経済学研究 *54-2*
北海道大学 2004.9

＜研究ノート＞

初等統計学の授業内容に関する幾つかの願望

園　信太郎

1．はじめに

統計学の入門講義へのささやかな願望を簡略に列挙したい。なお，これらの「個人的な」願望については当然厳しい批判が為され得るであろうが，一般に，「初等的な」教育において重視されるべきなのは，表層的な有用性ではなく，「論理」でなければならないというのが，筆者の判断なのである。

2．Kolmogorov system

Kolmogorov の公理系については，やはり正式に導入すべきである。その際，標本空間上の完全加法族（あるいは σ 集合体）の定義や可測空間の概念を明白に述べるべきである。また，可測空間上の（完全加法的な）確率測度を「確率」と呼ぶことをことわり，確率空間の概念を正式に導入すべきである。また，「Kolmogorov の拡張定理」にも言及すべきであろう。このようにして，「「確率」とは何か，またそれはどうあるべきか」という困難な「問い」に直接に接近するのではなく，数学的な諸形式によって困難な「問い」を迂回し，「確率」に関する解析学を構築する流儀を示しておくべきである。「確率」が，「面積」や「体積」と言うような「測度」の一つとして考察されることとなる。

3．空集合の「確率」と有限加法性

$P(\varnothing)=0$ は，$\varnothing=\varnothing\cup\varnothing\cup\varnothing\cup\cdots$ 及び P の完全加法性より，$P(\varnothing)=P(\varnothing)+P(\varnothing)+\cdots$ となるので，$P(\varnothing)$ が 0 とは異なる実数とすると右辺は発散し，結局 0 以外の値を取り得ないことより従う。ここで，P が実数値を取る関数であることは，いわゆる三公理の前提であり，完全加法性以外の非負性や（標本空間 Ω に関して）$P(\Omega)=1$ となることを持ち出すよりは，無駄が少ない。

互いに排反な「事象」A 及び B に対して，即ち $A\cap B=\varnothing$ ならば，$P(A\cup B)=P(A)+P(B)$ となることは，$A, B, \varnothing, \varnothing, \varnothing, \cdots$ が互いに排反な事象列であることと，完全加法性及び $P(\varnothing)=0$ より従う。これは有限加法性に他ならない。

4．加法法則及び包除原理

「確率」の完全加法性を認めるのならば，上の節で言及したように有限加法性が従うので，「確率」の加法法則は成立する。一方，有限加法性から従う，「任意の事象 A 及び B に対して，$P(A\cup B)=P(A)+P(B)-P(A\cap B)$」を加法法則と呼ぶ流儀もあるが，右辺には加法の他に減法も入っているので加法法則という呼称から「少し」ずれがある。むしろこの公式は，包除原理の基本的な例として言及すべきである。ここで包除原理とは，P を有限加法的な「確率」として，一般に，任意有限個（但し，二個以上）の事象に対して，

$$
\begin{aligned}
P(A_1\cup\cdots\cup A_n)=&\sum_{i=1}^{n}P(A_i)-\sum_{i_1<i_2}P(A_{i_1}A_{i_2})\\
&+\sum_{i_1<i_2<i_3}P(A_{i_1}A_{i_2}A_{i_3})-\cdots\\
&+(-1)^{r-1}\sum_{i_1<\cdots<i_r}P(A_{i_1}\cdots A_{i_r})+\cdots\\
&+(-1)^{n-1}P(A_1\cdots A_n),
\end{aligned}
$$

が従う，というものである。ここで，記号 ∩ は省略した。「証明」は，数学的帰納法によるのが良いと思う。また，右辺の項の総数が 2^n-1 であることも注意すべきである。

5．条件つき確率及び乗法法則

「確率」を「確率」で割るという自明でない操作によって，天下り式に「条件つき確率」の「定義」を行うのは，やはり得策ではない。「定義」の動機づけを明白にするためには，Bruno de Finetti のように，「主観確率」の立場からまず乗法法則を導き，その副産物として「定義」が従うことを示すのが，多分良いであろう。しかし多数派に従って，「$P(A)$ は 0 ではないと仮定して，$P(B|A)=P(A\cap B)/P(A)$ によって「A が与えられている場合の，B の（条件つき）確率 $P(B|A)$ を定義する」」という場合でも，この「定義」と乗法法則との間には段差があることは注意すべきである。ここで乗法法則とは，$P(A\cap B)=P(A)\times P(B|A)$ のことだが，「$P(A)=0$ の場合には，$P(B|A)$ を任意の値 x に各自固定する」と約束しておくと，$P(A)=0$ に対しては，$P(A\cap B)=0$ 及び $0\times x=0$ より，乗法法則は（$0=0$ より）やはり成立する。

