Les valeurs numériques et leur précision.

Généralités.

Un très grand nombre de calculs, voire tous, ont pour but de trouver une valeur numérique pour le résultat. Dans de nombreux exercices de mathématique, on cherche à établir une formule comportant des paramètres et des variables, et la dernière question est souvent nommée "application numérique". L'utilisation des calculettes a modifié l'intérêt de cette démarche, mais il reste le point le plus important : le résultat numérique que l'on écrit au bas de la page.

Sans entrer dans le détail des nombres, on les utilise sous deux formes

1- les nombres entiers qui servent aux comptages, indices etc.
2- les nombre réels qui servent à mesurer des quantités suivant une certaine unité. Ces nombres peuvent aussi représenter des éléments sans unité, par exemple le nombre π, des rapports de nombres, des pourcentages, des probabilités etc.

Je ne m'intéresserai ici qu'aux nombres réels.

Un exemple simple.

Soit un terrain rectangulaire longueur = 28.47 m. , largeur = 15.23 m. , quelle est son aire ? Le résultat de la multiplication donne 433.5981 m² Est-ce réaliste d'indiquer 4 chiffres après le point décimal ? Non, certainement pas. Imaginons que la largeur soit en réalité 15.225. On a arrondi à 15.23, ce qui est normal. Mais on aurait pu aussi arrondir à 15.22. Le résultat de l'aire serait alors 433.3134.

Donc, on décide d'arrondir le nombre 433.5981 à 433.598 puis à 433.60, puis à 434. m² ce qui est parfaitement normal. De la même façon, le nombre 433.3134 sera arrondi à 433. m². Lequel des deux résultat est bon ? Y a-t-il une faute ? La réponse est simple : le nombre 434 m² (ou 433 m²) représente l'aire du terrain. Ce n'est pas une valeur exacte, pas plus que la mesure de la longueur ou de la largeur.

Cet exemple sera repris de façon plus détaillée dans la suite.

Un peu d'histoire.

Il n'y a pas si longtemps, c'est à dire avant l'arrivée de l'informatique, les moyens de calcul étaient limités au crayon et aux machines mécaniques.
On utilisait aussi des outils considérés comme bizarres, la règle à calcul, la table de log. La finalité était simple : faire des calculs justes avec les moyens dont on disposait. Un calcul "juste" n'est pas un calcul "exact", c'est un calcul qui fournit un résultat cohérent avec les informations dont on dispose. En reprenant l'exemple du calcul d'aire, un résultat juste provenant d'un calcul juste est 434 m², mais aussi 433 m². La valeur "exacte" n'est pas connue.

Un mot sur la notion de précision. On l'exprime souvent en "nombre de chiffres significatifs", le nombre de décimales, c'est à dire la position de la virgule ne dépend que des unités choisies. Un exemple simple : la mesure d'une distance sur un plan. L'appareil utilisé est un triple décimètre appelé kutch. Il a une section triangulaire et donc on dispose de 6 échelles. On estime que le pouvoir séparateur d'un œil normal est 1/10 mm. Ce kutch mesure 300 mm,

on peut donc estimer qu'il y a 3000 positions distinguables par un œil normal, c'est à dire 3 ou 4 chiffres significatifs. Il en est de même avec une règle à calcul qui a la même dimension.
Pour mesurer des superficies, on utilise encore maintenant des planimètres. On estime que la précision de la mesure des aires est de l'ordre de 1/500, soit environ 3 chiffres significatifs.

1/7 = 0.14 ; vrai ou faux ?

Cette question a provoqué un long débat sur un forum.
Il y a une expression à gauche du signe '=' et un nombre décimal à droite.

Que représente le signe '=' ? Cela vaut la peine de se poser la question, mais une réponse complète nécessiterait probablement un chapitre entier.
Une réponse simple serait de dire : si on a A = B, alors on a aussi B = A. Une formulation plus mathématique serait de dire "Si A = B est vrai, alors B = A est vrai". En logique pure, une telle affirmation est presque toujours fausse. En effet, lorsqu'elle est vrai, cela signifie que A et B sont strictement interchangeables, si c'était le cas A et B contiennent la même information et la même formulation. Ce serait donc la même chose d'écrire A=A ou B=B, ce qui, on l'admettra, n'a pas grand intérêt. D'ailleurs, il ne semble pas que le signe '=' fasse partie des caractères utilisés en logique. Les caractères utilisés sont plutôt du genre '=>' qui se lit "implique". On peut remarquer aussi que le signe':=' est quelque fois employé. L'expression A:=B doit-elle être comprise comme "on déclare que A est égal à B" ? Ceci dépasse le cadre de ce papier.

