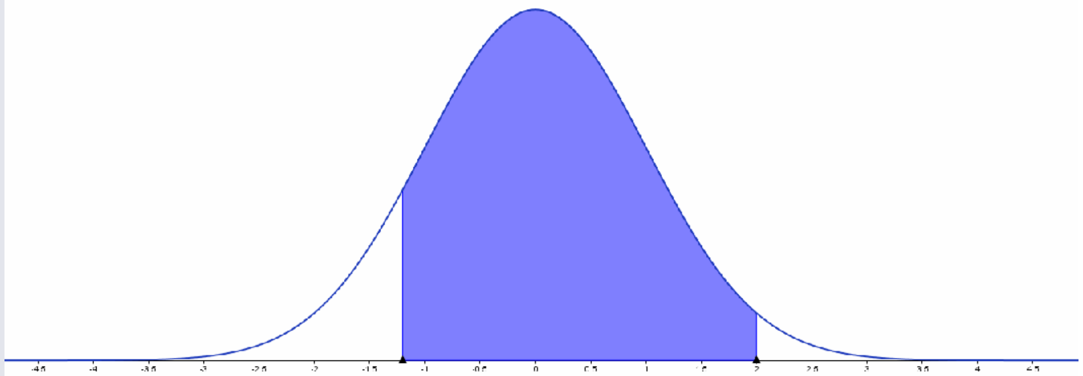
Atelier Probabilités et statistiques



Animation

Nouveau programme de Terminale

Mai 2012
Académie de Créteil

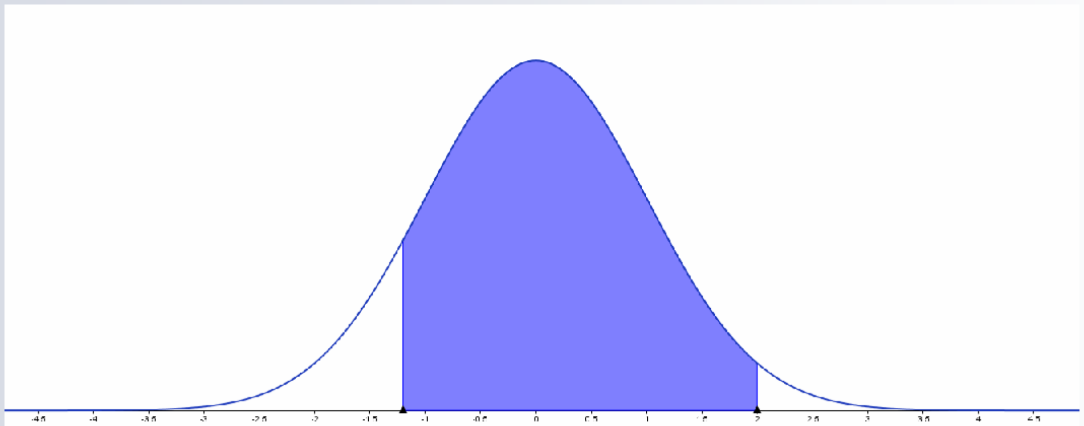
Quelques problèmes pour motiver

- ❑ Comment estimer la fiabilité d'une chaine de production ?
- ❑Peut-on déterminer la probabilité qu'un enfant porte des vêtements 3 mois dès la naissance ?
- ❑ Comment retrouver la trace des flux migratoires ?

Un outil: la loi normale N(0,1)

Une variable aléatoire X suit la loi normale centrée réduite notée N(0,1) si, pour tous réels a et b tels que $a \leq b$, on a :

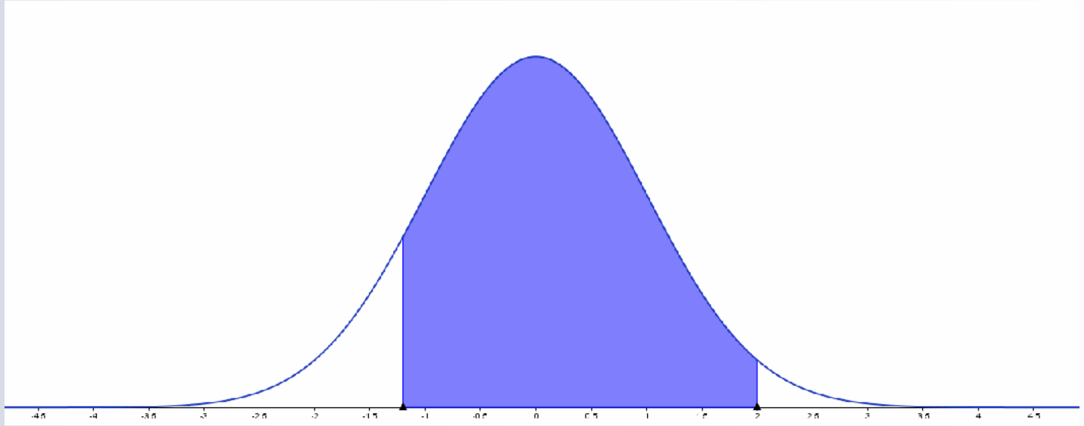
$$P(a < X < b) = \int_a^b f(x)dx \text{ où } f \text{ est définie sur } \Re \text{ par } f(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$$



•Densité de la loi normale

La loi normale N(μ, σ^2)

Une variable aléatoire X suit une loi N (μ, σ^2) si la variable aléatoire $\dfrac{X-\mu}{\sigma}$ suit la loi normale N (0,1).



•Densité de la loi normale

Valeurs particulières

Si X suit une loi N (μ, σ^2), alors

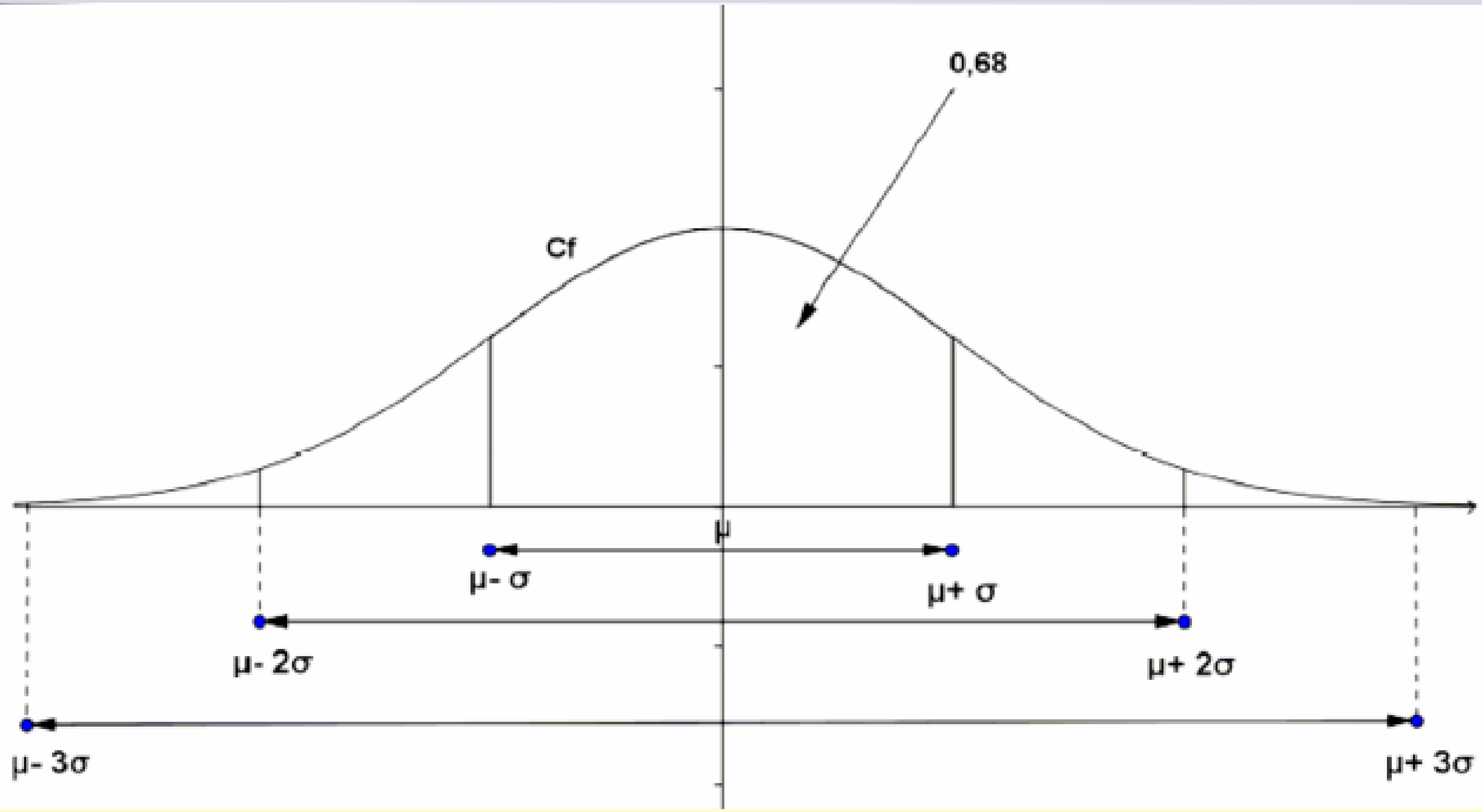
$$P(\mu - \sigma < X < \mu + \sigma) \approx 0,68 \text{ à } 10^{-2} \text{ près}$$

$$P(\mu - 2\sigma < X < \mu + 2\sigma) \approx 0,96 \text{ à } 10^{-2} \text{ près}$$

$$P(\mu - 3\sigma < X < \mu + 3\sigma) \approx 0,997 \text{ à } 10^{-3} \text{ près}$$

- Densité de la loi normale

Valeurs particulières



•Densité de la loi normale

La loi normale: exemple d'activité

Le périmètre crânien en cm d'un enfant de 3 ans suit une la loi normale d'espérance 49 cm et d'écart-type 1,6 cm.

