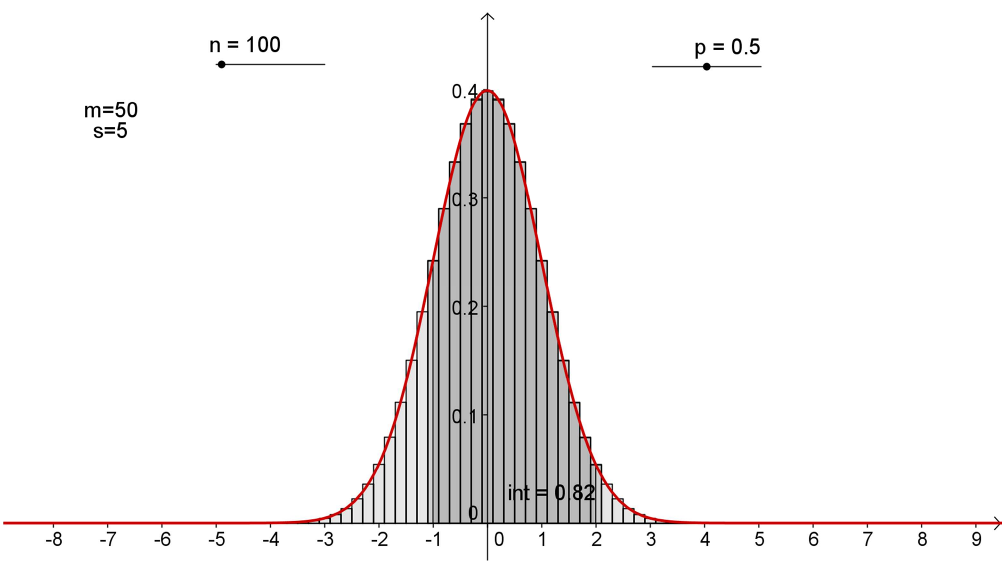
D'un histogramme à une courbe en cloche Nathan Hyperbole p399A1

Objectif : introduire la loi normale centrée réduite. X_n est une variable aléatoire qui suit une loi binomiale de paramètres n et p. On se propose d'étudier une propriété graphique de la variable aléatoire Z_n dite centrée réduite associée à X_n.

1. **Histogramme de X_n.**
   a) Avec le logiciel GeoGebra, créer un curseur n de 10 à 2000 avec un incrément de 1 et un curseur p de 0 à 1 avec un incrément de 0,01. Afficher l'espérance m et l'écart-type s de X.
   b) On associe la loi discrète de X_n à des aires de rectangles, comme on le fait pour l'histogramme d'une variable continue. Chaque rectangle de largeur 1 a pour hauteur $P(X_n= k)$. Pour cela, taper dans la zone de saisie binomiale(n, p) . Déplacer les curseurs et observer l'histogramme.
2. **Histogramme de Z_n.**
   On règle les curseurs n sur 100 et p sur 0,5.
   On considère la variable aléatoire
   $$Z_n = \frac{X_n - m}{s}, \text{ c'est à dire} Z_n = \frac{X_n - 50}{5}.$$
   a) Calculer l'espérance et l'écart-type de Z_n.
   b) Expliquer pourquoi Z_n prend ses valeurs entre -10 et $+10$. Quel est l'écart entre deux valeurs consécutives prises par Z_n ?
   c) Pour l'histogramme représentant Z_n, les rectangles d'aire $P(X_n = k) = P\left(Z_n = \frac{k-50}{5}\right)$ ont pour largeur 0,2 et pour hauteur $5 \times P(X_n = k)$.
   Avec GeoGebra, taper Séquence[(i - m) / s - 0.5 / s, i, 0, n + 1] (liste 1 des bornes) et Séquence[s Binomiale[n, p, k, false], k, 0, n] (liste 2 des hauteurs), puis Histogramme (liste 1, liste 2).
   d) Déplacer le curseur n .Plus n prend de grandes valeurs, plus les bords supérieurs des rectangles se rapprochent d'une courbe dont une équation semble être de la forme $y = ae^{-0,5x^2}$. Conjecturer la valeur de a. Tracer cette courbe avec GeoGebra.
3. **Approximation par une loi continue.**
   On règle à nouveau les curseurs n sur 100 et p sur 0,5.
   a) Expliquer pourquoi $P(45 \leq X_n \leq 60) = P(-1 \leq Z_n \leq 2)$.
   b) Sur l'écran GeoGebra précédent taper f(x) = 1 / sqrt(2π) exp(− 0.5 x²).
   c) Calculer l'aire sous la courbe entre – 1 et 2 en tapant : Intégrale(f, – 1, 2, true).
   d) Calculer la somme des aires de rectangles en tapant : SommeRectangles[f, -1.1, 2.1, 16, 0.5].



