

分析値の「不確かさ」の意味と FUMI 理論による評価

第一執筆者： 東京薬科大学 分析化学教室 小谷 明

第二執筆者： FUMI 理論研究所 林 謙

はじめに

国語辞書で「ふたしか」を引くと「確かでない」とあった。確かにといいながら、Oxford の辞書で「uncertain」を調べると、「feel doubt; likely to change」とあった。分析値を見ると、次は変わるだろうと思わず疑ってしまうのは、それが **uncertain** だからなのだ。分析科学では、分析値の不確かさの度合いは標準偏差 (SD) で表す。本稿は、不確かさが色濃く反映されている検出限界、判定限界、定量限界、精度プロファイル (精度対濃度プロット) を解説し、SD を推定する方法——くり返し実験と FUMI 理論——を比較検討する。

1. 検出限界と判定限界

ターゲット物質が検出されたかどうかを判定する基準は判定限界であって、検出限界ではない。以下で、ISO¹⁾ と JIS²⁾ による検出限界と判定限界の定義を紹介する。

1.1 判定限界の必要性

クロマトグラフィーにおいて、ターゲット物質の濃度をどんどん下げていくと、そのピークはしだいに小さくなって、しまいにはノイズと見分けがつかなくなる。そこで、信号雑音比 (S/N) を利用して、S/N=3 のピークを検出限界測定値 y_D 、 y_D を検量線 ($y=f(x)$) で変換した量 x_D を検出限界濃度と定義しよう。ある分析の測定値が y_D より十分に大きければ、濃度 x_D のターゲット物質は確かに検出されたと判断できそうである。問題は、測定値が y_D と同じくらいのときに、濃度 x_D のターゲットを何%の確率で検出できるかである。

ターゲットの検出が今の問題であるから、濃度 x_D のターゲットを同じ条件でくり返し測定する実験系を想定しよう。測定値 y が $y < y_D$ ならば「非検出」、 $y \geq y_D$ ならば「検出」という規則を設ければ、100 回のくり返し実験からは、約 50 の非検出と約 50 の検出が得られる。検出の観点からすれば、これではコイン投げの確率と変わらない。

50% よりずっと高い確率、たとえば 95% で検出したい。そこで、上の判定基準を y_D より小さい値 y_C に変更しよう。つまり、検出限界 (x_D と y_D) はそのまま、基準 y_C を新たに設定する。濃度 x_D のターゲットを検出するための新しい規則は「測定値 y が $y < y_C$ ならば非検出、 $y \geq y_C$ ならば検出」である。 y_C を判定限界 (decision limit) または限界値 (critical value) と呼ぶ²⁾。

1.2 ISO と JIS による検出限界と判定限界の定義^{1,2)}

図 1 の右側の曲線は検出限界濃度 x_D のサンプルをくり返し測定するときの測定値 y の分布(正規)である。この分布の平均は y_D であり、検量線で x_D に対応する。測定値 y が判定限界 y_C より小さい値を取る確率を β とする。図 1 では $\beta=5\%$ である。濃度 x_D の物質のくり返し測定を想像すれば、判定基準を y_D から y_C まで下げると、ターゲットの検出確率は 50% から 95% まで上がることが分かる。ところが、新たな問題が生じる。ブランクサンプル(ターゲットの濃度がゼロのサンプル)の測定値が判定限界 y_C より大きな値をとる可能性を否定できないことである。これは誤判定につながる。ブランクサンプルの測定値 y の分布の平均を y_0 、 y が判定限界 y_C より大きな値をとる確率を α とする(図 1、 $\alpha=5\%$)。

