Stochastic Approximation of the MLE for a Spatial Point Pattern

R. A. MOYEED and A. J. BADDELEY

Centre for Mathematics and Computer Science, Amsterdam

ABSTRACT. An iterative stochastic approximation to the maximum likelihood estimate is developed for the Strauss point process. We modify existing theorems to show that the approximation is consistent and asymptotically normal. It performs well in numerical tests.

Key words: Markov point processes, Monte Carlo inference, Robbins–Monro method, Strauss model

1. Introduction

Parameter estimation for two-dimensional point pattern data is difficult, because most of the available stochastic models have intractable likelihoods (see Ripley, 1977; 1988 & Diggle, 1983). An exception is the class of Gibbs or Markov point processes (Ripley & Kelly, 1977; Baddeley & Møller, 1989; Ripley, 1989), where the likelihood $f(x; \theta)$ typically forms an exponential family and is given explicitly up to a normalising constant $\alpha(\theta)$. However, the latter is not known analytically, so parameter estimates must be based on approximations (Ogata & Tanemura, 1984; Penttinen, 1984; Diggle & Gratton, 1984; Ripley, 1988).

In this paper we present a simple recursive approximation to the MLE $\hat{\theta}$, based on the Robbins–Monro stochastic approximation method. Estimates θ_k are updated by comparing the observed value of a sufficient statistic $S = s(X)$ with a random value simulated from the distribution specified by θ_k:

$$\theta_{k+1} = \theta_k + \frac{A}{k}(s(x) - s(X_k))$$

where x is the data and X_k is one simulated observation from $f(\cdot\,; \theta_k)$. Existing theorems (Hall & Heyde, 1980) can be modified to show θ_k is consistent and asymptotically normal (as an approximation to $\hat{\theta}$).

We use a very simple point process model, the Strauss process, to illustrate and test the method. Many other techniques are available for this case (see section 2) but we believe our method could be applied to complex models where the others cannot. The aim of the present paper is just to demonstrate that the Robbins–Monro method works in this simple case, and to assess its performance.

The next section gives some background details about the point process setting and previous work in the area. Section 3 studies the maximum likelihood estimator for the Strauss model, with a numerical example. Section 4 recalls existing theorems for the Robbins–Monro stochastic approximation method. Section 5 adapts them to our setting and gives a numerical example. Section 6 ends with some discussion.

2. Background

Assume we have observed a pattern of points $x = \{x_1, \ldots, x_n\}$ in a planar window W. In principle n is random. A *pairwise interaction process* is a stochastic model for x with density

$$f(x; \theta) = a(\theta) \prod_i b(x_i; \theta) \prod_{i<j} c(x_i, x_j; \theta) \tag{1}$$

where θ is the parameter vector, b and c are nonnegative functions, and a is the normalising constant. Formally $f(\cdot\,;\theta)$ is the Radon–Nikodym density of the distribution of x with respect to the distribution of a Poisson point process π of intensity 1 (say) in W. Alternatively, for each fixed n, equation (1) gives the joint probability density of the n points $x_1, \ldots, x_n$ up to a constant factor $e^{-w}w^n/n!$, where $w = \mathrm{area}(W)$.

A simple example is the Strauss process with parameter $\theta = (\beta, \gamma)$, in which $b(\cdot\,;\cdot) \equiv \beta$ and

$$c(x_i, x_j;\theta) = \begin{cases} \gamma & \text{if } \|x_i - x_j\| < r \\ 1 & \text{otherwise} \end{cases}$$

where the "interaction distance" $r > 0$ is assumed to be fixed and known. This model has density

$$f(x\,;\beta, \gamma) = a(\beta, \gamma)\beta^{n(x)}\gamma^{s(x)}, \quad \beta > 0, \quad 0 \le \gamma \le 1, \tag{2}$$

(taking $0^0 = 1$) where $n(x) = n$ is the number of points in x, $s(x)$ is the number of distinct pairs x_i, x_j with $\|x_i - x_j\| < r$, and again $a(\beta, \gamma)$ is the normalising constant.

The parameter γ represents interaction between points, with $\gamma = 1$ corresponding to a Poisson process (no interaction) and $\gamma = 0$ giving a so-called "hard core" process in which no pair of points can come closer than distance r. Parameter β in (2) varies the point intensity, e.g. if $\gamma = 1$ then the process is a Poisson process with intensity β. For further information see Kelly & Ripley (1976), Ripley (1981, 1988), Diggle (1983) and Baddeley & Møller (1989).

A numerical approximation to the MLE for some pairwise interaction models was developed by Ogata & Tanemura (1984), who approximated $Z(\theta) = a(\theta)^{-1}$ analytically using virial expansions. However, the approximation is accurate only for sparse patterns.

