



Paradox? What paradox?

R. Willink^{1,2}

Received: 16 October 2023 / Accepted: 24 January 2024
© The Author(s) 2024

Abstract

This article is a response to the preceding paper by Huang, who considers a recent result of Willink (Measurement: Sensors, 24:100416, 2022) and who describes the result as a paradox. The result implied that a set of information or a “state of knowledge” about a measurand cannot be identified with a unique probability distribution for the measurand, contrary to what seems suggested in the literature surrounding the revision of the *Guide to the Expression of Uncertainty in Measurement*. The result is restated and viewed in the context of CIPM Recommendation INC-1, which was foundational in the original development of the *Guide*. It is argued that the result is a proof, not a paradox, and that it will only appear paradoxical to those who have adopted an incorrect premise about probability. The idea of having “information” about the true value of a measurand is discussed and contrasted with the idea of having “belief” about it. The material supports the view that the analysis of measurement uncertainty is to be based on classical statistical principles.

Introduction

I am grateful to *Accreditation and Quality Assurance* for the opportunity to respond to this paper by Hening Huang, and I thank Dr Huang for his interest in the logical contradiction described in my earlier paper [1]. Dr Huang accepts the accuracy of the mathematical and logical result, but he and I seem to have a different view of its significance for the metrological community. He calls the contradiction a paradox, but I shall argue that it is not a paradox—it is just a proof, and it has profound implications for the role of probability in metrological data analysis.

A presumption

The contradiction in question follows from the presumption that a probability distribution about a constant can or does encode information about that quantity. In the language and notation of the literature relating to the *Guide to the Expression of Uncertainty in Measurement* (GUM), such a constant is an input quantity X or an output quantity Y . So the result

applies directly to the mode of reasoning advocated in the documents of Working Group 1 of the Joint Committee for Guides in Metrology (JCGM-WG1), which build upon and (arguably) reinterpret the original GUM [2] and which promote a non-classical, Bayesian, form of uncertainty analysis. For example, the first published supplement to the GUM [3, Introduction] states:

The PDF for a quantity expresses the state of knowledge about the quantity, i.e. it quantifies the degree of belief about the values that can be assigned to the quantity based on the available information.

So a probability distribution about a constant is being identified with a set of information about that constant. Thus, there is a premise which can be expressed as:

P: To any set of information about a quantity, there is a unique pdf to represent that quantity.

However, a logical contradiction ensues from this premise when we legitimately manipulate distributions that are deemed to correspond to different sets of information about the same quantity [1].

The contradiction

Specifically, we envisage two disjoint sets of information about a quantity X , with corresponding proposed probability density functions (pdfs) $f_{1X}(x)$ and $f_{2X}(x)$. It can be shown

✉ R. Willink
robin.willink@otago.ac.nz

¹ Measurement Standards Laboratory of New Zealand, Lower Hutt, New Zealand

² University of Otago, Wellington, New Zealand

that if we combine the two sets of information then the resulting pdf must obey the natural result $f_{12X}(x) \propto f_{1X}(x)f_{2X}(x)$ [1, Sec. 3.2]. Now suppose that the measurand is $Y = g(X)$. It is well known that $f_{jY}(y) = f_{jX}(x) |dx/dy|$, where $y = g(x)$. By combining the two pdfs before transforming from X to Y , we obtain $f_{12Y}(y) \propto f_1(x)f_2(x) |dx/dy|$, but by transforming before combining we obtain $f_{12Y}(y) \propto f_1(x)f_2(x) |dx/dy|^2$. The difference in the exponent means that the two resulting pdfs differ if g is nonlinear, as with $Y = \log(X)$. Thus, the presumption that a set of information can be accurately and meaningfully encoded as a probability distribution leads to an internal inconsistency [1, Sec. 3.3]. Therefore, this presumption must be incorrect. Huang calls the inconsistency the ‘Willink paradox’. I think of it simply as a proof that the premise of the analysis, P , is untenable.

Huang accepts the accuracy of the result, but his approach is to ask which of the pdfs $f_{12Y}(y) \propto f_1(x)f_2(x) |dx/dy|$ and $f_{12Y}(y) \propto f_{1X}(x)f_{2X}(x) |dx/dy|^2$ is to be used in a practical analysis. That is to miss the point, which is that any pdf would be wrong because it would be based on a wrong presumption. Neither of the two pdfs is to be used, and—to be logical—no other pdf can be used. The result simply shows that a *pdf does not encode information*. The metrology community cannot attribute pdfs to input quantities while retaining scientific credibility, and it does not matter which of the two pdfs is chosen here because they are both as bad as each other.

