統計数理研究所彙報
第30巻 第1号 (1982)

ある生物化学の現象解析のモデルと分布に関する覚え書

統計数理研究所 樋口伊佐夫

(1982年8月 受付)

1. はじめに

ここで述べる話の題材は，今から17〜8年ほど前に，筆者が東大生物化学教室の野田春彦先生の研究室に出入りして，教わったり相談を受けたりしていた頃の古い研究ノートから取出したものである．話題の背景にある科学的な問題は，生体内にある或る種の棒状に会合した高分子に超短波を照射すると節目で切れるが，放置すると復元するという現象を，機構的に立ち入って調べようとするもので，問題の設定とアプローチは，筆者の関知する側面を眺めただけでも，かなりの水準のものであったことが窺える．残念なことにいろいろの事情があってこの研究は完遂されずに終ったようである．

実はこの問題に関する筆者の仕事は，いろいろの分割モデルを作って，それらから計算される粒度分布（会合した分子を一つの粒子とみる）の時間変化のパターンを実測のパターンと比べて，どのような切断の仕方をするのかについて知見を得ようとするものであった．これについては一応の方法を用意したが，実験の精度が低く，決定的な決論を下すことができなかった．またモデルももも少し精密化してみる必要があったので論文にはしていなかった．しかし方法論としては面白いと思ったので，「情報処理と統計数理」を執筆した際に，かなり精しく紹介した [1]．この研究はその後発展させる機会を逸し，そのまゝになっている．

今回この覚え書きで述べるのは，そのモデル解析についてではなく，その仕事をしていた際に書きとめておいた，同じ現象に対する別のモデル群に関することである．このモデル群は実は上述の生物化学の問題の解決に役立てようという純粋で緊迫した動機から発想したものでなく，ある一般的で抽象的な問題を考えようとし，これをその素材にしようという邪まな魂胆に発するものである．その一般的問題というのは，科学的認識の際における異なった考えに基く種類（カテゴリー）の異るモデル群の得失を情報論的に評価して論じようというものである．ところが少し具体的に立入るとすぐわかることであるが，この題材と思いつきのモデル群との組合わせは，こうした一般論の雛型としては適当でない．それでこの研究もほんの着手のところで中止したまゝである．今筆者はこうした一般論には関心がないので，別の素材を探してまでモデル論を展開しようという意慾はない．それでかつてこんなことを考えたことがあるという意味で紹介しておきたい．ただモデルはあまりにもたわいないものであるから，動機の説明をもつけ加えさせていただく．

さらに問題の背景をなす科学的な研究はこうした数理統計学的発想のものとは趣きを異にする．その雰囲気を伝えておかないと，モデル解析の局面の話は非常に誤った印象を与える恐れがある．しかしこれはなかなかむつかしく，到底筆者の能くするところではない．そこで筆者の関与した範囲内でこの問題についての歴史的な点描をつけておきたい．筆者が野田研究室に出入りしていた頃，野田研究室の方からも若い人がつぎつぎと統計数理研究所に出入りして仕事をした．これは進歩のはげしい科学のある領域と統計数理との交流の一つの実例をなすもの

で，仕事の脈絡は書きとどめておく価値があると思う．

2. 一般論志向の契機

まず現象の具体的な説明をしておく．

実際の自然現象は複雑であるが，単純化して，つぎの情況を考える：

長さ L の棒状粒子にある種の刺激を加えると，節目のところで切れる．節目は3ヶ所あって等間隔に存在する．従って切れた後にできる粒子の長さは $L/4$, $2L/4$, $3L/4$ であるが，切れない場合もあるので，長さ L のものもある．簡単のために，これらをそれぞれ長さ k ($k=1, 2, 3, 4$) の粒子と呼ぶことにする．一般に粒子の両端は区別されるものとする．(たとえば分極している粒子の＋側を左側とする.) 節目には左から順に 1, 2, 3 の番号を付して，2番目の節目というように呼ぶ．測定されるものは，非常に多くの粒子の集団における粒度分布，すなわちこれらの長さの粒子の個数の割合である．——実際には体積濃度，すなわち粒度の重量分布が測定されるが，この場合一意的に重量分布から度数分布がでてくるので，度数分布が測定されるものと考えてよい．——

さて問題（知りたいこと）は，“どのような具合に切れるのか？”ということである．どのような具合にというのは，たとえば，まず中央から切れ，ついでまた二つに切れるとか，端から順次切れるとか，長さに比例した確率で切れるとか，そのようなたぐいのことである．

この問いに対する実際のアプローチは次のようにした．刺激の強さを一定に保ちつゝ，いろいろの時間 $t_1, t_2, \cdots$ の間刺激をかけた試料を測定し，この一連の結果を，連続して一定の刺激を加えているときの時刻 $t_1, t_2, \cdots$ における状況とみなし，粒度分布の変化と，切れ方をいろいろ仮定した，いろいろの確率過程モデルにおける粒度分布の変化のパターンとを比較する．実験結果とよく似たパターンをもつモデルにおいて仮定した切れ方を実際の切れ方と見做すわけである．

ここでモデルをつくるとき，つぎのような仮定をする．(i) 切断の現象は確率事象である．(ii) 粒子は互いに独立に切れる．(iii) 単位時間内に切れる確率は時間と共に変ることはなく，常に一定で，各粒子に共通している．

もしこれらの仮定がうまくないということがわかれば，その段階で改める．たとえば (ii) がまずいときは付和雷同性に関する何らかの仮定を入れ，(iii) がまずいときは，寿命分布を導入しなければならない．しかし実際にこれがあやしくなったときは，モデルだけをいじるのでなく，どういう実験をすれば疑問が解明できるかというアプローチの仕方を考えなおすのがこういった科学における普通の発想であろう．

それはさておき，筆者の用意したモデルはマルコフ鎖に基づく考えによる．本 [1] にくわしく書いてあるが，簡単に説明しておくと，粒子が切れて，切片がどこに行こうと，1つの粒子の状態が変ったものと考える．状態の種類は図1に示されたものとなる．n 番目のステップの直後の各状態の個数の期待値を A_n, B_n, C_n, D_n, D_n', E_n とすると，実験中再結合は起らないとして，

$$U_{n+1} = U_n Q^T$$

とあらわされる．ここで U_n はベクトル $(A_n, B_n, C_n, D_n, D_n', E_n)$ で $Q=\{q_{ij}\}$ は遷移確率の行列，Q^T は Q の転置行列で，再結合が起らないとすると $i<j$ の q_{ij} は0である．また $B\to C$, $C\to D$, $D\to D'$ も起らないから $q_{32}=q_{43}=q_{54}=0$ である．U_n は行列 Q の固有値 $\lambda_1, \cdots, \lambda_n$ と Q の要素により比較的簡単にあらわされる．さらにいろいろの切れ方を仮定することは q_{ij} および λ_i の間にいくつかの関係を仮定することになる．また1ステップの間に切れる

