

Bulletin

Groupe d'études et de recherche
 en analyse de décisions

GERAD

La **statistique** de la décision au GERAD

Pour les bâtisseurs du GERAD, tous des experts en recherche opérationnelle, la statistique devait occuper une place importante dans un centre de recherche dédié à l'analyse des décisions. Cette idée de combiner les deux domaines était, pour eux, naturelle, à l'image d'ailleurs des départements de méthodes quantitatives en gestion d'où, en majorité, ils provenaient.

Cette conception n'a pu se concrétiser, en force, que récemment. L'arrivée de jeunes statisticiens et probabilistes prometteurs et d'autres établis, notamment à HEC Montréal, a permis enfin au GERAD de réaliser sa vision originale. L'atteinte d'une masse critique ne manquera pas d'induire une accélération et un approfondissement de la contribution du GERAD à ce domaine. En voici quelques exemples prometteurs. Le Centre organise conjointement avec le Centre de recherches mathématiques (CRM) et l'Institut des sciences mathématiques (ISM) une série de colloques réguliers. Aussi, un groupe formé de deux chercheurs en recherche opérationnelle et de deux statisticiens vient d'obtenir une subvention équipe du FQRNT. Et un atelier thématique sur l'exploitation de données (*data mining*) a connu un succès dépassant nos prévisions les plus optimistes.



Suite à la page 2...

bulletin@gerad.ca

SOMMAIRE

Nouvelles brèves	2
La vie et la recherche se rencontrent	3
La sécurité routière sans la statistique serait comme une auto sans phares	5
Derrière les décisions la probabilité veut qu'il y ait... de la statistique	7
Le statisticien : le médecin des données	9
Statisticiens et hockeyeurs ont plus en commun que les longues fiches de statistiques . . .	11

- **Jean-François Cordeau** a obtenu une Chaire de recherche du Canada en logistique et transport pour la période du 1^{er} avril 2006 au 31 mars 2011 pour 100 000 \$ par année.
- **Pierre Hansen, François Bellavance, Gilles Caporossi** et **Denis Larocque** ont obtenu une subvention FQRNT-Équipe de 45 000 \$ par année (2006-2009) et une subvention pour équipement de 15 930 \$ par année (2006-2007).
- **Pierre Hansen** a été élu membre de l'International Academy of Mathematical Chemistry en mars 2006.
- Le GERAD a quatre nouveaux membres depuis le 20 janvier 2006 : **Michel Perrier**, École Polytechnique de Montréal, **Frank Galiana**, Université McGill, **Shie Mannor**, Université McGill, et **David McDonald**, Université d'Ottawa (membre associé).
- **Shie Mannor** est titulaire d'une Chaire de recherche du Canada en apprentissage automatique / *machine learning* de 2005 à 2010, à raison de 100 000 \$ par année.
- Les journées d'optimisation 2007 auront lieu du 7 au 9 mai 2007 à HEC Montréal. Le comité organisateur est composé de **Olivier Bahn, Gilles Caporossi, Pierre Hansen** et **Sihem Taboubi**.
- Le GERAD voit sa subvention du FQRNT renouvelée pour 364 000 \$ par année.
- Le GERAD a été l'un des six centres appelés à se présenter à la journée des regroupements stratégiques FQRNT le 23 février 2006. Voir un résumé de la présentation à <http://www.fqrnt.gouv.qc.ca/nateq/Actualite/PWP/GERAD-RolandMalhame.pdf>
- Le professeur **Gilbert Laporte** a été élu Fellow de l'Institute for Operations Research and Management Sciences.
- La compagnie Ex Pretio, entreprise essaimée GERAD/CRT de **Gilles Savard** et collaborateurs, adhère à la IATA (Association du transport aérien international). L'entreprise se spécialise en tarification dynamique des billets dans le transport ferroviaire et aérien.
- Le GERAD a été responsable de l'organisation logistique de l'Atelier sur les centres d'appel qui a eu lieu le 11 mai 2006. Il y a contribué financièrement aussi.
- La 18^e mini-conférence EURO sur le *Variable Neighbourhood Search* a eu lieu à Tenerife, Espagne, du 23 au 25 novembre 2005. Le VNS est une méthode créée par les professeurs **Pierre Hansen** et **Nenad Mladenovic**.
- Un livre du professeur **El-Kébir Boukas** intitulé *Stochastic Switching Systems, Analysis and Design* a paru chez Birkhauser.
- En mai 2006, **François Bellavance** a été nommé directeur des programmes de M. Sc. et de doctorat à HEC.



La statistique... (suite).

Cette édition du Bulletin présente cinq statisticiens qui appliquent leur savoir à l'analyse de décision dans des domaines aussi variés que la médecine, la sécurité routière, l'ingénierie financière, le marketing, et j'en passe. Chacun a son territoire théorique et appliqué de prédilection, mais ils ont tous en commun cette grande passion pour la recherche qu'ils transmettent généreusement à leurs étudiants. Ça promet pour le GERAD...

Bonne lecture!
Georges Zaccour

Bulletin du GERAD

Édité 2 à 3 fois l'an par le GERAD.

Directeur
Georges Zaccour
georges.zaccour@gerad.ca

GERAD
HEC Montréal
3000, chemin de la Côte-Sainte-Catherine
Montréal (Québec) Canada H3T 2A7
Téléphone : (514) 340-6053

Site Internet
www.gerad.ca
bulletin@gerad.ca

Rédacteur en chef
Robin Philpot
rphilpot@sympatico.ca

Traduction
Robin Philpot

Conception graphique
HEC Montréal

Dépôt légal : 2^e trimestre 2006
Bibliothèque nationale du Québec

Reproduction autorisée
avec mention de la source

La **statistique** au GERAD

La **vie** et la **recherche** se rencontrent



Brenda
MacGibbon

À l'instar de James Watt, dont les observations d'une bouilloire ont abouti à la mise au point de la machine à vapeur et à l'unité de mesure Watt, nombreux sont les chercheurs qui s'inspirent du quotidien dans leurs travaux. **Brenda MacGibbon**, professeure associée au Département de mathématiques à l'UQAM et première statisticienne au GERAD, est un exemple remarquable de l'influence sur le travail des joies et des peines de la vie quotidienne. Son histoire personnelle démontre également le potentiel des méthodes quantitatives appliquées à la décision en matière de santé.



François
Bellavance

Après avoir obtenu son doctorat en probabilité et analyse de McGill, suivi d'études postdoctorales à Paris, Brenda est revenue à Montréal au début des années 1970 où elle a donné naissance à une fille et à des jumeaux en moins de 18 mois. Ne travaillant pas, si on ose le dire ainsi, et le « seul adulte que je voyais, autre que mon mari, était le pédiatre qui, heureuse coïncidence, connaissait le meilleur généticien au Québec, le docteur Charles Scriver. Il était attaché à l'Hôpital de Montréal pour enfants qui, à ce moment-là, étudiait intensément le dépistage génétique de maladies telles que la fibrose kystique. Ayant accouché de jumeaux, je m'intéressais particulièrement à la génétique et l'hôpital avait un poste à temps partiel dans le département de la génétique. C'est là que j'ai appris la statistique et l'informatique. Vous savez, à ce moment-là on pouvait obtenir un doctorat en mathématiques sans être informaticien. »



Bruno Rémillard

Un poste à temps partiel à l'Hôpital de Montréal pour enfants ne menait évidemment pas à la permanence au niveau universitaire. Ainsi, pendant une année sabbatique de son mari, Brenda a décroché un poste de chercheur associé invité en théorie statistique de la décision à l'université Cornell. Elle a commencé à publier ses travaux et à contribuer à la statistique mathématique. Avant de se joindre au Département de mathématiques en 1988, toutefois, Brenda MacGibbon a enseigné à l'Université Carleton et la l'Université de Sherbrooke dans les départements de mathématiques et statistique ainsi qu'à Concordia dans le Département des sciences de la décision et des systèmes intégrés de gestion.



Denis Larocque

Ses travaux de recherche à l'UQAM l'ont amenée au GERAD avant même que l'UQAM n'y adhère officiellement. « J'essayais de résoudre un problème du minimax – minimiser le risque maximal – qui était plus difficile que le maximum ou le minimum tout court. Comme je n'y arrivais pas avec des méthodes d'optimisation locale, j'ai traversé le Mont Royal et, avec l'aide de Brigitte Jaumard, j'ai réussi à le résoudre en utilisant des méthodes d'optimisation globale. C'était le début d'une collaboration de plusieurs années. Le GERAD m'a acceptée comme membre associé jusqu'à ce que l'UQAM adhère formellement. »



Pierre Duchesne

Hormis le travail en génétique à l'hôpital, Brenda ne s'était pas encore penchée sur la statistique en santé. « Le sens de ma recherche a changé quand j'ai eu le cancer du sein en 1997, se rappelle-t-elle. Lorsqu'on est atteint d'une telle maladie, on est porté à penser à sa vie et à comment la changer. Dans mon cas, j'ai décidé de me concentrer davantage sur la biostatistique et moins sur la statistique mathématique. Une autre motivation : ma meilleure amie de Cornell était biostatisticienne à l'université de Southern California et elle m'avait tenue au courant des avancées dans le domaine. » Une concentration en biostatistique représentait une bonne façon de diriger ses étudiants, princi-

... **Brenda MacGibbon** (suite).

