

Self-Exciting 点過程モデルの最尤推定

統計数理研 尾崎 統

0. はじめに 地震のように突発的な事象や神経パルス系列のように離散的な確率現象は点過程と呼ばれ、その研究は主に日本以外の確率学者、通信工学者、統計学者達によって発展させられてきている。最近の通信工学におけるデジタル光通信や、放射線医学などの分野からの要請により、点過程の解明は今後連続過程に劣らず重要になってくると思われる。しかし近年の連続過程の理論応用両面にわたる著しい発展に比べるとその発展、特に統計的側面の発展は遅れている。その主な原因の一つは連続過程における AR モデルや ARMA モデルのように過去従属的 (history-dependent) な、一般的モデルで現実に適用可能なものがなかったことにある。点過程におけるこれまでの統計的な面での関心は主にスムージングによるスペクトル推定や、ワリアンスタインカーブの推定などのノンパラメトリックな面に注がれており、それを通して与えられた点過程のポアソン過程からの隔たりを議論す

りに留まっていた。統計的推定の対象になるパラメトリックモデルはポアソン過程や周期的 (periodic) ポアソン過程のように過去と現在の事象が独立なつまり (trivial) 場合に限られていた。本稿の目的は点過程においても一般的な過去従属的モデルによるパラメトリックな接近が可能であることを示すことにある。

1. 点過程のモデル化 今点過程 $N(t)$ は次の条件を満たすとする。

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

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

Hawkes (1971a) は intensity 関数 $\Lambda(t)$ が過去の事象に線型に従属する, 即ち

$$\Lambda(t) = \mu + \int_{-\infty}^t g(t-u) dN(u) \quad (1.1)$$

で与えられる一般的なモデルを導入した。このモデルは一般的点過程の線型近似としての意味を持つという点で連続過程の AR モデルに対応しているとみることもできる。我々はこれを Hawkes の self-exciting モデルと呼ぶことにする。又 (1.1) の $g(t)$ を応答関数と呼ぶ。この (1.1) モデルについての理論的研究は多くなされている。(例えば prob. gen. fl. については Hawkes & Oakes (1974), スペクトルについては Hawkes (1971a), (1971b), interval の分布については Adamopoulos (1975), intensity $\Lambda(t)$ の条件付期待値の動的な性質について

では Snyder (1975), Van Schuppen (1977) など) しかし奇妙なことにパラメータの推定値を求めることは困難とされ今までなされていなかった。以下 λ の方法と实例を順に述べていく。

2. 対数尤度 intensity $\Lambda(t|\theta)$ を持つ点過程 $N(t|\theta)$ から生成された標本 t_1, \dots, t_n の対数尤度は

$$\log L(\theta|t_1, \dots, t_n) = -\int_0^{t_n} \Lambda(t|\theta) dt + \int_0^{t_n} \log \Lambda(t|\theta) dN(t|\theta) \quad (2.1)$$

で与えられる (Rubin (1972))。

ここに $\theta = (\theta_1, \dots, \theta_r)$ 。 λ の gradient は

$$\frac{\partial \log L}{\partial \theta_i} = -\int_0^{t_n} \frac{\partial \Lambda(t|\theta)}{\partial \theta_i} dt + \int_0^{t_n} \frac{\frac{\partial \Lambda(t|\theta)}{\partial \theta_i}}{\Lambda(t|\theta)} dN(t|\theta) \quad (2.2)$$

($i=1, \dots, r$)

Hessian は

$$\frac{\partial^2 \log L}{\partial \theta_j \partial \theta_i} = -\int_0^{t_n} \frac{\partial^2 \Lambda(t|\theta)}{\partial \theta_j \partial \theta_i} dt + \int_0^{t_n} \left[\frac{\frac{\partial^2 \Lambda(t|\theta)}{\partial \theta_j \partial \theta_i} - \frac{\partial \Lambda(t|\theta)}{\partial \theta_i} \frac{\partial \Lambda(t|\theta)}{\partial \theta_j}}{\Lambda(t|\theta)^2} \right] dN(t|\theta) \quad (2.3)$$

