

# GUM改訂の動き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

# Outline

1. GUM改訂に関わる議論の背景
2. ベイズ統計とは
3. 改訂案の基本方針
4. モデル計算による4つの評価方法の比較
5. 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, 2nd ed. (1995) = JCGM 100 (2008)

# GUMの発展

- 1980 Recommendation INC-1
- 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:  
www.bipm.net/en/publications/guides/gum.html )
- 現在
  - GUM第1版から内容は実質上変わっていない
  - 基礎科学、計量標準の国際比較、試験・校正機関の認定、トレービリティの保証など、様々な局面で利用されるようになっている
- 2014? 改訂版ドラフト公開
- 201\*? 改訂GUM出版

# 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

# 国内でのGUM関連文書の出版

- 1996: 「計測における不確かさの表現のガイド」(日本規格協会) = GUMの翻訳 + VIM第2版
- 2012: 標準仕様書 TS Z 0033 「測定における不確かさの表現のガイド」
- 2012: 標準仕様書 TS Z 0032 「国際計量計測用語—基本及び一般概念並びに関連用語(VIM)」 = VIM第3版の翻訳

# JCGMで既作成／作成中の文書

## □ 国際計量計測用語

- 2007 JCGM 200 (= VIM第3版)

ベイズ統計  
に配慮



## □ GUM補完文書

- 2008 JCGM 101 (= GUM Supplement 1)

“モンテカルロ法を用いた分布の伝播”

ベイズ統計  
に沿う



- 2011 JCGM 102 (= GUM Supplement 2)

“出力量が複数ある場合への拡張”

ベイズ統計  
に沿う



- (審議中) JCGM 103 (= GUM Supplement 3) “モデリング”

## □ その他のJCGM文書

- 2009 JCGM 104 “GUMとその関連文書の紹介”

ベイズ統計  
に沿う



- 2012 JCGM 106 “適合性評価における測定不確かさの役割”

- (審議中) JCGM 105 “測定不確かさの評価のための概念, 原理, 方法”

- (審議中) JCGM 107 “最小二乗法の応用”

# GUM改訂に関する公開情報

## 論文

- 1) W. Bich, "How to revise the GUM?" *Accred. Qual. Assur.* 13 (2008) 271-275.
- 2) W. Bich, *et al.*, "Revision of the Guide to the Expression of Uncertainty in Measurement," *Metrologia* 49 (2012) 702–705.
- 3) W. Bich, "From Errors to Probability Density Functions. Evolution of the Concept of Measurement Uncertainty," *IEEE Trans. Instrum. Meas.* 61 (2012) 2153-2159.
- 4) H. Imai, "Expanding needs for metrological traceability and measurement uncertainty," *Measurement* 46 (2013) 2942–2945.

## 講演資料

- 1) 今井秀孝, 「計量計測関連国際文書類の改訂の動き –SI:単位系・VIM:用語・GUM:不確かさ–」, 不確かさクラブ第8回総会(2014年1月, 大阪)資料 (<https://www.nmij.jp/~measure-sys/metinfo/uncertainty/club8s.html>)

## BIPMアンケート調査

- 1) W. Bich, "Report on the GUM Online Survey" ([www.bipm.org/wg/JCGM/JCGM-WG1/Allowed/sub-committee\\_5/WG1-SC5-N12-14b\\_GUM\\_survey\\_report.pdf](http://www.bipm.org/wg/JCGM/JCGM-WG1/Allowed/sub-committee_5/WG1-SC5-N12-14b_GUM_survey_report.pdf))
- 2) JCGM Survey (GUM) Collated responses ([www.bipm.org/wg/JCGM/JCGM-WG1/Allowed/sub-committee\\_5/WG1-SC5-N12-15\\_JCGM\\_GUM\\_Survey\\_Collated\\_responses.pdf](http://www.bipm.org/wg/JCGM/JCGM-WG1/Allowed/sub-committee_5/WG1-SC5-N12-15_JCGM_GUM_Survey_Collated_responses.pdf))

# GUM改訂の背景（論文2等に基づく）

## 動機

現行GUMには

- 内部的不整合がある（Type AとType Bで確率の意味が異なる）
- 外部的不整合がある（GUMと、ベイズ統計を利用するSupplement 1 & 2、及びVIM3が不整合）

## 基本方針

- ベイズ統計の採用により、これらの不整合を解消する
- 内容の難易度を維持する



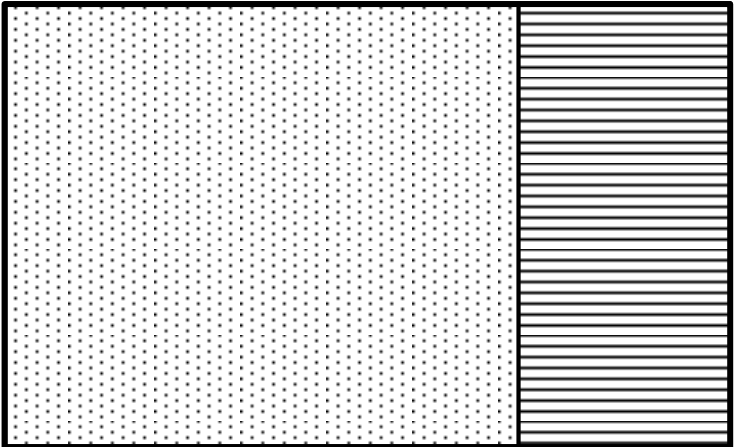
参考

# ベイズの定理

例) チームXの勝敗と天気

T1:晴れ  
(70%)

T2:雨  
(30%)



W:勝ち  
(31%)

L:負け  
(69%)



勝敗と天気の情報結びつけられていない状態では...

