

# 分析科学における確率論と統計学の違い

FUMI 理論研究所

「統計をとる」あるいは「統計を出す」という言葉から推し量れるように、統計解析の対象は実際のデータである。一方、ある事象に対して「確率が高い」と言えば、話者はその事象が発現する根拠を仄めかしている。統計学は実験データなしでは何かを言うことはできないが、それでも、確率論はきっとそれは起こるはずだと主張できる。統計学はある事象の発生機構は問題にしないが、確率論は、その機構が解明されなければ、打つ手がない。

科学はある自然現象の因果関係を定式化し、その因果法則から既知の事象を説明するあるいは未知の現象を予測する。量子力学や統計力学が典型例である。しかし、科学が解明できている因果関係は自然現象の中のほんの一部である。一方、統計学はほとんどすべての科学の問題に対処できる強力な方法論であり、実験データに基づいて仮説の検定を行うことも可能である。

本稿は確率論と統計学の科学における立ち位置の違いを、化学分析の検出限界を例として議論する。一つの話題は測定値の標準偏差をくり返し測定で求める際の問題点であり、統計学でカイ 2 乗分布として扱われている。この概念の基礎として、信頼区間、確率空間（母集団）と正規分布を解説する。もう一方の話題は、測定値の標準偏差をノイズの数学的性質から導出する確率論である。本稿は次の順に話を進める。

1 章 確率論と統計学の一般的な違い ( $n$ 、それとも  $n-1$ ?)

2 章 分析科学への応用 (検出限界を求めるときはどちらを使う?)

図 1 は本稿の要約であり、主張は「確率論は因から果へ、統計学は果から因へ」である。

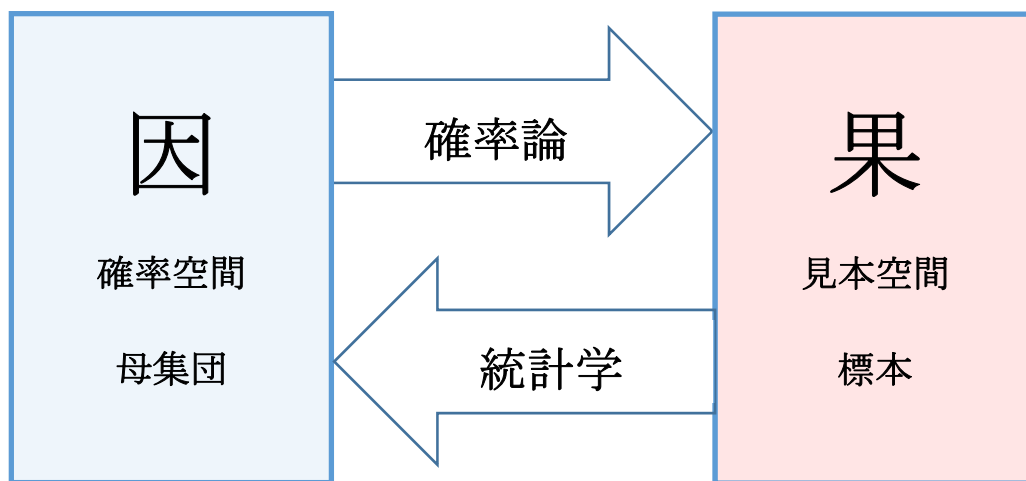


図 1 確率論と統計学の科学における違い

## 1章 確率論と統計学の一般的な違い

### 1.1 確率空間、母集団、標本空間

日本国語大辞典によれば

統計学は「集団に関する資料を整理し、それを特徴づけるいろいろな数値を算出して資料の示すものを知ろうとする記述統計学と、集団の様子をそれから抽出された標本から数学的に推測しようという推計統計学（推計学）の二つからなる」、

