HAL open science

Poursuite d'étude après un IUT STID : l'exemple du Cursus de Master en Ingénierie Statistique et Informatique Décisionnelle de Toulouse

Cécile Chouquet, Aurélien Garivier

▶ **To cite this version:**

Cécile Chouquet, Aurélien Garivier. Poursuite d'étude après un IUT STID : l'exemple du Cursus de Master en Ingénierie Statistique et Informatique Décisionnelle de Toulouse. Journées de Statistiques de la SFDS, Jun 2014, Rennes, France. hal-00977409

HAL Id: hal-00977409

https://hal.science/hal-00977409

Submitted on 11 Apr 2014

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

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

Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Poursuite d'étude après un IUT STID : l'exemple du Cursus de Master en Ingénierie Statistique et Informatique Décisionnelle de Toulouse

Cécile Chouquet[1] & Aurélien Garivier[2]

Institut de Mathématiques de Toulouse, Université Paul Sabatier
[1] *cecile.chouquet@math.univ-toulouse.fr* [2] *aurelien.garivier@math.univ-toulouse.fr*

Résumé.
Après obtention de leur DUT, de nombreux étudiants des IUT STID souhaitent poursuivre leurs études pour obtenir un Master qui leur donne la possibilité de travailler comme ingénieur-expert ou chef de projet. Une spécificité forte de la science des données est son aspect bi-disciplinaire : elle nécessite la maîtrise de concepts et d'outils tant informatiques (bases de données et informatique décisionnelle) que mathématiques (statistique et modélisation stochastique). La formation SID de l'Université Paul Sabatier offre, depuis une douzaine d'années, une filière de poursuite d'étude cohérente qui respecte cette double compétence. Anciennement Institut Universitaire Professionnalisé, elle est devenue cette année un Cursus de Master en Ingénierie. Tout en conservant une forte composante industrielle, elle s'ouvre ainsi à la formation par la recherche et continue à préparer aux métiers d'ingénieurs dans toutes leurs dimensions, en s'appuyant fortement sur les activités de mise en situation.

Nous présentons ici brièvement l'articulation entre l'IUT et le master pour l'enseignement de la statistique, ainsi que son évolution passée et à venir. Nous revenons sur les atouts que leur passage en IUT donne aux étudiants, ainsi que les difficultés spécifiques qu'ils rencontrent.

Mots-clés. enseignement, statistique, IUT, STID, poursuite d'étude, master

Abstract. After obtaining their DUT, many students of IUT STID wish to continue their studies to obtain a Master, which gives them the opportunity to work as an expert engineer or as a project manager. A high specificity of data science is its joint aspect: it requires mastery of concepts and tools in computer science (databases, Business Intelligence) as well as in mathematics (statistics and stochastic modeling). The SID training at Paul Sabatier University offers, for a dozen years, a continued study program consistent with this dual skill. Previously professionalized university institute, it became this year a Master of Engineering. While maintaining a strong industrial component, it opens to the research training and continues to prepare for careers as engineers in all their dimensions, while relying heavily on real-life professional situations.

We present here briefly the coordination between the IUT and the master regarding statistics, as well as its past and future. We return on the assets offered by the transition from the IUT to SID, as well as the specific difficulties they encounter.

Keywords. education, statistics, IUT, STID, continued studies, master

1 La formation SID comme prolongement des IUT STID

1.1 Vocation historique de la formation SID

La formation SID permet depuis une douzaine d'années à des élèves des IUT STID de poursuivre leurs études au niveau Master. Ancien Institut Universitaire Professionnalisé (IUP) et aujourd'hui Cursus de Master en Ingénierie (CMI), elle est orientée vers le monde professionnel par la présence de nombreux stages et de modules de préparations aux métiers de l'ingénierie dans toutes leurs dimensions. Sont ainsi abordés non seulement les matières techniques, mais aussi le management, la gestion d'entreprise et de projet, la communication et plus généralement la connaissance de l'entreprise. Du fait de cette ambition, le volume horaire des enseignements est important et le suivi des élèves assez intense, avec beaucoup d'activités de mise en situation. En ce qui concerne les enseignements techniques et spécialisés, la formation respecte un strict équilibre entre compétences relevant des mathématiques et de la statistique d'une part, et les enseignements d'informatique et notamment de bases de données d'autre part (même volume horaire, même nombre d'ECTS). Les intervenants non universitaires y sont nombreux des deux côtés.

