

# Entretien avec Jacques Bernier et Éric Parent



## Jacques BERNIER<sup>1</sup>

Ingénieur EDF à la retraite

## Éric PARENT<sup>2</sup>

AgroParisTech/INRAe, UMR Mathématique  
et Informatique Appliquées, Paris

## Gilles CELEUX<sup>3</sup>

Institut de mathématiques d'Orsay, Université Paris-Sud et Inria Orsay

### TITLE

Interview of Jacques Bernier and Éric Parent

### RÉSUMÉ

Jacques Bernier est un ancien ingénieur EDF et il fut un pionnier de l'enseignement de la décision statistique à l'ISUP. Je témoigne de la qualité de son enseignement que j'ai jadis suivi avec plaisir. Depuis son départ à la retraite en 1991, Jacques Bernier continue à être très actif scientifiquement et à publier des articles de recherche en statistique bayésienne appliquée. Il est l'un des scientifiques qui aura le plus contribué à la diffusion et à la démonstration du potentiel de l'approche bayésienne en hydrologie.

Son compère et disciple, Éric Parent, est ingénieur des Ponts, des Eaux et des Forêts. Il travaille aujourd'hui comme enseignant-chercheur en statistiques appliquées et en modélisation probabiliste pour l'ingénierie environnementale à AgroParisTech. Il a fait énormément pour la pénétration de bonnes pratiques bayésiennes en ingénierie par les nombreuses recherches appliquées qu'il orchestre.

**Mots-clés :** *statistiques bayésiennes appliquées, ingénierie, hydrologie, environnement.*

### ABSTRACT

Jacques Bernier was an engineer at the French Electricity Company (EDF). He pioneered a course on statistical decision at ISUP. I can witness the qualities of his course which I enjoyed when following as a former student of his. Since he retired in 1991, Jacques Bernier kept his scientific activities. He continues publishing articles in applied Bayesian statistics. He counts among the statisticians who contributed the most to the diffusion and the demonstration of the potential of a Bayesian approach in hydrology.

His disciple Éric Parent belongs to the French civil corps of Bridge, Road, Water and Forest Engineers. He presently works as a Research Teacher in statistics and probabilistic modeling applied to environmental engineering at AgroParisTech. He greatly contributed to the diffusion of good Bayesian practices in engineering through the numerous applied researches he conducted.

**Keywords:** *applied bayesian statistics, engineering, hydrology, environment.*

## Premières rencontres avec les concepts bayésiens

**G. Celeux :** Jacques Bernier, Éric Parent, vous avez écrit de nombreux ouvrages en français qui promeuvent la pensée de Thomas Bayes et en illustrent les applications à l'ingénierie (Bernier et al., 2000; Parent et Bernier, 2007; Boreux et al., 2010; Parent et Rivot, 2012). Quels ont été vos premiers contacts avec ce paradigme, peu répandu en France à vos débuts ?

1. jacques.bernier2@orange.fr
2. eric.parent@agroparistech.fr
3. gilles.celeux@inria.fr

**É. Parent :** On peut être nourri de cours de statistique classique durant les études, que ce soit Métivier et Neveu à Polytechnique en 1978, ou Deheuvels et Benzecri en DEA à Jussieu en 1983, et demeurer bien plus intéressé à la beauté des espaces hilbertiens qu'à comprendre de quel chapeau le statisticien peut bien sortir un estimateur. Ce n'est, durant ma thèse où Jacques m'a beaucoup guidé, que la lecture de son<sup>4</sup> livre avec Janine Ulmo me fait comprendre qu'un estimateur n'est finalement qu'une règle de décision particulière où l'espace des décisions peut s'identifier à celui des paramètres du modèle probabiliste. Dans les premiers chapitres de leur ouvrage encore timidement bayésien, Ulmo et Bernier (1973) présentent la théorie de la décision et le théorème de la classe complète pour l'admissibilité des règles de Bayes, mais la loi *a priori* reste vue comme une simple technique de convexification de l'ensemble des règles de décision, à la mode de Wald (1950) et non comme le fondement de la théorie subjective de la probabilité justifiant l'utilisation des concepts bayésiens (Savage, 1954). Sans le formuler d'une façon explicite, ils interprètent néanmoins le test du rapport de vraisemblance de Neyman et Pearson comme un facteur de Bayes (un rapport de crédibilités *a posteriori*) avec des lois *a priori* uniformes sur chacune des hypothèses en compétition. Voilà mon premier contact avec la cohérence du paradigme bayésien, la faute originelle à Jacques Bernier ! La lecture de Tribus (1969) dans sa traduction magnifique par Pezier (1972), m'apporte également beaucoup, je comprends que :

- la théorie de la décision de Ferguson (1967) permet de conduire un raisonnement argumenté pour construire une solution au projet d'engineering, ce qui me plaît en tant que futur ingénieur,
- la probabilité fait sens comme une réponse aux principes requis pour la construction d'un ordinateur à logique inductive selon les *desiderata* de Cox et Jaynes (1963) et Jaynes (1959), ce qui me plaît en tant que mathématicien.

Je m'aperçois ainsi qu'il existe une interprétation logique des probabilités indépendante de celle fondée sur la limite de fréquences, que tous mes professeurs, tant à l'Université que dans les Grandes Écoles, avaient passée sous silence. C'est cette interprétation subjective impersonnelle qui fournit aujourd'hui une justification à l'apprentissage automatique. À l'époque, mon incursion dans le contexte bayésien me permet de mieux comprendre le paradigme classique et de le positionner par rapport au cadre bayésien :

- L'approche fréquentiste, qu'on devrait plutôt appeler fréquentielle, décrit l'incertitude qui porte sur la valeur obtenue par une procédure d'estimation si on répétait celle-ci pour un grand nombre de fois pour des répliqués générés par le modèle probabiliste dont on connaîtrait les inconnues. Il s'agit donc de la fiabilité d'une procédure de calcul algorithmique en information parfaite. L'incertitude se manifeste dans le contexte classique au travers de la variabilité de répétition des observables conditionnellement à la connaissance des inconnues.
- L'approche bayésienne évalue l'incertitude qui reste sur les inconnues, conditionnellement à l'unique jeu de données observées, sachant que l'on est parti d'une croyance initiale sur leurs variations possibles. Il s'agit donc d'un jugement probabiliste direct sous la forme d'un pari quant à la valeur de ces inconnues.

