A Festschrift for Herman Rubin
Institute of Mathematical Statistics
Lecture Notes – Monograph Series
Vol. 45 (2004) 309–318

Maximum likelihood estimation for the contact process

Marta Fiocco[1] and Willem R. van Zwet[2]

Leiden University Medical Centre and University of Leiden

Abstract: The contact process—and more generally interacting particle systems—are useful and interesting models for a variety of statistical problems. This paper is a report on past, present and future of research by the authors concerning the problem of estimating the parameters of the contact process. A brief review of published work on an ad-hoc estimator for the case where the process is observed at a single (large) time t is given in Section 1. In Section 2 we discuss maximum likelihood estimation for the case where the process is observed during a long time interval $[0, t]$. We construct the estimator and state its asymptotic properties as $t \to \infty$, but spare the reader the long and tedious proof. In Section 3 we return to the case where the process is observed at a single time t and obtain the likelihood equation for the estimator. Much work remains to be done to find a workable approximation to the estimator and study its properties. Our prime interest is to find out whether it is significantly better than the ad-hoc estimator in Section 1.

It was a joy to write this paper for Herman Rubin's festschrift. To this is added the bonus that Herman will doubtless solve our remaining problems immediately.

1. Introduction

The contact process was introduced and first studied by Harris (1974). It is described as follows. At every time $t \geq 0$, every point (or site) x in the d-dimensional integer lattice Z^d is in one of two possible states that we shall call infected and healthy. The process starts at time $t = 0$ with a non-empty set $A \subset Z^d$ of infected sites. At time $t \geq 0$, the state of the site $x \in Z^d$ will be indicated by a random variable $\xi_t^A(x)$, given by

$$\xi_t^A(x) = \begin{cases} 1, & \text{if site x is infected at time t} \\ 0, & \text{if site x is healthy at time t .} \end{cases} \tag{1.1}$$

The function $\xi_t^A : Z^d \to \{0, 1\}$ describes the state of the process at time t and $\xi_0^A = 1_A$, the indicator function of the set A.

The evolution of this $\{0, 1\}$-valued random field is described by the following dynamics. A healthy site is infected independently and at rate $\lambda > 0$ by each of its $2d$ immediate neighbors that is itself infected. An infected site recovers at

[1]Department of Medical Statistics, Leiden University Medical Centre, P.O. Box 9604, 2300 RC Leiden, The Netherlands. e-mail: m.fiocco@lumc.nl

[2]Mathematical Institute, University of Leiden, PO Box 9512, 2300 RA Leiden, The Netherlands. e-mail: vanzwet@math.leidenuniv.nl

Keywords and phrases: contact process, supercritical contact process, maximum likehood, counting process.

AMS 2000 subject classifications: 62M30.

Figure 1: The process $\xi_t^{\{0\}}$ for $\lambda = 3$ and $\mu = 1$ after $30,000$ steps. Infected sites are represented by gray 1×1 squares. A darker gray level indicates a longer duration of the present infection.

rate $\mu > 0$. Given the configuration ξ_t^A at time t, the processes involved are independent until a change occurs. For $d = 2$ the contact process is a simplified model for the spread of an infection or, more generally, of a biological species in the plane. The growth of a forest is an example if diseased and healthy are interpreted as presence and absence of a tree in a square centered at the lattice site.

In Figure 1 we show the process that started with a single infected site at the origin with $\lambda = 3$ and $\mu = 1$ after 30,000 steps, i.e. 30,000 infections and recoveries. Infected sites are indicated by gray 1×1 squares. An additional feature of this figure is that for each infected site we have kept track of the number of steps since it was last infected and have indicated this by the gray level at that site: the darker the gray level, the older the present infection at a site. If we view the process as a model for the growth of a forest, then the gray level indicates the age of the tree. Obviously, the older trees are in the center of the picture away from the boundary.

