經濟論叢

第六十二卷 第一・二號

京都大學經濟學會

標本論の一般化の問題

——H. Levy and L. Roth の所論について——

青山秀夫

序

現代の統計理論は、その概念圖式に於ても、その問題構成に於ても傳統的な統計理論に比べて、著るしく——全然、と云つていいほど著るしく——面目を異にしている。こうした變化は大體イギリスの生物統計學者R・A・フィシャーの業績に由來する。一九二〇年代の始め、例えば論文"On the mathematical foundations of theoretical statistics," Philosophical Transactions of the Royal Society of London, Series A, Vol. 222, 1922. 或いは著書"Statistical methods for research worker", 1 st ed., 1925. などに於て、フィシャーはイギリス經驗主義の傳統に基きながら、統計的推論 statistical inference の論理と武器とを再構成することを試みたが、ここで彼が與えた新しい概念圖式と新しい問題構成とが現代統計理論の展開の基礎となつたのである。

さて此の理論統計學の飛躍的發展に當つてその一つの重要な基礎をなしたものは標本論 theory of sampling であつた。傳統的統計理論に於てかのガウス＝ラプラースの誤差法則ならびにそれに立脚する誤差論が重要な役割を演じていることは周知の通りであるが、一方に於ては、古典的誤差論が有した不合理な前提から自由に、合理

的に理論を構成するとともに、他方に於ては、統計的研究作業への確率論の適用範圍を飛躍的に擴大したものが此の標本論に他ならなかつた。フィシャーは、此の標本論をみづから展開しながら、然も、その成果が全幅的に利用されうるように、統計理論の概念圖式と問題構成とを再編成したのであり、彼の統計理論上の業績は此の標本論の展開と不可分に結びついているのである。たしかに彼の統計理論の基礎にあるものは、かの「ロヂツク・オブ・チャンス」(初版一八六六年)の著者ジョン・ヴェンなどによつて代表される「イギリス經驗主義的な確率論」(アルネ・フィシャー)であるが、此の經驗主義的な確率論を統計理論にまで結實させるに當つて標本論が決定的な役割を演じたことは否定し難いであろう。

現代統計理論に於て標本論の意義はかくの如く重大であり、從つてまた、現代統計理論の展開の一つの重要な中心は標本論にあるわけであるが、ここではH・レヴィとL・ロスとがその共著にかかる「確率論要論」("Elements of Probability", Oxford 1936.) に於て展開している標本論に關して、若干の解説と批判とを試みたいと思う。フィシャー以來、標本論は概ね正常分布 normal distribution を有する母集團 population, (parent) universe を前提して展開され來つたのに對して、此のLR理論(以下Levy-Roth の所論をこう呼ぶこととする)は、新しい觀點を加えることによつて、此の前提から自由に全く一般的に標本論を構成しようとしており、此の點に於て注目に値するからである。

かように本稿はLR理論の解説と批判とを根本の目標とするものであるが、然し、それに先立つて現代統計理論の內容を、必要な限りに於て、説明しておくのが便宜でないかと思う。吾國の學界の現状から云つて、成程、自然科學の方面では、多くの有能な數學者の努力によつて、現代統計理論は相當の普及を見るに到つたが、吾々

の社會科學の領域では、その必要がないわけではないにも拘わらず、此の新しい觀點は一般に親まれるにいたらず、從つて、此の點について或る程度豫備的說明を試みることが必要であると考えられるからである。此の意味に於て、先ず、「フィシャーに於て理論統計學の問題はどう再構成されたか、また、それに於て標本論はどういう地位を占め、どういう課題を負わされるか」、更に亦、「先に正常分布を有する母集團を前提して構成されるといつた現代標本論はどういう内容のものか、また、それは如何にして所與の課題を果すか」こういう點についてフィシャーの理論を概說し、然る後、LR理論に進むこととする。

一

先づ、問題の例示によつて、フィシャー的な――從つて現代統計理論の――概念圖式と問題構成とを說明しよう。

以下論ずるところは主としてフィシャーの上揭文献によるが、最近の英米の理論統計學の文献――それらは殆んど凡てフィシャーの線に沿つて展開される――に於ても吾々は殆んどつねに同樣の說明に遭遇する。例えば L. H. C. Tippet, The methods of statistics, 2 nd ed., London 1927. G. W. Snedecor, Statistical methods, Iowa 1937, W. D. Baten, Elementary mathematical statistics, New York 1939. P. R. Rider, An introduction to modern statistical methods, New York 1939. C. H. Goulden, Methods of statistical analysis, New York 1939. などを見よ。

問題例。或る電球商が製造業者から「平均の持續時間、五、〇〇〇時間」といわれる銘柄の電球を多量買い、念のため、此の中から何ら作爲するところなく九個取出して持續時間を檢査したところ、上記(1)の結果を得た。此の九個の標本の持續時間の平均は四、五〇〇時間(なお標準偏差は五七六時間)であるが、製造業者の「平均持續時間

五、〇〇〇時間」という主張は此の實驗結果から云つて、どこまで正しいであろうか。

さて此の問題例で問題となつているのは、明かに、電球商が製造業者から買つた電球全體の性質である。會社側はこれについてその持續時間の平均が五、〇〇〇時間であることを主張するが、電球商にとつては、これは「驗證」を要する假説である。ところで此の電球群全體について「眞」の平均持續時間を知る道は勿論存しない。吾々が知り得るの此の「母集團」から抽出した「標本」(sample) の持續時間だけであり、吾々は此の觀察所與 (obserbational data) に基いて、母集團の平均持續時間を見積り、所與の假定を吟味しようとするのである。

(1)

5,000時間	1個
5,500	3
4,500	2
4,000	2
3,500	1

統計的操作が直面する諸々の問題がこれと同樣の構造を有することは明かである。自然科學・技術の領域に於ては勿論、經濟統計・社會統計の領域に於ても、例えば坪刈りによる收獲高の推算・世調調査などに於て同樣の問題が存することは明かである。吾々は此等の統計の問題に共通な構造を、一層立入つて考えねばならぬ。

明かに此等の問題に於て吾々は、上記の問題例が明瞭に示すように、一方に於ては吾々の部分的な經驗事實から出發して一層一般的に妥當する知識に到達すること、他方に於ては一般的に妥當すべき任務を有する作業假設を個別的な經驗事實によつて驗證すること、このことを以て研究の課題としている。然しかように一般的法則を經驗的事實と結合することは、經驗科學的研究一般に課せられた課題であり、此の點だけに統計の問題の特徴を求めることは出來ない。統計の問題の特徴的構造は、こういう經驗科學一般を通ずる課題がそこで特に如何なる形を呈するか、という點に求められねばならぬ。

此の場合母集團は、完全に同質的なものから成り　從つて一個の個體について分析ずれば、それだけで此の母

集團の中の個體が凡て知悉し得られるというようなものではない。母集團を構成する個體は、相集つて一個の集團を形成し得るに足るだけ、同質的な側面をもつけれども、然し樣〻の偶發的 (accidental) な事情にもとづいてその間になお區別が可能である。此の區別は、例へば世論調査の「或る內閣に賛成かどうか」というように性質的になされることもあろうし、また、上記の問題例に於て電球が持續時間によつて區別されたように數量的になされることもあろうし、更にまた、後の場合に於ても數量的標識が不連續に變る場合と連續的に變る場合とが區別できるが、然し此の區別はどうでもよい。形式的には、區別の標識が連續的に變動する變數で表わされる場合を考えればそれで充分であろう。(他の場合はその特殊の場合と看做して大體差支えないからである)。以上母集團について敍べたと同樣のことは標本についても妥當するであろう。標本は上記のような母集團から抽出された一個の謂わば低次の集團である。從つて、それに於ても上記と同樣の(數量的)標識にもとづく區別が可能である。統計學ではかやうな(數量的)標識を變量 (variate) と呼ぶが、吾〻は此の變量の概念から出發して、次に分布(distribution)及び分布法則 (law of distribution) の概念をのべねばならぬ。

さて、上記の問題例に於て、變量は、明かに、電球の持續時間であるが、此の變量がそれの種〻の可能な數値—例えば五、〇〇〇時間—をとる回數は、上記の(1)が示すように、可能な數値に應じて夫〻異なつている。恰も富の分布によつて夫〻の個人が異つた富を有するように、度數もまた變量がとり得る夫〻の値に分布されて、數値は夫夫異る度數を有するが、これを分布と呼びその樣式を分布法則という。さて、此の分布法則として從來屢〻代表的と看做されたのは所謂正常分布 (normal distribution) である。即ち「變量の大さが x と $x+dx$ との間におちる個體の數が集團全體の大さに比較して如何なる割合を占めるか」が

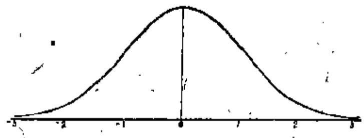
(2)　$$\frac{1}{\sqrt{2\pi}\,\sigma}\,e^{-\frac{(x-\theta)^2}{2\sigma^2}}\,dx$$

で與えられる場合(第一圖參照)である。勿論分布法則がかくの如き形をとる必然性はどこにもなく、分布曲線は種々さまざまの形をとり得るが、然しそれが如何なる形をとるにせよ、兎に角分布法則は、此の正常分布の例が示すように、上記の θ σ の如き若干個の集團構造係數をもつところの、變量 x の函數として表わされるであろう。

