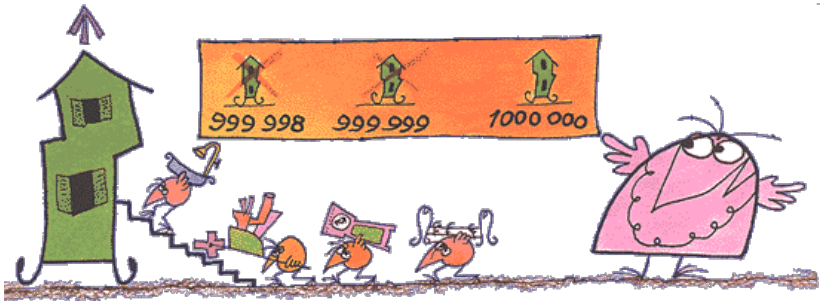
Calculs approchés

Ça peut coûter cher…

Pour un chiffre de plus…

Un dispositif de calibrage utilisé dans la fusée Ariane 4 avait été laissé actif, alors qu'il n'était pas utilisé dans Ariane 5.
Les conditions de vol étant différentes entre Ariane 4 et Ariane 5, la valeur d'une donnée traitée par ce dispositif se trouvait, dans Ariane 5, dépasser les limites prévues pour Ariane 4. Ce dépassement avait une probabilité jugée négligeable de survenir dans Ariane 4 et aucune récupération d'erreur n'était prévue pour le traiter.
En conséquence, selon une politique “normale” dans ce cas de figure, le module de navigation fut mis hors service pour erreur irrécupérable.
Le module actif et le module de secours étant identiques et contenant le même logiciel, les mêmes causes produisirent les mêmes effets sur les deux modules, donnant lieu à une situation non prévue.
En effet, le module de secours était destiné à compenser une erreur transitoire et aléatoire d'origine matérielle, erreur dont la probabilité était jugée suffisamment faible pour exclure dans la pratique une défaillance simultanée des deux modules.

La probabilité de réussir la mise sur orbite d'une fusée est d'une chance sur un million. Dépêchons-nous de rater 999 999 lancements !



La méthode Shadok

Vive le running

- Plage de mesure : 0 à 9999,9 m
- Erreur: ≤ 0,5%
- Diamètre de la roue : env. 32cm (12.6 ‘’)
- Périmètre de roue: env. 1 mètre
- Longueur totale : env. 101cm (39.8’’)
- Température appropriée: -10 ~ 45 °

Organisateur de courses à pied, vous devez mesurer le plus exactement possible les 42,195 km d’un marathon à l’aide d’une roue de mesure qu’un catalogue présente ci-contre…

Question :

On suppose que le diamètre de la roue a été mesuré à 32 cm avec une précision d’un dixième de mm. Peut-on être sûr de la longueur du marathon à 1 m près?

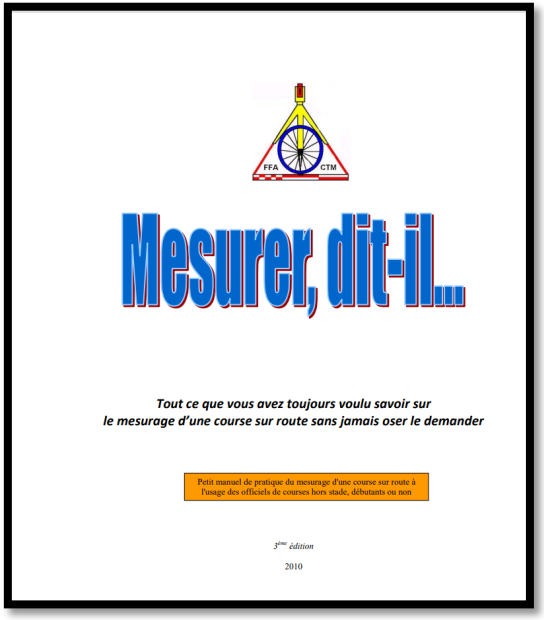
31,99 x 3,141592654 = 100,499549

Si on fait 42000 tours 100,499549x42 000 = 42 209,81…

32 x 3,141592654 x 42 000 = 42 223,005…

Moralité : on n’est sûr de rien. Comment mieux faire?

102 pages pour un marathon



Instructions pour l'organisation et la mesure de la longueur des courses pédestres

Une brochure de 102 pages présente les dispositions officielles prises par les autorités sportives pour garantir autant que possible la longueur d'un marathon. Il faut des compteurs Jones, plusieurs officiels sur plusieurs vélos, une piste d'homologation, des calculs statistiques, etc.

