ENTRETIEN AVEC UN-E STATISTICIEN-NE

Dans cette chronique « Entretien avec un-e statisticien-ne », nous partons à la rencontre de celles et ceux qui font la statistique, tant des praticiens qui l'utilisent comme un outil essentiel dans le cadre de leur activité professionnelle que des universitaires qui la développent dans le cadre de leurs travaux de recherche et d'enseignement. Une ou plusieurs série(s) de questions mettent en particulier en lumière le rapport à l'enseignement. Les praticiens sont appelés à exprimer les besoins de formation qu'ils jugent prioritaires, les compétences qu'ils apprécient chez ceux qu'ils recrutent. Les universitaires sont interrogés sur leur vision personnelle de la statistique et de son importance, sur leur art de la transmission du savoir statistique, entre théorie et applications, idéalisation professorale et réalités étudiantes.

MATHILDE MOUGEOT : LA STATISTIQUE CONNECTÉE

Mathilde MOUGEOT[1] et Gilles STOLTZ[2]

Mathilde Mougeot exerce actuellement comme maître de conférences à l'Université Paris Diderot. Elle avait d'abord occupé un poste de maître de conférences à l'Université Paris Ouest Nanterre La Défense[3], puis avait sollicité une mise en disponibilité pour participer à la création de Miriad Technologies et y travailler pendant plusieurs années. Elle a soutenu son habilitation à diriger des recherches début décembre 2015 ; son manuscrit est intitulé « Contribution to statistics and data science for industrial applications ; from neural networks to sparse linear models ».

L'entretien a commencé par téléphone le 23 juillet 2015 et s'est poursuivi par des échanges de courriels entre le 6 et le 25 septembre 2015.

Parcours

GS : *Mathilde, merci infiniment d'être le « proof of concept » de notre nouvelle rubrique ! (Nous reviendrons sur cette notion plus tard.) Pour l'heure, peux-tu présenter le début de ton parcours statistique en quelques mots ?*

MM : Je suis diplômée de l'AgroParisTech. J'ai validé ma dernière année d'école par le DEA[4] « Intelligence artificielle et reconnaissance des formes » à l'Université Pierre et Marie Curie. En

[1]Université Paris Diderot, France, mathilde.mougeot@univ-paris-diderot.fr
[2]HEC Paris, CNRS, Jouy-en-Josas, France, stoltz@hec.fr

[3]Dans la suite de l'entretien, nous nous référerons aux différentes institutions académiques par leur nom actuel, et pas par le nom qu'elles portaient à l'époque évoquée par Mathilde Mougeot.

[4]Diplôme d'Études Approfondies, actuellement master 2 recherche.

© Société Française de Statistique (SFdS), Décembre/December 2015

1987, mon stage de DEA a eu pour objet d'utiliser les réseaux connexionistes pour l'étude de la prévision de vent réel pour le bateau français de la course de l'*Admiral's Cup*. D'ailleurs, quatre ans plus tard, un bateau français remportait pour la première fois cette course !

A cette époque, j'ai rencontré Robert Azencott qui m'a proposé de m'encadrer en thèse CIFRE[5] en collaboration avec ce qui s'appelait à l'époque Thomson-CSF (Thales de nos jours). Ma thèse a eu pour objet l'apport des réseaux connexionistes pour la compression d'images et la modélisation du système visuel des mammifères. En 1992, j'ai accepté un poste de maître de conférences à l'Université Paris Ouest Nanterre La Défense et j'ai continué à travailler avec Robert Azencott, dans un cadre universitaire, sur des méthodes d'apprentissage statistique pour résoudre des problématiques industrielles.

La première application [1] a été réalisée en collaboration avec Olivier Catoni, un autre de ses anciens doctorants, pour la Société Européenne de Propulsion (la SEP, désormais incorporée au groupe Safran) ; elle a porté sur la création d'une méthode de surveillance du moteur Vulcain (celui de la fusée Ariane) à partir de signaux numériques acquis pendant les 12 premières secondes de démarrage du moteur, avant le décollage de la fusée.

Quelques années plus tard, j'ai pris une disponibilité pour exercer dans le secteur privé : pour participer à la création de la startup Miriad Technologies...