($i, j=1, \dots, r$)

で与えられる。ここに $\Lambda(t|\theta)$ は正則条件を満たす。

self-exciting モデルの具体的な型として最もよく議論されるものは $g(t) = \alpha e^{-\beta t}$ の場合であり、以下この場合について議論を進めていく。この場合尤度 (2.1) は

$$\log L(\theta|t_1, \dots, t_n) = -\mu t_n + \sum_{i=1}^n \frac{\alpha}{\beta} (e^{-\beta(t_n - t_i)} - 1) + \sum_{i=1}^n \log \{\mu + \alpha \cdot A(i)\} \quad (2.4)$$

ここに $A(i) = \sum_{t_j < t_i} e^{-\rho(t_i - t_j)}$. この場合の gradient, Hessian もパラメータを用いて直接表現できる (Ozaki (1977)).

3. 最尤計算 式 (2.4) にあるように self-exciting モデルの尤度はパラメータに對し非線型である為、最尤推定値を得るには非線型最適化が必要となる。非線型最適化の手法としては現在以下の三種類のタイプがある。

(i) 目的函数である尤度の他に λ の gradient と Hessian の評価を各繰り返し段階に用いることができる場合はよく知られた Newton-Raphson 法によって最適化することができる。

(ii) しかし各繰り返し段階において Hessian は gradient を用いて近似することができる。Hessian の直接的な評価は必ずしも必要でない。函数値と gradient だけを用いて行う最適化手法を gradient 法という。gradient 法の中で最も効率のよいものとして Davidon-Fletcher-Powell の方法である。(Davidon (1959), Fletcher & Powell (1963))

(iii) 函数値の最大化 (あるいは最小化) は各繰り返し段階において函数値だけを用いても行うことができる。たとえば Rosenbrock (1960), Powell (1964) などがある。函数値の一次微分と二次微分の評価が得られない時又は評価法が非常に複雑な場合はこの方法が有用である。

従って Hawkes の self-exciting モデルの最尤推定値は前節の程度, gradient, Hessian のパラメータによる直接表現と上記の最適化手法によって得ることが出来る。

4. シミュレーションの為の人工的データの発生 Algorithm.

最尤推定法の有効さを実際にチェックする為には人工的データが必要となる。過去従属的、点過程モデルの場合人工的データの発生は連続過程における AR モデルや ARMA モデルの時のように容易ではない。それは事象が起きる度に条件付 Hazard 関数に変化する為である。今、事象 t_1, \dots, t_k が与えられているとして次の事象 u ($u \geq t_k$) と t_k の間の間隔の条件付分布を $F(u|t_1, \dots, t_k, \theta)$ とし、その密度関数を $f(u|t_1, \dots, t_k, \theta)$ とする時、条件付 Hazard 関数は

$$\frac{f(u|t_1, \dots, t_k, \theta)}{1 - F(u|t_1, \dots, t_k, \theta)} = \lambda(u|t_1, \dots, t_k, \theta) \quad (4.1)$$

であるから

$$\begin{aligned} \log\{1 - F(u|t_1, \dots, t_k, \theta)\} &= - \int_{t_k}^u \lambda(t|t_1, \dots, t_k, \theta) dt \\ &= - \int_{t_k}^u \left\{ \mu + \sum_{i=1}^k g(t-t_i, \theta) \right\} dt \end{aligned} \quad (4.2)$$

を得る。 $1 - F(u|t_1, \dots, t_k, \theta)$ は $[0, 1]$ 上に一様に分布していることから結局 t_{k+1} の生成は一様乱数 U を生成し

$$\log U + \int_{t_k}^{t_{k+1}} \left\{ \mu + \sum_{i=1}^k g(t-t_i, \theta) \right\} dt = 0 \quad (4.3)$$

