Marche aléatoire

Énoncé

Un pion est placé sur la case de départ :

				Départ				

Le lancer d'une pièce bien équilibrée détermine le déplacement du pion.
- PILE, le pion se déplace vers la droite
- FACE, le pion se déplace vers la gauche

Un trajet est une succession de 4 déplacements. On s'intéresse à l'événement A : "le pion est revenu à la case départ après 4 déplacements".

A chaque lancer, on associe le réel +1 si le résultat est PILE et -1 si le résultat est FACE.

Étude expérimentale

1. Simuler à l'aide du tableur de 200 à 2000 trajets du pion et estimer la fréquence de l'événement A. Compléter le tableau suivant :

Nombre d'essais	200	400	600	800	1000	1200	1400	1600	1800	2000
Fréquence de A										

Appeler l'examinateur pour vérifier le tableau obtenu.

Étude mathématique

2. On appelle X la variable aléatoire qui prend pour valeur la somme des quatre réels.

   (a) En précisant la méthode choisie, calculer les valeurs possibles de X et le nombre de trajets possibles.

   Appeler l'examinateur pour contrôler la réponse et lui indiquer la démarche prévue à la question suivante

   (b) Calculer la probabilité de l'événement A à l'aide d'un schéma de Bernoulli et comparer avec l'estimation obtenue.

Production demandée

– Réaliser une simulation en utilisant les fonctions appropriées.

– Donner une réponse argumentée à la question 2.

Proposition de corrigé avec le Classpad 300

Étude expérimentale

1. Rappelons que l'expression `randList(`n`,`a`,`b`)` renvoie une liste de n entiers pseudo-aléatoires de l'intervalle $[a, b]$. Ainsi `randList(4,0,1)` renvoie une liste de quatre entiers pseudo-aléatoires égaux à 0 ou à 1.

   L'expression `2randList(4,0,1)-1` renvoie donc une liste de quatre entiers pseudo-aléatoires égaux à -1 ou à 1. On a ainsi une bonne façon de représenter un “trajet”.

   Si on ne s'intéresse qu'au retour sur la case de départ, il suffit de calculer la somme de la liste précédente : le résultat est nul si et seulement si l'événement A est réalisé.

   On appréciera ici comment le Classpad accepte des opérations arithmétiques sur les listes (sur notre exemple, multiplication ou soustraction terme à terme par un coefficient).

   Venons-en à la partie “Expérimentation” de l'énoncé.

   On entre dans l'application “Tableur” (icône ).

   On vide au besoin le contenu de la feuille de travail (« `Edit/Tout effacer` »).

   Dans le menu `Edit`, on sélectionne la fonction « `Remplir échelle` » (fig1).

   Cette fonction permet de remplir les cellules d'une zone rectangulaire au moyen d'une même formule. On place ici la formule `=sum(2randList(4,0,1)-1)` (qui renvoie le numéro de la case où se retrouve le pion à la fin de son trajet) dans la zone rectangulaire `A1:A200`, c'est-à-dire sur les deux cents premières lignes de la première colonne `A` : on réalise ainsi 200 expériences indépendantes simultanément.

   On voit le résultat fig2. On peut ensuite renouveler cette simulation de 200 trajets en utilisant la fonction « `Fich/Recalculer` ».