確率論は「確率の理論と応用を研究する数学の一分野」

である。

本稿では、統計学を上記の推計学の意味で捉える。つまり、実験データはある集団から抽出されると仮定し、その源の集団の本質を実験データから探る理論・方法論を統計学とする。標本は実験データ、見本、事象、サンプルなどを意味し、標本の母体となる集団は母集団と呼ばれる。標本は母集団からランダムに抽出（サンプリング）されるので、偶然に得られるものであるが、母集団はその標本が実現されるために必要なすべての情報と法則を含んでいる。推定の出発点である実験データは偶然に左右されるとしても、それから推定される母集団は確固とした確率的構造、どのような事象がどのような確率で出現するかという法則を有している。標本は稀な事象ではなく、その蓋然性は高いと信じて、標本から計算した平均と標準偏差を母集団の本質と認めることにする。そうすれば、その本質を足掛かりにして自分が意図している仮説を検定できる。仮説とは、たとえば、ある薬の効き目は確かにある、ある薬の効能は他の特定の薬より高い、ある中学校の能力テストの成績は全国平均と差があるかなどであり、仮説の信ぴょう性は母集団の分布という本質から数値として評価できる。実験データから始まって母集団の本質に至るのであるから、統計学は果から因に向かっていると考えられる（図1）。もちろん、結果はある法則に基づいて原因から生じる。

二点だけ補足しておく。

- 母集団（正規分布）の本質は二つだけの数値で記述できる。なぜならば、正規分布の数学的特質は平均と標準偏差だけで言い尽くされるからである。
- めったに起こらない標本が運悪く出現すれば、この標本に基づいた検定は外れることがある。検定の当たり外れを知ることは現実的には不可能であるから、自分の幸運を信じて疑わないのは随意である。しかし、不運は信頼区間の危険率（%）と同義である（後述）ので、シミュレーション（FUMI 理論研究所のホームページ「エクセルを用いた統計学実習」）で運命を占うことも一計である。

確率論に関する上記の辞書の説明は、色即是空を「色は即ちこれ空」と言っているようなもので、意味が分からない。そこで、理工系学部および大学院生を対象にした確率論の本を覗いてみよう。西尾真喜子著「確率論」（実教出版）の目次は

- 一章 離散型確率空間と確率論のモデル
- 二章 一般の確率空間
- 三章 確率変数と分布

四章 平均値

五章 特性関数

六章 独立確率変数の極限定理

七章 マルチンゲール

八章 マルコフ系列

である。まえがきの「確率とは全測度が 1 の測度である」から察せられることは、確率論は測度論、ルベーグ積分、関数解析などの現代数学を基に築かれていて、統計学とは礎石が異なることである。目次にあるように、確率論は確率空間から出発して、応用である極限定理、マルコフ系列に至る。

中心極限定理を例として、確率論を概観する。中心極限定理は同じ実験を繰り返して得られたデータの和は正規分布になると主張する。たとえば、サイコロを 1 回振って出た目の数の分布はどの目の確率も  $1/6$  の一様分布であるが、2 回振って出た目の和のヒストグラムは中心が盛り上がった山型になる。なぜならば、和 2 と和 12 の度数はどちらも 1 であるが、和 7 の度数は 6 だからである。サイコロを振る数を 2 回からどんどん増やしていくと、その目の和の分布が正規分布に収束するのである。中心極限定理によれば、あらゆる分布が正規分布に行きつくのだから、正規分布も中心極限定理も人間業とは思えないほどの畏怖である。上の説明では、サイコロの目の出現確率はあらかじめ分かっている。確率空間は推定されるものではなく、最初から定義されていて、中心極限定理はこの定義を嚆矢として証明される。こう考えると、確率論は因から果へ向かうことが納得できる (図 1)。

確率論の因は確率空間であり、確率空間はすべての事象の実現を支配する法則である。確率空間は 2 つの要素から成り、一つは標本の定義、他はその事象が起こる確率の定義である。正規分布においては、標本は正規分布の横軸  $x$  であり、確率は確率密度関数から求まる確率 (式 6) である。確率空間はすべての標本を出現させる母体であるから、ここから抽出されるすべての標本の集合を見本空間と呼ぶ。統計学では限られた数の標本から統計量を推定するのであるが、確率論ではすべてを含んでいる見本空間 (または確率空間) で平均と分散を計算するので、後者の値は推定値ではなく、真の値である。

統計学は帰納的な方法またはアポステリオリな方法であり、確率論は演繹的方法またはアプリオリな方法である。これが本項の結語である。科学における演繹的なアプローチの例を挙げてみる。量子力学では、Schrödinger 方程式の解である波動関数  $\varphi$  はある微少領域に粒子が見いだされる確率を表している。統計力学の Boltzmann 分布は系があるエネルギー状態をとる確率を記述している。どちらの確率も物理的な前提から理論的に導出されている。この確率に基づけば、実験データは予測可能となり、たとえば、ある系の状態はエネルギー、力などの変数から計算できる。統計力学は「統計」ではない。

