

ランダムなデータの解析 (I)

——生物リズムの研究の為に——

Analysis of Random Data (I)

——for the Study of Biological Rhythms——

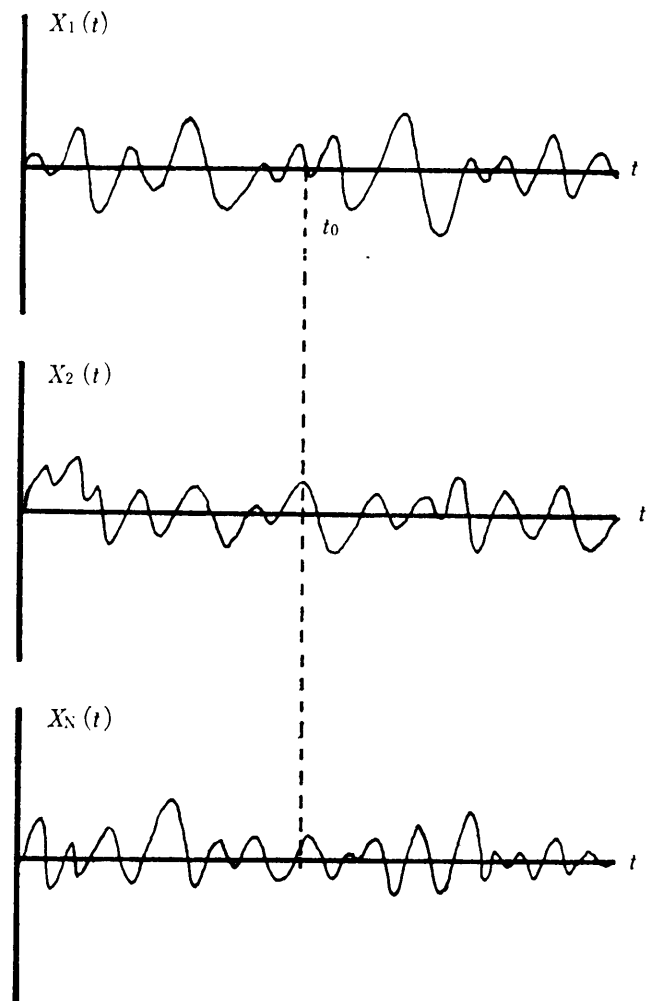
天 岸 祥 光 稲 村 欣 作
Yoshimitsu AMAGISHI and Kinsaku INAMURA

(Received Oct. 8, 1975)

1. はじめに

人間を含めた生物界のリズムに対する関心は最近かなり高まってきた。このリズムを実験的に明らかにしていくことは、それ自体生体のメカニズムの解明につながり、さらに医学的見地からもきわめて重要な問題提携になる。静岡大学においても、稲村等^{1),2)}によって、長年にわたって「人間のリズム」に関するデータがとられてきた。「人間のリズム」とは、生体のマクロな機能変動*であり、人間の精神面や身体面の機能がある周期で振動していることをいう。人によっては、精神作業の成績が20日から30日の周期性をもつことなどもその一例^{1),2)}である。このリズムは生体が本来もっている内因性のものと、外界からの影響(天体や生活環境の周期など)に基づく外因性のものとに大別されるであろう。これらのリズム—振動—は互いに相互作用をもち、また生体の中を伝搬していくことも考えられる。

さて、「人間のリズム」に関するナマのデータをみてもわかる通り、外界などからの非周期的な影響の為に、実際のデータにはランダム(random)成分がかなり含まれている場合が多い。またいくつもの周期が含まれている場合はさらにデータは複雑化する。このようなナマのデータから直接周期性をみいだすのは一般に不可能に近いであろう。また、データがある短い有限の長さしかなく、しかも何度もくりかえして測定、観測ができない



第1図

*生体の機能におけるマクロとは、単なる生理学的機能のよせあつめではなく身体作業など体力科学で扱うような、有機的な総合作用をいう。

場合は、データ解析は一段と難しくなる。実際に、「人間のリズム」に関するデータ採集には、何日も何百日もかかるものが多く、何度もくりかえして測定することが困難な場合が多い。したがって、いずれにせよ有限の長さのランダム・データの解析法を知らなくては、生物リズムの本質的解明は困難であろう。

この論文の目的は、「人間のリズム」に関するランダム・データをいかに解析したら、リズムの存在、伝搬などを科学的に明らかにすることができるかを検討することにある。但し、ここでは非数理系の研究者にも理解してもらえるように、できるだけ平易に、かつ重点的に、有限なランダム・データの解析法について述べ、問題点を明らかにしていきたい。また、この論文では「人間のリズム」の解析を念頭においているが、解説それ自体は一般的に述べられているはずであるから、他の分野のランダム・データ解析にも応用できることは無論である。

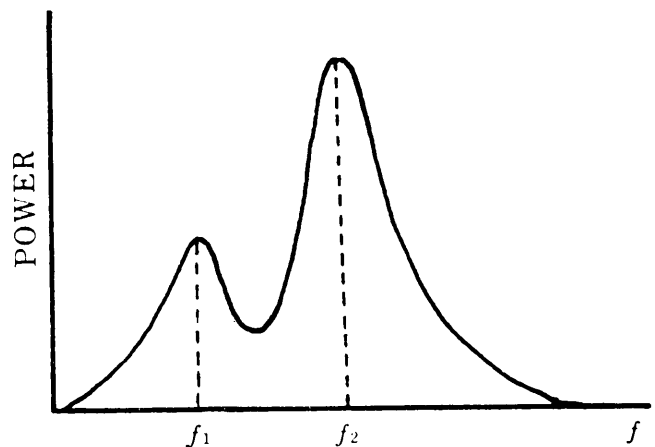
ここで考えているランダム・データとは、時間的に変動している不規則過程 (random process) の観測値 $x(t)$ —標本関数 (sampling function)— であって、しかも時間的にある有限の長さしかもっていないものをあつかう。もし何度も実験ができるなら、この標本関数の集合 $\{x_k(t)\}$ ($k=1, 2, 3, \dots$) が得られるであろう。一例を第1図に示す。もし時間を一定 ($t=t_0$) にして、タテに観測値の集合 $\{x_k(t_0)\} = (x_1(t_0), x_2(t_0), \dots, x_N(t_0))$ を考えればこれも1つの不規則変数 (random variable) になっている。確率統計学の理論は一般にこの不規則変数の集合—アンサンブル (ensemble)— を基礎に組立てられている。しかしながら我々が実験で得られるデータは多くの場合、このアンサンブルのうちのいくつか (最悪の場合1つ) の標本関数である。それでは一体このような標本関数と理論とをどのように組み合わせるかという問題がおこる。幸いに、もし不規則過程が定常 (stationary) であるなら、エルゴード性 (ergodic property) をもち、アンサンブル平均でもとめたものと、時間平均でもとめたものが等しくなる (4章)。この性質は大変都合がよく、標本関数が定常であれば、沢山の標本関数を得る必要はなく、1つだけで全体を知ることができることを意味している。しかも、いろいろな実験、観測で得られるデータには、このような定常不規則過程のものがかなり多い。

一方、ここではデータの長さが有限であるばかりでなく、観測値が離散的 (discrete) であるものについて考えていきたい。つまり、標本関数値が t_1, t_2, \dots, t_N といった離散的な時点でしかきまらないものを取扱っていく。実際連続的なデータでも、その解析がアナログ的にあつかえないかぎり、ある時間間隔で値を読んでいかなければならない。したがって、たとえ連続的であっても離散データとして取扱わなければならなくなる。よって、このようなデジタルな解析は応用範囲が大変広いことになる。

しかし、アナログ解析ではでてこないような問題点も含んでいるので注意を要する。

本論文の解析方法は次の2つである。

(i) 自己相関関数 (auto-correlation) をもとめ、そのフーリエ変換からパワー・スペクトル密度関数 (power spectrum density function) をもとめる。パワー・スペクトルは第2図に示すように振動数 f (周期の逆数) の関数であるから、どのような周期性のものがどの程度データ中に含まれているかを知ることができる。(ii) 相互相関関数



第2図

(cross-correlation function) およびその相互スペクトル密度関数 (cross spectrum density function) より2つの定常不規則過程の相関を調べ、その位相差から伝搬速度をもとめる。

ここでは以上の2点に焦点を合わせているので、確率論で重要な確率密度関数 (probability density function) をもとめ、分布の様子をみるといった検討を省いてある。さらにろ波器 (filter) の入力信号に対する応答などの問題にも一さい触れていない。

なお、相関関数関係の文献としては3) ~ 6) を参考にした。この方面の専門要語や定義などを短時間に調べたい場合は、文献6) を参考にするとうい。また生物リズムに関する文献などについては文献7) に詳しいのでそれを参照されたい。

第2章は一般的な相関関数について、第3章ではスペクトル密度関数を解説し、第4章でエルゴード性について述べる。第5章では、有限なデータ解析にともなう誤差を論じ、第6章で実際のデータ解析について述べる。第7章で、本論文にしたがった「人間のリズム」の解析例を示す。

2. 相関関数

標本関係のアンサンブルを $\{x_k(t)\}$ とするとある時刻 t におけるアンサンブル平均は

$$\mu_x(t) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N x_k(t) \tag{2.1}$$

であるが、確率密度関数 $p(x)^*$ をもちいて

$$\mu_x(t) = \int_{-\infty}^{\infty} xp(x)dx = \mathbf{E}[x_k(t)] \tag{2.2}$$

とあらわすこともできる。以後アンサンブル平均は $\mathbf{E}[\]$ の形で表わすことにする。これを不規則変数 $x_k(t)$ の期待値と呼ぶことにする。さらに添字 k は省略して単に $\mathbf{E}[x(t)]$ と書くことにしよう。また2つの不規則過程 $\{x_1(t)\}, \{x_2(t)\}$ の平均を表わす為 $\mu_i(t) = \mathbf{E}[x_i(t)]$, $i=1, 2$ などと書くが、この添字 i (或は j) は、上記の k とは意味が異なるので注意してほしい。

さて、一般には時刻が異なれば期待値も異なるから $t_1 \neq t_2$ ならば

$$\mu_i(t_1) \neq \mu_i(t_2) \tag{2.3}$$

である。さらに、ここで共分散関数 (covariance function) と呼ばれる関数を定義しよう。

$$C_{ij}(t, t+\tau) = \mathbf{E}[(x_i(t) - \mu_i(t))(x_j(t+\tau) - \mu_j(t+\tau))] \tag{2.4}$$

で、 $\tau=0$ とおくと

* ある変数 x の値がある値 x_1 を越えない確率を $P(x \leq x_1) = F(x_1)$ とかくとき、これを確率分布関数と定義する。次に、不規則変数が x_1 と $x_1 + dx$ の間にある確率は

$$\begin{aligned} P(x_1 \leq x \leq x_1 + dx) &= P(x \leq x_1 + dx) - P(x \leq x_1) = F(x_1 + dx) - F(x_1) = \frac{dF(x_1)}{dx} dx + 0^2(dx) \\ &= p(x_1) dx \end{aligned}$$

