

INTRODUCTION

*Ainsi, sans que jamais notre amitié décide
qui de nous deux remplit le plus utile emploi,
je marcherai pour vous, vous y verrez pour moi*
J.P. Claris de Florian - L'aveugle et le paralytique

Il ne passe pas un jour sans que l'on entende les médias commenter les statistiques du commerce extérieur, l'évolution du CAC 40, la cote de popularité du Président de la République, la somme gagnée lors du dernier tirage du loto, l'influence de l'alcoolémie sur les accidents de la route... Face à de telles annonces, on voit les individus naviguer entre la foi en l'objectivité du chiffre (« *les chiffres montrent que le président a vu sa cote de popularité augmenter de 2% depuis le mois dernier* ») et le plus complet scepticisme (« *on fait dire aux chiffres ce qu'ils veulent* »). En fait – et hors de tout contexte scientifique – on conviendra aisément que notre société est marquée par le nombre et ses très nombreux usages et mésusages qui ouvrent la voie aux charlatans de tous poils.

Les exemples qui précèdent sont quelques illustrations du fait que les probabilités¹ et les statistiques baignent notre société et touchent tous les domaines de la connaissance, y sont compris ceux ne relevant pas des classiques « sciences de l'ingénieur », par exemple économie, finance et bourse, sciences humaines, marketing, qualité, sécurité, sciences de la vie...

Pourtant, l'enseignement secondaire français attribuait jusqu'à une période récente une place très marginale à l'analyse de ces problèmes, faisant du bachelier français une victime facile des faiseurs d'opinion et analystes autoproclamés. Un heureux renversement de tendance s'est enfin produit mais, pour ce qui concerne la formation des ingénieurs, le passage en classes préparatoires aux grandes écoles d'ingénieurs constitue une parenthèse malheureuse pendant laquelle aucun enseignement de probabilités et/ou statistiques n'est proposé. Il faut donc attendre l'intégration dans une école² pour voir apparaître un enseignement systématique sur le sujet à la fois construit théoriquement et appliqué à des cas concrets. Or, comme nous le verrons tout au long de ce cours, l'apprentissage de ces questions nécessite une remise en cause assez fondamentale de mécanismes de pensée courants ; ce qui signifie que cette étude arrive bien tardivement dans la formation des personnes. Par voie de conséquence, les efforts demandés aux étudiants sont importants, et l'appropriation en profondeur des concepts et techniques clefs difficile. A travers ce discours, nous ne cherchons pas ici à faire le procès du système éducatif français afin de nous décharger de notre responsabilité éducative. Nous pensons au contraire, évidemment, que le défi mérite d'être relevé ; c'est la raison d'être de ce livre.

Notons enfin que la tradition française veut que l'on commence par étudier le calcul des probabilités avant de lui trouver un riche domaine d'applications à travers l'activité

¹ Les probabilités interviennent dans les questions précédentes puisque, par exemple, la hausse de 2% a été observée sur un échantillon de la population, à l'aide d'un sondage, et que le résultat de celui-ci dépend des personnes qui ont été interrogées (**au hasard**).

² On pense ici aux écoles ressemblant à l'école des mines de Saint Etienne pour leur recrutement.

statistique. La tradition anglo-saxonne est en tous points opposée puisqu'elle consiste à ne voir le calcul des probabilités que comme le substrat théorique sur lequel est bâtie la statistique ! On retrouve là l'opposition théorie/pragmatisme qui caractérise ces deux systèmes de formation. Pour un enseignement dispensé dans une optique professionnalisante, nous sommes convaincus de la pertinence d'un mode d'apprentissage basé sur l'induction, ne serait-ce que parce qu'il est le seul à correspondre au métier d'ingénieur, quel qu'en soit sa forme. Cependant, nous choisirons ici une voie médiane qui consiste à étudier en parallèle les deux matières et à profiter de ce fait en plein de leurs interactions. Cela permettra aux étudiants, habitués depuis longtemps à la seule approche déductive, de passer en douceur d'une culture à l'autre.

Venons-en à une présentation succincte de ces deux activités.










De façon schématique, on peut présenter le calcul des probabilités comme une science de l'incertain et de la variabilité. Comme telle et sans vouloir cultiver le paradoxe, elle cherche à effectuer des prévisions dans des situations où tout n'est pas parfaitement prévisible. Bien sûr, il faut s'entendre alors sur le sens du mot prévoir. Par exemple, une personne qui joue au 421 prévoit qu'elle a une chance sur 36 d'obtenir 421 en lançant 3 dés ; ce qui ne signifie évidemment pas qu'elle prévoit le résultat avant le lancer des dés. Ce type de problématique peut se retrouver dans toutes les sciences de l'ingénieur à travers des domaines aussi variés que les incertitudes de mesures, le contrôle de qualité, la fiabilité des matériels ou des communications, et plus généralement tout ce qui concerne la prévision en contexte incertain : financière, économique, ou même humaine.

Le but principal de la partie probabiliste de ce livre est de fournir les concepts fondamentaux nécessaires à une bonne compréhension du calcul des probabilités afin d'éviter les pièges dans lesquelles les bétotiens, fussent-ils décideurs, tombent malheureusement souvent³. Le déroulement de l'ouvrage ne sera pas analogue à celui de celui que l'on peut trouver dans tout bon livre de mathématiques et nombre de détails seront omis. L'accent n'est donc pas mis sur la capacité de l'étudiant à reproduire des techniques probabilistes mettant en œuvre les outils théoriques (tels que tribus, extensions de probabilités ...) mais plutôt sur une compréhension plus globale des techniques introduites et de leur portée pratique. Il s'agit, pour le lecteur de ces lignes, d'adopter une attitude d'ingénieur vis à vis d'une connaissance inconnue : ne pas se fourvoyer dans des mois d'étude que leur employeur ne leur accordera jamais mais ne pas - a contrario - utiliser des techniques nouvelles sans les avoir assimilées, en utilisant la recette préconisée par une société de services ou un vendeur de logiciels ! L'expérience montre que cet abord relativement nouveau des sciences est souvent déroutant après des classes préparatoires très encadrées mais cette nouvelle marche est indispensable à franchir pour le futur ingénieur.