(2)に於て變量を正規化 (normalize) して $u=(x-\theta)/\sigma$ と於けば(2)の函數は

(2')　$$\frac{1}{\sqrt{2\pi}}\,e^{-\frac{u^2}{2}}\,du$$

となり、最早 θ σ を含まなくなる。確率論ないし統計理論の教科書は殆んど凡て此の形に於て此の函數の積分の値を表示している。以下、本稿では一部の學者の呼方に依つて此の表を確率積分表と呼ぶことにする。



第一圖　正常分布曲線　$y=\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}$

ここに吾々が假に集團と呼ぶものは、變量(數量的標識)を屬性としてもつ個體の集りであつて、その內部で分布法則が考へられるもの、卽ちオスカア・アンダーソンの所謂 "Gesamtheit" (「數理統計學序說」一九三五年)であるが、明かに、上記の母集團も標本もともに此の意味の集團であり、從つて分布法則が考えられるわけである。ところで此の分布法則を規定する構造係數であるが、以下敍べる理由によつて、統計的研究に於て決定的重要性をもつのは此の集團構造係數であり、然も、此の集團構造係數について母集團のそれと標本のそれとを區別して

考えることが必要である。吾々は母集團の構造係數をパラメーター (parameter) と呼び、標本のそれを統計値 (statistiscs) と呼ぶこととする。直ちに明かなように統計値―例えば平均、標準偏差或は相關係數等―は與えられた標本から直ちにその數値が算出されるものであるのに對して、パラメーターは未知であり、推定する他なきものである。從つて統計値は、觀察所與から算出され此のパラメーターの見積りに役立つ如きものである。此の點後に詳論されるが予め注意を要する。

以上に於て若干の基礎概念が明かにされた。然らば標本にもとずいて母集團を推知するといふことは果して如何なる意味をもつか。吾々は科學的研究作業に於いて統計的操作がもつ意義にまで立歸つてこのことを考えねばならぬ。

リサーチ・ワーカーが統計的操作を試みる場合、彼は如何なる事態に直面しているであろうか。それは要するに觀察所與の簡單化 (reduction of data) ということに歸着するであろう。彼の前には觀察によつて得られた觀察値の集りが置かれている。此等の所與は謂わば數値の無秩序な、雜然たる集合である。研究者は觀察値に影響を與える諸事情を凡て彼の完全な統制下におくことは出來ない。從つて其等の觀察値は直ちに研究者が知ろうとする眞の値 (the true value) ではなく、多かれ少かれ様々の偶發的偏倚を含むものとなる。然も他方に於て研究者はかくの如き雜然たる數値の集りをそのまゝ用いて作業を進めることは出來ない。彼はこれを處理し易い形に變形しなければならぬ。即ち彼は觀察所與から小數の數値を導き出し、此の少數の數値によつて觀察所與が教える一切の有用な知識を表現すること (to express all the relevant informations contained in the mass by means of comparatively few numerical values.) を試みる。觀察所與は偶然的事情に影響されるがゆえに研究者にとつて無用な知識 (irrelevant information）をも含んでいる。此の不必要な知識を排除し、かくて觀察所與が含む有用な知識を餘すところなく抽出すること、かくの如き觀察所與の簡單化――或は濾過――が統計的操作という作業過程に於いてリサーチ・ワーカーが目標とするところである。

さて然らば觀察所與の簡單化、卽ち資料が含む重要ならざる知識を棄却し、重要なる知識だけを抽出するということは、如何なる手續きによつて行われるであろうか、要約的に云えば、統計値標本集團の構造係數を手懸りとしてパラメーターを見積る（estimate）という手續きによつてこのことはなされる。此の點稍立入つて論じることが必要である。

吾々は吾々の經驗ないし理論にもとずいて、母集團の分布法則についてそれが何らかの型―例えば上記の「正常分布」という如き型―に屬すると考え得るであろう。今假に母集團の分布が正常分布の法則に從うとしよう。若し、θ σ というパラメーターの値が確定するならば、母集團の分布構造は完全に知れるわけであり、吾々は、統計的操作に要求されている限りに於て、必要な知識の一切を得たことになろう。然し吾々に與へられたのは母集團ではなく、單に標本にすぎないのであるから、パラメーターの精確な値を知ることは斷念されねばならぬ。寧ろ標本から得られる樣々の統計値を利用して、パラメーターの最も合理的な見積りをつくること、次にまた、吾々はその見積をどれだけの信賴度をもつて利用し得るかを確定し、吾々が標本から引出し得る結論の確かさを明確に意識すること。これだけのことが吾々に課せられ、且つ、許された仕事である。要するに吾々は、パラメーターについて、適確に規定さるべき信賴度をもつた見積りをつくることによつて、母集團についての知識を標本から搾出するのであるが、觀察所與が敎へる有意義な知識はかくの如くにして吾々に把握されるにいたるので

ある。上記の問題例はかくの如き統計操作の課題を明瞭に示すものと云えよう。

以上に關連して尙附言すべきことがある。吾々は母集團の分布法則について單にその數學的形式を豫想するだけでなく、更にパラメーターの數値について何らかの豫想を立てることが出來る。例えば吾々は吾々自身の理論によつて此の大さについて何らかの理論的豫想を觀察所與によつて驗證することは當然行われねばならぬ科學的操作であろう。また、上記の問題例に於ては、謂わば外から與へられた假說の驗證が求められている。勿論時として表面的には何らの假說も設けられぬ場合もあるが、然し此の場合と雖も、觀察所與の完全利用のためには何らかの假說をもつてこれに臨むことが望ましいし、實際また、研究者は暗默のうちに何らかの假說を觀察所與の「拷問」にかけているであらう。兎に角、作業假說が何にもとづいて樹立せられるにせよ、吾々は常に作業假說―所謂「統計的假說」―の驗證のため吾々の觀察所與―標本―を利用している。かように吾々の作業假設が正しいかどうか、また、それがどれだけの確かさをもつかを、標本を通じて分析することを假設適合度の檢定 (test of goodness of fit) と呼ぶ。

統計的操作が直面する問題は、それが適用される分野の多種多樣性にもかかわらず、形式的にはかくの如く類似した構造を示す。理論統計學の課題は、かくの如き問題の共通の構造に著眼して、多樣多樣な適用範圍に共通する武器を研究者の統計操作に供給することにあるであろうが、その問題は、上記の如き問題構造に照應して次の三つの問題群に分つことが出來る。

(一) 特徴付けの問題 (Problem of Specification) ――。上記の如く吾々は先づ母集團の分布法則の型を數學的に選擇せねばならぬ。これが特徴付けの問題である。二項分布、正常分布、ポアソン分布などの理論的度數曲線の理

論、また、K・ピーアソンに於ける度數曲線の體系化の試みの如きは此の問題に關する貢獻と看做されよう。

（二）見積りの問題 (Problem of Esimation)――。或る未知のパラメーターに對してその見積りとして用ひ得る統計値は屢々多數存在する。此等の統計値相互間の優劣を決定する一般的原理を合理的に設定し、此の原理的な觀點から methodically に見積りとして最も適當な統計値を決定することが見積りの問題である。例えば所謂北歐學派に於ける數學的期望値の方法 (Methode der mathematischen Erwartung) は此の問題の一つの解答であるが、フイシャーは一層合理的な解決原理としてマキシマム・ライフリフッドの方法(Method of maximum likelihood) を始めた。

數學的期望値の方法については、例えば、チュプロオ「相關理論の基礎概念と根本問題」第六章、アンダーソン「數理統計學序說」一一八頁―一七〇頁などを見よ。また、マキシマム・ライクリーフッドの方法については、本稿は立入り得ないが、例えば、佐藤良一郎「數理統計學」四七二頁以下、河田龍夫「統計推理の基礎」（「數理統計學概論」所收）十三頁以下などによつてその大要を知ることが出來る。

（三）分布の問題 (Problem of Distribulion)――。吾々は母集團から無數に多くの標本を抽出することが出來る。さて今母集團から無作爲 at random に抽出された標本のそれぞれについて統計値を求めて、此の統計値の集團をつくるとすれば、此の集團内部に於て統計値は一定の法則に從つて分布されるであろう。此の分布法則を純粹に數學的に導出することがこゝに所謂分布の問題であり、標本論は此の問題の解答である。吾々は此の標本論が與える統計値の分布法則を用いて、パラメーターの見積りを合理的に行つたり、統計的假設の適合度を檢定したりすることができる。標本論に於て如何にして統計値の分布法則が與えられるか、また、それが如何にして此等の課題を果すか、此等の點は次節以下に於て明かにされるであろう。

吾々の統計的操作は、かくの如き理論統計學が提供する要具を利用しながら行われるであろう。從來統計的研究は直ちに歸納的研究方法の適用の如く看做され來つたが、以上の說明は寧ろ統計的方法も亦一般の科學的研究に於けると同樣に、演繹と歸納との交替的使用に基くことを示すであろう。吾々は先づ母集團の構造について假設を立て、その內容を精密に規定し、その論理的歸結を、標本論の適用に於て、全く演繹的に確認する。次に此の歸結を觀察された事實と對照する。若し此の二つのものが完全に一致するならば、吾々の假說は――すくなくとも新しい事實が現われない限り――正しかつたわけである。若し完全な一致が與えられぬならば、吾ゝは此の比較を通じて吾ゝの假說の成立可能の程度を標本論の適用に於て吟味するわけである。

