De la loi binomiale à la loi Normale

Dans une population, la proportion des personnes possédant le gène A actif est $p = 0{,}4$. On prélève au hasard un échantillon de taille n dans cette population (celle-ci étant assez grande pour considérer qu'il s'agit de tirages avec remise).

A. Histogramme normalisé de la loi binomiale centrée réduite

On désigne par X_n la variable aléatoire associant à chaque échantillon de taille n le nombre de personnes de l'échantillon possédant le gène A actif.

1. a. Justifier que X_n suit une loi binomiale dont on donnera les paramètres.

b. Déterminer, en fonction de n, l'espérance m et l'écart-type σ de X_n.

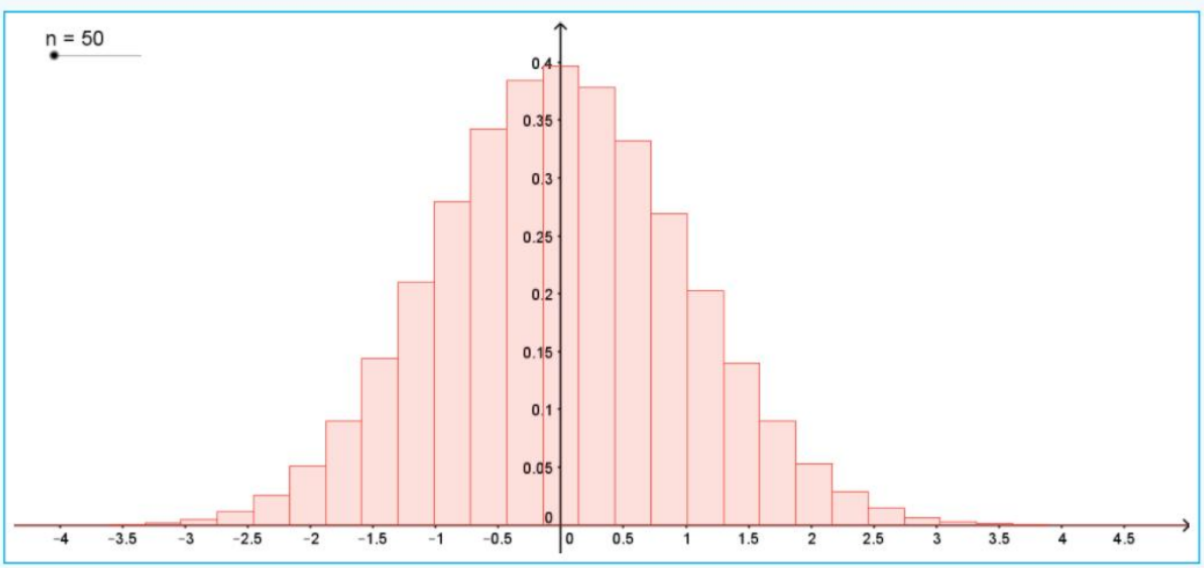
c. On considère l'événement E_n : « $m - \sigma \leqslant X_n \leqslant m + 2\sigma$ ».
Calculer $P(\mathrm{E}_n)$ pour $n = 50$ et $n = 200$.

2. Soit $Z_n = \dfrac{X_n - m}{\sigma}$, la variable aléatoire centrée réduite correspondant à X_n.

a. Écrire, à l'aide de Z_n, l'événement E_n.

b. De quelle forme sont les valeurs z_k, pour k entier allant de 0 à n, prises par la variable aléatoire Z_n ? Quel est l'écart entre deux valeurs consécutives z_k et z_{k+1} ?

3. On représente sur Geogebra (ici pour $n = 50$) l'histogramme dit « normalisé » de Z_n, c'est-à-dire tel que, pour chaque entier k compris entre 0 et n, l'aire du rectangle R_k centré sur z_k est égale à $P(Z_n = z_k)$.