Le défi consistera à garder la créativité en mettant à contribution la puissance de l'ordinateur.

pablement au deuxième cycle, qui, jusque là, avaient de la difficulté avec la statistique hautement théorique. « J'ai choisi la biostatistique et l'analyse de survie pour donner une direction à mes étudiants. »

Dans le Bulletin du GERAD de novembre 2005, le directeur, Roland Malhamé, a exprimé le désir de voir davantage de la recherche en biotechnologie et en biologie au GERAD. Rien ne pouvait plaire autant à Brenda MacGibbon car il s'agit d'un champ d'application vaste où les méthodes quantitatives gagnent en importance et en résultats. « En théorie de la décision, il y a toute une théorie sur l'efficacité des divers traitements disponibles. Sans ces nouvelles méthodes, les médecins doivent décider en se fiant sur leur seule intuition, bien que celle-ci soit informée. La théorie de la statistique de la décision peut compléter l'intuition. »

Il est facile de concevoir le nombre et l'ampleur des problèmes statistiques et mathématiques soulevés par l'analyse de survie. Parmi ceux-ci, les problèmes de données manquantes – le temps de survie étant connu pour certains, inconnu pour d'autres – de même que des problèmes de données censurées et tronquées découlant, par exemple, des délais des essais. Les travaux de Brenda portent principalement sur les différents types de cancer (du sein et de la peau) et sur l'insuffisance rénale terminale, qui touche à l'analyse de risque multidimensionnelle.

Après son diagnostic du cancer du sein, Brenda s'est jointe, à titre de défenseur des patients et de statisticienne, à l'agence subventionnaire du programme de recherche sur le cancer du sein de la Californie. Tous les membres chargés de l'examen par les pairs doivent venir de l'extérieur de la Californie. Ce travail a abouti à une collaboration percutante avec un chirurgien qui essayait de mettre au point des méthodes diagnostiques



moins envahissantes que la biopsie. « La biopsie, selon certains, aggrave le cancer, note Brenda. Nous avons voulu perfectionner l'outil diagnostique et en évaluer la fiabilité. »

Une autre expérience qui se prêtait bien à la biostatistique appliquée à l'UQAM se déroulait au CINBIOSE (Centre de recherche interdisciplinaire sur la biologie, la santé, la société et l'environnement) : la détection d'effets neurologiques précoces sur les Cris du Nord du Québec de niveaux de mercure élevés dans les poissons consommés ou chez les travailleurs exposés au manganèse. Anne Beuter et son équipe ont développé des machines ultra-sensibles pour mesurer des tremblements. À l'aide de méthodes statistiques, on pouvait déterminer la capacité de la machine de distinguer avec précision les tremblements normaux de tremblements de type anormal. Ainsi, le fil conducteur des travaux de recherche de Brenda MacGibbon est l'utilisation de méthodes quantitatives pour évaluer et perfectionner des outils de diagnostic.

Qui travaille en biostatistique travaille aussi, tôt ou tard, en génétique. « Oui, ajoute Brenda. Je suis retournée à mes premières amours, la génétique. La leucémie infantile soulève, par exemple, un grand nombre de problèmes mathématiques parce qu'il y a très peu de patients alors qu'il y a énormément de gènes. » En revenant à la génétique, Brenda s'est étonnée de l'impact des puissants ordinateurs et n'hésite pas à nous mettre en

garde contre l'éventuel étouffement de la pensée créatrice qui a marqué le travail en génétique dans la première moitié du 20^e siècle. Le défi consistera à garder la créativité en mettant à contribution la puissance de l'ordinateur.

À l'avenir, la statistique servira également à la mise au point de politiques en matière de médicaments et à la surveillance. Brenda MacGibbon est membre de la coalition canadienne de développement de la politique en matière de médicaments, réunissant des chercheurs canadiens souhaitant influencer le gouvernement fédéral dans la surveillance de médicaments lorsque ceux-ci sont approuvés et disponibles sur le marché. Il s'agit d'un aspect longtemps négligé. L'objectif sera, entre autres, de voir comment de grandes bases de données administratives peuvent servir à la surveillance de médicaments. Aurait-il été possible, par exemple, d'obtenir plus d'information sur le VIOXX et ses effets secondaires à l'aide de ces bases de données, comme celle de la RAMQ? Qu'en est-il des effets secondaires d'autres médicaments? Le cas échéant, les décisions sur l'utilisation de divers médicaments seront plus éclairées, plus précises, et le consommateur aura accès aux informations bien plus vite.

Dans chaque domaine de biostatistique appliquée, Brenda vise à développer autant de nouvelles méthodes que la théorie permettant de résoudre les problèmes. **G**

La sécurité routière sans la statistique serait comme une **auto sans phares**

Qui, en lisant des statistiques en manchette prévoyant un danger imminent, n'a jamais réagi comme le faisait le sceptique Mark Twain pour qui il y avait trois types de mensonges : les mensonges, les sacrés mensonges et les statistiques? Pourtant c'est à partir de statistiques que les gouvernements et les entreprises prennent des décisions importantes nous concernant. D'où l'exigence de rigueur, de discernement et de précision que la population impose aux statisticiens. **François Bellavance**, membre du GERAD et du CRT et professeur au Service de l'enseignement des méthodes quantitatives de gestion à HEC, qui se spécialise en statistique appliquée à la sécurité routière, connaît bien le rôle important des statistiques dans les décisions gouvernementales en cette matière.

Voici quelques statistiques qui pourraient défrayer nos manchettes et faire bouger nos décideurs. Des études montrent que vous avez cinq fois plus de chances de mourir si vous ne portez pas votre ceinture de sécurité lors d'un accident grave; une augmentation de 1 % du port de ceinture au Québec sauverait 25 vies par année; dans le monde entier, trois milles personnes meurent par jour à cause de collisions routières; il y a deux fois plus de risque de décès si un automobiliste est heurté par un véhicule utilitaire sport (SUV) que s'il est frappé par une autre automobile; l'utilisation d'un cellulaire au volant augmente de 400 % le risque de collision.

Selon François Bellavance, à l'exception de ce dernier énoncé, même s'il vient d'une étude réalisée à Toronto et publiée dans une revue médicale prestigieuse et qui été corroboré par une étude australienne, tous ces énoncés statistiques sont fiables. Le problème de l'énoncé sur l'utilisation du téléphone cellulaire, et l'erreur, se situe au niveau de la méthodologie. Or, si tous les gouvernements devaient se fier aux premières études – quatre fois plus de risque d'accident pour ceux qui conduisent en parlant au téléphone

– on aurait déjà partout des lois interdisant le téléphone cellulaire au volant.

En effet, en 2003, François Bellavance et ses collègues ont publié une étude sur l'utilisation de la téléphonie mobile et le

Le problème de l'énoncé sur l'utilisation du téléphone cellulaire, et l'erreur, se situe au niveau de la méthodologie.

risque d'accidents de la route démontrant que le risque d'accident n'était pas quatre fois plus élevé, mais seulement 1,38 fois plus élevé. L'étude a été réalisée avec la collaboration de la SAAQ (Société de l'assurance automobile du Québec) pour l'envoi de questionnaires, les dossiers de conduite et les rapports de police et avec des entreprises de téléphonie pour les dossiers d'appels téléphoniques. Elle trouve sa force dans l'importance de la cohorte étudiée et l'ampleur des données utilisées. Pour fins de comparaison, ils ont fait une étude parallèle avec les mêmes données mais en empruntant la méthodologie de l'étude de Toronto. Constat : quatre fois plus de risque avec le téléphone au volant.

François Bellavance tire quelques conclusions de cette expérience. D'abord,



... François Bellavance (suite).

il est très difficile de déterminer si un conducteur utilisait un téléphone lors d'un accident, en raison, entre autres, de l'imprécision de l'heure de la collision dans les rapports de police. Aussi, il est très difficile d'établir un lien causal entre l'utilisation du téléphone et les accidents d'automobile. On peut dire, toutefois, qu'il y a une augmentation réelle du risque d'accident, mais certainement pas de l'ordre de quatre fois, et que le sous-ensemble des gros utilisateurs de cellulaire en auto courent un risque d'accident encore plus élevé. Mais si le législateur adopte une loi interdisant le cellulaire au volant sur la foi du risque perçu quatre fois plus élevé, il risque de provoquer toutes sortes d'autres problèmes d'application. Malgré ce risque nettement plus faible, notons que François Bellavance ne touche jamais à un cellulaire dans l'auto! Ses études l'ont convaincu aussi de ne jamais rouler en auto sans que tout le monde soit attaché et de respecter rigoureusement la vitesse réglementaire. Ces statistiques ne mentent pas! (François n'a jamais eu d'accident de la route.)

La sécurité routière soulève un grand nombre de questions qui relèvent de la santé publique, dont des études statistiques peuvent fournir des réponses : l'impact sur le nombre et la gravité des accidents liés à la présence des panneaux publicitaires sur les autoroutes, l'impact d'un resserrement du processus d'obtention de permis pour les camionneurs – une étude en ce sens réalisée pour la SAAQ a confirmé l'effet positif de cette mesure –, l'impact de l'incompatibilité des véhicules utilitaires sport et des autos sur la gravité des accidents, etc. En outre, note François Bellavance, dès qu'on traite de la gravité des accidents avec blessés, l'étude devient plus difficile sur le plan méthodologique. « Les accidents avec blessés graves sont proportionnellement peu nombreux comparés à l'ensemble des accidents routiers. Il peut y avoir 700 décès alors qu'il y a des centaines de milliers d'accidents. C'est comme dans d'autres domaines pour les événements rares, les maladies rares. Le développe-

ment de méthodes statistiques appropriées pour l'analyse de ces données, et en particulier de méthodes non paramétriques, s'imposent. »

Avant d'arriver à HEC et de devenir le deuxième statisticien après Brenda MacGibbon à adhérer au GERAD, François avait dirigé le service de consultation statistique de l'Université Simon Fraser, à Vancouver, en Colombie-Britannique, en même temps qu'il terminait ses études de doctorat à l'Université de Montréal. Aussi, à titre de statisticien consultant à l'Institut de recherches cliniques de Montréal et plus tard comme biostatisticien à l'Université McGill et au Centre hospitalier St-Mary's, François a participé à de nombreuses recherches dans le domaine de la santé. Parmi celles-ci, signalons une étude sur des personnes âgées en perte d'autonomie. En effet, des personnes âgées aux prises avec un problème aigu tendent à se rendre à l'urgence où ils finissent par attendre longtemps,

ce qui leur occasionne du désagrément tout en surchargeant le réseau de santé. L'objectif de l'étude, qui portait sur des patients de quatre hôpitaux à Montréal, consistait à développer un outil pouvant identifier les personnes les plus à risque pour des rechutes et ainsi trouver des solutions autres que le retour à l'urgence, telles que l'envoi d'un travailleur social, l'intervention d'un CLSC ou le suivi plus serré par la clinique externe de l'hôpital. L'outil visé était un questionnaire simple que les patients à l'urgence pouvaient remplir sans trop de difficulté. Mais pour l'établir il a fallu notamment puiser dans les banques de données d'hospitalisation et de retours à l'urgence, faire des appels aux patients à différents intervalles. Donc, derrière le questionnaire simple, il y a des études qui ne sont pas simples du tout. Le résultat, selon François Bellavance, était un outil qui donnait un taux de prédiction relativement bon. **G**

François Bellavance ne touche jamais à un cellulaire dans l'auto! Ses études l'ont convaincu aussi de ne jamais rouler en auto sans que tout le monde soit attaché et de respecter rigoureusement la vitesse réglementaire. Ces statistiques ne mentent pas! Il n'a jamais eu d'accident.