を満足する u を求めることに帰着する。応答函数 $g(t) = \alpha e^{-\beta t}$ の場合, (4.3) は

$$\log U + \mu(u - t_k) - \frac{\alpha}{\beta} \left\{ \sum_{i=1}^k e^{-\beta(u-t_i)} - \sum_{i=1}^k e^{-\beta(t_k-t_i)} \right\} = 0 \quad (4.4)$$

に帰着する。この式は次のような帰納的表現を持つ。

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

ここに $S(1) = 1$ (4.5)

$$S(i) = e^{-\beta(t_i - t_{i-1})} \cdot S(i-1) + 1 \quad (i \geq 2)$$

従って応答函数 $g(t) = \alpha e^{-\beta t}$ の時の self-exciting モデルのデータ発生アルゴリズムは次のようになる。

Algorithm

1. $[0, 1]$ 上の一様乱数 U を生成する。
2. $t_1 = \log U / \mu$, $k=1$, $S(1)=1$ とする。
3. $[0, 1]$ 上の一様乱数 U を生成する。
4. (4.5) 式を u に関して解き解 u を得る。
5. $t_{k+1} = u$, $S(k+1) = e^{-\beta(t_{k+1} - t_k)} \cdot S(k) + 1$ とする。
もし $k+1 = n$ なる終了, そうでなければ k を 1 増やしてステップ 3 へ飛ぶ。

超越方程式 (4.5) は各段階において u_i を

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

によって更新する Newton 法を用いて数値的に解く。ここに

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

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

初期値 u_0 はプロセスが過去の事象に独立とした場合の解,

$$u_0 = t_k + \log U / \mu$$

を用いる。この時 $f(u)$ は単調増加関数で、 λ の解は一意的であり、解への収束は速い。

5. 数値実験例 前節の方法を用いて 500 点のデータを二組生成した。一つはモデル

$$\Lambda(t | \alpha, \beta, \mu) = \mu + \int_{-\infty}^t \alpha e^{-\beta(t-w)} dN(w)$$

において、 $\mu = 0.5$, $\alpha = 4.0$, $\beta = 5.0$, から、もう一つは、 $\mu = 0.5$, $\alpha = 0.8$, $\beta = 1.0$, から生成した。

よく知られているように self-exciting モデルのクラスターサイズ C は

$$C = \frac{1}{1 - \int_0^{\infty} g(t) dt} \quad (5.1)$$

で与えられる、平均レート (mean rate) は

$$\lambda = C \cdot \mu \quad (5.2)$$

で与えられる。従って上記の二つのモデルは同じクラスターサイズ、同じ平均レートを持つ。しかし β の違いから、両者の intensity 過程 $\Lambda(t | \alpha, \beta, \mu)$ のふるまいは図 1, 2 の如く

