

Lu pour vous

ROBERT BILINSKI
COLLÈGE MONTMORENCY

Sous la présente rubrique, vous trouverez ma première chronique à thème unique ! Ça fait longtemps que j'essaie d'en faire une. J'ai patiemment récolté et accumulé des livres sur un seul sujet pour les présenter dans un tout. Voici donc sept livres sur les statistiques... À peu près tous les styles sont couverts : des livres de vulgarisation grand public et savante, des livres scolaires pour historiens, pour statisticiens et pour ingénieurs, des livres pour praticiens et un livre d'histoire des statistiques. La seule chose qui manque (et c'est seulement un choix éditorial de ma part, car la recension est prête) est un livre sur les « erreurs statistiques ».

**Claudine Robert, *Contes & Décomptes de la statistique Vuiber*,
2003, 200 p., ISBN 2-7117-5320-4, environ 28 \$.**



Ce livre à la couverture florale nous invite à un voyage au pays des statistiques, mais surtout de la statistique. Le langage est simple et abordable. Les exemples sont forts nombreux et le texte est parsemé de nombreuses illustrations humoristiques. La matière présentée est celle d'un cours typique d'introduction aux statistiques de base (un peu de probabilité, statistiques descriptives, régression et intervalle de confiance). Par contre, l'approche est

loin d'être typique, et surtout, les exemples sont empreints d'expérience et brisent quelques carcans dans lesquels on tombe lorsque l'on enseigne les statistiques sans les pratiquer.

Une illustration de cette approche ne tarde pas à venir. Dès le premier chapitre, qui traite des histogrammes et autres représentations graphiques, l'auteur nous amène à utiliser un nuage de points pour déceler une structure sous-jacente aux données. Le nuage a ceci de particulier qu'il est fait avec une série statistique unidimensionnelle que l'on « garoche¹ » immédiatement dans tout « bon livre » de statistique dans un histogramme. Comme on sait, les histogrammes nous font perdre de l'information sur les données. On pense souvent à la perte de position à l'intérieur des classes, alors que l'on perd plus comme le démontre ce livre. À la page 13, on représente par un nuage des tirs à l'arc successifs. On se rend vite compte qu'en faisant l'histogramme on perd aussi l'ordre des tirs en comparant l'histogramme et le nuage de points des mêmes données. Pourtant je suis convaincu que cet exemple serait traité dans n'importe quel livre scolaire uniquement avec un histogramme.

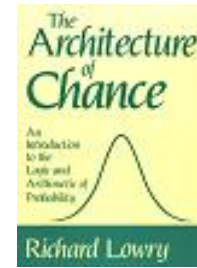
Dans chaque chapitre, on retrouve de ces exemples qui ouvrent l'esprit et font redécouvrir cette belle matière qui peut devenir routinière. J'ai vraiment eu un grand plaisir à lire cet ouvrage et je suis passé rapidement à travers. Même si le livre est « grand public », c.-à-d. avec beaucoup de texte et peu de formules, il reste tout de même plein de bonnes idées. C'est tout un changement par rapport aux textes de statistiques que l'on retrouve normalement.

Même si le livre est très bon dans son ensemble, je note quand même une petite faiblesse. En fait, les huit premiers chapitres sont extrêmement accessibles. Ce sont des textes travaillés, léchés et très lisibles. Vient alors le chapitre neuf qui est un choc par rapport aux premiers. L'auteur essaie de vulgariser en parlant des probabilités d'un point de vue de la mesure (elle parle presque de tribus et d'algèbres de Boole)! Cette partie contraste tant avec le reste... Ainsi, je vous recommande chaudement ce livre! Bonne lecture.

¹(Terme régional québécois) garocher : lancer violemment, ici : introduire violemment

Richard Lowry, *The architecture of chance*,

Oxford University Press, 1989, 180 p., ISBN 0-19-505608-6,
environ 45 \$.



J'ai un dada pour les probabilités, m'étant même « spécialisé » dans cette discipline. De plus, si je n'avais pas étudié en mathématique, j'aurais étudié en architecture. Le titre du livre avait donc tout pour m'attirer. Le résumé à l'arrière du livre présente son côté « bref et concis ». Il est certainement bref, mais est-il concis ? Il ne s'étale pas sur la théorie et s'en tient presque exclusivement aux applications. Ça le rend concis dans un certain sens.