The pseudo-likelihood estimator (Besag, 1977) could also be regarded as an analytic approximation to the MLE, and would generally be a good approximation (Ripley, 1988).

Monte Carlo techniques have also been suggested. Numerical estimates of $Z(\theta)$ can easily be formed for all θ by extensive simulations of the reference process only (or the process with some fixed parameter value θ_0). This would be computationally very expensive, and in practice cannot be performed "before" the data analysis because of its dependence on other parameters such as the geometry of the sampling window and the number of observed points.

Penttinen (1984) suggested solving the ML equation $\partial \log f(x;\theta)/\partial\theta = 0$ by Newton–Raphson iteration, after estimating the mean and mean derivative of the efficient score by simulation. When an approximate $\hat{\theta}$ is obtained, further simulations using this parameter value are used to refine $\hat{\theta}$ itself and to estimate its variance.

Diggle & Gratton (1984) advocated Monte Carlo parameter estimation by matching some chosen characteristic of the pattern, such as its reduced second-moment cumulative function $K(t)$, against the K function of the model, as evaluated through simulation. An iterative optimization technique is used to find a value of θ minimising (typically) the sum of squared deviations between the K functions of data and (simulated) model. The choice of K is appropriate because $K(r) \approx (w/n^2)\mathrm{E}[s(X)]$ ignoring edge effects.

3. MLE for the conditional Strauss model

In this paper we consider the very simple Strauss model (2) and further simplify it by conditioning on the observed number of points $n(x) = n$. We condition on n for simplicity in the analysis of the Robbins–Monro method, and because it is easy to generate simulations with fixed n by the discrete-time Markov chain method of Ripley (1979, 1987). The

conditional model has density

$$f_n(x; \gamma) = \frac{\gamma^{s(x)}}{Z_n(\gamma)}, \quad 0 < \gamma \leqslant 1 \tag{3}$$

with respect to the binomial process (n i.i.d. uniform random points in W). The case $\gamma = 0$ is examined separately below. Here $Z_n(\gamma)$ is the normalising constant, and $s(x)$ the number of pairs of distinct points closer than r units. The density (3) can be extended to $\gamma > 1$ which would correspond to a positive "clustering" of points. However, the unconditional model (2) is not integrable for $\gamma > 1$, so (3) could not be described as the "conditional" model with respect to (2). It will sometimes be convenient to allow $\gamma > 1$.

Clearly $S = s(X)$ is sufficient for γ, and has distribution

$$P_\gamma\{S = k\} = \frac{p_{nk}\gamma^k}{Z_n(\gamma)}, \quad k = 0, 1, \ldots \tag{4}$$

where $p_{nk} = P^{(n)}\{s(X) = k\}$ is the distribution of $s(X)$ for a binomial process of n points. This is an exponential family of discrete distributions, the only complication being that the reference distribution $\{p_{nk}, k = 0, 1, \ldots\}$ is not known analytically. A Poisson approximation to $\{p_{nk}\}$ is accurate when the data are sparse, i.e. when $n\pi r^2/w$ is small, see Ripley (1988). However, this approximation is poor in cases of real interest.

The normalising constant in (3), (4) is

$$Z_n(\gamma) = E^{(n)}[\gamma^{s(X)}]$$

$$= \sum_{k=0}^{m} p_{nk}\gamma^k \tag{5}$$

where $E^{(n)}[\ \cdot\ ]$ is expectation with respect to the binomial process, and $m = n(n - 1)/2$.

In principle, we could compute the likelihood by estimating $\{p_{nk}\}$ from extensive simulations of a binomial process, and calculating (4) for all values of γ. Equivalently, we could estimate (5) directly as the empirical probability generating function of $s(X)$. However, p_{n0} is typically close to zero, so if γ is small then both numerator and denominator of (4) are quite small and the simulation variance will be quite high. For small γ it would be better to estimate $p_{nk}(\gamma) = P_\gamma\{S = k\}$ using

$$p_{nk}(\gamma) = \frac{\left(\dfrac{\gamma}{\gamma_0}\right)^k p_{nk}(\gamma_0)}{\displaystyle\sum_{l=0}^{m} \left(\dfrac{\gamma}{\gamma_0}\right)^l p_{nl}(\gamma_0)} \tag{6}$$

for a reference value γ_0 that is close to γ. Similar problems were noted by Penttinen (1984).

A further complication is that when $\gamma = 0$ the conditional model in the form (3), (4) may be undefined. This occurs when $p_{n0} = 0$, i.e. when it is impossible (ignoring configurations of measure zero) to place n points $\{x_i\}$ in the window W with $\|x_i - x_j\| > r$ for all $i \neq j$, or equivalently, to place n discs of radius $r/2$ with centres in W and without overlap. If W is convex, a sufficient condition for this to occur is $n\pi r^2/4 > \text{area}(W) + r \text{ length}(\partial W) + \pi r^2$.