## 1.2 平均と分散

統計学による平均と分散の推定法と確率空間における平均と分散の定義を述べる。平均

と分散の推定値は、それぞれ、 $\bar{x}$ と $s^2$ 、真の値（定義）は $\mu$ と $\sigma^2$ と書く。

統計学で平均を推定する式は

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (\text{式 1})$$

であり、分散を推定する式は

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1} \quad (\text{式 2})$$

である。これらの式は1つの標本に基づいて行う推定であり、この推定の対象、つまり真の平均と分散は次のように定義されている。正規分布の確率密度関数 $f(x)$ （図1）は

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (\text{式 3})$$

であり、平均 $\mu$ は

$$\mu = \int_{-\infty}^{\infty} x f(x) dx \quad (\text{式 4})$$

分散 $\sigma^2$ は

$$\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx \quad (\text{式 5})$$

である。蛇足であると思うが、正規分布の分散の定義は式5であって、式2の $n = \infty$ の極限值ではない。

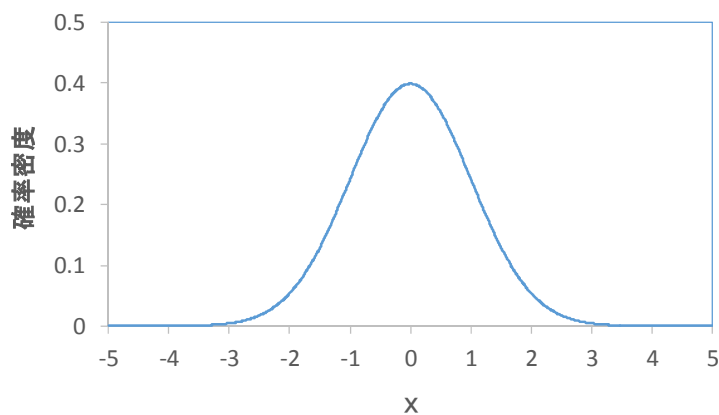


図2 平均0で分散1の正規分布（確率密度関数）

確率密度（式 3 と図 1 の縦軸）は文字通り確率ではない。一般に、密度を体積当たりの質量とすると、密度と体積の積がその体積中の質量となる。確率密度から確率を求める方法も同様であり、ある領域 $dx$ 内の事象 $x$ が出現する確率は、次の式で与えられる。

$$\text{領域 } dx \text{ 内の } x \text{ の出現確率} = f(x)dx \quad (\text{式 6})$$

確率は $x$ を含む領域の長さ $dx$ （微分）と確率密度 $f(x)$ （式 3）の積である。すべての確率の和は 1 であるから、確率密度（式 3）の無限積分（ $-\infty$ から $\infty$ での積分）は 1 である。

平均の一般的な定義は

$$\text{平均} = \int_{-\infty}^{\infty} (x \text{ の関数}) f(x)dx \quad (\text{式 7})$$

である。この定義により、 $(x \text{ の関数} = x)$  とすれば、正規分布の平均（式 4）が導かれる。同様に、 $x$ の関数 $= (x - \mu)^2$ とすれば、分散（式 5）が得られる。平均と分散はどちらも $x$ の関数の平均である。

正規分布の式の単位（ディメンション、次元）について述べておく。 $x$ の単位を[長さ]としよう。確率 $(f(x)dx)$ は定義により無次元であるから、確率密度 $(f(x))$ の次元は $1/[\text{長さ}]$ である。同様に、式 4 は $[\text{長さ}] \times \text{確率}$ であるから、平均の次元は $[\text{長さ}]$ であり、当然、 $x$ と同じ次元である。式 5 より、分散 $\sigma^2$ の次元は $[\text{長さ}]^2$ である。すると、標準偏差 $\sigma$ の次元は $[\text{長さ}]$ となる。分析科学で標準偏差が分散よりも頻繁に使われるのは、その次元が測定データの次元と同じで扱いやすいからである。また、相対標準偏差 $\sigma/\mu$ または $s/\bar{x}$ は無次元であり、こちらは異なった分析方法の精度を比べるときに便利である。