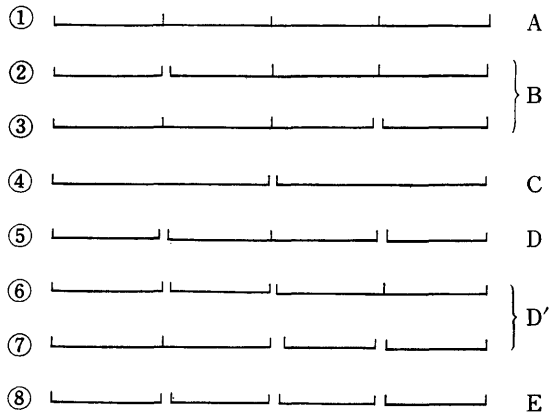

図　1

場所とその確率を指定すると，q_{ij} や λ_i をそれを用いて表わすことができる．1ステップの間の時間を Δt，その間におこる切断の諸確率は $k\Delta t+o(\Delta t)$ の形になっているものとし，$n\Delta t=t$ を固定して $n\to\infty$ にした極限を求めると U_n の各要素の極限は $e^{-\alpha kt}$ のいくつかの異なる α のものの一次結合となる．また粒度分布は U_n の要素の一次結合として容易に求まるので，時間を連続にしたときの確率過程モデルの粒度分布の時間変化が，いろいろの切断の仕方の仮定について計算される．

たとえば時間 Δt の間で3個の節目のうち1箇所でしか切れないとし，その確率はどの点も等しく $w_1=k\Delta t+o(\Delta t)$ とすると，粒度の重量分布は

$$\rho_1/\rho=\frac{1}{2}e^{-2kt}-\frac{3}{2}e^{-kt}+1$$

$$\rho_2/\rho=\frac{1}{2}e^{-3kt}-2e^{-2kt}+\frac{3}{2}e^{-kt}$$

$$\rho_3/\rho=-\frac{3}{2}e^{-3kt}+\frac{3}{2}e^{-2kt}$$

$$\rho_4/\rho=e^{-3kt}$$

のようになる．ここで ρ はもとの体積密度，ρ_i は長さが短い方から i 番目の粒子の体積密度である．なおこれは，次に述べる分割モデルの各点の切れる確率がすべて $p=1-e^{-kt}$ として独立とした場合の結果と同じである．

このようにしてつくられるモデル群を $\mathcal{A}$ であらわす．どこかで切れると次にきれやすいとか，微小時間に切れる確率はそのときの粒子の大きさに比例するとか，回転能率に比例するとかいった内容も遷移行列 Q の要素の関係に織り込むことができる．

ところで別のカテゴリーのモデルが容易に浮び上る．それはただ単に3箇所の節目で，切れるかきれないかの同時確率の組を与えることである．すなわち切れることを1，切れないことを0であらわし，1個の粒子について，第1と第2の節目が切れ第3の節目が切れない確率を P(110) のようにあらわすとき，P(000)，P(001)，…，P(111) の8個の値の組を与えることである．このようなものに対しても切れ方のモデルが或る程度つくれる．たとえば節目が1，2，3の順に切れるという場合には2，3が切れて1が切れないことは起らないから P (011)=

$P(010)=0$, 3が切れて，1，2が切れないことも起らないから $P(001)=P(101)=0$. すなわち8個の P のうち特定の4個が0であることにより特徴づけられる．このようなモデル群を $\mathscr{B}$ とあらわす．

ところで，この問題のモデルとしてはモデル群 $\mathscr{A}$ の方が $\mathscr{B}$ よりもぴったりしている感じがする．それは何故かというのが，問題なのである．$\mathscr{B}$ には時間が入っていないので直接は比較にならないが，パラメタとして入れて $P(000;t)$, $P(101;t)$ のようなものを考えることにする．こうしてもやはり $\mathscr{B}$ が問題に適合しないように見える．それは切れ方のパタンとして人々が心にもっている概念は時間的な構造も内包として含んでいて，それをモデルに表現するとき $\mathscr{B}$ の方はうまくできにくいということではないかと思う．

統計家の中にはしかし $\mathscr{A}$ のモデル群を嫌って $\mathscr{B}$ のモデルをよしとする人が多いのではないかと思う．そして t に関するレグレッション解析をすることを考えるかも知れない．それは $\mathscr{A}$ におけるいろいろの仮定，たとえば，マルコフだとか，寿命無視だとかいったことに対する批判や不安が根底をなすかも知れない．しかしこうした堅実ではあるが消極的な考えでいくと間々科学者や一般の普通人の知りたいことに鈍くしか対応できないことになろう．

たとえば，"大きさに比例して切れる" という命題に対応するモデルを $\mathscr{B}$ の中でみつけることはむつかしい．統計家が好んで用いるモデル群（モデルをつくる基本的考え方）の中には普通の人の知りたいこととピントがずれている場合が間々ある．こうしたモデル群の中で推定検定或いはモデル選択をいくらしっかりやっても統計家以外の人にはあまり恩恵にならない．ということは大分前（1960年頃——統計的モデル解析についての論説 [9] を書いたころ）から気になっていたのである．つけ加えておきたいことは昔は統計家が，理論計算の片手間に，或いは抽象的発想の下でつくった統計理論の例題として応用を行うことが多かったので特にそのような感じがした．最近ではこういう傾向が減少しているのであまり気にならない．なおモデル群 $\mathscr{B}$ を悪く言ったがそう思えるということで断定はできない．それこそこの本題の核心で，未解決の問題である．

さてこうしたモデル論を一般的にきちんと行うにはどうしたらよいか．知りたい命題族に対応するモデルの間の包含関係を見ることが一つの手がかりであろう．ここでモデルⅠがモデルⅡを含むということは，モデルⅡから出てくる分布はすべてモデルⅠからも導出され得ることを意味する．ある考えによるモデル群に属するモデルが，区別されるべき概念に関して互いに素であれば，この考えは問題に対して適確な答を用意しているが，重なるところが多ければ，こういう考え方は問題に対して最初からあいまいな対応しかしていない．そこでモデル群 $\mathscr{A}$ および $\mathscr{B}$ について包含関係をしらべようとした．ところでこの例題では測定される体積密度は4種の粒度に関して常に1となるので，自由度は3しかない．それで粒度分布といっても3個の数値をどう与えるかというだけであり，モデルを記述するのにどれ位こまかくするかつまりパラメタがいくつあるかが決定的にいろいろの性格をきめてしまう．しらべたいような性格はみなその影にかくれてしまうので，一般論を考えるときの雛型として適当でないことがわかった．それで調べることを放棄したのである．

