*Appl. Statist.* (1985)
34, *No.* 2, *pp.* 138–147

# A Kernel Method for Smoothing Point Process Data

By PETER DIGGLE

*CSIRO Division of Mathematics and Statistics, Canberra and Department of Statistics, University of Newcastle upon Tyne, UK*

[Received June 1984. Revised December 1984]

SUMMARY

A method for estimating the local intensity of a one-dimensional point process is described. The estimator uses an adaptation of Rosenblatt's kernel method of non-parametric probability density estimation, with a correction for end-effects. An expression for the mean squared error is derived on the assumption that the underlying process is a stationary Cox process, and this result is used to suggest a practical method for choosing the value of the smoothing constant. The performance of the estimator is illustrated using simulated data. An application to data on the locations of joints along a coal seam is described. The extension to two-dimensional point processes is noted.

*Keywords*: Cox process; Non-parametric density estimation; Point process data; Smoothing

## 1. Introduction

In this paper, we discuss a method of smoothing data which arise as a partial realization of a one-dimensional point process. Suppose that there are $n$ points $x_i$ in the interval $(0, T)$. A useful aid to preliminary interpretation of the data is an estimate of the "local intensity", $\lambda(x)$ say, of points $x_i$ in the neighbourhood of each point $x \in (0, T)$. We propose an estimator of $\lambda(x)$ which takes the general form

$$\hat{\lambda}_t(x) = \left\{ \sum_{i=1}^{n} \delta_t\ (x - x_i) \right\} / p_t(x). \tag{1.1}$$

In (1.1), $\delta_t(x) = t^{-1}\delta(t^{-1}x)$ is a p.d.f. symmetric about the origin, $t > 0$ determines the amount of smoothing and $p_t(x) = \int_0^T \delta_t(x-u)\,du$ is an end-correction. This defines a kernel estimator (Rosenblatt, 1956) although the end-correction is non-standard. The successful implementation of (1.1) rests on choosing an appropriate value of $t$, whereas the choice of the "kernel function" $\delta(x)$ is of very secondary importance (see, for example, Silverman, 1978).

In Section 2 we make precise the notion of "local intensity" by assuming that the underlying point process is a stationary Cox process. Using this as a working model, we then develop a method of choosing the smoothing constant $t$ with a view to minimizing mean square error. In Section 3, we illustrate various aspects of the performance of the estimator using simulated data. Section 4 presents an application to data on the locations of joints along coal seams. In one seam, there appears to be a gross increase in the local intensity of joints near a known geological fault, together with smaller variations away from the main fault zone. Qualitatively similar results are obtained for two further seams. Finally, Section 5 discusses briefly the difficulties which arise in extending our notion of local intensity to a general point process.

*Present address*: CSIRO Division of Maths and Statistics, P.O. Box 1965, Canberra 2601, Australia.


0035–9254/85/34138 $2.00

## 2. Development of the Estimator

### 2.1. *Definition of Local Intensity for a Stationary Cox Process*

We assume that the underlying point process is a stationary Cox process with rate process $\{\Lambda(x): x \in \mathbb{R}\}$. By this we mean that

(i) $\{\Lambda(x): x \in \mathbb{R}\}$ is a stationary, non-negative valued random process,
(ii) conditional on the realization $\lambda(x)$ of $\Lambda(x)$, the point process is an inhomogeneous Poisson process with rate function $\lambda(x)$.

See, for example, Cox and Isham (1980, pp. 70-75), who refer to such processes as "doubly stochastic Poisson processes". We call $\lambda(x)$ the *local intensity*, and regard (1.1) as an estimator for $\lambda(x)$.

Note that what is being estimated is a realization of a random process. In this respect, our problem has some similarity to that of kriging (Ripley, 1981, Ch. 4). A further similarity is that we shall measure the performance of our estimator for $\lambda(x)$ in terms of a mean square error in which the expectation is with respect to both the sampling distribution of the estimator and the distribution of $\lambda(x)$ itself. However, the technical details are quite different; in kriging, $\lambda(x)$ would be directly observable at a finite set of pre-determined sample points, whereas in the present context $\lambda(x)$ is observed indirectly over a continuum, via the realization of the underlying point process.