It is sometimes convenient to represent the state of the contact process at time t by the set of infected sites rather than by the function $\xi_t^A : Z^d \to \{0, 1\}$. Usually, this set is also denoted by ξ_t^A. Thus, by an abuse of notation, we write

$$\xi_t^A = \big\{x \in Z^d : \xi_t^A(x) = 1\big\}. \tag{1.2}$$

Let

$$\tau^A = \inf\big\{t : \xi_t^A = \emptyset\big\} \tag{1.3}$$

denote the time the infection dies out with the convention that $\tau^A = \infty$ if the infection survives forever. For a set $C \subset R^d$ and $a > 0$, we write $aC = \{ax : x \in C\}$. For sets C and D in R^d, $C \oplus D = \{x+y : x \in C, y \in D\}$ will denote their Minkowski sum and we define

$$H_t^A = \bigcup_{0 \le s \le t} \xi_s^A \oplus Q[-1/2, 1/2]^d \,. \tag{1.4}$$

Thus H_t^A is obtained from the set of sites that have been infected up to and including time t by replacing each site by a hypercube with sides of length 1 centered at this site.

The contact process has been the subject of extensive studies during the past decades. We list a few of its basic properties.

Property 1. If $\rho = \lambda/\mu$ exceeds a certain critical value ρ_d , then the infection will continue forever (i.e. $\tau^A = \infty$) with positive probability depending on the dimension d and the initial set A. This is called the supercritical case. On the other hand, if $\rho \leq \rho_d$, then the infection will eventually die out (i.e. $\tau^A < \infty$) with probability 1. We shall restrict attention to the supercritical case.

Property 2. In the supercritical case, there exist positive constants C and γ such that for every $t > 0$ and $A \subset Z^d$ with cardinality $|A|$,

$$\mathbb{P}\big(t < \tau^A < \infty\big) \leq Ce^{-\gamma t}, \qquad \mathbb{P}\big(\tau^A < \infty\big) \leq e^{-\gamma|A|}. \tag{1.5}$$

In particular, if A is infinite, then in the supercritical case the infection will survive forever.

Property 3. The distribution of the set ξ_t^A converges weakly to a limit distribution

$$\mathbb{P}\big(\tau^A < \infty\big)\delta_\emptyset + \mathbb{P}\big(\tau^A = \infty\big)\nu, \tag{1.6}$$

where $\delta_\emptyset$ denotes the measure that assigns probability 1 to the empty set and ν is the equilibrium measure depending only on ρ and the dimension d. Thus, given that the process survives forever—which is possible only in the supercritical case—it tends in distribution to ν. Here weak convergence coincides with convergence in distribution of the finite dimensional projections $\xi_t^A \cap F$, (i.e. $\{\xi_t^A(x) : x \in F\}$) for finite $F \subset Z^d$.

Property 4. There exists a bounded convex set $U \subset R^d$ with the origin as an interior point such that for every bounded $A \subset Z^d$, $\epsilon > 0$ and $t \to \infty$,

$$(1-\varepsilon)tU \subset \ H_t^A \ \subset (1+\varepsilon)tU, \tag{1.7}$$

eventually almost surely on the set $\{\tau^A = \infty\}$ where ξ_t^A survives forever. Thus if the infection persists, then for large t, $\mathrm{H}^A{}_t$ will grow linearly in t in every direction and $\mathrm{t}^{-1}\ \mathrm{H}^A{}_t$ will assume the shape of U. Moreover, on the set $\{\tau^A = \infty\}$ and for large t, the distribution of $\xi_t^A \cup (1-\epsilon)tU$ will approach its asymptotic distribution under the equilibrium measure ν in a sense that we shall not make precise here.

For these facts and other related matters the reader may consult Liggett (1985 & 1999).

The contact process and its many possible generalizations provide an interesting class of models for problems in spatial statistics and image analysis. In Fiocco & van Zwet (2003a & b) we began a statistical study of the supercritical contact process $\xi_t^{\{0\}}$ that starts with a single infected site at the origin and is conditioned on survival, i.e. on $\{\tau^{\{0\}} = \infty\}$. For this process we considered the simplest possible statistical problem, that is, to estimate the parameters of the contact process based on observing the set of infected sites at a single (large but unknown) time t. This corresponds to the realistic situation when one observes a large forest that has obviously been there for a long time without any knowledge when it began. On the

basis of such an observation it is clear that one can only estimate $\rho = \lambda/\mu$ but not λ and μ individually, as without knowing t, one cannot distinguish between observing the processes with parameters $c\lambda$ and $c\mu$ at time t/c for different values of $c > 0$. Equivalently, one may set $\mu = 1$ arbitrarily and estimate λ.

