ÉCHANTILLONNAGE – LE FIL ROUGE

Auteur : Boubakeur Kacimi TI-Nspire™ CAS

Mots-clés : statistiques, échantillonnage.

Fichier associé : intervalles de confiance.tns

1. Objectifs

• Rappeler le lien entre les intervalles donnés en Seconde, Première et Terminale.
• Sur un même exemple, traiter l'intervalle de fluctuation à 95 %, en Seconde, Première et Terminale et constater qu'il peut y avoir des interprétations différentes sous certaines conditions.
• Manipuler la TI-Nspire CAS.

2. Énoncé

Une urne contient 100 boules indiscernables au toucher, on sait que 30 boules sont bleues et 70 sont rouges. Si la boule tirée est bleue, le joueur a gagné.
Sur 50 personnes qui ont participé au jeu, seulement 8 ont gagné, peut-on raisonnablement soupçonner le jeu d'être truqué ?

Théorie statistique de l'estimation
On appelle épreuve de Bernoulli, une expérience qui n'a que deux issues possibles S et E.
On pose P(S) = p et P(E) = $1 - p = q$.

Si on analyse un échantillonnage de taille n approximativement normal ($n \geq 25$) et que l'on observe, à chaque fois, la fréquence d'apparition f de l'issue, on s'aperçoit que,
- lorsque la probabilité $p \in [0{,}2\ ;\ 0{,}8]$ est connue, au moins 95 % des fréquences se situent dans l'intervalle de fluctuation $I = \left[p - 1{,}96\sqrt{\frac{p(1-p)}{n}}\ ;\ p + 1{,}96\sqrt{\frac{p(1-p)}{n}} \right]$ (1).
- lorsque la probabilité p n'est pas connue, on peut affirmer que p est située dans l'intervalle de confiance $\left[f - 1{,}96\sqrt{\frac{f(1-f)}{n}}\ ;\ f + 1{,}96\sqrt{\frac{f(1-f)}{n}} \right]$ avec une probabilité d'erreur inférieure à 5 %.

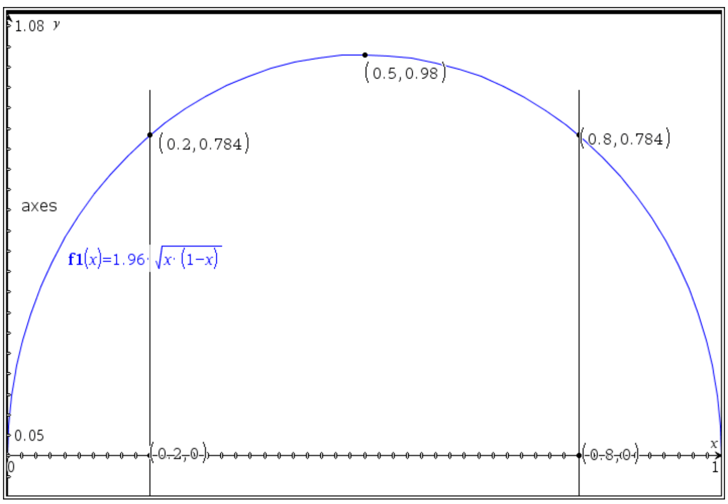
3. Commentaires

Depuis 2009, les programmes de mathématiques insistent sur les intervalles de fluctuations et de confiance aux trois niveaux.

Le programme de seconde donne l'intervalle de fluctuation
$I = \left[p - \frac{1}{\sqrt{n}}\ ;\ p + \frac{1}{\sqrt{n}} \right]$ (2).

Le lien entre les formules (1) et (2) vient du fait que dans (2) on majore $1{,}96\sqrt{p(1-p)}$ par 1.

On peut néanmoins constater que, lorsque p est situé dans l'intervalle [0,2 ; 0,8], alors $1{,}96\sqrt{p(1-p)}$ est dans l'intervalle [0,784 ; 0,98] ; autant dire que la majoration par 1 n'est pas mauvaise.



Le programme de première

On considère la variable aléatoire X dénombrant les succès obtenus au cours des n épreuves de Bernoulli. X suit une loi binomiale de paramètres n et p.

