



HAL
open science

Incertitude de mesure et probabilités : la confrontation des approches fréquentiste et bayésienne en métrologie

Fabien Grégis, Nadine de Courtenay

► To cite this version:

Fabien Grégis, Nadine de Courtenay. Incertitude de mesure et probabilités : la confrontation des approches fréquentiste et bayésienne en métrologie. Isabelle Drouet. Le bayésianisme aujourd'hui. Fondements et pratiques, Editions Matériologiques, 2016, Collection : Sciences & philosophie, 978-2-37361-078-9. halshs-03372276

HAL Id: halshs-03372276

<https://shs.hal.science/halshs-03372276>

Submitted on 10 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Incertitude de mesure et probabilités : la confrontation des approches fréquentiste et bayésienne en métrologie

Fabien Grégis & Nadine de Courtenay
Université Paris Diderot, Sorbonne Paris Cité,
laboratoire Sphere UMR 7219, F-75205 Paris, France
fabien.gregis@etu.univ-paris-diderot.fr
nadine.decourtenay@univ-paris-diderot.fr

Le bayésianisme aujourd'hui. Fondements et pratiques
Sous la direction d'Isabelle Drouet, Éditions Matériologiques, 2016, pp.317–391
(version de prépublication)

Table des matières

1 Erreurs de mesure et incertitude de mesure	5
1.1 Le GUM et la mission d'uniformisation du calcul des incertitudes	5
1.2 Variabilité des données issues de l'expérience et hypothèse des erreurs de mesure	7
1.3 Deux composantes de l'erreur de mesure	9
1.4 De l'identification des erreurs de mesure au calcul de l'incertitude de mesure .	13
1.5 Équation du mesurande et fonction de mesure	15
2 Traitement statistique des données expérimentales : approche fréquentiste, approche bayésienne	20
2.1 Approche fréquentiste	21
2.2 Approche bayésienne	29
3 Discussion	39
3.1 L'approche bayésienne, une solution épistémique au problème des erreurs systématiques.	40
3.2 Extension du modèle classique : priorité du social sur l'épistémique	48
4 Conclusion	58

Résumé

Le traitement statistique des données expérimentales est au centre d'une profonde mutation méthodologique. Les méthodes fréquentistes traditionnellement mobilisées dans la description des processus de mesure se trouvent aujourd'hui progressivement supplantées par des méthodes bayésiennes. Ce tournant fait l'objet de débats intenses, nourris par la publication, en 1993, du *Guide pour l'expression de l'incertitude de mesure* (GUM) dont le but était d'uniformiser le traitement des mesures au niveau international. Dans un premier temps, nous montrons pourquoi l'évaluation quantitative d'une « incertitude de mesure » est indispensable à la communication d'un résultat utilisable par tous. Nous examinons ensuite comment les nouvelles exigences métrologiques, dictées par le contexte scientifique et économique d'aujourd'hui, ont achoppé sur l'incapacité de l'approche fréquentiste à proposer un traitement probabiliste complet de l'incertitude de mesure, et ont facilité la promotion d'une approche entièrement bayésienne qui répond à cette tâche. Enfin, nous analysons la conception épistémique de la mesure qui accompagne cette solution bayésienne; mais nous montrons également qu'une extension de l'approche fréquentiste, proposée afin de pallier les limitations de l'approche fréquentiste classique, dessine une autre manière de répondre aux exigences actuelles de la métrologie en mettant l'accent sur la dimension, non plus épistémique, mais sociale des activités de mesure.

Abstract

The statistical treatment of experimental data is undergoing a profound methodological transformation. The traditional resort to frequentist methods in the description of measurement processes is today gradually overthrown by the call on Bayesian ones. This evolution stirs heated debates which have received much impetus from the publication, in 1993, of the *Guide to the expression of uncertainty in measurement* (GUM) whose aim was, originally, to supply, at an international level, a uniform handling of measurement. We first show why the evaluation of a “measurement uncertainty” is indispensable to the communication of results that can be used by everyone. We then turn to examine how the new metrological requirements imposed by today's scientific and economical context have stumbled against the incapacity of the frequentist approach to provide a complete probabilistic treatment of measurement uncertainty, and how they have facilitated the promotion of a fully Bayesian approach which is able to deal with this problem. We finally analyse the epistemic conception of measurement which goes along with this Bayesian solution; but we also show that an extension of the frequentist approach, which was put forward in order to overcome the limitations of the classical frequentist approach, sketches another way to respond to the current demands of metrology by stressing the social, rather than the epistemic dimension of measurement activities.

La façon dont les scientifiques conçoivent la mesure physique a subi de profondes transformations au cours de ces dernières décennies. Ces transformations touchent tout particulièrement la compréhension des méthodes statistiques utilisées dans l'analyse et le traitement des données expérimentales, ainsi que la compréhension des modèles probabilistes qui les sous-tendent. Après avoir largement dominé la science de la mesure, les méthodes statistiques fondées sur une interprétation fréquentiste des probabilités se trouvent en effet aujourd'hui fortement battues en brèche par les statistiques bayésiennes, fondées, quant à elles, sur une interprétation épistémique des probabilités. Dans ce qui suit, nous nous proposons d'examiner comment pareil tournant a pu en venir à se produire dans l'un des plus sûrs bastions de l'approche fréquentiste ; nous analysons ensuite les prolongements philosophiques de ce tournant.

La confrontation entre statistiques fréquentistes et bayésiennes s'inscrit dans une longue tradition de débats philosophiques¹. Elle se prolonge aujourd'hui avec, d'un côté, des ouvrages comme ceux de Howson et Urbach² et de Jaynes³, qui développent les principes et les modalités de l'application des méthodes bayésiennes dans la pratique scientifique ; et, de l'autre, ceux de Mayo⁴ et Spanos qui contestent la légitimité de ces dernières, et prônent une approche fondée sur l'interprétation fréquentiste des probabilités qu'ils nomment « erreur-statistique (error-statistical) »⁵. Nous souhaitons contribuer à cet échange de points de vue en adoptant un angle d'approche original. Nous nous proposons en effet d'explorer comment la question de l'interprétation des probabilités est apparue au sein même de la pratique de scientifiques quotidiennement engagés dans l'analyse des données de mesure. Le parti-pris adopté consiste à fonder notre étude sur les documents qu'ont produits ces scientifiques, et à les suivre de près, plutôt que de situer notre propos d'entrée de jeu sur le terrain philosophique. Notre matériau est ainsi constitué d'articles scientifiques, issus pour la plupart du domaine de la science de la mesure, ou « métrologie », qui est une discipline scientifique normative, fortement axée sur la méthodologie scientifique, dont l'une des tâches principales est d'assurer la comparabilité, la cohérence et la fiabilité des résultats de mesure obtenus par les différents acteurs scientifiques et économiques. Cette tâche requiert un travail tout à la fois scientifique, expérimental, technique et institutionnel dans lequel on pourrait voir la mise en œuvre d'une « technologie de l'intersubjectivité », en détournant quelque peu une expression de Hacking⁶. La métrologie a reçu beaucoup d'attention de la part des sociologues et des historiens, mais fort peu de la part des philosophes. Cette situation perdure malgré l'essor, au début des années 1980, d'une philosophie de l'expérimentation, avec, notamment, les travaux de Hacking⁷ et de Galison⁸,

1. Voir, par exemple, P.S. Bandyopadhyay & M.R. Forster (eds), *Philosophy of Statistics, Handbook of Philosophy of Science Volume 7 Philosophy of Statistics*, Elsevier, 2011.

2. C. Howson & P. Urbach, *Scientific Reasoning : The Bayesian Approach*, Open Court, 2006.

3. E.T. Jaynes, *Probability Theory : The Logic of Science*, Cambridge University Press, 2003.

4. D.G. Mayo, *Error and the Growth of Experimental Knowledge*, University of Chicago Press, 1996.

5. D.G. Mayo & A. Spanos, « Error Statistics », in Bandyopadhyay & Forster (eds), *Philosophy of Statistics, Handbook of Philosophy of Science Volume 7 Philosophy of Statistics, op. cit.*

6. I. Hacking, « Statistical Language, Statistical Truth and Statistical Reason : The Self-Authentication of a Style of Scientific Reasoning », in E. McMullin (eds.), *The Social Dimension of Science*, University of Notre Dame Press, 1992, p.152.

7. I. Hacking, *Representing and Intervening*, Cambridge University Press, 1983.

8. Galison, *Ainsi s'achèvent les expériences* [1987], traduit de l'anglais (américain) par B. Nicquevert, La Découverte, 2002.

et malgré l'accent que les écrits de Hon⁹ ont mis, par ailleurs, sur la fonction épistémologique que pouvait remplir l'étude de l'erreur dans les sciences. Notre investigation s'inscrit dans la lignée de ces travaux auxquels elle apporte une dimension supplémentaire en s'intéressant à des activités scientifiques de nature réflexive (méthodologiques et normatives).

Poussée par les avancées scientifiques et techniques du XX^e siècle, mais aussi par les exigences économiques de la mondialisation, la métrologie s'est engagée dans un vaste programme de réforme qui comprend, notamment, un profond remaniement du Système International d'unités (SI). Parmi les productions qui accompagnent ce mouvement de réforme, deux guides internationaux destinés à harmoniser les pratiques de mesure et la formulation des résultats de mesure concernent tout particulièrement notre propos ; il s'agit du *Vocabulaire International de Métrologie* (VIM), publié pour la première fois en 1984¹⁰, et du *Guide pour l'expression de l'incertitude de mesure* (GUM), dont la première et, à ce jour encore, unique édition date de 1993¹¹. La publication de ces deux documents a servi de caisse de résonance à des débats extrêmement nourris dans les revues scientifiques consacrées à la mesure, débats qui ont conduit au développement d'une véritable « philosophie appliquée » stimulée par les besoins de la pratique. On y trouve à la fois des analyses conceptuelles portant sur certaines des notions les plus fondamentales de l'activité scientifique (notamment les notions de grandeur et de valeur d'une grandeur, de modèle, de constante fondamentale), et des considérations qui relèvent de la philosophie des statistiques et des probabilités. Comme nous le verrons, le GUM est au nœud du retournement de situation que nous avons évoqué en ouvrant cet article. Les débats qu'il a aiguïsés sur les statistiques et les probabilités impliquées dans l'évaluation des « incertitudes de mesure » rallient à la perspective bayésienne un nombre croissant de membres de la communauté métrologique qui paraissait jusqu'alors foncièrement attachée à l'interprétation fréquentiste des probabilités. Notre analyse s'appuiera donc fortement sur le VIM, le GUM ainsi que sur les débats qui ont entourés ces deux publications.

Notre enquête ne visera pas à départager les approches fréquentiste et bayésienne. Elle

9. G. Hon, « Towards a typology of experimental errors : an epistemological view », *Studies in History and Philosophy of Science*, 20(4), 1989, 469-504; G. Hon, J. Schickore & F. Steinle (eds.), *Going Amiss in Experimental Research*, Springer, 2009.

10. Joint Committee for Guides in Metrology (JCGM), *Vocabulaire international de métrologie – Concepts fondamentaux et généraux et termes associés* (VIM), 2012.

http://www.bipm.org/utils/common/documents/jcgm/JCGM_200_2012.pdf

Le VIM a connu trois éditions différentes, en 1984, 1993 puis 2008, qui diffèrent substantiellement de l'une à l'autre. En 2012, le document a connu une révision mineure, portant sur des erreurs de forme : la version de 2012 demeure donc la troisième édition du VIM, couramment appelée « VIM3 ».

11. Joint Committee for Guides in Metrology (JCGM), *Évaluation des données de mesure – Guide pour l'expression de l'incertitude de mesure*, 2008.

http://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_F.pdf

Le GUM a d'abord été publié par l'ISO (Organisation Internationale de Normalisation). La première version, datant de 1993, a été très légèrement remaniée en 1995. En 1997, l'ISO a transféré au BIPM (Bureau International des Poids et Mesures) la charge du maintien et de la mise à jour du document. Le BIPM a publié en 2008 une version librement accessible, ne contenant que des modifications mineures par rapport à la version de 1995. Dans cet article, nous nous référons exclusivement au document de 2008. Le VIM et le GUM sont publiés simultanément en français et en anglais. Nous citerons dans la suite les versions françaises de ces documents.

s'efforcera plutôt d'analyser comment des questions philosophiques peuvent émerger et se déployer à partir de la pratique scientifique même. Il est en effet frappant de constater qu'au travers des débats suscités par le VIM et le GUM, des questions épistémologiques, voire métaphysiques, en sont venues à occuper une place non négligeable dans la littérature métrologique. De fait, la révision des critères d'acceptabilité de nos énoncés pratiques et théoriques, dans laquelle la métrologie s'est engagée afin de répondre aux nouvelles exigences scientifiques et économiques du XX^e siècle, semble appeler la remise en question de thématiques épistémologiques centrales. En liaison étroite avec les problèmes théoriques et pratiques rencontrés, les métrologues en sont venus à faire émerger, dans les pages des journaux scientifiques consacrés à la mesure, bien d'autres questions épistémologiques que celle qui consiste à s'interroger sur la façon dont il convient de trancher le débat entre fréquentistes et bayésiens. Nous nous attachons ici à cette philosophie développée par les métrologues et discutons les problèmes qu'elle soulève en nous efforçant de prendre acte de l'horizon pratique dans lequel elle se déploie. Nous serons ainsi amenés à inscrire, nous aussi, notre approche philosophique dans une perspective pratique, articulée aux débats que nous analysons.

Notre étude comporte trois moments. Dans la section 1, nous présentons les principes méthodologiques et épistémologiques qui sous-tendent l'analyse et la modélisation d'un processus de mesure. Nous montrons ainsi comment émergent les concepts d'« erreur de mesure », puis d'« incertitude de mesure » et, dans leur sillage, la question de savoir comment il convient de calculer l'incertitude associée à un résultat de mesure. Dans la section 2, nous montrons comment la méthode de calcul des incertitudes préconisée par le GUM a pu donner prise à la critique et conduire à promouvoir une approche entièrement bayésienne du calcul des incertitudes de mesure. Enfin, dans la section 3, nous présentons et discutons certains des prolongements proprement philosophiques qui émergent des débats présentés dans la section précédente.

1 Erreurs de mesure et incertitude de mesure

1.1 Le GUM et la mission d'uniformisation du calcul des incertitudes

Rédigé à l'initiative du Bureau International des Poids et Mesures (BIPM), le *Guide pour l'expression de l'incertitude de mesure* (GUM) se propose d'uniformiser les méthodes d'évaluation des résultats de mesure et de l'incertitude qui leur est associée en énonçant un ensemble de règles générales établies sur des bases probabilistes solides. Le projet du GUM se situe par là dans le sillage des travaux de Gauss et de Laplace sur la théorie des erreurs, qui, au début du XIX^e siècle¹², avaient commencé à mettre de l'ordre dans les différentes pratiques de traitement des mesures en usage, jusqu'alors dépourvues d'assise théorique.

Cependant, en cherchant à fournir aux scientifiques et aux industriels les principes de calculs clairs et aisément applicables qu'appelaient les développements techniques et économiques

12. S.M. Stigler, *The History of statistics. The measurement of Uncertainty before 1900*, Harvard University Press, 1986; M. Armatte, « La théorie des erreurs (1750-1820), enjeux, problématiques, résultats », in E. Barbin & J.-P. Lamarche, *Histoires de probabilités et de statistiques*, Paris, Ellipses, 2004, 141-160.

de la seconde moitié du XX^e siècle¹³, le GUM semble avoir mis à nu une difficulté centrale du traitement des erreurs de mesure dans les sciences de la nature. De fait, bien loin de rencontrer le consensus espéré, les recommandations du *Guide* ont au contraire suscité de nombreuses réactions critiques lors de leur publication en 1993. Celles-ci ont nourri, chez les métrologues, un mouvement de réflexion qui n'a cessé d'alimenter jusqu'à ce jour les pages des journaux scientifiques consacrés à la mesure. Parmi les critiques les plus discutées figure celle selon laquelle la méthode recommandée par le GUM fait intervenir des procédés qui relèvent tantôt d'une conception fréquentiste, tantôt d'une conception épistémique des probabilités, et juxtapose ainsi des points de vue dont le caractère réputé inconciliable jette le soupçon sur la validité de la méthode préconisée.

Les problèmes touchant à l'interprétation des probabilités ne sont guère discutés dans le GUM dont le propos se veut avant tout pratique. Ils donnent cependant lieu à quelques courtes remarques, reléguées dans l'une des annexes du document, où perce la tension dénoncée lors de la réception des recommandations :

(P)our certains traitements traditionnels de l'incertitude de mesure [...] le concept de probabilité est considéré comme s'appliquant *seulement* aux événements qui peuvent être répétés un grand nombre de fois, essentiellement dans les mêmes conditions, la probabilité p d'un événement ($0 \leq p \leq 1$) indiquant la *fréquence* avec laquelle se produit l'événement.

Par contraste avec ce point de vue de la probabilité fondée sur la fréquence, un autre point de vue aussi valable est que la probabilité est une mesure du *degré de croyance* en ce qu'un événement se produise. [...]

La recommandation INC-1 (1980) sur laquelle est fondé ce *Guide* adopte implicitement une telle approche de la probabilité [...]¹⁴

Cette brève incursion du GUM sur le terrain des fondements des probabilités se veut conciliante puisque les conceptions fréquentiste et épistémique sont présentées comme « aussi valable(s) » l'une que l'autre. La préférence est certes donnée au point de vue épistémique mais, en notant son caractère implicite, les auteurs minimisent l'importance de leur choix tout en suggérant que celui-ci pourrait être véhiculé par certains aspects techniques du calcul.

L'insuffisance de ces quelques lignes est patente ; non seulement parce que l'adoption d'une interprétation réputée « subjective » des probabilités, si longtemps exclue du champ des sciences de la nature et revendiquée, de surcroît, dans un document à vocation normative consacré au traitement des mesures – pierres de touche du processus de justification dans les sciences empiriques – mériterait davantage d'explication ; mais aussi parce que ce court passage suscite en effet une certaine confusion en laissant entendre que l'approche fréquentiste « objective » continue d'être utilisée dans certains « traitements traditionnels ».

L'analyse du document confirme que la méthode de calcul des incertitudes recommandée par le GUM conduit bel et bien à juxtaposer les deux approches. Des métrologues attentifs aux

13. JCGM, *Guide pour l'expression de l'incertitude de mesure*, op. cit., p.vi.

14. JCGM, *Guide pour l'expression de l'incertitude de mesure*, op. cit., p.59.

questions soulevées par l'interprétation des probabilités se sont rapidement engouffrées dans cette brèche conceptuelle. D'autres ont manifesté l'inquiétude qu'une approche mixte pourrait déboucher sur des incohérences techniques sérieuses. C'est ainsi que des préoccupations au départ méthodologiques, liées à des objectifs en grande partie pratiques, visant à améliorer la comparabilité et la fiabilité des résultats de mesure, ont pu entraîner certains métrologues à aborder des questions relevant de la philosophie des probabilités et, à travers elles, à des questions de philosophie des sciences.

Afin de montrer comment des problèmes touchant à l'interprétation des probabilités peuvent venir s'immiscer au cœur du traitement scientifique des données fournies par l'expérience, et occasionner le mélange des approches fréquentiste et bayésienne, il faut prendre en compte la complexité des élaborations conceptuelles et techniques qui entrent dans la production d'un résultat de mesure communicable et partageable par tous.

1.2 Variabilité des données issues de l'expérience et hypothèse des erreurs de mesure

L'énoncé d'un résultat de mesure se heurte au constat de la variabilité des indications numériques obtenues lorsqu'on effectue plusieurs mesurages¹⁵ de la même quantité¹⁶. Certes, cette variabilité ne se manifeste pas avec n'importe quel dispositif de mesure. Elle ne devient notoire qu'à condition d'utiliser des instruments ayant une sensibilité et une résolution suffisantes. Son étude ne s'est d'ailleurs imposée qu'avec l'amélioration des instruments au XVIII^e siècle et dans les domaines, longtemps limités à l'astronomie, où l'usage des résultats l'exigeait. Sa prise en compte est aujourd'hui devenue incontournable, aussi bien dans la recherche scientifique que dans l'industrie, pour assurer la fiabilité de l'infrastructure technologique sur laquelle reposent nos sociétés industrielles mondialisées, parcourues de réseaux de production, de télécommunication, de distribution et de services¹⁷.

15. Dans le vocabulaire de la métrologie, le terme spécifique de « mesurage » est employé pour désigner l'action de mesurer, à laquelle ne se limite pas le terme plus général de « mesure ». Voir JCGM, *Vocabulaire International de Métrologie, op. cit.*, p.xiv.

16. Les concepts de quantité et de grandeur sont souvent confondus dans l'usage courant comme dans l'usage scientifique. L'Encyclopédie de Diderot et D'Alembert les tient pour équivalents (voir, notamment, les articles « Grandeur » et « Quantité »). La distinction n'est pas davantage établie aujourd'hui dans le VIM où la confusion est largement entretenue par le fait que la version française utilise le terme de « grandeur » là où la version anglaise utilise celui de « quantity ». Nous ne pouvons exposer ici les raisons pour lesquelles nous pensons qu'il convient de distinguer ces deux concepts et, dans la suite, nous nous efforcerons de suivre (ce qui est souvent difficile à cause du pli qu'impose parfois l'usage) la suggestion de Russell d'utiliser le terme de « quantité » pour les quantités concrètes, engagées dans un système physique, objets d'un mesurage, et celui de « grandeur » pour les quantités abstraites, qui interviennent dans l'expression mathématique des relations physiques. B. Russell, *The Principles of Mathematics*, W. W. Norton & Company, Inc. Publishers, 1938 (1903), §151, p. 159. La distinction est reprise par L. Couturat, *Les Principes mathématiques*, Félix Alcan, 1905, p. 99. Pour une argumentation en faveur de cette distinction, voir N. de Courtenay, « The double interpretation of the equations of physics and the quest for common meanings », in O. Schlaudt & L. Huber (eds.) *Standardization in Measurement. Philosophical and Sociological Issues*, Pickering and Chatto Publisher, 2015, 53-68.

17. Sur le rôle majeur que joue la métrologie dans le monde technique d'aujourd'hui, voir T. Quinn, « Metrology, its role in today's world », in I. Lira, *Evaluating the Measurement Uncertainty. Fundamental and Practical Guidance*, Institute of Physics Publishing, 2002, 1-24.

La variabilité des indications numériques se manifeste de deux manières fort différentes. Elle se manifeste, d'une part, lorsqu'on effectue des mesurages répétés du même système physique dans des conditions que l'on juge identiques (conditions dites « de répétabilité »); on constate alors une dispersion des indications obtenues, c'est-à-dire qu'il y a non-unicité de la valeur mesurée. Elle se manifeste, d'autre part, lorsqu'on cherche à déterminer une même quantité par des expériences qui diffèrent par leur environnement, les instruments ou les principes de mesure mis en œuvre (conditions dites « de reproductibilité »). Dans ce second cas, lorsque l'incompatibilité des résultats obtenus dans les différentes expériences ne peut être raisonnablement attribuée à la dispersion constatée dans les conditions de répétabilité de chacune des situations expérimentales considérées, on est en présence d'un autre type de variabilité. Contrairement au premier, qui peut être repéré au sein d'un même dispositif expérimental soumis aux conditions de répétabilité, ce second type de variabilité échappe aux observations répétées au sein d'un même dispositif et suppose, pour pouvoir être dépisté, la confrontation des indications fournies par plusieurs dispositifs expérimentaux différents.

Comme l'ont souligné Campbell, puis Kyburg, l'idée selon laquelle la variabilité des données numériques signale l'existence d'erreurs de mesure n'est dictée ni par l'expérience ni par des « raisons *a priori* irréfutables »¹⁸. Elle émane de la confiance que nous accordons à des hypothèses ou des théories concernant le comportement des objets physiques que le développement scientifique a progressivement diversifiées et sophistiquées. Ainsi, « notre théorie vague et générale des objets physiques »¹⁹ nous porte-t-elle à penser que les propriétés des objets ne se prennent pas à changer sans raison et que, par conséquent, la quantité soumise à la mesure ne varie pas au cours de mesurages réalisés dans des conditions de répétabilité. On dira alors que si les données effectivement recueillies présentent des variations, c'est qu'elles s'écartent de la valeur « vraie » unique correspondant à la grandeur visée²⁰, et l'on *expliquera* ces écarts par des *erreurs de mesure*²¹. L'introduction de la notion d'erreur de mesure procède donc d'une tentative d'explication de l'expérience visant à rendre le constat de la variabilité des

18. H.E. Kyburg Jr., « Measuring Errors of Measurement », in C.W. Savage & Ph. Ehrlich (eds.), *Philosophical and Foundational Issues in Measurement Theory*, Lawrence Erlbaum, 1992, 75-91, p.77.

19. *Ibid.*

20. Nous verrons plus loin que la définition de la grandeur visée est toujours, nécessairement entachée d'un certain vague (il est en effet impossible d'en donner une description complète). De ce fait, la grandeur visée possède non pas *une* mais *plusieurs valeurs* (en fait une infinité), toutes aussi « vraies » les unes que les autres. Nombre de métrologues voient là un argument de taille (il y en a d'autres) pour renoncer tout à la fois à l'unicité de la valeur correspondant à la grandeur visée et à la notion de valeur « vraie ». Nous laisserons, quant à nous, ces questionnements de côté et chercherons surtout, ici, à clarifier les étapes qui interviennent dans l'établissement d'un résultat de mesure. Notre analyse se limitera pour cela à la démarche que l'on rencontre dans les exposés de métrologie classique, souvent largement méconnus des philosophes des sciences. Nous aurons cependant l'occasion de revenir sur les concepts de valeur vraie et d'erreur de mesure lors de la discussion de l'approche bayésienne (en particulier dans la section 3.1). Pour une discussion plus approfondie des problèmes philosophiques soulevés par la notion de valeur vraie, voir F. Grégis, « Can we dispense with the notion of true value in metrology? », in O. Schlaudt & L. Huber (eds.), *Standardisation in Measurement. Philosophical and Sociological Issues*, Pickering and Chatto Publishers, 2015, 81-93.

