GeoGebra et le calcul des probabilités

Valérie Henry

Mots clés : GeoGebra, probabilités, binomiale, normale, outils d'évaluation.

Résumé. *Depuis la version 4, GeoGebra propose une interface Calculs de probabilités accessible à l'ouverture du logiciel ou via le menu* `Affichage`*. Dans cet article, nous invitons le lecteur à une découverte de cette interface au travers de la résolution de deux problèmes issus des « Outils d'évaluation inter-réseaux ».*

Introduction

L'objectif de cet article est d'initier le lecteur aux fonctionnalités de l'interface *Calculs de probabilités*. Nous avons choisi pour cela de nous appuyer sur deux activités connues des enseignants puisqu'elles sont issues des outils d'évaluation inter-réseaux ([1]). Comme leur nom l'indique, ces problèmes ne sont *a priori* pas destinés à introduire les concepts mathématiques en jeu mais bien à vérifier leur acquisition. Dans cet article, nous ne discutons ni leur pertinence ni leur formulation dans quelque cadre que ce soit. Précisons également que nous ne développerons pas en détails les éléments de calcul des probabilités utilisés pour résoudre les exercices car ce n'est pas l'objet de l'article. Enfin, signalons que la lecture de cet article sera grandement facilitée si le lecteur effectue en parallèle les manipulations sur le logiciel.

1. L'interface

À l'ouverture de l'interface *Calcul des probabilités*, on obtient la fenêtre de la figure 1. Examinons quelque peu celle-ci.

Ce qui saute aux yeux de prime abord, c'est évidemment le graphique qu'on reconnait rapidement comme une courbe de Gauss sous laquelle une partie du plan a été mise en évidence. On constate que, sur les abscisses délimitant cette partie du plan, se trouvent des curseurs qui peuvent être déplacés et, en y regardant de près, on s'aperçoit que certaines valeurs du bas de la fenêtre s'adaptent au déplacement du ou des curseurs. Ainsi la ligne

$$P(x_1 \leqslant X \leqslant x_2) = p$$

se modifie en fonctions des valeurs x_1 et x_2 des curseurs et fournit une valeur p qui n'est autre que l'aire de la partie du plan sélectionnée ou encore la probabilité que les valeurs d'une variable aléatoire suivant une loi normale de moyenne $\mu = 0$ et d'écart-type $\sigma = 1$ ($X \sim N(0,1)$) soient comprises entre x_1 et x_2.

([1]) Disponibles sur `http://www.enseignement.be/index.php?page=24393&navi=2135`.

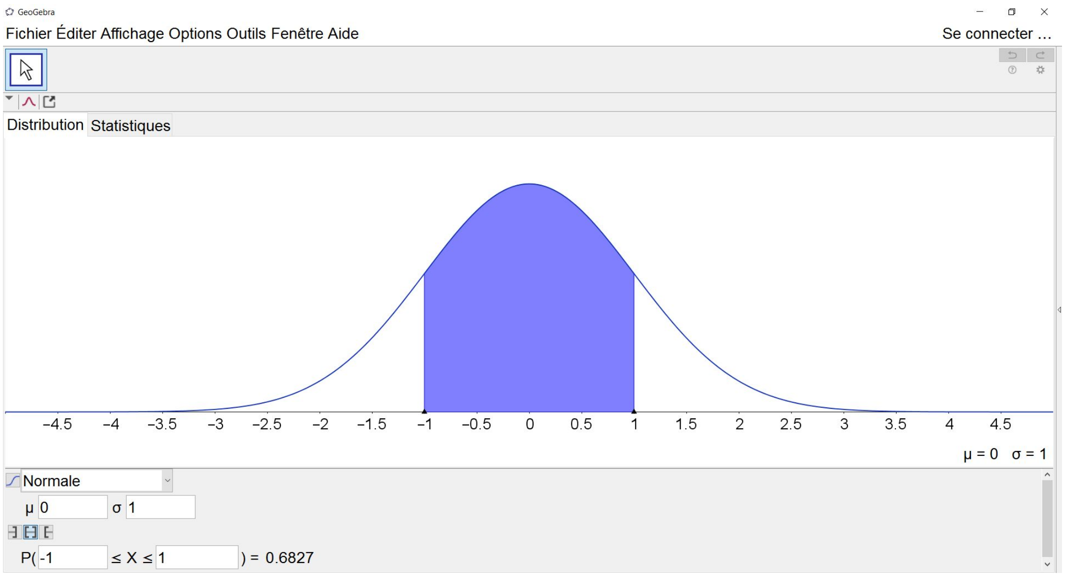


Fig. 1 : Interface « Calcul des probabilités »

Cela nous amène à considérer les autres informations présentes sous le graphique. L'indication `Normale` se trouve dans un menu déroulant qui, si on l'ouvre, dévoile toute une série de lois de probabilité dont `Binomiale` et `Poisson` que nous épinglons puisqu'elles font partie des programmes de l'enseignement secondaire. La sélection de `Binomiale` modifie considérablement l'interface (voir figure 2) qui s'adapte aux caractéristiques de la loi binomiale :

- les paramètres sont devenus n (nombre d'épreuves) et p (probabilité de succès) ;
- le graphique a été modifié ;
- les valeurs de la distribution sont affichées dans un tableau à la droite du graphique.

