REMARQUES SUR L'ENSEIGNEMENT DES PROBABILITÉS ET DE LA STATISTIQUE AU LYCÉE

Daniel PERRIN[1]

1 Introduction

Il peut paraître étrange qu'un mathématicien très ignorant en probabilités et statistique juge opportun d'émettre quelques idées sur leur enseignement au lycée et je ne l'aurais pas fait si cet enseignement n'était pas venu concurrencer, voire supplanter, d'autres domaines auxquels je suis fondamentalement attaché : la géométrie, les équations différentielles, etc. De ce point de vue, les choses ont bien changé. Il fut un temps (par exemple celui de mes études) où les disciplines de l'aléatoire n'existaient tout simplement pas dans l'enseignement, ni secondaire, ni supérieur : on pouvait être agrégé de mathématiques sans jamais avoir rencontré le mot *probabilités*. Je dis tout de suite que je considère que c'était un scandale, que le mépris de certains éminents mathématiciens de l'époque à l'égard des probabilités et de la statistique était stupide et qu'il fallait évidemment rectifier tout cela. Le problème c'est que, comme l'histoire le montre souvent, lorsqu'une communauté est opprimée, elle n'a de cesse que de devenir elle-même oppresseur. Il me semble que c'est un peu ce qui s'est passé dans les dernières années avec les probabilités et la statistique dans l'enseignement du second degré : elles étaient absentes, elles sont devenues dominantes. Je considère aussi que la façon dont cette mutation s'est faite (les réformes de programmes, pour parler clair) n'a pas été un modèle de démocratie et de transparence. C'est la raison initiale qui a fait que je me suis penché sur le sujet. On pourra évidemment me suspecter d'hostilité primaire à l'égard de tout ce qui touche à l'aléatoire. Je crois que ce n'est pas vrai pour au moins quatre raisons :

1. Je suis convaincu que probabilités et statistique sont un domaine essentiel des mathématiques qui, de surcroît, a beaucoup d'applications[2] et et que, par ailleurs, la place prise par les statistiques dans la vie courante nécessite une réflexion de tous les citoyens à leur propos.

2. Je trouve que la fabrication des programmes n'est pas suffisamment *rationnelle*. Fort de cette opinion, j'essaie honnêtement de ne pas être moi-même inutilement partisan.

3. J'ai fait l'effort d'apprendre un peu de probabilités et de statistique afin d'essayer de comprendre vraiment les tenants et aboutissants du sujet. En un mot, je suis ignorant, certes, mais je me soigne.

[1]Professeur à l'ESPE de l'Académie de Versailles, membre du comité scientifique des Instituts de Recherche sur l'Enseignement des Mathématiques, daniel perrin@math u-psud fr

[2]Mais ce n'est pas le seul !

4. Enfin, il se trouve que j'ai beaucoup d'amis, certains depuis fort longtemps, parmi les probabilistes et les statisticiens et, comme je ne veux pas me brouiller avec eux, je réfléchis à deux fois avant de porter des jugements trop catégoriques sur leurs disciplines.

Les points principaux que j'évoquerai ici concernent les intervalles de fluctuation (introduits en seconde) et de confiance (en terminale) et l'introduction de la loi normale (en terminale[3]).

2 La loi normale et le théorème de Moivre-Laplace

2.1 L'énoncé

L'objectif principal de cette introduction de la loi normale est de pouvoir utiliser le théorème de Moivre-Laplace, notamment pour approcher la loi binomiale. Je ne conteste pas que ce théorème est un magnifique résultat, essentiel en probabilités. Je le rappelle pour bien préciser les notations :

Théorème 2.1 (Moivre-Laplace). *Soit X_n une variable aléatoire qui suit la loi binomiale de paramètres n, p, avec $p \in]0,1[$. On pose $Z_n = \dfrac{X_n - np}{\sqrt{np(1-p)}}$ (variable centrée réduite associée). On a la formule :*

$$\lim_{n \to +\infty} \mathbf{P}(a \leq Z_n \leq b) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-t^2/2} dt.$$

Si l'on pose $N(z) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{z} e^{-t^2/2} dt$, le second membre est égal à $N(b) - N(a)$.

2.2 Intervalle de fluctuation

Cela étant, dans les programmes du lycée, quelle est la principale application de ce théorème ? Si on suit le document d'accompagnement, on lit : *Le théorème de Moivre-Laplace va permettre de donner un intervalle de fluctuation calculable directement, sous réserve que n soit assez grand.*

En particulier, il vise à justifier l'intervalle de fluctuation donné en seconde et que je rappelle, en citant le programme : *L'intervalle de fluctuation au seuil de 95%, relatif aux échantillons de taille n, est l'intervalle centré autour de p, proportion du caractère dans la population, où se situe, avec une probabilité égale à 0,95, la fréquence observée dans un échantillon de taille n. Cet intervalle peut être obtenu, de façon approchée, par simulation. Le professeur peut indiquer aux élèves le* **résultat**[4] *suivant, utilisable dans la pratique pour des échantillons de taille $n > 25$ et des proportions p du caractère comprises entre $0,2$ et $0,8$: si f désigne la fréquence du caractère dans l'échantillon, f appartient à l'intervalle $\left[p - \frac{1}{\sqrt{n}}, p + \frac{1}{\sqrt{n}}\right]$ avec une*

[3] J'ai beaucoup utilisé le document d'accompagnement publié par Eduscol : *Ressources pour la classe terminale, Probabilités et statistiques*. Même si je ne suis pas d'accord sur l'objectif, je trouve que ce texte est remarquable.