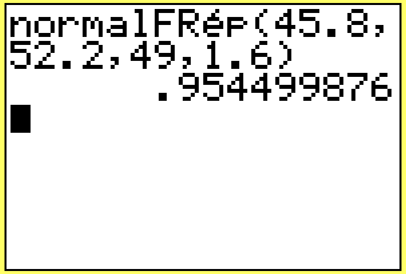
1)Quelle est la probabilité pour que le périmètre crânien d'un enfant de 3 ans soit comprise entre 45,8 et 52,2 cm ?

2)Quelle est la probabilité pour que le périmètre crânien d'un enfant de 3 ans soit inférieure à 48 cm ?

La loi normale: exemple d'activité à modifier

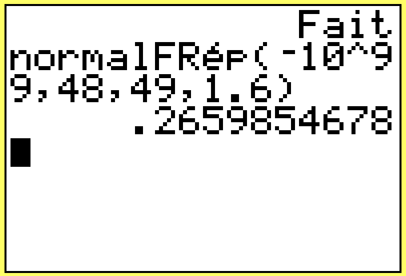
X suit une v.a. N(49;1,6²). Avec la calculatrice , on peut évaluer

1. $$P(45,8 \leq X \leq 52,2)$$

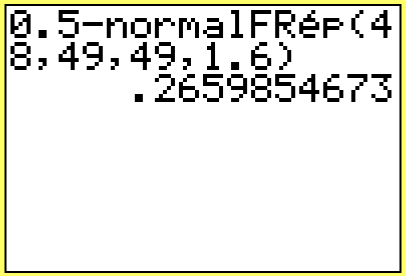


perimetrecranien.ggb

2. $$P(X \leq 48)$$



ou



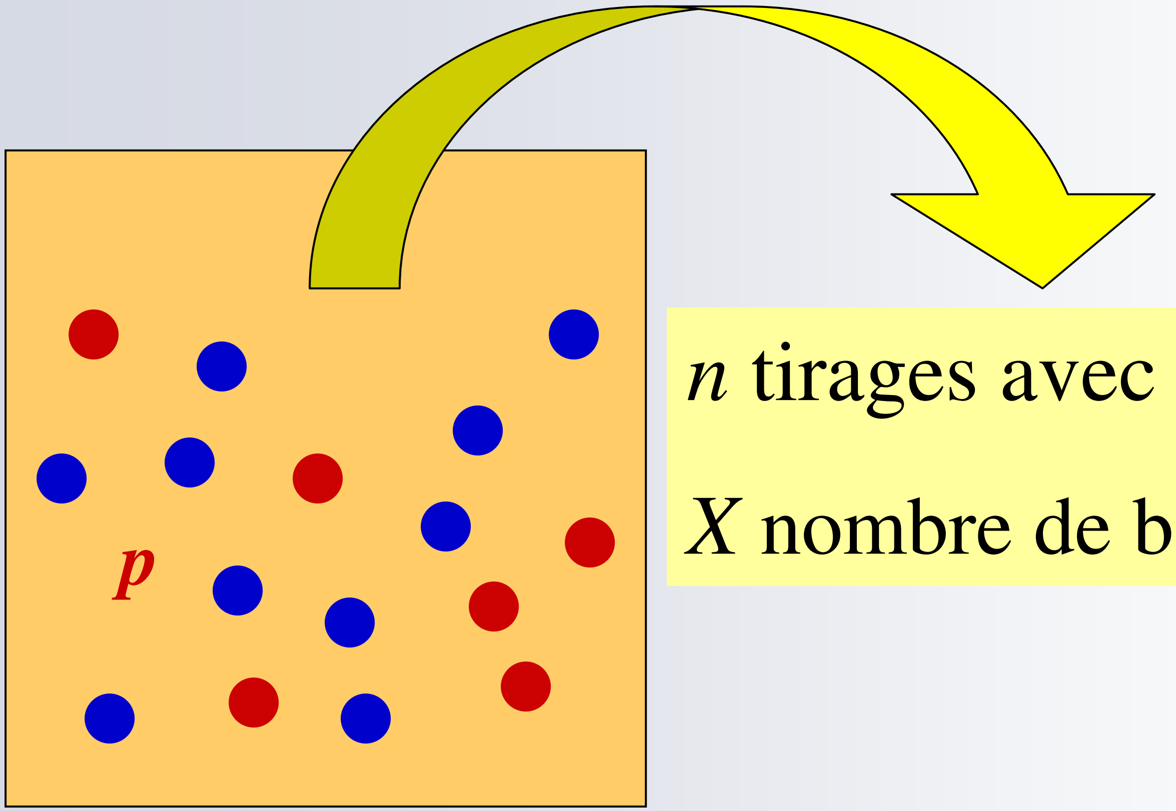
Comment introduire la loi normale en T^ale ?

Commentaires

Pour introduire la loi normale $\mathcal{N}(0,1)$, on s'appuie sur l'observation des représentations graphiques de la loi de la variable aléatoire $Z_n = \dfrac{X_n - np}{\sqrt{np(1-p)}}$ où X_n suit la loi binomiale $\mathcal{B}(n, p)$, et cela pour de grandes valeurs de n et une valeur de p fixée entre 0 et 1. Le théorème de Moivre Laplace assure que pour tous réels a et b, $P(Z_n \in [a,b])$ tend vers

$$\int_a^b \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}} \mathrm{d}x$$ lorsque n tend vers $+\infty$.

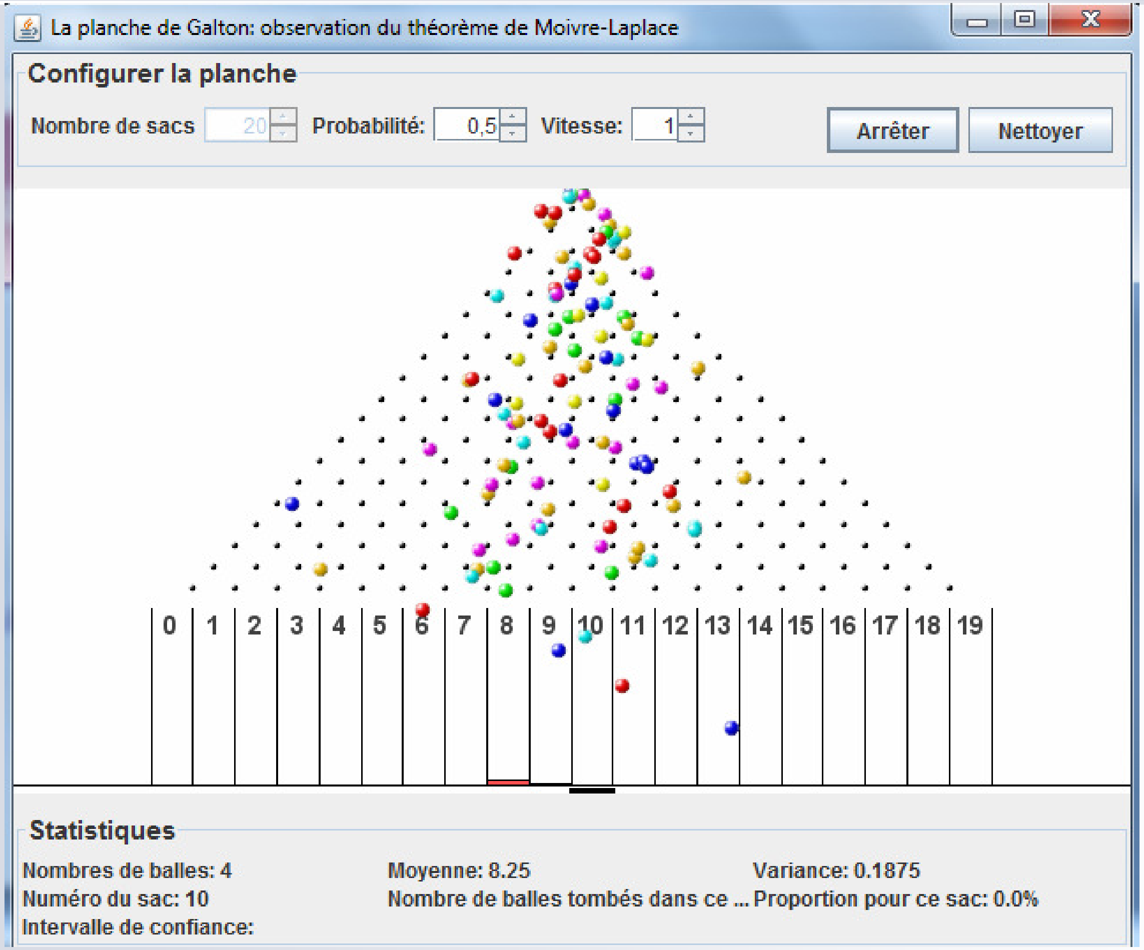
Retour à loi binomiale



n tirages avec remise.