Par contre en calcul numérique et en calcul formel, cette formulation est généralement utilisée. "A = B" doit être compris comme "A est connu, quelle qu'en soit la raison, on trouve une expression B qui lui est égale, éventuellement sous certaines conditions".
En calcul formel on peut estimer que, sauf exception, elle est toujours vraie, par exemple $(a+b)^2 = a^2 + 2ab + b^2$ est équivalent à $a^2 + 2ab + b^2 = (a+b)^2$.
Le logicien admettra facilement, je pense, que cette égalité est fausse : à gauche du signe égal, on a 3 termes associés par le signe '+' (addition), sans aucun rapport (sauf algébrique) avec ce qu'il y a à droite du signe '='.

En calcul numérique c'est clairement faux pour l'exemple cité (1/7 = 0.14).
Par contre 3/2 = 1.5 est vrai, puisque 1.5 = 3/2.

Etudions l'expression de gauche, on appelle cela une fraction. C'est une notation particulière et différente de 1:7. Clairement 1:7 est une expression qui représente une division. On s'attend à effectuer la division pour obtenir un nombre. On va diviser le nombre 1 par le nombre 7. Si on fait cette opération avec un ordinateur, dans le cas général, on obtiendra 0. La raison est simple, si on divise un nombre entier par un nombre entier, on obtiendra un nombre entier. Pour tourner la difficulté on va déclarer que 1 et 7 sont des nombres réels, par exemple en rajoutant un point décimal.
Que représente en mathématique l'expression 1/7 ? C'est une représentation codée pour dire "inverse du nombre 7".Dans le cas où le dénominateur est plus grand que le numérateur, le nombre équivalent est inférieur à 1. Cela ne peut être qu'un nombre réel, puisqu'il n'existe pas de nombre entier compris strictement entre 0 et 1.
Par définition, un nombre réel ne peut pas être exact. On en conclue que la représentation 1/7 est considérée comme la représentation simple (3 caractères) d'un nombre réel dont on ne connaît pas la valeur exacte.

Autres exemples de représentation de nombres réels sous forme symbolique rac(2), π, 22/7, e, 2/3 etc. On appelle cela "valeur exacte" ou "nombre exact", par opposition à respectivement 1.4142, 3.1416, 3.14, 2.71828, 0.67 qui sont des valeurs approchées.

Etudions l'expression de droite. C'est un nombre décimal. Le nombre de décimales est deux. Il est très courant d'exprimer les nombres décimaux avec deux décimales, par exemple les prix.
Dans la plupart des cas le nombre de "centièmes", c'est à dire la seconde décimale correspond à un arrondi ou éventuellement une troncature.
Exemple d'étiquette imprimée par une balance sur un produit de super-marché :
Prix 13.95€/kg
Poids net 0.212 kg
Prix à payer 2.96 €
Le résultat de la multiplication est 2.9574. Le nombre représentant le prix à payer est arrondi au centime le plus proche, c'est à dire 2.96. Ceci est-il faux ? Certainement pas.
Le prix au kg peut être considéré comme un nombre exact. C'est à dire le nombre réel dont tous les chiffres suivant le dernier indiqué sont des zéros. Plus rigoureusement, on a fixé le prix avec l'unité "centimes" par un nombre entier, puis on l'affiche avec l'unité € en positionnant une virgule (ou un point décimal) avant le deuxième chiffre en partant de la droite.
Le poids net est un nombre réel, avec 3 chiffres significatifs dans le cas présent, qui ne peut pas être un nombre exact. On ne connaît pas sa valeur. La balance l'a, probablement, arrondi au 1/1000 le plus proche. Enfin, le résultat de la multiplication a été arrondi au centime le plus proche.
En fait, on ignore si la balance a fait la multiplication du prix au kg et du poids mesuré, probablement avec 5 ou 7 chiffres significatifs, puis arrondi le résultat au centime le plus proche ou si au contraire, le nombre correspondant au poids mesuré a été arrondi au gramme le plus proche, puis multiplié par le prix au kg, considéré comme valeur exacte.