大きく違う。

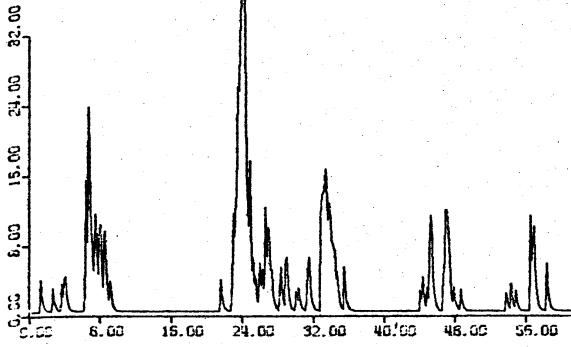


図1. モデル1の $\Lambda(t|\alpha, \beta, \mu)$

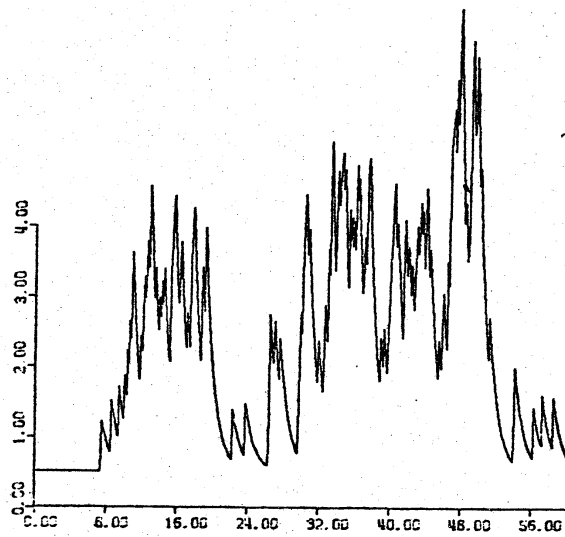


図2. モデル2の $\Lambda(t|\alpha, \beta, \mu)$

又、この場合のスペクトル $f(\omega)$ は、

$$f(\omega) = \frac{\lambda}{2\pi} \left\{ 1 + \frac{\alpha(2\beta - \alpha)}{(\beta - \alpha)^2 + \omega^2} \right\} \quad (5.3)$$

で与えられる(Hawkes (1971a)), 第一のモデルのスペクトルは図3のように、第二のモデルのスペクトルは図4のようになる。

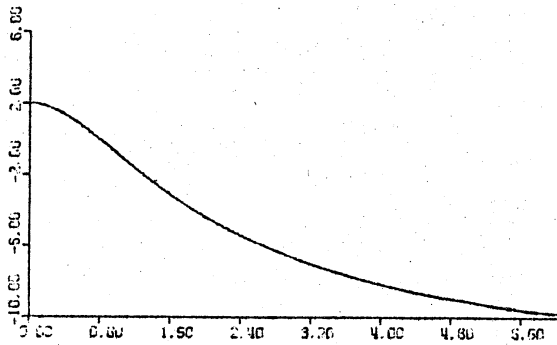


図3. モデル1の理論スペクトル

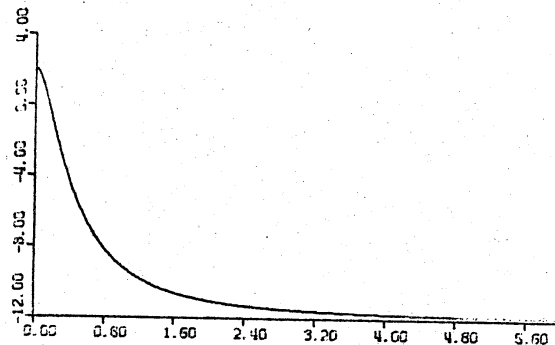


図4. モデル2の理論スペクトル

第一のモデルから生成したデータに最尤推定法を適用した結果が表1に、第二のモデルから生成したデータに適用した

結果が表2に示されている。

	α	β	μ
真値	4.0	5.0	0.5
推定値	3.968	5.174	0.459
分散- 共分散 行列	0.227×10^{-1}	0.845×10^{-2} 0.315×10^{-2}	0.590×10^{-2} 0.220×10^{-2} 0.153×10^{-2}

	α	β	μ
真値	0.8	1.0	0.5
推定値	0.684	1.018	0.672
分散- 共分散 行列	0.140×10^{-1}	0.194×10^{-1} 0.373×10^{-1}	0.502×10^{-2} 0.105×10^{-1} 0.153×10^{-1}

表1. モデル1の最尤推定値

表2. モデル2の最尤推定値

推定値 $\hat{\alpha}$, $\hat{\beta}$, $\hat{\mu}$ を (5.3) に代入して得られる推定スペクトル $\hat{f}(\omega)$ は第一のモデルの場合図5のようになり, 第二のモデルの場合図6のようになり両者とも真のスペクトルの形をよく復原する。図7と図8は各, 従来のピリオドグラムをスムースするやり方で求めた推定スペクトルである。