[4] C'est moi qui souligne.

probabilité d'au moins $0,95$. *Le professeur peut faire percevoir expérimentalement la validité de cette propriété mais elle n'est pas exigible.*

Si je traduis le paragraphe précédent, en interprétant le mot *résultat* comme *théorème*, il s'agit de l'assertion suivante :

Théorème 2.2. ([5] ? ?) *Soit X_n une variable aléatoire suivant une loi binomiale de paramètres n, p. On a la formule :*

$$\mathbf{P}\left(p - \frac{1}{\sqrt{n}} \leq \frac{X_n}{n} \leq p + \frac{1}{\sqrt{n}}\right) \geq 0,95$$

pourvu qu'on ait les conditions[6] *:* $n > 25$ *et* $0,2 < p < 0,8$.

2.3 Le lien avec la loi normale

On note d'abord que la probabilité qui intervient dans le programme de seconde s'exprime en termes de $Z_n = \dfrac{X_n - np}{\sqrt{np(1-p)}}$.

Proposition 2.1. *On a l'égalité :*

$$\pi_{n,p} := \mathbf{P}\left(p - \frac{1}{\sqrt{n}} \leq \frac{X_n}{n} \leq p + \frac{1}{\sqrt{n}}\right) = \mathbf{P}\left(-\frac{1}{\sqrt{p(1-p)}} \leq Z_n \leq \frac{1}{\sqrt{p(1-p)}}\right).$$

Comme p est dans $]0,1[$, on a $\frac{1}{\sqrt{p(1-p)}} \geq 2$, donc l'intervalle considéré contient $[-2,2]$ et la probabilité $\pi_{n,p}$ cherchée est $\geq \mathbf{P}(-2 \leq Z_n \leq 2)$ (ce cas correspond exactement à $p = 1/2$).

En négligeant l'écart entre la loi normale et la binomiale, on a donc :

$$\pi_{n,p} \geq N(2) - N(-2).$$

Or, on a $N(2) \simeq 0,9772$ et $N(-z) = 1 - N(z)$ d'où $N(-2) \simeq 0,0227$ et $N(2) - N(-2) \simeq 0,9545$. On voit qu'avec cette approximation on a bien le résultat annoncé en 2.2.

2.4 Retour sur le théorème 2.2

Le problème de ce résultat c'est qu'il est faux : pas très faux peut-être, mais faux tout de même.

En effet, pour tester le domaine de validité de l'assertion :

$$\mathbf{P}(np - \sqrt{n} \leq X_n \leq np + \sqrt{n}) \geq 0,95,$$

on peut utiliser n'importe quel logiciel de calcul formel qui permet d'afficher la fonction de répartition d'une loi binomiale, par exemple le logiciel généraliste *xcas* (ou les logiciels spécialisés de statistique comme R). On peut même utiliser une calculatrice. Par exemple, avec la

[5] Les points d'interrogation ont la même signification que dans la transcription des parties d'échecs ...

[6] Une variante utilisée en terminale impose $n \geq 30$, $np \geq 5$ et $n(1-p) \geq 5$.

Voyage 200, on trouve instantanément, pour $n = 41$ et $p = 1/2$:

$$\mathbf{P}(np - \sqrt{n} \le X_n \le np + \sqrt{n}) = \sum_{k=15}^{26} \frac{\binom{41}{k}}{2^{41}} \simeq 0,9404.$$

On voit que, bien que les conditions d'application de 2.2 soient largement remplies, la probabilité est $< 0,95$.

On va sans doute m'objecter que je chipote : 41 n'est qu'une valeur particulière et $0,94$ c'est presque $0,95$.

Sur le premier point, on peut faire le calcul avec *xcas*, en prenant $p = 1/2$, $1/3$, etc. Même avec $p = 1/2$, il y a 76 valeurs $n > 30$ pour lesquelles la probabilité en question est $< 0,95$, la plus grande d'entre elles (pour $n \le 5000$) étant 528. On a même une probabilité $< 0,945$ pour $n = 34, 41, 48, 55, 62, 71, 80, 89$. La pire valeur[7] pour $n \ge 30$ étant pour 41.

Si l'on fait varier p (je l'ai fait de $0,2$ à $0,8$ avec un pas de $0,05$), on trouve d'autres valeurs fautives ; par exemple, pour $n = 30$, $p = 0,45$, on trouve $0,935$. Au-delà de $n = 90$, on ne trouve plus de valeurs de $\pi_{n,p}$ en-dessous de $0,945$.

Alors, bien sûr, $0,94$ ou $0,95$, ce n'est pas très différent[8]. Mais, il faut savoir si l'on fait des mathématiques ou non. Il est des situations où une petite erreur peut entraîner des conséquences importantes[9]. Pour résumer, je trouve tout de même abusif de donner un tel résultat à des élèves sans dire clairement[10] qu'il n'est pas vraiment correct. **Attention**, je ne dis pas qu'il ne faut pas donner ce type de formule aux élèves, au contraire, voir §5 ce que je propose. Ce que je dis, c'est qu'il faut le faire sans les tromper sur la marchandise.