For any $x, y \in Z^d$ and $C \subset R^d$, let $|x - y| = \sum_{1 \le i \le d} |x_i - y_i|$ denote the L_1-distance of x and y, and define

$$k_t^{\{0\}}(x) = \big(1 - \xi_t(x)\big) \sum_{|x-y|=1} \xi_t^{\{0\}}(y), \tag{1.8}$$

$$n_t^{\{0\}}(C) = \sum_{x \in C \cap Z^d} \xi_t^{\{0\}}(x), \quad k_t^{\{0\}}(C) = \sum_{x \in C \cap Z^d} k^{\{0\}}(x). \tag{1.9}$$

Notice that $n_t^{\{0\}}(C)$ is simply the number of infected sites in C and $k_t^{\{0\}}(C)$ equals the number of neighboring pairs of infected and healthy sites, with the healthy site in C. For $x \in Z^d$, the flip rates at time t equal $\lambda k_t^{\{0\}}(x)$ and $\mu \xi_t^{\{0\}}(x)$ for the transitions $0 \to 1$ and $1 \to 0$ respectively and hence the number $n_t^{\{0\}}(C)$ of infected sites increases by 1 at time t with rate $k_t^{\{0\}}(C)$ and decreases by 1 with rate $n_t^{\{0\}}(C)$. In Property 4 above, we explained that on $\{\tau^{\{0\}} = \infty\}$ and at a large time t, the process will have progressed past the set $(1-\epsilon)tU$ and will be close to equilibrium there. This implies that the rate of increase of $n_t^{\{0\}}((1-\epsilon)tU)$ should approximately equal its rate of decrease, so that $\lambda k_t^{\{0\}}((1-\epsilon)tU) \approx \mu n_t^{\{0\}}((1-\epsilon)tU)$. Hence on $\{\tau^{\{0\}} = \infty\}$, $n_t^{\{0\}}((1-\epsilon)tU)/k_t^{\{0\}}((1-\epsilon)tU)$ should be a plausible estimator of $\rho = \lambda/\mu$, or of λ if one assumes $\mu = 1$, where it not for the fact that U is unknown. However, one can show that for every $\epsilon > 0$, the convex hull $\mathcal{C}(\xi_t^{\{0\}})$ of the set of infected sites $\xi_t^{\{0\}}$ satisfies

$$(1-\epsilon)tU \subset \mathcal{C}\Big(\xi_t^{\{0\}}\Big) \subset (1+\epsilon)tU, \tag{1.10}$$

eventually almost surely on $\{\tau^{\{0\}} = \infty\}$, so that $\mathcal{C}(\xi_t^{\{0\}})$ apparently approximates tU. If, for any $\delta > 0$, we define

$$C_t = (1-\delta)\mathcal{C}\Big(\xi_t^{\{0\}}\Big), \tag{1.11}$$

then for some $\epsilon > 0$, (1.10) ensures that $C_t \subset (1-\epsilon)tU$ eventually a.s. on $\{\tau^{\{0\}} = \infty\}$. Hence

$$\tilde{\rho}_t = \frac{n_t^{\{0\}}(C_t)}{k_t^{\{0\}}(C_t)}, \tag{1.12}$$

would seem to be a sensible estimator of ρ, given that the process $\xi_t^{\{0\}}$ will survive forever. Indeed we prove in Fiocco & van Zwet (2003b) that conditional on $\{\tau^{\{0\}} = \infty\}$, $\tilde{\rho}_t$ is a strongly consistent and asymptotically normal estimator of ρ, that is, as $t \to \infty$,

$$\tilde{\rho}_t \to \rho \qquad a.s. \qquad |C_t|_d^{1/2}\big(\tilde{\rho}_t - \rho\big) \xrightarrow{D} N\big(0, \tau^2\big) \tag{1.13}$$

Here $|C_t|_d$ denotes the cardinality of $C_t \cap Z^d$, or alternatively, the Lebesgue measure of C_t , and an explicit expression for τ^2 is available. For our purposes we merely note

that this implies that $(\tilde{\rho}_t - \rho) = \mathcal{O}_P(t^{-d/2})$ on $\{\tau^{\{0\}} = \infty\}$. Simulation confirms that the estimator behaves as predicted by the asymptotics (Fiocco (1997)).