**G. Celeux :** Je crois savoir que Jacques de vingt ans l'ainé d'Éric, partage plutôt la conception décisionnelle de la probabilité subjective, interprétée comme un pari personnel cohérent. Ce n'est pas étonnant puisque vous avez pu assister aux conférences données à l'IHP par Léonard Savaga (1954) et Bruno de Finetti (1937) dans les années cinquante. À EDF où vous avez fait l'essentiel de votre carrière, Jacques, avez vous ressenti à l'époque, l'impulsion de ce type de recherches mettant en application les travaux de Von Neumann et al. (1953) ou d'Allais (1953) ?

4. Heureuse bénédiction pour le thésard que j'étais, la bibliographie de Jacques était alors encore succincte...

**J. Bernier** : De formation universitaire en mathématiques et statistique (entre 1951 et 1955 – Sorbonne puis IHP), j'ai suivi parallèlement de 1953 à 1955 les cours de Probabilités (en partie avec Georges Darmois) et de l'ISUP (au départ avec Eugène Morice puis Georges Morlat). J'y ai bien entendu prononcé le nom de Bayes. Les commentaires de Darmois sur la mise en oeuvre de la formule de Bayes étaient assez réservés. Quant à Morlat, son objectif, en deuxième année de l'ISUP, était les tests et la théorie de la Décision Statistique de Wald (parus en 1950) et bien sûr le concept de « *règle de décision bayésienne* » y était présenté sans que le paradigme bayésien y soit bien développé. Il ne faut pas oublier que Wald lui-même était un « *fréquentiste* » et le raisonnement bayésien conçu par lui était un moyen pour construire des règles admissibles (propriété éminemment fréquentiste). Connaissant Morlat (1964) comme je l'ai connu par la suite, on pourrait s'étonner de sa prudence en ce qui concerne le paradigme bayésien mais on ne saurait oublier le contexte à l'ISUP à cette époque et même après ; le cours parallèle d'Indoudjian sur l'estimation était résolument fréquentiste et fisherien. Mais en septembre 1955, j'ai été engagé par Morlat dans la Division Statistique qu'il dirigeait, au Service de Recherches Hydrauliques d'EDF et là, tout a changé pour moi. Tout d'abord, je souhaite à tout jeune débutant (j'avais 23 ans) de travailler d'emblée avec un patron tel que Morlat, avec ses grandes qualités humaines, sa façon de transmettre son expertise ainsi que les nombreuses occasions offertes à ses subordonnés de développer leur environnement. Ainsi ai-je pu rencontrer Marcel Boiteux avec lequel Morlat avait des liens privilégiés. Je suis resté quatre ans avec lui jusqu'à la dissolution de la division statistique par un nouveau directeur dont l'idée était de regrouper dans un seul service tous ceux qui utilisaient tant soit peu les mathématiques : IMA, pour Informatique et Mathématiques Appliquées<sup>5</sup>. Le résultat : Morlat est parti rejoindre Boiteux et moi-même, avec quelques autres compagnons d'exil, je suis parti faire un séjour de 2 ans à Gaz de France avant de revenir à EDF, rappelé par Joseph Jacquet en 1962.

La Division Statistique de Morlat à EDF a une histoire qui nous ramène de fait à Bayes. Elle est l'émanation d'un groupe de statisticiens constitué autour du normalien Etienne Halphen, dès la nationalisation de l'Électricité de 1946 et du Service Hydrotechnique regroupant les études des anciennes compagnies d'électricité. C'est Pierre Massé, futur président d'EDF, qui l'avait accueilli dans ces services dès 1940 pour le protéger des occupants allemands. Les travaux très remarquables d'Halphen dans les domaines de la statistique et de ses applications à l'économie de l'électricité se sont poursuivis et développés, après la nationalisation, avec l'arrivée en 1948 de deux jeunes universitaires Georges Morlat et Lucien Le Cam (ce dernier est parti ensuite rejoindre Neyman à l'Université de Californie à San Francisco). Halphen a poursuivi ses travaux en liaison étroite avec Morlat jusqu'à sa mort prématurée en 1954 (à 44 ans).

Engagé en septembre 1955, je n'ai donc connu Halphen que par Morlat et aussi par les papiers, notes internes et articles publiés par Halphen. Son influence sur Morlat et sur moi-même est très importante. Morlat a rapporté qu'Halphen lisait peu, sa pensée était cependant originale : il était persuadé qu'un statisticien devait avoir un rôle éthique dans ses relations avec « *son client* » en lui transmettant la cohérence des principes statistiques. Il ne connaissait pas Savage et je ne pense pas qu'il ait jamais lu de Finetti. Cependant, il avait une conception de la probabilité proche de la conception subjective et refusait le fréquentisme au moins pour les problèmes qu'il avait à traiter. « *Je ne fais pas du contrôle de fabrication* », disait-il ! Morlat avait acquis cette conception enrichie par la lecture de de Finetti (1937) et de Savage (1954). Avec lui, j'ai pu écouter Savage en 1958 lors d'une invitation à l'IHP. Je rappelle qu'outre ses activités d'enseignement à l'ISUP, Morlat était très engagé dans les activités de la RO, Recherche Opérationnelle, alors en très fort développement (époque de la création de la SOFRO et de sociétés telles la SEMA). Il m'avait poussé à donner des cours de statistique décisionnelle au CIRO (Centre Interarmées de Recherche Opérationnelle au Fort de Montrouge). De ces premiers pas avec Morlat, date mon intérêt sur les méthodes bayésiennes. En ce qui concerne ma

