統計の基礎に関する Fisher の立場と Neyman の立場

石　田　正　次
植　松　俊　夫
崎　野　滋　樹

(1959 年 5 月 受 付)

On the Difference of Statistical Methodology between R.A. Fisher and J. Neyman

Masatugu Isida, Tosio Uematu and Sigeki Sakino

In studying the theory of statistics, it is of fundamental importance that we settle our standpoint as to the statistical methodology. It remains also true when we intend to apply the results of the study in practical problems. For the correct use of statistical method is possible only when we understand the true meaning and the limitation of the method in the light of our standpoints. To fulfil such a fundamental requirement, however, is not a very easy problem, and it needs much consideration about various respects.

In this situation it will be very helpful and suggestive to appreciate, without any prejudice, various thoughts and methodologies, upon which the existing theories in mathematical statistics have been developed.

This situation stimulated us to study the methodological difference between R. A. Fisher and J. Neyman, two representative contributors to modern mathematical statistics. Such a study is the main concern of the present paper.

In Section 2 we have a brief survey of Fisher's thought in connection with the concept of likelihood, the fiducial argument, the test of significance, and so on. In the same way we treat of Neyman in Section 3 mainly referring to the development of his theory of testing hypotheses. Then in Section 4 and 5 we discuss about the methodological difference between them by analyzing some papers in which they argued with each other about the mathematical foundation of statistics: in Section 4 about the inductive inference and inductive behavior which are the logical fundamentals of their theories respectively, and in Section 5 about the concepts of population and the mental processes of inference.

Especially we compare the confidence limits with the fiducial limits and discuss whether we should accept the notion of error of the second kind.

Institute of Statistical Mathematics

§1. ま え が き

統計学の研究に当つて，与えられた問題をどのような立場，考え方で取扱かつて行くべきかということは最も基本的なことである．このことを充分に考究しておくことは，統計学の研究の結果が正しく応用されるために極めて重要である．われわれはその一つの指針として，ここに R.A. Fisher と J. Neyman の間に取り交された論争を中心に両者の互に対立する立場を取上げ，その統計学に対する基本的考え方の相違を明らかにし，かつ対立の根本が奈辺にあるかについて論じたいと思う．

ここでわれわれの態度は，Fisher の統計学と Neyman の統計学を何れも対等に見，かつ公平に取上げていこうとするものである．統計学に関して一部に信じられている偏見——K. Pearson を継承して発展せしめられた Fisher の統計学にまつわる曖昧さ，非論理性が Neyman によつて正され，合理的かつ妥当な姿をとるに到つたという考え方——は統計学の本質から遊離した余りにも皮相な見方といわねばならない．

W.S. Gosset, K. Pearson などに始まつた近代統計学は，Fisher 及び Neyman の両者に承け継がれたが，両者の統計学に対する基本的態度の相違から，対立する二つの方向に分岐したとみるのが至当であろう．この二つの方向の何れが妥当であるかということは，理論の数学的精緻さ，応用面の広さといつたようなことだけから単純に律することはできない．そこには統計学の基礎に対する根本的な問題が潜んでいるのである．

この報告の目的とする所は，先ずわれわれの頭を白紙に戻し，虚心坦懐に Fisher 及び Neyman のそれぞれの主張を聴くことにより，その真意を理解し把握しようとすることである．勿論われわれは更に進んで，両者の主張を批判し，その取るべきは取り，捨つるべきは捨て，かくしてわれわれ独自の立場の決定に資することが肝要であるが，この報告はそのための一つの足場を与える事を目指しているものと了解されることを希望するものである．

§2. R.A. Fisher の統計学における立場

Fisher は統計学が自然科学の帰納的方法論であることを強調する．彼の中心問題は統計学の体系を数学的に完璧なものにすることではなく，科学の理論構成に当つてその推論の流れを，統計的手段により常に正しい方向に保つことにある．

それならば Fisher のいう統計学とは如何なるものであろうか．次に彼の考えを追つてみよう。

ある現象はある法則によつて律せられている．この法則は場所的にもまた時間的にも不変なものである．しかしその法則を直接に把握することはできない．それでわれわれはこの法則を見出すために，過去に得られた知識を基にして科学的仮説をたてる．この仮説のもとに実験，調査が企画され，データーが得られる．ここで実験は自然の姿をよく代表するように設計されなければならず，そこから得られるデーターは法則に関する典型的な標本でなくてはならない．データーは法則によつて律せられる部分と場所的また時間的その他の条件によつて変化する部分（つまり誤差）を含む．法則の把握のためには法則によつて律せられる部分の情報のみを，他の部分（つまり誤差と考えられる部分）から結晶化させ，これを抽出しなければならない．その為にわれわれはそのデーターが取られたであろうもとの姿を想定する．これを Fisher は仮説的無限母集団（infinite hypothetical population）と名づけた．そして，すべてのデーターはこの集団からのサンプルとみなされるのである．仮説的無限母集団はデーターから法則を導くための便宜的手段として設定されるものであり，各組のデーターに対して唯一つに定まるものでなければならない．また仮説的無限母集団はできるだけ少数のパラメーターによつて決定される必要がある．われわれの得た数多くの実験データーからこれらのパラメーターが推定できれば，データーに含まれる法則に関する有用な情報は，こ

の母集団の性質の中に要約されることになる．そして始めにたてた科学的仮説の真偽は，この母集団との対比によつて問われ，新な法則が導かれる．

データの要約のために仮説的無限母集団のパラメーターを推定するに当り，その推定値としてデータから計算すべき統計量に関しては，**Fisher** は **Student** を踏襲して **consistent, efficient, sufficient** であることを要求する．然し **unbiased** であることは積極的には望まない．これは **Fisher** の立場を決定する重要な点である．

次にこれらについて **Fisher** の与えた語句の定義を述べる．

Consistency ——統計量を母集団全体から計算した場合，その値がねらう母集団のパラメーターに一致するとき，この統計量は **consistency** の基準を満すという．（確率収斂ではない点に注意）

Efficiency——**efficiency** の基準とは標本数を大きくした場合，その統計量の分布が可能なかぎりで最少の標準偏差をもつ正規分布に近づくことである．

Sufficiency —— 同一標本から計算できる他のいかなる統計量も，求める母集団のパラメーターに関してある統計量以上に情報をもち得ない場合，その統計量は **sufficiency** の基準を満すという．

何故に **Fisher** はこのような条件を満す統計量を要求するのであろうか．それには二つの理由がある．その一つは彼の科学実験に対する基本的な態度であり，他の一つは **fiducial argument,** 検定論などにつながる彼の確率論の立場である．

Fisher の場合，同一条件のもとでの実験は唯一回限り可能であり，その操り返えしは存在しない．同じ袋の種子を同じ畑に何度かまいたとしても，そのたびに天候は相違し，種子，畑は時間的に条件変化をうける．また同じ袋の種子をいくつかの畑にまいたとすれば畑の条件が同一ではない．このような唯一回限りの実験（**unique sample**）について，一つの仮説的無限母集団が想定され，そこから一つの結論を摑まねばならない．この考えは彼の農業実験に対する根本的思想である．故に **Fisher** は，唯一回の実験データーから計算される統計量は目ざす母集団パラメーターに絶対量として少しでも近い（標準偏差最少＝**efficient**）ことを願い，しかもこの統計量の中にできるだけ多くの情報を盛り込もうとする（**sufficient**）．更に標本数を無限に大きくしてそれが全母集団と一致したとき，統計量はパラメーター自身に一致するという要求を付加するのである．一方同一実験の繰り返えしを否定するために **unbiased** なる概念は意味を持たない．以上が **Fisher** の実験と，そのデーターに関する考えの概略である．

consistency, efficiency, sufficiency の条件を満す統計量を得る手段として **Fisher** は，**K. Pearson** が起用したことのある **Helmert** の尤度の考えを選んだ．繰り返えしの存在しない **Fisher** の世界では，確率よりも唯一回の試行に意味ずけのできる尤度の方がより好ましいものである．つまりある標本（データ）が得られた場合，そのような標本が得られる尤度が最大になるような母集団のパラメーターの値を標本から計算するのである．この尤度を最大にするという方法が **Fisher** の欲する統計量を提供するかどうかは，はじめ彼自身も客観的に見通すことができなかつた．しかし彼はこの方法に確信と希望を持ちつづけ，終にこれがその有力な方法であることを明らかにしたわけである．これで母集団のパラメーターの推定の問題を一応解決した後，次に彼はこの推定値に対する保証の問題に取組んでいる．

母集団のパラメーターが与えられたとき，その母集団からの標本によつて計算された θ の推定量の分布が求められれば，これを手懸りとして，この推定量を用いた θ の推定の信頼性の評価を行わなければならない．**Fisher** は **Bayes** の定理的な考え（**inverse probability**）に非常な反感をもつており．これを用いずに推定量による θ の推定にある保証を与えようとする試みを **1930** 年に発表した．彼にとつては母集団のパラメーターの分布，つまり事前確率を仮定することは，母集団自身を既知とすることと同等であつて，このような場合においては推定ということが無意味となるの