2.5 Majorer l'erreur dans Moivre-Laplace ?

La raison de cette situation, c'est qu'il y a une vraie difficulté mathématique. On a vu que l'intervalle de fluctuation du programme de seconde a été obtenu en confondant la loi binomiale et son approximation normale. Son domaine de validité dépend donc de la majoration de l'erreur dans le théorème de Moivre-Laplace. Le problème c'est que les résultats sur ce sujet ne sont ni faciles, ni totalement satisfaisants. J'essaie de rassembler ici quelques éléments permettant de quantifier cette erreur. Mes sources principales (outre le livre de Feller *An Introduction to Probability Theory and Its Applications*) sont :

[7] Si l'on impose seulement $n \ge 25$, on trouve $0,9385$ pour $n = 29$.

[8] Je signale cependant que certains scientifiques mettent fortement en cause la borne $0,95$ en pensant qu'elle est trop faible pour garantir des résultats. Je cite un extrait du blog de Pierre Barthélémy : *Tout le travail de Valen Johnson a consisté à rapprocher les deux méthodes pour examiner si leurs critères de validation se recoupaient. Le résultat est fort instructif et pourrait ébranler le paradigme de « la valeur p inférieure à 0,05 ». A l'aune de l'approche bayésienne, ce seuil est tout simplement insuffisant. Pour résumer, une hypothèse qui passe de justesse sous cette barre n'a en réalité que de 3 à 5 chances contre 1 d'être vraie. Selon le statisticien, il se pourrait donc bien qu'une proportion non négligeable d'études se contentant de ce seuil soient tout simplement fausses. Valen Johnson estime qu'en étant optimiste, le phénomène pourrait concerner entre 17 et 25 % des articles en question ! Un taux qui, selon lui, serait cohérent avec la proportion de travaux dont on n'arrive pas à reproduire les résultats.*

[9] Voir par exemple, sur ma page web `http://www.math.u-psud.fr/~perrin/Conferences/logistiqueDP2.pdf` une discussion sur l'effet papillon.

[10] Je reconnais que le document d'accompagnement du programme de Terminale S explique cette difficulté de façon honnête (p. 25-27). En revanche, il dit un peu plus loin, p. 33, que lorsque $n \ge 30$, $np \ge 5$ et $n(1-p) \ge 5$ l'intervalle du programme de seconde est un intervalle de fluctuation au seuil de $0,95$, ce qui est faux.

[Suquet1] SUQUET Charles, *Initiation au calcul des probabilités*, http://www-gat.univ-lille1.fr/~suquet/ens/ICP/ICP0106.pdf ;

[Suquet2] SUQUET Charles, *Théorème limite central*, http://math.univ-lille1.fr/~suquet/Polys/TLC.pdf.

Examinons ce que donne le résultat qui, à ma connaissance, est le meilleur[11] connu à l'heure actuelle.

Théorème 2.3 (Berry-Esséen). *Soit X_n une variable aléatoire suivant la loi binomiale de paramètres n, p et soit $q = 1 - p$. On pose $F_n(x) = \mathbf{P}(X_n \leq x)$ et on note $N(x)$ comme ci-dessus la fonction de répartition de la loi normale centrée réduite. On a l'inégalité :*

$$F_n(x) - N\left(\frac{x-np}{\sqrt{npq}}\right) \leq \frac{0,4784(p^2+q^2)}{\sqrt{npq}}.$$

Remarque 2.1. Ce théorème n'est pas du tout trivial. La version initiale date de 1941 et la valeur $0,4784$ pour la constante est un résultat récent (Asov, 2011). Feller donne seulement une constante de l'ordre de 8.

Remarque 2.2. Dans la pratique, le théorème n'est pas encore bien fameux. Par exemple, dans le cas de l'approximation de la loi binomiale de paramètres $n = 30$ et $p = 1/2$, pour la valeur $F_n(x) = F_n(np + \sqrt{n})$ approchée par $N(2)$ il donne une erreur majorée par $0,4784/\sqrt{n} \simeq 0,0873$ alors que l'erreur effective ($\simeq 0,001363$) est 64 fois plus petite (on a $F_n(x) \simeq 0,9786$, $N(2) \simeq 0,9772$).

Corollaire 2.1. *On pose encore $q = 1 - p$. La différence entre $\mathbf{P}\left(p - \frac{1}{\sqrt{n}} \leq \frac{X_n}{n} \leq p + \frac{1}{\sqrt{n}}\right)$ et $N\left(\frac{1}{\sqrt{pq}}\right) - N\left(-\frac{1}{\sqrt{pq}}\right)$ est majorée par $\frac{0,9568(p^2+q^2)}{\sqrt{npq}}$.*

Remarque 2.3. Dans le cas $p = 1/2$ on a $\frac{1}{\sqrt{pq}} = 2$ et $p^2 + q^2 = 1/2$. L'erreur est majorée par $\frac{0,9568}{\sqrt{n}}$, ce qui n'est pas terrible. Pour avoir une erreur $\leq 0,0045$ (la différence entre $N(2) - N(-2)$ et le $0,95$ souhaité) il faut n de l'ordre de 45000 et non pas $n \geq 30$.