Let K be the minimum achievable value of $s(X)$:

$$K = \min\{k : p_{nk} > 0\}. \tag{7}$$

Then the limit as $\gamma \to 0$ of the Strauss conditional density (3) is

$$f_n(x; 0+) = \begin{cases} 1/p_{nK} & \text{if } s(x) = K \\ 0 & \text{else} \end{cases} \tag{8}$$

We henceforth use this distribution in place of (4) when $\gamma = 0$. The corresponding limit distribution of $s(X)$ is degenerate, $P\{s(X) = K\} = 1$.

Now consider maximum likelihood estimation of γ. Since γ is confined to $[0, 1]$ the maximum is attained either at $\gamma = 0$, $\gamma = 1$ or at a zero of the derivative of f. For the latter we have the usual ML estimating equations

$$s(x) - E_\gamma[s(X)] = 0 \tag{9}$$

where x is the observed point pattern, and the expectation is with respect to (4). Equivalently

$$M(\hat{\gamma}) = s(x) \tag{10}$$

where

$$M(\gamma) = E_\gamma[s(X)] = \gamma \frac{d \log Z_n(\gamma)}{d\gamma}. \tag{11}$$

From (11) and (5)

$$M(\gamma) = \frac{\gamma Z'_n(\gamma)}{Z_n(\gamma)} = \frac{\sum\limits_{k=0}^{m} k p_{nk} \gamma^k}{\sum\limits_{k=0}^{m} p_{nk} \gamma^k} \tag{12}$$

so that $M(0+) = K$. Also M is infinitely differentiable with

$$M'(\gamma) = \frac{Z_n Z'_n + \gamma Z_n Z''_n - \gamma(Z'_n)^2}{Z_n^2} \tag{13}$$

and in particular $M'(0+) = p_{n, K+1}/p_{nK}$.

Clearly $s(X)$ is a minimum variance bound estimator of $M(\gamma)$, and for $\gamma > 0$

$$\mathrm{Var}_\gamma[s(X)] = \gamma M'(\gamma) \tag{14}$$

where the subscript γ again means variance with respect to the distribution (4). This variance is nonzero for $\gamma > 0$, so

$$M'(\gamma) > 0, \quad 0 < \gamma \leqslant 1; \tag{15}$$

so $M(\gamma)$ is an increasing continuous function, and (10) has a unique solution for $K \leqslant s(x) \leqslant M(1)$. Note that $M(1) = E^{(n)}(s(X)) < m = n(n-1)/2$ and there is no solution to (10) for $s(x) > M(1)$.

Thus

$$\hat{\gamma} = \begin{cases} 0 & \text{if } s(x) = K \\ \text{solution of (10)} & \text{if } K < s(x) < M(1) \\ 1 & \text{if } s(x) \geqslant M(1) \end{cases} \tag{16}$$

and $\hat{\gamma}$ is otherwise undefined. This is the MLE for the parameter space $[0, 1]$; if we allow $\gamma > 1$ then $\hat{\gamma}$ is the solution of (10) for all $s(x) > K$.

To visualize the degree of difficulty in solving (10), Fig. 1 shows estimates of $M(\gamma)$, $M'(\gamma)$ and $\mathrm{Var}_\gamma[s(x)]$ obtained by extensive simulation. The window W was the unit square. We simulated 10^4 realizations of Strauss processes with $n = 25$ points, $r = 0.15$ and $\gamma_0 = 0.001$, 0.1, 0.25, 0.5 and 1. The distribution $\{p_{nk}(\gamma)\}$ of $s(X)$ at intermediate values of γ was obtained by extrapolation using (6). The value of K was found to be 0. The mean and variance of this distribution were then computed, giving $M(\gamma)$ and $\mathrm{Var}_\gamma[s(x)]$, and $M'(\gamma)$ was obtained from (14).
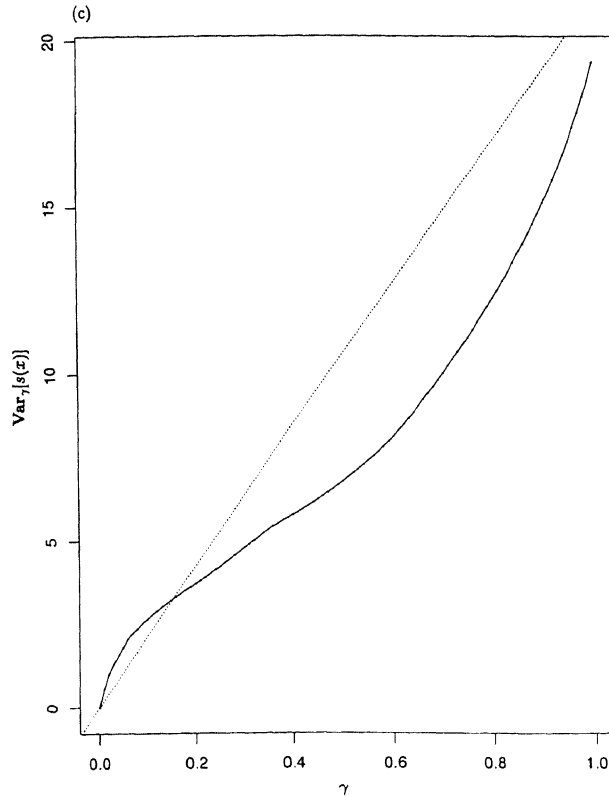
(a)

