

# Loi binomiale, courbe en cloche et tableur

Bernard Parzysz<sup>(\*)</sup>

L'étude de la loi binomiale figure au programme des classes terminales des séries S et ES, ainsi que dans le programme de spécialité de la série L. Il est bien sûr intéressant de chercher à la représenter graphiquement, et alors les tableurs-grapheurs fournissent à la demande un diagramme en bâtons, ce qui permet de constater que, pour  $n$  « assez grand » (comme on dit), les sommets desdits bâtons dessinent une « courbe en cloche » assez régulière. Cette notion de « courbe en cloche » est habituellement associée à la loi dite « normale », deux notions dont les élèves ont peut-être entendu parler par ailleurs, dans d'autres disciplines ou même hors du lycée. Cependant :

- d'une part, il ne suffit pas que la représentation graphique d'une fonction dérivable sur  $\mathbb{R}$  ait une allure de « cloche » pour que cette fonction soit la densité d'une variable distribuée normalement
- et, d'autre part, la loi binomiale est définie sur une partie finie de  $\mathbb{N}$  (d'où le diagramme en bâtons), tandis que la loi normale est définie sur  $\mathbb{R}$  (d'où la courbe).

Ce qui conduit à se poser la question suivante : *au-delà de l'opposition discret / continu, et comme le suggère la représentation graphique, la distribution binomiale a-t-elle quelque chose à voir avec la loi normale ?*

## 1. Une approche théorique

Commençons par rappeler que, étant donné un entier non nul  $n$  et un réel  $p$  compris entre 0 et 1, on dit qu'une variable aléatoire  $X_n$  suit la loi binomiale de paramètres  $n$  et  $p$ , notée  $\mathcal{B}(n ; p)$ , lorsqu'elle prend les valeurs entières  $k$  comprises

entre 0 et  $n$  avec les probabilités  $p_k = \binom{n}{k} p^k (1-p)^{n-k}$ , et que cette variable a pour

espérance  $m = np$  et pour variance  $\sigma^2 = np(1-p)$ . Rappelons aussi qu'on peut considérer  $X_n$  comme la somme de  $n$  variables de Bernoulli de même paramètre  $p$  et mutuellement indépendantes.

La réponse théorique à la question posée se trouve dans le fameux « théorème limite central », qui affirme en particulier que, étant donné une variable  $Y_n$  de loi

binomiale  $\mathcal{B}(n ; p)$ , la variable centrée réduite associée, soit  $Z_n = \frac{Y_n - np}{\sqrt{np(1-p)}}$ ,

« converge en loi »<sup>(1)</sup> vers la loi normale centrée réduite lorsque  $n$  tend vers l'infini.

(\*) parzysz.bernard@wanadoo.fr

(1) Voir plus loin note 4.

Ce résultat apparaît dans l'édition de 1756 de la *Doctrine of Chance* d'Abraham de Moivre<sup>(2)</sup>, qui le démontre dans le cas particulier  $p = 0,5$ . Pierre-Simon Laplace en fait plus tard (1812) une démonstration pour  $p$  quelconque, démonstration complétée par Markov en 1898. Quant au théorème limite central dans toute sa généralité, il est dû à Liapounov (1900), dont Paul Lévy améliora la démonstration en 1919<sup>(3)</sup>.

Voici un énoncé du théorème limite central :

Soit  $(X_n)_{n \in \mathbb{N}^*}$  une suite de variables aléatoires mutuellement indépendantes et de même loi d'espérance  $m$  et de variance  $\sigma^2$ . Notons  $\overline{X}_n$  la moyenne des  $n$  premières variables (qui a pour espérance  $m$  et pour variance  $\sigma^2/n$ ). Alors, lorsque  $n$  tend vers l'infini, la variable centrée réduite  $Z_n$  associée à  $\overline{X}_n$  (c'est-à-dire  $Z_n = (\overline{X}_n - m) \frac{\sqrt{n}}{\sigma}$ ) converge en loi vers une variable  $N$  suivant la loi normale centrée réduite<sup>(4)</sup>.

On peut remarquer qu'il s'agit d'un théorème puissant, car il affirme que, même si on ne connaît pas grand-chose sur la loi des variables  $X_n$  (seulement qu'elle a une espérance et une variance), la moyenne des  $n$  premières suit à peu près une loi normale lorsque  $n$  est assez grand.