Derrière les **décisions**, la probabilité veut qu'il y ait... de la **statistique**

« Les statistiques sont utilisées partout où on prend des décisions! » note tout simplement **Bruno Rémillard**, membre du GERAD et professeur titulaire en Méthodes quantitatives de gestion à HEC. Bruno Rémillard se concentre tout particulièrement sur deux domaines qui nous concernent tous, les probabilités et la statistique.

On y apprend, entre autres, qu'en décidant si on devait acheter un billet pour le 6/49, la possibilité de gagner le gros lot ne devrait même pas être considérée comme un facteur de décision!

tiquer les données et de vérifier des modèles. » À titre d'exemple, travaillant sur un sujet populaire s'il en est, les fluctuations des prix de l'essence à la pompe, il a démontré que, même si au cours des dernières années il y a eu des hausses et des baisses, les hausses ont été plus élevées que les baisses. (Ah, on le savait!). Pressé de nous expliquer pourquoi et à qui cela a profité, Bruno Rémillard répond, sourire en coin, que tel n'était pas l'objectif de l'étude.

L'article qu'il a tiré de cette recherche, publié dans la Revue canadienne de la statistique, a mérité à Bruno Rémillard et à ses collègues B. Abdous et K. Ghoudi, le prix du meilleur article de l'année en 2003 décerné par la Revue, qui a souligné la qualité exceptionnelle de sa présentation et de sa contribution méthodologique. (*Nonparametric weighted-symmetry tests*, Revue canadienne de la statistique, 2003). Les modèles ainsi mis au point servent aussi à constater et à évaluer d'autres phénomènes similaires. À titre d'exemple, Bruno Rémillard, qui s'intéresse à l'ingénierie financière, a relevé une asymétrie semblable dans la réaction de titres à la Bourse de Toronto et à celle de New York en cas de bonnes et de mauvaises nouvelles. Une bonne nouvelle, telle que l'amélioration de la cote de crédit d'une

entreprise, va engendrer une petite hausse de valeur du titre, alors qu'une mauvaise nouvelle, telle une décote, entraînera une baisse beaucoup plus importante. Il a démontré aussi que cette asymétrie dans l'impact des nouvelles est plus prononcée à la Bourse de New York qu'à Toronto. Encore ici, Bruno Rémillard précise que l'objectif n'est pas de dire pourquoi mais plutôt d'en mesurer l'impact.

Un autre domaine d'application en finance et en actuariat ainsi qu'en biostatistique est celui de la modélisation de la dépendance. Il peut s'agir de la dépendance d'un investissement à certains événements, de la relation de dépendance entre un prix à la bourse de Toronto et un prix à celle de New York, la relation entre deux ou plusieurs titres et ainsi de suite. « On regarde les deux pour déterminer s'il y a une dépendance et, le cas échéant, on essaie de la mesurer et de la modéliser. » Les banques et d'autres grandes entreprises sont évidemment très intéressées par ce genre de recherche.

Qui, dans sa vie courante, ne s'est pas posé la question sur la probabilité qu'un tel ou un tel événement survienne : qu'un avion s'écrase, qu'on attrape la grippe aviaire, qu'on gagne le gros lot et ainsi de suite. Le sujet est tellement populaire qu'un collègue probabiliste de Bruno Rémillard a publié, en 2005, un livre, devenu un best-seller, qui permet aux profanes de saisir les concepts de base de la probabilité et de se prémunir contre l'irrationalité dans la prise de décisions (*Struck by Lightning: The Curious World of Probabilities*, Jeffrey S. Rosenthal). On y apprend, entre autres, qu'en décidant si on devait acheter un billet pour le 6/49, la possibilité de gagner le gros lot ne devrait même pas être considérée comme un facteur de décision!

Tout en ayant d'innombrables applications, le travail de recherche de Bruno Rémillard se situe à un tout autre niveau. « Contrairement à plusieurs collègues statisticiens, je ne me préoccupe pas des méthodes de collecte de données. Je présume que celles dont je dispose sont fiables et j'essaie de construire des outils pour vérifier certaines hypothèses. Aussi, je construis de nouvelles façons de décor-

... Bruno Rémillard (suite).

Avant de rejoindre HEC, Bruno Rémillard a été professeur de mathématiques à l'Université du Québec à Trois-Rivières de 1988 à 2001, après des études postdoctorales à l'université Cornell. Il a fait ses études de premier et de deuxième cycles à l'Université Laval, à Québec, et son doctorat de l'Université Carleton, à Ottawa. Son arrivée à HEC coïncide avec un grand effort de la part de HEC de se doter d'une bonne équipe spécialisée en méthodes quantitatives. Aussitôt arrivé, Bruno adhère au GERAD, entre autres, parce que ses collègues statisticiens de HEC venaient, eux aussi, d'y adhérer, la statistique étant tout naturellement reliée à l'analyse de la décision. Il signale notamment « l'excellente dynamique des colloques du GERAD sur la statistique. » Notons que Georges Zaccour attribue l'essor de la statistique au GERAD justement à la présence d'une masse critique de statisticiens.

La biostatistique, même si elle ne fait pas partie de ses priorités de recherche, revêt un intérêt particulier à cause des problèmes captivants qu'elle soulève sur le plan statistique, telles les données tronquées, censurées et manquantes. « Dans la modélisation de la dépendance entre deux variables, une étape naturelle consiste à étudier les données censurées ou tronquées. Or, les données censurées proviennent principalement de la biostatistique. »

L'avenir de la statistique de la décision est très prometteur, selon Bruno Rémillard. « Les statistiques sont de plus en plus présentes dans les entreprises et dans tous les aspects de la vie courante. Les entreprises cherchent des solutions automatisées et rapides. Mais derrière ces solutions, il y énormément de recherches statistiques. Donc, plus on demandera la capacité d'automatiser les décisions, plus la statistique sera en demande. »

À titre d'exemple, lorsqu'une banque reçoit une demande de prêt, elle fait remplir un questionnaire simple mais avec beaucoup de questions. À partir de ce questionnaire, la banque va prendre une décision éclairée si oui ou non elle



L'objectif est de détecter la cible et de déterminer avec exactitude l'identité d'un autre avion : la solution, des algorithmes d'identification fondés sur la statistique.

accordera le prêt. La décision est éclairée par une grande quantité de données statistiques, analysées selon diverses méthodes, qui permettent d'établir avec précision la probabilité d'un emprunteur de faire forfait sur le prêt. Autre exemple, la reconnaissance de cibles aériennes non

coopérative ou reconnaissance d'avions amis et ennemis. L'objectif est de détecter la cible et de déterminer avec exactitude l'identité d'un autre avion : la solution, des algorithmes d'identification fondés sur la statistique. **G**

Le statisticien : le **médecin** des données

Denis Larocque, qui est membre du GERAD depuis qu'il s'est joint au Service de l'enseignement des Méthodes quantitatives de gestion à HEC en 2000, compare parfois son rôle de statisticien à celui d'un médecin : « quelqu'un vient me voir avec un problème, je l'écoute, lui pose des questions sur la nature des données et la façon dont elles ont été recueillies ainsi que sur les hypothèses et questions de recherches et, par la suite, je pose un diagnostic qui consiste à proposer les analyses appropriées ». La validité des conclusions dépend en grande partie de la qualité des données et donc notamment du plan d'échantillonnage initial. Ainsi, pour Denis Larocque, il est important que les statisticiens apportent leur expertise à toutes les étapes d'un projet de recherche et dans tout processus décisionnel faisant appel à l'analyse statistique.



La validité des conclusions dépend en grande partie de la qualité des données et donc notamment du plan d'échantillonnage initial.

La collaboration entre chercheurs opérationnels et statisticiens se construit quand on sait distinguer les deux domaines de recherche. Denis Larocque l'explique de façon simple. « En recherche opérationnelle, on débute souvent avec un problème précis qui comprend certaines hypothèses et contraintes et on tente de trouver une solution optimale, par exemple, en termes de coûts. En statistique, on travaille à partir d'un échantillon prélevé dans une population, on applique des méthodes statistiques pour tirer des conclusions qui seront généralisables, avec une certaine marge d'erreur, à l'ensemble de la population. »

Le potentiel de collaboration au sein du GERAD est important tant au niveau méthodologique qu'au niveau des appli-

cations, ajoute-t-il. Il cite en exemple l'exploitation de données (*data mining*), qui ressemble à l'orpaillage dans du sable mouvant, vu les quantités énormes de données disparates disponibles. En exploitation de données, les chercheurs opérationnels et statisticiens se complètent : la combinaison des connaissances en optimisation et méthodes de classification des uns et des connaissances en échantillonnage et méthodes statistiques des autres permet de produire des informations pertinentes et utilisables pour les décideurs. Au point de vue méthodologique, Denis Larocque signale un projet

amorcé avec Pierre Hansen, François Bellavance et Gilles Caporossi du GERAD sur le développement de méthodes d'arbres de classification pour des données longitudinales ou multivariées ainsi que la publication par le GERAD du livre intitulé *Statistical Modeling and Analysis for Complex Data Problems*, qui figure parmi les dix volumes produits dans le cadre du 25^e anniversaire du centre de recherche.

