0.66
143	0.49	0.45	0.49	0.50	0.50	0.48
144	0.59	0.75	0.60	0.60	0.60	0.62
145	0.20	0.13	0.20	0.20	0.20	0.22
146	0.27	0.11	0.28	0.27	0.27	0.37
147	0.21	0.12	0.22	0.21	0.21	0.24
Combined	0.29	0.53	0.24	0.29	0.29	0.19
161	0.04	0.07	0.04	0.04	0.04	0.04
162	0.98	0.86	0.97	0.98	0.98	0.91
163	0.00	0.00	0.00	0.00	0.00	0.00
164	0.05	0.00	0.05	0.05	0.05	0.09
165	0.40	0.79	0.39	0.40	0.40	0.31
166	0.00	0.00	0.00	0.00	0.00	0.00
167	0.86	0.73	0.86	0.86	0.86	0.86
Combined	0.00	0.00	0.00	0.00	0.00	0.00
181	0.28	0.59	0.27	0.28	0.28	0.18
182	0.91	0.86	0.91	0.91	0.91	0.96
183	0.05	0.05	0.05	0.05	0.05	0.05
184	0.80	0.54	0.81	0.80	0.80	0.87
185	0.28	0.25	0.28	0.28	0.28	0.27
186	0.13	0.11	0.13	0.13	0.13	0.14
187	0.89	0.99	0.88	0.89	0.89	0.89
Combined	0.02	0.02	0.02	0.02	0.02	0.02

Table 3

Significance levels for non-irradiated rats. The tests are defined in Table 1. "Combined" is the combined test statistic. The first leading digit in the rat number indicates the animals were not irradiated, the second number is the time of sacrifice, and the third is the rat number in its group.

rat	score	moran	scorecp	scores1	scores2	scoreJG
241	0.64	0.39	0.64	0.64	0.64	0.71
242	0.00	0.00	0.00	0.00	0.00	0.00
243	0.52	0.72	0.46	0.51	0.52	0.44
244	0.89	0.82	0.89	0.89	0.89	0.90
245	0.78	0.40	0.77	0.78	0.78	0.92
246	0.46	0.51	0.45	0.46	0.46	0.44
247	0.51	0.45	0.51	0.51	0.51	0.56
Combined	0.38	0.12	0.38	0.38	0.38	0.47
261	0.00	0.00	0.00	0.00	0.00	0.00
262	0.02	0.00	0.04	0.02	0.02	0.12
263	0.06	0.06	0.06	0.06	0.06	0.06
264	0.25	0.23	0.25	0.25	0.25	0.32
265	0.03	0.00	0.03	0.03	0.03	0.05
266	0.52	0.67	0.53	0.52	0.52	0.44
267	0.68	0.47	0.69	0.68	0.68	0.75
Combined	0.00	0.00	0.00	0.00	0.00	0.00
281	0.03	0.00	0.03	0.03	0.03	0.06
282	0.34	0.46	0.33	0.34	0.34	0.30
283	0.67	0.31	0.66	0.67	0.67	0.83
284	0.78	0.86	0.76	0.78	0.78	0.65
285	0.81	0.61	0.82	0.81	0.81	0.91
286	0.98	0.52	0.97	0.98	0.98	0.65
287	0.00	0.00	0.00	0.00	0.00	0.00
Combined	0.00	0.00	0.00	0.00	0.00	0.00

6. Discussion

We have described an important experiment in colon carcinogenesis, where the responses are binary and fall into a spatial alignment, with clear nonstationarity. One key question of interest is whether there is any spatial correlation: its existence would suggest that the response of interest, aberrant crypt foci, are localized in the colon, and thus that regions are affected by radiation and a carcinogen. Our analysis of the aberrant crypt foci experiment suggests that spatial correlation is present at 6–8 weeks after administration of the carcinogen, with or without radiation.

We developed a score-type test for this problem. The original motivation was the Matern class of correlation functions, although we also derived the same test using a particular form of the CAR model. The score-type test method requires no modeling of the correlation *per se*, and is easily computed. We also developed robust score-type tests that bound the influence of a few observations on the score test. The methods are shown via simulation to have test level near the nominal, and also to have, in some circumstances, more power than a modification of Moran's test.

