Ann. Inst. Statist. Math.
Vol. 40, No. 1, 29-39 (1988)

LIKELIHOOD ANALYSIS OF SPATIAL INHOMOGENEITY FOR MARKED POINT PATTERNS

YOSIHIKO OGATA AND KOICHI KATSURA

The Institute of Statistical Mathematics, 4-6-7, Minami-Azabu, Minato-ku, Tokyo, Japan

(Received November 12, 1986; revised August 17, 1987)

Abstract. An objective method is developed for estimations of both spatial intensity of the point locations and spatial variation of a characteristic parameter of the distributions for the attached marks. Its utility is demonstrated by means of analyses of seismological and ecological data sets.

Key words and phrases: Marked point patterns, intensity, distributions of marks, *B*-splines, smoothing, penalized likelihood, objective Bayesian method, magnitude frequency, *b*-value.

1. Introduction

For a data set of marked point pattern we have developed a method for estimating both the intensity rate of point locations and the spatial variation of a parameter which characterizes a distribution for the attached marks. The exponential form of 2-dimensional cubic *B*-spline function is used for estimating the both of these. Generally quite many parameters are required for representing a surface by the combination of *B*-spline bases, so that the maximum likelihood estimate for such case usually produces a rapidly fluctuated surface. Thus we had to resolve two conflicting aims in surface estimation, which are to produce a good fit to the data but to avoid too much rapid local variation. A log likelihood function is a measure of the goodness of fit, while a measure of the rapid local variation of a surface can be given by roughness penalties such as the integrated squared first or second derivatives. The weights of the penalties are considered as the hyperparameters of the prior probability of parameters in the likelihoods. The optimal hyperparameters are determined by using the objective Bayesian procedure suggested and developed by Good (1965) and Akaike (1979); that is, evaluating the type II maximum likelihood or minimizing Akaike Bayesian Information Criterion (ABIC). Since the calculation of the integral of the posterior with respects to parameters for the non-Gaussian likelihood is not easy, we approximate the posterior to a multivariate Gaussian distribution, following Ishiguro and Sakamoto (1983).

2. A statistical model for marked point pattern

We shall be concerned in this paper with a set of data in the form of point locations $X=\{(x_i, y_i);\ i=1, 2,\ldots, N\}$ with attached scalar marks $Z=\{z_i;\ i=1, 2,\ldots, N\}$ in a bounded planar region A. Assume that X is an inhomogeneous Poisson pattern and that z_i in Z are mutually independent but dependent only on the location (x_i, y_i). Let the intensity $\lambda(x, y, z)$ of the process is defined by

$$\text{(2.1)} \qquad \begin{aligned} &\mathrm{Prob}\{\text{event in a volume } \Delta x \times \Delta y \times \Delta z\} \\ &\quad = \lambda(x, y, z)\Delta x \Delta y \Delta z + o(\Delta x \Delta y \Delta z)\ , \end{aligned}$$

for small Δx, Δy and Δz. Then we can usually write it by

$$\text{(2.2)} \qquad \lambda(x, y, z) = \mu(x, y) f(z \mid x, y)\ ,$$

where $\mu(x, y)$ is the intensity for the point location in A, and $f(z \mid x, y)$ is the conditional probability density distribution of the marks.

Parameterizing μ and f by the vectors $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$, respectively, we have the following log likelihood function for $\theta=(\boldsymbol{\sigma}, \boldsymbol{\tau})$

$$\text{(2.3)} \qquad \begin{aligned} \log L(\theta) &= \sum_{i=1}^{N} \log \lambda_\theta(x_i, y_i, z_i) - \int_{-\infty}^{\infty} \int_A \lambda_\theta(x, y, z)\, dx dy dz \\ &= \log L_1(\sigma) + \log L_2(\tau)\ , \end{aligned}$$

where

$$\text{(2.4)} \qquad \log L_1(\boldsymbol{\sigma}) = \sum_{i=1}^{N} \log \mu_\sigma(x_i, y_i) - \int_A \mu_\sigma(x, y)\, dx dy\ ,$$

and

$$\text{(2.5)} \qquad \log L_2(\boldsymbol{\tau}) = \sum_{i=1}^{N} \log f_\tau(z_i \mid x_i, y_i)\ .$$

If the parameter vectors $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$ have no common components, then their maximum likelihood estimates are obtained independently by maximizing (2.4) and (2.5), respectively.

