Ann. Inst. Statist. Math.
31 (1979), Part B, 145-155

MAXIMUM LIKELIHOOD ESTIMATION OF HAWKES' SELF-EXCITING POINT PROCESSES

T. OZAKI

(Received Sept. 30, 1977; revised Jan. 12, 1979)

Summary

A maximum likelihood estimation procedure of Hawkes' self-exciting point process model is proposed with explicit presentations of the log-likelihood of the model and its gradient and Hessian. A simulation method of the process is also presented. Some numerical results are given.

1. Introduction

Recently problems of estimation, filtering and smoothing of point processes have been discussed by many authors (Vere-Jones [13], Snyder [12], Segal [11]). However, as is observed by Vere-Jones [13] no satisfactory solution of the parameter estimation problem for non-trivial point processes has emerged. In this paper we propose a maximum likelihood estimation procedure of a point process model called Hawkes' self-exciting point process model (or briefly Hawkes' model).

Let $N(t)$ be a point process such that

$$\Pr\{\Delta N(t)=1 \mid N(s)\ (s\leqq t)\}=\Lambda(t)\Delta t+o(\Delta t)$$

$$\Pr\{\Delta N(t)>1 \mid N(s)\ (s\leqq t)\}=o(\Delta t)\ .$$

Hawkes [3], [4] introduced a general point process model whose intensity function is given by

$$\Lambda(t)=\mu+\int_{-\infty}^{t} g(t-u)dN(u)\ ,$$

where $g(\cdot)\geqq 0$ and $\int_0^\infty g(u)du<1$. We call the function $g(t)$ the response function. In this paper we focus our attention on the Hawkes' model with a response function

$$g(t)=\alpha e^{-\beta t}\ .$$

We note that some probabilistic structure of the model is analyzed by Hawkes and Oakes [5].

2. Log-likelihood of the Hawkes' model

Given the occurrence observation $t_1, t_2, \cdots, t_n$ for an interval $[0, T]$ $(T \geqq t_n)$, the log-likelihood of a point process with a intensity function

$$\Lambda(t|\theta)=\mu+\int_{-\infty}^{t} g(t-u|\theta)dN(u)$$

is given (Rubin [10]) by

$$\log L(t_1, \cdots, t_n|\theta)=-\int_0^T \Lambda(t|\theta)dt+\int_0^T \log \Lambda(t|\theta)dN(t) ,$$

where $\theta=(\theta_1, \theta_2, \cdots, \theta_r)$. The gradient of the log-likelihood is given by

$$\frac{\partial \log L}{\partial \theta_i}=-\int_0^T \frac{\partial \Lambda(t|\theta)}{\partial \theta_i}dt+\int_0^T \left(\frac{\partial \Lambda(t|\theta)}{\partial \theta_i}\Big/\Lambda(t|\theta)\right)dN(t) \qquad (i=1, \cdots, r) ,$$

and the Hessian of the log-likelihood is given by

$$\frac{\partial^2 \log L}{\partial \theta_j \partial \theta_i}=-\int_0^T \frac{\partial^2 \Lambda(t|\theta)}{\partial \theta_j \partial \theta_i}dt+\int_0^T \left[\frac{\dfrac{\partial^2 \Lambda(t|\theta)}{\partial \theta_j \partial \theta_i}\Lambda(t|\theta)-\dfrac{\partial \Lambda(t|\theta)}{\partial \theta_i}\dfrac{\partial \Lambda(t|\theta)}{\partial \theta_j}}{\Lambda(t|\theta)^2}\right]dN(t) \qquad (i, j=1, \cdots, r) ,$$

where $\Lambda(t|\theta)$ is supposed to satisfy necessary regularity conditions.

We note that the Rubin's log-likelihood is defined under the assumption that the occurrence observation $t_1, \cdots, t_n$ are observed from the beginning of the process, i.e. the time zero, and the log-likelihood is given at the time T $(\geqq t_n)$. However in most identification problems, only $t_1, \cdots, t_n$ are given and T is not specified. Thus we assume $T=t_n$ in the rest of the present paper.