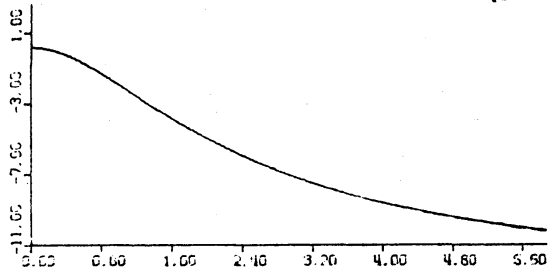


図5. モデル1の推定スペクトル

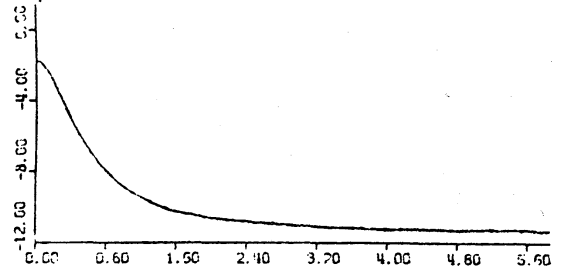


図6. モデル2の推定スペクトル

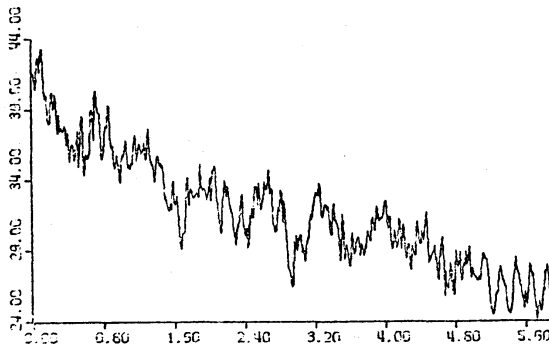


図7. モデル1のスムージングによる推定スペクトル。

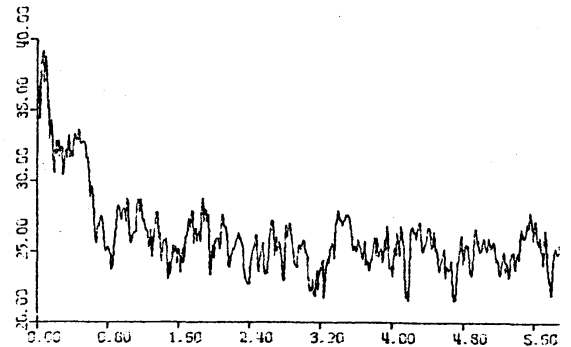


図8. モデル2のスムージングによる推定スペクトル。

6. より一般的な応答函数モデル. Hawkes は self-exciting

モデルの応答函数 $g(t)$ の一般的なものの一例として

$$g(t) = \sum_{j=1}^L \alpha_j e^{-\beta_j t} \quad (6.1)$$

を提案している。これはいくつかの β_j による減衰度の違い

によって応答函数の複雑な形を出せようとするものであるが、

実際に推定する場合 ill-posed 問題になることが多い。Ozaki

& Akaike (1977) はこれに代るものとして Laguerre 函数を

用い

$$g(t) = \sum_{m=0}^M \alpha_m t^m e^{-\beta t} \quad (6.2)$$

を提案している。但し

$$\lambda(t) = \mu + \int_{-\infty}^t g(t-u) dN(u) \geq 0$$

の条件からパラメータ $(\alpha_0, \dots, \alpha_M, \beta)$ は 制約条件

$$\int_0^{\infty} g(t) dt < 1$$

の他に $g(t) \geq 0$ を満たす必要がある。このモデルの尤

度, gradient, Hessian 及びマルコフ表現を用いたこれらの帰

納的計算法は Ozaki & Akaike (1977) に与えられている。

一般に応答函数 $g(t)$ の self-exciting 点過程モデルのスペク

トルは

$$P(\omega) = \frac{\lambda}{2\pi |1 - G(\omega)|^2} \quad (6.3)$$

ここに

$$G(\omega) = \int_{-\infty}^{\infty} \hat{g}(t) e^{-i\omega t} dt$$

$$\lambda = \frac{\mu}{1 - G(0)}$$

によって与えられる (Hawkes (1971b))。従ってモデル (6.2)