Q: 2013年9月3日が「雨」だった確率は？      ⇒ 30%

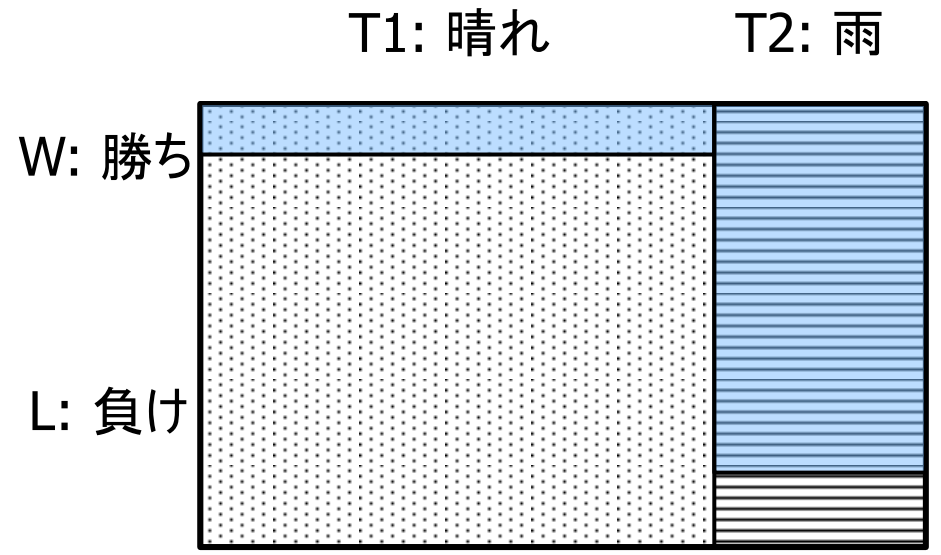
参考

# ベイズの定理(続き)

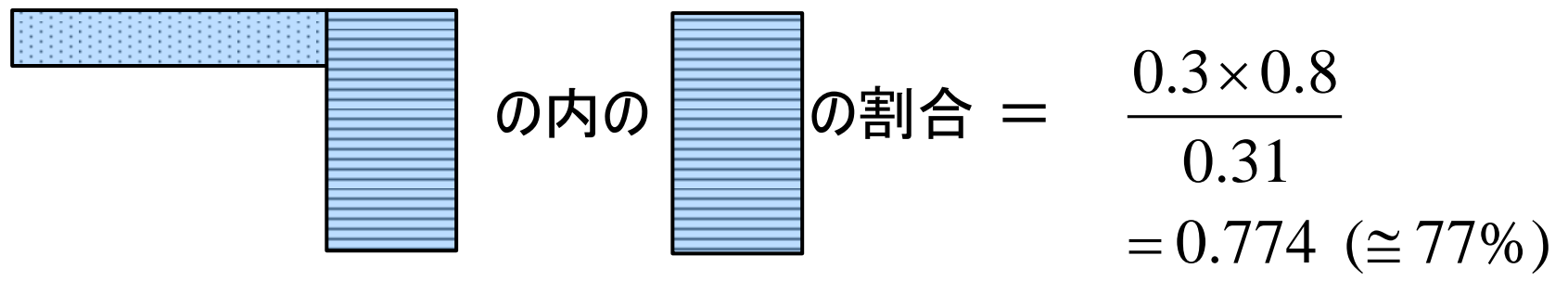
天気別の勝敗率がわかると...

	W: 勝ち	L: 負け
T1: 晴れ	10%	90%
T2: 雨	80%	20%

... チーム X は雨に強い



Q: 2013年9月3日は「勝ち」であった。その日が「雨」だった確率は？



# ベイズの定理(続き)

$$\Pr(T_2 | W) = \frac{\Pr(W | T_2) \Pr(T_2)}{\Pr(W)}$$

← ベイズの定理

事前確率

事後確率 (W が起こったとの条件下での  $T_2$  の確率)

- 因果関係を逆転した推論が可能

[天気 → 勝敗] ⇔ [勝敗 → 天気]

- 連続変数の場合

$$p(T | W) = \frac{p(W | T)p(T)}{\int p(W | T')p(T')dT'} \propto p(W | T)p(T)$$

- $p(\cdot)$  : 確率密度関数(PDF)
- $T$  依存性に関心があるので、右辺のように書くことが多い
- $p(W | T)$  を  $T$  の関数とみたとき、「尤度関数」と呼ばれる

# ベイズ統計

□ ベイズの定理を推論のエンジンとする

□ 確率の概念を広く捉える

(例) ・ 過去のある時点・ある場所の天気を確率変数とみなす

・ 基礎物理定数を確率変数とみなす

→ 主観確率 (“state of knowledge” を表す)

頻度主義的統計学(伝統的統計学)では...

- 同一条件下でチームXが何度も試合を繰り返せるならば、「勝利日はどんな天気か」は確率変数
- しかし、確定済み現象である2013年9月3日の天気を確率変数と考えることはできない

# タイプA評価へのベイズ統計の利用

ある物質の濃度  $C$  を繰り返し測定し、次のデータを得た。

$$\mathbf{x} = (x_1, x_2, \dots, x_n)$$

ただし  $x_i$  は互いに独立に正規分布  $N(C, \sigma^2)$  ( $C, \sigma$  は未知) に従うとする。 $C$  に対する標準不確かさ、及び包含区間は？

## 全体的方針

ベイズの定理を用いて事後分布  $p(C, \sigma | \mathbf{x})$  を計算した上で

$C$  に対する事後分布を次で求める

$$p(C | \mathbf{x}) = \int p(C, \sigma | \mathbf{x}) d\sigma \quad (\text{関心の無い } \sigma \text{ について"周辺化"})$$

