

不確かさガイド(GUM)に関わる 国際動向とベイズ統計の利用

産業技術総合研究所 計量標準総合センター
榎原研正

概要

1. GUMをめぐる国際的動き
 - GUM改訂ドラフト
 - 改訂の動機は？
2. ベイズ統計を利用した不確かさ評価
3. まとめ

*) GUM: Guide to the Expression of Uncertainty in Measurement (測定における不確かさの表現のガイド), ISO (1993).

GUMのメンテナンス

JCGM (Joint Committee for Guides in Metrology)

= 国際度量衡局 (BIPM)に事務局をおき、次の8機関をメンバー機関とする国際委員会

- BIPM: Bureau International des Poids et Mesures
- IEC: International Electrotechnical Commission
- IFCC: International Federation of Clinical Chemistry and Laboratory Medicine
- ILAC: International Laboratory Accreditation Cooperation
- ISO: International Organization for Standardization
- IUPAC: International Union for Pure and Applied Chemistry
- IUPAP: International Union for Pure and Applied Physics
- OIML: International Organization of Legal Metrology

□ JCGM-WG1: GUM担当

□ JCGM-WG2: VIM担当

日本からは今井秀孝氏(産業技術総合研究所客員研究員, 製品評価技術基盤機構客員調査員, 元産総研理事)が参加(ILAC代表)

GUM改訂をめぐる動き

- GUM … 内容は20年以上変わっていない
 - 1993 GUM 第1版
 - 1995 微修正
 - 2008 JCGM 100 (追加微修正, WEB上でオープン化)

- 統計学の立場から, GUMへの批判がある
 - … 「確率」の扱いに混乱がある (タイプA/タイプB評価で確率概念が異なる) [例えば L. J. Gleiser (1998)]

- 混乱を解消する幾つかの提案がある
 - … 多くは, ベイズ統計の利用による [例えば R. Kacker & A. Jones (2003)]

GUM改訂をめぐる動き

- GUM周辺文書へのベイズ統計の(部分的)導入
 - 2008: GUM補完文書1 (モンテカルロ法を用いた分布の伝播)
 - 2011: GUM補完文書2 (出力量が複数ある場合への拡張)
 - 2012: JCGM 106(適合性評価における測定不確かさの役割)

- GUM改訂ドラフトの策定・関係機関への配布・コメント収集
 - GUM改訂ドラフト (JCGM100 201x CD): 2014/12月
(www.affidabilita.eu/pdfeventi/JCGM_100_201X_CD.pdf)
 - コメント収集: 2014/12月- 2015/4月
 - 対象: JCGMメンバー機関(MOs) + 国家計量標準機関(NMIs)

GUM改訂ドラフト (JCGM100/101 201x CD)に対するコメントの集計 (2015/6) [C. Michotte (BIPM)による]

		Member Organizations	Institutes	
		6 MOs	25 NMIs	合計
本体 事例集	JCGM 100 (GUM rev.)	530	375	905
	JCGM 110 (exples)	83	85	168
				1073

... 改訂に否定的な意見が優勢であった

ISO/IEC ballots: « Do you agree to the circulation of the drafts as a DIS? »

	Yes	Yes+comments	No	Abstain	Result
ISO	9	1	2	3	✓
IEC / JCGM 100	6	4	3	1	✓
IEC / JCGM 110	6	4	2	2	✓

日本からの意見 (の一部)

- 現行GUMは当面このまま維持し、改訂ドラフトは別文書（例えば補完文書）として出版するのが妥当。なぜならば、
 - 現行GUMは、トレーサビリティ制度・試験所認定制度・工業規格など多くの社会的制度の中にすでに組み込まれているため、GUM改訂によりこのような制度の再構築に膨大なコストと労力が必要になる。
 - この点に配慮をせずに改訂を進めれば、これまでに構築されてきた「不確かさ」に対する社会的信頼を失い、不確かさのより一層の普及が阻害される可能性がある。
 - 別文書として出版することにより、GUMの改訂が現在の社会的制度に及ぼし得る影響を十分の時間をかけて調査することができる。