二

前節に於て吾ゝはフィシャー的な――從つて現代統計理論の――問題構成を見、統計理論が標本論に課している課題を明かにした。次に吾ゝは此の標本論の內容を明かにし、それが如何にして此の課題を果しているかを概觀しなければならぬ。

以下では、本稿の性質上、所論の範圍を標本の平均及び標準偏差の分布および分布に限るが、此の所謂「スツーデントの分布」は最初 Helmert（一八七五年）及び "Student" の匿名の下に W. S. Gosset（一九〇八年）によつて夫ゝ獨立に發見された（此の間の事情はアンダーソン上掲書九九頁以下などにも記されている）が、その證明を以下敍べるように洗錬したのもまた R. A. Fisher, Applications of "Student's" Distribution, Metron, vol 5 (1926), pp. 90—93. である。尙此の證明は、例えばケンネー「統計の數學」第二部一二八頁以下、北川敏男「小標本の理論」（雜誌「統計數理研究」第一卷第一號昭和十六年）などに於てもこれを見ることが出來る。

今母集團からn個の個體より成る標本を抽出する場合を考え、且つ、夫々の個體に番號を附することとする。即ち先づ第一番目の個體を取出して問題とする標識の大さを見、次に第二番目の個體を取出してこれを觀察し、以下同様にして、第n番目の個體の觀察をもつて標本の抽出を終るわけである。さて第一番目の個體の觀察値をx_1、一般に第i番目の個體のそれをx_iとすれば、ここに吾々は

(3)　$x_1,\ x_2,\cdots\cdots x_n$

という系列を惧る。一般に標本はかくの如き觀察値の系列として考えられる。上記の電球に關する問題例に於ては個體の觀察順序について言及しなかつたが、理論的に云えば、そこに觀察順序が附せらるべきであり、從つて度數分布(1)は(3)の如き、順序づけられた觀察値系列に整理を施したものと考えらるべきである。

さて此の觀察値の系列から

(4)　$\bar{x}=\frac{1}{n}\sum_1^n x_i$;　$s=\sqrt{\frac{1}{n}\sum_1^n(x_i-\bar{x})^2}$

を導出することは周知に屬するが、何れも標本に關するそれであるから、それぞれ標本平均 (sample mean) および標本標準偏差 (sample standard deviation) と呼ばれ、母集團に關するそれと區別される。上記の問題例では、$\bar{x}=4{,}500$;　$s^2=333{,}333$;　$s=$ 約567 であつた。

尙以下では標本に關する統計値はラテン文字で、母集團のパラメーターはギリシア文字で表わす(フィシャーの記號法)。また標本の凡ての個體について總和を取る場合には單に總和記號 Σ だけを記すこととする。

さて、吾々の問題では、母集團は有限個の個體から成るものであつたが、以下では、取扱いを一般的たらしめるため、それは無限に多くの個體から成ると考える。また、解析的處理の便宜のため、變量は連續的に變動する

と考える。從つて標本抽出を幾度となく繰返す間には、循環小數や無理數の觀察値も現われる、と想定するわけである。

上記の通り吾々はここで大さ n の標本しか考えぬのであるが、今此の一定の大さの標本の抽出を幾度となく繰返し、更に、此等の標本の夫々について平均及び標準偏差を算出するとすれば、ここに此等の統計値の集團が得られるわけであろう。實際此の標本抽出を無限に繰返した場合を考えれば、ここに統計値の母集團が成立するわけであり、吾々が吾々の觀察所與から算出した $\bar{x}$ 及び s はその標本――個體一つだけより成る標本――と考えられるが、ところで此の統計値母集團に於て平均及び標準偏差は如何なる分布法則に從うであろうか。これが當面の問題である。

先づ此の問題の解答を敍べよう。

記號的に云えば、$\bar{x}$ 及び s が

$$(5)\qquad \left\{\begin{array}{l} \bar{x}_0<\bar{x}<\bar{x}_0+d\bar{x} \\ s_0<s<s_0+ds \end{array}\right.$$

なる範圍内にある確率、謂わば $\mathrm{Pb}\,[\bar{x}_0<\bar{x}<\bar{x}_0+d\bar{x},\ s_0<s<s_0+ds]$ はどれだけか、此の問題が答えられれば、吾々の問題は解決されたことになろう。今、母集團内部の分布は(2)が示すような正常分布の法則に從うこと（謂わば正常母集團）また、標本の抽出はつねに無作爲（アトランダム）に行われること（謂わば無作爲標本抽出）を假定する。此の假定の下では、母集團の平均及び標準偏差を θ 及び σ とするとき、かのガンマ函數を用いて

(6)
$$\mathrm{Pb}[\bar{x}_0<\bar{x}<\bar{x}_0+d\bar{x},\ s_0<s<s_0+ds]$$
$$=\frac{1}{\sqrt{2\pi}\,(\sigma/\sqrt{n})}e^{-\frac{(\bar{x}-\theta)^2}{2(\sigma/\sqrt{n})^2}}d\bar{x}\times\frac{\left(\frac{1}{2}\right)^{\frac{n-3}{2}}}{\Gamma\{(n-1)/2\}}\frac{n^{\frac{n-1}{2}}}{\sigma^{n-1}}s^{n-2}e^{-\frac{ns^2}{2\sigma^2}}ds$$

となる。以下此の決定的に重要な公式を幾つかの段階に分つて説明することとする。

(6)の右邊は差當り $\bar{x}_0$ 及び s_0 という特定の値に應ずる確率を與えるものである。此の意味に於ては此の右邊の $\bar{x}$ s に添字 0 を附して――その $\bar{x}_0$ s_0 たる特定値をとらしめて――考えることが必要である。然し此の場合 $\bar{x}_0$ s_0 は全く任意の値であり、從つて、(6)の右邊はまた夫々の $\bar{x}$ 及び s に應ずる確率を現わすことが出來、此の意味に於て $\bar{x}$ 及び s の分布を與えるものである。吾々の目的は此の分布法則を與えることであつた。(6)の右邊に於て 0 をおとしたことは、此の分布法則を示すものとしての――夫々の $\bar{x}$ 及び s に應ずる確率を與えるものとしての――(6)の右邊の意味を明かにするに役立つであろう。以下かように、變數の任意の値について成立する性質を記す場合、それが變數の夫々の値について一般的に成立つことを示すために、屢々添字 0 を省略するであろう。

第一段。上記のように吾々は正常母集團の假定と無作爲標本抽出の假定との上に議論を進めるが、先づ此等の假定の意味を明瞭に把握しておくことが必要である。

便宜上、上記の問題例を利用しよう。先ず、正常母集團の假定は、母集團――電球商が買つた電球の全體――に於て $\bar{x}$ と $\bar{x}+d\bar{x}$ との間に持續時間をもつ電球が現われる度數の割合が結局に於て (in the long run) (2) となること、換言すれば、確率密度が、第一圖が示すように、母集團の平均値 θ に等しい持續時間に於て最も大であり、これを遠かるに從つて、漸次減退してゆくこと、を意味するが、正常分布については上記の通り確率積分表が利用出來るから、此の場合には此の假定の意味を一層具體的に示すことが出來る。今の場合 θ σ は勿論未知である

が假に $\theta=5,000$; $\sigma=500$ とすれば、持續時間が夫々の區間に落ちる確率は、確率積分表を用いて、次の如く計算できる

		確率密度	確率
4,950—5,000	5,000—4,050	0.3989	0.03989
4,900—4,950	5,050—5,100	0,3970	0.03970
4,850—4,900	5,100—5,150	0,3910	0.03910
4,800—4,800	5,150—5,200	0.3814	0.03814
⋮	⋮		
4,450—4,500	5,500—5,550	0.2420	0.02420
⋮	⋮		
3,950—4,000	5,000—6,050	0.0540	0.00540
⋮	⋮		
3,450—3,500	6,500—6,550	0.0044	0.00044
⋮	⋮		

此の表は次のことを意味する。持續時間が四、九五〇時間から五、〇〇〇時間までの電球、および、五、〇〇〇時間から五、〇五〇時間までの電球が現われるチャンスが最も多く、それは結局に於て、千回に四〇回の割合で現われる。此の割合は持續時間が五、〇〇〇時間を距るにつれて減小する。平均から標準偏差だけの距離のあるところ、卽ち、持續時間四、四五〇――四、五〇〇および五、五〇〇――五、五五〇の電球が現われるチャンスは千回に約二五回であり、更に標準偏差の三倍だけの距離があるところでは、一萬回に四・四回にまで減小する。上表の意味はかくの如くである、が正常母集團の假定は、夫々の區間の持續時間をもつ電球が、結局に於て、現われるチャンスが、母集團平均からの距離に從つて、かように變化することを意味する。