Ayant eu la « piqûre » pour la probabilité et la statistique dès son premier cours en la matière au CÉGEP et y voyant un potentiel énorme, il a gardé le cap depuis,

... Denis Larocque (suite).

faisant ses études en mathématiques et en statistique jusqu'au doctorat, en 1997, à l'Université de Montréal, des recherches postdoctorales avec le Groupe de recherche sur l'inadaptation psychosociale chez l'enfant, également à l'Université de Montréal, et des études postdoctorales à l'université de la Floride. Avant de se joindre à HEC Montréal où il est maintenant professeur agrégé, il a enseigné la statistique pendant un an à l'Université du Québec à Trois-Rivières.

Denis Larocque s'intéresse tout particulièrement aux méthodes statistiques non paramétriques multivariées. En termes simples, la validité des méthodes non paramétriques requiert moins de présupposés sur la nature des données, comme par exemple la normalité, que les méthodes paramétriques. Ces méthodes ont vu le jour dans les années 1940 et 1950 (Wolfowitz a utilisé pour la première fois le terme non paramétrique en 1942). « De nos jours, le terme non paramétrique englobe plusieurs familles de méthodes, note-t-il. L'une d'elles, l'inférence basée sur les rangs, possède l'avantage d'être valide même pour de petits échantillons et hérite naturellement d'une robustesse face aux valeurs aberrantes en ce sens que les résultats ne seront pas affectés par ces dernières. Une autre famille, les techniques de lissage, laisse les données parler plutôt que de les enfermer dans le carcan d'un modèle. »

À l'instar des chercheurs en optimisation combinatoire, Denis Larocque vante les percées réalisées grâce à la puissance de calcul des ordinateurs modernes qui ont révolutionné la recherche en statistique. « La puissance des ordinateurs nous permet d'utiliser des méthodes et d'ajuster des modèles qu'on ne pouvait songer à utiliser il y a quelques années. » Il cite en exemple le *bootstrap*, la métaphore étant inspirée de la légende selon laquelle le baron Münchhausen pouvait s'extraire d'un marécage ou de la mer en tirant sur les lacets de ses propres bottines. En rééchantillonnant à répétition à partir d'un échantillon original, il est possible d'esti-



La méthode *bootstrap* : comme le baron Münchhausen qui se serait tiré d'un marécage en tirant sur les lacets de ses bottines.

mer des caractéristiques d'intérêt (biais ou écart-type d'un estimateur) ou d'améliorer les performances prévisionnelles d'un classificateur (*bagging*). Il s'agit d'une méthode passe-partout qui permet, entre autres, d'obtenir des résultats qu'on ne peut obtenir de façon analytique. Mais sans l'ordinateur, ce serait impossible. La puissance de calcul devient donc un grand atout autant sur le plan théorique que pour les applications.

En plus de la recherche fondamentale en statistique, Denis Larocque collabore avec des chercheurs d'autres disciplines, notamment en recherche opérationnelle, en technologies de l'information, en marketing, en sécurité routière et en psychologie. L'intérêt pour la statistique dans le monde des affaires, comme dans d'autres secteurs, va grandissant, selon lui, qui

prend à témoin ses étudiants de maîtrise en Intelligence d'affaires (*Business Intelligence*). « Nos étudiants décrochent un poste avant même d'avoir terminé leurs études. La statistique n'a jamais été aussi vivante qu'aujourd'hui. Jamais il n'y a eu autant de journaux scientifiques et de colloques, le nombre de publications a explosé. »

Et le GERAD est au diapason! Denis Larocque se rappelle que c'est Georges Zaccour qui, en 2000, a proposé de créer, au GERAD, un noyau de statisticiens au même moment où HEC commençait à embaucher plusieurs. L'immense quantité et la complexité croissante des données à analyser perpétueront le besoin en médecins des données compétents pour les années à venir, conclut-il. **G**

Statisticiens et hockeyeurs ont plus en commun que les longues fiches de statistiques



Sachant que la première question posée à un chercheur porte souvent sur le caractère « théorique » ou « appliqué » de ses travaux, **Pierre Duchesne**, professeur agrégé au Département de mathématiques et statistique, à l'Université de Montréal, et membre du GERAD, a trouvé que la meilleure analogie vient du hockey : il serait donc un attaquant défensif (comme Guy Carbonneau), ou encore un défenseur offensif (comme Bobby Orr). Ainsi, Pierre Duchesne se décrit comme un statisticien appliqué qui s'aventure régulièrement dans la zone théorique ou encore un statisticien théoricien qui revient souvent dans la zone des applications.

Pierre Duchesne est un statisticien qui s'intéresse particulièrement aux séries chronologiques à la fois sous l'angle théorique et dans ses applications notamment en économétrie et en économétrie financière. L'économétrie financière est cette discipline qui a émergé pleinement ces dix dernières années où l'on applique et développe des méthodes statistiques et économétriques aux données issues de la finance. Les séries chronologiques comportent intrinsèquement une composante temporelle faisant en sorte que, contrairement aux techniques statistiques plus classiques, on ne peut présumer que les données recueillies forment un échantillon aléatoire, c'est-à-dire qu'elles sont prélevées indépendamment les unes des autres.

« Pour le genre de données sur lesquelles je travaille, il y a toujours cette notion de dépendance dans le temps, explique Pierre Duchesne. Les statisticiens et économètres se sont rendu compte de la nécessité de modèles visant à expliquer

ce type de données. De nombreuses séries économiques peuvent être modélisées adéquatement par des modèles de séries chronologiques. D'autres champs d'études, notamment l'hydrologie, ont utilisé avec grand succès ces modèles afin de prévoir par exemple la crue des eaux. Autre exemple, on pourrait étudier les séries chronologiques portant sur les prix de l'essence, sujet d'actualité s'il en est. Si je dispose de l'historique des prix de l'essence à Montréal, et si je dispose de d'autres variables temporelles que l'on pense reliées à ces données, je pourrais acquérir une meilleure compréhension des phénomènes en cause expliquant le prix. Autre exemple, je co-dirige une étudiante finissante au doctorat dans le programme de finance de HEC Montréal (maintenant professeure d'université) qui modélise des données financières à hautes fréquences, soit les transactions boursières. Des modèles de séries chronologiques permettent de les modéliser en tentant d'expliquer le délai temporel survenant entre ces transactions. »

Pierre Duchesne ajoute que les objectifs de ce genre d'études sont notamment la compréhension des fluctuations du marché et de la dynamique des événements, que ce soit une nouvelle transaction ou un certain changement dans le prix. « On essaie de cerner ce qui se passe au niveau de la microstructure du marché pour ensuite expliquer comment l'information est incorporée dans le prix, comment l'activité boursière varie au cours d'une journée, qu'est-ce qui détermine la volatilité et la liquidité du marché, etc. »

Ce genre d'analyse ne donne pas au statisticien ou à l'économètre une boule de cristal, mais il demeure que ces études

permettent, entre autres, une compréhension accrue des différentes structures d'échange, des stratégies de soumission des ordres par les investisseurs, etc. De plus, de nouvelles mesures quantitatives peuvent émerger, comme la construction de mesures de volatilité intra-journalières et des mesures de risque intra-journalières. Ainsi, le potentiel est fort important pour les régulateurs de marché afin de définir des politiques optimales. Pierre Duchesne note que, vu la disponibilité aujourd'hui de données boursières chronologiques précises – auparavant on devait se satisfaire de relevés aux 15 minutes ou plus – et grâce à la puissance des ordinateurs, on peut maintenant espérer expliquer des comportements boursiers sur différents horizons intra-journaliers.

Les modèles de séries chronologiques permettent d'établir des prévisions et de mesurer la qualité de ces dernières. Cette préoccupation du statisticien au GERAD, soit d'obtenir des estimations, des prévisions, ainsi que des mesures de la qualité de ces quantités statistiques, semble rejoindre celle des optimisateurs au GERAD, qui eux cherchent non seulement la solution optimale à un problème mais aussi la preuve qu'elle est optimale ou une mesure de l'écart de la solution optimale. Pierre Duchesne ne recule pas devant la comparaison. Pour lui, le mariage de la statistique et de l'optimisation des chercheurs en recherche opérationnelle est

beaucoup plus qu'un mariage de raison, la somme des deux étant plus grande que les deux parties. « Les statisticiens cherchent de bons estimateurs dans diverses situations, ajoute-t-il. Nous sommes amenés inévitablement à des problèmes statistiques d'estimation. Or, les estimateurs sont souvent obtenus selon des critères plus ou moins complexes, menant à une procédure d'optimisation, ce qui est au cœur de la recherche opérationnelle et de la théorie de l'optimisation. De plus, les techniques statistiques font souvent face à des problèmes numériques complexes, et une bonne stratégie pour les calculs de ces estimateurs n'est souvent pas une mince tâche. »

Selon Pierre Duchesne, tout reste à faire au GERAD en matière des statistiques de la décision, mais aussi tout est possible. « Le GERAD représente une plateforme incroyable, avec ses ressources, sa réputation, ses contacts. Comme statisticien, je pense que la masse critique est présente, mais le plein potentiel des ressources disponibles n'a peut-être pas mené à des contributions d'une envergure comparable à ce qu'ont obtenu les collègues spécialistes de la recherche opérationnelle. De plus, il y a des statisticiens de réputation mondiale à Montréal qui ne sont toujours pas membres du GERAD. Pour faire valoir tout le potentiel, il faut des chercheurs qui décident de "mener à bien l'attaque", poursuivant ainsi l'analogie du hockey, ce qui est parfois difficile pour les jeunes chercheurs qui doivent surtout consacrer les premières années de leur carrière à s'établir et à se faire reconnaître comme chercheurs. » Pierre Duchesne cite comme exemple à suivre la participation du GERAD depuis quelques années aux Colloques de Statistique CRM-ISM-GERAD organisés conjointement avec le Centre de recherche mathématique (CRM) et l'Institut des sciences mathématiques (ISM).