In the case of the Hawkes' self-exciting process model whose response function is $\alpha e^{-\beta t}$, the log-likelihood of the model is given by

$$\log L(t_1, \cdots, t_n|\theta)=-\mu t_n+\sum_{i=1}^{n} \frac{\alpha}{\beta}(e^{-\beta(t_n-t_i)}-1)+\sum_{i=1}^{n} \log \{\mu+\alpha A(i)\} ,$$

where $A(i)=\sum_{t_j<t_i} e^{-\beta(t_i-t_j)}$ for $i \geqq 2$ and t_i denotes the time of occurrence of the ith event, and $A(1)=0$. The gradients are

$$\frac{\partial \log L}{\partial \alpha}=\sum_{i=1}^{n} \frac{1}{\beta}(e^{-\beta(t_n-t_i)}-1)+\sum_{i=1}^{n}\left[\frac{A(i)}{\mu+\alpha A(i)}\right] ,$$

$$\frac{\partial \log L}{\partial \beta} = -\alpha \sum_{i=1}^{n} \left[\frac{1}{\beta}(t_n - t_i)e^{-\beta(t_n - t_i)} + \frac{1}{\beta^2}(e^{-\beta(t_n - t_i)}) \right] - \sum_{i=1}^{n} \left[\frac{\alpha B(i)}{\mu + \alpha A(i)} \right],$$

$$\frac{\partial \log L}{\partial \mu} = -t_n + \sum_{i=1}^{n} \left[\frac{1}{\mu + \alpha A(i)} \right],$$

where $B(i) = \sum_{t_j < t_i} (t_i - t_j)e^{-\beta(t_i - t_j)}$ for $i \geqq 2$ and $B(1) = 0$. The Hessian is defined by

$$\frac{\partial^2 \log L}{\partial \alpha^2} = -\sum_{i=1}^{n} \left[\frac{A(i)}{\mu + \alpha A(i)} \right]^2$$

$$\frac{\partial^2 \log L}{\partial \beta \partial \alpha} = -\sum_{i=1}^{n} \left[\frac{1}{\beta}(t_n - t_i)e^{-\beta(t_n - t_i)} + \frac{1}{\beta^2}(e^{-\beta(t_n - t_i)} - 1) \right] + \sum_{i=1}^{n} \left[\frac{-B(i)}{\mu + \alpha A(i)} + \frac{\alpha A(i) B(i)}{(\mu + \alpha A(i))^2} \right],$$

$$\frac{\partial^2 \log L}{\partial \beta^2} = \alpha \sum_{i=1}^{n} \left[\frac{1}{\beta}(t_n - t_i)^2 e^{-\beta(t_n - t_i)} + \frac{2}{\beta^2}(t_n - t_i)e^{-\beta(t_n - t_i)} + \frac{2}{\beta^3}(e^{-\beta(t_n - t_i)} - 1) \right] + \sum_{i=1}^{n} \left[\frac{\alpha c(i)}{\mu + \alpha A(i)} - \left(\frac{\alpha B(i)}{\mu + \alpha A(i)} \right)^2 \right]$$

where

$$c(i) = \sum_{t_j < t_i} (t_i - t_j)^2 e^{-\beta(t_i - t_j)} \qquad \text{for } i \geqq 2 \text{ and } c(1) = 0,$$

$$\frac{\partial^2 \log L}{\partial \mu^2} = \sum_{i=1}^{n} \left[\frac{-1}{(\mu + \alpha A(i))^2} \right], \qquad \frac{\partial^2 \log L}{\partial \alpha \partial \mu} = \sum_{i=1}^{n} \left[\frac{-A(i)}{(\mu + \alpha A(i))^2} \right],$$

$$\frac{\partial^2 \log L}{\partial \beta \partial \mu} = \sum_{i=1}^{n} \left[\frac{\alpha B(i)}{(\mu + \alpha A(i))^2} \right].$$

3. Likelihood maximization

As the log-likelihood of the Hawkes' model is non-linear with respect to the parameters, the maximization of the log-likelihood is performed by using non-linear optimization techniques. There are three types of non-linear optimization techniques available for this purpose.

(i) When we employ gradient and Hessian of the objective function (as well as the function value) at each updating stage of the function maximization, we can use the well-known Newton-Raphson method.