6．Bayes の公式

この公式は次の様式で導入するのが得策であると思う。

$$P(A|B)=\frac{P(A)P(B|A)}{P(B)},\ P(B)\neq 0.$$

ここで後半は「$P(B)$ は 0 ではない」という「仮定」である。通常多く見られるのは，分母を「分割公式」で表現したものであり，これは，全事象 Ω を A_i, $i=1, 2, \cdots, n$, によって，個数が 1 以上の有限個の事象に分割すると，

$$P(B)=\sum_{i=1}^{n}P(A_i)P(B|A_i)$$

となることを主張する。ただし，partition formula という呼称は少なくとも Leonard Jimmie Savage が「基礎論」で利用しているが，多数の教科書では，この公式は，「全確率の定理，theorem of total probabilities」などと呼ばれているようである。筆者は，「分割公式」という地味だが本質を突く呼称がはやることを願っている。

なおベイズ・ルールの「証明」だが，$A\cap B=B\cap A$ より $P(A\cap B)=P(B\cap A)$ となること，及び乗法法則より従う。

7．主観確率

例えば高等学校で「確率」を導入する場合には，「確率」の基礎づけは今日でもなお議論が続いていることをことわるべきである。さらにまた頻度論的な流儀のみでなく，「主観確率」の概念をも紹介すべきである。また，「不確定性に直面している「個人」の「合理的な」判断とはどうあるべきか」という「問い」に言及し，「主観確率」がこの「問い」に関ることを指摘すべきである。その際，Bruno de Finetti の議論が役立つであろう。また，大学での統計学の授業では，「主観確率」に基づく統計学が実際に利用されていることに言及すべきである。さらに，「主観確率」に基づくベイズ統計学の他に，「客観的ベイズ推論，objective Bayesian inference」も紹介すべきである。ベイズ接近は今日では広く利用されている手法であり，論理的な議論のみでなく，多数で多様な事例への膨大な応用例が堆積している。

8．確率変数及び期待値

統計学の入門段階では，「標本空間上の実数値を取る関数」として確率変数が導入されることが多いようだが，単なる関数ではなく，「可測な」関数であることを要請すべきである。つまり，既に導入されている可測空間 $(\Omega, \mathfrak{F})$ 上の可測な実数値関数として確率変数を「定義する」という，Kolmogorov system の流儀を

採用するのである。つまり，実数値関数 X に対して，「任意の実数 x に対して，$\{\omega \in \Omega \mid X(\omega) \leq x\} \in \mathfrak{F}\}$」をみたすことを要請するのである。そこで確率変数 X の期待値 $E(X)$ を，確率空間 $(\Omega, \mathfrak{F}, P)$ 上の Lebesgue（ルベーグ）積分で，

$$E(X) = \int_{\Omega} X(\omega) P(d\omega)$$

と「定義する」こととなる。初等的な講義では，受講者に対して Lebesgue 積分の基礎知識を前提とすることには，多分無理がある。しかし，値域を含む集合の方を分割して「近似和」を作り，分割を「細かく」することによって「得られるであろう極限」に言及して，その考え方を略述することはできるはずである。ここは，「イマジネーション」で切り抜けるのである。また，期待値が「存在」しないこともあると注意すべきである。

さらに期待値に関する公式の「正式の」読み方にも注意すべきである。例えば，$E(X+Y) = E(X)+E(Y)$ という公式は，「$E(X)$ 及び $E(Y)$ が共に存在するのならば，$E(X+Y)$ も存在して，この等号が成立する」と「読む」わけである。

また，g を実数直線上の Borel 可測な実数値関数として，$\int_{\boldsymbol{R}} g(x) P^X(dx)$ が存在すると仮定すると（但し，この P^X は X の P に関する確率分布である)，$E(g(X))$ も存在して，これらは等しくなることをも注意すべきである。

