

財団法人 国際超電導産業技術研究センター 〒135-0062 東京都江東区東雲 1-10-13 Tel: 03-3536-7283 Fax: 03-3536-7318

【隔月連載記事】

やさしい「不確かさ」のおはなし(その1)

産業技術総合研究所 計測標準研究部門 応用統計研究室 榎原研正

1. はじめに

測定の信頼性が問われる局面で、「不確かさ (uncertainty)」という指標を目にする機会が 増えています。不確かさが使われる機会が増 えた理由には二つあります。一つは、これま での「測定誤差」や「精度」といった指標に 代わるものとして不確かさが使われるように なったことです。測定データを図示する際に エラーバーをつけることは、従来から行われ てきましたが、そのエラーバーが何を意味す るかについて、統一的な決まりはありません でした。 $\pm \sigma$ (σ は標準偏差) であることも、 $\pm 2\sigma$ であることもあり、さらにその σ が単に 短時間での繰り返し測定のばらつきをあらわ すものなのか、あるいは他の要因も考慮した ものなのかといったことについて、十分な注 意が払われるケースは少なかったと言えます。

測定結果の信頼性を表す考え方が技術分野 や国によってまちまちであることは、科学技 術の発展や商取引の障害になり得ることは疑 いありません。この状況に警鐘をならしたの は CIPM (国際度量衡委員会)で、CIPM の決議 に基づき 1970 年代の後半から統一的で合意 された方法の確立にむけて国際的な努力が開 始されました。まず 1980 年に、INC-1 と呼 ばれる短い勧告が作成されました。後に採用 されるようになった用語はまだ使われていな いものの、不確かさ評価の考え方を最も手短 にまとめた資料として、INC-1 の内容を確認 しておくのは無駄ではありません(囲み記事 [1] 参照)。その後、ISO(国際標準化機構) の TAG 4 (度量衡) のグループにより詳細な ガイドの作成が行われ、最終的に、計測に関 与する 7 つの主要な国際組織 (BIPM, ISO,

[1] INC-1 勧告(1980)

─ 不確かさに関する最初の公式文書 ─

BIPM を事務局として組織された国際的な作業グループにより、「実験の不確かさの表現」の統一に向けた最初の文書である INC-1 と呼ばれる勧告が 1980 年にまとめられました。INC-1 は A4 用紙1ページに満たない短い資料です。以下にその内容を紹介します(一部抜粋)。

- 1) 測定結果の不確かさは、通常、幾つかの成分からなる。 各成分は、その大きさを評価する方法に応じて、次の2 カテゴリーに分類できる。
 - A. 統計的方法により評価されるもの
 - B. 他の方法によって評価されるもの

カテゴリーA 及び B への分類と、従来からの「偶然的」及び「系統的」不確かさへの分類の間には、単純な対応関係は必ずしも存在しない。「系統的不確かさ」という用語は紛らわしいので避けるべきである。

- 2) カテゴリーA の成分は、分散の推定値 s_i²(もしくは標準 偏差の推定値 s_i)と自由度 v_iにより特徴づけられる。
- 3) カテゴリーB の成分は、それに対応する分散が存在する と想定し、分散の近似的な値と考えられる量 u_i^2 により特 徴づけるべきである。 u_i^2 は分散のように扱い、 u_i は標準 偏差のように扱う。
- 4) 合成された不確かさは、分散を合成する通常の方法によって計算される値で特徴づけられる。合成された不確かさは、「標準偏差」の形で表現する。
- 5) 合成された不確かさに何らかの係数を掛けて総合された不確かさとして表す必要がある場合には、その係数を明示しなければならない。



財団法人 国際超電導産業技術研究センター 〒135-0062 東京都江東区東雲 1-10-13 Tel: 03-3536-7283 Fax: 03-3536-7318

IEC, IUPAP, IUPAC, OIML, IFCC)による調整と合意を経て、"Guide to the Expression of Uncertainty in Measurement" が 1993 年に出版されました $^{1),2}$ 。このガイドは世界的に GUM と略称されています。 GUM では、従来の様々な経緯を含んだ測定誤差や精度といった用語に代えて「不確かさ」を用いることとし、その評価方法の詳細な手順や背景にある考え方が示されました。 GUM に則って評価した不確かさの利用は、基礎物理定数の表記や計測標準の国際比較など、測定の信頼性が厳しく問われる分野でまず速やかに浸透しました。

不確かさが使われる機会が増えたもう一つの理由は、測定結果の信頼性の表記が要求される機会自体が増えたことにあります。国際的な商取引を円滑に進めるには、ある国の試験結果が別の国の試験結果と整合することが必要です。特に EU (欧州連合) の経済統合の過程において、このような要求が大変切実なものとなったと言われています。そのため、これまで自主的な活動として行われていた測定のトレーサビリティの確保が、第三者による審査などによって客観的に証明される必要性が高まり、国際的に共通の方法で測定の信頼性を評価・表現することが必要になりました。客観的な証明を可能にする制度として、国際的には各国の国家計量標準の同等性を確認するための仕組みが、また国内的には各国の国家計量標準へのトレーサビリティを確認するための仕組みが最近十数年ほどの間に急速に整備されて来ました。「不確かさ」は、このような目的にちょうど適うものだったため、測定のトレーサビリティの証明が必要である場合には、必須要件として不確かさが表記されるようになっています。

とはいえ、これまでに用いられてきた測定の信頼性表記の考え方や取り決めを、ただちに GUM の方式で置き換えることは必ずしも容易ではありません。技術分野によっては不確かさの採用が必ずしも順調に進んでいない分野もなお多くあります。国際単位系 (SI) の普及に時間がかかったように、不確かさの普及も時間をかけて地道に進めることが必要で、これは現在、計測分野においてとりわけ重要な課題となっています。例えば、BIPM は SI の普及の中心として活動してきた組織ですが、SI の普及が一段落した今、不確かさの利用の普及を SI 普及に替わる次の主要事業と位置づけています。

ここでは、測定の不確かさについての解説を数回にわたって本誌に掲載する機会を頂きましたので、できるだけ肩の凝らない説明を心掛けたいと思っています。不確かさの評価を一から解説するとなると、まず統計的手法に関わる数理を説明した上で、Aタイプ評価の方法(統計的方法)、Bタイプ評価の方法(統計的方法以外)を説明し、次に伝播則による不確かさの合成、最後に拡張不確かさの表記方法というのが、オーソドックスな順序と思われますが、ここではこの順序には必ずしもこだわらず、むしろ事例を含む heuristic な解説を行おうと思います。最小限の統計的概念や手法は必要であるだけでなく有用と思われますので、折に触れて数理的背景についても説明する予定です。補足的な内容については、囲み記事として記載します。

2. GUM の特色

測定の信頼性評価に関わる従来の考え方と比べたとき、GUMに特徴的な考え方は次の三点です。

- (1) 不可知量である「測定量の真値」や「測定誤差」といった概念を、評価手続きのルールの中に持ち込まない。
- (2) 統計的解析以外の方法で評価する不確かさ成分についても、「分散(あるいは標準偏差)」として評価する。
- (3) 測定器や測定方法ではなく、測定結果(つまり測定後に得られる特定の値)に対して評価する。

測定結果には多かれ少なかれ誤差が含まれており、測定量の真値を厳密に知ることは不可能です。



財団法人 国際超電導産業技術研究センター 〒135-0062 東京都江東区東雲 1-10-13 Tel: 03-3536-7283 Fax: 03-3536-7318

ある測定結果が真値とどれだけずれているかを確認するためには、別の測定を行わざるを得ず、その測定にもむろん誤差が含まれます。真値を厳密には知り得ないのは、測定という操作が原理的に持つ特徴と言えます。「測定誤差 = 測定結果 - 真値」という定義は確かにわかりやすいのですが、この素朴な定義に基づく誤差評価がしばしば混乱したため、GUMでは、不確かさ評価の手続きの中に不可知量を持ち込まず、評価可能な量だけで評価の手続きを書き下すことをポリシーとしました。不確かさに対する次の一見複雑な定義もこの考え方によります。

不確かさ = 測定結果に付随する、合理的に測定量に結びつけられ得る値のばらつきを特徴づけるパラメータ

この定義については項をあらためて説明します。なお、真値や測定誤差という概念自体が 無用になったわけではなく、これらは測定の 中でなお重要な概念であることに変わりはあ りません。

- (2) の特徴は、INC-1 の 3) に相当するもので、系統的成分(かたより)についても偶然的成分(ばらつき)と同じように扱うことを述べています。従来の誤差評価では、かたよりをどう扱うべきかが明瞭にされていないことが多かったので、(2) は GUM の最も重要なセールスポイントの一つと言えます。GUMのAnnex E は、何故、系統的成分を偶然的成分と同じ土俵で扱うことが妥当なのかを長々と説明しています。ここでは習うより慣れよで、連載の次号では簡単な事例について(2)の評価を実際に示したいと思います。
- (3) の特徴は、不確かさ評価は計測器の性能評価とはもともと立場が違うことを意味します。この点はしばしば見落とされていますが重要な点です。誤差というと、誤りとか狂いとかのニュアンスがあるため、その要因としては、測定器や測定方法など測定側システムに存在する何らかの理想的でない振る舞いが想定されることが多いと言えます。一方、被測定側が理想的な振る舞いをしないために生じる曖昧さも存在します。例えばヒトの最高・最低血圧は、一般に測るたびに結果が違いますが、これは血圧計の問題というよりはヒト側、被測定側の安定性の問題と考えられ

[2] GUM の国際文書としての地位

GUM は、度量衡、産業、物理学、化学、医薬学、法定計量 にわたる分野を代表する7つの国際機関がオーソライズし た文書であり、国際単位系(SI)の拠り所となっているSI文書 (SI Brochure; 最新版が www.bipm.org/ en/si/si_brochure/ から、またその日本語訳が www.nmij.jp/library/units/si/からダウンロード可能)と同格 の文書と考えられています。ISO や IEC によって発行される 規格の中では、不確かさを用いることになっており(ISO/IEC Directives, Part 2)、これらの規格の上位文書と考えられま す。ただし、文書自体は強制力のないガイドで、発行母体が 複数の機関にわたっていることもあり、文書の性格が一般 の人にはわかりにくいことも事実です。そこで ISO/IEC で は、GUM をあらためて ISO/IEC ガイドとして 2008 年に発行 しました(ISO/IEC Guide 98-3)。また国によっては、GUM を国内規格として採用している国もあるほか、7国際機関の それぞれで、独立のガイドとして(ただし中身はGUMを無修 正で)発行しようという計画も進んでいます。