21. N.R. Campbell, *Physics, the Elements*, Cambridge University Press, 1920; voir, en particulier, le chapitre XIV, p. 444; Kyburg Jr., « Measuring Errors of Measurement », *op. cit.*.

mesures compatible avec les attentes dérivées des hypothèses ou avec les théories auxquelles nous accordons foi et qui nous incitent à postuler une valeur visée unique. À ce titre, l'erreur de mesure se présente comme un *concept théorique* suggéré par nos théories qui jouent un *rôle normatif* en orientant l'analyse de l'expérience de mesure²². Le caractère théorique de l'erreur de mesure apparaît encore plus saillant dans le traitement de l'incompatibilité des données issues de mesurages effectués dans des conditions de reproductibilité. On rapporte en effet ces incompatibilités à des erreurs de mesure en faisant l'hypothèse selon laquelle les quantités se conservent ou peuvent être reproduites en différents lieu et temps, et qu'elles peuvent être déterminées par des lois, et donc des principes de mesure différents²³.

Que l'on se trouve dans des conditions de répétabilité ou de reproductibilité, l'erreur de mesure e est, dans les deux cas, définie comme l'écart entre l'indication numérique x fournie par le mesurage et la valeur numérique « vraie » v de la grandeur visée, posée comme étant la valeur inconnue visée par l'activité de mesure :

$$e = x - v \tag{1}$$

1.3 Deux composantes de l'erreur de mesure

Nous avons identifié deux types bien différents de variabilité des données de l'expérience. Les erreurs de mesure invoquées pour rendre compte de l'expérience sont considérées différemment selon qu'on les rapporte à l'un ou l'autre type de variabilité, c'est-à-dire qu'on les infère à partir de mesurages effectués dans des conditions de répétabilité ou de mesurages effectués dans des conditions de reproductibilité. En effet, elles ne semblent pas pouvoir relever du même type de traitement. Cette différence est à la racine des incohérences diagnostiquées dans le GUM.

Afin de rendre plus concret notre propos, nous développons ci-après un exemple – l'exemple (E) illustré figure 1 – que nous reprendrons à différentes reprises au cours de cet article²⁴. On mesure, au moyen d'un instrument comparateur, la longueur L d'un « calibre à bouts » par comparaison avec un étalon de longueur, un autre calibre à bouts dont la longueur L_E a été déterminée par un laboratoire d'étalonnage et inscrite dans le « certificat d'étalonnage » fourni avec l'étalon. Le comparateur indique la différence D entre les longueurs des deux calibres, ce qui permet, la longueur L_E étant connue, de déterminer la longueur L (figure 1).

22. On peut montrer que les concepts de grandeur et de valeur vraie sont également des concepts théoriques dont l'introduction est inséparable de celle du concept d'erreur de mesure. Le geste théorique que constitue l'introduction de l'erreur de mesure pour rendre compte de l'expérience joue ainsi un rôle crucial dans l'analyse de la mathématisation des sciences expérimentales. Voir, à ce sujet, Nadine de Courtenay, « Mesure et formation des concepts physiques. Rudolf Carnap et Norman Campbell », in J. Bouveresse & P. Wagner (dir.), *Mathématiques et expérience. L'empirisme logique à l'épreuve (1918-1940)*, 2008, 211-251.

23. Contrairement à l'opérationnalisme, selon lequel des principes de mesure différents définissent des concepts physiques distincts.

24. Nous nous inspirons d'un exemple présenté dans l'annexe H du GUM et que nous simplifions pour les besoins de l'exposition. JCGM, *Guide pour l'expression de l'incertitude de mesure, op. cit.*, pp. 82-88.

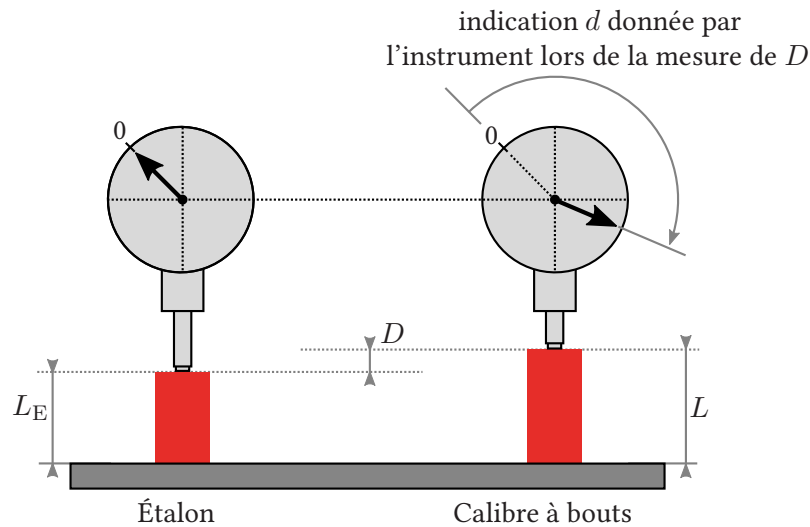


FIGURE 1 – *Exemple (E)* : mesure de la longueur d'un calibre à bouts par comparaison avec un autre calibre à bouts servant d'étalon, dont la longueur a été mesurée par un laboratoire d'étalonnage. La mesure s'effectue en deux temps. (i) Dans un premier temps, le comparateur est calé sur l'étalon, et on règle alors l'indication qu'il donne sur le zéro. (ii) Le comparateur est ensuite calé sur le calibre. Il indique alors par d la différence de longueur qu'il mesure entre le calibre et l'étalon.

La dispersion des données résultant d'une suite de mesurages effectués dans les mêmes conditions est constatable au sein même de l'expérience de mesurage. Des mesurages répétés effectués en positionnant le calibre dans le comparateur (dont le zéro a été préalablement réglé avec le calibre étalon) fourniront en effet des indications numériques différentes d_1, d_2, \dots, d_n , plus ou moins dispersées (selon la résolution du comparateur et le détail du protocole de mesure). En posant que les mesurages visent une grandeur fixe – la différence de longueur D entre les deux calibres²⁵ –, la dispersion est interprétée comme la manifestation d'erreurs de mesure dues à une multiplicité de causes inconnues et variables susceptibles d'affecter de façon imprévisible et incontrôlable le fonctionnement des instruments, l'environnement (variation aléatoire spatiale ou temporelle des grandeurs d'influence, ici la température), ou l'opérateur lui-même (si, par exemple, le relevé de l'indication de mesure nécessite une lecture qui pourrait être biaisée par un effet de parallaxe).

Depuis Gauss et Laplace, on a cherché à réduire cette dispersion en introduisant un traitement statistique fondé sur l'hypothèse supplémentaire selon laquelle les erreurs occasionnées par le processus de mesure sont *aléatoires* – on les appelle, de ce fait, erreurs « aléatoires ». Comme nous le montrerons dans la section 2.1, la réduction statistique des erreurs aléatoires n'est cependant jamais totale – elle est d'autant meilleure que le nombre de données expérimentales recueillies est grand – et se solde donc par une *incertitude* sur la valeur corrigée.

25. Cette caractérisation n'est pas suffisante; nous y reviendrons. Cette question touche au problème d'identification de la grandeur mentionné précédemment.

Si la variabilité des résultats obtenus dans les conditions de répétabilité suggère d'introduire l'hypothèse d'erreurs aléatoires, un mouvement supplémentaire porte à expliquer la variabilité qui se manifeste lorsqu'on confronte des indications issues de mesurages effectués dans des conditions de reproductibilité en faisant l'hypothèse d'erreurs qui ne changent qu'avec les conditions expérimentales : ces dernières sont donc constantes dans les conditions de répétabilité, et ne se signalent par aucun trait observable au sein d'un dispositif expérimental donné. Ces erreurs constantes sont imputées à l'intervention de causes perturbatrices qui demeurent fixes au cours de mesurages effectués dans des conditions de répétabilité et qui affectent toujours de la même manière les indications délivrées par un dispositif donné – on les appelle de ce fait erreurs « systématiques ». Elles introduisent un écart constant entre chaque donnée numérique fournie par l'expérience et la valeur cible de la quantité à mesurer, c'est-à-dire la valeur que l'on obtiendrait si l'on pouvait éliminer tous les phénomènes d'influence qui affectent l'issue de chaque mesurage ou en corriger numériquement les conséquences. Ces erreurs constantes expliquent ainsi l'écart entre les estimations calculées dans le cadre d'expériences différentes (entre lesquelles les conditions d'expérience varient), cet écart étant rapporté à des causes perturbatrices constantes dans un contexte expérimental donné, mais différant d'un contexte expérimental à l'autre.

Les erreurs systématiques ne se trahissant par aucun indice observable dans les données numériques fournies par l'expérience, leur correction, tout comme leur détection, doit faire appel à des *informations extérieures* à celles qu'un dispositif expérimental donné est susceptible de fournir dans des conditions fixées²⁶. De plus, parce qu'elles sont constantes, leur traitement ne peut faire appel aux méthodes statistiques utilisées dans le cas des erreurs aléatoires. Les causes constantes qui affectent les données peuvent avoir trois origines principales. Certaines causes ont pour origine (i) des effets physiques produits sur la propriété étudiée par les autres propriétés du système, par l'environnement ou par les instruments de mesure ; dans ce cas, la grandeur *effectivement mesurée* par le dispositif n'est pas la grandeur que l'on a l'intention de mesurer, appelée « mesurande » par les métrologues²⁷. La valeur du mesurande doit être inférée par une correction des facteurs d'influence. Cette correction, au travers de laquelle s'effectue l'identification du mesurande, mobilise des connaissances théoriques (c'est-à-dire la connaissance des

26. Les corrections faisant appel à des lois physiques, comme celles qui régissent la dilatation des calibres en fonction de la température, font bien intervenir des informations extérieures en ce sens.

27. Le mesurande est défini par le VIM comme la « grandeur que l'on veut mesurer », JCGM, *Vocabulaire International de Métrologie*, op. cit., 2.3, p.17. On remarquera que l'identification du mesurande supposerait, en toute rigueur, de fixer l'état de *tous les paramètres* dont la quantité à mesurer peut dépendre (voir JCGM, *Guide pour l'expression de l'incertitude de mesure*, op. cit., annexe D, pp. 51-52). Nous verrons plus bas que cela suppose, par exemple, d'indiquer à quelle température on souhaite mesurer la longueur du calibre de l'exemple (E) car, sans cette information, la quantité que l'on souhaite mesurer n'est pas bien définie : les matériaux se dilatent avec la température, la longueur du calibre peut donc prendre une infinité de valeurs si la température à laquelle on souhaite la mesurer n'est pas spécifiée – et ne plus même avoir de sens du tout au-delà de la température de fusion du matériau dans lequel est fabriqué le calibre. Mais il faudrait aussi indiquer, semble-t-il, la position du calibre dans le champ de pesanteur, les coordonnées du lieu où s'effectue le mesurage etc. On voit que la spécification de ces paramètres est inévitablement incomplète. On saisit mieux ici pourquoi le mesurande et, avec lui, les conditions de répétabilité ne peuvent être parfaitement définies. Cela ne pose cependant pas de difficulté si l'on se place du point de vue pratique, puisque l'utilisation que l'on projette de faire du résultat de mesure détermine jusqu'à quel niveau de détail il convient de poursuivre la description.

lois qui gouvernent les phénomènes et le comportement des instruments²⁸) ainsi qu'une expertise technique (lorsque les facteurs impliqués sont insuffisamment connus²⁹). Dans notre exemple, il faudra ainsi prendre en compte l'effet de la température sur la longueur du calibre et de l'étalon : si le mesurande est défini par la longueur L de la cale à $T_0 = 20^\circ\text{C}$ et que le mesurage s'effectue à $T = 23^\circ\text{C}$, il faudra tenir compte de la dilatation du calibre (ainsi que de l'étalon si la valeur fournie de celui-ci est celle de sa longueur à $T_0 = 20^\circ\text{C}$). D'autres causes constantes ont pour origine (ii) les défauts des instruments dont l'évaluation ne peut guère être faite à partir des résultats qu'ils fournissent dans le contexte de l'expérience en cours ; l'expérimentateur doit pour cela se référer aux spécifications des constructeurs (ici, les certificats d'étalonnage du comparateur et du calibre étalon). Enfin, l'on recense également (iii) les erreurs introduites par l'utilisation de résultats expérimentaux provenant d'autres expériences ou par l'utilisation de coefficients physiques fournis par les tables de constantes (par exemple, les coefficients de dilatation respectifs du calibre et de l'étalon, α et α_E) dont les valeurs ne sont, elles non plus, pas exactes.

Les erreurs systématiques, pas plus que les erreurs aléatoires, ne peuvent être complètement corrigées, mais pour des raisons différentes. D'une part, parce que le calcul des corrections fait intervenir des données (notamment des évaluations fondées sur l'expertise) ou des résultats numériques antérieurs à propos desquels subsiste également une incertitude ; ou encore parce que, pour des raisons pratiques, le calcul doit s'appuyer sur des expressions approchées des lois qui régissent les phénomènes responsables d'erreurs systématiques. D'autre part, parce que certaines causes d'erreur sont inconnues, les phénomènes qui en sont responsables n'ayant pas encore été identifiés. Le traitement des erreurs systématiques est par là tributaire de la connaissance que nous avons du processus et des conditions de mesure, de l'environnement, des instruments et des lois qui gouvernent leur comportement. Il est, de ce fait, impossible d'apprécier l'exactitude de la valeur obtenue après correction, c'est-à-dire l'étroitesse de l'accord entre la valeur corrigée et la valeur cible que l'on cherche à déterminer³⁰. L'exactitude est en effet suspendue à la confiance que l'on accorde à l'analyse du processus de mesure effectuée et, plus profondément, aux théories et aux modèles dont on dispose pour traiter les phénomènes impliqués. Il y a là une différence majeure avec les erreurs aléatoires puisque l'expérience de mesure donne *par elle-même* le moyen d'évaluer la *fidélité* de la mesure, c'est-à-dire l'« étroitesse de l'accord entre les indications ou les valeurs mesurées obtenues par des mesurages répétés du même objet [...] »³¹ : il suffit pour cela d'apprécier la dispersion des indications obtenues par répétition des mesurages. On remarquera que la fidélité d'une mesure (sa « précision », selon la terminologie ambiguë du langage courant que cherche à réformer le VIM) ne présume en rien de son exactitude.

28. Notons que l'on utilise souvent des expressions mathématiques approchées des lois décrivant l'effet des facteurs d'influence qui contribuent à l'erreur commise.

29. On peut aussi neutraliser certains facteurs d'influence par une conception astucieuse de l'expérience.

30. JCGM, *Vocabulaire International de Métrologie*, op. cit., 2.13, p.21.

31. JCGM, *Vocabulaire International de Métrologie*, op. cit., p. 22.

1.4 De l'identification des erreurs de mesure au calcul de l'incertitude de mesure

Il subsiste donc un doute irréductible sur la valeur obtenue après réduction des erreurs aléatoires et correction des erreurs systématiques : la valeur cible du mesurande et l'erreur commise restent inconnues et inaccessibles. Il est donc indispensable d'inclure dans le résultat fourni à l'issue du traitement une mention de l'incertitude *de mesure*, qui exprime quantitativement ce doute – mais dont la signification précise reste à déterminer. Laissant de côté la question de savoir si ce constat, en apparence négatif, justifie l'élimination de la notion de valeur vraie, comme le pensent certains métrologues, nous nous concentrerons sur l'apport des analyses précédentes. Celles-ci indiquent en effet comment il est possible de dégager des indications brutes fournies par l'expérience un *résultat* de mesure, c'est-à-dire une conclusion qui dépasse le cadre de l'expérience individuelle, possédant une signification consensuelle, publique qui la rend communicable et comparable à d'autres résultats, et de ce fait partageable, critiquable et utilisable par tous.

Les analyses présentées ci-dessus montrent, en premier lieu, comment tirer de la pluralité des indications issues de mesurages répétés un résultat de mesure exprimable sous la forme $Q = \{Q\} [Q]$, où « Q » désigne le mesurande (la longueur L du calibre à 20°C, dans notre exemple), et l'expression $\{Q\} [Q]$ désigne la *valeur du mesurande*, au sein de laquelle « $[Q]$ » est le nom de la grandeur de même nature que Q prise pour unité (le mètre, par exemple, désigné par « m ») et « $\{Q\}$ » la *valeur numérique* de Q relativement à l'unité choisie, indiquant le nombre de fois que l'unité $[Q]$ est contenue dans Q ³². Elles montrent également que si la comparaison du résultat ainsi obtenu avec d'autres résultats (produits dans des conditions de reproductibilité) *peut avoir un sens* c'est que ce résultat a fait l'objet de corrections. Le processus de correction participe en effet à l'identification de la grandeur visée à partir des indications brutes et constitue ainsi le terme de la comparaison.

Les analyses précédentes montrent, en second lieu, qu'un résultat de mesure exprimé sous la forme $Q = \{Q\} [Q]$ est inutilisable parce qu'incomplet. Pour que la comparaison avec d'autres résultats ait non seulement un sens mais soit aussi *possible* et ne se réduise pas à un constat de désaccord, auquel conduit inévitablement la confrontation de résultats ponctuels, le résultat doit être accompagné d'une *évaluation quantitative* de l'incertitude qui affecte sa détermination et tient compte des conditions particulières du mesurage. C'est seulement accompagné de l'évaluation de son incertitude qu'un résultat de mesure devient *effectivement comparable* à un autre résultat, à une prédiction théorique ou à une spécification technique (lorsqu'il s'agit, par exemple, de fabriquer les pièces d'un appareil destinées à être assemblées dans une autre chaîne de construction). On remarquera que, dans ces conditions, l'accord ou

32. Cette notation a été instituée par Maxwell, mais en raison de sa lourdeur, nous lui substituerons une notation plus conforme à celle qui est habituellement adoptée en physique aujourd'hui. Elle consiste à désigner les quantités et les grandeurs par des lettres majuscules (Q), leur valeur numérique par les lettres minuscules correspondantes (q) et l'unité par le nom de la grandeur de même nature prise pour unité. Lorsqu'on utilise les unités du Système international d'unités (SI), il suffit alors d'utiliser les notations normalisées correspondant au nom des unités de base (« m » pour le mètre, « s » pour la seconde, « K » pour le kelvin, etc.).

le désaccord entre résultats expérimentaux ne relève pas d'un constat de fait. L'accord ou le désaccord constitue plutôt une *hypothèse* dont l'examen fait appel aux probabilités et qui doit être tranchée par une décision³³ prise sur le fond de critères dépendant des risques encourus³⁴ – ainsi, le contrôle de la qualité des pièces fabriquées par une entreprise s'effectue au regard de la spécification de certaines tolérances définies typiquement par des impératifs économiques ou légaux.

Notre propos ne porte pas sur cette étape de prise de décision mais sur l'évaluation quantitative des incertitudes de mesure qui en constitue le préalable indispensable et fait l'objet des recommandations du GUM. En effet, si la comparaison et la combinaison des résultats de mesure requièrent l'évaluation des incertitudes associées, encore faut-il, pour répondre aux objectifs pratiques du GUM, que celles-ci soient définies et calculées par tous de la même manière³⁵. On a vu au 1.1 que le but du GUM était justement d'uniformiser l'évaluation des résultats de mesure et de leur incertitude associée. On vient de voir que l'enjeu de cette uniformisation était considérable puisqu'elle conditionne l'exploitation scientifique, technique et commerciale des résultats de mesure à une échelle devenue mondiale avec la globalisation des échanges. Ce qui précède permet, à présent, de saisir la difficulté à laquelle les rédacteurs du GUM se sont trouvés confrontés pour mener à bien leur tâche.

On l'a vu, l'estimation de la valeur du mesurande est obtenue à partir des données brutes fournies par l'expérience en effectuant deux processus d'inférence³⁶ : réduction statistique et correction. Ces deux processus d'inférence se fondent sur des informations, des modèles ou des théories de natures fort différentes : série d'indications expérimentales et modèle probabiliste pour la réduction des erreurs aléatoires ; informations issues des connaissances (scientifiques et techniques) dont on dispose sur le système de mesure pour la correction des erreurs systématiques. L'incertitude qui affecte l'estimation finale de la valeur du mesurande possède donc *deux* composantes à *première vue bien distinctes*, correspondant aux deux types d'inférence mis en œuvre pour tenter de remonter des données concrètes effectivement obtenues à la valeur de la grandeur qui les a produites (la valeur « vraie » v du mesurande). La quantification de l'incer-

33. Isabelle Drouet nous a fait remarquer que la perspective décisionniste était certes courante en statistiques mais nullement obligatoire, comme le montre l'approche de Fisher. Les méthodes statistiques utilisées dans le domaine de la mesure se rangent cependant du côté décisionniste. Là, comme dans bien d'autres disciplines, les manuels s'adressant aux étudiants ingénieurs présentent ce que Gigerenzer et Murray ont appelé une « théorie hybride », combinant les points de vue de Fisher et de Neyman-Pearson ; voir Gigerenzer *et al.*, *The Empire of Chance*, Cambridge University Press, 1989, chapitre 3. L'approche de Fisher est toutefois discutée par certains métrologues.

34. Par la mise en œuvre d'un test statistique, si l'on raisonne dans le cadre fréquentiste, ou par l'évaluation d'une vraisemblance, si l'on s'inscrit dans la perspective bayésienne.

35. Il suffirait, en toute rigueur, de savoir exactement comment chaque calcul d'incertitude a été effectué. Mais toute comparaison ou combinaison effective exigerait d'utiliser le détail des calculs effectués pour ramener les évaluations obtenues par des méthodes différentes à l'une d'entre elles. L'uniformisation est donc bien requise dans la pratique.

36. On rejoint ici les analyses de philosophes, comme Bogen et Woodward, montrant qu'un résultat de mesure ne renvoie à aucune des données particulières effectivement obtenues au cours de l'expérience : les théories sont confrontées à des résultats *inférés* à partir des données, et non aux données elles-mêmes qui sont trop « idiosyncratiques » pour pouvoir être expliquées ou prédites. J. Bogen & J. Woodward, « Saving the Phenomena », *The Philosophical Review*, 97(3), 1988, 303-352.

titude ne se présente pas de la même façon dans les deux cas. Le modèle probabiliste utilisé dans la réduction des erreurs aléatoires fournit une évaluation naturelle de la composante de l'incertitude correspondante dans le cadre fréquentiste : l'écart-type de la distribution de probabilité du modèle probabiliste des mesures, qui permet en effet de caractériser quantitativement la dispersion de ces dernières. Par contre, le traitement fréquentiste des erreurs systématiques ne semble guère proposer davantage, pour évaluer la composante associée, qu'un encadrement de l'erreur effectuée. Dans ces conditions, la question se pose de savoir si, compte tenu du rôle pris par la mesure dans le fonctionnement de nos sociétés, il est possible de faciliter le maniement des résultats en caractérisant la qualité d'un résultat de mesure par *un seul terme d'incertitude*. En d'autres termes, les deux composantes dégagées plus haut peuvent-elles légitimement être combinées et, si oui, comment ?

1.5 Équation du mesurande et fonction de mesure

Avant d'examiner plus en détails ces difficultés et les différentes solutions qui ont été successivement avancées pour les résoudre, il convient de mieux dégager les différentes relations mathématiques³⁷ à l'aide desquelles s'effectuent les deux types d'inférence présentés dans les sections 1.3 et 1.4.

La première relation mathématique qui vient naturellement à l'esprit est celle qui exprime la loi physique reliant la grandeur que l'on mesure (la grandeur D , dans l'exemple présenté à la section 1.3) et la grandeur que l'on veut mesurer (ℓ). Il s'agit de la relation qui exprime la loi gouvernant le principe de la mesure – c'est-à-dire le « phénomène servant de base au mesurage »³⁸ – dont l'expression mathématique, appelée « équation du mesurande » par les métrologues, permet de calculer la valeur de la grandeur d'entrée à partir de la valeur de la grandeur de sortie de la chaîne de mesure. Dans notre exemple, cette loi se réduit à la relation géométrique³⁹ :

$$L = L_E + D \quad (2)$$

37. Les méthodes mises en place par le GUM discutées ici ne s'appliquent qu'aux situations dans lesquelles grandeurs de sortie et grandeurs d'entrée de la chaîne de mesure peuvent être liées par des relations mathématiques. Il arrive souvent que cette condition ne soit pas remplie dans les systèmes qu'étudie l'ingénieur, qui peuvent être très complexes. Une autre méthode est alors adoptée que nous ne pouvons discuter dans les limites de cette étude. On pourra consulter à ce sujet Michèle Désenfant & Marc Priel, « Road map for measurement uncertainty evaluation », *Measurement*, 39, 2006, 841-848.

