IREM DE BASSE-NORMANDIE

UNIVERSITÉ DE CAEN
BASSE - NORMANDIE

IREM DE BASSE-NORMANDIE
CAMPUS II - SCIENCES 3 - BP.5186
Boulevard Maréchal Juin, 14032 - CAEN Cedex
Tél. : **02 31 56 74 02** - Fax. : **02 31 56 74 90**
Adresse électronique : irem@math.unicaen.fr
Site web - http ://www.math.unicaen.fr/irem/

LE MIROIR DES MATHS

NUMÉRO UN : Mars 2008

Sommaire

Éditorial : l'IREM, son miroir et son île.

L'IREM de Basse-Normandie, Institut de recherche sur l'enseignement des mathématiques de l'Université de Caen Basse-Normandie, est heureux de vous présenter le premier numéro du *Miroir des maths*, son nouveau magazine. Vous y trouverez des articles, des chroniques, des annonces : tout ce qui fait la vie de l'IREM. Lors de chaque parution, tous les professeurs de mathématiques de l'Académie de Caen seront informés grâce à l'alias géré par les inspecteurs pédagogiques régionaux, que nous remercions ici. La version pdf du *Miroir* pourra être visualisée et téléchargée sur le site de l'IREM ; une version papier existe aussi, principalement destinée aux établissements : dites-nous si vous souhaitez la recevoir. Ce premier numéro est un essai : peu à peu, le *Miroir* se musclera et embellira. Nous espérons qu'il donnera à de nouveaux collègues, quel que soit type d'établissement où ils enseignent, l'envie de rejoindre l'un de nos groupes : n'hésitez pas à nous contacter !

Dans le deuxième numéro, nous vous ferons part de nos projets d'avenir. Mais pour rendre compte des activités présentes de l'IREM, il m'a semblé intéressant de revenir sur notre week end de rentrée, qui s'est tenu cette année les 28 et 29 septembre 2007 sur l'île de Tatihou (Manche). Nous nous étions donné rendez-vous au port de Saint-Vaast-la-Hougue, d'où nous avons atteint l'île en bateau. Le lendemain, la marée avait découvert l'estran, et c'est en roulant entre les parcs à huîtres que le même bateau, étrange engin amphibie, nous a ramenés sur le continent. Nous avons eu le plaisir d'accueillir Françoise Guimier, directrice de l'IREM de Rennes, afin de mettre au point avec elle l'extension à la Bretagne de notre fameux rallye mathématique de Basse-Normandie, "dynamique et virtuel". Je rappelle au passage que la cinquième édition de ce rallye, destiné aux classes de Troisième et Seconde, aura lieu le vendredi 4 avril 2008.

Le programme des exposés fut d'une grande variété. Je vais essayer de le résumer en espérant que les orateurs reconnaîtront un peu ce qu'ils ont dit !

Jean-Pierre Le Goff a présenté le contenu et la méthode d'un des stages de formation continue qu'organisent les membres du "Cercle d'histoire des mathématiques". Dans ce stage, intitulé "De quelques problèmes originels", il est question de quadrature du... cercle, mais aussi de duplication du cube, de trisection de l'angle et de construction de l'heptagone régulier. Tous ces problèmes fameux sont impossibles à la règle et au compas, mais peuvent donner lieu à des solutions approchées. Le fascicule qui sert de support au stage rassemble des textes stimulants de Nicolas Chuquet, Léonard de Vinci, Albrecht Dürer, Charles de Bovelles, Rafaele Bombelli et Antoine Arnauld.

De méthodes de construction en géométrie, il fut aussi question dans l'exposé de Danielle Salles sur les pliages dans l'apprentissage de la géométrie. En suivant ses instructions, chacun a expérimenté, plus ou moins adroitement, comment la géométrie des pliages avec glissement (ou ajustement) rend possible et amusante la trisection de l'angle, impossible à la règle et au compas. A l'inverse, la géométrie des pliages simples (sans glissement) est moins puissante que la géométrie à la règle et au compas, mais suffit pour réaliser, par exemple, la bissection de l'angle et la construction du pentagone régulier.

Ruben Rodriguez, qui travaille aussi sur la géométrie des pliages simples, a montré comment on peut l'utiliser pour introduire la notion d'angle à l'école primaire. En complément, il a esquissé une étude comparative de l'enseignement des fractions et des décimaux, en s'appuyant en particulier sur des manuels hispano-américains.

Également sur l'apprentissage de l'arithmétique des fractions, Claudine Plourdeau a présenté une activité expérimentée dans une classe de Sixième. On part d'une bandelette de papier servant d'unité de longueur que la classe a baptisée le mouki (M). Après avoir réalisé, par pliage ou mesure, des bandelettes de longueur 1/2 M, 1/3 M, 1/4 M, 1/5 M, etc., on demande à l'élève de marquer divers segments, dont la longueur est une certaine fraction de Mouki, supérieure ou inférieure à 1. Il est ainsi amené à effectuer des décompositions additives de la fraction proposée, à remarquer l'égalité de certaines sommes, à choisir judicieusement une décomposition facile à construire.

Michel Levard a présenté la mise en œuvre informatique d'une intéressante activité géométrique pour le lycée : la création de fontes de caractères au moyen de courbes de Bézier. Si A, B, C sont trois points, la courbe de Bézier quadratique reliant A à C et contrôlée par B est un certain segment de parabole tracé dans le triangle ABC (donc un certain ensemble de barycentres de A, B et C). L'exposé nous a montré comment construire des courbes de Bézier quadratiques, comment les raccorder sans cassure, et finalement comment construire (sous le logiciel Géoplan) la lettre S majuscule (la cursive anglaise, avec sa belle boucle et ses volutes !)

Pendant les repas, Éric Ziad-Forest nous a présenté quelques-uns des tours de magie des cartes dans lesquels il excelle. La volonté acharnée de comprendre "le truc" était visible chez beaucoup d'Irémiens, et Éric a fini par livrer quelques secrets...

La visite de l'île est du plus haut intérêt et le temps nous a manqué pour l'approfondir. Le clou en est la fameuse tour Vauban, commencée en 1694 et due en fait à l'ingénieur de Combes ; on notera son intéressante forme géométrique : un tronc de cône de révolution de diamètre extérieur à la base 25 m, flanqué hors-œuvre d'une tourelle d'escalier de même forme coiffée d'une calotte hémisphérique. Voilà de quoi proposer de beaux calculs de volumes ! L'île abrita aussi à partir de 1889 le laboratoire maritime du Muséum d'histoire naturelle, dans l'ancien lazaret. Sa vie fut assez brève. Dès 1897, sa suppression était envisagée et les membres de la Faculté des sciences de Caen consultés sur l'intérêt qu'ils lui portaient (j'ai trouvé ceci aux Archives départementales du Calvados, série T 668) : s'ils avaient été plus volontaires, l'île de Tatihou serait peut-être aujourd'hui une bien agréable annexe de notre Université !

Pierre Ageron (directeur de l'IREM de Basse-Normandie)

Caen : les chiffres romains dans la ville.

Le système de numération des Romains reste assez bien connu de nos jours. Il ne s'agit pas d'un système positionnel comme notre système indo-arabe, arrivé en Europe au XII^{e} siècle, mais d'un système additif (comme le système grec), enrichi d'une dose de soustraction qui fait tout son charme. De nombreuses études ont été consacrées aux origines et aux développements de la numération romaine ; je renvoie par exemple à [3], et me borne ici à rappeler quelques principes (voir aussi l'article d'Eric Trotoux dans ce même numéro du *Miroir*). Pour représenter les nombres entiers de 1 à 4999, il suffit des sept chiffres I (1), V (5), X (10), L (50), C (100), D (500) et M (1000). Partant de la décomposition du nombre en puissances de dix, on écrit les chiffres nécessaires par ordre décroissant de la gauche vers la droite, de sorte que leur somme atteigne le nombre voulu. On évite quatre occurrences consécutives des symboles I, X ou C en plaçant une fois (et une seule) I devant V ou X, X devant L ou C, C devant D ou M ; de même, on s'interdit les séquences VIV, LXL et DCD qui doivent être respectivement remplacées par IX, XC et CM. Ces règles assez compliquées ne se sont standardisées qu'à l'époque moderne et n'ont pas toujours, loin s'en faut, été respectées.