One of our interests is to estimate the intensity rate $\mu_\sigma(x, y)$ of the point locations. Another interest is to know the spatial variation of the distribution $f_\tau(z \mid x, y)$. Specifically, a simple example can be the exponential distribution, that is

$$f_\tau(z|x, y) = \beta_\tau(x, y)e^{-\beta_\tau(x, y)z} , \tag{2.6}$$

where the parameter β of the exponential distribution is considered to be dependent on the location (x, y). Since the both $\mu_\sigma(x, y)$ and $\beta_\tau(x, y)$ are positive quantity, these are expressed in the exponential form,

$$C \cdot \exp\{h(x, y: \boldsymbol{c})\} , \tag{2.7}$$

of the following 2-dimensional spline function

$$h(x, y|\boldsymbol{c}) = \sum_{i=1}^{I+3} \sum_{j=1}^{J+3} c_{ij} F_i(x) G_j(y) , \tag{2.8}$$

where C is a suitable constant, $\boldsymbol{c}=\{c_{ij}\}$ are coefficient parameters identical with either σ or τ, and the functions F_i and G_j are cubic B-spline bases with equally spaced knots; see Section 4 for the detailed description. Sometimes the coefficients c_{ij} will be ordered lexicographically to treat $\boldsymbol{c}$ as a vector, that is, $\boldsymbol{c}=\{c_k\}$ such that $k=i(J+3)+j$.

3. The objective Bayesian method

Since quite many parameters $\{c_{ij}\}$ are required for representing μ_σ and β_τ, the maximum likelihood estimates for (2.4) and (2.5), respectively, usually produce a rapidly fluctuate surface. To overcome this difficulty we have to quantify the competition between the two conflicting aims in surface estimation, which are to produce a good fit to the data but to avoid too much rapid local variation. A measure of the rapid local variation can be given by a roughness penalty. Various roughness penalties for the continuous functions have been suggested and used (see Titterington (1985), for example), but we here use the combination of the followings (see Inoue (1986) and Meinguet (1979)) for the function h in (2.7) defined on an area $A \subset R^2$;

$$\Phi_1(h) = \int_A \left\{ \left(\frac{\partial h}{\partial x}\right)^2 + \left(\frac{\partial h}{\partial y}\right)^2 \right\} dxdy , \tag{3.1}$$

and

$$\Phi_2(h) = \int_A \left\{ \left(\frac{\partial^2 h}{\partial x^2}\right)^2 + 2\left(\frac{\partial^2 h}{\partial x \partial y}\right)^2 + \left(\frac{\partial^2 h}{\partial y^2}\right)^2 \right\} dxdy . \tag{3.2}$$

Using these we consider the penalized log likelihood (Good and Gaskins (1971)),

$$\log L - \{w_1 \Phi_1(h) + w_2 \Phi_2(h)\} , \tag{3.3}$$

where $\log L$ stands for either (2.4) or (2.5). To obtain the suitable weights w_1 and w_2 above we employ a Bayesian interpretation; that is to say, the sum of penalties in (3.3) are considered to be proportionate to the logarithm of prior distribution $\pi(\boldsymbol{c}|w_1, w_2)$ characterized by the hyperparameters w_1 and w_2. To avoid the difficulty in the case where the prior is improper, i.e., not a probability density, we divide the parameter $\boldsymbol{c}=(\boldsymbol{c}_1, \boldsymbol{c}_2)$ so that $\pi(\boldsymbol{c}_1, \boldsymbol{c}_2|w_1, w_2)$ is proper with respect to $\boldsymbol{c}_1$. Then we consider the marginal,