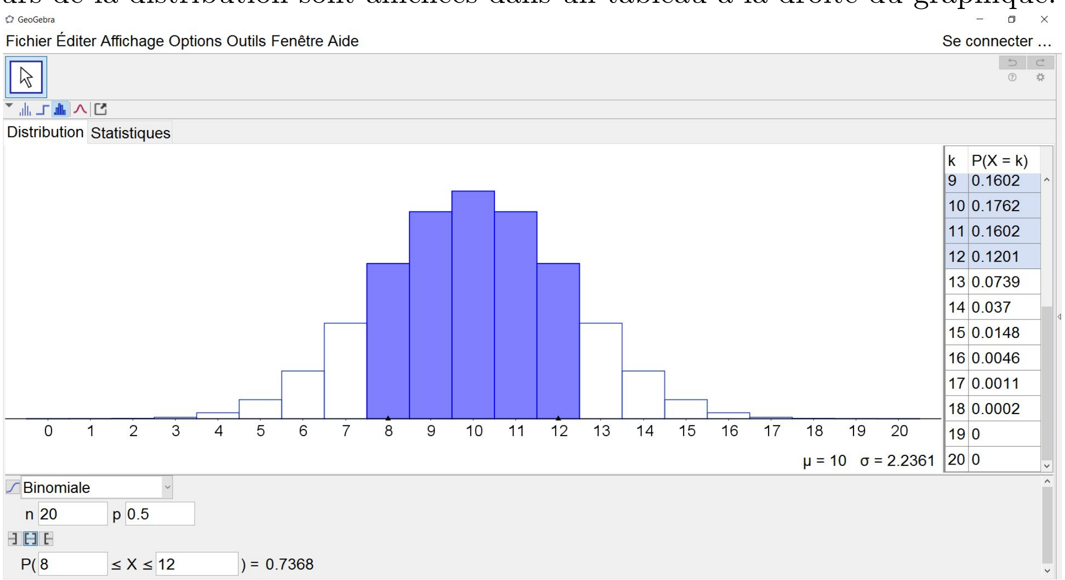
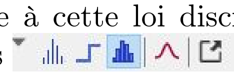


Fig. 2 : Loi binomiale

On est en droit de trouver la représentation en histogramme peu adaptée à cette loi discrète. Heureusement, il suffit de repérer, en haut du graphique, la liste de boutons  qui permet de choisir le type de graphique à utiliser : ici un diagramme en bâtonnets que l'on obtient en cliquant sur le premier bouton (voir figure 3).

Vous l'aurez compris rapidement, les paramètres de la loi sélectionnée peuvent être modifiés. Pour la binomiale, on voit que, par défaut, GeoGebra choisit $n = 20$ et $p = 0{,}5$, soit une loi symétrique.

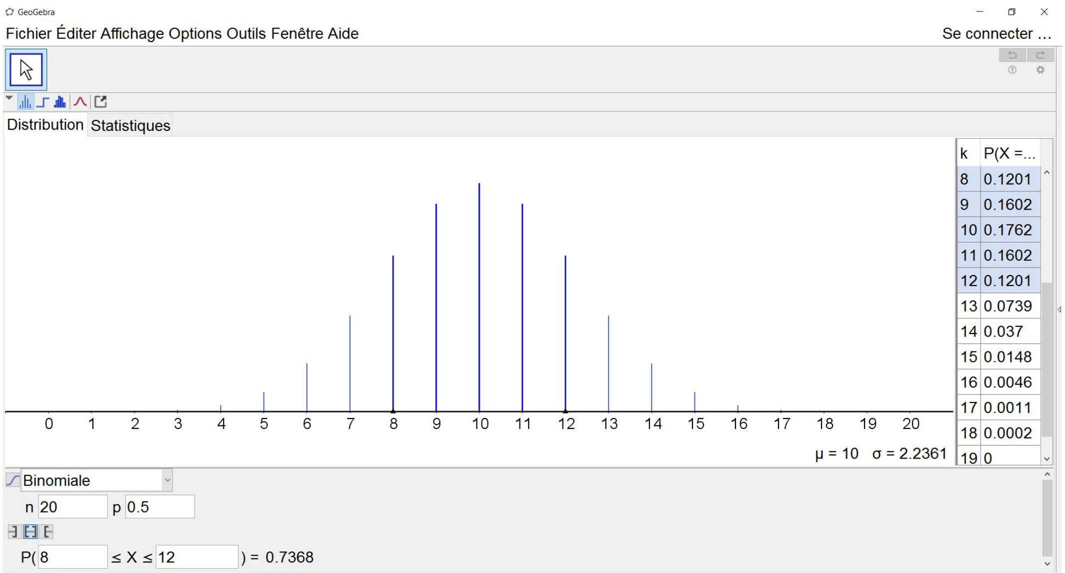


Fig. 3 : Diagramme en bâtonnets pour la loi binomiale

2. La loi binomiale et les tirs au but

Illustrons tout ceci avec un premier exercice tiré des outils d'évaluation inter-réseaux :

En étudiant les statistiques des tirs aux buts des joueurs de l'équipe première, l'entraîneur constate que, sur une série de 5 tirs, un des joueurs marque avec les probabilités suivantes :

Nombre de tirs réussis	5	4	3
Probabilité de réussite	0,2	0,5	0,3

Tout entraînement se termine par une épreuve de tirs au but. Celle-ci consiste en 2 séries de 5 tirs et l'entraîneur considère qu'elle est réussie si le joueur marque au moins 8 buts. Calculer la probabilité pour que, sur dix séances d'entraînement, ce joueur ne réussisse aucune épreuve. Commenter.