38. JCGM, *Vocabulaire International de Métrologie*, op. cit., p.17.

39. Rappelons que les lettres majuscules désignent les grandeurs physiques et les lettres minuscules, les valeurs numériques correspondantes dans le système d'unités choisi (la plupart du temps le SI) (voir la note 32). La répétition des mesures fournit plusieurs valeurs numériques d_i ; quant à la valeur numérique ℓ_E de L_E , elle est seulement connue avec une incertitude donnée, indiquée par le constructeur du calibre étalon. La détermination de cette incertitude est issue de l'étalonnage du calibre étalon, réalisé par un organisme accrédité, selon une méthode semblable à celle que décrit notre exemple. L'étalon utilisé au cours de cette phase d'étalonnage doit permettre de « raccorder » la valeur de la longueur du calibre du fabricant à l'unité du SI, le mètre, afin d'assurer la comparabilité *effective* des mesures réalisées à l'aide de L_E avec les mesures réalisées en d'autres lieux, avec d'autres calibres, produits par d'autres fabricants, mais également reliés au mètre par ce que l'on appelle une « hiérarchie d'étalonnage ». Chacune de ces phases d'étalonnage introduit une incertitude de mesure supplémentaire, et l'incertitude qui affecte la valeur de l'étalon est d'autant plus grande que celui-ci se situe plus bas dans la hiérarchie. Voir JCGM, *Vocabulaire International de Métrologie*, op. cit., 2.39 & 2.40, pp. 28-29.

Cependant, on l'a vu, les différentes valeurs numériques d_i que l'on obtient en effectuant plusieurs mesures de D conduisent à des équations numériques⁴⁰ :

$$\ell = \ell_E + d_i \quad (3)$$

qui se révèlent incompatibles, du moins si l'on considère que le mesurande possède une valeur numérique unique (valeur vraie ou valeur cible de la mesure⁴¹).

L'équation du mesurande ne constitue cependant qu'une partie de la « fonction de mesure »⁴² qui permet d'inférer le résultat de mesure à partir des données. Cette fonction de mesure décrit mathématiquement non seulement le principe mais aussi l'ensemble du dispositif et du processus de mesure. Elle peut être rapprochée de ce que certains philosophes désignent sous le nom de « modèle des données », à savoir une réalisation particulière d'une théorie des données liée à l'analyse du processus expérimental entier, et pas seulement à l'analyse du phénomène physique impliqué dans le principe de l'expérience⁴³.

En plus de l'équation du mesurande, la fonction de mesure doit donc comprendre, en premier lieu, les équations exprimant l'effet des grandeurs d'influence sur la grandeur étudiée. Le mesurande est, par exemple, la longueur L du calibre à la température $T_0 = 20^\circ\text{C}$, température qui est aussi celle à laquelle le calibre étalon a été étalonné par le fabricant – on notera L_E la longueur du calibre à T_0 , et $L_E(T)$ sa longueur à une température T quelconque. Si le mesurage est effectué à une température T différente, il faut corriger l'erreur systématique, due au fait que la longueur mesurée n'est pas la longueur à mesurer, en considérant les phénomènes de dilatation thermique qui affectent la longueur des deux cales⁴⁴ :

$$L_E(T) = L_E (1 + \alpha_E (T - T_0)) = L_E (1 + \alpha_E \Delta T) \quad (4)$$

$$L(T) = L (1 + \alpha (T - T_0)) = L (1 + \alpha \Delta T) \quad (5)$$

où $\Delta T = (T - T_0)$, et α et α_E désignent les coefficients de dilatation respectifs du calibre à mesurer et de l'étalon. Effectuer cette correction engage donc d'adjoindre au dispositif de

40. On voit ici que, comme le soulignent Bogen & Woodward, « Saving the Phenomena », *op. cit.*, l'équation du mesurande n'est pas satisfaite par les données brutes mais par les « résultats » de mesure, c'est-à-dire les estimations obtenues à l'issue des inférences décrites dans la section précédente. Les lois ne décrivent donc pas, pas plus qu'elles n'expliquent ou ne prédisent les observations, qui résultent du concours de trop de circonstances particulières.

41. R. Willink, *Measurement Uncertainty and Probability*, Cambridge University Press, 2013, p.6.

42. JCGM, *Vocabulaire International de Métrologie*, *op. cit.*, p.33.

43. P. Suppes, « Models of Data », in E. Nagel, P. Suppes & A. Tarski (eds.), *Logic, Methodology and Philosophy of Science : Proceedings of the 1960 International Congress*, Stanford University Press, 1962, 252-261. On pourrait aussi noter la parenté de la fonction de mesure avec les « lois phénoménologiques » de Cartwright qui en tire cependant des leçons philosophiques différentes ; N. Cartwright, *How the Laws of Physics Lie*, Oxford University Press, 1983.

44. À laquelle on eût rajouté d'autres corrections, comme la correction du champ de pesanteur, si le but de la mesure avait exigé de mesurer L avec un instrument d'une résolution supérieure ; L eût alors désigné un mesurande différent. On remarquera en outre que la relation utilisée pour corriger la dilatation thermique des cales n'est qu'une approximation et que la correction des effets thermiques est ainsi d'emblée entachée d'inexactitude.

mesure de D un dispositif de mesure de la température. L'équation du mesurande corrigée s'écrit alors :

$$L = \frac{L_E (1 + \alpha_E \Delta T) + D}{1 + \alpha \Delta T} = \mathcal{G}(D, L_E, \alpha, \alpha_E, \Delta T) \quad (6)$$

Dans la pratique, on linéarise cette expression par un développement limité au premier ordre et l'on obtient⁴⁵ :

$$L = L_E + \bar{d} - L_E \Delta \alpha \Delta T = \mathcal{F}(L_E, D, \Delta \alpha, \Delta T) \quad (7)$$

Le tableau 1 fournit une synthèse de ce travail de mise en équation. D'autres termes de correction d'erreur systématique pourraient être ajoutés. Ils peuvent être fournis par des expériences effectuées dans des laboratoires extérieurs; il pourrait s'agir, dans notre exemple, de corrections portant sur la valeur nominale de l'étalon, ou sur les graduations du comparateur, figurant sur les certificats d'étalonnage respectifs de ces deux instruments. Mais, dans d'autres cas, ces corrections pourraient aussi provenir de l'avis d'experts. Nous notons ici ces paramètres C_1, \dots, C_r . Dans le cas général, le mesurande Y est déterminé par p grandeurs d'entrée du modèle de mesure⁴⁶ X_1, \dots, X_p et les r autres paramètres C_1, \dots, C_r :

$$Y = \mathcal{F}(X_1, \dots, X_p, C_1, \dots, C_r) \quad (8)$$

Les expressions (7) (cas particulier de notre exemple) et (8) (cas générique) indiquent comment calculer le résultat de mesure, c'est-à-dire l'estimation (ici, ℓ) de la valeur vraie du mesurande. Il suffit, pour cela, de remplacer, dans l'expression fonctionnelle, les symboles dénotant les grandeurs par les estimations obtenues à l'issue du traitement des mesurages des quantités impliquées (ici, \bar{d} et $\overline{\Delta T}$), en introduisant des notations qui anticipent sur les sections suivantes) et les paramètres par les informations numériques externes lues dans les certificats et les tables; ce qui donne, dans le cas de notre exemple : où les symboles désignent alors des valeurs numériques.

$$\ell = f(\bar{d}, \ell_E, \overline{\Delta T}, \Delta \alpha) \quad (9)$$

Il reste cependant à préciser, d'une part, comment l'on obtient les estimations et les incertitudes associées de D et de ΔT à partir des mesurages effectués en laboratoire; d'autre part, comment il convient de combiner ces incertitudes avec les incertitudes des paramètres, tirées, quant à elles, des certificats d'étalonnages ou des avis d'experts, afin d'obtenir l'incertitude associée au résultat de mesure final. Le tableau 2 est une reproduction d'un « bilan d'incertitude » décrit dans le GUM, qui associe à chaque source d'incertitude identifiée une correction et la composante d'incertitude correspondante.

45. Cette approximation est d'ailleurs une source d'erreur supplémentaire, qui doit elle aussi faire l'objet, en toute rigueur, d'une analyse d'incertitude.

46. On notera que le mesurande, qui était la grandeur d'entrée de la chaîne de mesure, est devenue la grandeur de sortie du modèle de mesure, D et ΔT devenant les grandeurs d'entrée du modèle de mesure.

<p>équation du mesurande = loi physique générale décontextualisée</p> $L = L_E + D$	
<p>corrections</p>	<p>effets physiques affectant la mesure</p> <pre> graph TD A[effets physiques affectant la mesure] --> B[dilatation thermique : loi phénoménologique] A --> C[autres effets (négligés)] B --> D["Coefficients de dilatation α, αE (propriétés des matériaux)"] B --> E["température effective T (spécifique à cette mesure) température T0 à laquelle L est définie (définition)"] D --- F["Δα = α - αE"] E --- G["ΔT = T - T0"] </pre>
<p>fonction de mesure = loi contingente issue du modèle phénoménologique de l'expérience → loi en contexte</p> $L = \frac{L_E (1 + \alpha_E \Delta T) + D}{1 + \alpha \Delta T} = \mathcal{G}(D, L_E, \alpha, \alpha_E, \Delta T)$	
<p>fonction de mesure linéarisée = développement limité au premier ordre de la fonction de mesure</p> $L = \underbrace{L_E + D}_{\text{équation du mesurande}} - \underbrace{L_E \Delta \alpha \Delta T}_{\text{correction d'un effet systématique / biais}} = \mathcal{F}(D, L_E, \Delta \alpha, \Delta T)$	

Tableau 1 – Synthèse des différents éléments impliqués dans la modélisation de l'expérience pour l'exemple (E). L'« équation du mesurande » exprime la loi physique reliant la grandeur que l'on mesure (ici D) et la grandeur que l'on veut mesurer (L). Dans le cas présent, il s'agit d'une relation géométrique qui fait également intervenir une troisième grandeur, L_E , dont la valeur a été déterminée et communiquée par un laboratoire d'étalonnage. Cependant, l'équation du mesurande n'est pas vérifiée directement par les données expérimentales : il faut en effet effectuer les corrections liées aux effets systématiques. L'équation du mesurande corrigée devient la « fonction de mesure » \mathcal{G} , qui décrit mathématiquement non seulement le principe mais aussi l'ensemble du dispositif et du processus de mesure. Cette fonction de mesure met en relation une « grandeur de sortie » (ici L) avec différentes « grandeurs d'entrée » dont le nombre dépend des paramètres d'influence pris en compte, en fonction du niveau de détail accordé à la description du protocole. Pour pouvoir calculer l'incertitude de mesure sur L à partir des incertitudes de mesure des différentes grandeurs d'entrée, il se révèle souvent nécessaire de linéariser la fonction de mesure. Dans notre exemple, la fonction de mesure linéarisée \mathcal{F} fait réapparaître de façon immédiate l'équation du mesurande sous sa forme initiale, à laquelle est adjointe une correction des effets thermiques.

Composante de l'incertitude-type $u(x_i)$	Source de l'incertitude	Valeur de l'incertitude-type $u(x_i)$	$c_i \equiv \partial f / \partial x_i$	$u_i(l) \equiv c_i u(x_i)$ (nm)	Nombre de degrés de liberté
$u(l_S)$	Étalonnage du calibre étalon	25 nm	1	25	18
$u(d)$	Différence mesurée entre calibres	9,7 nm	1	9,7	25,6
$u(\bar{d})$	observations répétées	5,8 nm			24
$u(d_1)$	effets aléatoires du comparateur	3,9 nm			5
$u(d_2)$	effets systématiques du comparateur	6,7 nm			8
$u(\alpha_S)$	Coefficient de dilatation thermique du calibre étalon	$1,2 \times 10^{-6} \text{ } ^\circ\text{C}^{-1}$	0	0	
$u(\theta)$	Température du banc d'essai	0,41 °C	0	0	
$u(\bar{\theta})$	température moyenne du banc	0,2 °C			
$u(\Delta)$	variation cyclique de température de la salle	0,35 °C			
$u(\delta\alpha)$	Différence entre les coefficients de dilatation des calibres	$0,58 \times 10^{-6} \text{ } ^\circ\text{C}^{-1}$	$-l_S\theta$	2,9	50
$u(\delta\theta)$	Différence entre les températures des calibres	0,029 °C	$-l_S\alpha_S$	16,6	2
$u_c^2(l) = \sum u_i^2(l) = 1\,002 \text{ nm}^2$ $u_c(l) = 32 \text{ nm}$ $\nu_{\text{eff}}(l) = 16$					

Tableau 2 – Tableau tiré du GUM illustrant le détail d'un « bilan d'incertitude ». Le bilan porte sur l'exemple traité dans l'annexe H.1 du GUM, dont nous nous sommes inspirés pour développer l'exemple simplifié (E) présenté dans notre analyse.

2 Traitement statistique des données expérimentales : approche fréquentiste, approche bayésienne

Dans la section 1.5, notre présentation de la « fonction de mesure » appliquée à l'exemple (E) nous a permis d'aboutir à l'équation linéarisée suivante :

$$L = L_E + \bar{d} - L_E \Delta\alpha \Delta T \quad (10)$$

Par la suite, nous choisirons, pour des raisons de simplicité, de ne pas nous intéresser à la troisième composante de cette équation – elle exprime des corrections liées aux effets thermiques dont l'étude ne serait pas particulièrement instructive sur le plan épistémologique – et nous considérerons la fonction simplifiée suivante :

$$L = L_E + D \quad (11)$$

Dans ces conditions, la détermination de L nécessite la connaissance de L_E (dont la valeur estimée ℓ_E est fournie par un laboratoire d'étalonnage), ainsi que la mesure répétée de D (au moyen du comparateur) qui conduit à la collecte d'un ensemble de valeurs expérimentales d_i . Comme nous l'avons déjà mentionné, l'équation (11) indique comment attribuer une valeur à L en stipulant la relation entre valeurs numériques :

$$\ell = \ell_E + d \quad (12)$$

(où D est généralement la moyenne arithmétique \bar{d} des données d_i). Reste à établir la façon dont peuvent être déterminées les incertitudes correspondantes. C'est à cette étape que l'analyse statistique est requise.

En vertu de la relation (1), l'erreur de mesure associée à la mesure de la grandeur L de l'exemple (E) est définie par

$$e(\ell) = \ell - \lambda \quad (13)$$

où λ est la « valeur cible » ou « valeur vraie » du mesurande⁴⁷. Cependant, cette équation relie deux inconnues : valeur vraie et erreur. On ne peut inférer λ que s'il devient possible de connaître de quelque façon l'erreur commise : c'est en cherchant à évaluer l'erreur que l'on est amené à effectuer un constat d'incertitude. Au vu de la fonction de mesure simplifiée (11) et de la relation $\lambda = \lambda_E + \delta$ qu'elle implique entre les valeurs vraies des grandeurs concernées (à condition d'être considérée comme exacte), nous pouvons réécrire l'équation (13) de la façon suivante :

$$e(\ell) = (\ell_E - \lambda_E) + (d - \delta) = e(\ell_E) + e(d) \quad (14)$$

En isolant de la sorte les erreurs sur L_E et sur D , nous opérons une décomposition de l'erreur totale en deux composantes à propos desquelles nous disposons d'informations de nature différente, et auxquelles seront associées des méthodes de calcul d'incertitude distinctes. L'analyse

47. En complément du choix de notations décrit à la note 32, nous choisissons, dans la suite, d'écrire les valeurs vraies des grandeurs par les caractères grecs correspondants, c'est-à-dire ici λ pour L , λ_E pour L_E et δ pour D .

développée dans la suite de cette section montrera que l'on peut reconnaître dans ces deux composantes les deux types d'erreur que nous avons dégagés dans la section précédente et qui relèvent chacun d'un traitement spécifique. De fait, la première des deux composantes de l'équation (14) appelle une analyse d'erreurs dites « aléatoires », tandis que la seconde nécessite l'analyse d'une erreur dite « systématique ».

Dans ce qui suit, nous montrerons tout d'abord que le traitement qu'elle propose de ces deux types d'erreur empêche l'approche fréquentiste traditionnelle d'exprimer l'incertitude de mesure sous la forme d'une résultante unique et donc de répondre à l'objectif pratique visé. Nous verrons ensuite comment le GUM a cherché à remédier à ce défaut en conservant l'approche fréquentiste dans le traitement des erreurs de type aléatoire, mais en adoptant une approche que l'on peut qualifier de bayésienne pour aborder les erreurs systématiques. Certains métrologues ont accueilli cette solution comme un compromis insatisfaisant et ont préconisé une approche entièrement bayésienne des deux types d'erreur que nous exposerons dans la dernière partie de cette section.

2.1 Approche fréquentiste

Voyons pourquoi l'approche fréquentiste impose un traitement différent des deux composantes de l'erreur dégagées ci-dessus et aboutit à deux composantes d'incertitude qui, n'étant pas de même nature, ne peuvent être combinées en un seul terme. Nous commencerons par présenter l'étude de l'incertitude sur D , qui passe par une description des erreurs aléatoires. Puis, nous passerons au traitement que l'approche fréquentiste réserve à la valeur de L_E , qui est fournie par un laboratoire d'étalonnage, et à propos de laquelle est commise une composante systématique de l'erreur de mesure.

2.1.1 Traitement fréquentiste des erreurs aléatoires

Le traitement fréquentiste des erreurs aléatoires concerne les cas où l'on dispose, dans des conditions de répétabilité, d'une collection de données expérimentales dispersées – ici, une collection $\{d_i\}$ de mesures répétées de la différence D de longueur entre les deux calibres de l'exemple (E). Rappelons, à la suite des développements menés dans la section 1.2, que l'introduction d'un concept d'« erreur aléatoire » est une hypothèse théorique. Celle-ci découle d'un constat : les paramètres influant sur l'issue de l'expérience sont trop complexes pour être appréhendés individuellement. Comme l'explique Kyburg, le traitement de ce type d'erreurs suppose de saisir les erreurs *collectivement* par l'introduction d'un *modèle statistique*⁴⁸. À défaut de pouvoir connaître le mécanisme physique à l'origine de de l'erreur commise à chaque fois que l'on répète la mesure, ce mécanisme est traité comme une « boîte noire » dont on ne peut que très sommairement modéliser le contenu. Le modèle statistique constitue le cadre d'une inférence statistique qui permettra de remonter des données expérimentales à l'état de

48. Kyburg, « Measuring Errors of Measurement », *op. cit.*, p.76. Cela rejoint également les remarques de Bogen et Woodward, « Saving the Phenomena », *op. cit.*, sur le caractère idiosyncratiques des données expérimentales : il n'est pas possible de connaître *a priori* et dans le détail la relation causale entre l'état de la grandeur mesurée et les données brutes dans un contexte expérimental spécifique.

la grandeur elle-même.

Si le processus de mesure est suffisamment stable ⁴⁹, l'hypothèse fondamentale de la conception fréquentiste de la mesure consiste à modéliser la variabilité des données expérimentales comme si celles-ci étaient le résultat d'un processus *effectivement* aléatoire. Selon cette hypothèse, le résultat effectif d'une mesure individuelle est le produit du tirage aléatoire d'une valeur parmi une *population* ⁵⁰ statistique « parente ». On peut formuler l'hypothèse de la façon suivante :

(H1) *Le processus de mesurage de la grandeur \bar{d} peut être représenté mathématiquement par une variable aléatoire \hat{d} , qui porte sur les issues potentielles de la mesure, et dont la distribution de probabilité prend pour valeurs les fréquences relatives limites d'occurrence des données d_i .*

On parle pour \hat{d} de distribution statistique « parente ». Postuler l'existence d'une telle distribution permet de faire avancer l'enquête statistique pour peu que des hypothèses additionnelles soient formulées. Deux paramètres statistiques permettent en particulier de caractériser la variable aléatoire \hat{d} , son espérance (inconnue) μ et son écart-type (inconnu) σ . Comme nous l'avons vu, la description en termes d'erreurs de mesure consiste à considérer que des paramètres d'influence liés à certains aspects contingents de l'environnement expérimental, à l'expérimentateur lui-même ou aux instruments utilisés, perturbent la mesure, si bien que le résultat ne dépend pas uniquement de l'état du mesurande lui-même (ce qui eût été souhaitable). Si l'on considère que le système examiné ne varie pas sans raison, il est naturel de penser que la présence d'une dispersion dans les résultats de mesure provient de ces paramètres d'influence et non de l'état du mesurande. C'est l'écart-type (inconnu) σ de la variable aléatoire \hat{d} qui caractérise, dans le modèle, l'amplitude de la tendance qu'a le processus de mesure à produire des résultats variables. La mesure de \bar{d} peut, par ailleurs, être affectée par une erreur systématique (par exemple liée au comparateur); si tel est le cas, il faut chercher à la corriger. Mais si aucun élément accessible à l'examen n'a permis d'identifier une telle erreur, alors celle-ci doit être considérée comme nulle. Dans ces conditions, les erreurs aléatoires présentent, *par définition*, une espérance statistique nulle, c'est-à-dire qu'elles « s'annulent en moyenne » pour un nombre de mesures tendant vers l'infini. On peut alors relier l'espérance (inconnue) μ de \hat{d} et la valeur vraie (elle aussi inconnue) δ de D selon l'hypothèse suivante :

(H2) *Si aucune erreur systématique n'a été identifiée dans la mesure de D , on considère que l'espérance μ de la distribution statistique parente \hat{d} est égale à la valeur vraie δ de D .*

49. Eisenhart considère cette condition de stabilité comme un « postulat de la mesure », C. Eisenhart, « Realistic Evaluation of the Precision and Accuracy of Instruments Calibration Systems », *Journal of Research of the National Bureau of Standards – C. Engineering and Instrumentation*, 67C, 1963, 161-187, p.168. Nous pouvons noter ici la proximité entre ce « postulat de la mesure » et la notion de « conditions de répétabilité » déjà explicitées plus tôt.

50. On retrouve par exemple cette terminologie en 1908 chez William Gosset, dans les premières lignes d'un article publié sous le pseudonyme de Student, qui donnera son nom à la loi éponyme : « Toute expérience peut être considérée comme un individu dans une 'population' d'expériences qui pourraient être effectuées sous les mêmes conditions. Une série d'expériences est un échantillon tiré de cette population », Student, « The Probable Error of a Mean », *Biometrika*, 6, 1908, 1-25, p.1 (notre traduction).

L'hypothèse (H2) montre que l'analyse fréquentiste de D a pour cible l'espérance μ de la variable aléatoire \hat{d} . Ce faisant, l'analyse des erreurs *aléatoires* se réduit à un problème classique d'inférence purement statistique, que l'on peut formuler de la façon suivante : comment peut-on remonter, d'un échantillon fini de données issues d'une population statistique, à l'espérance de la population dans son ensemble ? La réponse fréquentiste consiste, dans un premier temps, à construire un « estimateur statistique » ; puis, dans un deuxième temps, à formuler le résultat sous la forme d'un « intervalle de confiance » probabiliste.

2.1.2 Estimateurs statistiques

Comme nous l'avons vu, la cible de l'inférence est l'espérance μ de la population statistique parente correspondant à la variable aléatoire \hat{d} . μ est inconnue, et il est seulement possible de l'évaluer imparfaitement par l'intermédiaire d'un *estimateur statistique*, c'est-à-dire une grandeur statistique qui présente les deux propriétés suivantes : (i) elle est *calculable* à partir des données expérimentales, et (ii) elle peut être considérée comme une bonne valeur approchée de μ .

La moyenne arithmétique \bar{d} d'un échantillon issu de la population parente tend vers l'espérance μ de \hat{d} lorsque la taille de cet échantillon tend vers l'infini. C'est pourquoi \bar{d} est un estimateur statistique classique de μ , en métrologie⁵¹ comme ailleurs. Cependant, la taille d'un échantillon réel est inévitablement *finie*. Par conséquent, \bar{d} diffère (en général) de l'espérance μ , qui demeure inconnue. S'il est certes possible que \bar{d} coïncide effectivement avec la valeur visée, il est toutefois impossible de s'en assurer : rien ne distingue, au regard de nos connaissances, un estimateur exact ($\bar{d} = \mu$) d'un estimateur inexact ($\bar{d} \neq \mu$) que la méthode statistique présente pourrait nous amener à calculer. À cet égard, il nous est donc impossible de savoir si un estimateur donné est exact ou non – c'est là, en somme, l'expression du caractère intrinsèquement faillible de toute connaissance. De façon plus générale, il nous est impossible de savoir si \bar{d} est proche de μ : l'écart $|\mu - \bar{d}|$ est inconnu. Cet état de fait est à l'origine d'une *incertitude* incontournable sur notre connaissance de μ , *quelle que soit l'exactitude effective de l'estimateur calculé*.

En quoi \bar{d} peut-il être considéré comme une bonne valeur approchée de μ ? Puisque nous ne connaissons pas l'écart $|\mu - \bar{d}|$, nous pouvons seulement raisonner en termes probabilistes. On peut montrer que, lors du tirage aléatoire d'un échantillon, \bar{d} aura d'autant plus tendance à se rapprocher de μ que (i) l'écart-type σ de \hat{d} est faible et (ii) la taille n de l'échantillon est grande (plus on répète la mesure, plus on améliore la qualité du résultat final). Mais σ est inconnu au même titre que μ : il se révèle donc également nécessaire d'en construire un estimateur. En pratique, on construit plutôt un estimateur de la variance σ^2 de \hat{d} , pour laquelle le meilleur candidat est généralement la « variance empirique » s^2 de l'échantillon de données d_i :

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (d_i - \bar{d})^2 \quad (15)$$

51. Voir la section 4.2.1 du JCGM, *Guide pour l'expression de l'incertitude de mesure*, op. cit., p.10.

Une synthèse des estimateurs statistiques est donnée dans le tableau 3.