Dans le cas particulier qui nous occupe :

- a) les  $X_n$  sont des variables de Bernoulli mutuellement indépendantes et de même paramètre  $p$ , et donc d'espérance  $p$  et de variance  $p(1-p)$ ,
- b) la somme  $Y_n$  des  $n$  premières de ces variables suit la loi binomiale  $\mathcal{B}(n; p)$ , d'espérance  $np$  et de variance  $np(1-p)$ ,
- c) leur moyenne  $\overline{X}_n (= Y_n/n)$  a pour espérance  $p$  et pour variance  $p(1-p)/n$ ,
- d) la variable centrée réduite associée à  $\overline{X}_n$  (donc aussi à  $Y_n$ ) est

$$Z_n = \frac{\overline{X}_n - p}{\sqrt{p(1-p)}} \sqrt{n}, \text{ soit encore } Z_n = \frac{Y_n - np}{\sqrt{np(1-p)}}.$$

Le théorème limite central affirme donc bien le résultat annoncé au début de ce paragraphe.

(2) Il s'y intéresse au nombre de piles obtenu à l'issue de 1800 lancers d'une pièce.

(3) Pour une vue globale de l'évolution des probabilités, on pourra consulter la Frise historique de Jean-François Pichard, in *Autour de la modélisation en probabilités*. Presses universitaires de Franche-Comté (Besançon 2001), p. 47-56.

(4) Que veut dire « converger en loi » ? Que si  $F_n$  est la fonction de répartition de  $Z_n$  et  $F$  celle de  $N$  (supposée continue) alors, pour tout  $x$ , la suite  $(F_n(x))$  converge – au sens usuel – vers  $F(x)$ .

## 2. Une approche visuelle

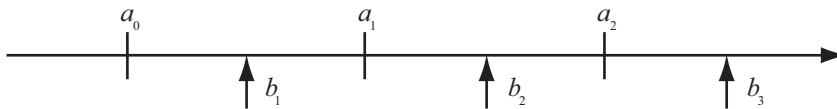
Le tableur-grapheur permet cependant, de façon plus concrète, d'une part de visualiser ce mode de convergence, et d'autre part d'estimer – toujours visuellement – sa « qualité ».

Prenons à titre d'exemple le cas  $n = 100$  et  $p = 0,2$  (d'où  $m = 20$  et  $\sigma = 4$ ).

Une variable  $X_{100}$  de loi  $\mathcal{B}(100 ; 0,2)$  prend toute valeur entière  $k$  comprise entre 0 et 100 avec la probabilité  $p_k$  indiquée plus haut. La variable centrée réduite  $Z_{100}$  associée à  $X_{100}$  prend alors les valeurs  $a_k = \frac{k-20}{4}$  avec les mêmes probabilités  $p_k$ .

Pour comparer le diagramme en bâtons de la distribution de probabilité de  $Z_{100}$  avec la densité d'une variable aléatoire  $N$  de loi normale centrée réduite, nous allons « discrétiser » la densité de celle-ci de la façon suivante.

Les 101 valeurs  $a_k$  sont régulièrement espacées entre  $a_0 = -5$  et  $a_{100} = 20$ , avec un pas de 0,25. Notons  $b_k$  le milieu de l'intervalle  $[a_{k-1} ; a_k]$  et cherchons, pour chacune des variables  $Y_{100}$  et  $N$ , la probabilité qu'elle a de prendre les valeurs de l'intervalle  $]b_k ; b_{k+1}]$ .



*Remarque.* Pour recouvrir  $\mathbb{R}$  tout entier, posons en outre  $b_0 = -\infty$  et  $b_{101} = +\infty$ .

Soit maintenant  $k$  un entier compris entre 0 et 100 :

- pour  $Y_{100}$  on a  $P(b_k < Z_{100} \leq b_{k+1}) = P(X_n = k) = p_k$ ,
- pour  $C$  on a  $P(b_k < N \leq b_{k+1}) = F(b_{k+1}) - F(b_k)$ , où  $F$  est la fonction de répartition de la loi normale centrée réduite.

Il s'agit donc de comparer ces deux probabilités pour toutes les valeurs de  $k$ .

*Remarque.* La fonction  $k \mapsto p_k$  n'est autre que la distribution de probabilité de la variable binomiale  $X_{100}$ , tandis que la fonction  $k \mapsto F(b_{k+1}) - F(b_k)$  est la distribution de probabilité de la variable  $4D + 20$ , où  $D$  est la variable discrète dont la loi « concentre » en chaque point  $k$  la loi normale centrée réduite.

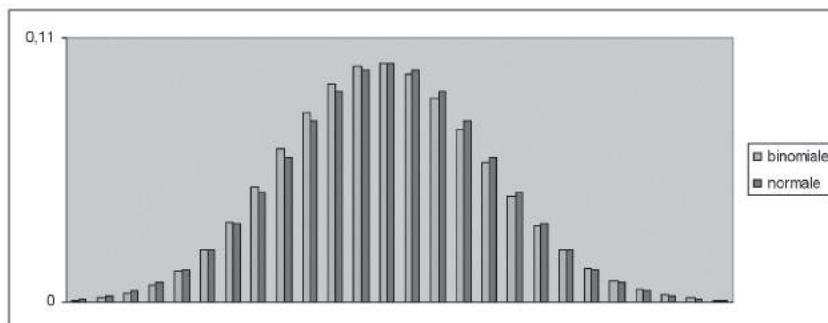
Le tableur nous fournit rapidement les valeurs souhaitées<sup>(5)</sup> (voir le tableau ci-dessous, qui a été tronqué en ne conservant que les valeurs de  $k$  pour lesquelles la probabilité n'est pas « négligeable »).

