KYBERNETIKA — VOLUME 36 (2000), NUMBER 4, PAGES 455-464

A MAXIMUM LIKELIHOOD ESTIMATOR OF AN INHOMOGENEOUS POISSON POINT PROCESS INTENSITY USING BETA SPLINES

PAVEL KREJČÍŘ

The problem of estimating the intensity of a non-stationary Poisson point process arises in many applications. Besides non parametric solutions, e. g. kernel estimators, parametric methods based on maximum likelihood estimation are of interest. In the present paper we have developed an approach in which the parametric function is represented by two-dimensional beta-splines.

1. INTRODUCTION

The problem of estimating the intensity of non-stationary two-dimensional Poisson point process is encouraged by the study of producing a risk map of tick-borne encephalitis and Lyme borreliosis in central Bohemia described in [10] and [4]. The data consisted of a record of the locations of events of the mentioned two diseases over last thirty years. The task was to estimate the risk map of the diseases using the available data. Recently, a Bayesian approach has been applied to this problem by [9].

The results given in [10] were carried out using a non-parametric kernel estimator, where the kernel function chosen was the density of Gaussian distribution. There was a problem with choice of the parameter called 'bandwidth'. The parameter estimated by using an optimalisation criteria provided a too smooth density estimator. Choosing the value approximately half of the theoretically optimal bandwidth, the results were more acceptable from practical point of view.

The problem of estimating the intensity of inhomogeneous point processes has been intensively studied over several last years. The basic concept is described in [3], further investigation concerning the maximum likelihood method was reported in [2, 5].

We have developed a method of estimating the intensity of a non-stationary Poisson point process based on the two-dimensional quadratic beta-splines and using the maximum likelihood method for the estimation of the parameters. The method can be applied after relaxing the assumption of Poisson property, too: in that case we use the approach of pseudo-likelihood as described in [1].

2. MAXIMUM LIKELIHOOD ESTIMATOR

The maximum likelihood estimator of parameters of a Poisson point process intensity function was recommended by [2]. The intensity function is a parametric function $\lambda_\theta(x)$, where $x \in B$, B is a Borel subset of $\mathbb{R}^d$, $\theta \in \Theta$ is the unknown parameter, Θ is the parametric space. The log-likelihood function is then of the form

$$L(\theta) = L(\theta; x_1, \ldots, x_k) = \sum_{i=1}^{k} \log \lambda_\theta(x_i) - \int_B \lambda_\theta(x)\, \mathrm{d}x, \tag{1}$$

where $x_1, \ldots, x_k$ is the sample realization of the Poisson point process.

Berman and Turner in [2] suggest a couple of parametric functions for λ, basically combinations of polynomials and exponential functions. Although the one dimensional results presented here are quite convincing and the possible extension to higher dimensions is proposed to be straightforward, we decided to use a different type of parametric function for the intensity prototype. The function is based on the quadratic beta splines. The main advantage of this choice is that the initial estimator of the parameters is graphical and provides a good starting position for the inevitable iteration process.

For a better understanding we have decided to start with one dimensional case and then to explain the two dimensional estimator. Thus the popular phrase that the extension to higher dimension is straightforward is avoided here.

2.1. Quadratic beta-splines

Let us explain first the nature of the quadratic beta-spline. The usage of spline curves in mathematics has been discussed many times, see e.g. [7] in numerical mathematics.

Basically, let $\boldsymbol{x} = (x, y) \in \mathbb{R}^2$ be a two dimensional point. The single quadratic spline is defined by three points. Namely by two end-points, say $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, and one control-point, say $\boldsymbol{c}$. The spline is then

$$f(t) = \boldsymbol{x}_1(1-t)^2 + 2\boldsymbol{c}t(1-t) + \boldsymbol{x}_2 t^2, \tag{2}$$

$t \in (0, 1)$, which assures that the curve starts in the point $\boldsymbol{x}_1$, ends in the point $\boldsymbol{x}_2$. $\boldsymbol{c} - \boldsymbol{x}_1$ is the tangent vector of the curve at the point $\boldsymbol{x}_1$, $\boldsymbol{x}_2 - \boldsymbol{c}$ is the tangent vector of the curve at the point $\boldsymbol{x}_2$.