となり、一般に $p(x)$ を確率密度関数という。

$$\int_{-\infty}^{\infty} p(x) dx = 1$$

$$\int_{-\infty}^{x_1} p(x) dx = F(x_1)$$

の関係がある。同様に2つの変数がそれぞれ x_1, x_2 の値をもつ確率密度関数は

$$p(x_1, x_2) = \frac{\partial^2 F(x_1, x_2)}{\partial x_1 \partial x_2}$$

である。

$$C_{ij}(t, t) = \mathbf{E}[(x_i(t) - \mu_i(t))(x_j(t) - \mu_j(t))] = \sigma_{ij}^2(t) \quad (2.5)$$

となる。特に $i=j$ のとき

$$C_{ii}(t, t) = \mathbf{E}[(x_i(t) - \mu_i(t))^2] = \sigma_{ii}^2(t) \quad (2.5)'$$

となり、これはいわゆる $\{x_i(t)\}$ の分散 (variance) である。 $i \neq j$ のときは、 $C_{ij}(t, t) = \sigma_{ij}^2(t)$ を $\{x_i(t)\}$, と $\{x_j(t)\}$ の共分散 (covariance) と呼ぶ ($\tau=0$ での)。

次に (2.2), (2.4) で次のような性質があるとき,

$$\begin{aligned} \mu_i(t) &= \mu_i \\ C_{ii}(t, t+\tau) &= C_{ii}(\tau) \quad (i, j=1, 2) \\ C_{ij}(t, t+\tau) &= C_{ij}(\tau) \end{aligned} \quad (2.6)$$

すなわち標本関数の平均が一定で、分散が時間差 τ のみの関数であるとき、 $\{x_i(t)\}$, $i=1, 2$ は広い意味での定常不規則過程 (stationary random process) という。

上記の2つの確率関数 (一次と二次) だけでなく、その他のあらゆる確率関数すべてが時間 t によらないという厳しい条件を満足する不規則過程 (例えばガウス分布など) を狭義の定常というが、ここでは、前者の広義の定常不規則過程のみをとりあつかう。

さて、2つの定常不規則過程 $\{x_i(t)\}$, $i=1, 2$ の $\mu_i = \mathbf{E}[x_i(t)]$, 及び $C_{ij}(\tau)$ の他に次の関数を定義する。

$$R_{ij}(\tau) = \mathbf{E}[x_i(t)x_j(t+\tau)] \quad (2.7)$$

$i=j$ のとき、すなわち $R_{ii}(\tau) = \mathbf{E}[x_i(t)x_i(t+\tau)]$ を自己相関関数 (auto-correlation function) といい、 $i \neq j$ のとき $R_{ij}(\tau)$ を相互相関関数 (cross-correlation function) という。この関数には

$$\begin{aligned} R_{ii}(-\tau) &= R_{ii}(\tau) \\ R_{ij}(-\tau) &= R_{ji}(\tau) \end{aligned} \quad (i \neq j) \quad (2.8)$$

の関係がある。また (2.4), (2.6), (2.8) より

$$C_{ij}(\tau) = R_{ij}(\tau) - \mu_i \mu_j \quad i, j=1, 2 \quad (2.9)$$

となるから、期待値がゼロなら、相関関数は共分散関数と一致する。相関関数には次の関係がある。

$$\begin{aligned} R_{ii}(0) &\geq |R_{ii}(\tau)| \\ R_{ii}(0)R_{jj}(0) &\geq |R_{ij}(\tau)|^2 \end{aligned} \quad (i, j=1, 2) \quad (2.10)$$

また

$$V_{ij}(\tau) = \frac{C_{ij}(\tau)}{\sqrt{C_{ii}(0)C_{jj}(0)}} \quad (2.11)$$

を 相関関数係数 (correlation function coefficient) という。相関関数の性質などについては次第に明らかになっていくであろう。

3. スペクトル密度関係

定常不規則過程 $\{x_i(t)\}$, $i=1, 2$ の各標本関数 $x_i(t)$ が次のようにフーリエ変換できたとする。

$$X_i(f) = \int_{-\infty}^{\infty} x_i(t) e^{-j \cdot 2\pi f t} dt \quad (3.1)$$

また、逆変換より

$$x_i(t) = \int_{-\infty}^{\infty} X_i(f) e^{j \cdot 2\pi f t} df \quad (3.2)$$

ここで、 f は振動数 (或は周波数といい、周期の逆数)、 $j = \sqrt{-1}$ である。一方 (2.7) より、二乗平均値 (mean square value) は自己相関関数を用いて表わすことができ

$$\mathbf{E}[x_i^2(t)] = R_{ii}(0) \tag{3.3}$$

となるが、振動数が f と $f+df$ 間にある二乗平均値を $P_i(f)df$ とすれば (63ページの脚注の確率 P とは異なる)

$$\mathbf{E}[x_i^2(t)] = \int_{-\infty}^{\infty} P_i(f)df \tag{3.4}$$

でなければならない。この $P_i(f)$ をパワー・スペクトル密度関数 (power spectrum density function) という。この関数と自己相関関数との関係は

$$R_{ii}(\tau) = \int_{-\infty}^{\infty} P_i(f)e^{j \cdot 2\pi f \tau} df \tag{3.5}$$

である。同様にして、相互スペクトル密度関数 (cross-spectrum density function) を $P_{ij}(f)$ とすれば ($i \neq j$),

$$R_{ij}(\tau=0) = \mathbf{E}[x_i(t)x_j(t)] = \int_{-\infty}^{\infty} P_{ij}(f)df \tag{3.6}$$

$$R_{ij}(\tau) = \int_{-\infty}^{\infty} P_{ij}(f)e^{j \cdot 2\pi f \tau} df \tag{3.7}$$

の関係が成立する。(3.5), (3.6) をフーリエ逆変換し

$$P_i(f) = \int_{-\infty}^{\infty} R_{ii}(\tau)e^{-j \cdot 2\pi f \tau} d\tau \tag{3.8}$$

$$P_{ij}(f) = \int_{-\infty}^{\infty} R_{ij}(\tau)e^{-j \cdot 2\pi f \tau} d\tau \quad (i \neq j)$$

を得る。 $P_i(f)$ は正の実数値で偶関数であり、 $P_{ij}(f)$ は一般に複素関数である。すなわち

$$P_i(-f) = P_i(f) \tag{3.9}$$

$$P_{ij}(-f) = P_{ij}^*(f) = P_{ji}(f)$$

但し (*) は複素共役を意味する。

パワー・スペクトル

(3.9) よりパワー・スペクトル密度関数は

$$\begin{aligned} P_i(f) &= \int_{-\infty}^{\infty} R_{ii}(\tau) \cos 2\pi f \tau d\tau \\ &= 2 \int_0^{\infty} R_{ii}(\tau) \cos 2\pi f \tau d\tau \end{aligned} \tag{3.10}$$

と書きかえることができる。無論 $P_i(f)$ が存在する為には数学的に

$$\int_{-\infty}^{\infty} |R_{ii}(\tau)| d\tau < \infty \tag{3.11}$$

でなければならない。さて、実際の実験では、振動数 f が負になることはないから、次のようなパワー・スペクトル密度関数 $G_i(f)$ を定義した方が現実的である。すなわち

$$G_i(f) = 2P_i(f) \quad (0 \leq f < \infty) \tag{3.12}$$

よって、 $G_i(f)$ と $R_{ii}(\tau)$ の間の関係は

$$G_i(f) = 4 \int_0^{\infty} R_{ii}(\tau) \cos 2\pi f \tau d\tau \tag{3.13}$$

$$R_{ii}(\tau) = \int_0^{\infty} G_i(f) \cos 2\pi f \tau df$$

パワー・スペクトル関数 $G_i(f)$ の物理的意味は、いろいろな振動数を含む標本関数 $x_i(t)$ の、振動数 f の成分の強度（二乗平均値）である。「パワー」の名は、例えば抵抗 r に電流（一般に変動している）を流したとき、単位時間あたりに消費される平均電力（パワー）は $\overline{I^2 r}$ であることからわかるように、二乗平均値に対してつけられたものである。

ランダムな信号などに含まれている固有振動数（又は周期）、その強度、スペクトルの広がり具合などを知るには、(3.13)式よりパワー・スペクトル密度関数 $G_i(f)$ をもとめるのが一番である。定常的な電気的信号の場合は、最近はずぐれたスペクトル分析器（spectrum analyzer）が普及しており、ブラウン管の上にアナログ的に表わすことができる。しかし、我々がここで取り上げているのは、時間的に有限な、離散データをいかに解析するかというデジタルな方法であるので、スペクトル分析器をそのまま用いることはできない。このような場合のパワー・スペクトルのもとめ方は第6章であたえられるであろう。しかし、(3.13)よりパワー・スペクトルをもとめるには、とにかく自己相関関数のもとめればよいということがわかった。

相互スペクトル

パワー・スペクトルの場合と同様に、相互スペクトル密度関数を新たに次のように定義する。

$$\begin{aligned} G_{ij}(f) &= 2P_{ij}(f) \\ &= 2 \int_{-\infty}^{\infty} R_{ij}(\tau) e^{-j \cdot 2\pi f \tau} d\tau \end{aligned} \quad (3.14)$$

$$= L_{ij}(f) - jM_{ij}(f), \quad (i \neq j) \quad (3.15)$$

$L_{ij}(f)$ と $M_{ij}(f)$ は実数で (2.8) を用いて、(3.14) より

$$L_{ij}(f) = 2 \int_0^{\infty} [R_{ij}(\tau) + R_{ji}(\tau)] \cos 2\pi f \tau d\tau \quad (3.16)$$

$$M_{ij}(f) = 2 \int_0^{\infty} [R_{ij}(\tau) - R_{ji}(\tau)] \sin 2\pi f \tau d\tau$$

ともとまる。すなわち、相互相関関数がわかれば、 $G_{ij}(f)$ がもとまることになる。一方 (3.15) は次のようにも表現できる。

$$G_{ij}(f) = |G_{ij}(f)| e^{-j\theta_{ij}(f)} \quad (3.17)$$

ここで

$$|G_{ij}(f)| = (L_{ij}^2(f) + M_{ij}^2(f))^{1/2}, \quad (3.18)$$

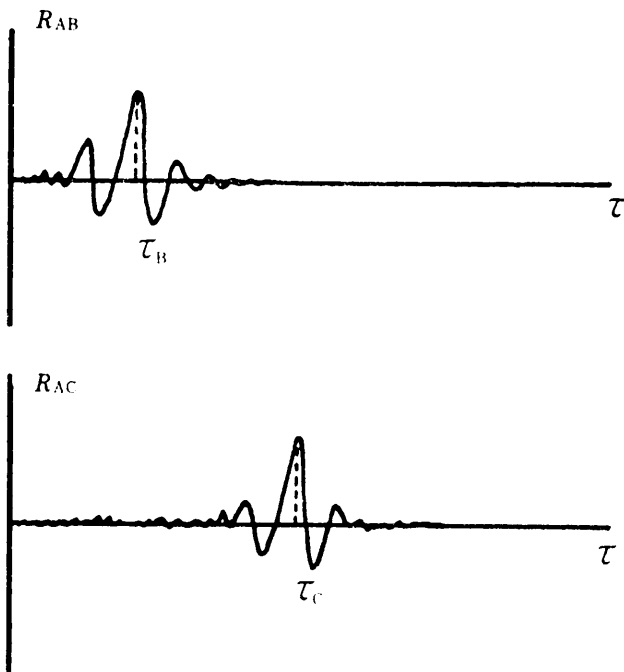
$$\theta_{ij}(f) = \tan^{-1} \left(\frac{M_{ij}(f)}{L_{ij}(f)} \right) \quad (3.19)$$