$k$	$a_k$	$b_k$	binomiale	normale	écart abs.	% écart
8	-3	-3,125	0,0006	0,0011	0,0006	48,9
9	-2,75	-2,875	0,0015	0,0023	0,0008	36,1
10	-2,5	-2,625	0,0034	0,0044	0,0011	24,3

(5) Du fait que la distribution de probabilité de la loi binomiale, ainsi que la fonction de répartition de la loi normale, font partie des fonctions statistiques du logiciel.

11	-2,25	-2,375	0,0069	0,008	0,0011	14,2
12	-2	-2,125	0,0128	0,0136	0,0008	6,24
13	-1,75	-1,875	0,0216	0,0217	0,0001	0,47
14	-1,5	-1,625	0,0335	0,0325	0,001	3,22
15	-1,25	-1,375	0,0481	0,0457	0,0023	5,1
16	-1	-1,125	0,0638	0,0605	0,0033	5,52
17	-0,75	-0,875	0,0789	0,0752	0,0037	4,86
18	-0,5	-0,625	0,0909	0,0878	0,0031	3,48
19	-0,25	-0,375	0,0981	0,0964	0,0016	1,7
20	0	-0,125	0,0993	0,0995	0,0002	0,18
21	0,25	0,125	0,0946	0,0964	0,0019	1,93
22	0,5	0,375	0,0849	0,0878	0,0029	3,35
23	0,75	0,625	0,072	0,0752	0,0032	4,28
24	1	0,875	0,0577	0,0605	0,0028	4,56
25	1,25	1,125	0,0439	0,0457	0,0019	4,05
26	1,5	1,375	0,0316	0,0325	0,0008	2,59
27	1,75	1,625	0,0217	0,0217	4E-06	0,02
28	2	1,875	0,0141	0,0136	0,0005	3,88
29	2,25	2,125	0,0088	0,008	0,0008	9,38
30	2,5	2,375	0,0052	0,0044	0,0007	16,8
31	2,75	2,625	0,0029	0,0023	0,0006	26,7
32	3	2,875	0,0016	0,0011	0,0004	39,6
33	3,25	3,125	0,0008	0,0005	0,0003	56,5

On peut constater sur ce tableau que les écarts absolus sont faibles. Quant au pourcentage de l'écart relatif entre les deux probabilités (colonne de droite du tableau), il montre qu'il est assez faible pour les valeurs de  $k$  situées autour de 20, c'est-à-dire de l'espérance de  $X_{100}$ , et correspond donc aux valeurs les plus probables de la loi binomiale. Par contre, pour les valeurs plus éloignées, il devient vite important. Mais, comme ces valeurs ont une faible probabilité, l'impact visuel n'est pas très perceptible, comme le montre la juxtaposition des deux diagrammes en bâtons (voir graphique 1 ci-dessous).

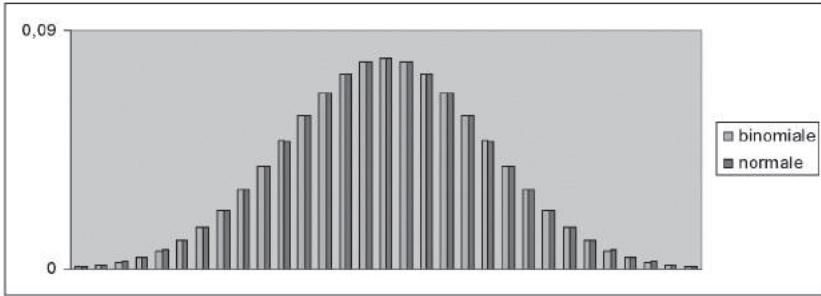


Graphique 1

Au-delà de cette ressemblance visuelle, on peut également chercher à se faire une idée de l'influence des deux paramètres  $n$  et  $p$  sur la « proximité visuelle » des deux diagrammes.

1° Cas  $n = 100$  et  $p = 0,5$ . On a cette fois  $m = 50$  et  $\sigma = 5$ . Le tableau tronqué ci-dessous montre que les écarts absolus comme les écarts relatifs sont nettement plus faibles que lorsque  $p = 0,2$  : on a une plage de 31 valeurs autour de 50 dans laquelle l'écart relatif est inférieur à 4 % (contre 5 valeurs lorsque  $p = 0,2$ ) ; en outre, les écarts absolus sont ici de l'ordre de  $10^{-5}$ , alors qu'ils étaient de l'ordre de  $10^{-3}$  dans le cas  $p = 0,2$ . Le graphique 2 confirme cette plus grande « proximité » des deux diagrammes.