ACKNOWLEDGEMENTS

Our research was supported by the National Cancer Institute (CA57030, CA61750, and CA82907), NASA (NPFR00202),

the Texas A&M Center for Environmental and Rural Health, the National Institute of Environmental Health Sciences (P30-ES09106), and the Australian Graduate School of Management. We think the associate editor and two referees for their detailed and helpful comments, and especially for the reference to Jacqmin-Gadda et al. (1997).

RÉSUMÉ

Dans une expérience pour comprendre la carcinogénèse du colon, tous les animaux ont été exposés au carcinogène tandis que la moitié des animaux ont aussi été exposés aux radiations. Spatialement, nous mesurons l'existence de ce que nous appellerons foyers cryptaux aberrants (ACF), c'est-à-dire des cryptes du colons morphologiquement modifiées qui sont connues pour être des précurseurs du développement du cancer du colon. La question biologique d'intérêt est des savoir si les localisations de ces ACF sont spatialement corrélées; si c'est le cas cela indique que le dommage sur le colon dûs au carcinogène et aux radiations est localisé. Statistiquement, les données prennent la forme de variables binaires (correspondant à l'existence d'un ACF) sur une grille régulière. Nous développons des méthodes de type score basées sur les modèles de Matern et les modèles d'autorégression conditionnelle (CAR) pour tester la corrélation spatiale dans de telles données, tout en autorisant la non-stationarité. A cause d'une particularité technique du test de type score, nous

développons aussi des versions robustes de la méthode. Les méthodes sont comparées à une généralisation du test de Moran pour les variables continues, et des simulations montrent qu'elles ont le potentiel d'une puissance plus grande. Appliquées à nos données, les méthodes indiquent l'existence d'une corrélation spatiale, et donc indiquent une localisation du dommage.

REFERENCES

- Abramowitz, M. and Stegun, I. (1965). *Handbook of Mathematical Functions*. New York: Dover.
- Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society, Series B* **36**, 192–236.
- Bird, R. P. (1995). Role of aberrant crypt foci in understanding the pathogenesis of colon cancer. *Cancer Letters* **93**, 55–71.
- Bird, R. P. and Good, C. K. (2000). The significance of aberrant crypt foci in understanding the pathogenesis of colon cancer. *Toxicology Letters* **112/113**, 395–402.
- Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* **88**, 9–36.
- Carroll, R. J. and Pederson, S. (1993). On robustness in the logistic regression model. *Journal of the Royal Statistical Society, Series B* **55**, 693–706.
- Cliff, A. D. and Ord, J. K. (1981). *Spatial Processes: Models and Applications*. London: Pion.
- Commenges, D. and Jacqmin-Gadda, H. (1997). Generalized score test of homogeneity based on correlated random effects models. *Journal of the Royal Statistical Society, Series B* **59**, 157–171.
- Cressie, N. (1993). *Statistics for Spatial Data*. New York: Wiley.
- Diggle, P. J., Moyeed, R. A., and Tawn, J. A. (1997). Model-based geostatistics. *Applied Statistics* **47**, Part 3, 299–350.
- Heagerty, P. J. and Lele, S. R. (1998). A composite likelihood approach to binary spatial data. *Journal of the American Statistical Association* **93**, 1099–1123.
- Jacqmin-Gadda, H., Commenges, D., Nejjar, C., and Dartigues, J. F. (1997). Tests of geographical correlation with adjustment for explanatory variables: An application to dyspnoea in the elderly. *Statistics in Medicine* **16**, 1283–1297.
- le Cessie, S. and van Houwelingen, J. C. (1994). Logistic regression for correlated binary data. *Applied Statistics* **43**, 95–108.
- McLellan, E. A., Medline, A., and Bird, R. P. (1991). Sequential analyses of the growth and morphological characteristics of aberrant crypt foci: Putative preneoplastic lesions. *Cancer Research* **51**, 5270–5274.
- Moran, P. A. P. (1948). The interpretation of statistical maps. *Journal of the Royal Statistical Society, Series B* **10**, 54–62.
- Morris, J. S., Wang, N., Lupton, J. R., Chapkin, R. S., Turner, N. D., Hong, M. Y., and Carroll, R. J. (2001). Parametric and nonparametric methods for understanding the relationship between carcinogen-induced DNA adduct levels in distal and proximal regions of the colon. *Journal of the American Statistical Association* **96**, 816–826.
- Morris, J. S., Wang, N., Lupton, J. R., Chapkin, R. S., Turner, N. D., Hong, M. Y., and Carroll, R. J. (2002). A Bayesian analysis of colonic crypt structure and coordinated response incorporating missing crypts. *Biostatistics* **3**, 529–546.
- Morris, J. S., Vannucci, M., Brown, P. J., and Carroll, R. J. (2003). Wavelet-based nonparametric modeling of hierarchical functions in colon carcinogenesis. *Journal of the American Statistical Association*, to appear.
- Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*. New York: Wiley.
- Richardson, S., Guihenneuc, C., and Lasserre, V. (1992). Spatial linear models with autocorrelated error structure. *Statistician* **41**, 539–557.
- Simpson, D. G., Ruppert, D., and Carroll, R. J. (1992). One-step GM-estimates and stability of inferences in linear regression. *Journal of the American Statistical Association* **87**, 439–450.
- Stein, M. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. New York: Springer.