式 1 と式 2 を統計学的式、式 4 と式 5 を確率論的式と呼ぶことにする。同じ概念を記述する式であっても、大きな違いがある。それを次にまとめる。

1. 統計学的式は有限個（ $n$ 個）の和（ $\Sigma$ ）であるが、確率論的式は連続な領域（ $-\infty$ から $\infty$ ）での積分（ $\int$ ）である
2. 確率論的式は確率 $(f(x)dx)$ を含んでいるが、統計学的式は確率を含んでいない。

統計学的式は限られた数の実験データからの推定であるから、 $n$ 個の和（ $\Sigma$ ）は当然であり、一方、確率論的式において $x$ に関する無限積分を行うのは、すべての実現可能な事象 $x$ を考慮するためである。目的が違うので計算法が違うのは当然だろうけれど、そうであっても、確率論の式と統計学の式はかけ離れている。統計学的式が、確率もなしに（ $1/n$ は確率ではない）、無限積分の結果を推定できるのは興味深い。

統計学的式は不偏推定量（unbiased estimator）である。不偏とは、その式を「平均」とすると、真の値（確率論的式、式 4 と式 5）に一致することを意味する。式 1 は平均であるから、不偏推定量は平均の「平均」である。後者の「平均」とは、式 1 を母集団（確率空間）で定義されているすべての事象について平均することである（式 7 の意味）。式 1 は偶然出

現した1つの事象を扱っているのです、これをすべての事象について「平均」すれば、「平均」値が式4の値 $\mu$ に一致するのである。具体的に述べる。ある $n$ 回の実験から式1の平均が得られたとしよう。この実験を1回のシリーズ実験と呼ぶ。 $N$ 回のシリーズ実験から、1式の平均が $N$ 個得られる。この $N$ 個の平均の「平均」は、1回のシリーズ実験の平均(式1)よりも、真の値 $\mu$ に近いと考えられる。そこで、 $N$ を無限大にすれば、式1の「平均」が式4の値に一致するのである。不偏推定の意味は $n$ を無限大にするのではない点に注意する必要がある。試しに、 $n=1$ として、式1(1つの標本)を「平均」(すべての標本についての平均)すれば、これが $\mu$ に等しいことは直観的に分かるだろう。分散(式2)の「平均」も同様に解釈できる。

平均あるいは分散の「平均」は信頼区間と同じ意味を持つ。次章で信頼区間を扱うので、ここで前もって議論しておく。 $n$ 回の実験から式1を計算し、 $t$ -分布を使うと、平均値の信頼区間が得られる。95%信頼区間は真の値 $\mu$ がこの信頼区間に含まれる確率が95%であると直感される。同じ $n$ 回の実験(1シリーズ)を100回くり返して(100シリーズ)、100個の95%信頼区間を得たとしよう。95%とは、95個の信頼区間が真値 $\mu$ を含み、5個の信頼区間が真値含まないことである(エクセルでのシミュレーションは次を参照：<http://www.fumi-theory.com/img/YH-Stat3.pdf>)。しかし、現実には、シリーズ実験を100回くり返すこと(合計 $n \times N$ )は不可能であるから、1個の信頼区間から、この区間内に $\mu$ が含まれる確率は95%であると言うのである。しかし、実験者には、1個の信頼区間が95%であろうが90%であろうが、それは $\mu$ を含んでいるかいないかの、二つに一つでしかない。

信頼区間の概念は宝くじに似ている。現実的ではないが、何かしらの賞に当たる確率が1%の宝くじを想定しよう。1枚の宝くじは1%の可能性があると云っても、その所有者にとっては当たりか外れのどちらかでしかない。しかし、この宝くじを100枚持っていれば、1枚当たるだろうから、1枚の宝くじに対して、1%の当選確率であると言ってよい。このように宝くじは信頼区間と似ている。ところが、大きな違いが一つある。それは、1枚の宝くじの当たり外れは抽選会で知らされるが、実験データの信頼区間の当たり外れは有限の実験からは証明できないことにある。実験データの信頼区間は抽選会のない宝くじである。そこで、科学は他の方法で抽選会を開催する。