確率過程をもっと拡大し，マルコフ鎖を仮定することは結果にどれだけの制限を加えているかなどを調べることはやはり興味はあることであるが，この特別の場合の計算がどれだけ意味をもつかわからない．

3. モデル群 $\mathscr{B}$ について

各粒子における切れる場所と生れる粒子の長さと個数の関係は表1のごとくなる．表1の○

表 1

切れる場所			粒子の長さ			
1	2	3	1	2	3	4
×	×	×	0	0	0	1
(×, ○)	×	(○, ×)	1	0	1	0
×	○	×	0	2	0	0
○	×	○	2	1	0	0
(○, ×)	○	×	2	1	0	0
○	○	○	4	0	0	0

は切れること，×は切れないことをあらわす．

切れることを1，切れないことを0であらわし，一つの粒子について，表1の×××の起る確率を $\mathrm{P}(000)$，××○の起る確率を $\mathrm{P}(001)$ 等とあらわす．切断に関して粒子間に相互作用はないものと考える．最初の粒子の個数を N，刺激を加えた後における長さ i の粒子の個数の期待値を N_i とすると，表1からただちに

$$N_1 = N\{\mathrm{P}(001)+\mathrm{P}(100)+2\mathrm{P}(101)+2\mathrm{P}(110)+2\mathrm{P}(011)+4\mathrm{P}(111)\}$$
$$N_2 = N\{2\mathrm{P}(010)+\mathrm{P}(101)+\mathrm{P}(110)+\mathrm{P}(011)\}$$
$$N_3 = N\{\mathrm{P}(001)+\mathrm{P}(100)\}$$
$$N_4 = N\,\mathrm{P}(000)$$

粒子数 N がきわめて大きいので，個数とその期待値を区別する必要はない．体積密度についても同様で，もとの体積密度を ρ，結果における長さ i の粒子の体積密度を ρ_i とすると，

$$\rho_i = (i/4)\cdot(N_i/N)\rho \qquad (i=1,\ 2,\ 3,\ 4)$$

となる．いう迄もなく，$\sum N_i$ は N に等しくないが，$\sum \rho_i$ は ρ に等しい．相対体積密度 ρ_i/ρ を C_i であらわす．$\sum C_i = 1$ である．以下 $\{C_i\}\,(i=1,\ \cdots,\ 4)$ を単に分布といい，割れる確率と，その結果生じる分布との関係を考察する．

まず各場所での切断が全部互いに独立の場合を考える．それぞれの場所での切断の確率を p_i $(i=1,\ 2,\ 3)$ とすると

$$N_1 = N\,(1+p_2)\,(p_1+p_3)$$
$$N_2 = N\,(2p_2-p_1p_2-p_2p_3+p_1p_3-p_1p_2p_3)$$
$$N_3 = N\,(1-p_2)\,(p_1+p_3-2p_1p_3)$$
$$N_4 = N\,(1-p_1)\,(1-p_2)\,(1-p_3)$$

となる．従って

$$N_1+N_2+N_3+N_4 = N\,(1+p_1+p_2+p_3)$$

切れることにより粒子数は当然増加する．

$$C_1 = \frac{1}{4}\,(1+p_2)\,(p_1+p_3)$$

$$C_2 = \frac{1}{2}\,(2p_2-p_1p_2-p_2p_3+p_1p_3-p_1p_2p_3)$$

$$C_3=\frac{3}{4}(1-p_2)(p_1+p_3-2p_1p_3)$$
$$C_4=(1-p_1)(1-p_2)(1-p_3)$$

この場合を IND 1 と記し，とくに $p_1=p_3$ のときを IND 2，さらに $p_1=p_2=p_3$ のときを IND 3 とあらわす．いう迄もなく

$$\text{IND}\,1\supset\text{IND}\,2\supset\text{IND}\,3$$

なる包含関係がある．

つぎに切断が場所に関して非独立の場合．

(D I) 1箇所が切れると他も切れる．

切れる確率を p とすると $P(000)=1-p$, $P(111)=p$, 其の他の $P=0$, 従って $N_1=4Np$, $N_2=0$, $N_3=0$, $N_4=N(1-p)$

$$\sum N_i=N(1+3p)$$
$$C_1=p,\ C_2=0,\ C_3=0,\ C_4=1-p$$

(D II) 先ず1と3が同時に切れる（方向性なし）．

$$P(000)=1-p-p^*,\ P(101)=p,\ P(111)=p^*,\ \text{その他の P は 0}$$

ここで $P(010)$ なることに注意．これは先に2が切れることがないことを示す．

$$N_1=N(2p+4p^*),\ N_2=Np,\ N_3=0,\ N_4=N(1-p-p^*)$$
$$\sum N_i=N(1+2p+3p^*)$$
$$C_1=\frac{1}{2}(p+2p^*),\ C_2=\frac{1}{2}p,\ C_3=0,\ C_4=1-p-p^*$$

(D III) どこか1箇所きれると，つぎに切れやすくなる（或いは切れにくくなる）．

ここで最初に切れる確率を p，次にきれる確率を p^*, ･･･ と確率はステップ毎に異るが，場所にはよらないとする．

$$P(100)=P(010)=P(001)=p$$
$$P(110)=P(011)=P(101)=p^*$$
$$P(111)=p^{**},\ P(000)=1-p^{**}-3p^*-3p$$
$$N_1=N(2p+6p^*+4p^{**}),\ N_2=N(2p+3p^*),\ N_3=2Np,$$
$$N_4=N(1-p^{**}-3p^*-3p)$$
$$\sum N_i=N(1+3p+6p^*+3p^{**})$$

よって $C_1=\frac{1}{2}(p+3p^*+2p^{**})$, $C_2=p+\frac{3}{2}p^*$, $C_3=\frac{3}{2}p$, $C_4=1-p^{**}-3p^*-3p$

(D IV) まず真中が切れる（方向性なしとする）．

$$P(010)=p,\ P(110)=P(011)=p^*,\ P(111)=p^{**},\ P(101)=P(100)$$
$$=P(001)=0,\ P(111)=p^{**},\ P(000)=1-2p^*-p-p^{**}$$

と定式化されるので

$$N_1=4N(p^*+p^{**}),\ N_2=2N(p+p^*),\ N_3=0,\ N_4=N(1-2p^*-p-p^{**})$$
$$\sum N_i=N(1+p+4p^*+3p^{**})$$
$$C_1=p^*+p^{**},\ C_2=p+p^*,\ C_3=0,\ C_4=1-2p^*-p-p^{**}$$

(D V)　端から順に切れる（方向性なしとする）.