5. je ne peux m'empêcher de rapporter le slogan qui s'est propagé ensuite dans ce service "*l'IMA, c'est pas le PEROU !*"

propre activité de recherche à EDF, je me suis résolument tourné vers la mise en oeuvre de la philosophie et des méthodes bayésiennes surtout après mon retour aux Etudes et Recherche à Chatou en 1962. Un problème important était, à cette époque, la planification et le contrôle statistiques des expériences d'augmentation artificielle de la pluie sur les bassins alimentant les réservoirs. Dès 1962, on y a développé à la fois des techniques de test classiques et tenté la mise en oeuvre de décisions bayésiennes. Ces travaux ont fait l'objet d'une présentation au 5<sup>e</sup> *Berkeley Symposium on Mathematical Statistics and Probability* de 1965. Je rappelle que ces rencontres étaient organisées par Neyman et Le Cam que j'ai eu l'honneur de rencontrer à ce moment. Neyman était particulièrement intéressé par la pluie artificielle et était à l'origine des séances consacrées au sujet. C'est là que le paradigme bayésien et son concept de probabilité subjective sont particulièrement bien adaptés et, dirais-je, même nécessaires dans les problèmes de décision face aux risques provoqués par les événements extrêmes. Crues de rivières, tremblements de terre, tempêtes maritimes, etc. qui peuvent menacer les ouvrages EDF; ces divers aspects ont fait l'objet d'études selon Bayes. C'est ainsi que les premières études bayésiennes de crue ont été présentées dès 1967 au symposium AIRH de Fort Collins.

Avec d'autres collègues d'EDF et d'instituts partenaires, nous avons consacré une part importante de nos activités à des développements méthodologiques mis en oeuvre dans ces problèmes ainsi que des questions environnementales diverses (propagation de la pollution des rivières et l'estimation des modèles mathématiques hydrodynamiques, etc.). Dans ces périodes, notre pierre d'achoppement était bien sûr le manque de méthodes adéquates de calcul bayésien pour les modèles d'extrêmes. Nous avons fait avec les moyens du bord : utilisation des propriétés asymptotiques et approximations (lorsque c'était possible) par les modèles gaussiens ou transformés (logarithme, etc.)

Je parle ici de cette période d'avant 1990, appelons la pré-Robert, où on ignorait en France les méthodes de calcul MCMC et autres Gibbs, Metropolis Hastings, ... et dont la succession des divers logiciels de calcul modernes, BUGS, JAGS, STAN, etc. provoque encore aujourd'hui une véritable libération bayésienne !

## Statistique et statistique bayésienne dans l'enseignement en France à la fin du 20e siècle

**G. Celeux :** Vous nous dites que la statistique bayésienne avait commencé à se répandre en France dans les applications de recherche et développement assez tôt. Il me semble pourtant que le milieu académique français soit resté longtemps insensible à la théorie bayésienne, quoi que de nombreux développements apparaissent outre-Manche (Jeffreys, 1961) et outre-Atlantique (Raiffa et Schlaifer, 1961).

**É. Parent :** Je peux témoigner de mon expérience d'enseignant en Écoles d'Ingénieur vers 1990. Chargé de mettre en place un cours de statistique pour les élèves-ingénieurs de l'École des Eaux et Forêts à la fin des années quatre-vingts, j'opte néanmoins pour le paradigme classique en ce qui concerne le cours principal, mais avec des modules optionnels bayésiens pour l'analyse du risque (contrôle de la qualité et calcul des valeurs de projets en hydrologie). Pourquoi pas complètement bayésien ? Sans doute mu par une volonté de compromis pour le réaliser avec mon ami Azaïs (qui deviendra reconnu parmi les fréquentistes, voir par exemple Azaïs et Bardet (2006)) et beaucoup de conformisme pour s'inscrire plus facilement dans le programme classique : échantillons et lois empiriques/Estimation/Tests/Modèles sous l'hypothèse linéaire générale. À la même époque, Jean-Pierre Raoult<sup>6</sup>, me sollicite dans son équipe pour renouveler un cours de statistique, qualifié pourtant de décisionnelle à l'école des Ponts et Chaussées, mais

6. pourtant incontestablement intéressé par le caractère fructueux de l'approche bayésienne dans sa propre recherche puisqu'il a été le directeur de thèse de Christian Robert.

là encore il suit la piste fréquentiste sauf en un minuscule et unique paragraphe du nouveau manuel (que je peux rédiger). Le cas de l'Institut National Agronomique est particulièrement intéressant : dès les années soixante-dix, Guy Lefort a tenté de proposer aux étudiants un cours de statistique bayésienne (Lefort, 1975). Par un choix astucieux de la loi *a priori*, il retombe sur les formules ordinaires des tests de comparaison de moyennes, de variances, etc., ce qui permet de ne pas surcharger la mémoire des éventuels amateurs de recettes d'une boîte à outils statistiques standardisée. À sa mort prématurée, le cours de statistique donné par Duby et Daudin, retourne sous une forme plus classique qui deviendra celui enseigné aujourd'hui (Daudin, 2015).