Plusieurs variables aléatoires peuvent être définies pour modéliser cette situation. Premièrement, on peut définir T comme le nombre de tirs réussis par épreuve. Les valeurs possibles pour T sont $\{6, \ldots, 10\}$ puisque le joueur exécute deux séries de 5 tirs et qu'il en réussit au moins 3 à chaque série (2). On a

$$P(T = 6) = 0{,}3 \times 0{,}3 = 0{,}09$$

puisqu'il réussira 6 tirs en tout uniquement s'il en réussit 3 à la première série et 3 à la deuxième série. De même,

$$P(T = 7) = 0{,}3 \times 0{,}5 + 0{,}3 \times 0{,}5 = 0{,}3$$

(2) Pour autant qu'on accepte de considérer que la loi de probabilité de l'énoncé est valable pour chaque série de tirs.

puisque, pour réussir 7 tirs en tout, il doit **soit** en réussir 4 à la première série **et** 3 à la deuxième **soit** le contraire. En continuant le raisonnement, on construit la loi de probabilité suivante pour T :

k	5	6	7	8	9	10
$p_k = P(T = k)$	0	0,09	0,3	0,37	0,2	0,04

Puisqu'une épreuve est considérée comme réussie lorsque le joueur marque au moins 8 buts, la probabilité de réussir une épreuve est $p = 0{,}37 + 0{,}2 + 0{,}04 = 0{,}61$. Chaque épreuve étant considérée comme indépendante des précédentes et les probabilités supposées constantes au cours des entraînements, la variable X définie comme le nombre d'épreuves réussies sur 10 entraînements suit une loi binomiale de paramètres $n = 10$ et $p = 0{,}61$, ce qui se note $X \sim B(10; 0{,}61)$ et modélise la situation. Telle qu'elle est posée la question ne nécessite cependant pas de connaissances sur la loi de X puisque la réponse peut être obtenue grâce aux propriétés de base des probabilités : la probabilité de ne réussir aucune épreuve est égale à $0{,}39^{10}$.

S'intéresser à la probabilité de réussir deux épreuves, par exemple, conduit tout naturellement à examiner plus globalement la loi binomiale $B(n, p)$. Lorsque les paramètres n et p valent respectivement 10 et 0,61 comme dans notre exemple, la distribution de probabilité s'écrit

k	0	1	2	3	4	5	6	7	8	9	10
p_k	$\simeq 0$	$\simeq 0$	0,01	0,04	0,1	0,19	0,25	0,22	0,13	0,05	0,01

et son illustration est obtenue aisément dans GeoGebra, comme à la figure 4, si on configure les paramètres de la loi avec les valeurs de notre exemple.

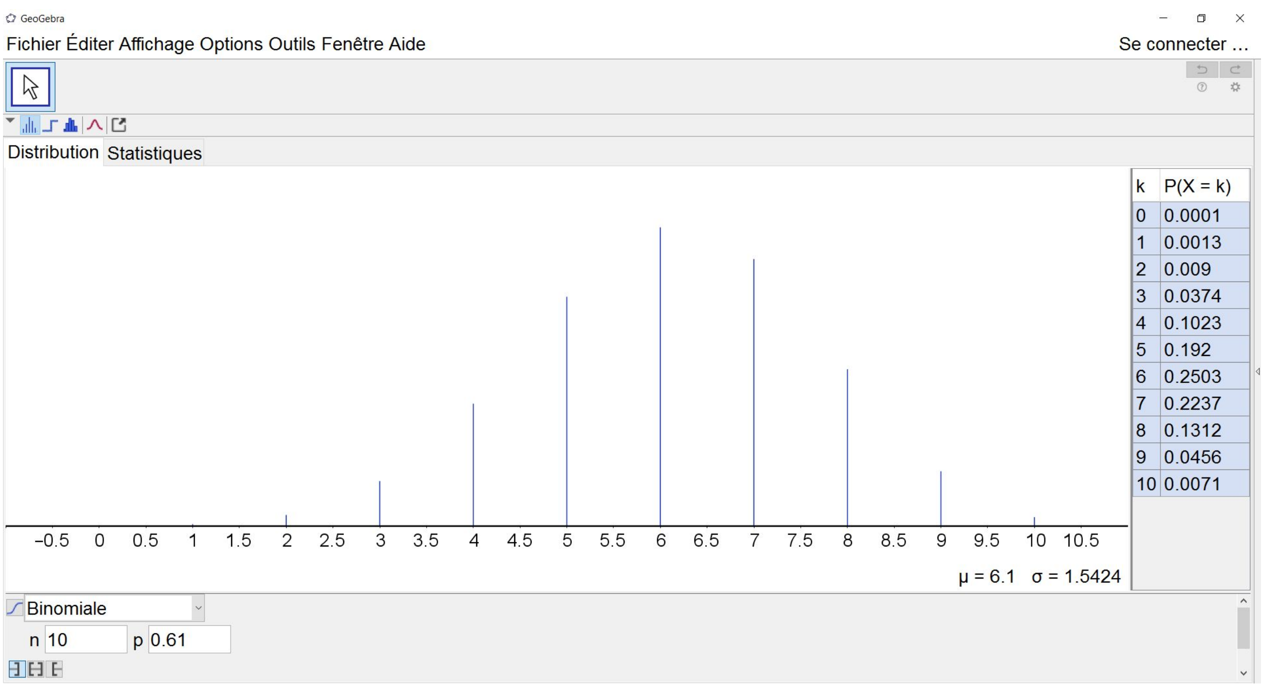


Fig. 4 : $X \sim B(10; 0{,}61)$

Le petit tableau à droite du graphique nous fournit automatiquement la réponse à la question précédente où on peut lire que $P(X = 2) = 0.009$. Cette information peut également être obtenue via la dernière ligne de l'écran (voir figure 5) puisque les cases blanches sont des cases modifiables.

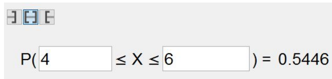


Fig. 5 : Ligne de calcul d'une probabilité

Pour obtenir $P(X = 2)$, on remplacera les deux nombres proposés par 2 puisque le logiciel ne propose pas l'égalité. On remarque encore trois petits boutons cliquables juste au-dessus de la ligne. Ceux-ci permettent d'obtenir une probabilité du type $P(X \leqslant x)$ ou $P(X \geqslant x)$. Notons que le premier type peut être illustré graphiquement via la fonction de répartition (voir figure 6) qui s'obtient en cliquant à gauche du nom de la loi sélectionnée Binomiale.