である。パワー・スペクトルと相互スペクトル関数の間には

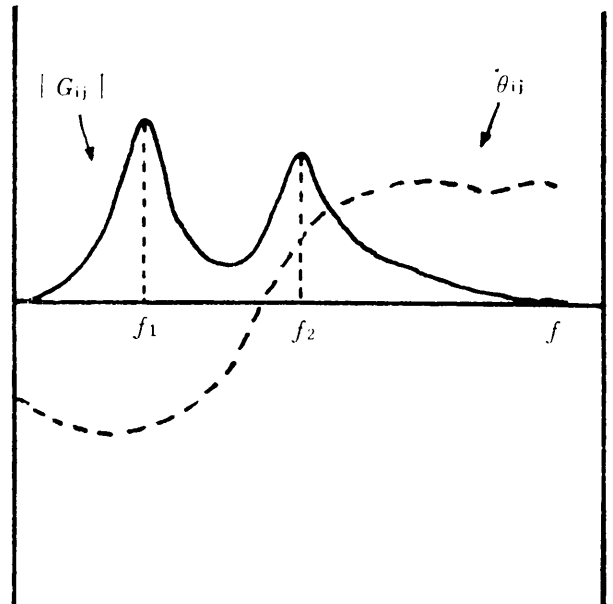
$$|G_{ij}(f)|^2 \leq G_i(f) G_j(f) \quad (3.20)$$

の関係がある。

以上の準備から相互相関関数及び相互パワー・スペクトルからどのようなことがわかるか検討してみよう。まず相互相関関数からは次のようなことがわかる。今仮にA点から発せられたパルスがB点、C点を通過して伝搬していくとする。しかし途中にいろいろ雑音があってB点、C点では直接パルスが観測できないとする。そこで、B、C両点での観測値（雑音を含んでいる） $x_B(t+\tau)$ 、 $x_C(t+\tau)$ とA点での値 $x_A(t)$ とから、2つの相互相関関数 $R_{AB}(\tau)$ 、 $R_{AC}(\tau)$ をもとめ、それが例えば第3図のようになったとする。相関関数の最大値を与える時刻 τ_B 、 τ_C がそのパルスがB点、C点を通過する時刻である。もしA、B、C間の距離がわかっているよ



第3図



第4図

うな場合は伝搬速度がもとまることになる (但し、以上のことは、後述のエルゴード性が成立するとしてである)。

しかしながら、いつも第3図のようにピークに対する時刻がはっきり決まるとはかぎらない。

また相互相関関数だけでは、それに関する振動数 f を決定することはできない。このような場合は (3.18) の相互スペクトルの絶対値及び (3.19) の位相角をもとめ、それをグラフにするとよい。例えば、それが第4図のようになったとしよう。まず $|G_{ij}(f)|$ のグラフからは互いに相関している振動数が f_1 と f_2 であることがわかる (パワー・スペクトルの方はこれ以外の振動数のピークがあってさしつかえない。2つの不規則過程で関係し合っているのが、 f_1 と f_2 だということである)。

一方位相差 $\theta_{ij}(f)$ の値より、波の伝搬に要した時間 (或は、おくれの時間 (delay time) といってもよい) がわかる。すなわち、位相差とおくれの時間 τ の間には

$$\theta_{ij}(f_k) = 2\pi f_k \tau \tag{3.21}$$

の関係があるから、もし $\theta_{ij}(f_k)$ が正 ($0 < \theta_{ij} \leq \pi$) ならば、 $x_j(t)$ の内の振動数 f_k の波は $x_i(t)$ のそれより $\theta_{ij}/(2\pi f_k)$ だけおくれて伝搬してきたことになる。 $\theta_{ij}(f_k)$ が負 ($-\pi < \theta_{ij} \leq 0$) ならば、逆に $x_j(t)$ の方が $x_i(t)$ より先に伝わっていることになる。

なお、ある振動数 f において、2つの不規則過程の間にどれほど相関があるかを表わす関数として、

$$\alpha_{ij}(f) = \frac{|G_{ij}(f)|^2}{G_i(f)G_j(f)} \tag{3.22}$$

を用いるとよい。これをコヒーレント関数 (coherent function) といい、(3.20) より $\alpha_{ij}(f) \leq 1$ である。等号は完全にコヒーレントの場合である。

4. エルゴード性

今までの話は、すべてアンサンブル平均をもとにして、相関関数をもとめてきたが、第1章で述べたように、実際の実験、観測においては多くの場合、データは時間の関数としてもとめられ、アンサンブル平均が得られるほど沢山のデータをあつめることは不可能な場合が多い。

しかしながら、エルゴード性 (ergodic property) というものが成り立つなら、アンサンブル平均のかわりに単に1つの標本関数の時間平均で計算すればよいことがわかる。以下、それについて説明しよう。

ある2つの不規則過程 $\{x_i(t)\}$, $i=1, 2$ を考え、このアンサンブルのうち k 番目の標本関数を $x_i(t; k)$ と表わそう。するとその平均値を時間平均でもとめれば

$$\mu_i(k) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T x_i(t; k) dt \quad (4.1)$$

であるが、これは一般に標本関数によって異なるはずだから k の関数になっている。一方共分散関数を時間平均でもとめると次のようになる。

$$\begin{aligned} C_{ij}(\tau; k) &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T [x_i(t; k) - \mu_i(k)][x_j(t+\tau; k) - \mu_j(k)] dt \\ &= R_{ij}(\tau; k) - \mu_i(k)\mu_j(k) \end{aligned} \quad (4.2)$$

ここで

$$R_{ij}(\tau; k) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T x_i(t; k)x_j(t+\tau; k) dt \quad (4.3)$$

である。これは時間平均からもとめた相互相関関数である。もし $i=j$ であるならば、自己相関関数になる。これらの時間平均からもとめたものとアンサンブル平均からもとめたものが互いに等しいとき、すなわち

$$\begin{aligned} \mu_i(k) &= \mu_i \\ C_{ii}(\tau; k) &= C_{ii}(\tau) \end{aligned} \quad (4.4)$$

のとき不規則過程 $\{x_i(t)\}$ はエルゴード性があるという。

逆にある不規則過程 $\{x_i(t)\}$ がエルゴード過程 (ergodic process) である為の十分条件は、①定常 (stationary) であること、②時間平均からもとめた平均値と相関関数がすべての標本関数について同じであること、の2つの条件を満足すればよいことが証明される。例えば平均値についていえば、前者の条件は (2.6) よりアンサンブル平均 μ_i が時間によらないことであり、後者の条件は $\mathbf{E}[\mu_i(k)] = \mu_i'$ である。この二つの条件より $\mu_i = \mu_i'$ が証明される。何故なら

$$\begin{aligned} \mu_i(k) &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T x_i(t; k) dt \\ \therefore \mathbf{E}[\mu_i(k)] &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \mathbf{E}[x_i(t; k)] dt = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \mu_i dt = \mu_i \end{aligned}$$

だからである。同様にして、相関関数についても、時間平均とアンサンブル平均が等しいことが証明される。

5. 統計的推定及び誤差

一般に不規則過程の母集団 $\{x(t)\}$ のパラメーター (例えば平均値とかパワー・スペクトルなど) の真の値をもとめることは不可能に近い場合が多い。従って我々にできることは、標本抽出によって取り出された標本 (sample) の値からパラメーターを推定 (estimate) し、誤差はどの位であるかをいう以外に方法はない。

そこで、 θ を不規則過程 $\{x(t)\}$ の未知のある真の値 (母数) とし、 $\hat{\theta}$ をある有限時間の標本関数 $x(t)$ からもとめた推定量 (estimator) としよう。 $\hat{\theta}$ は不規則変数 (random variable)

であるから、沢山の標本から推定量をもとめると、母集団 $\{\hat{\theta}\}$ ができる。もし

$$\mathbf{E}[\hat{\theta}] = \theta \quad (5.1)$$

であるなら、期待値が真の値 θ に一致し、このような場合の $\hat{\theta}$ をアンバイアス(unbiased)推定量という。もし $\mathbf{E}[\hat{\theta}] \neq \theta$ なら $\hat{\theta}$ をバイアス (biased) 推定量という。このバイアスの量を

$$\mathbf{B}[\hat{\theta}] = \mathbf{E}[\hat{\theta}] - \theta = \mathbf{E}[\hat{\theta} - \theta] \quad (5.2)$$

とあらわすことにしよう。このバイアスの量は真の値からのズレをあらわしているから (5.1) を満足するような推定量の母集団 $\{\hat{\theta}\}$ が最もよさそうに見えるが、必ずしもそうではない。というのはアンバイアス推定量であっても、真の値 θ のまわりにはかなりの範囲にわたって分布しているようなときは (つまり分散が大きい)、いくら (5.1) を満足しても優れた推定量とはいえない (第5図参照)。そこで $\hat{\theta}$ の二乗平均誤差

$$\mathbf{E}[(\hat{\theta} - \theta)^2] = \mathbf{E}[(\hat{\theta} - \mathbf{E}[\hat{\theta}])^2] + \mathbf{E}[(\mathbf{E}[\hat{\theta}] - \theta)^2] = \mathbf{V}[\hat{\theta}] + \mathbf{B}^2[\hat{\theta}] \quad (5.3)$$

を考える。右辺第一項の $\mathbf{V}[\]$ は分散をあらわす。すなわち $\hat{\theta}$ の二乗平均誤差は $\hat{\theta}$ の分散とバイアス (の二乗) による誤差との和になっている。

もしアンバイアスであるなら、二乗平均誤差は推定量の広がり具合、つまり分散だけで決まるが、標本の大きさを大きくしたとき (或は $T \rightarrow \infty$ でもよい) $\mathbf{E}[(\hat{\theta} - \theta)^2] \rightarrow 0$ となるような推定量 $\hat{\theta}$ を一致推定量 (consistent estimator) という。このような意味で、アンバイアス推定量は一致推定量の特別なものであることがわかる。