$$\mathscr{L}(w_1, w_2, \boldsymbol{c}_2) = \int L(\boldsymbol{c})\pi(\boldsymbol{c}_1, \boldsymbol{c}_2, |w_1, w_2)d\boldsymbol{c}_1 , \tag{3.4}$$

of the posterior to obtain w_1, w_2 and $\boldsymbol{c}_2$ which maximize $\mathscr{L}$ (Good (1965)). This is called the method of type II maximum likelihood. Akaike (1979) justified and developed this method based on the entropy maximization principle and defined

$$\mathrm{ABIC} = (-2) \max_{w_1, w_2, \boldsymbol{c}_2} \log \mathscr{L}(w_1, w_2, \boldsymbol{c}_2) . \tag{3.5}$$

In the next section we see that the penalties Φ_1 and Φ_2 are quadratic with respect to the parameters $\boldsymbol{c}=(c_{ij})=(c_k)$, and that the prior π is a multivariate Gaussian distribution. However, on the other hand, the likelihoods in (2.4) and (2.5) are certainly not Gaussian. We cannot obtain analytic solution of the integral in (3.4) unlike the case of Akaike (1979). Nevertheless, since the prior π is Gaussian, we may use the Gaussian approximation method suggested by Ishiguro and Sakamoto (1983): That is to say, the logarithm of the integrand in (3.4), assuming for simplicity that the prior π is proper

$$T(\boldsymbol{c}; w_1, w_2) = \log\{L(\boldsymbol{c})\pi(\boldsymbol{c}|w_1, w_2)\} , \tag{3.6}$$

is approximated by the quadratic form

$$T(\boldsymbol{c}; w_1, w_2) \cong T(\hat{\boldsymbol{c}}; w_1, w_2) - \frac{1}{2}(\boldsymbol{c} - \hat{\boldsymbol{c}})H(\hat{\boldsymbol{c}}; w_1, w_2)(\boldsymbol{c} - \hat{\boldsymbol{c}})^t , \tag{3.7}$$

where $\hat{\boldsymbol{c}}$ is the vector which maximizes T in (3.6) for fixed w_1 and w_2, and $H(\hat{\boldsymbol{c}}; w_1, w_2)$ is the Hessian matrix of the penalized log likelihood at $\hat{\boldsymbol{c}}$. Using this approximation the integral in (3.4) is obtained. Thus we have

$$\mathrm{ABIC} = (-2)T(\hat{\boldsymbol{c}}; w_1, w_2) + \log\{\det H(\hat{\boldsymbol{c}}; w_1,w_2)\} - K \log 2\pi , \tag{3.8}$$

where K is the dimension of the parameter $\boldsymbol{c}$. It should be noted that both the minimization in (3.8) with respect to the hyperparameters and maximization of T in (3.6) with respect to the parameters $\boldsymbol{c}=(c_k)$ for fixed hyperparameters w_1 and w_2 are non linear. We use the Davidon-Fletcher-Powell method (e.g., Fletcher and Powell (1963), and Akaike *et al.* (1979)) for maximizing T in (3.6) with respect to the vector $\boldsymbol{c}$ of parameters for fixed w_1 and w_2. We also use the same optimization method, but using numerically calculated gradients, for minimizing (3.8) with respect to w_1 and w_2. These are repeated by turns until the latter optimization converges. In optimizing T above with respect to $\boldsymbol{c}$ we found that the use of the Hessian $H(\boldsymbol{c}; w_1, w_2)$ in suitable stages (say, in the 5th or the 10th step) makes the convergence very rapid. This seems to support the goodness of the approximation (3.7) for (3.6).