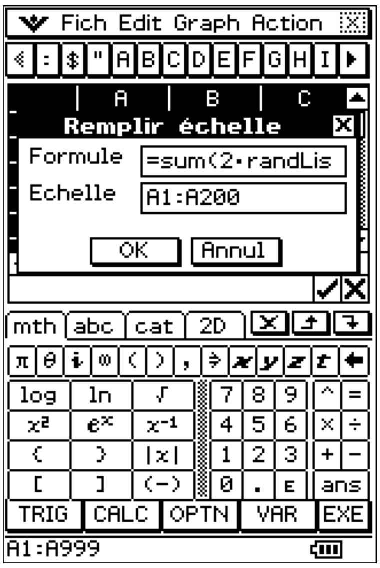

fig1 : effectuer 200 trajets

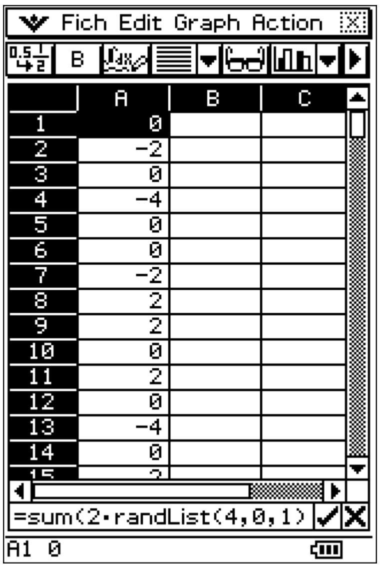

fig2 : les cases d'arrivée

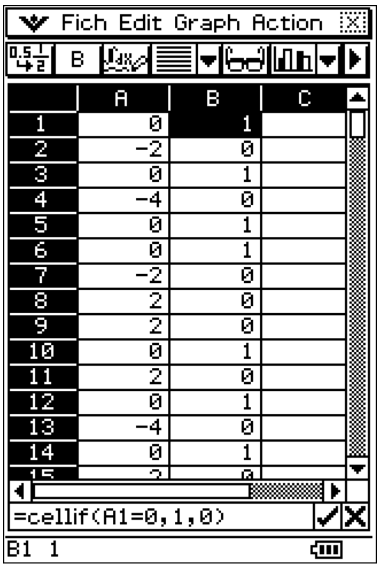

fig3 : 1 si A est réalisé

   Un rapide examen du contenu de la colonne `A` montre que le pion termine toujours son trajet sur une case de numéro pair ($-4, -2, 0, 1, 2$ ou 4). C'est normal dans la mesure où il y a un nombre pair de déplacements, que la position initiale est 0, et que chaque déplacement modifie la parité du numéro de la case sur laquelle se trouve le pion.

Puisqu'on s'intéresse à la réalisation de l'événement A, on va convertir les résultats précédents (les numéros des cases d'arrivée) en la valeur 1 si ce numéro est nul (c'est-à-dire si le pion est revenu à sa position de départ) et 0 sinon.

On place le curseur sur `B1`. Avec la fonction « `Edit/Remplir échelle` » on copie la formule `=cellif(A1=0,1,0)` dans la plage de `B1` à `B200`.

Pour calculer combien de fois le pion est revenu sur sa case de départ, il suffit d'effectuer la somme des 200 coefficients de la colonne `B`.

Pour cela, on entre la formule `=sum(B1:B200)` dans la cellule `C1`.

On effectue alors 10 rafraîchissements successifs avec « `Fich/Recalculer` ». À chaque fois, les colonnes `A` et `B` sont actualisées, de même que le contenu de la cellule `C1` (le nombre de retours à l'origine à l'issue des 200 trajets).

À chaque rafraîchissement, on recopie quelque part dans la colonne `C` (à partir de `C3` par exemple) le contenu de la cellule `C1`.

À la fin, on se retrouve avec dix résultats successifs (recopiés ici de la cellule `C3` à la cellule `C12`). Il suffit alors d'entrer la formule `=cuml(C3:C12)` dans la cellule `C13` pour y récupérer la somme cumulée des nombres de retours à l'origine (de quoi remplir le tableau demandé par l'énoncé).

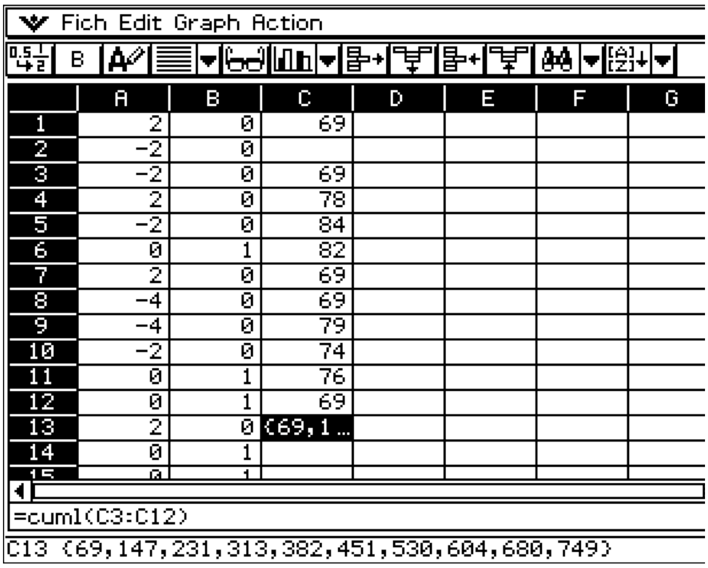

fig4 : on a effectué une simulation de 200 trajets, dix fois de suite

Voici donc comment nous remplirions le tableau avec les résultats obtenus ci-dessus :

Nombre d'essais	200	400	600	800	1000	1200	1400	1600	1800	2000
Réalisations de A	69	147	231	313	382	451	530	604	680	749
Fréquences de A	0.345	0.368	0.385	0.391	0.382	0.376	0.379	0.378	0.378	0.375

Remarque : pour obtenir les *fréquences* des réalisations de l'événement A, il suffit de placer la formule `=seq(k,k,200,2000,200)` dans la cellule `C14` (on obtient $\{200, 400, \ldots, 2000\}$), puis la formule `=C13/C14` dans la cellule `C15` (on obtient alors la liste des fréquences).

