Dossier : Calculatrices

Erreurs d'arrondis et Calculatrices

Christian Vassard et Didier Trotoux

L'idée de cet article nous a été inspirée par un exemple illustrant une contribution de Jean-Michel Muller ("Ordinateur en quête d'arithmétique" dans La Recherche spécial nombres juillet-août 1995), dont on trouve les bases dans "Arithmétique et ordinateur" du même auteur (Masson 1989). Nous nous appuyons par ailleurs sur un article de Daniel Saada ("Les réels et la calculatrice") paru dans le n° 339 du Bulletin de l'APMEP.

Dans certains calculs, les erreurs d'arrondis peuvent devenir si importantes qu'elles ôtent tout sens aux résultats obtenus : cela tient à la représentation des nombres-machine en virgule flottante, utilisée tant dans les calculatrices scientifiques de nos élèves que dans les ordinateurs les plus puissants. Cette représentation discrète du continu des nombres réels peut parfois poser de gros problèmes.

Pour preuve, l'histoire édifiante d'Alfred Logarithme qui, en bon père de famille, désire faire un placement à très long terme pour assurer l'avenir de sa descendance. Il se renseigne donc auprès du directeur de la Société Chaotique de Banque (SCB) qui lui propose son nouveau plan d'épargne en ces termes :

" Votre apport initial est de $e - 1$ francs. La première année, vous êtes perdant, on multiplie votre capital par 1, et l'on y prélève 1 franc pour frais de gestion. La deuxième année, c'est beaucoup mieux, on multiplie votre capital par 2 et l'on prélève toujours 1 franc pour frais de gestion. La troisième année, on multiplie votre capital par 3 et l'on prélève 1 franc, et ainsi de suite : la n-ième année, on multiplie votre capital par n et l'on prélève 1 franc. Au bout de 25 ans, vous pouvez retirer votre argent. Intéressant, n'est-ce pas ? "

Prudent Monsieur Logarithme décide de réserver sa réponse. Pouvons-nous l'aider à prendre sa décision ?

La situation se mathématise sans grande difficulté. Il s'agit tout simplement d'étudier la suite (u_n) définie par

- son premier terme $u_0 = e - 1$, où e est la base des logarithmes népériens ;
- la relation de récurrence : $\forall\, n \in \mathbb{N},\ u_{n+1} = (n+1)u_n - 1$

Qui plus est, seul le 25^e^ terme nous intéresse.

A première vue, c'est un calcul qui doit pouvoir s'effectuer sans grande difficulté sur une calculatrice scientifique : 25 multiplications et 25 soustractions, ce n'est pas le bout du monde ! C'est même une situation simple où l'écriture d'un petit programme s'impose.

I - Que dit la HP48S ?

Essayons avec la HP 48 S. Le programme est le suivant :

```
"→ C
"1 25 FOR I
I C * 1 -
'C' STO
NEXT"
C"
```

C est la variable contenant le capital d'Alfred; le calcul proprement dit est géré par une boucle FOR...NEXT et un compteur de boucle I. Attention aussi à l'utilisation de la notation polonaise !

Bilan : après 25 ans, le capital d'Alfred est de 14 809 554 868 500 francs, soit une fortune colossale de presque 15 000 milliards de francs en partant de presque rien.

II - Que dit la TI81 ?

Avant de nous réjouir trop vite (pour Alfred...), effectuons ce calcul avec une autre calculatrice, une TI81 par exemple.

```
Input C
0 → I
Lbl 1
I + 1 → I
IxC − 1 → C
If I < 25
Goto 1
Disp C
```

Même idée que précédemment, à ceci près que la gestion de la boucle se fait par test (il n'y a pas de FOR...NEXT sur une TI81).

Second bilan : après 25 ans, le capital d'Alfred s'élève maintenant à –701 655 174 800 francs soit une dette de plus de 700 milliards de francs ! Ce plan ne serait-il qu'une gigantesque escroquerie ?