周知のように、變量が連續的に變動する場合には、確率は區間について云われ、點については云われない。點について云われるのは寧ろ確率密度である。變量一個の場合について云えば、變量 X が x より小なる値を取る確率を $F(x)$ とするとき、此の函

數の x に於ける微係數 $f(x)$ を x に於ける確率密度 (density of probabilit) という。從つて X が x と $x+dx$ との間の値をとる確率 $\mathrm{Pb}[x<X<x+dx]$ は $f(x)dx$ で與えられる。

さて、夫々の持續時間をもつ電球がかような確率で現われるというのは、謂わば自由に放任した場合のことである。ところで吾々はこういう母集團から標本を抽出するのであるが、若し此の場合何らかの偏頗な抽出方法をとつたとすれば、どうであろうか、明かに何らかの癖をもつた標本が現われるに違いない。無作爲標本抽出の假定は、かくの如き偏つた抽出方法をとらぬことを意味する。從つて、標本を抽出する場合にも、持續時間五、五〇〇―五、五五〇の電球は〇・〇二四二の確率をもつて現われる。勿論、個々の標本の中にかういう電球が現れるかどうか、また、どういう度數をもつて現われるか、ということはわからない。然し、標本抽出を幾度となく繰返す場合、「結局に於て」、かういう電球は一千回に二四回の割合を以て現われるであらう。從つて標本の中の任意の――第何番目でもよい――個體の持續時間が區間 $(x,\ x+dx)$ 内にある確率が上記の(2)によつて與えられることになる。これが上記の無作爲標本抽出の假定が意味するところである。

第二段。標本空間 (sample space) 及び標本點 (sample point)。

明かに上記の觀察値の系列(3)は n 次元空間に於ける點$(x_1,\ x_2,\ \ldots\ldots,\ x_n)$で表すことが出來る。以下かような空間を標本空間と云い、かような點を標本點と呼ぶ。母集團から大さ n の標本を反復して抽出する場合、此等の標本は凡て此の標本空間の標本點に一一對應させられる(勿論構造が―個體の番號をも考慮に入れて―全く同じい標本は同一の標本點に對應する)わけであるが、以下では個個の標本を凡てかういう標本點として考える。

此の幾何的構成は標本論に於て極めて重要であるが、それは、此の標本空間が標本平均や標本標準偏差などの

統計値に對して、また、夫々の標本が現われる確率に對して、「場」(field) を形成するからである。吾々は經濟理論の選擇の問題に於て、所有量の組合せを點として考え、此の點の集り――謂わば所有量空間――を效用の場として考えるが、恰もそれと同樣に此の標本空間は統計値及び確率の場を形成するのである。選擇の理論は此の場の中に效用の無差別面を考えたが、それと同樣に、吾々は此の標本空間の中に標本平均および標本標準偏差の「無差別面」――統計理論プロパーの用語では等統計値の isostatistical 面――を考えることが出來る。此の無差別面は當面の問題に對して重要な意義を有するが、それに立入るに先立つて、先ず確率の場としての標本空間を考察しておくのが便宜である。

第三段、標本點 $(x_1, x_2, \ldots\ldots x_n)$ に於ける確率密度。

さて吾々は母集團からさまざまの標本を抽出し得るが、然し此等の標本點はその生起の可能性に於て決して均等ではない。或る標本點は頻繁に生起し得るであらうし、或る標本點は稀にしか生起しえない。さて然らばそれぞれの標本點は如何なる確率密度を有するか。

先づ標本そのものではなく、それを構成する個々の個體について考へよう。上記(第一段)によつて明かなように第 i 番目の個體―例えば第1番目の個體 $(i=1)$―が x_i^0 から $x_i^0+dx_i$ までの値をとる確率は、θを母集團平均とし、σを母集團標準偏差とすれば、

$$(7)\qquad \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x_i-\theta)^2}{2\sigma^2}} dx_i$$

で與へられる。今この明瞭な事實を手懸りとして、特定の標本點 P：$(x_1, x_2, \ldots\ldots, x_n)$ に於ける確率密度を考えよ

う。こゝで吾々は夫々の觀察値が互に獨立であることを前提する。具體的に云えば、例えば上記の問題例に於て、第一球目の持續時間がどれだけであるかは、第二球目の持續時間がどれだけであるかと無關係であることは明瞭な事實であるが、こゝで吾々は吾々が取扱う母集團、從つてまた標本は、凡てかくの如き性質を有すると假定するわけである。此の獨立性の假定の下に於ては x_1 が x_1^0 と $x_1^0+dx_1$ との間の値をとり、一般に x_i が x_i^0 と $x_i^0+dx_i$ との間の値をとることが、同時に實現される確率—卽ち $\mathrm{Pb}\ [x_1^0<x_1<x_1^0+dx_1,\ x_2^0<x_2<x_2^0+dx_2\cdots\cdots,x_n^0<x_n<x_n^0+dx]$ —は、確率論に於ける乘法定理によつて

$$\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{(x_i-\theta)^2}{2\sigma^2}}dx_i \tag{8}$$

である。(ここでは右肩の0を省略した。) 從つて特定の標本點 $(x_1, x_2, \cdots\cdots, x_n)$ に於ける確率密度 D は、$C=\left(\sqrt{2\pi}\sigma\right)^{-n}$ とおけば

$$D=C\Pi e^{-\frac{(x_i-\theta)^2}{2\sigma^2}}=Ce^{-\frac{1}{2\sigma^2}\Sigma(x_i-\theta)^2} \tag{9}$$

で與へられる。然るに $\sum(x_i-\theta)^2=\sum[(x_i-\bar{x})+(\bar{x}-\theta)]^2=\sum(x_i-\bar{x})^2+n(\bar{x}-\theta)^2$ であるから、ここに吾々は

$$D=Ce^{-\frac{1}{2\sigma^2}\left(\Sigma(x_i-\bar{x})^2+n(\bar{x}-\theta)^2\right)} \tag{10}$$

を得る。

第四段。標本平均及び標本標準偏差の「無差別面」。明かに標本はちがつても平均は一致する場合がある。かように平均が等しい標本の全體を考える場合、それに對應する標本點の集りはどういうものとなるか。また標本

が違つても標準偏差は等しい場合がある。かういう標準偏差の等しい標本點の集りは如何なるものであろうか。
先ず前者から考えよう。

今原點を通つて凡ての座標軸に對して等しい傾き(方向餘弦)をもつ直線、即ち

(11)　$x_1=x_2=\cdots\cdots=x_n$

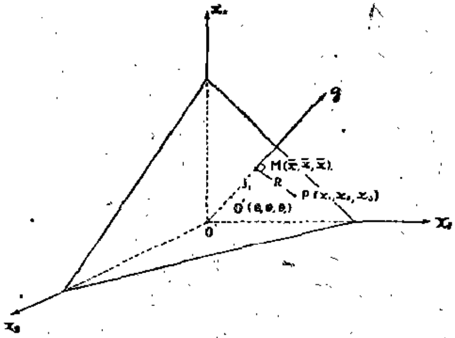
を考える。明かに此の直線の方向餘弦は凡て $1/\sqrt{n}$ に等しい。從つて

(12)　$\vec{n}=(1/\sqrt{n},\ 1/\sqrt{n},\cdots\cdots,\ 1/\sqrt{n})$

なる單位ヴェクトルを考えれば、Oを原點、Mを此の直線上の任意の點とするとき、

(11′)　$\overrightarrow{OM}=\xi\vec{n}$

であり、上記の等方向餘弦直線は此の方程式で定義される。明かにこれに於てξは線分OMの長さであり、Mの位置は此のξの大さによつて定まるであらう。



第二圖標本空間 ($n=3$)

さて次に、此の直線と點M——原點からの距離がξなる點M——に於て直交する平面を考へよう。此の平面の法線ヴェクトルは上記の$\vec{n}$に他ならないこと、また、此の平面が原點からξなる距離を有すること、此の二つの事情からして、此の平面の方程式は、P：$(x_1,\ x_2,\cdots\cdots,\ x_n)$ を此の平面上の任意の點とするとき、

(13)　$\vec{n}\cdot\overrightarrow{OP}=\xi$　或イハ　$(1/\sqrt{n})(x_1+x_2+\cdots\cdots+x_n)=\xi$

で與えられる。此の(13)の後式を標本平均 $\bar{x}$ を與える(4)と比較すれば、直ちに

(14)　$\xi=\sqrt{n}\,\bar{x}$

であること、即ち、ξは標本Pの標本平均の$\sqrt{n}$倍であること、が知れる。Pは此の平面上の任意の點であつたから、此のことは此の平面上の標本點が凡て線分OMの$\sqrt{n}$分の一に等しい標本平均を有することを示す。即ち、此の平面上の標本(點)は凡て等しい平均を有する。同時にまた、此の平面上の點のみが $\xi/\sqrt{n}$ だけの平均を有することも明かである。此の意味に於て吾々は等方向餘弦直線に直交する此の平面を等平均平面 (plane of constant means) と呼ぶ。選擇理論の用語法を用うるならば、それは標本平均の無差別面である。

さて無差別面という以上、指數が考えられる。即ち、それぞれの無差別面に夫々異る指數を與え、これによつて等平均平面を區別することが考えられる。明かに、標本平均 $\bar{x}$、或いは原點Oから平面までの距離 ξ は、何れも此の指數として利用され得るが、然しここでは