(b)

(c)
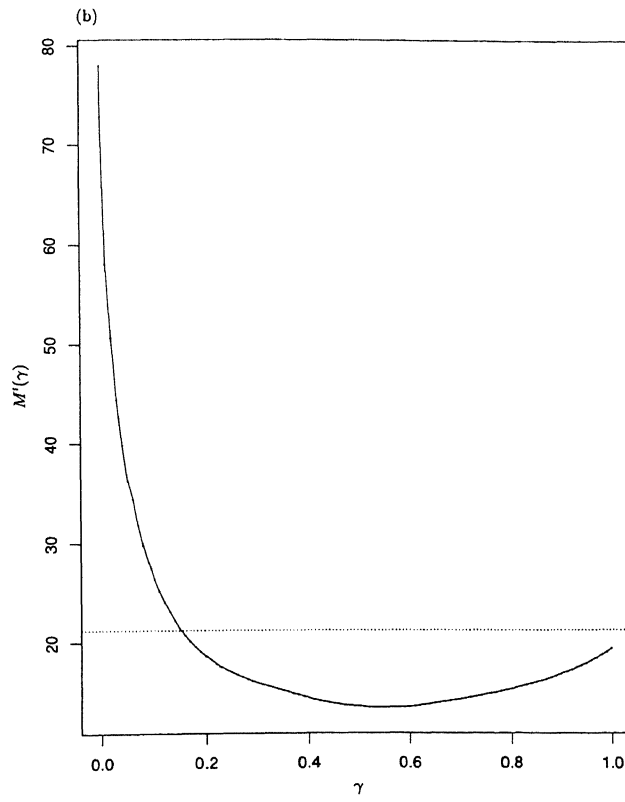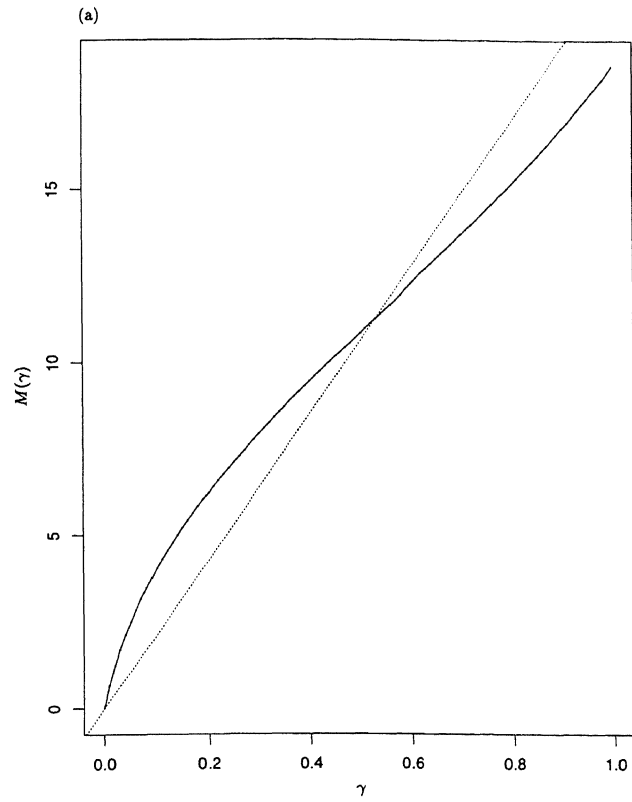
Fig. 1. Plots of (a) $M(\gamma)$; (b) $M'(\gamma)$; (c) $\mathrm{Var}_\gamma[s(x)]$ against γ, obtained through equation (6) by simulating from Strauss processes with $\gamma_0 = 0.001$, 0.1, 0.25, 0.5, 1 in a unit square with $n = 25$ and $r = 0.15$. Dotted line in each graph shows the Poisson approximation.