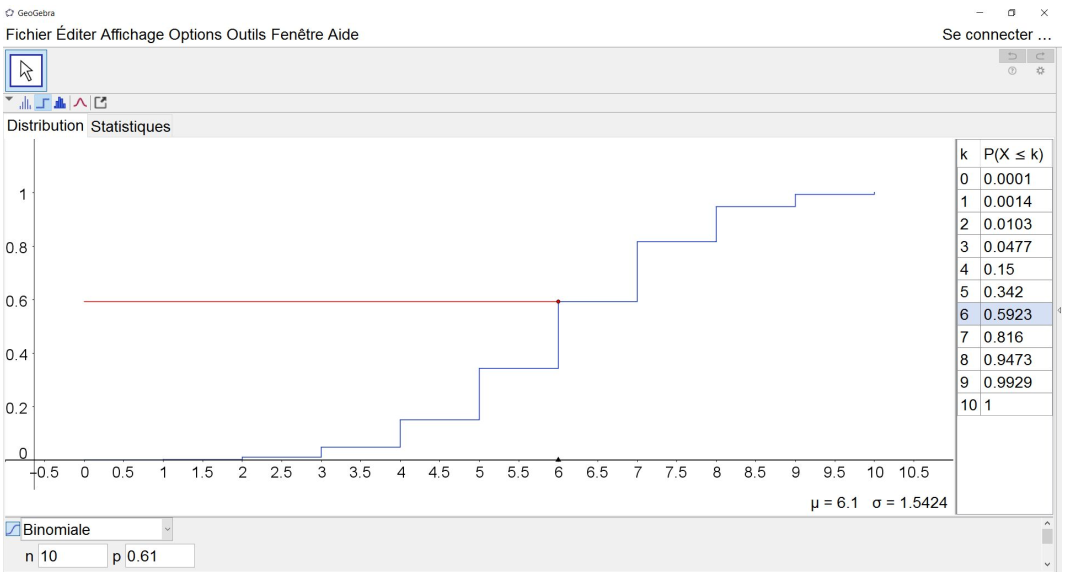


Fig. 6 : Fonction de répartition

On observe que le tableau à droite du graphique s'est adapté pour fournir les valeurs de la fonction de répartition au lieu de celles de la distribution de probabilité.

Venons-en maintenant à la dernière partie du problème proposé qui s'écrit comme suit.

Quel est le nombre minimal d'entraînements auquel doit participer ce joueur pour que sa probabilité de réussir au moins une épreuve soit supérieure à 0,99 ?

Bien sûr, la réponse s'obtient de manière classique en résolvant l'inéquation exponentielle suivante : $1 - (0,39)^n \geqslant 0,99$ par rapport à n (entier) mais l'interface peut utilement aider à se représenter ce que l'on cherche puisqu'elle permet de modifier les différents paramètres et d'observer les effets sur la loi de probabilité.

On s'intéresse donc à la probabilité $P(X \geqslant 1)$. On modifie alors la dernière ligne de l'écran en conséquence (attention, il faut revenir au graphique « distribution de probabilité » et plus « fonction de répartition ») comme à la figure 7.

On retrouve l'information déjà obtenue dans la première partie. Il reste alors à diminuer la valeur de n tant que la contrainte reste vérifiée. On observe que pour $n = 5$, la probabilité est de 0,991 alors qu'elle vaut 0,9769 pour $n = 4$. Le nombre minimal d'entraînements est donc de 5.

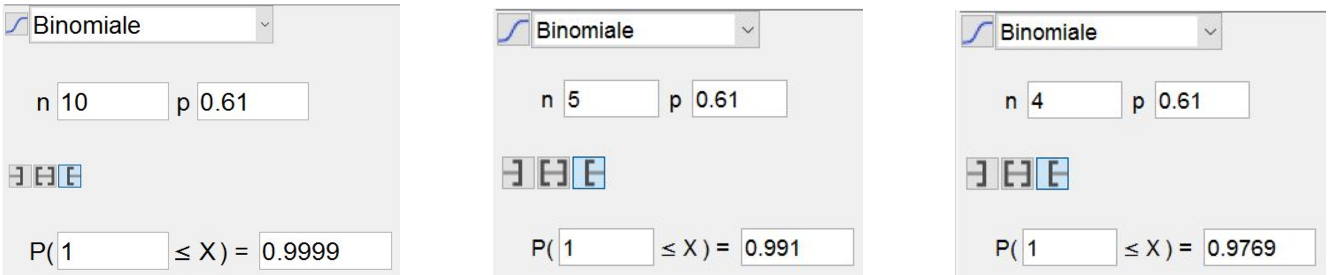
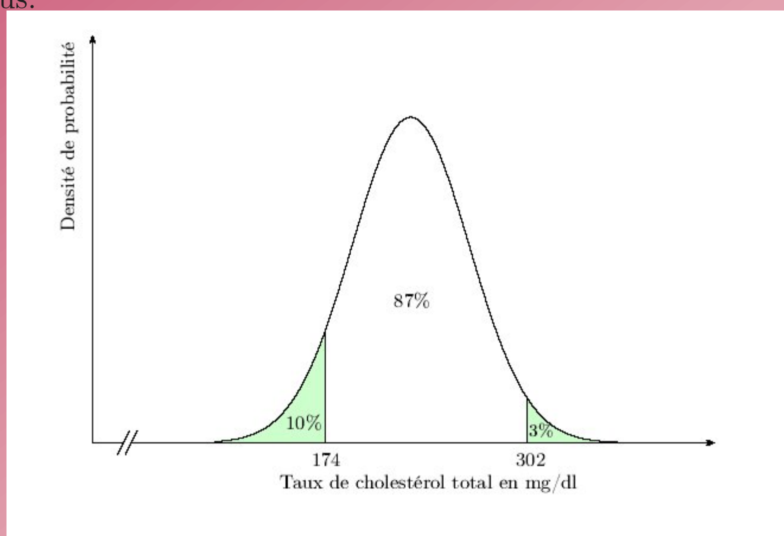


Fig. 7 : $P(X \geqslant 1)$ pour différentes valeurs de n

3. La loi normale et le taux de cholestérol

Revenons à la loi normale et intéressons-nous à un autre outil d'évaluation :

Un magazine médical présente le graphique suivant donnant la répartition du taux de cholestérol total dans une population comportant un grand nombre d'individus.