Tout ça pour tolérer une erreur de 1 pour 1 000, mais sur 42 km, cela fait 42 m, distance parcourue en environ 8 s.

Un compteur Jones sur un vélo →

Fausse proportionnalité

Mon automobile me ment

Mon automobile prétend me donner sa consommation « instantanée » et en déduire le kilométrage qu'elle peut encore assurer sans nouveau plein.

Mon ordinateur me ment

Observez la durée restante qu'indique votre ordinateur lors d'un téléchargement

« Le capitaliste, dit-on, a payé *les journées* des ouvriers ; pour être exact, il faut dire que le capitaliste a payé autant de fois *une journée* qu'il a employé d'ouvriers chaque jour, ce qui n'est point du tout la même chose. Car, cette force immense qui résulte de l'union et de l'harmonie des travailleurs, de la convergence et de la simultanéité de leurs efforts, il ne l'a point payée. Deux cents grenadiers ont en quelques heures dressé l'obélisque de Luqsor sur sa base ; suppose-t-on qu'un seul homme, en deux cents jours, en serait venu à bout ? »

Pierre-Joseph Proudhon 1840

Personne entre 90 et 100 km/h?

Extrait du tableau donnant les effectifs en fonction de la vitesse en m/s

4,81	12	10,25	94
4,83	16	10,89	129
4,89	13	11,58	1
4,94	18	11,61	177
4,97	17	12,42	263
5,03	19	13,36	36
5,08	16	13,39	312
5,14	8	14,47	578
5,17	19	14,50	14
5,22	25	15,78	305
5,28	15	15,81	931
5,33	17	17,36	3979
5,44	29	19,25	152
5,50	32	19,28	15680
5,56	24	19,31	4
5,61	25	21,64	1172
5,69	24	21,67	37333
5,81	29	21,69	11
5,86	27	24,69	1602
5,89	2	24,72	72459
6,00	33	24,75	937
6,08	25	24,78	1
6,14	41	28,78	43480
6,22	38	28,81	1062
6,31	25	28,83	15
6,44	29	28,86	2
6,53	51		

Le problème des données manquantes

Pour une étude commandée par une société d'autoroutes (Atlandes), on a mesuré les vitesses des véhicules empruntant une bretelle d'autoroute pendant un certain laps de temps. On place des capteurs et on mesure les vitesses de 180 000 automobiles. (étude réalisée par la Société de calcul mathématique, SCMSA, société de conseil aux entreprises)

On observe qu'il y a des « creux », 1 véhicule à 89 km/h 43 000 à 103 km/h, rien entre 90 et 100. Pourquoi?

L'étude tentait aussi de savoir si des véhicules prenaient la bretelle en sens interdit. Les capteurs sont très rapprochés (3,5 m) et ont une sensibilité de 0,02 s. Un véhicule roulant à 100 km/h met 0,126 s pour franchir 3,5 m. Un véhicule roulant à 90 km/h met, lui, 0,14 s. Le système ne distingue pas les vitesses intermédiaires.

Un calcul simple un résultat faux

Sylvie Boldo se penche, après W. Kahan, sur le calcul de l'aire d'un triangle à l'aide de la formule de Héron. Dans le cas d'un « triangle-aiguille », les causes d'erreur sont nombreuses

How to Compute the Area of a Triangle: a Formal Revisit with a Tighter Error Bound

Sylvie Boldo

a	b	c	Heron's Δ'	Accurate Δ
10	10	10	43.30127019	43.30127020
-3	4	2	**2.905**	Error
100000	99999.99979	0.00029	**17.6**	9.999999990
100000	100000	1.00005	50010.0	50002.50003
99999.99996	99999.99994	0.00003	**Error**	1.118033988
99999.99996	0.00003	99999.99994	**Error**	1.118033988
10000	5000.000001	15000	**0**	612.3724358
99999.99999	99999.99999	200000	**0**	Error
5278.64055	94721.35941	99999.99996	**Error**	0
100002	100002	200004	0	0
31622.77662	0.000023	31622.77661	0.447	0.327490458
31622.77662	0.0155555	31622.77661	246.18	245.9540000

Son but n'est pas de découvrir que les ordinateurs calculent faux, mais de créer des *programmes* de calcul qui vérifient, corrigent, contournent ces erreurs.