The usual quadratic beta-spline consists from n single beta-splines joining consequentively to each other in order to the whole curve is continuous and smooth. The quadratic beta-spline is then defined by two end-points and a sequence of n control points. For convenience, let the start-point be $\boldsymbol{x}_0$, the end-point $\boldsymbol{x}_n$ and the control points $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_n$. The ith quadratic spline is defined by end-points $\boldsymbol{x}_{i-1}$, $\boldsymbol{x}_i$ and the control-point is $\boldsymbol{c}_i$. The usual convention is that the intermediate end-points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{n-1}$ are calculated as

$$\boldsymbol{x}_i = \frac{\boldsymbol{c}_i + \boldsymbol{c}_{i+1}}{2}, \tag{3}$$

which is convenient but not necessary for the smoothness of the beta-spline. In fact, it is sufficient when the ith end-point lies anywhere on the line between ith and $(i+1)$st control-points.

2.2. One dimensional estimator

The equation (1) was used in [2] in the form with the right hand side integral replaced by a numerical approximation. We will show further, how beta splines can simplify the situation.

The maximum of (1) can be found from the series of equations:

$$\frac{\partial L(\theta)}{\partial \theta_j} = \sum_{i=1}^{k} \frac{\partial \lambda_\theta(x_i)}{\partial \theta_j} \frac{1}{\lambda_\theta(x_i)} - \frac{\partial}{\partial \theta_j} \int_B \lambda_\theta(x)\, \mathrm{d}x = 0, \quad j = 0, \dots, n+1, \tag{4}$$

where $n+2$ is the number of parameters to be estimated. One advantage of the spline method is that the partial derivatives of λ_θ do not depend on the estimated parameters at all. Another advantage of using splines is that the integral can be calculated directly in one dimensional case and, with some care, in higher dimensions as well.

Let us show a simple example. The simplest one is estimating the intensity of an inhomogeneous Poisson point process on the interval (a, b) using just one spline. Then the estimates of three parameters, say θ_0, θ_1 and θ_2 are needed. Let the parameter θ_1 be the control-point and the remaining two be the end-points of the spline. Naturally, the end points lie at the ends of the interval, we demand

$$\begin{aligned} \lambda_\theta(a) &= \theta_0, \\ \lambda_\theta(b) &= \theta_2 \end{aligned}$$

and then

$$\lambda_\theta(x) = \theta_0(1-t)^2 + 2\theta_1 t(1-t) + \theta_2 t^2, \tag{5}$$

where $x \in (a, b)$, $x = a(1-t) + bt$. The derivatives of the intensity function are:

$$\begin{aligned} \frac{\partial \lambda_\theta(x)}{\theta_0} &= (1-t)^2, \\ \frac{\partial \lambda_\theta(x)}{\theta_1} &= 2t(1-t), \\ \frac{\partial \lambda_\theta(x)}{\theta_2} &= t^2. \end{aligned}$$

The choice of the initial estimators is natural. Let $\Phi = \{x_1, \dots, x_k\}$ be the underlying point process. Let $\#(B \cap \Phi)$ denote the number of the process points in a Borel set B. Then the initial parameters are chosen as

$$\theta_0 = \frac{\#(\Phi \cap (a, a + (b-a)/4))}{(b-a)/4},$$

$$\theta_1 = \frac{\#(\Phi \cap (a+(b-a)/4, b-(b-a)/4))}{(b-a)/2},$$
$$\theta_3 = \frac{\#(\Phi \cap (b-(b-a)/4, b))}{(b-a)/4}.$$