Vers l'intervalle de fluctuation asymptotique : un sac de billes Bordas Indice p356A1

Objectif : approcher la notion d'intervalle de fluctuation asymptotique.

Un sac contient des billes, dont un cinquième de billes rouges. On tire au hasard avec remise un échantillon de n billes de ce sac. Soit X_n la variable aléatoire associée au nombre de billes rouges tirées.

1. a. Quelle loi suit X_n ? Quels sont ses paramètres n et p ?
   b. Soit F_n la variable aléatoire définie par $F_n = \frac{X_n}{n}$. Que représente-t-elle ?
2. Dans cette question $n = 100$.
   a. Calculer le réel $s = \frac{\sqrt{p(1-p)}}{\sqrt{n}}$ puis l'intervalle I = [p – 1,96 $\boldsymbol{s}$; p+ 1,96 $\boldsymbol{s}$].
   b. Montrer que la probabilité que F_n appartienne à I est égale à la probabilité de l'événement « $13 \leq X_n \leq 27$ ». A l'aide de la calculatrice, déterminer cette probabilité à 10^{-3} près.
2. Dans cette question, n est compris entre 100 et 800. Pour chaque valeur de n, on calcule le réel s comme à la question 2.a. puis on détermine l'intervalle I_n = [p – 1,96 s ; p + 1,96 s] et la probabilité de l'événement « $F_n \in I_n$ ». Effectuer les calculs avec un tableur puis représenter les points de coordonnées n et $P(F_n \in I_n)$.
   a. Quel est le premier point placé sur le graphique ?
   b. Donner un encadrement de ces probabilités pour n compris entre 100 et 800.
3. Dans cette question, n est supérieur ou égal à 100.
   a. Soit Z_n la variable aléatoire définie par $Z_n = \frac{X_n - 0{,}2n}{0{,}4\sqrt{n}}$. En utilisant le théorème de Moivre-Laplace déterminer la limite quand n tend vers + ∞ de la suite (a_n) où $a_n = P(-1{,}96 \leq Z_n \leq 1{,}96)$.
   b. Montrer que : $(-1{,}96 \leq Z_n \leq 1{,}96) \Leftrightarrow \left(0{,}2n - 0{,}784\sqrt{n} \leq X_n \leq 0{,}2n + 0{,}784\sqrt{n}\right)$.
   c. A l'aide de la question précédente, déterminer le réel u_n tel que :
   $$\lim_{n \to +\infty} P(0{,}2 - u_n \leq F_n \leq 0{,}2 + u_n) = 0{,}95.$$
   d. Observer sur tableur les intervalles $[0{,}2 - u_n ; 0{,}2 + u_n]$ pour n égal à 100, 500, 1000, 5000 et 10 000. Que constate-t-on ?

Comparaison intervalle de fluctuation asymptotique et intervalle de fluctuation vu en seconde Bordas Indice p357A3

On s'intéresse à une population dans laquelle la proportion d'un caractère donné est p. On extrait au hasard et avec remise de cette population de taille n.