である．そこで Fisher は fiducial argument と名づける概念を導入し，彼独自の統計的立場を確立した．彼の方法は次のようである．パラメーター θ の推定量として θ の充足統計量 T をとる。θ の任意の可能な値に対し，その θ の値の場合の T の分布函数が定まるが，今これを $F(T,\theta)$ とおく．Fisher は，$F(T,\theta)$ において T を固定して θ を動かすことにより，T を観測した場合において θ に関する確からしさを与える θ の分布函数が得られると考えるのである．その場合には T がパラメーターで θ が確率変数とみなされる．づまり $Pr\{\theta\geq\theta_0|T_0\}$ と $Pr\{T\leq T_0|\theta_0\}$ が同時的に成立する場が存在すると主張するのであつて，この前者を fiducial probability と名づける．更にいい直すならば，$F(T,\theta)$ が連続と考えて

$$\frac{\partial F}{\partial T}dT$$

が T の密度函数を表わすのに対し

$$-\frac{\partial F}{\partial \theta}d\theta$$

が θ の fiducial distribution を与えると考える訳である．これを用いて任意の有意水準 α における θ の fiducial limits を計算することになる．その操作は次の通りである．

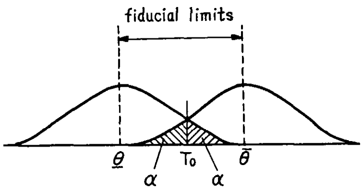
標本から統計量 T の標本値 T_0 を計算し，$F(T,\theta)$ を用いて，

$$Pr\{T\geq T_0|\underline{\theta}\}=\alpha$$

なるような $\underline{\theta}$ を求め，同様にして

$$Pr\{T\leq T_0|\overline{\theta}\}=\alpha$$

なるような $\overline{\theta}$ を計算すれば，この $\overline{\theta}$ と $\underline{\theta}$ は有意水準 α における θ の fiducial limits である．この fiducial limits の確率論的解釈は，後述する confidence limits と異り，直接的には相対頻度的なものによつて与えることはできない．limits をつくるときの一つの目標（信念）として確率が用いられているに過ぎないのである．有意水準 α も，確率論的には今述べたような意味の確率，つまり fiducial probability に関した数値である．



第　1　図

この方法においては T が充足統計量であることが要求されているが，このことは fiducial probability の考え方において特に重要なことである．

確信の測度として fiducial probability は一意にきまることが当然要求されるが，Fisher はこの一意性を sufficiency の概念に基ずいて導こうとする訳である．可能な限りにおいて最大限の information を盛りこんで得られた確信でなければ妥当性がないと彼は考えるのである．

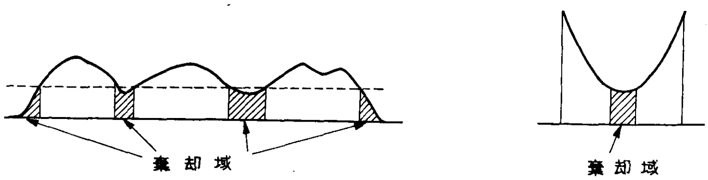
Inverse probability に対抗して作られた fiducial argument と呼ばれるもの，それに含まれる一連の概念，fiducial distribution, fiducial probability, fiducial limits 等は，今までの確率論における考え方とはかなり異なつたものであるために，多くの論争のもとになつたことは周知の通りである．つまり，これ等の概念は頻度的意味乃至は繰り返えし的な意味で把握されるのではなく，結論に対する直接的な信念の測度を与えるためのものとして受け取らなければならないのである．前にも述べたように Fisher は唯一回の実験の結論にその意味を持たせなければならないのであつて，数多くの実験を通してみたとき正しい結論がどのような割合で得られるかという観点から推論の方法を評価することは意味がないという．

以上母集団パラメーターの推定に対する Fisher の考え方を述べたが，次に統計的仮説検定の問題に関して述べよう．

Fisher の場合，仮説は予想される結論を否定するようにたてられ，検定は常に棄却のみを考え

る. データーが棄却域外に落ちた場合は仮説の真理からの隔りが直接仮説を棄却するほどは大きくはない. しかしそのデーターだけからは仮説の正しさを認めるわけにはいかないとして結論をさしひかえる.

棄却域は尤度が小さいところを選ぶ. これも, 繰り返えしを認めない唯一回限りの実験でものをいおうとするには, 確率よりは尤度の方が妥当な測度であるという Fisher の基本的考え方の現れである. 尤度だけで棄却域を定めようとすると下図のような場合, 常識とくいちがつて話に無理が生ずる.

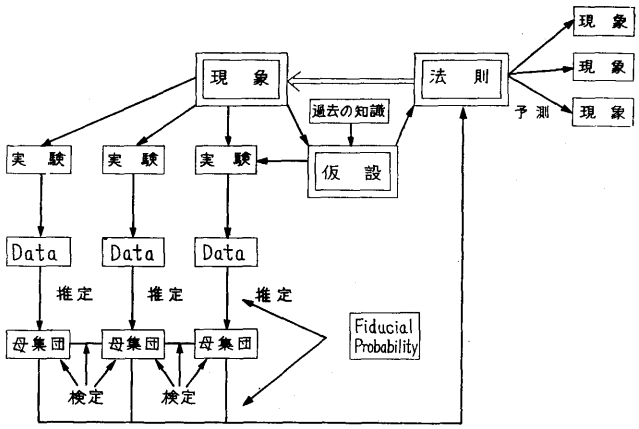


第 2 図

そこで統計量の分布が標本数が大きい場合正規分布に近づくという efficiency の性質が大切な意味をもつことになる. つまり分布が鐘型であれば棄却域は必然的にその裾の方にくるわけである.

Fisher は普遍妥当性のある原理の追求のみを目的とする. 従つてすべての客観性のあるデーターは何ものかからの標本である (その何ものかの中に目ざす真理が潜んでいる) と考える. 故に例えば, Neyman 的に考えれば全数検査とみられる場合であつても, 有意差検定を行うことが意味をもつのである. また科学実験においては通常対立仮説というものを考え得ないために第二種の誤差は Fisher の理論では考慮されていない (更に詳しくは §5 参照).

以上が Fisher が描く統計的論法つまり彼の inductive infernce である.



第 3 図

§ 3. J. Neyman の思想とその発展

Neyman の思想は, 彼の最も代表的業績のあげられた, 仮説検定論の分野においてよく見ることができる. われわれは主としてこの分野における彼の理論の発展を追い, 彼の主な考え方をたどつてみよう.

仮説検定論に対する彼の考え方が初めて明確に提示されたのは 1928 年である (論文 [19] 参照). この論文を読んでいくと, 彼がこの論文を書くに到つた動機となつたものは, 従来の検定論に付随する任意性という点にあつたように思われる.

先ず論文の冒頭において彼は次のように述べている:「ある仮説をたてた場合，この仮説を採択するかどうかは，取上げている問題に関して種々の考慮をなすことによりきめねばならない．この仮説に対してある検定をしてみて得られた結果も一つの手掛りであるが，それだけを考慮して結論を得ることはできない．この場合色々の考慮の結果を一つの数値に総合して，この尺度によつて判断するというようなことは不可能である．故に判断は結局個々の研究者に俟つ他はない．この場合検定ということは一つの道具であるから，人によつて，“よい”と考える検定の方法は異うかも知れない．客観的に最も良い方法はあり得ない．しかし検定が有用な道具として用い得るためには，種々の検定の方法*自身の意味及び限界をはつきりさせておく必要がある。」

このような考え方は後年の彼の考え方とはまるで違つているようである．しかし更に読み進んでいくと，彼がこの論文において既に，上述のような考え方に留まつていないことがわかる．即ち彼は例を用いて種々の検定の方法の意味を考えていつた後，検定法自体の好し悪し，更にできるだけ“良い”検定法を追求していくという方向に進んでいる．この論文においては，まだ，一定の基本原理の下に仮説検定論それ自体の展開によつて最良の検定法の決定を論ずることはなされていないが，検定に関する彼の考察の中には既に，彼が後年展開した仮説検定論において基本的役割を果すこととなつた諸概念——対立仮説，第1種及び第2種の過誤——が提示されている．

この論文で彼は検定における二つの行き方を述べている．第一は仮説の母集団から発して，実測されたサンプルがこの母集団からのものである確率を求めるという行き方で，在来の行き方はこれに含まれる．第二の行き方は実測されたサンプルから出発して，仮説の母集団が真の母集団である確率を求める行き方である．在来の検定論では専ら仮説にとつた母集団のみに注目している．所で上のどちらの行き方をとるにしろ，母集団について与えた仮説に対立する仮説が同時に考慮されていなくては，無意味であることを彼は強調している．この点の更に明決な所論は彼の 1937 年の書に見ることができる（[25] の p. 33-44 参照）．また第一の行き方においても，inverse probability についての何等かの考え方が含まれていなくては，議論が進められないと述べている．彼としては第二の行き方の方がより論理的であると考えているが，この点で彼は Fisher の考え方とは全く対立する．Fisher は Bayes 流の行き方，a priori な確率を考えることに全く反対している（§ 2 参照）．無論 Neyman は inverse probability の考え方にあらわれる a priori な確率に関する難点はよく承知しているが，それに留意した上で inverse method を発展させんとする試み**をこの論文の一部分で行つている．

この論文で最も興味あるのは，彼が既に第1種及び第2種の過誤の概念を提示していることである．彼の議論はすべてこの基本概念に基ずいて展開されている．ただし彼は第2種の過誤を確率によらず尤度的な思考によつて制御する方法をとつている．即ち，仮説の真なることについての確からしさを尤度によつて表現するのである．ここで対立仮説という考え方が重要なものである．今例えばサンプルから統計量 $\bar{x}, S$ を作つて点 $(\bar{x}, S)$ の空間内に棄却域を取る場合を考え，また種々の棄却域の境界がある曲線族をなす場合を考える．この時第1種の過誤はこの曲線族の充分外の方にある曲線をとりその外側を棄却域に用いることによつて，与えられた確率以内に押え得る．ただし第2種の過誤はその制御が難かしい．彼は次のように考える．実測したサンプル値から作つた $(\bar{x}, S)$ の値が内部の曲線上にある程，真の母集団のパラメーターの値（または，それに近いパラメ