III - Que dit la TI 92 ?

La dernière née de chez Texas Instrument permet deux approches : l'une qui donne un calcul approché et l'autre qui relève du calcul formel (très proche de DERIVE mais il y a des différences). Dans les deux cas, il n'est plus nécessaire d'écrire un programme car la calculatrice possède des instructions spécifiques (les calculs sont faits avec l'option MODE Display Digits FLOAT 12, c'est-à-dire avec le maximum de chiffres significatifs autorisés par la calculatrice).

♦ Un calcul approché peut être effectué en MODE SÉQUENCE ; la suite se définit alors avec la touche Y= : $u1 = n\ u1(n-1) - 1$;
$ui1 = e - 1$.

Dans WINDOW, on choisit l'indice du terme initial en faisant nmin = 0 (le premier terme de la suite est en effet u_0). TABLE permet alors d'afficher les valeurs successives de u_n, en pensant à élargir la taille des colonnes par F1, Format, Cell Width = 12.

Qu'observe-t-on ? La suite des valeurs affichées décroît vers 0 jusqu'à u_{15} puis continue de décroître en tendant vers $-\infty$: pour u_{25}, on obtient $-7{,}016552 \times 10^{11}$ qui correspond donc à un capital de moins de −701 655 174 800 francs (voisin de celui trouvé avec la TI81).

♦ Du point de vue formel, il suffit d'utiliser l'instruction When :

when(n=0,ℯ−1,n*u(n−1)−1) → u(n).

En quelques secondes, on obtient la valeur *exacte* de u(25) = u_{25} soit :

$$15511210043330985984000000\,e - 42163840398198058854693626$$

En approximant ce résultat ($\boxed{♦}$ $\boxed{\text{ENTER}}$), on obtient :

− 701 655 174 805

soit finalement quasiment le même résultat qu'avec la TI81.

IV - Que dit DERIVE ?

DERIVE, le "vrai" DERIVE des ordinateurs, peut voler au secours d'Alfred en effectuant un calcul avec autant de chiffres significatifs que l'on désire, ce que n'autorisait aucune des calculatrices précédentes (pas même la TI92). Notons que l'on ne peut pas utiliser la commande ITERE (ou ITERATION) car la suite (u_n) n'est pas de la forme $u_{n+1} = f(u_n)$, mais plus exactement de la forme $u_{n+1} = f(u_n, n)$.

On peut s'en sortir avec :

u(n) := si(n=0,ê−1,n*u(n−1)−1).

On obtient la même valeur qu'avec la TI92, mais c'est l'approximation (6 chiffres significatifs) qui diffère :

#3: 15511210043330985984000000-ê – 42163840398198058854693626

#4: 1.04653.10 ^{17}

De plus, on obtient des approximations extrêmement fluctuantes de u_{25} simplement en faisant varier le nombre de chiffres significatifs.

#5: ChiffrePrécision := 11

#6: – 7.6336323088.10 ^{11}

#7: ChiffresPrécision := 16

#8 : – 9.349023.10 ^{6}

#9: ChiffresPrécision := 21

#10: –200

#11: ChiffresPrécision := 26

#12: 0,043478260069565217391304347

#13: ChiffresPrécision := 31

#14: 0.03993871297242083758937691521961

#15: ChiffresPrécision := 36

#16: 0.0399387296729361008956226977478709465

On constate la grande instabilité du calcul pour les petites précisions : les approximations obtenues varient entre 10^{17} et -10^{11} ! Toutefois, à partir de 31 chiffres significatifs, le résultat semble se stabiliser autour de 0,0399 ce qui laisse à penser qu'Alfred récupèrera un capital s'élevant à un peu moins de 4 centimes, au bout de 25 années.

Une étude mathématique plus approfondie s'impose donc pour justifier rigoureusement le comportement de la suite (u_n) et donner à Albert le conseil judicieux qu'il attend.

