Fluctuation et confiance au lycée

Louis-Marie Bonneval(*)

Les nouveaux programmes de mathématiques du lycée font la part belle aux statistiques inférentielles.
La prise en compte de ce domaine était sans doute nécessaire, compte tenu de sa place grandissante tant dans les mathématiques que dans leurs applications et leurs enjeux citoyens.
Néanmoins la façon dont se fait cette évolution suscite de sérieuses critiques :
• elle se fait au détriment de la géométrie, dont la place n'a nullement diminué dans les mathématiques vivantes et leurs applications, et dont le rôle formateur est irremplaçable ;
• elle est d'une ambition démesurée compte tenu du niveau mathématique moyen des élèves du lycée ; par exemple c'est mettre la barre bien haut de parler de fluctuation d'échantillonnage à des élèves de Seconde qui maîtrisent mal les intervalles ou les proportions ; de loi binomiale à des élèves de Première qui découvrent les variables aléatoires ; de loi normale à des élèves de Terminale qui découvrent l'intégration.
• elle fait l'impasse sur la formation des enseignants, ce qui est dramatique quand on sait que la plupart des professeurs ont peu étudié cette branche des mathématiques dans leur formation initiale.
• elle se fait – comme souvent – dans l'urgence : les programmes de Seconde (respectivement de Première) on été mis en œuvre avant que les programmes de Première (respectivement de Terminale) ne soient connus ; les professeurs ont donc dû travailler à l'aveugle, sans disposer d'une vue d'ensemble.

Cette vue d'ensemble, nous l'avons désormais, et je peux maintenant être plus positif : la progression de la Seconde à la Terminale me paraît cohérente. C'est ce que je voudrais développer.

Entre réalité et modèle

Rappelons que les statistiques inférentielles comportent essentiellement deux branches : les tests et l'estimation.
– les **tests** visent à évaluer la pertinence d'un modèle probabiliste. Par exemple, lorsqu'on envisage le sexe d'un enfant à naître, on admet souvent l'équiprobabilité entre les deux issues, masculin et féminin : au vu des observations statistiques, ce modèle est-il adéquat ?
– l'**estimation** vise à déterminer les paramètres d'un modèle probabiliste adapté. Par exemple, si le test précédent conduit à rejeter l'équiprobabilité, quelles valeurs faut-il attribuer à la probabilité de naissance d'un garçon, d'une fille ?

(*) lm.bonneval@gmail.com

Ces deux démarches expliquent l'expression « statistiques inférentielles » : à partir de l'observation statistique d'un échantillon, on *infère* un modèle[(1)], c'est-à-dire qu'on le construit (estimation) et on le confronte à la réalité (test).
Ce va-et-vient entre réalité et modèle est général en Science, on le pratique aussi dans les domaines déterministes : ainsi en géométrie [2] quand on représente une chambre par un rectangle, ou en analyse quand on ajuste une fonction à des données expérimentales.
Ces comparaisons peuvent aider à comprendre que la réponse à un test ne peut être simplement oui ou non, et que la réponse à une estimation ne peut être une valeur exacte : l'une comme l'autre dépendent du *degré de précision* retenu. Mais ce qui est propre aux phénomènes aléatoires, c'est que se superpose à cette imprécision un second degré de « flou » : le *niveau de confiance* qu'on peut accorder à la réponse.

Les programmes du lycée limitent les tests et l'estimation à un seul paramètre, à savoir la probabilité d'un événement[(2)]. Cette limitation est sage :
– c'est la question la plus naturelle et la plus courante pour modéliser un phénomène aléatoire ;
– c'est déjà bien assez difficile pour des lycéens ;
– c'est suffisant pour exposer la problématique des statistiques inférentielles.
Mais, dans l'enseignement supérieur, les statistiques inférentielles traitent aussi d'autres problèmes : test d'équiprobabilité (qui figurait au programme précédent de terminale S), estimation des paramètres d'une loi normale, test d'indépendance de deux variables aléatoires, …

Observons que le mot « test » n'apparaît pas dans le libellé des programmes, sans doute par crainte de dérives théoriques excessives. Mais c'est bien de cela qu'on parle quand on écrit « prise de décision » : il s'agit de décider si le modèle est adapté ou non, plus précisément si la fréquence observée est compatible avec la probabilité supposée.