GS : *Pas si vite ! Je voudrais d'abord revenir sur ce que tu as fait en thèse, puis remonter plus loin dans le passé, aux sources de ton choix en faveur de la statistique. Tout d'abord, je ne connaissais pas le nom de « réseaux connexionistes » mais une recherche rapide me dit que ce terme se rapporte aux réseaux de neurones ? Peux-tu nous dire en quelques phrases d'où vient cet ensemble de techniques, quels sont ses succès pratiques et ses difficultés mathématiques, et également, je crois qu'il a connu une période de désert et connaît actuellement une résurgence phénoménale sous le vocable de* deep learning *?*

MM : Les réseaux connexionistes sont des modèles qui, historiquement, s'inspirent du comportement des neurones tel qu'observé en biologie. En mathématique et en informatique, il s'agit de modèles fondés sur une architecture composée d'unités élémentaires (les neurones), interagissant entre eux par des règles et réalisant, par exemple, des fonctions de régression ou de classification par l'optimisation d'une fonction de coût. Le premier modèle fondé sur des réseaux de neurones et ayant rencontré un certain succès [5] avait pour but la reconnaissance de caractères manuscrits. Il reposait plus précisément sur un modèle de neurones multi-couches associés à l'algorithme de rétropropagation de gradient. Ce modèle a eu pendant très longtemps les meilleurs taux de reconnaissance. Au fil des années, les *deep scattering neural networks* l'ont remplacé. Ces modèles ont souvent été boudés des mathématiciens, car leur succès était essentiellement lié à des performances obtenues sur des applications spécifiques et n'était pas soutenu par des propriétés mathématiques établies. Aujourd'hui, des mathématiciens comme par exemple Stéphane Mallat se penchent sur ces outils et tentent justement d'en expliquer les propriétés mathématiques.

Une autre activité de recherche également liée aux réseaux connexionistes a pour objet de construire des modèles d'évolution qui utilisent des règles d'apprentissage inspirées de la biologie. On peut citer la règle de Hebb et les travaux d'Elie Bienenstock et Stephen Grossberg [2, 4] pour expliquer l'organisation des systèmes biologiques.

[5] Conventions Industrielles de Formation par la REcherche : mode de financement d'une thèse où un industriel est l'employeur du doctorant et où un contrat de collaboration est passé entre le laboratoire des encadrants du doctorant et l'industriel.

© Société Française de Statistique (SFdS), Décembre/December 2015

Pendant ma thèse, je me suis intéressée à ces deux aspects, en étudiant dans un premier temps l'apport des modèles connexionnistes pour la compression d'images, et en m'intéressant ensuite à leur intérêt pour l'auto-organisation du système visuel des mammifères.

GS : *Revenons-en maintenant plus loin dans le temps : où étais-tu à 18 ans, comment se sont déroulées tes premières années d'études supérieures et surtout, où, quand, et avec qui tu as rencontré puis choisi la statistique ?*

MM : A 18 ans, j'étais en classes préparatoires à Versailles, dans ce qui correspond actuellement à la filière BCPST[6]. Je me rappelle encore mon premier cours de statistique, dispensé en deuxième année par Marie Cottrel : lumineux ! Plus tard, je suis arrivée (revenue !) à la statistique par le traitement d'images et les applications, plus particulièrement par la télédétection. Je voulais suivre l'évolution des cultures à l'aide d'images satellites et proposer (estimer) des indicateurs de suivi de performances. C'est donc par des besoins soulevés par les applications que j'ai repris contact avec la statistique. Je me suis rapprochée de Robert Azencott dans cette optique.

GS : *Tu as beaucoup cité Robert Azencott, et Marie Cottrel vient de faire son apparition dans ton parcours. Y a-t-il d'autres figures tutélaires ou des compagnons de route statistique qui t'ont marquée ?*

MM : J'ai une pensée pour Lucette Carter qui a été un merveilleux guide de pédagogie lorsque j'ai commencé mon métier d'enseignante à Paris Ouest Nanterre La Défense.

Je pense également à mes compagnons de route lors de la création de Miriad Technologies ; pour n'en citer que quelques-uns : Jérôme Lacaille (aujourd'hui expert émérite à Safran), Bruno Durand (désormais expert en fusion de données chez Renault), Charle-Albert Lehalle (actuellement gestionnaire R&D de fonds financiers), au milieu de bien d'autres.

De retour en poste à l'université, j'ai dans un premier temps collaboré avec Karine Tribouley sur des sujets plus « académiques ». Par la suite, en poste à l'Université Paris Diderot, j'ai travaillé avec Dominique Picard sur des sujets plus fondamentaux et sur des collaborations industrielles. J'ai alors pris conscience du fossé qui existait entre la recherche universitaire et la R&D que je menais à Miriad Technologies ! La réinsertion dans le milieu académique fut un travail de longue haleine...