$P(010)=0,\ P(100)=P(001)=p$

$P(110)=P(011)=p^*,\ P(111)=p^{**}$

$P(101)=0,\ P(000)=1-2p-2p^*-p^{**}$

$N_1=2N(p+2p^*+2p^{**}),\ N_2=2Np^*,\ N_3=2Np,\ N_4=N(1-2p-2p^*-p^{**})$

$\sum N_i=N(1+2p+4p^*+3p^{**})$

$$C_1=\frac{1}{2}p+p^*+p^{**},\ C_2=p^*,\ C_3=\frac{3}{2}p,\ C_4=1-2p-2p^*-p^{**}.$$

非独立の場合これらの標題のような文章的な内容が定式化されることは興味深いが，切れる場所が増加すると，このような対応は得られないので，この例は，そういう意味でも一般論の雛型としては不適当である.

さて

$$C_1+C_2+C_3+C_4=1,\quad C_i\geqq 0\ (i=1,\ \cdots,\ 4)$$

だから，粒度分布 $\{C_i\}$ は4次元標準単体の超斜面内の点である．これを C_1, C_2, C_3 の3次元単位立方体の点で表現し，この場合の包含関係をみよう．とくに，独立の場合は領域がどのような形をしているかは興味深い.

非独立の場合

(D I)　(一度に).

$C_2=C_3=0,\ 0\leqq C_1\leqq 1$ だから C_1 軸のみで他の点はない．(図2)

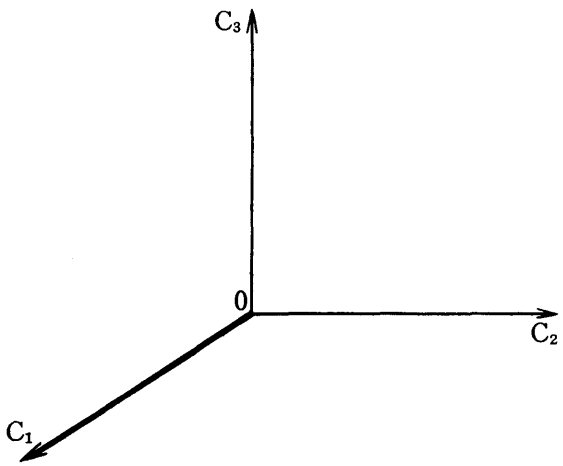

図　2

(D II)　(1と3が最初に同時).

$C_3=0$ の他に $p^*\geqq 0$ から $C_1\geqq C_2$，また当然 $C_1+C_2\leqq 1$. 従って領域は，C_1C_2-面上の (0, 0), (1, 0), $\left(\frac{1}{2},\ \frac{1}{2}\right)$ を頂点とする三角形領域内にある．(図3)

(D III)　(最初どこかが切れると切れやすく（または切れにくく）なる).

$C_1+C_2+C_3\leqq 1$ の条件のほかに，$0\leqq p\leqq 1,\ 0\leqq p^*\leqq 1,\ 0\leqq p^{**}\leqq 1$ の条件がある．これらはそれぞれ $0\leqq C_3\leqq 3/2,\ 3C_2\geqq 2C_3$ と $6C_2-4C_3\leqq 9$，および，$0<C_1-C_2+\frac{1}{3}C_3\leqq 1$ を

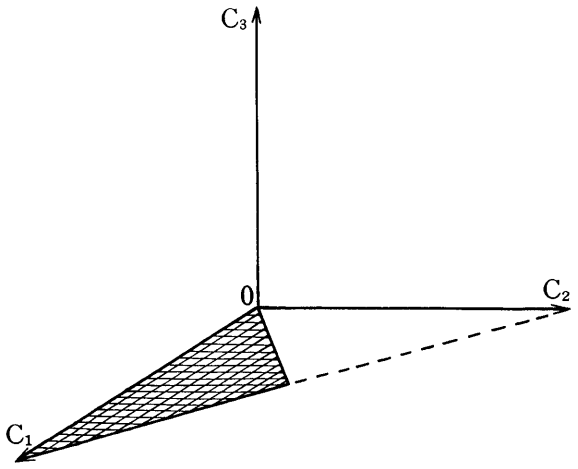

図 3

生じる．これらの諸条件の交りをとると，$C_1+C_2+C_3 \leqq 1$ なる標準単体の中でさらに $3C_2 \geqq 2C_3$ をみたす部分，すなわち標準単体の斜面，C_1C_2-面，C_2C_3-面，および C_1 軸を通り，C_1C_2-面とのなす面角の正切が 3/2 なる面でかこまれた4面体となる．(図4)

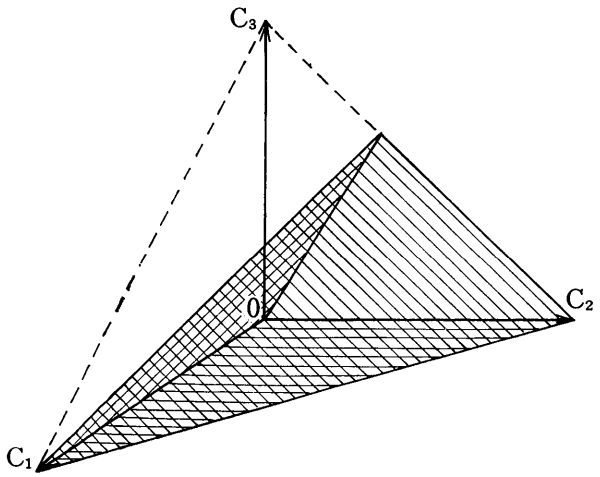

図 4

(D IV) (はじめに真中が切断)．

$C_3=0$ なる条件のため，領域は C_1C_2-面にある．$C_1+C_2 \leqq 1$ でなければならないので，C_1C_2-面内の原点を頂点とする二等辺直角三角形領域である．(図5)

(D V) (端から順)．

$p=(2/3)C_3$，$p^*=C_2$，$p^{**}=C_1-C_2-\frac{1}{3}C_3$ でこれらが0と1の間にあるという条件から，標準4面体内にあるという条件の他に $3(C_1-C_2) \geqq C_3$ をみたす必要がある．$3(C_1-C_2)$