GUM改訂案に対する代表的な意見

否定的意見

- 改訂のメリットが明瞭でない
- 改訂に伴う社会的コストが大きすぎる
- 内容がむづかしい

肯定的意見

- タイプAとタイプBの間の概念上の整合性が向上
- 知識の状態を適切に表現できる
- 改訂案のほうが理解しやすい
- 有効自由度の計算が不要になる

GUM改訂への動きの現状

- ❑ GUM改訂ドラフトに対する関係機関からの意見の多くは否定的であり、短い期間中での改訂の可能性はなくなったと考えられる
- ❑ ただし、現行GUM中の確率の扱いの不整合に関わる問題点は残っているため、これらの解消に向けたJCGMによる活動は継続されると考えられている
- ❑ JCGMと独立に、ベイズ統計やモンテカルロ法を利用した不確かさ評価を推進する動きもある

例: NIST Technical Note 1900, 2015

- ❑ GUMを改訂すべきかどうか、どのような方向で改訂すべきか等について、統計専門家だけでなく、計測関係者の立場での検討や意見形成が重要

モンテカルロ法を利用した分布の伝播のオンライン計算ツールの例

<http://uncertainty.nist.gov/>

The screenshot shows the NIST Uncertainty Machine web interface in a browser window. The page title is "NIST Uncertainty Machine". Below the title, there is a link to the user's manual and a list of instructions. The instructions include selecting the number of input quantities, changing quantity names, choosing distributions and parameters, selecting the number of realizations, writing the output quantity definition, and running the computation. The form fields are as follows:

- Random number generator seed: 61
- Number of input quantities: 3
- Names of input quantities: x0, x1, x2
- Update quantity names button
- Input quantity details table:

Quantity	Distribution	Parameter 1	Parameter 2	Parameter 3
x0	Gaussian (Mean, StdDev)	0	1	
x1	Student t (Center, Scale, No. of degrees of freedom)	1	1	2
x2	Rectangular (Left Endpoint, Right Endpoint)	-1	1	

- Number of realizations of the output quantity: 1000000
- Definition of output quantity (R expression): x0
- Symmetrical coverage intervals checkbox (unchecked)
- Correlations checkbox (unchecked)
- Run the computation button

JCGMによるGUM改訂の動機

(W. Bich他, Metrologia, 49(2012)702-705 等に基づく)

1. 現行GUMには、内部的不整合がある（タイプA評価とタイプB評価で確率の意味が異なる）
2. 外部的不整合がある（頻度主義統計に基づく現行GUMと、ベイズ統計を利用する補完文書1 & 2、及びVIM3が不整合）
3. ベイズ統計の採用により、これらの不整合を解消することができる

*) VIM3: International Vocabulary of Metrology – Basic and General Concepts and Associated Terms, 3rd edition. (国際計量基本用語集, 第3版)

タイプA評価とタイプB評価

タイプA評価 = 一連の観測値の統計的解析による評価方法

例) 家庭用血圧計による血圧の繰り返し測定値 ($x_1, x_2, \dots, x_n$) に基づき, 測定結果 $\bar{x}$ (平均値) の標準不確かさを次で評価する

$$u_A(\bar{x}) = \frac{s}{\sqrt{n}}$$

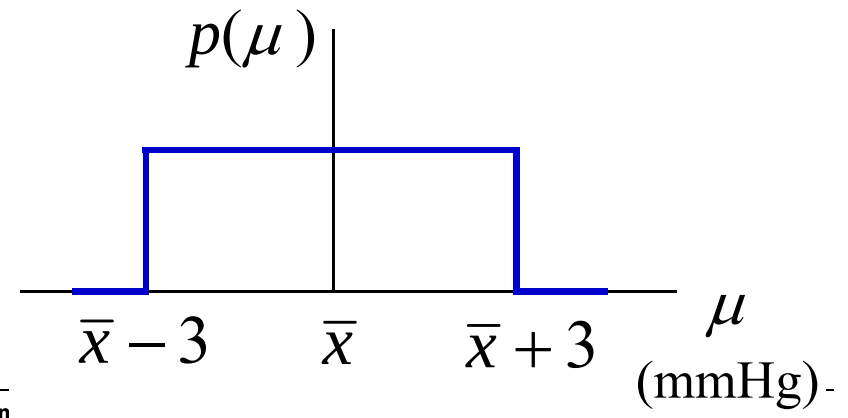
$$\text{ただし } s \text{ (実験標準偏差)} = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n-1}}$$