It is useful to get the estimation error of $\mu_{\hat{\sigma}}(x, y)$ and $\beta_{\hat{\tau}}(x, y)$ at each location (x, y). We know that joint error distribution of the parameters at $\hat{\boldsymbol{c}}=(\hat{c}_k)$ is approximately given by the multivariate normal $N(\hat{\boldsymbol{c}}, H^{-1})$, where $H^{-1}=(h^{k,k'})$ is the inverse of the same Hessian matrix $H=(h_{k,k'})$ as in (3.7). Since $h(x, y|\boldsymbol{c})$ in (2.8) is given by the linear combination with respect to c_k, this is also expected to be approximately normal and the variance-covariance matrix between (x, y) and (x', y') is given by

$$C(x, y\colon x', y') = \sum_k \sum_{k'} h^{k,k'} F_i(x) G_j(y) F_{i'}(x') G_{j'}(y') , \tag{3.9}$$

where $k=i(J+3)+j$ and $k'=i'(J+3)+j'$ for $i, i'=1, 2,\dots, I+3$ and $j, j'=1, 2,\dots, J+3$. Thus the standard error of $h(x, y|\boldsymbol{c})$ at (x, y) is

$$\varepsilon(x, y) = C(x, y\colon x, y)^{1/2} , \tag{3.10}$$

and the errors of $\mu_{\hat{\sigma}}$ and $\beta_{\hat{\tau}}$ are given by the corresponding log normal distribution owing to the relation (2.7).

4. Some technical comments

In this paper the area A is restricted to be a rectangle $[\xi_0, \xi_M]\times[\eta_0, \eta_N]$. Since 2-dimensional spline to be used is in the form of (2.8), we consider the segment $\Omega=[\xi_0, \xi_M]$ being extended to the segment $[\xi_{-3}, \xi_{M+3}]$, where $\{\xi_i; i=-3,-2,\dots, M+3\}$ are equally spaced knots in the distance of $d_x=(\xi_M-\xi_0)/M$. Inoue (1986) used the following cubic B-spline basis $\{B_i(r), i=1,2,3,4\}$ on $[0,1]$ such that

$$\begin{aligned} B_1(r) &= r^3/6 , \\ B_2(r) &= (-3r^3 + 3r^2 + 3r + 1)/6 , \\ B_3(r) &= (3r^3 - 6r^2 + 4)/6 , \\ B_4(r) &= (-r^3 + 3r^2 - 3r + 1)/6 . \end{aligned} \tag{4.1}$$

Thus for $(x, y) \in [\xi_k, \xi_{k+1}]\times[\eta_l, \eta_{l+1}] \subset A$ the function h in (2.8) is written by

$$h(x, y|\boldsymbol{c}) = \sum_{i=0}^{3} \sum_{j=0}^{3} C_{k+i,l+j} B_{4-i}(r_x) B_{4-j}(r_y) , \tag{4.2}$$

where $r_x=(x-\xi_k)/d_x$ and $r_y=(y-\eta_l)/d_y$. One of advantages of this spline is that the roughness penalties $\Phi_1(h)$ and $\Phi_2(h)$ in (3.1) and (3.2) reduce to the integrals $\int_0^1 B_i'B_j'\,dr$ and $\int_0^1 B_i''B_j''\,dr$, where the dashes indicate the derivatives with respect to r.

It is seen that the sum of the penalties in (3.3) is given by the quadratic form of the parameter $\boldsymbol{c}$, that is to say, for some non-negative definite symmetric matrix Σ we get $\boldsymbol{c}\Sigma\boldsymbol{c}^t/2=w_1\Phi_1(h)+w_2\Phi_2(h)$. Here it turns out that Σ is degenerated such that $r=\text{rank}(\Sigma)=(M+3)(N+3)-1$. Therefore the prior probability density π suggested in the previous section is constructed using the multivariate normal distribution characterized by w_1, w_2 and $c_{(M+3)(N+3)}$ in such a way that

$$(4.3)\qquad \pi(\boldsymbol{c}_1\,|\,w_1,\ w_2,\ c_{(M+3)(N+3)}) = (\det\Sigma_1)^{1/2}/(2\pi)^{r/2}\exp\{-\boldsymbol{c}\Sigma\boldsymbol{c}^t/2\}\ ,$$

where $\boldsymbol{c}=(\boldsymbol{c}_1,\ c_{(M+3)(N+3)})$, and Σ_1 is the cofactor of the last diagonal element of Σ.