La formation SID s'inscrit donc clairement dans la continuité des IUT STID, même si (comme nous le détaillons plus loin) la formation au niveau Master implique quelques différences quant aux compétences et à leur acquisition. De fait, la majorité des étudiants y sont admis avec un DUT, avec des origines géographiques très variées : si Paris et Carcassonne fournissent les plus gros effectifs, les flux sont également réguliers depuis Lisieux, Vannes, Pau, etc... Les étudiants ne provenant pas d'IUT STID sont admis avec (par effectifs décroissants) une licence 2 de mathématiques, un DUT d'informatique ou une licence 2 d'informatique. Cette variété constitue une richesse stimulante pour tous.

1.2 Organisation du cursus

Le cursus SID est organisé autour de la double compétence en informatique et statistique que les étudiants doivent renforcer de la Licence 3 au Master 2. Pour rendre cela possible, le programme est élaboré selon un équilibre raisonné du nombre d'heures d'enseignement

dédiées à l'informatique et aux mathématiques[1].

L'année de Licence 3 a pour objectif l'intégration et l'uniformisation des bases requises dans ces deux disciplines. En statistique, les principales notions sont vues ou revues en statistique descriptive, statistique inférentielle et modèle linéaire, complétées par des enseignements d'algèbre et d'analyse afin de consolider les bases en mathématiques nécessaires aux méthodes statistiques plus avancées. En informatique, la Licence 3 est axée sur les concepts fondamentaux des bases de données, de la programmation et du génie logiciel.

Une fois ces bases acquises, les années de M1 et M2 permettent d'étudier des méthodes plus avancées. En statistique, les enseignements partent des techniques élémentaires inférentielles (modèle linéaire) et descriptives (ACP), pour aller vers des modèles de plus en plus complexes : les modèles linéaires généralisés, les modèles gaussiens (données répétées, modèle mixte, ...), la modélisation bayésienne ou les séries chronologiques. Les étudiants sont amenés à découvrir l'importance de l'optimisation numérique. Dans une perspective plus instrumentale, ils développent leurs connaissances du *machine learning*, avec une sensibilisation au traitement des données massives. Ces connaissances sont approfondies par des études de cas et, en Master 2, par des enseignements plus spécialisés (biostatistique, fiabilité des matériaux, etc.).

Parallèlement, en informatique, ils abordent progressivement la protection et la qualité des données, la recherche d'information, le data-mining, les bases de données multidimentionnelles, objets ou réparties à grande échelle. Tous ces enseignements font l'objet de nombreux projets mis en œuvre par les étudiants et pouvant réunir plusieurs thématiques et donc plusieurs enseignants.

Au cours des trois années, les enseignements académiques sont complétés par des stages en milieu professionnel : au minimum trois mois en Licence 3, quatre mois en Master 1 et cinq mois en Master 2. Si les stages de Licence 3 (voire Master 1) sont orientés vers des sujets en statistique ou en informatique, la plupart des stages de Master 2 portent sur des sujets pluridisciplinaires où la double compétence est indispensable, permettant ainsi de mettre en valeur les capacités des étudiants en statistique et en informatique.

1.3 Agiles avec les outils, moins à l'aise avec les concepts

Bien que le cursus SID soit la suite logique de l'IUT STID, les enseignements ne sont pas élaborés dans la même optique : les étudiants en Licence 3 doivent apprendre à appréhender les méthodes de façon plus critique et plus conceptuelle qu'à l'IUT. Professionnellement parlant, cela traduit la différence entre le travail attendu d'un technicien et celui d'un ingénieur, ou entre un membre et un chef de projet.