の場合

$$P(\omega) = \frac{\mu}{2\pi \left(1 - \sum_{m=0}^{\infty} \frac{\alpha_m \cdot m!}{\beta^{m+1}}\right) \left|1 - \sum_{m=0}^{\infty} \frac{\alpha_m \cdot m!}{(\beta + i\omega)^{m+1}}\right|^2} \quad (6.4)$$

となる。

7. 地震データ解析への応用. 点過程の一つの応用に地震発生間隔の統計的解析がある。(Vere-Jones (1970), Adamopoulos (1973)) 日本においても河角(1970)による南関東地方における大地震69年周期説が一時ジャーナリズムをにぎわしよく知られている。この69年周期説は西暦818年から1930年までの南関東地方の合計33回の大地震の記録をもとに出されたものでピリオドグラム解析によったものである。(点過程の場合のピリオドグラム解析については Bartlett (1963), Lewis (1970) 参照) 上記の河角博士の解析した33回のデータ(以後河角データと呼ぶ)のピリオドグラムは図9のように69年とよの2倍, 3倍, 5倍の周期に大きなピークを持つ。

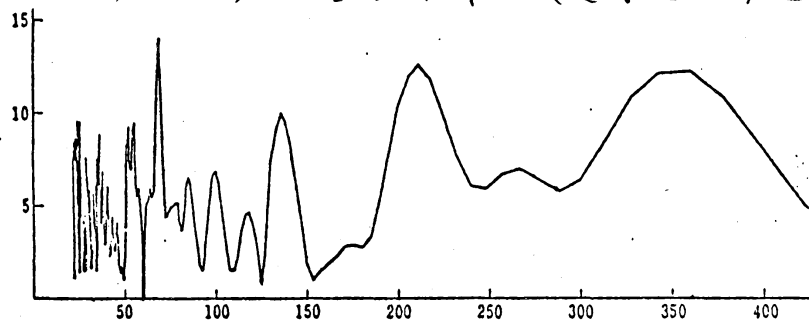


図9. 河角データのピリオドグラム

この河角説は点過程のパラメトリックな解析を通してみるとどうなるであろうか。今 intensity $\lambda(t)$ が

$$\lambda(t) = \exp\{\alpha + K \sin(\omega t + \theta)\}$$

の周期的ポアソン過程から生成されたとみなす。周期的ポアソン過程の尤度は

$$\log L = n\alpha - e^{\alpha} t I_0(K) + K \cos \theta \sum_{i=1}^n \sin \omega t_i \\ + K \sin \theta \sum_{i=1}^n \cos \omega t_i$$

で与えられる (Lewis (1970))。ここに $I_0(K)$ はオーダー 0 の modified Bessel 函数, t は今の時点 1977 年とする。この時の最尤推定値は $\hat{\alpha} = -3.8520$, $\hat{K} = 1.0580$, $\hat{\theta} = 1.3213$,

$\hat{\omega} = 0.0914$, $\log L = -139.4670$ となる。周期 $\hat{\tau}$ は

$2\pi/\hat{\omega} = 68.76$ となって周期的ポアソンとみた場合は周期は 69 年であるという河角説が裏づけられる。河角データが intensity $\lambda = e^{\alpha}$ のポアソンから生成されたとみるし左時は $\hat{\alpha} = -3.5896$, $\log L = -146.8665$ となる。モデルの良さの比較をエントロピー最大化原理 (Akaike (1976)) の立場に立ってやることにしネグントロピーの推定量として

$$AIC = (-2) \log(\text{maximum likelihood}) + 2(\text{パラメータ数})$$

を採用すると, 69 年周期モデルの AIC は 286.9340, ポアソンの場合 295.7330 となる, 69 年周期モデルのほうが良く合うモデルということになる。次に $\lambda(t) = \mu + \int_{-\infty}^t \alpha e^{-\beta(t-u)} dN(u)$ の self-exciting モデルをあてはめた場合は,

最尤推定値 $\hat{\alpha} = 0.1243$, $\hat{\beta} = 0.6387$, $\hat{\mu} = 0.2241 \times 10^1$ を得,