タイプB評価 = 統計的解析以外の方法による評価

例) 家庭用血圧計の説明書の記載「精度： ± 3 mmHg」にもとづいて、血圧計の目盛りのずれに起因する不確かさを次のように評価する。

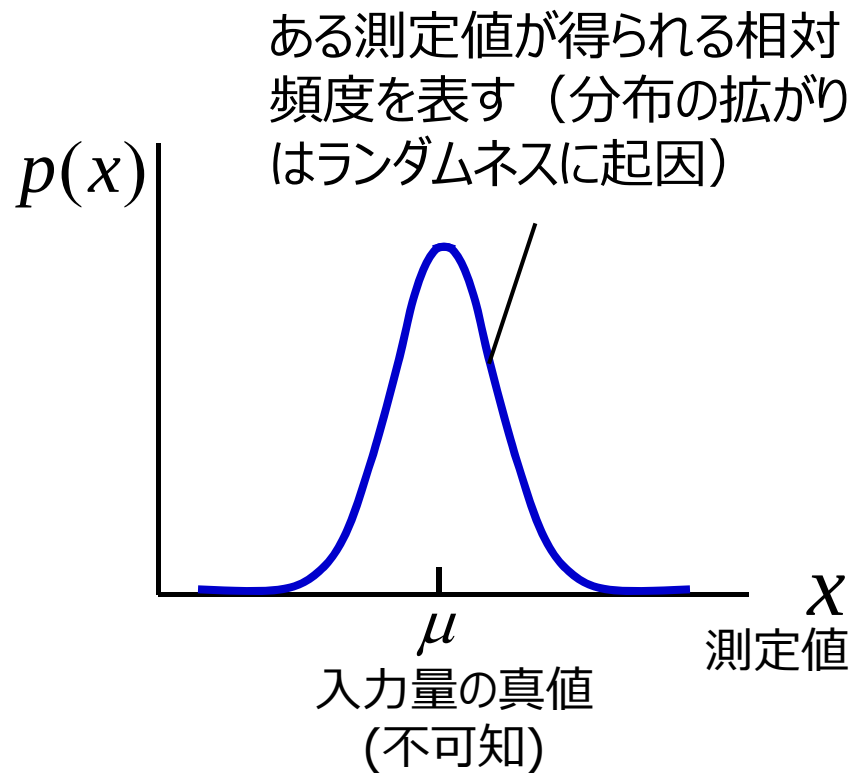
説明書の記載は、目盛りのずれの許容最大値を表すものと解釈し、血圧測定値 $\bar{x}$ に基づく真値 μ の確率分布を図のように想定する。
これから、血圧測定値の標準不確かさを次で評価する

