CONTRIBUTION à la DÉMONSTRATION D'UNE CONJECTURE de Louis-Marie BONNEVAL

(bulletin n°427 de l'APMEP)

Il faut rendre grâce à Louis-Marie BONNEVAL d'avoir examiné **[1, page 161]** , avec un œil aussi neuf que sagace, le libellé du nouveau programme de Seconde sur l'approximation des lois binomiales par les lois normales. Ce libellé se contente en effet de reproduire, mot pour mot, ce qu'il faut bien appeler le fatras des « règles usuelles » que seul l'usage reconduisait d'année en année. Mais, je le dis tout de suite, ces règles avaient des raisons majeures de perdurer : premièrement, on ne sait pas comparer, pour la relation d'ordre, une suite de lois binomiales X_n avec sa loi normale limite, deuxièmement, comment traduire dans la pratique que n tend vers l'infini ?

Il faut donc louer notre collègue de s'être débarrassé du corset de la loi normale (laquelle doit bien sûr rester présente à l'esprit du professeur) et d'avoir dégagé, au terme de ses expérimentions numériques, des conjectures claires et précises dont la plus générale est :

$$p(|f-p| \le \frac{1}{\sqrt{n}}) > 0{,}9 \quad \text{ou} \quad p(|f-p| > \frac{1}{\sqrt{n}}) \le 0{,}1 \qquad \textbf{(1)}$$

où f = X/n et X binomiale de paramètres n et p (en abrégé B(n,p)).

Je n'en livrerai ici qu'une démonstration partielle et je vais expliquer pourquoi.

Comme le remarque Louis-Marie Bonneval, l'inégalité de Bienaymé-Tchebychev démontre **(1)** pour des valeurs extrêmes de p ($p \le 0{,}1$ ou $p \ge 0{,}9$). Mais cette inégalité est susceptible de généralisations qui, dans certains cas, permettent de resserrer les « fourchettes » de p : soit X une variable aléatoire de moyenne m prenant les valeurs x_i avec les probabilités p_i : on a $p(|X-m| > t) < \frac{M_n(X)}{t^n}$, où $M_n(x) = \sum_i p_i |x_i - m|^n$. Or il est possible de calculer M_4 pour une X binomiale B(n, p) ; il vient, après un calcul laborieux :

M_4 = **npq(1+3(n-2)pq)**, où q =1-p, et donc $p(|f-p| > \frac{1}{\sqrt{n}}) \le \frac{pq(1+3(n-2)pq)}{n}$.

Pour $pq \le \frac{1}{6}$, cette fonction de n décroît ; comme sa limite est $3(pq)^2$, il en résulte que $pq \le \frac{1}{6}$ implique $p(|f-p| > \frac{1}{\sqrt{n}}) \le \frac{1}{12}$.

Conclusion : si p $\leq 0{,}2$ ou p $\geq 0{,}8$ alors $p(|f-p| \leq \frac{1}{\sqrt{n}}) > 0{,}915$.

Il est probable que l'emploi de moments M_n d'ordre supérieur donneraient des résultats de plus en plus intéressants, c.-à.-d. des intervalles de plus en plus étroits de centre $\frac{1}{2}$, mais leurs calculs tant à la main qu'à la machine devient acrobatique. Voilà pourquoi l'étude directe du cas limite p=0,5 m'a paru digne d'intérêt ; pour voir encore plus nettement le fondement de **(1)**, je me suis d'abord débarrassé des parties entières qui hérissaient la question de difficultés a priori secondaires.

Partie A : p = 0,5 et n de la forme $4n^2$.

Par symétrie, il s'agit de prouver que $p_n = p(0 \leq X \leq 2n^2 - 2n - 1) \leq 0{,}05$ **(2)**

quand X est binomiale B(1/2, $4n^2$). Pour n=1, c'est évident, **donc n sera > 1.**

A1- Une formule intégrale pour p_n .

Il existe une formule intégrale pour la fonction de répartition d'une loi B(n, p) :

$$\sum_{k=0}^{t} C_n^k p^k (1-p)^{n-k} = (n-t)C_n^t \int_0^{1-p} (1-x)^t x^{n-t-1}\, dx$$

pour t= 0,1,...,n-1, la démonstration se faisant par récurrence sur t et intégration par parties) ; cette formule est évidemment fausse quand t = n.