### 2.2. *Mean Square Error for a Uniform Kernel Function*

Let $N(a, b)$ denote the number of points of the underlying Cox process in the interval $(a, b)$. Write $E[\Lambda(x)] = \mu$ and $E[\Lambda(x)\,\Lambda(y)] = \nu(|\,x-y\,|)$. Let $K(t) = 2\mu^{-2} \int_0^t \nu(u)\,du$ be the corresponding reduced second moment measure. In (1.1), choose $\delta(x) = \frac{1}{2}$, $-1 \leqslant x \leqslant 1$. Ignoring the end-correction, this gives an estimator

$$\tilde{\lambda}_t(x) = (2t)^{-1}\, N(x-t, x+t).$$

The following result provides a method of estimating the value of $t$ which minimizes the *mean square error*, $MSE(t) = E[\{\tilde{\lambda}_t(x) - \Lambda(x)\}^2]$.

*Theorem.* $MSE(t) = \nu(0) + \mu\{1 - 2\mu K(t)\}/(2t) + \{\mu/(2t)\}^2 \int_0^{2t} K(y)\,dy$.

*Proof.* Note first that, by stationarity, $MSE(t)$ does not depend on $x$. Also, conditional on the realization of the rate process $\Lambda(x)$, $N(-t, t)$ follows a Poisson distribution with mean $\int_{-t}^{t} \Lambda(x)\,dx$. Thus,

$$E[\{\tilde{\lambda}_t(0) - \Lambda(0)\}^2 \mid \Lambda(x): x \in \mathbb{R}] = (2t)^{-2} \int_{-t}^{t} \Lambda(x)\,dx + \{(2t)^{-1} \int_{-t}^{t} \Lambda(x)\,dx - \Lambda(0)\}^2$$

$$= (2t)^{-2} \left\{ \int_{-t}^{t} \Lambda(x)\,dx + \int_{-t}^{t}\int_{-t}^{t} \Lambda(x)\,\Lambda(y)\,dxdy \right\}$$

$$- t^{-1}\,\Lambda(0) \int_{-t}^{t} \Lambda(x)\,dx + \{\Lambda(0)\}^2.$$

Now, taking expectations with respect to the rate process,

$$MSE(t) = (2t)^{-1}\,\mu + (2t)^{-2} \int_{-t}^{t}\int_{-t}^{t} \nu(|\,x-y\,|)\,dxdy - t^{-1} \int_{-t}^{t} \nu(|\,x\,|)\,dx + \nu(0).$$

To complete the proof, make the substitution $z = x - y$ in the double integral, note that $\nu(.)$ is an even function and incorporate the relationship between $K(.)$ and $\nu(.)$.

For a homogeneous Poisson process, $K(t) = 2t$ and $MSE(t)$ is monotone decreasing in $t$. This also follows directly from the fact that in this case $\Lambda(x)$ assumes a constant value $\mu$, $N(a, b)$ follows a Poisson distribution with mean $\mu(b - a)$ and $\tilde{\lambda}_t(x)$ is therefore unbiased with variance $(2t)^{-1}\,\mu$.

For a general stationary Cox process,

$$K(t) = 2t + \mu^{-2} \int_0^t \gamma(u)\, du, \tag{2.1}$$

where $\gamma(u) = \nu(u) - \mu^2$ is the covariance function of $\Lambda(x)$. To provide a specific illustration, we consider a "linear" Cox process (Bartlett, 1964), for which

$$\Lambda(x) = \{\mu/(\rho\sigma)\} \sum_{i=1}^{\infty} \phi\{(x - X_i)/\sigma\},$$

where $\phi(.)$ is the standard Normal probability density function and the $X_i$ are the points of a homogeneous Poisson process with intensity $\rho$. The covariance function of $\Lambda(x)$ is then

$$\gamma(u) = \{\rho^{-1}\,\mu^2 (4\pi\sigma^2)^{-1/2}\} \exp\{-u^2/(4\sigma^2)\}.$$

Figure 1 shows a typical set of curves of the function $M(t) = \mu^{-2}\{MSE(t) - \nu(0)\}$ for the case $(\rho, \sigma) = (20, 0.05)$ and each of $\mu = 200$, 600, 1000. Note that the minimizing value of $t$ decreases