For the estimator $\tilde{\rho}_t$ to perform well asymptotically as well as in simulations, it is essential that δ should indeed be positive in (1.11). At time t the process has spread approximately to tU, but beyond $(1-\epsilon)tU$ it is not yet in equilibrium and our argument fails. This is also intuitively obvious: having just reached the boundary of tU, the infected sites beyond $(1-\epsilon)tU$ should be less dense than they are closer to the origin where the infection arrived earlier and had time to achieve equilibrium. Beyond $(1-\epsilon)tU$ the fraction of infected sites should be too small, but among the infected sites the fraction with healthy neighbors should be too large. As a result $\tilde{\rho}_t$ should systematically underestimate ρ if δ is taken to be zero and simulation not only confirms this, but shows that in this case the estimator is bad. This effect also shows up asymptotically as $t \to \infty$. If $\delta = 0$, we can still prove consistency but no longer asymptotic normality. Shrinking the convex hull $\mathcal{C}(\xi_t^{\{0\}})$ to obtain the mask C_t for the estimator is essential for obtaining a satisfactory estimator.

Two minor problems are left. First, shrinking $\mathcal{C}(\xi_t^{\{0\}})$ towards the origin to obtain C_t is possible only if one knows where the origin is, i.e. where the infection has started. Generally this is not known: one sees the forest today, but not when or where it began. Of course one can estimate the origin in many different ways, for instance by averaging the locations of the infected sites. Shrinking towards this estimated origin will not influence the asymptotic behavior of the estimator. A more elegant solution is to replace the shrinking of $\mathcal{C}(\xi_t^{\{0\}})$ by another operation that removes the sites near the boundary of this set. Such operation is called peeling, where one removes layer after layer of sites on the boundary of the convex hull. In general, almost any reasonable type of shrinking will leave the asymptotic behavior of the estimator unchanged as long as the same fraction of sites is removed. Simulation suggests that this fraction should be around 20-30%, decreasing with increasing t.

Second, our analysis refers only to the behavior of the process - and hence of the estimator - on the set where $\tau^{\{0\}} = \infty$. Obviously, if $\tau^{\{0\}} < \infty$, there is not much to observe for sufficiently large t, since the infection will have died out. On the other hand, we can not know with certainty at any finite time t that we are indeed in the case where $\tau^{\{0\}} = \infty$, so one may wonder whether asymptotic results for $t \to \infty$ that are valid only on the set $\tau^{\{0\}} = \infty$ have any statistical significance. However, (1.5) ensures that having survived until a large time t, the infection will survive forever with overwhelming probability. Asymptotic results conditional on $\tau^{\{0\}} = \infty$ are therefore the same as those conditioned on $\tau^{\{0\}} \geq t$, that is, on the infection being present when observed.

2. Maximum likelihood for the fully observed process

Having briefly described the statistical results obtained for the contact process observed at a single time t, we now turn to the case where this process is observed continuously on the interval $[0,t]$ for a known (large) $t > 0$. In this case it should be possible to estimate λ and μ separately, rather than just their ratio $\rho = \lambda/\mu$. In fact we shall derive the maximum likelihood estimators of these parameters.

Let $0 < T_1 < \ldots < T_N$ denote the times when the contact process undergoes a change in the time interval $[0,t]$ and, for $i = 1, 2, \ldots, N$, let x_i denote the site at which the change occurs at time T_i. It will be convenient to write $T_0 = 0$ and

$T_{N+1} = t$ and $\xi_i = \xi_{T_i}^{\{0\}}$ for the configuration of the process at time T_i. Given the configuration ξ_{i-1} at time T_{i-1}, the rate of change at site x equals

$$r_i(x) = \begin{cases} \lambda \sum_{|x-y|} \xi_{i-1}(y) & \text{if } \xi_{i-1}(x) = 0 \\ \mu & \text{if } \xi_{i-1}(x) = 1, \end{cases} \tag{2.1}$$

and the total rate of change at any site is given by

$$R_i = \sum_{x \in Z^d} r_i(x) = \lambda k_{T_{i-1}}^{\{0\}}\big(Z^d\big) + \mu n_{T_{i-1}}^{\{0\}}\big(Z^d\big) \tag{2.2}$$

It follows that the likelihood of the observed process on $[0, t]$ is given by

$$L(\lambda, \mu) = \prod_{1 \le i \le N} R_i \exp\big\{-R_i[T_i - T_{i-1}]\big\}[r_i(x_i)/R_i] \exp\big\{-R_{N+1}[t - T_N]\big\}.$$

Hence

$$\log L(\lambda, \mu) = -\sum_{1 \le i \le N+1} R_i[T_i - T_{i-1}] + U_t \log \lambda + D_t \log \mu + h\big(\xi^{\{0\}}\big)$$

where U_t and D_t are the number of upward and downward jumps of the process on $[0, t]$, i.e.