Remarque 2.4. Quand on compare les valeurs de la loi binomiale, toujours dans le cas précédent, avec son approximation normale $0,9545$, on voit que la différence est nettement moindre que ce qu'annonce Berry-Esséen (pour $n = 1000$ la majoration donne une erreur de $0,03$. En réalité elle est au moins 10 fois moindre). Autrement dit, le calcul direct de l'intervalle de fluctuation est bien meilleur que ce qu'annonce la loi normale corrigée par Berry-Esséen. **D'un point de vue mathématique, ce n'est donc pas l'approximation par la loi normale qui peut justifier rigoureusement[12] l'intervalle de fluctuation donné en seconde.**

[11] Le lecteur vérifiera que le théorème d'Uspensky, voir [Suquet2] cor. 34, donne un résultat légèrement moins bon.

[12] Il n'y a rien de plus dur à justifier rigoureusement qu'un résultat faux !

3 Les arguments contre l'introduction de la loi normale en terminale

3.1 La définition

Je trouve désagréable de parler de choses que l'on n'a pas définies. Ici, il s'agit des intégrales impropres, indispensables pour parler de la loi normale. Certes, on admet assez facilement que l'aire sous la courbe en cloche est finie bien que l'intervalle soit infini, mais ce n'est tout de même pas évident pour des lycéens qui viennent d'avoir leur premier contact avec l'intégrale.

3.2 La preuve de Moivre-Laplace

Il est aussi un peu gênant que le théorème principal, celui qui gouverne tout le reste, soit hors de portée d'un lycéen. Cela conduit à l'appliquer comme une recette que l'on est obligé d'admettre et d'admirer sans la comprendre. Je sais qu'on peut donner des approches, notamment graphiques, du théorème. C'est bien. On peut même (et le document d'accompagnement a tenu cette gageure) en donner une preuve « élémentaire », soi-disant au niveau d'un élève de terminale. Mais, je considère que ce n'est pas sérieux. Pour qu'elle soit au niveau du lycée, les auteurs de cette preuve sont obligés de la découper en tout petits bouts et l'ensemble n'est pas vraiment compréhensible.

3.3 L'approximation

Je ne reviens pas sur ce point, déjà largement détaillé. Comme on ne contrôle pas la qualité de cette approximation on arrive à des affirmations incorrectes, comme celles dénoncées ci-dessus. Comme mathématicien, un peu borné, j'ai tendance à ne pas beaucoup aimer les théorèmes faux.

3.4 Du passé faisons table rase

En fait, dans les programmes, on n'utilise essentiellement le théorème que dans le cas $-2 \leq Z_n \leq 2$ et en parachutant les valeurs de $N(2)$ et $N(-2)$. Bien entendu, il est tout à fait possible de l'utiliser avec d'autres valeurs, mais cela nécessite de tabuler la fonction N.

Mais alors, que fait-on ? On approche la loi binomiale par la loi normale **dans un but calculatoire**, en utilisant le fait qu'on dispose d'une tabulation de la fonction de répartition de la loi normale. Mais ce souci est d'un autre âge ! D'abord, on ne s'interroge pas sur la manière de produire cette table de valeurs, et ensuite, c'est devenu inutile car il n'y a aujourd'hui, avec les outils actuels, aucune difficulté à déterminer les valeurs d'un intervalle de fluctuation pour une loi binomiale. En fait, les auteurs du programme font comme si le seul outil de calcul à notre disposition était la table[13] des valeurs de la loi normale. **C'est un peu comme si l'on**

[13] Je ne suis pas le seul à défendre cette position. Voilà ce que dit Bernard Prum dans l'avertissement de son livre *La démarche statistique*, Cépaduès 2010 : *Une des idées qui courent tout au long de cet ouvrage est la primauté sans cesse renforcée des « calculs exacts », par opposition aux approximations asymptotiques. Il est loin le temps*

s'en tenait à l'usage d'une table de logarithmes pour faire les calculs usuels d'analyse et de trigonométrie. D'ailleurs, le document d'accompagnement (qui sait par ailleurs l'accent mis sur l'utilisation des machines) ne prône pas vraiment l'utilisation de la loi normale pour le calcul :

Quand $n \geq 30$, $np \geq 5$, $n(1-p) \geq 5$, il est courant de faire les calculs impliquant une variable binomiale en la remplaçant par une variable suivant une loi normale de mêmes espérance et variance. Seul le programme de STI2D-STL mentionne cette pratique, qui ne doit donc pas être mise en œuvre dans les autres filières où tous les calculs de probabilités se font à la calculatrice en utilisant la loi exacte (au programme), quelle qu'elle soit.

Un autre argument en ce sens c'est que l'approximation normale est encore moins bonne lorsque p est très petit ou très grand, alors que le calcul direct avec la loi binomiale reste possible.