V - Etude mathématique de cette suite

On rappelle donc que (u_n) est définie par :

- son premier terme $u_0 = e - 1$, où e est la base des logarithmes népériens ;
- la relation de récurrence : $\forall n \in \mathbb{N},\ u_{n+1} = (n+1)u_n - 1$.

1) première expression de u_n : $\boxed{u_n = n! \sum_{k \geq n+1} \frac{1}{k!}}$ (1)

Démonstration :
Ceci se démontre par récurrence en utilisant le développement de e en série :

$$e = \sum_{k \geq 0} \frac{1}{k!}$$

• Pour $n = 0$, on retrouve bien $u_0 = e - 1 = 0!\, e = \sum_{k \geq 1} \frac{1}{k!}$

• Supposons la propriété démontrée au rang n.

Alors
$$\begin{aligned} u_{n+1} &= (n+1)u_n - 1 \\ &= (n+1)\, n! \sum_{k \geq n+1} \frac{1}{k!} - 1 \\ &= (n+1)! \left(\frac{1}{(n+1)!} + \sum_{k \geq n+2} \frac{1}{k!} \right) - 1 \\ &= (n+1)! \sum_{k \geq n+2} \frac{1}{k!} \qquad \text{CQFD.} \end{aligned}$$

La relation (1) peut encore s'écrire : $u_n = n! \left(e - \sum_{k=0}^{n} \frac{1}{k!} \right)$ (2)

Cette nouvelle écriture permet d'écrire un autre programme de calcul sous DERIVE, donnant le terme de rang n de la suite (u_n) :

u(n) := n! (ê – somme(1/k!,k,0,n))

où ê (obtenu par ALT e) est la base des logarithmes népériens. On obtient alors (noter la curieuse présence du facteur 494) :

$u_{25} = 494(31399210614030336000000\, e - 85351903640077042215979)$

(pour la TI92, c'est analogue, à très peu de choses près :

n! $(e - \Sigma(1/k!,k,0,n)) \rightarrow$ u(n)

et l'approximation (◆ ENTER)donne :

$-1{,}55112100433 \times 10^{12}$

2) Deuxième expression de u_n : $\boxed{u_n = \int_0^1 (1-t)^n e^t \, dt}$

Démonstration :

On utilise la formule de Taylor avec reste intégral, appliquée à la fonction exponentielle sur l'intervalle $[0\,;x]$ (x étant un réel strictement positif).

Il est clair que si l'on pose $f(x) = e^x$, alors pour tout entier naturel n de $\mathbb{N}$ f admet une dérivée d'ordre n et :

$$\forall x \in \mathbb{R},\ f^{(n)}(x) = e^x$$

En particulier, $f^{(n)}(0) = 1$ pour tout entier naturel n non nul.
La formule de Taylor donne alors :

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \ldots + \frac{x^n}{n!} + \int_0^x \frac{(x-t)^n}{n!} e^t\,dt$$

En faisant $x = 1$, il vient : $e = \sum_{k=0}^{n} \frac{1}{k!} + \frac{1}{n!} \int_0^1 (1-t)^n e^t\,dt$ et en remplaçant

dans (2), il vient : $$u_n = \int_0^1 (1-t)^n e^t\,dt$$

D'où une autre façon de calculer les termes successifs de cette suite avec DERIVE :

```
u(n):=INT((1-t)^n*EXP(t),t,0,1)
```

On obtient un résultat ayant la même forme que précédemment.

(Avec la TI92, on entre

```
u(n):=∫((1-t)^n*EXP(t),t,0,1)
```

et on obtient le résultat après un temps de calcul plus long qu'avec les formules précédentes, de l'ordre de 3 minutes).

3) Quelques conséquences peuvent être déduites de cette nouvelle expression.