Étude mathématique

2. (a) Soit L_k la variable aléatoire égale au réel (1 ou -1) associé au k-ième lancer.

   On a $X = L_1 + L_2 + L_3 + L_4$ et les variables L_k sont mutuellement indépendantes.

   Il y a autant de trajets qu'il y a de quadruplets (L_1, L_2, L_3, L_4), c'est-à-dire $2^4 = 16$.

   La variable $Y_k = \dfrac{1}{2}(L_k + 1)$ vaut 0 si le k-ème résultat est "face" et 1 si c'est "pile".

   Les variables indépendantes Y_k suivent donc une loi de Bernoulli de paramètre $\dfrac{1}{2}$.

   Puisque $L_k = 2Y_k - 1$, on a $X = 2Z - 4$ avec $Z = Y_1 + Y_2 + Y_3 + Y_4$.

   La variable Z (somme de quatre variables indépendantes de Bernoulli de paramètre $1/2$) suit la loi binomiale de paramètres $n = 4$ et $p = 1/2$ (elle compte le nombre de fois que le pion s'est déplacé vers la droite).

   Remarque : on peut aussi introduire Z (nombre de déplacements à droite), remarquer que $4-Z$ est le nombre de déplacements à gauche, et écrire $X = Z-(4-Z) = 2Z-4$.

   Puisque $Z(\Omega) = \{0, 1, 2, 3, 4\}$, on trouve $X(\Omega) = \{-4, -2, 0, 2, 4\}$.

   (b) Comme on l'a vu précédement, on a $X = 2Z - 4$ où Z (nombre de fois que le pion s'est déplacé vers la droite) suit la loi binomiale de paramètres $n = 4$ et $p = 1/2$.

   Pour tout k de $\{0, 2, 3, 4\}$, on a $p(Z = k) = \dfrac{1}{16}\dbinom{4}{k}$.

   En particulier $p(A) = p(X = 0) = p(Z = 2) = \dfrac{1}{16}\dbinom{4}{2} = \dfrac{3}{8} = 0.375$.

   On constate que les fréquences observées dans notre feuille de calcul sont assez proches de cette valeur dès qu'on considère un "grand" nombre de trajets successifs.

Complément (étude de la loi binomiale)

On suppose que le pion effectue n trajets successifs (avec $n \geqslant 200$).

On note A_k la variable de Bernoulli prenant la valeur 1 (avec la probabilité $p = 3/8$) si le k-ème trajet se termine sur la case "départ", et la valeur 0 sinon.

Les trajets étant indépendants les uns des autres, la variable $S_n = A_1 + A_2 + \cdots + A_n$ suit une loi binomiale de paramètres n, p.

Soit $f_n = \dfrac{S_n}{n} = \dfrac{A_1 + A_2 + \cdots + A_n}{n}$ la fréquence des retours à la case "départ" après n trajets.

Ainsi, pour tout k de $\{0, 1, \ldots, n\}$, on a $p\Big(f_n = \dfrac{k}{n}\Big) = p(S_n = k) = \dbinom{n}{k}p^k(1-p)^{n-k}$.

On peut facilement visualiser la loi de la variable S_n (donc celle de la loi f_n) avec le Classpad.

Pour cela, on passe dans l'application "Statistiques" (icône 

).

On choisit l'option `Distribution` dans le menu `Calc` (fig5). Dans la liste déroulante des distributions, on choisit `DP binomiale`. On peut également cocher la case `Aide` qui donne des indications sur la marche à suivre (fig6).

L'écran suivant (taper sur `Suiv.>>`) nous permet de calculer la probabilité pour qu'une variable aléatoire satisfaisant à la loi binomiale $\mathcal{B}(n, p)$ prenne une valeur x.

Nous choisissons les valeurs $n = 200$ (nombre d'essais) $p = 3/8$ (probabilité de succès à chaque essai) et $x = 75$ (valeur attendue) qui est ici l'espérance de la loi $\mathcal{B}(200, 3/8)$ (fig7).