La biostatistique représente aussi, selon Pierre Duchesne, un domaine qu'il aimerait voir se développer au sein du GERAD et du Département de mathé-

matiques et statistique de l'Université de Montréal. « Nous avons des étudiants qui aimeraient poursuivre dans cette direction et il faudrait pouvoir l'offrir. » Il reconnaît aussi que certains aspects de la biostatistique peuvent potentiellement donner lieu à des problèmes intéressants de séries chronologiques. Dès que des patients suivent des traitements médicaux dans le temps, ceci implique inévitablement une composante temporelle qui pourrait être modélisée avec des modèles plus ou moins sophistiqués de séries chronologiques.

Le parcours de Pierre Duchesne avant qu'il ne se joigne au Département de mathématiques et statistique de l'Université de Montréal est à l'image de « l'attaquant défensif » ou du « défenseur offensif ». Dès le Cégep, où il était en sciences humaines mais avec concentration en mathématiques, il se sentait attiré autant vers les applications que la théorie. Après un baccalauréat en mathématiques (orientation statistique) à l'Université de Montréal, il poursuit au niveau de la maîtrise en statistique, où son mémoire de maîtrise porte sur les aspects de la robustesse dans les modèles de régression. Au doctorat, ses contributions se situent dans la théorie de l'échantillonnage ainsi que dans l'analyse des séries chronologiques. Par la suite, il a fait un stage post-doctoral aux départements de sciences économiques et statistiques

à l'université Cornell. Professeur de statistique au Service de l'enseignement des méthodes quantitatives de HEC Montréal de 2000 à 2003, il était perçu par ses étudiants davantage comme un théoricien qu'un statisticien appliqué. Aujourd'hui, comme professeur au Département de mathématiques et de statistique à l'Université de Montréal, il enseigne notamment aux étudiants en actuariat qui, eux, peuvent être perçus comme des champions des applications des techniques statistiques à l'actuariat. Il supervise de nombreux étudiants à la maîtrise et au doctorat en statistique. Il est également responsable au département de mathématiques et statistique, conjointement avec le département de sciences économiques et le département d'informatique et de recherche opérationnelle, de l'option du programme de maîtrise et DESS en finance mathématique et computationnelle de l'Université de Montréal.

De ce parcours, il ressort le truisme que la théorie et les applications vont de pair. Qui plus est, le marché de l'emploi en témoigne de façon éloquente. Selon Pierre Duchesne, la demande de statisticiens est inouïe et le taux de placement des étudiants est de 100 %. Ils peuvent travailler dans plusieurs domaines et plusieurs milieux, notamment dans les centres de recherches, les hôpitaux, les banques, les compagnies d'assurances, pour ne nommer que ceux-là. **G**



Newsletter

Groupe d'études et de recherche
 en analyse de décisions

GERAD

Statistics in Decision Theory at GERAD

For those who built GERAD, all operations research experts, statisticians would necessarily be called on to play an important role in the research centre dedicated to decision theory and analysis. The idea of combining the two areas of research was perfectly natural for them. It jibed with the image of the departments of management sciences to which most of them belonged.

The idea was only recently able to gel, however. When a group of young statisticians and probability experts arrived, particularly at HEC Montréal, GERAD was able to pursue the vision of its founders. Achieving this critical mass will undoubtedly result in an acceleration and broadening of GERAD's contribution in this area. Here are some promising examples. GERAD is now organizing a series of regular seminars jointly with the Centre de recherches mathématiques (CRM) and the Institut des sciences mathématiques. In addition, a group comprising two operations research specialists and two statisticians recently obtained an FQRNT team research grant. And a thematic workshop on data mining was even more successful than the most optimistic among us predicted.



Please see Statistics on page 2...

bulletin@gerad.ca

SUMMARY

GERAD Update.....	2
When life experience and research meet	3
Road safety without statistics is like night-driving without headlights	3
Behind all decisions, in all probability there are... statistics!	5
The statistician as a doctor of data.....	9
Statisticians and hockey players have more in common than the long pages of stats.	11

GERADUPDATEGERADUPDATE

- **Jean-François Cordeau** received a \$100,000 per year Canada Research Chair in Transportation Logistics for the period from April 1, 2006, to March 31, 2011.
- **Pierre Hansen, François Bellavance, Gilles Caporossi** and **Denis Larocque** obtained a FQRNT team grant for \$45,000 a year (2006-2009) and an equipment grant of \$15,930 a year (2006-2007).
- **Pierre Hansen** was elected member of the International Academy of Mathematical Chemistry in March 2006.
- GERAD has had four new members since January 20, 2006: **Michel Perrier**, École Polytechnique de Montréal, **Frank Galiana**, McGill University, **Shie Mannor**, McGill University, and **David McDonald**, University of Ottawa (associate member).
- **Shie Mannor** holds a Canada Research Chair in Machine Learning / Apprentissage automatique for \$100,000 a year from 2005 to 2010.
- The 2007 Optimization Days will be held May 7 through 9, 2007, at HEC Montréal. The organizing committee includes Olivier Bahn, Gilles Caporossi, Pierre Hansen and Sihem Taboubi.
- GERAD has had its FQRNT grant renewed for an amount of \$364,000 a year.
- GERAD was one of six centres asked to make a presentation at the strategic grouping meeting held by the FQRNT on February 23, 2006. See a summary of the presentation at <http://www.fqrnt.gouv.qc.ca/nateq/Actualite/PWP/GERAD-RolandMalhame.pdf>.
- Professor **Gilbert Laporte** was elected Fellow of the Institute for Operations Research and Management Sciences.
- The firm Ex Pretio, a GERAD/CRT spin-off company led by **Gilles Savard** and his colleagues, joined the IATA (International Air Transport Association). Ex Pretio specializes in dynamic pricing for rail and air transportation tickets.
- GERAD was responsible for logistics at the Workshop on Call Centres held on May 11, 2006. GERAD also contributed financially.
- The 18th EURO mini-conference on the Variable Neighbourhood Search was held in Tenerife, Spain, November 23 to 25, 2005. The VNS is a method created by professors **Pierre Hansen** and **Nenad Mladenovic**.
- A book by Professor **El-Kébir Boukas** entitled *Stochastic Switching Systems, Analysis and Design* was published by Birkhauser.
- In May 2006, **François Bellavance** was appointed Head of the M.Sc. and PhD programs at HEC.

... Statistics from page 1.



In this issue of the GERAD Newsletter you will meet five statisticians who apply their knowledge to decision theory in such varied areas as medicine, road safety, financial engineering, marketing, and more. These five researchers are involved in theory and their preferred applications, but they all share the drive and passion for research and they generously transmit it to their students. For GERAD, this is very promising...

Enjoy reading the Newsletter!
Georges Zaccour

GERAD Newsletter

Published 2 to 3 times a year by GERAD.

Director
Georges Zaccour
georges.zaccour@gerad.ca

GERAD
HEC Montréal
3000, chemin de la Côte-Sainte-Catherine
Montréal, Québec, Canada H3T 2A7
Telephone : (514) 340-6053

Web site
www.gerad.ca
bulletin@gerad.ca

Editor
Robin Philpot
rphilpot@sympatico.ca

Translation
Robin Philpot

Graphic Design
HEC Montréal

Legal deposit: 2nd quarter 2006
Bibliothèque nationale du Québec

Reproduction authorized with
acknowledgement of source.

Statistics at GERAD

When **life** experience and **research** meet



Brenda
MacGibbon

Like James Watt, who perfected the steam engine after a boiling kettle piqued his curiosity and whose name became a household word, the scientific work of many researchers is determined by their personal experience and day-to-day observations. **Brenda MacGibbon**, Adjunct Professor in the UQAM Math Department and GERAD's first statistician, is an outstanding example of how life's joys and trials can influence a researcher. Her personal story is also proof of the potential of applying quantitative methods to decision-making in health questions.



François
Bellavance

After obtaining her PhD in Probability and Analysis at McGill and doing post-doc studies in Paris, Brenda was back in Montreal in the early 70s where she gave birth to a daughter and twin boys within a period of eighteen months. She wasn't "working" and "the only adult other than my husband I saw was the pediatrician who just happened to know the top geneticist in Quebec, Dr. Charles Scriver. He was attached to the Montreal Children's Hospital where they were working a lot on genetic screening for diseases such as cystic fibrosis. Having had twins, I was particularly interested in genetics and the Montreal Children's Hospital happened to have a part time job in the genetics department. That's how I learned statistics and computing, since in those days, you could get a doctorate in Math without knowing too much about computing."



Bruno Rémillard

The Montreal Children's Hospital position was not exactly a "tenure track" so during her husband's sabbatical, Brenda obtained a visiting Research Associate position in Statistical Decision Theory at Cornell University. At that point, she began publishing and making contributions in theoretical statistics. Before joining UQAM's Math Department in 1988, Brenda MacGibbon taught at Carleton University, the University of Sherbrooke in the Mathematics and Statistics Departments and Concordia in The Decision Sciences and Management Information Systems Department.



Denis Larocque

Her research work at UQAM brought her to GERAD even before UQAM itself had joined GERAD. "I was trying to solve a minimax problem (i.e., minimizing the maximum risk), which was harder than finding the simple maximum or minimum. I couldn't manage to compute it by local optimization methods so I crossed Mount Royal and, with Brigitte Jaumard, was able to solve the problem using global optimization methods. That was the beginning of several years of collaboration. GERAD made me an associate member until UQAM joined officially."



