

Research Report Department of Statistics Göteborg University Sweden

Outlying observations and their influence on maximum pseudo-likelihood estimates of Gibbs point processes

Aila Särkkä

Research Report 1996:5 ISSN 0349-8034

**					

OUTLYING OBSERVATIONS AND THEIR INFLUENCE ON MAXIMUM PSEUDO-LIKELIHOOD ESTIMATES OF GIBBS POINT PROCESSES

Aila Särkkä Department of Statistics Göteborg University

Abstract

Maximum pseudo-likelihood estimation method is an attractive method to estimate interaction parameters of Gibbs point processes. A drawback of the method is that it tends to overestimate interaction if there is strong repulsion between the points. We assumed that one reason for overestimation is that the method is sensitive to outlying points. Several techniques were used to detect outlying observations for the data of amacrine cells for which overestimation is suspected. Some strategies were then tested to take outliers into account in maximum pseudo-likelihood estimation.

Key words: Robustness, Tukey's sensitive curve, energy marks, EM algorithm, second order pseudo-likelihood, triplet interactions

1. Introduction

Gibbs point process is a model to describe spatial point patterns with interacting objects. The process is constructed by using a pair potential or an interaction function. The maximum pseudo-likelihood method can be applied to estimate the parameters characterizing the interaction function in order to reveal the scale and strength of interaction between the points (Besag, 1978). According to some empirical evidence, the method is not robust: it tends to overestimate interaction at least if repulsion between the points is strong (Särkkä, 1990; Diggle et al., 1994). We claim that the maximum pseudo-likelihood method is sensitive to the influence of a few individual outlying observations. Therefore methods to detect outlying points are needed.

Wartenberg (1990) considered exploratory spatial analysis the aim of which is to provide a quick and meaningful summary of both spatial and aspatial characteristics of a data set. He suggested ways to recognize outlying points. Stoyan and Grabarnik (1991) introduced a particular marking of points of Gibbs point process, so-called exponential energy marking, which can be applied in detecting local discrepancies.

In this paper Tukey's sensitive curve, exponential energy marking, distance to the neighbouring points and the number of neighbours, are suggested as tools for detecting outlying points of inhibitive Gibbs point process. Most of the methods can be applied for point processes in general; as an exception, exponential energy marking is characteristic for Gibbs point processes.

In order to study the influence of outlying points on the maximum pseudo-likelihood estimation the methods were applied to the data of amacrine cells (Diggle and Gratton, 1984). Interactions between the points of the data are strong and the maximum pseudo-likelihood estimates of the interaction parameters indicate stronger interaction than the corresponding maximum likelihood estimates which give a good fit. It appeared that isolated points cause the overestimation. We present several methods to take the isolated events into account in estimation to obtain better estimates. In addition, we summarize some additional trials that were made in order to correct the maximum pseudo-likelihood estimates.

2. Preliminaries

2.1. Gibbs point process

We consider a simple, stationary and isotropic Gibbs point process Φ in \mathbb{R}^2 observed in a bounded sampling window W (Stoyan et al., 1987). We concentrate on pairwise interaction processes, where a pair potential function $\phi:[0,\infty)\to\mathbb{R}$, depending only on the distance between two points, is applied to describe interactions between the points of the process. An interaction function $h=\exp(-\phi)$ can be used instead of the pair potential function. Especially for inhibitive processes it is convenient to use the interaction function because the values of it vary between zero and one and can be understood as probabilities to have points within a certain distance.

The so-called local energy, the energy needed to add the point x to the configuration φ is given by

$$E(x, \varphi) = \alpha + \sum_{y \in \varphi} \phi(||x - y||),$$

where α is a constant chemical activity connected to the intensity of the process and ||x-y|| is the distance between x and y.

The density function of the pairwise interaction process in a bounded window W is

$$f(\varphi_W) = \frac{1}{Z} \exp(-n\alpha - \sum_{i < j} \phi(||x_i - x_j||)),$$

where $\varphi_W = \{x_1, x_2, ..., x_n\}$ consists of the locations of n events in W and Z is a scaling factor (Ripley, 1988).

In order to avoid confusion with points of the process and points of W, we refer throughout this paper to *events* of the process or of the point pattern and to *points* of W.