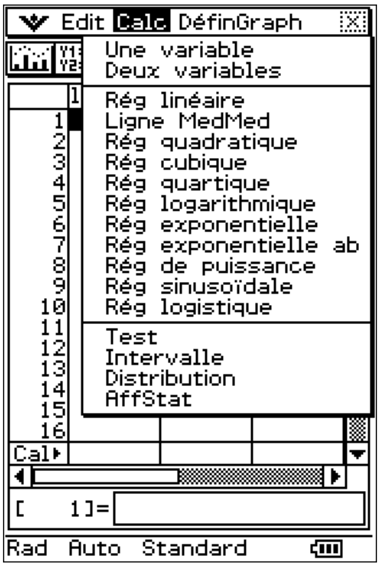

fig5 : choisir `Distribution`

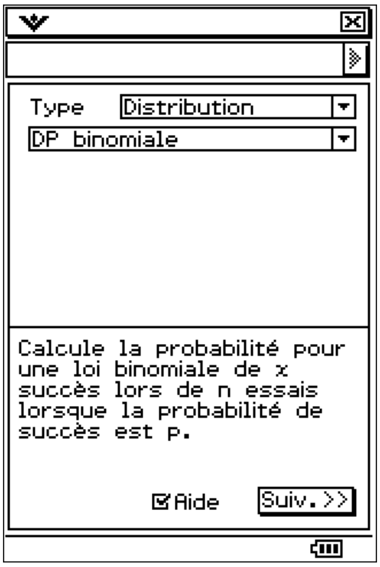

fig6 : l'aide est bienvenue

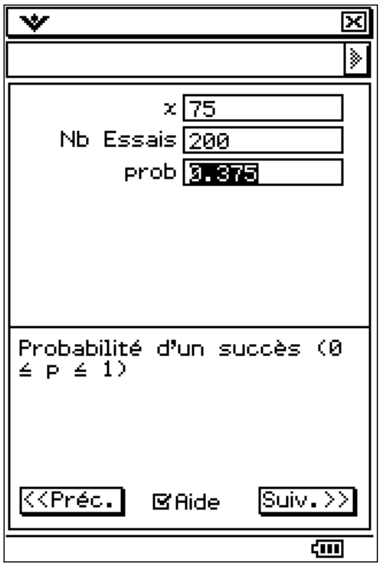

fig7 : choix des paramètres

Un appui sur Suiv.>> et nous obtenons la probabilité $p = 0.0581899$ que notre pion, après 200 trajets successifs, ait fini sa course exactement 75 fois sur la case "départ". Dans l'écran obtenu, si on tape l'icône , on obtient le tracé de la partie centrale de l'histogramme de la loi binomiale étudiée ici (fig8).

Par des appuis sur les touches de déplacement horizontal, on peut alors parcourir cet histogramme et lire les probabilités que notre variable aléatoire prenne telle ou telle valeur.

Mais revenons en arrière par des appuis répétés sur l'icône <<Préc.. Nous pouvons alors choisir `DC Binomiale` plutôt que `DP Binomiale`. Cela revient à rechercher la fonction de répartition de notre loi binomiale, donc les valeurs $p(S_{200} \leqslant k)$ plutôt que les valeurs $p(S_{200} = k)$.

Pour obtenir par exemple la probabilité que le nombre de retours à la case "départ" (toujours au bout de 200 trajets) soit inférieur ou égal à 80, on rentre les paramètres 80 (pour `x`), 200 (pour `Nb Essais`) et 3/8 (pour `prob`).