α を第一種の過誤 (false positive)、 β を第二種の過誤 (false negative) と呼ぶ^{1,2)}。

一般に、平均がゼロ、SD が σ の正規分布は、変数 x が 1.65σ より大きい値をとる確率は 5% である。図 1 より、正規分布の SD を σ_y 、 $\alpha=\beta=5\%$ とすると、判定限界と検出限界は、それぞれ、

$$y_C = 1.65\sigma_y \quad (\text{式 1})$$

$$y_D = 3.3\sigma_y \quad (\text{式 2})$$

と書ける。ここで、ブランク測定値の SD と検出限界濃度サンプルの測定値の SD は同じ (σ_y) としている(等分散の仮定)。検出限界濃度 x_D はゼロに近いので、等分散の仮定は妥当である。

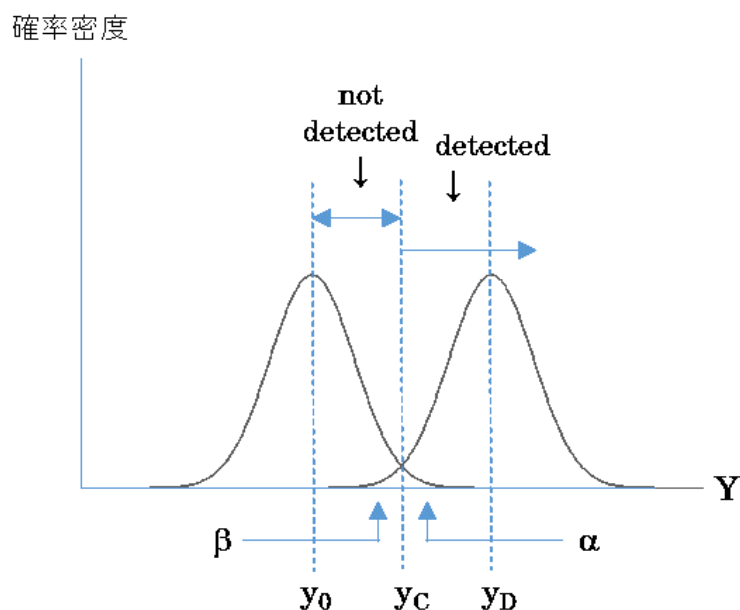


図 1 検出限界と判定限界の概念

ターゲット物質の測定結果 y から、分析者は次のように言うことができる。

5%の過誤をもって、

- $y < y_c$ ならば、濃度 x_D の物質は検出されなかった；
- $y \geq y_c$ ならば、濃度 x_D の物質は検出された。

y_D の検出・非検出の判定には基準 (y_c) と規則 ($<$ と \geq) が必要である。検出の判定基準が判定限界であり、検出の対象が検出限界である。検出限界より低い測定値からでも、検出限界濃度の物質を検出できることは注意すべきである。

これまでの議論では、ターゲットの濃度はゼロまたは x_D と仮定していた。しかし、実際の判定は濃度未知のサンプルに対して行う。そうではあっても、検出限界の課題は

濃度ゼロと濃度 x_D である二つの状態を前提として、対象のサンプルがどちらの状態に適合するかを決定する

ことである。対象物質の濃度は未知だからといって、勝手な濃度を議論しても意味はない。ゼロと x_D 以外の濃度は前提ではないからである。検出限界よりかなり大きな測定値が出現したときに、その物質の濃度を憶測するのは検出限界の範疇ではない。もっともっと検出したいと欲張って判定基準をむやみに変えたりするとバチが当たる。

2. SD の推定

検出限界と判定限界の数式は実にシンプルである (式 (1) と (2))。しかし、ある分析系の限界値を実際に決定するには、その系の測定値の SD を実験結果から推定しなければならない。

2.1 くり返し測定

最も簡単な SD の推定式は次である。

$$SD = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}} \quad (\text{式 3})$$

ここで y_i は i 番目の測定値、 \bar{y} は n 個の y_i の平均である。この式から求めた SD は実験のたびにばらつくことは分析者ならば誰しもが経験しているだろう。このばらつきの度合いはカイ 2 乗分布から見積もることができる。