Précision et confiance

Pour savoir si ma chambre est rectangulaire, je mesure ses côtés et ses angles (ou ses côtés et ses diagonales). Or l'instrument que j'utilise a une précision limitée, mes conclusions seront donc valables à ce degré de précision.
De même, pour tester ou estimer une probabilité, on va la mesurer. Comment peut-on mesurer la probabilité p d'un événement relatif à une épreuve e ? En répétant e le plus de fois possible, et en observant la fréquence d'apparition f de cet événement. On conçoit que plus le nombre de répétitions est élevé, plus est précise l'approximation de p par f. Ou du moins : plus elle a de *chances* d'être une approximation précise.

(1) ou seulement un paramètre de ce modèle
(2) que par commodité on convient d'appeler « succès » (même s'il s'agit d'un événement malheureux).

Formalisons un peu : appelons e^n l'épreuve qui consiste à effectuer n fois l'épreuve e (une issue de e^n s'appelle(3) *un échantillon de taille n* de l'épreuve e) ; notons F la variable aléatoire relative à e^n qui à chaque échantillon associe la fréquence de succès dans cet échantillon. Il est essentiel de bien distinguer la variable aléatoire F, relative à l'ensemble des échantillons possibles *avant* expérimentation, et sa réalisation f, observée sur l'échantillon obtenu *après* expérimentation.

Dire que f est une valeur approchée de p à ε près signifie que $|f-p|<\varepsilon$. Cela conduit à étudier l'événement $\{|F-p|<\varepsilon\}$ relatif à l'épreuve e^n. Posons $P(|F-p|<\varepsilon)=1-\alpha$: ε mesure la *précision* de l'approximation, $1-\alpha$ le *niveau de confiance* qu'il faut accorder à cette approximation.
On souhaite bien entendu ε le plus petit possible, mais aussi α le plus petit possible.
• Pour p et ε donnés, en augmentant n on peut faire baisser α, c'est-à-dire augmenter le niveau de confiance (c'est le théorème faible des grands nombres).
• Pour p et α donnés, en augmentant n on peut faire baisser ε (à niveau de confiance donné, plus on répète l'épreuve plus on gagne en précision).
• Mais, pour p et n fixés, ε et α varient en sens inverse (ce qu'on gagne en précision, on le perd en confiance, et inversement).

Qu'est-ce qui différencie le test de l'estimation ?

Dans les deux cas, on se donne une valeur de α suffisamment petite (souvent 0,05, parfois 0,01), seuil de probabilité en dessous-duquel un événement sera considéré comme *rare* ; et on observe sur un échantillon de taille n la fréquence f de succès.

– **Lors d'un test, *p* est connu**, ce qui veut dire qu'on a choisi un modèle probabiliste et on veut savoir s'il est pertinent. Par exemple on a supposé l'équiprobabilité entre garçons et filles à la naissance.

La condition $P(|F-p|<\varepsilon)=1-\alpha$ peut s'écrire $P(F\in\,]p-\varepsilon,p+\varepsilon[)=1-\alpha$, soit $P(F\notin\,]p-\varepsilon,p+\varepsilon[)=\alpha$: il est *rare* que F soit hors de l'intervalle $]p-\varepsilon,p+\varepsilon[$, dit *intervalle de fluctuation*(4) de F au niveau de confiance $1-\alpha$.
Par conséquent, si la fréquence observée f est hors de cet intervalle, on est fondé à rejeter l'hypothèse que p est la probabilité de succès, avec l'argument suivant : il s'est produit un événement qui serait rare sous cette hypothèse. C'est en quelque sorte un raisonnement par l'absurde, dans une logique probabiliste.
Le problème mathématique est de déterminer cet intervalle de fluctuation en fonction des données (probabilité de succès, taille de l'échantillon, niveau de confiance choisi), autrement dit de calculer ε connaissant α, n et p.