3.5 Où l'on marche vraiment sur la tête

On peut tout à fait discuter tous les arguments que je viens de donner et persister à vouloir introduire la loi normale en terminale, par exemple dans un but de culture : la gaussienne c'est une idée fondamentale[14], etc. Soit, mais là où ça ne va plus du tout c'est qu'ensuite, en classes préparatoires, les seules probabilités qu'on rencontre sont discrètes ! Inutile de dire que la loi normale sera totalement oubliée des étudiants s'ils ne la rencontrent plus. Pourtant, là, on disposerait de tous les outils pour prouver (par exemple) le théorème de Moivre-Laplace : les intégrales impropres, la formule de Stirling et même cet outil hyper-sophistiqué... l'intégration par parties !

4 Intervalles de fluctuation et de confiance

4.1 Intervalles de fluctuation

Les programmes actuels introduisent ces deux notions en seconde et en terminale, respectivement. J'ai déjà évoqué les intervalles de fluctuation ci-dessus, mais j'y reviens ici car il me semble que la racine des difficultés évoquées plus haut est la volonté de disposer d'intervalles de fluctuation **explicites** au sens suivant :

Définition 4.1. *Soit X une variable aléatoire qui suit une loi binomiale de paramètres n, p et soit α un réel de $]0,1[$. On appelle* **intervalle de fluctuation** *au seuil de $1-\alpha$ un intervalle de la forme $I = [A^-, A^+]$ qui vérifie $\mathbf{P}(A^- \leq X \leq A^+) \geq 1-\alpha$. Un intervalle de fluctuation* **explicite** *sera donné par des fonctions* **explicites** *$A^-(n,p,\alpha)$ et $A^+(n,p,\alpha)$, si possibles régulières (disons monotones et continues), à valeurs dans $[0,p]$ et $[p,1]$.*

où l'on devait consulter quelques tables sur papier (la gaussienne, les χ^2) ; aujourd'hui on peut – et bien plus rapidement – accéder à d'innombrables lois exactes grâce aux ordinateurs. Pourquoi alors délaisser une solution plus juste et plus rapide ?

[14] Comme le dit Francis Galton : *Elle règne avec sérénité et en toute abnégation au milieu de la confusion sauvage.*

(Je reprends essentiellement ici les notations de Jean-Pierre Raoult dans son billet paru sur le site *Images des mathématiques* :
http://images.math.cnrs.fr/Intervalle-de-confiance-pourquoi.html.)

Comme on le voit, la notion d'intervalle de fluctuation est conceptuellement assez simple. Le problème, comme on le reverra plus loin, c'est que la fonction de répartition de la loi binomiale est discontinue, de sorte que les intervalles de fluctuations naturels (i.e. minimaux) ne sont pas donnés par des fonctions régulières. Il faut donc se contenter d'intervalles plus grands, mais, à ma connaissance, il n'existe **aucune** fonction A^-, A^+ régulière vraiment satisfaisante, même en fixant $\alpha = 0,05$. Le programme de seconde donne $A^- = p - 1/\sqrt{n}$ et $A^+ = p + 1/\sqrt{n}$ mais on a vu que la probabilité correspondante peut être $< 0,95$. Le programme de terminale donne $A^-(n,p,0,05) = p - 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}}$ et $A^+(n,p,0,05) = p + 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}}$ et la probabilité est encore plus petite, puisque l'intervalle est plus petit.

4.2 Intervalles de confiance

Si la notion d'intervalle de fluctuation ne semble pas poser de problème conceptuel majeur, il n'en va pas de même de celle d'intervalle de confiance. Pour s'en convaincre on ira lire l'article polémique de Pierre Colmez (http://images.math.cnrs.fr/Mathematiques-post-modernes.html) sur le site *Images des mathématiques* et la savoureuse dispute qui s'ensuit dans les commentaires[15]. Pourtant, en discutant de ce sujet avec des collègues statisticiens, je me suis laissé convaincre qu'il s'agissait bien d'une notion essentielle. Je rappelle donc ici ce que j'en ai compris, en partant d'un exemple, et ce que j'en tire du point de vue de l'enseignement.

4.2.1 Définition

On suppose qu'on a une pièce avec une probabilité p de tomber sur pile et $1-p$ sur face. La probabilité p est parfaitement déterminée mais **inconnue**. On cherche à la déterminer par un sondage S qui consiste à lancer n fois la pièce et à observer le nombre de piles. La variable aléatoire sur l'ensemble de tous les sondages de cardinal n possibles associe à chaque S le nombre $nf(S) \in [0,1]$ de piles. On est donc dans la situation d'une loi binomiale de paramètres n,p.

La problématique de l'intervalle de confiance est la suivante. On considère **un** sondage S particulier et son résultat $f(S)$ et on appelle **intervalle de confiance à** 95% **associé à ce sondage** un intervalle $J(S)$ qui vérifie la propriété suivante : $\forall p \in [0,1], \ \mathbf{P}_p(J(S) \ni p) \geq 0,95$. Cela signifie que, si l'on regarde tous les sondages S possibles, dans 95% des cas, le p inconnu va être dans l'intervalle $J(S)$ fourni par le sondage.