図2は式 (3) で推定したSDの95%信頼区間とくり返し数 (n) の関係を示している。 n 回の実験を1シリーズの実験と呼ぶことにする。1シリーズの実験からは1個のSD推定値が得られる。95%信頼区間とは、100シリーズの実験から得られた100個のSD推定値の中の95個を含む区間であり、残りの5個は信頼区間の外に存在する。くり返し数が6の場合、図2の95%信頼区間の上限

と下限を見れば、SD 推定値の95%は真の値 (=1) の±60%の範囲にばらつくことが分かる。95%信頼区間を±20%に改善するには、40回以上のくり返し実験が必要であることも分かる。

SDの95%信頼区間が±60%であれば (n=6)、SD推定値がシリーズ実験ごとに2倍くらい違うこともありえる。すると、式 (1) と (2) より、判定限界と検出限界も同様にばらつく。これでは、ある毒物は検出されなかったと言われても、安心はできない。しかし、40回のくり返し実験は現実的にはほとんど不可能である。

カイ2乗分布は、SD推定値の確からしさと少ないくり返し数は両立しないことを定量的に示している。これは違反することができない数学の法則である。そうであっても、同じことを何度もくり返すのは嫌である。その代替策が冒頭に述べたS/Nである。しかし、多くの場合、S/Nから確率 α と β を理論的に導くことは不可能である。つまり、S/Nからは、数値的な危険率をもって検出・非検出を判定できないのである。

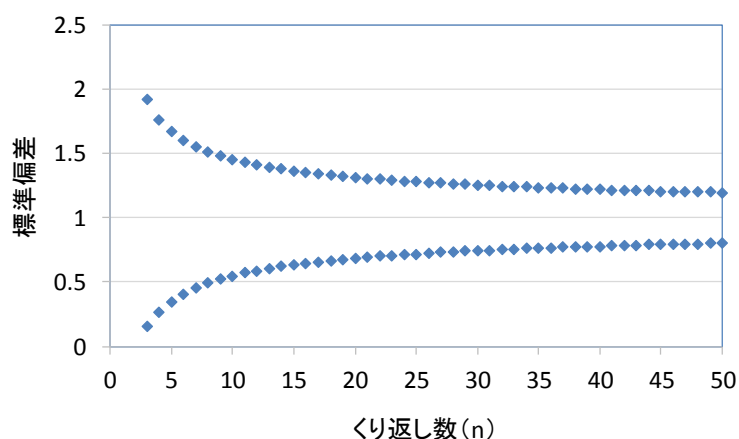


図 2 くり返し実験による SD 推定値の 95%信頼区間

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

が SD の 95%信頼区間を表す。 χ_{α}^2 と χ_{β}^2 はカイ 2 乗分布の確率 0.025 と 0.975 の χ^2 値である。

2.2 FUMI 理論

FUMI 理論はくり返し実験と S/N の欠点を補うことができる。FUMI 理論は確率論のモデルに基づいているため、くり返し測定なしに測定値の SD をかなり正確に予測できるが、反面、そのモデルの厳密性ゆえに、適用範囲が限定される。一方、統計的な SD 推定 (式 (3)) はどのような問題にも対処できるが、多くの労力(くり返し実験)なしには正確な推定値は得られない。FUMI 理論と統計的推定は相補的な関係にある。

FUMI 理論の目的は

- 測定値の SD ($=\sigma(k)$) をノイズの確率論的性質から予測する

(k はシグナルの幅である) であり、その根本的な仮定は¹⁻³⁾

1. 検出限界ほどの低濃度では、測定誤差の主原因はノイズである；

2. ノイズが作る面積が面積測定値の誤差に等しい；

3. ノイズは定常確率過程（ホワイトノイズと AR(1)（マルコフ過程）の和）で近似する；

である。そのため、この条件に合致しない分析法には適用できない。たとえば、機器分析以外の分析法、サンプリング誤差が大きい方法には FUMI 理論は無用である。

筆者らが現在までに確かめた FUMI 理論の適用範囲は——次項で述べる TOCO19 の場合——イソクラティック HPLC、グラジエント HPLC、内部標準法 HPLC である、GC および GC/MS は今後の課題である。