Les personnes dont le taux de cholestérol total est supérieur à 190 mg/dl doivent subir un examen complémentaire. Pour un échantillon extrait de la population étudiée, on constate que 770 personnes sur 1000 sont dans le cas. Cet échantillon est-il représentatif de cette population ?

Le problème proposé est la version « Mathématiques pour scientifiques » mais des versions adaptées existent pour les autres filières.

Tel qu'il est énoncé, ce problème n'est pas résoluble avec les ressources des compétences terminales puisqu'il requiert un minimum de notions sur la théorie de l'échantillonnage et sur les intervalles de confiance. Par contre, la question qui vient naturellement à l'esprit est « Quelle est la probabilité de devoir subir un examen complémentaire ? », ce qui revient à calculer $P(X \geqslant 190)$, si on pose X le taux de cholestérol.

À nouveau, la résolution « papier-crayon » de ce problème est parfaitement classique mais tentons d'utiliser l'interface proposée par GeoGebra pour mieux comprendre ce que l'on cherche. Revenons donc à la loi normale. Par défaut, c'est la loi normale centrée réduite qui s'affiche mais nous savons maintenant comment modifier les paramètres. L'idée est d'essayer de reproduire le graphique de l'énoncé et ainsi d'approcher les valeurs de μ et σ. L'observation du graphique de l'énoncé permet d'imaginer que la moyenne est d'au moins 200. On peut alors faire divers essais pour observer comment le graphique se modifie. On se rend compte rapidement que la valeur de l'écart-type est beaucoup trop faible par rapport à la situation de notre problème. Les figures 8, 9 et 10 fournissent quelques possibilités de paramétrisation de l'interface avec pour objectif de se rapprocher des valeurs possibles de μ et σ.