We expect $M(\gamma)$ to be approximately linear in γ whenever the Poisson approximation holds good, since if $\{p_{nk}\}$ were Poisson(λ) we would have $Z_n(\gamma) = \exp(\lambda(\gamma - 1))$ and $M(\gamma) = \gamma(d \log Z_n / d\gamma) = \lambda\gamma$. Under the sparse approximation $\lambda = n(n - 1)\pi r^2/(2 \text{ area}(W))$ and this approximation is plotted in Fig. 1.

4. Robbins–Monro procedure

The stochastic approximation procedure introduced by Robbins and Monro (1951) can be used to estimate the solution θ^* of an equation

$$F(\theta^*) = \phi$$

when there is very little information about the function F but where it is possible for any given θ to generate a random variable T_θ with expectation $\mathbf{E}(T_\theta) = F(\theta)$. The original application was to find the 50% survival probability point of a dose–response curve. For a recent survey see Dupač (1984).

The R–M procedure generates a sequence of estimates θ_k, $k = 1, 2, \ldots$ by

$$\theta_{k+1} = \theta_k + a_k\{\phi - T_{\theta_k}\}, \tag{17}$$

where $\{a_k\}$ is a decreasing sequence of positive numbers such that

$$\sum a_k = \infty, \quad \sum a_k^2 < \infty$$

and T_{θ_k}, $k = 1, 2, \ldots$ are independent random variables generated according to some parametric family of distributions $\{P_\theta : \theta \in I\}$ such that $\mathbf{E}(T_{\theta_k}) = F(\theta_k)$.

Following is a simplified list of regularity conditions on F condensed from Hall & Heyde (1980) and Nevel'son & Has'minskiĭ (1976):

(M1) F is a Borel-measurable function; $F(\theta^*) = \phi$ and

$$(\theta - \theta^*)(F(\theta) - \phi) > 0 \quad \forall \theta \neq \theta^*, \quad \theta \in I.$$

(M2) For some positive constants K_1 and K_2, and for all θ

$$K_1|\theta - \theta^*| \leq |F(\theta) - \phi| \leq K_2|\theta - \theta^*|.$$

(M3) $\sup_{\theta \in I} \mathbf{Var}[T_\theta] < \infty$.

(M4) F is differentiable at θ^* and $F'(\theta^*) > 0$.

(M5) For some $\eta > 0$

$$\sup_{\theta \in I} \mathbf{E}|T_\theta - F(\theta)|^{2+\eta} < \infty.$$

(M6) $\mathbf{Var}[T_\theta]$ is a continuous function of θ, at least at $\theta = \theta^*$.

It can be shown (see Hall & Heyde, 1980, p. 239) that under (M1)–(M3),

$$\theta_k \to \theta^* \quad \text{(a.s.)} \tag{18}$$

In the special case

$$a_k = A/k, \tag{19}$$

under conditions (M1)–(M6) we have

$$k^{1/2}(\theta_k - \theta^*) \to N\left(0, \frac{A^2\sigma^2}{2A\mu - 1}\right) \tag{20}$$

in distribution, assuming the constant A satisfies

$$A > \frac{1}{2\mu} \tag{21}$$

where

$$\mu = F'(\theta^*), \quad \sigma^2 = \mathbf{Var}[T_{\theta^*}]. \tag{22}$$

It is apparent from (20) that $A_{\mathrm{opt}} = 1/\mu$ is the value of A that minimizes the asymptotic variance of the normed R–M sequence $k^{1/2}(\theta_k - \theta^*)$; the minimal asymptotic variance being σ^2/μ^2.

In the case where θ is known to lie in a bounded open interval $I = (r_1, r_2)$, the algorithm is modified to a clipped version

$$\theta_{k+1} = [\theta_k + a_k\{\phi - T_{\theta_k}\}]_{r_1}^{r_2}, \tag{23}$$

where

$$[z]_{r_1}^{r_2} = \begin{cases} r_1 & \text{if } z \leqslant r_1 \\ z & \text{if } r_1 < z < r_2, \\ r_2 & \text{if } z \geqslant r_2. \end{cases}$$

One can see from Nevel'son & Has'minskii (1976, pp. 161–169) that the same limit results (18) and (20) hold with (M2) replaced by

(M2') $F(\theta)$ is bounded for $\theta \in I$.

5. The method

Returning to the conditional Strauss process model (3), our objective is to solve for $\hat{\gamma}$ in

$$M(\hat{\gamma}) = s(x),$$

where x is the observed data and $M(\gamma) = \mathrm{E}_\gamma s(X)$. This is an ideal application of the Robbins–Monro method since for any given γ we are able to generate a random variable T_γ with expectation $M(\gamma)$ by simulating a Strauss process X_γ with parameter γ and setting $T_\gamma = s(X_\gamma)$. The simulation is performed by running a birth-and-death process (Ripley, 1977, 1979; Baddeley & Møller, 1989) long enough to approach equilibrium.