母集団の SD (σ) を推定する点では FUMI 理論も統計的な方法も同じであるが、両者はそのアプローチが異なる。統計的な方法は測定値の集合（見本空間）から母集団の性質 (σ) を式 (3) によって推定する。測定値の集合はくり返し実験から構築する以外に手はない。一方、FUMI 理論は、母集団の性質 (σ) を記述する理論式 ($\sigma=\sigma(k)$) に含まれる確率的なパラメータを実際のノイズから推定する。つまり、 σ の推定は測定値のバラツキを起こす原因（ノイズ）から確率的なパラメータを抽出する演算である。そのため、FUMI 理論はくり返し測定を利用しない——もちろん、くり返しの結果 (SD) を予測するが。

3. 検出限界、判定限界、定量限界の実践

有用な理論であっても、簡便でなければ、現実には役に立たない。以下で、FUMI 理論の実践ソフトウェア (TOCO19、制作：FUMI 理論研究所) を使った精度解析を紹介する。図 4 から 7 はすべて TOCO19 の出力である。

3.1 不確かさを推定するソフトウェア³⁾

TOCO19 は 1 つのクロマトグラムを入力するだけで計算結果を出力する全自動ソフトウェアである。計算結果は検出限界、判定限界、定量限界、精度プロファイルである。

図 3 に TOCO19 の概要を示す。最初にクロマトグラムをノイズとシグナルに分け、それぞれの領域から情報を抽出し、その情報を理論式 (FUMI Theory) に代入して得られた測定値の SD、 $\sigma(k)$ 、に基づいて計算結果を出力する。

図 4 は TOCO19 に入力した有機酸の標準溶液のクロマトグラムである。その下の四角いブロックは自動認識されたピーク領域（幅）を示している。すべてのピークの幅が正確に検出されてい

ることが分かる。ピーク#2（リンゴ酸）は未知ピークと分離していないので、これらはまとめて検出されている。Y=-0.15にある水平な線分は自動認識されたノイズ領域を示している。ノイズに関する情報はこの領域だけから独占的に収集される。