Avec n = $4n^2$, p=0,5 , t= $2n^2$-2n-1 , on obtient :

$$p_n = (2n^2+2n+1)C_{4n^2}^{2n^2-2n-1} \int_0^{0,5} (1-x)^{2n^2-2n-1} x^{2n^2+2n}\, dx$$

qu'il est préférable d'écrire

$$p_n = (2n^2-2n)C_{4n^2}^{2n^2-2n} \int_0^{0,5} (1-x)^{2n^2-2n-1} x^{2n^2+2n}\, dx \ . \qquad (3)$$

A2- Une majoration fine de $C_{4n^2}^{2n^2-2n}$.

On utilise un encadrement très étroit de **n!** , **[3, page 54]** :

$$(\frac{n}{e})^n \sqrt{2\pi n}\; e^{\frac{1}{12n+1}} \leq n! \leq (\frac{n}{e})^n \sqrt{2\pi n}\; e^{\frac{1}{12n}} \ . \qquad (4)$$

D'où la majoration d'un C_n^k :

$$C_n^k = \frac{n!}{k!(n-k)!} \leq \frac{e^{\frac{1}{12n}}}{e^{\frac{1}{12n+1}+\frac{1}{12(n-k)+1}}} \cdot \frac{(\frac{n}{e})^n\sqrt{2\pi n}}{(\frac{k}{e})^k\sqrt{2\pi k}(\frac{n-k}{e})^{n-k}\sqrt{2\pi(n-k)}} .$$

Comme le quotient des deux exponentielles est inférieur à 1, il vient

$$C_n^k \leq \frac{1}{\sqrt{2\pi}}\sqrt{\frac{n}{k(n-k)}}(\frac{n}{k})^k(\frac{n}{n-k})^{n-k} \qquad (5)$$

et

$$C_{4n^2}^{2n^2-2n} \leq \frac{4^{2n^2}}{\sqrt{2\pi}}\sqrt{\frac{1}{n^2-1}}(\frac{n}{n-1})^{2n^2-2n}(\frac{n}{n+1})^{2n^2+2n} \qquad (6)$$

(rappelons que n > 1) , d'où

$$(2n^2-2n)C_{4n^2}^{2n^2-2n} \leq \frac{4^{2n^2}}{\sqrt{2\pi}}.2n.(\frac{n}{n-1})^{2n^2-2n}(\frac{n}{n+1})^{2n^2+2n} . \qquad (6\ bis)$$

A3- La suite $u_n = (\frac{n}{n-1})^{2n^2-2n}(\frac{n}{n+1})^{2n^2+2n}$ est majorée par sa limite e^{-2} .

Pour trouver la limite, il suffit d'écrire $\ln(u_n) = 2n^2\ln(\frac{n^2}{n^2-1}) + 2n\ln(\frac{n-1}{n+1})$ et

d'utiliser que ln(u) est équivalent à (u-1) quand u tend vers 1.

Pour établir la majoration, il suffit d'établir que la fonction

$y = (2x^2-2x)\ln(\frac{x}{x-1}) + (2x^2+2x)\ln(\frac{x}{x+1})$ est croissante sur $[2, +\infty[$, ce qui résulte d'une succession de petits miracles (ces miracles qui nous ont fait aimer les mathématiques) :

- y''' valant $\frac{8}{x(x^2-1)^2}$, y'' croît et, premier miracle, la limite de y'' en $+\infty$ est 0 ;
- y' est donc décroissante et, miracle encore, sa limite est nulle en $+\infty$.

La fonction y est donc bien croissante, sur $]1, +\infty[$.

<u>remarque 1</u> : moins élémentaire, mais beaucoup plus rapide, le développement en série entière de y conduit aux deux résultats à la fois. En effet, pour x > 1, on peut écrire :

$$y = -2x^2\ln(1-\frac{1}{x^2}) + 2x\left[\ln(1-\frac{1}{x}) - \ln(1-\frac{1}{x})\right] = -2 + \sum_1^\infty \frac{-2}{(n+1)(2n+1)}(\frac{1}{x})^{2n} .$$

On a donc : $$p_n \leq \frac{e^{-2}}{\sqrt{2\pi}}.2n.4^{2n^2}\int_0^{0,5}(1-x)^{2n^2-1}x^{2n^2+2n}\,dx. \qquad (7)$$

Un dernier effort va nous conduire enfin au point désiré.

A4- Les probabilités p_n sont majorées par leur limite $\frac{1}{\sqrt{2\pi}}\int_2^{+\infty} e^{-t^2/2}\,dt$.