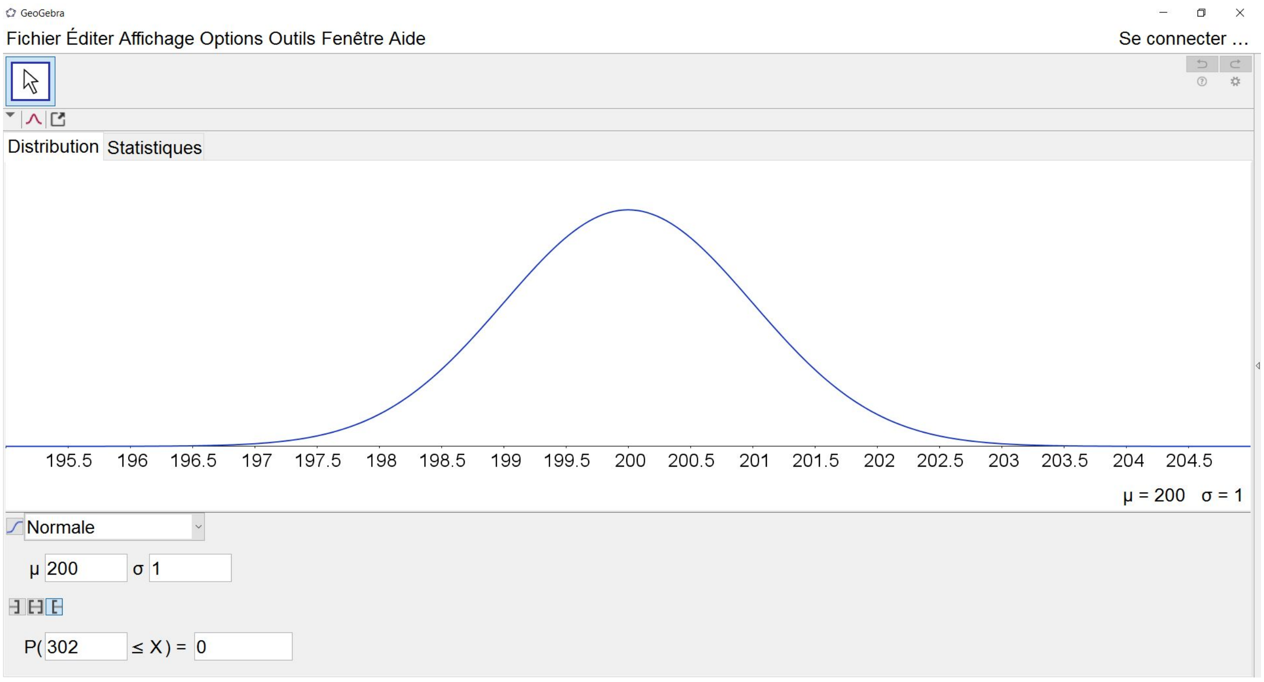


Fig. 8 : Pour $\mu = 200$ et $\sigma = 1$, $P(X \geqslant 302) = 0$

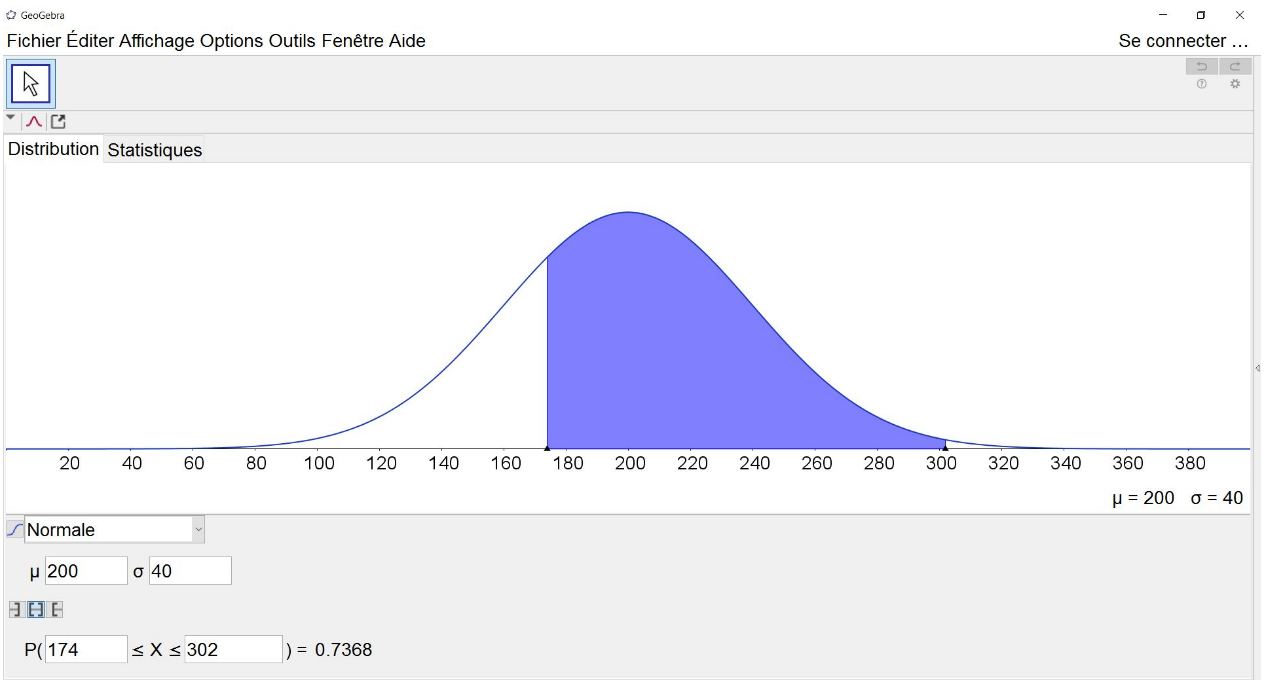


Fig. 9 : Pour $\mu = 200$ et $\sigma = 40$, $P(174 \leqslant X \leqslant 302) = 0{,}7368$

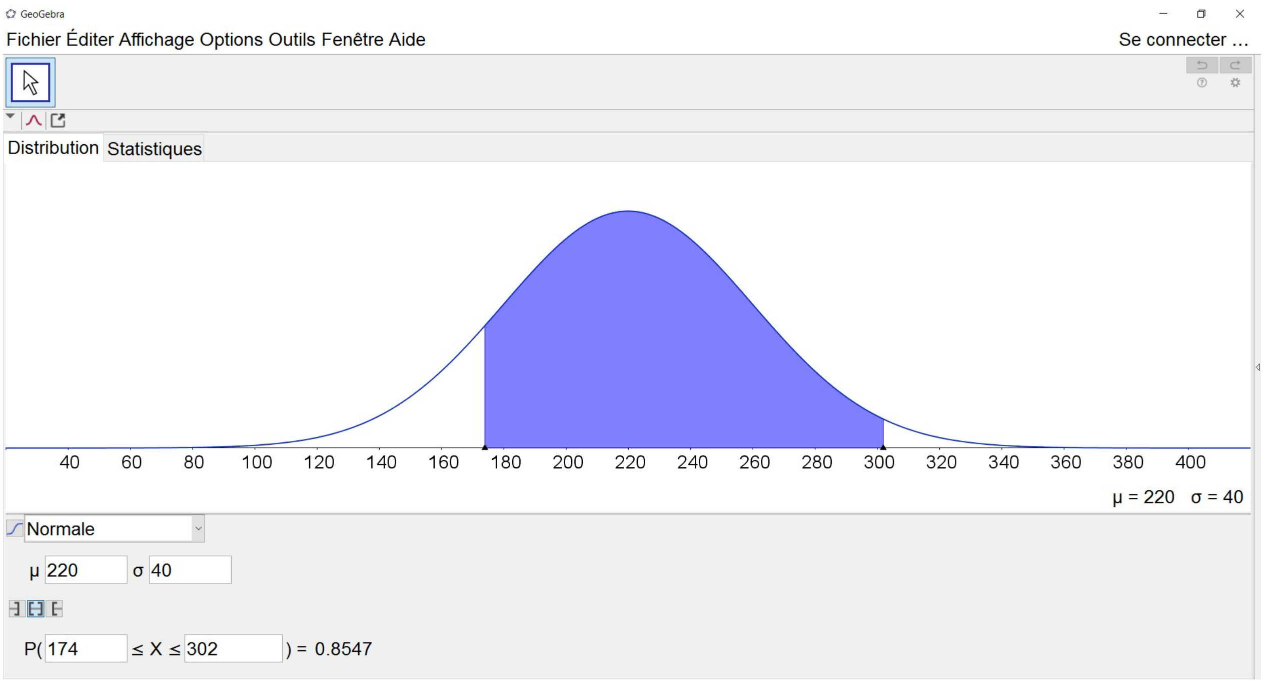


Fig. 10 : Pour $\mu = 220$ et $\sigma = 40$, $P(174 \leqslant X \leqslant 302) = 0{,}8547$

On se rend compte que ce tâtonnement peut se prolonger très longtemps (contrairement au problème des tirs au but) puisque d'une part, les valeurs cherchées ne sont pas nécessairement entières mais surtout parce que deux paramètres interviennent simultanément. Au-delà du problème en lui-même, cette recherche pousse, nous semble-t-il, à investiguer intelligemment l'impact des valeurs de la moyenne et de l'écart-type sur la distribution.