GS : *Où en es-tu actuellement ? As-tu atteint un point d'équilibre ?*

MM : L'équilibre à maintenir est permanent lorsque l'on souhaite travailler sur plusieurs fronts en même temps : à la fois la réalisation d'applications et une recherche plus académique... sans oublier l'enseignement ! Mon compromis actuel est de mener une recherche académique sur des contrats industriels en proposant essentiellement des études de faisabilité associées à des prototypes (légers) mais en écartant la réalisation de logiciels industriels qui nécessiteraient maintenance et support.

Expérience dans le privé

GS : *C'est le moment de revenir sur ta participation à la création de la startup Miriad Technologies : que faisait cette société, et peux-tu nous citer une de tes réalisations marquantes pour elle ?*

[6]Biologie, chimie, physique et sciences de la terre.

© Société Française de Statistique (SFdS), Décembre/December 2015

MM : La société avait pour objectif de commercialiser des logiciels pour proposer de l'aide à la décision à partir de données industrielles. Une petite dizaine de docteurs initialement dirigés par Robert Azencott se trouvaient enrôlés dans cette aventure. Miriad Technologies a été créée par des chercheurs, mais très rapidement, des investisseurs sont rentrés dans le capital... et un choc culturel s'est opéré.

Une des réalisations dont je suis le plus fière pendant mon séjour à Miriad Technologies est le développement d'un logiciel de surveillance et de diagnostic des compresseurs pour Air Liquide [3]. Après une preuve de l'intérêt de la méthode proposée (« proof of concept »), nous avons été amenés à développer un logiciel pilote qui a été installé pendant six mois à Dunkerque sur un site industriel. En visite sur ce site industriel, un responsable de la filiale Air Liquide America a déclaré : « I want it ». Six mois après, nous déployions ce logiciel sur le site opérationnel de Houston, qui supervise l'ensemble des compresseurs de la côte Sud-Ouest des Etats-Unis. Lors de la recette logicielle, le bon fonctionnement du logiciel a été testé à l'aide d'un plan d'expérience « grandeur nature ». Cela a été une expérience absolument unique, et la preuve par l'exemple que des logiciels fondés sur des méthodes d'apprentissage statistique pouvaient servir dans l'industrie et apporter une aide au diagnostic en temps réel.

GS : *« Recette logicielle » ? Tu as bien intégré le vocabulaire de l'entreprise et ses codes... Raconte à nos lecteurs à quel point tu les as vécus.*

MM : Lors de mes premières années à Miriad Technologies, je me suis très vite aperçue que pour convaincre de l'intérêt d'une solution technique (avant comme après une vente), il fallait la plupart du temps développer des arguments non techniques, facilement accessibles à nos interlocuteurs, qui n'étaient pas tous des statisticiens. L'utilisation exclusive d'arguments purement techniques pouvait avoir un effet nocif car ils rendaient la solution obscure. Parce que j'étais intéressée par ce côté rhétorique, j'ai été amenée à occuper partiellement un poste technique de soutien à la vente : mon rôle était de présenter et de proposer des solutions techniques à des interlocuteurs qui n'étaient ni statisticiens ni mathématiciens. Il y avait dans ce travail une forte similarité avec les aspects de pédagogie dans l'enseignement, ce qui d'ailleurs m'intéresse particulièrement.

GS : *Tu es donc bien placée pour développer ce que tu appelais le « choc culturel » dont tu parlais précédemment. Et d'ailleurs, profites-en pour nous offrir ta vision comparative des avantages et inconvénients de la profession universitaire que tu ré-exerces et du travail dans une startup.*

MM : Je suis rentrée à Miriad Technologies avec uniquement des compétences techniques, majoritairement en statistique et en traitement du signal. Quand j'ai quitté Miriad Technologies, j'avais travaillé étroitement pendant six ans avec le service informatique pour élaborer des solutions logicielles intégrant des bases de données, des interactions homme-machine, et des tests intensifs de programmes ; avec le service commercial pour la vente des études de faisabilité ; avec le service marketing pour réfléchir à la meilleure mise en valeur des solutions proposées ; avec le service de comptabilité et de gestion pour définir les recettes d'une livraison ; avec la direction pour l'orientation stratégique de la société ; etc. Pour avancer, il était vital de travailler en équipe, de communiquer avec tous les acteurs de la startup, qui avaient des profils et des logiques radicalement différents. Cela a été, au cours de ma carrière, une période d'une rare intensité et d'une grande richesse !