La majoration de $C_{4n^2}^{2n^2-2n}$ ne pouvant guère être améliorée, l'effort (si on avait la foi !) devait maintenant porter sur l'intégrale de **(7)**, notée I_n, que l'on commence par écrire $\int_0^{0,5}[x(1-x)]^{2n^2}.\frac{x^{2n}}{(1-x)^{2n+1}}dx$. Avec les changements de variable t = 2x puis x=1-t,

on obtient facilement : $4^{2n^2} I_n = \int_0^1 (1-x^2)^{2n^2}\frac{(1-x)^{2n}}{(1+x)^{2n+1}}dx$ et donc

$$p_n \le \frac{2e^{-2}}{\sqrt{2\pi}}\, n \int_0^1 (1-x^2)^{2n^2}\frac{(1-x)^{2n}}{(1+x)^{2n+1}}dx \le \frac{2e^{-2}}{\sqrt{2\pi}}\, n \int_0^1 (1-x^2)^{2n^2}\frac{(1-x)^{2n}}{(1+x)^{2n}}dx.$$

Enfin, en posant t = nx : $p_n \le \frac{2e^{-2}}{\sqrt{2\pi}}\int_0^n (1-\frac{t^2}{n^2})^{2n^2}(\frac{1-\frac{t}{n}}{1+\frac{t}{n}})^{2n}dt$. **(8)**

Il est connu que 1-u est majoré par e^{-u} sur R^+, un peu moins que $\frac{1-u}{1+u}$ y est majoré par e^{-2u} ; d'où $p_n \le \frac{2e^{-2}}{\sqrt{2\pi}}\int_0^{+\infty} e^{-2t^2-4t}dt = \frac{1}{\sqrt{2\pi}}\int_0^{+\infty} e^{-2(t+1)^2}dt$. En posant u = 2t+2, il vient enfin $p_n \le \frac{1}{\sqrt{2\pi}}\int_2^{+\infty} e^{-t^2/2}\,dt$ < 0,0228 (par le calcul numérique).

Ainsi **(2)** est démontrée, et au-delà, par l'Analyse, dans le cadre fixé.

A5- Une démonstration nouvelle pour la limite de p_n .

Il n'a pas échappé au lecteur que les majorations que nous avons obtenues en **(6)**, **(6bis)** et **(7)** sont aussi des **équivalents** : ainsi donc, quand n tend vers l'infini, $\mathbf{p_n}$ a même limite que $\frac{2e^{-2}}{\sqrt{2\pi}}\int_0^n (1-\frac{t^2}{n^2})^{2n^2}(\frac{1-\frac{t}{n}}{1+\frac{t}{n}})^{2n}dt$ .Comme la suite des fonctions positives $\mathbf{f_n}$(t) = $(1-\frac{t^2}{n^2})^{2n^2}(\frac{1-\frac{t}{n}}{1+\frac{t}{n}})^{2n}$, prolongées par 0 pour x > n, converge vers f(t) = e^{-2t^2-4t}

qui est intégrable sur $\mathbf{R}^+$ et majore les $\mathbf{f_n}$, on peut appliquer le **Théorème de convergence dominée de Lebesgue** :

$$\int_0^n \mathbf{f_n} = \int_0^{+\infty} \mathbf{f_n} \text{ converge vers } \int_0^{+\infty} f .$$

A6- Vérifications numériques.

Pour vérifier l'échafaudage ci-dessus, calculons numériquement quelques valeurs de $\mathbf{p_n} = \frac{1}{2^{4n^2}} \sum_{k=0}^{2n^2-2n-1} C_{4n^2}^k$; voici ce que Maple, par exemple, affiche en pourcentage pour n allant de 2 à 40 (mais une calculatrice programmable suffit amplement) :

1.0635, 1.4408, 1.6383, 1.7600, 1.8426, 1.9024, 1.9476, 1.9830,
2.0115, 2.0350, 2.0546, 2.0712, 2.0855, 2.0980, 2.1089,
2.1185, 2.1271, 2.1348, 2.1417, 2.1480, 2.1537, 2.1589,
2.1637, 2.1681, 2.1722, 2.1759, 2.1794, 2.1827, 2.1858,
2.1886, 2.1913, 2.1938, 2.1962, 2.1984, 2.2005, 2.2025,
2.2044, 2.2062, 2.2079.

D'où la **question** : la suite $\mathbf{p_n}$ est-elle effectivement croissante ?