All events of the point pattern are usually not observed, there are events outside of W that interact with the events of φ_W . There are several possibilities to take the edge effects into account (see e.g. Ripley, 1988). In this study periodic boundaries (opposite edges of W are identified) have been applied in calculations in order to avoid the boundary effects.

2.2. Maximum pseudo-likelihood estimation method

The idea of the maximum pseudo-likelihood (MPL) method is to construct a pseudo-likelihood (PL) function and maximize it with respect to the parameters. It is computationally easier than the maximum likelihood (ML) method since the likelihood function contains a scaling factor which cannot be calculated explicitly.

Jensen and Møller (1991) give a rigorous basis for using the maximum pseudo-likelihood method in estimation for spatial Gibbs point processes. According to their definition the PL function can be written as

$$\log PL(\theta; \varphi_W) = -\sum_{x \in \varphi_W} E(x, \varphi_W \setminus \{x\}) - \int_W \exp(-E(\xi, \varphi_W)) d\xi,$$

where $\theta = (\alpha, \beta)$ and β is the parameter of the pair potential function ϕ .

The estimation equations are obtained by maximizing the PL function with respect to the chemical activity and parameters of the interaction function. Therefore we obtain

$$n = \int_{W} \exp(-E(\xi, \varphi_W)) d\xi \tag{1}$$

by maximizing with respect to α and

$$\sum_{x \in \varphi_W} \sum_{y \in \varphi_W \setminus \{x\}} \frac{\partial \phi(||x - y||)}{\partial \beta} = \int_W \sum_{x \in \varphi_W} \frac{\partial \phi(||x - \xi||)}{\partial \beta} \exp(-E(\xi, \varphi_W)) d\xi \quad (2)$$

by maximizing with respect to β .

3. Methods to detect outlying events

Wartenberg (1990) studied outlying events in spatial data. He divided them into three groups: outliers, leverage events and influential events. An outlier is an observation which is relatively greater or smaller than all other observations. An observation that contributes disproportionately to a summary statistic is called a leverage event. If the deletion of a single observation effects a considerable change in the statistic being estimated the observation is said to be an influential event. Furthermore, he considered two types of

outliers, aspatial (global) and spatial (local). A value of a global outlier is unusual in any data set but a local outlier differs from the observations nearby it, not from all other observations. In this study we concentrate on influential events and local outliers and suggest methods to detect them. Most of the methods can be applied to point processes in general but calculating exponential energies is characteristic for Gibbs processes.

3.1. Tukey's sensitive curve

To detect possible influential events with respect to the parameter of interest we apply Tukey's sensitive curve

$$SC_n(x_i) = n(T_n(x_1,...,x_n) - T_{n-1}(x_1,...,x_{i-1},x_{i+1},...,x_n)),$$

where T_n is the estimate calculated from n events (Huber, 1981). The value of $SC_n(x_i)$ tells the influence of deletion of the event x_i on the estimate.

3.2. Exponential energy marking

Stoyan and Grabarnik (1991) suggested calculating exponential local energies

$$m(x_i) = \exp\left(E(x_i, \varphi_W \setminus \{x_i\})\right) = \exp\left(\alpha + \sum_{y \in \varphi_W} \phi(||x - y||)\right)$$

for the events of Gibbs process in order to make determining of moment characteristics easy for stationary and isotropic processes.

Exponential energies can be used in exploratory analysis of Gibbs point processes. After α and parameters of ϕ have been estimated the exponential energy can be calculated for each event. Differencies between the energies indicate local discrepancies: events or groups of events with extreme exponential energies indicate regions of irregularity or outliers. Considering inhibitive point patterns, for the events near to the other events the values of the pair potential function and therefore of the exponential energy are large compared to the others, and for the events far from the others the corresponding values are small. Therefore exceptionally large values indicate clustering while small ones indicate isolation. For attractive models it is the contrary.

3.3. Other methods

Distance to the nearest events (nearest, second nearest etc.) or number of neighbours are suitable statistics to be applied in detecting outliers. Roughly, long distances to small number of neighbours indicate isolation and small distances to huge number of neighbours clustering.

To detect isolated events Wartenberg (1990) suggested to pay attention to the nearest neighbour distance or average distance from one event to all other events. He also mentioned correlograms, Voronoi tessellations and area of Thiessen polygons.