(ii) However the Hessian can be approximated by using the gradient at each updating stage and direct evaluation of the second derivative of the function is not necessary at least in updating stages of the optimization procedure. Such a procedure which employs only gradient information and function values is called gradient method. One of the most efficient gradient method is Davidon's procedure (Davidon, W. C. [1], Fletcher, R. and Powell, M. J. D. [2]).

(iii) Maximization of a function can be performed employing only the function value at each updating stage. This is useful when the evaluation of the first or second derivative needs complicated computations. We call this kind of procedures direct method. There are many direct methods (Rosenbrock, H. H. [9], Hooke, R. and Jeeves, T. A. [6], Powell [8] etc.).

Although Powell's method employs only information of function values in the updating stages it is a kind of conjugate gradient method and the speed of convergence is high compared to other direct methods.

As we have overviewed, the maximum likelihood estimates of the Hawkes' model with exponentially decaying response function can be obtained by employing any one of the above procedures using the explicit representation of log-likelihood, gradient and Hessians given in the preceding section.

4. Simulation method

We need artificial self-exciting point process data to check the validity of the maximum likelihood method. Suppose that self-exciting point process data $t_1,\cdots,t_k$ are given. Let $F(t|t_1,\cdots,t_k,\theta)$ be conditional distribution of random variable of the interval between t_k and the next event t $(t\geqq t_k)$ of the process, and let $f(t|t_1,\cdots,t_k,\theta)$ be its probability density function.

The conditional Hazard function is given by

$$\frac{f(t|t_1,\cdots,t_k,\theta)}{1-F(t|t_1,\cdots,t_k,\theta)}=\Lambda(t|t_1,\cdots,t_k,\theta)\ ,$$

and we have

$$\begin{aligned}\log\{1-F(u|t_1,\cdots,t_k,\theta)\}&=-\int_{t_k}^{u}\Lambda(t|t_1,\cdots,t_k,\theta)dt\\&=-\int_{t_k}^{u}\left\{\mu+\sum_{i=1}^{k}g(t-t_i|\theta)\right\}dt\ .\end{aligned}$$

Since the time of the $k+1$st event u satisfies this equation and $1-F(u|t_1,\cdots,t_k,\theta)$ is distributed uniformly on $[0,1]$, generation of t_{k+1} is performed by generating a uniform random number U and solving the following equation with respect to u,

$$\log U+\int_{t_k}^{u}\left\{\mu+\sum_{i=1}^{k}g(t-t_i|\theta)\right\}dt=0\ . \tag{4.1}$$

In the case of exponentially decaying response function

$$g(t)=\alpha e^{-\beta t}$$

(4.1) reduces to the following transcendental equation

$$\log U+\mu(u-t_k)-\frac{\alpha}{\beta}\left\{\sum_{i=1}^{k} e^{-\beta(u-t_i)}-\sum_{i=1}^{k} e^{-\beta(t_k-t_i)}\right\}=0 . \tag{4.2}$$

This equation (4.2) has the following recursive expression which is suggested by Dr. H. Akaike,

$$\log U+\mu(u-t_k)+\frac{\alpha}{\beta} S(k)\{1-e^{-\beta(u-t_k)}\}=0 \tag{4.3}$$

where

$$S(1)=1$$

$$S(i)=e^{-\beta(t_i-t_{i-1})}S(i-1)+1 \qquad (i\geqq 2) .$$

The algorithm for the generation of Hawkes' self-exciting process data is thus described as follows:

Algorithm

1. Generate a uniform random number U on [0, 1].
2. Let $t_1=-\log U/\mu$.
3. Generate a uniform random number U on [0, 1].
4. Solve (4.3) with respect to U and get a solution $\boldsymbol{u}$.
5. Let $t_{k+1}=\boldsymbol{u}$ and

$$S(k+1)=e^{-\beta(t_{k+1}-t_k)}S(k)+1 .$$

If $k+1=n$ stop, otherwise increase k by 1 and jump to the stage 3.

The equation (4.3) can be numerically solved with respect to u by Newton's iterative method which updates u_i by the formula

$$u_{i+1}=u_i-\frac{f(u_i)}{f'(u_i)} ,$$

where

$$f(u)=\log U+\mu(u-t_k)+\frac{\alpha}{\beta} S(k)\{1-e^{-\beta(u-t_k)}\}$$

$$f'(u)=\mu+\alpha S(k)e^{-\beta(u-t_k)} .$$

The initial value of this iteration is taken to be

$$t_k-\log U/\mu$$

which is the solution of (4.2) when the process is Poisson.

