測定の不確かさ評価について

—4. 不確かさの活用—

城 野 克 広*

は じ め に

これまで4回に渡り，測定の不確かさについて解説しております．前回まで「不確かさとは何か」と「不確かさの算出手順」についてお話して参りました．最終回である今回は「不確かさの活用」についてお話していきます．最後に，全体を振り返ることができるように，演習問題も準備しました．

4·1 拡張不確かさ

この連載の初回に以下のような問題を提起しました．

弊社にとりまして重要なお客様である山田様が来社されます．12時50分から13時10分の間に受付にいらっしゃる確率が95％とのこと．お待たせするわけには行きません．私の時計のずれが5分以内である確率は95％です．この時計で何時何分になるまでに昼ご飯を食べ終え，受付に行けば，95％以上の確率で山田様をお待たせさせずに済むでしょうか？

初回に述べたように，この数字からだけでは，答は分かりません．しかし，不確かさ評価で使われるいくつかの約束を用いることで，この問題に一応の解を与えることができます．

一番のポイントは3·2節でお話した「正規分布で，95％という確率と幅の大きさが分かっているときには，その半幅を2で割ると標準偏差になる」ということです．厳密には，上の情報には一切正規分布ということは示されていません．その一方で，誤差というものが正規分布しやすいということは3·2節でお話した通りです．このため，不確かさの評価では95％の信頼の水準を持つ区間が与えられている場合，正規分布を用いることが多いのです．

山田様の来所時刻には，12時50分から13時10分という幅があります．この幅の半分である10分を2で割ったものは5分です．すなわち，山田様の来所時刻には13時ぴったりを中心として，5分の標準偏差を持つ正規分布に従うと考えることになります．同じく，私が受付に着く時刻は時計の示す時刻を中心として，2.5分という標準偏差を持つ正規分布に従うと考えることになります．

ここから，ちょっとややこしいので，じっくり考えてみて下さい．山田様が13時のX分前に受付にいらっしゃるとします．私が受付に着く時刻は，時計の指す時刻が13時のY分前のときとします．一方で，そのときの正しい時刻は，13時のZ分前とします．X, Y, Zはプラスの値になることも，マイナスの値になることもあります．$X=-2$の場合は，山田様が13時2分にいらっしゃったということです．

もちろん，私が13時よりどれだけ早く着いたか(Z)が，山田様のそれ(X)よりも大きいならば，私は山田様よりも先に受付に着くことになります．つまり，「ZからXを引いたもの$=Z-X$」が，プラスになっていれば，山田様をお待たせしないということになります．

山田様がいらっしゃる時刻は13時を中心にその前後です．私が受付に着く時刻は時計の指している時刻，つまり13時のY分前を中心に，その前後ということなります．ですから，私と山田様の受付への到着時間の差の平均はY分ということになります．例えば，時計が12時58分を指しているときに私が受付に到着するならば，到着時刻の差は平均としては2分ということです．すなわち，「$Z-X$」の平均はYなのです．さて，標準偏差はどうでしょう．

ここには2つの不確かさの要因があります．2・1節で述

* 産業技術総合研究所；物質計測標準研究部門主任研究員(〒305-8565 つくば市東 1-1-1)
Things about Measurement Uncertainty —4. How to Applicate It?—; Katsuhiro Shirono(National Institute of Advanced Science and Technology (AIST), Tsukuba)
Keywords: *Uncertainty, Guide to the expression of uncertainty in measurement* (*GUM*)
2014年12月10日受理[doi:10.2320/materia.54.462]

べたように，こういうときには，「2乗して足す」ということを思い出して下さい．すなわち，山田様の到着時刻の標準偏差5分と，時計の時刻のずれの標準偏差の2.5分を2乗して足して，正の平方根を取ります．これは，およそ5.6分ということなります．

「$Z-X$」の分布を正規分布とすると，その中心は Y，標準偏差は5.6となります．**図4.1**(a)を見ると分かるように，Y が9.2のとき，「$Z-X$」がマイナスになる確率は5％ということになります．