Compte tenu de ce qui précède, on peut déterminer un paramètre d'incertitude, appelé dans le vocabulaire moderne « incertitude-type » $u(d)$, par la relation :

$$u(d) = \sqrt{\frac{s^2}{n}} \quad (16)$$

L'incertitude-type $u(D)$ exprime la tendance qu'a \bar{d} à différer de μ , pour un échantillon de n données. Autrement dit, il semble que l'on puisse affirmer :

$$\mu = \bar{d} \pm u(d) \quad (17)$$

sans toutefois que le signe \pm employé n'ait encore de signification rigoureuse. Une seconde étape de l'inférence statistique vise à préciser les contours probabilistes de l'équation (17), et s'articule autour de deux concepts essentiels, l'« intervalle de confiance » I_p , et le « niveau de confiance » p qui lui est associé.

Valeur estimée (inconnue)	Estimateur (connu, calculé)
Espérance de \hat{d} $\mu = \int_{-\infty}^{+\infty} t f_{\hat{d}}(t) dt$	Moyenne empirique $\bar{d} = \frac{1}{n} \sum_{i=1}^n d_i$
Variance de \hat{d} $\sigma^2 = \int_{-\infty}^{+\infty} (t - \mu)^2 f_{\hat{d}}(t) dt$	Variance empirique $s^2 = \frac{1}{n-1} \sum_{i=1}^n (d_i - \bar{d})^2$

Tableau 3 – Synthèse des deux principaux estimateurs statistiques employés dans le GUM. L'espérance μ de la variable aléatoire \hat{d} est estimée par la « moyenne empirique » \bar{d} , qui est la moyenne arithmétique des données collectées d_i . La variance de \hat{d} est estimée par la « variance empirique » s^2 , et permet de déterminer la qualité de l'estimation de μ .

2.1.3 Niveau de confiance et intervalle de confiance

L'analyse précédente est incomplète car elle ne fournit pas une interprétation probabiliste de l'incertitude de mesure, ce qui constitue pourtant l'objectif de l'approche fréquentiste. Si l'on souhaite mener l'analyse à son terme, il faut introduire préalablement un « niveau de confiance » p requis (p étant une valeur comprise entre 0 et 1) et construire ensuite, pour ce niveau de confiance, un « intervalle de confiance » I_p pour lequel on pense pouvoir affirmer que

$$\mu \in I_p \quad (18)$$

Si l'on souhaite que les équations (18) et (19) soient interprétables en termes de probabilités, le niveau de confiance p doit correspondre à une probabilité fréquentiste. Deux questions se font alors jour. Comment construit-on un intervalle de confiance? Comment caractérise-t-on son niveau de confiance?

De façon remarquable, ce que les estimateurs statistiques \bar{d} et s^2 nous permettent de dire sur μ ne dépend d'aucune condition sur la forme de la distribution \hat{d} . Un corollaire de cette propriété est que ces estimateurs ne permettent pas de raisonner directement en termes probabilistes. Pour compléter l'information donnée par les estimateurs, et donc, par extension, par l'incertitude-type $u(d)$, il est nécessaire de raisonner sur la forme de la distribution \hat{d} , et donc, à défaut de connaître \hat{d} , d'exprimer une nouvelle hypothèse (H3), portant sur la forme de cette distribution. Une hypothèse courante consiste à considérer que \hat{d} est distribuée selon une loi normale, auquel cas l'hypothèse (H3) prend la forme suivante :

(H3) *Hypothèse gaussienne : \hat{d} est distribuée selon une loi normale*⁵²

$$f_{\hat{d}}(t) \propto \exp\left(-\frac{1}{2}\left(\frac{\mu - t}{\sigma}\right)^2\right) \quad (19)$$

où μ et σ sont l'espérance et l'écart-type (inconnus) de \hat{d} .

La complexité du problème est alors réduite à un nombre fini de paramètres. Dans le cas de l'hypothèse gaussienne, la distribution est en effet entièrement déterminée par l'espérance μ et l'écart-type σ .

L'hypothèse gaussienne constitue le cas le plus traditionnel, et celui qu'on retrouve dans le GUM. Depuis les travaux de Laplace et Gauss sur les probabilités, la loi normale s'est imposée comme distribution archétypale, soutenue en cela par le théorème central limite⁵³. La suprématie de la loi normale en tant que loi « naturelle » a, depuis, été régulièrement contestée⁵⁴ et son statut, notamment dans le GUM, est désormais celui d'un simple modèle, utile mais imparfait. La loi normale reste cependant l'une des distributions les plus usitées, notamment du fait de sa simplicité mathématique⁵⁵.

L'importance de l'hypothèse (H3) au sein de l'approche fréquentiste tient à ce qu'elle ouvre la voie à un calcul probabiliste *déductif*. C'est sur la base d'un tel calcul, qui fait intervenir (i)

52. Nous avons choisi de noter la variable de la fonction de distribution T et non D afin d'éviter toute ambiguïté; la lettre D désigne en effet la valeur numérique qui sera attribuée à la quantité fixe D .

53. Stigler, *The history of statistics, op. cit.*

54. *Ibid*, pp.326-361. Nous pouvons citer, à titre d'exemple, les critiques que Campbell formule à plusieurs occasions contre l'importance accordée à une telle distribution, Campbell, *Physics, the Elements, op. cit.*, chapitre 17

55. Ce qui rejoint le constat énoncé par Humphreys, « (D)errière une grande partie des sciences physiques, repose ce principe : *c'est l'invention et le déploiement de mathématiques traitables qui suscite une part considérable du progrès dans les sciences physiques.* [...] Il y a une réciproque à ce principe : *la plupart des modèles scientifiques sont spécifiquement taillés pour s'accorder avec, et sont donc contraints par, les mathématiques disponibles.* », (notre traduction, Humphreys souligne). P. Humphreys, *Extending Ourselves : Computational Science, Empiricism, and Scientific Method*, Oxford University Press, 2004, p.56.

l'estimateur statistique \bar{d} (ii) l'incertitude-type $u(D)$ tirée de l'estimateur statistique s^2 , et (iii) certaines informations supplémentaires sur \hat{d} , qu'on construit finalement, une fois choisi le niveau de confiance p souhaité, un intervalle de confiance I_p correspondant :

$$I_p = [\bar{d} - k_p u(d), \bar{d} + k_p u(d)] \quad (20)$$

où k_p est un facteur qui dépend (une fois p choisi) de la forme de la distribution \hat{d} . La demi-largeur $U_p(d) = k_p u(d)$ de l'intervalle I_p est interprétée comme une évaluation numérique de la confiance que l'on peut avoir dans l'inférence statistique effectuée, et, par suite, comme une évaluation de la *qualité* du résultat de mesure. En effet, conformément aux intentions initiales, I_p est construit de façon à être fortement susceptible de contenir l'espérance μ que l'on cherche à connaître.

Il reste cependant à comprendre ce que signifie ce que nous entendons par « fortement susceptible », c'est-à-dire à interpréter le niveau de confiance p en termes fréquentistes. Puisque I_p et μ sont tous deux bien définis et fixés, seuls deux cas sont possibles : soit μ est incluse dans I_p , soit elle en est exclue. Par conséquent, la procédure d'évaluation de μ , qui s'opère à chaque fois selon le même mode prédéfini, et dont l'aboutissement est la production de l'intervalle I_p , peut se solder soit par un succès, soit par un échec. Le niveau de confiance p quantifie le *taux de succès* escompté de cette procédure sur un nombre infini d'essais. Selon ces principes, on peut alors dire, en termes fréquentistes, que la procédure présente une probabilité p de produire un intervalle I_p qui contienne effectivement μ . Il convient donc d'insister sur le fait que le niveau de confiance p *ne correspond pas* à la probabilité que I_p contienne effectivement μ . De ce fait, l'approche probabiliste de la mesure proposée dans le cadre d'une interprétation fréquentiste des probabilités ne peut porter que sur la *procédure* dans son ensemble, et non sur le résultat final effectivement obtenu, lequel présente une « erreur », fixe mais inconnue. La formulation fréquentiste rigoureuse de la proposition « la probabilité que $\mu \in I_p$ est p », et donc du résultat, est la suivante :

Si l'on répète la procédure consistant à collecter un échantillon de n données à partir duquel on calcule un intervalle de confiance I_p , alors la fréquence limite d'occurrence des cas pour lesquels l'intervalle I_p contiendra effectivement le paramètre μ sera p .

Il convient de souligner le point suivant : la formulation probabiliste du résultat ne porte pas sur le paramètre μ lui-même, que l'on cherche à connaître, mais *sur la procédure de mesure dans son ensemble*. Cette formulation caractérise le taux de succès de la procédure par la capacité de cette dernière à produire des intervalles « corrects », c'est-à-dire contenant effectivement μ (il y a succès si $\mu \in I_p$ et échec sinon). On peut encore énoncer la proposition précédente de la façon suivante :

Dans le cadre de l'inférence statistique visant à estimer μ , la procédure consistant pour l'expérimentateur à identifier \bar{d} et μ mène à commettre une erreur finale $e = d - \mu$ qui sera, dans $100p\%$ des cas, inférieure (en valeur absolue) à la demi-largeur $U_p(d) = k_p u(d)$ de l'intervalle de confiance I_p .

Le terme $e = d - \mu$ est la composante de l'erreur finale due à la présence d'erreurs aléatoires. $U_p(d)$ est l'« incertitude élargie à $100p\%$ » attachée à la dispersion des résultats de mesure. Elle constitue une contribution à l'incertitude de mesure complète sur le résultat.

2.1.4 Prise en compte des erreurs systématiques et formulation du résultat final

Nous avons déjà eu l'occasion de laisser entendre que le traitement des erreurs systématiques différait de celui des erreurs aléatoires. Et, de fait, il apparaît bien ici que la méthode de *réduction* des erreurs aléatoires proposées à la section précédente ne peut pas être appliquée aux erreurs systématiques qui demeurent fixes lors de la répétition des mesurages. Ces erreurs peuvent par contre être appréhendées individuellement au moyen d'un modèle physique approprié qui permet d'en effectuer la *correction*. Cette correction, imparfaite, appelle une détermination de l'incertitude associée qui ne peut pas prendre la forme probabiliste exposée précédemment. Ce problème a connu des solutions variées avant même la publication du GUM en 1993. Jusqu'à la fin des années 1960 environ, l'on compte, parmi les solutions proposées, une approche fréquentiste traditionnelle qui fait appel à l'estimation de « limites crédibles de l'erreur »⁵⁶, visant à proposer un encadrement de l'erreur systématique par une analyse spécifique (susceptible de faire intervenir un jugement subjectif).

Reprenons l'exemple (E) de la figure 1 auquel correspond la fonction de mesure simplifiée suivante :

$$L = L_E + D \quad (21)$$

La mesure de L_E n'est pas effectuée par l'expérimentateur ; ce dernier se réfère aux valeurs indiquées par le laboratoire qui a procédé à l'étalonnage du calibre étalon. Par conséquent, si la valeur ℓ_E fournie par le laboratoire d'étalonnage diffère de la valeur vraie λ_E de l'étalon, cela introduit une composante fixe d'erreur ($\lambda_E - \ell_E$) : l'on est en présence d'une erreur systématique. Supposons ensuite que la valeur de L_E soit donnée par le laboratoire d'étalonnage sous forme d'un encadrement :

$$\ell_0 - h < \ell_E < \ell_0 + h \quad (22)$$

Un tel encadrement de l'erreur systématique ne constitue pas un intervalle probabiliste – ce n'est pas un intervalle de confiance, calculé à partir d'une variance statistique. Dans l'approche traditionnelle fréquentiste, l'analyse s'arrête ici, et la formulation finale d'un résultat de mesure se fait au travers d'un énoncé d'*exactitude* dans l'expression duquel interviennent deux composantes⁵⁷ : une composante aléatoire, étudiée précédemment, et une composante systématique. Or, ces deux composantes ne sont pas calculées de la même façon, et il n'est pas possible de les combiner mathématiquement. Par construction même du modèle fréquentiste, et en particulier du fait de la distinction entre deux types d'erreurs, aléatoires et systématiques, il n'est pas possible d'exprimer le résultat final au moyen d'une incertitude résultante unique. Au début des années 1970, ce constat est connu, compris et accepté par les tenants de l'approche traditionnelle fréquentiste, comme en témoigne la lecture du manuel de 1973 du National Physical Laboratory (NPL) :

À l'évidence, afin de tirer pleinement profit de cette approche, il est nécessaire de conserver les deux catégories d'incertitude séparées, non seulement dans la

56. Eisenhart, « Realistic Evaluation of the Precision and Accuracy of Instruments Calibration Systems », *op. cit.*, p.181. L'article d'Eisenhart peut être considéré comme le point culminant de cette approche traditionnelle.

57. Eisenhart, « Realistic Evaluation of the Precision and Accuracy of Instruments Calibration Systems », *op. cit.*, p.181.

combinaison progressive des incertitudes tout au long d'une expérience complexe, mais aussi dans la formulation du résultat final.⁵⁸

Il y a là une compétition entre deux exigences antagonistes. D'un côté, la richesse des résultats d'expérience est plus fidèlement conservée par un compte-rendu détaillé de chacune des composantes d'incertitude évaluée individuellement⁵⁹, ce qui permet de limiter la perte d'information le long de la chaîne de communication des résultats. D'un autre côté, pour certains types d'utilisations du résultat, l'objectif est de fournir un résultat sous la forme la plus simple possible, afin qu'il soit facilement utilisable et puisse, notamment, se prêter à des tests statistiques. Dans ce cas, il est essentiel de pouvoir synthétiser l'incertitude de mesure en une valeur résultante unique, c'est-à-dire de pouvoir combiner les différentes composantes d'incertitudes identifiées en amont.

C'est en grande partie en discutant du bien-fondé de ces deux exigences que les métrologues ont été conduits à projeter la rédaction du GUM. Entre 1977 et 1981, le BIPM convoque une assemblée d'experts pour organiser un débat autour de cette question. Les conclusions de ce débat sont publiées dans la « recommandation INC-1 »⁶⁰, qui établit une liste d'objectifs et de principes que le BIPM considère comme préférables pour l'expression commune de l'incertitude de mesure. Cette liste constituera la base sur laquelle se fonderont les futurs rédacteurs du GUM dans la décennie qui suivra. Le BIPM y adopte la position suivante : il est souhaitable que chaque méthode d'analyse d'incertitude puisse *toujours* permettre d'exprimer l'incertitude de mesure par une résultante *unique*. Cela ne signifie pas que l'incertitude de mesure doit toujours être formulée ainsi – seulement qu'elle puisse l'être. Dans ce cadre – *et uniquement en raison de cette exigence* – l'approche fréquentiste traditionnelle présente alors une lacune sérieuse.

Pour désigner les différentes composantes d'incertitude, les termes « incertitude aléatoire » et « incertitude systématique » avaient été proposés⁶¹ dans les années 1970, alors que l'approche traditionnelle commençait à s'étioler, et que la notion d'« incertitude » venait remplacer peu à peu celle d'« exactitude » dans le vocabulaire métrologique. Un résultat était alors formulé de la façon suivante :

$$L = \ell \pm u_{al} \pm u_{sys} \text{ m} \quad (23)$$

58. P. Campion, J. Burns & A. Williams, *A code of practice for the detailed statement of accuracy*, National Physical Laboratory, 1973, p.2 (notre traduction).

59. Ainsi trouve-t-on dans JCGM, *Guide pour l'expression de l'incertitude de mesure*, op. cit., p.26, le conseil suivant : « En pratique, la quantité d'information nécessaire pour documenter un résultat de mesure dépend de l'usage prévu ; cependant, le principe de base reste inchangé : lorsqu'on exprime le résultat d'un mesurage et son incertitude, il vaut mieux pécher par excès d'information plutôt que par défaut. ».

60. R. Kaarls, « Report of the BIPM Working Group on the statement of measurement uncertainties (1st meeting - 21 to 23 October 1980) to the Comité International des Poids et Mesures », Bureau International des Poids et Mesures, 1980, 13-14. La recommandation est reproduite dans P. Giacomo, « News from the BIPM », *Metrologia*, 17(2), 1981, 69-74, ainsi que dans JCGM, *Guide pour l'expression de l'incertitude de mesure*, op. cit., p.30.

61. Nous renvoyons à K. Protassov, *Analyse statistique des données expérimentales*, EDP sciences, 2002, pp.115-116 pour une explication plus détaillée de cette terminologie, en particulier sur le plan technique. Cette double appellation semble avoir été surtout utilisée pendant une décennie environ, de 1970 à 1980, durant laquelle elle a été régulièrement critiquée comme étant impropre, le caractère aléatoire ou systématique étant relatif aux erreurs correspondantes, non aux incertitudes elles-mêmes. Les termes appropriés pour décrire l'approche fréquentiste traditionnelle sont plutôt les termes de « fidélité », « justesse » et « exactitude ».

Dans la continuité, des méthodes *ad hoc* suggéraient alors une combinaison quadratique des différentes composantes⁶² :

$$u_{\text{res}} = \sqrt{u_{\text{al}}^2 + u_{\text{sys}}^2} \quad (24)$$

Cependant, une telle opération n'a aucune justification rigoureuse. C'est en cherchant à remédier à ces difficultés, attribuées à une insuffisance de l'approche fréquentiste, que le GUM a fait appel aux probabilités épistémiques.

2.2 Approche bayésienne

2.2.1 De l'approche fréquentiste à l'approche hybride du GUM

L'approche fréquentiste classique est insatisfaisante dès lors que l'on cherche à obtenir une résultante unique d'incertitude. S'il semble légitime d'affirmer, dans un premier temps, que les statistiques fréquentistes fonctionnent « naturellement » pour le traitement des erreurs aléatoires, leur limite se situe en revanche dans le traitement des erreurs systématiques, dont elles ne permettent pas une analyse statistique, comme le souligne Bich :

Les erreurs aléatoires sont considérées comme une population (typiquement gaussienne), à laquelle les outils statistiques classiques peuvent être appliqués. Par exemple, la variance V d'une population est généralement évaluée par la variance s^2 d'un échantillon. Cette procédure ne peut pas être utilisée pour les incertitudes liées à des biais, *car il n'y a pas de population derrière une erreur systématique* et, par conséquent, *il n'y a pas d'échantillon derrière un biais*.⁶³

La conception du GUM résulte d'un choix fondamental, selon lequel « les erreurs systématiques et aléatoires sont traitées sur la même base probabiliste »⁶⁴. Ce choix est revendiqué dans le corps du document :

(C)e *Guide* traite exactement de la même façon les composantes de l'incertitude provenant d'effets aléatoires et celles provenant des corrections estimées d'effets systématiques, lorsqu'on évalue l'incertitude du résultat d'un mesurage. [...] (T)outes les composantes de l'incertitude sont de même nature et doivent être traitées de manière identique.⁶⁵

Il est en particulier recommandé de ne plus employer les termes « incertitude aléatoire » et « incertitude systématique », auxquels le GUM substitue, sans établir d'équivalence stricte, les méthodes de « type A » et de « type B » d'évaluation de l'incertitude :

62. Bien que cette approche soit désormais considérée comme obsolète en métrologie, elle reste présente dans certains articles scientifiques. Voir par exemple l'article des neutrinos dits « supraluminiques », publié par Adam et al., « Measurement of the neutrino velocity with the OPERA detector in the CNGS beam », *ArXiv*, 2011, 1-24, p.22. On y retrouve les deux composantes d'incertitude, l'une statistique, l'autre systématique, d'abord exprimées séparément, puis combinées de façon *ad hoc*.

63. W. Bich, « From Errors to Probability Density Functions. Evolution of the Concept of Measurement Uncertainty », *IEEE Transactions on Instrumentation and Measurement*, 61, 2012, 2153-2159, p.2156 (nous soulignons; la traduction est également la nôtre).

64. C. Ehrlich, R. Dybkaer, W. Wöger, « Evolution of philosophy and description of measurement (preliminary rationale for VIM3) », *Accreditation and Quality Assurance*, 12, 2007, 201-218, p.205 (notre traduction).

65. JCGM, *Guide pour l'expression de l'incertitude de mesure*, op. cit., p.57.

L'incertitude d'un résultat de mesure comprend généralement plusieurs composantes qui peuvent être groupées en deux catégories d'après la méthode utilisée pour estimer leur valeur numérique :

A. celles qui sont évaluées à l'aide de méthodes statistiques,

B. celles qui sont évaluées par d'autres moyens.⁶⁶

Dans le GUM, les méthodes de type A conservent un fondement fréquentiste. Ce sont les méthodes dites « de type B » qui sont traitées au moyen de probabilités épistémiques. Cette portion du GUM n'est pas pour autant écrite dans un langage explicitement bayésien : les termes « Bayes » et « bayésien », en particulier, n'apparaissent pas. Cependant, au vu de l'insistance avec laquelle les probabilités y sont décrites comme des degrés de croyance, on comprend pourquoi le GUM a été largement interprété comme étant un texte au moins en partie bayésien, comme le revendiquent par exemple Lira et Wöger :

[Un effet systématique] constitue un élément pour lequel le traitement statistique classique échoue complètement. C'est principalement pour cette raison que le GUM a été écrit. Le GUM recommande d'évaluer les effets systématiques conformément à la procédure d'évaluation qu'on appelle « type B ». Bien que ce ne soit pas mentionné explicitement dans le corps principal du GUM, cette procédure est de nature essentiellement bayésienne.⁶⁷

Dans la section suivante, nous décrivons l'approche développée dans le GUM pour traiter les erreurs systématiques par les méthodes de type B. Nous dégagerons, dans le même mouvement, les ressorts bayésiens implicites de l'exposé du GUM en nous appuyant sur les commentaires de métrologues et statisticiens favorables à l'utilisation de statistiques bayésiennes.

2.2.2 Traitement bayésien des erreurs systématiques : méthodes de « type B »

Les méthodes de type B d'évaluation de l'incertitude visent à permettre la détermination, en termes probabilistes, d'une composante d'incertitude associée à une erreur systématique. Rappelons l'équation (11) fournissant la fonction de mesure simplifiée correspondant à notre exemple (E) :

$$L = L_E + D$$

Les informations fournies à propos de L_E sont identiques dans l'approche traditionnelle décrite au 2.1.4 et dans celle du GUM : on a ainsi l'encadrement

$$\ell_0 - h < \ell_E < \ell_0 + h$$

Cependant, la façon d'exploiter ces informations diffère d'une approche à l'autre. L'objectif de la méthode de type B, annoncé dans la recommandation INC-1, est de produire à propos de ℓ_E une incertitude équivalente à une variance statistique afin que cette incertitude puisse être combinée avec la ou les variances produites par les analyses fréquentistes des erreurs aléatoires, et ainsi conduire à l'expression d'une incertitude condensée en un seul terme. Le pivot de la

66. JCGM, *Guide pour l'expression de l'incertitude de mesure*, op. cit., p.viii.

67. I. Lira & W. Wöger, « Comparison between the conventional and Bayesian approaches to evaluate measurement data », *Metrologia*, 43(4), 2006, S249-S259, p.S253 (notre traduction).

description probabiliste de l'incertitude dans les méthodes de type B réside dans l'introduction d'une variable aléatoire, notée ⁶⁸ $\hat{\lambda}_E$, portant sur la valeur *fixe et inconnue* de L_E . Ce mouvement s'appuie de façon essentielle sur l'interprétation épistémique des probabilités : les probabilités ne décrivent alors plus la variabilité effective d'un processus de mesure, mais portent sur l'état de connaissance de l'expérimentateur en exprimant son degré (rationnel) de croyance quant à la valeur λ_E du paramètre décrit. Comme le soulignent Weise et Wöger, cela rend possible un traitement probabiliste d'un paramètre *constant* :

En interprétant la probabilité comme une description numérique d'un état de connaissance incomplète, sur la base de toute forme d'information rationnelle et pertinente possible, les statistiques bayésiennes offrent un énoncé de probabilité portant sur la vraie valeur, non aléatoire, inconnue, d'une quantité. ⁶⁹

La variable aléatoire ainsi employée est *construite* sur la base de toutes les informations disponibles. Une telle quantification probabiliste passe par l'expression d'une *distribution a priori*. Plutôt que de simplement statuer sur les « limites crédibles » de l'erreur (difficiles à interpréter) auxquelles doit se limiter la méthode fréquentiste classique, l'approche bayésienne de la mesure associe aux paramètres d'influence étudiés des variables aléatoires dont la distribution de probabilité peut être calculée à partir des informations disponibles, soit par des considérations élémentaires, soit au moyen de techniques bayésiennes usuelles (au premier rang desquelles figure le « principe d'entropie maximum », bien que ce principe soit absent du GUM lui-même). Ainsi peut-on mener l'analyse dans le cas de notre exemple : s'il n'existe à propos de ℓ_E aucune information autre que l'encadrement exprimé par l'équation (22), aucune raison ne pousse à privilégier une valeur plutôt qu'une autre parmi celles qui sont comprises dans l'encadrement. Cela suggère de choisir, pour la distribution de probabilité $\hat{\lambda}_E$, une loi rectangulaire centrée sur la valeur ℓ_0 , et de demi-largeur h (figure 2). Par analogie avec le cas fréquentiste, la racine carrée de la variance ainsi obtenue est appelée « incertitude-type » associée à ℓ_E , et s'écrit $u(\ell_E)$.