The integral in the formula (1) for calculating the maximum likelihood estimator can be evaluated as

$$\int_a^b \lambda_\theta(x)\,\mathrm{d}x = \begin{cases} (b-a)\frac{\theta_0+\theta_1+\theta_2}{3}, & n=1 \\ (b-a)\dfrac{\theta_0+2\theta_1+2\theta_2+\theta_3}{6}, & n=2 \\ (b-a)\dfrac{\theta_0+2\theta_1+3\sum_{i=3}^{n-2}\theta_i+2\theta_{n-1}+\theta_n}{3n}, & n>2. \end{cases}$$

To calculate the partial derivatives of the integral is straightforward.

For illustration of this process we tried to simulate a simple process on the line between points 0 and 10. The test process consisted of 20 points located as shown in the Table 1.

Table 1. The test point process.

No.	1	2	3	4	5	6	7	8	9	10
Pos.	3.21	3.45	4.12	5.28	6.11	6.12	6.56	6.82	7.11	7.22
No.	11	12	13	14	15	16	17	18	19	20
Pos.	7.98	7.99	8.98	9.01	9.10	9.13	9.23	9.28	9.34	9.52

Then we calculated the intensity as a single spline first. The iteration process is shown in Figure 1. The iteration was carried out by the method, where all the partial derivatives have been calculated and then the parameters shifted in the direction of its partial derivative multiplied by an appropriate small constant w. In our case we chose $w = 0.01$. The maxima for the log-likelihood function was reached after 4739 steps.

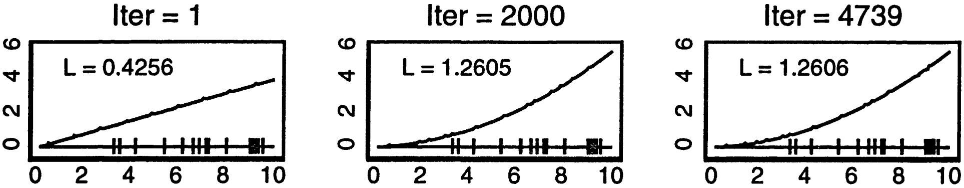

Fig. 1. Three steps of calculating the intensity as a single spline.

The graphs in Figure 1 show the estimated log-likelihood value L, so the difference in the last 2000 steps is not big. For practical purposes it would be probably sufficient to take the estimator after 1000 or 2000 steps.

Increasing the number of splines in the beta-spline, the theory is not more complicated. Given $x \in (a, b)$ and n the number of the splines, the first step is to calculate two variables, which can be done as follows:

$$\begin{aligned} i &= \left[\frac{nx}{b-a}\right] \\ t &= \frac{nx}{b-a} - i \\ &\text{if } i = n \text{ then } t = 1 \\ &\text{else } i = i + 1; \end{aligned} \tag{6}$$

The n means number of the splines in the beta-spline. Thus i denotes to which spline the point x belongs, and t means the offset from the beginning of the spline normalized to the interval $(0, 1)$. The value of the intensity is then given by

$$\lambda_\theta(x) = \begin{cases} \theta_0(1-t)^2 + 2\theta_1 t(1-t) + \frac{1}{2}(\theta_1 + \theta_2)t^2, & i = 1 \\ \frac{1}{2}(\theta_{i-1} + \theta_i)(1-t)^2 + 2\theta_i t(1-t) + \frac{1}{2}(\theta_i + \theta_{i+1})t^2, & i = 2, \ldots, n-1 \\ \frac{1}{2}(\theta_{n-1} + \theta_n)(1-t)^2 + 2\theta_n t(1-t) + \frac{1}{2}(\theta_n + \theta_{n+1})t^2, & i = n. \end{cases} \tag{7}$$

Note that the parameters θ_0 and θ_{n+1} correspond to the start and end values of the intensity respectively. The other points between are control points.