Souhaitant évaluer la place tenue par les chiffres romains dans le paysage et le patrimoine de Caen, j'ai consacré un week-end à les traquer dans les rues et monuments de la ville. Ils se sont avérés beaucoup plus rares que je ne l'imaginais ! Signes d'un lien avec un lointain passé, d'une sorte d'éternité des choses, je m'attendais en particulier à les trouver sur beaucoup d'inscriptions commémoratives ou de plaques tombales. Or les exemples ne sont pas si nombreux ; en voici quelques-uns.

Dans le chœur de l'abbatiale Saint-Étienne, une plaque de marbre gravée en 1801 rappelle le souvenir de Guillaume, duc de Normandie et roi d'Angleterre « qui obiit anno MLXXXVII ». Dans ce cas, on peut penser que c'est la rédaction en latin de l'épitaphe qui a imposé le recours aux chiffres romains. D'ailleurs, à quelques pas de la précédente, la plaque bilingue en mémoire des Caennais qui ont trouvé refuge dans l'église lors de la Seconde Guerre mondiale ne mentionne-t-elle pas l'année MCMXLIV côté latin et 1944 côté français ? D'autres inscriptions latines (une plaque dans une chapelle de l'église Saint-Pierre, deux tombeaux d'hommes d'Église au cimetière Saint-Pierre datant du début du XIX^{e} siècle) semblent confirmer la règle qui veut que les mots latins impliquent les chiffres romains ; de même que, plus récente, la pierre en façade de l'église Saint-André donnant la date de sa de consécration : « anno domini MCMLXII die VIII novembris ». Règle qui n'a cependant rien de mécanique : revenant à Saint-Étienne, on constate que la plaque tombale de dom Jean de Baillehache, bien que rédigée en latin, donne la date de sa mort en chiffres arabes (1644). À l'inverse, on trouve des millésimes en chiffres romains dans des inscriptions françaises, par exemple sur le tombeau d'un autre ecclésiastique au cimetière Saint-Jean. Dans ce cas, les quantièmes restent le plus souvent (mais pas toujours) en chiffres arabes. Ainsi la première pierre de reconstruction de l'université de Caen, dans l'aile naguère dévolue aux sciences, nous informe qu'elle « a été scellée (...) le 13 novembre MCMXLVIII ».

Un usage particulier du chiffre X se remarque plusieurs fois au cimetière des Quatre-Nations : l'abréviation X^{bre} désigne le mois de décembre (dixième mois de l'année dans le calendrier julien en usage jusqu'en 1582). Parmi les très rares tombeaux de laïcs qui portent des millésimes en chiffres romains, on note surtout ceux d'Arcisse de Caumont (cimetière Saint-Jean) et de Guillaume-Stanislas Trebutien (cimetière des Quatre-Nations, voir la photo gauche fig. 1) : il s'agit ici d'érudits, de passionnés d'antiquités, pour lesquels le recours à la numération romaine est presque naturel. Au cimetière des Quatre-Nations, la tombe de Charles Bouffey, conseiller à la cour, porte aussi des chiffres romains (voir la photo droite fig. 1) : dans ce cas, ils sont plutôt me semble-t-il des marqueurs de notabilité. On note d'ailleurs que son année de naissance (1791) est graphiée de manière déviante : MDCCLXXXXI au lieu de MDCCXCXI.

Inspectons maintenant, aidés du petit ouvrage de Gilles Henry [2], les maisons datées de Caen. Surprise : c'est presque toujours en chiffres arabes qu'elles le sont, même quand la date est intégrée dans une inscription latine. Sur les cent soixante-six recensées dans [2], du XVI^{e} siècle au XX^{e} siècle, deux

FIG. 1 – Cimetière des Quatre-Nations

seulement semblent porter une date en chiffres romains. À la maison dite « à la Tête de mort » (46, rue Caponière), une pierre portant une inscription se trouve engagée dans une autre aile, ne laissant lire que les derniers chiffres d'une date : xxxij, soit certainement 1532. L'intéressant est ici l'usage des minuscules, en harmonie avec l'écriture gothique du reste de l'inscription. Cet usage est paradoxal si l'on sait que les chiffres romains n'étaient pas, à l'origine, des lettres de l'alphabet, mais des signes de comptage. Dans la même rue, au numéro 149, la façade porte une niche abritant une Vierge à l'enfant et sur laquelle est gravée l'inscription « anno MDCDIII » (voir photo fig. 2) : il s'agit à nouveau d'une graphie déviante, la séquence DCD étant interdite. Gilles Henry est resté perplexe devant cette date, risquant un peu plausible 1608 sans même envisager le beaucoup plus vraisemblable 1903 (dont la graphie normale est MCMIII) : cela illustre bien les méprises que peut engendrer la numération romaine. On comprend pourquoi on l'a le plus souvent évitée.

L'horlogerie s'est montrée plus friande de chiffres romains. Il est vrai qu'elle n'a pas à dépasser le nombre XII ! L'horloge caennaise la plus monumentale, réalisée en 1744, se trouve dans le transept nord de l'abbatiale Saint-Étienne : elle offre la rare particularité d'associer chiffres romains pour les heures et chiffres arabes pour les minutes. On remarque aussi que le quatre romain y est noté IIII et non IV : ceci est un usage constant, comme le confirment tout près de là l'horloge du cloître, celle de l'Hôtel de ville et celle du palais de justice. On remarque encore que les chiffres romains de ces horloges suivent l'orientation du rayon vecteur, de sorte que le VI apparaît la tête en bas ; au contraire, les chiffres arabes d'une horloge comme celle de l'ex-gare Saint-Martin sont disposés parallèlement les uns aux autres, pour éviter la confusion entre 6 et 9. Reste à évoquer le cas des cadrans solaires : si ceux du Vieux Saint-Sauveur sont en chiffres arabes, celui qui fut peint en 1780 sur le clocher neuf de Saint-Michel de Vaucelles est en chiffres romains, et arbore non pas un IIII, mais un vrai IV !

Un autre usage traditionnel des chiffres romains est la numérotation des chapitres des livres bibliques, les chiffres arabes étant réservés aux versets (un usage que les orientalistes ont étendu au Coran, ce qui suscite d'ailleurs de nos jours d'inattendues controverses). De nombreuses tombes du cimetière protestant reproduisent une citation de l'Ancien ou du Nouveau Testament, en indiquant la source sous une forme comme « Ésaïe XL, 6-7 ». Une halte dans les fonds patrimoniaux de la Bibliothèque de Caen révèle l'utilisation très fréquente des chiffres romains pour les dates d'impression des livres jusqu'au XVIII^e siècle inclus, le plus souvent avec les variantes archaïques des chiffres IↃ et CIↃ des chiffres D et M : j'ai par exemple consulté un livre annoté par l'érudit caennais Samuel Bochart et édité à Leyde en CIↃ IↃ C XXIX, soit 1629.

Pour le marquage des bois, les charpentiers du Moyen Âge ont développé une variante de la numération romaine adaptée à leurs besoins. Apparue en Normandie vers 1200, avec notamment les charpentes des cathédrales de Lisieux et de Bayeux ([1]), elle se diffusa au cours du XIII^e siècle dans

FIG. 2 – Vierge à l'enfant - 149, rue Caponière

la moitié nord de la France, avec de nombreuses variantes régionales. On en voit un bel exemple à Caen, dans la cour du 10, rue Froide (voir images fig. 3). Il y a là une étroite maison à pans de bois du XVI^{e} siècle, tout à fait exceptionnelle à Caen parce qu'entièrement en bois, comme celles du Pays d'Auge, alors que la plupart des maisons à colombages de Caen, subsistantes ou disparues, sont en réalité des maisons de pierre à façade en bois. On distingue nettement sur les poteaux de bois (verticaux) et les écharpes de décharge (obliques) une numérotation au ciseau, dont le rôle est facile à comprendre si l'on sait qu'avant d'être levée, chaque colombe était travaillée à plat au sol pour l'adapter exactement aux irrégularités des ses voisines : les numéros permettaient aux levageurs de respecter scrupuleusement le plan élaboré. Le choix des chiffres romains peut s'expliquer par leur géométrie plus rectiligne : il est plus aisé, par exemple, de graver dans le bois un V ou un X qu'un 5 ou un 10. Cependant des confusions pouvaient naître lorsqu'on présentait une pièce dans le mauvais sens : par exemple, un XI lu à l'envers devient un IX, inconvénient dont les conséquences pouvaient être graves. D'où l'idée assez remarquable d'admettre plusieurs graphies équivalentes pour un même nombre. Dans ce système, le nombre cinq (V) peut tout aussi bien être graphié la tête en bas. Dans un nombre de plusieurs chiffres, l'invariance par symétrie joue globalement ou localement, ce qui implique l'existence de quatre graphies pour le nombre six : outre la graphie habituelle VI sont admissibles son image par symétrie gauche-droite, son image par symétrie haut-bas et son image par la composée des deux. En somme, le groupe de Klein opère sur l'ensemble des graphies possibles ! Le prix à payer est l'absence de notation soustractive comme dans le vrai système romain, et l'introduction de graphies nouvelles, comme celle du nombre neuf .