Proposition

Suppose x is an observed realization of a conditional Strauss process (3). Let $\gamma_0 \in (0, 1]$ be arbitrary. Recursively form a sequence of estimates of $\hat{\gamma}$ by

$$\gamma_{k+1} = \left[\gamma_k + \frac{A}{k} \{s(x) - s(X_{\gamma_k})\} \right]_0^1, \tag{24}$$

where X_γ is a random point pattern simulated from $f_n(\cdot\,;\gamma)$.

*Then $\gamma_k \to \hat{\gamma}$ (a.s.). Defining $\mu = M'(\hat{\gamma})$ and $\sigma^2 = Var_{\hat{\gamma}}[s(X)]$, if $A > 1/(2\mu)$ then γ_k is asymptotically normally distributed with mean $\hat{\gamma}$ and variance $A^2\sigma^2/(2A\mu - 1)$.*

Proof. We apply the results of section 4 with θ, θ^*, $F(\theta)$, ϕ and T_θ replaced by γ, $\hat{\gamma}$, $M(\gamma)$, $s(x)$ and $s(X_\gamma)$ respectively. Condition (M1) clearly holds since $M(\gamma)$ is a monotone increasing function by (15). Conditions (M2'), (M3) and (M5) are all implied by the boundedness of $T_\theta = s(x) \leqslant m$. Next (M4) is implied by the fact that $M'(\gamma) > 0$ for all $\gamma \in [0, 1]$ acccording to (15). Condition (M6) follows from the continuity of $\gamma M'(\gamma) = Var_\gamma[s(X)]$. Hence (M1)–(M6) hold, and we may apply the clipped version of the R–M procedure.   □

As a numerical example, Fig. 2(a) shows a simulated realization of a Strauss process having 25 points in a unit square with $r = 0.15$ and $\gamma = 0.1$. In this case $s(x) = 4$ and this is indicated in the figure by lines joining the pairs of points which are neighbours. An accurate value of the MLE is $\hat{\gamma} = 0.099$, obtained from the exhaustive simulations described in section 3.