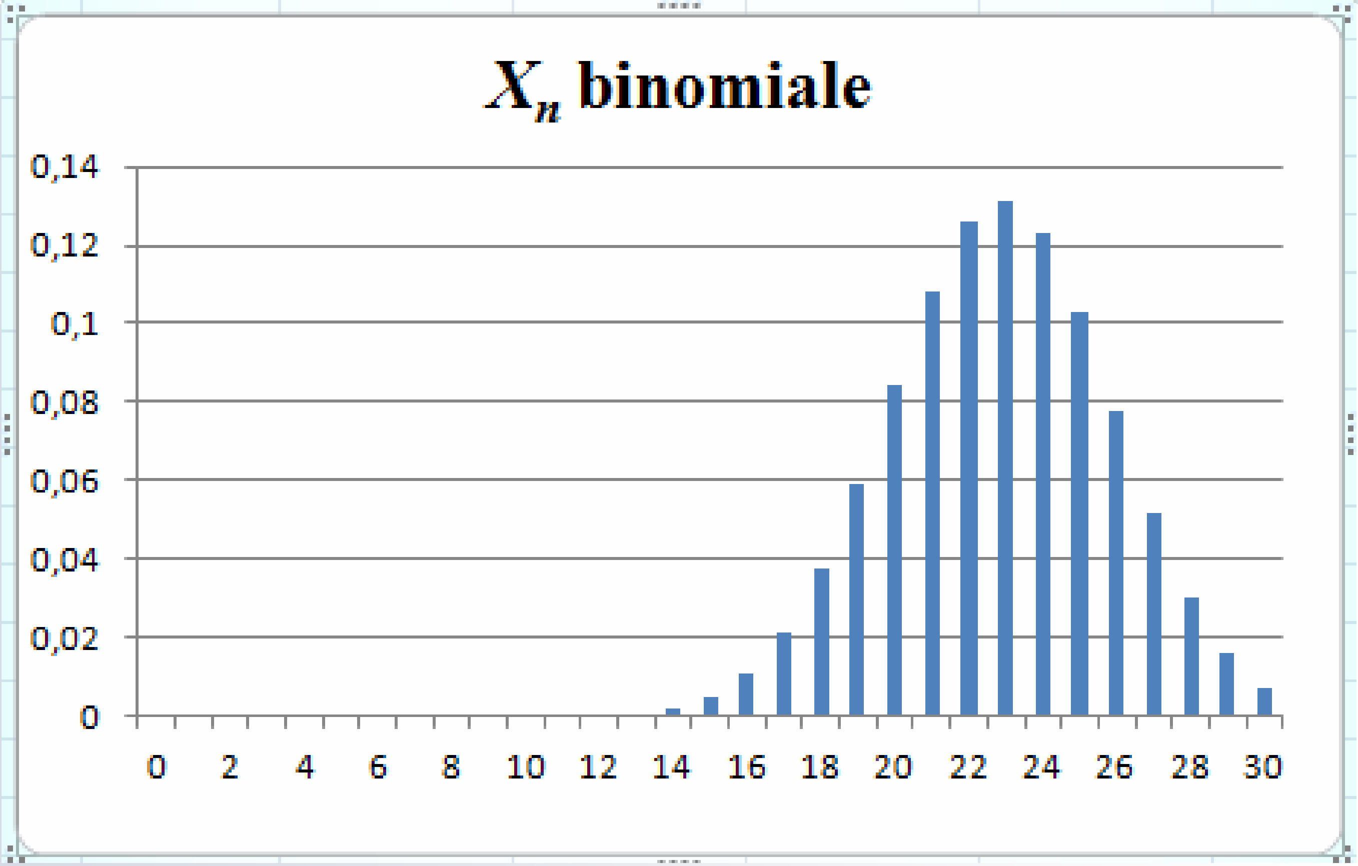
X nombre de boules rouges

La loi binomiale: visualisation



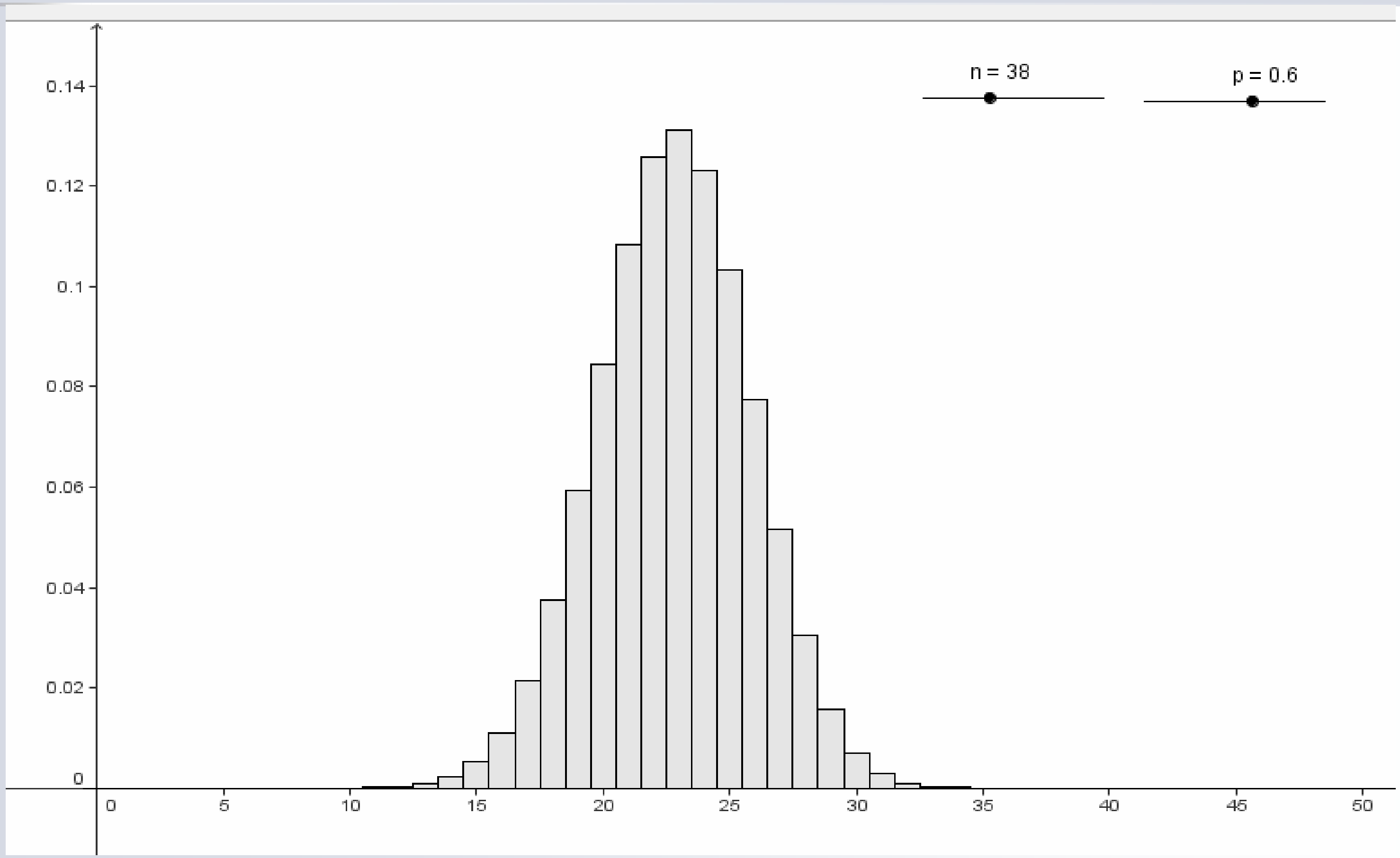
[Lancer la Planche de Galton]

La loi binomiale: visualisation



Distribution de B(38;0,6) avec le tableur

La loi binomiale: visualisation



Distribution de B(38;0,6) avec Geogebra4

X suit la **loi binomiale** de **moyenne** $E(X) = np = m$ et d'**écart type** $\sigma = \sqrt{np(1-p)}$.

On pose alors $Z = \dfrac{X - np}{\sqrt{np(1-p)}} = \dfrac{X - m}{\sigma}$.

$E(Z) = 0$ et $\sigma(Z) = 1$ sont **indépendants** de n et de p.

Z est appelée **binomiale centrée réduite**.

Avec $Z = \dfrac{X-m}{\sigma}$.