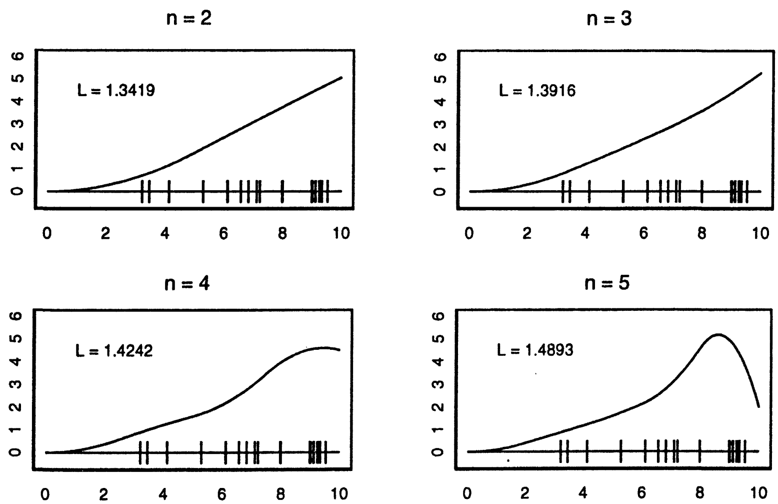

Fig. 2. The estimators of the intensity function for beta-splines containing more than one spline.

The calculation of intensity estimators using higher beta-splines is shown in Figure 2. The splines have been estimated for 2, 3, 4 and 5 divisions of the interval $(0, 10)$. The estimated maximum log-likelihood is shown in the graphs as 'L', as well as in Figure 1. Obviously the value of log-likelihood increases with the number of partitions of the interval. It would probably not decrease until the number of points of the process is reached. To obtain an optimal value for n, some penalty function should be introduced.

A note concerning the iteration process follows. The end-points should never be negative. On the other hand, the negative control point does not necessarily mean that some part of the spline is negative. But it would be certainly good to introduce a rule, by which the negative control point is reset to 0 when the sum with one of the neighbour control points is less than zero.

2.3. Two dimensional estimator

The extension of the quadratic beta-spline method of an inhomogeneous Poisson point intensity estimation from dimension one to dimension two is really not difficult when the window is a rectangle. The position of the parameters to be estimated is schematically shown in Figure 3. The window, which is a rectangular set B is partitioned into $m \times n$ subsets denoted as B_{ij}. They are emphasized be the solid rectangles in Figure 3.

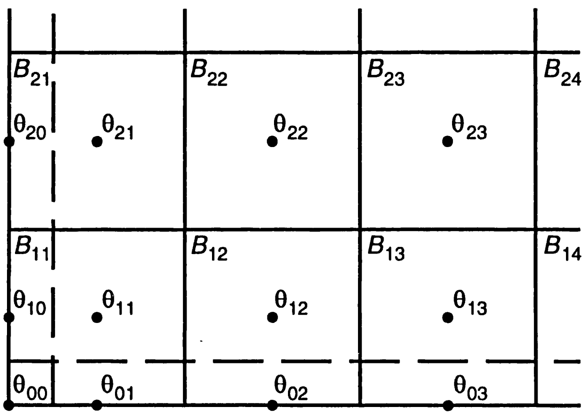

Fig. 3. The position of the parameter in the dimension two.

The parameters in the centers of the sets B_{ij} are the central control-points. The parameters on the boundary are the control-points, the parameters at the four corners are the end-points. The initial parameters are estimated as the number of the points of the process in the surrounding rectangle divided by the area of the rectangle. The rectangles correspond to the sets B_{ij} for the very central control-points. For the parameters on the boundary and their neighbouring central control-points, the rectangle is smaller as shown in Figure 3, marked by the dashed lines.