**J. Bernier :** Je peux témoigner pour ma part de mon expérience à l'ISUP où j'ai commencé dès les années 1960 comme chargé de TD puis chargé du cours de décision statistique et tests en succédant à Morlat en 1968 jusqu'en 1990. Dans cette période la pensée statistique à l'ISUP était surtout influencée par le fréquentisme et Fisher. Le cours sur l'estimation de Jeanine Ulmo, parallèle au mien, était de philosophie fréquentiste en suivant les ouvrages de Lehmann. Je rappelle qu'après 1968, les jeunes enseignants universitaires de statistique et probabilités se réunissaient régulièrement pour « lire le Lehmann », réputé comme un ouvrage de statistique particulièrement « propre » sur le plan mathématique : Bourbaki avait aussi fait des émules parmi les statisticiens ! Entraîné par le courant ambiant, avouons-le, j'ai donc moi aussi fait du Lehmann dans mon cours. Toutefois, je ne laissais pas mes auditeurs ignorer que pouvaient exister certaines méthodes bayésiennes. Encore faut-il préciser que ma référence à Lehmann et ses présentations séparées des tests et de l'estimation reste fortement relativisée au départ par l'exposé de la synthèse décisionnelle de Wald.

Cependant j'ai, à la même époque, l'occasion de faire quelques cours de statistique dans d'autres enceintes (Ecole Centrale, Ponts et Chaussées, Eaux et Forêts) et, là, pour en illustrer les exposés, je n'hésite pas à présenter quelques travaux de statistique appliquée effectués à EDF où la philosophie du comportement individuel du chercheur devant l'incertitude et les méthodes bayésiennes jouent le rôle essentiel.

Je ne connais le monde enseignant statistique post-1990 que grâce à Éric Parent et je suis entièrement d'accord avec lui sur l'enseignement de la statistique en France comparé à ce qui est fait dans le monde anglo-saxon. Hors de France j'ai cependant eu le loisir de propager les idées bayésiennes et les travaux d'Halphen par une longue collaboration avec Bernard Bobée de l'Université du Québec.

Il faut reconnaître l'efficacité du système éducatif anglo-saxon. Il en résulte une production méthodologique de qualité et des outils bayésiens informatiques remarquables. On en connaît nombre de réalisations depuis les travaux précurseurs de Spiegelhalter : les logiciels *Winbugs*, *Bugs*, les utilisations des graphes représentant les modèles hiérarchiques généraux ainsi que les multiples *Bayesian packages* en R et l'aboutissement *Jags*, logiciel solveur Gibbs, particulièrement clair et pratique. On ne saurait oublier O'Hagan et ses travaux sur l'élicitation des lois *a priori* et le traitement des incertitudes et l'émulation des codes complexes (MUCM group - Sheffield University). Je peux me tromper mais il me semble que les contributions françaises à ces développements de *softwares* restent assez limitées. J'avais beaucoup parié sur le développement par l'équipe de Bordeaux du logiciel BIIPS (Todeschini et al., 2014), fondé sur des algorithmes particuliers, qui ne s'est malheureusement pas diffusé à la mesure de son mérite, me semble-t-il.

L'école américaine n'est pas en reste avec Gelman et ses contributions au développement de logiciels tels que *Stan*, le compétiteur de *Jags*, reposant sur les algorithmes Hamilton-Metropolis-Hastings, les travaux de Higdon sur les émulateurs de codes numériques, ceux de Raftery, etc.

Mais il ne faut certes pas laisser de côté les contributions nombreuses à la pédagogie bayésienne de Christian Robert et son école aussi bien dans le domaine des publications méthodologiques générales (*The Bayesian Choice 2001*, etc.) que dans les multiples applications particulières (comme les modèles de mélange). Sur le plan philosophique, il me plaît de signaler sa contribution à un article coopératif important (avec Gelman et autres) de 2019 *Abandon Statistical Significance*, paru dans le numéro 73 de *The American Statistician*.

Pour en revenir à la pédagogie statistique, il faut malgré tout reconnaître le déclin, après 2000, de la statistique décisionnelle traditionnelle à la Wald qui a dominé la seconde moitié du 20<sup>e</sup> siècle, malgré des ouvrages tels que Parmigiani et Inoue (2009). Elle semble remplacée par la théorie de l'apprentissage à la façon de Vapnik (2013).

Il me semble que c'est la statistique à la Lehman qui reste majoritairement enseignée en France au niveau universitaire...

**G. Celeux :** On peut contester cela... Je dirais plutôt que c'est la statistique non paramétrique qui est maintenant privilégiée.

**J. Bernier :** Peut-être, en tout cas la statistique cohabite avec la nouvelle mode du *data mining*. Il y a aujourd'hui toute une panoplie d'outils bayésiens utilisés en *data mining* et *machine learning*. Il faut noter que dans tous ces outils, les fondements bayésiens invoqués appartiennent à la version de la probabilité personnelle comme extension de la logique binaire Cox-Jaynes, concurrent de la probabilité subjective décisionnelle. Cependant l'adjectif décisionnel y est souvent accolé (voir par exemple Tufféry (2012)). Il m'apparaît que les décisions impliquées ont pourtant peu à voir avec les formalisations statistiques traditionnelles. Celles-ci sont-elles obsolètes ?

Et finalement, je dois dire que je suis chagriné que l'ISUP, mon douar d'origine, ait abandonné la statistique traditionnelle, y compris sous sa forme bayésienne pour changer son nom en *Data Centre* afin d'adopter l'enseignement de ces techniques regroupant logique floue, probabilités diverses, et méthodes associées, analyses descriptives des données<sup>7</sup>, réseaux bayésiens, etc., inventaire dans lequel on s'attendrait à retrouver les ratons laveurs de Jacques Prévert !

## Des bayésiens isolés et maudits ?

**G. Celeux :** La perspective bayésienne ne s'implante donc pas dans le milieu académique et de recherche français autant qu'elle vous semble présente dans le monde anglo-saxon. Les individus bayésiens se sont-ils sentis isolés ?