Revenons à l'égalité, titre du paragraphe 1/7 = 0.14.
Le résultat "exact" pourrait s'écrire 1/7 = 0.142857142857142857142857 … le groupe "142857" se répétant indéfiniment.

Nota. Pour être tout à fait rigoureux, la question posée à des élèves de seconde était la suivante:
$1/7 \in \{0.15;0.14;5\}$?
Sans aucun autre commentaire.
La réponse attendue était NON. Il y a lieu de préciser que l'étonnement du professeur résidait dans le fait que les élèves qui se sont trompés ont effectué la division avec la calculette et ont arrondi à 0.14, ce qui leur a paru correct alors que le professeur attendait une réponse résultant du raisonnement "1/7 est un nombre réel, il ne peut pas être écrit sous forme décimale".

Avec l'informatique.

Comme je l'ai dit dans un paragraphe précédent, autrefois, on effectuait les calculs numériques avec une règle à calcul La règle à calcul donne le résultat sans tenir compte de la position de la virgule (point décimal). Autrement dit la manipulation pour calculer 689 x 468 ou 6.89 x 4.68 était la même. Le résultat était environ 322 et il incombait à l'utilisateur de positionner la virgule correctement.
Avec la table de logarithme, l'opération était plus facile, puisque le logarithme décimal se compose d'une mantisse et d'une caractéristique. La représentation dans une machine informatique est assez comparable. On parle généralement de nombre à virgule flottante. Il est composé d'une mantisse, la suite des chiffres composant le nombre et d'un exposant, ce qui correspond à la caractéristique d'une table de log.

Détaillons les problèmes liés à la précision.
La norme prévoit deux types de nombres scientifiques, le type "float" et le type "double".
Un nombre déclaré *float* est sur 32 bits, c'est à dire 4 octets. L'exposant est compris entre -38 et 38, soit environ 1 octet il reste donc 3 octets pour la mantisse, soit une précision de 7 chiffres significatifs.
Un nombre déclaré *double* est sur 64 bits, c'est à dire 8 octets. L'exposant est compris entre –308 et 308, soit environ 2 octets, il reste donc 6 octets pour la mantisse, soit une précision de 15 chiffres significatifs.
Ceci a des conséquences importantes.

1- comparaison de nombres "réel". Un nombre que l'on appelle réel en mathématique est impossible à représenter en informatique. Reprenons l'exemple de 1/7, la machine ne pourra mémoriser que 7 chiffres c'est à dire probablement 0.1428571
   Sur ma machine j'ai tapé les lignes suivantes:
   ```
   float A=1./7.;
   double B=A;
   printf("A=%0.8f B=%0.8f 1/7=%0.8f\n",A, B, 1./7.);
   ```
   Le résultat est le suivant
   A=0.14285715 B=0.14285715 1/7=0.14285714
   On observe la différence entre A qui est initialisé à 1./7. et l'impression de 1./7.
   Donc, si on avait testé l'égalité entre A et 1./7. le résultat aurait été FALSE.
   Il est donc indispensable de tester l'égalité entre flottants avec une petite tolérance.

2- Opération arithmétiques sur les flottants. Il est clair que la machine travaille et calcule avec les nombres dont elle dispose, c'est à dire 7 chiffres significatifs.
   Cas de l'addition : si on ajoute un petit nombre à un grand nombre, par exemple 12345.9 + 0.008, le résultat aura le nombre de chiffres significatifs possibles et le second terme sera probablement perdu.
   Cas de la multiplication : on connaît la formule de calcul d'aire d'un polygone à partir des coordonnées $A = ½ \Sigma((X_i+X_{i+1}) * (Y_i - Y_{i+1}))$. Compte tenu de ce qui a été dit, on constate l'importance de s'imposer de ne faire des calculs que sur des petits nombres, tout au moins de même ordre de grandeur

Il faut noter que les calculettes auraient tendance à masquer ces problèmes en faisant les calculs avec un grand nombre d'octets. Mais quels que soient le nombre de chiffres significatifs, ils sont en nombre fini.