(3) par référence à des tirages dans l'urne de Bernoulli qui modélise l'épreuve e (tirages qu'il faut supposer *avec remise*, de façon à répéter la *même épreuve* en garantissant l'*indépendance*).
(4) Cette dénomination à visée pédagogique est récente mais bienvenue.

– **Lors d'une estimation, *p* est inconnu**, ce qui veut dire qu'on cherche un modèle probabiliste. Par exemple on cherche la probabilité de naissance d'un garçon, d'une fille.

La condition $P(|F-p|<\varepsilon)=1-\alpha$ peut s'écrire $P(p\in\,]F-\varepsilon,F+\varepsilon[)=1-\alpha$, soit $P(p\notin\,]F-\varepsilon,F+\varepsilon[)=\alpha$: il est *rare* que l'intervalle aléatoire $]F-\varepsilon,F+\varepsilon[$ ne contienne pas le nombre p.

On est donc fondé à penser que l'intervalle $]f-\varepsilon,f+\varepsilon[$, dit *intervalle de confiance*(5) au niveau $1-\alpha$, contient la valeur p inconnue. On dispose donc pour p d'un encadrement plausible.
Le problème mathématique est de déterminer cet intervalle de confiance en fonction des données (fréquence observée, taille de l'échantillon, niveau de confiance choisi), autrement dit de calculer ε connaissant α, n et f.

Déterminer l'intervalle de fluctuation

Si α, n et p sont connus (cas du test), on peut en principe déterminer ε. En effet la loi de probabilité de F est connue : le nombre de succès nF suit la loi binomiale $B(n,p)$.
En pratique cette loi est malcommode, pour deux raisons :
- elle nécessite beaucoup de calculs, car on s'intéresse aux grandes valeurs de n ;
- sa fonction de répartition est en escalier, donc non inversible.

Le premier obstacle (difficulté des calculs) a conduit De Moivre en 1733 (20 ans après la publication de l'*Ars conjectandi* de Jakob Bernoulli) à rechercher des approximations utilisant l'exponentielle. Un siècle plus tard, Laplace a fait le lien entre ce travail et la loi de probabilité introduite par Gauss pour modéliser les erreurs de mesure : pour n suffisamment grand, la loi binomiale peut être approchée par la loi de Gauss, bien plus commode pour les calculs.
Aujourd'hui les ordinateurs ont aplani l'obstacle du calcul. La loi binomiale étant introduite en Première, on peut donc l'exploiter directement, à l'aide de la calculatrice ou du tableur.

Mais le deuxième obstacle (non inversibilité de la fonction de répartition) subsiste : pour le contourner, on est contraint à quelques contorsions techniques un peu délicates pour les élèves de Première. Au lieu de chercher ε tel que $P(|F-p|<\varepsilon)=1-\alpha$, on recherche le plus petit ε tel que $P(|F-p|<\varepsilon)$ dépasse $1-\alpha$; pour cela on doit revenir à la variable aléatoire nF, ce qui conduit, au lieu de chercher le rayon ε de l'encadrement, à en chercher séparément les deux bornes au moyen d'un algorithme. Par suite l'intervalle de fluctuation n'est pas toujours centré en p, et il n'a pas la même probabilité selon qu'il est ouvert, semi-ouvert ou fermé.

(5) Ce nom désigne aussi l'intervalle aléatoire $]F-\varepsilon, F+\varepsilon[$, mais au lycée cette notion serait un peu délicate…

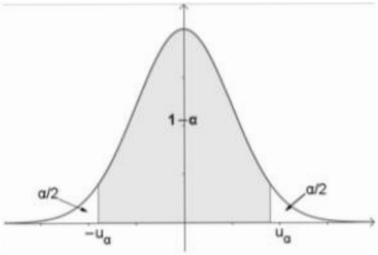
La loi normale élimine cet obstacle, ce qui – outre son intérêt théorique incontestable – justifie qu'on l'introduise en Terminale. Mais en contrepartie elle soulève une autre difficulté : elle n'est qu'une approximation de la loi binomiale, approximation dont la précision (qui dépend de p et de n) est difficile à quantifier.