4.2.2 Le calcul au lycée

Il s'agit de calculer $J(S)$, à partir de la seule information connue qui est $f(S)$. Autrement dit, on cherche à écrire $J(S) = [C^-(f(S)), C^+(f(S))]$ avec des fonctions C^- et C^+, si possible expli-

[15] Voir aussi, dans les numéros de janvier et avril 2014 de la revue Repères IREM, les articles intitulés *Calcul de risques de première et seconde espèce à travers un exemple* et *Le risque des statistiques*.

cites. Supposons qu'on dispose d'un intervalle de fluctuation explicite $I = [A^-(n,p), A^+(n,p)]$ à 95% près, associé à la loi binomiale $\mathrm{B}(n,p)$. Bien entendu, cet intervalle reste inconnu tant que p l'est. Par définition, la probabilité que $f(S)$ soit dans I est $\geq 95\%$. Si les fonctions $A^-(p)$ et $A^+(p)$ (à n fixé) sont croissantes et continues, elles ont des inverses $C^+(f)$ et $C^-(f)$, on a l'équivalence :

$$f \geq A^-(p) \Leftrightarrow p \leq C^+(f) \quad \text{et} \quad f \leq A^+(p) \Leftrightarrow p \geq C^-(f).$$

On obtient donc ainsi un intervalle de confiance $J(S) = [C^-(f(S)), C^+(f(S))]$. C'est ce que suggèrent les programmes et le cas facile (évidemment !) est celui du fameux intervalle du programme de seconde : si $A^-(n,p) = p - \frac{1}{\sqrt{n}}$, on a $C^+(f) = f + \frac{1}{\sqrt{n}}$ et de même de l'autre côté. L'intervalle de confiance est donc $J(S) = [f(S) - 1/\sqrt{n}, f(S) + 1/\sqrt{n}]$.

Les choses sont déjà plus ardues avec l'intervalle de fluctuation donné en terminale :

$$A^-(n,p,0.05) = p - 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}}$$

(et de même pour A^+), car le calcul des fonctions réciproques n'est pas si facile. En posant $k = 1,96$ on trouve[16] :

$$p = \frac{2fn + k^2 \pm k\sqrt{k^2 + 4nf(1-f)}}{2(n+k^2)}.$$

On en déduit un développement asymptotique :

$$p = f \pm k\frac{\sqrt{f(1-f)}}{\sqrt{n}} + o\left(\frac{1}{\sqrt{n}}\right),$$

ce qui donne la valeur utilisée « dans d'autres champs disciplinaires » comme dit le document ressource.

Là où les choses se compliquent c'est dans le traitement du domaine de validité de ces intervalles. On a vu que l'intervalle de fluctuation du programme de seconde en $1/\sqrt{n}$ n'est valable que sous certaines conditions. Admettons même qu'il le soit dans dans le domaine $n \geq 25$ et $0,2 \leq p \leq 0,8$. On en déduit l'intervalle de confiance $[f - \frac{1}{\sqrt{n}}, f + \frac{1}{\sqrt{n}}]$, mais sous quelles conditions ? La position courante des manuels est d'imposer la même condition sur f que sur p, par un « transfert » dont la justification est loin d'être claire. Le plus raisonnable est sans doute d'imposer *a priori* une limitation au domaine de variation de p. Je renvoie au billet de J.-P. Raoult sur ce sujet.

4.2.3 Le calcul direct

Reprenons une loi binomiale X de paramètres n, p. On considère sa fonction de répartition (pour x réel avec $0 \leq x \leq n$) :

$$\mathbf{P}(X \leq x) = \sum_{k \leq x} \binom{n}{k} p^k (1-p)^{n-k}.$$

Cette fonction dépend évidemment de n et p, mais on considère ici que n est fixé et on en prend la variante « en fréquence » $F(p,f) = \mathbf{P}(X \leq nf)$ qui est une fonction définie sur le carré $[0,1]^2$. On constate deux choses :

[16] Mais ce calcul est loin d'être évident pour des élèves de terminale.

- à p fixé, la fonction F est croissante au sens large, mais **discontinue** par rapport à f ;
- à f fixé, la fonction F est **continue** par rapport à p et **décroissante** de 1 à 0.

Si l'on fixe un seuil α, les deux équations fondamentales qu'on a à traiter sont $F(p,f) \leq \alpha$ et $F(p,f) \geq 1-\alpha$ et il y a deux solutions selon qu'on regarde f ou p.

1) On fixe p. On cherche un intervalle de fluctuation. **S'il existe un f tel que** $F(p,f) = \alpha$ (ce qui n'est pas évident car F est discontinue en f et non surjective), on le note $A^-(p)$, c'est une fonction croissante de p et on a la condition attendue : $F(p,f) \leq \alpha \Leftrightarrow f \leq A^-(p)$. Si α n'est pas dans l'image, il faut prendre pour $A^-(p)$ le sup des f tels que $F(p,f) \leq \alpha$, mais évidemment, cette fonction n'est pas continue : trouver un intervalle de fluctuation explicite avec des fonctions régulières c'est un peu la quadrature du cercle.