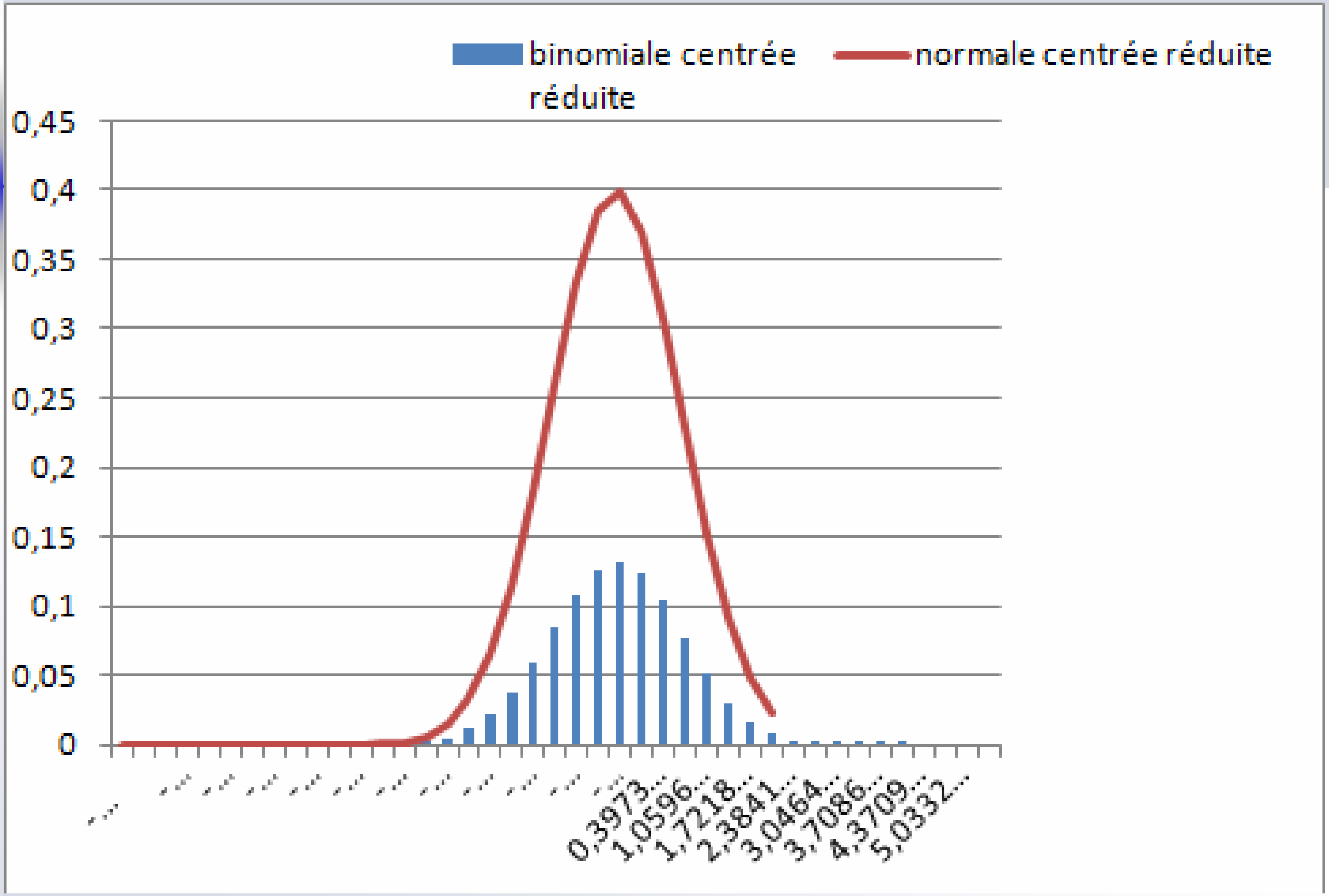
•X prend des valeurs entières k dans [0 ; n]

•Z prend des valeurs z régulièrement réparties

sur $\left[\dfrac{-m}{\sigma};\dfrac{n-m}{\sigma}\right]$ et l'écart entre deux valeurs

consécutives est $\dfrac{1}{\sigma}$.

La représentation graphique de Z est donc un diagramme en bâtons qui dépend donc de p et de n

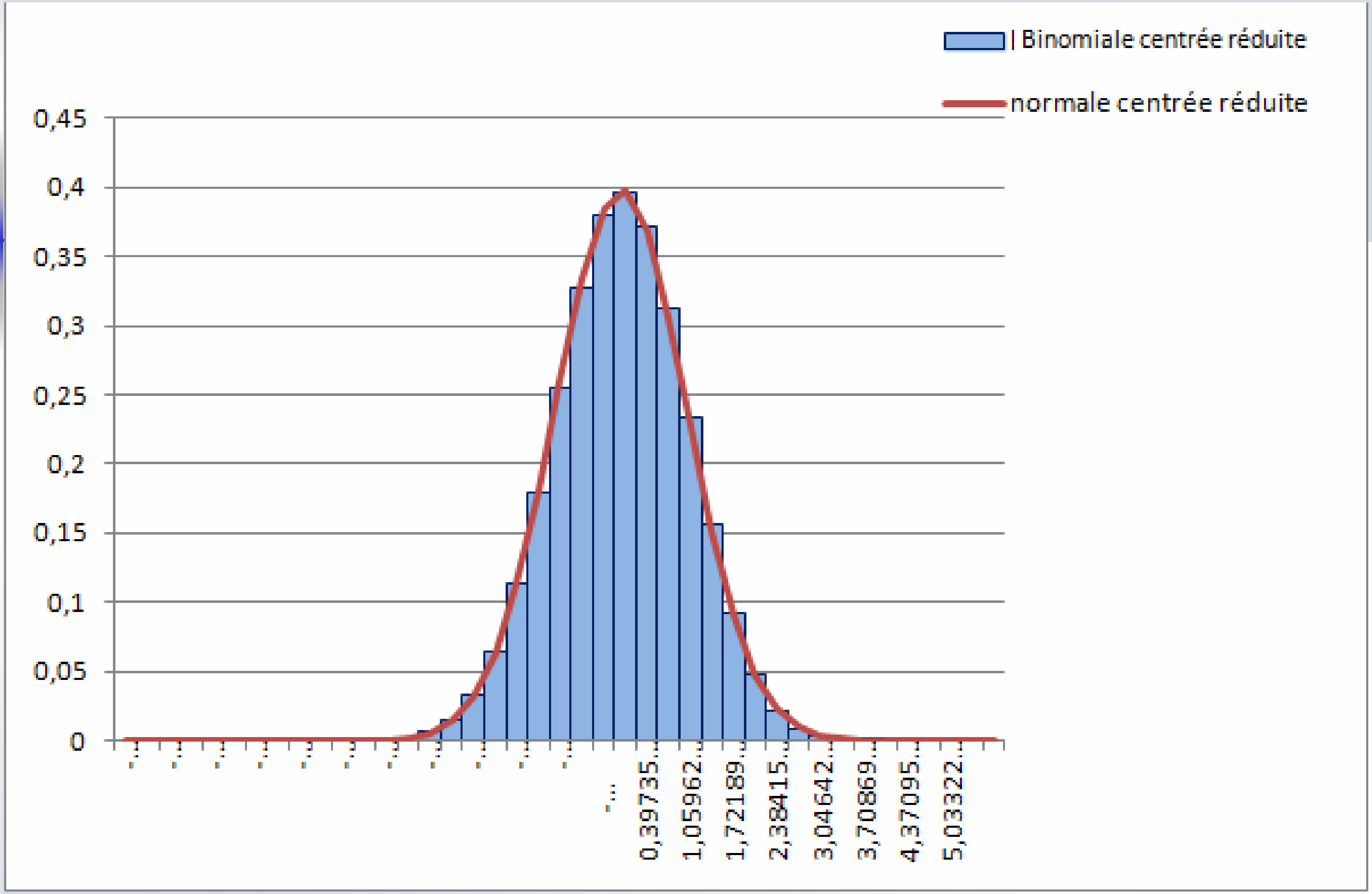


binomiale centrée réduite
normale centrée réduite
0,45
0,4
0,35
0,3
0,25
0,2
0,15
0,1
0,05
0
0,3973...
1,0596...
1,7218...
2,3841...
3,0464...
3,7086...
4,3709...
5,0332...

Pour comparer deux variables continues, et donc visualiser un histogramme, à chaque valeur de z on fait correspondre un rectangle vertical dont **l'aire est égale à *P(Z=z)*.**

- La base est un segment de l'axe horizontal de longueur $\frac{1}{\sigma}$ (écart entre deux valeurs consécutives prises par Z).

- La hauteur de ce rectangle est donc $\sigma \times P(Z = z)$.


Binomiale centrée réduite
normale centrée réduite
0,45
0,4
0,35
0,3
0,25
0,2
0,15
0,1
0,05
0
0,39735...
1,05962...
1,72189...
2,38415...
3,04642...
3,70869...
4,37095...
5,03322...

Ca marche !

La loi normale (T^{ale})

Commentaires

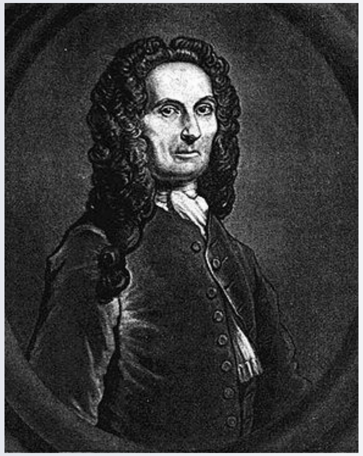
Pour introduire la loi normale $\mathcal{N}(0,1)$, on s'appuie sur l'observation des représentations graphiques de la loi de la variable aléatoire $Z_n = \dfrac{X_n - np}{\sqrt{np(1-p)}}$ où X_n suit la loi binomiale $\mathcal{B}(n, p)$, et cela pour de grandes valeurs de n et une valeur de p fixée entre 0 et 1. Le théorème de Moivre Laplace assure que pour tous réels a et b, $P\left(Z_n \in [a,b]\right)$ tend vers

$\displaystyle\int_a^b \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}} \mathrm{d}x$ lorsque n tend vers $+\infty$.

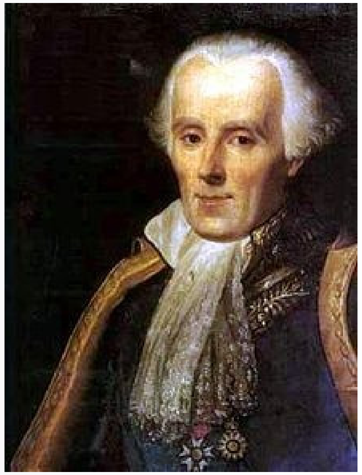
Un peu d'histoire…

Bernoulli
(1654-1705)

De Moivre
(1667-1754)



Laplace
(1749-1827)



Point commun: Volonté de mesurer aussi précisément que possible les fluctuations d'une variable aléatoire binomiale autour de son espérance.