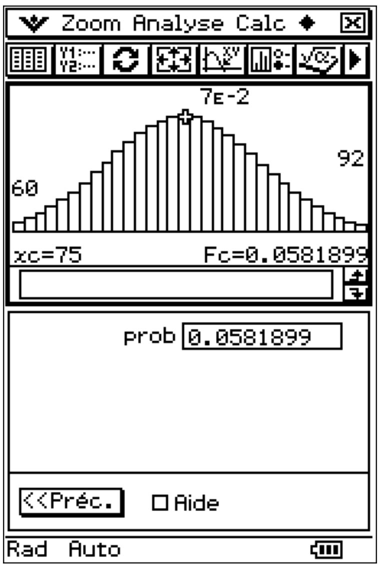

fig8 : $p(S_{200}) = 75$

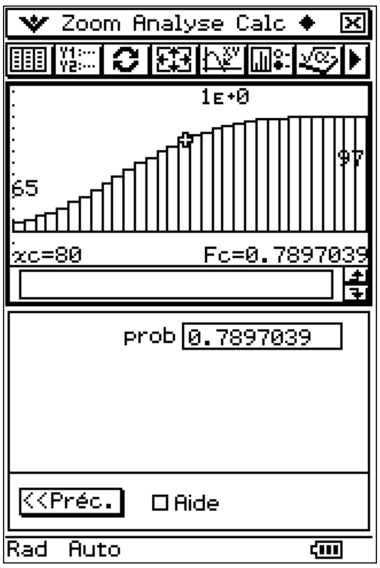

fig9 : $p(S_{200}) \leqslant 80$

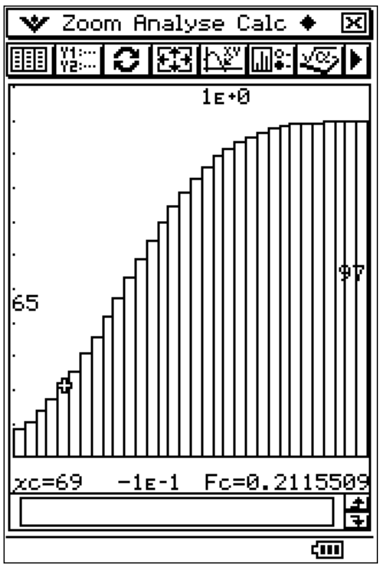

fig10 : $p(S_{200}) \leqslant 69$

On trouve alors $p(S_{200} \leqslant 80) \approx 0.7897039$, et il est possible là aussi de tracer l'histogramme des probabilités cumulées (fig9). On peut se déplacer (taper au besoin l'icône) sur la crête de cet histogramme et relever d'autres probabilités cumulées. Nous constatons par exemple que $p(S_{200} \leqslant 69) \approx 0.2115509$ (fig10, ici nous avons recadré la fenêtre de tracé avec).

Ce qui précède permet d'affirmer que :

$$p(70 \leqslant S_{200} \leqslant 80) = p(S_{200} \leqslant 80) - p(S_{200} \leqslant 69) \approx 0.578153$$

Autrement dit, il y a presque 60% de chances pour que le pion, à l'issue de ses 200 trajets, ait fini sa course sur la case "départ" un nombre de fois compris entre 70 et 80, c'est-à-dire pour que la fréquence de ces retours au départ soit comprise entre 0.35 et 0.4.

Rappelons que pour $n = 200$ notre simulation dans l'application Tableur a renvoyé 0.345 (nous étions donc dans les 40% restants...)

Complément (approximation par la loi normale)

Telle qu'elle apparait (fig8), la loi de S_{200}, c'est-à-dire la loi binomiale $\mathcal{B}(n,p)$ avec $n = 200$ et $p = 3/8$, évoque indéniablement la fonction densité d'une loi normale.

Rien d'étonnant à celà quand on sait que la loi $\mathcal{B}(n,p)$ peut être approchée par la loi normale de même espérance et de même écart-type, c'est-à-dire la loi $\mathcal{N}(np, \sqrt{npq})$ (où on pose $q = 1 - p$) quand les conditions suivantes sont réalisées : $n > 30$, $np \geqslant 15$, $nq \geqslant 15$.

Or ici $n = 200$ (ou davantage), $np = 75$, $nq = 125$: l'approximation est légitime (et elle le sera d'autant plus que n sera supérieur à 200).

Rappelons que $p(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}$ pour tout k de $\{0, 1, \ldots, n\}$.

Le calcul d'une telle probabilité devient très problématique pour les grandes valeurs de n (les coefficients binomiaux deviennent en effet très élevés, les coefficients $p^k(1-p)^{n-k}$ très petits, et leur produit, s'il ne provoque pas un débordement des capacités de la machine, conduit à des erreurs d'arrondi qui rendent le résultat douteux). On pourrait d'ailleurs constater que les calculs précédents (fig8 à fig10) ne sont plus possibles avec le Classpad dès que $n \geqslant 400$.

Dans la suite, on notera S'_n une variable suivant la loi normale $\mathcal{N}(np, \sqrt{npq})$, c'est-à-dire qui réalise une approximation de S_n.

On rappelle qu'en raison du caractère continu de S'_n, il faut opérer une *correction de continuité* en écrivant : $p(S_n = k) \approx p(k - 0.5 \leqslant S'_n \leqslant k + 0.5)$ pour tout entier k de $\{0, \ldots, n\}$.