4. Parameter estimation for contaminated data

In the following we present methods to take outlying observations into account in estimation. Let x_{i_o} denote the location of an outlier and φ_W^- the point pattern $\{x_1, ..., x_{i_o-1}, x_{i_o+1}, ..., x_n\}$. Methods are presented here for the case of one outlier but they can easily be applied to the case of several outliers, too.

4.1. Omitting an outlier

The simplest way to take an outlying observation into account in estimation is to omit it and estimate the parameters from the point pattern φ_W^- . The estimation equations (1) and (2) can be thought as equalities of sample means. However, omitting an outlier does not only mean that the mean values on the left sides of (1) and (2) are calculated as sums over n-1 observations instead of n observations. One event has effect also on the other terms of the sums via the pair potential function. Hence, to use this method is a little doubtful.

4.2. Simple replacement

More sophisticated approach than omitting an outlier is to replace it by a new event. A new event can be added according to a Gibbs model, here the estimated MPL model. Adding a new event can be thought as one iteration of the birth-and-death process, a simulation procedure of Gibbs point processes (Ripley, 1977). The outlier is omitted (death) and a new location $\xi \in W$ is suggested (birth). We accept the new randomly chosen event with probability proportional to $\exp(-E(\xi, \varphi_W^-))$. If it is not accepted we suggest another one and repeat this until acceptance. A new event is added several times and the parameters are estimated. The mean values of these estimates are then taken as the final ones.

4.3. Replacement by using EM algorithm

We can also use the EM algorithm in adding new events and estimating the parameter θ (Tanner, 1991). Let the initial value θ^0 be the MPL estimate calculated from the original point pattern φ_W , and θ^i denote the current value of the parameter.

Each iteration of the algorithm consists of two steps, expectation step and maximization step. First, we compute the expectation

$$\xi_n = \int_W \xi \cdot p_{\theta^i}(\xi \mid \varphi_W^-) \, d\xi \tag{3}$$

which gives the new location with respect to the conditional distribution $p_{\theta^i}(\xi \mid \varphi_W^-)$ proportional to $\exp(-E_{\theta^i}(\xi, \varphi_W^-))$. Then, we maximize the logarithm of $PL(\theta; \varphi_W^- \cup \{\xi_n\})$ with respect to θ to obtain θ^{i+1} . The algorithm is iterated until $\|\theta^{i+1} - \theta^i\|$ is sufficiently small.

Instead of predicting the new location ξ_n we can also predict a new value for the logarithm of the PL function, i.e. we compute the expectation

$$\int_{W} \log PL(\theta; \varphi_{W}^{-} \cup \{\xi\}) \cdot p_{\theta^{i}}(\xi \mid \varphi_{W}^{-}) d\xi \tag{4}$$

and maximize it with respect to θ .

4.4. Outlier regarded as measurement error

In both of the replacement approaches above the new event is allowed to be added in the whole study area with the same probability (taken the model into account). However, more natural would be to regard the outlier as measurement error. We can assume e.g. that the distribution of the error is Gaussian, i.e.

$$p_{\sigma^2}(x_{i_o} \mid \varphi_W^- \cup \{\xi\}) = \frac{1}{2\pi\sigma^2} \exp(-\frac{1}{2\sigma^2} ||x_{i_o} - \xi||^2),$$

where σ^2 is the variance and ξ the "true" value of incorrectly observed x_{i_o} . In estimation the EM algorithm can be applied. Now the expectation (3) or (4) is computed with respect to the conditional distribution $p_{\theta^i}(\xi \mid \varphi_W)$ proportional to $\exp(-E_{\theta^i}(\xi, \varphi_W^-)) \cdot p_{\sigma^2}(x_{i_o} \mid \varphi_W^- \cup \{\xi\})$.

5. Application: data of amacrine cells

To find out influence of outlying observations on MPL estimates we studied the data consisting of locations of 152 amacrine cells within a rectangular region of a rabbit retinal ganglion cell layer (Diggle and Gratton, 1984). The cells form a regular pattern (see Figure 1) and therefore we chose an inhibitive interaction model

$$h(r) = 1 - \exp(-\beta(r - r_0)^2),$$
 (5)

where r is the distance between two events, β is the parameter to be estimated and $r_0 = 15 \mu \text{m}$ is the fixed hard-core radius.