Received November 2002. Revised May 2003.

Accepted June 2003.

APPENDIX

A.1 Derivation of the Score Test Using the Matern Class

Recall that there are $k=1, \dots, N$ pairs of responses (Y_{1k}, Y_{2k}) , and that $Z_{ijk} = I(Y_{1k} = i, Y_{2k} = j)$. Note that $E(Z_{ijk}) = \pi_{ijk}$. Also, $\pi_{01k} = \pi_{\cdot 1k} - \pi_{11k}$, $\pi_{10k} = \pi_{1 \cdot k} - \pi_{11k}$, $\pi_{00k} = 1 - \pi_{01k} - \pi_{10k} - \pi_{11k} = 1 - \pi_{\cdot 1k} - \pi_{1 \cdot k} + \pi_{11k}$. In addition, it follows that $\partial \pi_{00k} / \partial \rho = \partial \pi_{11k} / \partial \rho$, $\partial \pi_{01k} / \partial \rho = -\partial \pi_{11k} / \partial \rho$ and $\partial \pi_{10k} / \partial \rho = -\partial \pi_{11k} / \partial \rho$. Then, the differentiation of the log likelihood, with respect to ρ , leads to

$$\frac{\partial \log L(\rho)}{\partial \rho} = \sum_k \left(\frac{Z_{00k}}{\pi_{00k}} - \frac{Z_{01k}}{\pi_{01k}} - \frac{Z_{10k}}{\pi_{10k}} + \frac{Z_{11k}}{\pi_{11k}} \right) \frac{\partial \pi_{11k}}{\partial \rho}.$$

The information can be shown to equal

$$E \left\{ -\frac{\partial^2 \log L(\rho)}{\partial \rho^2} \right\} = \sum_k \left(\frac{1}{\pi_{00k}} + \frac{1}{\pi_{01k}} + \frac{1}{\pi_{10k}} + \frac{1}{\pi_{11k}} \right) \times \left(\frac{\partial \pi_{11k}}{\partial \rho} \right)^2.$$

Recall that $\psi = \sigma_\lambda^2 / (1 + \sigma_\lambda^2) \Omega$. If $\pi_{11k} = \Phi_2\{\mu_{1k}^*, \mu_{2k}^*, \psi_k(\rho)\}$, then

$$\frac{\partial \pi_{11k}}{\partial \rho} = \frac{\partial \pi_{11k}}{\partial \psi_k} \frac{\partial \psi_k}{\partial \rho} = \phi_2\{\mu_{1k}^*, \mu_{2k}^*, \psi_k(\rho)\} \frac{\partial \psi_k}{\partial \rho},$$

where $\phi_2\{\mu_{1k}^*, \mu_{2k}^*, \psi_k(\rho)\}$ is a bivariate standard normal density function with correlation $\psi_k(\rho)$, evaluated at (μ_{1k}^*, μ_{2k}^*) . Therefore, the score is

$$\left(\frac{Z_{00k}}{\pi_{00k}} - \frac{Z_{01k}}{\pi_{01k}} - \frac{Z_{10k}}{\pi_{10k}} + \frac{Z_{11k}}{\pi_{11k}} \right) \phi_2\{\mu_{1k}^*, \mu_{2k}^*, \psi_k(\rho)\} \frac{\partial \psi_k}{\partial \rho}$$