L'introduction de ce livre donne rapidement le ton. En effet, dès la première page, on pose la question : « Est-ce que le jus de pomme augmente l'intelligence ? » On pourrait se demander quel est le lien avec les mathématiques. La réponse est donnée à la page suivante : « En sachant que sept personnes sur dix ayant pris du jus de pomme ont augmenté leur pointage dans un examen de Q.I. ... » On devine alors qu'il s'agit de statistique. On pourrait croire que ce livre est écrit par un statisticien. Un petit regard sur la couverture arrière nous indique qu'il s'agit plutôt d'un psychologue. De retour à la lecture, l'introduction du livre finit avec une déclaration de principe : « Ce livre a pour but de développer la logique et l'intuition des probabilités. » Assez surprenant ce livre, après tout... Un psy qui veut développer les maths ?!

Le résumé nous informe aussi sur la finalité de ce livre. L'auteur voulait donner la chance aux lecteurs qui suivent un cours obligatoire de statistique d'acquérir les préalables nécessaires aux probabilités. Les chapitres sont donc organisés dans l'ordre typique d'un cours de probabilité. On voit : les bases des probabilités, les règles d'addition et de multiplication, le concept d'indépendance, la loi binomiale à l'aide d'arbres, la loi normale qui est l'aboutissement d'une binomiale avec un grand n , calculs avec la loi normale et estimation à l'aide de la loi de Student. Le contenu est donc traditionnel, mais l'approche ne l'est pas. Elle nous permet de voir ce que les « autres » trouvent importants dans notre matière. De plus, les exemples sont assez nombreux et appliqués dans les derniers chapitres du livre (le début

a beaucoup de sous et de dés). Une autre surprise m'attendait quand j'ai lu la bibliographie : elle ne contient que des livres sur l'utilisation des statistiques, surtout l'utilisation abusive des statistiques !

Ce livre est donc concis et bref. C'est même un tourbillon. Je n'ai pu le digérer au complet et je vais probablement le relire (avec le livre *Elementary Probability with Applications* de Rabinowitz, recensé en mai 2005) avant d'enseigner Méthodes Quantitatives la prochaine fois. Ils me font penser à la question de la validité de l'approche au-delà de la validité de la matière. Comme disait ma collègue de bureau, Raymonde, ce matin : « On peut bien enseigner avec un mauvais livre et mal enseigner avec un bon livre ! » Le coup d'œil est différent, la matière est la même, la réflexion avance. . . Bonne lecture !

**Frédéric Saly-Giocanti, *Utiliser les statistiques en histoire*,
Collection Cursus Armand Colin, 2005, 191 p.,
ISBN 2-200-26545-X, environ 28 \$.**



J'ai trouvé ce livre dans un catalogue et le titre m'a tout de suite attiré d'un point de vue professionnel. J'espérais trouver, au mieux, un recueil pour enseigner en Méthodes Quantitatives (MQ) ou, au pire, un livre avec des idées pour des exemples afin d'enrichir en cachette mon enseignement dans ce cours. En effet, les exemples sont « faciles » à trouver en démographie, en géographie, en économie, en politique, etc., mais ceux en histoire ne courent pas les rues.

Par contre, en arrivant chez moi, ce livre a suscité une autre réaction : « Les historiens écrivent des livres de maths ! », me suis-je dis. Bien qu'un peu outré à l'idée initialement (vu les débats qui courent dans nos cégeps ces dernières années), je deviens de plus en plus habitué à cette mode « d'application » qui amène les éditeurs à aller chercher des pratiquants dans les domaines (espérons le « calé » en statistiques) plutôt que des statisticiens appliqués dans le domaine.

Le livre brise tout de suite avec la tradition des livres de statistiques. Le premier chapitre traite des séries chronologiques et de leurs représentations. Les statistiques descriptives ont été reléguées aux chapitres 3, 4 et 5. Entre les deux, on retrouve les distributions dans le chapitre 2. Dans le chapitre 6, on retrouve des représentations graphiques adaptées aux données historiques. Par la suite, on trouvera les pourcentages, indices (chap. 7 et 8), régression (chap. 9), graphique semi-log (chap. 10) et khi-deux (chap. 11).