A simple two dimensional quadratic spline, for example for the square B_{22} is then calculated as

$$\lambda_\theta(x,y) = c_{11}(1-s)^2(1-t)^2 + 2c_{12}s(1-s)(1-t)^2 + c_{13}s^2(1-t)^2$$

$$+2c_{21}(1-s)^2t(1-t)+4c_{22}s(1-s)t(1-t)+2c_{23}s^2t(1-t)$$
$$+c_{31}(1-s)^2t^2+2c_{32}s(1-s)t^2+c_{33}s^2t^2,$$

where s and t play the same role as t in (6) and

$$\begin{aligned}
c_{11} &= \tfrac{1}{4}(\theta_{11}+\theta_{12}+\theta_{21}+\theta_{22}) & c_{23} &= \tfrac{1}{2}(\theta_{22}+\theta_{23})\\
c_{12} &= \tfrac{1}{2}(\theta_{12}+\theta_{22}) & c_{31} &= \tfrac{1}{4}(\theta_{21}+\theta_{22}+\theta_{31}+\theta_{32})\\
c_{13} &= \tfrac{1}{4}(\theta_{12}+\theta_{13}+\theta_{22}+\theta_{23}) & c_{32} &= \tfrac{1}{2}(\theta_{22}+\theta_{32})\\
c_{21} &= \tfrac{1}{2}(\theta_{21}+\theta_{22}) & c_{33} &= \tfrac{1}{4}(\theta_{22}+\theta_{23}+\theta_{32}+\theta_{33})\\
c_{22} &= \theta_{22}
\end{aligned}$$

From the computational point of view, some attention is required when calculating all the partial derivatives correctly, but it is just a technical problem. Also the dummy points can be removed easily as well as in the one dimensional case, since the integral in the log-likelihood function follows the rule

$$\int_B \lambda_\theta(x)\,\mathrm{d}x = \frac{(b_x-a_x)(b_y-a_y)}{9mn}\begin{pmatrix} \theta_{00}+2\theta_{01}+3\theta_{02}+3\theta_{03}+3\theta_{04}+\dots \\ 2\theta_{10}+3\theta_{11}+6\theta_{12}+6\theta_{13}+6\theta_{14}+\dots \\ 3\theta_{20}+6\theta_{21}+9\theta_{22}+9\theta_{23}+9\theta_{24}+\dots \\ 3\theta_{30}+6\theta_{31}+9\theta_{32}+9\theta_{33}+9\theta_{34}+\dots \\ \dots \end{pmatrix} \tag{8}$$

where (a_x, b_x) are the bounds in x-direction, (a_y, b_y) are the bounds in y-direction. Of course, a modification of the above formula is necessary if one of the partition numbers m, n is less than three.

A more complicated situation occurs when the window is a non-rectangular area, which is the case in our application and in practice probably the most frequent case. The summation part of the log-likelihood equation is calculated without changes. The problem is how to estimate the integral. There would be again possible some solution using dummy points, but we looked for a better approximation of the integral.

When the window is a non-rectangular set, say A, then we take the smallest rectangle B containing the whole set A. The rectangle B is partitioned into sub-rectangles corresponding to the estimated parameters. The task is to calculate the integral of the intensity using the sub-rectangles. The spline over one sub-rectangle is calculated using 9 points, which are calculated from the beta-spline parameters as explained earlier. When the sub-rectangle is all contained in the set A, then the integral over this sub-rectangle is simply the area of the sub-rectangle multiplied by the average of all the nine points defining the spline.

What shall we do when the sub-rectangle is not all contained in the window A? Of course there are some methods how to estimate the integral anyway, but we need the estimator of the integral to be a linear combination of the nine points defining the spline. We have developed a simple rule how to estimate the integral. Say, we want to calculate the integral over the rectangle B_{ij}. First we need to estimate the area of $A \cap B_{ij}$. Then for many purposes we need to have a rule how to find out which point lies inside the set A and which lies outside. We will use this routine to determine which of the nine points lie inside A. Then we calculate weights for each particular configuration in the following way.