(15)　$\xi_1=\sqrt{n}\,(\bar{x}-\theta)$

で定義されたξ_1を此の指數として利用するのが便宜である。これについて若干說明を試みよう。

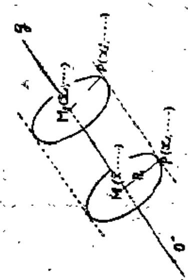
上記のように、線分OMは $\xi=\sqrt{n}\,\bar{x}$ だけの長さを有し、直線OMは何れの座標軸に對しても $1/\sqrt{n}$ の方向餘弦を有するのであるから、點Mの座標が

(16)　$x_1=x_2\cdots\cdots=x_n=\bar{x}$

で與えられることは明かである。さて今、此の直線上に

$$(17)\qquad x_1=x_2=\cdots\cdots\cdots\cdots=x_n=\theta$$

なる點O′を考える。明かに $OO'=\sqrt{n}\,\theta$ であり、從つて $O'M=\sqrt{n}\,(x-\theta)$ 即ち $=\xi_1$ である。これによつて ξ_1 の幾何的意味は明かであるが、ξ_1 がそれぞれの等平均平面によつて異ることも直ちに明かである。從つて、吾々は此の ξ_1 を此の平均の無差別面の指數として利用し得るが、以下O′からの距離が ξ_1 なる平面を ξ_1 平面と呼んでこれを表すこととする。



第三圖

以上のことが明かにされた以上、標本標準偏差の無差別面を考えることは容易である。今標本空間に於て任意に一點 $P_0:(x_1^0, x_2^0,\cdots\cdots x_n^0)$ を取り、此の點 P_0 から上記の等方向餘弦直線(11)に下した垂線の足を M_0、線分 P_0M_0 の長さを R_0、その標本標準偏差を s_0 とする。此の標本空間に於て、それから等方向餘弦直線に下した垂線PMが長さ R_0 に等しいような點Pは、凡て、此の P_0 と同じ標準偏差――即ち s_0 ――を有するであろう。換言すれば、等方向餘弦直線(11)を回轉の軸とし、半經が R_0 に等しい圓壔面上の點は凡て s_0 だけの標準偏差を有するであろう。以下このことを證明しよう。

上記の標本點 P_0 の平均を x_0 とする。明かに點 M_0 は $(x_0, x_0,\cdots\cdots, x_0)$ なる座標を有し、且つ、定義(4)により

$$(18)\qquad s_0=\sqrt{\frac{1}{n}\sum(x_i^0-\bar{x}_0)^2}$$

である。然も線分 P_0M_0 は、P_0 及び M_0 の座標から直ちに明かなように

なる長さを有するのであるから、

(19)　$(P_0M_0)^2=(x_1^0-\bar{x}_0)^2+(x_2^0-\bar{x}_0)^2+\cdots\cdots+(x_n^0-\bar{x}_0)^2$

(20)　$R_0=\sqrt{\sum(x_i^0-\bar{x}_0)^2}$

であり、從つて(18)により

(21)　$R_0=\sqrt{n}\,s_0$

である。さて、今、それから等方向餘弦直線(11)に下した垂線PMの長さがR_0に等しいような點の任意の一つをP: $(x_1, x_2, \cdots\cdots, x_n)$ とし、此の標本の平均及び標準偏差をそれぞれ$\bar{x}$ sとする。Mの座標は上記によつて(16)であるから、直ちに、PMについて(19)から添字 $_0$ を抜いた關係が成立ち、從つてここに

(22)　$\mathrm{PM}=R_0=\sqrt{\sum(x_i-\bar{x})^2}$

なる關係が成立する。これと此の標本點Pの標準偏差sの定義式(4)と組合せれば、明かに

(23)　$\mathrm{PM}=R_0=\sqrt{n}\,s$

を得る、(21)とこれとを比較すれば、點Pに於ける標準偏差sは點P_0に於けるs_0それに等しいことがわかる。かようにしてその標準偏差が或る特定の値sに等しいような標本點は凡て——然もそれのみが——等方向餘弦直線(11)を回轉の軸とし半經Rが

(24)　$R=\sqrt{n}\cdot s=\sqrt{\sum(x_i-\bar{x})^2}$

で與えられるところの圓壔面上にあることがわかる。此の標準偏差の無差別面についても指數を與えて夫々區別することが便宜であるが、以下吾々は此の半經Rを此の指數とし、指數Rの無差別面を單にR圓壔と略稱することにしよう。

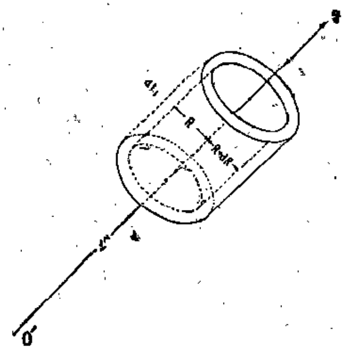
第五段。上記の指數$\xi_1^0 R$を確率密度(10)にとり入れてこれを變形しよう。任意の標本點P：$(x_1, x_2, \ldots\ldots, x_n)$に於ける確率密度は(10)で與えられるが、今、此の點が屬する平均及び標準偏差の無差別面の指數を夫々$\xi_1^0 R$とすれば、$\xi_1^0 R$は夫々(15)(24)で定義されるから、直ちにこれを確率密度(10)にもちこむことができ、ここに吾々は

$$D=Ce^{-\frac{1}{2\sigma^2}(\xi_1^2+R^2)}\equiv D(\xi_1,\ R) \qquad (\text{と置く}) \tag{25}$$

を得る。

第六段。以上のことが明かにされた以上、上記の(6)を證明することは容易である。吾々の問題は標本平均$\bar{x}$及び標本標準偏差sが(5)なる範圍にある確率を求めることであつた。ところで此の問題は、以上の標本空間の構想を利用するとき、次の如く云い改めることが出來よう。

(5)に於ける$\bar{x}_0$及びs_0に應じて$\xi_1^0 R$が定まる。此の$\bar{x}_0$及びs_0に應ずる$\xi_1^0 R$を、即ち$\bar{x}_0-\theta$及び$\sqrt{n}s_0$を$\xi_1^0 R_0$で表わすこととしよう。次に、ξ_1^0平面及び$(\xi_1^0+d\xi_1)$平面なる二つの平均の無差別面とR_0圓壔及び(R_0+dR)圓壔なる二つの標準偏差の無差別面とを考える(第四圖參照)。明かに此の二つの平面及び二つの圓壔は第四圖が示す如き一つの環—高さ$d\xi_1$、巾dRの環—を形成する。さて第四段の説明によつて此の環內にある標本點が何れも而してそれのみに限つて—條件(5)をみたすような標本平均及び標本標準偏差を有することは明かである。即ち、$\bar{x}$及びsが條件(5)の範圍にあるのは此の環內の何れかの標本點が現れる場合であり、然もそれだけに限る。從つて、求むる



第四圖

Pb $[\bar{x}_0<\bar{x}<\bar{x}_0+d\bar{x},\ s_0<s<s_0+ds]$ とは此の環內の何れかの標本點が現われる確率に他ならないのであり、此の後の確率が求められるならば、問題は解決されたこととなる。

さて、此の環內に於てそれぞれの標本點に於ける確率密度は $D(\xi_1^0,\ R_0)$ に等しいと看做され得る。($d\xi_1$, 及び dR は充分に小さいから)。從つて此の環內の何れかの標本點が現われる確率は

$$D(\xi_1,\ R_0)\times(\text{環の體積})$$

で與へられる。卽ち

$$\text{Pb}[\bar{x}_0<\bar{x}_0+d\bar{x},\ s_0<s<s_0+ds]=D(\xi_1^0,\ R)\times(\text{環の體積}) \tag{23}$$

である。然らば、此の環の體積は如何なる大さとなるか。明かに此の環の體積は 底面積×高さ$d\xi_1$ である。從つて底面積が知れるならば、容易に環の體積は算出されるわけである。さて、半徑 R_0 の $(n-1)$ 次元の球の體積は

$$R_0^{n-1}\pi^{\frac{n-1}{2}}/\Gamma((n+1)/2)$$

で與へられる(例えば、高木貞治「解析概論」昭和十三年五六八頁、グールサー「解析學教程」第一卷三六九頁參照)から、此の底面積はその微分に等しい。從つて求むる環の體積は

$$\frac{2R_0^{n-2}\pi^{(n-1)/2}}{\Gamma((n-1)/2)}dR\ d\xi_1$$

で與えられる。

此の結果を(26)にもちこめば、求むる確率は(添字を省略して)

$$(26') \qquad \frac{1}{\sigma\sqrt{2\pi}}e^{-\xi_1^2/(2\sigma^2)}d\xi\times\frac{(1/2)^{(n-3)/2}}{\Gamma((n-1)/2)}e^{-R^2/(2\sigma^2)}\frac{R^{n-2}}{\sigma^{n-1}}dR$$

となるが、更にこれに於て $\xi_1=\sqrt{n}(\bar{x}-\theta)$, $R=\sqrt{n}\,s$, $R^{n-2}=n^{\frac{n-2}{2}}s^{n-2}$, $dR=\sqrt{n}\,ds$ をとり入れるときは、上記の公式(6)に達する