Pierre Duchesne

Other than her earlier work in genetics, Brenda was not yet involved in health-related statistics. "What changed the direction of my research was that I got breast cancer in 1997," she recalls. "When you get such a disease yourself, it makes you think about your life and change it. One change I made was to do more biostatistics and less theoretical statistics. Another reason was that my best friend from Cornell days was a biostatistician at the University of Southern California and she had steadily kept me interested in the field." Brenda also saw concentration on biostatistics as a way to better direct her largely Master's level students who, until then, had difficulties with very theoretical statistics. "I chose biostatistics and survival analysis as a focus for my students."

... **Brenda MacGibbon** from page 3.

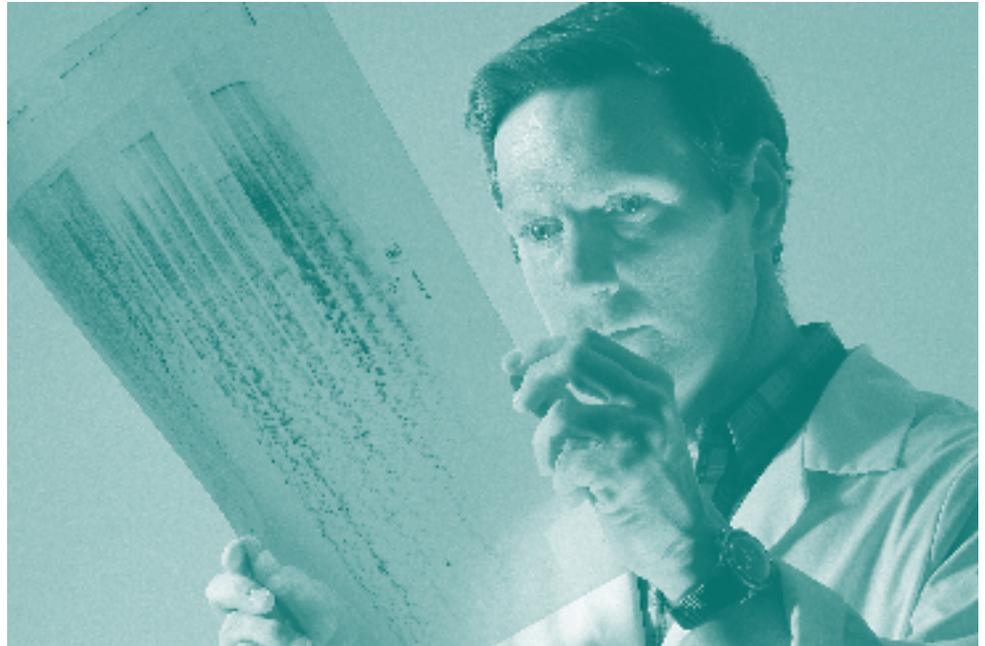
The challenge will be to maintain the creativity while harnessing the computer.

In the November 2005 Newsletter, GERAD director Roland Malhamé expressed his hope of increasing research in biotechnology and biology at GERAD. This was music to Brenda MacGibbon's ears since the field is vast and quantitative methods are increasingly important and productive. "There is a whole theory of statistics behind decisions made as to whether one treatment is better than another. Without these methods, doctors have to base their decisions on their informed intuition. Statistical decision theory can complement this intuition."

As can be imagined, survival analysis abounds with statistical and mathematical problems. These include missing data problems – for some the actual survival time is known, for others it is not known at all –, as well as censored and truncated data issues owing to trial time limits, for example. Brenda's work in survival analysis has focused mainly on different types of cancer (breast and skin cancer) and end-stage renal disease, which involves multidimensional risk analysis.

After she got breast cancer, Brenda became a patient advocate and statistician on the granting agency for the California Breast Cancer Research Program for three years. All members who did peer reviews had to be from outside California. This work resulted in a very interesting collaboration with a surgeon who was striving to develop breast cancer diagnosis methods that are less invasive than a biopsy. "According to some, biopsies can make the cancer worse," notes Brenda. "Our goal was to refine the diagnostic tool and to determine how reliable it is."

Another area where biostatistics could usefully be applied at UQAM was in CINBIOSE (interdisciplinary research centre on biology, health, society and the environment) where they were inter-



ested in the detection of early neurological effects on the Cree population in Northern Quebec of high mercury levels in the fish they eat or in workers exposed to manganese. Very sensitive machines were built by Anne Beuter and her team to measure tremors, and statistical methods could be used to determine how accurately the machine could distinguish between normal tremor and different abnormal types. One common thread therefore in Brenda's research interests is to use quantitative methods to evaluate and refine diagnostic tools.

Working in biostatistics leads sooner or later to genetics. "Yes," says Brenda, "I've now made a full turn back to my first love of genetics. Childhood leukemia, for instance, raises a vast number of mathematical problems because the number of patients is relatively very small, whereas the number of genes is enormous." In returning to genetics, she was struck by the impact of the exponential growth in computing power and warns against the danger that computers could stifle the imaginative and creative thinking that characterized the geneticists in the first half of the 20th century. The challenge will be to maintain the creativity while harnessing the computer.

In the future, statistics will also be used in drug policy and surveillance. Brenda MacGibbon is now on the Canadian Drug Policy development coalition, a group of Canadian researchers who desire to influence the way the federal government will monitor drugs once they have been approved and are on the market, an area that has long been neglected. The question will be to see how huge administrative databases such as the RAMQ, Quebec's drug insurance agency, could be used to support surveillance. Would it be possible, for example, to have obtained more information about VIOXX and its side effects earlier by using these data bases? Or about other drugs' side effects? If so, decisions made about the use of different drugs will be more enlightened, more accurate, and information to the consumer will be available much earlier.

In each of the areas in which Brenda does applied biostatistics, her aim is also to develop new methodology and theory to help solve the problems that arise. **G**

Road safety without statistics is like **night-driving without headlights**

On reading headline statistics forecasting imminent catastrophe, who has never reacted sceptically as did Mark Twain when he said there are three types of lies: lies, damn lies, and statistics? Statistics are nonetheless at the heart of important government and corporate decision-making on issues that concern each and every one of us. That is why the population demands the highest standards of rigour, accuracy, and judgment of statisticians. **François Bellavance**, who is a member of GERAD and CRT and Professor in HEC's Department of Management Sciences and who specializes in statistics applied to road safety, understands the crucial role of statistics in government decisions in this area.

Here are some statistics that could make headline material and move government decision-makers. Studies show that you are five times more likely to die if you are not wearing your seatbelt when you have a serious car accident. Twenty-five lives could be saved every year in Québec if seatbelt use could be increased by one percent. Throughout the world, three thousand people die every day due to road accidents. The risk of death is doubled if a car driver is hit by an SUV (sport utility vehicle) instead of by another ordinary car. The use of cell phones by drivers increases the risk of an accident by 400 percent.

According to François Bellavance, all of those statistical statements are reliable except for last one on cell phone use, even though it came from a study conducted in Toronto that was published in a prestigious medical journal and was corroborated by an Australian study. The problem with the cell-phone statistical statement, and the error therein, stems from the methodology. However, if all governments were to rely on the results of the first studies (i.e., four times greater risk for

those who use cell phones while driving), there would be laws everywhere prohibiting drivers from using cell phones.

The problem with the cell-phone statistical statement, and the error therein, stems from the methodology.

In 2003, François and his colleagues published a study on the use of mobile phones and the risk of road accidents in which they showed that the accident risk was not four times higher but rather 1.38 times higher. They conducted their study in collaboration with the SAAQ (Société de l'assurance automobile du Québec), for mailing of questionnaires, driver records, and police reports, and with the help of cell phone companies for telephone activity files. The strength of their study lies in the size of the cohort used and the volume of data brought to bear. For comparison purposes, they did a parallel study using the same data but following the methodology used in the Toronto study. In that parallel study, they too reached the conclusion of a four-fold increase in risk.



... François Bellavance from page 5.

François Bellavance draws certain conclusions from the experience. First, it is very difficult to determine whether or not the driver was on the phone at the time of the accident. One of the reasons is the vagueness of police reports. It is also very difficult to establish a cause and effect link between cell phone use and car accidents. It can nonetheless be stated that the risk of an accident does actually increase but it is not four times higher, and that the sub-group of heavy cell phone users among car drivers run a higher risk of accident than occasional users. However, if legislators were to outlaw cell phone use for drivers, they would likely create all kinds of problems in enforcing the law. It should be noted that, despite the much lower risk observed, François Bellavance never uses a cell phone while driving! His studies have also convinced him never to move the car until everybody has their seatbelt on and to always rigorously respect the speed limits. Statistics don't lie in this case. François has never had a car accident!

Road safety raises many questions related to public health and statistical studies can provide answers. These include the impact of billboards along highways on the number and seriousness of accidents, the impact of tightening up truck-driver licence granting procedures (a study conducted for the SAAQ confirmed the positive effect of such a measure), the impact of the incompatibility of SUVs and ordinary cars on the seriousness of accidents, and many more. Moreover, adds François Bellavance, whenever we deal with the seriousness of accidents, the study immediately becomes more complicated from a methodological standpoint. "Proportionally, there are not very many accidents with serious injuries as compared to all of the road accidents. For example, for a year during which there were 700 road accident casualties, there will have been hundreds of thousands of accidents. It is therefore similar to other fields where incidents occur rarely, such as rare diseases. The challenge is to

develop suitable statistical methods to analyse the data, and particularly non-parametric statistical methods."

Before joining HEC and becoming GERAD's second statistician (after Brenda MacGibbon), François Bellavance headed the Statistics Consultation Department at Simon Fraser University in Vancouver, British Columbia, and at the same time he was completing his doctorate at the Université de Montréal. In addition, as a consulting statistician at the Institut de recherches cliniques de Montréal and later as a biostatistician at McGill University and St. Mary's Hospital Centre, François Bellavance had a hand in several health-oriented research projects. One of these studies focussed on elderly people who were losing their autonomy.

Older people faced with acute problems tend to go emergency wards where they have long waits that inevitably upset them and also overload the health system.