確率空間はすべての定理および標本が導き出されるおおもとであるから、すべての結果の因果法則を含んでいる。確率空間から鳥瞰すれば、不偏推定量と信頼区間は平坦である。

### 1.3 サイコロ実験

サイコロの目の分散を計算する2つの式を比べる。1つの標本に対しては式2を使うが、確率空間では分散は次の式で与えられる。

$$\sigma^2 = \sum_{i=1}^n (x_i - \mu)^2 \frac{1}{n} \quad (\text{式 8})$$

式 2 と式 8 の違いは  $n - 1$  と  $n$  だけでなく、 $\bar{x}$  と  $\mu$  にもある。

サイコロを 6 回振る実験で、偶然に 1 から 6 の目が一回ずつ出たとしよう。平均は 3.5 であるから、統計学による分散の推定値 (式 2) は

$$s^2 = \frac{(1 - 3.5)^2 + (2 - 3.5)^2 + (3 - 3.5)^2 + (4 - 3.5)^2 + (5 - 3.5)^2 + (6 - 3.5)^2}{6 - 1} \quad (\text{式 9})$$

一方、式 8 の計算は次のように行う。

$$\sigma^2 = (1 - 3.5)^2 \frac{1}{6} + (2 - 3.5)^2 \frac{1}{6} + (3 - 3.5)^2 \frac{1}{6} + (4 - 3.5)^2 \frac{1}{6} + (5 - 3.5)^2 \frac{1}{6} + (6 - 3.5)^2 \frac{1}{6} \quad (\text{式 10})$$

式 9 の  $\bar{x} = 3.5$  はたまたまの値であるが、式 10 の  $\mu = 3.5$  は真値である。サイコロ実験の確率空間は、事象は 1 から 6 の整数、確率はそれぞれの事象で  $1/6$  と定義されている。確率空間での定義 (式 7) (積分を和で置き換える) を使って求めた  $(x - \mu)^2$  の平均が式 10 である。すると、式 9 の分母にある  $1/(6-1)$  は確率空間で定義した確率でないことが分かる。事実、サイコロを振る数をどんどん増やせば、 $1/(n-1)$  は確率  $1/6$  からますます離れていくが、それにもかかわらず式 9 は真の値にぐんぐん近づいていく。

分散の計算は分母が  $n - 1$  あるいは  $n$  の二通りがある。くり返し実験の数を  $n = 6$  と固定しよう。シリーズ実験から得られる  $N$  個の分散の「平均」は、式 2 (分母は  $n - 1$ ) を使えば、 $N$  の増大とともに真値  $\sigma^2$  に収束する。このとき、分母を  $n$  とすると (式 8)、上と同じシリーズ実験からの「平均」は真値  $\sigma^2$  には収束しない。これが不偏推定の意味である。一方、実験の確率空間 (すべての事象と確率  $1/n$ ) が既知ならば、分母は  $n - 1$  (式 2) ではなく、 $n$  (式 8) とすることが正しい。

## 2 章 分析科学への応用

本章は確率論と統計学の立場の違いを、分析機器の検出限界 (Detection Limit, DL) を例に挙げて議論する。以下で述べるように、検出限界を求めるための実験的な要求は平均と標準偏差 (SD) の推定である (式 11)。くり返し実験をすれば、これには何も問題はないと思われるかもしれない。確かに平均を精確に求めることは難しくはないが、SD の推定は厄介なのである。少数のくり返し実験から求めた SD の信頼性は高くない (ばらつく) ので、精度向上のためにくり返し数を増やすと、コスト、時間、労力などの実用面が障害となるからである。

DL は分析法の最小検出量あり、測定値のバラツキを考慮して次のように定義される。

$$y_D = y_0 + k\sigma(y_0) \quad (\text{式 11})$$

ここで、 $y_0$  と  $\sigma(y_0)$  は、それぞれ、ブランクサンプルのくり返し測定値の平均と SD である。ブランク平均  $y_0$  をゼロとすれば、DL はブランク標準偏差  $\sigma(y_0)$  の  $k$  倍となる。

DL は異なった分析機器にも適用される共通語であるから、必ず、分析の最終目的量（重さ、濃度など）の単位で記述される。最初に測定値の検出限界  $y_D$  を算出し（式 11）、次に検量線によって、これを最終目的量の単位での検出限界  $x_D$  に変換する。検出限界の統計学的な側面については拙著（FUMI 理論研究所のホームページ「検出限界、判定限界、定量限界の理論と事例」）に詳述してあるので、本稿では省略する。