1. a. Rappeler l'expression d'un intervalle de fluctuation asymptotique I_n au seuil de confiance 0,95 de la fréquence du caractère dans cet échantillon.
   b. Déterminer les variations de la fonction $\varphi : p \to p(1-p)$ sur $[0;1]$.
   c. En déduire que $\sqrt{p(1-p)} \leq \frac{1}{2}$ pout tout réel p de [0 ;1].
   d. En déduire que $I_n \subset J_n$ où J_n est l'intervalle $\left[p - \frac{1}{\sqrt{n}} ; p + \frac{1}{\sqrt{n}}\right]$.
   e. Peut-on dire que J_n est un intervalle de fluctuation asymptotique ?
2. On n'est pas assuré que la probabilité que $\frac{X_n}{n}$ appartienne à J_n soit supérieure à 0,95 pour toute valeur de n.
   On va faire une étude expérimentale, à l'aide d'un tableur, des valeurs prises par $v_n = P\left(\frac{X_n}{n} \in J_n\right)$.
   a. Pour p = 0,2 et p = 0,3, que peut-on conjecturer ?
   b. Pour p prenant les valeurs 0,35 puis 0,40 et 0,45, conjecturer à partir de quelle valeur n_0 on a $v_n \geq 0{,}95$ pour tout entier $n \geq n_0$.
   c. Conjecturer la valeur n_0 pour p = 0,5. Pourquoi est-ce pour cette valeur de $\boldsymbol{p}$ que le réel n_0 semble être le plus grand ?

A la recherche d'un intervalle de confiance ; introduction graphique Didier Math'x p435A2

Une urne contient des boules rouges et des boules bleues. On note p la proportion des boules rouges dans l'urne. On désigne pas F la variable aléatoire qui à tout prélèvement de 100 boules avec remise, dans l'urne, associe la fréquence f des boules rouges obtenue sur l'échantillon.

1. Supposons que p soit connu, que $100\,p \geq 5$ et $100(1-p) \geq 5$. Dans ce cas, un intervalle de fluctuation asymptotique au seuil de 95% de F est donné par :

$$\left[p - 0{,}196\sqrt{p(1-p)} ; p + 0{,}196\sqrt{p(1-p)}\right]$$

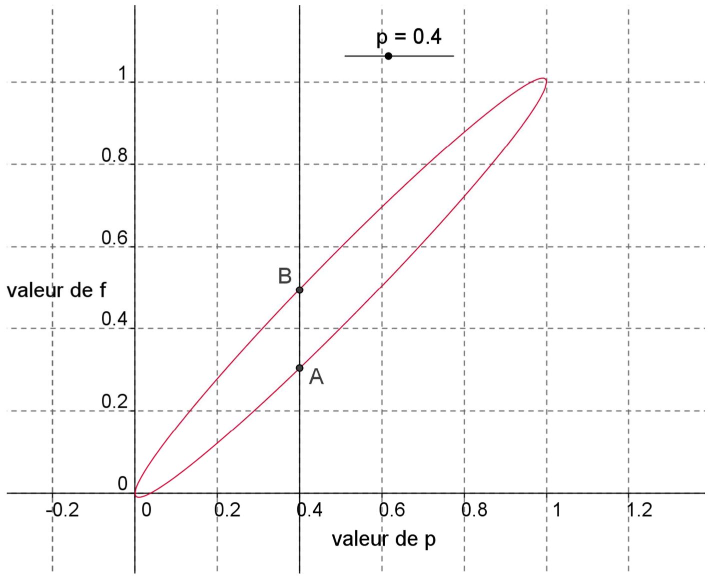
(Théoriquement, environ 95% des échantillons fournissent une fréquence f appartenant à cet intervalle).

a. A l'aide de GeoGebra, créer un curseur p et tracer les courbes C1 et C2 représentant les fonctions définies pour x dans l'intervalle [0 ;1] par $x \to x - 0{,}196\sqrt{x(1-x)}$ et $x \to x + 0{,}196\sqrt{x(1-x)}$.

Créer les points d'intersection de C1 et C2 avec la droite d'équation $x = p$.

b. Utiliser le fichier GeoGebra pour obtenir un intervalle de fluctuation asymptotique de F à 95% lorsque

$p = 0{,}4$; $p = 0{,}71$; $p = 0{,}02$. Que se passe-t-il dans ce dernier cas ?



2. Supposons maintenant que l'on ignore la valeur de p (on ne connait pas la proportion de boules rouges dans l'urne). On possède cependant la fréquence $f = 0{,}6$ des boules rouges après un prélèvement au hasard et avec remise de 100 boules. La recherche d'un intervalle « de confiance » à 95% pour p consiste en une inversion du point de vue.

a. Utiliser le fichier GeoGebra pour obtenir les abscisses x_1 et x_2 (avec $x_1 \leqslant x_2$) des points d'intersection des courbes C_1 et C_2 avec la droite d'équation $y = 0{,}6$.

b. Montrer que x_1 et x_2 sont solutions de l'équation $(0{,}6 - x)^2 = 0{,}196^2\, x\,(1 - x)$ et résoudre cette équation.