Que signifie la variable aléatoire ainsi employée pour décrire le biais que l'on cherche à caractériser ? Contrairement au cas du modèle fréquentiste, les valeurs prises par la distribution de probabilité ne sont plus des fréquences limites : il s'agit désormais de degrés de croyance. D'un point de vue conceptuel, la différence essentielle entre les approches fréquentiste et bayésienne de la mesure réside donc dans le fait que le domaine des probabilités n'est pas constitué par des événements mais par des propositions portant sur la valeur possible de la grandeur considérée. La probabilité n'a plus trait au processus physique de mesure, mais à la connaissance que l'expérimentateur estime avoir de la grandeur qu'il souhaite évaluer (ici L_E). Corrélativement, le propos ne porte plus sur la grandeur elle-même, mais sur ce que l'expérimentateur

68. Weise et Wöger, dans leur article séminal de 1992, parlent d'« estimateurs du mesurande ». « Un estimateur n'est pas identique au mesurande auquel il correspond, mais sert uniquement à estimer la 'vraie' valeur du mesurande, c'est-à-dire la valeur qui doit être déterminée », K. Weise & W. Wöger, « A Bayesian theory of measurement uncertainty », *Measurement Science and Technology*, 3, 1992, 1-11, p.3. Malgré cette remarque, ils désignent cependant par le même symbole le mesurande, son estimateur, et les valeurs que l'on peut attribuer au mesurande. Nous pensons, pour plus de clarté et pour éviter toute ambiguïté, qu'il est nécessaire de dénommer la variable aléatoire différemment de la grandeur et de la valeur de la grandeur, c'est pourquoi nous l'indiquons avec un chapeau pour signaler qu'il s'agit d'une fonction.

69. Weise & Wöger, « A Bayesian theory of measurement uncertainty », *op. cit.*, p.2 (notre traduction).

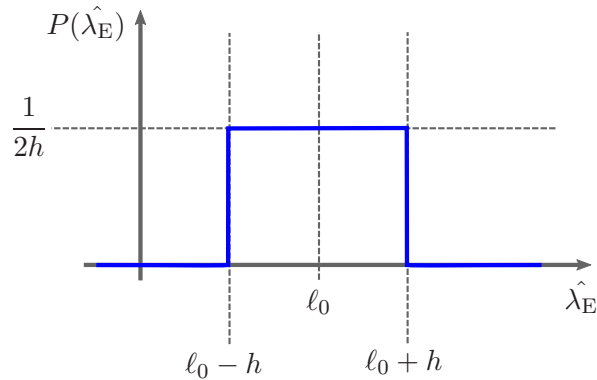


FIGURE 2 – Attribution d’une distribution de probabilité à la variable aléatoire $\hat{\lambda}_E$, en fonction des informations disponibles. Le certificat d’étalonnage indique que $\ell_0 - h < \ell_E < \ell_0 + h$. Aucune information ne porte à établir de distinction entre les valeurs de cet intervalle. L’expérimentateur est donc conduit à attribuer à toutes ces valeurs une même probabilité et à choisir pour la distribution a priori de λ_E une fonction de type « rectangulaire ».

peut raisonnablement croire à propos de cette grandeur ⁷⁰.

La façon dont sont ainsi mises en œuvre les méthodes de « type B » répond donc à l’objectif annoncé initialement : permettre le traitement sur des bases probabilistes de *toutes* les composantes d’incertitude. Or, les variables aléatoires fréquentistes et bayésiennes obéissent aux mêmes axiomes mathématiques ; par conséquent, l’approche du GUM permet de combiner les incertitudes évaluées selon les deux méthodes, type A et type B, en s’appuyant sur les propriétés mathématiques de l’algèbre des variables aléatoires communes aux deux méthodes. Le calcul de l’incertitude sur le mesurande (ici ℓ) passe donc par la combinaison des différentes composantes d’incertitude : c’est la formule dite de « propagation des incertitudes » ⁷¹. L’incertitude résultante est appelée « incertitude-type composée » ⁷² u_c . Dans le cadre de notre exemple, la propagation des incertitudes prend une forme simple (qui peut être plus complexe dans le cas général) :

$$u_c = \sqrt{u_1^2 + u_2^2 + \dots} \quad (25)$$

Soit, dans le cas de l’exemple (E) :

$$u_c = u(\ell) = \sqrt{u(d)^2 + u(\ell_E)^2} \quad (26)$$

On peut alors construire un intervalle d’incertitude pour L à partir de u_c et de la valeur ℓ adoptée. Cet intervalle est désigné dans le GUM ainsi que dans le *Vocabulaire International de Métrologie* (VIM) sous le nom d’« intervalle élargi » :

70. Joint Committee for Guides in Metrology (JCGM), *Evaluation of measurement data - An introduction to the “Guide to the expression of uncertainty in measurement” and related documents*, 2009, p.3.

71. JCGM, *Guide pour l’expression de l’incertitude de mesure*, op. cit., p.7.

72. JCGM, *Vocabulaire International de Métrologie*, op. cit., pp.26-27.

Intervalle élargi : intervalle contenant l'ensemble des valeurs vraies d'un mesurande avec une probabilité déterminée, fondé sur l'information disponible.⁷³

L'emploi des probabilités épistémiques pour les méthodes de type B semble donc permettre la « réunification », en une résultante unique, des composantes d'incertitude que l'approche fréquentiste, quant à elle, sépare par impossibilité de traiter de la même façon deux types d'information différents.

2.2.3 Extension de l'approche du GUM : traitement bayésien des erreurs aléatoires

Alors qu'elle résout en apparence un problème essentiel de la mesure, l'approche du GUM suscite dans le même temps une controverse importante. La formule de « propagation des incertitudes » de la méthode hybride du GUM consiste à combiner des objets dont la nature est fondamentalement différente, puisque les erreurs aléatoires sont traitées par des méthodes de « type A », fréquentistes, et les erreurs systématiques par des méthodes de « type B », bayésiennes. Les probabilités employées n'y sont donc pas les mêmes et ne portent pas sur le même objet : processus physique d'une part, état de connaissance de l'expérimentateur d'autre part.

Dans le GUM, l'accent est mis sur l'apparente cohérence mathématique de la méthode adoptée, et sur les problèmes techniques qu'elle résout – en particulier, la possibilité de « propager » (c'est-à-dire combiner) simplement les incertitudes. Cependant, la question de la nature des probabilités employées, qui prend en partie une forme philosophique, n'est pas attaquée de façon claire, et cela est source de confusions et de critiques. Pour certains, c'est l'utilisation des probabilités épistémiques tout court qui est problématique (nous reviendrons plus en détail sur les critiques fréquentistes adressées à l'approche bayésienne dans la section 3.1.1). À l'inverse, pour d'autres métrologues, l'incertitude composée ne saurait être réellement comprise que par l'interprétation épistémique des probabilités. En fin de compte, l'approche du GUM est souvent critiquée *des deux côtés* comme étant bancale, voire incohérente.

Certaines propositions récentes suggèrent de résoudre cette incohérence en adoptant une approche *entièrement bayésienne*, qui consiste à interpréter *toutes* les composantes d'incertitude de façon bayésienne et mène à revoir le traitement des méthodes de « type A » d'évaluation de l'incertitude. La proposition formulée par Kacker et Jones quelques années après la publication du GUM va dans ce sens :

[Le GUM] recommande les statistiques classiques (fréquentistes) pour l'évaluation des composantes d'incertitudes de Type A ; mais il interprète l'incertitude composée d'un point de vue bayésien. Cela est incohérent. Pour surmonter cette incohérence, nous suggérons que toutes les incertitudes de Type A soient évaluées par une approche bayésienne.⁷⁴

73. JCGM, *Vocabulaire International de Métrologie*, op. cit., p.27.

74. R. Kacker & A. Jones, « On use of Bayesian statistics to make the Guide to the Expression of Uncertainty Measurement consistent », *Metrologia*, 40(5), 2003, 235-248, p.235 (notre traduction).

Le « supplément 1 » du GUM, publié en 2008, propose une telle extension⁷⁵. Comment cela est-il mené à bien? Rappelons que, dans le cas de l'exemple (E), l'expérimentateur dispose d'une collection de valeurs expérimentales d_i issues de la mesure répétée de la grandeur D . L'approche fréquentiste détermine l'incertitude sur D à partir de la variance empirique de la collection. La méthode bayésienne ne met pas en œuvre les mêmes types d'inférence et ne probabilise pas sur les mêmes objets. Dans l'interprétation fréquentiste, les probabilités portent sur les données que le processus de mesure peut produire; elles caractérisent, dans le cadre bayésien, la connaissance que l'expérimentateur possède de l'inconnue visée.

Nous avons montré que le traitement bayésien des erreurs systématiques passe par l'introduction d'une variable aléatoire qui quantifie les degrés de croyance que l'expérimentateur entretient à l'égard des différentes valeurs qu'il pourrait attribuer au paramètre étudié. De la même façon, pour le traitement des erreurs aléatoires relatives à la grandeur D , l'expérimentateur peut construire une distribution *a priori* (ou *prior*) $\hat{\delta}$, sur la base des informations extérieures dont il dispose, et en mettant à contribution, si nécessaire, des outils mathématiques bayésiens. En ce sens, il s'agit d'une extension de la méthode de « type B » décrite dans le GUM.

Cependant, il s'agit ici d'analyser des résultats de mesures répétées; l'expérimentateur dispose donc de données expérimentales, les valeurs $\{d_i\}$ issues de la mesure de D . Le théorème de Bayes lui permet alors de procéder à une révision de ses connaissances en calculant la distribution *a posteriori* à partir du *prior* $\hat{\delta}$ et des d_i :

$$p(\hat{\delta} | \{d_i\}) \propto L(\hat{\delta} | \{d_i\}) \times p(\hat{\delta}) \quad (27)$$

où $L(\hat{\delta} | \{d_i\})$ est la fonction de vraisemblance qui exprime, sous forme d'une fonction de $\hat{\delta}$, la probabilité $p(\{d_i\} | \hat{\delta})$. L'approche bayésienne permet ainsi d'opérer une *inversion de probabilité*⁷⁶ : à partir d'une distribution de probabilité $p(\{d_i\} | \hat{\delta})$ portant sur les issues possibles d'une mesure, étant donnée la valeur de la grandeur mesurée, l'approche bayésienne va ainsi permettre l'expression d'une distribution de probabilité $p(\hat{\delta} | \{d_i\})$ portant sur la valeur de la grandeur mesurée, étant donnés les résultats de mesure effectivement obtenus.

On comprend ici en quoi l'adoption d'une l'interprétation épistémique des probabilités permet l'écriture de probabilités portant directement sur une valeur fixe (comme δ , valeur vraie de D , ou plus généralement λ , valeur vraie du mesurande ℓ), là où le modèle fréquentiste n'autorise de parler en termes probabilistes que des issues individuelles de mesure (ici les $\{d_i\}$), considérées comme les *réalisations* d'une variable aléatoire \hat{d} . La modélisation bayésienne formule cela à travers une *conditionnalisation* : bien que la valeur vraie δ de la grandeur D soit fixe,

75. Joint Committee for Guides in Metrology (JCGM), *Evaluation of measurement data – Supplement 1 to the “Guide to the expression of uncertainty in measurement” – Propagation of distributions using a Monte Carlo method*, 2008. http://www.bipm.org/utils/common/documents/jcgm/JCGM_101_2008_E.pdf

76. D'un point de vue technique, le déroulement des équations est un peu plus complexe car il mobilise aussi la variance inconnue des résultats individuels de mesure. Cette variance est considérée comme un « paramètre de nuisance » et est éliminé par le biais d'une opération de « marginalisation ». Pour une description technique plus exhaustive, nous renvoyons à Lira, *Evaluating the Measurement Uncertainty*, *op. cit.*, pp.175-180, ou Lira & Wöger, « Comparison between the conventional and Bayesian approaches to evaluate measurement data », *op. cit.*

la vraisemblance est exprimée comme une *fonction* des valeurs vraies *potentielles* de la grandeur. Il n'est possible de mener à bien cette méthode que si l'expression de la vraisemblance est connue. Celle-ci peut être l'objet d'une *hypothèse*, à rapprocher de l'hypothèse (H3) du modèle fréquentiste⁷⁷.

Finalement, pour peu que le calcul mathématique soit réalisable, la méthode permet d'aboutir à l'expression de la distribution de probabilité de $\hat{\delta}$ en fonction des données expérimentales $\{d_i\}$, c'est-à-dire uniquement en fonction de paramètres connus. Il devient donc possible, en particulier, de déterminer une incertitude-type relative à D , définie comme la variance de la distribution de $\hat{\delta}$.

2.2.4 Formulation du résultat final

Dès lors que l'on substitue une approche bayésienne aux méthodes de type A, la distinction entre méthodes de « type A » et « type B » n'a plus de raison d'être : il ne subsiste plus alors qu'une seule et même méthode, bayésienne, applicable à tous les paramètres. Celle-ci procède en deux étapes. Dans tous les cas, l'expérimentateur construit une distribution *a priori*. Ensuite, selon les cas, l'expérimentateur dispose ou non d'informations expérimentales lui permettant d'actualiser ses connaissances par l'entremise de la formule de Bayes ; autrement, la procédure s'arrête à la formulation de la distribution *a priori*. Kacker et Jones synthétisent ainsi le processus :

[La fonction de mesure] $Y = f(X_1, \dots, X_n)$ est bien définie si l'on suit le point de vue bayésien suivant. Les évaluations de Type A sont des paramètres de distributions bayésiennes *a posteriori* et les évaluations de Type B sont des paramètres de distributions bayésiennes *a priori*. Les deux ont la même interprétation probabiliste.⁷⁸

La démarche bayésienne consiste donc finalement, étant donné une fonction de mesure générique :

$$Y = F(X_1, \dots, X_p, C_1, \dots, C_r) \quad (28)$$

à établir des distributions de probabilité \hat{x}_j pour toutes les grandeurs X_j intervenant dans le modèle, à partir de toutes les informations disponibles, que ces distributions soient des distributions *a priori* ou *a posteriori*. Dès lors, il est possible d'associer à Y une variable aléatoire \hat{y} dont la distribution de probabilité est entièrement déterminée par application de la fonction F , connaissant les \hat{x}_j : ce calcul est appelé « propagation des distributions »⁷⁹. *Dans ce cas, le résultat de mesure n'est plus donné sous la forme d'une valeur numérique mais d'une distribution de probabilité.*

77. Le statut de la fonction de vraisemblance, et l'interprétation des probabilités qu'elle engage, mériteraient une attention plus poussée, tant cette fonction semble se situer à l'interface des probabilités « objectives » et « subjectives ». Cependant, c'est une analyse que nous devons laisser de côté dans le présent article.

78. Kacker & Jones, « On use of Bayesian statistics to make the Guide to the Expression of Uncertainty Measurement consistent », *op. cit.*, p.243 (notre traduction).

79. JCGM, *Evaluation of measurement data – Supplement 1*, *op. cit.*, p.vii.

De cette distribution, l'expérimentateur extrait usuellement un couple (valeur adoptée, incertitude), ou, alternativement, un intervalle d'incertitude. Ces expressions, bien que moins complètes que les distributions de probabilité elles-mêmes, sont plus aisées à communiquer. Par ailleurs, le calcul de \hat{y} peut se révéler extrêmement fastidieux (voire même impossible analytiquement). La démarche la plus simple consiste alors à ne pas appliquer la « propagation des distributions », mais plutôt à retrouver la formule classique de propagation des incertitudes utilisée dans le GUM :

$$u_c^2 = u_1^2 + u_2^2 + \dots \quad (29)$$

après avoir calculé les incertitudes-types relatives à chacune des distributions des grandeurs d'entrée. La propagation des incertitudes s'applique alors de façon non problématique (en tant que propriété de l'algèbre des variables aléatoires), puisque toutes les probabilités employées sont cette fois-ci de même nature. Notons cependant que depuis l'avènement du calcul numérique, les méthodes de simulation numériques permettent de réaliser directement la « propagation des distributions » pour le calcul de \hat{y} .

L'intervalle d'incertitude bayésien est un « intervalle de crédibilité » associé à un degré de croyance p choisi (figure 3). La probabilité que la valeur vraie du mesurande soit contenue dans l'intervalle $[a, b]$ est p . C'est une *probabilité épistémique* qui fait état des valeurs que l'expérimentateur pense pouvoir raisonnablement attribuer au mesurande, étant données les informations dont il dispose. L'approche bayésienne globale déployée dans le supplément 1 du GUM est schématisée figure 4.

Les statuts respectifs des intervalles d'incertitude fréquentiste (intervalle de confiance) et bayésien (intervalle de crédibilité) sont l'objet d'un profond désaccord entre les tenants des deux approches. Nous développons les origines et les enjeux de ce désaccord dans la section suivante qui se prolongera par une discussion des retentissements philosophiques de l'évolution des méthodes statistiques appliquées au calcul des incertitudes que nous venons d'analyser.

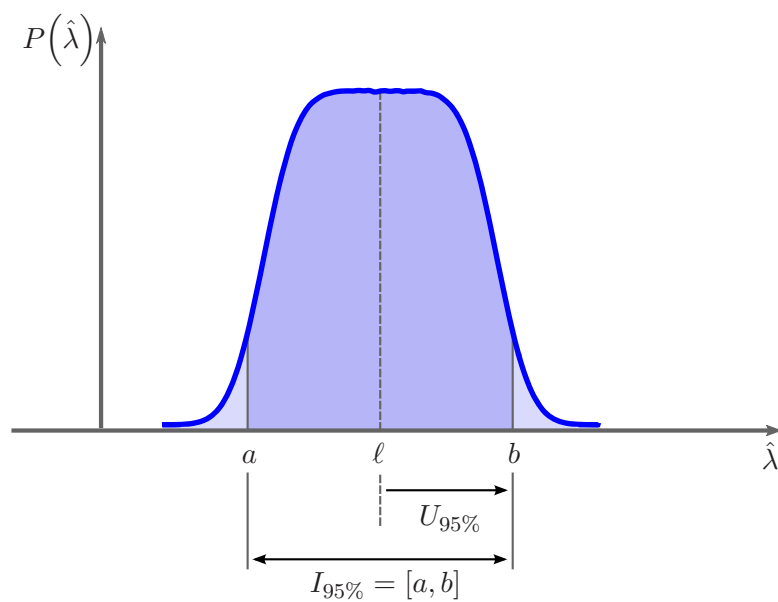


FIGURE 3 – Intervalle de crédibilité à 95% obtenu par la composition des distributions associées à $\hat{\delta}$ d'une part, et à $\hat{\lambda}_E$ d'autre part, correspondant à l'exemple (E) traité dans cet article. La courbe a été produite avec l'aide du logiciel « Uncertainty Machine » fourni par le National Institute of Standards and Technology (NIST, URL : <http://www.nist.gov/itl/sed/gsg/uncertainty.cfm>) en nous appuyant sur les valeurs données dans le GUM. Nous ne présentons ici que l'allure de la courbe.

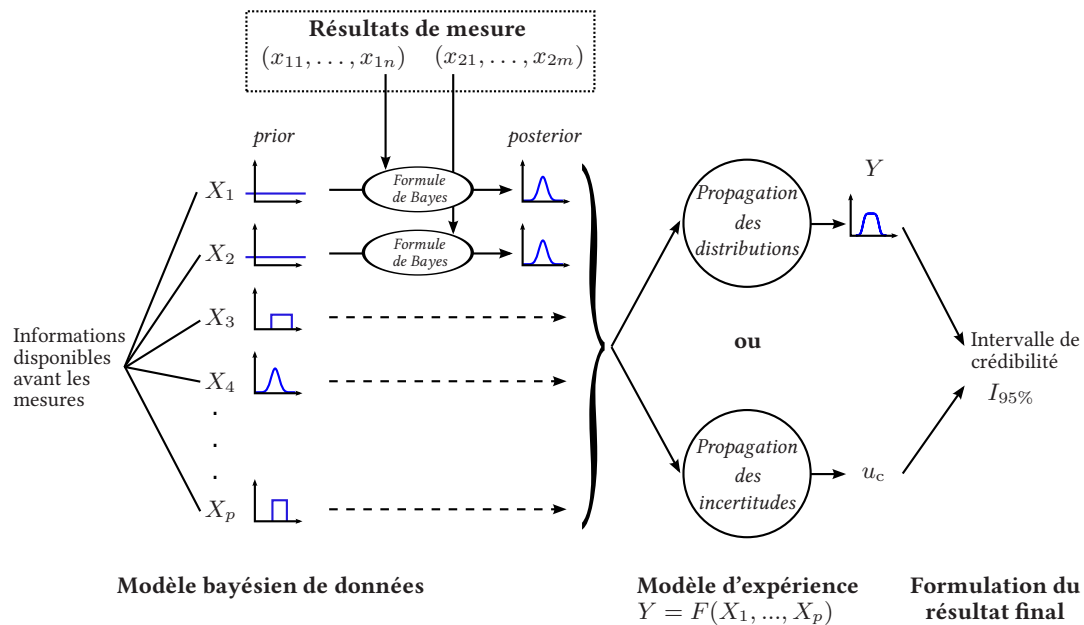


FIGURE 4 – Synthèse de l'approche bayésienne. L'expérimentateur part des informations disponibles, avant toute mesure, sur les différentes grandeurs d'entrée de la fonction de mesure. Il en tire les distributions *a priori* correspondantes, en utilisant une méthode à sa discrétion (par exemple le principe d'entropie maximum). Pour certaines des grandeurs d'entrée (X_1 et X_2 sur la figure), il disposera de résultats de mesure qui permettront d'actualiser les connaissances et produire des distribution *a posteriori*, en utilisant le théorème de Bayes. Pour les autres grandeurs (X_3 à X_p sur la figure), les distributions resteront inchangées. Ensuite, deux méthodes principales sont disponibles, selon qu'il dispose des moyens calculatoires de combiner directement les distributions (« propagation des distributions ») pour obtenir la distribution de Y ou qu'il extraie de chaque distribution un écart-type, puis qu'il combine ces écarts-types (« propagation des incertitudes ») pour obtenir une incertitude-type composée u_c . Dans tous les cas, il peut obtenir, ou reconstruire, un intervalle de crédibilité avec le seuil de crédibilité choisi – par exemple un intervalle de crédibilité à 95% – qui est l'expression du résultat final. Les formes des distributions sont données à titre d'exemple.

3 Discussion

Nous venons de montrer comment la métrologie a été le lieu d'une transition d'une approche traditionnelle des erreurs de mesure, fondée sur un traitement fréquentiste, à une approche épistémique, bayésienne. Dans la première approche, l'incertitude de mesure est reliée au taux de succès escompté de l'opération de mesure, celui-ci étant défini comme la capacité de cette dernière à effectivement produire des résultats qui s'approcheront au plus près de la « valeur vraie » de la grandeur mesurée. La seconde, en permettant le traitement probabiliste des erreurs systématiques, propose une approche unifiée en termes épistémiques des deux types d'erreurs dégagés dans la section 1. Elle conduit ainsi à relier l'incertitude de mesure à un degré de croyance – ou à un niveau de connaissance – relatif à la valeur du mesurande, étant donnés les informations disponibles et les résultats d'expérience.

Notre présentation des approches fréquentiste et bayésienne de la mesure en métrologie nous permet désormais de soulever plusieurs questions. Il est naturel, en premier lieu, de s'interroger sur les différences qui pourraient les séparer au niveau *pratique*. Les méthodes présentent-elles des différences qui vont au-delà des seules questions de *formulation* et d'*interprétation*? Ou bien sont-elles, au contraire, empiriquement équivalentes? Si, pour une même expérience, les deux méthodes aboutissaient aux mêmes conclusions, c'est-à-dire à une même valeur du mesurande, et une même incertitude associée, leur opposition semblerait alors se réduire à la seule querelle philosophique.

Si l'on entend comparer les résultats fournis, d'un côté, par la méthode fréquentiste exposée aux sections 2.1.3 et 2.1.4, et, de l'autre, par l'approche entièrement bayésienne présentée aux sections 2.2.3 et 2.2.4, l'on rencontre une difficulté. En effet, les termes dans lesquels s'expriment ces résultats ne sont pas les mêmes car ces résultats ne sont pas conceptualisés de la même manière : un résultat fréquentiste est exprimé sous la forme d'un énoncé d'exactitude qui ne peut pas aisément être comparé à un intervalle de crédibilité bayésien.

Si l'on se restreint au traitement des erreurs aléatoires, la comparaison devient possible et l'on peut montrer que les incertitudes fréquentiste et bayésienne ne s'accordent pas nécessairement mais convergent sous certaines conditions⁸⁰. Il s'agit là en partie de la conséquence d'une propriété bien connue, qui dépasse largement le cadre de la métrologie, stipulant que l'influence de la forme de la distribution de probabilité *a priori* devient de plus en plus ténue à mesure que la quantité de nouvelles informations disponibles augmente. C'est donc lorsque le nombre de mesures effectuées est faible que les méthodes sont vraiment susceptibles de diverger. Dans ce cas, l'approche fréquentiste devient difficilement applicable et certains métrologues considèrent qu'il s'agit là d'un avantage pratique qui milite en faveur de l'approche bayésienne ; en effet, il n'est souvent pas possible de réaliser un grand nombre de mesures (notamment dans le domaine industriel pour des raisons de temps et de coût). Cependant, le résultat de mesure

80. Kacker & Jones, « On use of Bayesian statistics to make the Guide to the Expression of Uncertainty Measurement consistent », *op. cit.*, p.235. Notons que Kacker et Jones utilisent pour leur part ce résultat afin de justifier la prévalence du modèle bayésien.

devient alors fortement dépendant du choix du *prior* – ce qui, pour certains, rend l’approche bayésienne particulièrement critiquable.