In this data set interactions between the events are strong (for the details, see Särkkä, 1990). The MPL method gives $\beta = 0.0002$ (Särkkä, 1990) while the corresponding maximum likelihood (ML) estimate obtained by using Monte Carlo approximation is $\beta = 0.0006$ (Penttinen, 1986). The estimated ML model fits well to the data (Penttinen, 1986) and the MPL model differs from it (see Figure 4). In the following we study whether the overestimation of the MPL method could be due to its sensitivity to outlying observations.

5.1. Outlying events

To find out the events that have most influence on the MPL estimate we calculated values of the Tukey's sensitive curve for each event (Figure 2). In the data of amacrine cells there were two events that had stronger effect than the others (squares in Figure 3). They appeared to be isolated events.

To detect isolated events we calculated the exponential energy for each event. Two of the events of this data set differed from the others because of their small exponential energies (triangles in Figure 3).

Distance to the nearest neighbours and the number of neighbours were also calculated. None of the events differed from the others by means of distance to the nearest or second nearest neighbour but two of them did when the distances to the third nearest events were measured. The same two events stand out when we calculated the number of neighbours (stars in Figure 3). Here the events closer than $100\mu m$ (interaction radius of the ML model) were regarded as neighbouring events.

5.2. Estimation

By studying the amacrine cells data we found three pairs of isolated events depending on the method we applied in detecting them. The new value of the MPL estimate appears to be the same for each pair.

Applying the three first of the methods presented in Section 4, omitting, simple replacement or replacement by applying the EM algorithm, the new MPL estimate became 0.0005, quite close to the ML one (see Figure 4). To confirm that the model fits to the data we calculated Ripley's L function and the upper and lower envelopes from 99 simulations (Figure 5). The model seems to be reasonable.

Next, we replaced the outliers with new events added near the original ones. We assumed that the outliers have been measured incorrectly and assumed the Gaussian distribution (with several values of the variance) for the error. However, this approach did not help for this data set. The MPL estimate remained 0.0002.

5.3. Some other trials to correct MPL estimates

We studied the influence of outlying observations on MPL estimates and presented some methods how to take them into account in estimation. Another possibility is to try to make the method itself less sensitive to outlying observations.

The MPL estimation equations (1) and (2) are equalities of sample means. Since sample means are sensitive to outlying observations we replaced them by sample medians, and calculated the MPL estimate of β . The MPL estimate of β remained 0.0002.

The MPL method fails because of strong interactions between the events. It seems that the method does not sufficiently take into account dependencies between the events. To solve this problem we applied the second order MPL method presented by Mase (1992). It consideres pairs of events instead of single events and therefore we assumed that it better takes into account dependencies between the events. The second order PL function can be written as

$$PL_{2}(\theta; \varphi_{W}) = - \sum_{x \in \varphi_{W}} \sum_{y \in \varphi_{W} \setminus \{x\}} E_{2}(x, y, \varphi_{W} \setminus \{x, y\})$$
$$- \int_{W} \int_{W} \exp(-E_{2}(\xi, \eta, \varphi_{W})) d\xi d\eta,$$

where

$$E_2(x, y, \varphi_W) = 2\alpha + \sum_{z \in \varphi_W} \phi(||x - z||) + \sum_{z \in \varphi_W} \phi(||y - z||) + \phi(||x - y||)$$

is the energy needed to add the events x and y, $x \neq y$ to the configuration φ_W . By maximizing the second order PL function with respect to β we obtained the estimate 0.0002. The method seems to have the same problem of overestimation as the MPL method of first order.

Our model is a pairwise interaction model, higher order interactions are not allowed to exist. To confirm that the model is rich enough we allowed triplet interactions between the events. The local energy E can now be written as

$$E(x, \varphi_W) = \alpha + \sum_{y \in \varphi_W} \phi(||x - y||) + \frac{1}{2} \sum_{y \in \varphi_W} \sum_{z \in \varphi_W \setminus \{y\}} \psi(x, y, z),$$

where ψ is a triplet potential function. Following Fiksel (1988) we suggested two different choices for the triplet interaction function,

$$\psi_1(x, y, z) = \begin{cases} 0 & \text{if } s(x, y, z) < r_0 \\ 1 & \text{if } S(x, y, z) \ge R \\ d & \text{otherwise} \end{cases}$$

