一般統計推論から (そのI)*

——信頼度について——

松下嘉米男

(1959年3月受付)

A Topic in Statistical Inference*

——On Weighing Reliability——

Kameo MATUSITA

This is an introductory paper to the problem of weighing reliability of results obtained by statistical methods. Bayes' formula, confidence interval, and fiducial limits are referred to.

現今，色々な分野で使われている統計の理論・方法の推論方式一般を凡て取扱うのが，一般統計推論であるが，その全般について述べることはいずれ他の機会にゆずることにし，本稿においては，その基本的な信頼度の問題について述べ，現代統計理論の持つている一つの問題点に触れようと思う．勿論，ここではそれに対して解答を与えるというわけではなく，ただこれにより，一般統計推論で取扱つている問題の一端でも紹介できたらと思う次第である．

先ず，統計では，その対象・問題に応じて，その対象の性質を適当に抽き出して，取扱い易くした集団を考え，実際の観測値はそのような集団から得られたものと考える，あるいはそのような集団について観測するとしてその方法を定めて観測する．そして得られた観測値から，その集団の性質を使つて結論を出す．この結論は勿論その集団についてのものであるが，それを実際の問題の解答と考えるのである．したがつて，上のような仮想的な集団の適否が，当然結果の精度，あるいはそれがどの程度に信頼するに足るものかという信頼度に大きく響いて来る．尤もこの辺の所は他の自然学科などにも見られる所であるが，統計においては更に独自な問題がある．それは，統計では，一般に斉一でないものの集団の性質を，その一部分を観測して，推し量るのであるから，その結果の絶対確実性というものを振りかざすわけにはいかない．それで，それぞれの方法は，斯々の意味で使えるというようなことをいう．その中で，結果に対する信頼度を‘確率’によつて表わそうとするものがある．これは統計推論においては，非常に基本的で，しかも重要な考えであり，これを唱える方法がまた論議の的となつて来たのである．以下このことについて簡単に述べようと思う．

話を簡単にするため，分布の一つのパラメーター θ を推定することを考える．この分布を今 $F(x, \theta)$ で表わす．このとき，大きさ n のサンプルをとるとし，n 変数の函数で，常に

$$t_1(x_1, x_2, \cdots, x_n) \leqq t_2(x_1, x_2, \cdots, x_n)$$

であるような，適当な t_1, t_2 を考え，観測したサンプル $(x_1, x_2, \cdots, x_n)$ に対し，

$$t_1(x_1, x_2, \cdots, x_n) \leqq \theta \leqq t_2(x_1, x_2, \cdots, x_n)$$

のように θ を推定する．そしてこれに対する信頼度というものを‘確率’で表わそうというので

* Prepared in connection with research sponsored by the Rockefeller Foundation at the Princeton University.

ある.

初めに，θ が色々な値をある確率を以てとると考えられる場合を考える．この分布函数を $\Phi(\theta)$ で表わす．また今，$F(x, \theta)$ はディスクリートの場合を考える．そしてサンプル $x=(x_1, x_2, \cdots, x_n)$ が得られる確率は $p(x, \theta)$ であるとする．そうすると，x が得られたときの，θ が区間 (α, β) の中に入る確率は

$$P(\alpha \leqq \theta \leqq \beta \mid x)=\frac{\int_{\alpha}^{\beta} p(x, \theta) d\Phi(\theta)}{\int_{-\infty}^{\infty} p(x, \theta) d\Phi(\theta)}$$

となる．ここで α, β を x に応じて，適当に定めて，$\alpha \leqq \theta \leqq \beta$ と推定すれば，その信頼度は上の確率によつて表わされる．α, β は一般に x の函数なる故，それぞれ $t_1(x_1, x_2, \cdots, x_n)$，$t_2(x_1, x_2, \cdots, x_n)$ の形に書ける．

これはベイズ (Th. Bayes) の定理と呼ばれて，よく知られたものであるが，これを適用するに当つては，ア・プリオリの分布と呼ばれる $\Phi(\theta)$ について知る必要があり，それがわからないときには，外のことからそれを推定しなければならない．したがつて，このア・プリオリの知識がある場合はよいが，それが無いときには，これを適用するわけには行かない．また，θ を確率変数と考えるわけにはいかない場合は，勿論，論外であるが，如何なる場合にそうであるかは，その時の考え方によるのであつて，客観的にどうのこうのといえるものではない．要は一つのパラメーターを推定するのであるから，そういう風に考えた方が都合がよければ，そうすればよいのである．

上の方法に対して，ア・プリオリの分布 $\Phi(\theta)$ がわからなくても，あるいは，θ は飽くまで未知の常数と考えて，その分布等問題にしなくてもできる方法がある．ここでは，普通この方法で考えられているように，θ は未知の常数として話を進める．それは，先ず α を $0<\alpha<1$ のような数として予め定めたとき，これに対し，n 変数の函数 t_1, t_2 で常に

$$t_1(x_1, x_2, \cdots, x_n) \leqq t_2(x_1, x_2, \cdots, x_n)$$

を満足し，かつ $X_1, X_2, \cdots X_n$ を $F(x, \theta)$ にしたがう独立な確率変数とするとき，θ の値が何であつても，

$$P(t_1(X_1, X_2, \cdots, X_n) \leqq \theta \leqq t_2(X_1, X_2, \cdots, X_n))=\alpha$$