詰まるところ，私は13時の9.2分＝9分12秒前，つまり12時50分48秒より前に受付に行けば，山田様をお待たせしない確率が95％以上となるわけです．不確かさのルールをうまく使うことで，まずまず納得感のある解にたどり着くことが出来たと言ってよいのではないでしょうか．

この例でもそうであるように，不確かさを使って何かを知りたいとなると，ある確率に対応するずれを知りたいということになるのではないでしょうか．今回は95％という確率に対応するずれである9分12秒という値を知りたいという問題だったと考えることもできます．

このようなことを見込んで，確率と結び付けられたずれのことを**拡張不確かさ**と呼びます．大抵の場合，その確率は90％とか，95％とか，99％という100％に近い大きな確率となります．（もっと言うと，ほとんどの場合は95％ですが…それについては次節にて．）ですから，拡張不確かさは標準偏差よりも大きな値になります．

今回の場合，図4.1(a)に見るように，「ゼロ〜無限大」という区間と95％という確率が対応しています．これはこれで，一応成立はしています．しかし，現実には平均を中心に左右対称の区間を考えたい場合が多いと言えます．今回の例でも，早く着きすぎて，昼ご飯の時間が削られるという損失も考えられます．図4.1(b)に「平均±9分12秒」という左右対称の区間を描きました．これは，左右のリスクを5％ずつ

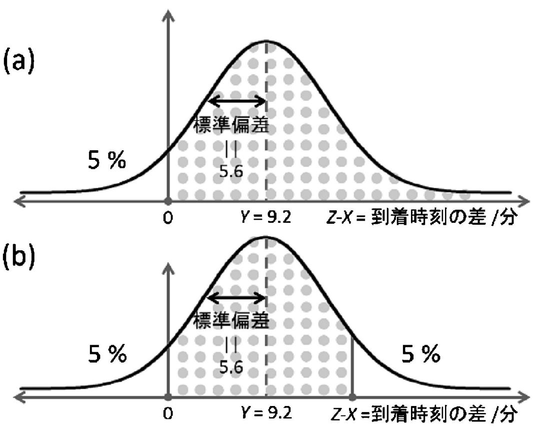

図4.1　時計の指す時刻が13時の9.2(＝Y)分前の時，受付に着いたときの，私と山田様の到着時刻の差($Z-X$)の分布(正規分布を仮定)．(a)到着時刻の差がマイナスになる確率は5％．(b)9.2分±9.2分(0分から18.4分)の間に90％の値が含まれている．

考慮し，90％の確率に対応した区間となります．拡張不確かさが左右対称の場合，「拡張不確かさは9分12秒である」という言い回しをよくします．

ちなみに，標準偏差に対応する不確かさは**標準不確かさ**と呼びます．これまでは，標準偏差と不確かさというのは，同じようなものかのようにお話してきましたので，混乱させてしまうかも知れません．不確かさには計算に便利な標準不確かさと，活用を考えたときに便利な拡張不確かさがあります．また，標準不確かさと拡張不確かさの比(＝拡張不確かさ/標準不確かさ)を**包含係数**と言います．

包含係数は，確率に応じて変わります．一方で，実のところ，95％という確率に対応して，包含係数は2とすることが極めて多いと言えます．この理由は次節で説明します．

4·2　技能試験と不確かさ

不確かさを何に使うかの一例は4·1節でお示ししましたが，ちょっと現実味のない応用でした．きちんとしたものをご紹介しましょう．もっとも重要な応用は，**技能試験**でしょう．技能試験とは，ある測定機関の測定した値と，別の測定機関が測定した値が同じ値になっているかどうかを確認することです．この測定は，同一の測定対象物に対して行われるものです．(**図4.2**)

測定の精度管理には，内部精度管理と外部精度管理があると言われます．内部精度管理とは，測定機関の内部のみで完結する精度管理のことです．例えば，測定マニュアルを整備したり，測定者の教育・研修を行ったりすることです．値がよく分かっている測定対象(標準物質)を測定することで，自分の測定方法の妥当性を確認することも含みます．