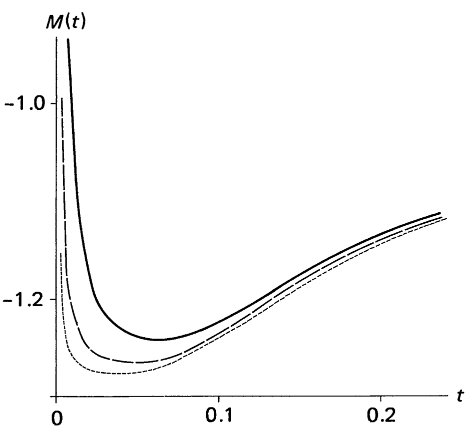


Fig. 1. The function $M(t)$ for three Cox processes with common correlation structure but different values of $\mu$. —— $\mu = 200$. – – – $\mu = 600$; . . . . . $\mu = 1000$.

as $\mu$ increases. This is the analogue of the familiar result from non-parametric probability density estimation that larger samples need less smoothing. However, we emphasize that in the present context the optimum value of $t$ does not change if the number of points simply increases in direct proportion to the length of the interval on which the points are observed.

2.3. *A Method for Choosing t*

In order to estimate the value of $t$ which minimizes $MSE(t)$ or, equivalently, $M(t)$, we require estimates of $\mu$ and $K(t)$. The natural estimator for $\mu$ is $\hat{\mu} = n/T$. An estimator for $K(t)$ is the one-dimensional analogue of Ripley's (1981, p. 159) estimator. This takes the form

$$\hat{K}(t) = T n^{-2} \sum_{i \neq j}\sum a(x_i, x_j, t),$$

where

$$a(x, y, t) = \begin{cases} 0 : & |x-y| > t \\ 1 : & |x-y| \leqslant t, \quad |x-y| \leqslant \min(x, T-x) \\ 2 : & |x-y| \leqslant t, \quad |x-y| > \min(x, T-x). \end{cases}$$

Substitution of $\hat{\mu}$ and $\hat{K}(t)$ into the theoretical expression for $M(t)$ defines an estimator $\hat{M}(t)$; the value, $t_0$ say, which minimizes this function can then be determined by direct inspection. In practice, we evaluate $\hat{K}(t)$ for a discrete set of values of $t$ determined by the resolution of the data, and approximate the integral term in $\hat{M}(t)$ by a numerical integration.

A useful by-product of this method for choosing $t$ is that $\hat{K}(t)$ can be used to confirm the significance of apparent aggregation in the data relative to a homogeneous Poisson process, for which $K(t) = 2t$. For a conditioned Poisson process, consisting of $n$ points independently and uniformly distributed on $(0, T)$, the method outlined in Lotwick and Silverman (1982) can be adapted to show that

$$\operatorname{var}\{\hat{K}(t)\} = T^2\{(n-1)/n^3\}\{4u - 5u^2 + 2(n-2)u^3/3\}, \tag{2.2}$$

where $u = t/T$. A useful diagnostic is a plot of $\hat{K}(t) - 2t$ together with tolerance limits defined by $\pm 2[\operatorname{var}\{\hat{K}(t)\}]^{1/2}$. Estimation of local intensity can only be useful if the data are markedly aggregated, and this is indicated by significantly positive values of $\hat{K}(t) - 2t$. Note also from (2.1) that for a Cox process with positive-valued $\gamma(u)$, $K(t) - 2t$ is an increasing function of $t$; thus the diagnostic plot gives some indication of whether the assumption of an underlying Cox process is reasonable.

2.4. *Extension to Non-uniform Kernel Functions*

Let $c = \int x^2 \delta(x)\, dx$. For the uniform kernel, $c = \frac{1}{3}$. In general, we suggest implementing (1.1) with $t = (3c)^{-1/2} t_0$, where $t_0$ minimizes $\hat{M}(t)$, i.e. we calibrate different kernels by equating their variances. In view of earlier remarks, we expect different kernels to give comparable results. Largely for aesthetic reasons, we have chosen to use a smooth kernel function,

$$\delta(x) = 0.9375\,(1 - x^2)^2 : -1 \leqslant x \leqslant 1, \tag{2.3}$$

for which $c = \frac{1}{7}$.

It would be preferable to evaluate $MSE(t)$ directly for a general kernel function. Unfortunately, this does not admit an explicit expression in terms of $\mu$ and $K(t)$, and therefore does not provide a practicable method for choosing $t$.