Quant au programme de Seconde, il propose une approche expérimentale appuyée sur la simulation. L'obstacle à ce niveau est d'ordre didactique : le passage de l'expérimentation à l'énoncé d'une règle empirique est une démarche de sciences expérimentales, inhabituelle en mathématiques.

Comparaison des trois approches

Je n'entre pas dans le détail des programmes, abondamment commentés dans les documents ressources. Les trois approches sont résumées dans le tableau ci-dessous, inspiré du document ressource de Terminale [5] (page 36).

	TEST : intervalle de fluctuation	ESTIMATION : intervalle de confiance
SECONDE	Si $\alpha = 0{,}05$, $n \geq 25$ $0{,}2 \leq p \leq 0{,}8$ $\varepsilon \approx \dfrac{1}{\sqrt{n}}$	« sensibilisation »
PREMIÈRE	algorithme (loi binomiale)	hors programme
TERMINALE	Si $n \geq 30$ $np \geq 5, n(1-p) \geq 5$ $\varepsilon \approx u_\alpha \sqrt{\dfrac{p(1-p)}{n}}$	Si $n \geq 30$ $nf \geq 5, n(1-f) \geq 5$ $\varepsilon \approx u_\alpha \sqrt{\dfrac{f(1-f)}{n}}$



Dans les formules de Terminale, u_α désigne le rayon de l'intervalle centré en 0 et de probabilité $1 - \alpha$ pour la loi normale centrée réduite ; c'est aussi l'antécédent de $1 - \dfrac{\alpha}{2}$ par la fonction de répartition de cette loi (fonction qui a le bon goût d'être continue et strictement croissante).

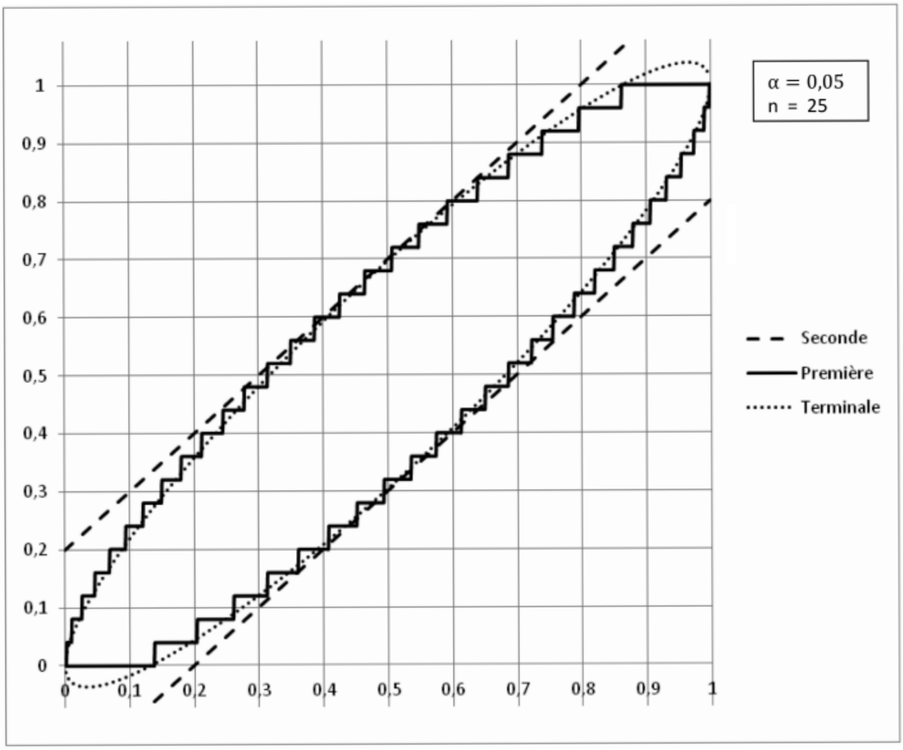
En particulier $u_{0{,}05} \approx 1{,}96$ et $u_{0{,}01} \approx 2{,}58$.