The role of probability in metrology

The logical error is associated with a misunderstanding of the role of probability and the nature of information. A probability can either represent a *relative frequency* or a *subjective degree of belief*, the former in the paradigm of frequentist statistics and the latter in the contrasting paradigm of subjective Bayesian statistics. But the contradiction shows that a probability cannot represent ‘rational belief’, ‘information’ or ‘knowledge’, which would correspond to a third possible role applicable in ‘objective Bayesian statistics’. Yet, in the documents that have appeared from JCGM-WG1 since it took over maintenance of the GUM, that illegitimate role for probability is what metrologists are being encouraged and compelled to accept.

Huang seems to have accepted that constants can and must be given probability distributions, for he writes about the two pdfs “We must choose one or the other.” To this I respond, “Not at all! What must be chosen is a new premise. You cannot in good conscience continue to assign probability distributions to constants believing it to be a scientific thing to do.” If metrologists want to retain credibility in their data analysis then they must return to the

classical paradigm of statistics, where distributions are not attributed to constants but are only used to model potential measurements and potential errors. That is the paradigm that brought us the ideas of confidence interval, level of confidence, minimum-variance unbiased estimation, least-squares linear regression, polynomial regression, chi-square tests of consistency and weighted-mean reference values, etc. That is the paradigm in which Type A evaluation of uncertainty has been carried out for many years and that is the paradigm in which Type B analysis was proposed via Recommendation INC-1 [4] [2, 0.7] and can be accommodated [5].

Recommendation INC-1 marked a watershed for our subject. A group of experts met in 1980 to discuss the difficult question of how to combine expressions of measurement uncertainty arising from systematic and random errors. How were systematic errors to be incorporated, given the existing formalism that dealt adequately with random, i.e., statistical, errors? In the report that accompanied Recommendation INC-1 [6, pp. 7, 8], we read:

The only viable solution to this problem, it seems, is to follow the prescription contained in the well-known general law of “error propagation”. The essential quantities appearing in this law are the variances (and covariances) of the variables (measurements) involved.

and

In these approaches it is necessary to make (at least implicitly) some assumption about the *underlying population*. It is left to the personal preference of the experimenter whether this is supposed to be for instance Gaussian or rectangular.

(Italicization added here.) From these and other sections of the report, it can be correctly inferred that the assumed variance (in what was to become known as a Type B evaluation) is the variance of an imagined population of measurement errors when measuring a constant, not the variance of a probability distribution attributed to such a constant. (So, the association of Type B evaluation with a Bayesian view of the role of probability seems to have followed a misinterpretation of Recommendation INC-1.)

A paradox of scientists?

In the result discussed by Dr Huang, there is no paradox. There is only a proof that constants cannot be attributed probability distributions in accurate response to information about them. But perhaps there is a paradox or two to be found in such practices. I cherished an influential 1969 book of Bevington [7] during my education. Bevington

died in 1980, but his book lives on in later editions co-written by a former colleague, who has himself since died. In the third edition [8, p.63], but not in the first edition, we read:

Similarly, if we were to repeat the entire experiment many times, ... we should expect that approximately 68% of our determinations of \bar{x} should fall within the range $(\mu - s_\mu) < \bar{x} < (\mu + s_\mu)$ [Then] we make a slight logical leap to state that *there is approximately 68% probability that the true value of the mean μ lies in the range $(\bar{x} - s_\mu) < \mu < (\bar{x} + s_\mu)$* ...

(Emphasis added here.) The new author seems to have joined the party in which probability statements are to be made about constants! But in appearing to treat the constant μ as a random variable with a specified probability of actively lying in a specified numerical range he unwittingly compromises the rest of the book, which is firmly based on classical principles, (which do not involve that practice). In the phrase “we make a slight logical leap”, he implies discomfort with this step, but he proceeds. The logical leap seems to be either (a) to knowingly use misleading language or (b) to replace the classical concept of probability (where the mean of the relevant distribution is the unknown parameter μ) by the fiducial concept of probability (where the experimenter feels entitled to reposition and reorientate the distribution around the estimate \bar{x} , making μ the subject). Whichever was the case, why was that step thought acceptable? The fact that it was thought appropriate is much more paradoxical than the contradiction in my paper. *Just where is the evidence that a probability statement about a constant makes scientific sense?*