三

以上に於て分布法則(6)の證明という當面の目標は達せられたわけであるが、此の機會にこれから導かれる若干の分布法則を敍べておこう。

明かに分布法則(6)は標本平均 $\bar{x}$ と標本標準偏差 s との同時的分布 (joint distribution) を與へるものである。ところでこれから平均 $\bar{x}$ だけの分布法則及び標準偏差 s だけの分布法則を導くことは極めて容易である。

先づ Pb $[\bar{x}_0<\bar{x}<\bar{x}_0+d\bar{x}]$ を考えよう。これは、標準偏差 s がどうあるかに關係なく、平均 $\bar{x}$ が $\bar{x}_0$ と $\bar{x}_0+d\bar{x}$ との間にある確率である。換言すれば ξ_1^0 平面と $\xi_1^0+d\xi_1$ 平面とに狹まれた、薄板のような、部分空間内に標本點の何れかが現れる確率である。これは (26′) を R について 0 から ∞ まで積分することによつて得られるであらう。今この計算を實際にやつて見ると

$$\int_0^\infty e^{-R^2/(2\sigma^2)}\left(\frac{1}{2}\right)^{(n-3)/2}\frac{R^{n-2}}{\sigma^{n-1}}dR=\int_0^\infty e^{-R^2/(2\sigma^2)}\left(\frac{R^2}{2\sigma^2}\right)^{\{(n-1)/2\}-1}\frac{R}{\sigma^2}dR$$

であり、これに於て $u\equiv R^2/(2\sigma^2)$, $du\equiv(R/\sigma^2)dR$ とおけば

$$=\int_0^\infty e^{-u}u^{\{(n-1)/2\}-1}du\equiv\Gamma((n-1)/2)$$

となるから、求むる確率 Pb $[\bar{x}_0<\bar{x}<\bar{x}_0+d\bar{x}]$ は 0 をおとして

(27)　$$\frac{1}{\sigma\sqrt{2\pi}}e^{-\xi^2/(2\sigma^2)}d\xi_1\quad\text{或イハ}\quad\frac{1}{\sqrt{2\pi}(\sigma/\sqrt{n})}e^{-(\bar{x}-\theta)^2/\{2(\sigma/\sqrt{n})^2\}}d\bar{x}$$

で與えられる。明かにこれはθを平均とし $\sigma/\sqrt{n}$ を標準偏差とする正常分布である。同様にして標準偏差 s の分布法則を求むることが出來る。即ち式(26′)をξ_1についてマイナス∞からプラス∞まで積分すればよい。然るに、(72)の前式(式(26′)の第一因子)をマイナスの∞からプラスの∞まで積分したものは、周知の通り $\sigma\sqrt{2\pi}$ であるから、Pb$[s_0<s<s_0+ds]$ は0をおとして、

(28)　$$\frac{(1/2)^{(n-3)/2}}{\Gamma((n-1)/2)}e^{-R^2/(2\sigma^2)}\frac{R^{n-2}}{\sigma^{n-1}}dR\qquad\text{或イハ}\qquad\text{式(6)右邊ノ第二因子}$$

によつて與えられる。これは最早正常分布ではない。

最後に上記の分布法則(6)ないし(26′)を用いて所謂 t 分布を導いておこう。今

(29)　$$z=(\bar{x}-\theta)/s$$

なる變數 z を考え、先づその分布

(30)　$$\text{Pb}[z_0<z<z_0+dz,\ s_0<s<s_0+ds]\equiv D(z_0,\ s_0)\ dzds$$

を求めよう。このために、先ず

(31) $z_0 \equiv (\bar{x}-\theta)/s_0, \quad d\bar{x}=s_0 dz$

を満足する如き$\bar{x}_0$及び $d\bar{x}$ を考え、次に $\bar{x}_0$, $d\bar{x}$ をかくの如きものとして、上記と同様に、條件(5)を滿足する$\bar{x}$及び s を有する標本點の集り卽ち第四圖の如き環を考えよう。明かにかくの如き標本點の何れも一然もそれのみが $-z_0$ と z_0+dz との間の z、及び、s_0 と s_0+ds との間の s を有することは明かである。從つてかくの如き標本點の何れかが現はれる確率が求むる確率(30)に他ならない。從つて、

(32) $D(z_0, s_0)\, dzds=$上記(6)の右邊

であり、更に(31)をとり入れてこれを書改めれば、こゝに吾々は

$$=\frac{1}{\sqrt{\pi}\,\Gamma((n-1/2))}\left(\frac{1}{2}\right)^{(n-2)/2} n^{n/2}\frac{s^{n-1}}{\sigma^n}e^{-\{ns^2(1+z^2)\}/(2\sigma^2)}dsdz$$

を得る。これが z と s との同時的分布法則である。

次に s と無關係な、z だけの分布を示す法則を求めよう。このためには(32)の最右邊の微分を s について0から∞まで積分すればよい。ところで此の場合

$$\int_0^\infty (1/2)^{(n-2)/2} n^{n/2}\frac{s^{n-1}}{\sigma^n}e^{-\{ns^2(1+z^2)\}/(2\sigma^2)}$$

$$=(1+z^2)^{-n/2}\int_0^\infty e^{-\{ns^2(1+z^2)\}(2\sigma^2)}\left[ns^2(1+z^2)/(2\sigma^2)\right]^{(n/2)-1}\frac{ns^2}{\sigma^2}(1+z_2)ds$$

であるか、これに於て $\{ns(1+z^2)/\sigma^2\}ds=d\{ns^2(1+z^2)/(2\sigma^2)\}$ なることに注意すれば

$$=(1+z^2)^{-n/2}\,\Gamma(n/2)$$

となる。從つてこゝに吾々は

$$\text{(33)}\qquad \mathrm{Pb}\,[z_0<z<z_0+dz]=\int_0^{\infty}(\text{微分式}(32))$$

$$=\frac{\Gamma(n/2)}{\sqrt{\pi}\,\Gamma((n-1)/2)}(1+z^2)^{-n/2}dz$$

を得る(添字の0を省いた)。尚これに於て、$\Gamma(1/2)=\sqrt{\pi}$ に注意すれば、ベーター函數を用いて

$$=\frac{1}{B((n-1)/2,\ 1/2)}(1+z^2)^{-n/2}dz$$

と書くことも出來る。

以上で z の分布が明かにされたが、實際役に立つのは z そのものよりも導ろ

$$\text{(34)}\qquad t=\sqrt{n-1}\,z \qquad \text{即チ} \qquad =(\bar{x}-\theta)\div(s\sqrt{n-1})$$

で定義された t の分布である。此の t 分布は(33)から直ちに導かれる。即ち

$$\text{(35)}\qquad \mathrm{Pb}[t_0<t<t_0+dt]=\frac{1}{\sqrt{n-1}\,B((n-1)/2,\ 1/2)}\left(1+\frac{t^2}{n-1}\right)^{-\frac{n}{2}}dt$$

である。此の函數の數値も現代統計理論の敎程に一般に表示されている(例えばR・Aフィシャー上掲書)が、此の場合、標本が相當大きい(例えば三十以上の個體から成る)ときは上記の分布(35)は正常分布に近づくから、こういう大標本の場合は、所謂「確率積分表」が轉用できる。

四

前三節の證明は現代標本論の內容をつくすものではない。然し、その基本的な點は以上によつて略々概觀し得たであろう。次に吾々は最初の問題例を中心に、此の標本論が實際の統計的處理にどう應用されるかを概說しておこう。これも勿論現代統計理論に於ける標本論の應用を全部的に示すものではなく、「標本論が上記の如きその課題を如何にして果すか」を例示せんがためのものに他ならない。

[I] 上記の問題例にかえろう。此の場合母集團は商人が買うた電球群であり、標本は(1)が示すような分布を有する九個の電球より成つていた。標本平均は四、五〇〇(時間)であるが、問題は此の觀察所與から見て、母集團は平均五、〇〇〇時間の壽命を有する電球より成るか、また、此の觀察所與から見て母集團の眞の平均持續時間はどう見らるべきか、である。

以下の考察を通じて吾々はつねに、(一)母集團が正常分布を有すること、(二)標本の抽出が完全に無作爲(アトランダム)に行われたことを前提しなければならぬ。前前節の說明から明かなように、此の前提の下に於てのみ、上記の標本論は適用可能となるからである。

さて當面の問題に於て母集團の標準偏差は未知である。母集團の標準偏差を未知としながら、然も問題を處理し得る點に現代標本論の特徴があるのであるが、然しこゝでは、最初、標本標準偏差を以て母集團のそれに代用しながら議論を進めよう。卽ち、母集團も、標本と同樣に、五七六(時間)の標準偏差を有することとする。