When we use as the correlation function, $\Omega_k(\rho)$, a member of the Matern class with $\nu = m + \frac{1}{2}$, the correlation function is of the form $\exp(-d_k/\rho)$ times a polynomial in $-d_k/\rho$ degree m (Abramowitz and Stegun, 1965, Section 10.2.15). The derivative of such correlation function is of the form $\exp(-d_k/\rho)$ times a polynomial in $-d_k/\rho$ degree $m + 2$. Using L'Hopital's rule, one can prove that $\exp(-d_k/\rho)(-d_k/\rho)^l \rightarrow 0$ as $\rho \rightarrow 0$ for $l \geq 0$. Hence, $\partial\psi_k(\rho)/\partial\rho = \sigma_\lambda^2/(1 + \sigma_\lambda^2)\partial\Omega_k(\rho)/\partial\rho \rightarrow 0$ as $\rho \rightarrow 0$; as a result, the score evaluated at $\rho=0$ goes to 0 as well. The solution to that technical difficulty, we propose, is to focus on pairs that are exactly the same distance apart, in which case $\psi_k(\rho)$ is going to be the same for all k . We reparameterize $\psi_1(\rho) = \psi_2(\rho) = \dots = \psi_N(\rho) = \psi$. Notice that $\psi=0$ if and only if $\rho=0$. So, we can reformulate the null hypothesis of no spatial correlation in terms of a new parameter, $H_0 : \psi=0$. The new score that will be used in constructing the test is also recalculated in terms of a new parameter ψ

$$s_k(\mu_{1k}^*, \mu_{2k}^*, \psi) = \left(\frac{Z_{00k}}{\pi_{00k}} - \frac{Z_{01k}}{\pi_{01k}} - \frac{Z_{10k}}{\pi_{10k}} + \frac{Z_{11k}}{\pi_{11k}} \right) \phi_2\{\mu_{1k}^*, \mu_{2k}^*, \psi\} \tag{A.1}$$

This shows that the reparametrization allows us to eliminate the term $\partial\psi_k/\partial\rho$, which is equal to 0 when $\rho=0$. Notice that $\pi_{i,jk}|\psi=0 = (-1)^{i+j}(1-i-\pi_{1.k})(1-j-\pi_{1.k})$ and $\phi_2\{\mu_{1k}^*, \mu_{2k}^*, 0\} = \phi(\mu_{1k}^*)\phi(\mu_{2k}^*)$. Hence, the score evaluated at $\psi=0$

$$\begin{aligned} \mathcal{G}_k(\mu_{1k}^*, \mu_{2k}^*) &= s_k(\mu_{1k}^*, \mu_{2k}^*, 0) \\ &= \frac{(Y_{1k} - \pi_{1.k})(Y_{2k} - \pi_{1.k})\phi(\mu_{1k}^*)\phi(\mu_{2k}^*)}{\pi_{1.k}(1 - \pi_{1.k})\pi_{1.k}(1 - \pi_{1.k})}. \end{aligned}$$

The variance of $\mathcal{G}_k(\mu_{1k}^*, \mu_{2k}^*)$ under the null hypothesis is

$$\text{var}\{\mathcal{G}_k(\mu_{1k}^*, \mu_{2k}^*)\} = \frac{\{\phi(\mu_{1k}^*)\phi(\mu_{2k}^*)\}^2}{\pi_{1.k}(1 - \pi_{1.k})\pi_{1.k}(1 - \pi_{1.k})}.$$

Let the μ s depend on covariates X and a parameter β_* , i.e., $\mu(X, \beta_*)$ with the property that for any constant c , $c\mu(X, \beta_*) = \mu(X, \beta_{**})$ for some β_{**} . Recall that under the null hypothesis, the Y s are independent, and $\text{pr}(Y=1|X) = \Phi\{\mu(X, \beta_*)/(1 + \sigma_\lambda^2)^{1/2}\} = \Phi\{\mu(X, \beta)\}$. Then, the score can be written as

$$\mathcal{H}_k(\beta) = \frac{[Y_{1k} - \Phi\{\mu(X_{1k}, \beta)\}][Y_{2k} - \Phi\{\mu(X_{2k}, \beta)\}]\phi\{\mu(X_{1k}, \beta)\}\phi\{\mu(X_{2k}, \beta)\}}{\Phi\{\mu(X_{1k}, \beta)\}[1 - \Phi\{\mu(X_{1k}, \beta)\}]\Phi\{\mu(X_{2k}, \beta)\}[1 - \Phi\{\mu(X_{2k}, \beta)\}]}$$