Finally it is found that the standard error $\varepsilon(x,\ y)$ in (3.10) for the case where $h(x,\ y)=$constant is slightly cyclic with spatial period $(d_x/2,\ d_y/2)$. This is due to the quadratic form of (3.9) with respect to the spline bases. To remove the cyclicity we use $\{\varepsilon(x+d_x/4, y+d_y/4)+\varepsilon(x-d_x/4, y-d_y/4)\}/2$ instead of $\varepsilon(x,\ y)$.

5. Numerical performance

The source of the seismicity data examined was the catalogue (*The Seismological Bulletins*) compiled by the Japan Meteorological Agency (JMA). Here the data set for 65 years from 1926 through 1980 was considered, and the region used was a rectangular from 141° E to 145° E and from 36° N to 42° N where the area of Off Tohoku District is included (see Fig. 1). The earthquakes not less than magnitude 5.0 were taken to keep the data set homogeneous. The number of events was 1736. Most of shocks occurred in the boundary between the continental plate and the subducting oceanic plate. The area was divided by the equally spaced knots with $M=8$ and $N=15$, which needs 198 parameters for $\boldsymbol{c}=(c_{ij})$ in (2.8) or (4.2).

Before describing the estimate of spatial variation of $\beta(x,\ y)$ we should mention about the definition of the b-value of the magnitude frequency, which is based on the empirical law of the number $N(m)$ of earthquake having magnitude m or larger such that $\log_{10}N(m)=a-bm$, observed by Gutenberg and Richter (1944). Here we may be interested in estimating spatial variation of b-values. The relation of the $b(x,\ y)$ and $\beta(x,\ y)$ defined in (2.6) is therefore given by $b(x,\ y)=\beta(x,\ y)\log_{10} e$. Thus we parameterized β_τ by

$$(5.1)\qquad \beta_\tau(x,\ y) = \exp\{h_\tau(x,\ y)\}/\log_{10} e\ ,$$

where $h_\tau(x, y)$ is in (2.7). Using the numerical procedure stated in the preceding sections we got the optimal estimate $b_{\hat{\tau}}(x, y)$ with the minimum ABIC in (3.8), which is shown by contour lines in Fig. 1(a). The contour lines of standard error $\varepsilon_{\hat{\tau}}(x, y)$ of $\log b_{\hat{\tau}}(x, y)$ are given in Fig. 1(b). These suggest that the spatial variation of b-values from the east to the west of the area seems significant. However we cannot conclude here that whether this is due to the genuine geophysical effect in all of the area, or not: although northern part of the area is nearly surrounded by several JMA seismic stations in Tohoku and Hokkaido, for the rest the location of seismic stations are only one sided, which might cause a bias for the spatial variation of the b-values. On the other hand, some seismologists claim that the cut-off magnitude 5.0 is large enough to guarantee the homogeneity of data set for the area and the time span. If this is shown to be true, then our estimate here will be a reflection of genuine geophysical effect.

We also got Fig. 1(c) and 1(d) for the optimal intensity estimate $\mu_{\hat{\sigma}}(x, y)$ of the location of earthquakes and $\log \mu_{\hat{\sigma}}(x, y)$, respectively.

Finally the data of natural stands of seedlings and saplings of the Japanese black pine with number of 204 in a 10 metres$\times$10 metres area (Numata (1964)) is considered. For each pine two numbers, height by centimetres and age, are attached. This is formerly analyzed in Ogata and Tanemura (1985, 1986), for example, and Fig. 2 in the former paper also includes the full information of the data set. Here we simply assume marked non-homogeneous Poisson, i.e., no interactions among the locations themselves. The area was divided by the equally spaced knots with $M=N=10$. Figure 2(a) and (b) provide the estimated intensity $\mu_{\hat{\sigma}}(x, y)$ of the Poisson models for the location of the pines and standard errors of $\log \mu_{\hat{\sigma}}(x, y)$, respectively. The estimated intensity is quite similar to Fig. 2 in Ogata and Tanemura (1986) where AIC is used to select the optimal order of polynomial. The histogram of the height of the pines (see Fig. 5 in Numata (1964)) seems to be distributed according to the exponential function on the whole. This leads us to examining the spatial variation of the exponential coefficients by the similar way to the b-value analysis for the magnitude frequencies. The minimum ABIC is attained when the both penalties w_1 and w_2 in (3.3) are very large; that is, $b_{\hat{\tau}}(x, y)$ is constant. This means no spatial variation with b-values and it seems that the density of the pine location does not affect the distribution of their heights, while we saw opposite effect, from the heights of pines to the locations, in Ogata and Tanemura (1985) and in Stoyan (1986).