Un schéma me semble très éclairant(6). Il est suggéré dans les documents ressources (Première, page 47-48), et développé dans [5]. Voir page suivante.

Sur ce graphique, α et n sont fixés : on s'est donné un niveau de confiance $1 - \alpha$

(6) pour nous enseignants, pas pour les élèves !

(ici $1 - \alpha = 0{,}95$), et on connaît la taille n de l'échantillon (ici $n = 25$). En abscisse est indiquée la valeur de p (probabilité de succès), en ordonnée la valeur de f



(fréquence de succès observée sur l'échantillon).
Trois courbes sont dessinées, dont la construction est expliquée plus loin :

– une paire de droites parallèles : elle provient du rayon $\dfrac{1}{\sqrt{n}}$ et correspond au programme de Seconde ;
– une ligne brisée fermée : elle provient de la loi binomiale, et correspond au programme de Première ;
– une ellipse : elle provient de la loi normale, et correspond au programme de Terminale.
Chacune de ces courbes délimite une zone, que j'appellerai respectivement *zone de Seconde*, *zone de Première*, *zone de Terminale*.

On peut exploiter ce schéma à la manière d'un abaque, de deux façons :
- **Test** : p est connu, on peut tracer la droite verticale d'abscisse p : son intersection avec la zone de Première détermine l'intervalle de fluctuation.
- **Estimation** : f est connu, on peut tracer la droite horizontale d'ordonnée f : son intersection avec la zone de Première détermine l'intervalle de confiance.

Notons que le schéma comporte des points pour lesquels f est extérieur à [0,1]. Bien entendu ces points n'ont pas de sens dans notre contexte. On les a néanmoins dessinés pour montrer les limites des approximations exposées plus loin.

Intervalle exact ou approché ?

Dans les deux cas, l'intervalle exact est fourni par la zone de Première. Les deux autres zones en fournissent une approximation(7).

On constate que cette approximation est meilleure au centre du graphique que sur les bords : c'est ce qui motive les critères $np \geq 5$, $n(1-p) \geq 5$ indiqués dans le programme de Terminale. Ces critères ont un côté arbitraire, ils dépendent de ce qu'on considère comme une bonne approximation, et certains livres en proposent d'autres. Rappelons que, pour n donné, quand p est proche de 0 ou de 1, la loi de Poisson fournit une meilleure approximation de la loi binomiale que la loi normale.

Il est intéressant d'écrire cette double condition $np \geq 5$, $n(1-p) \geq 5$ sous la forme $p \in \left[\frac{5}{n};1-\frac{5}{n}\right]$, car on voit que l'amplitude de cet intervalle de validité augmente avec n. Pour $n = 25$, cet intervalle est [0,2 ; 0,8], soit celui du programme de Seconde : cela explique sans doute la valeur 25 indiquée comme plancher dans ce programme (pour p dans cet intervalle, si l'approximation est bonne quand $n = 25$, elle l'est a fortiori quand n est plus grand). On peut s'étonner que ce plancher ne soit pas le même qu'en Terminale ; c'est sans doute dû au manque de visibilité que j'ai dénoncé au début ; mais il faut aussi noter qu'on ne parle pas de la même approximation : en Terminale, il s'agit d'approcher la ligne brisée par l'ellipse, pour tout α ; en Seconde, il s'agit d'approcher la ligne brisée par la bande, dans le seul cas où $\alpha = 0{,}05$.

On observe aussi qu'au centre du graphique la zone de Première déborde légèrement par endroits la zone de Seconde : cela indique que, $\mathrm{P}\left(|\mathrm{F}-p| < \frac{1}{\sqrt{n}}\right)$ peut être inférieure à 0,95 pour certaines valeurs de p proches de 0,5. Cette propriété subsiste pour des valeurs de n assez grandes (de l'ordre de 500). En revanche, sur les bords du graphique, la zone de Seconde déborde largement la zone de Première : autrement dit pour p proche de 0 ou de 1, $\mathrm{P}\left(|\mathrm{F}-p| < \frac{1}{\sqrt{n}}\right)$ peut être bien supérieure à 0,95.