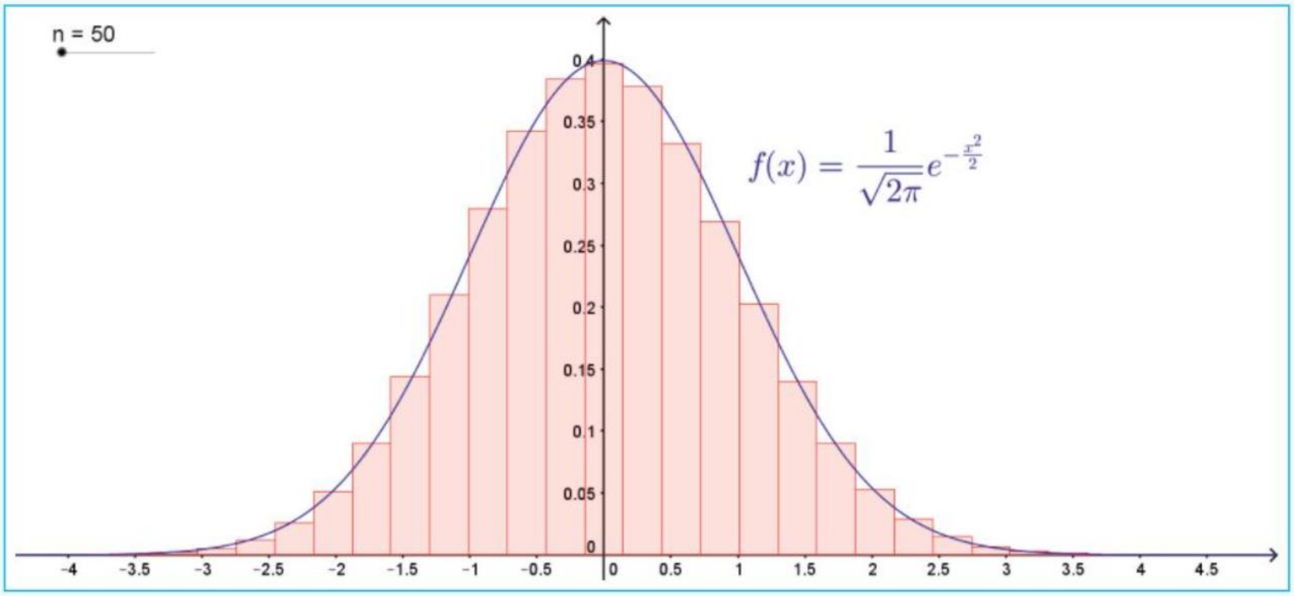
a. Quelle est la largeur commune des rectangles ? Quelle est la hauteur de R_k ?
Quelle est la somme des aires des rectangles R_k ?

b. Interpréter graphiquement $P(\mathrm{E}_n)$.

c. Qu'observe-t-on pour l'histogramme lorsque n augmente ?

B. Courbe de Gauss et loi normale centrée réduite

On introduit la fonction f définie sur $\mathbb{R}$ par $f(x) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.



1. Que constate-t-on ?

2. Calculer, à l'aide d'une calculatrice ou d'un logiciel, une valeur approchée de $I = \int_{-1}^{2} f(x)\,\mathrm{d}x$.

3. Interpréter graphiquement cette intégrale. Comparer I et $P(E_n)$ quand n augmente.

Lien vers le fichier geogebra : [lien]

De la loi binomiale à la loi Normale – Correction

A. 1. a. La variable aléatoire X_n compte le nombre de « succès » (gène A actif) observés sur un échantillon de n personnes dont le prélèvement s'effectue avec remise. X_n suit donc la loi binomiale $\mathcal{B}(n\,;0,4)$.

b. $E(X_n) = 0,4n\,;\sigma(X_n) = \sqrt{0,4 \times 0,6 \times n} \approx 0,49\sqrt{n}$.

c.

- Pour $n = 50$, $m = 20$ et $\sigma \approx 3,46$. Donc $m - \sigma \approx 16,54$ et $+ 2\sigma \approx 26,92$.

Ainsi $P(E_{50}) = P(17 \leqslant X_{50} \leqslant 26) \approx 0,8125$ (à la calculatrice ou avec GeoGebra).

- Pour $n = 200$, $m = 80$ et $\approx 6,93$. Donc $m - \sigma \approx 73,07$ et $+ 2\sigma \approx 93,86$.

Ainsi $P(E_{200}) = P(74 \leqslant X_{200} \leqslant 93) \approx 0,7994$ (à la calculatrice ou avec GeoGebra).

2. a. $m - \sigma \leqslant X_n \leqslant m + 2\sigma$ s'écrit aussi $-1 \leqslant \dfrac{X_n - m}{\sigma} \leqslant 2$ soit $-1 \leqslant Z_n \leqslant 2$.

b. $z_k = \dfrac{k - m}{\sigma}$ avec k entier $0 \leqslant k \leqslant n$.

$z_{k+1} - z_k = \dfrac{1}{\sigma}$.

3. a. La largeur commune des rectangles R_k est $\dfrac{1}{\sigma}$.

La hauteur de R_k est donc $\sigma \cdot P(Z_n = z_k) = \sigma \cdot P(X_n = k)$.

La somme des aires des rectangles R_k est égale à 1.

b. $P(E_n) = P(Z_n \in [-1\,; 2])$. Donc sur l'histogramme normalisé de Z_n, $P(E_n)$ est égal à la somme des aires des rectangles dont les centres sont compris entre – 1 et 2.

c. Quand n augmente, l'histogramme « se lisse » et prend l'allure d'une forme en cloche.

B. 1. On constate que lorsque n augmente, l'histogramme se lisse et tend à se confondre avec $\mathcal{C} : y = f(x)$.

2. À l'aide de GeoGebra (par exemple), on trouve que $\int_{-1}^{2} f(x)\,dx \approx 0,8186$.

3. I est l'aire du domaine limité par l'axe des abscisses, la courbe $\mathcal{C}$ et les droites d'équations $x = -1$ et $x = 2$.

Lorsque n augmente, $P(E_n)$ est très voisin de I.