平均値の推定量

エルゴード性をもつある不規則変数の標本 $x_i(t)$ を有限時間で平均をとった、標本平均値 (sample mean value) は (推定量のことである)

$$\hat{\mu}_i = \frac{1}{T} \int_0^T x_i(t) dt \quad (5.4)$$

一方、真の平均値 (期待値) は

$$\mu_i = \mathbf{E}[x_i(t)] \quad (5.5)$$

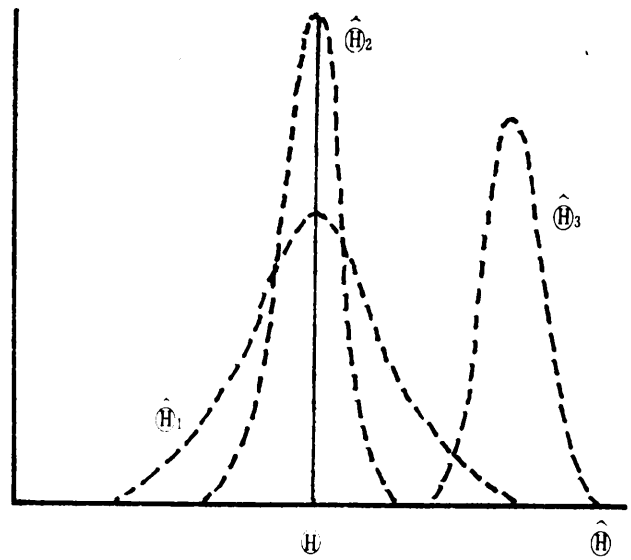
であるから、推定量 $\hat{\mu}_i$ の期待値をもとめると

$$\mathbf{E}[\hat{\mu}_i] = \frac{1}{T} \int_0^T \mathbf{E}[x_i(t)] dt = \frac{1}{T} \int_0^T \mu_i dt = \mu_i \quad (5.6)$$

となるから (5.1) より $\hat{\mu}_i$ はアンバイアスである。従って $\hat{\mu}_i$ の二乗平均誤差は $\hat{\mu}_i$ の分散のみによって決まる。すなわち

$$\mathbf{E}[(\hat{\mu}_i - \mathbf{E}[\hat{\mu}_i])^2] = \mathbf{E}[(\hat{\mu}_i - \mu_i)^2] = \mathbf{V}[\hat{\mu}_i] = \mathbf{E}[\hat{\mu}_i^2] - \mu_i^2 \quad (5.7)$$

途中の計算は省略するが、平均値 μ_i がゼロでない、バンド巾が B のホワイトノイズ (white noise) では近似的に ($|\tau| \ll T$ という条件で)



第5図 $\hat{\theta}_1, \hat{\theta}_2$ はアンバイアス推定量
 $\hat{\theta}_3$ はバイアス推定量

$$\mathbf{V}[\hat{\mu}_i] \simeq \frac{C_{ii}(0)}{2BT} = \frac{\sigma_{ii}^2}{2BT} \quad (5.8)$$

となる。従って規格化二乗平均誤差は

$$\varepsilon^2 = \frac{\mathbf{V}[\hat{\mu}_i]}{\mu_i^2} = \frac{1}{2BT} \left(\frac{\sigma_{ii}}{\mu_i} \right)^2 \quad (5.9)$$

となり、 ε を規格化標準誤差 (normalized standard error) という。

二乗平均値の推定量

同様にして、二乗平均値の推定量は

$$\hat{\phi}_i^2 = \frac{1}{T} \int_0^T x_i^2(t) dt \quad (5.10)$$

であり、真の二乗平均値 (期待値) は

$$\Phi_i^2 = \mathbf{E}[x_i^2(t)] \quad (5.11)$$

で時間 t に依存しない。そこで $\hat{\phi}_i^2$ の期待値は

$$\mathbf{E}[\hat{\phi}_i^2] = \frac{1}{T} \int_0^T \mathbf{E}[x_i^2(t)] dt = \Phi_i^2 \quad (5.12)$$

となるから、 $\hat{\phi}_i^2$ はアンバイアスであることがわかる。従ってこの二乗平均的誤差は

$$\mathbf{E}[(\hat{\phi}_i^2 - \Phi_i^2)] = \mathbf{V}[\hat{\phi}_i^2] = \mathbf{E}[\hat{\phi}_i^4] - \Phi_i^4 \quad (5.13)$$

となる。ここでも計算は省略するが、バンド巾 B のホワイトノイズについては ($|\tau| \ll T$)

$$\mathbf{V}[\hat{\phi}_i^2] \simeq \frac{C_{ii}^2(0)}{BT} + \frac{2}{BT} \mu_i^2 C_{ii}(0) \quad (5.14)$$

となる。もし $\mu_i = 0$ ならば

$$\mathbf{V}[\hat{\phi}_i^2] \simeq \frac{R_{ii}^2(0)}{BT} = \frac{\Phi_i^4}{BT} \quad (5.15)$$

であるから規格化標準誤差は

$$\varepsilon = \sqrt{\frac{\mathbf{V}[\hat{\phi}_i^2]}{\Phi_i^4}} = \frac{1}{\sqrt{BT}} \quad (5.16)$$

4

相関関数の推定量

相関関数の推定量は

$$\hat{R}_{ij}(\tau) = \frac{1}{T} \int_0^T x_i(t) x_j(t+\tau) dt, \quad (i, j=1, 2) \quad (5.17)$$

であるから、この推定量の期待値も真の値に等しくなることから、 $\hat{R}_{ij}(\tau)$ もアンバイアス推定量である。二乗平均誤差は従ってこの分散だけからもとまり、前と同じようなバンド巾 B のホワイトノイズ ($\mu_1 = \mu_2 = 0$ として) については

$$\varepsilon^2 = \frac{\mathbf{V}[\hat{R}_{ij}(\tau)]}{R_{ij}^2(\tau)} \simeq \frac{1}{2BT} \left(1 + \frac{R_{ii}(0)R_{jj}(0)}{R_{ij}^2(\tau)} \right) \quad (5.18)$$

となる。 $i=j$ であれば自己相関関数についての誤差になる。

スペクトル密度関数の推定量

パワー・スペクトル密度の推定量 $\hat{G}_i(f)$ を得るのに、 B_e のバンド巾のフィルターを用いたとしよう。ここでいうバンド巾とは、 $x_i^2(t)$ のうち振動数が $(f - B_e/4)$ から $(f + B_e/4)$ までの間にあるものについての時間平均する為のものであり、分解能に相当するものである。つまり振動数の差が $(1/2)B_e$ 以下では二つの振動数は区別できないことを意味する。従って、今までの B 、つまり標本関数全体のバンド巾とは別のものである。そこで、入力 $x_i(t)$ を振動数 f でバンド巾 B_e の狭いフィルター (narrow band filter) を通した結果の値を $x_i(t; f, B_e)$ とかくことにしよう。するとこの出力の二乗平均の推定量は

$$\hat{\Phi}_i^2(f, B_e) = \frac{1}{T} \int_0^T x_i^2(t; f, B_e) dt \tag{5.19}$$

である。パワー・スペクトル密度の推定量は定義より (5.19) を B_e で割ったものであるから

$$\hat{G}_i(f) = \frac{\hat{\Phi}_i^2(f, B_e)}{B_e} = \frac{1}{B_e T} \int_0^T x_i^2(t; f, B_e) dt \tag{5.20}$$

そこで $\hat{G}_i(f)$ の期待値を (5.12) を用いてもとめると

$$\mathbf{E}[\hat{G}_i(f)] = \frac{1}{B_e} \mathbf{E}[\hat{\Phi}_i^2(f, B_e)] = \frac{\Phi_i^2(f, B_e)}{B_e} \tag{5.21}$$

一方

$$G_i(f) = \lim_{B_e \rightarrow 0} \frac{1}{B_e} \Phi_i^2(f, B_e) \tag{5.22}$$

であるから (5.21) と (5.22) より

$$\mathbf{E}[\hat{G}_i(f)] \neq G_i(f) \tag{5.23}$$

よって、 $\hat{G}_i(f)$ は一般にバイアス推定量である。

次に二乗平均誤差はバイアスの効果がでてきて (5.3) より

$$\mathbf{E}[(\hat{G}_i(f) - G_i(f))^2] = \mathbf{V}[\hat{G}_i(f)] + \mathbf{B}^2[\hat{G}_i(f)] \tag{5.24}$$

となる。 $G_i(f)$ がフィルターのバンド巾 B_e にわたって一定とすると

$$\mathbf{V}[\hat{G}_i(f)] \simeq \frac{G_i^2(f)}{B_e T} \tag{5.25}$$

$$\mathbf{B}[\hat{G}_i(f)] \simeq -\frac{B_e}{24} G_i''(f)$$

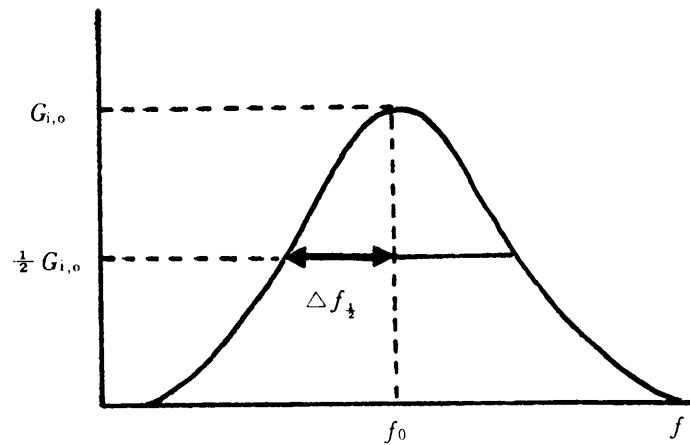
が得られる。従って規格化標準誤差 ε は

$$\varepsilon = \frac{\mathbf{E}[(\hat{G}_i(f) - G_i(f))^2]}{G_i^2(f)} \simeq \frac{1}{B_e T} \left(\frac{B_e^2 G_i''(f)}{24 G_i(f)} \right)^2 \tag{5.26}$$

からもとまる。 $G_i(f)$ を $G_{ij}(f)$ に置き換えれば相互スペクトル密度関数についての誤差になる。

さて (5.26) の右辺第二項はバイアスの効果によって生じた誤差であるから、スペクトルの振動数に対するズレの程度をあらわしている。これはバンド巾 B_e と真のスペクトルの広がり具合によって決まることがわかる。すなわち第6図のように真のスペクトルの半値巾を $4f_{\frac{1}{2}}$ とすれば

$$\begin{aligned} \frac{1}{2} G_i(f_0) &= G_i(f_0 \pm 4f_{\frac{1}{2}}) \\ &= G_i(f_0) \pm G_i'(f_0) 4f_{\frac{1}{2}} - \frac{1}{2} G_i''(f_0) (4f_{\frac{1}{2}})^2 \pm \dots \end{aligned}$$



第6図

であるが、 $G'(f_0) = 0$ であるから半値巾は

$$\Delta f_{\frac{1}{2}} \approx \left| \frac{G(f_0)}{G''(f_0)} \right|^{\frac{1}{2}} \quad (5.27)$$

となる。従って (5.26) は

$$\varepsilon^2 \approx \frac{1}{B_c T} + \frac{1}{576} \left(\frac{B_c}{\Delta f_{\frac{1}{2}}(f)} \right)^4 \quad (5.28)$$

と書きかえられる。これからスペクトルの巾 $\Delta f_{\frac{1}{2}}$ が大きいときは分解能 B_c をあまり小さくともなくてもバイアス誤差は小さくてすむが、かなりシャープなスペクトル分布のときは、 B_c をそれに応じて小さくしなければならない。一方(5.28)の右辺の第一項は分散による誤差であるからランダム誤差といってもさしつかえないものであるが、これはパワー・スペクトルの滑らかさの程度を表わしている。この誤差が大きいと、スペクトル分布が滑らかに変化せずばらつきが大きくなる。この誤差を小さくする為にはバンド巾 B_c を大きくしなければならない(データの長さ T は一定として)から、結局適当な B_c を選んで両方の誤差の和が最小になるようにするのが望ましい。 $\Delta f_{\frac{1}{2}}$, T は一定として、この最小値を与えるバンド巾 B_{c0} は(5.28)より

$$B_{c0} = \left(\frac{144 \Delta f_{\frac{1}{2}}^4}{T} \right)^{\frac{1}{5}} \quad (5.29)$$