Pour finir ce petit inventaire, il m'a semblé que les chiffres romains les plus fréquents et visibles dans les rues sont ceux qu'arborent les plaques émaillées des débits de boisson, indiquant, conformément à une loi de 1941, la nature de leur licence : il faut par exemple une licence II pour servir du cidre, une licence III pour du pommeau et une licence IV pour du calvados !

Quelles conclusions tirer de cette promenade romaine dans l'Athènes normande ? En ce qui concerne les petits nombres, disons jusqu'à 20 ou 30, les chiffres romains ont eu quelques usages spécifiques solidement enracinés jusqu'au milieu du XX^{e} siècle, sans doute en voie de disparition. Les grands nombres en revanche, et en particulier les millésimes, paraissent dès le XVI^{e} siècle extrêmement rares hors du contexte savant : casse-tête pour le commun des mortels, ils fonctionnent alors comme marqueurs

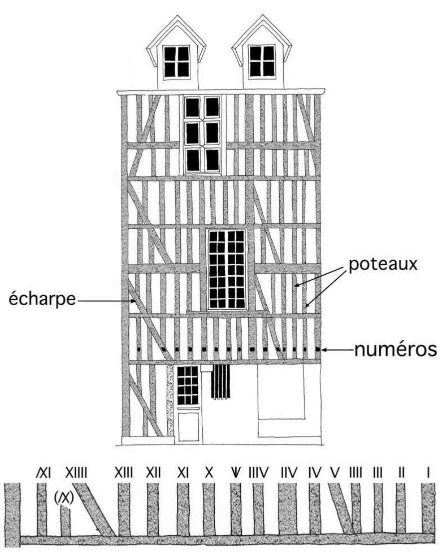


FIG. 3 – Maison à pans de bois - 10, rue Froide

d'une élite qui n'est elle-même pas toujours en mesure de les écrire correctement, et ils sont évités lorsqu'il importe d'être compris. Faut-il, dans ces conditions, continuer à les enseigner, dès la deuxième année du Cours élémentaire ? Oui parce que l'étude de la variété des systèmes de numération, disparus ou non, constitue une initiation à la diversité des civilisations. Oui encore, parce que cette variété démontre de manière frappante que les nombres en eux-mêmes sont des objets mathématiques indépendants de toute représentation graphique, mais aussi que les algorithmes qui permettent d'explorer l'univers mathématiques requièrent une réflexion attentive sur la manière la plus judicieuse de représenter ces objets. Oui enfin, parce que les clefs de lecture de notre patrimoine doivent être offertes à tous les citoyens, et non réservées à l'élite. Alors III fois oui aux chiffres romains !

texte : Pierre Ageron clichés et dessin : Patrice Gourbin

bibliographie.

[1] Frédéric Épaud, *De la charpente romane à la charpente gothique en Normandie*, Publications du CRAHM, Caen, 2007

[2] Gilles Henry, *Les belles dates du Caen jadis*, Charles Corlet, Condé-sur-Noireau, 2001

[3] Henry Plane, « La numération romaine », in : *Contribution à une approche historique de l'enseignement des mathématiques*, IREM des Pays de la Loire, Nantes, 1997, 177-185

Une approche globale des systèmes de numération.

L'article sur la découverte des chiffres romains dans la ville de Caen m'a amené à dépoussiérer un exposé préparé pour l'oral du CAPES de mathématiques (années 1975).
Commençons par citer un extrait des éléments d'histoire des mathématiques de N. Bourbaki (p.64).

> L'histoire et l'archéologie nous font connaître un grand nombre de systèmes de numération ; leur but initial est d'attacher à chaque entier individuel (jusqu'à une limite qui dépend des besoins de la pratique) un nom et une représentation écrite, formés de combinaisons d'un nombre restreint de signes, s'effectuant suivant des lois plus ou moins régulières. Le procédé de beaucoup le plus fréquent consiste à décomposer les entiers en sommes d' « unités successives »$u_1, u_2, u_n, \dots$ dont chacune est un multiple entier de la précédente ; et si en général $\dfrac{u_n}{u_{n-1}}$ est pris égal à un même nombre b (la « base »du système, le plus souvent 10), on observe mainte exception à cette règle comme chez les Babyloniens, où $\dfrac{u_n}{u_{n-1}}$ est tantôt égal à 10, tantôt à 6, dans le système chronologique des Mayas, où $\dfrac{u_n}{u_{n-1}}$ est égal à 20 sauf pour $n = 2$, et où $\dfrac{u_2}{u_1} = 18$, et dans le système romain où $\dfrac{u_n}{u_{n-1}}$ est tantôt égal à 5, tantôt à 2. Quant à l'écriture correspondante, elle doit indiquer le nombre d' « unités »u_i de chaque ordre i ; dans beaucoup de systèmes (comme chez les Égyptiens, les Grecs et les Romains), les multiples successifs $k.u_i$, (où k varie de 1 à $\dfrac{u_{i+1}}{u_i} - 1$) sont désignés par des symboles qui dépendent à la fois de k et de i. Un premier et important progrès consiste à désigner tous les nombres $k.u_i$ (pour la même valeur de k) par le même signe : c'est le principe de la « numération de position », où l'indice i est indiqué par le fait que le symbole représentant $k.u_i$, apparaît à la i-ème place dans la succession des « tranches »constituant le nombre représenté. Le premier système de cette nature se rencontre chez les Babyloniens, qui, sans doute dès 2 000 avant J.-C., notent par un même signe tous les multiples $k.60^{\pm i}$ correspondant à des valeurs quelconques de l'exposant i .

La présentation du système b-adique (qui généralise la numération décimale de position et où b est un entier supérieur ou égal à 2) se trouve fréquemment détaillée dans les ouvrages de mathématiques élémentaires. Nous nous proposons d'examiner ici une présentation moins courante, qui généralise le système b-adique en formalisant les situations « uni et multi-base » évoquées dans l'extrait précédent. Sautons donc joyeusement dans le formalisme « bourbakien »

Représentation générale d'un entier

On se donne une suite d'entiers **non nuls** $(b_i)_{i \in I}$ avec soit $I = [\![1, l]\!]$ où $l \in \mathbb{N}^*$(cas fini), soit $I = \mathbb{N}^*$(cas infini). Pour chaque $i \in I$, on considère un sous-ensemble d'entiers S_i fixé et constitué de b_i représentants pour la congruence modulo b_i. S_i peut être $(0, 1, ..., b_i - 1)$ mais aussi tout autre système de représentants, même négatifs.
soit m un entier rationnel.

On appelle développement de m dans le système $(S_i, b_i)_{i \in I}$ toute suite finie d'entiers $(a_j)_{0 \leq j \leq k}$ $(a_0, a_1, \dots, a_k)$ telle que

- $0 \leq k \leq l$ dans le cas fini, $k \geq 0$ sinon
- Si $0 \leq j \leq k - 1$ alors $a_j \in S_{j+1}$ et si $k + 1 \in I$ alors $a_k \in S_{k+1}$ (a_k est un entier quelconque lorsque S_{k+1} n'est pas défini)
- $k \neq 0$ entraîne $a_k \neq 0$
- $m = a_0 + a_1 b_1 + a_2 b_1 b_2 + a_3 b_1 b_2 b_3 + \cdots + a_k b_1 b_2 \cdots b_k$

Exemples

Système de « Cro-Magnon » La suite $(b_i)_{i \in \mathbb{N}^*}$ est la suite constante valant 1. $S_1 = \{0\}$ et $S_i = \{1\}$ pour $i \geq 2$. Tout entier m positif admet le développement (0,1,1,1,. . ., 1) ; c'est une liste de longueur $m + 1$ et $k = m$.