- **La suite (u_n) est positive** car : pour tout x de $[0;1]$, $(1-t)^n e^t \geqslant 0$
- **La suite (u_n) est strictement décroissante.**

En effet :

$$u_{n+1} - u_n = \int_0^1 (1-t)^{n+1} e^t\,dt - \int_0^1 (1-t)^n e^t\,dt = \int_0^1 -t(1-t)^n e^t\,dt < 0$$

(inégalité stricte car la fonction définie sur $[0;1]$ par $g(t) = -t(1-t)^n e^t$ est négative et continue sur $[0;1]$ sans être partout nulle sur cet intervalle : donc son intégrale entre 0 et 1 ne peut pas être égale à 0)

- **La suite (u_n) est donc convergente,** car strictement décroissante et minorée par 0.

• **Pour tout n de $\mathbb{N}$, $\dfrac{1}{n+1} \leqslant u_n \leqslant \dfrac{e}{n+1}$**

En effet, pour tout t de $[0;1]$, $\quad 1 \leqslant e^t \leqslant e$.

D'où : $(1-t^n) \leqslant (1-t)^n e^t \leqslant e(1-t)^n$ (car $(1-t)^n \geqslant 0$ sur $[0\,;\,1]$).

Soit : $\displaystyle\int_0^1 (1-t)^n\,dt \leqslant u_n \leqslant e\int_0^1 (1-t)^n\,dt$

Un calcul simple donne : $\displaystyle\int_0^1 (1-t)^n\,dt = -\left[\frac{(1-t)^{n+1}}{n+1}\right]_0^1 = \frac{1}{n+1}$

ce qui prouve bien que $\dfrac{1}{n+1} \leqslant u_n \leqslant \dfrac{e}{n+1}$

• **La suite (u_n) converge vers 0,** selon le théorème du gendarme, ou du sandwich, (ou du sandwich au gendarme !), car elle est encadrée par deux suites convergeant vers 0.

Le pauvre Alfred n'a donc aucun espoir de s'enrichir avec ce placement douteux : bien au contraire, plus le temps passe et plus son capital se rapproche de 0 franc.

4) influence du premier terme sur le comportement de la suite.

Etudions maintenant la suite (v_n)définie par :

- son premier terme $v_0 = a$, où a est un réel quelconque ;
- la relation de récurrence :

$$\forall\, n \in \mathbb{N},\ v_{n+1} = (n+1)v_n - 1.$$

Je dis que : $\boxed{v_n = u_n + n\,! \times (a+1-e)}$

La démonstration peut se faire par récurrence.

Ceci est vrai pour $n = 0$ car $v_0 = a = u_0 + 0! \times (a+1-e) = e - 1 + a + 1 - e$.

On suppose la propriété démontrée pour un entier n.

$$\begin{aligned} \text{Alors } v_{n+1} = (n+1)v_n - 1 &= (n+1)(u_n + n! \times (a+1-e)) - 1 \\ &= (n+1)u_n - 1 + (n+1)!(a+1-e) \\ &= u_{n+1} + (n+1)!(a+1-e) \qquad \text{CQFD} \end{aligned}$$

L'étude précédente permet de conclure que :

$$\begin{cases} \text{si } a > e-1, & \lim\limits_{n \to +\infty} v_n = +\infty \\ \text{si } a < e-1, & \lim\limits_{n \to +\infty} v_n = -\infty \end{cases}$$

Plus précisément, il est clair que $v_n \sim n!\,(a + 1 - e)$ pour n tendant vers $+\infty$ donc on peut s'attendre à ce que v_n tende très rapidement vers $\pm\infty$, étant donné la présence de $n!$.

5) Ceci permet de justifier les résultats obtenus, sachant que e est peu différent de 2,71828182845904523536 avec 20 chiffres après la virgule.