9. 独立性の定義

事象の列に関する独立性の「定義」は，「対ごとに」独立ではあっても全体では独立でない例が容易に構成できるので，やはり明白に提示しておくべきである。独立性は P に依存する性質である。事象の列，$A_1, A_2, A_3, \cdots,$ が P に関して互いに独立であるとは，空でない任意の有限部分列 $A_{i_1}, A_{i_2}, \cdots, A_{i_r},$ に対して，常に $P(\bigcap_{k=1}^{r} A_{i_k}) = \prod_{k=1}^{r} P(A_{i_k})$ となることである。より一般的な事象の族に対しても，任意の「有限の」部分族に対して同様の性質を考えれば，問題の「定義」は（形式上は）容易に拡張できる。また，確率変数の列 $X_1, X_2, X_3, \cdots,$ の独立性は，「実数直線における Borel 集合の任意の列 $B_1, B_2, B_3, \cdots,$ に対して，事象の列 $\{X_1 \in B_1\}, \{X_2 \in B_2\}, \{X_3 \in B_3\}, \cdots,$ が（P に関して）互いに独立となる」ということである，と「定義」しておく。

10. 分布関数

確率変数 X の（累積）分布関数 F は「任意の実数 x に対して，$F(x) = P(X \leq x)$」によって「定義」されるが，定義式の右辺を $P(X < x)$ とする流儀もあることを，ことわるべきであろう。また変数 X の分布関数を F_X と表記すると，変数列 $X_1, X_2, \cdots, X_n,$ が互いに独立であることと，任意の実数列 $x_1, x_2, \cdots, x_n,$ に対して，$P(X_1 \leq x_1, X_2 \leq x_2, \cdots, X_n, \leq x_n) = \prod_{i=1}^{n} F_{X_i}(x_i),$ とは同値であることを述べておくべきである。ベクトル値の確率変数と多次元の分布関数とが既に導入されているのならば，この等式の左辺は $F_{(X_1, \cdots, X_n)}(x_1, \cdots, x_n)$ という様式で書けるので，「元の多次元の分布が各周辺分布の積に分解する」ことが，独立であることの必要十分条件であることがわかる。

11. 確率密度関数

確率密度関数を持つ確率変数を連続的と言う流儀が広く行われているが，本来，変数 X が連続的とは，$P(X = a) = 0, \ \forall a \in \boldsymbol{R}$ ということである。初等的な段階においても，「絶対連続」という術語に明白に言及すべきであり，「連続な分布関数が密度関数の積分として表現される」ことの非自明性（「連続」ではあっても「密度」を持つとは限らない）に，注意を促すべきである。また，密度関数の値を Lebesgue 測度が零である任意の集合上で任意に変更しても，得られる分布関数は変化しないことにも言及すべきであろう。

12. 条件つき期待値

これは本来ならば，Radon-Nikodýmの定理に基づき導入されるべきだが，初等的な段階では，通常為されているように，条件つき期待値を導入する前に，例えば，多次元の密度関数 $f_{(X,\,Y)}(x,\,y)$ を周辺の密度関数 $f_X(x)$ で割って得られる $f_{(X,\,Y)}(x,\,y)/f_X(x),\ f_X(x)\neq 0,$ を，「$X=x$ が与えられている場合の，Y の条件つき密度関数」，$f_{Y|X}(y|x)$，を導入することとなるであろう。但し，これらの「定義」が「もっともらしい」ことは，「変数の値が微小部分に属することとなる「確率」」などを利用して説明しておくべきであり，天下り式に「定義」を導入することは避けるべきである。また，「条件つき分布」が導入されると，「条件つき期待値」が「定義」されることとなる。だが一方，Radon-Nikodýmの定理による流儀にも「おおまかに」言及しておくべきである。

13. Chebyshevの不等式

この際この不等式を，

$$P(|X|\geq\varepsilon)\leq\frac{E(|X|^2)}{\varepsilon^2},\ \forall\varepsilon>0,$$