一方の外部精度管理は，技能試験の結果に基づいて，測定プロセスを改善する管理の仕方です．内部精度管理において標準物質を使ったのとは異なり，技能試験では値がよく分からないものと比較することが前提にあります．このため，より実物に近い測定対象について，測定パフォーマンスの評価が可能になります．

話そのものは，単純かと思います．同じものを，違う人が測って同じになれば，うまく行っているような気がしますし，そうでなければうまく行っていない気がします．ただ，その評価をするときに，「どの程度，値がそろっていたらうまく行っていると言えるのか？」という問題があります．この場面で，不確かさを使うというわけです．

典型的な技能試験の形態の一つは，2試験所間の比較です．このとき，ある測定機関の実力の方が，もう一つの測定機関の実力よりも高いことが想定して行います．この技能試験は，実力が低いとされる機関のために実施されるということになります．実力が高いとされる測定機関の測定値を X_{ref}，評価対象の測定機関の測定値を X_{lab} とします．X_{ref} と X_{lab} の95％の確率に対応する拡張不確かさを U_{ref} と U_{lab} とします．つまり，「$X_{\mathrm{ref}}\pm U_{\mathrm{ref}}$」および「$X_{\mathrm{lab}}\pm U_{\mathrm{lab}}$」という区間が，それぞれに95％という確率に対応するということ

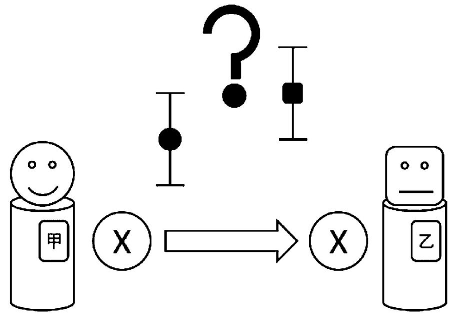

図4.2　技能試験では，同じものを別の機関で測定したときの結果から，測定のパフォーマンスを確かめる．瓶に小分けした試料など，厳密に同じものではなく，同じ値が出ると思われる試料を配布して，一斉に測定することもある．

です．

これらの値を用いて，以下の $\boldsymbol{E}_{\mathrm{n}}$ **数**と呼ばれるスコアを計算します：$E_{\mathrm{n}}=(X_{\mathrm{lab}}-X_{\mathrm{ref}})/\sqrt{U_{\mathrm{lab}}^2+U_{\mathrm{ref}}^2}$．測定機関の認定の分野では，$E_{\mathrm{n}}$ 数の絶対値が 1 以下であれば，この技能試験における測定機関のパフォーマンスは「満足」であるというのが，一定の基準となっています[1]．1 を超えた場合には，「不満足」と評価されます．記号がたくさん出てきてしまって申し訳ありませんが，よく見てみれば，計算はそれほど複雑なものではないと思います．

ここで言う「満足」というのは，もちろん，不確かさが十分に小さいという意味ではありません．決定した不確かさの評価と，実際の測定の結果のかい離が妥当な範囲で，測定プロセスがよくコントロールされているということです．実際に，十分に小さい不確かさかどうかは，定量化された不確かさの大小から直接に判断することになります．

技能試験は，様々な団体で実施されています†．ただし，不確かさを使わない技能試験の評価方法もあり，そちらの方がよく行われているというのが現状です．不確かさの情報をうまく使った技能試験データの解析は，現在進行形の研究課題とも言えます．

4･1節に述べた95％という確率に対応して，包含係数を 2 とすることが極めて多いのは，測定の不確かさを活用する先が，事実上，技能試験であるためと思います．技能試験では 95％の拡張不確かさ以外は使われないと言ってよいでしょう．

4･3　許容差と不確かさ

不確かさの活用ということを考えると，必ず出てくる話題が，許容差と不確かさの関係です．

例えば，あるせんべい屋で，一袋 100 g のせんべいを売ることを考えます．このとき，実際の一袋の内容量は 95 g～105 g に入るように管理したいとします．許容差というのは，狙いの 100 g からどのくらいずれていてもよいかという値のことです．この場合の許容差は 5 g です††．