De nombreux articles techniques reviennent en détail sur différents points de désaccord qui montrent que les deux approches ne se distinguent pas seulement sur le plan de l’interprétation⁸¹. Le partage entre les deux méthodes constitue donc, au-delà des questions philosophiques, un enjeu scientifique et pratique avéré. Choisisant, pour notre part, de laisser ici de côté les questions qui relèvent du travail technique, nous tenterons dans ce qui suit de dégager certains des enjeux philosophiques soulevés par cette confrontation. En clarifiant les significations différentes qu’accorde chacune des approches à l’expression des résultats finaux de mesure, nous analyserons les enseignements philosophiques qui émergent de la solution épistémique proposée par les défenseurs de l’approche bayésienne en métrologie. Nous montrerons ensuite que cette solution épistémique n’est pas la seule envisageable en présentant l’approche fréquentiste « élargie » récemment avancée par un métrologue insatisfait de la solution bayésienne. Cette approche élargie nous servira d’instrument critique pour rappeler l’importance des conditions sociales de possibilité de l’établissement du savoir expérimental par la mesure.

3.1 L’approche bayésienne, une solution épistémique au problème des erreurs systématiques.

3.1.1 Signification des intervalles d’incertitude

Lors de notre description des approches fréquentiste et bayésienne en métrologie, nous nous sommes attachés à montrer que les deux interprétations différaient quant à la nature des objets sur lesquels portent les probabilités. Les probabilités utilisées dans l’approche fréquentiste caractérisent le processus de mesure, et portent de ce fait sur un élément du « monde extérieur ». Dans l’approche bayésienne, les probabilités portent sur la connaissance du sujet relative à la grandeur mesurée. C’est donc sans grand étonnement que l’on retrouve dans les débats soulevés par le calcul des incertitudes l’opposition classique entre probabilités objectives et subjectives. De fait, on retrouve, dans les débats internes à la métrologie, quantité d’objections semblables à celles que Mayo a formulées à l’encontre du subjectivisme bayésien⁸². En effet, dans son acception traditionnelle, l’objectif de la mesure physique est de fournir des informations sur le monde lui-même, et non sur un état personnel de connaissance ou de croyance d’un individu.

81. Kacker & Jones, « On use of Bayesian statistics to make the Guide to the Expression of Uncertainty Measurement consistent », *op. cit.*; I. Lira & W. Wöger, « Bayesian evaluation of the standard uncertainty and coverage probability in a simple measurement model », *Measurement Science and Technology*, 12, 2001, 1172-1179; Lira & Wöger, « Comparison between the conventional and Bayesian approaches to evaluate measurement data », *op. cit.*; Bich, « From Errors to Probability Density Functions », *op. cit.*; Willink, *Measurement Uncertainty and Probability*, *op. cit.*; R. Willink, « Difficulties arising from the representation of the measurand by a probability distribution », *Measurement Science and Technology*, 21, 2010, 1-11. De manière générale, la revue *Metrologia* présente régulièrement un grand nombre de débats sur le sujet.

82. Mayo, *Error and the Growth of Experimental Knowledge*, *op. cit.*, p.76. Il convient de préciser qu’à notre connaissance, les travaux de Mayo ne sont pas directement cités dans la littérature spécialisée de métrologie sur laquelle nous avons porté notre attention.

Le débat entre point de vue objectif et subjectif, tel qu'il se présente au sein de la métrologie, comporte cependant quelques spécificités. Il se cristallise autour de la question de savoir si une notion statistique, le « taux de succès », entendu comme mesure de la performance d'une méthode statistique, est pertinente ou non. Rappelons tout d'abord ce qu'est le « taux de succès » d'une procédure, déjà évoqué dans les développements techniques de la section 2. Nous avons vu que l'expression d'un résultat de mesure passe par l'écriture d'un *intervalle d'incertitude* associé à une probabilité p . Ce peut être un intervalle de confiance (p est alors le niveau de confiance), un intervalle de crédibilité (où p est le degré de croyance), ou encore un « intervalle élargi » hybride, à mi-chemin entre les deux approches – rappelons que c'est ce dernier choix qui est fait dans le GUM. Quelle que soit la nature de l'intervalle, deux possibilités se présentent et deux seulement : l'intervalle peut contenir la « valeur vraie » de la grandeur soumise à la mesure, ou ne pas la contenir. Autrement dit, le résultat obtenu peut être correct ou incorrect (figure 5).

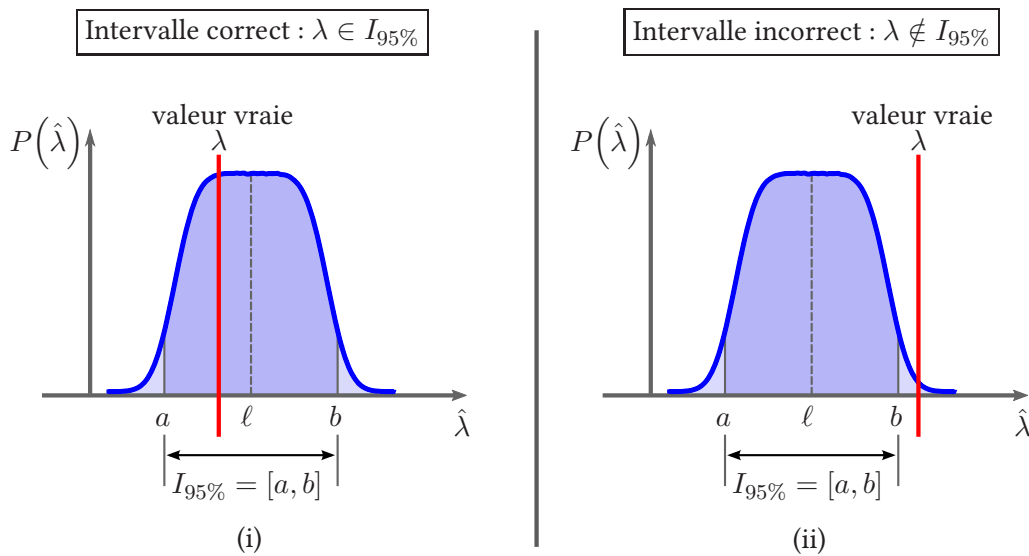


FIGURE 5 – Illustration du « taux de succès » d'une procédure. Pour un résultat donné, deux possibilités sont envisageables. (i) La valeur vraie est comprise dans l'intervalle d'incertitude obtenu. Dans ce cas, la procédure est considérée comme un succès. (ii) La valeur vraie est située hors de l'intervalle d'incertitude. Il s'agit alors d'un échec. Le « taux de succès » correspond à la fréquence limite avec laquelle le processus produirait un succès s'il était répété à l'infini.

Le taux de succès d'une procédure est défini comme la fréquence (limite) avec laquelle cette procédure produit des résultats corrects. Or, le métrologue et statisticien Robin Willink soutient que nous sommes en droit d'attendre que la probabilité p associée aux intervalles d'incertitude produits par la procédure corresponde au taux de succès de cette dernière (ou le dépasse), sans quoi elle peut être considérée comme défectueuse. À ce titre, il considère, d'une part, le taux de succès d'une procédure comme la mesure de sa performance *effective*; et, d'autre part, la probabilité p associée à un intervalle d'incertitude comme une affirmation sur

la performance *attendue* de la procédure. Comme nous allons le voir, le taux de succès est le nœud d'un désaccord profond entre statisticiens fréquentistes et bayésiens.

L'exigence de Willink se heurte certes à une difficulté majeure, puisqu'il n'est jamais possible de garantir qu'un résultat est correct (ou incorrect), la valeur cible n'étant jamais parfaitement connue – ce qui est justement la raison pour laquelle il est jugé nécessaire de construire un intervalle d'incertitude. Il est donc également impossible de mesurer le taux de succès effectif d'une procédure. Willink concède ainsi que, dans sa propre conception de l'approche fréquentiste, celle-ci, bien qu'objectiviste, aboutit également à la formation de croyances⁸³. Mais, rétorque Willink, il demeure possible de calculer un taux de succès théorique en raisonnant en amont sur la procédure et la méthode employées⁸⁴ (par des simulations numériques, ou en testant celle-ci sur des cibles connues avec une bonne précision). Il affirme ainsi que la qualité du résultat n'est pas attachée à la comparaison, impossible, entre le résultat obtenu et la valeur vraie, mais à la possibilité de contrôler de façon indépendante la fiabilité de la procédure. Le taux de succès devient par là l'outil nécessaire à toute analyse d'incertitude digne d'intérêt, car il apporte l'assurance que nous pouvons avoir confiance, à un degré bien identifié, dans les résultats que nous produisons – si bien que le contrôle du taux de succès garantit l'objectivité de l'entreprise de mesure. Willink admet et revendique qu'il suspend ainsi l'acceptabilité d'une méthode à un critère fréquentiste⁸⁵. Le pivot de la position de Willink est alors le postulat suivant : un expérimentateur peut avoir dans le résultat qu'il a effectivement obtenu la *même* confiance que celle qu'il accorde à la procédure qu'il emploie pour produire ses résultats :

La confiance que nous pouvons placer, après la mesure, dans l'idée que nous *avons réussi* est la même que celle que nous avons placée, avant la mesure, dans l'idée que nous *allions réussir*.⁸⁶

Willink parle de correspondance entre une « assurance pré-mesure » et une « assurance post-mesure », et souligne, en filigrane, une différence de perspective avec l'approche bayésienne subjectiviste. Cette dernière est centrée sur le présent : elle vise la meilleure utilisation possible

83. Willink n'admet pas pour autant une interprétation subjective des probabilités. Les probabilités induisent des degrés de croyance, mais ne sont pas des degrés de croyance : « bien qu'une probabilité implique un degré de croyance, un degré de croyance n'implique pas une probabilité », Willink, *Measurement Uncertainty and Probability*, *op. cit.*, p.38. « (C)e n'est pas le concept de croyance qui distingue les statistiques bayésiennes des statistiques fréquentistes. C'est plutôt la nature des objets sur lesquels porte cette croyance. », R. Willink, « Probability, belief and success rate : comments on 'On the meaning of coverage probabilities' », *Metrologia*, 47, 2010, 343-346, p.344, note 1.

84. Willink, *Measurement Uncertainty and Probability*, *op. cit.*, p.50.

85. Willink affirme en particulier que seule la notion d'espérance statistique permet de comprendre la façon dont on calcule les probabilités attachées à des degrés de croyance en s'appuyant sur la propension à s'engager dans un pari. Or, il rattache l'espérance statistique à l'interprétation fréquentiste au travers de la loi faible des grands nombres. C'est pourquoi l'interprétation épistémique lui semble en définitive dépendre de l'interprétation fréquentiste. Willink, « Probability, belief and success rate : comments on 'On the meaning of coverage probabilities' », *op. cit.*, p.344.

86. Willink, *Measurement Uncertainty and Probability*, *op. cit.*, p.51 (notre traduction). Il nous semble - sans prendre le temps de développer plus avant cette idée - que la position de Willink, et son idée selon laquelle des probabilités objectives peuvent induire un degré de croyance, peut être rapprochée de l'interprétation propensionniste des probabilités appliquée aux événements singuliers, décrite, par exemple, dans D. Gillies, *Philosophical Theories of Probabilities*, Routledge, 2000, pp.119-126.

des informations disponibles au moment présent, qui permettent d'actualiser un état antérieur de connaissances. L'approche fréquentiste est quant à elle tournée vers le futur : elle vise à produire un résultat dont on a de bonnes raisons de penser qu'il sera effectivement le meilleur possible ; de ce fait, ce résultat engage un test futur qui le rend falsifiable.

Le taux de succès est ainsi présenté comme le garant de l'objectivité de l'approche fréquentiste. Dans la continuité, les critiques de l'approche bayésienne mettent en avant le caractère subjectif et personnel attribué par cette dernière aux probabilités, et donc aux résultats de mesure. Selon Willink, l'intervalle de crédibilité bayésien ne peut pas être confronté à l'expérience, puisque les probabilités impliquées portent sur l'état de croyance de l'expérimentateur et non sur l'objet mesuré lui-même⁸⁷. Là réside le point d'achoppement de la méthode bayésienne dans son application à la mesure : que représentent, par rapport à la réalité, les croyances ainsi formulées, si ce n'est justement un simple état de croyance ? On pourrait aller jusqu'à dire qu'une probabilité bayésienne intervenant dans l'expression d'un résultat de mesure n'est par nature jamais erronée : quelle que soit effectivement la probabilité associée à une valeur ou à un intervalle, celle-ci fait bel et bien état de la croyance de l'expérimentateur (que cette dernière soit erronée ou non). Il est, dès lors, impossible de soumettre le résultat bayésien à un test empirique. Pour un défenseur de l'approche objective tel que Willink, ce constat est inacceptable. Il pourrait être écarté si l'on admettait de nouveau que les intervalles produits ont pour but de contenir la valeur de la grandeur soumise à la mesure, et que le critère de performance de la méthode bayésienne est fondé sur son taux de succès. Cependant, certains bayésiens ne tiennent pas cette exigence pour nécessaire⁸⁸, d'autant plus qu'elle conduirait à subordonner les méthodes bayésiennes à une interprétation fréquentiste.

Les défenseurs de l'approche bayésienne en métrologie répondent à ces objections en mettant en avant que la subjectivité est inhérente aux résultats de mesure⁸⁹. Ils font remarquer que l'approche fréquentiste incorpore elle-même de nombreux éléments de subjectivité à tous les niveaux de l'expérience, tant au niveau du choix des hypothèses que de la construction des « limites crédibles » d'erreur, mais que cette subjectivité est masquée⁹⁰. Ils contestent de plus que la subjectivité soit une faille de l'approche bayésienne. Face aux fortes critiques exprimées par un certain nombre de fréquentistes à l'égard des distributions de probabilité *a priori*⁹¹, spé-

87. Willink, « Measurement Uncertainty and Probability », *op. cit.*, p.38 puis p.234. Cela rejoint ce qu'affirme également Mayo : « il n'y a aucune raison de supposer qu'il existe un degré de croyance correct à entretenir », Mayo, *Error and the Growth of Experimental Knowledge*, *op. cit.*, p.83.

88. I. Lira, « On the meaning of coverage probabilities », *Metrologia*, 46, 2009, 616-618.

89. « L'attitude [objectiviste] est basée sur l'illusion que la subjectivité peut être totalement évitée dans la mesure, alors qu'elle en imprègne une grande partie. », Bich, « From Errors to Probability Density Functions. Evolution of the Concept of Measurement Uncertainty », *op. cit.*, pp.2155-2156.

90. Voir la remarque de Mayo citant Good : « Les bayésiens objecteront qu'il est impossible d'exclure les opinions, et que les bayésiens ont au moins le mérite de mettre ces dernières en évidence au lieu de les 'cacher sous le tapis', pour paraphraser I. J. Good », Mayo, *Error and the Growth of Experimental Knowledge*, *op. cit.*, p.82 (notre traduction). Voir aussi D'Agostini : « Ces méthodes qui sont annoncées comme étant 'objectives' tendent en réalité à cacher les hypothèses sur lesquelles elles sont fondées. », D'Agostini, *Bayesian Reasoning in Data Analysis*, World Scientific, 2003, p.79 (notre traduction).

91. Voir par exemple les critiques formulées par Mayo, *Error and the Growth of Experimental Knowledge*, *op.*

cifiques à l'approche bayésienne, des bayésiens comme Lira et Wöger réaffirment la légitimité d'une description subjectiviste des jugements probabilistes :

La détermination du *prior* est peut-être le point le plus controversé des statistiques bayésiennes. Les fréquentistes défendent que cette attribution, étant 'subjective', assujettit l'ensemble du processus d'inférence aux caprices de la personne qui effectue l'évaluation. Cela serait en effet le cas si cette attribution était effectuée arbitrairement.⁹²

L'approche bayésienne ne laisse pas l'expression d'un résultat à la fantaisie de l'expérimentateur mais vise à ce qu'il exploite *de la meilleure façon possible* l'ensemble des informations dont il dispose, en lui permettant de toutes les intégrer à une évaluation probabiliste de son état de connaissance.

En dépit de la profusion d'arguments visant à défendre ou à critiquer l'une ou l'autre des approches, il ne nous semble pas que les arguments présentés parviennent véritablement à les départager. De fait, on constate que le débat est fondé sur l'adoption ou non d'un critère de performance, qui apparaît comme un choix de *principe* philosophique et méthodologique, qui, par conséquent, ne peut pas être justifié lui-même. Les arguments fournis constituent alors plutôt une somme de remarques critiques qui viennent nourrir la réflexion sur les différentes pratiques mises en jeu et redéfinir les cadres d'action de ces dernières. Par ailleurs, ce débat ne se limite pas à une simple querelle de statisticiens ; il possède des ressorts spécifiques à la mesure. La notion de « taux de succès » véhicule une valeur épistémologique forte car elle est intimement attachée aux concepts de « valeur vraie », d'« erreur de mesure », et ainsi d'« exactitude de mesure ». Or, parallèlement à la poussée des statistiques bayésiennes en métrologie, l'on constate que les textes internationaux comme le GUM laissent de côté l'exactitude de mesure pour se focaliser sur une conception pleinement épistémique de l'« incertitude de mesure ». Cette dernière n'est pas directement attachée à la notion de valeur vraie d'une grandeur, mais décrit la connaissance exprimée du sujet, indépendamment de son statut de vérité. Nous explicitons cet aspect dans la sous-section suivante.

3.1.2 Un « tournant épistémique » en métrologie

Notre analyse du GUM s'est jusqu'à présent focalisée sur des questions liées à l'interprétation des probabilités. Or, le succès des méthodes bayésiennes s'accompagne d'une évolution plus générale de la conception de la mesure en métrologie. Cette évolution pourrait être qualifiée de « tournant épistémique », tant elle témoigne d'un glissement d'intérêt entre deux objectifs : décrire « ce qui est », et décrire un « état de connaissance » à propos de « ce qui est ».

cit., pp.75-83 et pp.102-127. Une réponse classique consiste à renvoyer au principe d'entropie maximum, interprété comme la procédure de choix la plus objective possible pour rendre compte de l'ensemble des informations disponibles, sans créer artificiellement d'information supplémentaire. Cependant, certains bayésiens contestent la nécessité d'imposer une telle contrainte et défendent l'intérêt d'une expression (rationnelle) libre des *prior*. Voir Howson & Urbach, *Scientific Reasoning : The Bayesian Approach*, *op. cit.*, pp.276-288, ou encore D'Agostini, *Bayesian Reasoning in Data Analysis*, *op. cit.*, p.81.

92. Lira & Wöger, « Comparison between the conventional and Bayesian approaches to evaluate measurement data », *op. cit.*, p.S255 (notre traduction).

Le premier constat est celui d'un déplacement de la visée de la mesure : si l'investigation expérimentale porte toujours sur les objets du monde extérieur, l'accent est mis sur la connaissance que nous pouvons avoir de ces objets. Mari considère que, de ce point de vue, les métrologues prennent leur distance à l'égard de la perspective métaphysique des scientifiques du XIX^e siècle⁹³. Ils comprennent avant tout la mesure comme une opération *utile*, servant à atteindre un objectif concret bien défini : « les mesures sont les résultats d'opérations reconnues comme adéquates par rapport à leur objectif, à savoir d'obtenir des informations sur les objets mesurés »⁹⁴. Dès lors, la mesure n'est plus considérée comme le moyen d'accéder à « ce qui est », mais comme un outil qui permet d'exprimer l'état de notre connaissance sur une situation que nous avons délimitée en fonction de nos objectifs à un instant donné. Giordani et Mari voient dans ce déplacement l'écho du mouvement analogue qui se produit en philosophie et en sociologie des sciences dans le courant du XX^e siècle⁹⁵. Cependant, l'influence grandissante du rôle de la mesure dans des domaines appliqués tournés vers la technologie nous semble avoir contribué à ce déplacement de façon autrement plus décisive. Les aspects décisionnels et pratiques occupent, dans ces domaines, une place essentielle, et l'utilisation des résultats expérimentaux par un opérateur devient un enjeu central. Nous pensons que l'orientation tout à la fois épistémique et pragmatique prise par la métrologie doit davantage à la transformation des objectifs liés à l'activité de mesure et à l'intensification des enjeux de la métrologie dans le monde technique⁹⁶ qu'au rayonnement des travaux de sociologie et de philosophie comme le suggèrent parfois les deux auteurs. De ce point de vue, le tournant épistémique de la métrologie résulterait de l'exacerbation de l'activité technique et non de questionnements philosophiques qu'il faudrait bien plutôt considérer comme une conséquence de ce tournant.

Les métrologues ont identifié certaines caractéristiques de ce que nous avons appelé « tournant épistémique » et ont cherché à les comprendre. Ainsi font-ils état, dans le VIM, d'une transition qui se serait opérée entre une « approche erreur » et une « approche incertitude » de la mesure :

Le changement dans le traitement de l'incertitude de mesure, d'une approche « erreur » (quelquefois appelée approche traditionnelle ou approche de la valeur vraie) à une approche « incertitude », a conduit à reconsidérer certains des concepts correspondants qui figuraient dans la deuxième édition du VIM. L'objectif des mesurages dans l'approche « erreur » est de déterminer une estimation de la valeur vraie qui soit aussi proche que possible de cette valeur vraie unique. [...] En général il est seulement possible d'estimer une limite supérieure de la valeur absolue de l'erreur totale, appelée parfois abusivement « incertitude ». L'objectif des mesurages dans l'approche « incertitude » n'est pas de déterminer une valeur vraie le mieux possible. On suppose plutôt que l'information obtenue lors d'un mesurage

93. L. Mari, « Epistemology of measurement », *Measurement*, 34, 2003, 17-30, p.21 (notre traduction).

94. *Ibid.*, p.24. Mari, « Epistemology of measurement », *op. cit.*, p.24.

95. A. Giordani & L. Mari, « Measurement, Models, and Uncertainty », *IEEE transactions on instrumentation and measurement*, 61, 2012, 2144-2152, p.2145.

96. Quinn, « Metrology, its role in today's world », *op. cit.*

permet seulement d'attribuer au mesurande un intervalle de valeurs raisonnables, en supposant que le mesurage a été effectué correctement.⁹⁷

Dans l'approche dite « erreur », le résultat est décrit en termes d'erreurs *de mesure* et le processus en termes d'« exactitude »⁹⁸, où l'exactitude caractérise la propension du processus à aboutir à une faible erreur de mesure finale. Ces deux concepts renvoient à une adéquation entre la visée de la mesure et le résultat obtenu. Dans l'approche « incertitude », qui vient se substituer à l'approche « erreur », ces deux concepts ne sont pas nécessairement irrecevables, mais ils ne sont plus utilisés. L'accent est plutôt mis sur l'« incertitude de mesure », que l'on tient pour un concept épistémique décrivant un état de connaissances, en harmonie avec l'interprétation épistémique des probabilités. Ce glissement est particulièrement sensible dans le GUM, où le contraste avec la notion d'erreur est clairement marqué :

Le concept d'incertitude comme attribut quantifiable est relativement nouveau dans l'histoire de la mesure bien que l'erreur et l'analyse des erreurs soient des concepts depuis longtemps pratiqués dans la science de la mesure, c'est-à-dire en métrologie.⁹⁹

L'incertitude de mesure est généralement décrite comme caractérisant le doute que l'expérimentateur entretient à l'égard du résultat qu'il a obtenu, mais il convient de noter que cette notion n'est calculée qu'à partir d'une analyse des erreurs que l'expérimentateur a effectivement identifiées. Il est tout à fait possible que l'incertitude de mesure évaluée se révèle à terme largement inférieure à l'erreur effectivement commise, en particulier si certaines sources d'erreur systématiques sont inconnues – la découverte de l'effet Josephson en 1962 et son application à la mesure de précision en physique conduisit à mesurer des valeurs des constantes physiques significativement différentes de celles obtenues jusqu'alors, ce qui fut interprété comme la preuve de l'existence d'une erreur systématique non identifiée dans les résultats antérieurs¹⁰⁰. L'expérimentateur pourrait alors être tenté de systématiquement étirer l'incertitude estimée, de façon artificielle, afin de limiter un tel risque et garantir ainsi un résultat plus « sûr ». Cependant, il s'agit là d'une attitude « sécuritaire » qui ne correspond pas au rôle de l'incertitude de mesure¹⁰¹. Celle-ci doit être déterminée à partir de toutes les informations dont l'expérimentateur dispose, et uniquement de celles-ci. En ce sens, comme l'exprime Bich, par exemple, elle

97. JCGM, *Vocabulaire International de Métrologie*, *op. cit.*, p.x.

98. L'article de 1962 de Churchill Eisenhart, que l'on peut légitimement considérer comme l'une des formes les plus poussées du fréquentisme traditionnel en métrologie, est centré autour du concept d'« exactitude » ; celui d'« incertitude de mesure » n'y figure pas. Eisenhart, « Realistic Evaluation of the Precision and Accuracy of Instruments Calibration Systems », *op. cit.*

99. JCGM, *Guide pour l'expression de l'incertitude de mesure*, *op. cit.*, p.vii.

100. Taylor, Parker et Langenberg discutent de la conséquence de la découverte de cet effet sur la mesure de la constante e/h , rapport de la charge élémentaire et de la constante de Planck. B.N. Taylor & W.H. Parker & D.N. Langenberg, « Determination of e/h , Using Macroscopic Quantum Phase Coherence in Superconductors : Implications for Quantum Electrodynamics and the Fundamental Physical Constants », *J. Phys. Chem. Ref. Data*, 41, 1969, 375-496.

101. Une telle vision « sécuritaire » de l'incertitude a été critiquée à plusieurs occasions dans les comptes rendus portant sur l'ajustement des constantes de la physique. Voir par exemple E.R. Cohen & J.W.M. DuMond, « Our Knowledge of the Fundamental Constants of Physics and Chemistry in 1965 », *Reviews of Modern Physics*, 37, 1965, 537-594, pp.540-541.