Le contenu est visiblement très mathématique. L'auteur n'a pas peur de mettre les formules bien qu'elles n'ont pas la place que l'on trouve dans les livres écrits par les mathématiciens et les statisticiens. Il démontre à maintes reprises sa maîtrise de la matière. Les exemples, basés sur des données réelles et d'intérêt pour un historien, sont nombreux et variés. Par contre, les exercices sont peu nombreux : on en retrouve un ou deux par chapitre, avec 20 exercices d'intégration ou projets de complexité variable dans le douzième chapitre du livre. Chaque chapitre contient des instructions pour appliquer les connaissances avec un tableur comme Excel. Par contre, il faut au préalable avoir une connaissance de ce logiciel pour pouvoir suivre les instructions car celles-ci sont en texte sans illustration.

En lisant ce livre, j'admets avoir appris plusieurs nouvelles choses (la médiale était une mesure qui ne m'était pas connue. . .). J'ai bien aimé le changement de format. Au minimum, il permet de réfléchir à une autre manière de présenter les statistiques. Un bémol est le format compact de la présentation : les calculs sont intégrés dans les paragraphes qui sont souvent longs. Je ne suis pas convaincu que la mise en page permettra aux étudiants de mieux comprendre ; personnellement, cela ne m'a pas dérangé. Il aurait probablement fallu prendre quelques pages de plus pour soigner le côté esthétique. Par contre, le côté compact (le livre est en format de poche) et la couverture molle du livre m'ont permis de me promener un peu partout avec et de le lire dans le bus et les cafés du coin. En somme, j'ai bien aimé ce livre et je le recommande pour ceux qui enseignent MQ ou en sciences humaines. Bonne lecture!

Collectif de la SFdS (Société Française de Statistique), *Modèles statistiques pour données qualitatives*,

Éditions Technip, 2005, 292 p., ISBN 2-7108-0855-2, environ 92 \$.



Dans les mathématiques comme dans tout autre domaine, on retrouve des modes. En statistiques, ces temps-ci, on retrouve beaucoup de livres traitant de génétique, de simulations ou de marchés financiers (séries chronologiques, processus stochastiques, équations différentielles stochastiques). Alors la sortie de ce livre sur les données qualitatives m'a intrigué tout simplement par son non-conformisme à la mode. De plus, je me suis dit que ce type de variables est très couramment rencontré et que, justement de ce fait, elles sont négligées. Je voulais donc voir si je ne pouvais pas me ressourcer un peu et découvrir quelques nouvelles connaissances.

Ce livre est le neuvième dans une lignée de livres qui rendent compte de « Journées d'Études en Statistiques » organisées par la Société Française de Statistique avec le concours de la Société Mathématique de France. Puisque c'est un collectif, les chapitres ne se ressemblent pas forcément dans leurs styles, car chaque auteur en ayant un qui lui est propre. Ainsi, le premier chapitre, qui présente l'historique et les notions de base dans le domaine, ne contient à toutes fins pratiques pas d'exemple. Le côté historique est fort intéressant, mais le côté « liste de formule » paraît un peu stérile. En contraste, les chapitres 2 et 3, écrits par un autre auteur, sont plus vivants, car les concepts sont adéquatement illustrés par des exemples. Un bon point pour ce livre est la qualité du français, je dirais même la légèreté du français... et ce, pour tous les auteurs. J'ai aussi énormément aimé l'interprétation géométrique des connaissances statistiques, qui manquait complètement aux cours que j'ai suivis.

Le contenu du livre dépasse le niveau collégial. Il faut un minimum de connaissances statistiques « fraîchement » acquises pour comprendre le jargon à partir du chapitre 5 (probit, logit, gompit, ...). Par contre, je le recommanderais à quiconque s'embarque dans un cours de consultation statistique ou de laboratoire statistique à l'université. En effet, on retrouve des chapitres sur des sujets comme la régression logistique (simple, multiple, robuste et

PLS, ...) et la régression de Poisson (modèles de comptage avec application en assurance) comme sur les classes latentes, les méthodes de discrimination ainsi que sur les modèles log-linéaires et linéaires généralisés. Ainsi, c'est un bon livre de référence pour identifier rapidement un concept ou une approche à un problème donné. Par contre, pour avoir le côté « mécanique » ou « pratique », il faudra ensuite utiliser un autre livre même si les chapitres successifs deviennent de plus en plus appliqués. Par chance, le livre contient une bibliographie assez longue de douze pages, ce qui facilitera ce genre de perfectionnement. De plus, les références à la bibliographie sont nombreuses dans le texte, ce qui permet des explorations rapides en bibliothèque au moment opportun (où on ne comprend pas). Dans le même esprit, je crois que ce livre pourrait former un bon ajout dans les livres de références d'un consultant statistique ou d'un professeur d'université. En passant, pour les fiers Québécois, Trois-Rivières et Montréal sont mentionnées dans un exemple à la page 190.