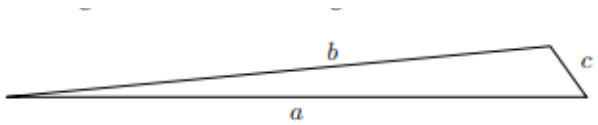


Fig. 1. A Needle-Like Triangle

The common formula to compute the area is two millennia old and is attributed to Heron of Alexandria:

$$\Delta = \sqrt{s\,(s-a)\,(s-b)\,(s-c)}$$

$$\text{where } s = \frac{a+b+c}{2}$$

This formula is known to be inaccurate using floating-point arithmetic since the 80s. This has been first studied by Kahan [1]. He gave examples of incorrect computations: either the result was very wrong or the computation was stopped due to a negative square root, created by round-off errors. Kahan also proposed an algorithm

Les causes d'erreur

- Erreur de principe : on ne fait **pas le bon calcul**, on fait faire le calcul par **un logiciel « de confiance »** ;
- Le résultat du calcul repose sur des **mesures ;**

« Toutes les mesures du monde ne valent pas un seul théorème par lequel avance authentiquement la science de la vérité éternelle » C.-F. Gauss

- On croit à la **proportionnalité** en toutes circonstances ;

Calculer juste, avec quels nombres?

Les nombres des mathématiques

- Les décimaux (dont les entiers) ;

- Les décimaux relatifs ;

Tous les couples (a, b) de décimaux qui donnent la même différence sont **un** relatif (qui a le signe de $a - b$)

- Les rationnels ;

Soient a et b des relatifs tels que $b \neq 0$. Tous les couples (p, q) de relatifs tels que $aq - pb = 0$ sont **un** rationnel.

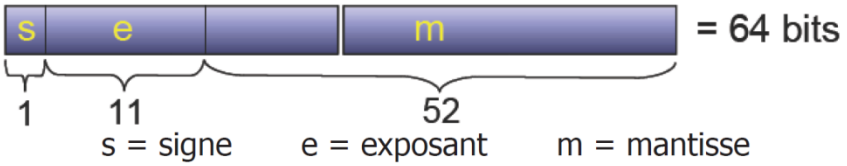
- Les irrationnels ;

Tous les autres nombres réels (parce que ce n'est pas fini).

Les nombres « virgule flottante »

La norme IEEE-754

Double précision = 2 X 32 bits



Nombre = $(-1)^s \times 1.m \times 2^{e-1023}$

Valeur absolue min = $2,225... \, 10^{-308}$

Valeur absolue max = $1,797... \, 10^{308}$

Erreur relative = $2,220446049 \times 10^{-16}$

soit 16 chiffres maximum de précision

Mais alors, lesquels choisir?

Dans un calcul littéral, on traite de nombres qui sont représentés par des lettres ou des symboles ou écrits explicitement. Au moment d'écrire un **résultat :**

- **Si on peut écrire ce résultat comme un nombre explicite** on n'a qu'à respecter les règles d'écriture des nombres ;

- **Si le résultat est un irrationnel**, on l'écrit sous forme interprétable ($\frac{1+\sqrt{5}}{3}$, $e^{\pi\sqrt{163}}$, etc.) ; dans le cas où une ***valeur indicative*** est nécessaire, on en précise le *type* (**arrondi, troncature, valeur approchée**) et la *qualité*. C'est à ce niveau qu'intervient l'écriture à virgule flottante (N.B. Virgule flottante se dit floating point en étatsunien)

Approximation et chiffres significatifs

Point de vue du physicien

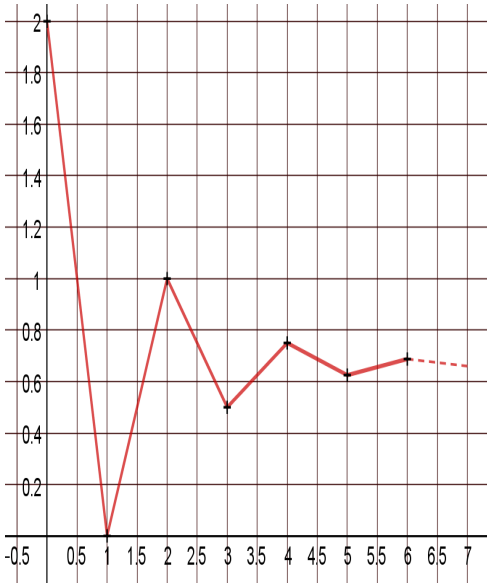
Une grandeur expérimentale n'est jamais parfaitement connue, il existe une certaine incertitude sur une mesure. Par exemple si vous mesurez une longueur avec une règle graduée en millimètres la mesure sera au mieux connue au mm. En mathématique les nombres sont supposés parfaitement connus, si on écrit 1/3 on connaît une infinité de chiffres après la virgule 0,3333333… . En math $\pi = 3{,}141592\ldots$, en sciences physiques on écrira 3,14 , 3,1 ou 3 suivant la précision recherchée.
Ainsi rien ne sert d'indiquer tout les chiffres, donnés par une calculatrice par exemple. Vous voulez découper une bande de papier de L = 1 m en neuf parties égales. Pour vos découpes d=11,1 cm (indiqué au mm car c'est la précision de votre règle, nous n'écrirons pas d = 11,11111111 cm).