L'intervalle $[x_1 ; x_2]$est un intervalle de confiance de p à 95% de confiance.

3. En seconde, un intervalle de fluctuation de F au seuil de 95% est donné par $\left[p - \frac{1}{\sqrt{100}} ; p + \frac{1}{\sqrt{100}}\right]$.

Préciser et représenter les fonction définies sur [0 ;1] par $x \to x - \frac{1}{\sqrt{100}}$ et $x \to x + \frac{1}{\sqrt{100}}$. Les exploiter, comme précédemment, pour obtenir un intervalle de confiance de p lorsque $f = 0{,}6$.

Intervalle de confiance et fourchettes de sondages   document d'accompagnementet Nathan Transmath p451TP19

Le 18 avril 2002, l'institut IPSOS effectue un sondage dans la population en âge de voter.
On constitue un échantillon de 1000 personnes (inscrites sur les listes électorales) que l'on suppose choisies ici de manière aléatoire. Les résultats partiels en sont les suivants : sur les 1000 personnes :
135 ont déclaré vouloir voter pour Jean-Marie Le Pen
195 ont déclaré vouloir voter pour Jacques Chirac
170 ont déclaré vouloir voter pour Lionel Jospin.
Déterminer l'intervalle de confiance au niveau de confiance de 95% pour chacun des 3 candidats.
Que peut-on à priori imaginer sur cette élection ?
Les résultats le jour de l'élection étaient 16,9% pour M Le Pen, 19,9% pour M Chirac et16,2% pour M Jospin. Comparer ces résultats avec les intervalles de confiance.
Afin de mettre en évidence la difficulté de prévoir l'ordre des candidats à l'issue de premier tour, il est possible de réaliser une simulation d'un sondage aléatoire sur 1000 personnes à partir des résultats définitifs.
Comment réaliser une telle simulation ?

Intervalles de confiance, approfondissement : comparaison de deux proportions (semis)

Un maraîcher achète un lot de semences de tomates. Il lui reste des semences de l'année passée, dont il doit contrôler le taux de germination pour pouvoir les utiliser avec les autres.
En effet, des taux de germination trop différents provoquent des trous dans les plates-bandes, ce qui génère un coût de manutention plus élevé. Il faut donc comparer les taux de germination des semences des deux années.
Pour faire cette comparaison, le maraîcher prélève aléatoirement dans les semences de l'année un échantillon de 200 graines qu'il met à germer. Il constate que 185 graines germent.
Il prélève ensuite aléatoirement dans les semences de l'année précédente un échantillon de200 graines qu'il met à germer. Il constate que 150 graines germent.
Déterminer un intervalle de confiance au niveau de confiance de 95% de chacun des taux de germination, puis conclure.

Utilisation de la calculatrice (voir fiches calculatrices) ou du tableur

Sécurité d'un ascenseur.

1. On considère la variable aléatoire Z qui, à tout adulte, usager d'ascenseurs, choisi au hasard, associe son poids en kg.
On suppose que ***Z suit la loi normale d'espérance mathématique 70 kg et d'écart type 15 kg.***
a) Calculer, à 10^{-2} près, la probabilité pour qu'un adulte, usager d'ascenseurs, choisi au hasard,

- *pèse moins de 90 kg.*
- *pèse entre 55 et85 Kg*
- *pèse plus de 60kg*

b) Compléter avec une calculatrice (et selon le modèle proposé) les phrases suivantes

- *« 75% de la population pèse moins de …….. kg »*
- *« 30% de la population pèse plus de …….. kg »*

2. Un ascenseur peut supporter 500 kg avant la surcharge. Les normes de sécurité spécifient que la probabilité de surcharge ne doit pas dépasser 0,0001.
On admet que le poids total de n usagers adultes d'ascenseurs, dont les poids sont indépendants, est une variable aléatoire S_n qui suit la loi normale d'espérance mathématique 70 n et d'écart type 15 racine (n).
2. Calculer les probabilités de surcharge p_5 lorsqu'il y a 5 adultes dans l'ascenseur et p_6 lorsqu'il y a 6 adultes dans l'ascenseur. En déduire le nombre maximal de personnes autorisées à emprunter l'ascenseur.