« peut être vue comme la réciproque logique de l'état de connaissance ». ¹⁰² C'est en cela que l'incertitude de mesure se distingue nettement du concept d'« exactitude », qui correspond à un objectif d'encadrement de l'erreur, et ne saurait jamais être qu'imparfaitement *estimée*. C'est pourquoi, en reprenant à notre compte, en les adaptant, certaines remarques formulées par Willink ¹⁰³, l'on devrait distinguer en toute rigueur *l'estimation* d'une « exactitude » inconnue et approchée, concept central d'une formulation *métaphysique* de la mesure, de la *détermination* d'une « incertitude » connue, concept central d'une caractérisation *épistémologique* de la mesure. Le remplacement progressif, en métrologie, de l'exactitude par l'incertitude dans l'expression des résultats de mesure (ou encore de l'analyse d'erreur par l'analyse d'incertitude) résulte du refus de plus en plus marqué de faire appel à des entités que l'on ne pourra jamais connaître qu'imparfaitement, comme en témoigne cette remarque de Bradley en 1982 :

On apprécie l'absurdité qu'il y a à présumer de la valeur exacte de la grandeur à mesurer, et pourtant nous continuons à utiliser le mot précision. En fait, ce que l'on tente de faire est d'exprimer ensemble la valeur observée et la finesse des limites atteintes. Ainsi exprimons-nous la crédibilité de nos conclusions plutôt que leur précision. ¹⁰⁴

La différence entre les deux conceptions de la mesure, appelées ici « métaphysique » et « épistémologique », peut également être saisie en contrastant la formulation, traditionnelle, du National Physical Laboratory :

[Le signe +/-] indique la plage de valeurs du résultat final dans laquelle la vraie valeur de la quantité mesurée est censée se trouver (is believed to lie). ¹⁰⁵

avec celle que l'on trouve dans le GUM :

L'incertitude de mesure peut ainsi être décrite comme la mesure de la façon dont on croit que l'on connaît la valeur vraie essentiellement unique du mesurande. ¹⁰⁶

Quelle est la conséquence de ce « tournant épistémique » en métrologie ? L'exigence d'exactitude tend à être remplacée par une exigence de rationalité, laquelle guide l'établissement d'un accord intersubjectif qui justifie nos prises de décision. Quinn, notamment, a souligné le danger que comporte un tel glissement et insiste sur l'importance de la notion d'exactitude ¹⁰⁷ – nous y reviendrons plus loin. Selon Mari, cette inquiétude manifeste une tension entre deux conceptions de la science : la science considérée comme idéal de représentation d'une réalité indépendante de l'esprit, et la science considérée comme activité instrumentale orientée vers une application ou un but concret donné ¹⁰⁸ :

102. Bich, « From Errors to Probability Density Functions », *op. cit.*, p.2155.

103. Willink, *Measurement Uncertainty and Probability*, *op. cit.*, p.xvii.

104. D. Bradley, « Expressing Doubt about Uncertainty ? », *European Journal of Physics*, 3, 1982, 58-59, p.58 (notre traduction).

105. Campion & Burns & Williams, *A code of practice for the detailed statement of accuracy*, *op. cit.*, p.1 (notre traduction).

106. JCGM, *Evaluation of measurement data – An introduction*, *op. cit.*, p.3.

107. Quinn, « Metrology, its role in today's world », *op. cit.*, pp.11-13.

108. C'est une tension que l'on retrouve dans les discussions autour de la philosophie des « technosciences », voir, notamment, B. Bensaude-Vincent & al., « Matters of Interest : The Objects of Research in Science and Technoscience », *Journal for General Philosophy of Science*, 42, 2011, 365-383.

l'ontologie et le critère de vérité ont été remplacés par l'information et un critère d'adéquation [...] (L)a mesure est-elle alors juste une évaluation dont les résultats sont subjectivement adéquats à des objectifs donnés? ¹⁰⁹

On peut se demander s'il est possible d'articuler ces deux conceptions de la science. En particulier, dans quelle mesure l'« incertitude de mesure », interprétée comme description d'un niveau de connaissance, peut-elle effectivement correspondre à un constat d'exactitude, c'est-à-dire à une évaluation de l'erreur de mesure finale? Tal a soulevé cette question sur un plan philosophique à partir d'une étude de la définition de la seconde, et a suggéré que la possibilité d'effectuer la jonction entre incertitude et exactitude était suspendue à une « condition de robustesse » ¹¹⁰. Nous n'aborderons pas ce problème touchant au réalisme scientifique de front, mais de façon indirecte, en nous appuyant sur une alternative fréquentiste à l'approche bayésienne que nous présentons dans la section suivante.

3.2 Extension du modèle classique : priorité du social sur l'épistémique

L'un des attraits exercé par l'approche bayésienne tient à son traitement homogène des composantes aléatoires et systématiques de l'incertitude de mesure. Willink conteste que cet avantage soit l'apanage des statistiques bayésiennes; il avance qu'il est tout aussi possible d'harmoniser les deux composantes de l'incertitude en restant fidèle à la perspective fréquentiste. Selon lui, cette dernière est à même de fournir un traitement des erreurs systématiques, à condition de procéder à une extension de l'approche traditionnelle. Bien que la proposition de Willink ne paraisse pas avoir attiré l'attention d'un grand nombre de métrologues jusqu'à présent, elle présente à nos yeux un intérêt philosophique, car les aménagements que ce chercheur néozélandais suggère pour parvenir à un traitement unifié des incertitudes dans le cadre fréquentiste nous semblent explorer une voie toute différente de la voie privilégiée par l'approche bayésienne : une voie qui met en relief non pas la dimension épistémique mais la dimension sociale des pratiques de mesure. Nous verrons que cette autre voie suggère, quant à elle, une réponse négative à la question posée par Mari.

3.2.1 Une approche fréquentiste élargie

Comme nous l'avons vu, les erreurs systématiques demeurent fixes lors de la répétition des mesurages ¹¹¹. Elles ne semblent donc pas pouvoir se prêter à une modélisation probabiliste de type fréquentiste. De ce point de vue, Willink tombe d'accord avec les bayésiens pour dire que l'approche fréquentiste traditionnelle est adaptée aux problèmes d'inférence qui se présentent couramment dans les sciences économiques et sociales, le commerce, la démographie, la santé publique (où l'on dispose de populations identifiables et d'un grand nombre de données) mais qu'elle est trop limitée pour s'appliquer au domaine des mesures physiques où l'existence et le

109. Mari, « Epistemology of measurement », *op. cit.*, p.25 (notre traduction).

110. E. Tal, « How Accurate is the Standard Second? », *Philosophy of Science*, 78, 2011, 1082-1096.

111. Comme nous l'avons vu dans l'exemple traité section 2.3, on compte également parmi les erreurs systématiques les écarts à la valeur cible qui, tels les effets de la dilatation thermique du calibre, varient au cours du mesurage selon une loi connue et qui peuvent donc être corrigées.

traitement des erreurs systématiques revêtent une importance capitale.

Willink n'en tire pas pour autant la conclusion que le domaine des mesures physiques doit être abandonné à un traitement épistémique et par là réduit à la sphère intersubjective. Une raison parmi d'autres, sur laquelle il revient à de nombreuses reprises dans les pages de la revue *Metrologia*, tient à ce que l'interprétation de l'intervalle de crédibilité couramment avancée par les bayésiens ne fournit aucun moyen d'en tester la validité. La notion de validité doit bien sûr être entendue ici dans son sens fréquentiste qui, selon Willink, est le seul à lui donner un contenu pratique en renseignant l'utilisateur du résultat sur la fiabilité de la procédure qui a été suivie pour obtenir l'intervalle – la fiabilité s'exprimant (comme nous l'avons vu dans la section précédente) en terme de critère de performance ou de taux de succès de la procédure. Garantir aux utilisateurs la fiabilité des résultats qui leur sont communiqués suppose donc d'adapter l'approche fréquentiste aux activités de mesurage telles qu'elles se déroulent dans la pratique et, en particulier, au traitement des erreurs systématiques – du moins, si l'on endosse l'objectif du GUM d'exprimer l'incertitude associée à un résultat de mesure par un seul terme.

L'étape essentielle de cette extension consiste à adopter une conception élargie du processus de mesure. Celui-ci est désormais appelé à comprendre, en plus de l'expérience de mesurage elle-même, toutes les activités antérieures, situées à l'arrière-plan, dans d'autres laboratoires ou sites industriels, au cours desquelles les erreurs systématiques impliquées dans le mesurage ont été engendrées : acquisition et étalonnage des instruments, calcul des meilleures estimations des constantes physiques utilisées dans le calcul du résultat de mesure, adoption des valeurs de référence, qualification du personnel ¹¹².

Cet élargissement conduit à reconnaître que les informations figurant dans l'analyse d'une expérience de mesure peuvent être envisagées selon différents points de vue, selon que l'on est producteur ou seulement utilisateur. Willink attire ainsi l'attention sur le fait que l'univers de référence d'un usager qui utilise des informations relatives aux erreurs systématiques des instruments qu'il a acheté, ou qui utilise, en bout de chaîne, des résultats de mesure, n'est pas le même que celui des laboratoires ou groupes industriels qui lui fournissent ces informations ou résultats. Cette remarque est à l'origine d'un second élargissement qui concerne cette fois la notion de répétition des mesurages dont la conception orthodoxe, très étroite, limite l'application de l'approche fréquentiste. Ce second élargissement consiste à introduire un taux de succès « pratique » reflétant le contexte d'usage des résultats de mesure. Le calcul du taux de succès pratique (que nous ne détaillerons pas ici) suppose cependant que l'utilisateur puisse choisir ses fournisseurs parmi une variété suffisamment large de laboratoires proposant le même type de prestation, mais travaillant avec des techniques et dans des contextes différents.

Si cette condition est remplie, un laboratoire utilisateur pourra traiter l'erreur systématique fixe, qui intervient dans ses mesurages du fait de son choix de fournisseur, comme la réalisation d'un tirage dans la distribution mère des différentes erreurs systématiques produites par l'ensemble des laboratoires qui lui étaient accessibles au départ. Du point de vue du laboratoire

112. Willink, *Measurement Uncertainty and Probability*, op. cit., tableau 4.1, p.61.

utilisateur situé en bout de chaîne, les erreurs systématiques auxquelles il a affaire pourront ainsi, selon Willink, être traitées de façon probabiliste. Dans ce cas, le taux de succès n'est pas un taux de succès calculé à partir de mesurages répétés à l'identique selon une procédure unique, mais un taux de succès « pratique » déterminé « par rapport à un grand nombre de problèmes potentiels où chaque erreur systématique peut prendre des valeurs différentes »¹¹³.

L'adoption d'une conception élargie du processus de mesure débouche donc sur une révision de la notion de répétition des mesurages. Elle permet à un laboratoire utilisateur situé en bout de chaîne de traiter les erreurs systématiques dans le cadre d'une approche fréquentiste assouplie, davantage en harmonie avec les conditions d'application pratiques que l'approche orthodoxe exposée dans les manuels. Un raisonnement similaire peut s'appliquer à la valeur d'un mesurande dont un laboratoire fournit l'estimation à un usager. On peut considérer que l'utilisateur d'un résultat de mesure est alors concerné par un taux de succès relatif à l'ensemble des mesurages qui sont susceptibles d'être effectués par les laboratoires auxquels il a accès et qui peuvent résoudre le type de problème de mesure qui l'intéresse. Du point de vue de l'utilisateur, la valeur du mesurande n'apparaît alors plus comme une valeur inconnue et fixe « isolée de tout contexte plus vaste »¹¹⁴ mais comme la réalisation d'une variable aléatoire décrivant la distribution des fréquences des valeurs cibles susceptibles d'être réalisées par les différents laboratoires fournisseurs auxquels il a potentiellement accès pour le type de problème de mesure concerné. Willink fait remarquer que, dans cette approche fréquentiste « pratique », la valeur du mesurande visé en vient donc à être traitée comme une variable aléatoire mais d'une tout autre façon que dans le cadre bayésien. L'intervalle de confiance pertinent (du point de vue de l'usager) est alors un intervalle de confiance « moyen »¹¹⁵.

3.2.2 Dimension sociale de l'approche fréquentiste élargie

Nous avons présenté les grandes lignes de la solution avancée par Willink parce qu'elle nous semble tenter de répondre aux objectifs du GUM en mettant en avant, non pas les aspects épistémiques des activités de mesure, mais leur ancrage social. Ce faisant, cette proposition invite à aborder le débat traditionnel entre fréquentistes et bayésiens sous un nouvel angle. Avant de la confronter à l'approche bayésienne, précisons donc comment les aménagements de l'approche fréquentiste qui viennent d'être présentés peuvent, selon nous, se rapporter au caractère collectif des pratiques de mesure, et des activités scientifiques et techniques en général.

Inclure les opérations situées en amont de l'expérience de mesurage proprement dite dans une conception élargie du processus de mesure revient, en effet, à prendre explicitement en considération le fait que tout mesurage appelle à utiliser des résultats obtenus par d'autres¹¹⁶.

113. Willink, *Measurement Uncertainty and Probability*, *op. cit.*, p.71.

114. *Ibid.*, pp.72-73.

115. *Ibid.*, p.73. Nous ne visons ici qu'à donner l'idée directrice de la solution proposée par Willink. Le lecteur intéressé pourra se reporter aux chapitres 4 et 5 de son ouvrage *Measurement Uncertainty and Probability*.

116. L'utilisation d'une simple règle graduée engage déjà, à l'arrière-plan, tout l'ensemble de mesurages et d'étalonnages qu'il a fallu réaliser pour la fabriquer. On peut toujours, bien sûr, utiliser des étalons naturels, main, doigt,

Si l'on adopte le vocabulaire des épistémologues anglo-saxons, cela revient à prendre acte de la « dépendance épistémique »¹¹⁷ dans laquelle se trouvent les membres des sociétés modernes les uns à l'égard des autres pour tout ce qui relève de la connaissance et de l'utilisation des quantités physiques. Il en résulte un élargissement de l'expérience à l'ensemble des acteurs, des équipements et des infrastructures directement et indirectement mobilisés dans les activités de mesurage qui conduit à faire intervenir, au sein même de l'analyse scientifique, des considérations sociales liées à la division du travail : la différence de point de vue des acteurs qui doivent entrer en relation pour produire un résultat de mesure et qui sont situés dans des cadres professionnels et institutionnels distincts se traduit concrètement au niveau de la conceptualisation (l'univers de référence) et du choix des approximations pertinentes (taux de succès pratique, intervalle de confiance moyen) du calcul des incertitudes. Enfin, la question des conditions d'applicabilité de ces concepts et approximations conduit à faire entrer en ligne de compte des circonstances qui dépassent le cadre des considérations théoriques et pratiques généralement invoquées au cours des analyses scientifiques et techniques. La possibilité de procéder à un traitement unifié des différentes composantes de l'incertitude dépend en effet de l'existence et du maintien de conditions sociales et institutionnelles susceptibles d'assurer aux usagers l'accès à un choix suffisamment diversifié de laboratoires utilisant des méthodes et des dispositifs de mesure différents – pareil accès est d'ailleurs également nécessaire au travail de dépistage des effets systématiques des scientifiques et métrologues situés en amont. Si les usagers ne peuvent faire appel qu'à un seul type de laboratoire, utilisant une seule technique ou un ensemble limité d'instruments, les conditions requises pour le traitement probabiliste des erreurs systématiques ne seront pas remplies ; selon Willink, les contributions aléatoires et systématiques de l'incertitude de mesure devront alors être exprimées séparément¹¹⁸ (comme indiqué dans la section 2.1.4).

On notera que cette extension s'inscrit bien dans la continuité de l'approche fréquentiste classique. On se souvient que, dans cette dernière, les probabilités portent sur les réalisations (événements, résultats, tirages) potentielles d'expériences effectuées dans des conditions fixées. Elles sont liées au déroulement d'un processus expérimental et dépendent des caractéristiques et des performances physiques des instruments, ainsi que des procédures et des manipulations mises en œuvre. Partant de là, la prise en considération des infrastructures matérielles et sociales qui sous-tendent ces processus physiques et ces procédures se présente comme un prolongement naturel, sinon aisé à mettre en œuvre, de l'approche fréquentiste traditionnelle. La qualité des mesures reflèterait ainsi non seulement le comportement à long terme des processus physiques directement impliqués mais également celui de l'organisation sociale plus large qui les produit, les maintient et en assure le développement dans une société scientifique et technique donnée – des circonstances dont il est certes difficile de tenir compte dans le détail, mais dont l'examen est nécessaire, on l'a vu, pour justifier le calcul de taux de suc-

bâton trouvé au cours d'une promenade ; mais l'utilisation du mesurage sera alors limité à un usage privé et ne pourra pas vraiment s'inscrire dans un réseau de communications et d'actions publiques, et, ce faisant, être coordonnée à l'activité d'autrui.

117. Nous reprenons l'expression de J. Hardwig, « Epistemic Dependence », *The Journal of Philosophy*, 82(7), 1985, 335-349.

118. R. Willink, « On the validity of methods of uncertainty evaluation », *Metrologia*, 47, 2010, 80-89, p. 89.

cès pratiques. L'approche fréquentiste élargie en vient ainsi à incorporer au niveau même de la conceptualisation scientifique (notamment du taux de succès pratique et de ses conditions d'application) le champ de réflexion qu'explorent les différents courants de l'« épistémologie sociale »¹¹⁹ lorsqu'ils s'interrogent sur la manière dont les institutions scientifiques devraient être conçues pour promouvoir le développement d'une connaissance qui soit le moins possible exposée à des biais.

3.2.3 Du tournant épistémique au tournant social : le problème de la dépendance épistémique

Si la manière dont Willink cherche à résoudre les difficultés soulevées par le GUM n'a, pour l'instant, guère de portée pratique, elle retient notre attention pour sa portée critique. La piste sociale que nous y avons décelée met en avant à quel point les problèmes métrologiques s'opposent à la conception individualiste de la connaissance qui a si longtemps constitué l'assise commune au débat philosophique entre fréquentistes et bayésiens. Elle suggère également que ce débat ne peut plus être formulé dans les termes de l'opposition classique entre point de vue objectif et point de vue subjectif discutée plus haut, et qu'il serait plus approprié de le reconsidérer dans le cadre d'une épistémologie attentive à la dimension sociale de la connaissance, et de l'activité scientifique en particulier. Sans nous aventurer dans le détail des différents courants de l'épistémologie sociale contemporaine, nous voudrions, avant de conclure, attirer l'attention sur la manière fort différente dont les deux approches rivales considèrent la condition de « dépendance épistémique » que les questions métrologiques liées à la globalisation des échanges du tournant des XX^e et XXI^e siècles révèlent de façon particulièrement saillante.

Il convient tout d'abord de remarquer que l'élaboration du GUM, et la métrologie en général, ne constituent pas une illustration de plus du rôle que jouent les « témoignages » et les relations de confiance dans le travail scientifique, ni une invitation de plus à réviser, en conséquence, l'épistémologie individualiste dominante¹²⁰. Avec le GUM, la métrologie assume ouvertement pour l'une de ses tâches le développement d'une méthodologie scientifique dont la formulation, soit implicitement par ses objectifs (approche bayésienne), soit plus explicitement (approche fréquentiste élargie), entérine la pluralité des acteurs scientifiques et techniques qui la mettent en œuvre. Cette méthodologie prend acte, tout à la fois, de l'absence d'autonomie

119. Pour un aperçu de ces différents courants, on pourra consulter A. Goldman, « Social Epistemology », in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2010 Edition), URL=<http://plato.stanford.edu/archives/sum2010/entries/epistemology-social/>; et H. Longino, « The Social Dimensions of Scientific Knowledge », in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2013 Edition), URL=<http://plato.stanford.edu/archives/spr2013/entries/scientific-knowledge-social/>.

120. On peut rappeler les exemples de Hardwig montrant que dans les sciences expérimentales, et de plus en plus souvent aussi en mathématiques, les scientifiques ne disposent individuellement ni de l'expertise, ni des données expérimentales dont ils ont besoin pour établir de nouveaux résultats, et donc acquérir de nouvelles connaissances. Le développement scientifique est ainsi suspendu à l'existence et au bon fonctionnement de réseaux de collaborations qui sont devenus particulièrement visibles au XX^e siècle avec la formation d'équipes de recherches parfois très étendues (notamment en physique des particules). À moins de contester la rationalité des sciences exactes elles-mêmes, Hardwig conclut à la nécessité de procéder à une révision de l'épistémologie individualiste. J. Hardwig, « The Role of Trust in Knowledge », *The Journal of Philosophy*, 88(12), 1991, 693-708.

des acteurs et des bénéficiaires que ces derniers peuvent retirer de leur situation de dépendance – puisque, on l’a vu, la communication et la mise en commun des résultats ouvrent sur le travail critique et permettent d’atteindre des niveaux de fiabilité inaccessibles aux résultats individuels. De ce point de vue, la métrologie se présente comme une discipline très particulière ; ses investigations méthodologiques comportent une dimension réflexive et normative qui la pousse à conjuguer ses activités scientifiques classiques, liées au mesurage proprement dit, à des activités légales, organisationnelles, institutionnelles au sein desquelles le sujet qui fait la science est d’emblée considéré comme un acteur social.

Cette dimension normative est manifeste dans le GUM, puisque ce dernier cherche à établir comment un résultat de mesure doit être constitué pour pouvoir être considéré comme partageable et comment il convient de caractériser sa qualité. Dans le cadre de l’approche fréquentiste révisée de Willink, un résultat est considéré comme partageable s’il a été établi par un processus que l’on estime satisfaisant à certaines conditions de fiabilité fondées, d’une part, sur l’étude fréquentiste (classique) des processus et des manipulations physiques impliqués et, d’autre part, sur l’inspection (moins classique) de l’environnement technique et institutionnel qui oriente les interactions potentielles des acteurs. On cherche en outre, à partir de ces études, à quantifier la confiance que l’on peut accorder au résultat obtenu de manière à pouvoir calculer le risque d’échec encouru – la prise d’une mauvaise décision, par exemple – lors de l’utilisation des résultats dans d’autres conditions que celles dans lesquelles il a été obtenu. La dépendance épistémique des acteurs se présente alors comme l’expression des médiations physiques, disciplinaires, professionnelles, économiques, institutionnelles successives qui s’interposent entre l’usager (situé à un nœud du réseau) et le mode de production des résultats qu’il utilise. La connaissance des acteurs se révèle par là fortement conditionnée à la fois par le monde physique et social.

La dépendance épistémique des acteurs se trouve de la sorte explicitement corrélée, à la fois à la position que ces derniers occupent dans le réseau, et à un état particulier d’avancement théorique, technique et institutionnel. Le caractère explicite que prend cette corrélation au sein de l’approche fréquentiste élargie marque bien que les résultats sont toujours rapportés à une configuration particulière. L’approche fréquentiste élargie donne par là le moyen d’établir la distinction entre reproductibilité et exactitude sur laquelle Quinn insiste avec ardeur. De fait, la reproductibilité peut être assurée au sein d’une configuration scientifique, économique et institutionnelle donnée sans pour autant que l’exactitude le soit.

Des données qui sont reproductibles sans être exactes peuvent conduire à déformer la représentation de la nature d’une multitude de façons. Une chose est sûre ; il serait imprudent de construire des avions sur la base de données reproductibles mais qui ne sont pas exactes car l’on ne pourrait garantir leur sécurité. [...] Si leur exactitude n’est pas fermement établie, il n’y a pas moyen de savoir si des résultats apparemment reproductibles sont constants dans le temps. ¹²¹

Avoir égard à la distinction entre reproductibilité et exactitude passe par le fait de reconnaître

121. Quinn, « Metrology, its role in today’s world », *op. cit.*, p.13 (notre traduction).

explicitement que des biais collectifs peuvent être rendus invisibles à l'intérieur d'une configuration (scientifique, technique, institutionnelle) donnée par des circonstances aussi bien physiques que sociales. Or, l'approche fréquentiste révisée, parce qu'elle s'attache aux processus (physiques et sociaux), peut prendre les configurations particulières pour objet (au lieu d'évoluer à l'intérieur d'un état de connaissance, dans la sphère de l'intersubjectivité) ; et parce qu'elle est orientée vers le futur (comme toute approche fréquentiste), porte à anticiper la falsification possible des résultats obtenus dans la configuration en cours et à envisager des variations de cette dernière. De ce point de vue, l'élargissement de la perspective fréquentiste fait bien droit à l'exigence d'exactitude, en sus de la simple reproductibilité, car il prend acte, derrière le *statu quo*, de la sollicitation, toujours pendante, à remanier activement une configuration donnée d'un point de vue scientifique, mais aussi social. L'environnement, les médiations (professionnelles, économiques etc.) peuvent constituer autant de volants susceptibles, comme la découverte de nouveaux principes physiques, de modifier une configuration et mettre au jour des biais indécélables dans la situation de départ. En introduisant au sein de la méthodologie scientifique des éléments d'analyse sociologiques, voire politiques qui peuvent influencer sur la capacité à déceler des biais physiques, à développer de nouveaux principes de mesure, ou à garantir la diversification des méthodes effectivement disponibles, l'approche fréquentiste révisée se dote d'un volant d'action supplémentaire. Celui-ci se révèle essentiel pour garantir l'exercice d'une *rationalité critique entière*, capable de répondre de la fiabilité des résultats de mesure et de donner les moyens de progressivement rectifier, sans pouvoir toutefois jamais prétendre les faire disparaître, les biais dus à la dépendance épistémique attachée à telle configuration cognitive et sociale particulière susceptible de grever nos décisions.