Système « horaire » I est fini, $l = 3$. Nous fixons les valeurs suivantes : $b_1 = b_2 = 60$, $b_3 = 24$, $S_1 = S_2 = \{r \in \mathbb{Z}/0 \leq r < 60\}$ et et $S_3 = \{r \in \mathbb{Z}/0 \leq r < 24\}$. Toute durée exprimée en secondes peut s'exprimer en jours, heures, minutes et secondes. Par exemple m=4617087 se décompose selon le développement(27,31,10,53) soit 53 jours, 10 heures 31 minutes et 27 secondes. Ici $k = 3$ et $m = \mathbf{27} + \mathbf{31} \times 60 + \mathbf{10} \times 60 \times 60 + \mathbf{53} \times 60 \times 60 \times 24$. Comme S_4 n'est pas défini $a_3 = 53$ n'a pas de contrainte à satisfaire.

Système « décimal » La suite $(b_i)_{i\in\mathbb{N}^*}$ est la suite constante valant 10. $S_i = \{r \in \mathbb{Z}/0 \leq r < 10\} = \{0, 1, 2, \ldots, 8, 9\}$ pour tout $i \in \mathbb{N}^*$. $m = \overline{12379}^{10} = \mathbf{9} + \mathbf{7} \times 10 + \mathbf{3} \times 10^2 + \mathbf{2} \times 10^3 + \mathbf{1} \times 10^4$.

Système « Romain limité » I est fini, $l = 6$. On choisit la suite $(b_i)_{1\leq i\leq 6}$ telle que $b_i = 5$ pour i impair et $b_i = 2$ pour i pair. $S_i = \{-1, 0, 1, 2, 3\}$ pour i impair et $S_i = \{0, 1\}$ pour i pair.
$m = 3773 = \underline{3}+\underline{0}\times5+\underline{2}\times5\times2+\underline{1}\times5\times2\times5+\underline{2}\times5\times2\times5\times2+\underline{1}\times5\times2\times5\times2\times5+\underline{3}\times5\times2\times5\times2\times5\times2$
Le développement de m est (3,0,2,1,2,1,3). $k = 6$ et $a_6 = 3$.
En utilisant les sept symboles (I,V,X,L,C,D,M) pour chaque tranche d'unités on en déduit l'écriture de type additif $m =$ **MMMDCCLXXIII**. $a_1 = 0$ se traduit par l'absence du symbole V dans l'écriture. Pour $m = 14$ nous avons $14 = -\mathbf{1} + \mathbf{1} \times 5 + \mathbf{1} \times 5 \times 2$ qui donne $m =$ **XIV** avec la convention d'écriture soustractive classique. Ici $a_0 = -1, a_1 = 1, a_2 = 1$. Pour $m = 99$ le développement est donné par (-1,0,0,0,1) qui produit l'écriture $m =$**IC**. Cette représentation ne respecte pas la convention « pas de soustraction d'une lettre plus de dix fois supérieure » et on écrit plutôt $m =$**XCIX** qui ne cadre pas avec notre type de décomposition (l'entité X est à la fois ajoutée et retranchée). Ces règles sont restées assez floues jusqu'au Moyen-Âge, et pas toujours appliquées d'ailleurs.

Système « Romain généralisé » I est $\mathbb{N}^*$. On choisit la suite $(b_i)_{i\in I}$ telle que $b_i = 5$ pour i impair et $b_i = 2$ pour i pair. $S_i = \{-1, 0, 1, 2, 3\}$ pour i impair et $S_i = \{0, 1\}$ pour i pair.
Ce système ne présente pas d'intérêt pratique car avec le principe additif de l'écriture symbolique, il nous faudrait un ensemble infini de symboles pour écrire tous les entiers.

Examinons maintenant l'existence et l'unicité de la décomposition des entiers dans ces systèmes $(S_i, b_i)_{i\in I}$.

Théorème 1

Tout entier m a un développement unique dans tout système $(S_i, b_i)_{i\in I}$ fini, $I = [\![1, l]\!]$.

Une remarque préliminaire : Soit $(a_0, a_1, \ldots, a_k)$ un développement de m dans $(S_i, b_i)_{i\in I}$. Nous avons $a_0 \equiv m[b_1]$ et $m - a_0 = b_1 m_1$. Nous en déduisons $m_1 = a_1 + a_2 b_2 + a_3 b_2 b_3 + \cdots + a_k b_2 \cdots b_k$ où les a_i satisfont aux conditions d'un développement dans le système $(S_i, b_i)_{i\in I\setminus 1}$; $(a_1, a_2, \ldots, a_k)$ est un développement de m_1 dans $(S_i, b_i)_{2\leq i\leq l}$.
Réciproquement, soit m un entier et $a_0 \in S_1$ tel que $m \equiv a_0[b_1]$. Posons $m_1 = \frac{m-a_0}{b_1}$. Si $(a_1, a_2, \ldots, a_k)$ est un développement de m_1 dans $(S_i, b_i)_{2\leq i\leq l}$ alors $(a_0, a_1, \ldots, a_k)$ est un développement de m dans $(S_i, b_i)_{i\in I}$.

Nous prouvons alors par récurrence sur la longueur l du système $(S_i, b_i)_{i\in I}$ que tout entier a un développement unique dans tout système de longueur l.

- Pour $l = 1$, $m = a_0 + a_1 b_1$ où a_0 est le représentant de S_1 congru à m modulo b_1 et $a_1 = \frac{m-a_0}{b_1}$. D'où l'existence et l'unicité de la décomposition de tout entier dans ce cas.
- D'après la remarque préliminaire, l'existence de la décomposition est une propriété héréditaire à partir de $l = 1$. Montrons que l'unicité de la décomposition l'est aussi :
  Nous posons $k = l + 1$ et envisageons deux décompositions.
  $m = a_0 + a_1 b_1 + a_2 b_1 b_2 + a_3 b_1 b_2 b_3 + \cdots + a_k b_1 b_2 \cdots b_k$ et
  $m = \alpha_0 + \alpha_1 b_1 + \alpha_2 b_1 b_2 + \alpha_3 b_1 b_2 b_3 + \cdots + \alpha_k b_1 b_2 \cdots b_k$.
  Ces décompositions entraînent $a_0 - \alpha_0 \equiv 0[b_1]$ d'où $a_0 = \alpha_0$ puisque a_0 et α_0 sont dans S_1. Nous avons alors $a_1 + a_2 b_2 + a_3 b_2 b_3 + \cdots + a_k b_2 \cdots b_k = \alpha_1 + \alpha_2 b_2 + \alpha_3 b_2 b_3 + \cdots + \alpha_k b_2 \cdots b_k$ et l'unicité pour la longueur l (hypothèse de récurrence) entraîne l'égalité $a_i = \alpha_i$ pour tout $i \in [0, l + 1]$.

Nous concluons selon le principe de récurrence que tout entier m a une décomposition unique pour tout système fini de longueur l, quel que soit l'entier l non nul.

Abordons maintenant le cas infini où $I = \mathbb{N}^*$.

Théorème 2

Avec les hypothèses plus restrictives suivantes :
Pour tout i de $\mathbb{N}^*$, $\quad b_i > 1$ et $S_i \subset \{r \in \mathbb{Z} / -b_i < 2r < 2b_i\}$
nous pouvons affirmer que tout entier m strictement positif admet un développement unique dans le système $(S_i)_{i\in I}$. C'est aussi le développement dans le système fini $(S_i, b_i)_{1\leq i\leq m}$. De plus le développement de $m = 0$ ne peut être que $(0)(k = 0$ et $a_0 = 0)$. Celui-ci existe ssi $0 \in S_1$.

Preuve

Étape 1 Pour $m \geq 0$, nous prouvons que dans tout système fini extrait de $(S_i)_{i\in I}$ en se limitant à $1 \leq i \leq l$ $(l \geq 1)$, le dernier élément du développement a_k vérifie $a_k \geq 0$.
$m = a_0 + a_1b_1 + a_2b_1b_2 + a_3b_1b_2b_3 + \cdots + a_kb_1b_2\cdots b_k$. Or pour $i < k$, $a_i \in S_{i+1}$ entraîne $a_i < b_{i+1}$ ou $a_i \leq b_{i+1} - 1$. Il vient alors par majoration
$m \leq (b_1 - 1) + (b_2 - 1)b_1 + (b_3 - 1)b_1b_2 + (b_4 - 1)b_1b_2b_3 + \cdots + (b_{k-1} - 1)b_1b_2\cdots b_{k-1} + a_kb_1b_2\cdots b_k$
et en développant et simplifiant $m \leq -1 + b_1b_2\cdots b_k + a_kb_1b_2\cdots b_k$, d'où $m \leq -1 + (a_k + 1)(b_1b_2\cdots b_k)$.
Finalement $m < (a_k + 1)(b_1b_2\cdots b_k)$ ce qui pour $m \geq 0$ permet de conclure $a_k \geq 0$.
Si $a_k = 0$, on a $k = 0$ et $m = a_0 = a_k = 0$. Donc ce cas ne présente pas lorsque $m > 0$ ou $k \neq m$.