としてしまっても良いのではなかろうか。ここで X は複素数値でも良いし，右辺の期待値が発散しても当然成立する。また証明は，$E(|X|^2, |X|\geq\varepsilon)\geq\varepsilon^2P(|X|\geq\varepsilon)$ に注意すれば，ほとんど自明である。X の期待値が存在する場合，それを μ として，上の不等式の X を $X-\mu$ で置き換えれば，通常の「Chebyshevの不等式」が得られる。「Chebyshevの不等式」を「平均からの隔たりについての不等式」と（固定的に）見てしまうと，かえって「その自明性」が見えなくなるであろう。

なお，$E(X,\,A)=\int_A XdP=\int_\Omega I_A(\omega)X(\omega)P(d\omega)$ は，確率変数 X の事象 A 上の「半期待値，partial expectation」とでも呼べば良いのではなかろうか。

14. 母集団という言葉

これについては，始めは「標本抽出の対象となる集団」というように漠然と規定しておき，議論が進むに従って，目標母集団，抽出母集団，そしてユニバース（universe）というような言葉を，場合に応じて導入して行くのが得策であろう。だが，例えば，小学6年生の男子の全体からなる集団に対して，例えば「身長」を測定すると想定して，この「身長」によって，問題の集団をヒストグラム的に分類することによって得られる集団を，母集団と呼ぶ流儀もある。この場合，元の集団に対して今度は「体重」を測定すると想定するのならば，「異なる」母集団が得られることとなるであろう。この流儀は，母集団分布の概念を「数学的に」導入する際には便利であるかもしれないが，「同一の「母集団」を，異なる研究者が，それぞれの立場から分析する」という表現が事実上できなくなるので，「実際上は」不便である。この流儀では，一人の研究者が，異なる幾つかの指標（例えば，「身長」や「体重」など）を問題とする際には，指標が異なれば「異なる」母集団を問題とすることとなるはずだが，それらの指標からベクトルを構成して多次元の分布を取り扱うとすれば，多次元の「一つの」母集団を問題にしているのだと，弁明ができるかもしれない。しかし，少なくとも初等的な授業においては，「数学的に」再構成された「集団」のみを母集団とするこの流儀は，「母集団」という元来は「実際的な」概念を失念させることとなるであろう。研究者らが問題としている「同じ」集団を，つまり数学的諸形式に組み込まれていない「元の」集団を，彼らが「「一つの」母集団」と呼ぶとしても，決して異様ではなく，むしろ健全である。

さらに，頻度論的な「「仮説」の有意性検定」におけるように，結局いかなる母集団が問題となっているのかが明白ではない場合でも，黙って進むのではなしに，直面している頻度論的な状況を説明した上で，率直に，「この場合の母集団を明確に規定することは（頻度論的には）

今のところ困難である」と述べておくべきである。

15. 事後的有意水準，つまり P 値

正規母集団で分散が既知の場合，その「未知かつ固定されている」母平均が特定の値 μ_0 に等しいかより大であるかを「検定する」状況を想定してみる。これは，論点を明白にするための（状況の）単純化である。母平均に対応するパラメータ（つまり母数）を μ と表現すると，「仮説」（「帰無仮説」という言葉は敢えて用いない）$H_0 : \mu = \mu_0$ を，「対立仮説」$H_1 : \mu > \mu_0$ に対して「検定する」こととなる。検定統計量（この場合通常は標本平均）の実現値 t が，μ_0 よりも「「有意に」大である」のならば，ひとまず「 H_0 は棄却され，H_1 が採択される」と判断されるのである。しかしここで，「「有意に」大である」とはいかなる状況なのであろうか。通常は，「 H_0 のもとでの分布が，t 以上の部分に配分する「確率」（つまり t の P 値）が，「「有意に」小である」」ということである。ここで，どの値を「「有意に」小」とするかは，本来は，「検定」を試みる当事者が「主観的に」定めるはずのことなのである。だが実際には，「 0.05 以下」とか「 0.01 以下」とかが多用される。しかし，このような「有意性」検定の手順には「論理的に」困難があることが，長年（少なくとも 77 年間）にわたって指摘されてきたのである。