* 与へられる仮説の検定の為に考えられる色々の統計量のとり方及び色々の棄却域のとり方に対応して，種々の検定の方法がきまるわけである．

** a priori な確率そのものを用いることは避けている．彼はこの後に述べる尤度比検定法に於ける考え方と，或意味で同等な別の考え方を用いて得られる検定法を導いて，これを a priori な確率の分布を考えた場合の a posteriori な確率の場と比較する．この時，彼の導いた検定法は，式の形の上では a priori な確率を一定として得られる a posteriori な確率の場で考えることと同じことになる．この意味でこの検定法は一つの inverse method であると彼は考えている．このような意味づけは少々こじつけのような気がする．

ーターの値）を持つ母集団の仮説の尤度が高く，外側の曲線上にある程，真の母集団のパラメーターの値に遠いパラメーターの値を指定する対立仮説の尤度が高い――というように $(\bar{x}, S)$ 空間の曲線族を取ることが若しできれば，それによつて第2種の過誤の制御も考慮することができる．この考え方から出発して彼は，更に一般の場合に，与えられた仮説の尤度と対立仮説の尤度の比較によつて，第2種の過誤の制御をも考慮に入れた検定の方式として，尤度比検定法*を提示した．

論文 [20] において彼の検定論におけるる考え方は明確な形をとる．彼の検定論に含まれる数学的内容はこの論文において初めて明確な形に **formulate** された．この場合われわれは，彼が検定ということを行動 (**behavior**) のための一つの合理的手段であると考える立場にあることを明らかに窺い知ることができる．彼は犯罪容疑者の処罰の問題を例として，第1種及び第2種の過誤の考慮の必然性とそれに対する統計家としての立場を述べている．この例の場合には有罪か無罪かが必ず決定されねばならないが，そうすれば第1種及び第2種の過誤が互に相容れないものとして現われる．われわれが誤りを犯す可能性を考慮しつつ，而も有罪か無罪のある場合には極めて悲劇的な結果を生じ得る何れかに決定することは，社会的及び人道的考慮の何れがより重んじられるかによる．このような考慮は数学の関知する所ではない．数学の役割はこの場合，誤りが存在しながらもそれをできるだけ小さくするための一つの合理的手段として考えられねばならない．合理性だけで割切れない部分は数学が解決し得るものではない．数学は合理性の上からはこうすればよいという方法を与えるためのものである．かくして彼は検定論に関して第1種及び第2種の過誤を基とした考え方を明確に **formulate** する．即ち第1種の過誤の確率を与えておいて，第2種の過誤の確率を最小にするような棄却域（最良棄却域）を用いる検定の方式を最良のものとして提出する．前の論文 [19] に於ける所論はここに到つて更に発展ぜしめられているのである．前の論文で提出された尤度比検定法も，新らしい観点からは，やや曖昧な規準から導かれたものとして更にその妥当性の検討を要する訳である．極端にいえば，尤度比検定法も無数に考え得る検定方式のうちの一つに過ぎないので，その妥当性は更に，上に導入された新らしい原理に徴して究めなくてはならないのである．この論文において彼は最良棄却域に関する一般的研究を行つている．就中，単純仮説でかつ対立仮説も単純な場合，または対立仮説が複合の場合でもすべての対立仮説に対し共通の最良棄却域が存在する場合には，尤度比検定法の棄却域が最良棄却域であることを示し，それにより尤度比検定法の妥当性の根拠を与えている．彼は更に与えられた仮説が複合仮説の場合にも一般的議論を行い，ある条件の下に任意に一つとつた対立仮説に関する最良棄却域が存在すること，また更にある条件を課する時，多くの対立仮説に対し共通の最良棄却域の存在することを示している．しかし複合仮説の場合の上述の諸条件は可成り厳しいものであつて，その適用外にある実用上の問題も少くないという点での難点が存在する．例えば考えるべき棄却域は，与えた複合仮説に属するどの仮説が真の場合でも，棄却域にサンプルが落ちる確率が与えられた第1種の過誤の確率に等しいようなものでなくてはならない．即ちいわゆる **similar region** を考えていかねばならない．然るに応用上重要な幾つかの場合に **similar region** が存在しないか，または存在しても実用にならないということが起り得る（論文 [8] 参照)．この場合 **similar region** の考え方をせずに，母集団に関し **a priori** な分布を考えることにより難点を克服することは **Neyman** は行つていない**．

第1種及び第2種の過誤の考慮を簡明に表現するために彼は検定の **power function** なる概念を導入した．単純仮説かつ複合対立仮説の場合の上の彼の結果は **power function** に関していえば，若し許容される対立仮説のすべてに対する一様に **most powerful** な検定法があれば，それは

---

* admissible な仮説を Ω，また test したい仮説（Ωの或る部分集合）を ω とする時，
$\lambda=\dfrac{P(\omega_{\max})}{P(\Omega_{\max})}$ の小さい値を有意とみる検定法である．

** Neyman は a priori な分布を考えることに消極的のようである．然し彼が a priori の分布を考えることに否定的である訳ではない（論文 [31] 参照)．積極的に a priori な分布を考えていくことは von Mieses あたりが始めたようである（例へば論文 [17]，[18] 参照)．

尤度比検定法で与えられるということになる．所が応用上の重要な場合において，このような一様に most powerful な検定法が存在しない場合が数多くある．ここにおいて彼は最良の検定法の意味を，このような場合にどのように与えればよいかを研究した．この結果が論文 [21], [26], [27] である．これ等の論文において彼は不偏検定なる考え方を導入し，棄却域のとり方にある合理的な制限を与えることを考えた．こうして不偏検定の範囲で考えた場合に一様に most pomerful な検定法が存在するか否かを研究するために，A 型及び A_1 型の不偏な棄却域の存在のための条件を論じている．ここで注意すべきは，一様に most powerful な検定法がない場合，一様に most powerful な不偏検定法を用いることが妥当であると彼は絶対的な意味で考えているのではないことである．これに関連して彼は次のように述べている．「統計を応用しようとする場合のやり方は，先ず現象に適したモデルを構成し，次にそれに基ずいて数学的議論を行つて結論を出すという形をとる．したがつて検定の応用においても二つの段階がある．数学理論によつて構成される段階においては何人にも異論の起り得る余地はない．しかし現象の formulation の段階では個人の主観が入り得る余地が大いにある．一様に most powerful な不偏検定を採用するかどうかという場合にも事情は同様である．」と．(勿論この検定法には相当の根拠が与えられており，彼自身はこれをとりたいと思つている)．この所見からも，彼がその理論の展開において数学的完璧性に極めて留意しているにも拘らず，やはり統計家としての立場を守つていることが知られる．ただ注意すべきことは，上述の彼の言において，種々の現象を統計的な乃至確率論的な場として把握することができさえすれば，それ以後は数学的理論の展開だけで充分な結論が出せるという考え方が強く出ているように思われる点である．この点は Fisher の考え方とはつきり対立する点である．

彼の理論は以後の統計学において主流をなすもので，以後のほとんどの研究者が同じ立場をとつて更に発展させていくこととなつた．その代表的人物は A. Wald であつた．Wald は decision の理論によつて Neyman の理論を更に発展させている．

以上 Neyman の仮説検定論の発展について概観したが，次にわれわれは彼の理論構成（それは現代のほとんどすべての統計家についても同じである）の基礎となつている確率及びその利用の場について触れてみよう．任意事象に対し彼の考える確率は，同一条件下において実験を繰り返えした時，この事象の起る相対頻度が実験回数を無限にふやした時近づく極限値という意味で実際面と結びつけられるものである．故に例えば母集団に対する仮説 H が真の時にこれを棄却する確率が 0.01 であるとは，H が真の時，母集団からサンプルをとつて検定を行うという操作を若しも極めて多数回例えば一万回行えば，そのうち約百回の場合に H が棄却され，残りの場合に H が棄却されないということを意味する，即ち唯一の不変の母集団からの無限回のサンプリングということを仮想する．この点で Fisher とは真向から対立する (§4 参照)．両者の間の論争のうち最も花々しいもの即ち confidence limits の考え方と fiducial limits の考え方をめぐる論争も，母集団に対する上述のごとき考え方の相違の一つの必然的帰結とも考えられる (§ 4 参照)．

Neyman の母集団に対するこのような考え方は confidence limits の考え方に端的に示されている．サンプル x の函数 $\overline{\theta}(x), \underline{\theta}(x)$ を用い信頼区間により母集団のパラメーター θ の幅づけを $\underline{\theta}(x)<\theta<\overline{\theta}(x)$ のごとく行う場合，信頼度が α 即ち $P\{\underline{\theta}(x)<\theta<\overline{\theta}(x)\}=\alpha$ とは次のごとき意味である．サンプル x から $\underline{\theta}(x), \overline{\theta}(x)$ を作れば母集団のパラメーターの真の値は区間 $(\underline{\theta}(x),\overline{\theta}(x))$ に入るか入らぬかどちらかである．故にこれだけでは上のごとき α なる確率を「θ が区間 $(\underline{\theta}(x), \overline{\theta}(x))$ に含まれる」という事柄に付与することはできない．これが実際に確率 α を持つという意味は，母集団からのサンプルの極めて多数回の繰り返えしの抽出を考えて始めて与えられる．即ちこの多数回の繰り返えしにおいて，「θ が区間 $(\underline{\theta}(x),\overline{\theta}(x))$ に含まれる」ということが実際に起る割合が α になるという意味で与えられるのである．これは θ の起り方の確率ではなくて，θ について上のような幅づけをする場合，それが成功する確率に他ならない*.