$$U_t = \#\big\{0 \le i \le N-1 : \xi_{i-1}(x_i) = 0\big\} = \#\big\{1 \le i \le N : \xi_i(x_i) = 1\big\}, \tag{2.3}$$

$$D_t = \#\big\{0 \le i \le N-1 : \xi_{i-1}(x_i) = 1\big\} = \#\big\{1 \le i \le N : \xi_i(x_i) = 0\big\}, \tag{2.4}$$

and $h(\xi^{\{0\}})$ depends on the process $\{\xi_s^{\{0\}} : 0 \le s \le t\}$, but not on the parameters λ and μ.

Define

$$A_t = \int_0^t k_s^{\{0\}}\big(Z^d\big)\, ds, \qquad B_t = \int_0^t n_s^{\{0\}}\big(Z^d\big)\, ds. \tag{2.5}$$

As $n_s^{\{0\}}(Z^d)$ and $k_s^{\{0\}}(Z^d)$ are constant for $s \in [T_{i-1}, T_i)$ and $T_{N+1} = t$, (2.2) implies that

$$\sum_{1 \le i \le N+1} R_i[T_i - T_{i-1}] = \lambda A_t + \lambda B_t$$

and hence

$$\log L(\lambda, \mu) = -\lambda A_t - \mu B_t + U_t \log \lambda + D_t \log \mu + h\big(\xi^{\{0\}}\big). \tag{2.6}$$

Differentiating with respect to λ, and μ we find that the maximum likelihood estimators $\hat{\lambda}_t$ and $\hat{\mu}_t$ of λ, and μ are given by

$$\hat{\lambda}_t = \frac{U_t}{A_t}, \qquad \hat{\mu}_t = \frac{D_t}{B_t}. \tag{2.7}$$

The maximum likelihood estimator of $\rho = \lambda/\mu$ therefore equals

$$\hat{\rho}_t = \frac{U_t B_t}{D_t A_t}. \tag{2.8}$$

As in the previous section we can prove that conditional on $\{\tau^{\{0\}} = \infty\}$, these estimators are strongly consistent and asymptotically normal, but converge to the parameter to be estimated at the faster rate $\mathcal{O}(t^{-(d+1)/2})$. Thus conditional on $\{\tau^{\{0\}} = \infty\}$ and as $t \to \infty$,

$$\hat{\lambda}_t \to \lambda \quad a.s., \quad \hat{\mu}_t \to \mu \quad a.s. \quad \hat{\rho}_t \to \rho \quad a.s., \tag{2.9}$$

$$\begin{aligned} t^{(d+1)/2}(\hat{\lambda}_t - \lambda) &\to N(0, \sigma_\lambda^2), \\ t^{(d+1)/2}(\hat{\mu}_t - \mu) &\to N(0, \sigma_\mu^2), \\ t^{(d+1)/2}(\hat{\rho}_t - \rho) &\to N(0, \sigma_\rho^2), \end{aligned} \tag{2.10}$$

again with explicit expressions for the variances being available. The proof is long and involved and will be given elsewhere.

There are two different ways of looking at these maximum likelihood estimators heuristically. First we may observe that the counting process U_t has compensator λA_t and since $A_t \to \infty$ if $\tau^{\{0\}} = \infty$, $\hat{\lambda}_t = U_t/A_t$ should approximate λ. Similarly, μB_t is the compensator of D_t and $\hat{\mu}_t = D_t/B_t$ should approximate μ on $\{\tau^{\{0\}} = \infty\}$. Hence $\hat{\lambda}_t$, $\hat{\mu}_t$ and $\hat{\rho}_t$ are plausible estimators of λ, μ and ρ.