なるほど，t が μ_0 から隔たる程，H_0 への「疑念」が強まるであろうが，実際に得られた値は t であり，それよりも大の値ではない。つまり，H_0 は実現値 t のみでなく，それよりも大きい，従って H_0 にとっては「より不利」な，「空想的な値たち」に基づいて「裁かれている」のである。H_0 の「弁護者」は，これらの「空想的な値たち」を組み入れない，実現値 t のみによって条件づけられた「判決」を強く求める「権利」がある。しかも，0.05 という多用される値が，実際には「あてにならない」ことが，「主観確率」とベイズ・ルールとによって長年にわたって指摘されてきたのである。つまり，この値を境界とする検定は H_0 にとっては「不利」であり，本来の「有意性，significance 」を反映しているとは言い難いのである。

このように，「主観確率」に基づく判断と有意性検定による選択とが食い違う状況の（単純化された例による）分析は，Lindley’s paradox の名の下に議論されてきたのである。

16. Behrens-Fisher 問題

これは，二つの正規母集団の母平均の差に関する検定で，未知の母分散の同等性に疑念が持たれる場合への言及である。頻度論的には Welch（ウェルチ）の検定で「近似的に」対応する状況である。また，Welch の検定の様式や考え方が「論理的に」明白ではないとしても，少なくともその結果は述べておくべきである。しかし，（非報知事前分布を利用する）客観的ベイズ推論においては，母平均の差に対する事後分布が（いわゆる Behrens-Fisher 分布として），明白な手順によって導出される。またこの分布は，事前の「主観確率」が漠然としている場合の事後分布の「近似」でもある。つまり，ベイズ接近の立場からすれば，この「問題」は「難問」などではない。採用する方法論に依存するこのような「手順」及び「結論」の違いは，本来は授業中に言及すべきことである。

17. 「誤り」の「確率」

（頻度論的な）仮説検定の枠組では，「二つの過誤」が問題となる。慣例では，「仮説」H_0（「帰無仮説」という言葉は敢えて用いない）が「真である」のにそれを棄却して「対立仮説」H_1 を採択する「誤り」を，「第一種の誤り」とかタイプ I（one）エラーと呼び，「対立仮説」H_1 が「真である」のにそれを棄却して「仮説」H_0 を採択する「誤り」を，「第二種の誤り」とかタイプ II（two）エラーと呼んだりする。さらに，利用する検定統計量 T に関する（元来

は各自が主観的に定める）棄却域を C とし，$\alpha = P(T \in C; H_0)$ かつ $\beta = P(T \notin C; H_1)$ と置く。但しここで $P(\cdot; H)$ とは，「仮説」H が「真である」場合の母集団分布である。ここで注意すべきなのは，α を「第一種の誤りを犯す確率」，そして β を「第二種の誤りを犯す確率」と呼ぶ事がよくあることなのである。しかしここで「確率」と呼ばれているのは，「仮説」H が特定された後に定まる，結局「「母数」が特定された後に定まる」，値のことであり，「誤りを犯す確率」そのものではない。一方，「仮説」の適否を問題としている研究者の「主観確率」を Prob とすると，乗法法則により Prob（第一種の誤りを犯す確率）＝Prob(H_0)×Prob$(T \in C | H_0)$ となる。ここで研究者が Prob$(T \in C | H_0) = P(T \in C; H_0)$ と「見なす」のならば，Prob(第一種の誤りを犯す確率)＝Prob$(H_0) \times \alpha$ となる。また「第二種の誤り」に対しても，同様に，但し Prob$(T \notin C | H_1) = P(T \notin C; H_1)$ と見なして，Prob（第二種の誤りを犯す確率）＝Prob$(H_1) \times \beta$ となる。つまり頻度論では，（未知ではあるが固定されていると想定されている）「仮説」H に対して，「主観確率」Prob(H_0) 及び Prob(H_1) を配分することができないので，これらが欠落している値 α 及び β を，そのまま「「誤り」の「確率」」として強引に導入してしまうのである。「H が与えられている場合の，「誤り」の「確率」」と，本来の「「誤り」の「確率」」とは当然異なるのであり，この差異は（初等的な授業でも）注意されるべきであろう。

18. ヒストグラム

ヒストグラムと棒グラフとは当然異なる。例えば二項分布のような離散的な分布を，いわゆる確率関数によって，棒グラフ的に表現する場合と，（各点を（長さ１の）底辺の中点として，その点に対応する「確率」を底辺上の長方形の面積として表現し，）ヒストグラムとして表現する場合とは異なるのである。ヒストグラムとして表現すると，二項分布に対する「正規曲線による「確率」の近似」を行う際に，なぜいわゆる 1/2（＝0.5）修正を行うと良いのかが明白になるであろう。この方が，曲線下の面積と（問題となっている）「確率」との対応がうまく行くのである。天下り的に修正を行うよりは，「近似」のヒストグラム的状況を説明した方が授業として得策ではなかろうか。