5. Numerical results

Employing the method of generation of Hawkes' self-exciting process stated in the preceding section we generated artificial data and applied the three methods of the maximum likelihood estimation procedure which are mentioned in the previous section to check the validity of the estimation procedure. We generated two sets of data $\{t_i;\ i=1, 2, \cdots, N\}$, each with $N=500$. One is generated from the model whose parameters are $\alpha=4.0$, $\beta=5.0$ and $\mu=0.5$ (which we abbreviate as (4.0, 5.0, 0.5)), and another is from the model $(\alpha, \beta, \mu)=(0.8, 1.0, 0.5)$. As is well-known the clustering size of the Hawkes' model is given by

$$c=1\Big/\left(1-\int_0^{\infty} g(t)\,dt\right)$$

and the mean rate λ of the whole process is given by $\lambda=c\mu$. The two sets have the same clustering size and the same mean rate but the decaying behaviour of the intensity differs significantly. The difference is clearly seen in the behaviour of the intensity process $\Lambda(t\,|\,\alpha, \beta, \mu)$ (see Fig. 5 and Fig. 7).

The maximum likelihood estimates and their estimated variance-covariance matrix of the first model (4.0, 5.0, 0.5) is shown in Table 1 and those of the second model (0.8, 1.0, 0.5) in Table 2. Since the

Table 1. Maximum likelihood estimates and variance-covariance matrix of the estimates of the model (4.0, 5.0, 0.5)

	α	β	μ
true	4.0	5.0	0.5
estimated	3.968	5.174	0.459
variance-covariance matrix	0.227×10^{-1}	0.845×10^{-2} 0.315×10^{-2}	0.590×10^{-2} 0.220×10^{-2} 0.153×10^{-2}

Table 2. Maximum likelihood estimates and variance-covariance matrix of the estimates of the model (0.8, 1.0, 0.5)

	α	β	μ
true	0.8	1.0	0.5
estimated	0.684	1.018	0.672
variance-covariance matrix	0.140×10^{-1}	0.194×10^{-1} 0.373×10^{-1}	0.502×10^{-3} 0.105×10^{-1} 0.153×10^{-1}

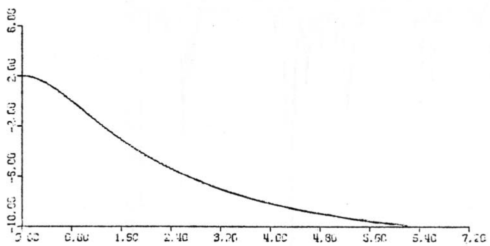

Fig. 1. Theoretical spectrum of the model $(\alpha, \beta, \mu)=(4.0, 5.0, 0.5)$.

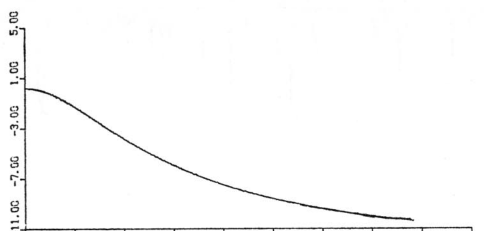

Fig. 2. Estimated spectrum through the estimated model $(\hat{\alpha}, \hat{\beta}, \hat{\mu})=(3.97, 5.17, 0.46)$.

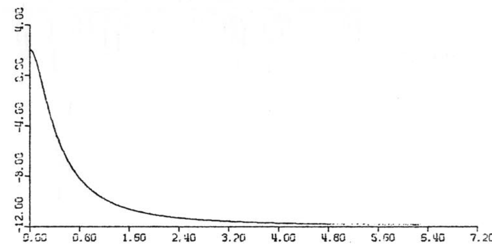

Fig. 3. Theoretical spectrum of the model $(\alpha, \beta, \mu)=(0.8, 1.0, 0.5)$.