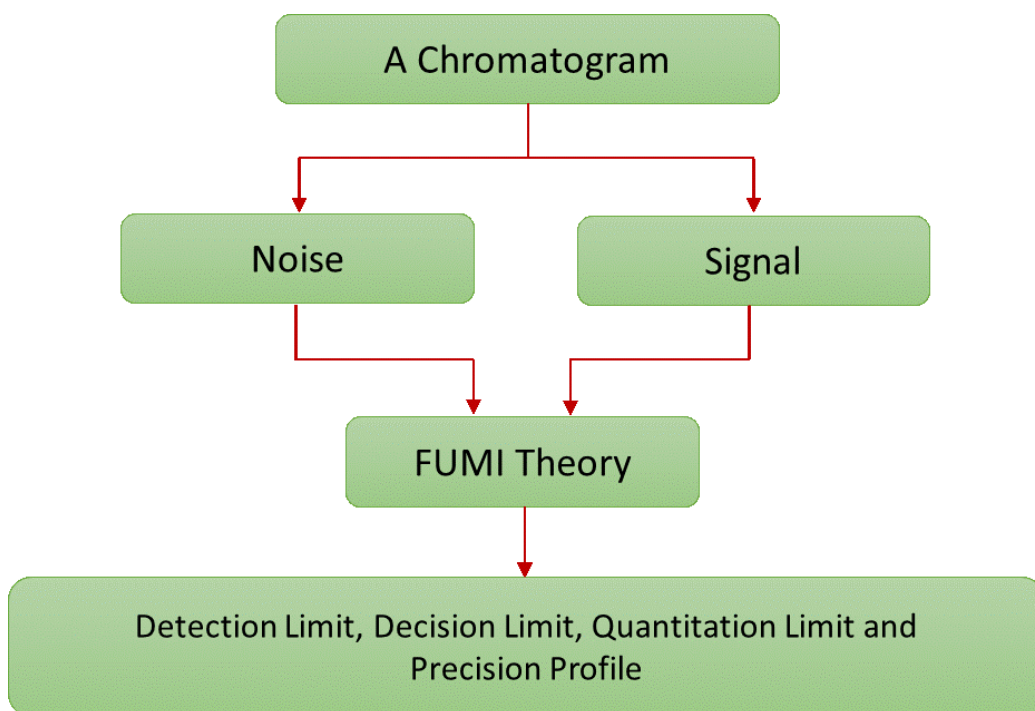


図 3 TOCO19 の精度推定アルゴリズム

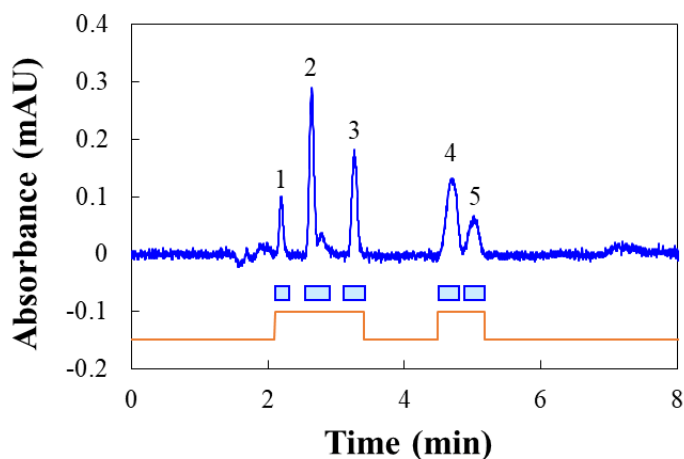


図 4 有機酸の標準溶液のクロマトグラム³⁾

ピーク：1, システムピーク；2, リンゴ酸；3, 酢酸；4, クエン酸；5, コハク酸（濃度はそれぞれ 10mg/L）。HPLC 条件：移動相, 20 mM リン酸；カラム, YMC-Triart C18 (150×3.0 mm i.d., 3 μm)；カラム温度, 37°C；流速, 0.425 mL/min；測定波長, 220 nm；データ取込間隔, 0.2 s/point.

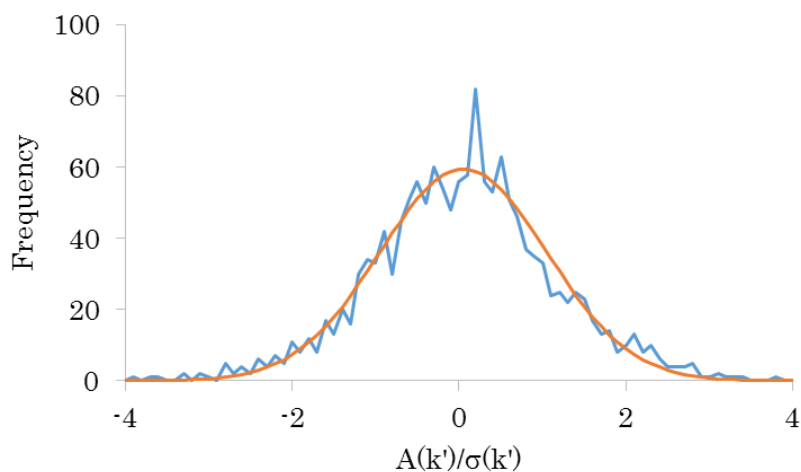


図 5 面積測定値のばらつきヒストグラム(誤差の分布、ノイズが作る面積の分布)

図 5 はノイズが作る面積 ($A(k')$) のヒストグラムである。ギザギザの線は図 4 のノイズ領域から実際に収集された面積値の分布であり、滑らかな線は理論曲線 (正規分布) である。理論と実践がぴったりと合っている。この一致があつて、初めて FUMI 理論の精度予測が成功したと言える。なぜならば、「測定値の誤差=ノイズが作る面積」(2.2 の根本仮定 2) が有効である限り、図 5 は測定値のばらつきヒストグラムと等価だからである。 $A(k')$ は幅 (k') を 300 ポイントに固定した仮想ピーク面積であり、横軸は $A(k')$ の分布の SD ($=\sigma(k')$) で標準化してある。