Figures 2(b) and 2(c) show typical runs of the algorithm with $A = 1$ and $A = A_{opt} = 1/M'(\hat{\gamma}) = 0.04$ respectively. As expected, variability is much reduced when A_{opt} is used. The theoretical asymptotic standard errors of γ_k were $0.23/\sqrt{k}$ and $0.06/\sqrt{k}$ respectively. Sample fluctuations of the γ_k sequence die down much more rapidly in the optimal case, at roughly the rate suggested by the ratio of asymptotic variances $\mu^2/(2\mu - 1) = 0.23^2/0.06^2 = 14.7$. Similar results were obtained with other values of n, r and $s(x)$.
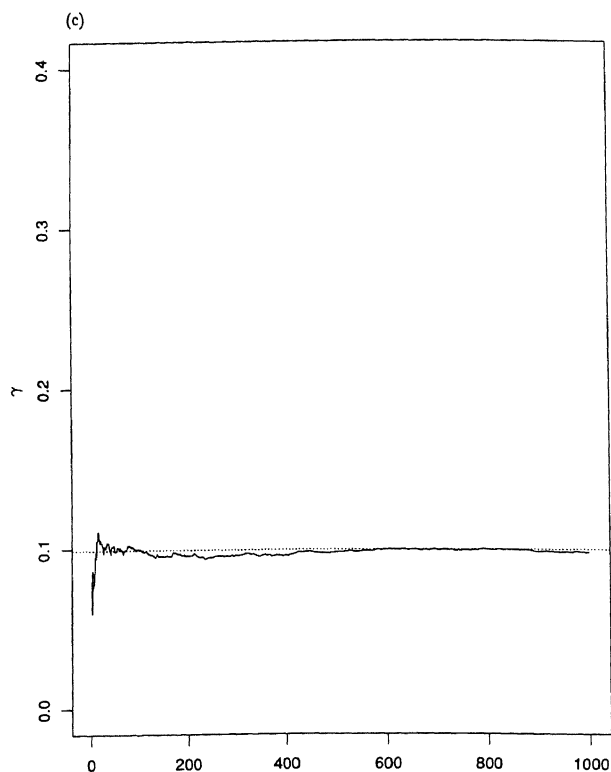
(a)
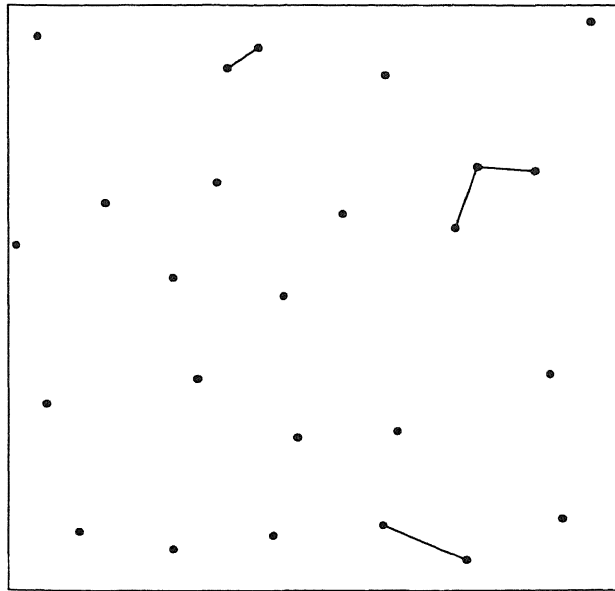
(b)
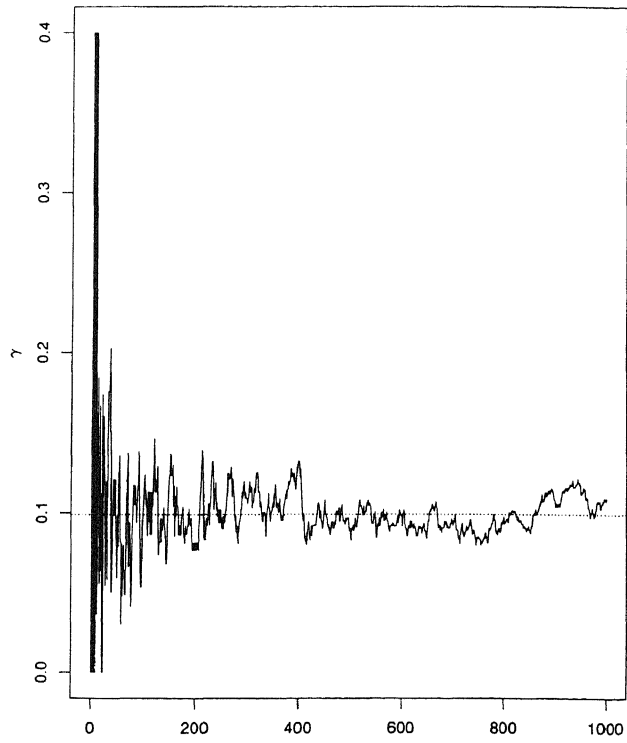
Fig. 2. (a) Simulated realization of a Strauss process, 25 points in the unit square with $r = 0.15$ and $\gamma = 0.1$. (b) Sequential plot of the R–M estimate of γ for the simulated data in Fig. 2(a) with $s(x) = 4$ and $A = 1$. Dotted line shows the MLE of $\gamma = 0.099$. (c) As in (b) with $A = A_{\mathrm{opt}} = 0.04$.

6. Conclusions and discussion

The use of the optimal rate constant in the R–M method can dramatically improve performance. Note that the larger the value of μ, the greater this improvement.

The starting value γ_0 is arbitrary, but should be set to an initial approximation such as that holding in the sparse case (i.e. derived from the Poisson approximation to $\{p_{nk}\}$, Ripley 1988)

$$\gamma_0 = \frac{2s(x)\,\mathrm{area}(W)}{n(n-1)\pi r^2}.$$

This approximation should also be used to provide initial values of μ and σ^2 for choosing the value of A and estimating the asymptotic variance of γ_k. Thus A_{opt} would be estimated as

$$\frac{2\,\mathrm{area}(W)}{n(n-1)\pi r^2}.$$

It is also possible to obtain consistent estimates of μ and σ^2 during the R–M iteration step, using the simulated data; see Dupač (1984). This would allow us to adjust A dynamically to its optimal value, and to estimate the standard error of γ_k consistently. This would probably turn out to be cheaper than Penttinen's (1984) method, where one is required to simulate

from the Strauss process to get estimates of the mean and mean derivative of the efficient score for every update of the parameter.

However, the **number** of iterations of the R–M procedure can be decided in advance. The variance of $\hat{\gamma}$ itself is approximately

$$\mathbf{Var}(\hat{\gamma}) = \frac{\gamma^2}{\mathbf{Var}_\gamma[s(x)]}$$

using the delta method (note that we are not in a position to apply usual asymptotic results for MLE's). Our results give for the optimal A

$$\mathbf{Var}_{\mathrm{opt}}(\gamma_k) = \frac{\gamma^2}{k\,\mathbf{Var}_\gamma[s(x)]},$$

so that after k iterations (with the optimal parameter settings) the variance due to stochastic approximation is approximately a fraction $1/(k+1)$ of the total variance of estimation. Twenty iterations would often suffice.