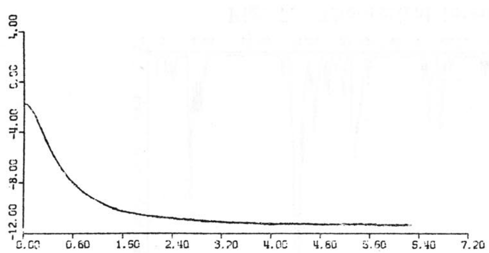

Fig. 4. Estimated spectrum through the estimated model $(\hat{\alpha}, \hat{\beta}, \hat{\mu})=(0.68, 1.01, 0.67)$.

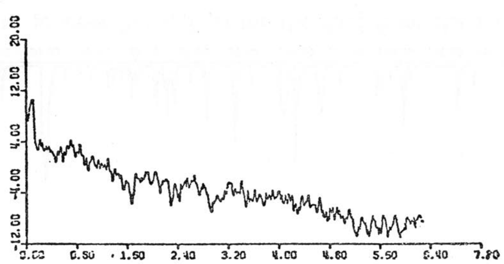

Fig. 5. The spectrum estimated by smoothing the periodogram.

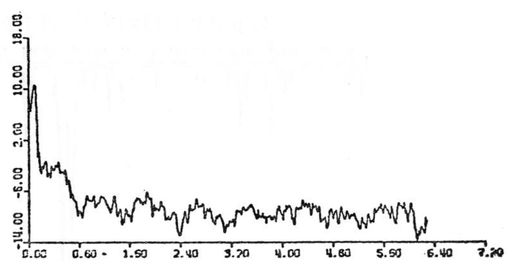

Fig. 6. The spectrum estimated by smoothing the periodogram.

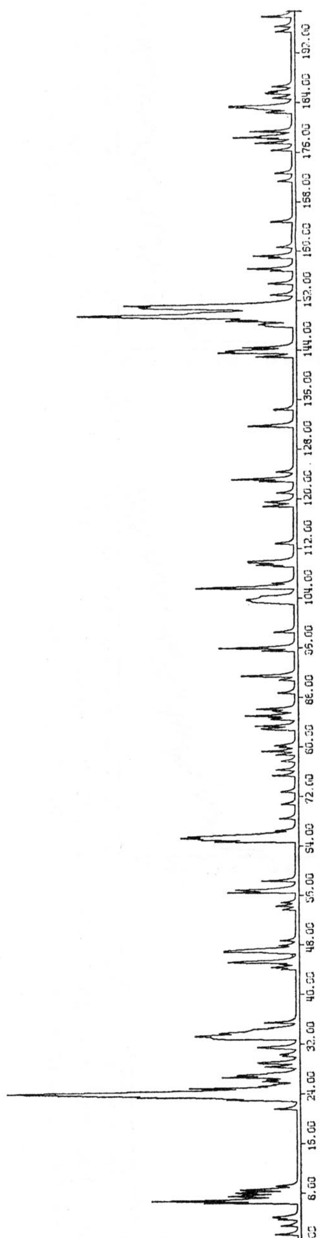

Fig. 7. Theoretical intensity process $\lambda(t\,|\,\alpha, \beta, \mu)$ for the data from the model $(\alpha, \beta, \mu)=(4.0, 5.0, 0.5)$.

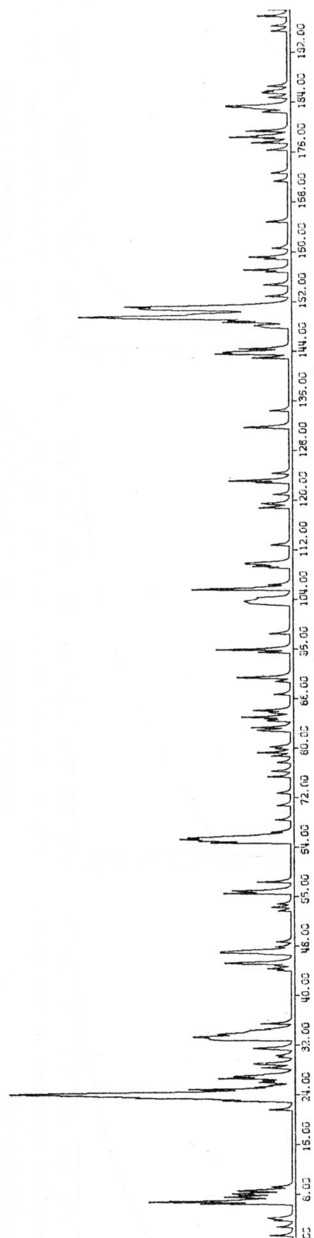

Fig. 8. Estimated intensity process $\lambda(t\,|\,\hat{\alpha}, \hat{\beta}, \hat{\mu})$ through the model $(\hat{\alpha}, \hat{\beta}, \hat{\mu})=(3.97, 5.17, 0.46)$.