The variance of $\mathcal{H}_k(\beta)$ under the hypothesis of no spatial correlation is

$$\mathcal{V}_k(\beta) = \frac{[\phi\{\mu(X_{1k}, \beta)\}\phi\{\mu(X_{2k}, \beta)\}]^2}{\Phi\{\mu(X_{1k}, \beta)\}[1 - \Phi\{\mu(X_{1k}, \beta)\}]\Phi\{\mu(X_{2k}, \beta)\}[1 - \Phi\{\mu(X_{2k}, \beta)\}]}$$

Then our test statistic is

$$\frac{\sum_k \mathcal{H}_k(\beta)}{\left\{ \sum_k \mathcal{V}_k(\beta) \right\}^{1/2}} \tag{A.2}$$

There is an important subtlety in (A.2), namely, that while not independent, under the null hypothesis, the terms $\mathcal{H}_k(\beta)$

are mutually uncorrelated, and hence (A.2) indeed has mean 0 and variance 1.

That (A.2) is asymptotically normally distributed, with mean zero and variance one under the null hypothesis of independence, follows from a result of Commenges and Jacqmin-Gadda (1997), under conditions that govern the behavior of the terms $\mu(X, \beta)$.

Of course, β is not known. The result of Commenges and Jacqmin-Gadda (1997) can be used to show that when a $n^{(1/2)}$ -consistent estimate $\hat{\beta}$ is substituted into the test statistic in place of the true value β , the limit distribution of the test statistic is unaffected. Because of page limits, we do not provide details, but the essential point is the orthogonality of the numerator of the test to β . In other words, it may be shown that, under the null hypothesis of independence,

$$E \left\{ \frac{\partial \mathcal{H}_k(\beta)}{\partial \beta} \right\} \Bigg|_{\psi=0} = 0.$$

Therefore, our test statistic

$$\frac{\sum_k \mathcal{H}_k(\hat{\beta})}{\left\{ \sum_k \mathcal{V}_k(\hat{\beta}) \right\}^{1/2}}$$

is asymptotically standard normal under the null hypothesis of independence.

A.2 Score Test for the Conditionally Autoregressive Model

We now show that our score test is also the score test for the conditionally autoregressive model, see Besag (1974) and Richardson et al. (1992). Let λ be the vector of the λ 's. For the Gaussian CAR model, $\lambda = \text{Normal}\{0, \sigma_\lambda^2 B(\rho)\}$, where $B(\rho) = (I - \rho C)^{-1}$, and where C is chosen to be a neighborhood matrix whose (i, j) th element is equal to 1 if region i and region j ($i \neq j$) are neighbors. Let us assume that, for the set of locations $(1, \dots, n)$, we observe a binary vector (Y_1, Y_2, \dots, Y_n) . Define the likelihood function

$$\begin{aligned} L(\rho) &= \text{pr}(Y_1 = y_1, Y_2 = y_2, \dots, Y_n = y_n) \\ &= \int \dots \int \phi\{x_1, x_2, \dots, x_n; I + \sigma_\lambda^2 B(\rho)\} dx_1, dx_2, \dots, dx_n, \end{aligned}$$

where $\phi(x_1, x_2, \dots, x_n; V)$ is the n -variate normal density, with mean 0 and covariance matrix V , and the integral with respect to x_i is from $-\infty$ to μ_i if $Y_i = 1$, and from μ_i to ∞ if $Y_i = 0$.