Il paraît intéressant de préciser une application de ces notions. Soit un logiciel qui traite de représentation géographique, par exemple le cadastre.
On sait que les coordonnées géographiques sont dans le système légal Lambert 93. Les coordonnées de l'origine sont X=700 000 m, Y=6 600 000 m. C'est à dire 6 ou 7 chiffres significatifs pour les mètres. Il semble donc illusoire d'espérer calculer quoi que ce soit avec une précision meilleure que le mètre. La solution consiste à faire un changement de variable. Par exemple pour la région de Strasbourg, on fera une translation
DX=-1070000. m. et DY=-6880000 m.
Autre astuce, puisqu'on travaille avec des valeurs de précision homogène. c'est à dire qu'on va faire les calculs au millimètre près pour être sûr de la précision du cm., on pourra multiplier toutes les valeurs par 1000 ou 10000 et calculer en nombres entier. On "gagne" ainsi 2 à 3 chiffres significatifs.

Précision des calculs à partir de mesures.

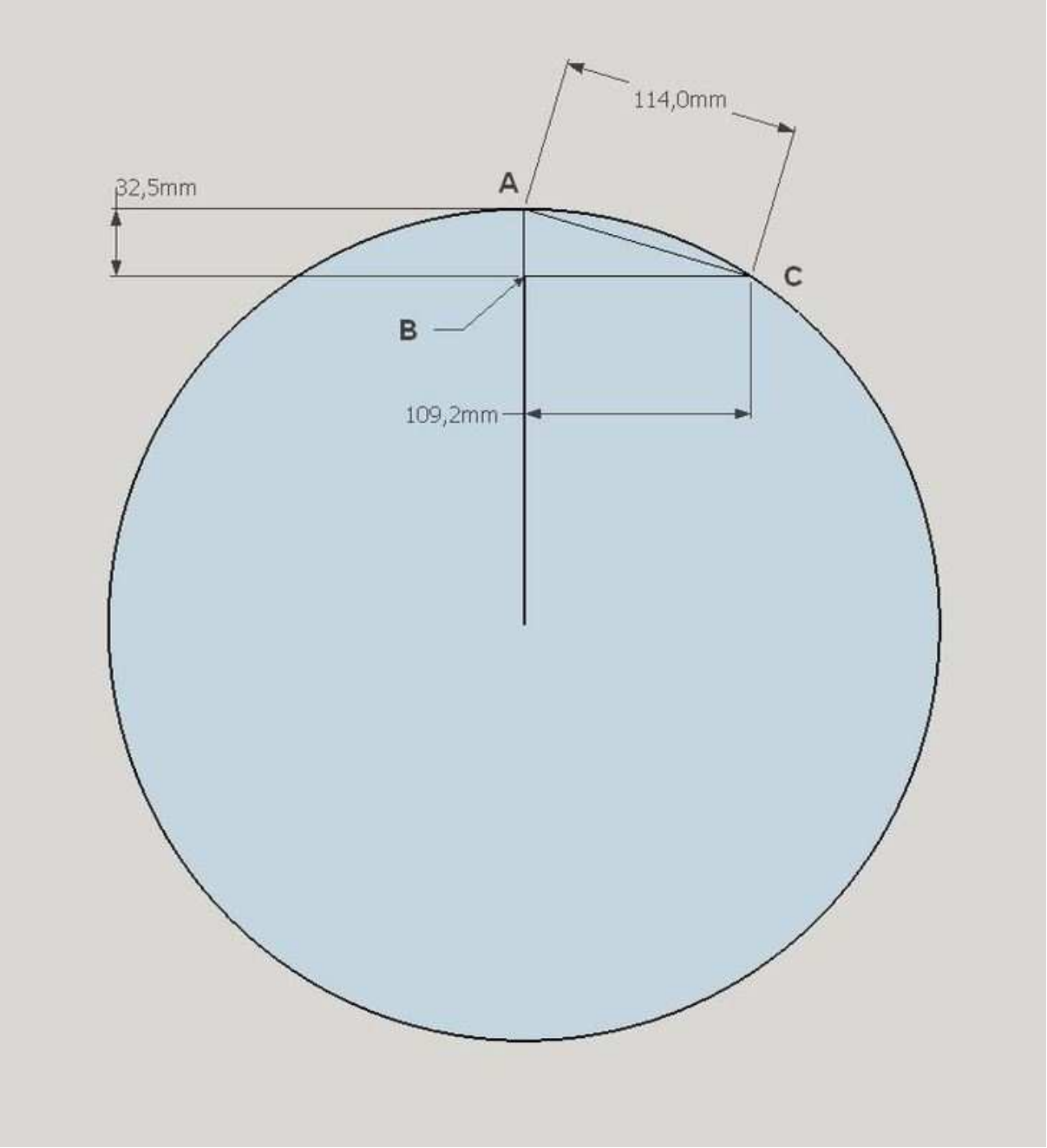
Ce paragraphe est plus compliqué et entre dans le cadre du calcul d'erreur. Je partirai d'un exemple du cours de JJ. Levallois et donnerai les explications au fur et à mesure.