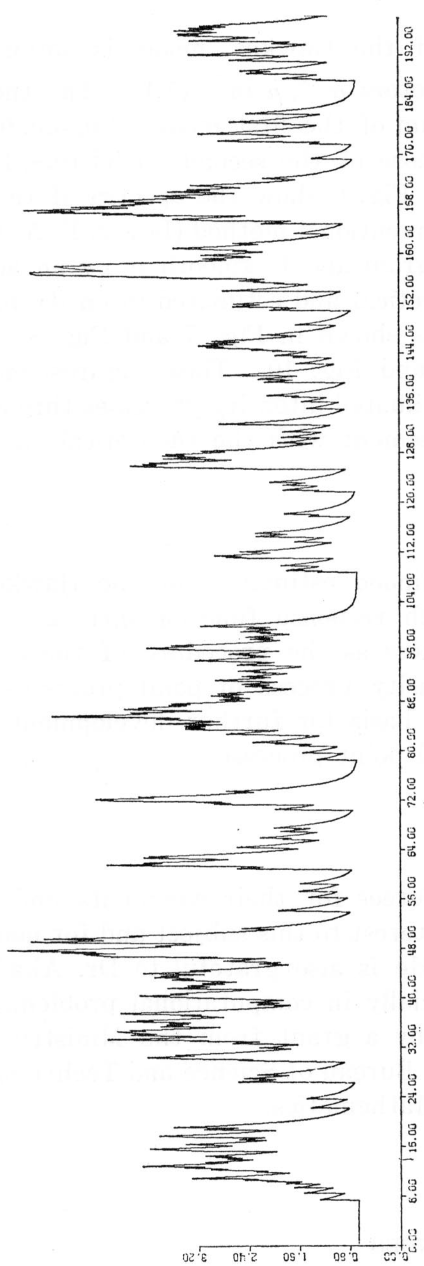

Fig. 9. Theoretical intensity process $\lambda(t \mid \alpha, \beta, \mu)$ for the data from the model $(\alpha, \beta, \mu)=(0.8, 1.0, 0.5)$.

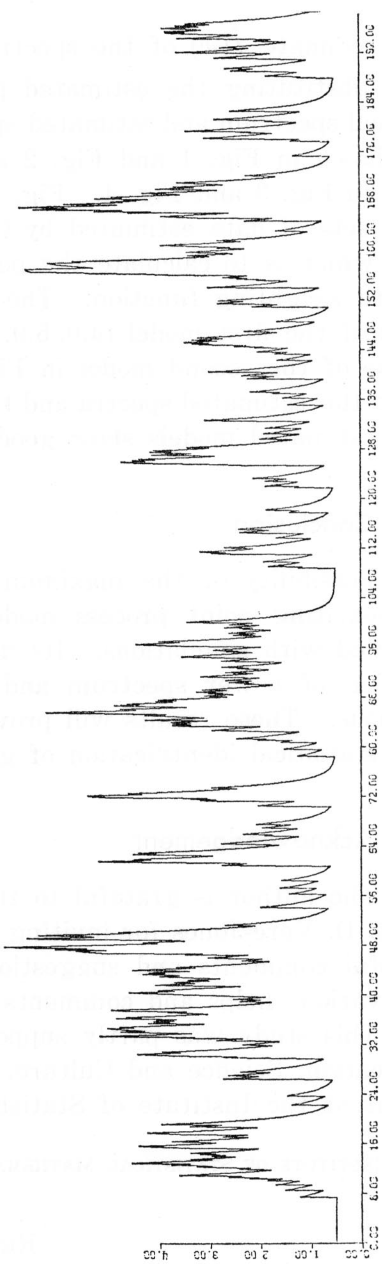

Fig. 10. Estimated intensity process $\lambda(t \mid \hat{\alpha}, \hat{\beta}, \hat{\mu})$ through the model $(\hat{\alpha}, \hat{\beta}, \hat{\mu})=(0.68, 1.01, 0.67)$.