Ce que fit Bernoulli…

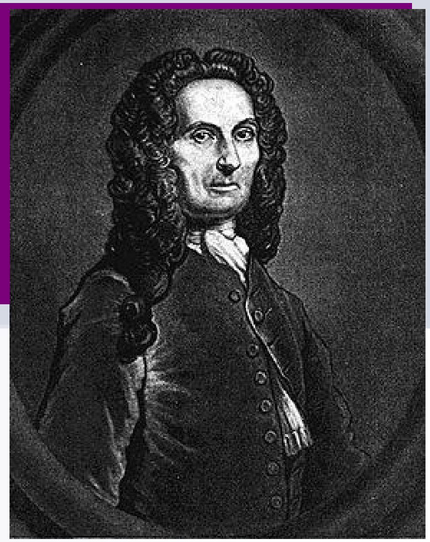
Jacob Bernoulli (1654-1705)

Il est l’un des premiers à aborder l’approche fréquentielle des probabilités.

"Même pour le plus stupide des hommes, l’instinct naturel, et sans aucune instruction, ce qui est remarquable, le conduit à être convaincu que plus on fait d’observations, moins le danger de dévier de son but est grand ».

Il établit la première version de la loi des grands nombres.

Un peu d'histoire…



Abraham De Moivre (1667-1754)

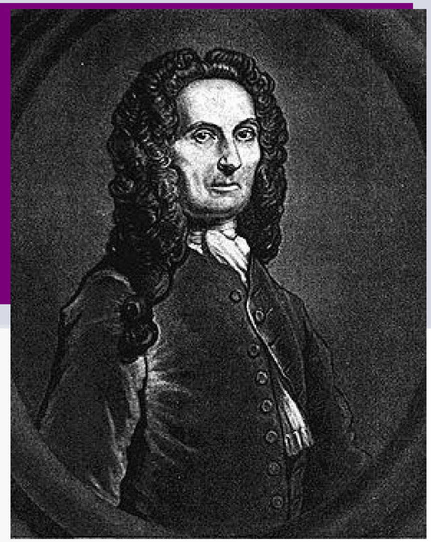
De Moivre est un protestant qui fut emprisonné suite à la révocation de l'Edit de Nantes. Puis, il émigre en Angleterre où il rencontre James Stirling.

Formule dite de Stirling (à tort ?): $$n! \underset{+\infty}{\sim} \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$$

C'est la première rencontre avec la loi normale.

De Moivre, qui fréquenta Leibniz et Newton, utilise **les techniques de calcul infinitésimal.**

Un peu d'histoire…



Abraham De Moivre (1667-1754)

Il utilise l'aire sous la courbe de $x \mapsto e^{-\frac{2x^2}{n}}$

Et obtient que

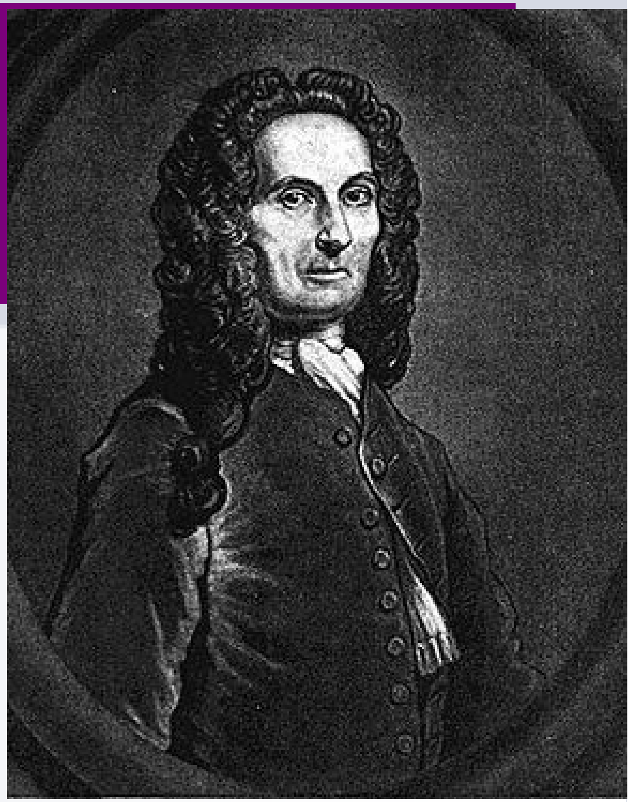
Si X_n suit $B\left(n;\frac{1}{2}\right)$ alors

$$\lim_{+\infty} P\left(\left|\frac{X_n}{n}-\frac{1}{2}\right| \leq \frac{1}{2\sqrt{n}}\right)=0,682688$$

$$\lim_{+\infty} P\left(\left|\frac{X_n}{n}-\frac{1}{2}\right| \leq \frac{1}{\sqrt{n}}\right)=0,95428$$

« environ 68% des observations sont au plus à un écart-type ».
« environ 95% des observations sont au plus à deux écart-types ».
Densitenormale.ggb

Un peu d'histoire…



Abraham De Moivre

Cette démonstration figure dans The doctrine of chances (1718).

Il généralise ses résultats sans démonstration au cas où $p \neq \frac{1}{2}$.

Un peu d'histoire…

Pierre-Simon De Laplace (1749-1827)

Il démontre que la moyenne arithmétique des erreurs d'observations commises lors de *n* mesures se comporte approximativement comme une "loi normale"; et l'approximation est d'autant plus précise que *n* est grand. C'est que l'on appelle aujourd'hui **le théorème central-limite**, dont le théorème de Moivre –Laplace est un cas particulier.

Ce théorème rend la loi normale…centrale !

Applications

- Prendre une décision à partir d’un échantillon.
- Estimer une proportion

INTERVALLE DE FLUCTUATION
INTERVALLES DE CONFIANCE

	Intervalle de fluctuation *p connue*	**Intervalle de confiance** *p inconnue*
SECONDE	$n \geq 25$ et $0{,}2 \leq p \leq 0{,}8$, seuil 95% $\left[p-\frac{1}{\sqrt{n}}, p+\frac{1}{\sqrt{n}}\right]$	Sensibilisation
PREMIÈRE	Avec la loi binomiale	
TERMINALE	$n \geq 30$ et $np \geq 5$ et $n(1-p) \geq 5$ Asymptotique au seuil $1-\alpha$ $I_n = \left[p - u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}, p + u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$	Au niveau de confiance 95% $\left[f-\frac{1}{\sqrt{n}}, f+\frac{1}{\sqrt{n}}\right]$

Pourquoi un nouvel intervalle de fluctuation en Terminale ?

L'intervalle de fluctuation vu en première est certes exact mais présente un défaut majeur : **pas de formule donnant ses extrémités en fonction de *n et p !***

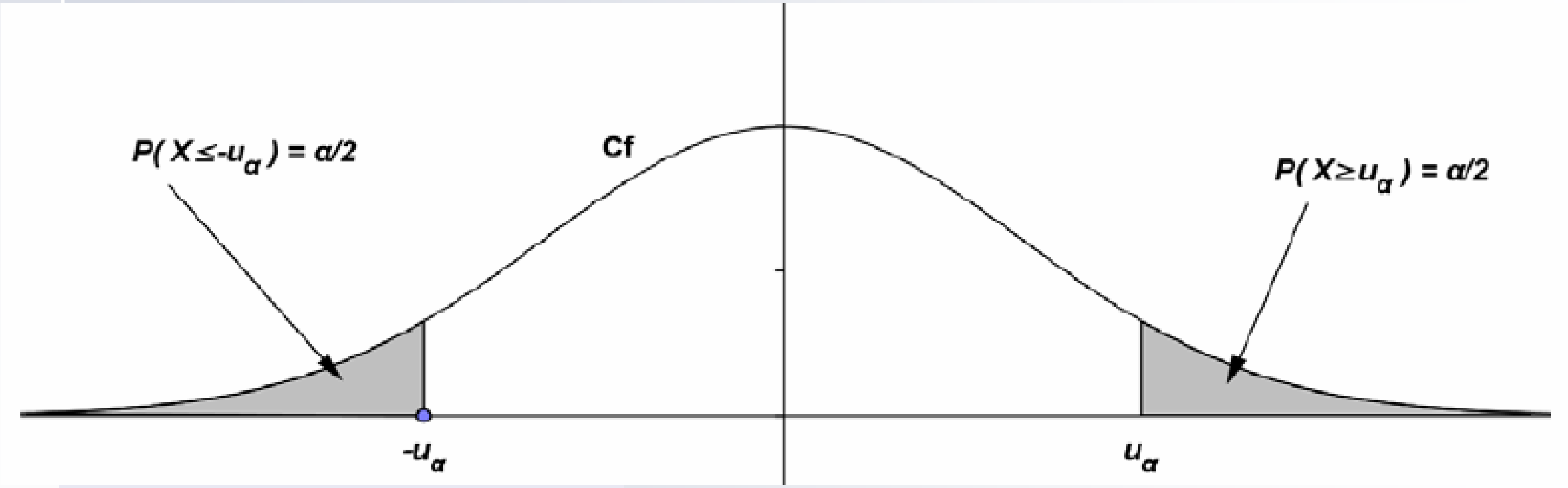
- Il faut donc le déterminer **au cas par cas ,** soit à l'aide d'un tableur soit à l'aide d'un algorithme.
- Cela peut s'avérer très long, voire impossible, lorsque *n est grand* et *np (l'espérance) aussi.*

  exemple :en 2010, en France, sur les 802224 naissances, 410140 étaient des garçons. Tester l'hypothèse $p=0.5$.
- Pas de formule, donc pas de formule à inverser pour obtenir un intervalle de confiance !