**É. Parent :** Tu as raison. Rouanet s'est alarmé par exemple que le raisonnement statistique perde sa cohérence (Rouanet et al., 1991), et je suis sûr, pour l'avoir rencontré, qu'il en souffrait presque d'un point de vue personnel. Lecoutre (1997) ne comprend pas qu'un statisticien bien éduqué ne puisse pas se rendre compte que le raisonnement bayésien se glisse diaboliquement sous les pierres de fondation de la statistique classique. En Angleterre, ce phénomène s'est produit plus tôt. Face à la prééminence de Fisher et de ses disciples, le bayésien Jeffreys (1961) a eu bien du mal à trouver sa place, mais à partir des années 60, c'est terminé : Lindley diffuse et affirme la pensée bayésienne, à la fois en théorie (Lindley, 1965) et la vulgarise efficacement (Lindley, 1991, 2006). Hors de France se fédère ensuite une communauté bayésienne anglo-saxonne forte autour des disciples de Lindley (Bernardo, Berger, Dawid, O'Hagan...) qui peut parfois donner à l'extérieur une impression quasisectaire. Ce monde anglo-saxon a exercé une

7. triomphe ultime de Benzecri !

forte attraction sur nombre de bayésiens français qui y ont fait ou y font encore de longs séjours : Christian Robert à Cornell puis à l'université de Warwick, Nicolas Chopin à l'université de Bristol, Arnault Doucet et Judith Rousseau à Oxford, moi-même à Cambridge. La statistique bayésienne ne se développe véritablement qu'au début des années 1990 – en France comme ailleurs – avec l'avènement des méthodes de simulations fondées sur les algorithmes d'importance sampling et de chaînes de Markov (Robert et Casella, 1998). Associées à la constante augmentation de la puissance de calcul de nos ordinateurs, elles ont complètement libéré, la puissance modélisatrice de l'approche bayésienne (Brooks, 2003). Spiegelhalter et son équipe développent puis offrent au domaine public le premier outil générique d'inférence bayésienne, le formidable logiciel gratuit *Bugs* (Spiegelhalter et al., 1996). En France, c'est avec l'école de Robert et al. (2004) et de Marin et Robert (2007) que se diffuse largement la perspective bayésienne dans l'université à partir des niveaux Master (Paris-Dauphine, Montpellier, etc.). Pour prendre le recul nécessaire avec cette vision trop focalisée sur les cinquante dernières années, on trouvera dans McGrayne (2011) l'intéressante épopée historique des protagonistes de la démarche bayésienne, *cette théorie qui ne voulait pas mourir*.

L'engouement actuel pour l'informatique a certainement fait basculer le débat du champ philosophique « *Quel caractère scientifique accorder à la probabilité bayésienne, revendiquée comme une quantification de la relation entre un individu et les manifestations incertaines de son objet d'études* » au cadre pratique de l'implémentation algorithmique « *Éviter de calculer une énorme intégrale afin d'évaluer une probabilité conditionnelle* ». Et, sous le poids de la nécessité technique partagée aujourd'hui avec d'autres, semblent être dissipées les aigreurs des querelles statisticiennes passées.

**J. Bernier :** Je ne suis pas d'accord sur ce soi-disant isolement des bayésiens dans la période pré-1990. Certes les difficultés calculatoires bayésiennes cessent à partir des années 1990, en France comme ailleurs. Mais elles n'ont pas empêché une production d'applications notables même partielles. Un ouvrage essentiel *Applied Statistical Decision Theory* de Pratt, Raiffa et Schlaifer en 1965 a marqué cette époque. On y trouve notamment la définition et une mise en oeuvre des distributions conjuguées naturelles dont l'importance s'est trouvé confirmée avec les algorithmes de Gibbs. Ces auteurs avaient publié dès 1964 dans JASA un article *The Foundations of Decision Under Uncertainty : An Elementary Exposition* exposant les bases élémentaires de l'axiomatique de Savage. Bien qu'assez tôt la théorie de la décision ait été l'objet de critiques notamment de la part de Maurice Allais concernant ceux qu'il appelait « *l'école américaine* » et l'axiome dit « *sure thing principle* » nécessaire à la mise en oeuvre de la formule de Bayes, ces discussions n'ont nui en rien aux développements des applications de la théorie, des méthodes bayésiennes et du critère de l'utilité espérée, au moins hors de France.

Sur le plan de la pratique et parmi d'autres applications, je citerais *Bayesian decision theory applied to design in hydrology* par Davis, Kisiel et Duckstein dans *Water Resources Research* en 1972, dont mes contacts après la parution m'ont permis de commencer une longue amitié avec Lucien Duckstein (Davis et al., 1979). Les bayésiens de cette époque n'étaient certes pas si isolés et maudits que cela !

## Quel avenir à la statistique bayésienne ?

**G. Celeux :** Nous voilà donc à l'époque du *Big Data*, de l'apprentissage automatique et de l'intelligence artificielle... Que pensez vous de l'avenir de la statistique bayésienne dans ces conditions peu propices ?

**J. Bernier** : L'importance<sup>8</sup> de l'informatique dans le développement des méthodes d'IA aujourd'hui ou celles des méthodes computationnelles bayésiennes au tournant du siècle ne doivent pas faire oublier la nécessité de la rigueur du langage, en particulier quand il s'agit d'incertitudes. Le paradigme bayésien s'impose naturellement lorsqu'un individu doit raisonner et agir face à son incertitude concernant les conséquences de ces actes. Le recueil d'informations et leurs traitements par modélisation probabiliste et algorithmes informatiques ne sont que contingents à l'objectif de base. On ne peut justifier Bayes ou n'importe quel autre paradigme uniquement sur des arguments de facilité de calcul.

L'incertitude ne décrit qu'un état psychologique d'un individu face à un environnement particulier où il doit agir. Pour cela il devra résoudre son problème de décision avec des données et au moyen de modèles traités par des algorithmes mobilisant des informations.

On ne peut donc analyser le tryptique *problème incertain/ données/ modèles & algorithmes de calcul* dans n'importe quel sens. Or il est regrettable que la tendance existe souvent dans la littérature technique où les choix dictés par les besoins de calcul l'emportent sur les choix, disons philosophiques, liés à la signification de l'incertitude et à son application au problème en main.