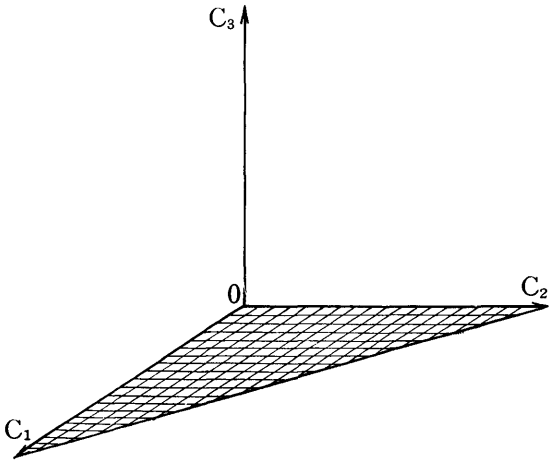

図　5

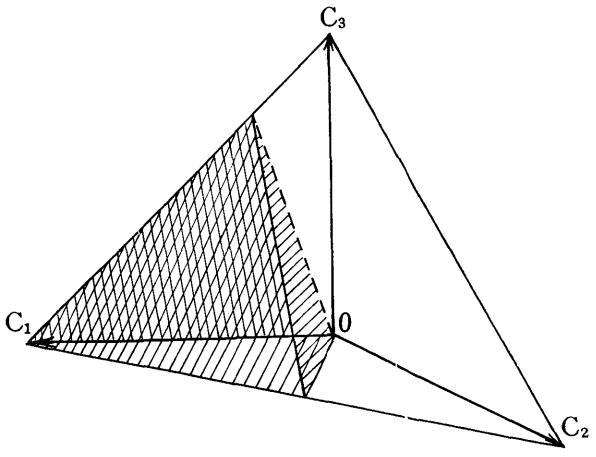

図　6

$=C_3$ は原点を通る面で C_1C_2-面では $C_1=C_2$, C_1C_3-面では $3C_1=C_3$ なる直線である．従って，領域は標準4面体の $C_1=1$ を尖端とする部分をこの面で切りとった4面体である．（図6）

独立の場合

(IND 3)　(すべての p が等しい)．

この場合　$C_1=\frac{1}{2}p\,(1+p)$, $C_2=\frac{1}{2}p\,(2-p-p^2)$, $C_3=\frac{3}{2}p\,(1-p)^2$, $C_4=(1-p)^3$ である．p の各値に対し C_i の点が一つきまるので，領域は曲線をなす．この $\{C_1, C_2, C_3\}$ はやはり標準4面体の中にある空間曲線である．図7の原点から発し (1, 0, 0) にいたる曲線がこれである．

この曲線の C_2C_3-面への射影は，C_2, C_3 の式から p を消去して

$$2\,(3C_2+C_3)^3=27C_3\,(3C_3-2C_3)$$

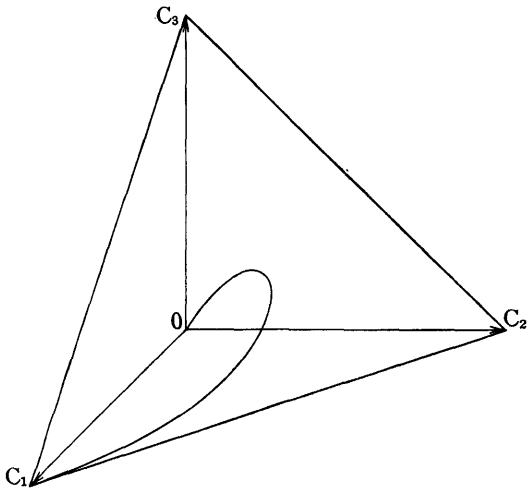

図 7

なる曲線である．この両辺を C_2 について微分した式から

$$\frac{dC_3}{dC_2}=\frac{3(1-3p)(1-p)}{(3p+4)(1-p)-(2+p)}\leqq\frac{3}{2}$$

が得られる．また，この曲線の点においては

$$3C_2-2C_3=\frac{9}{2}p^2(1-p)\geqq 0$$

となる．従って IND 3 は完全に D III の中にある．

DV の中においては $\frac{1}{3}C_3+C_2-C_1\leqq 0$ でなければならない．ここにおける IND 3 は，この関係をパラメク p で表現すると，$p-2p^2\leqq 0$ となる．すなわち $p\geqq\frac{1}{2}$ でなければならない．

(IND 2) $(p_1=p_2)$.

$$C_1=\frac{1}{2}p_1(1+p_2),\quad C_2=p_2(1-p_1)+\frac{1}{2}p_1^2(1-p_2),\quad C_3=\frac{3}{2}p_1(1-p_1)(1-p_2)$$

これから p_1, p_2 を消去する．やゝ面倒な計算の結果

$$\begin{aligned}&36C_1-12C_3-45C_1^2-3C_1C_3-2C_3^2+9C_1^3\\&\quad+12C_1^2C_3+C_1C_3^2=(36C_1+12C_3-9C_1^2-3C_1C_3)C_2\end{aligned}$$

を得る．標準4面体の中で IND 2 の領域はこの曲面である．これがどんな曲面かを見るため，まず C_1C_2-面への射影を見てみる．パラメタ p_1, p_2 をもとにして p_1=一定の曲線群と p_2=一定 の曲線群をえがいたのが図8である．射影は C_1C_2-面の直角二等辺三角形 $C_1+C_2\leqq 1$ を全部覆うのではなく，0, 1 の両端をのぞいて C_1 軸に近い方で一つの連結した領域が空になっている．p_1=一定 の線群を $p_1=0$ の方から見てゆくと，$p_1=0.5$ の曲線より p_1 の大きい側