Étape 2 Nous prouvons que $a_k > 0$ entraîne $m > k$.
$2m = 2a_0 + 2a_1b_1 + 2a_2b_1b_2 + 2a_3b_1b_2b_3 + \cdots + 2a_kb_1b_2\cdots b_k$. Pour $0 \leq i < k \quad 2a_i > -b_{i+1}$, ce qui équivaut à $2a_i \geq -b_{i+1} + 1$. Nous déduisons que
$2m \geq 1 - b_1 + (1 - b_2)b_1 + (1 - b_3)b_1b_2 + (1 - b_4)b_1b_2b_3 + \cdots + (1 - b_k)b_1b_2\cdots b_{k-1} + 2a_kb_1b_2\cdots b_k$. En simplifiant après calculs, il nous reste $2m \geq 1 + (2a_k - 1)b_1b_2\cdots b_k$. Comme $a_k > 0$, $2a_k - 1 \geq 1$ d'où $2m \geq 1 + b_1b_2\cdots b_k > b_1b_2\cdots b_k$. Sachant que $b_i \geq 2$ nous en tirons $2m > 2^k \geq 2k$ ce qui implique $m > k$. Cela montre que le développement de m est alors le même dans le système $(S_i)_{1\leq i\leq m}$.

Étape 3 Pour $m > 0$ choisissons $l = m$. m admet un développement dans le système fini (théorème 1) extrait de $(S_i)_{1\leq i\leq m}$ et nous avons $k < m$. Cela fournit donc aussi ($a_k \in S_{k+1}$ car $k + 1 \leq m$) un développement dans le système infini $(S_i)_{i\in I}$. Ainsi m possède un développement.

L'unicité découle aussi du théorème 1 car deux développements de m dans le système infini $(S_i)_{i\in I}$ sont aussi des développements dans des systèmes finis extraits donc aussi dans $(S_i)_{1\leq i\leq m}$ dans lequel ils sont nécessairement identiques.

Pour $m = 0$, nous avons nécessairement $a_k = 0$ pour tout développement, puisque $a_k > 0$ entraîne $m > k \geq 0$. Dès lors $m = a_0 = 0$. Ce développement existe ssi $0 \in S_1$.

Corollaires

les hypothèses du théorème 2 étant remplies :

Système b-adique $(b \geq 2)$ Tout entier $m \geq 0$ a un développement unique :
$m = a_0 + a_1b + a_2b^2 + \cdots + a_kb^k$. Les a_i sont dans $S = \{r \in \mathbb{Z}/0 \leq r < b\}$.

Système romain généralisé Tout entier $m > 0$ a un développement unique :
$m = a_0 + a_1 \times 5 + a_2 \times 10 + a_3 \times 50 + \cdots + a_k \times u_k$ où $u_k = 10^{k/2}$ si k pair et $u_k = 5 \times 10^{(k-1)/2}$ si k est impair. Les (a_i) sont dans $S = \{-1, 0, 1, 2, 3\}$ pour i pair et dans $S = \{0, 1\}$ pour i impair.

Éric Trotoux

Quartiles : programme officiel, calculatrice, tableur... ou Wikipedia

0. Introduction.

Aujourd'hui la notion de quantile est passée dans le langage courant, même si le mot n'est pas prononcé.

La définition en semble simple a priori :

Définition : Soit p un nombre réel appartenant à $]0,\ 1[$. Soit une série statistique de données réelles que l'on supposera pour simplifier ordonnées par ordre croissant : $(x_1, x_2, ..., x_n)$.

Le nombre réel q_p est un quantile d'ordre p si $100p\%$ des données sont inférieures ou égales à q_p et $100(1-p)\%$ des données sont supérieures à q_p.

Certains quantiles sont connus sous d'autres noms.

- $Me = q_{0.50}$ = quantile d'ordre 1/2 = médiane.
- $Q_1 = q_{\frac{1}{4}}$ = quantile d'ordre 1/4 = premier quartile
- $Q_3 = q_{\frac{3}{4}}$ = quantile d'ordre 3/4 = troisième quartile
- $D_1 = q_{\frac{1}{10}}$ = quantile d'ordre 1/10 = premier décile

Trois exemples issus pour les deux premiers de revues économiques et le dernier d'un site médical.

- Exemple 1 : En 2003, le salaire médian, à temps complet était de 1849 euros c.-à.-d. la moitié des salariés à temps complet gagnait moins de 1849 euros et l'autre moitié plus de 1849 euros. Alors $1849 = Me =$ médiane.
- Exemple 2 : En 2003, les 10% des ménages les plus pauvres gagnaient moins de 609 euros par mois. Alors $609 = D_1 =$ premier décile.
- Exemple 3 : Dans une étude sur 342 secteurs de gardes de la zone Vénissieux et Cours-la Ville, le quart de ces secteurs concernaient une population de moins de 34.41 h/km2 et le quart de ces secteurs une population de plus de 203.76 h/km2. Alors $34.41 = Q_1 =$ premier quartile et $203.76 = Q_3 =$ troisième quartile.

Il paraît important, à partir de données, de pouvoir faire calculer effectivement aux élèves certains de ces quantiles, et en particulier les quartiles qui sont utilisés pour établir une mesure de la dispersion des données - l'écart-interquartile - et dans l'établissement des « boîtes à moustaches », dites encore boîtes de Tukey, (box-plot en anglais).

La difficulté rencontrée est la suivante : à partir d'un même jeu de données, suivant la méthode utilisée, on peut aboutir à des résultats différents.

Dans une **première étape**, nous illustrerons ceci à partir des quatre jeux de données, et de trois méthodes les plus utilisées : la méthode dite "du programme officiel", la méthode utilisant les calculatrices, la méthode utilisant les tableurs classiques et les logiciels plus sophistiqués comme R.

Dans une **deuxième étape**, nous rappellerons les définitions mathématiques des concepts utilisés. Dans une **troisième étape**, nous tenterons d'expliquer ce qui "justifie" l'utilisation de chacune de ces méthodes. Dans une **quatrième étape**, nous amorcerons un débat sur les éventuelles conséquences sur les élèves, de la diversité de ces méthodes. Enfin nous irons voir ce que dit Wikipedia.

Ces réflexions sont inspirées de l'article Quartiles, déciles et tutti quantiles (*Jean-Claude Girard*) tiré de la brochure APMEP n°156 produite par la Commission Inter-IREM *Statistique et Probabilités*, et tentent de compléter cet excellent article.

1. Quatre jeux de données et trois méthodes.

Durée de vie d'ampoules de 100W

Jeu 1 : 6 ampoules : $46, 270, 293, 382, 630, 952$

Jeu 2 : 7 ampoules :$49, 90, 198, 302, 387, 547, 763$

Jeu 3 : 8 ampoules : $34, 47, 71, 263, 282, 622, 667, 968$

Jeu 4 : 9 ampoules : $39, 174, 196, 252, 331, 401, 456, 637, 944$

Calcul des quartiles

1 er quartile

	Programme	Calculatrice	Tableur
Jeu 1	270	270	275.75
Jeu 2	90	90	144
Jeu 3	47	59	65
Jeu 4	196	185	196

3ème quartile

	Programme	Calculatrice	Tableur
Jeu 1	630	630	568
Jeu 2	547	547	467
Jeu 3	622	644.5	633.25
Jeu 4	456	546.5	456

1. Définitions mathématiques des concepts utilisés.
Le cas d'une variable aléatoire "continue" :

Soit X variable aléatoire réelle qui suit une loi de densité f non nulle sur I.
Soit F la fonction de répartition définie par :

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$$

Soit $p \in]0, 1|$. **Le quantile d'ordre p est le réel q_p appartenant à I tel que $F(q_p) = p$.**
On écrit aussi $q_p = F^{-1}(p)$.

Un exemple :

Une ampoule est vendue avec l'indication : durée moyenne = 1 an = 365 jours.

X est la variable aléatoire qui mesure (en jours) la durée de vie de l'ampoule dont la loi est modélisée par la densité exponentielle

$f(x) = \lambda e^{-\lambda x}$ si $x \geq 0$ et $f(x) = 0$ si $x < 0$.