The goal of the study, centring on patients at four Montreal hospitals, consisted in developing a tool that would help identify those who were most likely to experience a relapse, and then find solutions other than a return to the emergency. A solution could involve sending a social worker to the person's home, having a local clinic intervene, or a getting the hospital out-patient clinic to monitor patients more closely. The target tool was a simple questionnaire that emergency patients could fill out without difficulty. To establish it, however, it was first necessary to dig into hospitalization data bases, verify returns to the emergency ward, and make calls to patients at different intervals. In other words, behind an apparently simple questionnaire, there are studies that are by no means simple. The result, said François Bellavance, was a tool that offered a relatively accurate prediction rate. **G**

François Bellavance never uses a cell phone while driving! His studies have also convinced him never to move the car until everybody has their seatbelt on and to always rigorously respect the speed limits. Statistics don't lie in this case. François has never had a car accident!



Behind all **decisions**, in all probability there are... **statistics!**

"Statistics are used wherever decisions are made!" observes **Bruno Rémillard** straightforwardly. Bruno is member of GERAD and Full Professor in the Department of Management Sciences at HEC Montréal. The main focus of Bruno Rémillard's work is in probability and statistics, two fields that concern everybody.

One lesson is that if you are deciding whether or not to buy a lottery ticket, the possibility of winning the jackpot should not be a factor in making the decision!

Who in daily life has never wondered what the odds are that such or such an event will occur, be it a plane crash, catching avian flu, winning the sweepstakes, and so on? The subject is so popular that one of Bruno Rémillard's colleagues in probability published a book that became a 2005 bestseller. The book popularizes the basic concepts of probability and aims help readers fight irrationality in decision-making (*Struck by Lightning: The Curious World of Probabilities*, Jeffrey S. Rosenthal). One lesson is that if you are deciding whether or not to buy a lottery ticket, the possibility of winning the jackpot should not be a factor in making the decision!

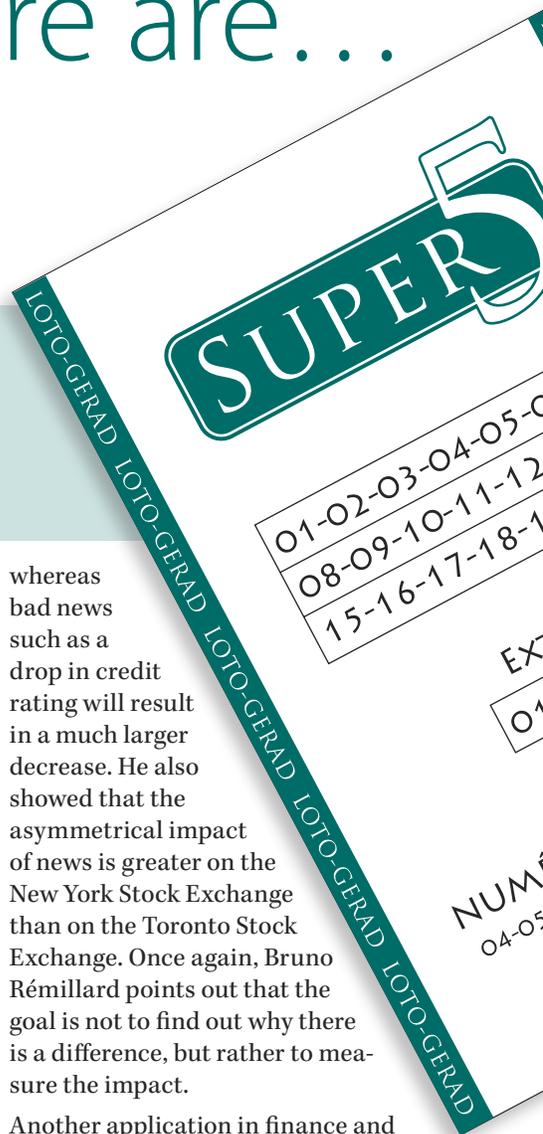
While Bruno Rémillard's work has many and varied applications, his research work is a different level from that of his colleagues in statistics. "Unlike many statisticians, I do not concern myself with the methods used to gather data. I assume that the data I have are reliable and I try to build tools to verify certain hypotheses. In addition, I build new ways of breaking data down and verifying models." A case in point is the much dis-

cussed issue of consumer gasoline price fluctuations. Bruno Rémillard has shown that even though over the past few years prices have both risen and dropped, the price increases have been higher than the decreases. (What we've all felt is therefore true.) When asked who has profited from the increases, Bruno smiles but remains coy: "that was not the goal of the study."

The article Bruno and his colleagues B. Abdous and K. Ghoudi wrote about the gasoline price fluctuations was published in *The Canadian Journal of Statistics* and won the award for the best article of the year in 2003. The Journal underscored the outstanding quality of presentation and the contribution to methodology. (Nonparametric weighted-symmetry tests, *Canadian Journal of Statistics*, 2003). The models developed also help to observe and evaluate other similar phenomena. For example, Bruno Rémillard, who is particularly interested in financial engineering, also observed a similar asymmetry in the reaction of securities on the Toronto New York stock exchanges to good news and to bad news. Good news, such as a company's improved credit rating, will generate a small increase in the value of the security,

whereas bad news such as a drop in credit rating will result in a much larger decrease. He also showed that the asymmetrical impact of news is greater on the New York Stock Exchange than on the Toronto Stock Exchange. Once again, Bruno Rémillard points out that the goal is not to find out why there is a difference, but rather to measure the impact.

Another application in finance and also in actuarial sciences and biostatistics is the modelling of dependence. This can mean the dependence of an investment on certain events, the relationship of dependence between stock prices on the Toronto and New York stock exchanges, the relationship between two or many different securities, and more. "We look at both variables to determine whether or not there is dependence and, as the case may be, we measure and model that dependence." Banks and other



...Bruno Rémillard from page 7.

large corporations are obviously very interested by this type of research.

Before joining HEC Montréal, Bruno Rémillard was Professor of Mathematics at the Université du Québec à Trois-Rivières from 1988 to 2001, after having completed post-doctoral research at Cornell. He did his B.Sc. and Master's at Université Laval in Québec City and his PhD at Carleton University in Ottawa. He arrived at HEC Montréal at a moment when the Montreal business school was building its team of specialists in quantitative methods. As soon as he arrived, Bruno joined GERAD for a number of reasons including the fact that his HEC colleagues in statistics were all members and that statistics are naturally linked to decision theory. He was particularly impressed by the dynamism of the seminars on statistics organized by GERAD. Georges Zaccour has also attributed the growing importance of statistics at GERAD to the presence of a critical mass of statisticians.

Though biostatistics is not among Bruno's research priorities, it presents a particular interest because of the captivating problems that are raised from a statistics standpoint. These include truncated, censored, and missing data problems. "When dependence between two variables is modelled, a natural stage involves studying censored or truncated data, and biostatistics represents the main source of censored data".

Statistical decision theory has a very promising future according to Bruno Rémillard. "Statistics are more and more at the heart of corporate operations and everywhere in daily life. Companies are looking to obtain quick and automatic decision-making solutions. Vast statistical research, however, underlies all these solutions. In other words, the greater the demand for automated decision making, the greater the need for advanced statistics."

For instance, when a bank receives a loan request, customers are requested to fill out a questionnaire that is simple but



The goal is to detect the target so as to accurately determine the identity of the other plane. The solutions will involve identification algorithms based on statistics.

that has many questions. Based on the answers, the bank will make an enlightened decision on whether or not to grant the loan. The decision is enlightened by a large quantity of statistical data that have been analysed using a variety of methods and that can accurately establish the probability of a borrower defaulting on

the loan. Another example is the recognition of uncooperative aerial targets or friend-enemy plane recognition. The goal is to detect the target so as to accurately determine the identity of the other plane. The solutions will involve identification algorithms based on statistics. **G**

The statistician as a **doctor** of data

Denis Larocque, who has been a member of GERAD since he joined the HEC Montréal Management Sciences Department in 2000, compares his role as a statistician to that of a doctor: “Somebody comes to me with a problem. I listen, ask questions about the nature of the data and how they were gathered as well as about the hypotheses and research questions, and then I provide a diagnosis that consists in proposing the appropriate analysis of the problem.” The validity of the conclusions depends to a large extent on the quality of the data and specifically on the initial sampling plan. For Denis Larocque, it is important, therefore, for statisticians to bring their expertise to bear at all stages of a research project and throughout any decision-making process calling on statistical analysis.

Collaboration between operations research specialists and statisticians can be built when one knows how to distinguish the two areas of research. Denis Larocque explains the difference quite simply. “In operations research, people often begin with a specific problem that comprises certain hypotheses and constraints and then try to find an optimal solution, in terms of cost for instance. In statistics, however, we work with a sample taken from an overall population and we apply statistical methods to draw conclusions that can be generalized to the entire population with a certain margin of error.”

The potential for collaboration at GERAD is significant from both a methodological and applications standpoint, he adds. A



The validity of the conclusions depends to a large extent on the quality of the data and specifically on the initial sampling plan.

case in point is data mining, which is like panning for gold in quicksand, considering the vast quantities of disparate data available. In data mining, operations researchers and statisticians complement one another. The knowledge in optimization and classification methods of the former combined with the knowledge of sampling and statistical methods of the latter can provide decision-makers with relevant and highly usable information. In the area of methodology, Denis Larocque refers to a project initiated with

Pierre Hansen, François Bellavance, and Gilles Caporossi at GERAD on the development of classification tree methods for longitudinal or multivariate data. He also notes the publication by GERAD of the book entitled *Statistical Modeling and Analysis for Complex Data Problems*, one of the ten volumes produced to mark the research centre’s 25th anniversary.

Denis Larocque got the “bug” for probability and statistics during his first course in the subject at Cegep. He immediately saw the huge potential and has

...Denis Larocque from page 9.

pursued in the same direction ever since. He studied mathematics and statistics at the Université de Montréal where he obtained his PhD, followed by post-doctoral research with the Groupe de recherche sur l'inadaptation psychosociale de l'enfant, also at U de M. He then did post-doctoral studies at the University of Florida. Before joining HEC Montréal, where he is now Associate Professor, he taught statistics for a year at the Université du Québec à Trois-Rivières.

