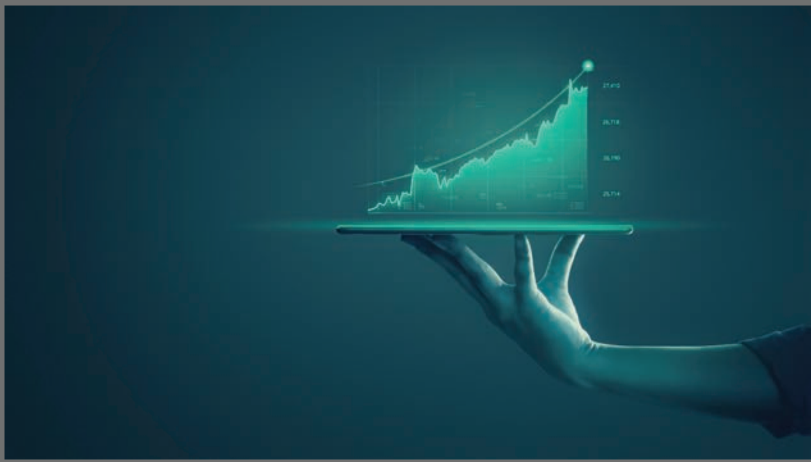
Statistique bayésienne

Nous continuons notre tour d'horizon des technologies et des avancées techniques qui sont en train de remodeler le paysage de la métrologie, en l'obligeant parfois à se remettre en cause, à s'adapter et dans tous les cas à évoluer. Le groupe de travail Creative Metrology et les experts qui le constituent, contribuent à ces changements et assurent leur diffusion à la communauté des métrologues et aux industriels. Aujourd'hui, c'est Nicolas Fischer du LNE qui explique les fondements et le rôle que la statistique bayésienne est amenée à jouer à différents niveaux du processus de mesure.



Historique

La statistique bayésienne connaît ses origines au 18e siècle avec les travaux de deux hommes, le révérend britannique Thomas Bayes et le scientifique français Pierre Simon Laplace. De manière quasi concomitante et sans avoir connaissance des travaux l'un de l'autre, ils s'intéressent à la théorie des probabilités et énoncent ce qui restera comme la loi de Bayes. En s'intéressant au lien entre causes et conséquences en astronomie notamment, cette formule exprime comment calculer la probabilité d'un événement conditionnellement à une observation ou à un autre évènement. Dans un premier temps elle fut très utilisée pour évaluer dans les jeux de hasard les probabilités de gain pour un scénario de jeu donné. Elle est depuis utilisée dans des domaines très variés, du décryptage du code Enigma des nazis durant la seconde guerre mondiale à la compréhension des mécanismes cérébraux de nos jours.

Définition

La formule de Bayes, considérant deux événements A et B, s'énonce comme suit :

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

- **P(A|B)** représente la probabilité conditionnelle de A sachant B, elle est également nommée probabilité *a posteriori* de A, c'est-à-dire une fois B observé
- **P(A)** représente la probabilité *a priori* de A, c'est-à-dire avant toute connaissance sur B, elle est également appelée probabilité marginale de A
- **P(B)** représente la probabilité de B ou probabilité marginale
- **P(B|A)** représente la **fonction de vraisemblance de B**

La formule de Bayes peut donc être vue comme un processus de mise à jour de l'information sur A, c'est-à-dire de la probabilité que l'événement A se réalise, avant et après avoir observé B. L'inférence bayésienne est la méthode de mise à jour de l'information sur A utilisant la formule de Bayes. Notons de plus que le cadre bayésien introduit une définition alternative des probabilités à celle de la statistique classique. En effet, du point de vue de la statistique dite classique ou fréquentiste, la probabilité est vue comme la fréquence d'apparition d'un événement tandis qu'en statistique bayésienne, la probabilité représente l'état de connaissance de l'événement étudié.

Métrologie

Procéder à un mesurage, évaluer un résultat de mesure, c'est exprimer quantitativement l'information dont on dispose sur la grandeur d'intérêt, le mesurande. Rappelons que selon le *Vocabulaire International de Métrologie* (VIM) un résultat de mesure peut s'exprimer selon une loi de probabilité. Or celui-ci est la transcription de notre état de connaissance du mesurande une fois mis en œuvre le processus de mesure qui lui correspond. L'interprétation bayésienne de la loi de probabilité comme une fonction de connaissance du mesurande (appelée plus souvent fonction de croyance dans ce cadre) semble donc parfaitement convenir au cadre de la métrologie où le praticien cherche à quantifier/qualifier l'information, la connaissance qu'il obtient sur une grandeur suite à des observations.