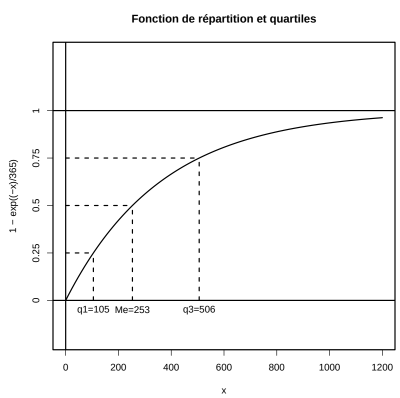
Puisque moyenne $= E(X) = \frac{1}{\lambda}$, $\lambda = \frac{1}{365}$ et $F(x) = 1 - e^{-\lambda x} = 1 - e^{-\frac{x}{365}}$ si $x \geq 0$ et $F(x) = 0$ si $x < 0$.

q_p vérifie $1 - e^{-\lambda q_p} = p$ donc $q_p = -\dfrac{\ln(1-p)}{\lambda} = -365\ln(1-p)$

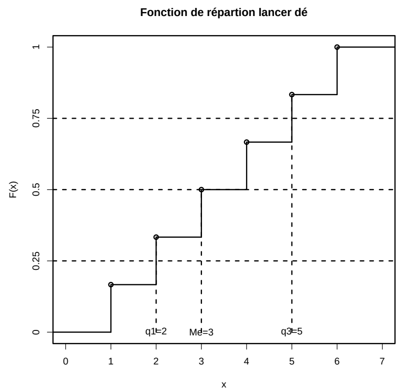
$Me = q_{0.50} = -365\ln(1 - 0.50) = 253$

$Q_1 = q_{0.25} = -365\ln(1 - 0.25) = 105$

$Q_3 = q_{0.75} = -365\ln(1 - 0.75) = 506$



Le cas d' une variable discrète à partir d'un exemple
X =numéro tiré après tirage d'un dé équilibré
$P(X = i) = \frac{1}{6}$ pour $i = 1, 2, 3, 4, 5, 6$.
$F(x) = P(X \leq x) = 0$ si $x < 1$
$F(x) = P(X \leq x) = \frac{1}{6}$ si $1 \leq x < 2$
$F(x) = P(X \leq x) = \frac{2}{6}$ si $2 \leq x < 3$
....
$F(x) = P(X \leq x) = \frac{5}{6}$ si $5 \leq x < 6$
$F(x) = P(X \leq x) = 1$ si $x \geq 6$
Définition :
$q_p = \min\{x \;\; /F(x) \geq p\}$

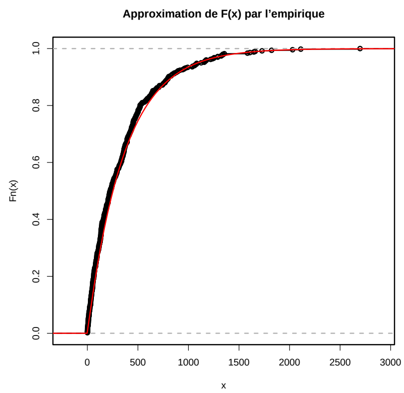


Remarque :
Pour une variable continue, on peut appliquer la même définition car
$\forall x \geq q_p \quad F(x) \geq F(q_p) = p$
et de plus F est continue au point q_p d'où la définition générale :

$$q_p = \min\{x \;\; /F(x) \geq p\}$$

Fonction de répartition empirique
$x_1, x_2, ..., x_n$ suite de données classées par ordre croissant. La fonction de répartition empirique $\hat{F}$ est définie ainsi :
$\hat{F}(x) = \frac{1}{n}Card\{i \;\; / \; x_i \leq x\}$ = fréquence des données inférieures ou égales à x.
La fonction de répartition empirique construite à partir d'un ensemble de n données est une bonne approximation de F, fonction de répartition de la variable aléatoire qui a engendré les données.

Exemple :
Simulation de 500 durées de vie pour des ampoules de durée moyenne 1 an



2. Justification du calcul pour chacune des trois méthodes :

2.1. Méthode du programme officiel :

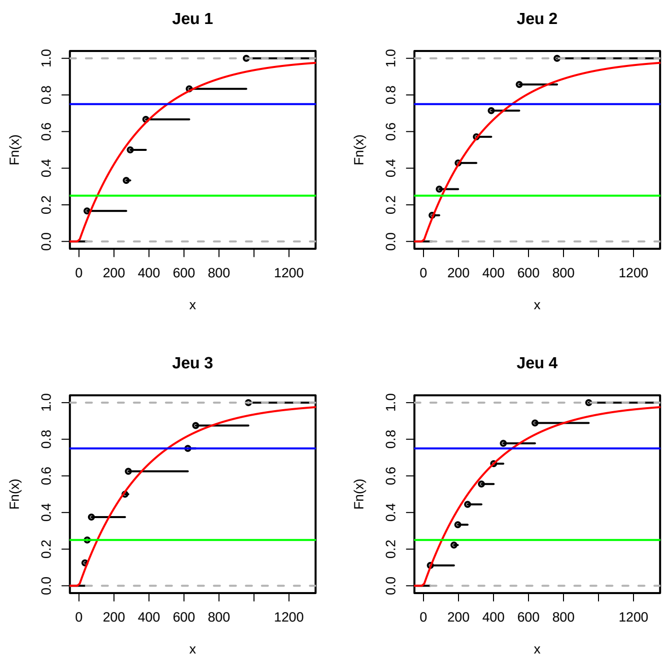
Nouveaux programmes de première :

Le premier (respectivement le troisième) quartile est le plus petit élément q_1 (respectivement q_3) des valeurs de la série ordonnée par ordre croissant, tel qu'au moins 25%(respectivement 75%) de ces valeurs soient inférieures ou égales à q_1 (respectivement q_3).

Jeu 1 : $46, 270, 293, 382, 630, 952$
$q_1 = 46$? Il y a 1 valeur inférieure ou égale à 46 soit $\frac{1}{6} = 16,66\%$
$q_1 = 270$? Il y a 2 valeurs inférieures ou égales à 270 soit $\frac{2}{6} = 33,33\%$
Donc $q_1 = 270$

Jeu 3 : $34, 47, 71, 263, 282, 622, 667, 968$
$q_1 = 34$? Il y a 1 valeur inférieure ou égale à 34 soit $\frac{1}{8} = 12,5\%$
$q_1 = 47$? Il y a 2 valeurs inférieures ou égale à 47 soit $\frac{2}{8} = 25\%$
Donc $q_1 = 47$

Fonction de répartition empirique pour les quatre jeux de données :



Un constat :

$\hat{q}_{0.25}$ = 1er quartile calculé suivant programme. Alors : $\hat{q}_{0.25} = \min\{x \ \ /\hat{F}(x) \geq 0.25\}$
$\hat{q}_{0.75}$ = 3ème quartile calculé suivant programme. Alors : $\hat{q}_{0.75} = \min\{x \ \ /\hat{F}(x) \geq 0.75\}$

La justification mathématique de cette méthode parait évidente. Puisque la fonction de répartition empirique $\hat{F}$ est une bonne approximation de la fonction de répartition F, pour obtenir $\hat{q}_{0.25}$ (resp. $\hat{q}_{0.75}$), estimateur du quantile $q_{0.25}$(resp. $q_{0.75}$), il suffit de remplacer dans l'expression générale d'un quantile d'ordre p, à savoir $q_p = \min\{x \ \ /F(x) \geq p\}$, la fonction de répartition F par son estimation $\hat{F}$ et ainsi on retombe sur la définition du programme officiel.

Remarque importante : avec cette méthode très générale puisqu'elle s'applique quelle que soit la valeur de $p \in]0, 1[$, on devrait pouvoir définir la médiane Me comme le plus petit élément des valeurs de la série ordonnée par ordre croissant, tel qu'au moins 50% de ces valeurs soient inférieures ou égales à Me. Or dans le programme officiel, il n'en est rien et c'est une autre définition qui est proposée (voir méthode de la calculatrice).

3.2. Méthode de la calculatrice :

Rappel : *« On ordonne la série des observations par ordre croissant ; si la série est de taille* $2n+1$*, la médiane est la valeur du rang* $n+1$ *dans cette série ordonnée ; si la série est de taille* $2n$ *la médiane est la moyenne des valeurs des termes de rang* n *et* $n+1$ *dans cette série ordonnée »* (Document d'accompagnement des programmes de première).