2) On fixe f. On cherche un intervalle de confiance. Cette fois, plus de problème, comme F est continue et décroissante en p, il existe un réel $C^+(f)$ tel que $F(C^+(f),f) = \alpha$, la fonction C^+ est croissante et continue et on a l'équivalence $F(p,f) \leq \alpha \Leftrightarrow p \geq C^+(f)$. Bref, il n'y a pas de problème pour trouver une fonction[17] $C^+(f)$ (contrairement à A^- sa soi-disant réciproque). La difficulté est peut-être que cette fonction n'est pas en stock sur les logiciels de calcul usuels (par exemple *xcas*). Cependant, en utilisant la fonction de répartition de la binomiale et un solveur d'équations (ou deux lignes de programme) on peut la calculer aisément de manière approchée et cette difficulté s'estompe si l'on utilise de vrais logiciels de statistique.

4.2.4 Conclusion

À mon avis, la source d'un certain nombre de difficultés est dans le paradoxe suivant : la notion d'intervalle de fluctuation est simple à comprendre, mais, dans le cas d'une fonction de répartition discontinue, il ne peut y avoir de bonne fonction régulière qui le donne. La notion d'intervalle de confiance est plus subtile, mais, au moins en théorie, le calcul ne présente pas de difficultés. Là où l'on en introduit, c'est en voulant ramener le calcul de l'intervalle de confiance à celui de l'intervalle de fluctuation, ce qui revient à chercher une brave fonction continue comme réciproque d'une fonction discontinue.

On comprend alors la tentation de recourir à la loi normale, qui est continue, elle. J'ai critiqué ce choix plus haut et la solution que je préconise est plutôt d'utiliser plus et mieux les outils modernes de calcul. C'est d'ailleurs l'un des objectifs des programmes et on a ainsi des exemples de leur utilisation moins tristes que ceux qu'on trouve d'ordinaire.

5 L'importance du $\sqrt{n}$

5.1 L'intervalle de fluctuation vu en seconde

Je propose de garder l'assertion 2.2 relative à ce fameux intervalle, mais de manière **plus prudente** et plus honnête en disant (par exemple) :

[17] Bien connue des statisticiens qui l'appellent *fonction de vraisemblance cumulée*.

Pour des échantillons de taille $n > 25$ et des proportions p du caractère comprises entre $0,2$ et $0,8$: si f désigne la fréquence du caractère dans l'échantillon, f appartient à l'intervalle $\left[p-\frac{1}{\sqrt{n}}; p-\frac{1}{\sqrt{n}}\right]$ avec une probabilité **voisine** *de $0,95$.*

5.2 Pourquoi conserver le $\sqrt{n}$ et comment le justifier ?

Il me semble important en effet de conserver ce « résulta », ne serait-ce que pour attirer l'attention des élèves sur l'importance de l'ordre de grandeur $\sqrt{n}$ en probabilités[18]. Il est essentiel, en particulier, pour la formation du citoyen, de savoir que lorsqu'un sondage sur $n = 1000$ personnes donne un score de 51% des voix pour une élection entre deux personnes, la marge d'erreur étant de l'ordre de $\sqrt{n}$, donc de 31 voix, donc de 3%, le résultat final est loin d'être assuré.

Il se pose donc la question de la justification de ce $\sqrt{n}$, parachuté en seconde. Bien sûr, le théorème de Moivre-Laplace la donne, mais on peut aussi, en étant un peu moins ambitieux, comprendre l'irruption de ce terme grâce à la loi faible des grands nombres[19]. Là aussi, on utilise un résultat, mais incomparablement plus facile que Moivre-Laplace et qui ne nécessite pas d'utiliser de probabilités continues, ni d'intégrales impropres. Je résume la démarche pour montrer qu'elle est relativement élémentaire. On prouve l'inégalité de Bienaymé-Tchebychev, valable pour tout a (essentiellement, ce n'est rien d'autre que l'aire du rectangle) :

$$\mathbf{P}(\ X-E(X)\ > a) \leq \frac{V(X)}{a^2}\cdot$$

Cette inégalité se renverse en donnant :

$$\mathbf{P}(E(X)-a \leq X \leq E(X)+a) \geq 1-\frac{V(X)}{a^2}\cdot$$

On voit apparaître la variance, ou sa racine, l'écart-type $\sigma(X)$. Si on applique la formule avec $a = 2\sigma(X)$, on a déjà :

$$\mathbf{P}(E(X)-2\sigma(X) \leq X \leq E(X)+2\sigma(X)) \geq \frac{3}{4} = 0,75.$$

On peut ensuite appliquer l'inégalité avec comme variable la somme $X_n = Y_1 + \cdots + Y_n$ de n variables indépendantes et de même loi, disons celle d'une variable de Bernoulli Y, ce qui donne la loi faible des grands nombres :

$$\mathbf{P}\left(E(Y)-\frac{a}{n} \leq \frac{X_n}{n} \leq E(Y)+\frac{a}{n}\right) \geq 1-\frac{nV(Y)}{a^2}\cdot$$

En analysant cette formule, on voit qu'il y a deux exigences antagonistes :

[18]Je cite B. Prum : ... *sous de très larges hypothèses incluant l'indépendance, la variance d'une moyenne décroît en $1/n$ (n = taille de « l'échantillon »), et « la précision » est liée à l'écart-type, qui décroit donc en $1/\sqrt{n}$. C'est une notion importante pour comprendre l'imprécision des sondages, mais aussi la « précision » d'une mesure faite en TP de physique (la constante J de Joule, l'indice de réfraction de l'eau, la capacité d'un condensateur) : pour avoir 2 fois plus de précision, il faut 4 fois plus de mesures ; pour avoir 10 fois plus de précision, 100 fois plus de mesures.* Pour ma part, j'ai compris ce phénomène dans les années 1970 en tentant de calculer π par une simulation de l'expérience de l'aiguille de Buffon sur l'une des premières calculatrices programmables. Il fallait alors beaucoup de patience pour avoir deux décimales.