Fig. 1(a). Spatial variation of the b-values for the shallow seismic activity in East Off Tohoku, Japan for 1926–1980, contour lines and bird's-eye view. The values in "HEIGHT=" give the range of the corresponding contour lines. The mark + indicates the epicentre of an earthquake with magnitude $M \geq 5.0$.

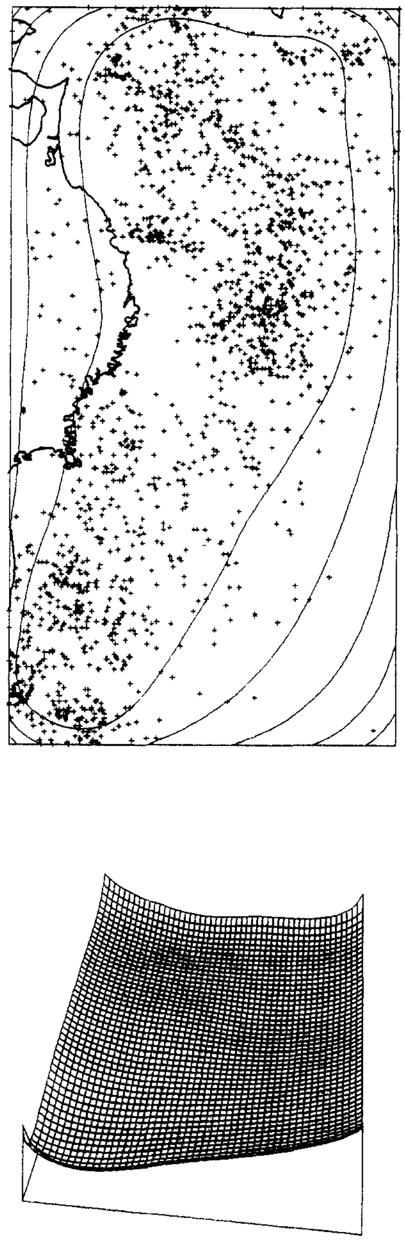


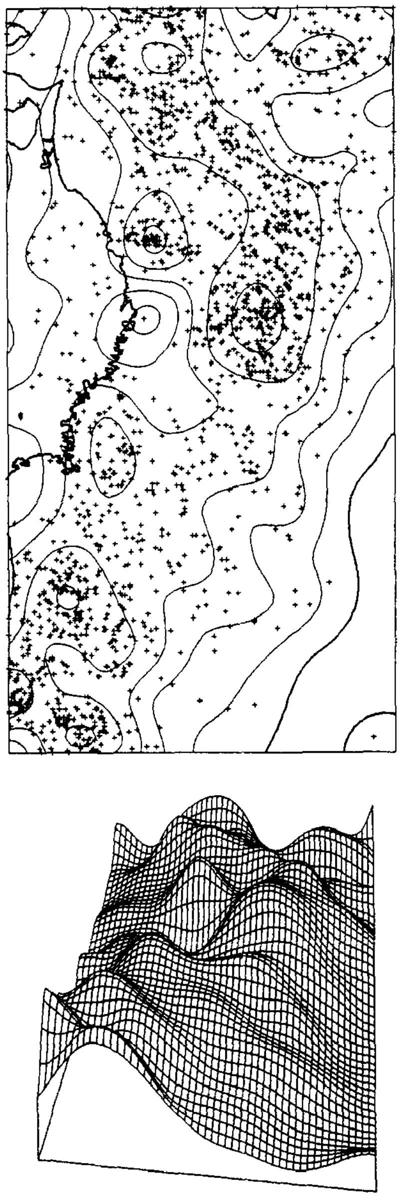
Fig. 1(b). Standard error surface for the logarithm of the estimate of the spatial b-value in Fig. 1(a), contour lines and bird's-eye view.