2.5. *Connection with Probability Density Estimation*

There is an obvious similarity between our estimator (1.1) and the kernel estimator of a probability density estimation (pdf) introduced by Rosenblatt (1956). Operationally, the only non-trivial distinction is the end-correction in (1.1). A more subtle distinction concerns the genesis of the data and the implications of this for the choice of the smoothing constant. In estimation of a pdf, the domain of the data is assumed to be the whole of $\mathcal{R}$ or, less commonly, a *fixed* subset thereof. For point process data, the restriction to an interval $(0, T)$ is essentially arbitrary in that the process exists, but is unobserved, outside $(0, T)$. In conjunction with an assumption of stationarity, this implies that any method for choosing the smoothing constant should be essentially independent of $T$, as is the case for the method developed in the present paper.

## 3. Simulation Studies

We now illustrate several aspects of our proposed method, using simulated data. The basic simulation model is the linear Cox process defined in Section 2.2.

### 3.1. *Bias and Variability in Estimation of the Optimal t*

It is of some interest to compare the theoretically optimal $t_{opt}$ for a linear Cox process with values $t_0$ obtained by applying the method of Section 2.3 to simulated data. For the model parameters, we fixed $\mu = 200$ and took various combinations of values for $\rho$ and $\sigma$ as indicated in Table 1. We then simulated 25 realizations of each process for each of $T = 1.0, 2.5$ and $5.0$, corresponding to average sample sizes $n = 200$, 500 and 1000. Table 1 gives the corresponding values of $t_{opt}$ and the sample means and standard deviations of $t_0$. Note that the bias in $t_0$ is generally positive and that, as would be expected, the standard deviation of $t_0$ decreases as $n$ increases. The degree of bias in $t_0$ is such as to produce only a mild deterioration in the performance of the estimator. For example, taking $(\rho, \sigma) = (20, 0.05)$ and $T = 1.0$, $MSE(\bar{t}_0)/MSE(t_{opt}) = 1.09$. At least for relatively small $n$ and large $\sigma$, the variability in $t_0$ is potentially more serious. Again taking $(\rho, \sigma) = (20, 0.05)$ and $T = 1.0$, then $\bar{t}_0 + SD(t_0) = 0.118$ and $MSE(0.118)/MSE(t_{opt}) = 1.90$.

TABLE 1

*Bias and variability in estimation of $t_{opt}$. For $\mu = 200$ and each combination of $\rho$ and $\sigma$, the table gives $t_{opt}$ and the sample means and standard deviations (in parentheses) of $t_0$ from 25 simulations, for $T = 1.0, 2.5$ and $5.0$.*

| | | $\rho$ 10 | 20 | 40 |
|---|---|---|---|---|
| $\sigma$ | 0.025 | 0.024 | 0.030 | 0.036 |
| | | 0.025 (0.009) | 0.033 (0.009) | 0.046 (0.022) |
| | | 0.028 (0.007) | 0.035 (0.008) | 0.042 (0.009) |
| | | 0.027 (0.006) | 0.033 (0.005) | 0.038 (0.009) |
| | 0.05 | 0.050 | 0.061 | 0.074 |
| | | 0.050 (0.020) | 0.076 (0.042) | 0.101 (0.052) |
| | | 0.058 (0.023) | 0.064 (0.024) | 0.078 (0.025) |
| | | 0.059 (0.015) | 0.059 (0.016) | 0.083 (0.022) |
| | 0.1 | 0.103 | 0.122 | 0.148 |
| | | 0.120 (0.045) | 0.130 (0.054) | 0.146 (0.047) |
| | | 0.095 (0.038) | 0.124 (0.042) | 0.145 (0.048) |
| | | 0.114 (0.034) | 0.126 (0.037) | 0.156 (0.030) |

### 3.2. *Illustrative Examples*

Our first example is a simulation of the linear Cox process with $\mu = 500$, $(\rho, \sigma) = (20, 0.05)$ and $T = 1.0$. Figure 2a shows the empirical function $\hat{M}(t)$, from which we read off $t_0 = 0.045$. Figure 2b compares two versions of $\hat{\lambda}_{0.045}(x)$, using the uniform kernel and the quartic kernel (2.3), along with the realization $\lambda(x)$ of $\Lambda(x)$. The estimate using the quartic kernel looks more attractive in that it reflects the analytic smoothness of $\lambda(x)$. Although this improvement over the uniform kernel is largely cosmetic, we shall use only the quartic kernel in the sequel.