A remaining problem with the R–M procedure (and all simulation-based procedures) is that, for a stochastic model which can only be simulated and not treated analytically, it is not possible to identify the regions of the sample space that correspond to values of the maximum likelihood estimate lying on the boundary of the parameter space. In our case, the values K and $M(1)$ in (16) are not known analytically, and would have to be estimated by simulation.

This is not a serious problem for the upper limit $M(1)$, since we may run simulations *of the conditional model* for $\gamma > 1$, and simply allow γ_k to exceed 1, instead of clipping to [0, 1]. A final value of γ_k greater then 1 indicates $\hat{\gamma} = 1$ for the parameter space [0, 1]. If γ_k is much greater than 1, the practical interpretation might be that r is too large, or that the model (3) is wrong.

However, if the lower limit K is positive, and the R–M process gives $\gamma_k = 0$ for some k, then the standard birth-and-death process simulations (see Ripley 1977, 1979 and Baddeley & Møller, 1989) will go into an infinite loop because they use the rejection method in a situation where the success probability is zero. In case W is convex, a sufficient condition for $K > 0$ is that $n\pi r^2/4 > \mathrm{area}(W) + r\,\mathrm{length}(\partial W) + \pi r^2$, but there are marginal values of r below this limit where $K > 0$ still. One response would be to set an upper limit to the number of rejections that would be allowed during the birth of a new point in the spatial birth-and-death process simulations. When this limit is exceeded, γ_k may be put slightly above zero and one can then continue with the iterations. This would be in the style of (24) but we have not yet investigated it. An alternative would be to recast the entire estimation problem in terms of $\zeta = \log \gamma$ thereby avoiding simulations of the hard core process.

Another unanswered question about the simulations is whether the long (theoretically infinite) time required for running a spatial birth-and-death process to equilibrium could be avoided by using shorter runs. This would introduce bias at each step, and dependence between steps if the birth/death simulations of X_{γ_k} started with $X_{\gamma_{k-1}}$ as the initial state.

We have also ignored important questions (raised e.g. by Ripley, 1988) about the desirability of maximum likelihood estimators, the applicability of Markov point process models to real data, and the role of edge effects, since these do not bear on the technique itself.

Acknowledgements

References

Baddeley, A. J. & Møller, J. (1989). Nearest-neighbour Markov point processes and random sets. *Internat. Statist. Rev.* **57**, 89–121.

Besag, J. (1977). Some methods of statistical analysis for spatial data. *Bull. Internat. Statist. Inst.* **47**(2), 77–92.

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

Diggle, P. J. & Gratton, R. J. (1984). Monte Carlo methods of inference for implicit statistical models (with discussion). *J. Roy. Statist. Soc. Ser B* **46**, 193–227.

Dupač, V. (1984). Stochastic approximation. In *Handbook of statistics, Vol.* 4 (eds. P. R. Krishnaiah and P. K. Sen), pp. 515–529. Elsevier Science Publishers.

Hall, P. & Heyde, P. (1980). *Martingale limit theory and its application.* Academic Press, London.

Kelly, F. P. & Ripley, B. D. (1976). On Strauss's model for clustering. *Biometrika* **63**, 357–360.

Nevel'son, M. B. & Has'minskiĭ, R. Z. (1976). *Stochastic approximation and recursive estimation.* Translations of Math. Monographs, Vol. 47, American Math. Society, Providence.

Ogata, Y. & Tanemura, M. (1984). Likelihood analysis of spatial point patterns. *J. Roy. Statist. Soc. Ser. B* **46**, 496–518.

Penttinen, A. (1984). Modelling interactions in spatial point patterns: parameter estimation by the maximum likelihood method. *Jyväskylä Studies in Computer Science, Economics and Statistics* 7.

Ripley, B. D. (1977). Modelling spatial patterns (with discussion). *J. Roy. Statist. Soc. Ser. B* **39**, 172–212.

Ripley, B. D. (1979). Simulating spatial patterns: dependent samples from a multivariate density. *Applied Statistics* **28**, 109–112.

Ripley, B. D. (1981). *Spatial statistics.* Wiley, New York.

Ripley, B. D. (1987). *Stochastic simulation.* Wiley, New York.

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

Ripley, B. D. (1989). Gibbsian interaction models. In *Spatial statistics: past, present and future* (ed. D. A. Griffiths), Image, New York.

Ripley, B. D. & Kelly, F. P. (1977). Markov point processes. *J. Lond. Math. Soc.* **15**, 188–192.

Robbins, H. & Monro, S. (1951). A stochastic approximation method. *Ann. Math. Statist.* **22**, 400–407.

Rana A. Moyeed, School of Mathematical Sciences, University of Bath, Claverton Down, Bath BA2 7AY, United Kingdom.