Local scoring rules for spatial processes

Philip Dawid 1 and Monica Musio 2

1 Statistical Laboratory, University of Cambridge
2 Department of Mathematics, University of Cagliari

Introduction

Maximum likelihood estimation of a spatial process can be computationally demanding because of the need to manipulate the normalisation constant of the joint distribution. Besag introduced the method of pseudo-likelihood to sidestep this problem. Although this has traditionally been considered as an approximation to the full likelihood, the method can be justified in its own right, as leading to an unbiased estimating equation. Other methods, constructed from proper scoring rules, have similar justification and properties, and supply useful alternatives.

Proper Scoring Rules

A scoring rule is a loss function \( S(x, Q) \) measuring the quality of a quoted probability distribution \( Q \) for a random variable \( X \) in the light of the realised outcome \( x \) of \( X \). \( S \) is proper if the expected score \( E(Q) := \mathbb{E}(x \to S(X, Q)) \) is minimised by quoting \( Q = P \). An example is the log score, \( -\log q(x) \), where \( q \) is the density of \( Q \).


evaluated

Given any proper scoring rule \( S \) and a smooth parametric statistical model \( P = \{P_\theta\} \) for \( X \), let

\[
s(x, \theta) = \frac{\partial S(x, P_\theta)}{\partial \theta}.
\]

Then we can estimate \( \theta \) by \( \hat{\theta}_Q \), the root of the estimating equation

\[
s(x, \hat{\theta}_Q) = 0.
\]

When \( S \) is the log score, (1) is the likelihood equation, and \( \hat{\theta}_Q \) is the maximum likelihood estimate. More generally, for any differentiable scoring rule and any smooth statistical model, (1) is an unbiased estimating equation [6]. In particular it will typically deliver a consistent estimator in repeated sampling. We can then choose \( S \) to increase robustness or ease of computation.

Spatial process In the context of a spatial process \( X = \{X_v : v \in V\} \), we can define a useful class of proper scoring rules [7] by

\[
S(x, Q) = \sum_v S(x_v, Q_v),
\]

where \( Q_v \) is the conditional distribution of \( X_v \) given the values \( x_{\neq v} \) for the variables \( X_{\neq v} \) at all sites other than \( v \), and \( S_v \) is a proper scoring rule for the state at a single site. In particular, if \( Q \) is Markov on a graph \( G \), then \( Q_v \) only depends on the values \( x_{v(a)} \) at the sites neighbouring \( v \). This avoids the need to evaluate the normalising constant of the full joint distribution \( Q \).

Corresponding to (2) we have estimating equation

\[
\sum_v s(x_v, P_{\hat{\theta}_Q}) = 0
\]

with each term in the sum having expectation 0. When \( S_0 \) is the log score, (3) gives the (negative log) pseudo-likelihood [3]. For \( X_v \) binary and \( S_0 \) the quadratic (Brier) score, it yields the method of ratio matching [8].

Missing data Missing data are readily dealt with. Let \( A_v = 0 \) if any value at \( \{x\} \) or any of its neighbours is missing, else 1. Then so long as the data are missing completely at random, \( s(x_v, P_\theta) \times A_v \) has expectation 0, so we can just omit incomplete terms from (3) while retaining an unbiased estimating equation.

Example: Phytophthora data

The pathogen Phytophthora capsici Leonian causes lesions on the crown, stem, and leaves of bell pepper, and rapidly causes the plant to die. Figure 1 refers to the presence or absence of such disease in pepper plants in a regular 20 \( \times \) 20 grid [4].

![Pepper data](image)

**Fig. 1:** Presence (1) and absence (0) of pathogen in bell pepper plants

We model the data as a stationary first-order Markov process with respect to the grid, which thus follows the autologistic model [1, 2]:

\[
\logit \pi_{x_v} = \alpha + \beta(x_{v-1,j} + x_{v+1,i}) + \gamma(x_{v-1} + x_{v+1})
\]

where \( x_v = 1 \) if the disease is present and \( x_v = 0 \) if absent.

Results

Table 1 displays the results of fitting the model (4) by pseudo-likelihood (PL) and by ratio matching (RM).

<table>
<thead>
<tr>
<th>Method</th>
<th>Intercept, ( \alpha )</th>
<th>WE, ( \beta )</th>
<th>NS, ( \gamma )</th>
</tr>
</thead>
<tbody>
<tr>
<td>PL</td>
<td>-2.4396</td>
<td>1.0644</td>
<td>0.6266</td>
</tr>
<tr>
<td>RM</td>
<td>-2.3524</td>
<td>1.5864</td>
<td>0.5375</td>
</tr>
</tbody>
</table>

**Tab. 1:** Coefficients estimated by pseudo-likelihood (PL) and ratio matching (RM)

Discussion

The pseudo-likelihood and ratio matching methods, as well as others derived from different proper scoring rules, all involve solving an unbiased estimating equation. In the example studied, the estimates from PL and RM are broadly in line. Further theoretical and experimental work is needed to explore and compare their accuracy, efficiency and robustness properties.

References