La médiane détermine deux sous-séries de même taille constituées des observations antérieures.

. sous-série 1 = observations $\leq$ médiane

. sous-série 2 = observations $\geq$ médiane

Si taille de l'échantillon $= 2n+1$, on exclut dans chaque sous-série une et une seule donnée : la donnée centrale de la série initiale (c.-à-d. la médiane), sinon on n'exclut rien.

1er quartile = médiane de la nouvelle sous-série 1.

3ème quartile = médiane de la nouvelle sous-série 2.

Exemples :

Jeu 1 : $46, 270, 293, 382, 630, 952$

Taille $= 2n = 2 \times 3$. Médiane $= = \dfrac{x_3 + x_4}{2} = \dfrac{293+382}{2} = 337.5$

Sous-série 1 : $46, 270, 293$. 1er quartile $= 270$

Sous-série 2 : $82, 630, 952$. 3ème quartile $= 630$

Jeu 4 : : $39, 174, 196, 252, 331, 401, 456, 637, 944$

Taille $= 2n+1 = 2 \times 4 + 1$. Médiane $= = x_5 = 331$

. sous-série 1 $= 39, 174, 196, 252, 331$

. sous-série 2 $= 331, 401, 456, 637, 944$

Taille impaire :

. nouvelle sous-série 1$=$ $39, 174, 196, 252$. 1er quartile $= \dfrac{174+196}{2} = 185$

. nouvelle sous-série 2 $= 401, 456, 637, 944$. 3ème quartile $= \dfrac{456+637}{2} = 546.5$

La justification apparaît ici aussi évidente. Après avoir ordonné les données, il suffit de les "couper" en quatre parts. La médiane sert à "couper" la série ordonnée d'abord en deux parties. Pour obtenir le premier quartile (resp. de troisième quartile), il suffit de "couper" de nouveau en deux les données inférieures à la médiane (resp. supérieures à la médiane), et donc d'en prendre à nouveau la médiane.

3.3. Méthode utilisée par les tableurs Excel, Open Office et le logiciel R

Calcul du quantile d'ordre p.

On construit d'abord le tableau suivant :

Ordonnée	x_1	x_2	$\cdots$	x_i	x_{i+1}	$\cdots$	x_n
Abscisse	0	$\frac{1}{n-1}$	$\cdots$	$\frac{i-1}{n-1}$	$\frac{i}{n-1}$	$\cdots$	1
Indice	1	2	$\cdots$	i	$i+1$	$\cdots$	n

Pour $i = 1, 2, ..., n-1$, on relie par un segment de droite les points $(\frac{i-1}{n-1}, x_i)$ et $(\frac{i}{n-1}, x_{i+1})$.

Ce polygone est donc le graphe de la fonction $H : [0,1] \longrightarrow \mathbb{R}$:

$H(p) = (1-\lambda_p)x_i + \lambda_p x_{i+1}$ si $\frac{i-1}{n-1} \leq p \leq \frac{i}{n-1}$ avec i défini ainsi :

$i = [r]$ où $r = (n-1)p + 1$

$\lambda_p = (n-1)p + 1 - i = r - [r]$

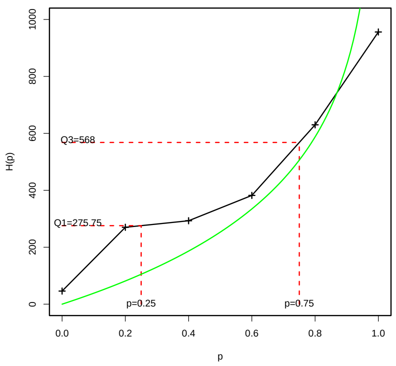
Le quantile d'ordre p est alors donné par : $q_p = H(p)$

Exemple :

Jeu 1 : $46, 270, 293, 382, 630, 952 \quad n = 6$

Ordonnée	46	270	293	382	630	952
Abscisse	0	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{3}{5}$	$\frac{4}{5}$	$\frac{5}{5}$
Indice	1	2	3	4	5	6

On a ajouté le graphe de F^{-1}.



Recherche des quartiles :

Lorsque le graphe de H a été tracé, le quantile d'ordre p est l'ordonnée du point du graphe de H correspondant à l'abscisse p.

On peut aussi le trouver en utilisant la suite de calculs suivants :

1er quartile : $Q_1 = q_{0.25} = ?$
- on calcule $r = (n-1)p + 1 = (6-1) \times 0.25 + 1 = 2.25$
- alors $i = [r] = [2.25] = 2$ et $\lambda_{0.25} = r - [r] = 2.25 - 2 = 0.25$
- on cherche $x_i = x_2 = 270$ et $x_{i+1} = x_3 = 293$
- on trouve $q_{0.25} = H(0.25) = (1-\lambda_p)x_i + \lambda_p x_{i+1} = (1-0.25) \times 270 + 0.25 \times 293 =$ donc $Q_1 = 275.75$.

3ème quartile : $Q_3 = q_{0.75} = ?$
- on calcule $r = (n-1)p + 1 = (6-1) \times 0.75 + 1 = 4.75$
- alors $i = [r] = [4.75] = 4$ et $\lambda_{0.75} = r - [r] = 4.75 - 4 = 0.75$
- on cherche $x_i = x_4 = 382$ et $x_{i+1} = x_5 = 630$
- on trouve $q_{0.75} = H(0.75) = (1-\lambda_p)x_i + \lambda_p x_{i+1} = (1-0.75) \times 382 + 0.75 \times 630 = 568$ donc $Q_3 = 568$.

La justification paraît aussi évidente. Puisque $q_p = F^{-1}(p)$, il suffit d'obtenir une approximation de l'inverse de la fonction de répartition de F. Puisque la fonction de répartition empirique $\hat{F}$ est une "bonne" approximation de F, pourquoi ne pas prendre l'inverse de cette fonction de répartition empirique ? Ce n'est pas possible car $\hat{F}$ est une fonction "en escalier". Pour une ordonnée fixée, il peut exister plusieurs abscisses qui correspondent à cette ordonnée... ou aucune ! Elle n'est donc pas inversible. On utilise alors une approximation de F^{-1} par une fonction affine par morceaux construite de telle manière que pour chaque p, il puisse exister une valeur pour cette fonction. C'est la raison pour laquelle on prend les points d'abscisses $\dfrac{i}{n-1}$ et non $\dfrac{i}{n}$ ce qui paraîtrait pourtant plus naturel. La justification qu'en donne J.P.Girard utilise une généralisation de "rang" mais qui donne le même résultat. D'autres tableurs Minitab,... et logiciel SAS peuvent utiliser encore d'autres méthodes mais toutes reviennent à utiliser une approximation de F^{-1}.

4. Début de discussion.

Il n'est pas évoqué une méthode pourtant universellement utilisée. C'est celle où les données sont groupées en classes et où, à partir de ces classes, on construit la fonction représentative du polygone des fréquences cumulées qui permet ensuite par interpolation d'obtenir les quantiles désirés. Cette méthode qui utilise l'interpolation a très longtemps été la seule enseignée. Lorsque les données sont livrées effectivement, "en classes" avec pour chaque classe la fréquence de la classe, c'est la seule méthode utilisable.

Lorsque les données sont fournies sous la forme d'une série de nombres, les regrouper en classes (quelles classes ?) fait perdre de l'information et ceci d'autant plus que la taille de la série est petite. De toute façon, le calcul des quartiles pourra donner par cette méthode un résultat encore différent des trois autres méthodes.

Rappelons aussi qu'historiquement, la médiane a été définie en 1757 par le moine serbe Boscovitch comme une valeur m qui minimise $\sum_{i=1}^{n} |x_i - m|$. On sait que si n est pair, cette valeur peut ne pas être unique.

L'étudiant ne risque-t-il pas d'être dérouté par le fait que, pour un même jeu de données, différentes méthodes puissent donner des résultats différents pour une même définition ?

L'enseignant peut répondre que le quartile obtenu n'est qu'une approximation - estimation - du "véritable" quartile, et qu'il peut y avoir, pour un paramètre inconnu à estimer à partir d'un échantillon de taille de toute façon finie, plusieurs approximations possibles, mais ceci entraîne une interrogation supplémentaire, quelle est la "réalité", sinon mathématique, de ce paramètre inconnu, le "véritable" quartile ? Cette interrogation ne nous amène-t-elle pas à une autre : la frontière entre le fini et l'infini ?