Our second example has some bearing on the practical implications of the stationarity assumption. It takes $\mu = 500$, $(\rho, \sigma) = (10, 0.1)$ and $T = 1.0$, and leads to $t_0 = 0.11$. Figure 3 shows $\hat{\lambda}_{0.11}(x)$ and the realization $\lambda(x)$ of $\Lambda(x)$. Neither $\lambda(x)$ nor its estimate resemble the conventional image of a stationary random function, albeit a highly non-Gaussian one. Given only a single realization $\lambda(x)$, the question of whether its genesis is stochastic or deterministic is a rather sterile

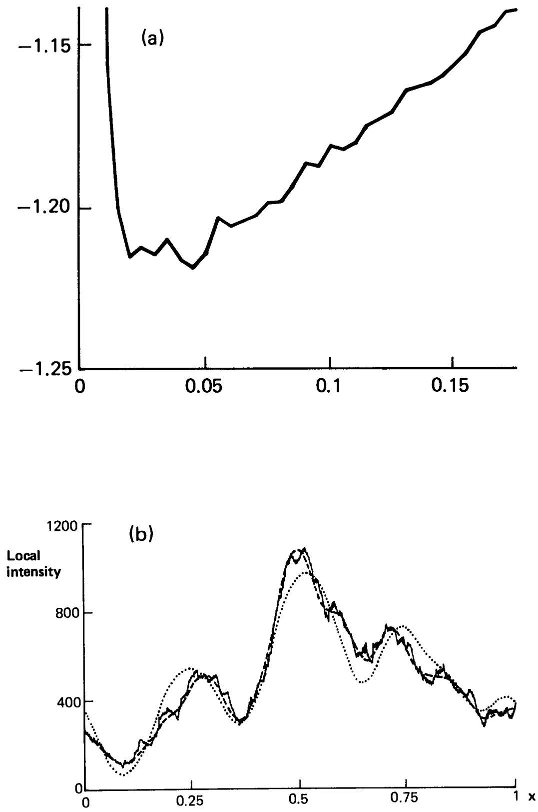


Fig. 2. Simulation of a Cox process with $\mu = 500$, $\rho = 20$, $\sigma = 0.05$, $T = 1.0$. (a) The empirical function $\hat{M}(t)$; (b) the local intensity and two estimates. —— $\hat{\lambda}_{0.045}(x)$, uniform kernel; – – – $\hat{\lambda}_{0.045}(x)$, quartic kernel; . . . . . $\lambda(x)$.

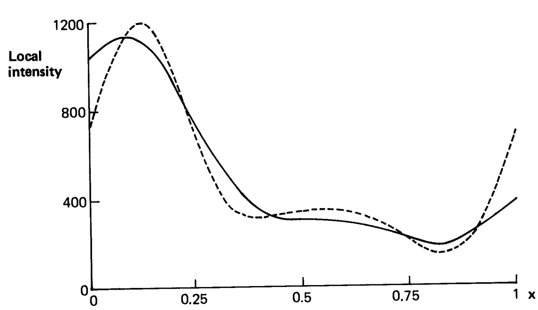


Fig. 3. Simulation of a Cox process with $\mu = 500$, $\rho = 10$, $\sigma = 0.1$, $T = 1.0$
—— $\lambda_{0.11}(x)$; – – – $\lambda(x)$.

one. We regard the stationarity assumption as a pragmatic working model which often gives a sensible answer; note in particular that a function $\lambda(x)$ which has few turning points within $(0, T)$ will typically be associated with a relatively large value of $t_0$.

Our final example anticipates a feature of some data which we shall examine in Section 4. It consists of a concatenation of four simulations, each with $(\rho, \sigma) = (20, 0.025)$ and $T = 0.5$, but with $\mu$ alternating between 300 and 700. The function $\hat{M}(t)$ for these data, given as Fig. 4a, has two distinct local minima at around $t = 0.03$ and $t = 0.2$. Figure 4b shows the two corresponding