HEIGHT= 50. 100. 150. 200. 250.
300. 350. 400. 450. 500.
550. 600. 650. 700. 750
800.

Z0=0.0 ZN=1000. ZW=5.0

Fig. 1(c). Graphs of contour lines and bird's-eye view for the intensity rate of the seismic activity in East Off Tohoku, Japan. The values in "HEIGHT=" mean the corresponding frequency rate of contour lines of seismicity with magnitude $M \geq 5.0$ per the unit area, 1° longitude×1° latitude for 1926–1980.

HEIGHT=1.0 2.0 3.0 4.0 5.0 6.0



Z0=0.0 ZN=10.0 ZW=5.0

Fig. 1(d). Graphs of contour lines and bird's-eye view for the intensity rate of the seismic activity in East Off Tohoku, Japan. The same as that in Fig. 1(c) but the heights are plotted in logarithmic scale.

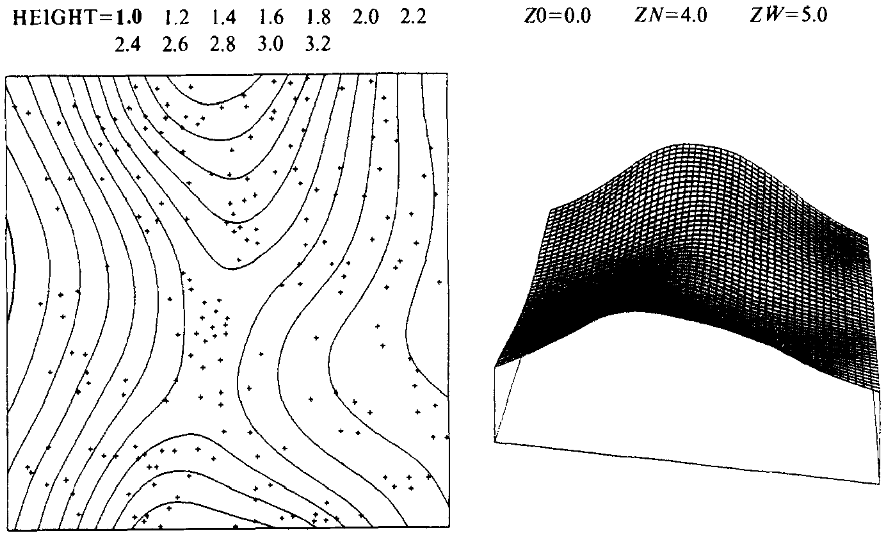


Fig. 2(a). Graphs of contour lines and bird's-eye view for the intensity rate of the location of the black pines (mark +). The values in -HEIGHT- provide the range of the corresponding contour lines of frequency per unit area; 1 square metres.