ともとまるから、規格化標準誤差の最小値は

$$\varepsilon_{min} = \frac{\sqrt{5}}{24} \left(\frac{B_{c0}}{\Delta f_{\frac{1}{2}}} \right)^2 = \frac{1}{2} \left(\frac{5}{TB_{c0}} \right)^{\frac{1}{2}} \quad (5.30)$$

となる。

実際にはスペクトルの真の半値巾 $\Delta f_{\frac{1}{2}}$ がわからない場合が多いから、 ε_{min} 或は B_{c0} をもとめることは難しい。そこで許されるかぎり分解能をあげて、バンド巾を小さくしバイアス誤差が無視できるような状態で解析を行うと、このときの誤差は

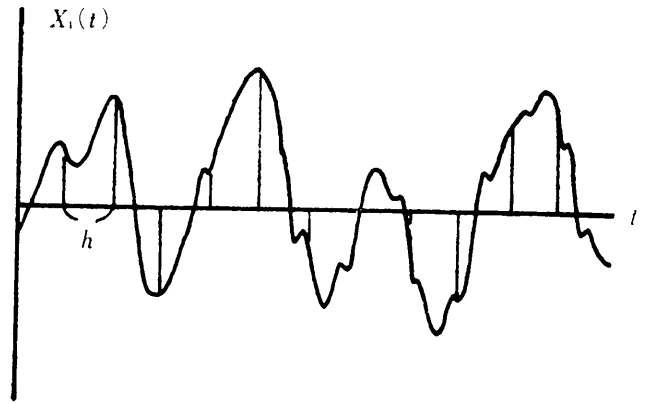
$$\varepsilon \approx \frac{1}{\sqrt{B_c T}} \quad (5.31)$$

となる。一般には分解能をあげてスペクトルをもとめるのが普通だから、(5.31)が成り立つ。しかしこうするとランダム誤差が目立ってくるはずであるから、スペクトルを平滑 (smooth-

ing) しなければならない。これについては第6章でのべるが、有限の長さのデータ解析では是非この平滑法をとり入れる必要がある。

6. 実際のデータ解析

第1章でのべたように、ここではアナログ的な解析でなく、標本関数 $x_i(t)$ をある時間間隔 h で分割して、その時の値 $x_i(nh)$ を読むことによって、もともとのパラメーター（相関関数とかスペクトル密度）を推定するデジタルな方法について述べる。この方法は応用範囲が広いが、注意すべき問題点を含んでいる。その代表的なものがエーリアシング (aliasing) である。まずそれについて述べよう。



第7図

第7図のように、連続的なデータを時間 h で分割したとする。従って、 $2h$ がこのランダムなデータの中に含まれる波の最小の周期でなければならない。或は、このデータの中の波の最大の振動数 f_N は

$$f_N = \frac{1}{2h} \tag{6.1}$$

でなければならない。エーリアシングの問題は f_N 以上の振動数が実際に存在するとき起こってくる。つまりパワー・スペクトルをもとめたとき、計算では f_N 以上の振動数に対しては $\hat{G}_i(f \geq f_N) = 0$ となるが、実際に f_N 以上にスペクトル分布があるときは、その周波数のうち

$$(2f_N \pm f), (4f_N \pm f), (6f_N \pm f), \dots, (2mf_N \pm f) \tag{6.2}$$

(但し、 $0 \leq f < f_N, m = 1, 2, 3, \dots$)

の振動数に対するパワー・スペクトルの値が、振動数 f のところにかさなってくる。つまり

$$\hat{G}_i(f) \rightarrow \hat{G}_i(f) + A \tag{6.3}$$

ここで $A = \hat{G}_i(2f_N - f) + \hat{G}_i(2f_N + f) + \hat{G}_i(4f_N - f) + \hat{G}_i(4f_N + f) + \dots$

となる。何故なら、今離散的な $x_i(t)$ の値を

$$x_i(nh)_1 = \cos 2\pi f \cdot nh = \cos 2\pi f \frac{n}{2f_N} \tag{6.4}$$

($n = 0, 1, 2, 3, \dots$)

とおくと、振動数 $(2mf_N \pm f)$ のときの値は

$$x(nh)_2 = \cos 2\pi(2mf_N \pm f) \frac{n}{2f_N} = \cos 2\pi f \frac{n}{2f_N} = x(nh)_1$$

であるから、(3.13) からわかるように $\hat{G}_i(f)$ と $\hat{G}_i(mf_N \pm f)$ の区別がつかなくなる。従って (6.3) のようになり、 f_N 以上のスペクトルの一部が f_N 以下の部分に折れかさなってくる (folding) ことになる。したがって、 f_N 以上の振動数が実際にはないように f_N を選ばなければならない。しかしデータによっては標本時間間隔 h をあまり小さくできない (特に不連続なデータでは、 h の最小値はおのずから決まってしまう) 場合もあり、この時は f_N 以下にすべての振動数を含めることが不可能になることもありうる。逆に、このようなデジタルな解

析では、あらかじめ、予想される振動数がわかれば、それからいかに実験をすべきか（どのような h をとるか）が強制されることになる。また、経験的に、或は物理的に f_N 以上の振動数は存在しないだろう、又はあっても極めて小さいだろうと推定できる場合もかなり多い。いずれにせよ f_N についての検討は、まず第一に行なわれなければならないことである。

6-1 自己相関関数とパワー・スペクトル

第5章で述べた、スペクトルの分解能に関するスペクトル巾 B_e は、このようなサンプリング法による解析では h の何倍かの値から決まるはずである。つまり

$$B_e = \frac{1}{mh}, \quad m = 1, 2, 3, \dots \quad (6.5)$$

従って、パワー・スペクトルの誤差もこの m の値に左右されることになる。標本のサイズを N とすれば、データの長さ T は

$$T = Nh \quad (6.6)$$

であるから、 m を大きくとり、 N に近づけると (6.5) より B_e が小さくなり分解能があがりパワー・スペクトルのバイアス誤差は小さくなるが、逆にランダム誤差は大きくなる。今、ランダム誤差の方がバイアス誤差よりずっと大きいとして（スペクトルがかなり巾をもっている場合もあてはまる）、スペクトルの規格化標準誤差は (5.31), (6.5), (6.6) より

$$\varepsilon = \frac{1}{\sqrt{B_e T}} = \sqrt{\frac{m}{N}} \quad (6.7)$$

となる。よって、バイアス誤差の心配がないときは m はできるだけ小さい方が望ましいことになる。 $m \sim (1/10)N$ 程度が望ましいが、データによってはこの条件を満足できない場合がかなりある。そのときは m をある程度増すことはやむをえないが、ランダム誤差が大きくなり、時にはパワー・スペクトル密度の値が負になったりすることがある。このような不安定性 (instability) をおさえる為に、後述の平滑 (smoothing) が必要なのである。

さて、以上のことを念頭において、 m が一応決まったとすると自己相関関数の推定量は ($\mu_i = 0$ として)

$$\hat{R}_i(rh) = \frac{1}{N-r} \sum_{n=1}^{N-r} x_i(nh) x_i((n+r)h) \quad (r = 0, 1, 2, \dots, m) \quad (6.8)$$

となる。このフーリエ変換 ((3.13) より) よりパワー・スペクトルは $0 \leq f \leq f_N$ として

$$\tilde{G}_i(f) = 2h \left[\hat{R}_i(0) + 2 \sum_{r=1}^{m-1} \hat{R}_i(rh) \cos\left(\frac{\pi r f}{f_N}\right) + \hat{R}_i(mh) \left(\frac{\pi m f}{f_N}\right) \right] \quad (6.9)$$

となる。しかし周波数も離散的であるから

$$f = k \frac{f_N}{m} \quad (k = 0, 1, 2, \dots, m) \quad (6.10)$$

とおけば (6.9) は

$$\tilde{G}_{i,k} = \tilde{G}_i\left(\frac{k f_N}{m}\right) = 2h \left[\hat{R}_i(0) + 2 \sum_{r=1}^{m-1} \hat{R}_i(rh) \cos\left(\frac{\pi r k}{m}\right) + (-1)^k \hat{R}_i(mh) \right] \quad (6.10)'$$

となる。最後に平滑する方法として最もよく使われるのがハニング (Hanning) 法と呼ばれて

いる方法で、次のようにして $\hat{G}_i\left(\frac{kf_N}{m}\right) = \hat{G}_{i,k}$ をもとめる。

$$\begin{aligned} \hat{G}_{i,0} &= 0.5\tilde{G}_{i,0} + 0.5\tilde{G}_{i,1} \\ \hat{G}_{i,k} &= 0.25\tilde{G}_{i,k-1} + 0.5\tilde{G}_{i,k} + 0.25\tilde{G}_{i,k+1}, \quad (k=1, 2, \dots, m-1) \\ \hat{G}_{i,m} &= 0.5\tilde{G}_{i,m-1} + 0.5\tilde{G}_{i,m} \end{aligned} \quad (6.11)$$

なお $0 \leq f \leq f_N$ での二乗平均値の和は

$$\hat{R}_{ii}(0) = \int_0^{f_N} \tilde{G}_i(f) df = \frac{f_N}{m} \left[\frac{1}{2} \tilde{G}_{i,0} + \sum_{k=1}^{m-1} \tilde{G}_{i,k} + \frac{1}{2} \tilde{G}_{i,m} \right] \quad (6.12)$$

であるから、これを用いて (6.12) を規格化すると便利である。

6-2 相互相関関数と相互スペクトル

相互相関関数の推定量も (6.8) と同様に

$$\hat{R}_{ij}(rh) = \frac{1}{N-r} \sum_{n=1}^{N-r} x_i(nh)x_j((n+r)h) \quad , \quad (i, j=1, 2) \quad (6.13)$$

としてもとまる。

よって、相互スペクトル ((3.14) 参照) の実数部分と虚数部分の推定量は (3.16) から次のように推定できる。

$$\tilde{L}_{ij}(f) = 2h \left[\hat{A}_0 + 2h \sum_{r=1}^{m-1} \hat{A}_r \cos\left(\frac{\pi r f}{f_N}\right) + \hat{A}_m \cos\left(-\frac{\pi m f}{f_N}\right) \right] \quad (6.14)$$

$$\tilde{M}_{ij}(f) = 2h \left[2 \sum_{r=1}^{m-1} \hat{B}_r \sin\left(\frac{\pi r f}{f_N}\right) + \hat{B}_m \left(\frac{\pi m f}{f_N}\right) \right]$$

ここで

$$\hat{A}_r = \hat{A}_{ij}(rh) = \frac{1}{2} [\hat{R}_{ij}(rh) + \hat{R}_{ji}(rh)] \quad (6.15)$$

$$\hat{B}_r = \hat{B}_{ij}(rh) = \frac{1}{2} [\hat{R}_{ij}(rh) - \hat{R}_{ji}(rh)]$$