$$u_B(\bar{x}) = \text{確率分布の標準偏差} = \frac{3}{\sqrt{3}} = 1.7 \text{ (mmHg)}$$

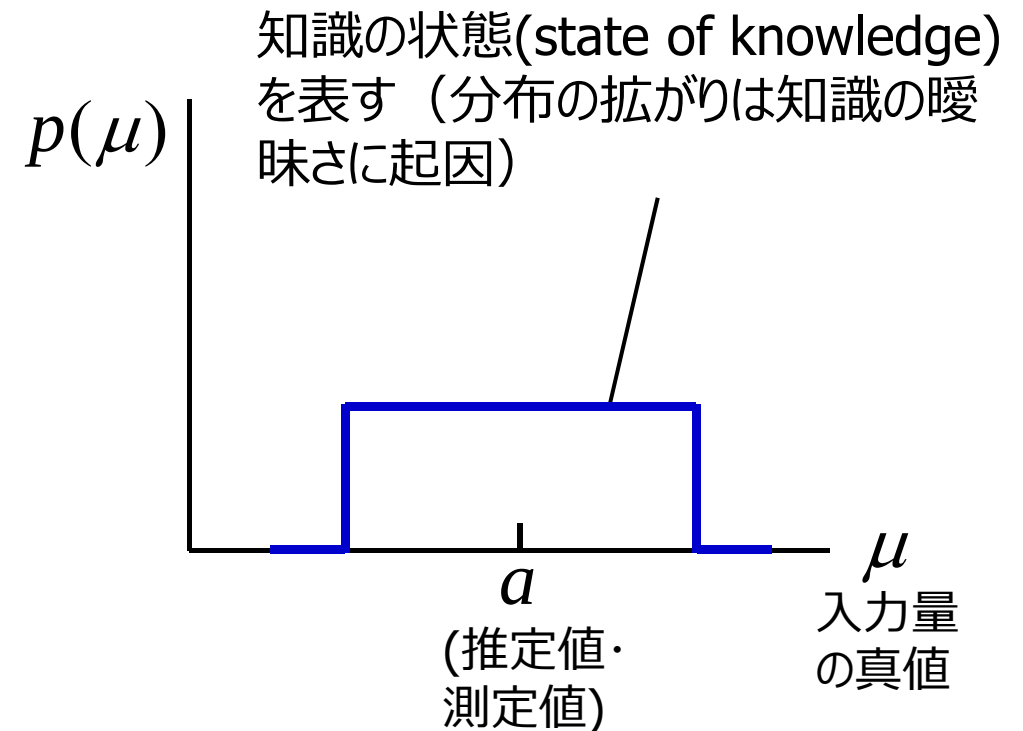


タイプA／タイプB 評価における確率概念の違い

タイプA評価



タイプB評価



タイプA／タイプB 評価における確率概念の違い

	タイプ A 評価	タイプ B 評価
確率	相対発生頻度	確信度 (degree of belief)
確率変数	測定値	量の真値
確率分布の 拡がりの原因	偶然的ばらつき (ランダムネス)	知識の曖昧さ
確率概念の 基盤	頻度主義統計学 (伝統的統計学)	ベイズ統計学

現行GUMの統計的基盤

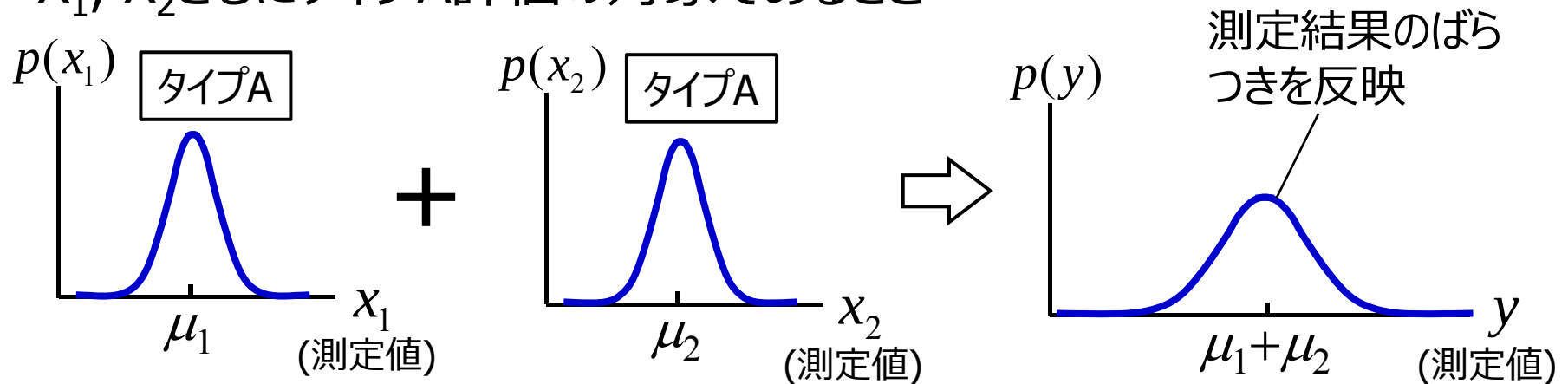
- タイプA評価：頻度主義
- タイプB評価：
 - 確率概念はベイズ的
 - 標準不確かさの評価にベイズ統計の推論方法（ベイズの定理）を用いているわけではない
 - 標準不確かさに自由度を（強引に）付加して、頻度主義の中で再解釈しようとしている
- 拡張不確かさの計算：頻度主義
 - 有効自由度の計算（Welch-Satterthwaiteの近似式の利用）
 - t分布を用いた包含区間の計算 = 頻度主義における信頼区間の計算