3.2 検出限界、判定限界、定量限界のシグナル

TOCO19 は 3 つの限界値を数値的に出力するのは勿論である (図 6 のレジェンドを参照) が、これらの限界値に関する視覚的な情報も提供する。図 6 は酢酸の定量分析 (図 4) における検出限界 (A)、判定限界 (B)、定量限界 (C) の S/N の様子を示す。シグナルはガウスピーク (正規分布) で代替してある—もちろん、その幅 ($\pm 3\sigma$) は酢酸のピークの幅 (k) と同じである。一方、ノイズは図 4 の実際のノイズの一部である。検出限界と判定限界は式 (1) と (2) の定義を採用しているが、定量限界 (QL) は、日本薬局方⁴⁾に従って、 $QL=10\sigma_y$ としている。式 (1) と (2) の σ_y は母標準偏差であるので、TOCO19 では、FUMI 理論で予測される $\sigma(k)$ に置き換えている。

検出限界と定量限界の良く知られている定義は、それぞれ、 $S/N=3$ と $S/N=10$ である。TOCO19 で求めた検出限界 (図 6A) と定量限界 (図 6C) の S/N はほぼこれらの値と同じである。まったく異なる方法—S/N と FUMI 理論—が同等な結果を与えることは大変興味深い。理由は分からないが、偶然ではないだろう。

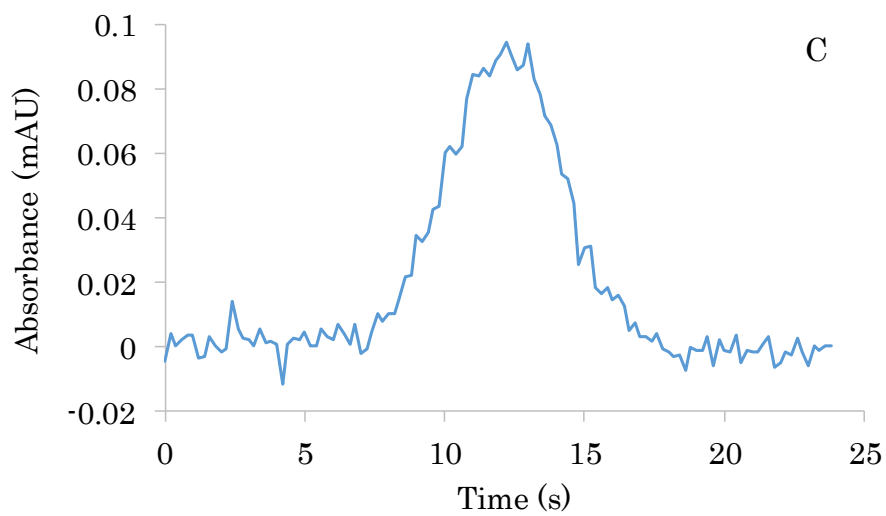
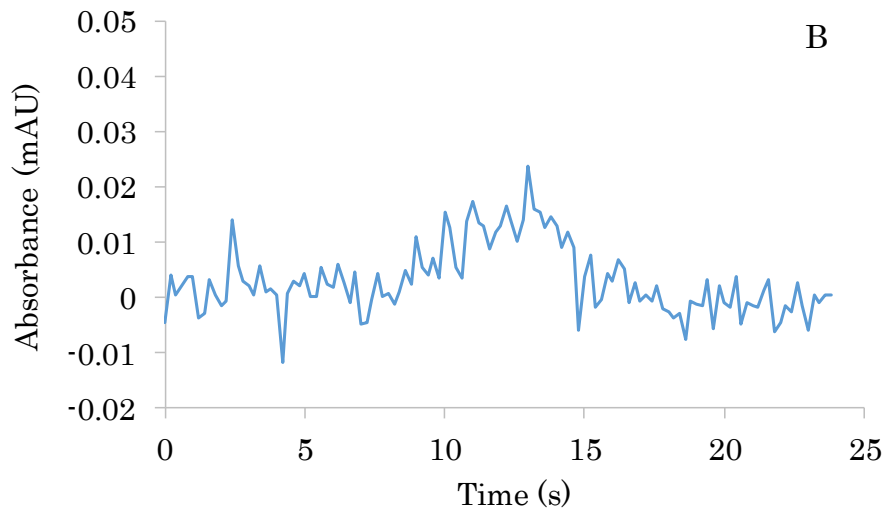
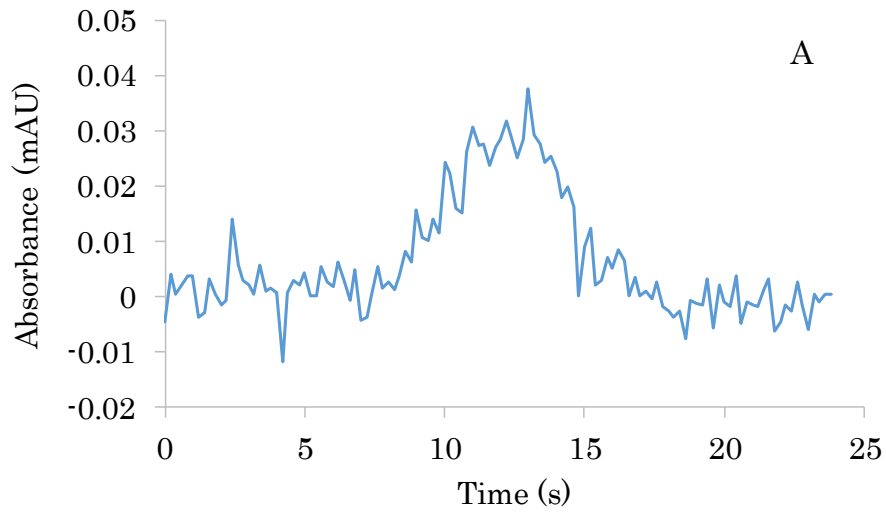