The base is the situation, when only one point from nine lies inside the set A. There are three basic possibilities. The point is a corner, an edge or the center. For each situation, we have established the weights shown in Figure 4. If more than one point is contained in the window, the weight for each point is the sum of weights created by the superposition of the elementary situation with appropriate rotations.

a)

1369	370	37
370	37	10
37	10	1

b)

962	1628	962
260	440	260
26	44	26

c)

676	1144	676
1144	1936	1144
676	1144	676

Fig. 4. The weights for the three basic situations of the window intersection with the rectangle. a) the corner is in the window, b) the edge is in the window, c) the center is in the window.

Then all nine weights are normalized to the sum 1 and the integral is estimated as the weighted sum of the nine points multiplied by the area of $A \cap B_{ij}$. This rule does not seem to be simple, it really requires quite complicated programing. On the other hand, given the partition of the rectangle B, the weights are calculated once and then just appear in the derivatives of the integral.

3. APPLICATION

The described two dimensional method has been applied in the mapping of the disease-risk study. The problem of mapping risk of tick-borne encephalitis and Lyme borreliosis in Central Bohemia region has been introduced in [10]. The data consist of cases of both the diseases over last thirty years.

The result of calculating an inhomogeneous Poisson intensity estimator in the dimension two using the method of quadratic beta-splines is shown in Figure 5.

The general theory of estimation of the disease risk was discussed by Stern et al [9] and then by Machek [6]. The problem of tick-borne disease-risk mapping has been solved using Bayesian approach in [8], where an additional information about the forest density in Central Bohemia region is also taken into the account.

There is another additional information included in [10], namely the population density in Central Bohemia region. In order to calculate the correct map of risk, Zeman takes into the account only the cases, when the inhabitants have been infected in their living place or in its close neighbourhood. This approach, together with the population density, allows to estimate the total risk as explained in [6], i. e. if the population density is estimated as $p(\boldsymbol{x})$ the total risk is given by

$$r(\boldsymbol{x}) = C\frac{\lambda(\boldsymbol{x})}{p(\boldsymbol{x})}, \qquad \boldsymbol{x} \in B, \tag{9}$$

where C is a normalisation constant.

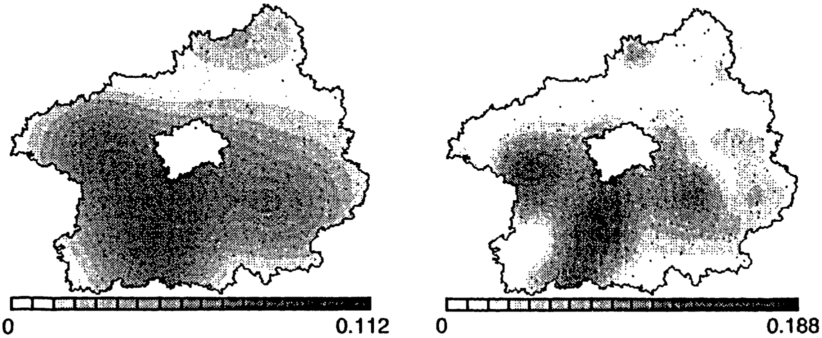

Fig. 5. Two intensities of tick-borne encephalitis cases in central Bohemia calculated by the spline method. The first intensity is calculated using the grid 6 × 5 splines, the second is produced by the grid 12 × 10. Both intensities are calculated by ten steps of iterations. The gray scales show the intensities, the numbers correspond to estimated mean number of cases per square kilometer.