Un *calcul d'incertitude* indique ce qu'on peut attendre du résultat. Il est donc inutile d'ajouter des chiffres qui ne seraient pas *significatifs* (attention : 0 peut être un chiffre significatif).

On veut des définitions (1)

Valeur approchée : on dit que a est une valeur approchée de x à ε près ssi $|x - a| \leq \varepsilon$.

Par exemple, pour la fonction représentée (partiellement) ci-contre, on peut établir que pour tout ε positif, il existe M positif tel que $x > M \Rightarrow \left|f(x) - \frac{2}{3}\right| < \varepsilon$



« valeur approchée » porte sur la *distance*

« π est une valeur approchée de 3, 14 à 0, 002 près » est une phrase juste, même si son utilité ne saute pas aux yeux.

On veut des définitions (2)

Les nombres sont écrits dans la numération de base q, dont les chiffres sont 0, 1, 2, etc. Si l'écriture de N (éventuellement illimitée) est : $N = \alpha q^n + \beta q^{n-1} + \cdots + \sigma q^{-m+1} + \tau q^{-m} + \omega q^{-m-1} + \cdots$

(où m et n sont des entiers naturels),

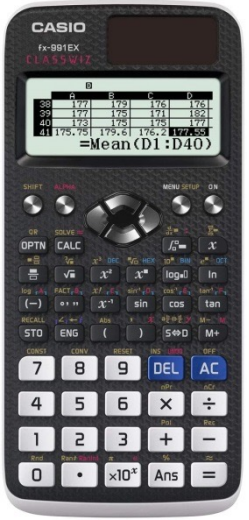
La **troncature** de N au rang $-m$ est $T_m(N) = \alpha q^n + \beta q^{n-1} + \cdots + \sigma q^{-m+1} + \tau q^{-m}$.

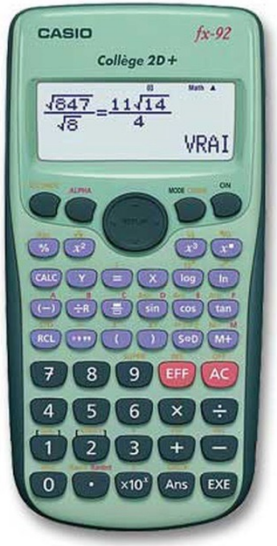
L'**arrondi** au rang $-m$ de N est $A_m(N) = T_m(N)$ si $\omega < \frac{q}{2}$ et $A_m(N) = T_m(N) + q^{-m}$ si $\omega \geq \frac{q}{2}$.

Arrondi et troncature portent sur les *formats* d'écriture

Moralité

- Les calculs littéraux, quand ils sont justes, sont parfois insuffisants ou inutilisables pour le but qu'on poursuit ;
- Tout **affichage** produit par une calculatrice ou un ordinateur ne peut être utilisé tel que : il doit être **interprété** (une réflexion préalable est nécessaire)

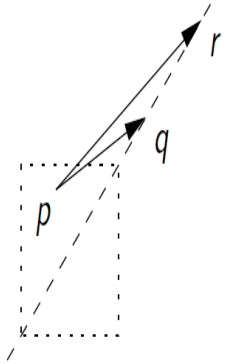




L'ordinateur calcule parfois faux

(G. Melquiond)

