Explainability and interpretability of the GUM methodologies

Alistair Forbes[1]

[1]National Physical Laboratory, UK
Data Science Department

MathMet Conference, 2–4 November 2022, Paris

NPL National Physical Laboratory

Motivation

- Uncertainty quantification (UQ) playing a bigger role in scientific endeavour
- Role of metrology in helping the more general scientific community in evaluating and working with (measurement) uncertainty (30 years or so of the GUM)
- Developing UQ methodologies associated with new measurement modalities, new data analysis algorithms, machine learning, AI, etc.
- Supporting trust, interoperability and interpretability through frameworks for traceability
- UQ: uncertainty quantification of what and how is it quantified?

NPL National Physical Laboratory

Various and varied GUM methodologies

- GUM(LPU): law of propagation of uncertainty (first and second moments), through an input-output model
- GUM(MCM): propagation of distributions, through an input-output model
- GUM(W-S): Welch-Satterthwaite formula and coverage/confidence intervals
- GUM(Bayes): Bayesian inference applied to measurement uncertainty evaluation, measurement equation/observation equation approach, forward problems/inverse problems
- Subjective models, objective inference
- Computational feasibility

NPL National Physical Laboratory

Probabilistic inference

- Product rule:

$$p(\boldsymbol{a}, \boldsymbol{b}) = p(\boldsymbol{a}|\boldsymbol{b})p(\boldsymbol{b})$$

- Change of variable formula: if

$$\boldsymbol{b} = \boldsymbol{f}(\boldsymbol{a}), \quad \boldsymbol{a} = \boldsymbol{g}(\boldsymbol{b}), \quad p_A(\boldsymbol{a}),$$

then

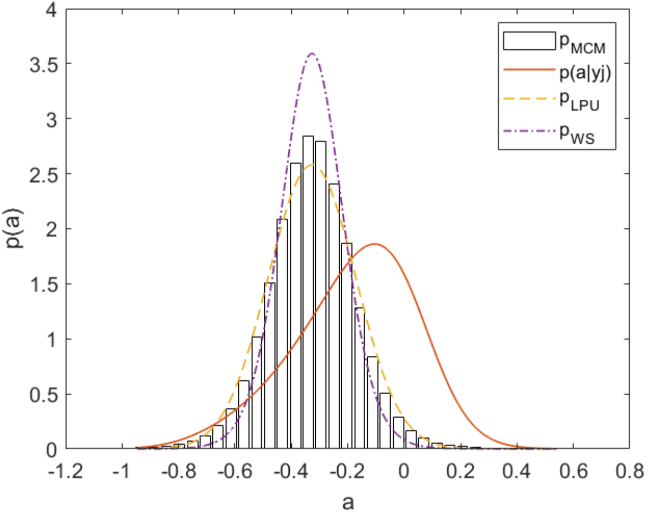
$$p_B(\boldsymbol{b}) = p_A(\boldsymbol{g}(\boldsymbol{b}))|J|$$

- Marginalisation:

$$p(\boldsymbol{a}) = \int p(\boldsymbol{a}, \boldsymbol{b})\mathrm{d}\boldsymbol{b}$$

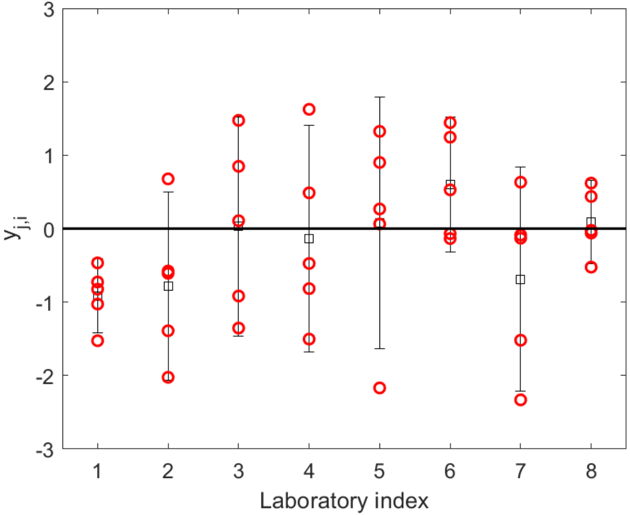
National Physical Laboratory

The body in the library



National Physical Laboratory

Simulated ILC data involving repeated measurements



Repeated measurements of single quantity

- Suppose observations $\boldsymbol{y} = (y_1, \ldots, y_m)^\top$, arise in according to the model

$$\boldsymbol{y}|a, \sigma^2 \sim \mathrm{N}(\boldsymbol{e}a, \sigma^2 I), \quad \boldsymbol{e} = (1, 1, \ldots, 1)^\top$$

- Sample mean $\bar{y}$ and sample variance:

$$\bar{y} = \frac{1}{m}\sum_{i=1}^{m} y_i, \quad s^2 = \frac{1}{m-1}\sum_{i=1}^{m}(y_i - \bar{y})^2.$$

- In classical inference, a is estimated by $\hat{a} = \bar{y}$ and σ^2 is estimated by s^2.

