Local scoring rules for spatial processes

Philip Dawid
University of Cambridge <apd@statslab.cam.ac.uk>

Monica Musio
Università di Cagliari <mmusio@unica.it>

Abstract: We display pseudo-likelihood as a special case of a general estimation technique based on proper scoring rules. Such a rule supplies an unbiased estimating equation for any statistical model, and this can be extended to allow for missing data. When the scoring rule has a simple local structure, as in many spatial models, the need to compute problematic normalising constants is avoided. We illustrate the approach through an analysis of data on disease in bell pepper plants.

Keywords: proper scoring rule, pseudo-likelihood, ratio matching, unbiased estimating equation

1 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 (1975) developed the method of *pseudo-likelihood* to sidestep this problem. This has traditionally been considered as an approximation (of unknown quality) to the full likelihood. However, as we describe below, 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.

2 Proper scoring rules

A *scoring rule* $S(x, Q)$ is a loss function measuring the quality of a quoted probability distribution Q for a random variable X, in the light of the realised outcome x of X — see *e.g.* Dawid (1986). It is *proper* if, for any distribution P for X, the expected score $S(P, Q) := \mathrm{E}_{X\sim P} S(X, Q)$ is minimised by quoting $Q = P$. A prominent example is the *log score*, $-\log q(x)$, where q denotes the density or probability mass function of X.

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

$$s(x, \theta) := \frac{\partial S(x, P_\theta)}{\partial \theta}.$$

Then we can estimate θ by $\widehat{\theta}_S$, the root of the *estimating equation*

$$s(x, \theta) = 0. \tag{1}$$

When S is the log score, this is just the likelihood equation, and $\widehat{\theta}_S$ is the maximum likelihood estimate. More generally, for any differentiable scoring rule and any smooth statistical model, $\mathrm{E}_\theta\{s(X, \theta)\} = 0$, *i.e.* (1) is an unbiased estimating equation (Dawid and Lauritzen 2005). In particular it will typically deliver a consistent, if not necessarily efficient, estimator in repeated sampling. We can then choose S to increase robustness or ease of computation.

In the context of a spatial process $X = (X_v : v \in V)$, we can define a useful class of proper scoring rules (Dawid *et al.* 2011) by

$$S(x, Q) = \sum_v S_0(x_v, Q_v), \tag{2}$$

where Q_v is the conditional distribution of X_v, given the values $x_{\setminus v}$ for the variables $X_{\setminus v}$ at all sites other than V, and S_0 is a proper scoring rule for the state at a single site. In particular, if Q is Markov on a graph $\mathcal{G}$, then Q_v only depends on the values $x_{\mathrm{ne}(v)}$ 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

$$\sum_v s_0(x_v, P_{\theta,v}) = 0 \tag{3}$$

with each term in the sum having expectation 0. When S_0 is the log score, (3) gives the (negative log) *pseudo-likelihood* (Besag 1975). For X_v binary and S_0 the quadratic ("Brier") score, it yields the method of *ratio matching* (Hyvärinen 2007).

Missing data are readily dealt with (although with some loss of efficiency). Let $A_v = 1$ if any value in $\{v\} \cup \mathrm{ne}(v)$ is missing. Then so long as the data are missing completely at random, $s_0(x_v, P_{\theta,v}) \times A_v$ has expectation 0, so we can just omit incomplete terms from (3) while retaining an unbiased estimating equation.

3 Phytophthora data

Figure 1 displays the presence or absence of the pathogen *Phytophthora capsici* Leonian in bell pepper plants on a regular 20×20 grid (Chadoeuf *et al.* 1992).

We model the data as a stationary first-order Markov process with respect to the grid, which thus follows the autologistic model (Besag 1972; Besag 1974; Gumpertz *et al.* 1997):

$$\operatorname{logit} \pi_{ij} = \alpha + \beta(x_{i-1,j} + x_{i+1,j}) + \gamma(x_{i,j-1} + x_{i,j+1}) \tag{4}$$

where π_{ij} is the probability of $X_{ij} = 1$, given all other values.