Dans cette optique, la mienne, on ne peut dire, par exemple, que l'incertitude se manifeste par la variabilité de répétition des observables. Cette variabilité là, dite source d'incertitude pour le statisticien, n'est qu'un état naturel des données<sup>9</sup>. Méfions nous des noms ou adjectifs fallacieux comme *l'intelligence* des données, autre nom d'usage du *data mining*.

**É. Parent** : Et même si on se revendique rigoureux en s'appuyant sur le calcul des probabilités, tout ne part pas de Kolmogorov ! Hájek (2003), comme tous les philosophes de la connaissance, distingue bien la question : *How should probability theory be formalized ?* qui est une question mathématique (à laquelle Kolmogorov apporte une réponse) et doit être séparée de : *What do statements of probability mean ?* qui est une question philosophique dont la réponse doit venir d'autre part, hors du domaine des mathématiques, c'est-à-dire de l'homme ou de son environnement naturel, et là, les réponses sont de nature subjectiviste ou fréquentiste. C'est aussi tout ce qui fait la différence entre Calcul de Probabilités et Statistique comme l'exprimait Halphen dans les années 1950 : « *Il est acquis que la statistique n'est scientifique que dans la mesure où elle se tient en liaison avec le calcul des probabilités. Mais elle ne se réduit pas à ce dernier, et ne s'y réduira jamais, pour la raison fort simple que le calcul des probabilités est une construction idéale vide de tout contenu réel : un mathématicien peut faire de la géométrie Euclidienne, Riemannienne, Hilbertienne... a priori : il faut un physicien pour savoir laquelle de ces géométries correspond à la réalité. On peut donc dire réciproquement que le calcul des probabilités ne sera fécond que s'il se tient en liaison avec la statistique : sinon il résoudra de beaux problèmes qu'il se sera donnés à lui-même pour son amusement, mais qui n'ont aucun intérêt* ». Kadane (2011) exprime la différence autrement : « *To be a good statistician requires grounding in each of the disciplines we rely on : mathematics, computing and philosophy* ».

**G. Celeux** : Cependant, l'inférence bayésienne bénéficie aussi des nouveaux moyens de calcul et la recherche en algorithmique bayésienne est particulièrement active.

**J. Bernier** : Je me réjouis de la nouvelle multiplicité des moyens de calcul à disposition du bayésien : algorithmes, logiciels (*Stan, Jags, etc.*), *packages* R et autres systèmes, actuels et en devenir. Il est clair que les statisticiens classiques en profitent aussi largement : ne peut-on pas

8. Son caractère ludique ne pourrait-il pas y être aussi pour quelque chose ?

9. État normal mais l'adjectif peut être trompeur car en météorologie par exemple, la moyenne de température a été affublée du nom de « normale » alors que la variabilité est un état tout aussi normal des processus météorologiques

considérer que l'algorithme EM n'est qu'un cas particulier de l'algorithme de Bayes variationnel (Keribin, 2010), si on étend la famille des approximations aux lois de Dirac ? Toutefois je pense qu'il est vital de ne pas laisser tomber les fondements bayésiens dans les décombres du passé, de même que les enseignements de la théorie de la décision statistique, seule façon de justifier le caractère dit décisionnel de certaines procédures à la mode. Comme le dit Éric : « *Tout ne part pas de Kolmogorov !* ». Mais l'axiomatique mathématique de la probabilité a un poids si fort aux yeux de certains qu'elle paraît incompatible avec « l'interprétation aléatoire des paramètres des modèles probabilistes ». Curieusement dans la préface de l'ouvrage d'Éric et de ses collègues (Albert et al., 2015), Jean-Michel Marin nous propose alors de discuter l'idée selon laquelle l'approche bayésienne se situerait à un autre niveau sémantique qui ne remet pas en cause cette modélisation de la réalité, en supposant que la loi *a posteriori* est un outil de résumé de l'information disponible sur les paramètres, sans remettre en cause également l'existence de ces paramètres inconnus certes mais non aléatoires. Malgré toute mon admiration pour cet ami d'Éric incontestablement bayésien, et, de surcroît, notre actuel président de la SFdS, je trouve cette position assez difficile à comprendre car ce niveau sémantique n'est jamais bien défini. Il me semble plus clair d'admettre avec Hajek et Halphen, comme le rapporte Éric Parent, que la statistique ne peut se réduire au calcul des probabilités désincarnées de Kolmogorov et que l'interprétation concrète de la probabilité demande une formulation adaptée concrètement au type d'incertitude dans laquelle est placée la personne concernée. S'agissant donc d'interprétation du concept de probabilité, deux positions cohérentes sont proposées. La première est l'interprétation logique nécessaire de Cox-Jaynes, extension de la logique binaire fondée sur des axiomes de comportement compatibles, après calculs, avec l'axiomatique de Kolmogorov. Nous n'avons pas suivi cette voie, la logique du robot raisonneur, car sur le plan de la mise en oeuvre, elle perd de sa cohérence au niveau de l'élicitation des lois *a priori* au moins pour de nombreux problèmes. Cette cohérence, comme demandée par Lindley (2006) est assurée par l'approche subjective décisionnelle de de Finetti, Savage, Raiffa, ... à laquelle Éric et moi adhérons.