Confidence intervals

- Sampling distributions for $\hat{a}$ and s^2:

$$\hat{a}|a,\sigma^2 \sim \mathcal{N}(a,\sigma^2/m), \quad s^2|\sigma^2 \sim \mathcal{G}(\alpha,\beta), \quad \alpha = \frac{m-1}{2} \quad \beta = \frac{m-1}{2}\frac{1}{\sigma^2}$$

- Setting $u = \sqrt{s^2/m}$,

$$z = \frac{\hat{a}-a}{u} \sim t_\nu(0,1), \quad \nu = m-1.$$

- If k represents the 0.975 percentile point the of t-distribution CDF, then

$$[\hat{a}-U, \hat{a}+U], \quad U = ku,$$

involving sampled $\hat{a}$ and U will contain a 95 % of the time.

- Classical inference does not/cannot assign a t-distribution state of knowledge distribution for a.

National Physical Laboratory

Bayesian state of knowledge distribution

- Assuming a non-informative prior distribution $p(a, \sigma^2) \propto 1/\sigma^2$, then

$$p(a, \sigma^2 | \boldsymbol{y}) = p(a|\sigma^2, \boldsymbol{y}) p(\sigma^2 | \boldsymbol{y})$$

  with

$$a|\sigma^2, \boldsymbol{y} \sim \mathcal{N}(\hat{a}, \sigma^2/m), \quad \sigma^2|\boldsymbol{y} \sim \mathcal{IG}((m-1)/2, (m-1)s^2/2).$$

- Marginalisation:

$$a|\boldsymbol{y} \sim t_{m-1}(\hat{a}, s^2/m).$$

- Note that $\hat{a}$ is not a sample from a t-distribution centred on a; the model has $\hat{a} \in \mathcal{N}(a, \sigma^2/m)$.

ILC involving repeated measurements

- Suppose the jth laboratory, $j = 1, \ldots, n$, provides data $\boldsymbol{y}_j$ according to the model
$$\boldsymbol{y}_j \in \mathcal{N}(\boldsymbol{e}_j a, \sigma_j^2 I).$$
- For non-informative priors $p(a) \propto 1$, $p(\sigma_j^2) \propto 1/\sigma_j$,
$$a|\boldsymbol{y}_j \sim t_{m_j-1}(\hat{a}_j, s_j^2/m_j)$$
- Can marginalise the joint posterior $p(a, \{\sigma_j^2\}|\{\boldsymbol{y}_j\})$ analytically with respect to $\{\sigma_j^2\}$:
$$p(a|\{\boldsymbol{y}_j\}) \propto \prod_{j=1}^{n} p_t(\hat{a}_j|a, s_j^2/m_j, m_j - 1),$$
a product of t-distributions.

Least squares analysis of ILC data

- Suppose
$$x_j \in \mathcal{N}(a, u_j^2), \quad j = 1, \ldots, n,$$
- a can be estimated by
$$\hat{a} = \sum_{j=1}^{n} c_j x_j, \quad c_j \propto 1/u_j^2, \quad \sum_j c_j = 1.$$

  with
$$u^2(a) = \sum c_j^2 u_j^2, \quad \hat{a} \in \mathcal{N}(a, u^2(a)).$$
- For $p(a) \propto 1$,
$$a|\boldsymbol{x} \sim \mathcal{N}(\hat{a}, u^2(a)).$$

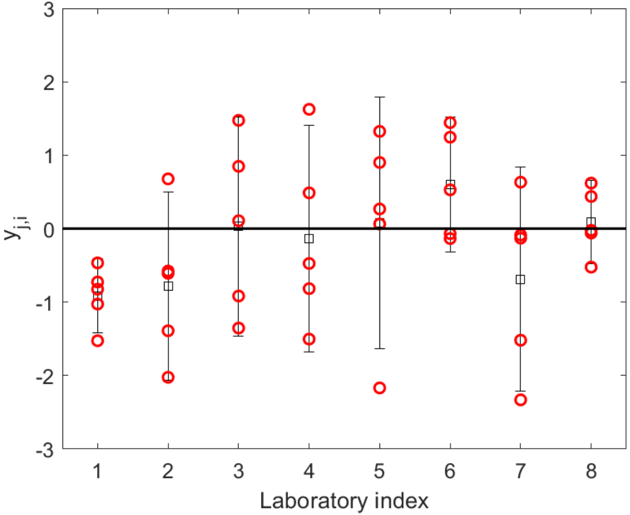
NPL National Physical Laboratory

Weighted least-squares analysis

	1	2	3	4	5	6	7	8
x_j	-0.91	-0.78	0.03	-0.14	0.08	0.60	-0.69	0.09
u_j	0.25	0.64	0.75	0.77	0.86	0.46	0.76	0.29
c_j	0.38	0.06	0.04	0.04	0.03	0.11	0.04	0.29