さて今吾々の母集團は、假設通り五、〇〇〇の平均持續時間を有するとしよう。卽ち $\theta=5,000$, $\sigma=576$ であ

る。さて此の前提の下に於ては、前々節の標本論は、標本平均が四、五〇〇時間以下となる確率を教える。特に標本平均は上記(27)が示すように正常分布法則に從うのであるから、吾々の此の確率を所謂確率積分表によつて直ちに知ることが出來る。實際、今の場合、標本平均の標準偏差 $\sigma/\sqrt{n}$ は一九二(時間)であり、$x=4{,}500$ は $\theta=5{,}000$ から此の標準偏差の約二・六倍距つている。今確率積分表に依つて計算すれば、θ からこれ以上の距離を有する標本平均が現われ得る確率は 0.0047 である。即ち千回のうち約五回という勘定である。

以上は母集團の平均は五、〇〇〇時間という假定だけから全く抽象的に導き出された結論である。然らば此の抽象的推理から吾々は如何なる實際的歸結を導き得るか。

母集團から無作爲に標本を抽出した場合、その平均は母集團の平均と充分異り得る。從つて兩者が一致しないからといつて該貨物が不良品だということは出來ない。此の不一致が大きく、無作爲抽出の場合に生ずる狂い以上になつた場合に始めて此の狂いは「意味を持つ」("significant")わけである。ところで上記の問題例に於ては、若し母集團の平均が五、〇〇〇時間であるとすれば、千回に五回しかおこらないような狂いが現れたのである。或いは此の事實にも拘わらず、此の狂いはチャンスに基くものであり、母集團の平均は、やはり五、〇〇〇時間と看做してよいと考える「暢氣な」人があるかも知れない。然しまた、「狂いがチャンスだけに基いて生ずるということは、千回に五回という極めて稀有なことであるから、此の狂いは最早偶然的事情だけに基くとは考えられない。母集團の電球そのものが平均五、〇〇〇時間の壽命を有しないのだ」と考へる人もあろう。

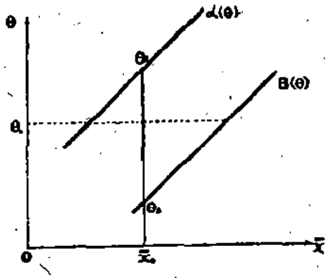
何れにせよ、かくの如き檢定は狂いが有意か否かを調べるものであり、此の意味に於て「有意度檢定」(test of significance) と呼ばれるが、此の有意度檢定に當つて如何なる基準によつて有意、非有意を決するかは全く恣意的

である。然し一般には狂いがそれ以上となる確率が 0.05 以下の場合「有意である」と云い 0.01 以下となる場合には「高度に有意である」(highl significant) と云ている。此の習慣的基準から云えば、此の問題例に於ける開きは「有意」であり、從つて、他の新しい資料が現われぬ限り、母集團の平均壽命は五、〇〇〇時間以下と考えられねばならぬ。

[II] 然らば以上の觀察結果よりして吾々は母集團の平均について如何なる推定をなし得るであろうか。傳統的な理論に於ては、かのベイズ (Bayes) の定理から出發し、パラメーター θ がその可能な樣々の値をとり得る確率―所謂事前確率―がその可能な値に均等に分配されている (事前確率 $g(\theta)d\theta=\text{const.}$) という假定にもとずいて此の問題を解決しようとするが、現代統計理論は此の不合理な構成から完全に離脱しようとする。上記の問題例を利用しながら、現代統計理論に於ける此の問題の取扱いを述べよう。これまでと同樣、母集團の標準偏差が既知であり、然も五七六(時間)であり、標本平均の標準偏差は、從つて、一九二(時間)であるとしよう。さて、今假設に從つて $\theta=5,000$ とするとき、確率積分表によれば、標本平均が 5,394 時間 (卽ち $5,000+2,05\times\sigma/\sqrt{n}$) 以下である確率は 0.98 である。パラメーターの或る値が與へられた場合、かように、統計値がそれ以下の値をとる確率が μ であるとき、此の値をその統計値の μ―パーセンティル (μ-percentile) という。上記に於て、五、三九四時間は $\theta=5,000$ に對する $\bar{x}$ の 0,98 パーセンティルである。明かに、これに對して、$\bar{x}=4,606$ は 0.02 パーセンティルである。ところで若し $\theta=5,000$ であるならば、x は五、三九四時間と四、六〇五時間との間にあると見た場合、謬りをおかす割合は、結局に於て、一〇〇回に四回ということになろう。かように或る危險率 ε とが與へられた場合、$\varepsilon/2$―パーセンティルと $\{1-(\varepsilon/2)\}$ パーセンティルとの間の區間を誤謬危險率 ε を有するパーセ

ンティル區間 (percentile interval with error risk ε) と呼ぶ。

さて以上ではθに特定の値を與へながら考察を進めた。然し、θについて何ら特別の値を考へ得ないというのが一般の場合である。ところで、θに何ら特別の値を與え得ぬとしても、一定の誤謬危險率εが與へられた場合には、吾々は夫々のθの値に應じて $(\varepsilon/2)$ー パーセンティル及び $\{1-(\varepsilon/2)\}$ パーセンティルを決定し得るはずであり、これをθの函數として夫々 $\alpha(\theta)$, $\beta(\theta)$ で示すとき、吾々は$\bar{x}\theta$平面上に$\alpha(\theta)$ 及び $\beta(\theta)$ を表示する曲線を描き得るわけである。さて或る標本よりして、從つて、$\bar{x}$の特定の値$\bar{x}$を得たとする。此の$\bar{x}$に對して $\bar{x}=\alpha(\theta)$, $\bar{x}=\beta(\theta)$ を滿足する如きθの値を夫々θ_1, θ_2 とすれば、かように特定の標本平均$\bar{x}$を得たことから母集團平均が(θ_1, θ_2)なる區間內にあると考えた場合、結局に於て、謬りをおかす危險はεに過ぎないであらう。吾々はかくの如き區間を信賴區間 (Confidence interval) と呼ぶ。現代統計理論はかくの如く、事前確率といふ如きものを考えることとなく、然も、結局に於て (in the long run) の誤謬の頻度を考える仕方に於て、問題を解決するのである。



第五圖

[III] 以上では標本標準平均を以て直ちに母集團の標準偏差と見做して考察を進めた。此の前提の下に確率積分表を利用して議論することは既に傳統的理論が用いた論法であり、此の點については、その概念圖式に若干の改良はあるにしても、現代理論の特徴はあまり多く出てこない。現代理論の特徴は母集團の平均のみならず、標準偏差をも未知とし、然も標本標準偏差の分布が正常でない小標本をも對象として問題を處理するところから始ま

る。次にその處理法の一端を敍べよう。

標本標準偏差は未知であるとしても、前節で敍べた係數tは、母集團平均θについて或る豫想値が與へられるならば算出出來る。實際上記の問題例に於て $\theta=5{,}000$ とすれば $s=576$, $n-1=8$, $\bar{x}=4{,}500$ であるから $\bar{x}=4{,}500$ に應ずるtの値をt_0とすれば

$$t_0=\frac{4.500-5{,}000}{576/\sqrt{8}}=-2{,}5$$

となる。今t分布の表によつて

$$t=\frac{\bar{x}-5{,}000}{576/\sqrt{8}}$$

が絶對値に於てt_0以上になる確率を求めれば $0{,}05$ よりも尙小さい。このことについて上記と同樣の推論を用いると、有意度檢定の慣習的基準を用いる限り、こゝに現われた開きは有意であり、此の母集團の平均持續時間は五、〇〇〇時間以下と見なければならぬことになる。

五

前節で標本論の應用の一端を敍べた。勿論現代統計理論に於ける標本論の應用は以上に盡きないのみか、寧ろ以上はその一小局部に過ぎない。然し、かのカイ自乘分布と假設適合度檢定へのその應用、幾つかの標本が與へられた場合のその統計値（平均、標準偏差など）の違ひの有意度の檢定、所謂ヴェリアンスの分析、更に回歸係數・相關係數などへの標本論の應用等々の廣汎な諸領域に立入ることは此の小稿の能くするところではない。L.R.理

論の解説及び批判といふ當面の目標に必要な範圍では、現代統計理論、特に標本論の内容は明かになつたと思はれるから、次にLR理論そのものに考察を轉ずることにしよう。（以下頁數だけを記しているのは、上記の Levy and Roth 共著の「確率論要説」からの引用である。）

統計的操作を歸納と演繹との結合と見る點に於てLR理論はこれまでのフィシャー的觀點と同樣である。觀察所與が與へられそれぞれについて假設がたてられる。數學的分析は此の假設から論理必然的に結果する歸結を導く。此の歸結は觀察所與によつて吟味され承認あるいは否認される。統計操作をかように歸納と演繹との結合と看做す點に於てLR理論はフィシャーの思想を繼承するが、此の結果として統計操作が含む三つの要素が先づ見出される。第一は、觀察所與としての「標本」であり、第二は「母集團」であり、第三は、母集團の構造を規定すると暫定的に看做される「假設」乃至「假設的法則」である。（一四六頁）

以上に關する限り、LR理論は何ら特色を有しない。然しLRは以上三つの要素の上に第四の要素を考へる。母集團から標本抽出（sampling）が行はれる過程がこれである。母集團の上に標本抽出乃至選出という手續（process of selection or sampling）が加わることによつてここに標本が生れる。母集團をP、選出手續きをS、標本をsとすれば、標本sはPにSといふオペレーターが作用して生れるものであり、記號的には $s=S(P)$ で表現される如きものである。かくの如き「選出過程」という第四の要素を導入する點がLR理論の最大の特徴である。（以上については二七頁以下、一三三頁以下、一四六頁以下參照）然し此の著眼が生産的であるためには、此の著眼自體が標本論の數學的取扱いのなかに具體化されねばならぬが、彼等は如何にして此の選出過程を標本論にとり入れるか。