En 2005, j'ai souhaité reprendre mon poste d'enseignant-chercheur avec l'objectif d'approfondir mon activité de recherche par rapport à mon activité de développement ; en effet, le travail

© Société Française de Statistique (SFdS), Décembre/December 2015

à Miriad nécessitait d'enchaîner de plus en plus les études de faisabilité et les moments de recherche étaient devenus de plus en plus rares. Cela dit, comparé à mon quotidien à Miriad Technologies, mon travail d'enseignant-chercheur est moins soumis à des stimulations extérieures, mais pas plus reposant pour autant !

Questionnaire à la Proust

GS : *Avant de nous intéresser à la seconde partie de ta carrière, lorsque tu en reviens donc à un poste universitaire, et à la vision nouvelle que tu ne manques pas d'avoir sur l'enseignement de la statistique, je voudrais nous proposer un moment de détente : un questionnaire de Proust (ou en tout cas, sa version statistique). La règle du jeu est qu'il faut donner une réponse courte, à tout le moins pour la plupart des réponses...*

Quel est ton résultat de statistique préféré (théorème ou application) ?

MM : Le théorème de la limite centrale.

GS : *Un jour, un collègue m'a fait remarquer qu'il fallait traduire* central limit theorem *par « théorème limite central », parce que ce théorème limite universel était central aux calculs des probabilités et à la statistique, et que cela n'avait rien à voir avec le fait que la loi limite était centrée. Penses-tu que ce genre de considérations intéressent nos étudiants ?*

MM : Pourquoi pas ! Tu me l'apprends. Je pense que les deux points de vue sont justes. Le second point de vue a forcément été donné bien plus tard, bien après les premières preuves du théorème. On avait alors tout le recul pour caractériser l'importance « centrale » de ce théorème.

GS : *Quel est ton manuel de statistique préféré ?*

MM : *All of Statistics* de Larry Wasserman [6] parce que je trouve que c'est un livre qui se lit comme un roman et qui aborde plein de notions fondamentales.

GS : *Fréquentiste ou bayésienne ?*

MM : Fréquentiste par formation, mais qui sait...

GS : *Statistique ou statistiques ?*

MM : Statistique.

GS : *Quelle est la faiblesse principale de la statistique ?*

MM : Les données ! Tout serait si simple dans un monde d'hypothèses.

GS : *Quelle est ta vertu préférée en statistique ?*

MM : Le rapport au réel.

GS : *Quelle est notre principal défaut en tant que statisticiens ?*

MM : Mon principal défaut : vouloir tout expliquer à partir des données, passer un temps indécent à explorer de nouveaux modèles (en *clustering*, en régression) afin d'améliorer les performances de l'application considérée.

GS : *Pourquoi la recherche mathématique est-elle masculine et le monde de la statistique est-il plus mixte ?*

MM : Joker ! Je crains de n'énoncer que des lieux communs.

© Société Française de Statistique (SFdS), Décembre/December 2015

GS : *Connais-tu la SFdS, que t'apporte-t-elle ?*

MM : Un congrès annuel, des conférences thématiques intéressantes, des rencontres sympathiques... Je recommande l'adhésion !

Enseignement aux non-spécialistes

GS : *Nous arrivons à ce qui est peut-être le cœur de cet entretien : ta vision de l'enseignement de la statistique. Peux-tu commencer par nous brosser un tableau rapide des différentes filières où tu as été amenée à enseigner, et quels cours de statistique tu as pu donner ?*

MM : A l'Université Paris Ouest Nanterre La Défense, j'ai enseigné dans la filière MASS (Mathématiques Appliquées aux Sciences Sociales) aux niveaux L3 et M1, et dans le cursus ISIFAR (Ingénierie Statistique et Informatique de la Finance, de l'Assurance et des Risques) aux niveaux M1 et M2.

Aujourd'hui, à côté de cours de statistique classiques (estimation, tests), j'enseigne le *data mining* (la fouille des données) et la *data science* (science des données) dans le M2 de modélisation aléatoire (c'est-à-dire en statistique, probabilités et finance) de l'Université Paris Diderot, et en troisième année de l'Ecole centrale de Paris. J'ai été également amenée à dispenser des formations aux professeurs du secondaire dans la perspective de présenter des exemples d'applications métiers où les statistiques peuvent intervenir.