Ainsi pour $n \geq 25$, la condition $0{,}2 \leq p \leq 0{,}8$ n'est ni nécessaire ni suffisante pour que

(7) La valeur $n = 25$ est intéressante d'une part parce que $\frac{1}{\sqrt{n}} = 0{,}2$, d'autre part parce que les deux approximations (de Seconde et de Terminale) sont convenables sans être excellentes : on voit bien ce qui se passe.

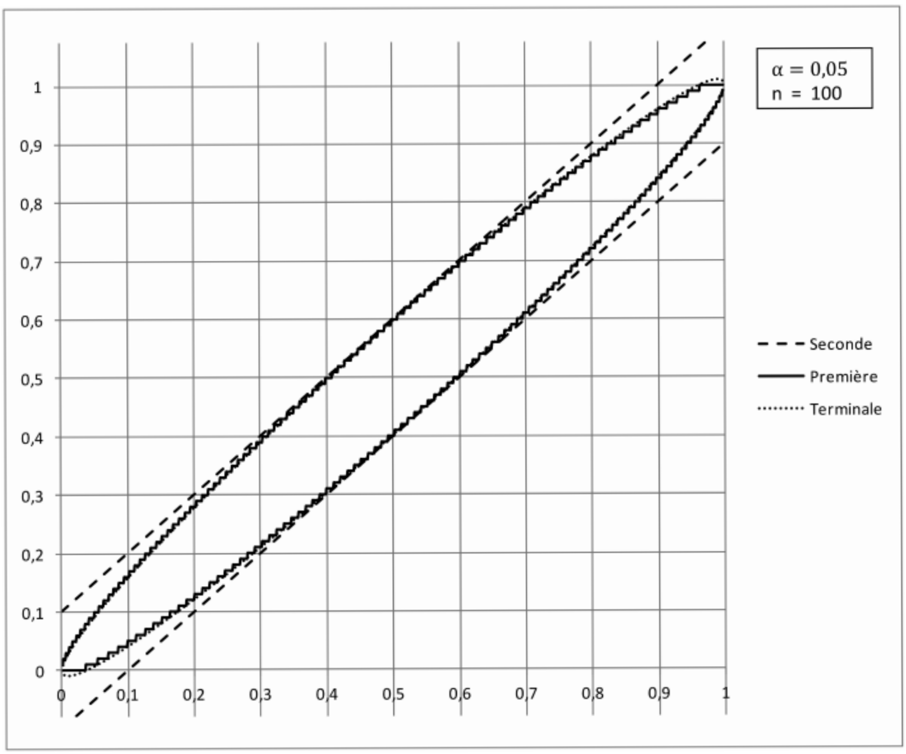
$P\left(|F-p|<\dfrac{1}{\sqrt{n}}\right)$ *dépasse* 0,95 [1][9]. En revanche, et c'est ce qui importe, elle permet d'affirmer que cette probabilité *est voisine* de 0,95.

Expression de l'intervalle de confiance

La première bissectrice du repère n'est pas l'axe de symétrie de l'ellipse (elle ne contient pas les sommets), mais elle en est très proche, d'autant plus proche que n est grand (on le constatera un peu plus loin en augmentant n).
Or, comme l'indique la dernière cellule du tableau vu plus haut, le passage de l'intervalle de fluctuation à l'intervalle de confiance se fait usuellement en permutant simplement les rôles de p et f. Cela revient à remplacer l'ellipse par sa symétrique par rapport à cette première bissectrice. C'est à nouveau une approximation(8), dont on voit graphiquement qu'elle est bonne au centre du graphique mais mauvaise sur les bords.