# 参考 タイプA評価へのベイズ統計の利用(続き)

□ 事前分布:  $p(C, \sigma) \propto 1/\sigma$  と仮定(無情報事前分布)

□ 測定データのモデル分布:

$$p(\mathbf{x} | C, \sigma) = \prod_i p(x_i | C, \sigma) = \left( \frac{1}{\sqrt{2\pi}\sigma} \right)^n \exp\left( -\frac{1}{2\sigma^2} \sum_i (x_i - C)^2 \right)$$

$$\propto \frac{1}{\sigma^n} \exp\left( -\frac{(n-1)s^2 + n(\bar{x} - C)^2}{2\sigma^2} \right) \quad (s: \text{実験標準偏差})$$

□ 事後分布:  $p(C, \sigma | \mathbf{x}) \propto \frac{1}{\sigma^{n+1}} \exp\left( -\frac{(n-1)s^2 + n(\bar{x} - C)^2}{2\sigma^2} \right)$

□  $\sigma$  について周辺化:

$$p(C | \mathbf{x}) \propto \int \frac{1}{\sigma^{n+1}} \exp\left( -\frac{(n-1)s^2 + n(\bar{x} - C)^2}{2\sigma^2} \right) d\sigma = T_{n-1}\left(\frac{C - \bar{x}}{s/\sqrt{n}}\right)$$

自由度  $n-1$  の Student の t 分布

# タイプA評価へのベイズ統計の利用(続き)

□ 測定結果  $y =$  期待値  $E[C] = \bar{x}$

□ 標準不確かさ  $u(y) = \sqrt{\text{分散} V[C]} = \sqrt{\frac{n-1}{n-3}} \frac{s}{\sqrt{n}}$

$t$  分布の標準偏差

□ 95%信用区間(credible interval)<sup>注1)</sup>

$$C = \bar{x} \pm t_{n-1}(0.95) \frac{s}{\sqrt{n}}$$

注1: 頻度主義における信頼区間(confidence interval)と意味が違いため、このように呼ばれる

注2:  $p(C | \mathbf{x}) = T_{n-1}\left(\frac{C - \bar{x}}{s/\sqrt{n}}\right)$  において、 $C$  が確率変数,

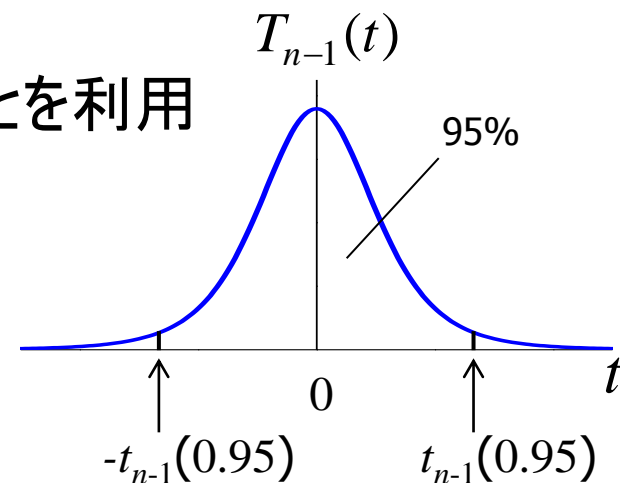
$\mathbf{x}$ ,  $\bar{x}$ ,  $s$  は確定値として扱われている

# 頻度主義からのアプローチ

$\mathbf{x} = (x_1, x_2, \dots, x_n)$  の平均値を  $\bar{x}$ 、実験標準偏差を  $s$  とすると

$t = \frac{\bar{x} - C}{s/\sqrt{n}}$  が自由度  $n-1$  の  $t$  分布に従うことを利用

$$\Pr\left(-t_{n-1}(0.95) < \frac{\bar{x} - C}{s/\sqrt{n}} < t_{n-1}(0.95)\right) = 0.95$$



$$\Rightarrow 95\% \text{信頼区間} : C = \bar{x} \pm t_{n-1}(0.95) \frac{s}{\sqrt{n}}$$

注1: この例は、信用区間と信頼区間が一致する例。一般には一致しない。

注2: "95%" は  $C$  がこの区間に含まれる"確率"を表すものではない。  
 何度もこの方式で区間推定したときの"成功割合"を表す。



# 頻度主義統計 vs ベイズ統計

## □ 頻度主義統計学(伝統的な統計学)

- 頻度にもとづく、「確率」の厳格な定義
- 推定対象の母数  $\Rightarrow$  固定値(不可知)、 データ  $\Rightarrow$  確率変数
- 適用できる問題に限られる(「Kennedy大統領の殺害犯がL.H. Oswaldである確率」は対象外)
- 科学的妥当性は広く認められている

## □ ベイズ統計学

- 主観確率を許容(確信度) ... 日常感覚とは合う
- 推定対象の母数  $\Rightarrow$  確率変数、(取得後の)データ  $\Rightarrow$  固定値  
(頻度主義と逆)
- 柔軟で、適用範囲が広い
- 事前分布の設定にしばしば曖昧さが生じる
- 科学的妥当性は、なお論争の対象となっている

# 信頼水準／信頼の水準／信用水準

- 信頼水準 (Confidence level) … 頻度主義  
同じ方式で何回も区間推定したときの成功割合 (何らかの確率変数に結びつけられた「確率」とは解釈されない)
- 信頼の水準 (Level of confidence) … GUM
  - 「合理的に測定量に含まれ得る値」が、ある区間に含まれる割合
  - 頻度主義の論理に沿って計算されるが、タイプBの確率概念が混在するため、包含確率として説明
- 信用水準 (Credibility level) … ベイズ統計  
推定対象母数が、ある範囲に含まれる主観的確率