となるようなものを見付ける．そこで観測したサンプル $(x_1, x_2, \cdots, x_n)$ に対し，

$$(*) \qquad t_1(x_1, x_2, \cdots, x_n) \leqq \theta \leqq t_2(x_1, x_2, \cdots, x_n)$$

と推定する．これが普通信頼区間（法）と呼ばれるもので，α はその信頼係数といわれる．（上のような函数 t_1, t_2 の組が，一つの α に対し，幾つか得られるときは，その数だけの信頼区間が得られるわけで，その中より何れを選ぶべきかということが，次の問題として起るが，本稿の話には直接関係はないと思われる故，ここではその問題には立入らぬことにする．）

さて，ここで注意を要するのは，(*) においては，$t_1(x_1, x_2, \cdots, x_n)$，$t_2(x_1, x_2, \cdots, x_n)$ は常数である．それ故，θ を一つの常数と見做す限り，(*) の関係は成立するか，あるいは成立しないかの何れかである．したがつて，観測したサンプルに対して，(*) が成立する確率といつても，それは 0 かまたは 1 とするより外はなく，θ の値がわからない故，(*) が成立する確率を予め規定したり，知つたりするわけにはいかない．上の場合 (*) 自体は α とは直接関係はない．ただ確率を相対頻度として考えるとき，上の方法を何回も適用すれば，正しい結果が得られる率が α に近いというだけである．ところで，一般に，サンプルの値 $x_1, x_2, \cdots, x_n$ が得られたときには，人はその外の値が得られたときのことなどにはあまり関心はない．得られた値に基いて結論を出したいので

ある．これが上の方法の考え方に対する非難の論点である．α というのは，まさに，t_1, t_2 なる函数を用いて推定する方法自体の信頼度である．したがつて，上の結果に対しては，斯々の信頼度を持つた尺度で測つたのであるから，その程度において，それを信用するとするより外に仕方がない．勿論大量のものの受入れ検査のように，同じ方法を何回も繰り返えして行う場合には，上の考え方が大いに役立つわけである．問題はただ一回の，あるいは数少い実験，観測から結論を出そうとする場合である．

上に述べた二つの方法に対して，もう一つの考え方がある．それはフィッシャー (R. A. Fisher) のフィドゥーシャル・リミッツ (fiducial limits) と呼ばれるものである．話をわかり易くするために，更に具体的な例を引いて説明する．今ガウス分布 $N(m, \sigma^2)$ の平均値 m を推定することを考える．ここで σ^2 は予めわかつていないとする．このとき，サンプルを代表する確率変数 (X_1, X_2, $\cdots$, X_n) に対して，

$$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i,$$

$$S^2=\frac{1}{n(n-1)}\sum_{i=1}^{n}(X_i-\overline{X})^2$$

とおき，

$$T=\frac{\overline{X}-m}{S},\qquad \text{ここに } S=\sqrt[+]{\frac{1}{(n-1)n}\sum(X_i-\overline{X})^2}$$

を考えると，これは自由度 $n-1$ の t-分布にしたがう．これを

$$m=\overline{X}-ST$$

と書き換える．次に，観測したサンプル (x_1, x_2, $\cdots$, x_n) に対して，$\overline{X}$, S を計算し，それを $\bar{x}$, s と記す．そこで，自由度 $n-1$ の t-分布にしたがう確率変数 T^* をとり，

$$\mu=\bar{x}-sT^*$$

とおく．この μ は，その値として，m として考えられる値凡てをとる．また μ は，$\bar{x}$ を平均値に持ち，t 分布を少し変えたような分布にしたがう．このように μ を定めた上，予め定めた，$0<\alpha<1$ なる α に対して，

$$P(|T^*|>t_{1-\alpha})=1-\alpha$$

なる $t_{1-\alpha}$ を定め，求める平均値 m は，区間

$$[\bar{x}-st_{1-\alpha},\ \bar{x}+st_{1-\alpha}]$$

の中にあると推定するのである．これが今の場合のフィドゥーシャル・リミッツである．この中に確率変数 μ が入る確率は α である．この α がフィドゥーシャル・プロバビリティー (fiducial probability) とも呼ばれるもので，その結果に対する信頼度を表わすものである．この論法の根拠は $\bar{x}$, s の値が定まつたときに，そのようなサンプルが観測された場合の m はこの辺の値をとつているであろうと思われる程度，あるいは信頼度の分布が，上記の μ によつて与えられると考える所にある．この考え方が如何なる意味において是認されるか，これが問題である．実際，μ の平均値は $\bar{x}$ であつて，$\bar{x}$ を中心とする区間で推定するのであるから，$\bar{x}$ と m は当然近くにあることを期待しての話である．そのために，上のような区間を作るときに使う統計量というのは，サンプルから得られる統計的なことがらは全部含んでいると考えられる，所謂サフィシェント (sufficient) なものを使い，サンプルから知られることがらを総動員して，推定に役立たせようとするのである．それで，例えば，上記 x_1, x_2, $\cdots$, x_n の外に y_1, y_2, $\cdots$, y_m というサンプルが得られたときには，