⇒ 現行GUMの全体的枠組みは、頻度主義統計学

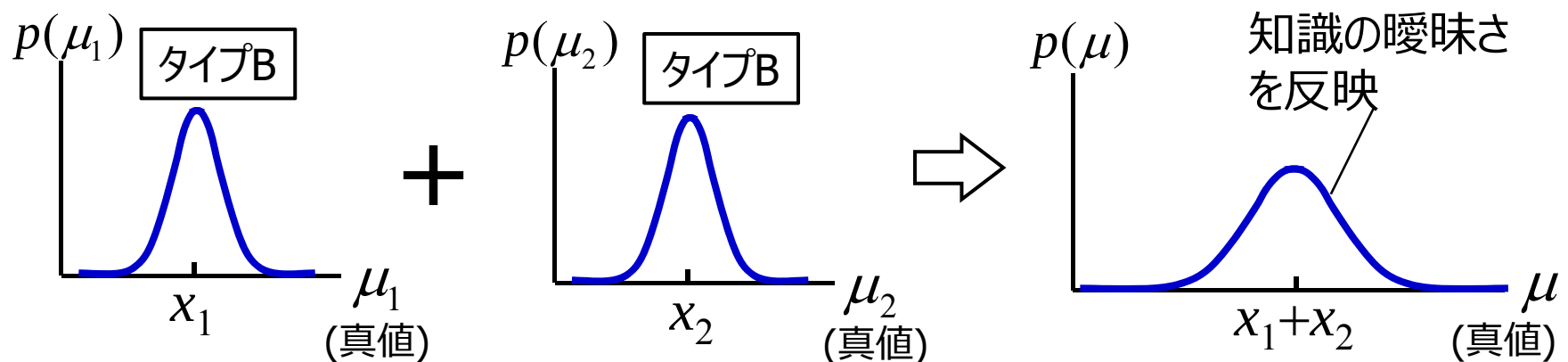
タイプA評価／タイプB評価の不整合について

例) 測定モデルが $Y = X_1 + X_2$ のとき, Y の確率分布を推定する

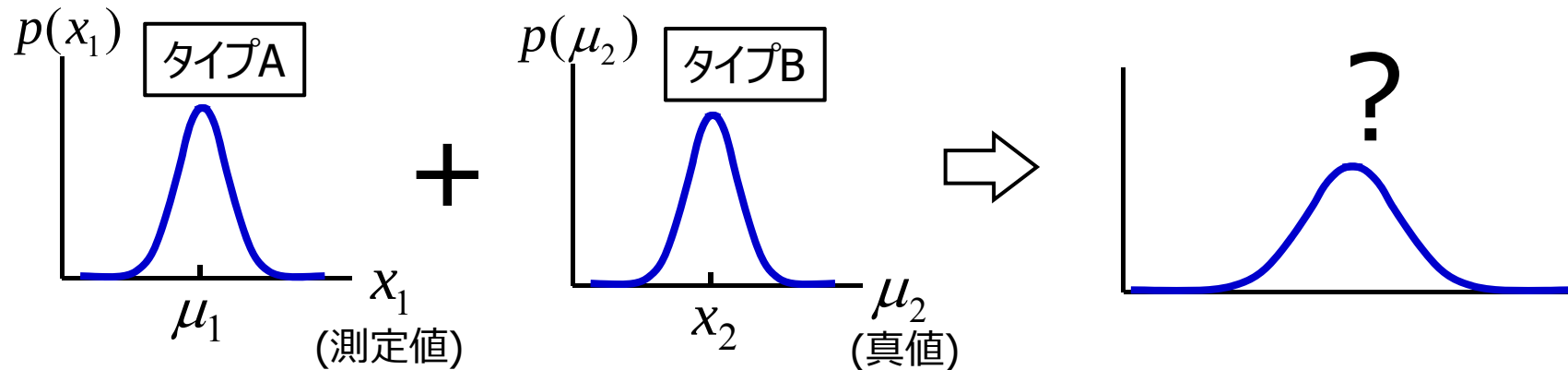
- X_1, X_2 ともにタイプA評価の対象であるとき



- X_1, X_2 ともにタイプB評価の対象であるとき



- X_1 :タイプA評価, X_2 :タイプB評価の対象であるとき



- (モンテカルロ法などにより) 確率分布を合成しようとするとき、確率変数の意味の不整合が顕在化する
- 不整合を放置したときに生じる問題
 - ・ 包含確率や包含区間の厳密な意味づけがしづらい
 - ・ ただし、単に標準不確かさを合成する範囲（分散の足し算）では、問題は顕在化しない（不整合がないわけではない）