# ベイズ統計における自由度

## 頻度主義

□ 自由度 = 実験分散の(実験毎の)ばらつきを表す尺度

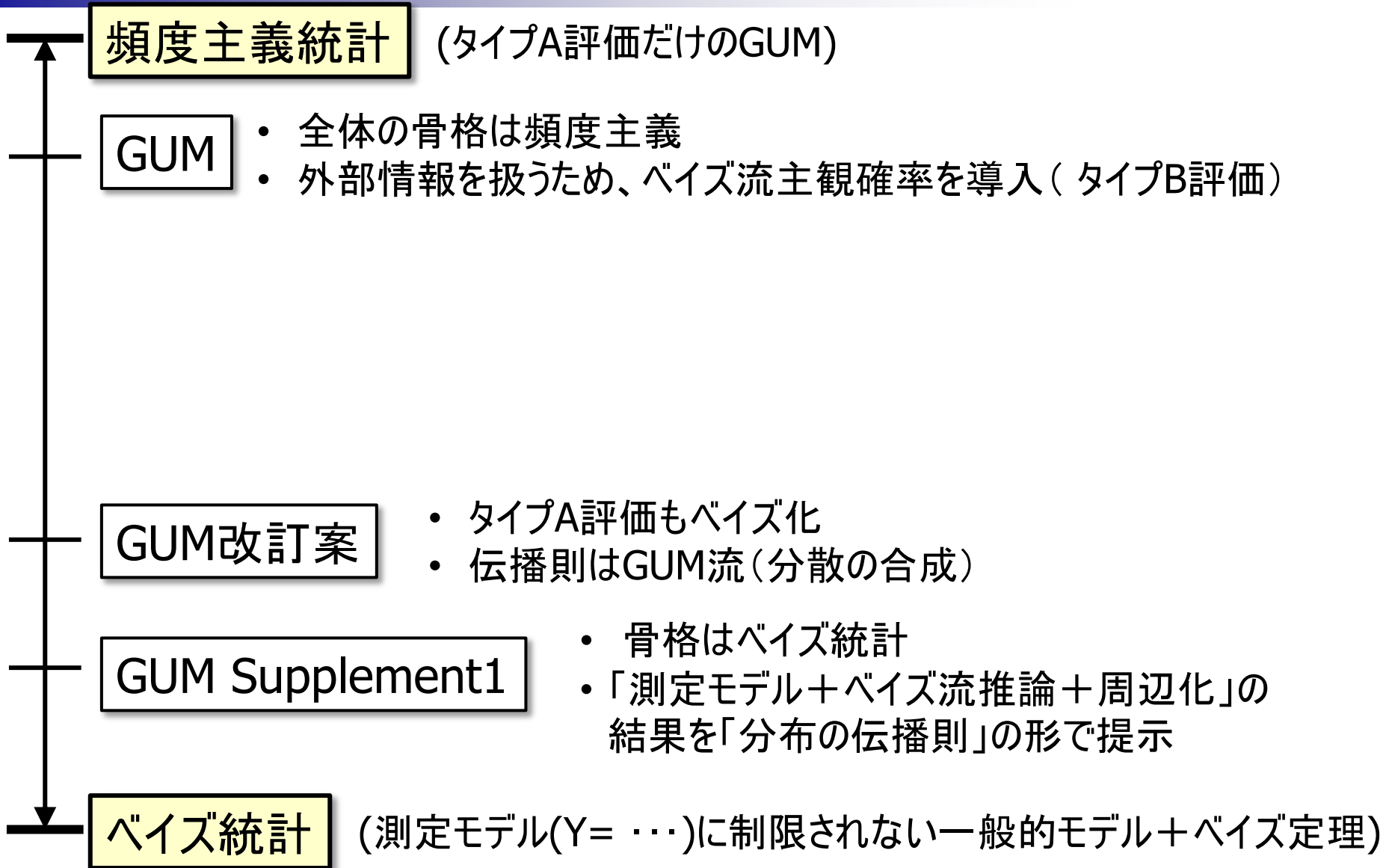
## ベイズ統計

□ 自由度の概念は現れない

□ 手持ちの情報はすべて確率分布として表現され、情報の曖昧さは、確率分布自体に盛り込まれる

- ただし、Aタイプ評価の対象量の周辺事後分布として、自由度(n-1)の t 分布が自然に現れることがある (統計モデルに正規分布を仮定し、無情報事前分布  $p(\mu, \sigma) \propto 1/\sigma$  を使った場合)
- この際の (n-1)は単に t分布のパラメータとしての意味しかもたず、情報の曖昧さの指標ではない

# 「ベイズっぽさ」のレベル



# VIMにおける不確かさ

## VIM第2版(1993)

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

(Parameter, associated with the result of a measurement, that characterizes the dispersion of the values that could reasonably be attributed to the measurand)

## VIM第3版(2007)

用いる情報に基づいて、測定対象量に帰属する量の値のばらつきを特徴付ける負ではないパラメータ

(non-negative parameter characterizing the dispersion of the quantity values being attributed to a measurand, based on the information used)

ベイズ統計への配慮がみられる

# GUM改訂案の概要

□ タイプA評価(繰り返し数  $n$ )における標準不確かさ

$$\frac{s}{\sqrt{n}} \text{ (現行)} \Rightarrow \sqrt{\frac{n-1}{n-3}} \frac{s}{\sqrt{n}} \text{ (ただし } n \geq 4 \text{ )}$$

- $n \leq 3$  では t分布の分散が求まらない。この場合の対応策は現時点で不明。
- Kacker & Jones (Metrologia, 2003)には、「t分布の95%信頼限界半幅 /1.96」で代用との提案がある。

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

# GUM改訂案の概要(続き)

□ 不確かさ伝播則は、継続して使用

□ 拡張不確かさ(包含区間)の計算手続きについては、現時点で不透明

○ 可能性1) GUM Supplement 1を引用 (⇒ 数値計算が必要。普及困難?)