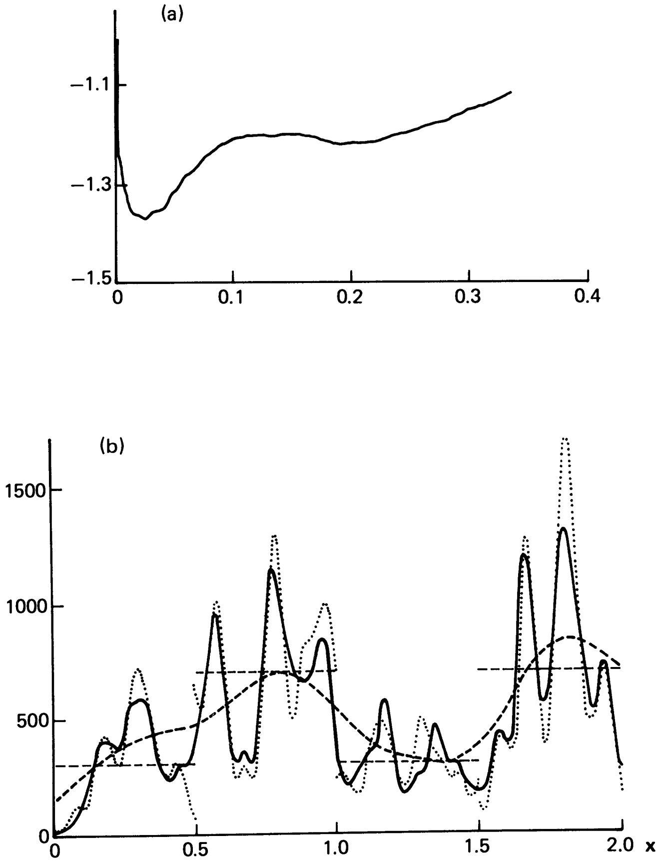


Fig. 4. Simulation of Cox process with alternating values of $\mu$ over four consecutive stretches. (a) The empirical function $\hat{M}(t)$. (b) the local intensity and two estimates —— $\hat{\lambda}_{0.03}(x)$; --- $\hat{\lambda}_{0.20}(x)$; ..... $\lambda(x)$ (alternating values of $\mu$ shown as thin dashed horizontal lines).

estimates $\hat{\lambda}_t(x)$, also the alternating values of $\mu$ and the realization $\lambda(x)$. The estimate with $t = 0.03$ successfully captures the small-scale fluctuations in $\lambda(x)$, whilst that with $t = 0.2$ better describes the larger-scale changes in $\mu$, although it cannot of course reflect the discontinuous changes in $\mu$ at the three break-points. Further simulations suggest that the appearance of distinct local minima in $\hat{M}(t)$ is a reasonable indication that the underlying $\lambda(x)$ incorporates at least two separate "scales of variation". We suggest that for an exploratory analysis, estimates using values of $t$ to correspond to the various local minima of $\hat{M}(t)$ highlight different features of the data, and together provide a useful summary description.

## 4. Application to Locations of Joints along a Coal Seam

Data from Shepherd *et al.* (1981) give the locations of 246 joints along a 168 metre stretch of coal seam. One question of interest concerns the possibility that the local intensity of joints increases in the vicinity of a geological fault. Early detection of such faults during the working of the seam has potentially important implications for the safety of the mining operation. Shepherd *et al.* present a $k$-nearest neighbour estimate of local intensity (Loftsgaarden and Quesenberry, 1965), but give no objective method of choosing the smoothing constant, $k$.

Figure 5a shows $\hat{M}(t)$ for these data, whilst Fig. 5b shows the two estimates $\hat{\lambda}_t(x)$ for values

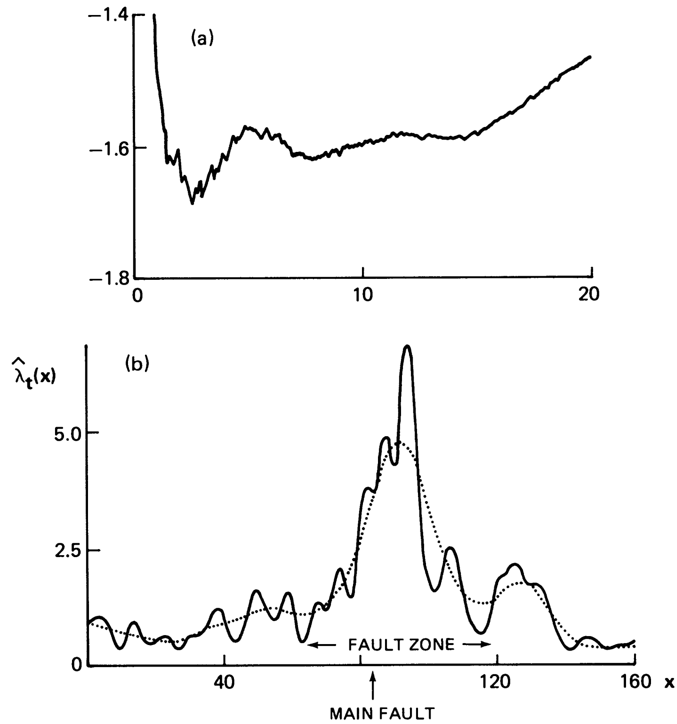


Fig. 5. Joint locations along a coal seam (Ivanhoe Colliery). (a) The empirical function $\hat{M}(t)$; (b) Two estimates of local intensity —— $\hat{\lambda}_{2.7}(x)$; . . . . . $\hat{\lambda}_{8.1}(x)$.