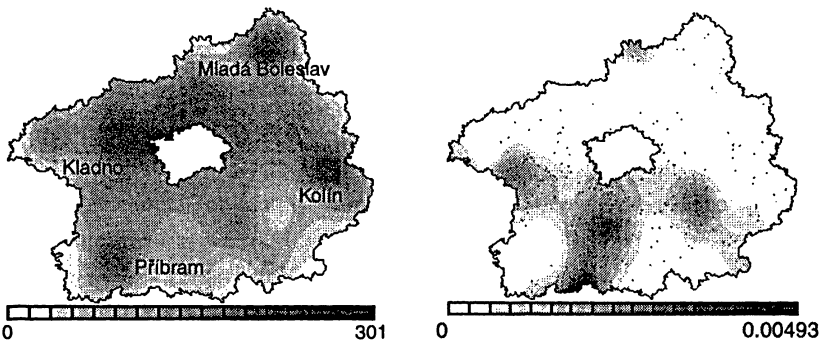

Fig. 6. The map of estimated population density (left) and risk of tick-borne encephalitis (right) in central Bohemia calculated by the spline method. The maps were produced by the grid 12 × 10 splines. The gray scale of population density corresponds to the number of inhabitants per square kilometer. The big towns of central Bohemia region are matched with the dark spots. The empty place in the middle of the region is the capital Praha, which was not considered in this study. The unit of the risk map is the probability of developing the disease.

The population density $p(x)$ can be introduced into the present method as follows (Figure 6, left). The data consisted of spatial point (coordinates of a town or a village) and corresponding number of inhabitants. So it could be treated as a non-simple process or as a marked process. In fact, there is no limitation in the presented method, as the log-likelihood function (1) can be modified by adding the mark in front of the logarithm in the sum. The method described in this paper allows to calculate after some modification also the risk map. The result is shown in Figure 6 (right).

ACKNOWLEDGEMENT

The research was supported by the Grant Agency of the Czech Republic under projects no. 201/98/0090 and 201/99/0269.

(Received November 9, 1999.)

REFERENCES

[1] A. Baddeley and R. Turner: Practical Maximum Pseudolikelihood for Spatial Point Patterns. Research Report No. 1998/16, Department of Mathematics and Statistics, University of Western Australia 1998.

[2] M. Berman and T. R. Turner: Approximating point process likelihoods with glim. Appl. Statist. *41* (1992), 1, 31–38.

[3] D. J. Daley and D. Vere-Jones: An Introduction to the Theory of Point Processes. Springer, New York 1988.

[4] M. Daniel, J. Kolář, P. Zeman, K. Pavelka and J. Sádlo: Prediction of sites of enhanced risk of tick *Ixodes ricinus* attack and developping tick-borne encephalitis in Central Bohemia using satelite data. Epidemiol. Mikrobiol. Imunol. *47* (1998), 1, 3–11.

[5] C. J. Geyer: Likelihood inference for spatial point processes. In: Stochastic Geometry, Likelihood and Computation (O. E. Barndorff-Nielsen et al, eds.), Chapman and Hall, London 1999, pp. 79–140.

[6] J. Machek: On statistical models for the mapping of disease-risk. In: International Conference on Stereology, Spatial Statistics and Stochastic Geometry (V. Beneš, J. Janáček and I. Saxl, eds.), Union of the Czech Mathematicians and Physicists, Prague 1999, pp. 191–196.

[7] G. I. Marčuk: Methods of Numerical Mathematics. Academia, Prague 1987.

[8] M. Mašata: Assessment of risk infection by means of a bayesian method. In: International Conference on Stereology, Spatial Statistics and Stochastic Geometry (V. Beneš, J. Janáček and I. Saxl, eds.), Union of the Czech Mathematicians and Physicists, Prague 1999, pp. 197–202.

[9] S. H. Stern and N. Cressie: Inference for extremes in disease mapping. In: Methods of Disease Mapping and Risk Assesment for Public Health Decision Making (A. Lawson et al, eds.), Wiley, New York 1999.

[10] P. Zeman: Objective assesment of risk maps of tick-borne encephalitis and lyme borreliosis based on spatial patterns of located cases. Internat. J. Epidemiology *26* (1997), 5, 1121–1129.

Mgr. Pavel Krejčíř, Ph.D., Institute of Computer Science – Academy of Sciences of the Czech Republic, Pod vodárenskou věží 2, 182 07 Praha 8. Czech Republic.
e-mail: krejcir@cs.cas.cz