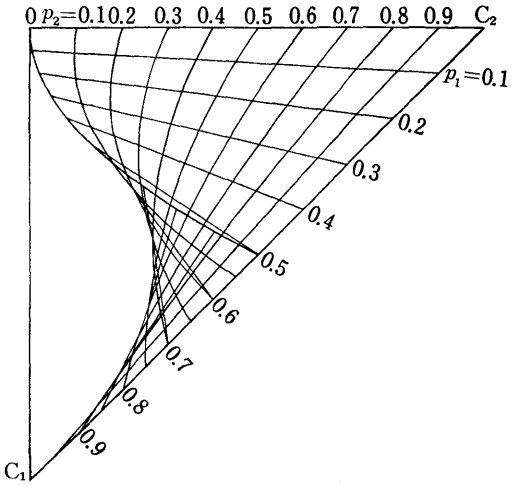

図　8

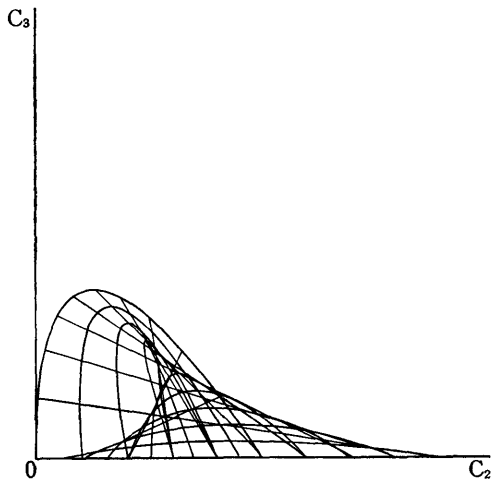

図　9

では重複している．すなわち同一の (C_1, C_2) 点に対して二組の (p_1, p_2) 点が対応する．(p_1, p_2) の組には C_3 の値が一意に対応するので，この領域では，曲面は二重になっているものと思われる．図9は曲面における p_1= 一定，p_2= 一定 の曲線群の C_2C_3-面への投影である．図で山の形をしたのは p_2= 一定 の曲線群で，斜めに上ったままで途中で終っているのは p_1= 一定の曲線群である．曲面の構造は，これだけではわかりにくいが，位相的に言うと次のようなものである．図10は四辺形のシートで伸縮するものとする．B のところをまくりあげて，B が E に来るようにし，BC と EC が一致するようにする．この部分で p_1= 一定をあらわす横縞一つの間隔に p_2= 一定のたて縞は二つずつ対応するようにする．

こうしてできた袋状のシートの A 点は C_1, C_2, C_3 空間の原点　C 点は(1, 0, 0)点，D 点は(0, 1, 0)点になるものとする．対角線 CD では C_3=0 であるが他ではふくらみがある．図

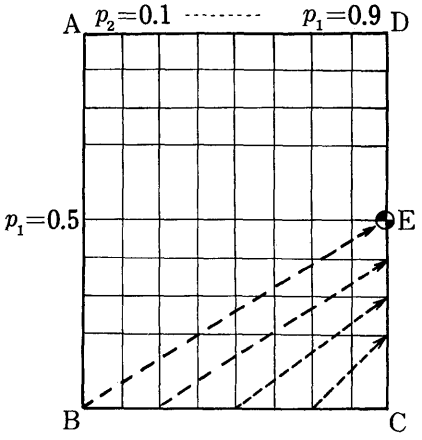

図 10

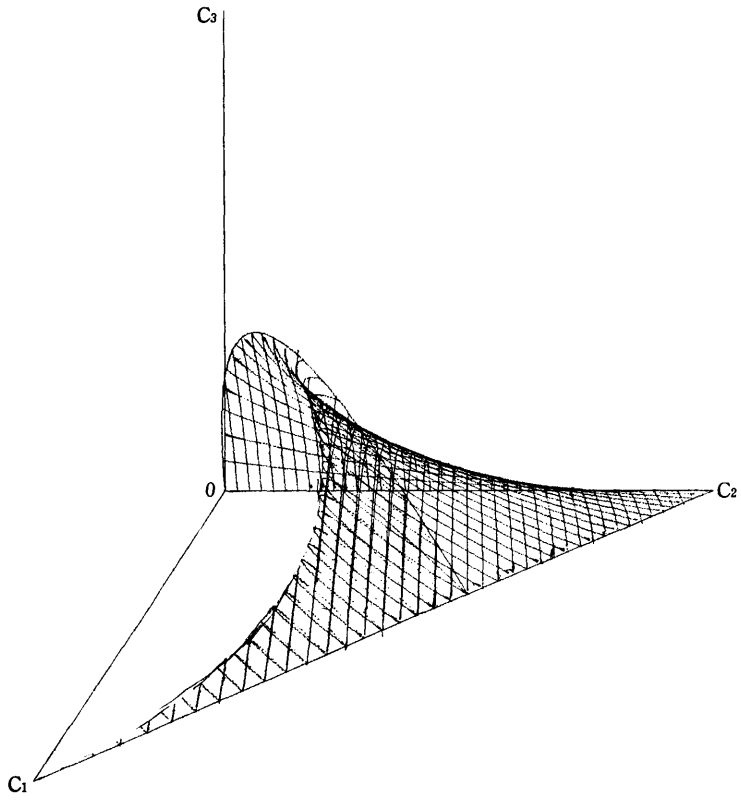

図 11

11 はその概略をあらわす.

(IND 1) (p_i がすべて異る).

$\hat{p}=\frac{1}{2}(p_1+p_2)$, $\Delta=|p_1-p_3|$ とおくと,

$C_1=\frac{1}{2}(1+p_2)\dot{p}$, $C_2=p_2(1-\dot{p})+\frac{1}{2}\dot{p}^2(1-p_2)-\frac{1}{8}\Delta^2(1-p_2)$

$C_3=\frac{3}{2}\dot{p}(1-\dot{p})(1-p_2)+\frac{3}{8}(1-p_2)\Delta^2$ となる.

当然 $\Delta=0$ のときは IND 2 に一致する. $p_1\neq p_3$ のいろいろの組合せがあるために, IND 2 の曲面が厚みをもつことになる. $C_1=$一定 の断面でみると, $\dot{p}=p_1$ とした場合の IND 2 の点に対し C_2 が $\frac{1}{8}\Delta^2(1-p_2)$ $(\geqq 0)$ だけ減少し, C_3 はその 3 倍増大する点が対応する. IND 2 の曲線の勾配は 3 をこすことがないので, Δ を一つきめたときの IND 1 の曲線は IND 2 の場合の上側 (C_3 の大きい方) にある. 従って領域の下の面は IND 2 の p_1 を $\dot{p}$ とした式であらわされる. Δ^2 は $p_1>p_3$ のとき $p_3=0$ $p_1=2\dot{p}$ のとき最大となる. ただし $\dot{p}>\frac{1}{2}$ では $C_1+C_2+C_3\leqq 1$ という条件から $\Delta=\min(1, 2(1-\dot{p}))$ であることがわかる. 結局領域の上の面 (C_3 が大きい方) は $\dot{p}<\frac{1}{2}$ では $C_1=\frac{1}{2}(1+p_2)\dot{p}$, $C_2=p_2(1-\dot{p})$, $C_3=\frac{3}{2}\dot{p}(1-p_2)\dot{p}\geqq\frac{1}{2}$ のところでは $C_1+C_2+C_3=1$ すなわち標準単体の斜面となる.

このようにしてみると独立の場合は全体の中で非常に小い部分しか占めないことがわかる.