くり返し測定によって得られる SD 推定値のバラツキの程度を見てみよう。図 3 は SD 推定値の 95%信頼区間のくり返し数  $n$  への依存性を描いている。この図はタイトル下のカイ 2 乗分布の式によって作成されている。くり返し数が 6 の場合、95%信頼区間の上限と下限を見れば、SD 推定値は真の値 (=1) の  $\pm 60\%$  の範囲にばらつくことが分かる。つまり、100 回のシリーズ測定（合計  $6 \times 100$ ）から 100 個の SD 推定値を得れば、そのうち 5 個は真の値の  $\pm 60\%$  の範囲から外れることになる（エクセルでのシミュレーションは次を参照：<http://www.fumi-theory.com/img/YH-Stat4.pdf>）。二つの SD 推定値が、2 倍くらい違っても不思議ではないくらいのバラツキである。しかし、くり返し数を大幅に増やせば（30 から 50 回）、95%信頼区間は  $\pm 20\%$  となるので、SD 値のバラツキは 1.5 倍以下には十分に抑えられるだろう。

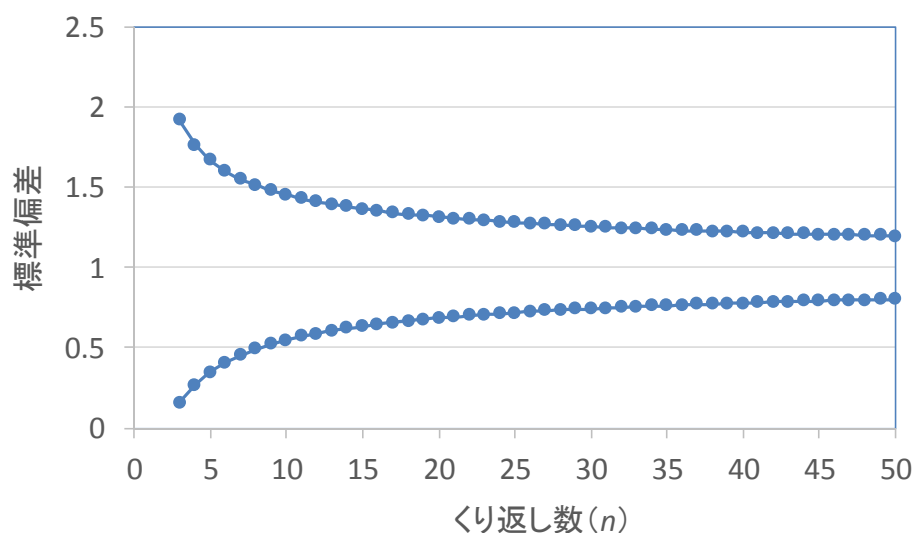


図 3 くり返し実験による SD 推定値の 95%信頼区間

95%信頼区間は分散の信頼区間の式

$$\sqrt{\frac{\chi_{\beta}^2}{n-1}} \geq \frac{s}{\sigma} \geq \sqrt{\frac{\chi_{\alpha}^2}{n-1}}$$

から作成した。 $\chi_{\alpha}^2$  と  $\chi_{\beta}^2$  は、それぞれ、カイ 2 乗分布の確率 0.025 と 0.975 の  $\chi^2$  値である。



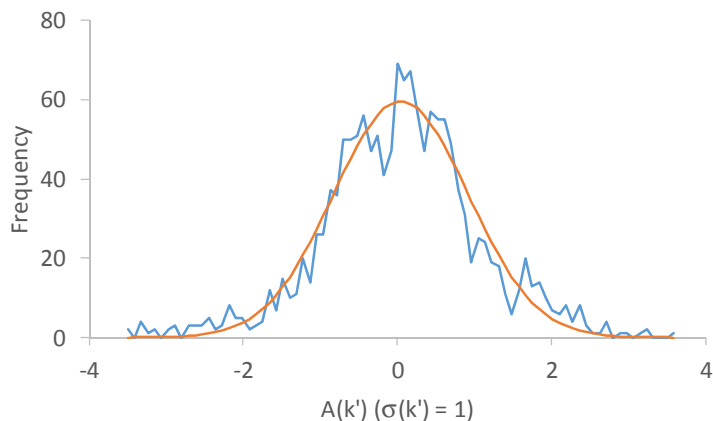
式 11 から分かるように、この SD 推定値のバラツキはそのまま DL のバラツキに反映される。くり返し数が 6 であれば、DL が 2 倍くらい違ってても不可思議ではないが、この DL の差異が重大な社会問題を引き起こすかもしれないという懸念はある。DL のバラツキはせいぜい 1.3 倍くらいに収めたい。ところが、そのためには、少なくとも 30 回のくり返し測定が必要であるが、コスト、サンプル量、時間、分析者の労力などの点から、これはほぼ実現不可能である。カイ 2 乗分布によれば、くり返し測定から SD 推定値を求める限り、推定値の統計的信頼性を上げることとくり返し数を減らすことは二律背反であるから、現実問題と DL の信頼性を勘案した妥協点を探さなくてはならない。カイ 2 乗分布は数学の理論であり、断じて抜け出すことはできない。