○ 可能性2) 分布形非依存の包含係数を利用

区間  $y \pm k u(y)$  は任意の分布に対して少なくとも  $(1 - 1/k^2)$  を包含

(Chebyshev不等式) → 95% 包含係数  $k = 4.47$  (他にGauss不等式など)

(⇒ 包含区間が無駄に大きくなりすぎる...)

○ 可能性3) 中心極限定理を援用し、正規分布を仮定:  $k \cong 2$  (⇒ 近似の一般的妥当性?)

○ 可能性4) 上記を併記

○ 可能性5) その他(?)

# モデル計算による4つの評価方法の比較

## □ 評価方法

[1] GUM (GUM)

[2] GUM改訂案 (GUM2)

[3] GUM補完文書1 [Monte Carlo法による分布の伝播] (MC)

[4] 本来のBayes (Bayes)

## □ 対象モデル

$$X = B \times Y$$

測定器の応答 (例: スペクトル強度)      感度      測定量 (例: ある物質の濃度)



# モデル計算(続き)

## □ 入力量

$X$  : タイプA評価  $x_i = 99.71, 104.66, 96.26, 97.81, 105.87$

$$\Rightarrow \begin{cases} \bar{x} = 100.86 \text{ (平均)} \\ s = 4.22 \text{ (実験標準偏差)} \end{cases}$$

$B$  : タイプB評価

$B \sim$  正規分布(中心 1.0, 標準偏差  $u_B$ )

$$u_B = \begin{cases} 0.004 & \text{(タイプB < タイプA)} & \cdots & \text{[case 1]} \\ 0.02 & \text{(タイプB} \approx \text{タイプA)} & \cdots & \text{[case 2]} \\ 0.1 & \text{(タイプB > タイプA)} & \cdots & \text{[case 3]} \end{cases}$$

# モデル計算(続き) … [1] GUM

$$Y = \frac{X}{B} \quad \dots \text{測定の数学的モデル}$$

□ 測定結果:  $y = \frac{100.86}{1.0} = 100.86$

□ 不確かさの伝播則:  $\left(\frac{u(y)}{y}\right)^2 = \left(\frac{u(\bar{x})}{\bar{x}}\right)^2 + \left(\frac{u(b)}{b}\right)^2 = \left(\frac{4.22/\sqrt{5}}{100.86}\right)^2 + \left(\frac{0.02}{1.0}\right)^2$

□ 有効自由度 (Welch-Satterthwaiteの式):

$$\frac{[u(y)/y]^4}{\nu_{eff}} = \frac{[u(\bar{x})/\bar{x}]^4}{\nu_{\bar{x}}} + \frac{[u(b)/b]^4}{\nu_b} = \frac{[1.89/100.86]^4}{4} + \frac{[0.02/1.0]^4}{\infty}$$

□ 包含係数:  $k = t(\nu_{eff} = 18.3, 95\%) = 2.10$  (Case 2の場合)

# モデル計算(続き) … [2] GUM2

$$Y = \frac{X}{B} \quad \dots \text{測定の数学的モデル}$$

□ 測定結果:  $y = \frac{100.86}{1.0} = 100.86$

□ 不確かさの伝播則: GUMと異なる部分

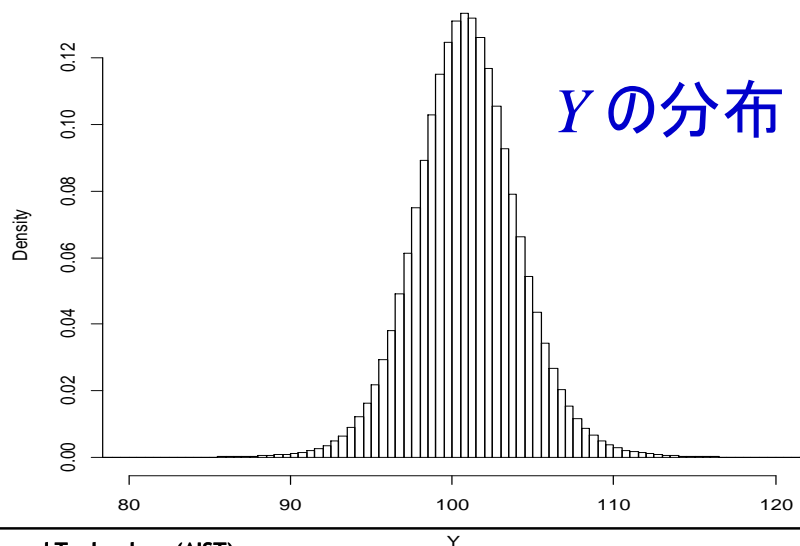
$$\left( \frac{u(y)}{y} \right)^2 = \left( \frac{s \sqrt{\frac{n-1}{n-3}}}{\sqrt{n} \bar{x}} \right)^2 + \left( \frac{u(b)}{b} \right)^2 = \left( \frac{4.22 \sqrt{\frac{4}{2}}}{\sqrt{5} \cdot 100.86} \right)^2 + \left( \frac{0.02}{1.0} \right)^2$$

□ 包含係数: 中心極限定理を根拠に、 $Y$  が正規分布に従うとみなすことにするならば、 $k = 1.96$

# モデル計算(続き) … [3] MC

$$Y = \frac{X}{B} \quad (\text{測定の数学的モデル}) \quad \dots (1)$$

- $X \sim$  (ベイズ統計における周辺事後分布として)中心を  $\bar{x} = 100.86$  に shiftし、分布幅を  $s/\sqrt{5} = 1.887$  倍した、自由度4の scaled and shifted t-分布に従う乱数
- $B \sim$  正規分布(中心 1.0, 標準偏差  $u_B$ )に従う乱数
- 式(1)に代入して得られる  $Y$  の分布 (右図)から標準偏差と95%信頼区間を求める