Rôle de la taille de l'échantillon

Si, avec le même niveau de confiance 0,95, on fait passer à 100 la valeur de n, on obtient le nouveau schéma ci-dessous :



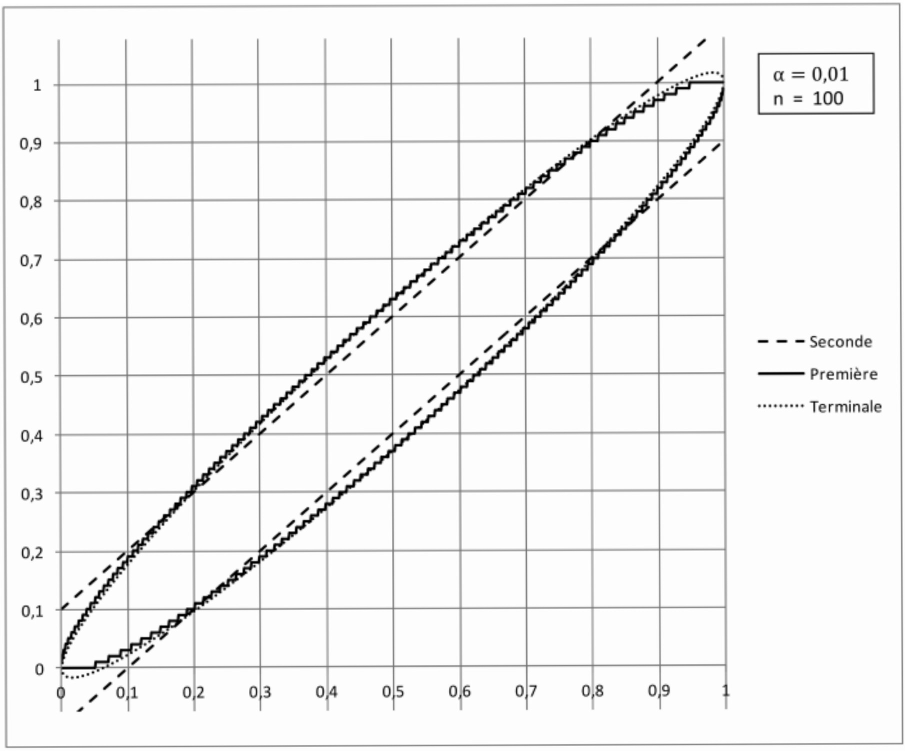
(8) On peut bien sûr traiter le problème par le calcul : voir par exemple [10], p. 306-307.

On constate un resserrement des intervalles : en quadruplant n, on a doublé la précision.

On constate aussi que l'approximation de la loi binomiale par la loi normale est bien meilleure : la ligne brisée tend à se confondre avec l'ellipse. Si on passait à $n = 1000$ (taille habituelle d'un échantillon lors d'un sondage d'opinion) on ne distinguerait plus les deux courbes.

Rôle du niveau de confiance

Si maintenant, en gardant $n = 100$, on choisit $\alpha = 0{,}01$, autrement dit un niveau de confiance 0,99, on obtient le schéma suivant :



Les deux zones de Première et de Terminale, tout en restant très proches l'une de l'autre, se sont dilatées : on a perdu en précision ce qu'on a gagné en confiance.

Quant à la zone de Seconde, elle n'est plus adéquate : elle ne l'était qu'au niveau de confiance 0,95. Si on voulait l'adapter, il faudrait se rappeler que $\dfrac{1}{\sqrt{n}}$ est une approximation (par excès) de $\dfrac{u_{0,05}}{2\sqrt{n}}$, et donc le remplacer par $\dfrac{u_{0,01}}{2\sqrt{n}}$, soit $\dfrac{1,3}{\sqrt{n}}$.

Comment ont été construites ces courbes au tableur[9] ?

Les zones de Seconde et de Terminale sont faciles à construire, puisqu'elles résultent de formules explicites. On peut d'ailleurs les construire à l'aide d'un traceur de courbes comme Geogebra.

- La zone de Seconde est définie par la double inégalité $p-\frac{1}{\sqrt{n}}<f<p+\frac{1}{\sqrt{n}}$:

elle est limitée par les deux droites d'équation $f=p-\frac{1}{\sqrt{n}}$ et $f=p+\frac{1}{\sqrt{n}}$.