不整合問題への可能な対応策

- (1) 不確かさ評価の全体を，ベイズ統計の立場から再構築する
 - ・・・改訂GUMドラフトやGUM補完文書 1, 2等の立場
- (2) タイプB評価の手順を，頻度主義の立場で再解釈する
 - ・・・今すぐには困難？
- (3) 放置する
 - ・・・確率解釈の不整合性(内部的不整合)が許容範囲かどうか
 - ・・・GUMとその周辺文書の不整合性(外部的不整合)が許容範囲かどうか

ベイズ統計適用の幾つかの考え方

(1) GUM改訂ドラフト

- タイプA評価にベイズ統計を適用
- 現行の不確かさの伝播則をそのまま使う

(1') GUM補完文書 1 (モンテカルロ法による分布の伝播)

- タイプA評価にベイズ統計を適用
- 不確かさの伝播則に替えて, 確率分布の伝播

(2) ベイズ統計の全面的適用(Full Bayes)

- タイプB評価で仮定する確率分布を事前分布とし, それをタイプA評価で得たデータで「情報更新」して, 事後分布を計算

ISO/TR 13587:2012 Three statistical approaches for the assessment and interpretation of measurement uncertainty

タイプA評価のベイズ統計の立場での書き直し

- n 回の繰り返し測定データ： $x_1, x_2, \dots, x_n$
(平均 $\bar{x}$, 実験標準偏差 s)
- 各 x_i は正規分布 $N(\mu, \sigma^2)$ に従うと仮定 (μ : 測定量の真値)
- μ, σ^2 について, 事前の知識がないと仮定

↓ (ベイズ定理)

- μ の確率分布 = 拡がりを $s/\sqrt{n}$ 倍し, 中心を $\bar{x}$ に
ずらした自由度 $(n-1)$ の t 分布

- この分布の標準偏差 = $\sqrt{\frac{n-1}{n-3}} \frac{s}{\sqrt{n}}$ → これを標準不確かさとする

タイプA評価: 現行GUM vs. 改訂ドラフト

現行(頻度主義)

標準不
確かさ

$$\frac{s}{\sqrt{n}}$$

その意味

s^2/n は $\bar{x}$ の母分散 σ^2/n の推定値

自由度

$$n-1$$

その意味

s 自体の曖昧さ

改訂ドラフト(ベイズ)

$$\sqrt{\frac{n-1}{n-3}} \frac{s}{\sqrt{n}}$$

- 1 より大
- $n \rightarrow \infty$ で 1

真値 μ に関する知識を表す確率分布の標準偏差

ベイズ統計では自由度の概念はない（知識の曖昧さは全て確率分布の中に織り込む）

GUM改訂ドラフトの概要

- タイプA評価(繰返し数 n)における標準不確かさ

$$\frac{s}{\sqrt{n}} \text{ (現行GUM)} \quad \Rightarrow \quad \sqrt{\frac{n-1}{n-3}} \frac{s}{\sqrt{n}} \text{ (改訂案)}$$

- (有効) 自由度の概念は消滅。Welch-Satterthwaiteの式は無用に

- 不確かさ伝播則は、継続して使用

- 包含係数 $k = 2$ が使いやすくなる

(繰返し数 n が小さくとも, $U = k u(y)$ において, $u(y)$ 自体が大きくなるため, 単純に $k = 2$ としても $\pm U$ の信頼の水準が95 %より小さくなりにくい)

不確かさ評価の例：現行GUM vs. 改訂ドラフトの比較

家庭用血圧計で最高血圧を5回繰り返し測定し、 $(x_1, x_2, \dots, x_5) = (128, 132, 123, 121, 125)$ [mmHg] を得た。

血圧計の説明書に、「精度： ± 3 mmHg」（目盛りのずれの最大を表すと解釈）の記載があった。

測定結果 $y = \bar{x} = 125.8$ mmHg の不確かさは？

＊）現実の最高血圧は時々刻々変化している可能性があるが、健康の指標として利用することが可能な、少なくとも短時間の間は一定の「最高血圧の真値」というものが存在するここでは考える。血圧測定の目的は、これを知ることである。