Mais le pire reste à venir avec les statistiques ! En effet, l'objectif poursuivi ici est de permettre à l'individu de déterminer sa position de la manière la plus raisonnée possible en fonction du contexte. Prenons un exemple qui donnera une idée de l'immensité des situations concernées. Supposons que l'on cherche à comparer les effets de deux médicaments hypotenseurs. Que peut-on conclure si sur 10 personnes traitées avec le médicament A, la tension moyenne est de 12,2 alors que pour le même nombre de personnes traitées avec le médicament B, on obtient une tension moyenne de 13,4 ? Un instant de réflexion montre que le seul fait que ces deux chiffres soient différents ne suffit pas à conclure car nous avons peut-

³ La consultation attentive de la presse écrite montre par exemple que la notion très élémentaire de probabilité conditionnelle est souvent source d'erreurs nombreuses.

être manqué de « chance » en observant la tension des 10 personnes traitées avec le médicament B. 10 autres personnes traitées avec ce même médicament auraient peut-être eu une tension moyenne inférieure à 12,2. De même, on sent bien intuitivement que la décision sera d'autant plus facile à prendre que l'écart entre ces deux moyennes est important. Tout cela n'est que du bon sens, mais comment rationaliser notre démarche et prendre une décision éclairée ? Pour répondre à ce type de questions, deux aspects doivent être abordés. Le premier a trait à la manière dont un chiffre doit être interprété. Un cours sur ce sujet est de notre point de vue impossible et nous nous contenterons de donner quelques exemples illustratifs. Le second point concerne le traitement des chiffres proprement dits et constitue l'essentiel de l'enseignement de statistiques. On verra qu'une approche scientifique existe, quoique son caractère inductif déconcerte quelque peu l'habitué des activités déductives, comme peuvent l'être les mathématiques. Nous pensons que cette approche constituera une bonne illustration de la citation en exergue, à savoir que l'aveugle calcul des probabilités et la paralytique statistique peuvent difficilement vivre séparées !

Chaque chapitre s'achève par une collection d'exercices. Ceux-ci sont d'ampleur et de difficulté très inégales. Plus la difficulté augmente, plus le temps se couvre, les symboles correspondants étant , , , . Les exercices de niveau  sont en général des exercices assez faciles d'application des notions vues dans le chapitre concerné. Ceux allant de  à  sont d'une plus grande ampleur et les exercices de niveau  demandent un investissement personnel conséquent. Ils sont alors souvent énoncés avec un vocabulaire se rapprochant du langage courant, le premier travail de résolution⁴ consiste alors à poser les questions avant de les résoudre ! Les exercices nécessitant l'utilisation d'un ordinateur sont indiqués par le symbole . Les corrections des exercices sont données en deux étapes, des indications de solution sont fournies à la suite des énoncés alors que des corrigés détaillés sont regroupés en annexe.

Enfin, pour ce qui concerne les activités statistiques, nous ne serions pas cohérents si nous ne proposons pas de traitements de jeux de données numériques. Pour ce faire, nous avons choisi d'utiliser deux logiciels, de nature très différente mais accessibles à tous.

Le premier est le logiciel R. Il s'agit d'un logiciel libre développé initialement par John Chambers aux laboratoires Bell, qui est à la fois un langage et un environnement pour faire du traitement de données et des statistiques. On peut trouver tous les détails sur le site www.r-project.org, qui indique également comment télécharger le logiciel (fonctionnement dans les environnements Windows, Mac OS et Linux). L'apprentissage de R n'est pas immédiat mais la présence de très nombreuses fonctionnalités préprogrammées et son orientation traitement de données le rendent précieux lorsque les besoins dépassent les techniques élémentaires.

Le second est beaucoup plus connu du grand public puisqu'il s'agit de Excel[®]. Cela surprendra peut-être le lecteur statisticien. Mais la raison d'un tel choix est en fait très simple : l'observation que nous avons pu faire à de nombreuses reprises montre que pour beaucoup d'ingénieurs, le seul outil de calcul disponible dans leur milieu professionnel est Excel[®], voire le seul outil accepté par les collègues, et que celui-ci permet de mener de nombreuses études statistiques élémentaires, quitte à mobiliser le langage de programmation associé Visual Basic[®]. Dans un contexte professionnel, nous recommandons l'usage des deux

⁴ Il ne s'agit d'ailleurs pas dans ce cas d'une résolution au sens étymologique du terme.

logiciels en même temps : R pour le traitement et l'analyse statistique de données, puis Excel pour la présentation des résultats.

En conclusion de cette... introduction, nous signalerons que ce cours est le résultat de plus de 10 ans d'essais variés, effectués au sein du département *méthodes et modèles mathématiques pour l'industrie* de l'école des mines de Saint-Etienne, dans une ambiance fort agréable au sein d'une équipe à la fois motivée et modeste, en remise en cause permanente. En conséquence, l'approche utilisée ci-après représente un équilibre précaire, atteint après beaucoup d'années de réflexion, entre formalisme et intuition, déduction et induction, questions précises et problèmes ouverts...