**É. Parent :** Quel que soit le paradigme, d'accord, la modélisation a donc recours au calcul des probabilités.

Mais je pense qu'il nous faut refuser l'oecuménisme : même si la sauce mélangeant bayésien et classique a bien belle apparence, elle n'est pas digeste. Tandis que l'inférence classique requiert des techniques d'optimisation, la recherche de la loi *a posteriori* n'engendre pas une telle rupture technique : la loi *a posteriori* est elle-même un objet du calcul des probabilités et les algorithmes de simulation de Monte-Carlo font également largement recours à la mathématique des probabilités. Voilà où un même outil mathématique est à toutes les sauces et il y a là un véritable problème de compréhension et de communication, car un grand nombre d'exposés, vus sur le Net, au-delà d'une présentation mathématique axiomatique à la Kolmogorov, justifiant les concepts de variables aléatoires  $X$ , de densités de probabilités  $f(x, \theta)$  et de paramètres, ne font que distinguer verbalement :

- l'approche fréquentiste :  $\theta$  certain,
- et l'approche bayésienne où  $\theta$  incertain est supposé appartenir à une sigma-algèbre d'événements munie d'une densité  $\pi(\theta)$ .

Hélas, la démarche bayésienne y est (plus ou moins elliptiquement et très verbalement) *expliquée* par la simple distinction technique entre aléa sur  $X$  et incertitudes dites épistémiques sur  $\theta$ .

**J. Bernier :** Pour les futurs bayésiens, la vie sera plus simple : de Finetti a prouvé l'existence conjointe de la loi *a priori* et de la vraisemblance, et il s'agira simplement de les utiliser à bon escient. L'approche subjective décisionnelle est basée sur deux piliers :

- l'élicitation des probabilités *a priori*, formalisée par Savage (1971) avec la technique des

paris et justifiée par un corps de principes de comportement devant l'incertitude dont essentiellement le principe dit « *d'évitement des pertes sûres* » et le principe de prévision dit « *sure thing principle* ».

- l'acceptation du théorème de de Finetti, dit de représentation, reposant sur l'hypothèse d'échangeabilité simple ou partielle des observations. Cette hypothèse remplace l'hypothèse d'indépendance et de distributions identiques des observations sur laquelle se fonde la statistique fréquentiste. Le théorème est fondamental pour le comportement du bayésien devant l'information et justifie, de fait, l'interprétation aléatoire des paramètres des modèles statistiques. Ce théorème devrait être un outil de modélisation fondamental pour tout bayésien. Il faut noter que des auteurs comme Robert ou Gelman y accordent peu d'importance ou ne le citent même pas. Dans son livre fondateur de 1954, Savage, après de Finetti, le présente sans lui donner la même importance qu'à son système d'axiomes. Mais cela lui a permis, l'année 1955 suivante, de généraliser, avec Hewitt, le théorème aux processus d'information les plus généraux et qui est connu maintenant sous le nom de théorème de de Finetti-Hewitt-Savage.

Personnellement, j'ai évolué dans ma perception de la théorie subjective décisionnelle de la probabilité en déplaçant sa pierre angulaire pour les applications, de l'élicitation vers le théorème de représentation en accord, d'ailleurs, avec le jugement de Bernardo (1996). Bien entendu, mon opinion personnelle ne préjuge pas des orientations que suit ou suivra le monde des statisticiens du XXI<sup>e</sup> siècle, ma personne étant trop limitée à cet égard. Simplement mon jugement reflète une expérience assez longue de ma pratique en même temps que celle des autres, ceci grâce en soit rendue à Éric Parent qui me sollicite souvent pour me sortir de mon état de retraité.

**É. Parent :** La statistique, c'est ce que font les statisticiens. Si nous nous sentons heurtés aujourd'hui par la vague du *Big Data* et de l'intelligence artificielle, c'est que la dialectique qu'elle emploie nous met face à deux de nos travers.

- Nous travaillons sans doute trop seuls : c'est pour celà que le *data scientist*, comme ils disent, ressemble à un homme-orchestre, à la fois statisticien, informaticien, gestionnaire de données, analyste numérique et psychologue. Je suis surpris, par exemple, par l'irruption du bayésien dans les sciences cognitives contemporaines : nos modèles mentaux du monde verraient-ils leur plausibilité vis-à-vis des observations réévaluées selon la formule de Bayes (Seriès, 2016) ? Dehaene (2013) y voit même les fondements d'une inférence probabiliste chez le bébé!
- Nous nous faisons comprendre sans doute trop mal : c'est pour celà que la nouvelle vague veut changer la donne en changeant les termes : on ne fera plus de modélisation hiérarchique mais on construira un profond réseau de neurones ; on ne fera plus d'inférence, mais de l'apprentissage statistique voire un apprentissage automatique, algorithmique, adaptatif...

Mais sous ce grand et futile chamboulement linguistique, ne percevez-vous pas finalement la réminiscence de la formule de T. B. ? Une petite formule, jetée il y a bientôt trois siècles dans l'océan des connaissances mathématiques qui a généré des ondes qui pourraient bien s'achever en un tsunami.

## Références

Albert, I., S. Ancelet, O. David, J.-B. Denis, D. Makowski, E. Parent, A. Rau, et S. Soubeyrand (2015), *Collectif BioBayes : Initiation à la statistique bayésienne - Bases théoriques et applications en alimentation, environnement, épidémiologie et génétique*, Références sciences. Edition Ellipses, URL <https://hal.archives-ouvertes.fr/hal-01194144>.