GS : *Ma question la plus importante serait : qu'est-ce qui a changé dans ta manière d'enseigner après ton passage dans le secteur privé ?*

MM : Aujourd'hui, je ne peux plus introduire un concept de statistique sans mettre en perspective les applications industrielles que j'ai côtoyées et sans évoquer aux étudiants les différents métiers et domaines où un travail – passionnant ! – sur les données les attend. Je cite quelques exemples de telles applications. Quels sont les facteurs influents qui caractérisent la qualité d'une poudre cosmétique de luxe (à partir de valeurs scalaires, de données fonctionnelles) ? Comment suivre le procédé de distillation du cognac pour anticiper l'instant de coupe lié à sa qualité ? Et, plus sérieusement, comment mettre en place des méthodes de surveillance automatique d'équipements industriels ?

Dans mon travail d'enseignante, je simule le plus souvent des données qui reflètent des applications sur lesquelles j'ai travaillé, et je les propose aux étudiants en parallèle d'une question métier et d'une illustration statistique sous-jacente.

Je m'efforce également de présenter aux étudiants des applications potentielles (en devenir) : des applications qui peuvent se développer autour des données, comme par exemple les données liées à l'« Internet des objets » où les métiers de demain sont à construire.

GS : *« Internet des objets » ? Tu as plutôt intérêt à nous raconter cela en détails, dans la partie suivante de mise en perspective temporelle ! Laisse-moi pour l'instant continuer sur une perspective un peu générale : dis à nos lecteurs pour quels cours et quelles filières cette manière d'enseigner à partir de données et d'applications te semble particulièrement cruciale. Parce qu'en L3 MASS, cela éveille peut-être la curiosité des étudiants, mais ce n'est pas une nécessité...*

MM : Je pense avant tout à un cours de M2 créé en 2006, au sein du Master ISIFAR de Paris Ouest Nanterre La Défense : un cours de consultance, aujourd'hui dénommé « statistiques indus-

© Société Française de Statistique (SFdS), Décembre/December 2015

trielles ». Ce cours est essentiellement motivé par des applications industrielles réelles, fruits de mon expérience. Les méthodes statistiques sont justifiées à partir des problématiques réelles et appliquées, pour le coup, sur des données réelles.

GS : *Peux-tu nous donner un exemple d'astuce métier que tu transmets à ces étudiants ?*

MM : En entreprise, on distingue les métiers de la recherche et de la production. Un modèle statistique est mis au point dans un service de recherche. S'il a montré de bonnes performances, ce modèle a pour but de quitter le service de recherche où il a été conçu le plus souvent avec grand soin pour aller vivre sa vie en production. Il me paraît alors essentiel de garantir la justesse des sorties calculées par le modèle dans un environnement de production qui peut être en constante évolution. C'est pourquoi j'incite toujours mes étudiants à mettre en place, en parallèle par exemple d'une sortie calculée par un modèle de prévision, un outil de *scoring* (d'évaluation) qui fournisse en même temps un indicateur sur la pertinence de la sortie calculée par rapport aux données qui ont servi à la calibration du modèle. Il faut informer l'utilisateur sur un possible mésusage de la solution statistique !

GS : *Si tu me permets de changer un peu de sujet, je sais que tu as monté un cours qui sort de l'ordinaire, destiné aux étudiants littéraires de l'Ecole normale supérieure... et pour cause : l'aventure m'avait été proposée, mais j'avais bien trop peur de la charge de travail afférente ! Comment as-tu trouvé l'expérience ?*

MM : Ce cours, « Outils statistiques pour littéraires », a été créé avec Claire Zalc, qui est historienne et a écrit un ouvrage [7] de méthodes quantitatives pour historiens ; le tout, à l'initiative et à la demande de Magali Reghezza et de Claude Viterbo, alors directeurs des enseignements respectivement des départements de géographie et de mathématiques de l'Ecole normale supérieure.

Ce cours conduit certes à introduire des notions statistiques mais il vise également à convaincre des étudiants littéraires de l'intérêt d'utiliser la statistique pour leurs études quantitatives. Par exemple, le théorème limite central(e) est présenté aux étudiants dans sa version mathématique, mais des arguments non techniques sont utilisés pour le faire comprendre et motiver son utilité pour des perspectives d'application, liées aux problèmes de recherche de ces étudiants.