De son côté, l'approche bayésienne a été amenée à se développer afin de dépasser sa perspective individualiste traditionnelle. Cette évolution s'est opérée dans le cadre général des études bayésiennes, qui ont cherché à étendre leur traitement des probabilités au point de vue intersubjectif en s'intéressant au processus d'agrégation des opinions¹²². Les métrologues favorables à l'approche bayésienne n'ont pas manqué de se saisir de ces travaux qui fournissaient des instruments pour aborder les aspects collectifs des pratiques de mesure.

Cependant, dans le contexte des travaux métrologiques qui nous occupent ici, le point de vue épistémique individuel est resté premier. L'extension de l'approche bayésienne consiste en effet à combiner les fonctions de densité de probabilité décrivant les degrés de croyance de différents experts, ou les densités de probabilité obtenues par différents laboratoires concernant un effet systématique ou une quantité particulière, afin d'obtenir une densité de probabilité résultante *agrégée* représentant le degré de croyance d'un groupe d'experts ou d'un groupe de laboratoires. Ce type d'agrégation se révèle en particulier indispensable pour traiter les résultats des campagnes de mesure internationales qui ont pris une importance considérable avec la signature, en 1999, de l'Accord de reconnaissance mutuelle (MRA) mis en place par le Comité international des poids et mesures (CIPM)¹²³. Certaines de ces campagnes, les campagnes de

122. Une référence classique, déjà ancienne, est Ch. Genest & J.V. Zidek, « Combining Probability Distributions : A Critique and an Annotated Bibliography », *Statistical Science*, 1(1), 1986, 114-135.

123. Pour un exposé retraçant les étapes qui ont conduit à l'Accord de reconnaissance mutuelle, consulter T. Quinn,

« comparaisons clés interlaboratoires », permettent en effet d'assurer plus fermement la compatibilité des mesures au niveau international en établissant avec quel degré d'incertitude l'on peut considérer que les étalons nationaux maintenus par les différents Instituts nationaux de métrologie (INM) des pays signataires sont équivalents entre eux.

L'approche bayésienne n'est donc pas plus figée que l'approche fréquentiste. Ses méthodes sont en plein essor au sein de la métrologie, et différentes manières d'opérer l'agrégation des fonctions densité de probabilité (et de traiter les résultats des campagnes de comparaison) sont explorées avec assiduité¹²⁴. La contestation à laquelle fait aujourd'hui face l'hégémonie des experts de la part des associations d'utilisateurs semble aussi constituer un facteur non négligeable d'évolution.

Cependant, dans l'approche bayésienne révisée, la dépendance épistémique des agents n'intervient pas de façon explicite et n'est pas liée à une position ou à un environnement particulier. Sans doute pareil lien est-il congénital au point de vue épistémique en tant que tel, si bien qu'il peut sembler inutile de le rappeler. La dépendance épistémique intervient tout au plus à l'arrière-plan, en incitant au processus de révision des croyances de façon informelle et générale. Mais il semble, du même coup, que l'élargissement de la perspective bayésienne ne porte pas à remettre en cause la conception du rapport entre sphères scientifique et sociale qui est habituellement véhiculée par les méthodes scientifiques. L'approche bayésienne révisée continue de faire fond sur une confiance socialement instituée qu'elle ne se préoccupe ni de justifier, ni même de rendre explicite. En acceptant l'information associée à une probabilité interprétée comme degré de croyance (issue ou non d'une opération d'agrégation), l'utilisateur endosse sa dépendance épistémique à l'égard de la communauté dont il fait partie et dont il partage inévitablement les biais sans toutefois que cette position de dépendance soit explicitement reconnue ni scientifiquement thématifiée au sein du calcul.

Cette confiance présumée de la part des utilisateurs, la lettre de la méthode semble aussi se contenter de la présumer du côté des experts et des laboratoires puisqu'elle ne donne, par exemple, aucune indication sur les conditions que doit vérifier l'environnement pour garantir la *rationalité du processus de révision*. De fait, l'attrait de l'approche bayésienne tient à ce que la révision des croyances et donc des résultats répondant à de nouvelles informations peut s'effectuer de façon systématique suivant une règle, la formule de Bayes (désormais susceptible d'être prise en charge par les ordinateurs), qui formalise, dans l'esprit d'une logique inductive, les opérations mentales sur la base desquelles s'effectue la révision. Or, l'initiative d'enclencher le processus de révision en faisant appel à de nouvelles informations repose sur un

From Artefacts to Atoms. The BIPM and the Search for Ultimate Measurement Standards, Oxford University Press, 2012, 328-340; ainsi que le site du BIPM : Bureau International des Poids et Mesures, L'Arrangement de reconnaissance mutuelle (CIPM MRA), URL=<http://www.bipm.org/fr/cipm-mra/>, page consultée le 10 septembre 2014.

124. Dans le domaine de la métrologie, l'agrégation des opinions est étudiée dans le cadre de l'analyse des comparaisons clés interlaboratoires. Nous nous contenterons de citer, à titre d'exemple : A. G. Steele, K. D. Hill & R. J. Douglas, « Data pooling and key comparison reference values », *Metrologia*, 39(3), 2002, 269-277 ; P. Ciarlini *et al.*, « The use of a mixture of probability distributions in temperature interlaboratory comparisons », *Metrologia*, 41(3), 2004, 116-121.

travail critique de confrontation des résultats – obtenus, par exemple, par la mise en œuvre de principes de mesure différents dans le cadre scientifique, ou encore par des mesures « citoyennes »¹²⁵. Tout comme le dépistage des erreurs systématiques, le processus de révision est tributaire d'une organisation sociale et institutionnelle, voire associative, qu'il faut préserver ou dont il faut favoriser le développement. Il conviendrait notamment d'examiner à quelles conditions les experts sont à même de percevoir la nécessité de recueillir de nouvelles informations – l'appartenance des experts à un groupe industriel peut, en effet, aisément biaiser leur appréciation en les induisant, par exemple, à éviter les coûts qu'entraînerait l'actualisation des résultats antérieurs. Dans un monde dont le fonctionnement dépend de plus en plus de l'activité scientifique et technique et mobilise de façon complexe toutes les couches de l'organisation sociale, il ne paraît plus guère possible de s'en remettre à la « nature » comme à une providence bienveillante qui aurait soin d'assurer la rationalité du processus de révision bayésien.

Aussi, pourrait-on penser que les métrologues en charge des aspects fondamentaux de la science de la mesure devraient revendiquer, sur la base d'arguments méthodologiques, d'étendre leur activité au-delà du domaine scientifique et technique « classique » pour participer aux décisions susceptibles d'avoir un impact sur les infrastructures qui sous-tendent les conditions de possibilité sociales et politiques de l'exercice d'une rationalité conçue comme critique et comme capacité de révision.

3.2.4 Quels enseignements tirer de l'approche fréquentiste élargie ?

Que peut retenir le philosophe de la riposte fréquentiste à l'avancée du bayésianisme et quel éclairage critique peut-il en tirer ? Un premier trait remarquable de l'extension fréquentiste a déjà été souligné. Il tient, si l'on accepte notre analyse, à la prise en compte explicite de la sphère sociale dans l'établissement des résultats scientifiques. On ajoutera que cette inscription de la sphère sociale dans l'appareil méthodologique des sciences a ceci de marquant que, loin de signaler une forme de relativisme, elle est, tout au contraire, inspirée par la défense des notions de validité et d'exactitude.

Cette extension présente par ailleurs un intérêt touchant aux rapports entre développement scientifique et développement philosophique. Elle montre en effet, comment l'interprétation philosophique attachée à une méthode qui a fait ses preuves dans un contexte donné peut être amenée à évoluer sous la pression des procédés que l'on est conduit à mettre en œuvre pour adapter cette méthode à des situations et des buts nouveaux. De fait, nous avons vu comment l'extension proposée par Willink devait recourir à des approximations et des moyennes pour parvenir à rendre l'approche fréquentiste apte à répondre au cahier des charges inspiré par le renforcement des exigences métrologiques liées à la mondialisation – l'expression de l'incertitude associée à un résultat par un seul paramètre est, il faut le rappeler, une exigence dictée par

125. Nous remercions A. Dolez d'avoir porté à notre connaissance le mouvement associatif des « capteurs citoyens » qui entreprend des mesures *in situ* de certains paramètres environnementaux sensibles afin de les confronter aux mesures publiées par l'expertise officielle ; F. Chateauraynaud & J. Debaz, « De la métrologie en démocratie. La nouvelle vague des capteurs citoyens. » Article en ligne, URL=<http://socioargu.hypotheses.org/4505>, page consultée le 23 novembre 2014.

la pratique. Et nous avons montré comment cette extension pouvait se comprendre, sur le plan philosophique, comme une extension de l'interprétation « objective » classique des probabilités incorporant la prise en compte des conditions sociales de l'activité scientifique.

Ces observations invitent à aller plus loin. Elles mettent en effet en relief que l'élargissement de l'approche fréquentiste est motivé par la volonté de réaliser certaines fonctions (jugées désirables) dans un contexte et un environnement donnés. La démarche d'extension semble ainsi davantage relever d'une activité d'ingénierie, qui se préoccupe de la résolution de problèmes effectifs tels qu'ils se présentent en situation, que d'une activité qui serait exclusivement orientée par la recherche d'une méthode unitaire à visée universelle – la métrologie n'est-elle pas, après tout, comme le soulignait Giacomo, ancien directeur du BIPM, à la fois la science et l'art de la mesure ¹²⁶ ? Qu'un fréquentiste aussi convaincu que Willink recherche des solutions de compromis ¹²⁷ et ne refuse pas de s'atteler lui-même au développement de méthodes bayésiennes conforte cette idée. Les difficultés d'application déjà évoquées donnent d'ailleurs à penser que l'approche fréquentiste révisée ne peut vraiment prétendre être utilisée seule et serait sans doute davantage appelée, même par ses plus fervents défenseurs, à être combinée avec l'approche bayésienne, peut-être comme méthode de contrôle. C'est le lieu de noter que lorsqu'on se place dans ce registre orienté vers l'action, l'exigence réaliste d'exactitude qui anime l'approche fréquentiste, telle qu'elle se révèle dans ses potentialités de révision, semble moins relever d'une prise de position métaphysique, comme le suggèrent certains métrologues, que de la vigilance à l'égard des *risques* d'erreur *non calculables* toujours possibles. Celle-ci est liée à l'orientation caractéristique de l'approche fréquentiste vers le futur que nous avons déjà eu l'occasion de souligner (section 3.1.1).

De leur côté, les partisans de l'approche bayésienne semblent plus enclins à accorder une visée universelle à leur mode d'analyse des résultats de mesure, car ils considèrent que ce dernier est justifié par la conception épistémique des probabilités. Celle-ci se trouve d'ailleurs à leurs yeux consolidée par sa capacité à proposer une interprétation unifiée des deux composantes de l'incertitude. Il faut cependant remarquer que l'attrait de l'approche bayésienne doit beaucoup aux avancées techniques qui ont permis d'accroître la puissance de calcul des ordinateurs, sans lesquelles, par exemple, la « propagation des distributions » mentionnée dans la section 2.2.4 serait presque toujours impraticable. De ce point de vue, le développement de l'approche bayésienne touche, lui aussi, aux sciences de l'ingénieur.

Au travers de ces réflexions, l'extension de l'approche fréquentiste prend une fonction critique de mise en garde contre la tentation de prolonger trop avant la posture théorique dans le domaine de la pratique. Elle invite par là, nous semble-t-il, le philosophe des sciences à faire

126. P. Giacomo, « Mesure et Métrologie », *Bulletin du Bureau National de Métrologie*, 101, 1995, cité par M. Himbert, « Unités, références, incertitudes pour les mesures », *La pluridisciplinarité dans les enseignements scientifiques. Tome 2 : La place de l'expérience. Actes de l'université d'été, 9-13 juillet 2001, Cachan*. L'article est téléchargeable à partir à l'adresse :

http://www.udppc.asso.fr/national/attachments/article/537/uescience_himbert_111926.pdf

127. R. Willink & I. Lira, « A united interpretation of different uncertainty intervals », *Measurement*, 38, 2005, 61-66. Lira a, par la suite, renié cette tentative de conciliation.

retour sur sa propre activité. Il y a assurément longtemps que celui-ci a renoncé à son rôle prescriptif – peut-être pas encore tout à fait à celui d’arbitre. Nous voudrions, pour terminer, indiquer une direction que le philosophe des sciences pourrait être incité à prendre à la lumière du travail réflexif effectué par les métrologues, dont la discipline se situe à l’interface des préoccupations « purement » scientifiques et des buts que nos sociétés se proposent de poursuivre.

4 Conclusion

Plusieurs éléments invitent le philosophe qui se proposerait de prendre parti pour l’une ou l’autre des deux approches à la prudence. Il convient tout d’abord de souligner que le spectre des méthodes d’analyse d’incertitude qu’utilisent métrologues et scientifiques en général ne se limite pas strictement aux deux approches que nous venons d’évoquer dans cet article. Nous avons choisi de nous limiter aux deux méthodes présentées, d’une part parce qu’elles figurent parmi les plus utilisées, et, d’autre part, parce que ce sont elles qui apparaissent en filigrane dans le GUM et qui sont donc les plus débattues. Il n’existe cependant pas, dans la pratique, *une* méthode fréquentiste unique, ni *une* méthode bayésienne unique ; les méthodes que nous avons présentées constituent plutôt deux approches limites entre lesquelles se déploie un spectre d’approches différentes qui croisent entre elles des concepts et des termes provenant de ces deux extrêmes, si bien que l’on est en présence, comme le notent Ehrlich *et al.*, d’« un mélange de concepts et de terminologies »¹²⁸, et non de méthodes « pures ». Aussi, pour ressaisir les enjeux philosophiques des débats déclenchés par le GUM, avons-nous dû recourir, en partie, à des reconstructions conceptuelles qui ne correspondent pas toujours aux pratiques effectives des scientifiques. Il importe par ailleurs d’insister sur le fait que les approches explorées ne se réduisent pas non plus à des variations autour des thèmes fréquentistes et bayésiens. De nombreuses autres propositions existent qui vont parfois jusqu’à s’écarter de l’approche probabiliste. Ainsi, par exemple, les tenants de la « théorie des possibilités » considèrent-ils que, pour parvenir à un traitement cohérent des composantes aléatoire et systématique de l’incertitude, il faut se tourner vers d’autres mathématiques, les mathématiques du flou¹²⁹.

À côté des développements mathématiques, des développements techniques contribuent aussi à diversifier les méthodes. Avec l’amélioration considérable de la puissance de calcul des appareils informatiques, les méthodes de calcul numérique dites de « Monte-Carlo »¹³⁰ sont devenues de plus en plus fréquemment utilisées en métrologie à partir des années 1990. Le supplément 1 du GUM¹³¹, publié en 2008, est d’ailleurs consacré à leur utilisation dans le cadre

128. Ehrlich, Dybkaer, Wöger, « Evolution of philosophy and description of measurement », *op. cit.*, p. 201.

129. L.A. Zadeh, « Generalized Theory of Uncertainty (GTU) – Principal Concept and Ideas », *Computational Statistics & Data Analysis*, 51, 2006, 15-46.

130. « Les simulations MC (Monte Carlo) sont des algorithmes implémentés sur ordinateur qui utilisent le hasard pour calculer les propriétés de modèles mathématiques, les modèles étudiés ne présentant pas eux-mêmes de caractère aléatoire. » E. Winsberg, « Computer Simulations in Science », in E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, Fall 2014 Edition, URL=<http://plato.stanford.edu/archives/fall2014/entries/simulations-science/>, section 2.4.

131. Voir note 75.

de l'analyse d'incertitude. Les méthodes de Monte-Carlo présentent plusieurs avantages sur les techniques traditionnelles du GUM, fondées sur des modèles linéaires, l'omniprésence de la loi normale, et des modes de calcul analytique. Elles offrent tout d'abord un spectre d'applicabilité plus large car, contrairement aux méthodes traditionnelles, elles s'appliquent tout autant aux cas linéaires qu'aux situations présentant une non-linéarité. Elles permettent par ailleurs d'automatiser certaines étapes de l'analyse d'incertitude par l'application directe d'un algorithme de calcul, la tâche étant désormais aisée grâce aux moyens informatiques modernes. Initialement, l'application des méthodes de Monte-Carlo en métrologie ne représentait qu'une alternative de secours, qui pouvait se révéler utile si l'on atteignait les limites classiques du GUM. Aujourd'hui, l'on envisage aisément qu'elles puissent devenir le cadre standard de l'analyse d'incertitude, dont le GUM ne formerait alors plus qu'un cas particulier d'application (des remarques vont déjà dans ce sens dans certains documents récents)¹³².

L'objectif d'uniformiser l'expression des incertitudes de mesure a donc non seulement conduit les deux approches classiques les plus débattues à réinterpréter ou enrichir leur perspective philosophique initiale ; il a aussi stimulé une pluralité de propositions. La diversité des modes de conceptualisation et de calculs explorés et discutés par les métrologues (ainsi que par d'autres scientifiques) afin d'évaluer les incertitudes de mesure dans un très grand nombre de domaines et de contextes est propre à jeter le soupçon sur toute analyse philosophique qui prétendrait pouvoir, à elle seule, trancher entre ces différentes directions, ou même énoncer un principe unique qui devrait les orienter – tout comme cela peut être le cas dans les conflits auxquels se trouvent confrontés les ingénieurs, le conflit entre exactitude et simplicité, par exemple, ne peut être résolu qu'au regard du contexte.

La percée récente et le succès des méthodes de Monte-Carlo sont là pour rappeler à quel point ce sont avant tout des considérations pratiques, en particulier liées à la faisabilité et à la performance¹³³ des méthodes utilisées qui, en fin de compte, conditionnent la préférence accordée à une méthode plutôt qu'une autre. Les choix conceptuels et philosophiques opérés sont alors, en partie, la *conséquence*, et non la cause, de l'adoption de la méthode. On retrouve le caractère déterminant, si fortement souligné par Humphreys¹³⁴, que joue dans le choix des orientations scientifiques l'impératif d'obtenir des résultats effectifs susceptibles d'être appliqués. Il faut donc éviter de tomber dans le cercle qui consiste à consacrer une méthode sous prétexte qu'elle est légitimée par une position philosophique que des avantages pratiques ont contribué à sélectionner et à adopter. Encore une fois, l'approche bayésienne est portée en avant par l'exigence toute pratique d'exprimer l'incertitude de mesure par un seul terme. Plutôt que de chercher à promouvoir l'une ou l'autre des interprétations classiques, il conviendrait peut-être d'insister sur la nécessité de disposer de méthodes différentes qu'il s'agit à chaque fois d'appliquer de façon appropriée selon la situation et le groupe d'acteurs impliqué. Contrairement

132. Voir, par exemple, I. Lira, « On the long-run success rate of coverage intervals », *Metrologia*, 45, 2008, 21-23.

133. Weise et Wöger, dans leur article développant les fondements de l'approche bayésienne appliquée à la métrologie, défendaient justement cette dernière en affirmant que « le critère principal pour choisir une méthode d'inférence est sa performance », Weise et Wöger, « A Bayesian theory of measurement uncertainty », *op. cit.*, p.2.

134. Humphreys, *Extending ourselves. Computational Science, Empiricism, and Scientific Method*, *op. cit.*

au projet unitaire affiché au départ¹³⁵, il paraît difficile de prétendre que le GUM et ses suppléments proposent les fondements d'une méthode à visée universelle, destinée à régler toutes les activités et valant dans tous les contextes – les laboratoires à la pointe de la recherche ne sauraient, quant à eux, se contenter d'appliquer les recommandations avancées car ils explorent des territoires limites où ces dernières sont susceptibles d'être mises à l'épreuve. Dans les faits, ces recommandations paraissent essentiellement appelées à régler la pratique de groupes d'acteurs spécifiques : les scientifiques, les groupes industriels et techniques, qui peuvent (dans des circonstances « ordinaires » qu'il revient cependant à chacun d'apprécier!) se décharger du soin de veiller aux conditions susceptibles d'assurer la rationalité de leur conduite sur la société plus large dans laquelle s'inscrit leur activité – tout en ayant conscience du caractère faillible de cette conduite.

Nous avons souligné les aspects épistémologiques du travail des métrologues qui, en s'interrogeant sur les modalités de validation des mesures adoptent une attitude réflexive. Cependant, les débats et les décisions des métrologues sont aussi soumis à des *contraintes temporelles* caractéristiques des activités pratiques, qui ne s'imposent guère de la même façon à la réflexion du philosophe des sciences. Encore une fois, le projet du GUM se situe à la croisée de réflexions d'ordre conceptuel et pratique. S'il vise, d'un côté, à clarifier les conditions de la preuve empirique, il a, de l'autre, pour mission de répondre aux besoins engendrés par la mondialisation des échanges en énonçant des recommandations permettant de déterminer concrètement, de la façon la plus simple possible, la qualité des résultats de mesure et des produits. Or, les activités scientifiques, techniques et commerciales sont certes en attente de recommandations plus claires et mieux étayées, surtout depuis la signature du CIPM MRA, mais elles ne peuvent attendre que les débats conceptuels soient *entièrement* résolus. La publication de la nouvelle version du GUM sur laquelle le JCGM travaille afin de remplacer l'édition tant critiquée de 1993 est attendue et se doit de paraître dans un délai raisonnable. L'objectif des métrologues n'est pas seulement cognitif. Leurs débats doivent aboutir à une prise de décision, une action qui consiste ici à publier un texte de recommandations dont l'impact sur la sphère publique est considérable. La mission des métrologues ne dépasse pas seulement les préoccupations épistémologiques, elle dépasse la conception classique que l'on se fait de la science, attachée à l'exploration des phénomènes, pour prendre en charge (scientifiquement), au sein de la société même, les situations induites par les développements technoscientifiques et commerciaux que les sciences ont contribué à engendrer. Elle participe de la sorte à ce que Beck appelle un mode de « scientificisation réflexi(f) »¹³⁶; mais, nous l'avons vu, en courant le danger de donner à penser qu'il pourrait y avoir une *science* de la pratique. Il nous semble que cette conjoncture doit être accompagnée par une évolution parallèle de la philosophie des sciences susceptible de maintenir visible la pluralité des approches possibles. La philosophie des sciences viendrait alors s'opposer à l'illusion théorique dénoncée plus haut, qui peut conduire à figer une mé-

135. « Ce Guide établit les règles générales pour l'évaluation et l'expression de l'incertitude pour les mesurages qui peuvent être effectués à des niveaux variés d'exactitude et dans de nombreux domaines - de la boutique du commerçant à la recherche fondamentale », JCGM, *Guide pour l'expression de l'incertitude de mesure*, *op. cit.*, p.1.

136. U. Beck, *La Société du risque. Sur la voie d'une autre modernité* [1986], traduit de l'allemand par L. Bernardi, Flammarion, 2001. En particulier le chapitre 7, « Une science au-delà de la vérité et de la rationalité émancipatrice? Réflexivité et critique de l'évolution scientifico-technique », pp. 341-398.

thode répondant aux besoins du monde technico-scientifique comme si elle était scientifiquement fondée au risque de modeler, voire bloquer en retour l'activité scientifique elle-même.

D'un côté, la philosophie des sciences doit continuer de pouvoir s'inscrire dans la temporalité et la posture classiques propres à l'attitude cognitive afin d'assurer un rôle critique qui consiste, ici, à analyser les différentes approches proposées, voire en développer de nouvelles. Mais elle doit aussi, d'un autre côté, accomplir un tournant pratique qui ne consiste pas seulement à *décrire* ces phénomènes que sont les pratiques scientifiques (pour éventuellement mettre en avant la pluralité des méthodes), ni même à s'appliquer à résoudre des problèmes concrets – des attitudes qui relèvent, somme toute, encore de la posture théorique de l'observateur –, mais plutôt à renouer avec la *philosophie pratique* pour *travailler* de pair avec les autres agents scientifiques, techniques, politiques, à la construction d'un monde commun, et se donner les moyens de « transform(er) le processus de modernisation lui-même en processus d'apprentissage »¹³⁷.

Remerciements. Nous remercions Isabelle Drouet et Thierry Martin de leur invitation au sein du séminaire « Probabilités, Décision, Incertitude » de l'IHPST dans lequel nous avons exposé, en novembre 2012, une première version de notre travail. Nous remercions vivement, une fois encore, Isabelle Drouet de ses nombreuses remarques qui nous ont permis de clarifier bien des points de ce chapitre. Nous exprimons notre gratitude aux membres du département « Mesures, analyse, qualité et ingénierie des matériaux (Maquim) » du Conservatoire national des arts et métiers (Cnam), et tout particulièrement à Saida Guellati-Khélifa, Marc Himbert, Patrick Juncar, Michel Lecollinet et Annick Razet, de nous avoir fait bénéficier de leur profonde connaissance de la métrologie. Notre reconnaissance va aussi à Michèle Désenfant, Nicolas Fisher, Luca Mari, François Nez et Marc Priel pour les nombreuses discussions enrichissantes que nous avons eues avec eux. Nous remercions également les participants de la Journée internationale d'étude "The taming of measurement uncertainties : Theoretical, practical, and philosophical issues", que nous avons co-organisée avec Olivier Darrigol et Oliver Schlaudt, et qui s'est déroulée le 4 décembre 2012 à l'université Paris Diderot.

137. U. Beck, *La Société du risque. Sur la voie d'une autre modernité*, op. cit., p. 392.