# モデル計算(続き) … [4] Bayes

□  $X = B \cdot Y$  に対する測定値  $x_i$  のモデル分布:

$$p(x_i | X, \sigma) = \text{正規分布(中心 } X, \text{ 標準偏差 } \sigma)$$

□  $B$  の事前分布:  $p(B) = \text{正規分布(中心 } 1.0, \text{ 標準偏差 } u_B)$

□  $Y$  の事前分布:  $p(Y) \propto \text{const. (無情報事前分布)}$

□  $\sigma$  の事前分布:  $p(\sigma) \propto 1/\sigma$  (無情報事前分布)

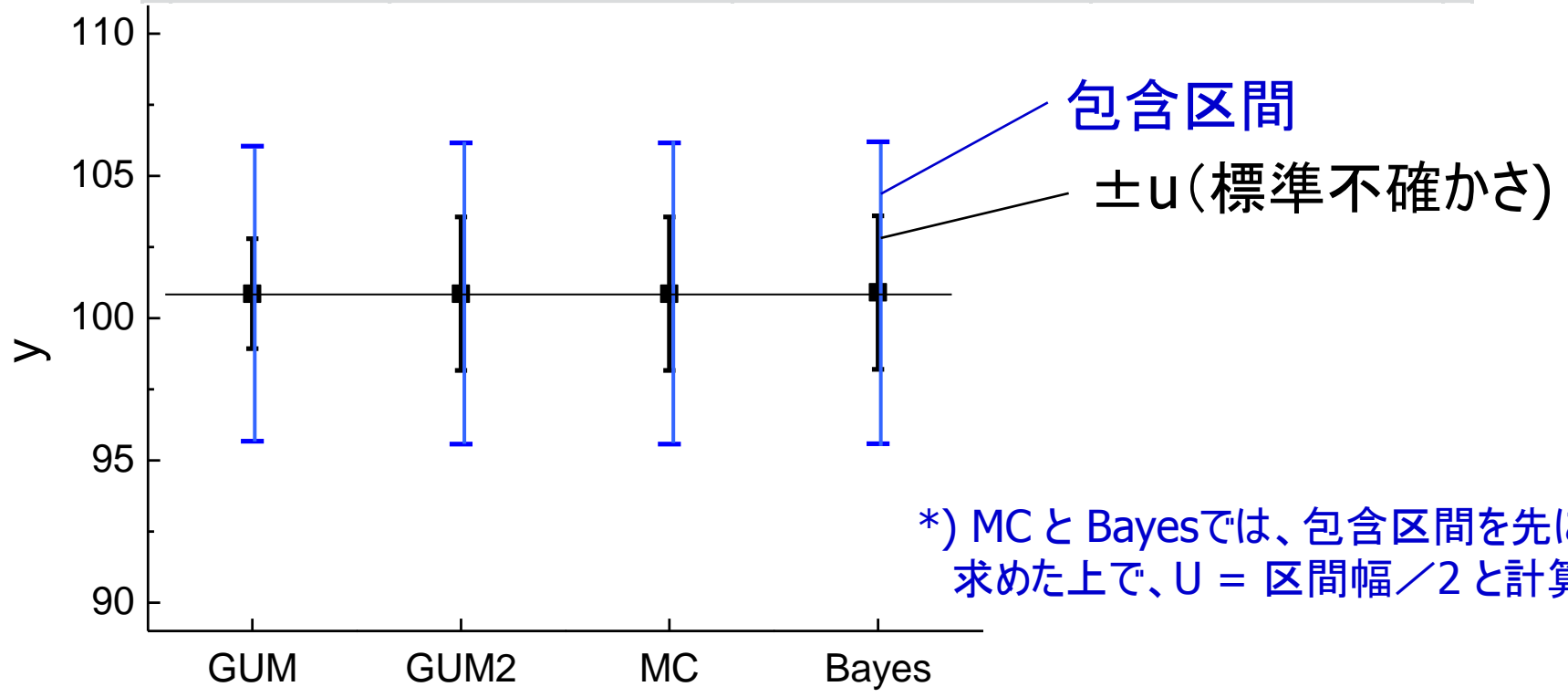
$$\text{事後分布 } p(Y, B, \sigma | \{x_i\}) \propto \left[ \prod_i p(x_i | BY, \sigma) \right] p(B) p(Y) p(\sigma)$$

$$\text{周辺事後分布 } p(Y | \{x_i\}) = \iint p(Y, B, \sigma | \{x_i\}) d\sigma dB$$

(実際の計算は WinBUGS Ver. 1.4.3による)