Je te confirme que la préparation de ce cours a été chronophage, mais je ne t'en tiens pas rigueur ! J'ai passé pas mal de temps à préparer des exemples et des données qui pouvaient motiver et interpeller ce public d'étudiants en géographie, en histoire et en sciences sociales. Mais, à l'arrivée, quelle récompense ! J'ai eu la satisfaction d'observer quelques étudiants littéraires (issus d'un baccalauréat L) « embrasser » le quantitatif et la notion d'aléatoire. Je pense notamment à une étudiante qui avait mené toute une étude quantitative descriptive puis de modélisation sur les maîtres orfèvres à partir de données d'archives, progressivement récoltées. Ce n'était pas du *big data* mais son rapport de recherche était passionnant et son travail, époustouflant !

GS : *Dans cette veine de la transmission parfois source de défis : quelle est la notion ou quel est le point de méthode que tu trouves la ou le plus difficile à enseigner (pour toi) ou à recevoir (pour les étudiants) ?*

MM : J'éprouve le plus de plaisir à enseigner des méthodes que j'ai croisée dans la réalisation d'applications, car je sais les mettre en perspective. Au contraire, j'ai toujours eu un du mal à motiver mes étudiants à exploiter les tests statistiques sur les coefficients pour les modèles linéaires.

© Société Française de Statistique (SFdS), Décembre/December 2015

GS : *Je voudrais conclure cette partie en revenant à la présentation initiale de tes activités d'enseignement. Tu disais que tu avais également réalisé des formations aux professeurs du secondaire : peux-tu nous en dire davantage à ce sujet ?*

MM : Laisse-moi préciser le contexte, tout d'abord : les enseignants du secondaire sont aujourd'hui chargés d'introduire des notions de statistique alors que pour une grande majorité d'entre eux, ils n'ont pas reçu de formation à ce sujet pendant leurs études. Mes interventions ont eu lieu dans des journées de programmes académiques de formation (à Caen et en région parisienne) et avaient pour objet de motiver l'intérêt de la statistique pour la résolution de problématiques industrielles... j'en reviens toujours au même ! Mais sur le plan technique, mon propos se limitait aux notions présentées dans l'enseignement secondaire, comme par exemple, la loi gaussienne et les intervalles de confiance.

Plus précisément, j'avais présenté à ces occasions les méthodes SPC (*Statistical Process Control* et « Six Sigma », très utilisées dans l'industrie pour la surveillance et la maîtrise de la qualité produit, par exemple pour des pièces manufacturées. J'avais également évoqué la méthode statistique mise en place dans le logiciel CANARY[7], qui illustre l'utilisation de la loi binomiale et du théorème de Moivre-Laplace pour la surveillance de la qualité de l'eau aux États-Unis.

GS : *Merci Mathilde pour ce nouvel aperçu de la vie en entreprise, qui revient toujours au galop. Il pique à nouveau ma curiosité : peux-tu décrire rapidement l'objet et le fonctionnement des méthodes SPC et « Six Sigma » ?*

MM : Avec plaisir, et je vais tenter une brève description comme celle utilisée pour les enseignants en lycée. Le SPC a pour but d'anticiper des défaillances dans la production de produits manufacturés. Cette méthode produit des résultats graphiques facilement interprétables par des non-statisticiens sur l'évolution au cours du temps d'un critère de qualité important (par exemple, la longueur d'une pièce ou son poids). Dans un cas très simple de SPC, on fait l'hypothèse que le critère de qualité suit une loi normale ; on a par ailleurs en tête une distribution statistique attendue (une distribution de référence, qui modélise le fait que le degré de qualité désiré est atteint et donc que le processus est sous contrôle). Des « cartes de contrôles » relatent alors l'adéquation des pièces produites à cette distribution de référence. Formellement, il peut s'agir de l'indication de scores z, indiquant à combien d'écarts-types de la moyenne de référence se situent les différentes pièces.

La méthode « Six Sigma » est une démarche de suivi et d'amélioration de la qualité, qui repose largement sur le SPC. Elle a bien un lien avec 6σ, où σ serait l'écart-type de la distribution de référence, ... mais il serait trop long de commencer à détailler cela. Je renvoie le lecteur intéressé à une recherche Internet sur le sujet : de nombreuses pages, y compris des pages maintenues par des universitaires, parlent de cette méthode.

En tout état de cause, dès les tout premiers cours de statistique, on peut ainsi illustrer des concepts statistiques simples réellement utilisés en pratique dans l'industrie.

Mise en perspective temporelle

GS : *Dans cette partie de conclusion, je voudrais te faire mettre en perspective passé, présent et avenir. Et revenir notamment comme promis sur l'« Internet des objets » comme source fu-*

[7] Voir https://software.sandia.gov/trac/canary.