図 6 酢酸の検出限界 (A)、判定限界 (B)、定量限界 (C) のシグナル
 (上から) 1.68 mg/L、0.84 mg/L、5.10 mg/L の酢酸のピークに相当。

検出限界濃度 x_D の酢酸を測定したときの平均的な見かけのピークが (A) である。濃度 x_D の酢酸をくり返し測定すれば、(A) のピークより小さいピークも大きいピークも 50% の確率で出現するが、判定限界 (B) よりも小さいピークは 5% (β) でしか出現しない。一方、濃度ゼロの酢酸の (ブランク) 測定では、(B) より大きいピークは 5% の確率で出現する。つまり、バックグラウンドノイズは (B) よりも大きいうねりを 5% の確率 (α) で自ら作る。

判定限界シグナル (B) は検出の判定には小さすぎると直感する読者がいるかもしれない。 α は成功する確率ではなく、失敗する確率であることを思い出していただきたい。仕事で 20 回に 1 回の間違いをする人は同僚たちからは疎まれるだろうし、上司からはにらまれるだろう。 α はそれくらいに高い確率である。物質の検出を調べている分析者が煩わしく思うくらい頻繁に、ノイズは (B) ほどのうねりを作る。判定限界のピーク (B) がノイズとそれほど変わらないように見えるのは、この確率 (5%) のためである。逆に考えると、これくらいのリスクを冒さなければ、検出限界ほどの極少の物質を 95% もの高い確率で検出することはできないのである。虎穴に入らずんば虎子を得ずの事例と言える。