On considère un rectangle de côté x, y mesurés au double-décamètre ; on demande l'erreur moyenne quadratique (écart-type) caractérisant la précision de la surface.
Nous avons $S = x \cdot y$
$dS = x\,dy + y\,dx$.
On appelle E^2 indice x le carré de l'écart moyen quadratique, qu'on appelle aussi variance.
Alors $E^2_s = x^2\,E^2_y + y^2\,E^2_x$. En effet, on applique le principe de l'indépendance des erreurs et les produits de petits de second ordre sont négligeables.
Soit ε l'erreur moyenne quadratique d'une portée de double-décamètre
Le côté x contient n portées (y compris l'appoint).
Le côté y contient n' portées (y compris l'appoint).
On a par conséquent $E^2s = \varepsilon^2.n$; $E^2y = \varepsilon^2.n'$.
D'où $E^2s = x^2\,\varepsilon^2.n' + y^2\,\varepsilon^2.n$
Application numérique x=120m. n=6 ; y=160m. n'=8 ; ε=±1cm
$E^2s = 120^2 \times 8 \times (0.01)^2 + 160^2 \times 6 \times (0.01)^2$
Es = ± rac((1.20)² x 8 + (1.60)² x 6) ≈ ± rac(27) = ± 5.1 m²
" On voit donc combien il est illusoire de fixer le chiffre des centiares dans les évaluations de surfaces, où le centiare ne peut jouer que le rôle de décimale << psychologique >>."

Exemple

Cet exemple a été pris sur un forum de mathématique généraliste et de haut niveau, c'est à dire qu'il n'a pas pour objectif principal l'aide aux lycéens.
La question posée était "comment calculer de rayon du cercle".
Géométriquement, ce problème est simple, d'autant qu'il était accompagné d'une figure.
Pourtant, il a provoqué plusieurs échanges entre des ténors de ce forum.

Examinons la figure, on a un triangle ABC. AC est une corde du cercle cherché, et B apparaît bien comme la projection orthogonale de C sur le rayon passant par A. En d'autres termes, telle que la figure a été dessinée, le triangle ABC est rectangle en B.



Les longueurs des 3 côté du triangle sont 114.0 ; 32.5 ; 109.2. Si on applique le théorème de Pythagore, on obtient $114.0^2 \sim 32.5^2 + 109.2^2$, soit 12996.0 ~ 1056.25 + 11924.64
Ce qui pourrait laisser penser que ce triangle n'est pas rectangle.
La racine carrée de la somme des deux côtés donnerait une hypoténuse égale à 113.98 au lieu de 114.0, soit une différence de 2/100 mm. D'après les données, les mesures ont été faites avec la précision du 1/10 mm.
On a vu que l'on peut admettre l'hypothèse que le triangle ABC est rectangle en B. Il y donc une valeur en sur-nombre.
Si on admet comme juste les longueurs des 2 côtés, alors l'hypoténuse mesure 113.98
De la même façon le grand côté, admettant l'hypoténuse et le petit côté serait 109.27
De la même façon le petit côté, admettant l'hypoténuse et le grand côté serait 32.73
Les erreurs de mesures seraient respectivement 0.02, 0.07 et 0.23.

N'ayant aucune information sur la méthode de mesure, il est impossible de décider. Pour mémoire, les différents résultats trouvés sont 199.71, 199.94 et 198.52.

Conclusion.

Lors de l'énoncé d'un résultat numérique, il est d'usage de considérer que le nombre indiqué est "bon". Ce qualificatif "bon" ne signifie pas "exact" mais signifie "résultat obtenu avec les informations dont on dispose".
Un nombre exagéré de décimales n'apporte rien et surtout diminue la crédibilité de ce résultat..

Pierre Dolez, le 22 septembre 2015