© Société Française de Statistique (SFdS), Décembre/December 2015

ture d'applications prometteuses ! Mais avant cela, essayons de situer le présent. Comment résumerais-tu l'évolution de notre métier depuis le début de ta carrière, quelles évolutions te frappent le plus ?

MM : Lorsqu'on enseigne à un niveau où la grande majorité des étudiants cherchent un travail directement après leur formation, on ne peut plus faire l'impasse sur les cours appliqués de type *data science* et donc, sur des séances de travaux pratiques. La plupart des entreprises cherchent des étudiants ayant un bon niveau d'agilité dans le traitement des données (au sens large) et dans l'utilisation des logiciels permettant de mettre en place rapidement des algorithmes de fondements statistiques. Je pense à `R` ou `Python` par exemple. Cela dit, il me semble important d'introduire également dans les formations de statistique des notions de programmation et de complexité algorithmique, et pas uniquement la manipulation de logiciels.

GS : *Constates-tu une désaffection actuelle des étudiants envers la statistique, comme cela semble être le cas pour l'ensemble des disciplines et carrières scientifiques ?*

MM : Pas du tout ! Je constate au contraire un afflux très important d'étudiants dans les cours liés à la statistique. Lorsqu'avec Nicolas Vayatis nous avons commencé en 2009 un cours de fouille des données (*data mining*) à l'Ecole centrale de Paris, il y avait 25 étudiants inscrits dans l'option de mathématiques appliquées de la troisième et dernière année de scolarité. Depuis ces cinq dernières années, plus de soixante-dix étudiants souhaitent chaque année suivre cette option : ils sont pour la plupart très motivés et ont parfois travaillé pendant une année de césure sur des problématiques de fouille des données.

J'ai pu observer une augmentation similaire des effectifs dans les cours de statistique du master de modélisation aléatoire depuis que j'enseigne à l'Université Paris Diderot.

GS : *A quoi attribues-tu cela ? J'avais promis que nous reviendrions sur le concept d'« Internet des objets » (IoT :* Internet of things*), je parie que c'est le moment !*

MM : Pour définir l'IoT à nos lecteurs qui ne connaissent pas ce concept, je vais citer l'introduction de la page `Wikipédia`[8] qui y est consacrée :

> L'Internet des objets [...] représente l'extension d'Internet à des choses et à des lieux du monde physique. Alors qu'Internet ne se prolonge habituellement pas au-delà du monde électronique, l'Internet des objets représente les échanges d'informations et de données provenant de dispositifs présents dans le monde réel vers le réseau Internet. L'Internet des objets est considéré comme la troisième évolution de l'Internet, baptisée Web 3.0 (parfois perçu comme la généralisation du Web des objets mais aussi comme celle du Web sémantique) qui fait suite à l'ère du Web social. L'internet des objets est en partie responsable de l'accroissement du volume de données générées sur le réseau, à l'origine du *big data*. L'Internet des objets revêt un caractère universel pour désigner des objets connectés aux usages variés, dans le domaine de la e-santé, de la domotique ou du *Quantified Self*[9].

Je travaille depuis quelques mois sur la mise en place d'un cours où les étudiants travailleraient sur leurs propres données, acquises par l'intermédiaire de capteurs mobiles qu'ils porte-

[8] Page https://fr.wikipedia.org/wiki/Internet_des_objets, consultée le 23 septembre 2015.

[9] Traduction des auteurs : la« mesure de soi » ou l'« auto-mesure », par exemple au cours d'une journée, de ses caractéristiques physiologiques comme sa température ou sa pression artérielle, ou du nombre de pas effectués.

© Société Française de Statistique (SFdS), Décembre/December 2015

raient, peu onéreux car par exemple inclus dans leurs *smartphones* (ce serait l'idéal). J'envisage d'introduire les concepts statistiques à l'aide de ces données pour des études individuelles ou des études de cohortes. Depuis deux à trois ans, j'ai pu observer que les étudiants sont de plus en plus intéressés, à la sortie de leur M2, par travailler dans de toutes petites entreprises ou même de monter leur propre structure. Il est aujourd'hui facile de récupérer des données issues de capteurs spécifiques et, par ailleurs, beaucoup de données circulent sur Internet, parfois librement (ou sont issues d'Internet). Ainsi, une idée, un concept, du travail, et une petite touche en plus peuvent permettre de transformer la *data* en service à forte valeur ajoutée. Certains étudiants l'ont compris et n'hésitent pas à se lancer dans l'aventure !