α を，もはや，$x_1, x_2, \cdots, x_n$ のみより作つた上記区間に，m が入るという信頼度を表わすフィドゥーシャル・プロバビリティーとは考えないようである．この点，信頼区間の方は，外にサンプルが得られようが得られまいが，同じ信頼係数を持つた信頼区間に留る．こんなわけで，フィドゥシャル論においては，サフィシェントという概念が非常に重要になる．なお序乍ら，ここにおいては，偏りが無い（unbiased）などという概念は問題にはならない．むしろ，コンシステント（consistent）という考えの方が大切になる．

さて，上に挙げた例で，信頼区間を考えると，その結果はフィドゥーシャル・リミッツと一致する．勿論，信頼区間では，個々の区間の中に推定すべきパラメーターが入る確率等ということはいえない．しかし，上のように，信頼区間とフィドウーシャル・リミッツが，形の上で一致する場合は割合にある．実際，信頼区間というものはフィドゥーシャル・リミッツの一つの根拠付けであるという見方もある位である．併し，そうはいつても，この両者の考え方は根本的に違つている．そして，次のような場合には，その具体的な解も異つて来るのである．それは二つのガウス分布 $N(m_1, \sigma_1^2)$, $N(m_2, \sigma_2^2)$ で，$m_1, m_2, \sigma_1^2, \sigma_2^2$ が共にわかつていないとき，m_1 と m_2 が同じであるかどうかを調べる問題 (Behrens-Fisher の問題) に関係してである．先ず，$x_1, x_2, \cdots, x_n$ を $N(m_1, \sigma_1^2)$ から，$y_1, y_2, \cdots, y_m$ を $N(m_2, \sigma_2^2)$ から観測されたサンプルとし，

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n} x_i, \quad s_1^2=\frac{1}{n(n-1)}\sum_{i=1}^{n}(x_i-\bar{x})^2,$$

$$\bar{y}=\frac{1}{m}\sum_{i=1}^{m} y_i, \quad s_2^2=\frac{1}{m(m-1)}\sum_{i=1}^{m}(y_i-\bar{y})^2$$

とおく．更に，T_1^*, T_2^* をそれぞれ独立な自由度 $n-1$, $m-1$ の t-分布にしたがう確率変数として，

$$\mu_1=\bar{x}-s_1T_1^*,$$

$$\mu_2=\bar{y}-s_2T_2^*$$

とおき，

$$\mu_1-\mu_2=(\bar{x}-\bar{y})-(s_1\cdot T_1^*-s_2\cdot T_2^*)$$

を考える．そうすると，これの分布は，フィドゥーシャル論の立場からはたやすく計算できる．（実際には次のようにすると便利である．今

$$s_1^2+s_2^2=r^2,$$

$$\frac{s_1}{r}=\sin\theta, \quad \frac{s_2}{r}=\cos\theta$$

とおくと，

$$\mu_1-\mu_2=(\bar{x}-\bar{y})-r(\sin\theta\cdot T_1^*-\cos\theta\cdot T_2^*)$$

とかける．最後の括弧内の

$$\sin\theta\cdot T_1^*-\cos\theta\cdot T_2^*$$

の分布は，T_1^*, T_2^* の分布から計算でき，これに関する数表もできている．）このようにして，$\mu_1-\mu_2$ に対するフィドゥーシャル・リミッツがわかり，上の問題を取扱うことができる．これに対して，m_1-m_2 に対する信頼区間というのは，複雑なものとなる．このようなことは，リグレッション推定をするときにも見られる．

以上述べて来たように，その結果に対する信頼度を‘確率’で表わそうとする推定方法は，何れ

も現在なお議論の余地がある状態である．得られる結果に対して，すつきりした信頼度を与える推定方法を得るためには，信頼度を表わす‘確率’の概念に立戻つて考える必要があるように思われる．実際フィドゥーシャル・プロバビリティーという概念には，普通相対頻度として把握される確率の概念に加わるに他の何物かが含まれているように見える．また，サンプルから得られるものは余す所無く使うという考え，斯々のサンプルを得たればこそ，斯く推定するのであるという考え方など如何なる場合にも役立つものである．これ等のことを考慮しつつ，その結果に，すつきりした信頼度を与える推定方法を考え出すことが，現在一般統計推論における一つの課題である．

統計数理研究所

参考文献

[1] R. von Mises, On the correct use of Bayes' formula, *Ann. Math. Stat.*, Vol. 13 (1942).
[2] J. Neyman, *Lectures and conferences on mathematical statistics and probability*, 1952.
[3] R. A. Fisher, *Statistical methods and scientific inference*, 1956.