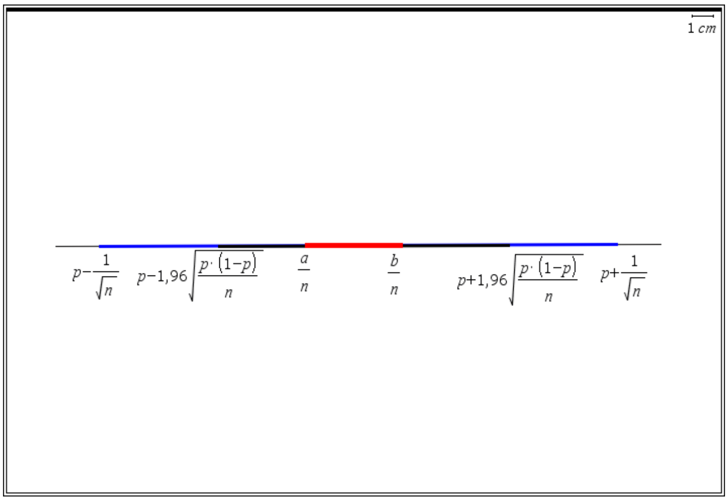
p étant connue, dans ce cas on parle d'intervalle de fluctuation à 95 % : $I = \left[\frac{a}{n} ; \frac{b}{n}\right]$, avec a et b tels que :

- a est le plus petit entier vérifiant $P(X \leq a) > 0{,}025$ (3)
- b est le plus petit entier vérifiant $P(X \leq b) \geq 0{,}975$.

Le lien entre les formules (1) et (3) vient du fait que, dans (3),

d'une part, on majore : $p - 1{,}96\,\frac{\sqrt{p(1-p)}}{\sqrt{n}}$ par $\frac{a}{n}$ et, d'autre part, on minore : $p + 1{,}96\,\frac{\sqrt{p(1-p)}}{\sqrt{n}}$ par $\frac{b}{n}$.

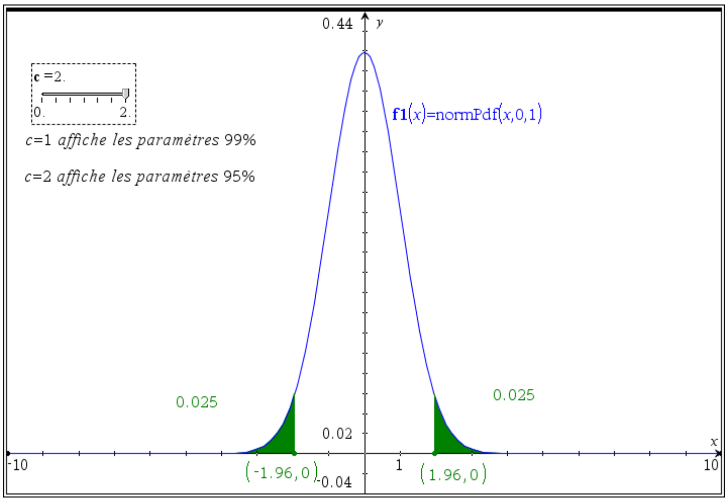
L'image mentale à retenir :



4. Conduite de l'activité

L'activité est plutôt à mener avec le logiciel TI Nspire CAS, avec un affichage mode ordinateur.

1) Dans un premier temps, on peut constater que pour une loi normale centrée réduite, on retrouve bien les paramètres 1,96 pour un seuil de 95 % et 2,58 pour un seuil de 99 %.

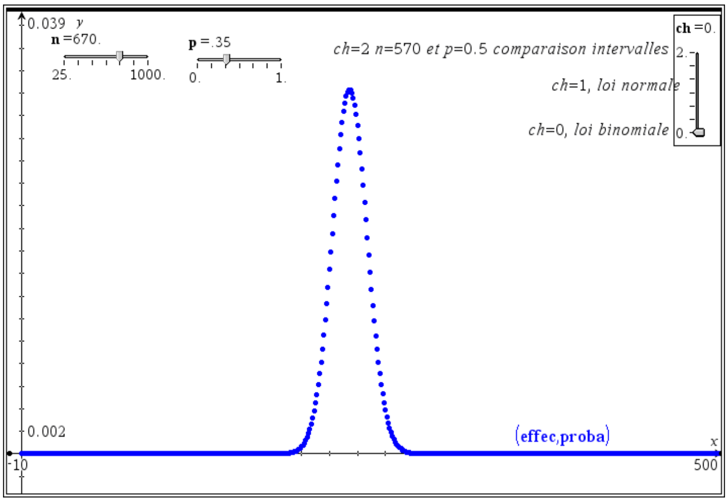


2) On poursuit le déroulement par :

a. l'animation d'une loi binomiale de paramètres *n* et *p*

On profitera de cette animation pour constater :