Ce livre est quand même difficile à classer :

- 1) il expose de la théorie (des formules surtout) sans les démonstrations en général mais avec une « discussion » ;
- 2) il a des exemples mais qui sont en partie inventés de toute pièce ou tirés d'autres livres ;
- 3) il est appliqué, mais avec une approche « européenne » (rigoureuse), car il énonce systématiquement les conditions théoriques d'application (ce qui n'arrive pas toujours dans les livres appliqués américains) ;
- 4) il offre un survol intéressant de la littérature récente sur le sujet, en soulevant certaines polémiques et contradictions que l'on retrouve dans les « discussions » (par exemple p. 155 sur la robustesse de l'EMV en régression logistique par rapport aux explosions).

Avis à l'éditeur : en ajoutant plusieurs exemples et une série d'exercices à la fin de chaque chapitre, vous pourriez transformer rapidement ce livre de référence en manuel scolaire (n'oubliez pas une section de solutions à la fin). J'ai beaucoup apprécié la lecture de ce livre qui a éclairé quelques concepts que j'avais rencontrés au préalable. Il ne s'adresse malheureusement pas à tout le monde, mais pour ceux que j'ai mentionnés (étudiants en stats, professeurs d'université, consultants, bibliothécaires, personnes intéressées par les statistiques « actives », ...), ce livre est à considérer comme un bon achat. Bonne lecture !

Anders Hald, *History of probability and statistics and their applications before 1750*,

Wiley Interscience, 2003, 594 p., ISBN 0-471-47129-1, environ 92 \$.



Pour continuer ce tour d'horizon sur les statistiques, je voudrais aussi mentionner un autre livre que j'ai trouvé fort intéressant et que j'ai consulté à plusieurs reprises ces derniers mois. Par contre, je vais commencer ma recension avec un mea-culpa. J'avoue ne pas l'avoir lu au complet. Ce n'est pas parce qu'il est mal écrit ou inintéressant. Bien au contraire! C'est plutôt à cause de sa « taille imposante ». En effet, ce livre fait officiellement 594 pages (en fait, si on enlève les biographies et les pubs à la fin, on obtient à 548 pages). Mais les pages sont remplies et, de surcroît, la police est petite. . . Ainsi, le nombre de pages devrait être plus grand!

D'après les dires mêmes de l'auteur, ce livre offre quelques caractéristiques intéressantes. Les fils conducteurs sont les sujets statistiques et non les statisticiens. Ainsi, chaque chapitre est consacré à un problème, un théorème ou à un ensemble de concepts, et l'on découvre dans l'ordre les contributions des divers mathématiciens qui y ont travaillé. Il va sans dire que cela est plus intéressant pour le « praticien » que pour « l'historien ». Autre point intéressant, les contributions sont fidèles aux originaux des travaux, mais traduites dans les notations modernes. Ceci facilite grandement la lecture et rend ce livre utile aux mathématiciens et aux statisticiens en plus d'être utile aux historiens des mathématiques. Ça me rappelle mes lectures des *Éléments* d'Euclide en traduction littérale; la lecture était fascinante mais pas simplifiée par l'incompatibilité des écritures. C'est donc à mes yeux un plus pour ce livre.

Pour avoir lu les chapitres de ce livre qui ont attiré mon attention et ceux qui étaient pertinents à mes recherches de l'automne dernier, je peux dire qu'en tout et pour tout, j'en ai lu la moitié et feuilleté l'autre « en diagonale ». Ainsi, je pense avoir saisi ce livre à son plein potentiel en tant que mathématicien, professeur ou statisticien. Par contre, je ne peux en dire autant du côté historique. Même si le domaine m'intéresse grandement, je ne m'y suis pas spécialisé (j'ai par contre assisté à beaucoup de conférences et d'ateliers en

histoire des mathématiques, notamment pendant les congrès de la SMC). Donc, ce que je vais dire sera sous toute réserve et les historiens des mathématiques qui me lisent pourront me corriger s'ils le jugent nécessaire : je pense que ce livre est une bonne source historique également. Bonne lecture!