GS : *Outre cet impact d'IoT sur la taille et la motivation de notre auditoire, quelles évolutions prévois-tu dans notre manière d'enseigner ? Ou, à défaut de prévisions, as-tu des points d'inquiétude ou d'attention à soulever ?*

MM : L'enseignement de la statistique nous offre la possibilité de motiver nos étudiants en présentant très facilement des applications sur des méthodes enseignées dans les domaines qui rejoignent leur curiosité, comme par exemple les réseaux sociaux. Cela étant, pour ces données et pour celles de l'IoT, des compétences informatiques sont nécessaires pour les récupérer, pour les archiver dans des bases de données, parfois pour les traiter sur des architectures distribuées. Pour pouvoir former des étudiants habiles avec les données et sachant notamment les récupérer, nous avons besoin de proposer des formations jointes avec nos collègues informaticiens.

Par ailleurs, et c'est là une inquiétude liée à la situation budgétaire des universités, il faut prendre garde à ce que l'enseignement de la statistique aux non-spécialistes ne nous échappe pas. Pour les autres départements d'enseignement de nos universités, rapatrier un cours de statistique, c'est donner plus d'heures de service à des enseignants-chercheur de ces départements, au détriment des départements de mathématiques. C'est pourquoi il faut tout particulièrement soigner nos cours et entrer en discussion, au sein de nos cours, avec les disciplines concernées, afin d'élaborer des cours pertinents et adaptés.

Le mot de la fin

GS : *Y a-t-il un sujet sur lequel tu aurais voulu t'exprimer et que j'ai oublié ? Ou as-tu un message à faire passer à nos lecteurs ? Bref, que veux-tu écrire comme fin à cet entretien ?*

MM : En relisant l'interview, la statistique m'apparaît aujourd'hui encore plus qu'hier comme un vecteur exceptionnel permettant d'interagir naturellement avec différents métiers, d'échanger avec des collègues d'autres disciplines et de dispenser des enseignements « connectés ». Je pense qu'il ne faut pas hésiter à pousser ces portes parfois vers l'inconnu mais qui peuvent également soulever de nouvelles questions de recherche.

Message à l'attention des futurs interviewés. J'ai pris beaucoup de plaisir à échanger sur ces sujets d'enseignement souvent peu abordés entre nous, mais qui pourtant, en tant qu'enseignants-chercheurs, nous occupent largement. Alors chers collègues, n'hésitez pas, lancez-vous, vous passerez un bon moment !

Gilles, merci pour ce temps d'échange et de réflexion.

GS : *J'espère que nos lecteurs seront aussi enthousiastes que toi vis-à-vis de cette nouvelle chronique. Pour sa deuxième édition, elle recueillera le témoignage d'un praticien : d'un statis-*

© Société Française de Statistique (SFdS), Décembre/December 2015

ticien travaillant dans l'industrie, devant recruter d'autres statisticiens issus de nos formations et, idéalement, donnant lui aussi des cours dans l'enseignement supérieur. Il sera intéressant d'avoir son point de vue critique sur nos enseignements !

Références

[1] Azencott, R., O. Catoni, et M. Mougeot (1994), Diagnostic neuronal multicapteurs : application au démarrage du moteur Vulcain, CNES, Rapport technique, Miriad Technologies.

[2] Bienenstock, E. L., L. N. Cooper, et P. W. Munro (1982), Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex, *The Journal of Neuroscience*, **2**(1), 32–48.

[3] Cadet, O., C. Harper, et M. Mougeot (2005), Monitoring energy performance of compressors with an innovative auto-adaptive approach, in *Proceedings of ISA Expo*, ISA – the Instrumentation, System, and Automation Society, Chicago.

[4] Carpenter, G. A. et S. Grossberg (1987), A massively parallel architecture for a self-organizing neural pattern recognition machine, *Computer Vision, Graphics, and Image Processing*, **37**(1), 54–115.

[5] LeCun, Y., B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, et L. D. Jackel (1989), Backpropagation applied to handwritten ZIP code recognition, *Neural Computation*, **1**(4), 541–551.

[6] Wasserman, L. (2004), *All of Statistics: A Concise Course in Statistical Inference*, Springer.

[7] Zalc, C. et C. Lemercier (2008), *Méthodes quantitatives pour l'historien*, La Découverte.

© Société Française de Statistique (SFdS), Décembre/December 2015