GUM は SI 文書のように、誰もが自由にダウンロードできる状態になっていることが理想です。現在の GUM の著作権は、ISO/TAG4 が執筆を担当した歴史的経緯から、ISO が保有しており、当然ながら ISO 事務局は GUM のオープン化には抵抗しています。著作権問題がクリアされれば、機関毎のガイドについて、自由なダウンロードが可能となる可能性があります。

ます。機械的ポンプの脈動の最高・最低圧を測定する場合は、もっと安定した結果が得られるからです。(3) の帰結として、不確かさ評価では、このような被測定側の要因も含め、測定後の我々の知識の曖昧さに寄与する要因はすべて考慮しなければならないことになります。従来からの誤差評価は、しばしばこの点が曖昧のまま行われてきました。



財団法人 国際超電導産業技術研究センター 〒135-0062 東京都江東区東雲 1-10-13 Tel: 03-3536-7283 Fax: 03-3536-7318

今回は GUM に関する歴史的経緯とその考え方の大枠を紹介しました。次回は簡単な事例について実際に不確かさ評価を行いたいと考えています。なお、不確かさ評価に関する理屈や能書きを述べた解説は(この解説を含めて)少なからずありますが、不確かさ評価の実際の事例についてのまとまった情報は多くは存在しません。このため、2009 年 3 月 9 日東京において、様々な分野における不確かさ評価事例を集めた発表会の開催を計画しています。

詳細については、産業技術総合研究所 計測標準研究部門の不確かさ

WEB (http://www.nmij.jp/mprop-stats/stats-partcl/uncertainty/club2.html)

を参照下さい。

参考文献

- 1) International Organization for Standardization (Geneva, Switzerland.), "Guide to the Expression of Uncertainty in Measurement" (初版 1993、修正版 1995)
- 2) 飯塚幸三監修: ISO 国際文書 「計測における不確かさの表現のガイド」 日本規格協会 (1996) (文献 1 の翻訳)



財団法人 国際超電導産業技術研究センター 〒135-0062 東京都江東区東雲 1-10-13 Tel: 03-3536-7283 Fax: 03-3536-7318

【隔月連載記事】

やさしい「不確かさ」のおはなし(その2)

産業技術総合研究所 計測標準研究部門 応用統計研究室 榎原研正

3. 不確かさ評価の例 ― 家庭用血圧計

3.1 評価の実際

習うより慣れよの精神で、簡単な例で不確かさ評価を実行してみます。家庭用血圧計を用いて、最高血圧 Ph を測定し、128 mmHg の値が出たものとします。その不確かさ u(Ph) を評価します。u(Ph) という記号は、不確かさが Ph の関数であることを意味するのではなく、「測定結果 Ph に対する標準不確かさ」を意味する GUM 流の表記です。

血圧は測るたびにばらつきます。まずこのばらつきを評価するため、引き続き測定を繰り返して、次のデータを得たとします。

128, 132, 123, 121, 125 (mmHg)

最初のデータ 128 mmHg (= q_1 とします) はもとの測定結果、残りの 4 つのデータ 132~125 mmHg (q_2 ~ q_4) は不確かさ評価を行うために追加測定したデータです。これらの 5 個のデータの実験標準 偏差(囲み記事 [3] 参照)は s=4.32 mmHg となります。そこで u(Ph) のうち、測定のばらつきによる成分を $u_1(Ph)$ と書くことにすると、

$$u_1(Ph) = s = 4.32 \text{ mmHg}$$
 (1)

[3] 実験標準偏差

n個の測定データの組 $\{q_1, q_2, \cdots, q_n\}$ があるとき、これに対する実験標準偏差sは、次のように求めます。 [Step 1]

平均値:
$$\overline{q} = \frac{q_1 + q_2 + \cdots + q_n}{n}$$

[Step 2]

平方和:
$$S = (q_1 - \overline{q})^2 + (q_2 - \overline{q})^2 + \dots + (q_n - \overline{q})^2$$

[Step 3]

実験分散:
$$V = \frac{S}{S \text{ O} \hat{\theta} + \hat{\theta} \hat{\theta}} = \frac{S}{n-1}$$

[Step 4]

実験標準偏差: $s = \sqrt{V}$



財団法人 国際超電導産業技術研究センター 〒135-0062 東京都江東区東雲 1-10-13 Tel: 03-3536-7283 Fax: 03-3536-7318

平方和の計算に出てくる $(q_i - \overline{q})$ という量は全部でn個ありますが、これらを全部足すと必ずゼロになります。

n個のうち(n-1) 個の値がわかっていれば、最後の 1 個は自動的にわかるため、情報の個数としては(n-1) 個しかありません。これが Step 3 の計算で n-1 が自由度と呼ばれている理由です。

後で説明する予定ですが、 \mathbf{q} の母分散を \mathbf{s}^2 と書くと、 \mathbf{V} の期待値が \mathbf{s}^2 となります。従って、 $\mathbf{\sigma}^2$ のかたよりのない推定値として \mathbf{V} が使えます。これが \mathbf{S} / \mathbf{n} でなく、 \mathbf{S} (\mathbf{n} -1) によって \mathbf{V} を定義する理由です。

なお、「実験標準偏差 (experimental standard deviation)」というのは GUM 用語です。通常の統計学の教科書にいう「試料標準偏差」や「不偏分散の平方根」と同じ意味です。ただし試料標準偏差が $\sqrt{S/n}$ を表す流儀もあるため、混乱を避けて「実験標準偏差」という用語が使われていると考えられます。

一方、血圧計の目盛りがずれている可能性があります。家庭用血圧計はデジタル表示ですから外から見える目盛りはありませんが、ここでは内部的な目盛りのずれを考えています。「かたより」に相当する成分です。血圧計の説明書の仕様の欄を

見ると「精度±4 mmHg」の記述があります。これをどう解釈するかは単純ではありませんが、「目盛りのずれは±4 mmHg を超えることはない」と解釈することにします。測定結果は 128 mmHg ですから、正しい目盛りで測定したと想定するときの測定結果は 128 ± 4 mmHg の範囲内にあると考えられます。この範囲の中で、どれが最も確からしいという情報はないので、どの値も同じように確からしいと考えるのは不合理ではありません。この状況を確率分布として表現したものが図 1 です。

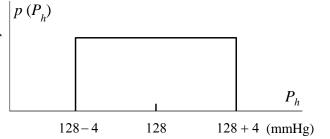


図1正しい目盛りの血圧計で得られると予想される測定結果に対して確率分布を想定する

目盛りのずれを考慮することによって、測定結果に図 1 の確率分布であらわされるようなぼけ(あるいは拡がり)が出てきたことになります。このぼけの大きさは、確率分布 p(Ph) の標準偏差(囲み記事 [4] 参照)で表すことができます。図 1 のような確率分布は一様分布と呼ばれ、その標準偏差は半幅/ $\sqrt{3}$ で計算できます。半幅は 4 mmHg ですから、目盛りのずれによる不確かさの成分を $u_2(Ph)$ と書くことにすると、

$$u_2(Ph) = \frac{4}{\sqrt{3}} = 2.31 \text{ mmHg}$$
 (2)

となります。

全体としての不確かさ $u_c(Ph)$ は、 $u_1(Ph)$ と $u_2(Ph)$ を次の二乗和のルールで合成して求めます。

$$u_c^2(P_h) = u_1^2(P_h) + u_2^2(P_h)$$
 (3)



財団法人 国際超電導産業技術研究センター 〒135-0062 東京都江東区東雲 1-10-13 Tel: 03-3536-7283 Fax: 03-3536-7318

これは統計学で分散の加法性といわれる定理に相当しています。不確かさ評価の文脈では、不確かさの伝播則と呼ぶルールの最も単純な場合に相当します。式 (3) 右辺の $u_1(Ph)$ 、 $u_2(Ph)$ は標準偏差相当の大きさとして表現した不確かさという意味で、いずれも標準不確かさ(standard uncertainty)と呼ばれます。左辺の $u_c(Ph)$ は合成標準不確かさ(combined standard uncertainty)です。

式 (3) で実際に計算すると、

$$u_c(P_h) = \sqrt{4.32^2 + 2.31^2}$$

= 4.898 ··· \(\times 4.9\) (mmHg) (4)

不確かさは最終的に 2 桁で表すことが一般的なので、4.9 と丸めます。測定結果の最小桁を不確かさの最小桁に合わせ、結局、測定結果は 128.0 mmHg、その合成標準標準不確かさは 4.9 mmHg となります。

不確かさの最終報告は、このように合成標準不確かさを使うのが一つの方法です。例えば基礎物理定数のデータベース(http://www.codata.org/)では合成標準不確かさが使われています。もうひとつの方法は合成標準不確かさに一定の数因子k(包含係数と呼ばれます)をかけた拡張不確かさとして表すものです。包含係数としてしばしば、k = 2 が用いられます。これを用いると、不確かさの最終報告は、「測定結果 128.0 mmHg の拡張不確かさは 9.8 mmHg (包含係数k= 2)」となります。拡張不確かさを使うときは必ず包含係数の値を併記することになっています。拡張不確かさUを用いる理由は、 $\pm U$ を信頼区間として用いることができるからです。kの選択の考え方についても後にあらためて説明する予定です。

[4] 確率分布の標準偏差

確率密度関数 *p(x)*で表される確率分布に対して、 期待値 *E[x]、*分散 *V[x]、*及び標準偏差□は次のように定義されます。

期待値: $E[x] = \int x p(x) dx$

分散: $V[x] = \int (x - E[x])^2 p(x) dx$

標準偏差: $\sigma = \sqrt{V[x]}$

囲み記事[3]の計算が有限個のデータに対する ものであったのに対し、データが連続で無限個 ある場合への自然な拡張がこのような定義にな ります。

3.2 検討すべき課題

不確かさの評価は、上で見たように、個別の不確かさ成分を数え上げ、それぞれを標準偏差相当の大きさとして評価し、分散 (標準偏差の二乗) の加法性を使って合成するという手順で行います。 複雑な手順ではありませんが、次の点について注意しておく必要があります。