ならば、カイ 2 乗分布を迂回し、バラツキの原因から SD をダイレクトに記述しようという動機が生じる。分析機器の測定誤差の原因は意外に単純である。つまり、分析機器の測定値のバラツキの主原因はノイズである。HPLC 測定の場合、分析対象物質を含むサンプルをだんだん希釈していくと、目的ピークがしだいに小さくなり、ついにはバックグラウンドノイズに埋もれてしまう。DL はピークとノイズが区別できる限界である。面積測定の場合、測定値はピークの端から端までの領域におけるピークの積分値（和）であるから、測定誤差はノイズがこの領域で作る面積に相当する。

ノイズが作る面積の SD を求めるためには、まず、ノイズを数学的に記述する必要がある。確率過程論を利用すれば、分析機器のノイズはホワイトノイズと AR(1) (first order autoregressive process、マルコフ過程の一種) という二つの確率過程の和として近似できる。ただし、これらの確率過程に含まれているパラメータ（三個）を現実のノイズから決定する理論（parametrization）が必須である。以前はフーリエ変換がパラメトリゼーションに利用されていたが、最近他の方法が提案された。この SD 推定理論は FUMI 理論（FUnction of Mutual Information の略）と呼ばれていて、現在では、ISO の国際規格として採用されている（ISO 11843-7）。

FUMI 理論はホワイトノイズと AR(1)に基づいているので、測定誤差またはノイズが作る面積は正規分布に従うと仮定されている。図 4 は実際の HPLC のバックグラウンドから得られた測定誤差のヒストグラムである。ギザギザの線が観測値であり、滑らかな線がパラメトリゼーションによって決定された理論線である。これらの一致の良さから、例として取り上げた HPLC 測定系の測定誤差の分布は正規であり、FUMI 理論の前提条件を満たしていると結論できる。この正規分布の平均と標準偏差が測定系の確率空間を構成するので、この確率空間から演繹的に任意のピークの検出限界が予測できる。

図 4 の実験条件では、クロマトグラムの長さは 10 分であるので、6 回のくり返し測定から SD を求めるには 60 分必要であるが、FUMI 理論に必要なデータの収集は 5 分で足りる。しかも、FUMI 理論から求めた SD の統計的信頼性は、ノイズの確率論的性質やピークの幅にもよるが、30 から 50 回のくり返し測定に匹敵する。



**図4 HPLCのバックグラウンドノイズが作る面積（測定誤差）のヒストグラム**  
**SD = 1** として表示。「検出限界、判定限界、定量限界の理論と実例」から転載。  
 全度数は約 1500、測定時間は 5 分である。

統計学の式 2 をサイコロの分散の推定に使うことはありえないだろう。なぜならば、サイコロの目は 1 から 6 の自然数であって、それぞれの出現確率が分かっているのだから、確率空間が構築できるからである。既存の知識を最大限に活用して、正確な結果を確率空間（式 8）から導き出すことが在るべき在り方である。知らぬ顔の半兵衛を決め込んで、式 2 を使えば、最良の結果は得られない。式 2 はもともとサイコロの目も確率も知らないときに使う式であるから、例えば、歪んでいる二十面体のサイコロにも有効であるほどの広帯域のスペクトルをもつが、残念ながら、推定はぼんやりした玉虫色である。一方、狂いのない精度推定を謳う FUMI 理論は融通がきかない。ホワイトノイズと AR(1) で記述できないノイズにはお手上げだからである。そうではあっても、不確かさの因果関係を包含した確率空間の適用範囲内においては、くり返し測定に依らないで標準偏差を推定する方が、実践的には実り豊かだろうし、理論的には透徹している。

林 譲  
 令和元年 5 月 1 日