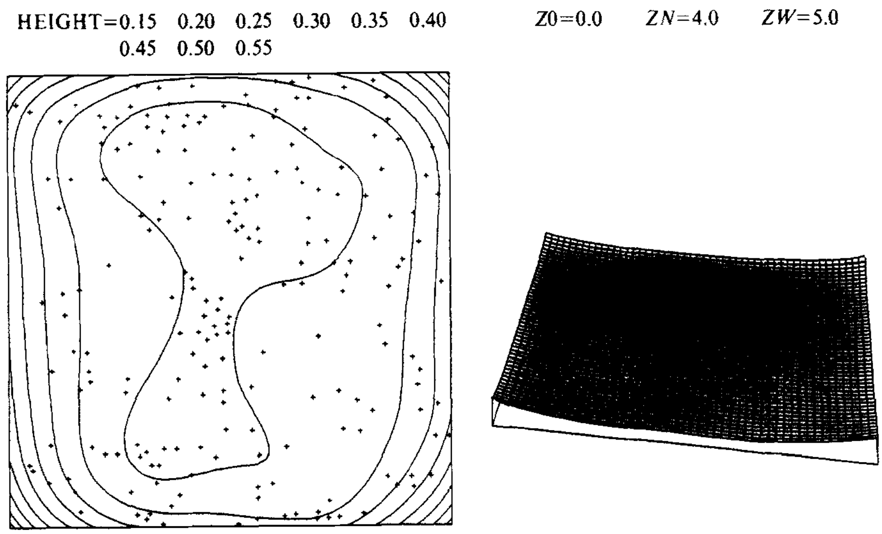


Fig. 2(b). Standard error for the estimate of logarithm of the intensity rate corresponding to Fig. 2(a), contour lines and bird's-eye view.

Acknowledgements

We are grateful to Hiroshi Inoue for supplying us subroutines for the cubic spline function and its related calculation evaluating the penalties which are published in Inoue (1986). We also thank our colleagues Makio Ishiguro and Genshiro Kitagawa for helpful comments in performing the numerical calculations. This research was partly supported by Grant-in-Aid for Scientific Research 60530016 from the Ministry of Education, Science and Culture of Japan.

REFERENCES

Akaike, H. (1979). Likelihood and Bayes procedure, *Bayesian Statistics*, (eds. J. M. Bernard *et al.*), University Press, Valencia, Spain.

Akaike, H., Kitagawa, G., Arahata, E. and Tada, F. (1979). TIMSAC-78, *Comput. Sci. Monographs*, No. 11, Institute of Statistical Mathematics, Tokyo.

Fletcher, R. and Powell, M. J. D. (1963). A rapidly convergent descent method for minimization, *Comput. J.*, **6**, 163-168.

Good, I. J. (1965). *The Estimation of Probabilities*, M.I.T. Press, Cambridge, Massachusetts.

Good, I. J. and Gaskins, R. A. (1971). Nonparametric roughness penalties for probability densities, *Biometrika*, **58**, 255-277.

Gutenberg, B. and Richter, C. F. (1944). Frequency of earthquakes in California, *Bull. Seismol. Soc. Amer.*, **34**, 185-188.

Inoue, H. (1986). A least squares smooth fitting for irregularly spaced data: Finite element approach using cubic *B*-spline basis, *Geophysics*, **51**, 2051-2066.

Ishiguro, M. and Sakamoto, Y. (1983). A Bayesian approach to binary response curve estimation, *Ann. Inst. Statist. Math.*, **35**, 115-137.

Meinguet, J. (1979). Multivariate interpolation at arbitrary points made simple, *Z. Angew. Math. Phys.*, **30**, 292-304.

Numata, M. (1964). Forest vegetation, particularly pine stands in the vicinity of Choshi-flora and vegetation at Choshi, Chiba Prefecture, VI, *Bulletin of the Choshi Marine Laboratory*, No. 6, 27-37, Chiba University (in Japanese).

Ogata, Y. and Tanemura, M. (1985). Estimation of interaction potentials marked spatial point patterns through the maximum likelihood method, *Biometrics*, **41**, 421-433.

Ogata, Y. and Tanemura, M. (1986). Likelihood estimation of interaction potentials and external fields of inhomogeneous spatial point patterns, *Proc. PACIFIC STATISTICAL CONGRESS-1985*, (eds. I. S. Francis, B. F. J. Manly and F. C. Lam), 150-154, North-Holland.

Stoyan, D. (1988). Statistical analysis of spatial point processes: A soft-core model and cross-correlations of marks, *Biometrical J.* (to appear).

Titterington, D. M. (1985). Common structure of smoothing techniques in statistics, *Internat. Statist. Rev.*, **53**(2), 141-170.