A) 不確かさは測定結果の属性

上では血圧の測定結果 Ph は、ただ 1 回の血圧測定値 q_1 であるとしました($Ph=q_1$)。しかし不確かさ評価のために 5 回の繰り返しデータ $q_1\sim q_5$ をとっています。もしこれらの平均値 $\bar{q}=125.8$

mmHg を測定結果として使う($P_h'=\bar{q}$)ならば、不確かさは違ってきます。具体的には、式(1) に換えて

$$u_1(P_h') = \frac{s}{\sqrt{5}} = 1.93 \text{ mmHg}$$
 (5)



財団法人 国際超電導産業技術研究センター 〒135-0062 東京都江東区東雲 1-10-13 Tel: 03-3536-7283 Fax: 03-3536-7318

とばらつきの寄与が小さくなります。 $\sqrt{5}$ の 5 は、平均値をとったデータの個数です。式 (5) の由来については機会をあらためて説明します。

ここで重要なポイントは、測定器や測定方法が決まっても不確かさは決まらないという点です。 同じ方法を使っていても、測定にうまいへたがある場合もあり、測定量の大きさが違うと不確かさ が違うこともあり得ます。すなわち、不確かさは測定器や測定方法ではなく、測定結果(つまり特 定の値)の属性ということになります。

B) ぼやけた真値

5つの血圧測定値のばらつきの主な原因は、血圧計の不安定性ではありません。むしろ脈拍毎に、あるいは短時間の間に最高血圧の値が変動しているためと考えられます。また「最高血圧」という測定量がどのように定義されているかもここでは問題になります。それは、「1回の脈拍に対して」なのか、「5分間の平均値」なのかといったことです。つきつめて考えると、ほとんどの測定でこの種の定義の曖昧さが存在します。曖昧さが少ないに越したことはありませんが、それは測定系の設計の問題であって、不確かさ評価の問題ではありません。

つまり、測定量の真値はただ一つの値ではなく、測定対象の時間的変動や測定量の定義の曖昧さなどによってぼやけていることがしばしばあり、不確かさ評価ではこれらも考慮する必要があるということです。「誤差」というと、測定器や測定方法が理想的でないふるまいをすることに起因するというニュアンスがあるため、従来の誤差評価では、被測定側の要因をどこまで考慮するかには曖昧な点がありましたが、不確かさ評価では全部考慮します。

C) 不確かさ成分の分割

上では不確かさの成分を、ばらつきとかたよりの二つの成分に分けて評価しました。上の例では これは自然な分割と思われますが、このような分割にはしばしば注意が必要です。これについて次 節であらためて検討します。

4. かたよりとばらつき

4.1 かたよりとばらつきは区別できるか

測定におけるかたより(系統誤差)とばらつき(偶然誤差)は意外にやっかいな概念です。ここではかたよりとは、長さ測定における熱膨張の補正、質量測定における浮力補正など、補正可能な既知の効果はすべて補正した後になお残る効果のことを意味するものとします。計測標準を用いた校正が可能であれば、そのような校正も行った後に残る効果です。その値を知ることは原理的に不可能なものがかたよりです。

かたよりとばらつきをどのように切り分けられるかがまず問題になります。前節の5つの血圧測定データの相互の違いは、原因はどうあれ、ばらつきと呼ぶのは妥当そうです。図2の場合はどうでしょうか。この図は、温度目盛りの国際的な整合性を確認するため、水の三重点(気体、液体、固体が共存する温度)を実現する三重点セルという装置を複数の国の国家計量標準研究所(NMI; National Metrology Institute)の間で回付して測定したラウンドロビンテストの結果の例です。縦軸は参加した20 NMI の測定結果の単純平均(これが真値というわけでは決してありません)からの差を示しています。

財団法人 国際超電導産業技術研究センター 〒135-0062 東京都江東区東雲 1-10-13 Tel: 03-3536-7283 Fax: 03-3536-7318

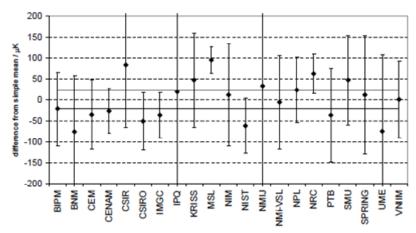


図 2 水の 3 重点に対する国際比較の例(M. Stock et al. "Final Report on CCT-K7: Key comparison of water triple point cell", Metrologia 43, 03001 (2006))

わが国の NMIJ の測定結果の結果は平均値よりすこし上にずれていますが、このずれはばらつきでしょうか、かたよりでしょうか。 NMIJ については、長いエラーバー(±拡張不確かさで表わした信頼区間を意味しています)の中に平均値が含まれているので、ばらつきと解釈するのが妥当かも知れません。しかしこのエラーバーの長さの大部分は系統的な効果を安全側に(過大に)評価したためであることがわかっています。従って次の測定を行った場合、やはり平均値より上に出る可能性は低くありません。とすると、かたよりと解釈してもよさそうです。 MSL はニュージーランドの NMI ですが、エラーバーを考慮に入れても、平均値より上に明確にずれていますので、MSL の平均値からのずれは概ねかたよりと解釈できそうです。

一方、参加国のデータ全部をながめたとき、どの国が最も信頼性の高いデータを出しているか判断することは不可能です。その点では、上の血圧測定のデータと大きく変わりません。国毎のデータはばらついており、平均値や標準偏差の計算といった、通常ばらつきに対して行われる統計的演算にも意味があります。

このように考えると、測定におけるかたよりとばらつきの区別は予想以上に微妙であることがわかります。これをもう少し整理するためには、ばらつきの中にある構造を考える必要があります。これは次回に検討します。

4.2 多数決原理と真値

図2の結果について、MSL及びNMIJの名誉のためにも、一言追加しておきます。図2の中で異端にみえる MSL は、かたよりの少ない測定を行っていたと現在では考えられています。これは当時必ずしもよく認知されていなかった、セル中の水の同位体比及び不純物濃度の効果が MSL による測定では補正されていたことによります。多数決原理によっては必ずしも真値に近づけないことを示す例と言えます。

幾つかの国は、これらの効果を補正しないかわりに、不確かさの一部として評価していました。 NMIJ のエラーバーが長いのはそのせいで、測定能力が低いということでは必ずしもありません。補正も行わず、不確かさ成分としての考慮もしていなかった NMI もたくさんあったことがわかっています。現在では、この補正をどのように行うべきか知られており、図 2 の状況は過去のものとなっています。



財団法人 国際超電導産業技術研究センター 〒135-0062 東京都江東区東雲 1-10-13 Tel: 03-3536-7283 Fax: 03-3536-7318

【隔月連載記事】

やさしい「不確かさ」のおはなし (その3)

産業技術総合研究所 計測標準研究部門 応用統計研究室 榎原研正

4. かたよりとばらつき (つづき)

前回は、家庭用血圧計の不確かさを、かたよりとばらつきに分けて評価し、それらを合成するという手順で全体の不確かさを求めました。しかし、かたよりとばらつきの区別は、場合によってはかなり微妙であることも指摘しました。今回は、この話題の整理のため、ばらつきの中に存在する構造について説明します。なお、前回も述べましたが、かたよりとは、補正可能な既知の効果はすべて補正した後になお残る効果のことを、ここでは意味しています。

4.3 ばらつきの階層構造

同一の測定条件で同一対象を短時間の間に複数回測定することをここでは「繰り返し(repetition)」とし、測定日、測定者、測定装置などのどれか一つ以上の測定条件を変えた複数の測定を「反復 (replication)」と呼ぶことにします。日を変えると、前日のデータと傾向が違っていたということは少なからずあります。測定を日常的に行っている人は、繰り返しのばらつきよりも反復のばらつきの方が一般には大きいことを経験されていると思います。測定日、測定者、測定装置などによるばらつきは、繰り返しによるものとは、大きさ、原因、統計的ふるまいなどが一般に異なります。

真値 τ の測定量を、5 日間にわたって毎日 10 回繰り返し測定したものとします。第 i 日目の第 j 繰り返しのデータを y_i とすると、測定データは次のようにモデル化することができます。

$$y_{ij} = \tau + \delta + \gamma_i + \varepsilon_{ij}$$

= $\mu_i + \varepsilon_{ij}$ (6)

このような表現は、統計学でデータの構造模型と呼ばれています。 δ はこの測定法がもつ固有のかたより、 γ_i は日による違いを表す確率変数、 α_i は繰り返しのばらつきを表す確率変数です。

この状況は模式的に図 3(a)のように表せます。第 i 日目の楕円は、その日の測定値の「母集団」、つまりその日の測定データとして可能性のある全ての値の集合を表すものとします。これは、その日に測定を無限回繰り返したと想定したときに得られる値の集合です。実際に測定を無限回繰り返すことは不可能ですから、母集団というのはあくまで仮想的な集合です。一方、現実に手元にある 10 個のデータは、この母集団からランダムに抽出して得た「試料」と考えます。「母集団」と「試料」の二段構えは、測定値を統計学的に取り扱う際にしばしば利用される考え方です。

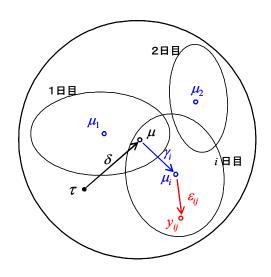
母集団としては、測定日毎の母集団以外に、測定日を特定しない全体の母集団があります。図 3(a)では、最も外側の円がこれに対応します。この全体の母平均(母集団の平均値)を図では μ としています。一方、式(6)の μ , は、第i日の測定値の母平均です。日を特定しない場合、繰り返しのばらつき α , 以外に、日によるばらつき α , があり、ばらつきが階層構造をもっていることがわかります。測定者や測定装置などの要因も入ってくると、さらに多重の階層構造となります。

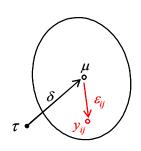
なお、ここでは不可知量はギリシャ文字で、可知量はアルファベットで表しています。この記法は統計学でしばしば用いられるもので、例えば同じ「平均」であっても、母平均(不可知量)なのか、試料平均(測定値の平均で可知量)なのかについての無用の混乱を避けることができます。ここでは τ 、 δ 、 γ 、 κ はすべてその厳密な値は知り得ない不可知量で、測定値 γ が可知量です。



財団法人 国際超電導産業技術研究センター 〒135-0062 東京都江東区東雲 1-10-13 Tel: 03-3536-7283 Fax: 03-3536-7318

また、式(6)のような表現はあくまでモデルであることに注意しておきたいと思います。現実の測定に生じるばらつきは、ここまでが測定日によるばらつき、ここからが繰り返しによるばらつきと明瞭に区別できるとは限りません。特に、ランダムウオークのように元来は小さい偶然的変動が継続的に集積することにより一定のかたよりが生じるような場合、式(6)のような表式は必ずしも適当ではありません。





(a)日によるばらつきを考慮する場合

(b)日によるばらつきを考慮しない場合

図3 測定データの母集団

4.4 母集団の選択

 δ をかたより、 ϵ_{ij} をばらつきと呼ぶのは自然な呼び方と思われますが、 γ をどう呼ぶかは微妙です。第 i 日の母集団の中だけで見ると、その中では一定値であり、かたよりですが、全体の母集団からはばらつきに見えます。すなわち、かたよりかばらつきかは、母集団をどう捉えるかに依存しています。ところで図 3(b)は、日毎のばらつきが無視できるような状況で想定される母集団ですが、測定は一日で終了し、二日以上に及ぶことはない場合、不確かさ評価の際に母集団として想定すべきはこの図 3(b)でよいでしょうか?