19. 分布関数の収束

変数 X の分布関数 F は，x を任意の実数として，$F(x) = P(X \leq x)$ によって「定義」される。（しかし，この $\leq$ を $<$ で置き換えたものを「定義」として採用する流儀もあると，注意すべきである。）また経験的（empirical）分布関数 $F_n(x;\omega) = \frac{1}{n}\sum_{i=1}^{n} I_{(-\infty,x]}(X_i(\omega))$，（但し，変数列 $\{X_i\}_{i=1}^{\infty}$ は P に関して独立であり，これらは同一分布 F に従うとする）に対して，$F_n(x;\omega)$ は，確率１で $F(x)$ に収束する。但しこの x は，任意に指定されている実数である。だが，この概収束が x に関して一様であるか否かが問題となるはずである。しかし，$\sup_{x \in \boldsymbol{R}} |F_n(x;\omega) - F(x)|$ は ω に関して可測であり，n が限りなく増大するに従って，確率１で０へと収束することが知られている。この事実は，Glivenko-Cantelli の定理と呼ばれることがある。

また一般に，分布関数の列 $\{F_n\}_{n=1}^{\infty}$ が分布関数 F へと法則収束する場合，つまり，F の各連続点 x に対して $F_n(x) \to F(x)$, $n \to \infty$, が成立する場合，F が実数直線上で連続であるのならば，$F_n(x) \to F(x)$, $n \to \infty$, は $x \in \boldsymbol{R}$ に関して一様な収束である。即ち，$\sup_{x \in \boldsymbol{R}} |F_n(x) - F(x)| \to 0$, $n \to \infty$, が従う。これは Pólya（ポーヤ）による定理であり，その証明は困難ではない。

20. 区間推定

区間の端点に対応する確率変数を導入して,「信頼区間」を構成するのだが, この「区間」は, 区間を値として取る変数である。また, その際の「信頼係数」は, 未知固定の母数がこの「信頼区間」に属する（頻度論的な）「確率」に他ならない。一方, この「区間」を構成する各確率変数をその「実現値, realized value 」で「置き換える」ことによって得られる区間も信頼区間と呼ばれ, 実現値が得られるより「まえ」の「信頼係数」が（「信頼係数」という言葉と共に）そのまま流用されるのである。しかし, この「事後の」（つまり, 各実現値が得られた「あと」の）「信頼係数」の解釈は（少なくとも頻度論的には）不明である。また, 実現値に基づく「推論」は,「変数を「その実現値」によって「置き換える」」ことによって得られる「区間」ではなく,「現実の「世界」からもたらされる「その実現値」によって「条件づけられる」確率」によって為されるべきである。とにかく授業においては,「「事後の」信頼係数」の内容が, 少なくとも頻度論的には不明であることを明瞭にことわるべきであり, できれば, 事後分布に基づいて（例えば）区間推定をおこなうベイズ接近の論理的一貫性に言及すべきである。

21. 中心極限定理

これは, 独立で同一の分布に従う変数列に関する古典的なものをまず述べて, その証明の概略を示すこととなるであろう。（また, 古典的な中心極限定理の最も一般的なものとして, Lindeberg-Feller の中心極限定理に言及して良いであろう。）初等的統計学で良く言及されるのは, 同一の分布に従う独立な Bernoulli 確率変数（取る値は０か１かの二つのみ）の列に関するものであり, 二項分布の正規近似とも関りがあるもので, de Moivre-Laplace の定理と呼ばれるものである。

結局, 独立同一分布の変数列で, その分布に正の分散が存在し,「「第 n 項までの和として得られる変数」を標準化（standardization）したもの」を第 n 項とする変数列 $\left\{Z_n\right\}_{n=1}^{\infty}$ に対して,

$$P(Z_n \leq z) \to \int_{-\infty}^{z} \frac{\exp(-\frac{u^2}{2})}{\sqrt{2\pi}} du,\ n \to \infty, \forall z \in \boldsymbol{R},$$