On peut illustrer cela en citant l'enseignement des bases de données : à l'IUT, il est principalement axé sur l'apprentissage des outils ACCESS et Visual Basic (ce qui

[1] De nombreuses précisions sur la formation sont disponibles en ligne sur son site web : `https://cmisid.univ-tlse3.fr/`

est très utile en pratique pour les étudiants). Au niveau Master, l'enseignement des bases de données avancées (multidimensionnelles, objet, réparties, voire noSQL) nécessite que les étudiants maîtrisent de façon approfondie les théories sous-jacentes aux bases de données : algèbre relationnelle, modèles de conception, etc. Parallèlement, en statistique, les étudiants arrivent d'IUT avec une très bonne maîtrise des logiciels statistiques (SAS et/ou R en particulier), mais ils ont souvent du mal à se détacher de l'outil au détriment de la compréhension des méthodes statistiques. Ce phénomène semble se renforcer au fil du temps. Par exemple, la statistique inférentielle ne faisait pas partie des premiers programmes du cursus SID car les étudiants diplômés STID maîtrisaient bien les concepts de tests, intervalles de confiance, etc... Mais il a été nécessaire, voici quelques années, de les réintroduire en Licence 3, ce qui a produit des effets dont les bénéfices ont été très rapidement visibles.

Par ailleurs, il est bien évident que la formation n'aborde pas, pendant le cursus, tous les logiciels potentiellement utiles : les étudiants sont parfois amenés à utiliser des programmes qu'ils ne maîtrisent pas ; cela les déstabilise souvent, en particulier au début de leurs stages. Toutefois, leurs expériences professionnelles leur montrent qu'avec leur niveau d'étude, ils doivent être capables de s'adapter très rapidement aux outils de l'entreprise, et plus généralement à leurs problématiques.

Enfin, on attend des étudiants futurs ingénieurs qu'ils sachent argumenter le choix d'une méthode, et qu'ils prennent du recul sur les méthodes qu'ils appliquent aussi bien en statistique qu'en informatique – ce qui nécessite pour eux un apprentissage.

2 Vers le Master : forces et faiblesses des étudiants issus d'IUT STID

Le département de mathématiques héberge, en plus du Master SID, un autre Master Professionnel d'Ingénierie Mathématique (IMAT) dont une des deux spécialisations est consacrée à la modélisation stochastique et statistique. Contrairement à ceux de SID, ses étudiants ont, pour une écrasante majorité, validé une licence après le baccalauréat et n'ont pas connu l'IUT. A l'issue de leur diplôme, ils sont parfois amenés à postuler sur des stages et des emplois semblables (par exemple comme chargé d'étude statistique, etc.). Il est donc intéressant de comparer ces deux populations étudiantes, afin de bien saisir la spécificité d'un parcours STID+SID.

2.1 Double compétence mathématique-informatique et professionalisation

La différence la plus flagrante, qui est très appréciée sur le marché du travail, est la réelle double compétence en statistique et en gestion informatique des données : il ne s'agit pas seulement d'une coloration informatique donnée à un parcours de mathématiques (ou

l'inverse), mais bien de connaissances équilibrées entre les deux. Cette double compétence apparaît de plus en plus comme indispensable dans ce que l'on nomme maintenant les *sciences des données*, comme en témoignent notamment les gros efforts pluridisciplinaires lancés autour des très en vogue *big data*. Pourtant, elle reste encore relativement rare, probablement du fait du découpage des enseignements universitaires en départements autonomes et peu habitués à collaborer de façon équilibrée. Même au sein des écoles d'ingénieurs, parfois un peu plus souples sur ce point, l'offre reste relativement segmentée.

Par ailleurs, les étudiants de SID ont au cours de leur cursus des enseignements de professionnalisation (techniques de communication, gestion, connaissance de l'entreprise, marketing, etc) plus nombreux, tout comme les projets et les stages plus longs et plus précoces : leur intégration en entreprise en est facilitée. Les contreparties sont des périodes d'enseignement plus ramassées dans l'année et des volumes horaires plus lourds, laissant moins de temps au travail et aux approfondissements personnels.