- Allais, M. (1953), «Le comportement de l'homme rationnel devant le risque : critique des postulats et axiomes de l'école américaine», *Econometrica : Journal of the Econometric Society*, pp. 503–546.
- Azaïs, J.-M. et J.-M. Bardet (2006), «Le modèle linéaire par l'exemple régression, analyse de la variance et plans d'expériences illustrations numériques avec les logiciels r, sas et splus».
- Bernier, J., E. Parent, et J.-J. Boreux (2000), *Statistique Pour L'environnement. Traitement Bayésien de L'incertitude.*, Lavoisier, TEC et DOC, Paris.
- Boreux, J.-J., E. Parent, et J. Bernier (2010), *Pratique du calcul bayésien*, Springer Science & Business Media.
- Brooks, S. P. (2003), «Bayesian computation: A statistical revolution», *Trans. Roy. Statist. Soc., series A*, vol. 15, pp. 2681–2697.
- Cox, R. T. et E. T. Jaynes (1963), «The algebra of probable inference», *American Journal of Physics*, vol. 31, no 1, pp. 66–67.
- Daudin, J.-J. (2015), *Le modèle linéaire et ses extensions : Modèle linéaire général, modèle linéaire généralisé, modèle mixte, plans d'expériences*, Technosup (Paris), Ellipses.
- Davis, D. R., L. Duckstein, et R. Krzysztofowicz (1979), «The worth of hydrologic data for nonoptimal decision making», *Water Resources Research*, vol. 15, no 6, pp. 1733–1742.
- de Finetti, B. (1937), *La Prévision : ses Lois Logiques, ses Sources Subjectives*, Institut Henri Poincaré, Paris.
- Dehaene, S. (2013), «Les principes bayésiens de l'apprentissage : sommes-nous des scientifiques dès le berceau ?», in «<https://www.college-de-france.fr/site/stanislas-dehaene/course-2013-01-08-09h30.htm>», Collège de France.
- Ferguson, T. (1967), *Mathematical Statistics, a Decision Theoretic Approach*, Academic Press.
- Hájek, A. (2003), «Interpretations of probability», in «The Stanford Encyclopedia of Philosophy (Zalta)», CiteSeer.
- Jaynes, E. T. (1959), *Probability theory in science and engineering*, 4, Socony Mobil Oil Company Field Research Laboratory.
- Jeffreys, H. (1961), *Theory of Probability*, Oxford : Clarendon Press.
- Kadane, J. B. (2011), *Principles of Uncertainty*, Texts in Statistical Science, Chapman & Hall.
- Keribin, C. (2010), «Méthodes bayésiennes variationnelles : concepts et applications en neuroimagerie», *Journal de la Société Française de Statistique*, vol. 151, no 2, pp. 107–131.
- Lecoutre, B. (1997), «C'est bon à savoir ! Et si vous étiez un bayésien qui s'ignore ?», *Modulad*, vol. 18, pp. 81–87.
- Lefort, G. (1975), *Cours d'Introduction à la théorie de la décision et à la statistique bayésienne*, Institut National Agronomique Paris-Grignon.

- Lindley, D. V. (1965), *Introduction to Probability and Statistics from a Bayesian Viewpoint*, University Press.
- Lindley, D. V. (1991), *Making Decision*, John Wiley & Sons, INC, New York.
- Lindley, D. V. (2006), *Understanding Uncertainty*, John Wiley & Sons, INC, New York.
- Marin, J.-M. et C. Robert (2007), *Bayesian core: a practical approach to computational Bayesian statistics*, Springer Science & Business Media.
- McGrayne, S. B. (2011), *The Theory That Would Not Die*, Yale University Press.
- Morlat, G. (1964), «Statistique et théorie de la décision», *Revue de Statistique Appliquée*, vol. 12, no 2, pp. 5–13, URL [http://www.numdam.org/item/RSA\\_1964\\_\\_12\\_2\\_5\\_0](http://www.numdam.org/item/RSA_1964__12_2_5_0).
- Parent, E. et J. Bernier (2007), *Le raisonnement bayésien : modélisation et inférence*, Springer Science & Business Media.
- Parent, E. et E. Rivot (2012), *Introduction to hierarchical Bayesian modeling for ecological data*, Chapman and Hall/CRC.
- Parmigiani, G. et L. Inoue (2009), *Decision theory: Principles and approaches*, vol. 812, John Wiley & Sons.
- Pezier, J. (1972), *Décisions Rationnelles dans l'incertain*, Masson, traduction de "Rational Descriptions, Decisions and Designs" par M. Tribus.
- Raiffa, H. et R. O. Schlaifer (1961), *Applied statistical decision theory*, Graduate School of business administration-Harvard.
- Robert, C. P. et G. Casella (1998), *Monte Carlo Statistical Methods*, Springer-Verlag.
- Robert, C. P., N. Chopin, et J. Rousseau (2004), «Harold Jeffreys's theory of probability revisited», *Statistical Science*, vol. 24, pp. 141–172.
- Rouanet, H., M.-P. Lecoutre, M.-C. Bert, B. Lecoutre, et J.-M. Bernard (1991), «L'inférence statistique dans la démarche du chercheur», *Publications universitaires européennes. Série VI : Psychologie*.
- Savage, L. J. (1954), *The Foundations of Statistics*, Dover Publications, New York.
- Seriès, P. L. (2016), «Le cerveau est-il une machine bayésienne ?», in «Le Bayésianisme aujourd'hui», Éditions Matériologiques, Paris.
- Spiegelhalter, D. J., A. Thomas, N. G. Best, W. Gilks, et D. Lunn (1996), «Bugs: Bayesian inference using gibbs sampling», *Version 0.5,(version ii)* <http://www.mrc-bsu.cam.ac.uk/bugs>, vol. 19.
- Todeschini, A., F. Caron, M. Fuentes, P. Legrand, et P. Del Moral (2014), «Biips: software for Bayesian inference with interacting particle systems», *arXiv preprint arXiv :1412.3779*.
- Tribus, M. (1969), *Rational Descriptions, Decisions and Designs: Pergamon Unified Engineering Series*, Pergamon.

Tufféry, S. (2012), *Data mining et statistique décisionnelle : l'intelligence des données*, Editions Technip.

Ulmo, J. et J. Bernier (1973), *Éléments de décision statistique*, P.U.F.

Vapnik, V. (2013), *The nature of statistical learning theory*, Springer science & business media.

Von Neumann, J., O. Morgenstern, et al. (1953), *Theory of games and economic behavior*, Princeton University Press.

Wald, A. (1950), *Statistical Decision Functions*, Wiley.