- La zone de Première n'a pas une équation aussi simple. Avec Excel ou Open Office, la formule =CRITERE.LOI.BINOMIALE(n,p,u) renvoie le plus petit x pour lequel $\mathrm{P}(\mathrm{Z}\leq x)\geq u$. En donnant à u respectivement les valeurs $\frac{\alpha}{2}$ et $1-\frac{\alpha}{2}$, on obtient pour nF les bornes inférieure et supérieure de l'intervalle de fluctuation (ouvert à gauche, fermé à droite) ; en divisant ces deux nombres par n, on obtient les bornes de l'intervalle de fluctuation pour F.

- La zone de Terminale est définie par la double inégalité

$$p-u_\alpha\sqrt{\frac{p(1-p)}{n}}<f<p+u_\alpha\sqrt{\frac{p(1-p)}{n}},$$

où le nombre u_α est fourni par la formule

$$=\text{LOI.NORMALE.STANDARD.INVERSE}\left(1-\frac{\alpha}{2}\right).$$

Cette zone est limitée par les deux courbes d'équations $f=p-u_\alpha\sqrt{\frac{p(1-p)}{n}}$ et $f=p+u_\alpha\sqrt{\frac{p(1-p)}{n}}$, dont la réunion a pour équation $(f-p)^2=\frac{u_\alpha^2}{n}p(1-p)$. En remplaçant p et f par x et y, on obtient $(y-x)^2=\frac{u_\alpha^2}{n}x(1-x)$: on reconnaît une ellipse.

Cette ellipse a pour centre de symétrie $\left(\frac{1}{2},\frac{1}{2}\right)$, comme d'ailleurs les deux autres zones : cela traduit le fait que la situation est équivalente si on remplace p par $1-p$ et f par $1-f$ (autrement dit si on permute les rôles du succès et de l'échec).

(9) Les fichiers *fluctuconf.xlsx* (Excel) et *fluctuconf.ods* (Open office) sont disponibles sur le site de l'APMEP en complément à cet article.

On peut remarquer qu'en ses deux points d'abscisse $\frac{1}{2}$, cette ellipse admet des tangentes[10] d'équations respectives $y = x - \frac{u_\alpha}{2\sqrt{n}}$ et $y = x + \frac{u_\alpha}{2\sqrt{n}}$. Or pour $\alpha = 0{,}05$, $u_\alpha \approx 1{,}96 \approx 2$: on retrouve les deux bords de la zone de Seconde, extérieurs à l'ellipse mais très proches d'elle au centre du graphique.

Observons pour conclure que la géométrie, et en particulier l'étude des coniques, rend service même là où on ne l'attend pas, au cœur des statistiques inférentielles…

Bibliographie

[1] BONNEVAL Louis-Marie, *Intervalle : de confiance ?* BV 427.

[2] BONNEVAL Louis-Marie, *Géométrie ou probabilités : une même démarche de modélisation ?* BV 456.

[3] Document ressource Seconde
http://media.eduscol.education.fr/file/Programmes/17/9/Doc_ressource_proba-stats_109179.pdf

[4] Document ressource Première
http://media.eduscol.education.fr/file/Mathematiques/59/6/Ressource_Statistiques_Probabilites_1eres_208596.pdf

[5] Document ressource Terminale
http://media.eduscol.education.fr/file/Mathematiques/11/5/LyceeGT_ressources_Math_T_proba-stat_207115.pdf

[6] DUCEL Yves et SAUSSEREAU Bruno, *La prise de décision de la Seconde à la Première*, Repères-IREM n° 85.

[7] GRES (ENFA), *Abaques et intervalles de confiance*, Bulletin n° 5 (juin 1997)
http://www2.enfa.fr/r2math/math/res-ped/GRES/B5/Ab-int-conf.pdf

[8] GRIHON Pierre, *De la fluctuation à la confiance, une synthèse des nouveaux programmes de probabilités au lycée*, deux conférences sur le site de l'APMEP,
http://www.apmep.asso.fr/De-la-fluctuation-a-la-confiance

[9] MAILLARD Christian : compléments à l'article [1], BV 436 p.732-733.

[10] SAPORTA Gilbert, *Probabilités, analyse de données et statistiques* (Technip, 2000).

(10) Les deux diamètres d'équations $y = x$ et $x = \frac{1}{2}$ sont conjugués.