Nous pouvons maintenant nous tourner vers les outils du calcul des probabilités pour obtenir les valeurs cherchées. Ainsi, on sait que $X \sim N(\mu, \sigma)$. Les informations du graphique fournissent le système suivant :

$$\begin{cases} P(X \leqslant 174) & = & 0{,}1 \\ P(X \geqslant 302) & = & 0{,}03 \end{cases}$$

On pose alors, de façon classique, $Z = \dfrac{X - \mu}{\sigma} \sim N(0,1)$ et le système devient

$$\begin{cases} P(Z \leqslant \dfrac{174 - \mu}{\sigma}) & = & 0{,}1 \\ P(Z \geqslant \dfrac{302 - \mu}{\sigma}) & = & 0{,}03 \end{cases}$$

Le retour à l'interface de GeoGebra (pour une $N(0,1)$) permet d'obtenir le système suivant

$$\begin{cases} \dfrac{174 - \mu}{\sigma} & = & -1{,}2816 \\ \dfrac{302 - \mu}{\sigma} & = & 1{,}8808 \end{cases}$$

en imposant cette fois, dans la dernière ligne de l'écran, la probabilité à obtenir (voir figures 11 et 12).

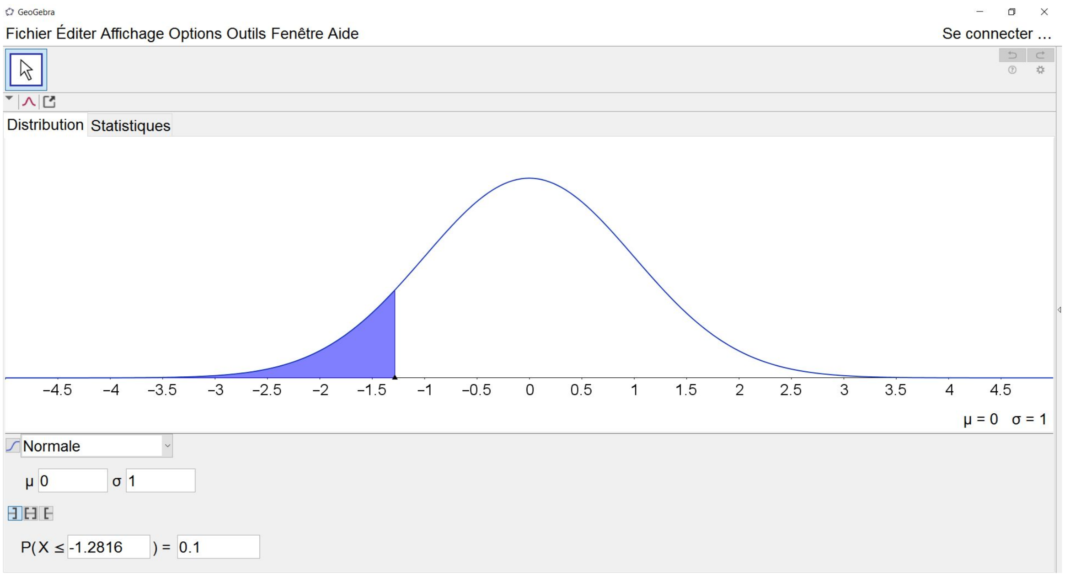


Fig. 11 : $N(0,1)$ et $p = 0{,}1$

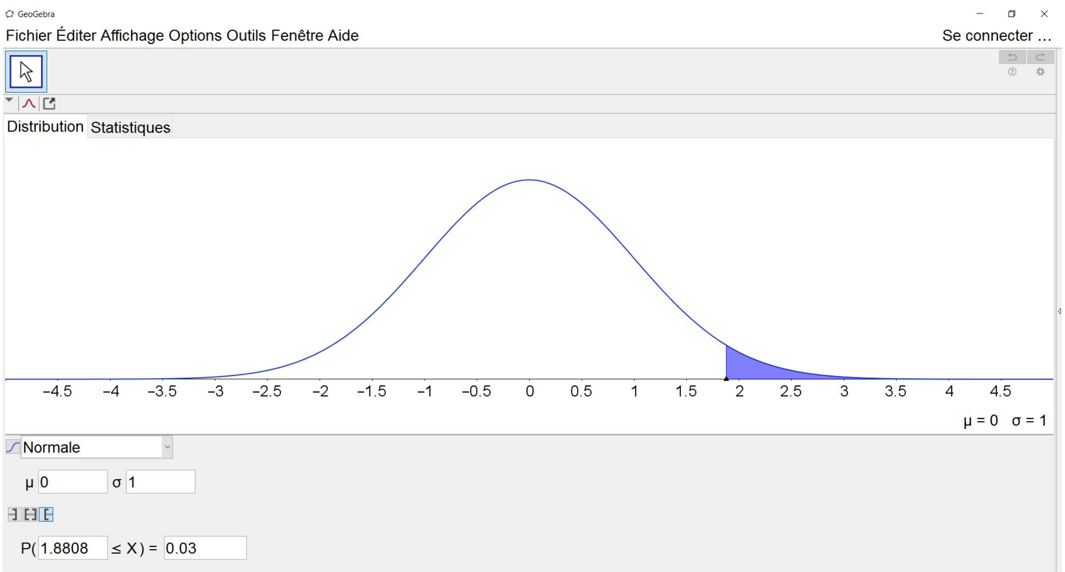


Fig. 12 : $N(0,1)$ et $p = 0{,}03$

La résolution du système se fait aisément à la main ([3]) et fournit comme solution $\mu = 225{,}87$ et $\sigma = 40{,}48$. On ne manque pas d'utiliser ces valeurs dans l'interface de GeoGebra et de vérifier qu'elles sont bien cohérentes avec les informations du problème (voir figure 13).