However, one may also be interested in a comparison of the maximum likelihood estimator $\hat{\rho}_t$ based on the fully observed process $\{\xi_s^{\{0\}} : 0 \le s \le t\}$, and the ad-hoc estimator $\tilde{\rho}_t$ of Section 1, which is based on observing $\xi_t^{\{0\}}$ at the single time t. We assume throughout that $\tau^{\{0\}} = \infty$. First of all, (B_t/A_t) in (2.8) estimates the same quantity ρ as $\tilde{\rho}_t = n_t^{\{0\}}(C_t)/k_t^{\{0\}}(C_t)$ in (1.12). On the one hand, B_t/A_t averages information over the interval $[0, t]$ and should therefore have a variance of a smaller order than $n_t^{\{0\}}(C_t)/k_t^{\{0\}}(C_t)$. On the other hand B_t/A_t uses the entire set of infected points and its healthy neighbors, and we have argued in Section 1, that without shrinking this set, this will lead to underestimating ρ. The factor U_t/D_t in (2.8) now serves to correct this negative bias. In equilibrium, the number of upward and downward jumps should approximately cancel out, but near the boundary of the set of infected points, equilibrium has not yet set in. In fact, the number of infected sites $U_t - D_t + 1$ grows roughly as a constant factor times the Lebesgue measure of tU, that is, at the rate of t^d. Individually, both U_t and D_t are counting processes and easily seen to be of order t^{d+1}. Hence $(U_t/D_t) - 1$ is positive and decreases at the rate t^{-1}, so that the factor U_t/D_t in (2.8) does serve to correct the negative bias which does indeed decrease like t^{-1}.

The asymptotic results (1.13) and (2.10) imply that the estimators $\tilde{\rho}_t$ and $\hat{\rho}_t$ of ρ have random errors of orders $\mathcal{O}(t^{-d/2})$ and $\mathcal{O}(t^{-(d+1)/2})$ respectively. Hence the maximum likelihood estimator $\hat{\rho}_t$ based on observing the entire process $\{\rho_s^{\{0\}} : 0 \le s \le t\}$, is asymptotically an order of magnitude better than the ad-hoc estimator $\tilde{\rho}_t$ based on a single observation of $\xi_t^{\{0\}}$. In Figure 2 we show a single run of simulated values of both estimators after $500, 1,000, 1,500, \ldots, 20,000$ jumps of the process for $\lambda = 0.8$ $\mu = 1$, and hence $\rho = 0.8$. For the ad-hoc estimator, the shrinking of the convex hull of infected sites $\mathcal{C}(\xi_t^{\{0\}})$ to obtain the mask C_t has been achieved by peeling rather than multiplication by $(1 - \delta)$ as is done in (1.11). Peeling fractions of 30%, 50% and 70% were used. It appears that the maximum likelihood estimator is indeed superior.

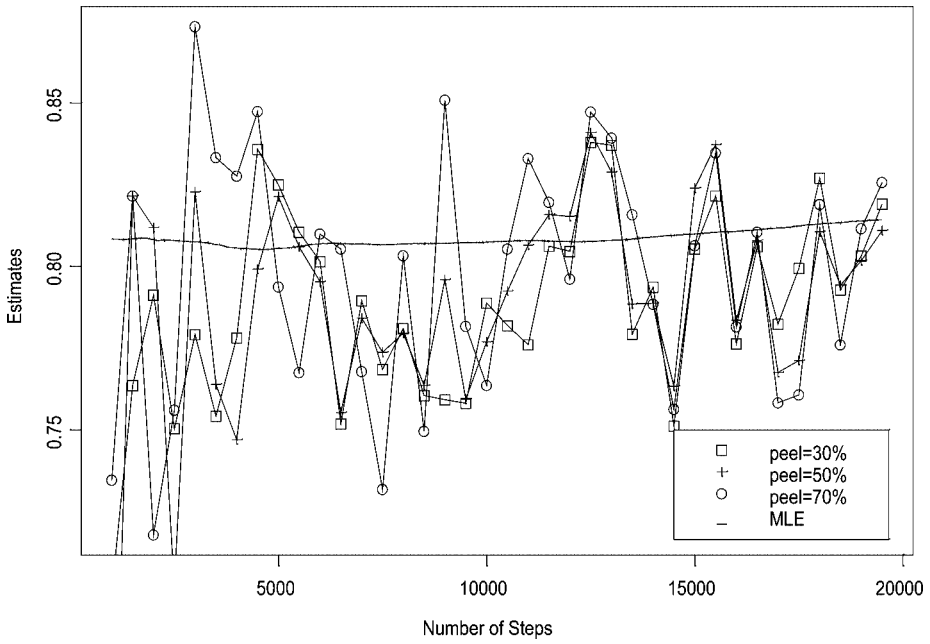

Figure 2: Maximum likelihood estimator $\hat{\rho}_t$ and the ad-hoc estimator $\tilde{\rho}_t$.