NPL National Physical Laboratory

Simulated ILC data involving repeated measurements



Least squares analysis based on $p(a|\boldsymbol{y}_j)$, $j = 1, \ldots, n$

- Our repeated measurements model yields

$$a|\boldsymbol{y}_j \sim t_{m_j-1}(\hat{a}_j, s_j^2/m_j)$$

- Assert that $\hat{a}_j \in t_{m_j-1}(a, s_j^2/m_j)$.
- Assuming $m_j > 3$, $j = 1, \ldots, n$, set

$$u_j^2 = \frac{m_j - 1}{m_j - 3}\frac{s_j^2}{m_j}$$

and determined least squares estimate $\hat{a}$, and Gaussian distribution associated with a.

National Physical Laboratory

Propagation of distributions approach

- Assert that $\hat{a}_j \in t_{m_j-1}(a, s_j^2/m_j)$.
- Least squares estimate $\hat{a}$ is a linear combination $\sum_j c_j \hat{a}_j$ of the $\hat{a}_j$
- Use MCM to propagate input t-distributions to output distribution for $\hat{a}$, a convolution of t-distributions
- Two possible implementations:

$$x_{j,q} \in t_{m_j}(\hat{a}, s_j^2/m_j), \quad x_{j,q} = \hat{a} + \delta_{j,q}$$

or

$$x_{j,q} \in t_{m_j}(\hat{a}_j, s_j^2/m_j), \quad x_{j,q} = \hat{a}_j + \delta_{j,q}$$

- Evaluate

$$\hat{a}_q = \sum_j c_j x_{j,q} = \sum_j c_j(\hat{a} + \delta_{j,q}) = \sum_j c_j(\hat{a}_j + \delta_{j,q}),$$

since $\hat{a} = \sum c_j \hat{a}_j$ and $\sum c_j = 1$.

Application of Welch-Satterthwaite

- Assert that $\hat{a}_j \in t_{m_j-1}(a, s_j^2/m_j)$.
- LS estimate defined in terms of a sum of independent t-variates $\hat{a} = \sum_{j=1}^n c_j \hat{a}_j$.
- Set

$$u_j^2(a) = c_j^2 s_j^2/m_j, \quad u^2(a) = \sum_j u_j^2(a), \quad \nu_{eff} = \frac{u^4(a)}{\sum_j u_j^4(a)/(m_j - 1)}$$

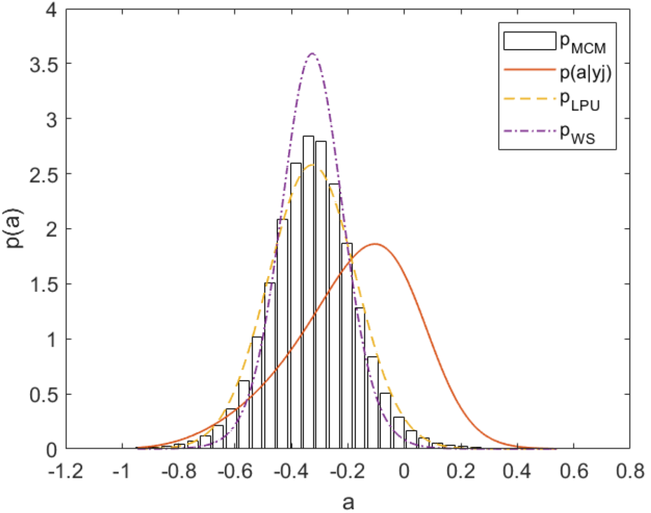
- GUM(W-S): a is associated with the t-distribution $t_{\nu_{eff}}(\hat{a}, u^2(a))$
- W-S provides a gamma approximant to the sampling distribution for $u^2(a)$, a sum of gamma variates, given $\{\sigma_j^2\}$, while

$$\hat{a} \in \mathcal{N}(a, \sigma^2), \quad \sigma^2 = \sum_j c_j^2 \sigma_j^2/m_j$$

and $p(\sigma^2|\{\boldsymbol{y}_j\})$ is a sum of inverse gamma variates

NPL National Physical Laboratory

Mystery solved!



Summary

- GUM methodologies correspond to different (implicit) models, different populations
- State of knowledge distributions are not sampling distributions, in general
- Welch-Satterthwaite is not consistent with LPU (or MCM or Bayes), and does not produce a state of knowledge distribution, in general, (but can deliver confidence intervals quite effectively in some cases)
- A statement of uncertainty can only be interpreted correctly if the underlying models and methodology are made explicit

National Physical Laboratory