（以下については一四七頁以下參照）

今問題とする數量的標識——例へば上記の問題例に於ては電球の持續時間——をt、母集團に於けるその分布を$V(t)$標本におけるその分布を$U(t)$とする。選出過程をとり入れる最も自然な方法は、

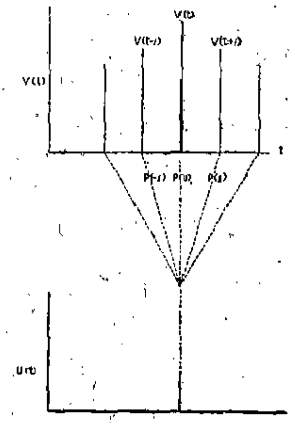
$$U(t)=p(t)V(t) \tag{36}$$

と置く方法であらう。卽ち、標本は母集團から所謂「無作爲」に抽出されるのではない。その理由は如何ともあれ、兎に角母集團のうち或る大さの標識をもつ部分に對してはその大さに應じた或る「重み」を附して抽出が行はれる。かように考へることであらう。

然しＬＲの觀點はこれと異る。容ろ彼等に從えば「觀察者は、例へばt_1に於ける彼の『讀み』に$t_1\pm1$, $t_1\pm2$等々に於ける『讀み』を混入せしめるといふのが、觀測者が實際に行ふところである。觀測者はtの特定値に於ける正確な觀察値を得ているといふ假定にもとづいて行動するから、かように誤つた觀察値を混入するとは彼の統制範圍外にある」(一四七頁)。かようにしてＬＲは或る特定の法則に從つて觀察値の誤つた混入が行はれると假定する。卽ち、觀察者が特定値t_0に於ける觀察値$U(t_0)$と考へるものの中には、勿論$V(t_0)$も入つているが、然しまた、その近傍の$V(t+s)$も$p(s)$だけの割合をもつて混入されており、結局第六圖が示すように

$$U(t)=\sum_{s=-\infty}^{\infty} V(t+s)p(s) \tag{37}$$

なる關係に立つと考へる。勿論問題とするtの近傍を除いては



第六圖

注意深い觀察者にとつては $p(s)=0$ と看做されねばならぬ。尚、彼に於ては、此の $p(s)$ なる函數――選出過程を特徴づける函數――が t に無關係なものとして取扱はれていることが、注意されねばならぬ。

ＬＲ理論はかような形に於て標本論の問題を取扱はうとする。從つて「標本分布 $z(t)$ が與へられ、また、或る選擇の方法 $p(s)$ が前提せられた場合、もとの母集團 $V(t)$ について吾々は如何なる歸結を導出し得るか」これが標本論の問題となる。さて此の問題の取扱ひに當つてＬＲは變量が不連續に變動する場合と連續的に變動する場合とを區別して敍べている。以下この二つの場合について彼の取扱ひを要約的に敍べよう。

さて變量が不連續的に變動する場合については先ずオペレーター E を $Ef(t)=f(t+1)$, $E^s f(t)=f(t+s)$ で定義する。このオペレーターを用いると(37)は

(38) $$U(t)=\sum_{-\infty}^{\infty} E^s V(t)\, p(s)=\sum_{-\infty}^{\infty}\left\{E^s p(s)\right\} V(t)$$

となるが、若し此の無限級數が $\phi(E)$ として表現され得るならば

$$=\phi(E)V(t)$$

となり、從つて

(39) $$V(t)=\phi^{-1}(E)U(t)$$

が此の方程式の一つの特殊解となる。今これに於て定差法の記號をとり入れれば $Ef(t)=f(t+1)-f(t)=\Delta f(t)+f(t)=(\Delta+1)f(t)$ となるが、これを用いるときは $\phi^{-1}(E)\equiv\phi^{-1}(1+\Delta)$ は $A_0+A_1\Delta+A_2\Delta^2+\cdots\cdots$ なる形に展開される。從つて求むる函數 $V(t)$ は所與の函數 $U(t)$ とその定差によつて表現され得ることとなる。特に $U(t)$ が t

の多項式の場合には、或る巾以上の定差は消えるから、便宜である。以上は(37)の一つの特殊解に過ぎない。一般解は此の特殊解と

$$(40)\qquad \varphi(E)V(t)=0$$

の解とから成るであらう。ところで此の(40)の解たる函數は、例えば若干個の任意常數を含み、此の限りに於て恣意的である。從つて一般解は決して一義的には定まらない。即ち、一定の選出函數$p(s)$が前提されたとしても、所與の標本を可能ならしめる母集團としては無數に多くのものが可能である。

ＬＲは(40)の解が或る程度任意となることを若干の例を用ひて示しているが、ここには立入らない。

次に、變量が連續的に變動する場合も取扱ひはこれと同樣である。選出函數 $p(s)$ は、$a\leqq s\leqq b$ 以外では零に等しいとすれば(37)の代りに

$$(41)\qquad U(t)=\int_a^b V(t+s)p(s)ds$$

が與へられ、所與の $U(t)$, $p(s)$ に對してこれを滿足するような $V(t)$ を求めることが問題となる。ＬＲは此の場合については $p(s)$ が正常分布の場合、即ち、h を所謂精度として $p(s)=\pi^{-1/2}he^{-h^2s^2}$ である場合に限つて、二三の例題についてこれを論じている。然しここには立入つて論ぜず、たゞ解が此の場合には前と違つて任意性をもたなくなつていることを注意するに止める。

六

ＬＲ理論の特徴的な點は以上に明かにした如くである。先に明かにしたように、現代標本論は正常母集團及び

無作爲標本抽出の假定から出發するが、これに對して、LR理論が此等の前提から一擧に離脱しようとしていることは注目に値すると言はなければならぬ。然しLR理論は此の試みに於てどこまで成功しているであらうか。卒直に言つて、失敗に終つているように思はれる。以下その理論を批判しながらこのことを明かにしよう。

上記の通り、LR理論の特色は、標本論の第四の要素として選出過程乃至標本抽出過程をとり入れるところにある。此の構想そのものは一應是認されてよいであらう。然しそれを具體化するに當つて彼が用いた選出函數$p(z)$はどうであらうか。此の函數について注目すべきことは、第一にそれが觀察値の謬つた混入を示すこと、第二にそれが變量tと無關係なこと、換言すれば變量tは變つても此の混入を支配する法則は不變とされていること。これである。念のために此の構想を具體的な例を設けて解說すればかうである。今何百人かの壯丁について身長を調べる。此の場合、例へば眞の身長が五尺二寸のもののうち何割かは誤つて身長五尺三寸のものとして取扱はれ、五尺三寸の員數に編入される。同樣にして五尺四寸のものの何割かも五尺三寸のものの員數に編入される。「何割が誤つて他の身長の員數に編入されるか」の法則を與へるものが$p(z)$である。然も此の混入の割合は謂はば混入を供出するものの大いさと混入を受入れるものの大いさとの開きだけで定まり、混入を受入るものの大いさとは無關係である。卽ち身長五尺三寸にかかわらず五尺四寸と見誤られる割合と五尺四寸にかかわらず五尺五寸と見誤られる割合とは等しい、といふ風に考へられる。

此のLRの構成が不合理なことは以上の具體的な例を通じて極めて明瞭である。觀察値が誤讀されるといふ場合は殆どないし、また、若干の誤讀があつたとしてもそれは許され得る誤差の内部に於て生じ別に重大な結果を生ずることはないであらう。LRが如何なる具體的な場合を念頭に置いているかは明かでないが、尠くとも、吾

々が問題とする社會統計・經濟統計に關してはかくの如き誤讀の可能性は少ない。吾々が取扱ふ殆ど凡ての場合を通じて

$$(42)\qquad p(s)=1,\ \text{if}\ s=0;\quad p(s)=0,\ \text{if}\ s\neq 0$$

と考えてよいであらう。ところでかように特殊化した場合、以上のLR理論は如何なる結果をもたらすか。明かに $V(t)=U(t)$, 即ち、標本分布は母集團分布に同じといふ結果である。これ程不合理な結果はないが、その由來する所は、LRが標本選出を誤讀と同一視するの餘り、標本選出の消極的作用、即ち標本選出によつて母集團の個體のあるものは、個々の標本の中に現れ得なくなるといふ事實、を看過することにある。

かようにしてLR理論は、選出函數 $p(s)$ の構造が不合理であるために、最初の注目すべき意圖にもかかわらず、殆んど失敗に終つているように思はれる。實際標本論そのものの實質的展開に於てもLRは上記の觀點から何らの貢獻も殘し得ず、有意度檢定の問題などの取扱ひに於ては、寧ろ從來の——本稿前半が取扱つたような——標本論に全面的に依存している(一六四頁以下)が、此の事實はその企圖の失敗を最も明瞭に示すものと言えよう。

(昭和二十三年一月八日)