Mario Lefebvre, *Cours et exercices de statistique mathématique appliquée,*

**Presses Internationales Polytechniques, 2004, 512 p.,
ISBN 2-553-01139-3, environ 68 \$.**



Je n'ai pas utilisé ce livre dans un cours. Ainsi, ma recension ne se base pas sur une expérience directe avec ce livre dans une classe. Mais, je suis en train de faire une réflexion sur les statistiques depuis quelque temps. Je vous fais donc part de mes réflexions en tant que « probabiliste » qui aime les statistiques et qui enseigne au cégep des cours de statistiques et de méthodes quantitatives dans des programmes allant des sciences humaines à l'architecture et l'informatique.

Le livre commence avec 204 pages couvrant les bases des probabilités : de la définition empirique aux principales lois continues. Ensuite, on retrouve 276 pages consacrées aux statistiques : des statistiques descriptives au contrôle de qualité, en passant par la régression et les tests d'hypothèses. Les 30 dernières pages sont remplies de tables, de réponses aux questions paires et d'un index.

Le premier chapitre sert d'introduction globale, voir historique au sujet. Le livre continue alors sur la « vraie » introduction aux probabilités et à la statistique, soit la combinatoire. Par contre, les exercices se démarquent nettement des livres auxquels je suis habitué au cégep. On y retrouve la fiabilité de composantes électroniques, des balbutiements de contrôle de qualité et de la gestion environnementale en plus des traditionnels dés et jeux de cartes.

Dans le troisième chapitre, on aborde les variables aléatoires et leurs lois. Ce chapitre se démarque des autres livres de probabilité que j'ai lus par la complétude des lois énoncées (on

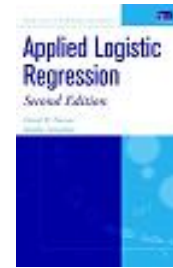
retrouve la Weibull, la bêta, la chi-deux, etc.). Par contre, on remarque aussi un problème de présentation puisque les sections ne se distinguent pas. En effet, sur une même page, on peut retrouver deux lois sans qu'il y ait de démarcation. . . D'ailleurs, je ne sais pas si c'est la police ou un autre élément, mais la finition ne m'accroche pas. Est-ce une question de taille de marché ? Un plus pour le livre est le style d'écriture fort lisible de l'auteur. L'auteur poursuit alors en statistique.

Chaque chapitre contient la théorie, puis une trentaine de questions résolues, une cinquantaine de questions dont seulement les réponses aux questions paires se trouvent à la fin, et une dizaine de questions à choix multiples. Je reviens donc encore sur le contenu des exercices : chimie, physique, électronique, gestion de l'environnement. . . Je n'ai probablement pas fouillé assez dans d'autres livres (attention pour d'autres chroniques), mais il me semble que les exercices sont très scientifiques par rapport à ce que je suis habitué à voir depuis quelques années (peut-être que j'enseigne trop de MQ ?).

Dans le collège où j'enseigne, un des cours d'intégration en sciences de la nature est un cours de statistiques. En lisant ce livre, l'idée m'est venue que, lorsque mon tour viendra pour l'enseigner, je le ferai peut-être avec celui-ci. Mais, le livre semble un peu trop fort (théorique) pour une utilisation en classe. Ceci dit, avec les réformes dans les programmes techniques au cégep, on retrouve des cours de statistique par-ci par-là (à Montmorency, il y en a en architecture et en génie civil maintenant). Même si le niveau est trop élevé pour ces cours, le fait qu'il ait été écrit pour des ingénieurs en fait un bon livre de référence pour les gens qui enseigneraient les statistiques.

Durant mon baccalauréat, on utilisait le livre de Ross, soit une traduction suisse d'un ouvrage américain. Il est peut-être bon de signaler qu'il existe maintenant des alternatives faites ici, au Québec : le livre « Théorie des Probabilités » de Reischer, Leblanc et Rémillard aux Presses de l'Université du Québec, et le livre de Mario Lefebvre qui fait l'objet de cette recension. Ce dernier pourra vous servir soit en classe, soit comme livre de référence. Bonne lecture !