Neyman の場合の確率や母集団の考え方は上述のごとく極めて明確である．しかしながらそれは数学上の概念として明確に定義されているという意味においてである．実際問題に統計を応用する場合にも Neyman の考え方が常に正しい考え方だとは必ずしもいえないのではないであろうか．これについてはまた後に議論する．勿論 Neyman の考え方を用いなくてはならない応用分野も広いことをわれわれは認識している．

以上の 2 節で Fisher と Neyman について個別に述べたが，この後の 2 節においてわれわれは両者の論争を中心にして，その根本的相違を明らかにして行く積りである．

§ 4. Inductive inference と Inductive behavior

統計学を応用する際における「推論の論理的基礎」に関して既に，Fisher と Neyman の間にはその立場の上ではなはだしい相違がみられる．それは inductive inference と inductive behavior の二つの立場の相違である．この差違は二人の間の論争にあつて常にその基礎をなすものである．それ故に初めにそれについて述べるのが適切であろう．(以下のことについては論文 [16] 及び [31] を参照)．

論理的観点からみて，統計学を応用する場合の方法に関して二人の間に共通する点がない訳ではない．二人共データーからの帰納ということを基本的であると考えている．それは統計学を用いる上からは当然のことである．ただデーターからの帰納ということの意味に関して両者の考え方ははつきりと対立するのである．

Fisher においてはデーターからの帰納という意味は自然科学においてよく知られている帰納法そのものである (彼は inductive inference という言葉を用いる)．彼は自然科学の領域において真理を探求し，自然についての知識をより進歩させて行こうとすることにのみ関心をもち，そのための方法として inductive inferenc を用いていく訳である．彼は別の領域即ち工業や商業の分野にも統計学が活用されて，色々の有用な結果が得られるということ，またその場合彼と違つた思考の論理が用いられることには全然賛成である．彼が反対するのは，このような思考論理が例えば test of significance の場合にまで持ちこまれることである．Neyman の仮説検定論から初まつて Wald に到つて明瞭な形に formulate させた decision の考え方は，Fisher の関係する学問の分野にまで持ちこまれてはならないものであるという訳である．Fisher によれば decision の考

---

* θ の巾づけに対するこのような方法と対立するものは，Bayes の公式を用いる古典的方法である．この場合には θ についての a priori な分布が必要であるが，この様な a priori な分布を用いる事を避け得る方法として Neyman は confidence limits の方法を考え出した訳である．必然の結果として，θ の巾づけに対する確率の意味が違つて来る．confidence limits の場合は本文に述べられている通り，$\underline{\theta}(x) < \theta < \bar{\theta}(x)$ なる巾づけが成功する確率であるが，Bayes の公式を用いる方法の場合には，x というサンプルが得られた条件の下で，θ の起り方の確率，θ が前もつて与えられた或る巾に入る確率が考えられるわけである．Bayes の公式による古典的方法は von Mieses により復活され，その方向の研究を生ぜしめた (論文 [17], [35] 参照)．この方法では，a priori な分布を消去する点が難かしいが，その為には集積された過去の観測値をうまく利用する訳である．von Mieses の confidence limits に対する批判は，confidence limits の場合には x のすべての起り方に関して巾づけの一つのシステムを考えるという点である．所が真に知りたいことは，x が例えば x_0 という値をとつたと知つた時，θ についてどれだけのことが云へるかということである．この場合 x が他の色々の値をとるかも知れないこと，その場合に起り得ることは問題に関係がないことである．

統計的推論に関する von Mieses のこのような考え方は Fisher の考へ方と一致している．但し用いる方法は両者は全く違う．例えば母集団のパラメーター推定の問題では，von Mieses は前述の如く，Bayes の公式を基とするが，Fisher はあく迄も a priori な分布を用いることを認めない．その場合 Fisher は fiducial probability を用いる訳であ．即ち両者が推定の信頼度を表現する為に用いる確率が全然異る．von Mieses の場合の θ の起り方の確率 (x を固定した条件下における確率) は，相対頻度の極限として意味づけられるが，Fishcr の場合はそうではない．Fisher の θ に対する fiducial probability は，von Mieses の意味の確率とは概念的に全然異つた信頼度の測度である．

え方は acceptance procedure であつて，その行き方はできるだけうまく acceptance procedure を企画することにより，人間の行動においてできるだけ得をしよう（あるいはできるだけ損をすまい）とするものである．それに反し自然科学の研究は損得の問題ではない． Fisher は次のようにいう.「自然科学の研究は人間の知識を進歩させることを目的とするものである．知識を進歩させることによつて，人間や社会が色々の目的を遂行することをより容易にさせるというのがこの場合の意図である．この様な行き方をした時，上述のごとき意図がどれだけの成果をあげるかを評価することはやらないしまたできもしない．結果として得られる知識がどのような意図で用いられるかということに関係なく，万人に納得のいく推論の方法をとるのが自分の行き方である．」

これが Fisher の意見である．

Fisher と異つて Neyman の場合のデーターからの帰納は，彼のいう inductive behavior の意味においてである．この場合彼もまたその出発点を自然科学の研究に置いている．自然科学の研究における論理として，帰納法はその内容が正当に理解されていないと彼は考える．

彼のいう所にしたがえば，先ず科学の理論は自然現象の種々のモデルである．各々のモデルは，現象研究の適切な要素として案出された entities に関して考え出された幾つかの仮説の集りである．われわれはモデルを構成する仮説の結論が，観測されたデーターに一致するものであることを期待する．もしデーターから検討したあげく充分な程度の一致がみられると考えられれば，われわれはモデルが適切なものであると考えるし，そうでなければモデルは適切なものでないと考える．モデルによつては適切なものもそうでないものもあり得る訳であるが，あまり適切なものでなくても，限られた目的のためには役に立つこともあり得る．

モデルには deterministic なものと indeterministic なものとがある．後者に関して現象を研究するのが数学的武器としての統計学であつて，その基礎にあるのは確率論である．その場合確率は，経験的に観測できる相対頻度を理想化したものである．（このような意味の確率論のきちんとした基礎づけは von Mieses によつてなされている）．

以上のような意味で考えられた科学の理論において，仮説に対して何等かの結論を下す場合に用いられる方法が帰納法である．この場合の思考過程は次のようにまとめることができる．

(i)　研究せんとする現象に関する仮説の組の幾つかの可能なものを具体的に表現する．

(ii)　これ等の仮説の組の各々から演繹を行う．

(iii)　データーを得てそれに基づき，上述の仮説の幾つかの組の各々に対しとるべき態度あるいは行動を決定する．これは act of will である．

上記三つの過程のどこにも普通の帰納法にいうごとき帰納的な推理というものはみられない．(i) は推理と全然関係ないし，(ii) は演釈的推理である．(iii) も推理では全然なく，これは decision であり，act of will である．

このように考えて Neyman は inductive reasoning という語で表わされている帰納法を，その内容からみて inductive behavior という方が適切であると主張する．

彼が inductive reasoning を標榜する統計学者のとる態度に反対するのは，これ等の人々が，人間の belief を調整してくれ規準を与えてくれるような客観的公式を作り出すことができるとしている点，及び人間の行動の主な原動力となるのは belief であると考えていて，色々の行動をとつた時の色々の結果については何等顧慮していない点である．

第一の点について Neyman は次のように述べる．この立場をとる場合の科学者のやり方は，一組の仮説及び彼の得ている限りの知識及びデーターの下で，先ず，彼の belief を得られている知識に合うように調整してくれる規準の公式を計算し，しかる後その結果に基ずき行動するということになる．所がこの場合 belief の調整の規準に関して科学者の間では必ずしも一致しない． belief の絶対的な測度というものがないのである．科学者は互に自分の測度の妥当性を主張することになるが，それは客観的な裏づけはなく，全く各々の科学者の主観に関する事柄である．

第二の点について Neyman は，彼の考え方を分り易く説明するために，我々は我々の belief に反した行動をとることもしばしば起るということに関し次の例をあげている．即ち飛行機で旅行する場合，人は墜落事故にあうことはまずないと確信している．それにも拘らず人は事故が起つた場合のことを考えて保険に入るという行動をとる．今取上げている問題においてこの例が適当であるか否かはともかく，Neyman の立場をこの例はよく示している．

彼の立場では，研究の目的は，考えられる幾つかの行動の過程のうちのどれをとるのがよいかを選択するための拠り所を与えることである．その場合最もよい方法を与えることは必ずしもできない．最もよい方法は，現実のその場の事情によるし，またそれは個人の好みとか，belief とかにもよるからである．しかしわれわれの行動を観測された結果に基ずいて調整するための色々の可能な方法の全体を研究し，それにより誰でもが，そのめいめいの場合に，それぞれの好みや belief に応じて最もよいと考える方法を選べるようにしてやることはできる．

以上のような Neyman の立場は behaviorist の立場であり，帰納法をその立場から考える訳である．彼の inductive behavior の考え方は自然現象を含めて広く一般の現象に亘つてとり入れることができるもので，Fisher のいうようにそれが工業や商業の分野における一個の道具たるに留まるべきものだとは Neyman は絶対に考えないのである．