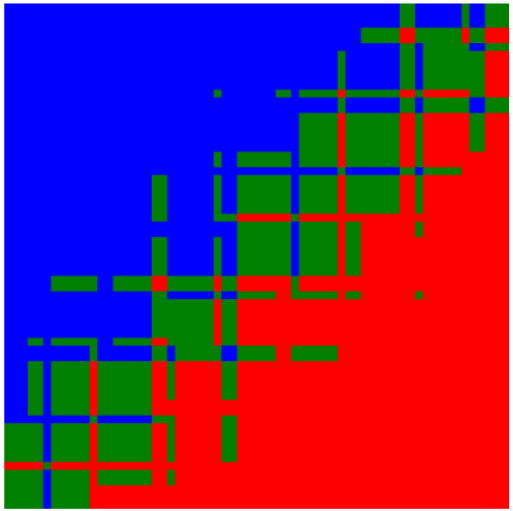
Étant donnés 3 points du plan p, q et r. On veut savoir si pqr sont alignés dans le sens horaire ou dans le sens anti-horaire.



$$\text{orient}_2(p,q,r) = \text{signe} \begin{vmatrix} q_x - p_x & r_x - p_x \\ q_y - p_y & r_y - p_y \end{vmatrix}$$

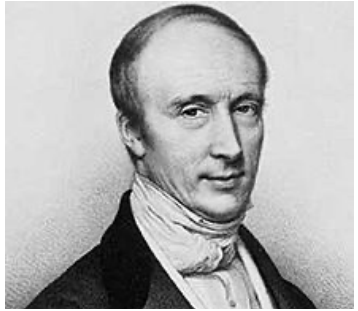
```
float det = (qx - px) * (ry - py)
          - (qy - py) * (rx - px);
if (det > 0) return POSITIVE;
if (det < 0) return NEGATIVE;
return ZERO;
```

Avec $q = (8{,}1; 8{,}1), r = (12{,}1; 12{,}1)$ on regarde les résultats obtenus en faisant varier p autour de $(1{,}5; 1{,}5)$



Des calculs sans erreur?

Limiter les erreurs humaines



Augustin Louis Cauchy (1789-1857)

Mémoire de 1840 : « Sur les moyens d'éviter les erreurs dans les calculs numériques »

Un exemple :

	4	2	3	$\overline{4}$	$\overline{2}$	C'est 42 258
+	1	$\overline{5}$	$\overline{4}$	5	$\overline{3}$	C'est 4 647
=	5	$\overline{3}$	$\overline{1}$	1	$\overline{5}$	C'est 46 905

Autre exemple : $\frac{1}{7} =$

$0{,}142857\ 142857\ 142587\ \ldots$

S'écrit aussi $\frac{1}{1\overline{3}} = 0,143\ \overline{1}\ \overline{4}\ \overline{3}\ \ldots$

Avizienis redécouvre sans le savoir le système de Cauchy. Ses travaux sont utilisés dans les processeurs. Ces systèmes évitent la propagation de retenues.

Algirdas Antanas Avizienis (né en 1932)

Arrondir les troncatures ?

Dans le système ternaire dit « équilibré » les chiffres sont $1, 0$ et $\overline{1}$ comme chez Cauchy, mais il y en a moins… Pour passer d'un entier écrit en base 3, on remplace simplement les 2 qui apparaissent dans l'écriture par $\overline{1}$, en ajoutant 1 aux chiffres situés à leur gauche.
Par exemple : le nombre $1\,789$ s'écrit $2\,110\,021$ en base 3 et $1\overline{1}\,110\,1\overline{1}\,1$ dans le système ternaire équilibré.
Dans ce système, l'arrondi et la troncature coïncident !

En Union soviétique, on a amorcé à la fin des années 1950 la construction d'ordinateurs obéissant à une logique ternaire. Elle fut peu à peu abandonnée en tant que voie industrielle.

Ordinateur Setun
URSS 1958

Les victoires du calcul littéral

Idée reçue :pour multiplier deux nombres entiers de deux chiffres, il faut 4 produits d'un chiffre par un chiffre, deux sommes et éventuellement deux sommes et l'ajout de trois retenues.

Mais : $(10a+b)(10c+d) = 100\times ac + 10\times\big(ac+bd-(a-b)(c-d)\big) + bd$

Ce calcul ne demande que **trois** produits de nombres de un chiffre (souvent inférieurs aux chiffres de départ).