3. Maximum likelihood for the singly observed process

As we pointed out in Section 1, one will rarely have the opportunity to observe the process throughout a time interval $[0, t]$. In most cases one will have to be content with a single observation of the process at a (large but unknown) time t. For the latter situation we reported on the study of an ad-hoc estimator $\tilde{\rho}_t = n_t^{\{0\}}(C_t)/k_t^{\{0\}}(C_t)$ of ρ, and noted that it is essential to choose the mask C_t well inside the convex hull $\mathcal{C}(\xi_t^{\{0\}})$ of the set of infected points in order to avoid underestimating ρ. Of course, we are still interested in finding and studying the maximum likelihood estimator for this case, if only to see whether or not it will improve substantially on the ad-hoc estimator.

Obviously this is going to be a difficult assignment. In Section 2 we studied the maximum likelihood estimator for the fully observed process and discovered two things. First of all this estimator uses the ratio of (the integrals of) $n_s^{\{0\}}(Z^d)$ and $k_s^{\{0\}}(Z^d)$ and we conclude that the use $n_t^{\{0\}}/k_t^{\{0\}}$ in the ad-hoc estimator was a good idea. Second, the bias correction was achieved by the correction factor U_t/D_t, which is a rather more subtle way to achieve this than by discarding a sizeable fraction of the data, as is done for the ad-hoc estimator. It therefore seems plausible that the maximum likelihood estimator for the singly observed process will also depend on $n_t^{\{0\}}/k_t^{\{0\}}$, and that conditional expectations of the numbers of upward and downward jumps in $[0, t]$ given $\xi_t^{\{0\}}$ will also play a part.

Since we observe $\xi_t^{\{0\}}$ at an unknown time t and have no information about the times of any of the jumps, we may discard the time element entirely and view the process as a sequence of configurations $\xi_1^{\{0\}}, \xi_2^{\{0\}}, \ldots, \xi_{n-1+2k}^{\{0\}}$ after the first, second,$\ldots$, $(n-1+2k)$th jumps that take place consecutively at sites $x_1, x_2, \ldots, x_{n-1+2k}$ during the time interval $[0, t]$. The final configuration $\xi_{n-1+2k}^{\{0\}}$ equals

the observed configuration $\xi_t^{\{0\}}$. For some k, $(n-1+k)$ of the jumps are upward (i.e. $n_t^{\{0\}}$ increases by 1 at this jump) and k are downward. Hence the total increase $n_t^{\{0\}}(Z^d) - n_0^{\{0\}}(Z^d) = n_t^{\{0\}}(Z^d) - 1$ of the number of infected points must equal $(n-1+k) - k = n-1$, so that we must have $n = n_t^{\{0\}}(Z^d)$. Finally, we write $n_{i-1}^{\{0\}}(x_i)$ and $k_{i-1}^{\{0\}}(x_i)$ for the values of $n_t^{\{0\}}$ and $k_t^{\{0\}}$ after the time of the $(i-1)st$ jump at the site x_i where the next jump will occur, and $n_i^{\{0\}}(Z^d)$ and $k_i^{\{0\}}(Z^d)$ for the values of $n_t^{\{0\}}(Z^d)$ and $k_t^{\{0\}}(Z^d)$ immediately after the $i-th$ jump. The probability of $(n-1+k)$ upward and k downward jumps consecutively at sites $x_1, x_2, \ldots, x_{n-1+2k}$ equals

$$\prod_{1\le i\le n-1+2k} \frac{\lambda k_{i-1}^{\{0\}}(x_i) + \mu n_{i-1}^{\{0\}}(x_i)}{\lambda k_{i-1}^{\{0\}}(Z^d) + \mu n_{i-1}^{\{0\}}(Z^d)}$$
$$= \lambda^{n-1-k}\mu^k \prod_{1\le i\le n-1+2k} \frac{k_{i-1}^{\{0\}}(x_i) + n_{i-1}^{\{0\}}(x_i)}{\lambda k_{i-1}^{\{0\}}(Z^d) + \mu n_{i-1}^{\{0\}}(Z^d)}$$

because either $k_{i-1}^{\{0\}}(x_i)$ or $n_{i-1}^{\{0\}}(x_i)$ vanishes. It follows that the likelihood is given by