4. 精度プロファイル

精度プロファイル (precision profile) は JIS²⁾ と ISO¹⁾ の用語である。JIS における定義は「応答変数又は正味状態変数の CV であり、正味状態変数の関数で表した数学的記述」²⁾ である。CV は変動係数 (標準偏差を平均で除したもの) である。応答変数と正味状態変数は奇妙に響くかもしれないが、それぞれ、測定値と濃度である。つまり、検量線 ($y=f(x)$) の y と x である。

精度プロファイルは CV の濃度依存性を表しているのだから、対象の分析系の不確かさを鳥瞰できる。特に、検出限界は CV=30% に相当する濃度、定量限界は CV=10% に相当する濃度 (以下参照) として、精度プロファイルから視覚的に探り当てることができる。

図 7 は、図 4 のクロマトグラムにある酢酸のピークに対する精度プロファイルである。この精度プロットにある \rightarrow から、検出限界は 1.68 mg/L、定量限界は 5.10 mg/L であることがアナログ的に分かる。

TOCO19 およびくり返し測定により 10 mg/L 酢酸のピーク面積の CV を求めたところ、それぞれ 6.1% (N=1)、6.4% (N=6) であった。念のために統計的検定を行うと、くり返し測定で求めた CV (N=6) の 95% 信頼区間は 4.0~15.9% であるので、TOCO19 で求めた CV を含むことが分かる。吸光検出 HPLC の精度評価に TOCO19 が適用できることを示す事例である。なお、この CV の 95% 信頼区間は、平均はばらつかないと仮定して求めた区間推定であり、95% の意味は「100 シリーズの実験から得られた 100 個の信頼区間の中で、95 個が真の値を含み、5 個は含まない」である。

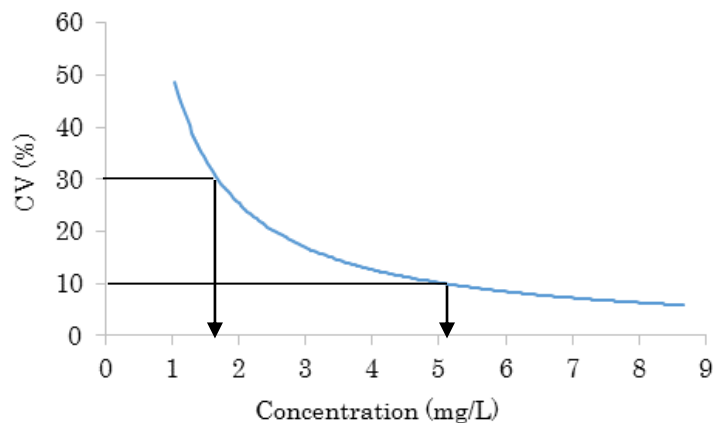


図 7 酢酸の精度プロファイル

検出限界濃度のターゲットの測定値が CV=30%であることは、式 (2) を次のように変形すれば理解できる²⁾。

$$\sigma_y / y_D = 1 / 3.3 = 30 \% \quad (\text{式 4})$$

左辺は標準偏差を平均で除したものであるから、CVである。この CV が 30%に等しいのである。ただし、この平均と SD は検出限界濃度サンプルの測定値の統計量である。同様に、定量限界では CV=10%が導かれる。

以上、分析値の不確かさに基づく概念である検出限界、判定限界、定量限界、精度プロファイルの基礎的な理論を説明し、具体例を示した。本稿は FUMI 理論の数学的な基礎を解説する場ではなかったが、それは別の機会としたい。不確かさの基礎的な考え方が具体例とともに読者に伝わったのであれば、著者らには望外の喜びである。

参考文献

- 1) ISO 11843-7, Capability of detection - Part 7: Methodology based on stochastic properties of instrumental noise, ISO, (2012)
- 2) JIS Z 8462-7, Capability of detection - Part 7: Methodology based on stochastic properties of instrumental noise, JIS, (2018)
- 3) A. Kotani, S. Tsugu, H. Hakamata, Y. Hayashi, J. Chromatogr. A (2020, in press)
- 4) 第 17 改正日本薬局方, 参考情報: 分析法バリデーション, 厚生労働省 (2016), p2343-2345