モデル群 $\mathcal{A}$ の場合と比較することも興味があるが, 計算をしていないまゝである. この例は, 自由度があまりに小さいために, 数値の組合せの可能性の少ない中でいろいろの結果を包含していることになる. そのためモデルの性格を文章的に表現すると大きな差異となってくるものと思われる. いわゆる言葉での表現のちがいにすぎないことが多くあらわれる. たとえば, 積率でしらべてみても, モデルが含むパラメタ (確率) の個数だけの積率で分布がきまってしまうので, 何ら新たに知るものはない. 一般論の雛型になり得ないのはこういう点にある. マルコフ型のモデル群 $\mathcal{A}$ に関しても, モデルの文章的ほんやくの内容は異るにしてもこの事情はかわらない.

4. 背景の点描

この題材の背後にある分子生物学の研究を筆者がはじめて知ったのは, 昭和36～38年科研費総合研究「統計的モデル解析についての総合研究」を林知己夫所長 (当時第2研究部長) が主宰され, 筆者も参加し, 野田先生にも加わっていただいた時のことと思う. 分担課題を各自持ちよって説明をする会があった. そのとき野田先生が持って来られた課題は「蛋白質分子の会合状態の分布について」であった. これは分子生物学の一つの問題であって, 統計的モデル解析とどこでどのように関連するのか当時の筆者には見当もつかなかった. 総合研究の分担課題としては異質的なものの感じがした. しかし今となって考えると, この課題は統計的モデル解析の新しい手法の開発を促す契機を多分に含んでいたのである. 野田先生の慧眼には全く感服させられる.

余談であるが, この総合研究のことは今まで忘れていて, これを書くにあたって機縁をふりかえってみて思い出したのである. この総合研究は, でき上ったものをまとめたというような種類ではなく, いろいろの分野の人が集って発展の萌芽を作るようなものであったと思う. ここで述べる話はただ一例にすぎず, 筆者の知らないところで, 他の人達が, いろいろの問題について, いろいろの形で多大の成果を生んでいるものと思う. そしてその大部分は「統計的モ

デル解析」とは関係のなさそうなものであろう．主唱者の林先生は恐らく，「それでよい」とお考えのことと思う．実際こういうことを可能にしたのは，御世辞ではなく，林先生の包容力と組織力に負うものと考えられる．

ともあれ，綜合研究を契機として筆者は野田研究室によく出入りするようになった．そして知識よりはむしろ「科学の方法」「科学する心」を教わった．（口頭でなく実践を通じて．）ここでは筆者の接した範囲で，この研究の経過の大略をのべる．

生体内の高分子というのは，実はコラーゲンだったと思う．カンガルーの尻尾に多いということであった．これの会合についてのいろいろ面白い性質は当時どれほどわかっていたのか筆者は知らないが，とに角，野田先生がこれに目をつけられ，はっきりさせたいとして居られたことは事実である．方法としては粒度分布(分子が棒状に会合したものでありその長さの分布)をしらべることで，方法としては流動複屈折現象を用いることが考えられていた．

流動複屈折は長い分子が溶けている粘性流体を流して剪断応力をかけると，川の中の藻草のように流れの方向に分子の向きが揃えられ，そのため光学的には結晶と同じような性質をもつに到り，方解石のように複屈折をおこす現象のことである．流動複屈折は1930年代によく研究されたものでその頃の Zeitschrift für Physik にいくつか論文が載っている [4]．粒度分布を測るというのは，分子が揃うのは，揃わせないようにするブラウン運動に抗してのことなので，応力を大きくしないと小さい粒子までは揃って来ない．濃度と屈折率とは関係があるので，流速をかえては屈折率を測るということにより粒度の重量累積分布が測定される．

粒子統計の分布推定の見地からみた筆者の分類によると，このような測定法は，不完全分離間接測定に属する [2] [3]．原理的には測定可能であっても，実際には極度の注意と技輌を以てしても，いわゆる clean cut ができにくく，きわめて鈍い結果しか得られないものであった．実はさきに述べたマルコフ過程によるモデル解析を適用するには測定はあまりにも粗く，かつ誤差が大きい．そのため明澄な決論は得るには到らなかった．

そこで野田教授の発想によって回転電場による，回転拡散分布測定法の開発がすゝめられた．これはきわめて粗雑に言うと，液体を動かす代りに，粒子だけを電場によって動かそうというものである．棒状分子は液体中で分極して双極子となっているので，電場を回転すると，それに従って回転する．あまり速く回転すると，今度は大きい粒子はついてゆけない，回転を次第に速くすると，ついて回るのは次第に小さい方だけになる．

今度は，光学的に結晶様になった溶液の偏光面が外部電場の回転にともなって回転するのであるから，測定にあたってはセンサーの光軸を位相のみをずらせて電場にあわせて回転させることになる．

さて，原理はこのようなものであるが，実際に器械を製作し，さらに測定を行うには，定量的な理論が必要である．この測定理論をつくるには，電磁力学，電磁光学，分子光学，流体力学，分子運動論（拡散理論）等古典物理学のほとんどすべての領域の知識を従横に駆使し，しかも現象で無視できる項（効果）を見極めて定式化しなければならない．さらにただの理論でなく測定に使える式にするには，特別の解析学的工夫を要するものである．筆者の手に負えるものではないが，協同研究の関係上やらねばならないかと思って困っていたところ，天文学の出身で，当時野田研に居られた船越浩海博士（現東大教養学部）が見事完成した．彼のこの骨の折れる仕事は，統計数理研究所の Annals の supplement に発表されている [5]．測定装置自体はいろいろの事情でつくられなかったのではないかと思う．そのため船越氏の労作も理論だけに終り，しかも論文を統計数理研の方に出していただいたため，化学物理の専門家の目にとまらず埋もれてしまったとすると大変悲しいことである．

測定がすゝむにつれて，切れる場所が正確には等分点でなく，いろいろのところがあって，長さにさらにゆらぎを考える必要があるということになった．これに応じて筆者の分割モデルも修正しなければならなくなった．このような知見を与えるのに恐らく一役買ったと思われるのは，当時大学院生であった大内正俊氏のフーリエ解析による電子顕微鏡写真の画像処理である [6]．そこで用いられた手法は現在では全く巨大科学の一つとなってしまった画像処理の中心をなすものであるが，画像の鮮明化にディジタル計算機を用いようという考えは，1967～8年頃まだ珍らしく，とくに当時のコンピュータの性能では実用は無理のようであった．大内氏は統計数理研究所の当時の HIPAC 103 を用いて殆んど独力でこれをやってのけた．(あのコンピュータで出来たことは信じがたい再現されるべき画像が直線の帯状に並んだ分子縞模様であったために，可能となったものと思う．) 大内氏の論文は修士論文であり，しかも生物化学科に提出されているため，情報処理関係者の目にとまらなかったかも知れないが，今 Ernest L. Hall の本 [8] の文献をみると，コンピュータを利用した画像処理に関する論文は，彼の論文の翌年1970年になってからわっと出ている．「情報処理と統計数理」の中には大内氏の仕事と写真とともに簡単に紹介してある [7]．この本の初版は1970年だから当時としては割合いい線を行っていると思う．