である。しかし、パワースペクトルの場合と同様に、振動数は離散的であるから (6.10) より

$$\tilde{L}_k = \tilde{L}_{ij,k} = \tilde{L}_{ij}\left(\frac{kf_N}{m}\right) = 2h \left[\hat{A}_0 + 2 \sum_{r=1}^{m-1} \hat{A}_r \cos\left(\frac{\pi r k}{m}\right) + (-1)^k \hat{A}_m \right] \quad (6.16)$$

$$\tilde{M}_k = \tilde{M}_{ij,k} = \tilde{M}_{ij}\left(\frac{kf_N}{m}\right) = 4h \sum_{r=1}^{m-1} \hat{B}_r \sin\left(\frac{\pi r k}{m}\right)$$

とすべきである。

最後にハニング法により平滑を行い、

$$\begin{aligned} \hat{L}_0 &= 0.5\tilde{L}_0 + 0.5\tilde{L}_1 \\ \hat{L}_k &= 0.25\tilde{L}_{k-1} + 0.5\tilde{L}_k + 0.25\tilde{L}_{k+1} \quad (k=1, 2, \dots, m-1) \\ \hat{L}_m &= 0.5\tilde{L}_{m-1} + 0.5\tilde{L}_m \end{aligned} \quad (6.17)$$

及び

$$\begin{aligned} \hat{M}_0 &= 0.5\tilde{M}_0 + 0.5\tilde{M}_1 \\ \hat{M}_k &= 0.25\tilde{M}_{k-1} + 0.5\tilde{M}_k + 0.25\tilde{M}_{k+1} \quad (k=1, 2, \dots, m) \end{aligned} \quad (6.18)$$

$$\hat{M}_m = 0.5\hat{M}_{m-1} + 0.5\hat{M}_m$$

これらと (3.17), (3.18), (3.19) の関係式から相互スペクトル密度及び位相角の推定量がもとまる。

$$\begin{aligned} \hat{G}_{ij}\left(\frac{kf_N}{m}\right) &= \hat{G}_{ij}\left(\frac{kf_N}{m}\right) e^{-j\theta_{ij}(kf_N/m)} \\ \hat{G}_{ij}\left(\frac{kf_N}{m}\right) &= (\hat{L}_k^2 - \hat{M}_k^2)^{\frac{1}{2}} \quad (k=0, 1, 2, \dots, m) \\ \theta_{ij}\left(\frac{kf_N}{m}\right) &= \tan^{-1}\left(\frac{\hat{M}_k}{\hat{L}_k}\right) \end{aligned} \quad (6.19)$$

6-3 定常性の検定

これまでの話はすべてエルゴード性の成り立つ定常不規則過程を考えてきた。すなわち (2.6) であたえられたように、期待値 (又は時間での平均値) が時刻 t に依存しないこと、さらに相関関数が時間差だけの関数であることの二点が成立しなければならない。実際のデータがこれらの条件を満足しているかどうかの検定は解析の際にまず行なわれなければならないが、定常性を破るもっともよくある原因は測定値の長期的傾向 (trend), すなわち時間とともにゆっくりと一方向へ変化していく傾向や、長周期の正弦波などが含まれている場合などである。その他振幅が変化してるもの、振動数が変わるものなどがある。

もし、沢山の標本関数 (長い時系列から任意に標本関数が抽出できるもの、又は何度も同じ実験が行なえるものなど) がもとまるときは、これらの検定は比較的楽であろう。すなわち、各標本関数の平均値と二乗平均値をもとめ、それらが標本によらないかどうかをみればよい (その検定法としてはランテスト、トレンドテストなどが便利である) からである。何故なら、相関関数の $\tau = 0$ での値は二乗平均値であるから ((3.3) 参照) この値が定常でなければ相関関数は定常でないはずだからである。

しかし、我々が今考えているのは、このような恵まれた条件の下ではない、一本の有限なデータの定常性である。このようなデータの定常性をきちんと検定するのは実際不可能に近い。したがって、しばしば長年のカンで見極める場合が多いのも事実である。ただ、平均値がゼロでない場合 (これが非定常性だというわけではない) や、トレンドを予想して、前もって自動的に差引いて相関関数をもとめる便利な方法はある。無論この方法は、平均値とトレンドを差引いたものが定常だという仮定のもとでの話である。また、トレンドや長周期の波が非定常性の原因であるなら、パワー・スペクトルをもとめたとき振動数がゼロの近傍でパワーがゼロにならないで、ある分布を示すであろう。つまり、パワー・スペクトルをまずもとめてみれば、このようなトレンド等の存在が推定できる訳で、パワー・スペクトルが定常性の検定にも使える場合がある。

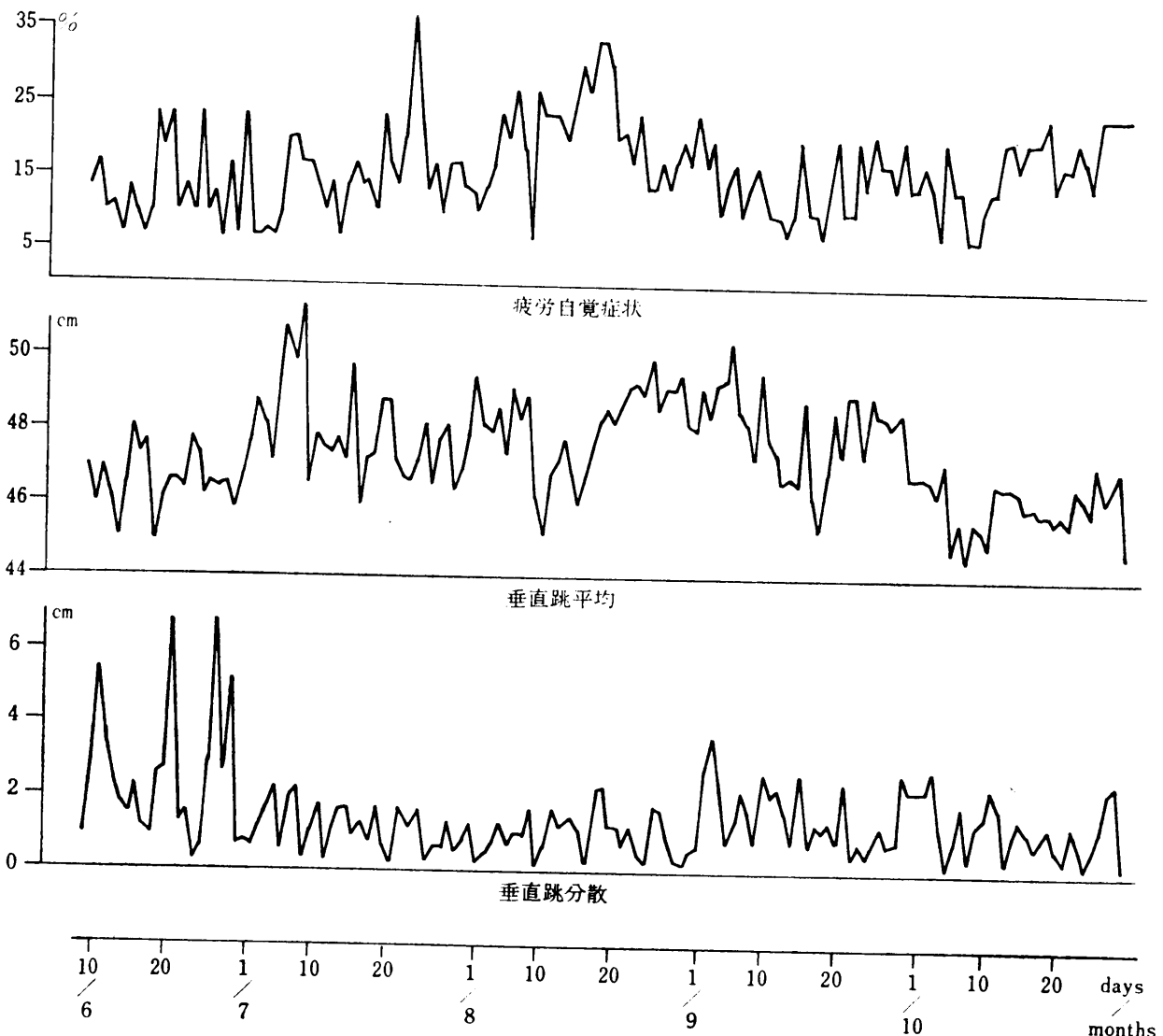
ここでは、以上簡単に述べるにとどめ、具体的検討等については他に譲ることにする。

7. 解析例 一疲労の自覚症状と垂直跳一

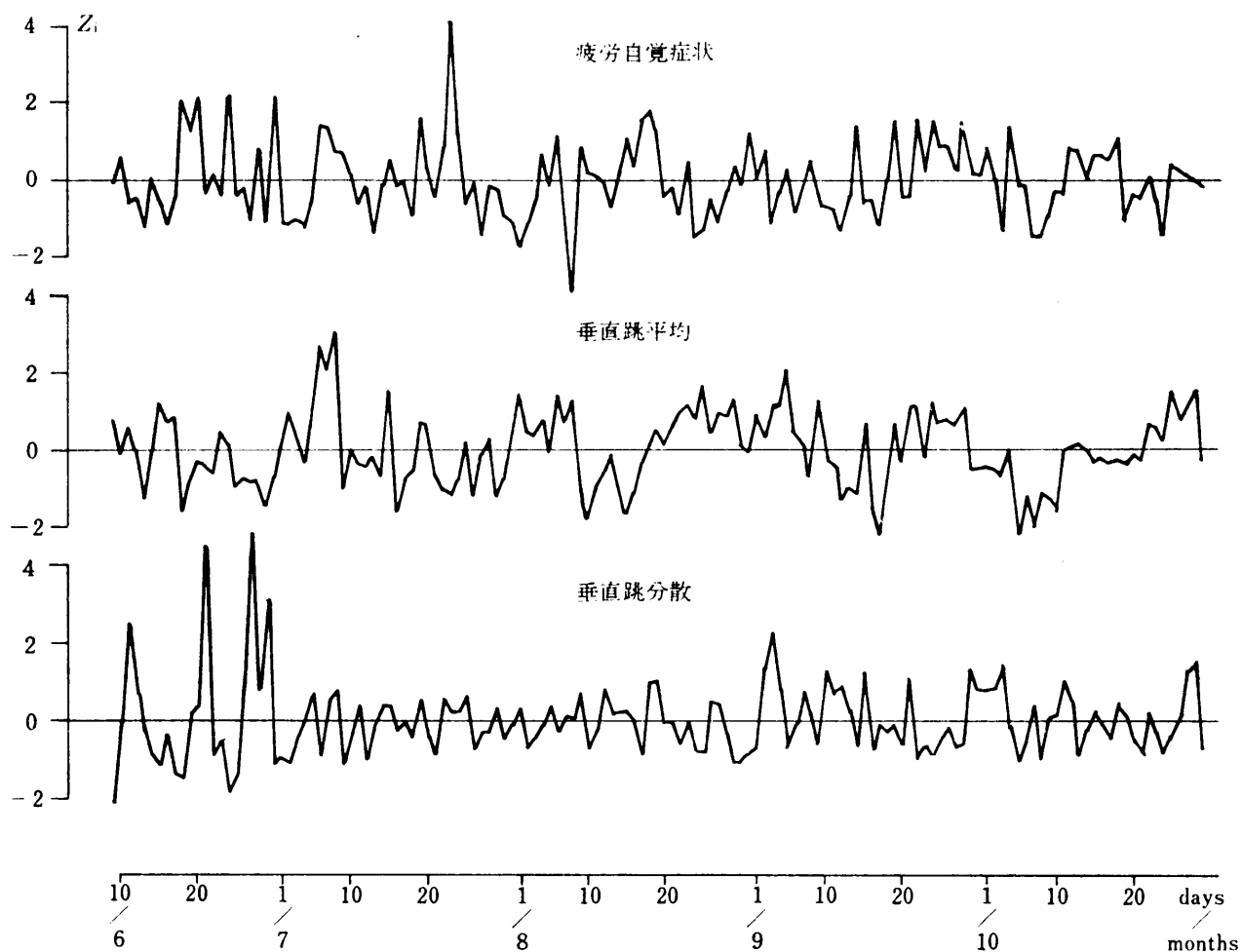
稲村等^{1),2)}は、自己相関係数(系列相関係数)のコレログラムにおける波動の頂点の間隔を利用し、男女各1名の「人間のリズム」をみいだしている。しかしそこでは、離散データの解析に対する検討がなされていなかったため、不備な点が幾つか生じている。例えば自己相関係数を求めるとき、 m を大きくとりすぎているので、コレログラムの波動に「不安定性」が生じていることである。そこで、その中から男子被験者における疲労自覚症状の出現頻度、および垂直跳の平均と分散を引用し、本論文の方法で解析してみよう。