Plus généralement, pour tous j, k de $\{0, \ldots, n\}$: $p(j \leqslant S_n \leqslant k) \approx p(j - 0.5 \leqslant S'_n \leqslant k + 0.5)$.

Nous allons maintenant voir comment appliquer cette approximation dans le cas de $n = 200$ trajets (on espère donc retrouver les résultats obtenus avec S_{200}). On rappelle que l'espérance de S_{200} ; est $\mu = np = 75$, et son écart-type est $\sigma = \sqrt{npq} = 5\sqrt{30}/4 \approx 6.846531969$.

On va vérifier la qualité de l'approximation $p(70 \leqslant S_{200} \leqslant 80) \approx p(69.5 \leqslant S'_{200} \leqslant 80.5)$.

On va donc dans l'application "Statistiques" (icône Statistiques), où on choisit l'option `Distribution` dans le menu `Calc`. On choisit `DC normale` pour étudier la fonction de répartition (la densité serait obtenue par le choix `DP normale`). On choisit la valeur inférieure 69.5, la valeur supérieure 80.5, l'écart-type $\sigma = 5\sqrt{30}/4$ et l'espérance $\mu = 75$ (fig11).

On constate (fig12) que $p(69.5 \leqslant S'_{200} \leqslant 80.5) \approx 0.5782139$, ce qui est tout de même très proche du résultat $p(70 \leqslant S_{200} \leqslant 80) \approx 0.578153$ trouvé précédemment.

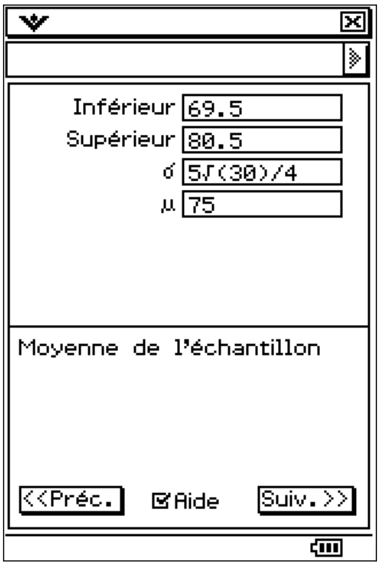

fig11 :

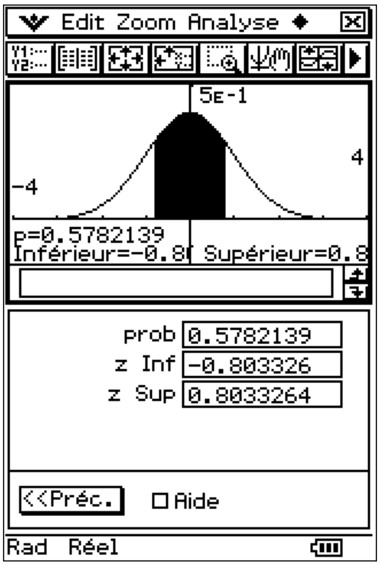

fig12 :

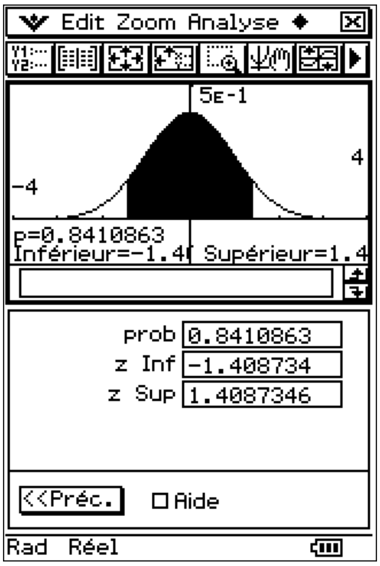

fig13 :

On voit (fig13) le cas de $n = 2000$ trajets (donc $\mu = np = 750$ et $\sigma = \sqrt{npq} = 25\sqrt{3}/2$) :

On a vérifié l'approximation $p(720 \leqslant S_{2000} \leqslant 780) \approx p(719.5 \leqslant S'_{2000} \leqslant 780.5) \approx 0.8410863$.

Autrement dit, après 2000 trajets, il y a environ 85% de chances pour que la fréquence des retours à la case "départ" soit comprise entre $\dfrac{720}{2000} = 0.36$ et $\dfrac{780}{2000} = 0.39$.