One relevant factor is ambiguity with the symbol \bar{x} . In the first sentence of the quoted text the symbol \bar{x} describes something that varies, while in the second sentence it seems to indicate a particular number, which is what the reader would expect. This is an example of where the careful use of notation to distinguish a random variable from a realization of that random variable would have helped. Thus, it is now common for statisticians to use the upper-case symbol \bar{X} to indicate the random variable, i.e., the entity with the property of randomness, and to use the corresponding lower-case symbol \bar{x} to indicate the actual number that resulted. So there is another explanation for the misleading text: perhaps the new author intended the symbol \bar{x} in the second sentence to mean the random quantity. However, the subject of the probability statement, i.e., the entity in possession of the probability, should then have been the random interval $[\bar{X} - s_\mu, \bar{X} + s_\mu]$, not the constant μ , and (in the original notation) the phrase would have been better written as ‘there is approximately 68% probability that the range $(\bar{x} - s_\mu) < \mu < (\bar{x} + s_\mu)$ covers the true value of the mean μ ’.

What is information?

Our context is the (mis)use of probability distributions to describe information about fixed quantities. But what is ‘information’? Envisage having information about an unknown constant number X . Any actual piece of information about X , i.e., any fact about X , involves one or more of the relationships =, > and <, because the only attributes of X are magnitude and sign. So any statement of information about X is a statement of the form “ $0 \leq X \leq 4$ ”, say. The contrasting claim that “ X is more likely” to be in one interval than in another is not informative because X does not possess the attribute of ‘being likely’. Rather, that is a claim of personal belief. (If that claim imparts information at all then it is information about the speaker’s belief about X , not information about X .) From this, we can conclude that there is no actual information about X found in a general pdf for X . It is true that if we knew that $X \not\leq 0$ ($X \geq 0$) and thought it appropriate to attribute X a pdf then we would choose a pdf with zero density at negative values of the dummy variable x . But the knowledge gained from the statement “ $X \not\leq 0$ ” gives us no right to choose any form for that pdf in the feasible positive region. We see that information and pdfs do not go together. A pdf for a constant describes degree of subjective belief, not objective information.

So the two pdfs featuring in the contradiction do not represent genuine information about the measurand Y . Huang chooses between these pdfs by appealing to the idea of entropy found in ‘information theory’, which is a field that grew out of the work of Shannon [9] in communication theory. As explained briefly in my book [10, Section 13.2], Shannon was concerned with maximizing the rate of transmission of letters of a discrete alphabet. So the original context of information entropy related to *information rate in the realization of many discrete random variables*. Yet, by some unclear argument, others later claimed that this could also apply with *information content in the realization of a single continuous random variable*, which is Huang’s context. Like the misuse of the term ‘probability’, this was a misuse of another common word ‘information’. The ‘information’ of information theory and the meaningful information that metrologists have about their measurement techniques are not the same thing: ‘information’ is a qualitative colloquial word.

So Dr Huang’s proposal appeals to a dubious principle in an attempt to find a non-existent solution. Like a set of epicycles in a Ptolemaic model of the solar system, his suggestion acts as paper over a very big crack—a crack that is getting wider and wider. There is no solution to be found without a new starting point, the new but old starting point of classical statistics.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Willink R (2022) On revision of the Guide to the expression of uncertainty in measurement: proofs of fundamental errors in Bayesian approaches. *Meas: Sens* 24:100416
2. BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, and OIML. Guide to the expression of uncertainty in measurement (1995)
3. JCGM (2008) JCGM 101:2008 Evaluation of measurement data—Supplement 1 to the “Guide to the expression of uncertainty in measurement”—Propagation of distributions using a Monte Carlo method. <https://www.bipm.org/en/committees/jc/jcgm/publications>
4. Giacomo P (1981) News from the BIPM. *Metrologia* 17(2):69–74
5. Willink R (2022) An enhanced method for the evaluation of measurement uncertainty. *Meas: Sens* 24:100373
6. Kaarls R (1980) Report of the BIPM working group on the statement of uncertainties (1st meeting—21 to 23 October 1980) to the Comite International des Poids et Mesures
7. Bevington PR (1969) Data reduction and error analysis for the physical sciences. McGraw-Hill, New York
8. Bevington PR, Robinson DK (2003) Data reduction and error analysis for the physical sciences, 3rd edn. McGraw-Hill, New York
9. Shannon CE (1948) A mathematical theory of communication. *Bell Syst Tech J* 27(379–423):623–656
10. Willink R (2013) Measurement uncertainty and probability. Cambridge University Press, Cambridge

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.