測定が一日で終わるとしても、その一日が図 **3(a)**のどの日に相当するのかは不明ですから、日による変動があり得ると思われる場合には、図 **3(a)**の全体の母集団(一番外側の円)を想定する必要があります。測定者や測定装置などについても同様で、「この測定は担当者一人で実施しているので、人によるばらつきは考えなくて良い」ということにはなりません。

逆に図 3(b)のような母集団の想定が妥当となるのは、日毎の変動が無視できることが確認されている場合、毎朝 1 回測定標準を用いて測定装置を校正することにより日毎の変動を除いている場合 \dagger 、あるいは日毎の変動をかたより δ の一部として評価する場合です。

なお図 3(a)の状況で、日毎のばらつきの大きさと繰り返しのばらつきの大きさを分離して評価する方法として、分散分析と呼ばれる統計的手法があります。分散分析を用いた不確かさ評価の例に

_

[†]厳密にいうと、この場合は、校正作業の不確かさを考慮する必要があります。図 3(b)にはこの成分は描かれていません。



財団法人 国際超電導産業技術研究センター 〒135-0062 東京都江東区東雲 1-10-13 Tel: 03-3536-7283 Fax: 03-3536-7318

ついてもいずれ紹介したいと考えています。

4.5 「偶然」の意味 ― 繰り返しのばらつきについて

一般に、ばらつきは偶然誤差、かたよりは系統誤差とも呼ばれます。偶然誤差と言われると、原因不明で対策無しと思わず納得してしまいそうですが、偶然というのはどういう意味でしょうか。例えば、マイクロメータでシリンダーの直径を繰り返し測定した際のばらつきはどうでしょうか。シリンダーの測定位置の微妙なずれ、測定力のばらつき、手から伝わる熱などによる温度の微小な時間的変化など、なんらかの物理的原因によりばらつきが生じていると考えられます。明確な原因は今は不明にせよ、努力すれば原因が判明し、ばらつきを減らすことができる可能性は潜在的にはあります。本当に人が制御できないばらつきも存在し、それは不確定性原理で表されるような量子的な効果によるものですが、我々が実際に目にするデータのばらつきが量子的な限界に達しているケースは多くはありません。

つまり偶然誤差の「偶然」というのは、統計的な一つの見方にすぎず、物理的には何らかの原因による必然的な結果であると(量子効果を除けば)言えます。測定の繰り返しによるばらつきは偶然誤差と呼ばれることが多いのですが、これを原理的に原因不明で対策が無いものと安易に仮定するのは適当ではありません。無論、その低減対策は現実には難しいのですが、「偶然」という言葉には十分の警戒が必要です。

4.6 かたよりは減らせるか

ばらつきの階層構造を上に昇るにつれ、かたより的な性質が増えてきます。このような固定的な傾向がある成分は、その原因を追求し対策をとることが比較的容易にできそうだと思われるかも知れません。しかし現実にこれが実行可能であることはあまり期待できません。例えば、測定日によるばらつきは、温度や湿度の日による違いが原因なのではないか、それらを揃えればばらつきは減るのではないかと、机上では想像できるかもしれません。しかし日毎の違いが弊害となっている状況ならば、容易に想像できる原因についての対策はすでにとられていると考えるのが自然です。十分な努力による対策がとられ、可能な補正はすべて施した上でなお残っているかたよりが本当の問題なのです。原因がわかっていてもコストのために対策がとれないということもあります。コストは重要な技術要素の一つですから、実際上対策がとれないという点では同じです。前回、水の三重点の国際比較の例で紹介したように、データのばらつきの主要な原因の一つが判明し、その対策が可能になるというのは大変珍しい例で、このようなことがあればそれ自体科学技術の大きな成果といえます。日常の測定の中に含まれる変動成分については、それが上の方の階層でかたより的性質が強いものでも、原因を特定して低減対策をとることは一般に難しいと言えます。

ガス器具のエネルギー効率試験の不確かさ評価に協力したことがあります。もともと難しい試験で、試験結果のばらつきも大きいのです。同一のガス器具を複数のラボで試験するラウンドロビンテストを行ったところ、ラボ間のばらつきは、同一ラボ内での繰り返し試験のばらつきより明白に大きいことが判明しました。これに対して、ある統計学の専門家が、ラボ間のばらつきは繰り返しのばらつきのレベルまで下げるべきであり、これを放置したまま試験を続けるのはモラルの欠如とまで発言されたのを聞いて、大変驚いたことがあります。繰り返しのばらつきは偶然的効果であって対策は不可能であるが、それより大きな変動を持つ効果はその原因を特定して対策をとることができるはずであるという二重の誤解がここにはあると思います。ラボ間のばらつきをラボ内での繰り返し試験のばらつきと同程度の大きさに押さえることは一般には困難です。ラボ間のばらつきに対して私たちが期待できるのは、各ラボ毎に評価された不確かさの大きさと同程度かそれ以下であるということです。

かたよりはくみしやすし、ばらつきはくみし難しとの理解が、しばしば見受けられますが、実際



財団法人 国際超電導産業技術研究センター 〒135-0062 東京都江東区東雲 1-10-13 Tel: 03-3536-7283 Fax: 03-3536-7318

にはむしろ逆のことが多いと言えます。かたよりの大きさについての手掛かりは、自分の測定データをいくら眺めても見つからないのです。測定結果に含まれているかたよりの大きさを知ることは大変難しいのです。かたよりは、科学技術の新たな発展、別のグループとの測定結果の比較、あるいは別の測定原理との比較などによって初めて目に見える形で姿を現すことになります。

一方、ばらつきはそれが実際に統計学で想定されている偶然誤差のように振る舞うならば、その対策はむしろ容易です。まず、その大きさは自ら採取可能なデータによって見積もることができます。その低減が必要であれば、測定方法や測定環境に対して何らかの対策をとろうということになるでしょう。それが技術的に難しいことがわかったときには、最後の対策として、多数個のデータをとってその平均を求めるという手段が残されています。n 個のデータの平均値はもとのデータと

比較するとばらつきが $1/\sqrt{n}$ 倍に減少するというのが統計学の教えるところです。

不確かさ評価では、ばらつきやかたよりの低減が目的ではなく、それらの大きさをありのまま評価することが重要です。かたよりはばらつきよりも評価しにくいという観点からは、かたより的な成分もできるだけばらつき的な成分として評価すればよいことになります。母集団としてできるだけ広いものを想定し、単なる繰り返しのばらつきだけでなく、測定日や測定装置などに類する様々な要因をとりあげた反復測定のばらつきを評価する実験を行うのが望ましいということが言えます。

4.7 まとめ

以上、あらためて整理すると、

- (1) ばらつきには一般に階層構造がある。繰り返し測定のばらつきと測定法固有のかたよりの間の中間的な変動成分が存在する。
- (2) それがかたよりかばらつきかは測定データの背後に想定する母集団の選択に依存する。
- (3) 現実の測定で固定されている要因であっても、不確かさ成分として考慮しなければならないことがある。
- (4) 「かたよりは補正可能、ばらつきは対処不可能」という考え方は適当ではない。現実には、かたよりはばらつきより、評価も低減対策も難しいことが多い。
- (5) 不確かさ評価では、繰り返し測定だけでなく、反復測定を行うことにより、評価の難しいかたより的成分を、統計的に評価可能なばらつき的成分として評価することができる。

なお、GUM では、かたよりとばらつきを系統誤差、偶然誤差とは呼ばず、系統効果(systematic effect)、偶然効果(random effect)と呼んでいます。かたよりとばらつきの区別は母集団の想定次第で変わり得るため、誤差をその性質に応じて固定的に区別することを避ける意図があるものと考えられます。



財団法人 国際超電導産業技術研究センター 〒135-0062 東京都江東区東雲 1-10-13 Tel: 03-3536-7283 Fax: 03-3536-7318

【隔月連載記事】

やさしい「不確かさ」のおはなし(その4)

産業技術総合研究所 計測標準研究部門 応用統計研究室 榎原研正

5. 標準不確かさの A タイプ評価

標準偏差相当の大きさとして表現した不確かさを標準不確かさと呼ぶことはすでにお話ししました。標準不確かさの評価の方法を、GUM は次の 2 種類に分類しています。

A タイプ評価 = 一連の観測値に対する統計的解析による方法

Bタイプ評価 = 統計的解析以外による方法

A タイプ、B タイプというのは身も蓋もない呼び方ですが、ばらつき、かたよりという不確かさの性質によらず、評価の方法で分類している点は重要なポイントと考えられます。評価方法で分類したのは、前節でやや長々と述べたようにばらつきとかたよりの区別を厳密に行うことはできないことと、GUM は評価の手続きを定めることを目的に書かれた文書であることから来ていると考えられます。従来の「誤差評価」において、誤差をばらつき(偶然誤差)とかたより(系統誤差)に分類しているのとは、分類の考え方が違っています。

家庭用血圧計の例(3 節)でいうと、繰り返し測定のばらつきは、複数のデータから実験標準偏差を求めているため A タイプ評価、目盛りのずれは外部情報にもとづいているため B タイプ評価に分類されます。今回は A タイプ評価の方法について解説します。

5.1 平均値の不確かさ

測定量Xに対して、n回測定を繰り返して得た一連のデータを $[x_1, x_2, \dots, x_n]$ とします。これらの平均値と実験標準偏差をそれぞれ \bar{x} ,sとします:

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{1}$$

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2}$$
 (2)