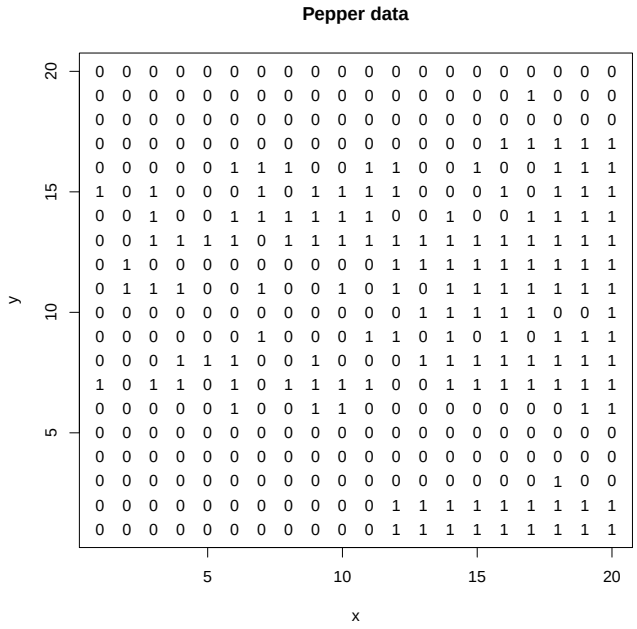

Figure 1: Presence (1) and absence (0) of pathogen in bell pepper plants

To fit by maximum pseudo-likelihood (PL), we simply proceed as if the (X_{ij}) were all independent, and maximise the resulting "likelihood". This can be done by a standard generalized linear model analysis, readily implemented in standard software such as `R`, using the binomial family and (default) logit link function.

Alternatively, and possibly more robustly, we could apply ratio matching (RM), based on the Brier scoring rule, which leads to the least-squares recipe: minimise $\sum(x_{ij} - \pi_{ij})^2$. Again this can be implemented in standard GLM software, treating the data as if they were normal with constant variance, and using the logit link function (in `R` this is effected using the `glm()` command with option `family=quasi(link=logit,variance=constant).`)

Note however that, although it is easy to compute the estimates, the associated "standard errors" output by the software will be inappropriate, since they do not take account of the dependence in the data.

4 Results

Table 1 displays the results of fitting the model (4) by pseudo-likelihood (PS) and by ratio matching (RM). Values at sites on the boundary of the grid, which do not have four observed neighbours, are not used as responses, though they are used as covariate values for their neighbouring interior sites. There are thus $18 \times 18 = 324$ data-points used to fit the model.

Method	Intercept, α	WE, β	NS, γ
PL	-2.4390	1.6514	0.6266
RM	-2.2654	1.5864	0.5375

Table 1: Coefficients estimated by pseudo-likelihood (PS) and ratio matching (RM)

5 Concluding remarks

The PL and RM 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. However further theoretical and experimental work is needed to explore and compare their accuracy, efficiency and robustness properties.

References

Besag, J. E. (1972). Nearest-neighbour systems and the auto-logistic model for binary data. *Journal of the Royal Statistical Society. Series B (Methodological)*, **34**, 75–83.

Besag, J. E. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, **36**, (2), 192–236.

Besag, J. E. (1975). Statistical analysis of non-lattice data. *Journal of the Royal Statistical Society. Series D (The Statistician)*, **24**, 179–95.

Chadoeuf, J., Nandris, D., Geiger, J., Nicole, M., and Pierrat, J. (1992). Modelisation spatio-temporelle d'une epidémie par un processus de Gibbs: Estimation et tests. *Biometrics*, **48**, 1165–75.

Dawid, A. P. (1986). Probability forecasting. In *Encyclopedia of Statistical Sciences*, (ed. S. Kotz, N. L. Johnson, and C. B. Read), pp. 210–8. Wiley-Interscience.

Dawid, A. P. and Lauritzen, S. L. (2005). The geometry of decision theory. In *Proceedings of the Second International Symposium on Information Geometry and its Applications*, pp. 22–8. University of Tokyo.

Dawid, A. P., Lauritzen, S. L., and Parry, M. (2011). Proper scoring rules on discrete sample spaces. `arXiv:1104.2224`.

Gumpertz, M. L., Graham, J. M., and Ristaino, J. B. (1997). Autologistic model of spatial pattern of phytophthora epidemic in bell pepper: Effects of soil variables on disease presence. *Journal of Agricultural, Biological, and Environmental Statistics*, **2**, 131–56.

Hyvärinen, A. (2007). Some extensions of score matching. *Computational Statistics and Data Analysis*, **51**, 2499–512.