§ 5. Fisher と Neyman の論争

Fisher と Neyman の間の論争の萠芽は大体 1934 年頃に現われたと見得るようである．最初その中心課題となつたのは fiducial argument の是非であつた．Neyman によつて発展せしめられた confidence limits の考え方は，それが発表された初めの頃（1934 年）においては，Fisher がその少し前に提唱した fiducial argument（論文［11］に発表）と同等なものと思われていた。Neyman 自身 1934 年に confidence limits の論文を発表した当時においては，両者の研究の方法ははなはだしく違つてはいるが取扱つていることは本質的に同じものであり，自分の理論は Fisher のものを更に発展させたものだと思つていたのである．それはその当時までは，何れの方法によつても同一の結果の式が得られたためであつた．Fisher のとつた手順や概念（fiducial distribution 等）は Neyman には理解のできないものがあつたが，これは何か Fisher のいい損いだ位に Neyman は思つていたのである．所が彼がこの論文の内容を Royal Statistical Society において発表し，それについて討論がなされた時，学会に出ていた Fisher の考え方をきく機会を持つたが，Neyman の予期に反し Fisher は fiducial probobility と fiducial distribution が Fisher の理論では本質的な部分であると述べ，更にこれらが用いられるのは，充足統計量が存在する場合に限るべきだと述べた（Neyman の理論では統計量の充足性は必要としない）．ここにおいて Neyman は二人の理論が全く異るものであるという確信を持つようになつた．その後 Fisher と Neyman はそれぞれ独立に自己の理論を発展させた（Fisher の論文［13］，Neyman の論文［24］等参照）．所が fiducial argument と confidence limits は同じものか全く異つたものかということをめぐつて色々の人々が論じたが，その中には両者が本質的には同じものであるという主張もあつた．就中 Bartlett が論文［5］及び［6］において両者が同じものであるという意見を述べたことに対して Fisher は一連の論文において反論している（例えば［14］参照）．

このような背景のもとに Neyman はいよいよ立場を明確にする必要性を感じ，fiducial argument と confidence limits の考え方が，概念的にも導かれる結果の上でも全く異るものであることを論ずる論文を発表した（論文［28］参照）．この論文において先ず Neyman は両者の相違を理論的に立証している．即ち彼は Behrens-Fisher の問題*を取上げて，この問題における fiducial argument は confidence limits の条件を満足しないことを証明している．更に彼は二つの方法

* 二つの正規母集団についての等平均値の仮設の検定の問題である．詳しい事は 38 頁を参照せよ．

の論拠について触れているが，特に彼は fiducial argumaent の基礎にある確率についての考え方が彼にとつては全く理解できないものであることを述べている．

ここにおいてわれわれは，Fisher と Neyman の考え方の相違と両者の論争の根本が奈辺にあるかを，漠然とではあるが暗示されるように思うのである．われわれは両者の考え方の差を明らかにするために，以下両者の主張及び論争について詳しく述べていく．

Neyman の確率は §3 にも述べたように，同一条件の下での実験の繰り返えしにおいて相対頻度として意味づけられなくてはならないものである．同一条件の下での実験の繰り返えしは勿論それが直ちに現実に実現できるものではない．しかしそれは少くとも仮想的に常に考えられなくてはならない．即ち彼は同一母集団からの繰り返えしのサンプリングを常に頭に置いているのである．これに反し Fisher はこのような実験の繰り返しは少くも彼が関係している分野においては現実の立場からは無意味であると考える．彼にあつては確率は実験を繰り返えしていくことによりその値が把握されるものでなく，確率分布そのものが既に現象の研究において取上げた問題に付随して存在すると考えているように思われる．しかし彼は現象そのものが数学としてはある確率論的構造で表現されると考えている訳ではなく，現象の学問的研究に対して，研究の各段階において統計学を inductive inference のための手段として考え，このような立場から統計学を適用しようとする場合の考え方として確率の場を設けるのである．推論の結果の妥当性は常に統計学だけでなく，それ以外の種々の証拠にも多くの考慮を払わなくてはならないと彼は主張する．彼の行き方の基調をなすものとしての induction inference を常に頭に置いて初めて彼の考え方は了解される．彼の確率の考え方を規定する今一つの要素は，彼の研究の対象となる分野が生物現象にあるということである．確率を実験の繰り返えしに結びつけて考えることは，彼の研究における実際上の考慮からすれば無意味である．生物学の分野では実験は任意に数多く繰り返えせるものではない．この制約の下で而も自然科学者の立場にあつて，その理論展開のある段階においてたてた仮説の妥当性をデーターから判断せんとする場合，彼はその段階における何等かの合理的結論を欲するものであつて，「若し同じ条件の下に実験を繰り返えしたならば」というような仮定の下での結論は無意味なのである．

このようにして Neyman にとつては確率はその数字それ自体としてい空虚なもので，それが真に意味を持ち得るのは実験の繰り返しにおける相対頻度としての意味づけができるからである．これに反して Fisher では，確率がそれ自体事象の確からしさを表現するある何ものかであつて（彼もまた相対頻度とししして考えているようであるが，それは実験の繰り返しにおいて見出される意味でなくて，現実に母集団のうちの問題にしている部分がその割合においてあるという意味で考えている（論文 [9] 参照）．この尺度に照らして仮定の妥当性を判断していこうというのである．このようにして Fisher の思考過程においては確率について言及する時常に，確率を一個の尺度（而もその妥当性は繰り返えしの実験の場において確かめられるようなものではない）として考えているようである．その妥当性は Fisher の場合確証によらず信念によつているようにさえ思われる．あるいは彼は彼の考え方の妥当性は結局真理の探求における彼の理論の展開の結果に照らして確証されると思つているのかも知れない．とにかく統計理論自体がそれ自身完成したものとなつていることよりも，真理の探求において彼が正しい結果に導かれること，その場合に統計学が有用な一つの方法として使えることが Fisher にとつては重要であるということだけはいえよう．

以上のような基礎的な考え方の差から母集団とそれからとられたサンプルに対する二人の考え方には結局次のような違いがあるといえよう．Neyman の場合はサンプルを考える場合には常に同一母集団から繰り返えしのサンプリングを想定する．勿論その場合色々の可能な母集団を考えていく（それは現象をできるだけよく表現するもとなるように考えていく）．而してサンプルはこれ等の母集団の中の真なるある一つのものからとられたとみえるが，この唯一回のサンプリングだけを考えることは決してない．一方 Fisher の意見では（論文 [16] 参照），同一の母集団からの繰り返えしのサンプリングが考えられるためには，サンプルがとられるべき母集団が客観的に実在する

のでなければならぬ．それは例えば工業生産における品質管理のごとき問題の場合には可能である．しかしわれわれが *unique sample* を以て *significance test* を行わんとする場合には，このような母集団の存在は統計家の空想に過ぎない．実際にはこのような場合には，この段階までの知識からサンプルの元として妥当なものとして設定し得る母集団の可能な群があるということのみに基ずいて考えていかねばならない．その何れかを想定してそれから仮想される繰り返えしのサンプリングを想定する考え方は，われわれの知りたい真の知識の獲得には役立たぬ．このように Fisher は考えるのである．彼の立場は多数の母集団（*multiplicity of population*）の考え方が基本的なわけである*.

---

* (イ) 統計的推論ということが，得られたデーターからその元の母集団に関して推論することを意味するとすれば，Neyman の如く繰返しのサンプリングに於てどのようなことが起るかということに基づいて考える考え方，従つて power だけの考慮だけで推論をするやり方は合理的でないということについての D.R. Cox の議論は，上述の Fisher の反論よりも理解し易い．以下に Cox の与えた一つの例について述べてみる（論文 [7] 参照）．

今正規母集団の平均値 θ を問題にしているとす．但しこの場合の母集団は，平均値が共に θ で分散が夫々 σ_1^2, σ_2^2 且つ $\sigma_1^2 \gg \sigma_2^2$ なる二つの正規母集団 Σ_1，Σ_2 の何れかが，夫々 1/2 づつの確率で現われるものとす．又サンプル S がとられる時，それがどちらの母集団からとられたものであるかは判るものとす．この時仮説 $\theta=0$ を，対立仮説 $\theta=\theta'(\simeq\sigma_1)$ に対し検定する問題を出来るだけ Neyman 流の formulation で考えてみる．

繰り返えしのサンプリングを考えてゆくとすれば，サンプル空間は Σ_1 と Σ_2 を 1/2 づつの確率で混合して得られるものである．われわれはこの全体的サンプル空間で考えて most powerful な検定法を決定することができる．一方 power の点ではおとる次のような検定法が考えられる．即ち Σ_1 又は Σ_2 の中で考えて，それ等のサンプル空間に関してそれぞれ most powerful な棄却域 W_1, W_2 をきめておき，若し S が Σ_1 から得られた場合には W_1 に基づき，Σ_2 から得られた場合には W_2 に基づき検定をする．

第1の方法の場合の考え方は，S が例えば Σ_1 から得られたと判つた場合でも，考察を Σ_1 に限定しない．即ち今は S が Σ_1 からとられたが，われわれが繰り返しのサンプリングを行つてゆけば Σ_2 の方からサンプルがとられることも起り得る．このことを考慮しない第2の方法では power が劣る．power だけで検定法の良否をきめるとすれば当然第1の方法の方がよい訳である．

しかし統計的推論の意味を前述の如く考える時，繰り返しのサンプリングの間に Σ_2 の方も起り得るということはデーターの翻訳のためには無関係なことである．即ち S が Σ_1 から得られたサンプルであると知つた場合には，S と同質のサンプルの色々の起り方に関しての推論をしたいのであつて，従つて Σ_1 に限定して考えなくてはならない．同様にもし S が Σ_2 から得られたと判つたなら Σ_2 に限定して考えるべきである．このように単に power のみを考えて検定法の良否をきめることは誤りである．

(ロ) 繰り返しのサンプリングの考え方が常に妥当だとは限らないことを特に強調する研究者として我々は Barnard をあげることが出来る．一例について彼の考え方の一端を述べてみる．