[19]Un autre pilier des probabilités, bizarrement à peine évoquée dans les programmes.

- on veut avoir un intervalle de fluctuation dont la largeur tende vers 0 avec n, donc avoir a infiniment petit par rapport à n ;
- on veut avoir une probabilité la plus grande possible donc le terme n/a^2 le plus petit possible, donc a^2 grand par rapport à n.

La solution est évidemment de prendre un a de l'ordre de $\sqrt{n}$. Par exemple, pour avoir une probabilité $\geq 0,75$, il suffit de prendre $a = 2\sigma(Y)\sqrt{n}$:

$$\mathbf{P}\left(\frac{X_n}{n} - E(Y) \ \leq \frac{2\sigma(Y)}{\sqrt{n}} \right) \geq 0,75.$$

On voit ainsi que le terme $\sqrt{n}$ apparaît naturellement, ainsi que le paramètre de dispersion $\sigma(Y)$. Alors, bien sûr, avec la loi binomiale de paramètre p, on trouve seulement :

$$\mathbf{P}\left(p - 2\frac{\sqrt{p(1-p)}}{\sqrt{n}} \leq \frac{X_n}{n} \leq p + 2\frac{\sqrt{p(1-p)}}{\sqrt{n}} \right) \geq 0,75,$$

au lieu du fameux 0,95, mais ce n'est déjà pas si mal[20] et cela présente trois avantages :

- on peut **démontrer** cette minoration en terminale ;
- on met en évidence le rôle de l'écart-type et on voit apparaître naturellement le terme $\sqrt{n}$;
- on peut justifier 2.2 en observant que, dans le cas de la binomiale, les bords décroissent particulièrement vite, et le tester expérimentalement.

6 Conclusion

Si j'ai écrit cet article dans *Statistique et enseignement*, c'est avant tout pour poser aux probabilistes et aux statisticiens des questions sur l'enseignement de leur discipline, dans l'espoir qu'eux-mêmes y réfléchissent sereinement en oubliant un peu de se comporter en groupe de pression.

Je ne cache pas mon opinion : je pense qu'il est souhaitable[21] de diminuer **un peu** la part des probabilités et de la statistique dans l'enseignement secondaire pour remettre à la place un peu plus d'analyse notamment[22]. Bien sûr, les modifications doivent être modérées, eu égard à l'investissement important des collègues des lycées dans le domaine de l'aléatoire.

[20] Comme Pascal Gamblin me l'a signalé, on obtient une majoration un peu meilleure (environ 0,86) en utilisant une astuce élémentaire de Bernstein, voir [Suquet1] p. 150.

[21] Comme disait un illustre mathématicien que le lecteur reconnaîtra aisément : *J espère qu on me croira sans peine si j ajoute que je n ai aucun intérêt personnel dans ces questions et qu il me chaut fort peu de savoir s il y aura une réforme ni ce qu en seront ses modalités. J ai simplement voulu verser au dossier un exemple de ce qu on pourrait faire si l on cherchait à agir de façon* **rationnelle**.

[22] Est-il normal qu'au lycée on n'entende plus parler d'équations différentielles, d'intégration par parties, de rotations et d'homothéties (planes), de barycentres ?

Ma proposition essentielle est de renvoyer à l'université (ou aux classes préparatoires) l'étude de la loi normale, de se concentrer au lycée sur les probabilités discrètes[23] (lois binomiale, de Poisson, géométrique ...) et d'utiliser ces connaissances de probabilités pour faire aussi de la statistique. À partir de ces exemples on peut en effet parler d'intervalle de fluctuation et d'intervalle de confiance[24] de manière pertinente. En ce qui concerne le calcul on a alors deux voies : l'utilisation des logiciels ou l'approximation en $\sqrt{n}$, essentielle, mais avec des limites abondamment évoquées ci-dessus.

En tous cas, s'il y a une chose dont je suis convaincu, c'est que la loi normale a toute sa place à l'université et en classes préparatoires. Je trouve vraiment incompréhensible que les nouveaux programmes n'aient introduit que des probabilités discrètes à ce niveau, alors que les programmes de Terminale faisaient la part belle aux lois continues. Il me semble qu'il aurait fallu faire exactement l'inverse.

Remerciements

Des versions préliminaires de ce texte ont été le support d'une « dispute » épistolaire très riche et fructueuse avec Bernard Prum et Jean-Pierre Raoult où nous avons pu mettre à plat nos nombreuses convergences et nos quelques divergences. Je les en remercie vivement. Je remercie aussi Michèle Artigue et Pascal Gamblin pour leurs remarques et leurs suggestions.

[23] Je ne pleurerais pas si l'on n'abordait plus les lois continues au lycée.
[24] Mais en étant conscient que cette notion est plus difficile.