and

$$\psi_2(x,y,x) = \begin{cases} 0 & \text{if } s(x,y,z) < r_0 \\ 1 & \text{if } S(x,y,z) \ge R \\ \exp(-b\exp(-\gamma\alpha(x,y,z))) & \text{otherwise} \end{cases},$$

where s(x, y, z) is the minimal and S(x, y, z) the maximal distance between the events x, y and z, $\alpha(x, y, z)$ is the maximal angle of the triangle formed by x, y, and z, $r_0 = 15\mu$ m as before and $R = 150\mu$ m is the fixed interaction radius. We applied the MPL method to estimate the pairwise interaction parameter β and the parameter d of ψ_1 and the parameters b and γ of ψ_2 . Estimate of β remains 0.0002 and the estimates of the parameters of both ψ_1 and ψ_2 indicate that there are no triplet interactions between the events.

One reason for existency of isolated events may be that all events of the point pattern are not observed. We tested this possibility by adding one or two events near to the isolated events of amacrine cells data according to the estimated MPL model. However, this procedure seems not to help for this data set.

5.4. Light-off amacrine cells

In fact, the data of amacrine cells consist of two types of cells: 152 light-on and 142 light-off cells (Diggle, 1986). The data set that was studied in Sections 5.1–5.3 is the one of cells prosessing light-on information. Diggle (1986) concludes that the point patterns formed by light-on and light-off cells are generated by very similar mechanisms. Both patterns are regular (see Figure 6). Therefore a natural assumption is that the MPL method overestimates interaction between the light-off cells, too.

We estimated the parameter β of the model (5) with the hard-core radius $12\mu m$ (minimum interpoint distance) by the MPL and ML methods. The MPL and ML estimates we obtained are 0.0002 and 0.0004, respectively.

Furthermore, we applied the methods presented in Section 3 to detect possible outliers of the light-off data. We found one influential event and two outliers by using exponential energy marking, number of neighbours or distance to the nearest neighbour as a criteria (see Figure 7). The outliers detected by using the last two methods were the same. Applying the first three methods given in Section 4 we obtained a new MPL estimate of β in each case. If the influential event was considered an outlier we got 0.0004 as the new estimate. In the other cases the value of β became 0.0005. According to the L function study presented in Figure 8 both of the models, with $\beta = 0.0004$ and $\beta = 0.0005$, seem to be reasonable.

6. Discussion

The maximum pseudo-likelihood estimation method tends to overestimate

interaction if it is strong between the events. The aim of this paper was to find out the reason for the overestimation. The study of the light-on and light-off amacrine cells data showed that for these inhibitive data sets the overestimation was caused by isolated events. The result was reasonable since isolated events lie far from other events and therefore seem to inhibite the other events strongly.

We presented several methods to detect isolated events and methods to take them into account in estimation. First, we suggested to omit the isolated events in estimation. However, omitting them is questionable method since the number of events changes. The events that are omitted interact with the other events and have effect on them, e.g. on the values of the local energies. Therefore a better way is to replace the outlying events by new events added according to some rule. The replacement methods given in Section 4 can be applied also when values of Tukey's sensitive curve are calculated.

Only the influence of single outliers was studied here, no attention was paid to groups of outliers. Similar kind of problems may appear if, for example, the point pattern is regular in general but on one part of the study area it is clustered.

We applied the methods suggested for detecting outliers and estimating given the outliers only to one data set. However, this study gives some empirical tools to study robustness of the maximum pseudo-likelihood method.

Acknowledgements

I would like to thank Antti Penttinen, Michel Goulard and Peter Diggle for their valuable comments.