Cette réflexion est toutefois relativement récente dans la communauté métrologique. Un projet de révision du GUM a été proposé en 2014 afin de donner un cadre bayésien à l'évaluation de l'incertitude de mesure, présentant les avantages de ce paradigme et expliquant que son supplément 1 est d'ores et déjà compatible avec ce cadre. Même s'il a été abandonné, les travaux du *working group 1 du Joint Committee for Guides in Metrology* (JCGM) en charge de faire évoluer le GUM visent à promouvoir, dans des documents existants ou à venir, la statistique bayésienne que ce soit pour l'évaluation de l'incertitude, mais aussi le traitement de données de comparaisons interlaboratoires, l'évaluation de la conformité ou l'estimation de fonctions d'étalonnage.

La littérature scientifique du domaine, citons la revue *Metrologia*, voit se multiplier les articles traitant de méthodologie bayésienne et exemples d'applications dans divers domaines de la mesure. Des projets de recherche récents, notamment dans le cadre du *European Metrology Programme for Innovation and Research* (Empir) d'Euramet visent à fournir des documents de bonnes pratiques dans le domaine, des exemples illustratifs et des outils logiciels.

Avantages

- L'inférence bayésienne étant un processus de mise à jour de la connaissance, elle donne toute sa place à la connaissance disponible *a priori*, c'est-à-dire avant mesurage. Or les experts d'un domaine ont bien souvent une connaissance préalable des grandeurs auxquelles ils s'intéressent : mesures réalisées antérieurement, expertise du domaine, données constructeurs, etc. Il est donc particulièrement intéressant de pouvoir la formaliser sous forme de loi de probabilité *a priori* et d'en tenir compte dans le processus d'évaluation d'un résultat de mesure. Cette information disponible contribuera ainsi à réduire l'incertitude associée à la grandeur évaluée.
- Dans beaucoup de procédés, la grandeur d'intérêt n'est pas observée directement. Elle est obtenue par inversion d'un modèle à partir de mesures d'une grandeur dont elle dépend, par exemple : les propriétés thermiques d'une paroi d'un bâtiment sont estimées à partir de mesures de températures ou de flux. Or, la formule de Bayes travaille justement sur les probabilités conditionnelles d'observer des conséquences à partir des causes. Autrement dit la résolution de problèmes inverses se fait tout naturellement dans ce cadre.
- La statistique bayésienne repose sur un cadre probabiliste solide mathématiquement qui permet, à partir des lois de probabilités *a priori*, de représenter la connaissance associée à des phénomènes de natures extrêmement variées.

Applications

Il existe un champ très vaste d'applications possibles dès lors que des résultats de mesure sont obtenus avec des estimations associées et ceci dans presque tous les secteurs d'activité.

On peut citer :
- évaluation d'incertitude de mesure ;
- estimation de fonctions d'étalonnage ;
- imagerie médicale : estimation de paramètres biologiques, physiologiques à partir d'observations ;
- évaluation de la conformité sous incertitude : estimation de risques de mauvaises décisions de conformité tenant compte de l'incertitude associée à un mesurage ;
- sûreté de fonctionnement , fiabilité : estimation de probabilités conditionnelles de dépassement d'un seuil critique (sécurité d'un ouvrage d'art, présence de polluants, détection de substances toxiques...)
- Bien d'autres encore.

Verrous

La mise en œuvre calculatoire de l'inférence bayésienne peut s'avérer complexe. Dans bon nombre de situations, la probabilité *a posteriori* recherchée n'a pas de forme analytique et ne peut qu'être approchée numériquement. Le plus souvent ce sont des algorithmes de type Monte Carlo Markov Chain (MCMC) qui sont utilisés, hors ceux-ci requièrent des compétences certaines en statistique pour être implémentés. De nombreuses bibliothèques d'algorithmes sont certes disponibles sous R, Python, matlab, etc., mais leur utilisation en elle-même n'est pas si évidente. Afin de favoriser la diffusion de ces méthodes, il convient donc de faire monter en compétences les personnes concernées (par la formation notamment), mais également de fournir des logiciels spécifiques pour des besoins précis.

Citons à ce titre le projet de recherche EMPIR CAsoft qui a pour objectif de développer et diffuser un logiciel interfacé d'évaluation de la conformité sous incertitude et de rendre ainsi accessible au plus grand nombre la mise en œuvre de formules de calculs de probabilités conditionnelles associées à l'évaluation des risques de mauvaises décisions ●

✎ **Nicolas Fischer,**
Responsable du département sciences des données et incertitudes, LNE

pour aller plus loin

C. Elster et al. A Guide to Bayesian Inference for Regression Problems, 2015 https://mathmet.org/publications/guides/index.php#regression

C. Yardin, S. Amar, N. Fischer, M. Sancandi and M. Keller, Bayesian estimation of a polynomial calibration function associated to a flow meter, Advanced Mathematical and Computational Tools in Metrology and Testing XI, Décembre 2018, pp 417 - 426