	現行GUM	改訂ドラフト
実験標準偏差 s	4.32 mmHg	
タイプA成分 $u_A(\bar{x})$	$\frac{s}{\sqrt{5}} = 1.93 \text{ mmHg}$	$\sqrt{\frac{5-1}{5-3}} \frac{s}{\sqrt{5}} = 2.73 \text{ mmHg}$
タイプB成分 $u_B(\bar{x})$	$\frac{3}{\sqrt{3}} = 1.73 \text{ mmHg}$ (一様分布を想定)	
合成標準不確かさ $u_c(y)$	2.59 mmHg	3.23 mmHg
有効自由度 ν_{eff}	13 (Welch-Satterthwaiteの式による)	—
95%包含係数 k	2.16 (自由度13の t分布による)	約 2 (合成後の分布を正規分布と仮定)
拡張不確かさ U	$2.16 \times 2.59 \cong 5.6 \text{ mmHg}$	$2 \times 3.23 \cong 6.5 \text{ mmHg}$

不確かさの違い：現行GUM vs. 改訂ドラフト

- 現行GUMと改訂ドラフトの評価結果は一般に異なる。違いの大小は、状況による
- 合成標準不確かさは、改訂ドラフトの方が必ず大きくなる
- 拡張不確かさでは、その差は縮小する（包含係数は改訂ドラフトの方が小さくなるため）
- 改訂ドラフトで包含区間を厳密に決めるには、モンテカルロ法による分布の伝播の計算が薦められている（この例ではモンテカルロ法で求まる包含係数は、 $k = 1.91$ ）

[有効自由度の計算が不要である点は、改訂ドラフトのメリット]

まとめ

- GUMが内包する，確率解釈に付随する問題点の解消のため，ベイズ統計を取り入れたGUM改訂の動きがある
- 最初の改訂ドラフト(2014/12)に対する各機関の意見は改訂に否定的であったため，短期間での改訂の可能性は低い
が，JCGMによる改訂に向けた長期的な動きは継続すると考えられる
- GUMのこのような問題点を放置することの是非や，GUM改訂をベイズ統計にもとづいて行うことの是非を，統計専門家だけでなく，計測関係者の視点で検討・意見形成することが重要

参考資料

□ GUMの改訂に関わるもの

1. Bich, W., How to revise the GUM? *Accred. Qual. Assur.*, 13 (2008) 271–275.
2. Bich, W. From Errors to Probability Density Functions. Evolution of the Concept of Measurement Uncertainty, *IEEE Tran. Instrument. Meas.* 61 (2012) 2153-2159.
3. W. Bich, *et al.*, “Revision of the Guide to the Expression of Uncertainty in Measurement,” *Metrologia* 49 (2012) 702–705.
4. H. Imai, "Expanding needs for metrological traceability and measurement uncertainty," *Measurement* 46 (2013) 2942–2945.
5. 今井秀孝, 「JCGMの最新動向：GUM及びVIMの現状と将来」, 計測標準と計量管理, 65, No.2(2015) 52-59.
6. BIPM Workshop on Measurement Uncertainty (June 15-16, 2015)資料
[www.bipm.org/en/conference-centre/bipm-workshops/measurement-uncertainty/]

□ 不確かさ評価へのベイズ統計の利用

1. Lira, I. and Wöger, W., Bayesian evaluation of the standard uncertainty and coverage probability in a simple measurement model, *Meas. Sci. Technol.* 12 (2001) 1172–1179.

2. Kacker R. and Jones, A., On use of Bayesian statistics to make the Guide to the Expression of Uncertainty in Measurement consistent, *Metrologia* 40 (2003) 235–248.
3. Kacker, R. Toman, B., and Huang, D., Comparison of ISO-GUM, draft GUM Supplement 1 and Bayesian statistics using simple linear calibration, *Metrologia* 43 (2006) S167–S177.
4. Rabinovich S., Towards a new edition of the "Guide to the expression of uncertainty in measurement," *Accred. Qual. Assur.* 12 (2007) 603–608.
5. Guthrie, et al., Three Statistical Paradigms for the Assessment and Interpretation of Measurement Uncertainty, pp.71-115, in "Data Modeling for Metrology and Testing in Measurement Science," Pavese, F. and Forbes, A.B. (eds.), (2009, Birkhauser, Boston).
6. Forbes, A.B. and Sousa, J.A., The GUM, Bayesian inference and the observation and measurement equations, *Measurement* 44 (2011) 1422–1435.

□ ベイズ統計の利用への異論

1. White, D. R., In pursuit of a fit-for-purpose uncertainty guide, *Metrologia* 53 (2016) S107–S124.
2. Attivissimo, F., Giaquinto, N., Savino, M., A Bayesian paradox and its impact on the GUM approach to uncertainty, *Measurement* 45 (2012) 2194–2202.