Nouvel intervalle de fluctuation

propriété (propriété et preuve uniquement en Terminale S !)

Lorsque X suit une loi normale $\mathcal{N}(0,1)$,pour tout $\alpha \in]0;1[$, il existe un unique réel positif u_α tel que : $P(-u_\alpha \leq X \leq u_\alpha) = 1 - \alpha$



pour α=0.05, $u_\alpha \approx 1.96$
pour α=0.01, $u_\alpha \approx 2.58$

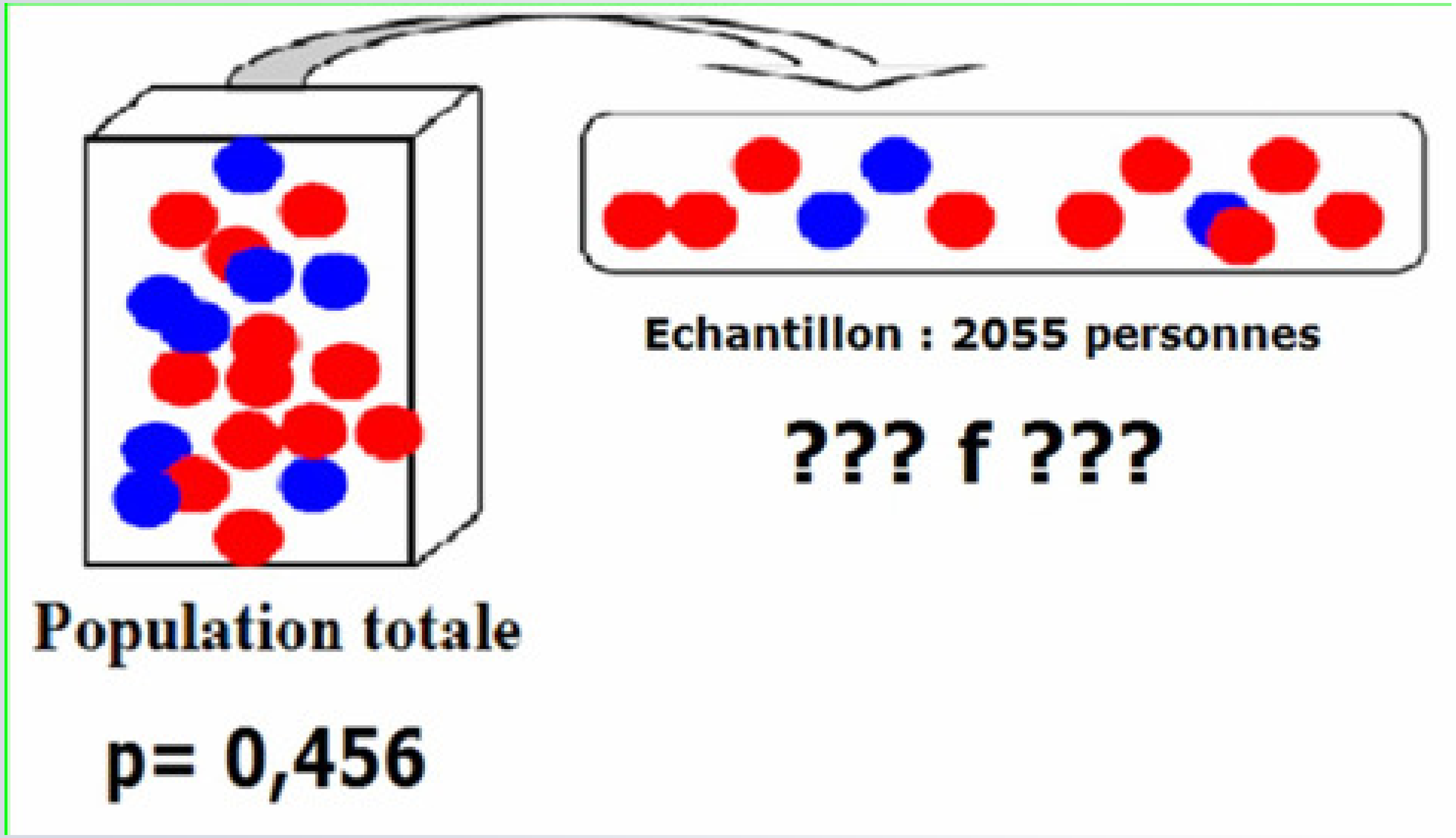
Application de l'étude de l'intervalle de fluctuation en lien avec SVT

- Dans l'ouest Armorico-Vendéen 45,6% des habitants ont le groupe sanguin O.

- Une étude (*) portant sur n =2055 marins bretons: f = 51% sont du groupe O contre p = 0,456 !!!

- (*) <u>Source</u>: http://www.persee.fr/web/revues/home/prescript/article/bmsap_0037-8984_1953_num_4_1_2610

Utilisation de l'intervalle de fluctuation: la prise de décision



Utilisation de l'intervalle de fluctuation: la prise de décision

Ici, n =2055 et p = 0,456.

Les conditions d'approximation par la loi normale

$n \geq 30$, $np \geq 5$ et $n(1-p) \geq 5$ sont vérifiées.

La probabilité que f soit dans l'intervalle de fluctuation

$$\left[p - 1.96\frac{\sqrt{p(1-p)}}{\sqrt{n}}; p + 1.96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right] \approx [0,434;0,478]$$

est proche de 0,95.

Utilisation de l'intervalle de fluctuation: la prise de décision

La probabilité que f soit dans l'intervalle de fluctuation

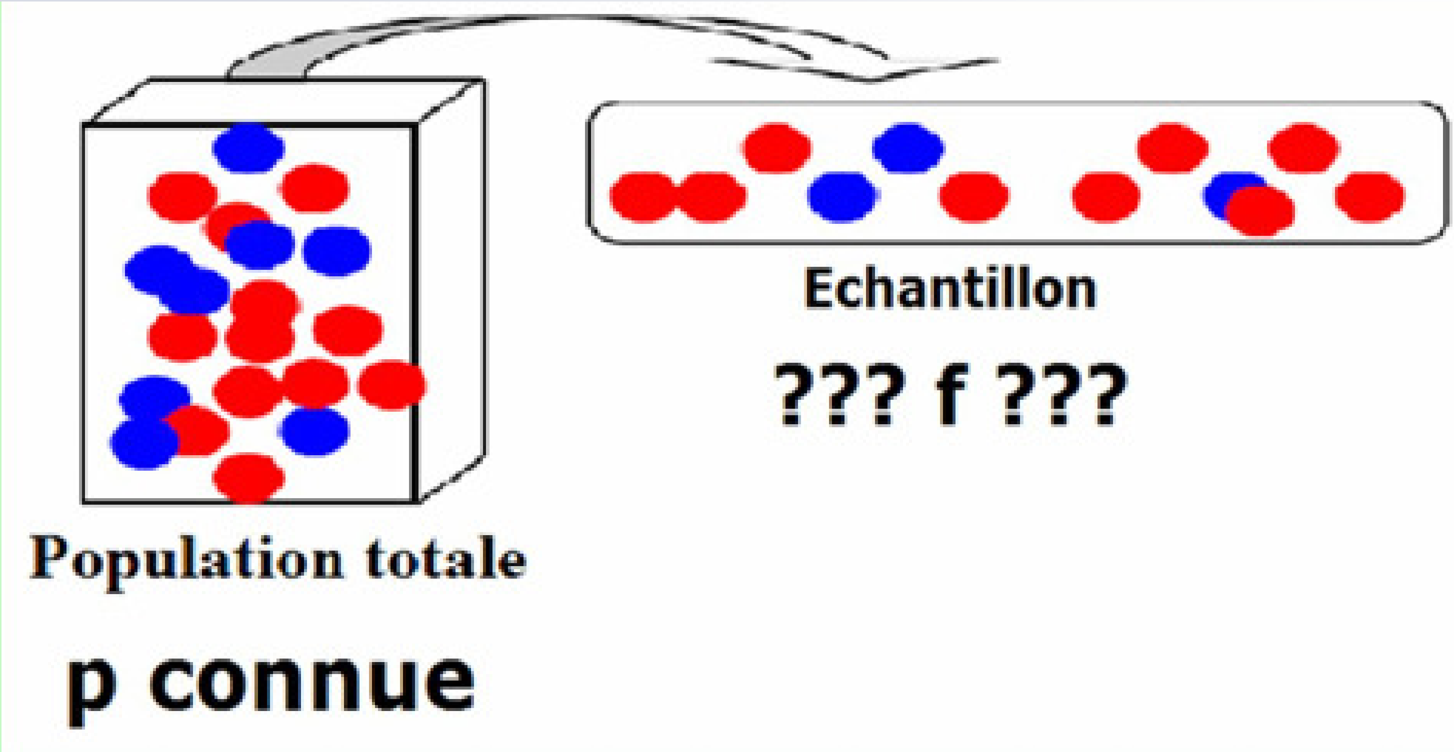
$$[p - 1.96\frac{\sqrt{p(1-p)}}{\sqrt{n}} ; p + 1.96\frac{\sqrt{p(1-p)}}{\sqrt{n}}] \approx [0,434;0,478]$$

est proche de 0,95.

Ayant observé $f = 0,51$ dans le groupe particulier des marins bretons, on rejette l'hypothèse:

« Les marins bretons, ont le même groupe sanguin que le reste de la population » au risque de 5%.