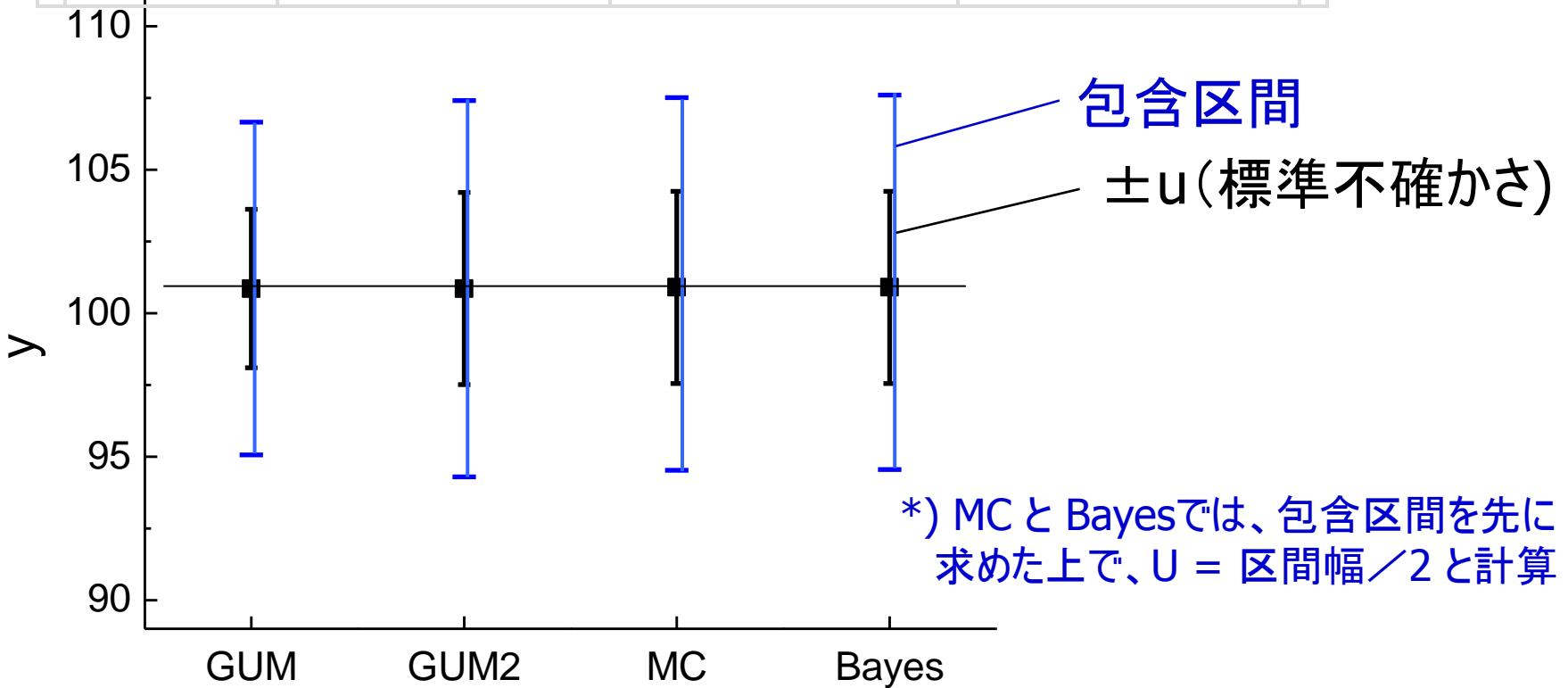
# モデル計算結果 [Case 1 (タイプB < タイプA)]

	測定結果 $y$	標準不確かさ $u$	拡張不確かさ $U$ *)
GUM	100.86	1.93	5.19
GUM2	100.86	2.70	5.29
MC	100.86	2.70	5.30
Bayes	100.90	2.70	5.31



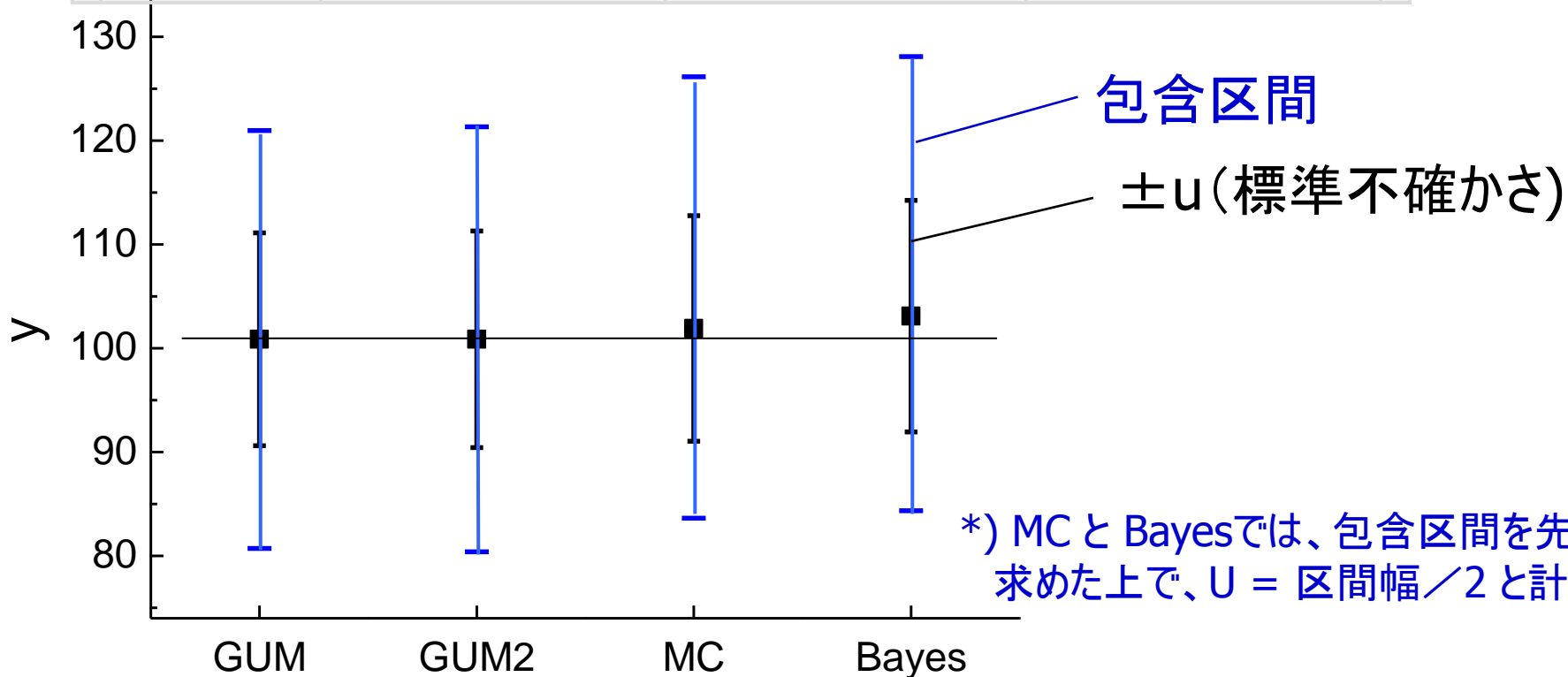
# モデル計算結果(続き) [Case 2 (タイプB ≈ タイプA)]