2.2 Les exigences du niveau Bac+5

Ce dernier point n'est peut-être pas sans conséquence sur une autre différence que l'on peut observer, et dont les causes sont probablement multiples. Les étudiants du cursus SID ont une maîtrise reconnue des outils de statistique ou d'informatique décisionnelle, qu'ils ont acquise grâce aux projets et aux stages. Toutefois, ils rencontrent plus de difficultés pour formaliser les problèmes, présenter des concepts ou une méthodologie, plutôt que des procédures logicielles pré-définies.

Ces étudiants ont une vision très procédurale et algorithmique du travail : leur façon d'appréhender un problème est souvent basée sur un catalogue des méthodes qu'ils connaissent en fonction des données du problème. Cela met en évidence le manque de recul ou de regard critique dont ils sont capables vis-à-vis de leur travail, entraînant également des difficultés sur les méthodologies plus innovantes. L'expérience montre que ces difficultés perdurent pendant tout le cursus, les étudiants restant attachés à la façon d'aborder les problèmes en IUT ; peut-être que ce cap pourrait être franchi plus facilement si les étudiants y étaient préparés très en amont.

3 L'évolution de la formation vers le Cursus de Master en Ingénierie

Depuis la rentrée 2013, la formation SID est devenue un Cursus de Master en Ingénierie (CMI). Ce nouveau format de formation est labellisé par le Réseau FiGuRe (Formation à l'Ingénierie par des Universités de Recherche), un groupement d'universités mettant en place des formations d'excellence aux métiers de l'ingénierie avec un cahier des charges contraignant qui les rapproche du modèle des *Master of Engineering* anglo-saxons :

- formation cohérente, progressive et exigeante de la sortie du baccalauréat jusqu'au master,
- une formation professionnalisante s'appuyant sur un grand centre de recherche (implication des laboratoires IMT et IRIT du LabEx CIMI),
- équilibre entre enseignements de spécialité, fondamentaux scientifiques, ouverture socio-économique et sciences de l'ingénieur,
- implication importante des professionnels dans la formation (enseignements, conférences, pilotage à travers le comité de perfectionnement),
- place centrale pour les activités de mise en situation et les stages.

Dans son dossier de labellisation, la formation a insisté pour montrer que les étudiants en provenance des IUT STID et continuant dans le CMI SID satisfont toutes les recommandations et tous les équilibres demandés pour les CMI. Il apparaît en effet que le cahier des charges des CMI est en grande cohérence avec celui des IUT, à un point près : l'importance de l'*initiation à la recherche* dès les premières années. Pour cela, un effort particulier est fait en L3 à destination des étudiants venant d'IUT pour rattraper leur retard (projet de recherche, invitations aux séminaires des laboratoires et aux événements mixtes comme la Journée Big Data du 8 octobre 2013[2].).

Cette mutation accompagne les évolutions pédagogiques des enseignements, en accélérant l'introduction de modules d'auto-apprentissage (un partenariat avec SAS permettra des expérimentations à la rentrée prochaine, des contacts avec Oracle sont également en cours), ou de classes inversées (pour l'apprentissage du JAVA par exemple). L'approche qui fait consensus dans l'équipe pédagogique consiste à recentrer les enseignements traditionnels (cours magistraux) sur les aspects les plus conceptuels tout en renforçant les capacités d'auto-formation aux outils et aux technologies au travers de modules-projets.

Enfin, l'importance socio-économique du *big data*, relayée fortement par le comité de perfectionnement, pousse à la mise en place d'un module dédié dès l'année prochaine en Master 2 (Nosql, RHaddop) ainsi qu'au renforcement, dès le Master 1, des enseignements de machine learning.

[2]Voir `http://perso.math.univ-toulouse.fr/bigdata/`