が成立するのである。ここで注意すべきなのは, 上の第 19 節の末尾で言及した「 Pólya の定理」により, この収束は z に関して一様だということである。つまり中心極限定理は, 各点ごとの収束のみでなく, 標準正規分布からの全体的な「ずれ」が, 0 に収束することを主張しているのである。

22. 一致性

点推定を行う際に推定量が導入されるが, それは通常は変数列としての「標本」$\left\{X_i\right\}_{i=1}^{n}$ の関数, つまり統計量であり, ここで n は標本の「大きさ」と呼ばれる。初等的な授業では, しばしば,「母数」θ を指定すると定まる母集団分布 $P(\cdot\,;\theta)$ に関して,「標本」は独立で同一分布に従うと想定されている。そこで問題の推定量が, 未知固定の（真の）「母数」に対する「良い」推定量であるかを, $P(\cdot\,;\theta)$ に基づいて調べることとなり, さらには,「良い」推定量を探査することとなる。しかしここで, 推定量の「良さ」とは自明なものではなく, 異なる個人の間では当然異なり得るのである。頻度論的な接近では,「母数」θ に関する推論を θ の「事後分布」に基づいて組織的に行うことなどは無理である。

「良い」推定量が満たすべき性質としてしばしば言及されるのが「一致性, consistency 」である。だがこれは, 推定量から成る（無限的な）列についての性質である。つまり, $P(\cdot\,;\theta)$ に関して独立で同一の分布に従う変数列 $\left\{X_i\right\}_{i=1}^{\infty}$ を想定し, 各 n に対して $\left\{X_i\right\}_{i=1}^{n}$ は

大きさ n の「標本」であるとし，$T_n = T_n(X_1, \cdots, X_n)$ を θ に対する推定量とする。そこで，推定量の(無限的な)列 $\{T_i\}_{i=1}^{\infty}$ が一致性を満たすとは，$P(\cdot\,;\theta)$ に関して，$\{T_i\}_{i=1}^{\infty}$ が θ に確率収束することであり，つまり，$P(|T_n - \theta| \geq \varepsilon;\theta) \to 0,\ n \to \infty, \forall\varepsilon > 0$, ということである。

しかしこの一致性の規準は，このままでは「常識的に合理的」とは言い難いのである。実際，$\{T_i\}_{i=1}^{\infty}$ が一致性を満たすとすると，$T_i \equiv 0,\ i = 1,2,\cdots,I,\ I = 10^J,\ J = 10^K,\ K = 10^{10}$, と変更しても，一致性は満たされるわけである。

23. 方法論の教授

特定の方法論を教授する際には，その問題点をも率直に述べておくべきである。例えば，「確率」に対する頻度論的見解を保持する場合，その際の統計学的諸手法にはいかなる利点がありいかなる難点があるのかを，折にふれて，しかし率直に述べるべきなのである。このことは，上の第 5，7，14，15，16，17，20，及び 22 節とも関りがある。方法論を教授する際に要求されるのは，やはり「フェアネス，fairness」であろう。

24. おわりに

以上の文章は単なる願望に過ぎない。しかし，統計学の授業を担当することとなった（あるいは「なってしまった」）教師の頭脳を多少なりとも（生産的に）刺激することとなれば幸いである。なお，「統計学の基礎」については，既に「基礎論」 Savage（1954, 1972）がある。また，園（2001 年 12 月）が参考になるであろう。さらに，Leonard Jimmie Savage の「統計学」及び「世界」に対する基本的な態度をうかがうには，園（2004 年 6 月）が役に立つかもしれない。

2004 年 6 月 16 日（水）

参考文献

Savage, Leonard Jimmie, *The Foundations of Statistics*, Wiley, New York, 1954. *Second Revised Edition,* Dover, New York, 1972. これは「基礎論」であり，統計学へのサヴェジ氏の偉大な貢献である。

園信太郎,『サヴェジ基礎論覚書』，岩波出版サービスセンター，東京，2001 年 12 月 20 日。「基礎論」への要約，注釈，及び「読み」を提示している。

園信太郎，「サヴェジ氏による 1971 年の公開講義について」，『經濟學研究』（北海道大学），第 54 巻第 1 号，109（109）- 140（140），2004 年 6 月。