- Pour e, la TI81 donne 2,718281828**459**, après avoir pris soin de mettre en évidence les trois chiffres de réserve : $e - 1$ est donc une valeur approchée par défaut et l'on est donc dans le cas où la suite tend vers $-\infty$.
- C'est la même valeur approchée de e qui est utilisée par la TI92, avec cette fois un seul chiffre de réserve : les conclusions sont donc les mêmes que pour la TI81.
- Pour e, la HP48 donne 2,71828182846 : $e - 1$ est cette fois-ci une valeur approchée par excès, ce qui explique que la suite obtenue tende vers $+\infty$.
- Pour DERIVE, la justification est de même nature : tout dépend, suivant la précision choisie, de la valeur approchée de e (et donc de $e - 1$) retenue. Si c'est une valeur approchée par défaut, la suite doit tendre vers $-\infty$ et si c'est une valeur approchée par excès, elle doit tendre vers $+\infty$. Une différence de fonctionnement essentielle cependant : DERIVE mémorise de façon interne tous les nombres soit comme des entiers, soit comme des fractions réduites d'entiers. Ainsi, avec une précision de 6 chiffres significatifs, la valeur de e retenue par DERIVE est : 2,71828 (entrer e puis approX) mais cette valeur est mémorisée de façon interne comme de 23225/8544 (il suffit de demander une simplification - commande Simplifie - du 2,71828 précédemment obtenu) soit environ 2,718281835 qui est bien une valeur approchée par excès de e. Ce qui explique que l'on obtienne un résultat tendant vers $+\infty$.
  A l'inverse, avec une précision de 11 chiffres significatifs, e est mémorisé comme : 15749589/5793950, soit encore 2,718281828458996 qui est une valeur approchée par défaut ; on obtient cette fois-ci une suite qui tend vers $-\infty$.

Qu'en est-il de l'erreur commise, en fonction de la précision retenue ?
Si on conserve p chiffres après la virgule, en supposant, pour simplifier que les résultats sont tronqués, l'erreur commise sur $e-1$ est de 10^{-p}.

Cette erreur est certes faible en général :

- pour la TI81,p vaut 12 en comptant les chiffres de réserve (3) ;
- pour la HP48, p vaut 11 (pas de chiffres de réserve) ;
- pour la TI92, p vaut 12 (1 chiffre de réserve en MODE Display Digits FLOAT 12) ;
- pour DERIVE, p est modulable comme on l'a vu.

Mais quand on veut calculer u_{25}, cette erreur est multipliée par 25! qui est un très grand nombre (soit environ $1{,}55\times10^{25}$) et les résultats obtenus peuvent devenir complètement fantaisistes pour de trop petites valeurs de n (ce qui est le cas avec toutes les calculatrices).

Pour DERIVE, les résultats sont fortement entachés d'erreurs jusqu'à $n = 26$ mais à partir de $n = 31$, les premières décimales n'évoluent plus.
C'est l'exemple typique d'un calcul instable, sensible à de toutes petites perturbations initiales (le fameux effet "papillon").

En conclusion, c'est DERIVE qui donne le bon résultat car il autorise le calcul avec plus de chiffres significatifs, ce qui permet de maîtriser l'explosion des erreurs d'arrondis. Il est tout à fait remarquable dans cette situation d'avoir affaire à une suite qui converge pour une seule valeur de son terme initial, qui plus est, une valeur *irrationnelle*, qui ne peut donc pas être représentée exactement dans une arithmérique de type finie !

Finalement, quelle réponse peut-on apporter à notre ami Albert ?
Qu'il propose à son banquier de placer 2 francs plutôt que $e - 1$ francs. Là, toutes les calculatrices et autres ordinateurs se réconcilient (et pour cause !) et donnent comme résultat après 25 ans de placement

$$4{,}369789732 \times 10^{24} \text{ francs.}$$

Si son banquier accepte, c'est la faillite de la SCB...

BIBLIOGRAPHIE :

"*Ordinateur en quête d'arithmétique*" de Jean-Michel Muller dans le numéro spécial de la Recherche sur les nombres (n°278 juillet-août 1995) ;

Les réels et la calculatrice : article du *Bulletin* l'APMEP n°339 de Daniel SAADA ;

pour approfondir, "*Arithmétique des ordinateurs*" de Jean-Michel Muller collection Etude et Recherche en Informatique (Masson 1989).