さて，普通の人は，はかりにせんべいを乗せたときの表示値が 95 g～105 g に入っているものを 100 g と称して売れば，狙い通りの品質管理ができていると考えることでしょう．しかし，ここまでこの入門講座を読み進められてきた皆様には，違う考えをお持ちであって欲しいところです．

皆様に目を付けていただきたいのは，はかりが示した値がずれているかも知れないということです．はかりが示した値が本当の値からどのくらいずれているか？これが，測定の不確かさです．**図4.3**に示す通り，不確かさを考えると，はかりの表示値のみに着目しても，狙い通りの品質管理ができないのです．

この問題については，色々なアプローチが可能です．しかし，もっともシンプルなのは，許容差に対して，不確かさが無視できるほど小さいという状況を考えることです．不確かさがほぼゼロとみなせるならば，はかりの表示が 95 g～105 g のときに，そのせんべいを一袋として売れば，ほぼ狙い通りの品質管理ができていると言えましょう．

すると問題は「不確かさが許容差よりどのくらい小さければ，無視できるほど小さいと言えるか？」ということになります．ここに私なりの結論を述べましょう．それは，ケタ違いに小さい，つまり，10倍小さければ，無視できるというものです．今回の場合，許容差が 5 g ですから，測定の標準不確かさが 0.5 g より小さければ，無視してよいということです．

この問題について述べた文書で有名なものには，ANSI（米国国家規格協会）/NCSL Z 540.3[2]があります．その中で

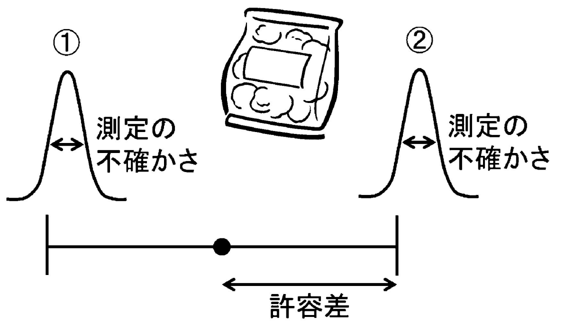

図4.3　許容差と不確かさの関係．平均が許容差の範囲内でも，測定の不確かさのために，測定値は許容差外にはみ出してしまうことがある（①）．逆に，平均が許容差の範囲外でも，測定値が許容差内に入ってしまうこともある（②）．このような誤りを無視できる不確かさとは？

† 読者の皆様が関連しそうなもので，民間の認定団体が主催する技能試験の Web サイトを紹介しておきます：公益財団法人日本適合性認定協会　技能試験の提供，http://www.jab.or.jp/service/proficiency_testing/（2014年12月 1 日最終確認）．金属の化学分析，硬さ試験，引張試験の技能試験を提供しています．ただし，これらの技能試験では，不確かさを使わない評価方法が行われています．不確かさの妥当性確認は各自行うということになるでしょう．

†† せんべいを例にしておきながら何ですが，これは日本の商品量目制度とは何の関係もない話だということは，悪しからずご了承下さい．

は，「許容差」と「測定のずれ」が4対1くらいの関係であればよかろうということが書いてあります．この4対1を導くのに，厳密な数字に基づいた検討があったとは思えません．むしろ，このくらいが現場の相場感として，納得のいくものなのだろうと理解しています．

ANSI/NCSL Z 540.3を理解するためのハンドブック[3]によれば，「測定のずれ」は95％の信頼の水準に対応する区間と言いかえられます．正規分布を想定するならば，標準偏差の2倍ということですね．つまり，許容差と測定の標準不確かさの2倍の比が4対1ならば，不確かさは無視して良いと解釈できそうです．ということは，許容差が測定の標準不確かさの8倍が目安になるということです．私としては，8も10も変わらんだろうし，きりがよいという理由で，10倍と考えています．かなり大ざっぱな話なので，より厳密に扱いたい方はANSI/NCSL Z 540.3やハンドブックにあたることをお勧めします．

ただし，ANSIはアメリカ合衆国の規格であり，国際規格でも日本の規格でもありません．この許容差と不確かさの問題に関しては，今後，不確かさの普及に伴って，研究やルール作りが進んでいくようにも思います．実際の測定をされている皆様，測定されたモノを使っている皆様に，それぞれの場で適切な不確かさの使い方というものを，是非ご提案していただきたいと思っています．