**D. Hosmer et S. Lemeshow, *Applied logistic regression*,
Wiley, 2000, 380 p., ISBN 0-471-35632-8, environ 160 \$.**



En discutant avec des amis statisticiens, j'ai essayé de trouver des livres « incontournables » en statistiques. En voici un qui est à la fois « de base » et « avancé ». On n'enseigne pas ce contenu dans les cours de tronc commun. Il faut avoir persévéré un peu en statistiques pour le voir. Je vais donc essayer d'en vulgariser le contenu pour ensuite m'étendre sur la manière avec laquelle les auteurs le présentent.

La régression logistique permet d'estimer les facteurs de risque en effectuant une régression avec des variables dichotomiques du type :

$$I_{\text{propriété}}(x) = \begin{cases} 0 & \text{si le patient } x \text{ n'a pas la propriété} \\ 1 & \text{si le patient } x \text{ a pas la propriété.} \end{cases}$$

En présence de ce type de variable, la régression « normale » ne fonctionne pas, car on a deux nuages de points parfaitement horizontaux aux droites $y = 0$ et $y = 1$. En modélisant le lien entre les variables à l'aide d'une fonction logistique, on obtient assez naturellement ce que l'on appelle le « odds ratio » qui désigne la quantité de fois qu'il est plus probable d'avoir la maladie si on a la propriété. Par exemple si la variable étudiée est « avoir le cancer des poumons » et le facteur est « être fumeur », alors un odds ratio de deux indique qu'un fumeur a deux fois plus de chances d'avoir le cancer des poumons qu'un non-fumeur.

L'approche de ce livre contraste avec celui de la chronique précédente. En effet, l'autre auteur a inséré le mot « mathématique » entre statistique et appliquée pour ne pas confondre ce livre, comme il le disait lui-même dans sa préface, avec « un livre de statistique bourré de sorties de logiciels ». Ce qui est un peu le cas avec ce livre. Mais ce n'est pas un défaut. Ainsi, on retrouve quelques caractéristiques propres aux livres de statistiques appliquées :

- 1) les grandes lignes des raisonnements sont présentes, mais on ne voit pas les preuves au complet ;
- 2) il y a des sorties de logiciels comme SAS ou SPSS ;
- 3) les exemples sont sortis de « vrais problèmes », et ne sont donc pas « inventés » ;

- 4) il y a beaucoup plus de texte que de formules ;
- 5) chaque chapitre contient souvent un seul exemple étiré très longtemps ;
- 6) on insiste beaucoup sur l'interprétation des résultats obtenus et leur sens dans la vraie vie.

En plus, je dois dire que ce livre est très bien écrit et peut quasiment se lire comme un ouvrage de vulgarisation. Par contre, une chose qui m'a rebuté et déçu au plus haut point est la totale absence de réponse, de solution ou d'indice aux exercices. Pour les voir, il faut doubler la mise et acheter le solutionnaire qui vient dans un volume séparé. Vive le « marketing » !

En somme, ce livre n'est pas considéré sans raison comme la référence « appliquée » dans ce domaine. Il est cité à plusieurs reprises dans le livre français recensé plus haut (Méthodes statistiques pour variables qualitatives) dans les chapitres au milieu du livre. Il est bien écrit et servirait probablement à être inclus dans la bibliothèque d'un praticien ou d'un étudiant en statistiques. Pour tous ceux qui veulent se perfectionner sans trop souffrir (hormis votre portefeuille, car il vous faudra alors le solutionnaire), ce livre s'adresse à vous. Pour le cégep, on peut s'en inspirer en intégration peut-être pour de futures biologistes, pharmaciens ou médecins... Mais ils ne devront pas avoir peur des mathématiques et il faut avoir accès à un logiciel de statistiques plus performant qu'EXCEL. Bonne lecture !

À venir :

En français : Promenades mathématiques, L'empire des nombres, Les mathématiques, Tests de logique, etc.

En anglais : The Mathematical Traveler, 200% of nothing, The hidden unity in nature's laws, etc.

Robert Bilinski
Collège Montmorency
rbilinski@gmail.com

Vous venez de lire un ouvrage qui vous a passionné ? Ou qui vous a choqué ? Nous attendons vos commentaires : un bref texte que vous postez à Robert Bilinski, Dép. de Maths, 475, boul. de L'avenir, Laval (Québec), H7N 5H9. Vous pouvez aussi utiliser le courrier électronique (rbilinski@cmontmorency.qc.ca).