of $t$ corresponding to the two most obvious local minima in $\hat{M}(t)$. Also shown in Fig. 5b are the location of the main fault subsequently identified, and a geologist's definition of the associated fault zone. The increase in local intensity near the main fault seems unequivocal, and is summarized by the smoother of the two estimates. The other estimate describes smaller scale fluctuations in local intensity.

The complete data-set shows overwhelmingly significant departure from a homogeneous Poisson process. For example, $\hat{K}(t) - 2t$ increases steadily in the range 0 to 5 metres, from zero to about 6.45, whilst according to (2.2), the standard deviation under the Poisson assumption increases from zero to about 0.24. Less obviously, the data away from the fault zone exhibit significant aggregation. Let $\hat{K}_1(t)$ and $\hat{K}_2(t)$ denote estimates of $K(t)$ from the $n_1$ and $n_2$ points in the two sub-intervals before and after the declared fault zone. Pool these to give $\tilde{K}(t) = \{n_1\hat{K}_1(t) + n_2\hat{K}_2(t)\}/(n_1 + n_2)$. Figure 6 shows $\tilde{K}(t) - 2t$ together with tolerance limits of plus and minus two standard deviations deduced from (2.2). This strengthens the justification for our interpretation in terms of two distinct scales of variation in local intensity, although we acknowledge that the boundaries of the declared fault zone may be imprecise. Note also the significantly small estimates $\tilde{K}(t)$ near the origin, suggesting small-scale inhibitory interactions between points.

Qualitatively similar results were obtained for two further sets of data containing 504 and 405 joints in stretches of approximately 230 metres and 170 metres respectively. These are summarized by Fig. 7. In each case, the two values of $t$ used for $\hat{\lambda}_t(x)$ again correspond to obvious local minima in $\hat{M}(t)$.

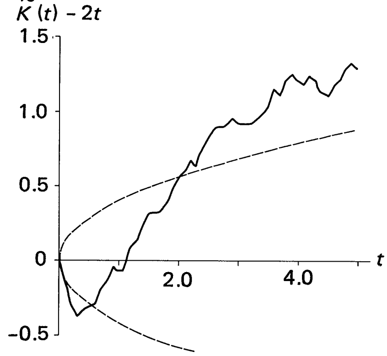


Fig. 6. Pooled estimate of $K(t)$ for data in two sub-intervals away from declared fault zone (Ivanhoe Colliery). —— $\tilde{K}(t) - 2t$; – – – plus and minus two standard deviations, deduced from (2.2).