PARTIE B : p = 0,5 et n quelconque

Il s'agit toujours de majorer $(\mathbf{n}-\mathbf{t})C_\mathbf{n}^\mathbf{t} \int_0^{0,5} (1-\mathbf{x})^\mathbf{t} \mathbf{x}^{\mathbf{n}-\mathbf{t}-1}\,\mathbf{dx}$ avec l'entier t **strictement** inférieur à $\frac{n}{2} - \sqrt{n}$ **.** Dans toute la suite nous poserons $\frac{n}{2} - \sqrt{n} = \lambda$.

Pour n = 1,2,3,4 cette probabilité vaut 0 ; **on supposera donc n > 4.**

B1- Une majoration de $(\mathbf{n}-\mathbf{t})C_\mathbf{n}^\mathbf{t}$ en fonction du réel λ .

L'idée est simple mais sa mise en œuvre est, curieusement, délicate : comme u(t)=$(\mathbf{n}-\mathbf{t})C_\mathbf{n}^\mathbf{t}$ vaut par exemple $(\mathbf{t}+1)C_\mathbf{n}^{\mathbf{t}+1}$, u est une fonction croissante de t (tant que t reste < n/2) ; il paraît donc légitime de majorer u(t) par $(\mathbf{n}-\lambda)C_\mathbf{n}^\lambda$, $C_\mathbf{n}^\lambda$ étant évidemment à définir. Et c'est là qu'il nous faut faire intervenir la fonction Γ qui, on le sait, prolonge n ! aux réels, mais avec un petit décalage puisque Γ (n) = (n-1) ! Aussi poserons-nous F(t) = Γ (t+1), de façon que F(n)=n !

Pour prouver que $(n-t)C_n^t = \frac{(n-t)n!}{F(t)F(n-t)} \leq \frac{(n-\lambda)n!}{F(\lambda)F(n-\lambda)}$, on utilise d'abord F(t) = t F(t-1) : tout revient alors à prouver que F(t)F(n-t-1) décroît, et après dérivation (F est $\mathbf{C}^\infty$), que F'(t)F(n-t-1) ≤ F(t)F'(n-t-1) , ou encore $\frac{F'(t)}{F(t)} \leq \frac{F'(n-t-1)}{F(n-t-1)}$; en définitive, comme t ≤ n-t-1, il s'agit de montrer que F'/F est croissante (on dit que F est logarithmiquement convexe).

Or ceci résulte d'une formule sur la fonction Γ **[2, page 476] :**

$$\frac{F'(t)}{F(t)} = \frac{\Gamma'(t+1)}{\Gamma(t+1)} = -\gamma - \frac{1}{t+1} + \sum_{n=1}^{\infty} \frac{t+1}{n(n+t+1)}$$, où γ est la constante d'Euler.

Ensuite le lecteur doit admettre que **(4)** s'étend aux réels :

$$\left(\frac{t}{e}\right)^t \sqrt{2\pi t}\; e^{\frac{1}{12t+1}} \leq F(t) \leq \left(\frac{t}{e}\right)^t \sqrt{2\pi t}\; e^{\frac{1}{12t}}$$

et donc que **(5)** s'applique. En conclusion (rappelons que n > 4) :

$$(n-t)C_n^t \leq \frac{\sqrt{n}}{\sqrt{2\pi}} \sqrt{\frac{\sqrt{n}+2}{\sqrt{n}-2}}\, 2^n \left(\frac{\sqrt{n}}{\sqrt{n}-2}\right)^{\frac{n}{2}-\sqrt{n}} \left(\frac{\sqrt{n}}{\sqrt{n}+2}\right)^{\frac{n}{2}+\sqrt{n}} .$$

Enfin, comme en **A3** , $\left(\frac{\sqrt{n}}{\sqrt{n}-2}\right)^{\frac{n}{2}-\sqrt{n}} \left(\frac{\sqrt{n}}{\sqrt{n}+2}\right)^{\frac{n}{2}+\sqrt{n}}$ est majoré par sa limite $\mathbf{e}^{-2}$:

il suffit en effet de poser x = $\frac{\sqrt{n}}{2}$ pour retrouver la fonction y étudiée dans ce paragraphe.

D'où $$p_n \leq \frac{\sqrt{n}}{\sqrt{2\pi}} \sqrt{\frac{\sqrt{n}+2}{\sqrt{n}-2}}\, 2^n e^{-2} \int_0^{0,5} (1-x)^t x^{n-t+1} dt .$$

B2- Une majoration intégrale de p_n.