Xに対する最終的な測定結果を \bar{x} とする場合、繰り返しのばらつきに起因するその標準不確かさは次の式で求められます。

$$u(\bar{x}) = \frac{s}{\sqrt{n}}$$
 (A タイプ評価の基本式) (3)

これが A タイプ評価の最も基本的な式です。この式右辺の $1/\sqrt{n}$ と式 (2) の $1/\sqrt{n-1}$ の因子は少し 紛らわしいので注意が必要です。前者は \bar{x} が n 個のデータの平均であることから、後者は平方和 $\sum (x_i - \bar{x})^2$ の自由度が (n-1) であることから来ています。

現実の測定では、必ずしも繰り返しが行われるとは限りません。1回の測定で得たデータ x1をそ



財団法人 国際超電導産業技術研究センター 〒135-0062 東京都江東区東雲 1-10-13 Tel: 03-3536-7283 Fax: 03-3536-7318

のまま測定結果とすることはよくあります。その場合、式 (3) の分母ではn=1とおけばよいのですが、データ 1 個では実験標準偏差 s を求めることはできません。また、n が 1 でなくとも、2 や 3 といった小さな数の場合には、求まった s は信頼のおける値になりません。このような場合、s を求めるための実験をあらかじめ別に行うことになります。あらかじめ行ったm回の繰り返し測定で得られたデータを $[q_1, q_2, \cdots, q_m]$ とするとき、実験標準偏差 s はこれらから、

$$s = \sqrt{\frac{1}{m-1} \sum_{i=1}^{m} (q_i - \overline{q})^2}$$
 (4)

これを式 (3) に代入することにより

$$u(\bar{x}) = \sqrt{\frac{1}{n(m-1)} \sum_{i=1}^{m} (q_i - \bar{q})^2}$$
 (5)

となります。これは、測定結果 \bar{x} を求めるためのデータセットと不確かさを求めるためのデータセットが異なる場合に、式 (3) を一般化したものと考えることができます。信頼性の高い \bar{s} を求めるためには繰り返し数 \bar{m} を大きくすれば良いのですが、どれ程大きければ良いかについては囲み記事 [5] を参照下さい。

5.2 因子 $1/\sqrt{n}$ について

基本式 (3) における $1/\sqrt{n}$ の意味について考えます。いま、血圧測定のばらつきを調べようとしているものとし、血圧のデータを次の 2 つの方法で記録するものとします。

[方法 A] 1回毎の測定データをそのまま記録する

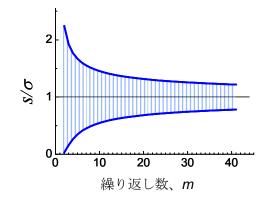
[方法B] 4回繰り返し測定したときの平均値を1回分のデータとして記録する

[5] A タイプ評価における測定の繰り返し数について

実験標準偏差を求めるための測定の繰り返し数mをどの程度大きくとれば良いかについて、定説は実はありません。容易に繰り返し測定ができる場合には、大まかな目安としてm=10程度をここでは推奨したいと思います。

無限回測定を行ったと仮想したときのデータの集合(母集団)の標準偏差が母標準偏差(σ)です。実際に無限回測定は行えませんので、母標準偏差は不可知量です。実験標準偏差sは母標準偏差sの推定値として用いられます。sはsのまわりでばらつきますが、sが小さいほど、このばらつきが大きく、従って推定の信頼性が低くなります。

sがどの程度ばらつくかを見るため、sの分布の 95% を含む範囲を繰り返し数 m の関数として示したものが下の図です。縦軸は s/σ で、この値が 1 であれば s が σ を正しく推定していることになります。図の下側の曲線が 2.5 パーセンタイル、上側が 97.5 パーセンタイル、両曲線の間が 95%をカバーする区間で、区間の上下幅が狭



図A1 実験標準偏差sの95%存在範囲



財団法人 国際超電導産業技術研究センター 〒135-0062 東京都江東区東雲 1-10-13 Tel: 03-3536-7283 Fax: 03-3536-7318

いほど、s の信頼性が高いといえます。およそ m=10 程度までは、m が増えるにつれて区間幅は速やかに狭くなっていきますが、m=10 を超えるあたりでそのスピードが鈍ります。これ以上 m を増やしても労多くして功少なしとなることが、m=10 程度が妥当ではないかと考える理由です。

m を繰り返し数と呼びましたが、何らかの測定条件を変えた反復の回数についても事情は同じです。単純な繰り返しと比べると反復の回数を大きくとることは一般に難しく、*m* = 10 を実現することは難しいことも多いと考えられます。このようなときには、不確かさ評価結果の信頼性とそのために必要となる労力を天秤にかけて、*m* の大きさを決める必要があります。

図 1 にこの様子を示します。記録を無限回行ったと仮想したときの値の集合は測定値の母集団で、その分布は図 1 下部のような確率密度関数 p(x)で表すことができます。方法 A の母平均を μ 、母標準偏差を σ とすると、方法 B では母平均は μ のままで、母標準偏差が $\sigma/\sqrt{n}=\sigma/2$ と半分になることが証明できます(囲み記事 [6] 参照)。

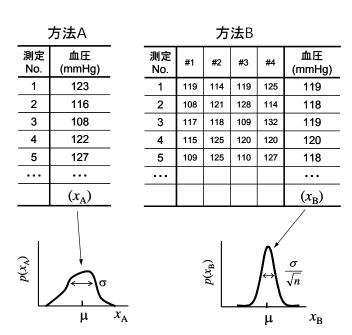


図1 生データの分布と平均値の分布

[6] 試料平均の母平均と母標準偏差

個々の測定データ $x_i(i=1\sim n)$ は、母平均を μ (= $E[x_i]$) として次のように表すことができます。

$$x_i = \mu + \varepsilon_i \tag{A1}$$

 ε_{i} は誤差で、互いに独立、期待値はゼロ、分散は ε_{i}^{2} と仮定します($E[\varepsilon_{i}]$ =0, $V[x_{i}]$ = $V[\varepsilon_{i}]$ = ε_{i}^{2})。 $E[\varepsilon_{i}]$ がゼロでないときは、それを μ に繰り込んだ残りを ε_{i} と再定義すれば良いので、 $E[\varepsilon_{i}]$ = 0 は妥当な仮定です。なお、 $E[\varepsilon_{i}]$ の記号については、3節の囲み記事で説明したように p(x)を確率変数 x の確率密度関数として、



財団法人 国際超電導産業技術研究センター 〒135-0062 東京都江東区東雲 1-10-13 Tel: 03-3536-7283 Fax: 03-3536-7318

 $E[x] \equiv \int x \, p(x) \mathrm{d}x$

$$V[x] \equiv E[(x - E[x])^2] = \int (x - E[x])^2 p(x) dx$$

がそれぞれxの期待値と分散を表します。 x_i の試料平均は、式(A1)から、

$$\overline{x} = \mu + \sum_{i} \varepsilon_{i} / n \tag{A2}$$

で、これに対して期待値を求めると、 $E[\bar{x}] = \mu$ が容易に得られます。これは平均値をとっても元のデータと母平均は変わらないことを意味しています。一方、 \bar{x} の分散は、

$$V[\bar{x}] = E[(\bar{x} - \mu)^2] = E[(\sum_i \varepsilon_i)^2]/n^2 = E[\sum_{i,i'} \varepsilon_i \varepsilon_{i'}]/n^2 = \sigma^2/n$$
(A3)

となります。ここで、 ε_i が相互に独立との仮定から、 $i \neq i'$ に対しては $E[\varepsilon_i \varepsilon_{i'}]=0$ となることを使っています。従っ

て \bar{x} の母標準偏差は

$$\sqrt{V[\overline{x}]} = \sigma / \sqrt{n} \tag{A4}$$

となり、もとのデータの母標準偏差の $1/\sqrt{n}$ 倍になることがわかります。

平均値をとるとその母標準偏差が小さくなるというのは重要な性質です。これが何故起きるかですが、図1の例で例えば108 mmHg というのは方法 A におけるデータの中で極端に小さい値です。 方法 B で 108 mmHg という値が記録されるためには、4回の測定で連続して108 mmHg 付近の値が出る必要があります。極端な値が4回連続する確率は当然小さく、結果として平均値は中心付近の値となる確率が高くなります。このため確率分布の分布幅が狭くなる、すなわち母標準偏差が小さくなることになります。

以上から、n個のデータの平均値 \bar{x} の標準不確かさとして $u(\bar{x}) = \sigma/\sqrt{n}$ とすれば良いのですが、 母集団のパラメータである σ は不可知量ですので、 σ をその推定値 \bar{s} で置き換えた式 (3) が基本式として用いられるのです。これが式 (3) 右辺における因子 $1/\sqrt{n}$ の由来です。

平均値をとることのもう一つの効果として、もとの確率分布の形状が(ほぼ)どのようなものであっても、nが大きくなるとともに平均値の分布形状は正規分布に近づくという性質があります。これは統計学において中心極限定理として知られている有名な定理です。測定データを統計的に解析する場合に、しばしば正規分布が仮定される根拠はこの定理にあります。

5.3 ばらつきに構造がある場合の A タイプ評価

短期間の繰り返し測定のばらつきをε、日によるばらつきをγとして、測定データが

$$\mathbf{x} = \mu + \gamma + \varepsilon \tag{6}$$

とモデル化できる場合を考えます。ばらつきに階層構造がある例です。ここで μ は母平均です。 γ と ϵ の期待値はゼロ、分散はそれぞれ σ_r^2 , σ_ϵ^2 と表記します($V[r] = \sigma_r^2$, $V[\epsilon] = \sigma_\epsilon^2$)。r 日にわたり、各日ごとにn回の繰り返し測定を行ったときの平均値を測定結果として用いるものとします。個々の測定データ x_i は

$$x_{ij} = \mu + \gamma_i + \varepsilon_{ij}$$
 $(i = 1 \sim r, j = 1 \sim n)$ (7)



財団法人 国際超電導産業技術研究センター 〒135-0062 東京都江東区東雲 1-10-13 Tel: 03-3536-7283 Fax: 03-3536-7318

と表せます。nr個全部のデータの平均値は

$$\overline{\overline{x}} = \frac{1}{nr} \sum_{i=1}^{r} \sum_{j=1}^{n} x_{ij} = \mu + \frac{1}{r} \sum_{i=1}^{r} \gamma_i + \frac{1}{nr} \sum_{i=1}^{r} \sum_{j=1}^{n} \varepsilon_{ij}$$
(8)