この後分子生物学の爆発的な発展がなされたことは周知の通りである．分割モデルの改良も筆者の個人的事情で怠っているうちに，この種の問題は流行から外れていった．すなわち，蛋白質分子についてはもっとこまかいいろいろの性質がわかるようになったため，人々の関心はその方に移っていった．もち論マクロレベルの問題は，ミクロレベルのことがわかったからとて，それから演繹できるわけのものではないので問題が解決されたわけではないが，こうした研究をする人が少くなった．とくに電子顕微鏡の画像処理技術が進んだため，流動複屈折とか超遠心分離などの面倒な方法を使って粒度分布を測定し，それにより間接的にさぐりを入れるというようなことはすたれた．(電子顕微鏡で見る状態は乾燥された状態だから流動複屈折などの方法が全く不心要になったとはいえないが)，何れにせよ1970年代は生物化学にとって新しい忙しい時代であった．野田研究室からは長谷川政美博士を統計数理研究所に迎えることになり，新しい交流が始まった．そして直接の接触からは筆者は退いた．ミクロのレベルが異ると世界が全くちがうのでここで述べた問題と長谷川氏の生物化学における仕事とは全く断絶となる．進歩のはげしい科学の世界では過去を顧みるひまもないのであろう．誰も己の過去の仕事にこだわるような未練がましいことはしない．しかし筆者は分野がちがうので，やはり残念に思う．自分のお手伝いの仕事において，論文にしていない確率過程モデルによる解析法は前述の本に書いた．今回はさらにその残滓を引きずり出したわけである．これらは本当に走り書きのようなものであるから，わざわざ発表するほどのものでないが，船越氏の仕事，さらに遡(さかのぼ)って野田先生のアイディアが実際の方法として実現していないのは残念である．しかし思わぬところで復活する可能性は十分にある．それを期待したい．大内氏の研究における進取性はまた高く評価されるべきである．この方はしかし現在の日本の企業における情報処理技術の優秀性に直接むすびついているので，まずは慶祝すべきである．

あ と が き

統計の理論家を横からながめると，全局的な最適値でなく，局所的な最適値を熱心に追っているように見えることが時々ある．まじめな人こそこういう風になりやすい．これに対し，不まじめに“もっと広い視点に立ってみては？”というようなことをつい言いたくなるが，説得力をもつためには，一般論もどきのものを作らねばならないと思う．それで時々一般論をつく

ってみようという邪念をいだくことになる．しかし，反省してみれば，特殊な問題にぴたり役に立つ仕事も出来ないくせにどうして一般に役立つ仕事が出来ようかと思う．そういう風に考えると一般論を考えようなどという気持は霧散してしまう．

最近世界的な傾向として，統計数理をとりまく研究の雰囲気が急速によくなりつつあるように思える．これは生物科学の急速な進歩にともなって起っている多様な問題に統計学者が柔軟に対応しているからではないかと思う．昨年来相ついで統計数理研究所に来所した外人研究者の話からも何となくうかがえる．(たとえば [10].) 今や統計学者のイメージは変わりつつあるのではないかと思われる．統計学者はかつてのように他分野の下働きとして単にデータ解析をする専門家ではなく，データ解析をして科学的知見を増していく開拓的科学者そのものに変貌しつゝあるようである．われわれの周囲はすでにそういった空気であるが，筆者もじめじめした教訓的一般論にこだわっているひまがあれば，一つでも具体的な問題を解決しないと世界の趨勢にもおくれそうな気がする．

参 考 文 献

[1] 林知己夫，樋口伊佐夫，駒沢勉 (1970)．情報処理と統計数理 (初版)，産業図書，160-169.

[2] 樋口伊佐夫 (1964)．粒子統計における二，三の研究，統計数理研究所彙報，**12**，316-348.

[3] Higuti, I. (1976). Some remarks on the concepts of size-distribution in the study of small particles, *Contributions to Applied Statistics*, Birkhäuser Verlag, 63–68.

[4] Peterlin, A. (1938). *Zeit. Physik*, **111**, 232–363.

[5] Funakoshi, H. (1968). Theory of the birefringence caused by the orientation of particles in rotating electric field, *Ann. Inst. Statist. Math. Supplement*, **V**, 45–66.

[6] 大内正俊 (1969)．東京大学大学院修士論文.

[7] 上掲 [1] 214.

[8] Hall, E.L. (1979). *Computer Image Processing and Recognition*, Academic Press.

[9] 樋口伊佐夫，横田紀男，鈴木重夫 (1959)．統計における模計解析について，統計数理研究所彙報 **7**，81-92.

[10] Griffiths, D. and Sandland, R. Fitting generalized allometric models to multivariate growth data, (投稿中).

A Note on Some Sets of Models and Distribution for the Analysis of a Biochemical Phenomenon

Isao Higuti

(The Institute of Statistical Mathematics)

This note is an extract from an old memorandum scribbled by the author when he was engaged in a certain research collaborating with Professor H. Noda of Tokyo University in 1960's.

The main problem was to find the mode of splitting of a sort of associated protein molecules caused by the application of supersonic waves and as a source of information the experimental data of change of size-distribution of perticles (multiplicity of association) was to be used. This was intended by providing a set of stochastic models which is of Markov type. The detail of the method of analysis based on this set of models has already been mentioned in [1].

There is, however, possibility of using anther sets of models. As one of these sets we might use the set of those models which only assume the probabilities of splitting at respective nods. For the purpose of the research it seems very natural that we prefer the set of Markovian process models to the latter. The motivation of thinking of such different categories of model is that we might use these as a prototype of more general problem: Why a special category of models is preferable to the other for a certain definite scientific research ? As easily seen, this example is too simple to contain essential characters for general problems. So the aother has abandoned to develope such motifs concieved.

Since it seems to be interesting to compare these sets to each other, an explanation of latter set which has never been mentioned yet is gives in this note.

Also a brief sketch of the contact of the author with the bio-chemical problems is added. This would show itself as an example of real activities of collaboration between the field of statistical mathematics and other rapidly developing fields of science.