Une autre réponse possible est d'enseigner ce qui n'est "pas contraire" à l'outil de calcul utilisé. Dans le cas où c'est la calculatrice qui est utilisée, prendre la définition de la calculatrice. Dans le cas où c'est le tableur, prendre la définition du tableur. Mais n'est-ce pas dans ce cas se soumettre aux fabricants de matériel ?

" Quant au calcul fait par Excel, il peut être intéressant d'expliquer aux élèves que l'on ne maîtrise pas toujours les résultats donnés par les fonctions standard d'un logiciel." C'est une réponse trouvée dans un document de l'Inspection pédagogique régionale de mathématique de l'Académie de Strasbourg intitulé : "Les difficultés soulevées par la définition de la médiane et des quantiles" !

5. ...et Wikipedia ? L'encyclopédie participative en ligne Wikipedia est en quelques années devenue une ressource documentaire très usitée.

Citons l'article « quartile » tiré de cette référence.

> « En statistique descriptive, un quartile est chacune des 3 valeurs qui divisent les données triées en 4 parts égales, de sorte que chaque partie représente 1/4 de l'échantillon de population.
>
> Calcul des quartiles
>
> Voir à quantile pour les méthodes. Le quartile est calculé en tant que 4-quantile. Donc :
> – le 1er quartile sépare les 25% inférieurs des données ;
> – le 2e quartile est la médiane de la série ;
> – le 3e quartile sépare les 75% inférieurs des données.
>
> Méthode :
> – Dans le cas continu on utilise la fonction représentative du polygone des fréquences cumulées. (voir à Statistiques élémentaires continues)
> – Dans le cas discret on range les données par ordre croissant ensuite : Le quartile inférieur est la valeur du milieu du premier ensemble, dans lequel 25% des valeurs sont inférieures à Q1 et 75% lui sont supérieures. Le premier quartile prend la notation Q1. Le quartile supérieur est la valeur du milieu du deuxième ensemble, dans lequel 75% des valeurs sont inférieures à Q3 et 25% lui sont supérieurs. Le troisième quartile prend donc la notation Q3
>
> Exemple :
>
> Les valeurs dans l'ordre ascendant 1, 11, 15, 19, 20, 24, 28, 34, 37, 47, 50, 57.
>
> Q1 est entre 15 et 19 donc : Q1 = 17
>
> Q2 est entre 24 et 28 donc : Q2 = 26 (c'est la médiane)
>
> Q3 est entre 37 et 47 donc : Q3 = 42 »

Quelques critiques... entre autres !

- Dans le cas discret, l'"ensemble" dont il est question est sans doute le sous ensemble $E =]15, 19[$ car 25% des valeurs de la série sont inférieures à n'importe $x \in E$ et 75% sont supérieures à $x \in E$. Pour le Jeu 3, on retrouve la définition des calculatrices. De toute façon, définition pas claire !

- Qu'en est-il pour le Jeu 2 : $49, 90, 198, 302, 387, 547, 763$?
  $E =]90, 198[$ ou bien $E =]49, 90[$?
  Dans le premier cas, plus de 25% des valeurs inférieures à $x \in E$ et moins de 75% sont supérieures à $x \in E$: Q1= 144 ;
  Dans le deuxième cas, moins de 25% des valeurs inférieures à $x \in E$ et plus de 75% sont supérieures à $x \in E$: Q1= 69.5. Définition inapplicable dans l'état si $n \neq 4k$.

Pour poursuivre la polémique, voir le site MathemaTex :
`http ://forum.mathematex.net/mathematiques-f5/mediane-quartiles-t2127.html`

Jean LEJEUNE

Permanence « Statistiques » à l'IREM

Vous vous posez des questions concernant les programmes, vous cherchez des compléments d'information, ...
Jean LEJEUNE peut vous rencontrer à l'IREM si vous le souhaitez.
Prendre contact par messagerie : jeannotlejeune@orange.fr

Un nouveau livre

Arts et sciences à la Renaissance sous la direction d'Évelyne Barbin, Paris, éditions Ellipses, 2007, 320 pages.

Trois des treize contributions de cet ouvrage collectif sont dues à des membres de l'IREM de Basse-Normandie : Didier Bessot et Jean-Pierre Le Goff.

Résumé et sommaire

Les différents chapitres de cet ouvrage s'intéressent à " l'art-science ", c'est-à-dire au mélange de sciences et d'arts dans la peinture, l'architecture, la sculpture, la musique ou la poésie à la Renaissance. Ils expliquent les investigations que les artistes vont entreprendre pour rendre compte exactement du réel et pour imiter parfaitement la nature. Les peintures de Léonard de Vinci sont instruites par les pratiques de la dissection et les connaissances anatomiques. Les peintres et les architectes inventent des méthodes de représentation en perspective. Les anamorphoses, au contraire, leur servent à déformer curieusement le réel. Dans son traité de sculpture, Alberti donne un tableau des mesures humaines et propose un appareil qui repère chaque point de la statue par des coordonnées. Vincenzo Galilei entreprend des expériences sur les cordes de son luth, qui inaugurent une nouvelle physique. Johannes Kepler établit un lien entre consonances et polygones constructibles à la règle et au compas. Jacques Peletier du Mans, invente la " poésie scientifique ". Tandis que les ouvrages de botanique, de zoologie, de géométrie ou de fortification contiennent des images aussi fidèles que possible à la réalité, mais aussi des images d'êtres mythiques.

Peinture, architecture et sculpture — Léonard De Vinci et sa vision de la main, par Dominique Le Nen – De points de distance en points de fuite, une histoire non linéaire de la perspective, par Jean-Pierre Le Goff – La perspective géométrique dans la tradition française : Jean Pèlerin, dit Viator et Jean Cousin, dit le Vieux, par Jean-Pierre Le Goff – Géométrie des anamorphoses : distorsions perspectives ou réseaux déformateurs ?, par Didier Bessot – Alberti et la technique de l'imitation, par Jean-Yves Boriaud

Musiques et sciences — Musique et science : les aléas d'un modèle, par Philippe Vendrix – Gioseffo Zarlino-Vincenzo Galilei : mathématiques et musique, un nouveau regard ?, par Anne Boyé – Musiques et sciences à Florence, à la fin du XVe siècle : la Camerata et les Galilei, par François Baskevitch – Kepler, théoricien de la musique : un savant original et inclassable, par Patrice Bailhache

Arts et mathématiques — L'artillerie à la Renaissance entre art et science : l'exemple anglais, par Pascal Brioist – Jacques Peletier du Mans, poète et mathématicien de la Renaissance, par Myriam Papin

Images des sciences — La Renaissance des images naturalistes, par Sylvène Renoud – Images de la nature, images de l'homme dans les ouvrages de géométrie pratique, par Frédéric Métin – Le bestiaire d'Ambroise Paré entre mythe et réalité, par Philippe Massé

Annonces de colloques IREM

Épistémologie, Géométrie La figure et la lettre. Nancy, 23 et 24 mai 2008 (en collaboration avec les Archives Henri Poincaré)

Formation des enseignants La notion de fonction. L'évaluation par compétences. Antony, 19 et 20 juin 2008

Collège, statistiques et probabilités Les dés sont-ils à jeter ? Périgueux, 19, 20 et 21 juin 2008

Pour plus de détails, consultez le site des IREM : http ://www.univ-irem.fr/

La revue Repères

La revue des instituts de recherche sur l'enseignement des mathématiques
Sommaire du numéro 70, janvier 2008

Pour consulter le site Web de la revue Repères IREM et les articles en ligne : Accédez au site du réseau des IREM par http ://www.univ-irem.fr/ puis cliquez sur REPERES (dans bandeau supérieur horizontal), ensuite sur CONSULTER. (Les nombres en rouge indiquent qu'au moins un article de ce numéro a été mis intégralement en ligne ; les nombres en vert indiquent que tous les articles de ce numéro sont intégralement en ligne)

Pour vous abonner à Repères IREM ou acheter séparément des numéros, contacter : Topiques-Éditions, 3, place Jeanne d'Arc, 57000 METZ, France, Téléphone & télécopie : 03 87 75 20 59, adresse électronique : topiquesed@aol.com
Prix d'un abonnement (4 numéros par an) : Métropole : Établissements, 40 euros ; Particuliers, 32 euros DOM-TOM ou Étranger (par avion) : Etablissements, 51 euros ; Particuliers, 43 euros
Prix au numéro : 11 euros + frais d'expédition si envoi par avion

---