で、これを用いると \bar{x} の分散が、囲み記事 [6] と同じ考え方により、次のように求まります。

$$V[\overline{\overline{x}}] = \frac{{\sigma_{\gamma}}^2}{r} + \frac{{\sigma_{\varepsilon}}^2}{nr}$$
 (9)

 σ_r^2 , σ_ϵ^2 はいずれも母集団のパラメータで不可知量なので、これらを何らかの方法により推定したものをそれぞれ \mathbf{s}_r^2 , \mathbf{s}_ϵ^2 と表記します。 \overline{x} の標準不確かさの内、ばらつきに起因する成分は以上から次のようになります。

$$u(\overline{\overline{x}}) = \sqrt{\frac{{s_{\gamma}}^2}{r} + \frac{{s_{\varepsilon}}^2}{nr}}$$
 (10)

これは、ばらつきに 2 層の階層構造がある場合の A タイプ評価の基本式です。 \bar{x} は、日間ばらつきについては r 個の、繰り返しばらつきについては nr 個の平均になっているため、式 (10) 右辺の分散のそれぞれの分母に、対応する因子が現れることがポイントです。

式 (10) に則って A タイプ評価を行うためには、 $s_r^2 \geq s_\epsilon^2$ を分けて評価する必要があります。単純に x_i の実験分散を計算してもこれらを求めることはできず、分散分析といわれる統計的手法を用いる必要があります。分散分析は x_i を対象に行うことも可能ですし、n, r の大きさが十分でない場合には、別途行った実験で求めたデータ q_i によることも可能です。これは式 (3)と(5) の関係に対応します。分散分析については今回は説明する余裕がありませんが、いずれ機会があれば触れたいと思います。

財団法人 国際超電導産業技術研究センター 〒135-0062 東京都江東区東雲 1-10-13 Tel: 03-3536-7283 Fax: 03-3536-7318

【隔月連載記事】

やさしい「不確かさ」のおはなし(その5)

産業技術総合研究所 計測標準研究部門 応用統計研究室 榎原研正

6. 標準不確かさの B タイプ評価

6.1 B タイプ評価

標準不確かさの2つの評価方法のうち、今回はBタイプ評価について説明します。Bタイプ評価とは、統計的解析以外の方法によるものと GUM は定義しています。すなわち、Aタイプでないものすべて、ということになります。

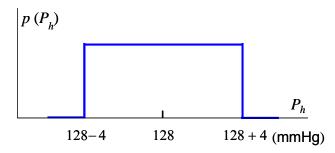


図1 正しい目盛りの血圧計で得られると予想さる測定結果に対する確率分布

家庭用血圧計の例(3節)に戻ると、血圧計の仕様書における「精度±4 mmHg」の記述を「目盛りのずれは±4 mmHg を超えることはない」とまず解釈しました。これから、もし厳密に正しい目盛りによって測定したと仮定したときの測定結果は、実際に得られた値 128 mmHg を中心とする±4 mmHg の範囲にあると考えることができます。この範囲内でどの値がより確からしいかの情報はないため、これを表現する確率分布として図1(再掲)のような一様分布が想定できます。この分

布の標準偏差(一様分布の場合、半幅/ $\sqrt{3}$)から、血圧の測定結果 P_h の標準不確かさ $u_2(P_h)$ を

$$u_2(P_h) = \frac{4}{\sqrt{3}} = 2.31 \text{ mmHg}$$

と求めることができます。このように、何らかの入手可能な情報から確率分布を想定し、その標準 偏差を求めるというのが、典型的な B タイプ評価の方法です。

情報源としては、仕様書以外に、測定器の校正証明書、理科年表の類のデータ集、測定系や測定対象に関して専門家が有する経験や知識、過去のデータなどがあり得ます。このような情報は、不確かさ評価で用いられることを前提に提供されているものばかりではないので、それを確率分布に置き換える際に主観が入ったり、曖昧さが残ったりする場合があるのはある程度やむを得ません。例えば、血圧計の仕様書の「精度」の中に、血圧計側に原因があるばらつきの効果が含まれているかも知れません。すると A タイプ評価した不確かさと部分的に二重カウントしている可能性があり



財団法人 国際超電導産業技術研究センター 〒135-0062 東京都江東区東雲 1-10-13 Tel: 03-3536-7283 Fax: 03-3536-7318

ます。

この対策として、Aタイプ評価の際に、このようなばらつきの効果を、血圧自体の変動とは分離して評価する実験を行うことが考えられます。これにより、二重カウントした場合とそうでない場合で結果にどのような差が出るかは少なくともわかります。実際にこのような実験する価値があるかどうかは、不確かさをどこまで厳密に評価する必要があるか、対象となっている不確かさ成分が他の成分と比べて無視し得ない寄与を持ちそうかどうかの専門家の予測などにもとづき判断することになります。血圧計の場合は、血圧計側のばらつきよりも被測定側の血圧自体の変動の方が大きいと考えられ、血圧計のばらつきが仮に二重カウントとなっていても、大きな過大評価にはつながらないと考えられます。

6.2 B タイプ評価で用いられる確率分布

一様分布は、不確かさ評価ではよく用いられる確率分布です。上下の限界値で不連続な飛びを有する点で、現実の何らかの現象を表現する分布としては不自然で、むしろ手元にある知識の状態を表現したものと考えるべきものです。数学的理屈をつければ、確率変数が上下限の範囲内にあるという制約条件のもとで情報量のエントロピーを最大にする分布が一様分布となります。これは最大エントロピー法という考え方の適用例です。

一様分布がよく使われるのは、考え得る最も簡単な確率分布であることと、限界値 $\pm \Delta$ から求めた標準偏差 σ が、他の一般的な分布(例えば、三角分布ならば $\sigma = \Delta/\sqrt{6}$ 、 $\pm 3\sigma$ を限界値とみなした正規分布ならば $\sigma = \Delta/3$)と比べて大きめとなり、安全側に評価したい場合には好都合という事情によると思われます。ただし不確かさの過大評価が望ましいわけではありません。

Bタイプ評価では、一様分布以外に図 2 に示すような確率分布が想定されることもあります。外部からの情報は、しばしば $\pm \Delta$ 以内という存在範囲の限界の形で解釈できることが多いため、図では確率分布の特徴的な点と存在範囲 $\pm \Delta$ を関係づけて示しています。

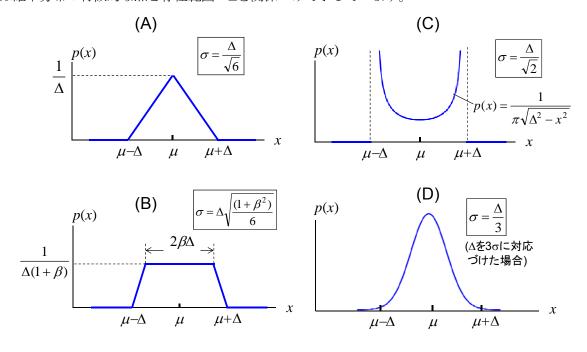


図2 Bタイプ評価における代表的な確率分布



財団法人 国際超電導産業技術研究センター 〒135-0062 東京都江東区東雲 1-10-13 Tel: 03-3536-7283 Fax: 03-3536-7318

図 **2(A)** の三角分布は、中心に近い点ほど可能性が高いと思われる場合に適当な分布で、一様分布のような不連続な飛びが生じません。標準偏差の大きさは $\Delta/\sqrt{6}$ です。

図 **2(B)** は台形分布で、一様分布と三角分布の中間的な分布です。 β は上底の下底に対する比で、 β = **1** では一様分布、 β = **0** では三角分布に一致します。一様分布において、限界値 Δ の値自体の情報が曖昧であるとして Δ の値に一様分布を想定し、一様分布二つの畳み込み積分を行うことからもこの分布が出てきます。標準偏差は $\Delta\sqrt{(1+\beta^2)/6}$ です。

図 2(C) は U 字型分布で、例えば温度がフィードバック制御されているなどのために時間的に正弦波的にふるまう場合、温度の分布はこのような形で表されます。 $\pm \Delta$ の限界値に近づくほど大きくなるのは、正弦波では最大振幅付近の滞在時間が長いためです。標準準偏差は $\Delta/\sqrt{2}$ です。

正規分布の場合 $\pm 3\sigma$ の範囲をはずれる確率は0.3%以下なので、 $\pm 3\sigma$ を「ほぼ確実にこの範囲内に入る限界値」と解釈して $\pm \Delta$ と同一視すると $\sigma = \Delta/3$ となります。図2(D)はこのような正規分布を表しています。

情報として与えられている値が何シグマに相当するかが最初から明瞭な場合には、必ずしも確率 分布の形を想定する必要はありません。拡張不確かさが記載された、測定器の校正証明書などがこ のような場合に相当します。

同じように確率と言っても、A タイプ評価に付随して考える確率分布(例えば第 4 回の図 1)と B タイプ評価で想定する上図のような確率分布では、確率の意味が微妙に違います。この点につい ては、囲み記事[7]で説明します。

6.3 A タイプ評価と B タイプ評価の区別

不確かさ評価がAタイプなのかBタイプなのかを厳密に分類することに実際上の意義はありませんが、過去のデータを用いた評価がAタイプなのかBタイプなのかがときどき議論の対象になりますので、ここで整理しておきます。GUMに厳密に従うと、

- (1) 測定結果 \bar{x} を求めるためのデータセット $\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n\}$ と $u(\bar{x}) (= s/\sqrt{n})$ のs (実験標準偏差) を求めるデータセットが同じ場合が本来のAタイプ評価
- (2) これに準じるケースとして、別のデータセットでsを求めているが、 \bar{x} とsは、管理状態にあることが確認された同一の測定系で得られている場合もAタイプ評価

で、それ以外はBタイプ評価に分類されます。つまり、Aタイプ評価は、測定結果を出すためのデータの背後に想定される母集団と同じ母集団からサンプリングされたと考えられるデータを用いて実験標準偏差を計算している場合に限定されます。

単に「実験を行った場合」にAタイプと呼ぶということではない点にも注意しておきたいと思います。例えば後述する感度係数(偏微分係数)を計算によらず実験的に求める場合がありますが、それとAタイプかどうかは無関係です。