$\log L = -141.2952$ を得る。従ってその AIC は 288.5909 となり、このモデルはポアソンよりは良いが 69 年周期モデルには及ばないということになる。次に $M=4$ の場合の Ozaki-Akaike 型のモデル、 $\lambda(t) = \mu + \int_{-\infty}^t \sum_{m=0}^M \alpha_m (t-u)^m e^{-\beta(t-u)} dN(u)$ の場合は $\hat{\alpha}_0 = 0.2535 \times 10^{-1}$, $\hat{\alpha}_1 = -0.1126 \times 10^2$, $\hat{\alpha}_2 = 0.2666 \times 10^4$, $\hat{\alpha}_3 = -0.3857 \times 10^6$, $\hat{\alpha}_4 = 0.2306 \times 10^8$, $\beta = 0.2997 \times 10^1$, $\hat{\mu} = 0.9667 \times 10^2$ $\log L = -128.6333$ を得る。この場合の AIC は 271.2666 で 69 年周期モデルよりよくフィットしていると言える。

以上の解析から河角データについていえることは 69 年周期モデルによってポアソン説を否定することはできても 69 年周期があると断定することはできないということである。

付記 以上は筆者の仕事の他、筆者と Victoria 大学の Vere-Jones 教授との共同の仕事(未発表)と、筆者と統計数理研究所の赤池弘次博士との共同の仕事の一部をまとめたものである。

参考文献

1. Hawkes, A.G. (1971a) "Spectra of some self-exciting and mutually exciting point processes" *Biometrika*, 58, 1, 83-90.
2. Hawkes, A.G. (1971b) "Point spectra of some mutually exciting point processes" *J.R.S.S.*, B 33, 438-443.
3. Hawkes, A.G. & Dakes, D. (1974) "A cluster process

- representation of a self-exciting point processes", J. Appl. Prob. 11, 493-503.
4. Adamopoulos, L. (1975) "Some counting and interval properties of the mutually exciting processes" J. Appl. Prob. 12, 78-86.
 5. Snyder, D. L. (1975) "Random point processes" John Wiley & Sons.
 6. Van Schuppen, J. H. (1977) "Filtering, prediction and smoothing for counting process observations, a Martingale approach". SIAM. J. Appl. Math. 32, 3, 552-570.
 7. Rubin, I (1972) "Regular point processes and their detection" IEEE, IT-18, 5, 547-557.
 8. Ozaki, T. (1977) "Maximum likelihood estimation of Hawkes' self-exciting point processes" Research Memo. no 115.
 9. Ozaki, T. & Akaike, H. (1977) - to appear.
 10. Davidon, W. C. (1959) "Variable metric method for minimization" Argonne Natl. Lab. Report 5990.
 11. Fletcher, R & Powell, M. J. D. (1963) "A rapidly convergent descent method for minimization", The Comp. J., 6, 163-168
 12. Rosenbrock, H. H. (1960) "An automatic method for finding the greatest or the least value of a function" The Comp. J. 3, 175-184.

13. Powell, M. J. D. (1964) "An efficient method finding the minimum of a function of several variables without calculating derivatives", *The Comp. J.* 7, 155-162.
14. Vere-Jones, D. (1970) "Stochastic method for earthquake occurrence" *J. R. S. S.*, B 32, 1-62.
15. Adamopoulos, L. (1973) "Statistical analysis of earthquake data", Ph. D. Thesis Univ. of Durham.
16. 河角 広 (1970) "関東南部地震 69年周期の証明とその発生 of 緊迫度並びに対策の緊急性と問題点" *地学雑誌* 79, 3, 115-137.
17. Bartlett, M. S. (1963) "The spectral analysis of Point Processes" *J. R. S. S.*, B 25, 264-296.
18. Lewis, P. A. W. (1970) "Remarks on the theory, computation, and application of the spectral analysis of series of events" *J. Sound Vib.*, 12, 3, 353-375.
19. Akaike, H. (1977) "On entropy maximization principle" *Proc. Symposium on Applications of Statistics*, P. R. Krishnaiah ed., North-Holland, Amsterdam, to appear.