	測定結果 $y$	標準不確かさ $u$	拡張不確かさ $U$ *)
GUM	100.86	2.76	5.80
GUM2	100.86	3.35	6.56
MC	100.90	3.35	6.50
Bayes	100.90	3.35	6.53



# モデル計算結果(続き) [Case 3 (タイプB > タイプA)]

	測定結果 $y$	標準不確かさ $u$	拡張不確かさ $U^*$
GUM	100.86	10.26	20.12
GUM2	100.86	10.43	20.45
MC	101.90	10.87	21.25
Bayes	103.10	11.16	21.87

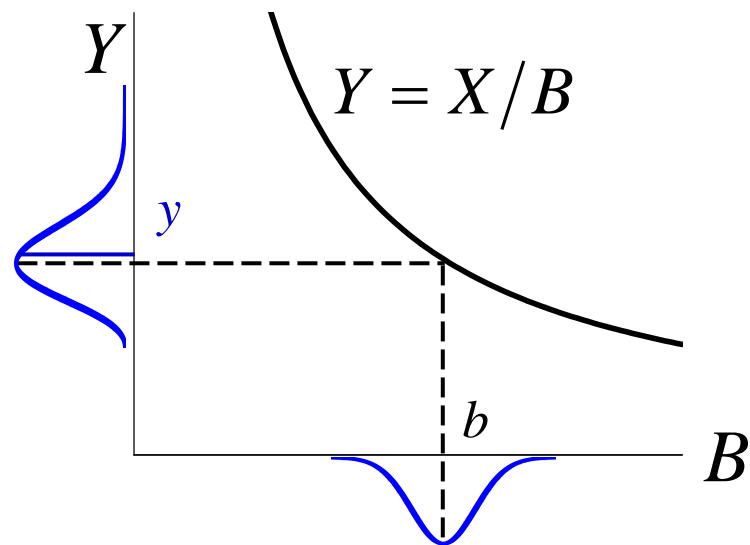


\*) MC と Bayesでは、包含区間を先に求めた上で、 $U = \text{区間幅} / 2$  と計算



# 4つの評価方法の比較 — まとめ

- 一般に、評価方法によって不確かさの大きさは異なる
- GUM改訂案を含めベイズ統計を利用する方法では  $u$  (標準不確かさ) がGUMと比べて顕著に大きくなることもある (自由度の小さいタイプA不確かさが支配的な場合) が、 $U$  (拡張不確かさ) ではその違いは縮小する
- MC と Bayesでは、(不確かさが大きい場合には) モデルの非線形性が不確かさだけでなく、測定結果にも反映される



# ベイズ統計導入に慎重な意見もある

## □ ベイズ統計導入のメリット

- タイプA評価とタイプB評価で、一貫した確率概念に基づいて評価できる
- それは、日常的な「確率概念」に近い
- 妥当な算定が困難だった有効自由度の計算が不要となる

## □ ベイズ統計導入のデメリット

- “真の値” を確率変数とみなすことの不合理や違和感  
例) 「プランク定数が確率変数(ベイズ)」  $\leftrightarrow$  「プランク定数は固定されており、その測定データが確率変数(頻度主義)」
- 統計学の専門家の間でも、ベイズ統計の妥当性になお議論がある
- 推定の成功率(Long-run success rate)において頻度主義に劣ることがある
- 計算機による数値計算が必要となることが多い
- 頻度主義統計ほどよく知られていない
- まだ普及途上にあるGUMの今後の普及が阻害される可能性がある

# Comments (by the speaker)

- 統計学(Statistics)と計量学(Metrology)は別物。統計学上の整合性を、実学である計量学でも要求すべきかどうかは、検討の余地がある。
- 現行GUMでは、不確かさの第一義的表現は標準不確かさ(例: 基礎物理定数データベース)。標準不確かさの評価では、確率概念の不整合は大きな問題とならない。拡張不確かさ(or 包含区間)が産業界等で本当に必要とされているかどうかは、あらためて検討の余地がある。

# まとめ

- ベイズ統計を取り入れた、GUMの改訂作業が進められている(早ければ2014年中に草稿が公開?)
- 予想される主要な変更点
  - タイプA評価(標準不確かさに因子  $\sqrt{\frac{n-1}{n-3}}$  がかかる)
  - 自由度が不要になる
  - [不確定要素あり] 包含区間(拡張不確かさ)の決め方
  - 一貫性のある確率概念
- 変わらない点
  - タイプB評価
  - 不確かさの伝播則
- 改訂GUMの受容にあたって…
  - ベイズ統計の導入により産業界等が受けるメリット・デメリットの冷静な見極め
  - ベイズ統計の導入に反対する意見があることについての配慮
  - 過渡期の混乱を避ける工夫

# 謝辞

今井秀孝様 (産業技術総合研究所、  
製品評価技術基盤機構)

小池昌義様 (産業技術総合研究所)

田中秀幸様 (産業技術総合研究所)

城野克広様 (産業技術総合研究所)