A タイプとするか B タイプとするかは、単に定義の問題ですので、繰り返しますが、これらを厳密に分類することに大きな意義はありません。

6.4 A タイプ評価から B タイプ評価へ

A タイプ評価と B タイプ評価のどちらが信頼性が高いとか、どちらが望ましいとかいうことが一般論としてあるわけではありません。特定の不確かさ成分に着目したとき、それをどちらの方法で



財団法人 国際超電導産業技術研究センター 〒135-0062 東京都江東区東雲 1-10-13 Tel: 03-3536-7283 Fax: 03-3536-7318

評価すべきかは、多くの場合ほぼ自動的に決まります。もしどちらの方法も利用可能ならば、得られる不確かさ自体の信頼性と評価のコストを勘案して検討することになります。

不確かさ評価の事例が蓄積され、社会的に共有されるようになれば、B タイプ評価で利用できる情報が増えるはずです。これにより A タイプの不確かさ成分の割合が少ない(あるいは全く含まれない)評価(ただし最小限の性能チェックはむろん必要でしょうが)が行える可能性が増えると考えられます。評価コストは一般に A タイプの方が高いので、生産現場での日常測定のように、不確かさ評価を厳密に行うより、コストがかからない簡便な評価を行うことが重要であるような局面でも不確かさが利用されるようになるためには、B タイプ評価が行いやすい状況が拡がることが望まれます。

基礎物理定数を集めた CODATA データベース (http://www.codata.org/) はこの点では優等生で、リストされているすべての基礎物理定数に不確かさが付与されています。また、最近の測定器の校正証明書では、校正値に対する不確かさが記載されたものが多くなっています。一方、理科年表などの古くからあるデータ集では、不確かさの手掛かりになるような情報はまだ手薄です。今後改善されてゆくことを期待したいと思います。

6.5 過大評価の問題

B タイプ評価では、手持ちの情報をどのように料理するかについて、何らかの主観的判断が関与することがしばしばあります。その際、安全側評価との観点で、不確かさが過大に評価されることがよくあります。逆に、稀なケースですが、市場競争の厳しい測定器のメーカの立場で競争を有利に進めるなどのため、不確かさを小さめに評価するケースもあります。

健康診断で、血液のある成分の分析値が用いられている場合を想定します。分析値には誤差が含まれ得るため、本来なら正常とすべき人を異常ありと判定する誤り(あわて者の誤り、あるいは第一種の誤りとも呼びます)と、本来ならば異常ありとすべき人を正常としてしまう誤り(ぼんやり者の誤り、あるいは第二種の誤り)の2通りの誤りが生じ得ます。ぼんやり者の誤りの方が罪が深いことは明かです。従って、分析の不確かさの大きさに応じて、正常と判定する範囲を不確かさと同程度の大きさだけ狭い範囲に設定することは合理的です。不確かさの過大評価が安全側というのはこの意味です。しかし必要以上に狭くすると、少しのことで再検査に回される人が増え、無駄な再検査が増えるだけでなく、検査の信頼性が失われて、異常との結果が出ても再検査にやってこない人が増える可能性があります。判定限界値は、2種類の誤りで生じる損失のバランスを考慮したリスク解析により設定すべきで、これを合理的に行うには、リスク算定の基礎となる不確かさの大きさは過大でも過小でもなく評価されている必要があります。不確かさ自身に不確かさがあることへの配慮も、リスク解析の中で行うのが合理的です。

原子炉の熱出力の拡張不確かさは現在 2 %程度と言われており、このうちのほとんどは、冷却水の流量を求める流量計の不確かさに起因すると言われています。この不確かさの存在を見込んで、可能な最大出力から少しだけ低めに実際の出力が設定されます。流量計の不確かさを低減することができれば、熱出力をそれだけ増加させることが可能になります。流量計の不確かさが 1 %に減らせれば日本全体で原子炉 1 基の半分ぐらいの出力をここから生み出すことが可能と考えられています。原子炉 1 基の建設費用が数千億円ですので、不確かさの低減が莫大な経済効率の向上につながることがわかります。過大でも過小でもない、妥当な不確かさ評価を行うことが、効率化を進めるためにはたいへん重要になります。



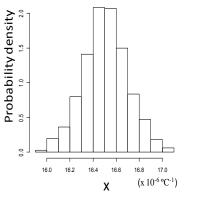
財団法人 国際超電導産業技術研究センター 〒135-0062 東京都江東区東雲 1-10-13 Tel: 03-3536-7283 Fax: 03-3536-7318

[7]古典統計とベイズ統計 — 頻度基準の確率と主観確率

A タイプ評価では、測定量の真の値は(不可知ではあっても)一意的に定まった値であり、逆に測定データは何らかの確率分布に従う確率変数と考えます。多数の測定データをとってヒストグラムを作ることにより、ある測定値はどのような確率で生じるかを実験的に確認することができます。測定データの個数を無限個とした極限で、ヒストグラムから確率密度関数が得られます(図 A-1)。この意味で A タイプ評価で考えている確率は、「頻度基準の確率」と呼ばれます。無限回の繰り返し測定は不可能なので、実際には図 A-1 右側のような確率分布は不可知ですが、頑張れば原理的には実証可能というタイプの確率です。真値は(不可知の)確定数、測定値は確率変数という点がポイントで、これは古典統計学の考え方の中心となる枠組みです。

一方、B タイプ評価で想定される、本文図 1 や図 2 のような確率分布は、実験による頻度分布から推定したものではなく、手持ちの情報をもとに一般に主観をまじえて想定したものです。このような確率は、「主観確率」あるいは「ベイズ流の確率」と呼ばれ、対応する分布は先験的分布と呼ばれます。本来一意的であるはずの真値をも確率変数として扱う点がポイントで、このような確率はベイズ統計学において取り扱われているものです。

A タイプ評価と B タイプ評価で確率の数学的意味がこのように違いますが、このために現実の不確かさ評価で大きな混乱が生じるということはありません。現実の測定には、統計学で取り扱える部分とそうでない部分が存在することを反映して、GUM の中に二つの考え方が妥当なバランスで共存していると思います。



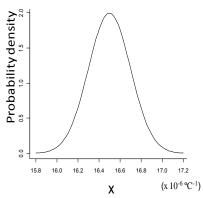


図 A-1 ヒストグラムと確率密度関数



財団法人 国際超電導産業技術研究センター 〒135-0062 東京都江東区東雲 1-10-13 Tel: 03-3536-7283 Fax: 03-3536-7318

【隔月連載記事】

やさしい「不確かさ」のおはなし(その6)

産業技術総合研究所 計測標準研究部門 応用統計研究室 榎原研正

7. 一般的な不確かさの合成を含む例 一液体の体積測定

第2回目に紹介した家庭用血圧計の例では、ばらつきとかたよりという不確かさの成分二つを二乗和のルールで合成することを説明しました。この例では、二つの成分がともに血圧の不確かさ(すなわち「血圧」の次元で表される)となっており、不確かさの合成という点では最も単純な例です。 最終回となる今回は、もっと一般的な合成の方法を含む例として、次の液体の体積測定の例について不確かさ評価を行います。

ある液体の体積 v を、その質量と密度から求めることとします。まず、液体の質量を天秤で 5 回繰り返し測定し、測定データ [100.0 g, 100.3 g, 99.9 g, 99.7 g, 100.1 g] を得ました。またこの天秤の内蔵分銅の値の不確かさが、天秤の校正証明書に記載された拡張不確かさとして U = 0.10 g (k = 2) とわかっているものとします。一方、液体の組成は既知であり、その密度に関する次の情報がデータブックに記載されているものとします。「この液体の密度値は 2.00 g/cm³ である。この値の誤差は \pm 0.01 g/cm³ を越えない。」

なお、この例は説明のためのモデルで、実際の測定での数値がこのような値になるということではありません。また現実の不確かさ評価では、天秤測定での風袋の効果や液体密度の温度依存性など、他にも考慮すべき要因があり得ますが、話を簡単にするためにこれらは考えないこととします。 不確かさ評価は次の順序で進めます。

7.1 測定の数学的モデルの設定

液体の体積 v は、質量の繰り返し測定値 m_i ($i=1\sim5$) の平均値を \overline{m} 、データブックに記された密度値を ρ_0 として、次の式で計算することになります。

$$v = \frac{\overline{m}}{\rho_0} \tag{1}$$

最終的な測定量を、その値を導くために使用する別の量 (GUM では入力量と呼ばれます)の関数として表したこのような式を、「測定の数学的モデル」と呼んでいます。測定の数学的モデルは、測定方法や手順に応じて違ったものとなります。例えば、液体の体積をメスシリンダで測定する場合、式 (1) とは異なるモデルとなります。測定の数学的モデルの設定は、大変重要なステップで、モデルがうまく設定できていると、後の評価をスムーズに進めることができます。

現実の測定では、数学的モデルをうまく設定することは意外に難しいことが多く、例えば、観測データを非線形の最小二乗法で理論曲線にあてはめることにより最終的な測定量が求まるといった複雑な場合のように、測定量を入力量の陽関数として表せないことがあります。また逆に、1回の測定の読み値をそのまま測定結果として使う(ただし測定誤差の要因としてはいろいろ考えられる)といった一見単純そうなケースでも、数学的モデルをどう書き下すかにとまどう場合があります。ここでは詳述する余裕がありませんが、このような場合についても数学的モデルの設定について一定の処方が存在します。



財団法人 国際超電導産業技術研究センター 〒135-0062 東京都江東区東雲 1-10-13 Tel: 03-3536-7283 Fax: 03-3536-7318

 \overline{m} = 100.0 g, ρ_0 = 2.00 g/cm³ ですから、式 (1)より測定結果は v = 50.0 cm³ と求まります。不確かさ評価は、この 50.0 という数値にどの程度の曖昧さがあるのかを定量的に知ることが目的です。

7.2 不確かさの伝播則

測定の数学的モデルが一般に

$$y = f(x_1, x_2, ..., x_N)$$
 (2)

で表されるとき、測定結果 y の合成標準不確かさ $u_c(y)$ と入力量の値 x_i の標準不確かさ $u(x_i)$ の間に次の関係が成立することが示せます(囲み記事 [8] 参照)。

$$u_c^2(y) = \sum_{i=1}^N \left[\frac{\partial f}{\partial x_i} \right]^2 u^2(x_i)$$
 (3)

この式は、従来から誤差の伝播則として知られていたものと本質的に同じものです。この式が成立する条件については囲み記事 [8] を参照下さい。

なお $c_i = \partial f/\partial x_i$ のことを GUM では「感度係数」と呼んでいます。f の関数形がわかっているときには偏微分をすれば求まりますが、関数形が簡単な形で書き下せない場合には、数値微分や感度解析の実験によって感度係数を求める手もあります。