$$L^*(\lambda,\mu) = \sum_{0\le k<\infty} {\sum}^* \lambda^{n-1+k}\mu^k \prod_{1\le i\le n-1+2k} \frac{k_{i-1}^{\{0\}}(x_i) + n_{i-1}^{\{0\}}(x_i)}{\lambda k_{i-1}^{\{0\}}(Z^d) + \mu n_{i-1}^{\{0\}}(Z^d)}, \tag{3.1}$$

where $\sum^*$ denotes summation over all possible sequences $\xi_1^{\{0\}}, \xi_2^{\{0\}}, \ldots, \xi_{n-1+2k}^{\{0\}}$ for which $\xi_{n-1+2k}^{\{0\}}$ is the first configuration equaling $\xi_t^{\{0\}}$, and $n = n_t^{\{0\}}(Z^d)$. As we noted in Section 1 we can only estimate $\rho = \lambda/\mu$, but not λ and μ separately as t is unknown. However, we can still maximize the likelihood L^* as a function of λ and μ, but we shall find that both likelihood equations are identical. If U and D denote the number of upward and downward jumps until the configuration equals $\xi_t^{\{0\}}$ for the first time, then differentiation with respect to λ and μ yields the likelihood equations

$$E(U|\xi_t^{\{0\}}) = E\left[\sum_{1\le i<U+D}\left[\frac{\lambda k_{i-1}^{\{0\}}(Z^d)}{\lambda k_{i-1}^{\{0\}}(Z^d) + \mu n_{i-1}^{\{0\}}(Z^d)}\right]|\xi_t^{\{0\}}\right], \tag{3.2}$$

$$E(D|\xi_t^{\{0\}}) = E\left[\sum_{1\le i<U+D}\left[\frac{\mu n_{i-1}^{\{0\}}(Z^d)}{\lambda k_{i-1}^{\{0\}}(Z^d) + \mu n_{i-1}^{\{0\}}(Z^d)}\right]|\xi_t^{\{0\}}\right]. \tag{3.3}$$

Adding these two equations yields the identity $E(U+D|\xi_t^{\{0\}}) = E(U+D|\xi_t^{\{0\}})$, so (3.2) and (3.3) are equivalent to the difference

$$E(U-D|\xi_t^{\{0\}}) = E\left[\sum_{1\le i<U+D}\left[\frac{\lambda k_{i-1}^{\{0\}}(Z^d) - \mu n_{i-1}^{\{0\}}(Z^d)}{\lambda k_{i-1}^{\{0\}}(Z^d) + \mu n_{i-1}^{\{0\}}(Z^d)}\right]|\xi_t^{\{0\}}\right],$$

and since $U - D = n_t^{\{0\}}(Z^d) - 1$, this reduces to

$$E\left[\sum_{1\le i<U+D}\left[\frac{\lambda k_{i-1}^{\{0\}}(Z^d)-\mu n_{i-1}^{\{0\}}(Z^d)}{\lambda k_{i-1}^{\{0\}}(Z^d)+\mu n_{i-1}^{\{0\}}(Z^d)}\right]|\xi_t^{\{0\}}\right]=n_t^{\{0\}}(Z^d)-1. \tag{3.4}$$

Even though this last step removes the dependence on the conditional expectation of $U-D$, this is no great help since the conditional behavior of $U+D$ still enters through the range of the summation.

Thus, as expected the maximum likelihood estimator of $\rho=\lambda/\mu$ presumably depends on both $n_t^{\{0\}}(Z^d)/k_t^{\{0\}}(Z^d)$ and the conditional behavior of U given $\xi_t^{\{0\}}$. Obviously there are two different possibilities to study the maximum likelihood estimator, namely asymptotic approximation of the estimator and simulation. Work in the former direction is in progress.

References

[1] Fiocco, M. (1997). *Statistical Estimation for the Supercritical Contact Process.* PhD thesis, University of Leiden, The Netherlands.

[2] Fiocco, M. and van Zwet, W. R. (2003a). Decaying correlations for the supercritical contact process conditioned on survival. *Bernoulli*, **9**(5), 763–781. MR2047685

[3] Fiocco, M. and van Zwet, W. R. (2003b). Parameter estimation for the supercritical contact process *Bernoulli*, **9**(6), 1071–1092. MR2046818

[4] Liggett, T. (1985). *Interacting Particle Systems.* Springer-Verlag, New York. MR776231

[5] Liggett, T. (1999). *Stochastic Interacting Systems: Contact, Voter and Exclusion Processes.* Springer-Verlag, New York. MR1717346