今 A, B なる二つの処理が或る植物の苗の成長に効果があるかを見る問題を考える．簡単のために処理の結果が成長を "助長" するか "妨害" するかの二つに判定されるとし，之等のことをそれぞれ I, II であらわす．この時 "A, B なる二種類の処理と苗の成長とは無関係である" という仮説を検定するものとす．一つのやり方としてN本の苗に対して次のような実験を施す合を考える．N本のうちのm本には処理Aを，n本には処理Bを場（$m+n=N$）施した後，実際の苗の成長を見る．但しAを施すべき苗とBを施すべき苗はランダムに決める．このようにして実験した結果は，次のような表であらわされる．

処理＼結果	I	II	計
A	a	c	m
B	b	d	n
計	r	s	N

この場合の Neyman 流の考え方は次の通りである（論文 [32] 参照）

先ず問題の上の如き取り上げ方の場合，実験の対象としてのN本の苗は固定されている．故に出る結論はこの特定のN本の苗に関する結論である．次に我々は推論のために実験の繰り返しを考えるが，それは二種類の処理の割当てに当つてのランダマイゼイションに関してである．即ち若し実験が初めから繰り返されるとすれば，実験毎にA乃至Bの割当ての模様は変動する．勿論この場合実験の繰り返しは仮想的なものである．考えているN本の苗のうち I なるものの数が r 本，II なるものの数が s 本（$r+s=N$）とするに，仮説が真であれば上の表のような結果を得る確率は $m!n!r!s!/N!a!b!c!d!$ になる．われわれは特定のN本の苗を考えているから，仮説が真であれば r, s は常数であるとしてよい．又その数値は実験の結果から知り得る．上の如く計算される確率は，上に述べた意味の仮想的実験の繰り返しに於て起る相対頻度としての意味を持つものである．

若し我々が特定のN本に関してでなく，更に広い母集団に関する推論をしたいとすれば，問題の上のような取り上げ方ではよくない．その場合には苗の母集団を想定して，そこからランダムに抽出されたN本であると考えるとか，或はA又はBのそれぞれの処理を施した苗の母集団を想定して，その各々からそれぞれ m 本及び n 本ずつランダムに抽出されたと考えるとか，いつた考え方をしなくてはならない．その場合には実験の繰り返しはランダマイゼイションに関して考えるのではなく，われわれの目的に適うように想定された母集団からの繰り返しのサンプリングと考える訳である．

この考え方に対する Barnard の反対は次のような点にある．上の取り上げ方の場合ランダマイゼイションに関して実験の繰り返しを考えるというのは，あく迄も仮想的なものである．もしやろうと思えば繰返しが可能であるものならまだ話は判る．今のような場合に確率を実験の繰り返しに結びつけて考えることは無意味である．N本の苗を特定のものに固定せず，その変動を考えていくという後に述べられている取り上げ方ならば，確率を実験の繰り返しに結びつけて考えることも納得できる．但しその場合にも，そのように確率を考えて行くことが許されるだけの，条件の恒常性について充分の保証がなければならぬ．例えば工業生産のような場合ならば，同一条件の保持が出来る問題が多いであろうし，繰り返しのサンプリングを考えてゆくことは大体妥当であろう．しかし，一般の場合にも常に確率を繰り返えしのサンプリングに結びつけて考えてゆくという事はよくない．

これが Barnard の考え方である．このような考え方から彼は，Fisher の考え方，尤度や fiducial probability に重きを置くやり方を見直さなくてはならないという方向に進んでいる（論文 [1], [2], [3], [4] 参照）．論文 [4] に於て彼は尤度比をもとにして仮説の間で確からしさを比較する或る考え方の数学的 formulation を試みている．しかしながらこの論文は問題の取上げ方が余りに形式的に過ぎて，彼の方法の内容的な面でわれわれを納得させるには程遠いように思われる．

このような Fisher の考え方をより明らかにするために次に具体的に一例をあげてそれに対する Fisher のやり方をみていくことにしよう（論文 [6] 参照）．今変数 (x, y) についてのN個の観測値がデーターとして得られているとし，x の各々の値に対し y の分布が分散 σ^2，平均が $E_x(y)=\alpha+x\beta$ の正規分布であることが別の知識から分つているとす．この時 β の推定量として統計量 $b=\sum(x-\bar{x})(y-\bar{y})/\sum(x-\bar{x})^2$ をとることができる．同一の $A=\sum(x-\bar{x})^2$ を持つサンプルに対し，b は平均が β，分散 σ^2/A の正規分布に従う（A の値を単に固定して考える）．故に仮説 $\beta=\beta_0$ を検定するには $t=(b-\beta_0)\dfrac{\sqrt{A}}{s}$ が自由度 $N-2$ の t–分布にしたがうことを用いればよい．これがこの場合の Fisher のやり方である．ここで肝腎なのは変数 x に関する部分 A を固定してしまう点である．而もそれは条件付きの分布を考えるという意味ではなく，頭から固定してしまうのである．この場合 Fisher はxが確率変数でないとしている訳ではない．xがまたある分布にしたがつていることが知られてもよい．ただxが確率変数であるという知識は推論に当つて無関係な知識であり，xに関して彼が使うのは，(x, y) の観測値において x がしかじかの値として観測されたという事実のみである．即ち b が平均値 β のまわりに分散 σ^2/A をもつて分布するということは acceptance procedure としてのサンプリングの過程（その場合は勿論xの観測値したがつてAが観測ごとに変動していく）に対応するものではない．それは観測されたサンプルに対し，凡ゆる関連する点を考慮した上でそれの元の母集団と考えられるものに対応するのである．観測値Aは補助の変数であつて，この場合Aを固定しても，xの観測値 $x_1 \cdots\cdot$, x_N そのものを固定してもどちらでもよい．これがこの例の場合における Fisher の考え方である．このような考え方に反対して Neyman がどのように考えるかは明らかであろう．Neyman の場合はxも飽くまで確率変数として取扱かわれなくてはならず，xの観測値を固定する場合は，それ等が固定されたという条件の下におけるyの観測値の分布を考えていくことであらねばならぬ．彼の場合には推論に先立つて確率変数 (x, y) に対応する母集団が考えられていて，以後の推論の過程においてはあくまでもこの母集団を基礎とする確率の場が保持され，それからの繰り返しのサンプリングの概念に基ずいて議論が進められねばならないのである*．

---

* 同じ問題に対する取上げ方が違うので，この場合導き出される結論の意味が両者で違つて来る．Fisher の場合には，得られたデーターに対し，このようなデーターを与えるような（特に変数xについては，その観測された値と同じ値を与えるようなものという制限をつける）母集団として，仮説 $\beta=\beta_0$ で指定される母集団が適はしいものか否かが結論される．これに反し Neyman の場合には，現在得られたデー

fiducial argument が **Neyman** にとつて全く理解できないのも，二人の立脚点が上述のように全く対立しているためである．次に簡単な例について **fiducial argument** についての二人の意見の相違を示してみよう（論文 [16] 及び [31] 参照).

今，平均値 μ の正規母集団からのサンプル $x_1,\cdots\cdot x_n$ が得られた時，それをもとにして μ の **fiducial probability** を出す場合の **Fisher** 思考の論理をおつてみよう．このサンプルから $\bar{x}=\sum x_i/n$, $s^2=\sum(x_i-\bar{x})^2/(n-1)$ を作り，更に$t=(\bar{x}-\mu)\sqrt{n}/s$ を作ればこれは t–分布にしたがう．故に信頼度 95% を与えた時，対応する $t/\sqrt{n}$ の値を用いれば，$P_r\{\bar{x}-t_0s<\mu<\bar{x}+t_0s\}=95\%$ である．ここで $\bar{x}$, s の値に観測値を入れればμに対する評価の確率として例えば $P_r\{91.9<\mu<93.0\}=95\%$ を得る．即ち母集団のパラメーター μ が **91.9** と **93.0** の間にあることが95%の確率でいえる．これが **Fisher** の考え方である．

Neyman の考え方にしたがえばこれは明らかにおかしい．第一の式は意味を持つが第二の式は全く無意味である．$\bar{x}, s$ は確率変数であり，したがつて $\bar{x}-st_0$, $\bar{x}+st_0$ もまたそうである．それ等の変動を考えることにより第一の式の確率が意味を持つているのである．所がそれ等に数値を入れ，かくしてそれ等を常数とみた場合に，それ等に関して何等かの確率を以々することは無意味である．しかるに第二の式でこのようなことがなされている．又 μ は初め常数として考えられたものだが，第二の式ではこれが確率変数となつている．実際は μ の値は常数であつて母集団のパラメーターの真の値として考えられている．故に $91.9<\mu<93.0$ という事柄は起るか起らないか何れかであつて，即ちその確率は **1** か **0** である．**Neyman** の考え方では第一の式の確率の意味は同じ条件の下における繰り返えしのサンプリングにおいて $\bar{x}-st_0<\mu<\bar{x}+st_0$ なることが本当に起る割合である．所が第二の式においてはこのような意味づけはできない．

これに反して **Fisher** は次のように反論する．「母集団のパラメーター θ と統計量 T に関し，$P_r\{T<\theta\}=5\%$ なる関係により **significance test** を行う場合，この式の意味として (a) “θ は Tをこえる確率 5% を持つ” といういい方と (b) “T は θ より小さい確率 5% を持つ” というい方は全然同等なものであり，共にTとθの間に存在する関係を表したものである．この場合これ等のいい方は変数 (θ, T) の母集団に関したものとして意味を持つている．上のいい方ではこの母集団において θ とTのどちらかが常数であるともいつてないし，さればといつて一方を固定して他を動かして考えることがよくないともいつていない．それで前述のような通常の **significance test** の場合には，計算に当つては θ を固定してTの変動に関し確率を計算すればよい．実際この場合 θ の変動はあるのだがそれは計算には関係してこない．このことは丁度 **regression** の問題において独立変数の分布が計算に関係しないと同様である．