式(3)を(1)に適用すると次が得られます。

$$u_c^2(v) = (\frac{1}{\rho_0})^2 \cdot u^2(\overline{m}) + (\frac{\overline{m}}{\rho_0^2})^2 \cdot u^2(\rho_0)$$
 (4)

[8]不確かさの伝播則について

伝播則を導出するとともに、伝播則のパリエーションについて説明します。測定の数学的モデル $y=f(x_1, x_2, ..., x_N)$ において、 x_i が真値(GUM に忠実に言うと x_i の期待値) μ_i から微少量 α_i だけ離れている($x_i=\mu_i$ + α_i)として μ_i のまわりに Taylor 展開し、 α_i について 1 次まで残すことにより、

$$\varepsilon_{y} = \sum_{i} \frac{\partial f}{\partial x_{i}} \varepsilon_{i} \tag{A1}$$

が得られます。ただし、 ε_r は y と真値 μ_y (= $f(\mu_1, \mu_2, ..., \mu_N)$) の差(つまり測定誤差)です。式(A1)の両辺を二乗し、その期待値をとります。ここで、 ε_r 及び ε の分散をそれぞれ σ_y^2 , σ^2 とし、 ε どうしは互いに独立(これは期待値の記号を $E[\]$ とすると、 $i \neq j$ について $E[\varepsilon_i \varepsilon_i] = 0$ を意味します)とすると

$$\sigma_y^2 = \sum_i \left(\frac{\partial f}{\partial x_i}\right)^2 \sigma_i^2 \tag{A2}$$

が得られます。分散 σ^2 は不可知量ですが、これをその推定値 $u^2(x_i)$ でおきかえることにより、本文の式(3) の伝播則が得られます。

以上からもわかるように、式(3)の伝播則は厳密には、(i) 各siが十分に小さくて $f(x_1, x_2, ..., x_N)$ の 1次までの Taylor 展開が良い近似になっている、(ii) 各siが(従って xiが)独立、の二つの条件のもとで成立します。 xi どうしが互いに独立でないときには、式(3)にかえて次の伝播則が得られます。



財団法人 国際超電導産業技術研究センター 〒135-0062 東京都江東区東雲 1-10-13 Tel: 03-3536-7283 Fax: 03-3536-7318

以上からもわかるように、式(3)の伝播則は厳密には、(i) 各 ϵ :が十分に小さくて $f(x_1, x_2, ..., x_N)$ の 1 次までの Taylor 展開が良い近似になっている、(ii) 各 ϵ :が(従って x:が)独立、の二つの条件のもとで成立します。 x:どうしが互いに独立でないときには、式(3)にかえて次の伝播則が得られます。

$$u_c^2(y) = \sum_{i=1}^N \left[\frac{\partial f}{\partial x_i} \right]^2 u^2(x_i) + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} u(x_i) u(x_j) r(x_i, x_j)$$
(A3)

ここで、 $r(x_i, x_i)$ は入力量間の相関係数の推定値を表します。

実際の測定では、測定の数学的モデルが、 $c \ge p_i (i=1, 2, ..., N)$ を定数として、

$$y = c \cdot x_1^{p_1} \cdot x_2^{p_2} \cdots x_N^{p_N}$$
 (A4)

のように書けるケースがしばしば生じます。実際、液体の体積測定(本文式(1))もこのような例の一つです。 この場合、伝播則(式(3))は次の形に変形できます。

$$\left(\frac{u_c(y)}{y}\right)^2 = \sum_i \left(p_i \frac{u(x_i)}{x_i}\right)^2 \tag{A5}$$

これは相対不確かさの伝播則と呼ばれています。これが利用できる場合には、通常の伝播則よりも計算が楽 で、見通しも良くなります。

7.3 質量の標準不確かさ $u(\overline{m})$

質量の測定結果 \overline{m} の不確かさ $u(\overline{m})$ は、天秤測定のばらつきによる成分 $u_{r}(\overline{m})$ と、内蔵分銅の値の不確かさから生じる成分 $u_{w}(\overline{m})$ からなると考えられます。前者は、血圧計の場合のAタイプ評価と同様の考え方で、5 個のデータ [100.0 g, 100.3 g, 99.9 g, 99.7 g, 100.1 g] から求まる実験標準偏差 s=0.223 g を使って

$$u_{\rm r}(\overline{m}) = \frac{0.223}{\sqrt{5}} = 0.100 \text{ (g)}$$

となります。分母の $\sqrt{5}$ は \overline{m} が 5 個のデータの平均値であることから来ています。一方、 $u_{\rm w}(\overline{m})$ は、内蔵分銅の値に対する拡張不確かさが包含係数を k=2 として U=0.10 g とわかっていますので、

$$u_{\rm w}(\overline{m}) = \frac{0.10}{2} = 0.05$$
 (g) (6)

となります。

ちなみに、個々の測定データ m_i に対する不確かさ $u(m_i)$ (これは実際には評価の必要はありませんが)と $u(\overline{m})$ の関係は、それぞれの成分について $u_{\mathrm{r}}(\overline{m})=u_{\mathrm{r}}(m_i)/\sqrt{5}$ 、 $u_{\mathrm{w}}(\overline{m})=u_{\mathrm{w}}(m_i)$ となりま



財団法人 国際超電導産業技術研究センター 〒135-0062 東京都江東区東雲 1-10-13 Tel: 03-3536-7283 Fax: 03-3536-7318

す。後の式は、内蔵分銅の誤差はかたよりとしてどの m_i にも同じように働くことを意味します。

 \overline{m} の全体としての不確かさは、二乗和ルール $u^2(\overline{m}) = u_r^2(\overline{m}) + u_w^2(\overline{m})$ で合成して、

$$u(\overline{m}) = 0.111 \quad (g) \tag{7}$$

と求まります。

なお、 $u_i(y) = |\partial f/\partial x_i| u(x_i)$ を定義すると、 $u_i(y)$ は x_i の不確かさにより生じる yの不確かさ成分と

解釈することができ、式(3)は $u_c^2(y) = \sum u_i^2(y)$ となります。この形で書くと、伝播則はすでに述べた二乗和ルールと本質的に同じであることがわかります。不確かさの合成では数多くの成分が出てきて記号が複雑になることがあるので、記号の表し方を統一しておくことは重要です。原則として次のような表記を用いることが勧められます。

- ・標準不確かさには、小文字の u を用いる
- ・必ず u(x), u(y)のように「何の」標準不確かさかを明示する
- ・u(x)が幾つかの成分からなるとき、uに添え字をつけて区別する; $u_1(x)$, $u_{rep}(x)$ など

7.4 密度の標準不確かさ $u(\rho_0)$

密度の標準不確かさ $u(\rho_0)$ はデータブックの情報をもとに求めます。密度 ρ の真の値は、 ρ_0 = 2.00 g/cm³ を中心に 1.99 から 2.01 g/cm³ の範囲内にあると考えられます。他に情報がないため、 ρ の先験的分布として、この範囲内の一様分布を想定することにすると、次が得られます。

$$u(\rho_0) = \frac{0.01}{\sqrt{3}} = 0.00577 \text{ (g/cm}^3)$$
 (8)

7.5 不確かさの合成

あとは一本道で、式 (4) にこれまでに求めた値を全部代入することで $u_c(v)$ を求めることができます。この計算の過程をしばしば表 1 のように整理します。このような表は不確かさのバジェット表 (budget sheet)と呼ばれています。

バジェット表では、(i) 感度係数を含めて全ての数値の単位を明らかにしておくこと、(ii) 質量のように、入れ子の構造(合成で求めた u(m) がさらに $u_c(v)$ の合成のために用いられる)がわかるようにしておくこと、の2点がポイントです。入れ子構造が複雑な場合には、表を分割するか、骨格部分のみをバジェット表にします。

最後に有効数字 2 桁までとって、合成標準不確かさが $u_c(v) = 0.15$ cm³ と求まります。

表 1 液体体積測定における不確かさのバジェット表

要因	記号	標準不確かさ $u(x_i)$	感度係数 $c_i = \partial y / \partial x_i$	測定量の標準不確かさ $u_i(y) = c_i u(x_i)$	備考
質量測定の不確かさ	u(m)	0.111 g	$0.500 \text{ cm}^3/\text{g}$	0.0559 cm^3	_
繰り返し測定のばらつき	$u_r(m)$	0.100 g	—	—	5回の繰り返し測定(Aタイプ評価)
内蔵分銅の不確かさ	$u_w(m)$	0.050 g	_	_	天秤の校正証明書(Bタイプ評価)
密度の不確かさ	$u(\rho_0)$	0.00577 g/cm	$-25.0 \text{ cm}^6/\text{g}$	0.144 cm^3	理科年表(Bタイプ評価)
合成標準不確かさ	$u_c(v)$	_	_	0.154 cm^3	



財団法人 国際超電導産業技術研究センター 〒135-0062 東京都江東区東雲 1-10-13 Tel: 03-3536-7283 Fax: 03-3536-7318

7.6 不確かさの報告

すでに述べたように、最終的な不確かさは合成標準不確かさ $u_c(v)$ 、もしくは拡張不確かさ U のいずれかで報告します。U を使う場合は、「拡張不確かさは、 $U=0.31~{\rm cm}^3$ (ただし包含係数 k=2)」のように表します。用いた包含係数の値は必ず明示するのがルールです。

8. 終わりに

これまで6回にわたって測定の不確かさについて解説する機会を頂きました。当初予定していたにもかかわらず説明の時間がなくなってしまった話題も二、三ありますが、これまでの解説だけで、実際に不確かさ評価を行える測定の例は少なくないと考えています。ただし、測定が複雑になってくるとさらに高度な考察が必要な場合も生じることは当然予想されます。例えば、次のような場合が考えられます。

- ・回帰分析を含む測定 -測定が何らかのパラメータの関数として行われる場合
- ・ばらつきに階層構造があるため、分散分析により各ばらつきを分離して評価する場合
- ・包含係数を決めるのに、t-分布を利用したい場合

これらは、不確かさ評価に固有の問題というより、むしろ一般的な統計的処理の問題です。これらについても解説させて頂く機会があると良いと思いますが、それまでは、一般的な統計学のテキストを参考にしつつ、このような問題にも取り組んで頂けることを期待しています。