Denis Larocque is particularly interested in nonparametric multivariate statistical methods. Simply speaking, the validity of nonparametric methods requires fewer assumptions about the nature of the data, such as their normality, than do parametric methods. These methods were developed in the 1940s and 1950s (Wolfowitz used the term nonparametric for the first time in 1942). "Today, the term nonparametric encompasses several families of methods," he points out. "One of them, inference based on ranks, has the advantage of being valid even for small samples, and it naturally inherits robustness with regards to outliers in that the results will not be affected by them. Another family, smoothing techniques, allows the data to speak rather than have them confined to a model."

As with researchers in combinatorial optimization, Denis Larocque underscores the breakthroughs that have been made as a result of today's increasingly powerful computers that have revolutionized research in statistics. "The power of computers enables us to use methods and adjust models that we never dreamed of using only a few years ago." As an example, he cites the bootstrap method. The metaphor is inspired by the legend of Baron Munchausen who was supposed to have pulled himself out of a swamp or the sea by pulling on the straps of his own boots. Similarly, by repeatedly re-sampling an original sample, it is possible to estimate the interesting features (an estimator's bias or standard deviation) or improve the forecasting performance



He cites the bootstrap method. The metaphor is inspired by the legend of Baron Munchausen who was supposed to have pulled himself out of a swamp or the sea by pulling on the straps of his own boots.

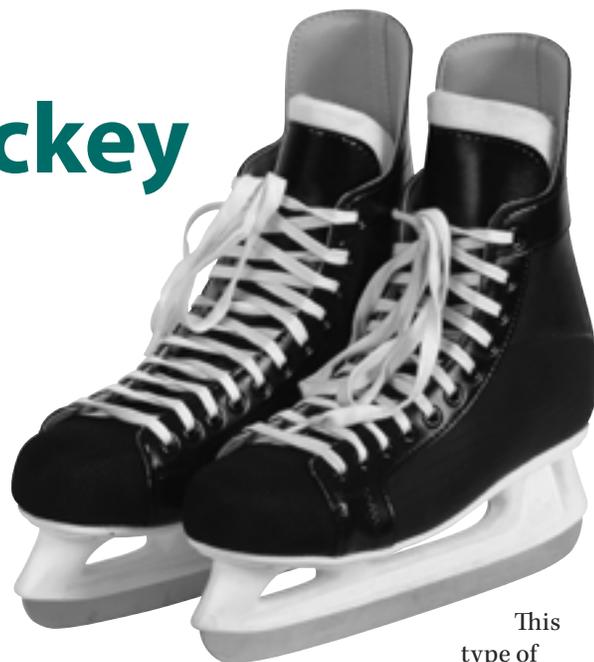
of a classifier (bagging). It is a multi-use method that, for instance, is able to yield results that cannot be obtained analytically. Without the computer, however, it would be impossible. The computational power thus becomes a tremendous boon for both theory and applications.

As well as doing fundamental research in statistics, Denis Larocque collaborates with researchers in other disciplines, including operations research, information technology, marketing, road safety, and psychology. The interest in statistics in the business world, as in other sectors, is steadily increasing, according to Denis Larocque. To make the point, he refers to his master's students in Business Intelligence. "Our students are offered

jobs even before they finish their degrees. Statistics has never been as lively as it is today. There have never been so many scientific journals and seminars. The number of publications has mushroomed."

What's more, GERAD is in the right place at the right time! Denis Larocque recalls that, in 2000, Georges Zaccour was the one who proposed that GERAD develop a nucleus of statisticians at a time when HEC Montréal was in fact hiring professors in statistics. He concludes that the huge amounts of data to be analysed combined with their ever increasing complexity means that the demand for competent data doctors will continue over the years to come. **G**

Statisticians and hockey players have more in common than the long pages of stats



Knowing that the first question asked of a researcher is often about the “theoretical” or “applied” nature of their work, **Pierre Duchesne**, Professor in the Université de Montréal’s Department of Mathematics and Statistics and a GERAD member, found the best analogy in hockey. Just as some hockey players can be “defensive forwards” (like Guy Carbonneau) or “offensive defensemen” (like Bobby Orr), Pierre Duchesne describes himself as an applied statistician who often forays into the theoretical zone or a theoretical statistician who is known to rush back to the applications zone.

Pierre Duchesne is particularly interested as a statistician in time series analysis from both a theoretical standpoint and in their applications, especially in econometrics and financial econometrics. The latter is a discipline that has fully emerged over the past ten years and involves the development of statistical and econometrical methods and their application to financial data. Chronological series intrinsically comprise a temporal component meaning that, unlike more classical statistical techniques, it cannot be assumed that the data gathered represent a random sample (i.e., that they have been collected independently from each other).

“For the type of data I’m working with, there is always the notion of dependence in time,” explains Pierre Duchesne. “Statisticians and econometricians real-

ized it was necessary to have models designed to explain this type of data. Many economic series can be modelled adequately using time series models. Other fields of study, such as hydrology, have used these models very successfully to predict flooding, for example. Another example is the study of time series based on gasoline prices, which is always a good subject. If I know the background of gas prices in Montreal and if I have other time-based variables that could be related to these data, I would be able to acquire a better understanding of the phenomena explaining the price. I am also co-directing a doctoral student (now a university professor) who is completing her work in the finance program at HEC Montréal. She models high-frequency financial data, namely stock exchange transactions. Time series models enable her to model the transactions by attempting to explain the time lapse between them.”

Pierre Duchesne adds that the goals of this type of study include a better understanding of market fluctuations and the dynamics behind the events that occur, be it a new transaction or a change in price. “We try to determine what happens at the microstructure level in the market so as to explain how the information is incorporated in the price, how stock exchange activity varies in the course of a day, what determines market volatility and liquidity, and so forth.”

This type of analysis does not give the statistician or the econometrician a crystal ball, but there is no doubt that these studies provide an enhanced understanding of the different exchange structures and of investors’ strategies when submitting orders, etc. What’s more, new quantitative measures can emerge, such as construction of intraday volatility measurements and intraday risk measurements. The potential is very significant for market regulators as they devise optimal policies. Pierre Duchesne observes that, considering the availability of detailed chronological stock exchange data now – before readings taken every 15 minutes or more were the best available – and thanks to modern computing power, it is now in the realm of possibility to explain stock exchange behaviour over different intraday horizons.

Time series models make it possible to forecast and to measure the quality of the forecast. This concern of statisticians at GERAD to obtain estimations and forecasts as well as to measure the quality of the statistical quantities, tends to coincide with the concern of optimizers at GERAD who not only wish to find the optimal solution but also the proof that it is optimal or else a measurement of how much it differs from the optimum. Pierre Duchesne sees the comparison as more than a coincidence. Bringing statistics together with optimization done by oper-

...Pierre Duchesne from page 11.

ations research people results in much more than the sum of the two parts. “Statisticians are seeking good estimators for different situations,” he points out. “We are inevitably confronted with statistical estimation problems. Moreover, the estimators are often obtained using somewhat complex criteria that lead to optimization procedures, and that is at the heart of operations research and optimization theory. Statistical techniques also have to deal with complex numerical problems, and finding a good strategy for calculating these estimators can be a challenge.”

Pierre Duchesne believes that GERAD is really just beginning in the field of statistics, but that the potential is enormous. “GERAD represents an unbelievable platform, with its resources, its reputation, and its contacts. As a statistician, I think that it has the critical mass necessary, but the full potential of the resources available has yet to produce results comparable to what our operations research colleagues have obtained. There are also world class statisticians in Montreal who are not yet members of GERAD. In order to develop the potential, some researchers will have to ‘go on the offensive’, to use the hockey analogy again, which can be difficult for young researchers who must dedicate their first years to establishing themselves and building their reputations in research.” Pierre Duchesne cites as an example to follow GERAD’s participation in recent years in the CRM-ISM-GERAD seminars on statistics organized jointly by the Centre de recherches mathématiques (CRM) and the Institut des sciences mathématiques (ISM).

Biostatistics also represents an area that Pierre Duchesne would like to see develop both at GERAD and the Mathematics and Statistics Department at the Université de Montréal. “We have students who want to pursue work in that direction and we should be offering it.” He also recognizes that certain aspects of biostatistics can potentially give rise to interesting time series problems. Once patients are follow-

ing treatments over time, there is inevitably a temporal component that could be modeled with time series models of varying degrees of sophistication.

Before he joined the Mathematics and Statistics Department at the Université de Montréal, Pierre Duchesne’s career was like that of the “defensive forward” or the “offensive defenseman”. In Cegep, he studied humanities but majored in math and was interested in both applications and theory. After obtaining his B.Sc. in math with a statistics orientation at the U. de M., he went on to do a Master’s in statistics and his thesis focused on aspects of robustness in regression models. At the PhD level, his contributions were in sampling theory and in time series analysis. He then did post-doctoral research in the Department of Statistical Science and the Department of Economics at Cornell. At HEC Montréal from 2000 through 2003 as Professor of Statistics in the Department of Management Sciences, he was seen by his students more as a mathematical statistician than as an applied sciences stat-

istician. Now as Professor of Mathematics and Statistics at the U. de M., he teaches many actuarial students who are known to be champions in statistical applications in actuarial sciences. He also supervises many Master’s and PhD students in statistics. Within his department, he is jointly responsible, along with the economics and computer science and operations research departments, for the Master’s program and the DESS (a higher education degree) options in mathematical and computational finance at the Université de Montréal.

His background thus reinforces the truism that theory and applications are inseparable. Moreover, the job market further confirms it. Pierre Duchesne points out that the demand for statisticians is overwhelming, with the student placement rate being 100 percent. They can work in many fields and many surroundings, including research centres, hospitals, banks, and insurance companies, to name only a few. **G**