Intervalle de fluctuation

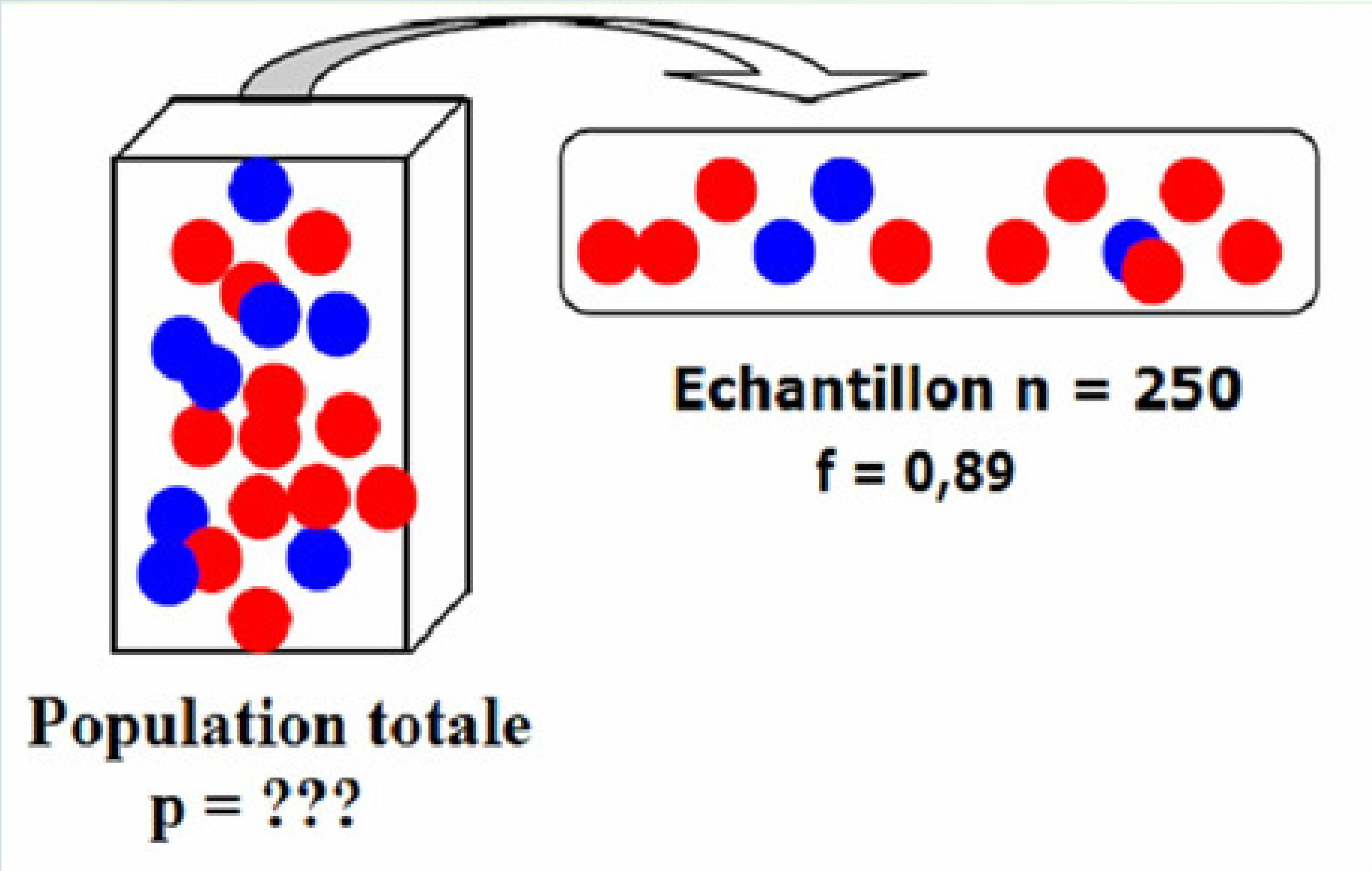


Intervalle de confiance

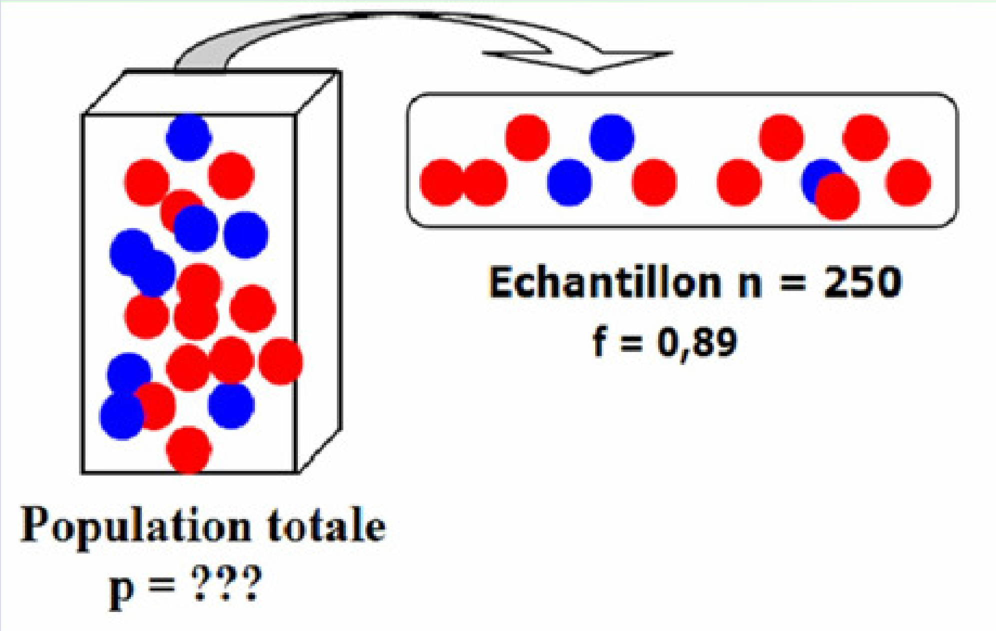
- Un fabricant de feux d'artifices souhaite contrôler la qualité de sa production: il en déclenche 250 choisis au hasard.
- 89% fonctionnent correctement.

Que peut-on en conclure sur la qualité de la production ?

Intervalle de confiance

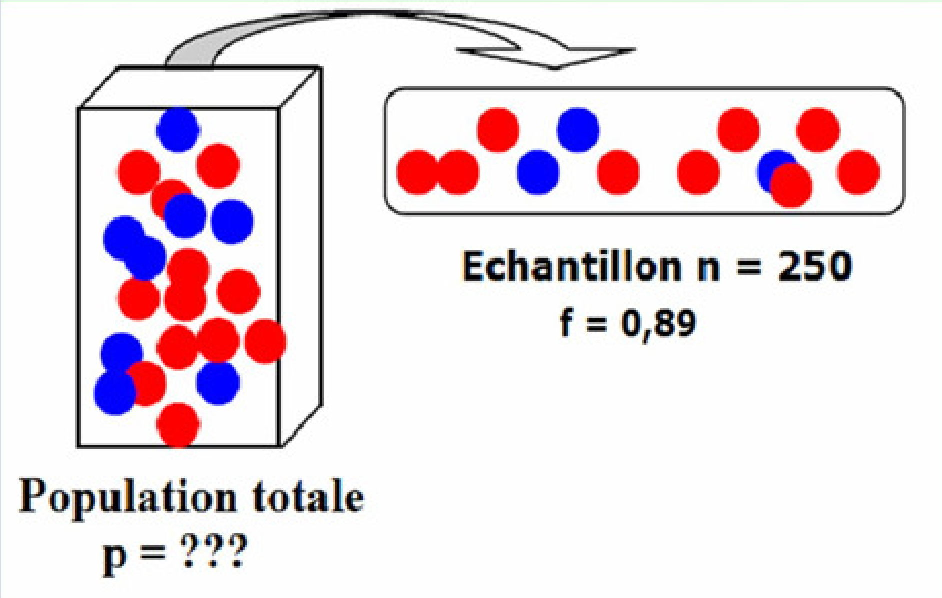


Intervalle de confiance



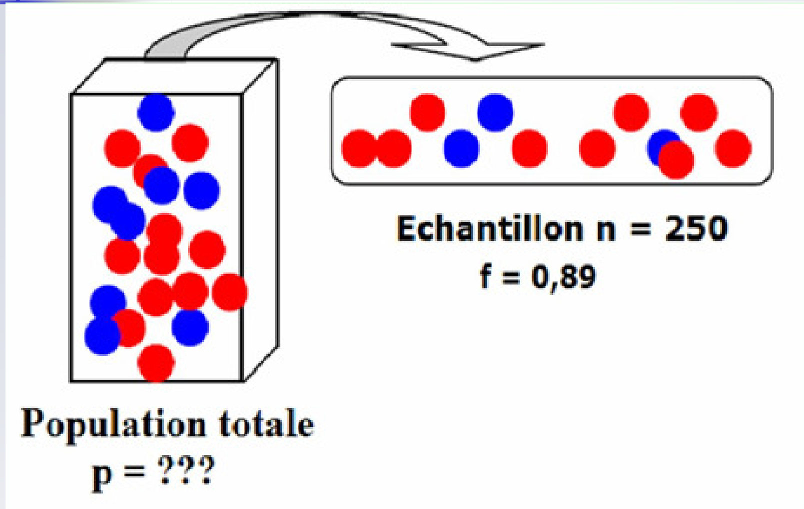
$$p \in \left[f - \frac{1}{\sqrt{n}} \,;\, f + \frac{1}{\sqrt{n}} \right]$$