第8図は引用したデータである。このデータには、トレンドが含まれているので注意を要する。第9図はこのデータからトレンドを除いたものであり、ここではそれを定常不規則過程の標本関数とみなし、その解析を試みた。

本論文の方法では、このデータについての分解能を 0.07 ($N=143$ 日, $h=1$ 日, $f_N=0.5c/day$, $m=15$) 程度しかとれない。しかしそれではこの被験者の「人間のリズム」がもつ $20\sim 30$ 日の周期、すなわち $0.05c/day$ の振動を分解できない。そこで分解能を $0.02c/day$ (この場合 $m=50$) 程度まであげる必要があり、ランダム誤差に注意を払いながら、 $m=10$ から $15, 20, 25, 50$ の場合を求めた。



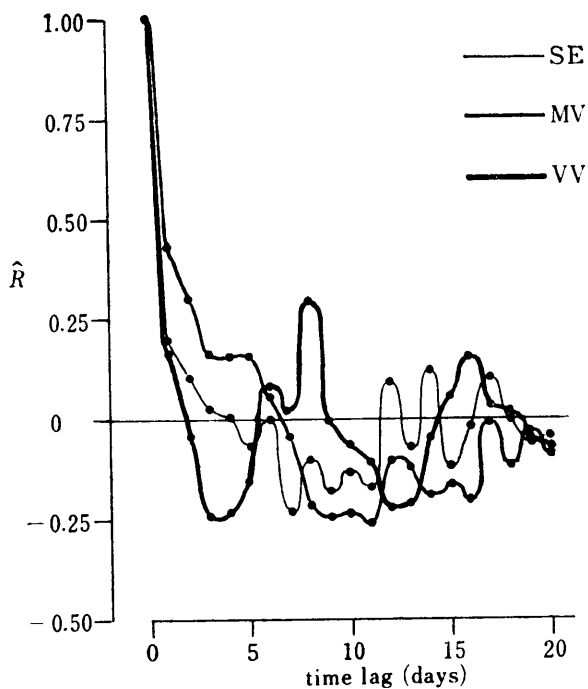
第8図



第9図 $Z_i = (x_i - \mu_i) / \sigma$, ここで μ_i は平均値, σ は標準偏差

文献1)と2)は、引用したデータについて(i)3種目とも20~30日の周期をもっている、(ii)疲労自覚症状におけるそのリズムは、複数の周期成分の合成波と予測できる、(iii)垂直跳の分散には6~8日の周期も含まれている、(iv)20~30日周期のリズムは、互いに相関があり位相差をもっている、などを明らかにしている。ここでは以下考察の便の為に、疲労の自覚症状を SE、垂直跳の平均を MV、その分散を VV と呼ぶことにしよう。

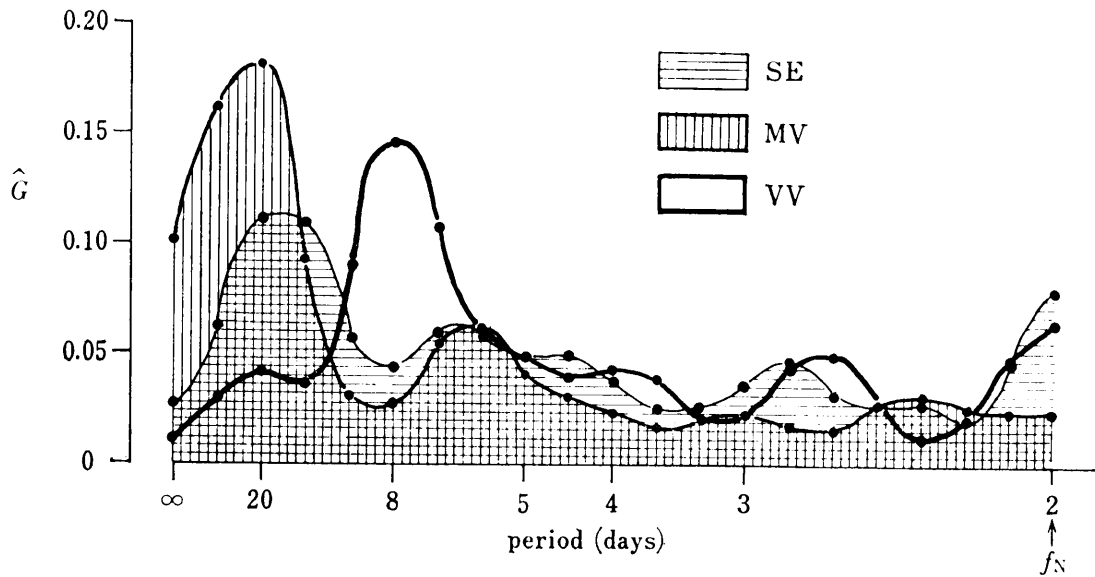
第10図は、第9図のデータから求めた $m=20$ の場合の自己相関関数である。また第11図は同じ場合のパワー・スペクトルである。パワー・スペクトルは $m=50$ の場合まで求めたが、それでも周期の判定には支障がなかった。第11図の f_x 、すなわち周期にして2日のピークは、「エーリアシング」によるものではないかとも考えられる。そこで、このデータの測定条件では最も強いパワーを持つと思われる1日の周期について計算をしてみた。しかし結果は該当しなかった。実際にこのデータでは、それは1日1回同時刻のサンプリング条件でカットされているはずである。



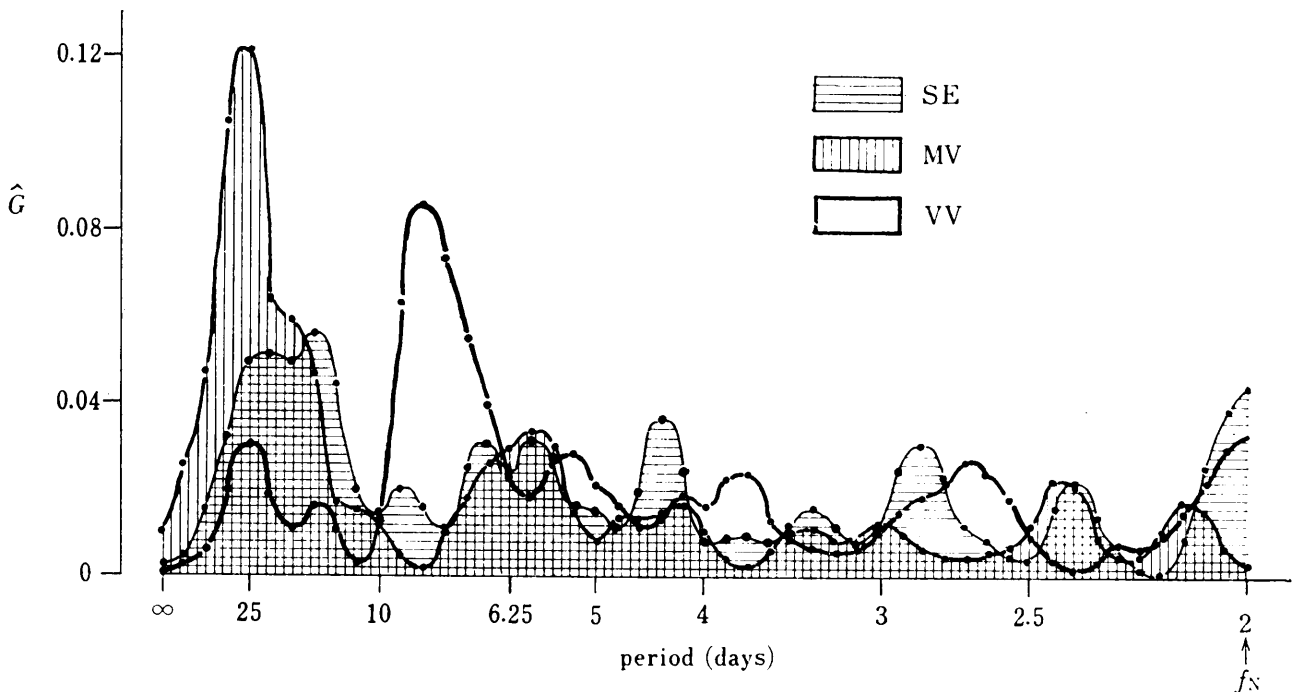
第10図

さて、第11図では周期20日のところに3種目とも、パワー・スペクトルのピークがある。したがって、これまでの結果(i)は再確認されたわけである。その周期の範囲は16.0~26.7日で、振動の強さの順は、MV、SE、VV である。また VV は8日 (7.3~8.9日) のピークであるが、これは結果(iii)に相当するものである。さらにこの図には、SE と MV に6日前後(5.3~7.3日)、SE と VV には2.7日前後 (2.6~3.0日) のピークがみられる。この2つはこれまで解析し得なかったものであるが、今後検討を要する。