この様な **Fisher** の考え方にはわれわれにどうも納得のいかぬ点がある．成程 T のみならず θ の変動も考えて変数 (θ, T) の母集団を考えるのであれば $P_r\{91.9<\mu<93.0\}=95\%$ の様な式は意味を持ち得る．然しこれ等の数値が μ を固定して計算した $P_r\{\bar{x}-st_0<\mu<\bar{x}+st_0\}=95\%$ なる式に $\bar{x}$, s の数値を単に入れた結果として導かれることは，どう解釈してよいのか分らない．

regression の問題との類似を云々することも何の意味か分らない．(θ, T) に関する母集団を考える以上Tを固定した時の θ の確率の計算では少くとも θ の **a priori** な分布が考慮に入れられるのでなければ到底理解できない．固定されていた θ が以後動かして考えられるためには，この過程の途中においてどこかにこの θ の考え方の変化を可能にする思考がなければならないであろう．

このように上の例では **Fisher** の考え方にはわれわれの普通の思考からすれば何か理解し難いものがある．故に **Fisher** が今の場合 **fiducial probability** を母集団のパラメーターに対する推論

ターたけでなく，繰り返しのサンプリングに於てデーターに関しみられるであろう変動を考慮（x も変動する）した上で，現在のデーターの元である母集団が，仮説 $\beta=\beta_0$ で指定される母集団であるとすることが妥当か否かが結論される．

の確からしさを与えるものと主張する時，われわれはそこに単なる彼自身の信念を見るように思うのである．しかしそれでは fiducial probability あるいは fibucial distribution の考え方が全くおかしいものかといえば一概にそうはいえないであろう．これ等の概念は母集団のパラメーターの値について推論の確からしさをあらわす一つの（勿論絶対的なものではない）測度と考える事は出来る（Quenoulli の本 ct）．confidence limits の概念も同様に一つの測度である．工業生産における問題のような場合には勿論後者が遙かにすぐれた測度であるが，同一の母集団からの繰り返えしのサンプリングの考え方が妥当でない場合（取上げた問題によつてはそのような場合が確かにあり得る）においては，confidence limits の考え方はよい測度であるとはいえない．勿論それが直ちに fiducial limits のかかる問題における妥当性を意味しはしないが，一つの測度として考えることは許されるであろう．

Fisher と Neyman の論争に関して今一つ，前に述べたような Fisher の考え方を示す著しい例として，Behrens-Fisher の問題をあげる．これは論文 [28] において Neyman が取上げたものである（Fisher の論文 [13] も参照されたし）．大さ n 及び n' の二種の互に独立な観測値がそれぞれ二つの正規母集団の各々からのサンプルである時，これらが同じ平均値を持つ母集団からのサンプルであるか否かについての significance test の場合である．今二つの母集団の平均値を μ 及び μ'，また分散を σ^2 及び σ'^2 とす．これ等のサンプルの各々のサンプル平均を $\bar{x}, \bar{x}'$ とし，また $s^2=\sum(x-\bar{x})^2/n(n-1)$，$s'^2=\sum(x'-\bar{x}')^2/n'(n'-1)$ とおく．しかる時は $t=\mu-\bar{x}/s$，$t'=\mu'-\bar{x}'/s'$ は互に独立にそれぞれ自由度 $n-1$ 及び $n'-1$ の t–分布にしたがう確率変数である．容易に $\{(\mu'-\mu)-(\bar{x}'-\bar{x})\}/\sqrt{s^2+s'^2}=t'\cos R-t\sin R$ となることが分る．ただし $R=s/s'$ とす．仮説が真であれば $(\bar{x}-\bar{x}')/\sqrt{s^2+s'^2}=t'\cos R-t\sin R$ となる．故に $t'\cos R-t\sin R$ の分布が分れば，$(\bar{x}-\bar{x}')/\sqrt{s^2+s'^2}$ を用いることにより，与えられた test を行うことができる．ここで Fisher は R の値を観測値に基ずき固定する．即ち $t'\cos R-t\sin R$ を t と t' の変動のみに関して変動させて考える．而もこの t と t' は上に導いた互に独立な t 分布にしたがう変数として考えていく．このようにすれば勿論 $t'\cos R-t\sin R$ の分布は容易に導くことができる．ここで問題になるのは，推論の途中においてRは知られた量であるとして固定し，かくして t と t' に関する分布のみから $t'\cos R-t\sin R$ の分布を出す点である．Neyman の考え方では，Rを固定するのは，この値が得られたという条件のもとで考えることであり，上述の Fisher のようなやり方で求めた $t'\cos R-t\sin R$ の分布は間違つていることになる．実際に彼は R を固定したという条件の下で (t, t') の同時分布を求め，この条件付き分布では t と t' の独立性は失われること，及びこの分布には Fisher が消去したとしている母集団のパラメーター σ^2 及び σ'^2 が現われて来ることを示している．このように普通の確率の計算法からすれば，Fisher のやり方は明らかに間違つていることになる．Fisher 自身そのことはよく知つているのである．彼がRの値乃至 s, s' の値を固定する時，条件付き確率を考える訳ではないのである．s, s' の値を観測した上で彼はその状況に最も適切な推論の場として，t と t' が互に独立な t–分布の変数として変動するような母集団を考えていくのである．この過程は母集団が観測値に基ずき一歩一歩ごとに（この場合二段にわたつて）設定し直されるという風にいい表わせるであろう．

Fisher と Neyman の論争における今一つの代表的問題は第2種の過誤をめぐる対立である．次にこれについて述べよう．この対立は結局 inductive inference と inductive behavior の二つの考え方の対立から来ている．

Neyman にあつては behaviorist の立場から，仮説検定の問題における第2種の過誤の考え方は必然のものである．仮説の採択は behavior ということに関しての一つの可能性であり，この可能性の与える結果を評価しておくことはこの場合の inductive behavior のための必要事項の一つである．またこの場合，二つの可能な行動——仮説の棄却と採択——は互に対立するものであり，

その各々から起る過誤は違つた種類のものと考えられる．この区別をあらわすために Neyman は第1種の過誤と第2種の過誤を用い，それ等を互に対立させるのである．

一方 Fisher の方は inductive behavior に反対する立場から必然的に第2種の過誤を考えることに反対する．Fisher は次のようにいう．

（1） 仮説の検定が decision の一種と了解された場合，異る行動のそれぞれから起る結果に対しその確率を評価しなくてはならないが，そのためには a priori な確率が知られるか．あるいは過去の経験により，行動の結果の起り方の確率が知られることが必要である．例えば品質管理のような問題ではこれは可能となるが，自然科学の研究においてはほとんどの場合この様な a priori な知識はないかまた役に立たない．

（2） またたとえ a priori な知識があつたとしても，第2種の過誤は実際には評価できるものではない．何となれば第2種の過誤を単にその起る確率によつて評価しても無意味である．真の評価であるためには，帰無仮説以外の個々の仮説が帰無仮説に非常に近いあるいは遠いという点に関する評価をも含まなくてはならないが，このようなことは数値で評価することは不可能である．

（3） 最後に帰無仮説が“誤つている場合にそれが採択される”といういい方は，研究者が現実のこのような場合にとる態度を無視するものである．test of significance において帰無仮説が棄却されなかつた場合，現実の研究者はこの仮説の採択を行いはしない．彼はこの結果の強弱に応じてこの仮説が真であることについて色々な程度に確信を強めはするが，機械的に採択を行うことはない．(このような確信が積み重ねられ，また他の色々な考慮の下に仮説の真なることが極めて強く確信された場合に初めて仮説を採択するというのが Fisher の考え方であると思われる．ただしこの場合でも仮説の採択は仮の説の一部としてとり入れるという意味でなされるもので，その意味であくまでも暫定的なものである)．これに反し acceptance procedure（仮説が誤つている場合にそれが採択されるということは acceptance procedure でなければ起り得ない）においては，前もつて与えておいた数学的な規則に基ずき行動を機械的に行うもので，そのようなやり方は現実の研究者においてはみられない．

これが第2種の過誤の考え方に反対する Fisher の意見である*.

（1） については，Neyman の仮説検定論は a priori な分布の考慮が必要でないように作られており，実際一様に most powerful な検定法のような規準が用いられるという反論ができる**.