Intervalle de confiance



$$p \in [0,82 ; 0,96]$$

Intervalle de confiance



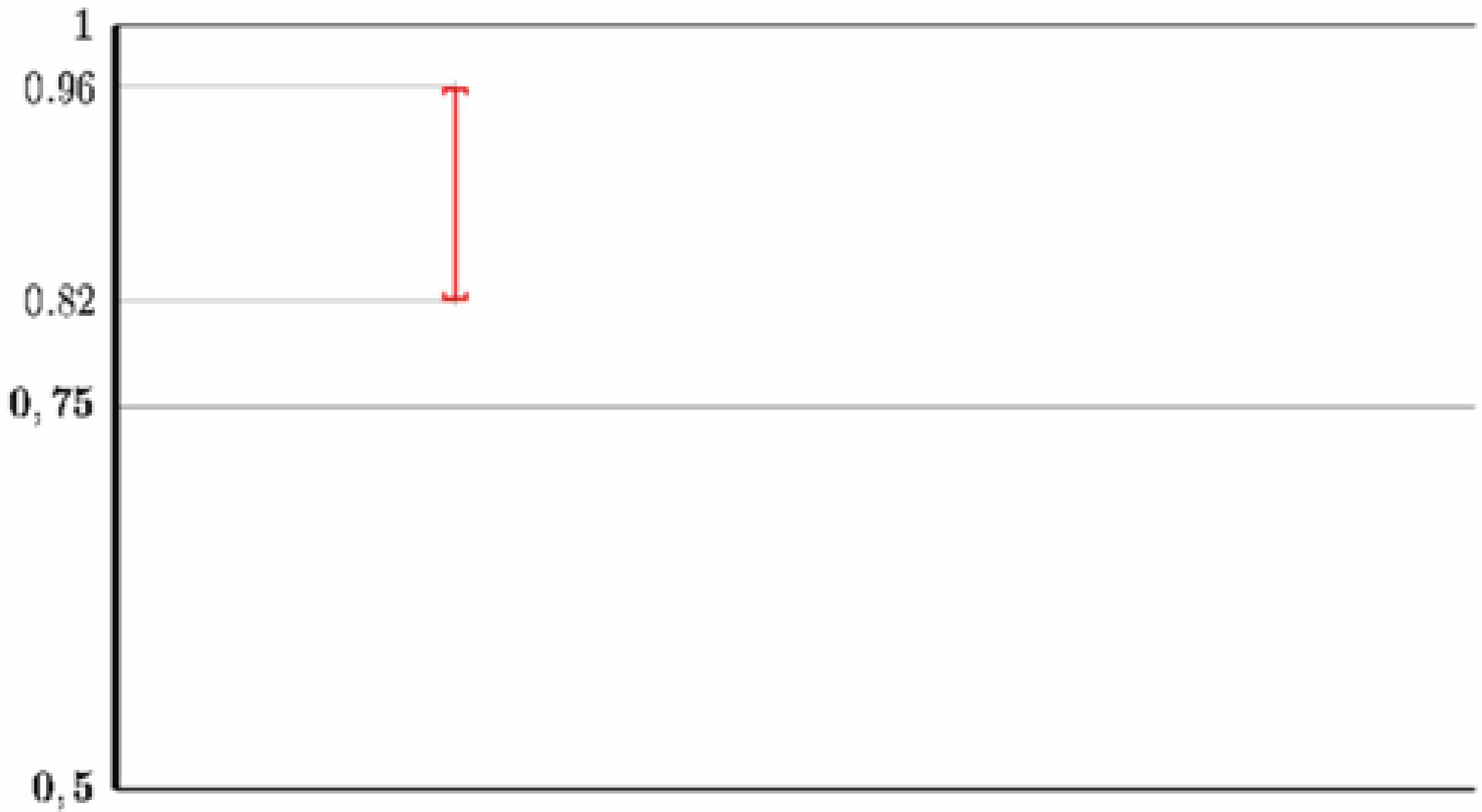
On dit que l'intervalle [0,82 ; 0,96] est l'intervalle de confiance au niveau de confiance 0,95 pour la proportion *p.*

Intervalle de confiance

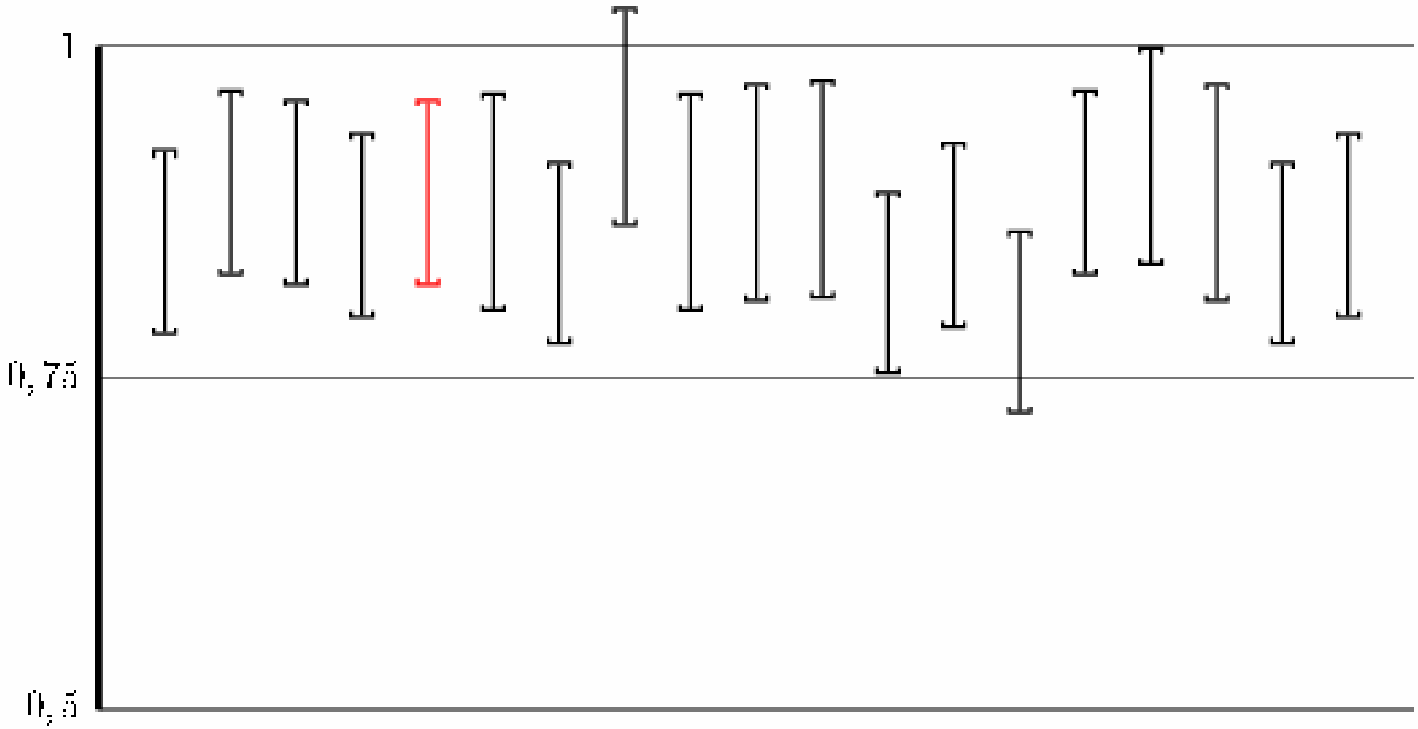
- Une fois que l’échantillon est connu, il n’y a plus d’aléatoire: p est ou n’est pas dans l’intervalle [0,82 ; 0,96]
- <u>Rédaction:</u>

« p est dans l’intervalle [0,82 ; 0,96]
avec un niveau de confiance de 0,95 ».

Intervalle de confiance



Intervalle de confiance



Intervalle de confiance

- Avant de fabriquer un intervalle de confiance, je peux dire qu’il aura une probabilité de 0,95 de contenir p

- Une fois l’échantillon prélevé, l’intervalle de confiance est connu, il n’y a plus de hasard.

Intervalle de confiance

<u>Evolution d'une fréquence</u>

- Après un changement de procédé de fabrication, le producteur souhaite savoir s'il a réussi à améliorer la qualité.
- Un nouveau test sur 250 produits donne 93% de fonctionnement correct.
- L'intervalle de confiance associé est [0,86 ; 1].

Intervalle de confiance

<u>Evolution d'une fréquence</u>

- Avant changement de procédé de fabrication: [0,82 ; 0, 96]
- Après un changement de procédé de fabrication [0,86 ; 1].
- On ne peut pas affirmer qu'il y ait eu un progrès car il se peut par exemple que p=0,9 *avant* et *après* le changement !

Deux objectifs différents

Intervalle de fluctuation	**Intervalle de confiance**
permet de prendre une décision	permet d'estimer une proportion
Rejet de l'hypothèse au risque de 5%:  « Les marins bretons ont le même groupe sanguin que le reste de la population ».	La proportion p de feux fonctionnant correctement est dans l'intervalle [0,82 ; 0, 96] **avec un niveau de confiance de 0,95.** 

Deux objets différents

Intervalle de fluctuation	Intervalle de confiance
On considère une population ayant un caractère p connu ou sur laquelle on formule une hypothèse.	On considère une population ayant un caractère ***p*** **inconnu**
Si on prélève plusieurs échantillons alors la fréquence f du caractère va varier.	On réalise un intervalle de confiance à partir d'une fréquence f observée sur un échantillon.
L'intervalle de fluctuation est l'intervalle centré sur p pour lequel on a environ 95% de chances d'avoir f	Un intervalle de confiance est aléatoire et dans cette expérience aléatoire on a environ 95% de chance d'avoir p