Sur l'intervalle [0, $\frac{1}{2}$], $(1-x)^t x^{n-t-1} \leq (1-x)^{t'} x^{n-t'-1}$ dés que $t \leq t'$; il en résulte que

$$\int_0^{0,5} (1-x)^t x^{n-t+1} dt \leq \int_0^{0,5} (1-x)^{\frac{n}{2}-\sqrt{n}} x^{\frac{n}{2}+\sqrt{n}-1} dt .$$

En incorporant le $\mathbf{2^n}$ dans l'intégrale :

$\mathbf{p_n} \leq \frac{\sqrt{n}}{\sqrt{2\pi}}\sqrt{\frac{\sqrt{n}+2}{\sqrt{n}-2}}\, e^{-2}\int_0^{0,5}[2x(2-x)]^{\frac{n}{2}}\frac{x^{\sqrt{n}-1}}{(1-x)^{\sqrt{n}}}\,dt$; comme en **A4,** avec t=2x et x=1-t ,

$\mathbf{p_n} \leq \frac{\sqrt{n}}{\sqrt{2\pi}}\sqrt{\frac{\sqrt{n}+2}{\sqrt{n}-2}}\, e^{-2}\int_0^{1}(1-x^2)^{\frac{n}{2}}\frac{(1-x)^{\sqrt{n}-1}}{(1+x)^{\sqrt{n}}}\,dx$, puis avec t= $x\sqrt{n}$

$\mathbf{p_n} \leq \frac{1}{\sqrt{2\pi}}\sqrt{\frac{\sqrt{n}+2}{\sqrt{n}-2}}\, e^{-2}\int_0^{\sqrt{n}}(1-\frac{t^2}{n})^{\frac{n}{2}}\frac{(1-\frac{t}{\sqrt{n}})^{\sqrt{n}-1}}{(1+\frac{t}{\sqrt{n}})^{\sqrt{n}}}\,dt$ et avec toujours les outils de **A4**

$$\mathbf{p_n} \leq \frac{1}{\sqrt{2\pi}}\sqrt{\frac{\sqrt{n}+2}{\sqrt{n}-2}}\, e^{-2}\int_0^{+\infty} e^{-\frac{t^2}{2}}e^{-2t\frac{\sqrt{n}-1}{\sqrt{n}}}\,dt = \frac{1}{\sqrt{2\pi}}\sqrt{\frac{\sqrt{n}+2}{\sqrt{n}-2}}\int_2^{+\infty} e^{-\frac{t^2}{2}+2\frac{t-2}{\sqrt{n}}}\,dt \qquad \textbf{(9)}$$

B3- La plus grande probabilité p_n est $p_8 < 3{,}52\%$.

Comme les suites $\sqrt{\frac{\sqrt{n}+2}{\sqrt{n}-2}}$ et $\int_2^{+\infty} e^{-\frac{t^2}{2}+2\frac{t-2}{\sqrt{n}}}\,dt$ sont décroissantes, pour n ≥ 100

$\mathbf{p_n} \leq \frac{1}{\sqrt{2\pi}}\sqrt{\frac{12}{8}}\int_2^{+\infty} e^{-\frac{t^2}{2}+\frac{t-2}{5}}\,dt$ < 3,1%. Un calcul programmé assure ensuite que

le maximum des $\mathbf{p_n}$ pour n allant de 5 à 99 est $\mathbf{p_8} = \frac{9}{2^8}$ et donc , avec les notations de **(1) :**

$$\boxed{p\left(\left|f-\frac{1}{2}\right| \leq \frac{1}{\sqrt{n}}\right) > 92{,}9\% > \frac{13}{14}} \qquad \textbf{(10)}$$

<u>remarque 2</u> : (9) s'écrit aussi $\mathbf{p_n} \leq \sqrt{\frac{\sqrt{n}+2}{\sqrt{n}-2}}\frac{1}{\sqrt{2\pi}}\int_{2-\frac{2}{\sqrt{n}}}^{+\infty} e^{-\frac{t^2}{2}}\,dt$; en appelant p l'intégrale limite de $\mathbf{p_n}$, on montre sans peine qu'il existe une constante K telle que $\mathbf{p_n} \leq \mathbf{p} + \frac{K}{\sqrt{n}}$.

---FIN---

Bibliographie :

1. Louis-Marie Bonneval, ***Intervalles : de confiance ?,*** Bulletin n° 427 de l'APMEP, mars-avril 2000.

2. Jean Dieudonné, ***Calcul Infinitésimal,*** 2$^{\mathbf{e}}$ édition, Hermann, 1980.

3. William Feller, ***An introduction to probability theory and its applications,*** 3$^{\mathbf{e}}$ édition, volume 1, Wiley International Edition.

---

e-mail : dsaada@yahoo.fr