4･4　おわりに

ようやく，この入門講座のゴールです．ここまでお付き合い下さりありがとうございました．なるべく平易に解説するよう努めたつもりではありますが，分からない部分もあると思います．特に，専門外であるために，金属の測定の具体的なお話は全くできませんでした．ご質問などありましたら，直接筆者(k.shirono@aist.go.jp)にご連絡下さい．

最後に，全体を通した事例のご紹介をしていないので，下に演習問題を付します．ちょっと難しいですけれども，前号までの内容もご覧になりながら，ご確認下さい．

演習問題

ばねの伸びから物体Xの質量を測定し，その不確かさを求める．下の(A)～(E)に入る数値を答えよ．ただし，最終的に2桁の不確かさを報告するため，途中の計算では3桁を残している．

使用するばねについて，伸び1 cm当たりの質量は，97 g～103 gの範囲であると分かっている．このばねを用いて，物体Xを3回吊るし，伸び3.80 cm, 4.00 cm, 4.20 cmを得た．

ばねの伸びを質量に変換する係数(伸び1 cmあたりの質量)は97 g/cm～103 g/cmである．この係数について，他の情報がないため，100 g/cm±3 g/cmの一様分布を考える．係数の値100 g/cmに対して，標準不確かさは3/(A)g/cm＝1.73 g/cmと定まる．

ばねの伸びの平均値は4.00 cmである．3回の測定結果の平均値からのずれは，－0.20 cm, 0.00 cm, ＋0.20 cmである．これらのずれの2乗和を(測定回数－1)の2で割って，正の平方根を取った0.200 cmは一つの測定値の標準偏差である．平均値の標準偏差はこの0.200 cmをさらに(B)で割ったもので，0.115 cmとなる．

伸びと質量の変換係数100 g/cmとばねの伸びの平均値4.00 cmを掛け，測定対象量である物体Aの質量の測定値400 gが求まる．変換係数に対する感度係数は4.00 cmである．これを標準不確かさ1.73 g/cmに掛けて得られる6.92 gは，物体Xの質量の不確かさの1つの要因となる．また，ばねの伸びの平均値に対する感度係数は(C)g/cmである．これを標準不確かさ0.115 cmに掛けて得られる11.5 gは，同じく，物体Xの質量の不確かさの1つの要因となる．

ここに，2つの不確かさ要因があり，その標準不確かさは6.92 gと11.5 gである．測定対象量である質量の標準不確かさは，これらを(D)乗して，足して，正の平方根を取ったものである．これを計算すると13.4 gとなる．

この測定対象量の値の分布は厳密には複雑なものである．しかし，それを正規分布に近似するならば，標準不確かさ13.4 gを(E)倍して，およそ95％の信頼の水準に対応する拡張不確かさを26.8 gと求めることができる．最終的な不確かさの報告は多くても2桁で十分であるから，3桁目を四捨五入し，27 gとする．

以上のことから，物体Xの質量の測定値は400 g，その95％の信頼の水準に対応する拡張不確かさは27 gと報告する．

(完)

文　　献

(1) JIS Q 17043：2011，適合性評価―技能試験に対する一般要求事項．
(2) ANSI/NCSL Z 540.3：2006, Requirements for the calibration of measuring and test equipment.
(3) NCSL International, Handbook for the application of ANSI/NCSL Z 540.3–2006―Requirements for the calibration of measuring and test equipment, (2009).


城野克広

★★★★★★★★★★★★★★★★★★★★★★★
2007年　東京大学大学院新領域創成科学研究科博士課程修了
2009年　産業技術総合研究所研究員
2013年3月・現職
専門分野：不確かさの評価の技術開発・比較試験の統計処理
◎活動内容などは下記Webサイトをご覧ください．
(https://staff.aist.go.jp/k.shirono/)
★★★★★★★★★★★★★★★★★★★★★★★

解答　A：$\sqrt{3}$，B：$\sqrt{3}$，C：100，D：2，E：2