Heureusement, cette façon de faire se généralise à des nombres plus grands, selon le principe « diviser pour régner »

$$(10^n a+b)(10^n c+d)$$
$$= 10^{2n}\times ac + 10^n\times\big(ac+bd-(a-b)(c-d)\big) + bd$$

Le produit par une puissance de la base (que ce soit 10 ou un autre entier n'a pas d'importance) n'entraîne qu'un déplacement de virgule, ne coûte rien en puissance ou temps

Anatoli A. Karatsuba
1937 - 2008

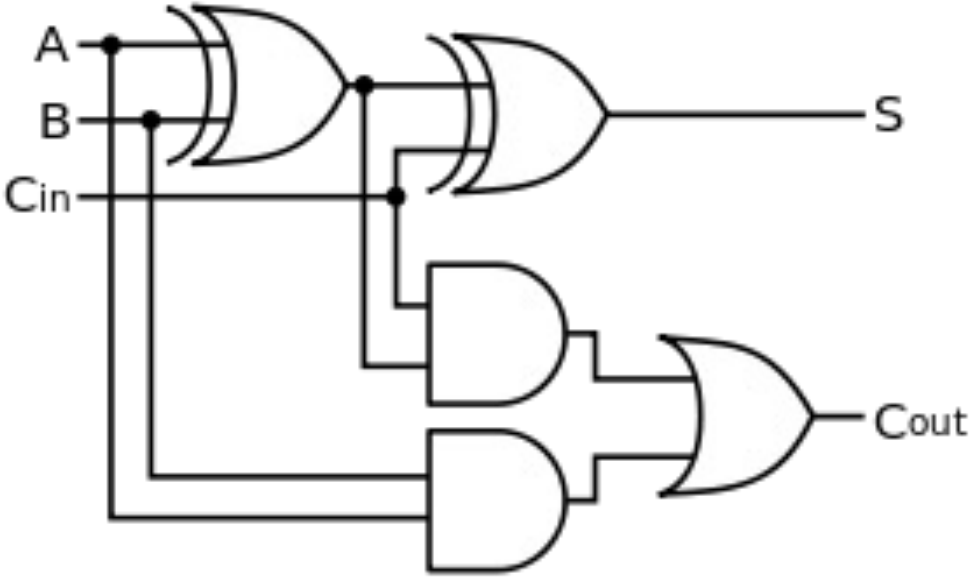
Maîtriser les retenues (1)

Un simple **additionneur binaire** doit à chaque pas gérer une retenue (carry)

Les portes XOR donnent 1 en sortie si et seulement si les entrées sont différentes.

Les portes ET donnent 1 en sortie si et seulement si les deux entrées sont 1, 0 sinon.

Il y a bien d'autres façons d'anticiper ou retarder l'apparition des retenues.



George Boole (1815 – 1864) et Claude Shannon (1916 – 2001)

Maîtriser les retenues (2)

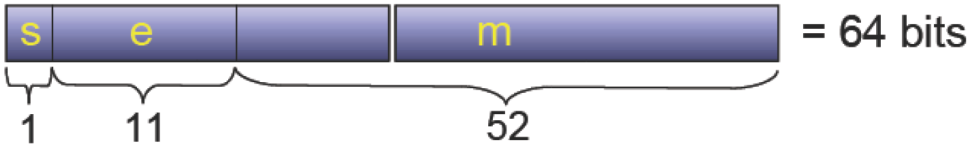
Une vieille histoire…

Calculer par référendum

Avant les calculateurs électroniques, [les calculs] étaient réalisés par des équipes, essentiellement des femmes, réunies dans de grandes salles de calcul qui résolvaient à la chaîne des problèmes arithmétiques simples. Les calculs complexes étaient décomposés et exécutés par deux équipes. Si le résultat était identique, il était validé, sinon, ils répétaient le processus. Les superviseurs archivaient les réponses correctes.

La solution virgule flottante

Double précision = 2 X 32 bits



Nombre = $(-1)^s \times 1.m \times 2^{e-1023}$

Valeur absolue min = $2,225... 10^{-308}$

Valeur absolue max = $1,797... 10^{308}$

Erreur relative = $2,220446049 \times 10^{-16}$

soit 16 chiffres maximum de précision

Les nombres « possibles » avec ce système sont appelés des flottants. Un changement de l'exposant modifie la distance entre deux flottants voisins. Il y a un plus petit *nombre machine*, un plus grand et un *epsilon machine* (Ne jamais faire un test sous la forme $= 0$?)
Les ***arrondis*** sont eux aussi des nombres machine…

Document issu de la présentation de Claude Gomez Pépinière avril 2011

… et la norme IEEE 754

Pour les curieux… qui ont raison de l'être

. Le site https://interstices.info en choisissant les articles par niveau

. À l'adresse http://perso.ens-lyon.fr/jean-michel.muller/goldberg.pdf l'article « What every computer scientist should know about floating point arithmetics »

Sans oublier **Sylvie Boldo** Présidente du jury de la première agrégation d'informatique