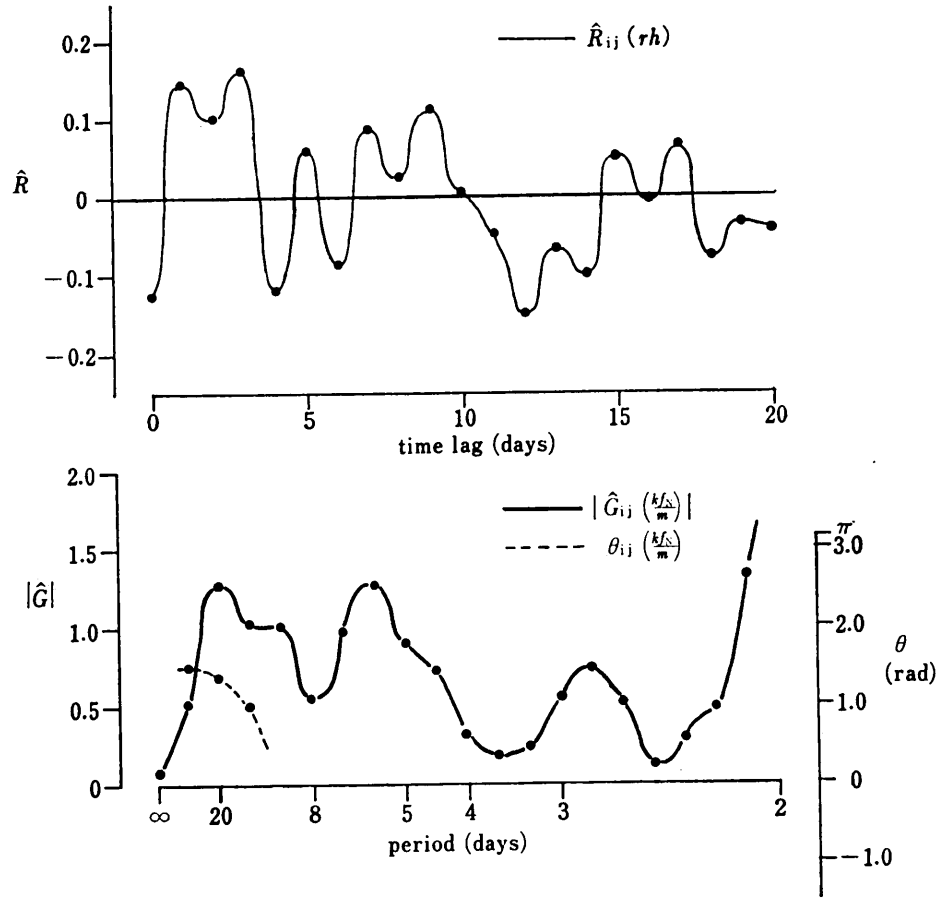
第12図は、 $m=50$ の場合のパワー・スペクトルである。この図から、先の20日のピークが、25日 (22.3~28.6日) と14.3日 (13.3~15.4日) の二つの周期成分から成立っていることを知ることができる。稲村等の結果(ii)に加えて、ここではMV、VV についても知ることができた。またさらに、SE における2つの成分は同じ程度のパワーをもち、MV と VV においては25日の成分の方が強いことを知ることができる。



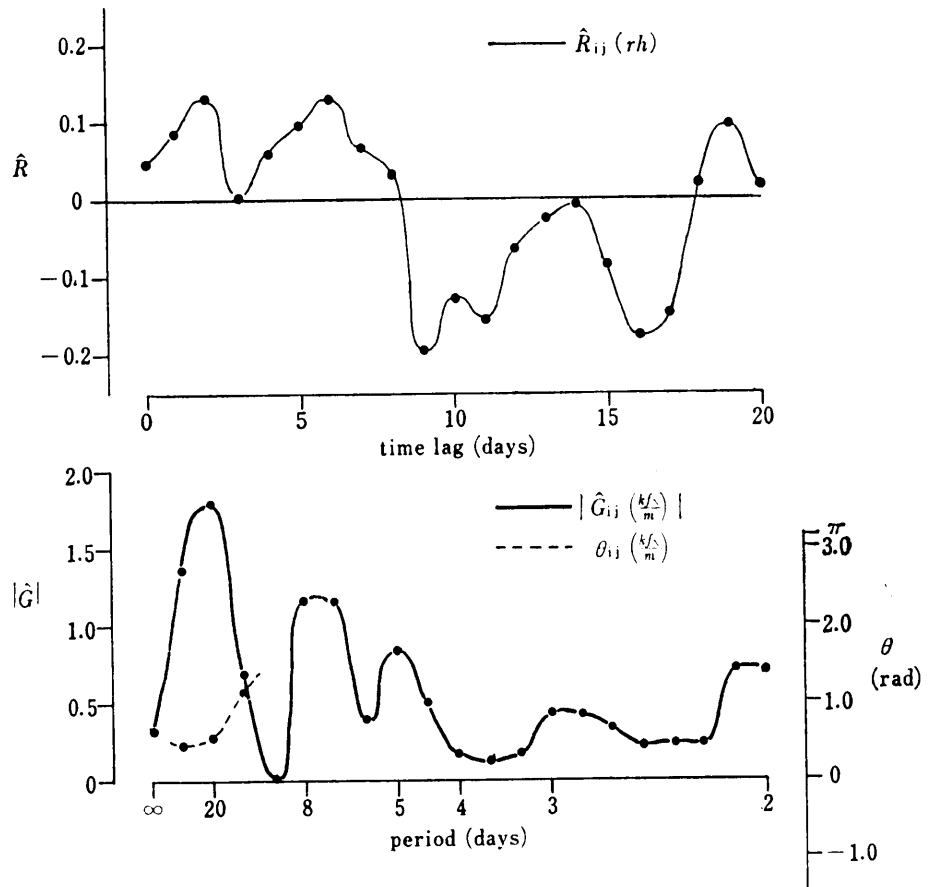
第11図



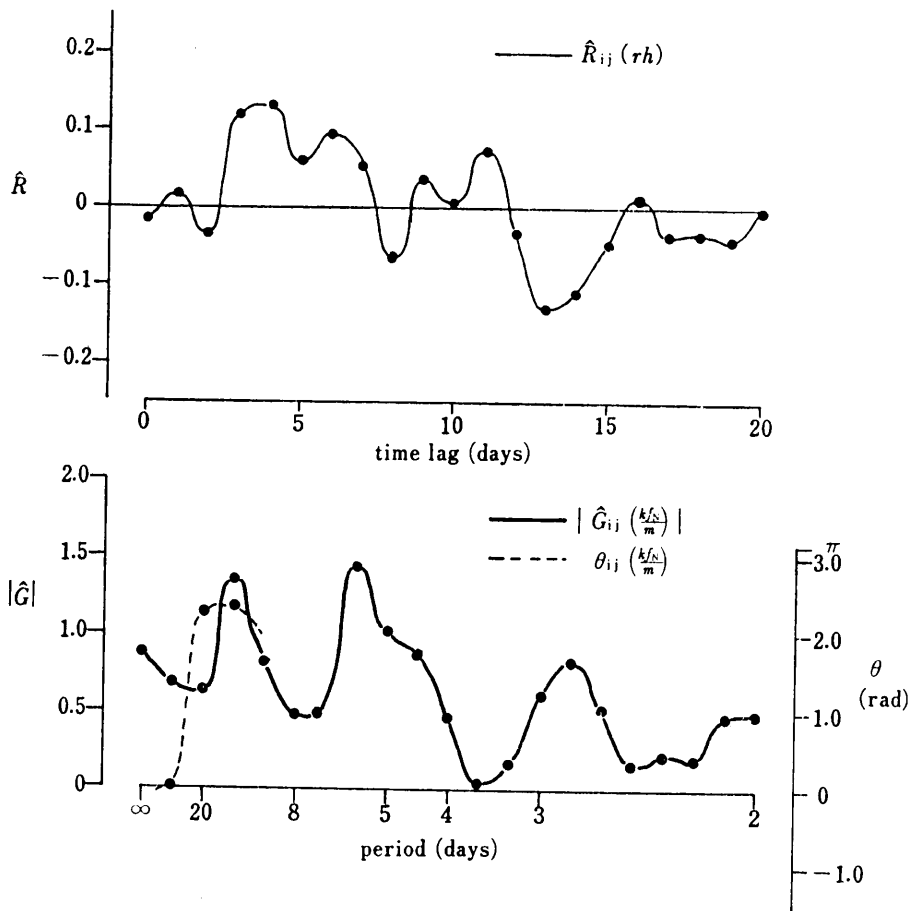
第12図



第13图 x_i —SE x_j —VV



第14图 x_i —VV x_j —MV



第15図 x_i —SE x_j —MV

ここで、3種目に共通する20日のピークの相互関係を、相互スペクトルで解析してみよう。注意しなければならないことは、MVのデータの符号を逆転しなければならないことである。なぜなら、垂直跳の記録が良い時と疲労症状が多い時の関連性を求めても、体力科学ではあまり意味を持たないと思われるからである。

第13図はSEを x_i 、VVを x_j 、第14図はVVを x_i 、MVを x_j 、第15図はSEを x_i 、MVを x_j として求めた相互相関関数と相互スペクトルである。ここでは第13図と第14図に20日のピークがみられる。したがってSEとVV、VVとMVでは相互関係があるとみられる。また第15図では20日のところにピークがなくなり、相関が弱くなっている。位相差は、時間で計算したところ、VVはSEに4.5日遅れ、MVはVVに1.7日遅れている。そこで念のため、SEとMVの間を計算すると、MVがSEに7.2日遅れ、0.8日の誤差はあるが、一応つじつまは合っている。したがって、ここでは稲村等の結果(iv)に加えて、3種目の周期がSE、VV、MVの順で時間的にならんでいることを知ることができた。

さて、詳細な生理学的、体力科学的検討は他に譲るとしても、3種目における20日の周期の順番が、SE→VV→MVとなっていることは大変興味あることである。何らかの原因で疲労の自覚症状が現れると、それから4日程して垂直跳の記録が不安定になりちらばる。そして2日程してその記録が落ちてしまう。これだけのデータで、本当にリズムが伝搬しているとは結論できない。だが、もし伝搬しているとすれば、SEとMVの相関が弱くなることは、伝搬していくうちにそのリズムが変形してしまうことを示しているのではなからうか。

一方、比較的短時間の作業による疲労では、まず疲労感が起り、次に作業が不正確になり、最後に本当に疲れて作業成績が落ちるといった現象がしばしばみられる。SE→VV→MVの関係はこの経過とよく似ている。このデータの場合の疲労症状は、長期間通常的生活をしていた

間の中枢性疲労症状であるから、一般の作業による疲労とは異なる。したがって、この結果から両者の関係を論ずることは早計であるが、もし性質の異なる疲労の長期変動と短期変動とが、同じ経過をもつとしたら、それは生理学・体力科学にとって大変興味のある問題である。

参考分献

- 1) 稲村欣作ほか;「人間のリズム」について, 静岡大学教養部研究報告 No. 2~No. 8 (1966~1972)
- 2) 稲村欣作ほか;「精神身体機能の定常時系列における長周期リズム」, 静岡大学教養部研究報告 No. 10 (1974)
- 3) W. B. Danvenport *et al.*; “An Introduction to the Theory of Random Signals and Noise” McGRAW-Hill Book Company, Inc.
- 4) R. B. Blackman *et al.*; “The Measurement of Power Spectra”, Dover Publications, Inc.
- 5) J. S. Bendat *et al.*; “Measurement and Analysis of Random Data” John Wiley & Son, Inc.
- 6) 滝 保夫編; 岩波講座基礎工学3「確率統計現象Ⅰ, Ⅱ」, 岩波書店
- 7) A. sollberger; “Biological Rhythm Research”, Elsevier Publishing Company, Inc.