spectrum of the Hawkes' model (α, β, μ) is given (Hawkes [3], [4]) by

$$f(w)=\frac{\mu}{2\pi}\frac{\beta}{\beta-\alpha}\left\{1+\frac{\alpha(2\beta-\alpha)}{(\beta-\alpha)^2+w^2}\right\}, \tag{5.1}$$

the estimate $\hat{f}(w)$ of the spectrum of the Hawkes' model is obtained by substituting the estimated parameters $\hat{\alpha}$, $\hat{\beta}$, $\hat{\mu}$ into (5.1). The theoretical spectrum and estimated spectrum of the first model (4.0, 5.0, 0.5) is shown in Fig. 1 and Fig. 2 and those of the second model (0.8, 1.0, 0.5) in Fig. 3 and Fig. 4. Fig. 5 and Fig. 6 show the spectra of those two sets of data estimated by the conventional method (Lewis, P. A. W. [7]) which is to calculate the periodogram and to smooth it by an adequate weighting function. The theoretical and estimated intensity process of the first model (4.0, 5.0, 0.5) is shown in Fig. 7 and Fig. 8 and those of the second model in Fig. 9 and Fig. 10. These figures show that the estimated spectra and the estimated intensity processes through the estimated models show good agreement with the theoretical ones.

6. Conclusion

Possibility of the maximum likelihood estimation of the Hawkes' self-exciting point process model with response function $g(t)=\alpha e^{-\beta t}$ is verified with simulations. Its usefulness as the procedure of the estimation of power spectrum and intensity process of point processes is obvious. These results will provide a basis for further development of the statistical identification of general point processes.

7. Acknowledgement

The author is grateful to the referees for their comments and to Prof. D. Vere-Jones for inviting his interest to this subject and for many helpful comments and suggestions. He is also grateful to Dr. Akaike for various helps and comments especially in computational problems.

This study was partly supported by a grant from the Ministry of Education, Science and Culture, and a Bureau of Science and Technology grant at the Institute of Statistical Mathematics.

THE INSTITUTE OF STATISTICAL MATHEMATICS

REFERENCES

[1] Davidon, W. C. (1959). Variable metric method for minimization, *Argonne Natl. Lab. Report*, No. 5990.

[2] Fletcher, R. and Powell, M. J. D. (1963). A rapidly convergent descent method for minimization, *Computer J.*, **6**, 163–168.

[3] Hawkes, A. G. (1971a). Spectra of self-exciting and mutually exciting point processes, *Biometrika*, **58**, 83-90.

[4] Hawkes, A. G. (1971b). Point spectra of some mutually exciting point processes, *J. R. Statist. Soc.*, B, **33**, 438-443.

[5] Hawkes, A. G. and Oakes, D. A. (1974). A cluster process representation of a self-exciting process, *J. Appl. Prob.*, **11**. 493-503.

[6] Hooke, R. and Jeeves, T. A. (1961). Direct search solution of numerical and statistical problems, *J. Ass. Compt. Mach.*, 8, 212-229.

[7] Lewis, P. A. W. (1970). Remarks on the theory, computation and application of the spectral analysis of series of events, *J. Sound Vid.*, **12**, 353-375.

[8] Powell, M. J. D. (1963). An efficient method of finding the minimum of a function of several variables without calculating derivatives, *Computer. J.*, **7**, 155-162.

[9] Rosenbrock, H. H. (1960). An automatic method for finding the greatest or the least value of a function, *Computer J.*, **3**, 175-184.

[10] Rubin, I. (1972). Regular point processes and their detection, *IEEE Trans. Inf. Theory*, IT-**18**, 547-557.

[11] Segal, A. (1976). Recursive estimation from discrete-time point processes, *IEEE Trans. Inf. Theory*, IT-**22**, 422-431.

[12] Snyder, D. L. (1972). Smoothing for doubly stochastic Poisson processes, *IEEE Trans. Inf. Theory*, IT-**18**, 558-562.

[13] Vere-Jones, D. (1975). On updating algorithms and inference for stochastic point processes, *M. S. Bartlett Memorial Volume.*