---

* その他にも例えば有意水準（Neyman の場合は第1種の過誤の確率）に対する考え方に関して，両者は次の様に対立する．Fisher の場合は有意水準を例えば5%にとるか1%にとるかは，取上げている個々の問題に対する考慮からきめるべきものと考え，確率だけの考慮からきめるのではない．一方 Neyman の方は，第1種の過誤の確率を5%に与えるか1%に与えるかは第2種の過誤がからんでくる問題であり，この様な過誤の確率の綜合的考慮の下にきめられるべきものだと考える．従つて彼にあつては5%に与えるのと1%に与えるのとではその意味が全く異なつてくる．

** かかる反論は決して充分なものではない．即ちわれわれは次のような点で一様に most powerful な検定法の考え方が欠点を持つていると思う．

先ず第一に一様に most powerful な検定法に対する第2種の過誤の確率が相当に大きい場合が起つた時，実際問題としてわれわれは何もすることが出来ないであろう．inductive behavior の考え方はこの場合問題の解決に役立つものとは思われない．

第二に繰返しのサンプリングを考えていく場合に母集団がランダムに変動すると考えなくてはならない問題も実際には多いであろう．このような場合は母集団の変動に対する a priori な確率の分布を想定しなくては理論上確率が計算出来ない．勿論このような場合でも一様に most powerful な検定法を考える限り，a priori な確率の分布を実際に知ることは必要でない．即ちすべての可能な a priori の分布のうちのどれが実際に起つているとしても，一様に most powerful な検定法は他の検定法よりも第2種の過誤の確率を小にする．しかしこの場合の過誤の確率の値の実際の大きさは（第1種，第2種の何れも）a priori な分布が判らぬ限り実際に知り得ない．a priori の分布が分らなければ，過誤の確率（即ち第1種及び第2種の過誤の確率を a priori な確率を weight として加えたもの）を評価するには，それを知られている知識から実際に可能と考えられる a priori の分布の全部について考え，その最大値をとるより他ないが，この場合この過誤の確率の最大値は，a priori の分布についての知識が全然

(2) の場合は loss というものを現実の問題でどう評価するかという難点のために，一応 Fisher のいう所を認めない訳にはいかぬ．(3) については，Neyman の立場でも数学的規則をたてておいて機械的に行動する訳ではない．仮説が棄却されなかつた場合直ちに採択するとは限らない．この場合直ちに採択するというやり方も一つの方式であるがその他の方式も勿論あり得る．即ち仮説が採択されなかつた場合，更にデーターをとるなど別の確証を得るという行動を行動の一つの可能性としてとり入れた方式も考え得る．どのような方式をとる場合でも，行動の最終的結果の可能性の一つとして，“仮説が誤つている場合にそれが採択される”という可能性が考えられる．Neyman の仮説検定論は色々の decision の rule の一つの方式に関するもので，この方式のもとで最も良いものを研究しているのである．Neyman を代弁すれば以上のようになるであろう*.

以上 Fisher と Neyman の論争の主な点を大体述べた積りである．われわれは両者の立場をできるだけ公平に理解しようと努めた．ただ Neyman についてはその考え方は何人にも容易に理解できるが，Fisher の考え方は難解を極める点が多い．この解り難い Fisher の考え方の理解を少しでも進めようと意図した結果，自然 Fisher の意見に関する部分が多くなることを免かれなかつた．また色々な点でわれわれの思い違いも多いであろうが，それについては読者の御批判に俟つものである．

§6. むすび

以上 Fisher と Neyman の統計学に対する立場を述べてきたが，この両者の対立は単に論理的な面ばかりではなく，その科学に対するイデオロギーの相違にも大きな問題があると考えられる．

Fisher は「米国においては統計学が生産の向上と企業の拡充のための道具として扱われている．こうした背景のもとにでき上つているのが Neyman 流の統計であつて，これは純粋科学の道ではない」と断言する．この言葉が果して Neyman の立場の正しい見方であるかどうかは疑問であるが，その一面を指摘していることは事実であろう．確かに，ある目的をもつて構成された母集団に関して測定され，そして決定される Neyman のいう行動はその母集団，測定法を離れては意味をもたない．これは絶対の真理を求めようとする Fisher の立場からみればいかにも不純なものとして写ることであろう．しかし Fisher の立場も科学の研究の方法論として完全なものとはいえない．そこには数学的疑問，実践的意味の薄弱さが存することは明白である．

現在の統計の研究はほとんどすべて Neyman 流に集中され，Fisher 的行き方は今や忘れ去られようとしている．しかし科学の方法論として両者をみた場合必ずしも Neyman 一辺倒とばかりはいくまい．第2種の過誤の難点，科学研究における loss function, cost function の意味づけの困難性など基本的問題がのこされている．

われわれは統計学において今後方法論としてより筋の通つた立場とその理論を求めねばなるまい．

統計数理研究所

文献

[1] Barnard, G.A. (1945). A new test for 2×2 tables. *Nature*, **156**, No. 3954, 177.

ないために凡ゆる a priori の分布を考えなくてはならぬ場合には，極めて普通な或る条件の下で常に 1/2 を越えることが vor Mises により示されている (論文 [18] 参照)．故に一様に most powerful な検定法であつても，上述のような場合には評価された過誤が大きすぎて実際上使いものにならない．本当は過誤が過大に評価されているのかも知れない訳であるが，正確な過誤の評価のためには a priori な分布の知識が必要とされる．

* このような弁明には又反論の余地は大いにある．色々の方式の中には，その数学的 formulation が極めて困難なものがあるであろう．又数学的 formuletion ができるものでも，formulete された loss が余り現実的に妥当な意味を持ち得ないものもあろう．そのような場合色々の方式の何れを採用するかは結局主観の問題となつてしまう．

[2] Barnard, G.A. (1947). Significance tests for 2×2 tables. *Biom.*, **34**, 123.

[3] Barnard, G.A. (1947). The meaning of a significance level. *Biom.*, **34**, 179.

[4] Barnard, G.A. (1949). Statistical inference. *J. R. S. S.*, *B*. **11**, 115.

[5] Bartlett, M.S. (1936). The information available in small samples. *Proc. Camb. Phil. Soc.*, **32**, 560.

[6] Bartlett, M.S. (1939). Complete simultaneous fiducial distribution. *Ann. Math. Stats.*, **10**, 129.

[7] Cox, D.R. (1958). Some problems connected with statistical inference. *Ann Math. Stats.*, **29**, 357.

[8] Feller, W. (1938). Note on regions similar to the sample space. *Stat. Res. Mem.*, **2**, 117.

[9] Fisher, R.A. (1921). On the mathematical foundations of theoretical statistics. *Phil. Trans. Roy. Soc. Lond.*, *A*, **222**, 309.

[10] Fisher, R.A. (1925, 1944). *Statistical Methods for Research Workers*. (Ist edn. 1925. 9th edn. 1944). Oliver and Boyd, Edinburgh.

[11] Fisher, R,A, (1930). Inverse probability. *Proc. Comb. Phil. Soc.*, **26**, 528.

[12] Fisher, R.A. (1933). The concepts of inverse probability and of fiducial probability reffering to unknown parameters. *Proc. Roy. Soc.*, *A*, **139**, 343.

[13] Fisher, R.A. (1935). The fiducial argument in statistical inference. *Ann. Eyg.*, **6**, 391.

[14] Fisher, R.A. (1937). On a point raised by M.S. Bartlett on fiducial probability. *Ann. Eug.*, **7**, 370.

[15] Fisher, R.A. (1940). A note on fiducial inference. *Ann. Math. Stats.*, **10**, 383.

[16] Fisher, R.A. (1955). Statistical metheds and scientific induction. *J.R.S.S.*, *B*, **17**, 69.

[17] von Mieses, R. (1942). "On the correct use of Bayes' formula," *Ann. Math. Stats.*, **13**, 156.

[18] von Mieses, R. (1943). On the problem of testing hypotheses. *Ann. Meth. Stats.*, **14**, 238.

[19] Neyman, J., and Pearson, E.S. (1928). On the use and interpretation of certain test criteria for purpose of statistical inference. *Biom.*, *A*, **20**, 175 and 263.

[20] Neyman, J., and Pearson, E.S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Phil. Trans. Roy. Soc. Lond.*, *A*, **231**, 289.

[21] Neyman, J., and Pearson, E.S. (1936). Contributions to the theory of testing statistical hypotheses: I. Unbiassed critical regions of Type A. *Stat. Res. Mem.*, **1**, 1.

[22] Neyman, J., and Pearson, E.S. (1936). Sufficient statistics and uniformly most powerful tests of statistical hypotheses. *Stat. Res. Mem.*, **1**, 113.

[23] Neyman, J., and Tokarska, B. (1936). Errors of the second kind in testing 'Student's' hypotheses. *J. Am. Stat. Ass.*, **31**, 318.

[24] Neyman, J. (1937). Outline of a theory of statistical estimation based on the classical theory of probability, *Phil. Trans. Roy. Soc.* Lond., *A*, **236**, 333.

[25] Neyman, J. (1937). *Lectures and Conferences on Mathematical Statistics*. Lectures delivered at Graduate School, U.S., Department of Aggriculture,

[26] Neyman, J., and Pearson, E.S. (1938). Contribution to the theory of testing statistical hypotheses: II. Certain theorems on unbiassed critical regions of Type A. *Stat. Res. Mem.*, **2**, 25.

[27] Neyman, J., and Pearson, E.S. (1938). Contribution to the theory of testing statistical hypotheses: III. Unbiassed tests of simple statistical hypotheses specifying the values of more than one unknown parameter. *Stat. Res. Mem.*, **2**, 36.

[28] Neyman, J. (1941). Fiducial argument and the theory of confidence intervals. *Biom.*, **32**, **128**

[29] Neyman, J. (1942). Basic ideas and some recent results of the theory of testing statistical hypotheses. *J.R.S.S.*, **105**, 292.

[30] Neyman, J. (1956). Note on an article by Sir Ronald Fisher. *J.R.S.S.*, *B*, **18**, 288.

[31] Neyman, J. (1957). "Inductive behavior" as a basic concept of philosophy of science. *Rev.*

Intern. Stat. Inst., **25**, 17.

[32] Pearson, E.S. (1947). The choice of statistical tests illustrated on the interpretation of data classed in a 2×2 table. *Biom.*, **34**, 139.

[33] Pearson, E.S. (1955). Statistical concepts in their relation to reality. *J.R.S.S.*, **17**, 204.

[34] Quenoulli, M.H. (1958). *The Fundamentals of Statistical Reasoning*. Charles. Griffin & Company, Ltd., London.

[35] Robbins, H. (1956). An empirical Bayes approach to statistics. *Proc. 3rd Berk. Symp.*, **1**, 157.

[36] Wald, A. (1950). *Statistical decision function*. J. Wiley and Sons, New York.