References

- Besag, J. (1978). Some methods of statistical analysis for spatial data. *Bull. Internat. Statist. Inst.*, 47, 77–92.
- Diggle, P. J. (1986). Displaced amacrine cells in the retina of a rabbit: analysis of a bivariate spatial point pattern. *Journal of Neuroscience Methods* 18, 115–125.
- Diggle, P. J. & Gratton, R. (1984). Monte Carlo methods of inference for implicit statistical models. J. Roy. Statist. Soc. B 46, 193–227.
- Diggle, P.J., Fiksel, F., Grabarnik, P., Ogata, Y., Stoyan, D. & Tanemura, M. (1994). On parameter estimation for pairwise interaction point processes. *Int. Statist. Rev.* **62**, 99–117.
- Fiksel, T. (1988). Estimation of interaction potentials of Gibbsian point processes. *statistics* 19 1, 77–86.
- Huber, P.J. (1981). Robust Statistics. New York: Wiley.
- Jensen, J.L. & Møller, J. (1991). Pseudolikelihood for exponential family models of spatial point processes. *The Ann. of Appl. Prob.* 3, 445–461.
- Mase, S. (1992). Maximum pseudo-likelihood estimator of second order for Gibbsian spatial point process and its strong consistency. Technical Report 308, Hiroshima University.

- Penttinen, A. (1986). The estimation of pair potentials of spatial point patterns. Statistic seminar, Universities of Glasgow and Strathclyde, Scotland (unpublished).
- Ripley, B.D. (1977). Modelling spatial patterns. J. Roy. Statist. Soc. B 39, 172–212.
- Ripley, B.D. (1988). Statistical inference for spatial processes. Cambridge University press.
- Särkkä, A. (1990). Applications of Gibbs point processes: pseudo-likelihood estimation method with comparisons. Reports from the Department of Statistics, University of Jyväskylä, 4.
- Stoyan, D., Kendall, W.S. & Mecke, J. (1987). Stochastic geometry and its applications. Chichester-New York-Brisbane-Toronto-Singabore: John Wiley & Sons.
- Stoyan, D. & Grabarnik, P. (1991). Second-order characteristics for stochastic structures connected with Gibbs point processes. *Math. Nachr.*, **151**, 95–100.
- Tanner, M.A. (1991). Tools for statistical inference. Berlin-Heidelberg: Springer-Verlag.
- Wartenberg, D. (1990). Exploratory spatial analyses: outliers, leverage points, and influence functions. In: *Proceedings from a Symposium of Spatial Statistics: Past, Present and Future* (ed. D.A. Griffith), Institute of Mathematical Geography, Syracuse University, 133–162.

Labels of figures

- Figure 1. Locations of 152 amacrine cells within a $1070\mu\text{m}\times600\mu\text{m}$ rectangular region.
- Figure 2. Influence of the events on the MPL estimate of β . The size of the circles are related to the values of Tukey's sensitive curve. Events with small, unessential influence are denoted by crosses only.
- Figure 3. Locations of amacrine cells: squares, triangles and stars indicate outlying events with respect to the values of Tukey's sensitive curve, exponential energy and number of neighbours, respectively.
- Figure 4. Interaction functions: corrected MPL estimate (solid), ML estimate (dashed) and original MPL estimate (dotted).
- Figure 5. L function calculated from the data (solid) and the upper and lower envelopes from 99 simulations of the corrected MPL model (dashed).
- Figure 6. Locations of 152 amacrine cells light-on cells (crosses) and 142 light-off cells (circles) within a $1070\mu\text{m}\times600\mu\text{m}$ rectangular region.
- Figure 7. Locations of light-off amacrine cells: squares, triangles and stars indicate outlying events with respect to the values of Tukey's sensitive curve, exponential energy and number of neighbours, respectively.
- Figure 8. L function calculated from the data of light-off cells (solid) and the upper and lower envelopes from 99 simulations of the model (5) with $\beta = 0.0004$ (dashed) and $\beta = 0.0005$ (dotted).