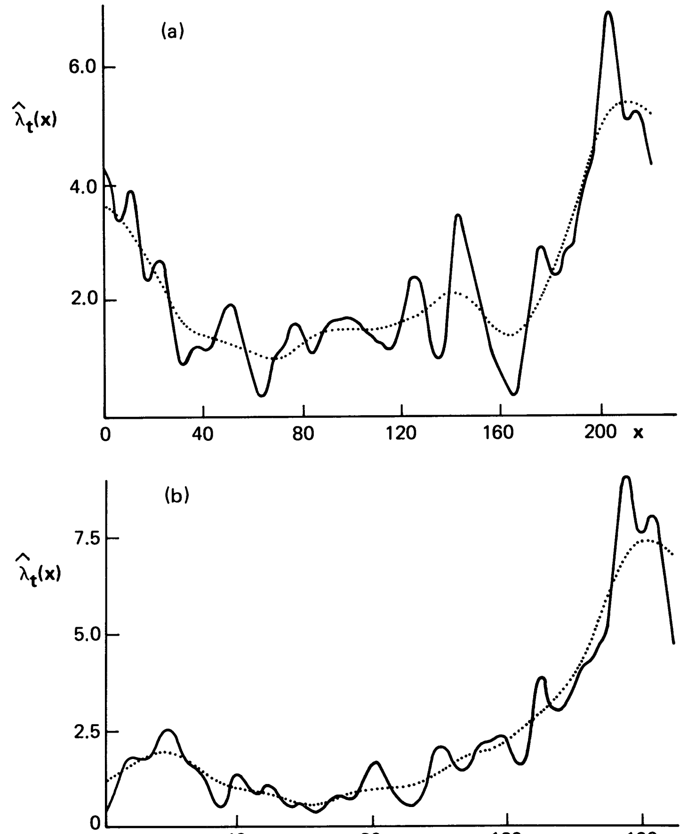


Fig. 7. Two further sets of joint locations (Wallerawang Colliery) (a) No 3 mine haulage road, $n = 504$ —— $\hat{\lambda}_{4.2}(x)$; . . . . . $\hat{\lambda}_{14.9}(x)$; (b) No. 3 mine travelling road, $n = 405$ —— $\hat{\lambda}_{3.7}(x)$; . . . . . $\hat{\lambda}_{11.6}(x)$.

The above analysis is clearly exploratory in nature, and does not answer the important practical question concerning early detection of faults. However, we suggest that it gives some food for thought and may therefore help in the development of more sophisticated, and more specifically directed, methods of analysis. In this respect, the similarity of the results obtained from the three separate sets of data is encouraging.

## 5. Discussion

The formal justification of our method of estimating local intensity relies on the assumption that the underlying point process is a Cox process. Informally, we regard it as a useful tool in the exploratory analysis of large, potentially heterogeneous sets of point process data. For a general point process, there is a real difficulty in sustaining the formal notion of local intensity as a realization of a stationary random process. The difficulty is acute in the case of clustering phenomena, because clustering of points may be mathematically indistinguishable from variation in local intensity. Indeed, the linear Cox process described in Section 2.2 has precisely this dual interpretation (Bartlett, 1964). Small-scale inhibitory interactions between points may also occur in practice, and were apparent in the data analysed in Section 4. However, these present fewer difficulties of interpretation. Their net effect is to include a local regularity into the pattern of points which, unlike clustering, is qualitatively different from variation in local intensity.

The method extends in principle to point patterns in two or more dimensions. The two-dimensional analogue of the Theorem in Section 2.2, based on counting the number of data-points in a disc of radius $t$, is

$$MSE(t) = \nu(0) + \mu\{1 - 2\mu K(t)\}/(\pi t^2) + (\pi t^2)^{-2} \iint \nu(\| \mathbf{y} - \mathbf{z} \|)\, d\mathbf{y}\, d\mathbf{z},$$

where $\| \cdot \|$ denotes Euclidean distance, each integral is over the disc with centre the origin and radius $t$ and $K(t) = 2\pi\mu^{-2} \int_0^t \nu(u)\, u\, du$. Unfortunately, the double integral cannot be reduced to an explicit formula involving only $\mu$ and $K(\cdot)$. This raises a practical difficulty of implementation since $\nu(\cdot)$ is a density whose estimation itself involves the solution to a non-trivial smoothing problem. However, a relatively crude estimate of $\nu(\cdot)$ may suffice for reasonably accurate estimation of the integral, except for small $t$. This possibility is currently being investigated.

## Acknowledgements

I thank Nick Fisher and Mark Westcott (CSIRO Division of Mathematics and Statistics) for helpful discussions, John Creasey (CSIRO Division of Mineral Physics) for supplying the data on joint locations and an anonymous referee for discovering a mistake in the original statement of the Theorem.

## References

Bartlett, M. S. (1964) Spectral analysis of two-dimensional point processes. *Biometrika*, **51**, 299–311.

Cox, D. R. and Isham, V. (1980) *Point Processes*. London: Chapman and Hall.

Loftsgaarden, D. O. and Quesenberry, C. P. (1965) A non-parametric estimate of a multivariate density function. *Ann. Math. Statist.*, **36**, 1049–1051.

Lotwick, H. W. & Silverman, B. W. (1982) Methods for analysing spatial processes of several types of points. *J. R. Statist. Soc.* B, **44**, 406–413.

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

Rosenblatt, M. (1956) Remarks on some non-parametric estimates of a density function. *Ann. Math. Statist.*, **27**, 832–7.

Shepherd, J. Creasey, J. W. and Fisher, N. I. (1981) Statistical analysis of spacings between geological discontinuities in coal mines, with applications to short-range forecasting of mining conditions. *Austral-Coal-Geol.*, **3**, 71–80.

Silverman, B. W. (1978) Choosing the window width when estimating a density. *Biometrika*, **65**, 1–11.