Let the elements of $B(\rho)$ be $B_{ij}(\rho)$. Make the change of variables $z_i = x_i/\{1 + \sigma_\lambda^2 B_{ii}(\rho)\}^{1/2}$, so that $L(\rho) = \text{pr}(Y_1 = y_1, Y_2 = y_2, \dots, Y_n = y_n) = \int \dots \int \phi\{z_1, z_2, \dots, z_n; \Sigma(\rho)\} dz_1, dz_2, \dots, dz_n$, where Σ has elements Σ_{ij} , with $\Sigma_{ii} = 1$ and $\Sigma_{ij}(\rho) = \sigma_\lambda^2 B_{ij}(\rho)/\{[1 + \sigma_\lambda^2 B_{ii}(\rho)]\{1 + \sigma_\lambda^2 B_{jj}(\rho)\}\}^{1/2}$, and the integral with respect to z_i is from $-\infty$ to μ_i^* if $Y_i = 1$, and from μ_i^* to ∞ if $Y_i = 0$, where $\mu_i^* = \mu_i/\{1 + \sigma_\lambda^2 B_{ii}(\rho)\}^{1/2}$. The

first derivative of the log likelihood, with respect to ρ , is

$$\begin{aligned} \frac{\partial \log L(\rho)}{\partial \rho} &= \frac{1}{L(\rho)} \frac{\partial L(\rho)}{\partial \rho} \\ &= \frac{1}{L(\rho)} \left(\sum_{i < j} \frac{\partial L(\rho)}{\partial \Sigma_{ij}} \frac{\partial \Sigma_{ij}}{\partial \rho} + \sum_i \frac{\partial L(\rho)}{\partial \mu_i^*} \frac{\partial \mu_i^*}{\partial \rho} \right) \\ &= \frac{1}{L(\rho)} \sum_{i < j} (2Y_i - 1)(2Y_j - 1) \\ &\quad \times \int \dots \int \phi\{z_1, \dots, \mu_i^*, \dots, \mu_j^*, \dots, z_n; \Sigma(\rho)\} \\ &\quad \times \prod_{k \neq i, j} dz_k \frac{\partial \Sigma_{ij}}{\partial \rho} \\ &\quad - \frac{1}{L(\rho)} \sum_i \frac{\partial L(\rho)}{\partial \mu_i^*} \frac{\mu_i}{2\{1 + \sigma_\lambda^2 B_{ii}(\rho)\}^{3/2}} \sigma_\lambda^2 \left(\frac{\partial B}{\partial \rho} \right)_{ii}. \end{aligned}$$

Now, note that when $\rho = 0$, $B = I$. Also, $\partial B(\rho) / \partial \rho|_{\rho=0} = C$, with diagonal elements equal to 0. This means that

$$\left. \frac{\partial \Sigma_{ij}}{\partial \rho} \right|_{\rho=0} = \frac{\sigma_\lambda^2 C_{ij}}{1 + \sigma_\lambda^2}.$$

Now note that $\Sigma(\rho = 0) = I$, the identity matrix. Using the previous results, it follows that

$$\begin{aligned} \left. \frac{1}{L(\rho)} \frac{\partial L(\rho)}{\partial \Sigma_{ij}} \right|_{\rho=0} &= \frac{(2Y_i - 1)(2Y_j - 1) \phi(\mu_i^*) \phi(\mu_j^*) \prod_{k \neq i, j} \text{pr}(Y_k = y_k)}{\prod_k \text{pr}(Y_k = y_k)} \\ &= \frac{(2Y_i - 1)(2Y_j - 1) \phi(\mu_i^*) \phi(\mu_j^*)}{\text{pr}(Y_i = y_i) \text{pr}(Y_j = y_j)} \\ &= \frac{(Y_i - \pi_i)(Y_i - \pi_i) \phi(\mu_i^*) \phi(\mu_j^*)}{\pi_i(1 - \pi_i) \pi_j(1 - \pi_j)}. \end{aligned}$$

Hence,

$$\begin{aligned} \left. \frac{\partial \log L(\rho)}{\partial \rho} \right|_{\rho=0} &= \sum_{i < j: C_{ij} \neq 0} \frac{(Y_i - \pi_i)(Y_i - \pi_i) \phi(\mu_i^*) \phi(\mu_j^*)}{\pi_i(1 - \pi_i) \pi_j(1 - \pi_j)} \\ &\quad \times \frac{\sigma_\lambda^2}{(1 + \sigma_\lambda^2)}. \end{aligned}$$

The common term $\sigma_\lambda^2 / (1 + \sigma_\lambda^2)$ can be omitted from the score test statistic, leading to our test, as claimed.