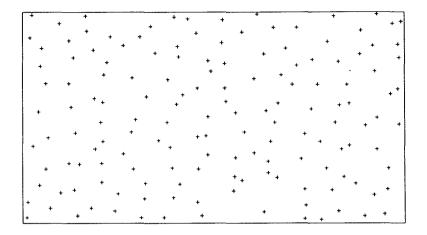


Figure 1

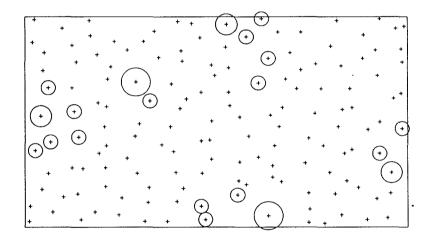


Figure 2

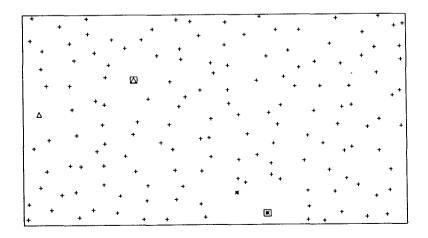
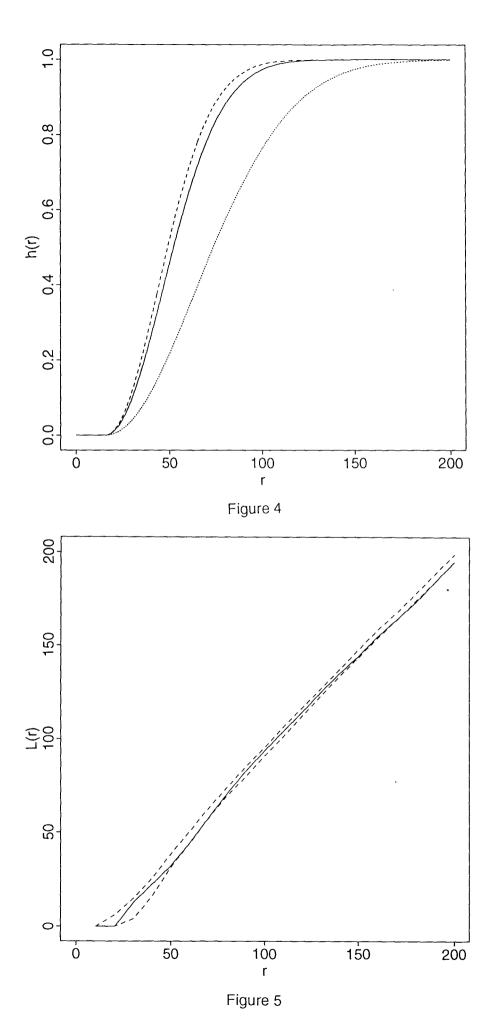


Figure 3



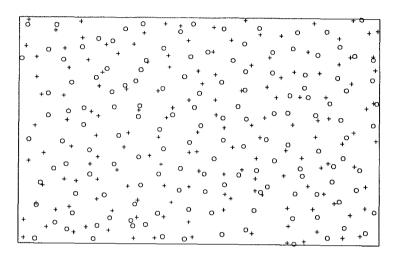


Figure 6

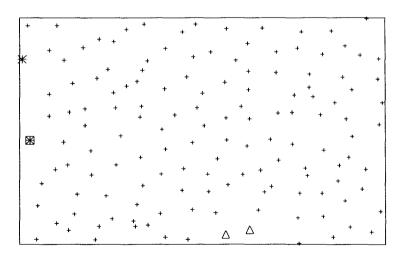
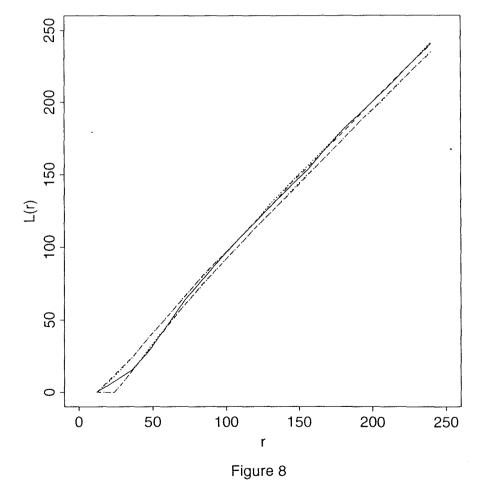


Figure 7



Research Report

1995:1	Arnkelsdottir, H	Surveillance of rare events. On evaluations of the sets method.
1995:2	Sveréus, A	Detection of gradual changes. Statistical methods in post marketing surveillance.
1995:3	Ekman, C	On second order surfaces estimation and rotatability.
1996:1	Ekman, A	Sequential analysis of simple hypotheses when using play-the-winner allocation.
1996:2	Wessman, P	Some principles for surveillance adopted for multivariate processes with a common change point.
1996:3	Frisén, M. & Wessman, P	Evaluations of likelihood ratio methods for surveillance.
1996:4	Wessman, P & Frisén, M	Evaluation of univariate surveillance procedures for some multivariate problems.