Avec ces paramètres pour la loi normale, on obtient immédiatement que $P(X \geqslant 190) = 0{,}8122$.

Terminons en signalant qu'il nous semble impossible d'interpréter correctement, avec les outils du programme de l'enseignement secondaire, l'écart entre la probabilité théorique ($p = 0{,}8122$) et la fréquence observée sur l'échantillon ($f = 0{,}77$) et donc de statuer sur la représentativité de l'échantillon.

([3]) L'interface de calcul formel de GeoGebra peut aussi être utilisée pour résoudre le système mais nous ne le développons pas ici.

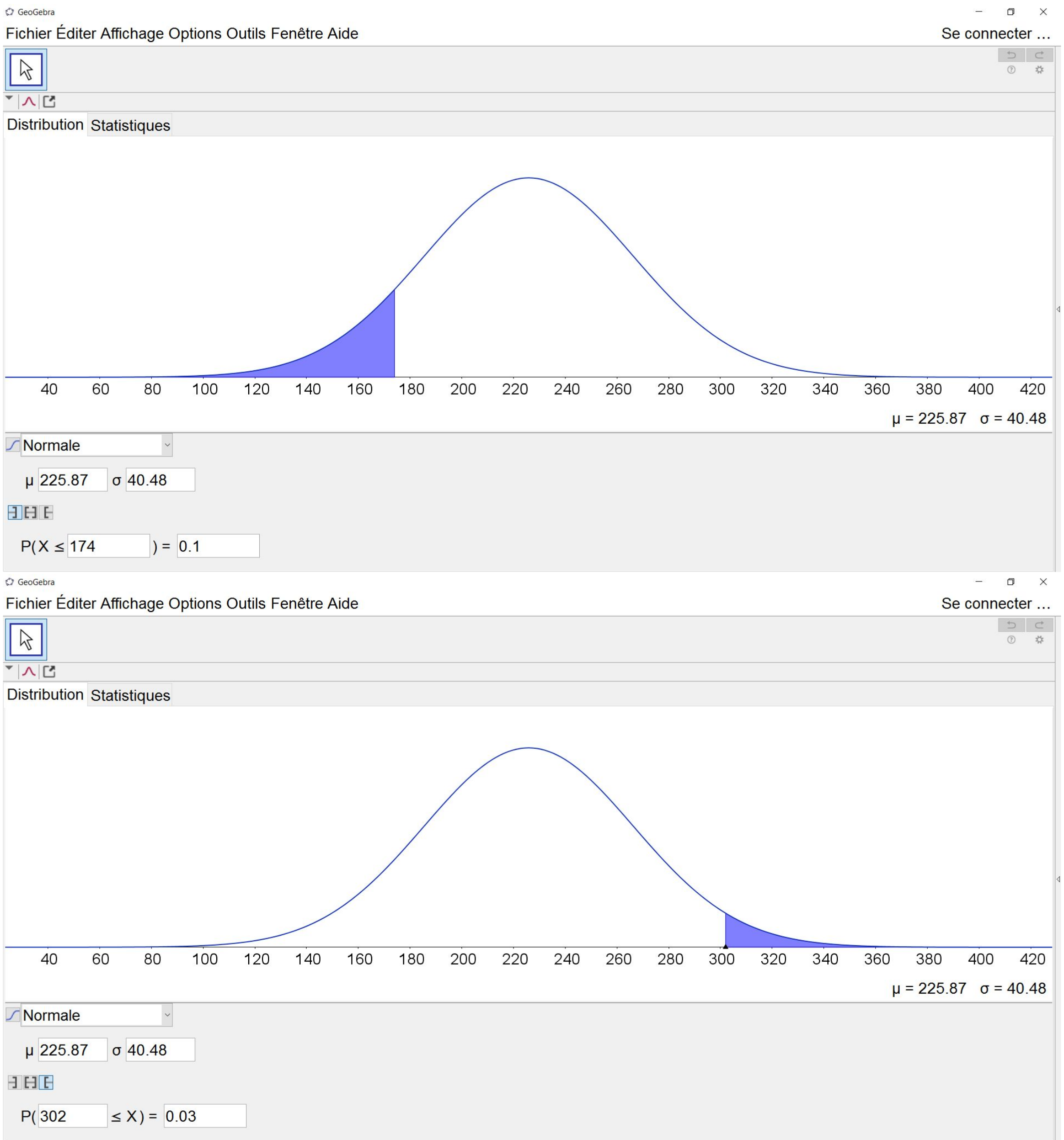


Fig. 13 : Vérification des valeurs obtenues pour μ et σ

4. Conclusions

Avec cet article, nous souhaitions simplement initier le lecteur à un outil de GeoGebra qui est peu utilisé et nous avons pensé que l'illustrer au travers de deux exemples avait plus de sens que de le décrire. Cela a conduit à certains choix, en particulier, certaines fonctionnalités n'ont pas été décrites mais nous pensons qu'une fois familiarisé avec l'interface, l'enseignant curieux sera à même de tester lui-même d'autres potentialités du logiciel.

Cet article ne constitue certainement pas une proposition d'activités pour la classe mais bien des situations de découverte du logiciel pour l'enseignant qui maîtrise déjà les notions mathématiques sous-jacentes. Même si nous pensons que les idées développées peuvent être réinvesties en classe, un travail didactique important est encore à fournir pour en faire des séquences de cours, travail qui dépend des connaissances des élèves, du matériel disponible, . . . , autant d'éléments fondamentaux qui n'ont pas été pris en compte ici.

Valérie Henry est présidente de la SBPMef. ✉ losanges@sbpm.be