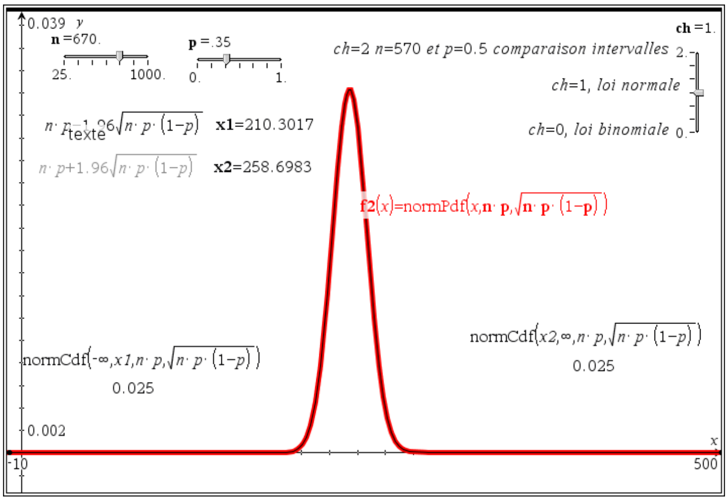
- l'allure gaussienne de ce nuage de points,
- la relative symétrie du nuage de points par rapport à l'axe d'équation $x = \mu = n \times p$,
- la cohérence du déplacement de cette cloche lorsque n et p varient, l'abscisse du sommet étant $n \times p$ l'amplitude de la cloche est : $\binom{n}{\mathrm{Ent}(n \times p)} p^{\mathrm{Ent}(n\times p)} (1-p)^{n-\mathrm{Ent}(n\times p)}$.



b. l'animation d'une loi normale de paramètres $\mu = n \times p$ et $\sigma = \sqrt{n \times p \times (1 - p)}$

On mettra l'accent sur le fait que la loi n'est autre que la fonction de densité

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$



c. la comparaison des intervalles de fluctuations

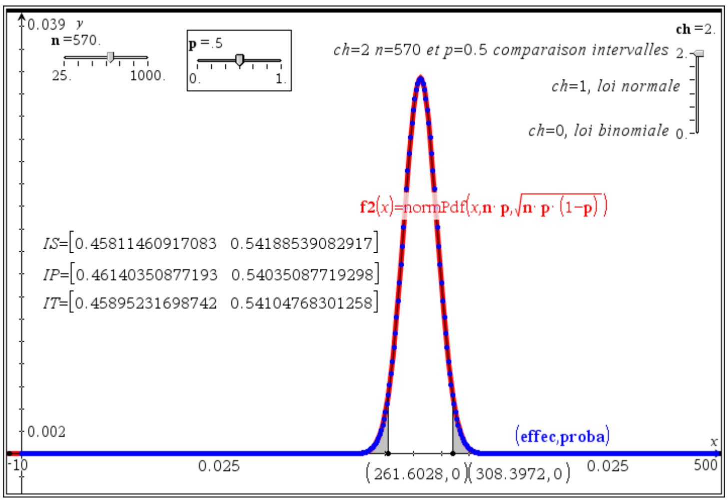
On veillera à faire attention aux valeurs n et p pour l'illustration de la situation.

Notations utilisées :

IS pour intervalle seconde,

IP pour intervalle première,

IT pour intervalle terminale.



Conséquence

On peut trouver des situations :

- validées en seconde mais rejetées en première et en terminale,
- rejetées en première mais validées en terminale.

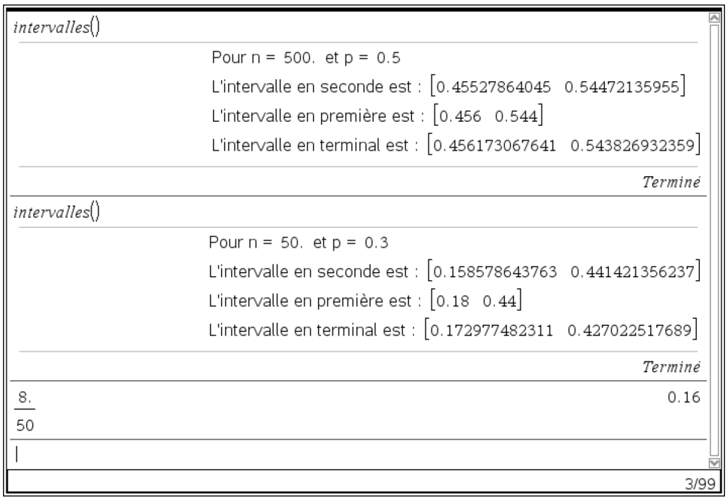
On peut avantageusement mettre en évidence l'interactivité des différentes parties du logiciel.

5. Retour au problème initial

Dans notre cas la probabilité de gagner est connue, p = 0,3 ($p \in$ [0,2 ; 0,8]), n = 50, les conditions sont remplies pour déterminer l'intervalle de fluctuation ; pour cela, il suffit de :
- régler les curseurs aux bonnes valeurs soit : n = 50 et p = 0,3
- exécuter le programme « intervalles »,
- faire l'analyse

« 0,16 $\in$ *IS*, en seconde le jeu n'est pas truqué au seuil de 95 %.
0,16 $\notin$ *IP* et 0,16 $\notin$ *IT* … »

On retrouve bien entendu les inclusions de l'image mentale.



Complément

En exécutant le programme « intervalles », on peut constater que la différence entre les trois intervalles devient d'autant plus faible que les valeurs de n sont grandes.