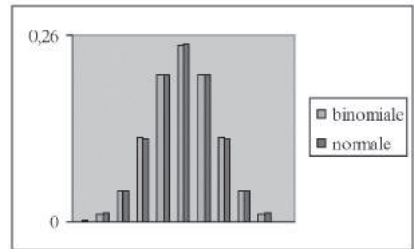
$k$	$a_k$	$b_k$	binomiale	normale	écart abs.	% écart
35	-3	-3,1	0,0009	0,0009	3,44E-05	3,82
36	-2,8	-2,9	0,0016	0,0016	4,14E-05	2,59
37	-2,6	-2,7	0,0027	0,0027	4,47E-05	1,63
38	-2,4	-2,5	0,0045	0,0045	4,15E-05	0,92
39	-2,2	-2,3	0,0071	0,0071	2,95E-05	0,41
40	-2	-2,1	0,0108	0,0109	8,27E-06	0,08
41	-1,8	-1,9	0,0159	0,0158	2,01E-05	0,13
42	-1,6	-1,7	0,0223	0,0222	5,05E-05	0,23
43	-1,4	-1,5	0,0301	0,03	7,53E-05	0,25
44	-1,2	-1,3	0,039	0,0389	8,7E-05	0,22
45	-1	-1,1	0,0485	0,0484	8,03E-05	0,17
46	-0,8	-0,9	0,058	0,0579	5,49E-05	0,09
47	-0,6	-0,7	0,0666	0,0666	1,65E-05	0,02
48	-0,4	-0,5	0,0735	0,0736	2,41E-05	0,03
49	-0,2	-0,3	0,078	0,0781	5,48E-05	0,07
50	0	-0,1	0,0796	0,0797	6,66E-05	0,08
51	0,2	0,1	0,078	0,0781	5,48E-05	0,07
52	0,4	0,3	0,0735	0,0736	2,41E-05	0,03
53	0,6	0,5	0,0666	0,0666	1,65E-05	0,02
54	0,8	0,7	0,058	0,0579	5,49E-05	0,09
55	1	0,9	0,0485	0,0484	8,03E-05	0,17
56	1,2	1,1	0,039	0,0389	8,7E-05	0,22
57	1,4	1,3	0,0301	0,03	7,53E-05	0,25
58	1,6	1,5	0,0223	0,0222	5,05E-05	0,23
59	1,8	1,7	0,0159	0,0158	2,01E-05	0,13
60	2	1,9	0,0108	0,0109	8,27E-06	0,08
61	2,2	2,1	0,0071	0,0071	2,95E-05	0,41
62	2,4	2,3	0,0045	0,0045	4,15E-05	0,92
63	2,6	2,5	0,0027	0,0027	4,47E-05	1,63
64	2,8	2,7	0,0016	0,0016	4,14E-05	2,59
65	3	2,9	0,0009	0,0009	3,44E-05	3,82



Graphique 2

2° Cas  $n = 10$  et  $p = 0,5$ . Le tableau ci-dessous et le graphique 3 ci-dessous font apparaître une grande proximité des deux distributions, malgré la faible valeur de  $n$  : hormis les 4 valeurs extrêmes (de probabilité inférieure à 1 %), l'écart relatif ne dépasse pas 1 %.

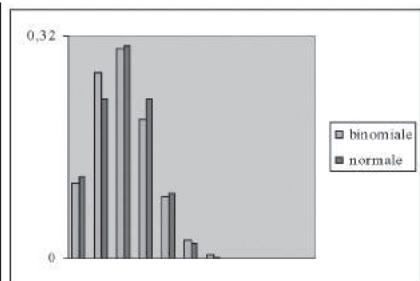
$k$	$a_k$	$b_k$	binomiale	normale	% écart
0	-3,162	-2,846	0,001	0,0022	56
1	-2,53	-2,214	0,0098	0,0112	13
2	-1,897	-1,581	0,0439	0,0435	1
3	-1,265	-0,949	0,1172	0,1145	2,4
4	-0,632	-0,316	0,2051	0,2045	0,3
5	0	0,316	0,2461	0,2482	0,8
6	0,632	0,949	0,2051	0,2045	0,3
7	1,265	1,581	0,1172	0,1145	2,4
8	1,897	2,214	0,0439	0,0435	1
9	2,53	2,846	0,0098	0,0112	13
10	3,162	3,479	0,001	0,0003	287



Graphique 3

2° Cas  $n = 10$  et  $p = 0,2$ . Le tableau ci-dessous et le graphique 4 montrent cette fois des différences assez sensibles puisque, à l'exception de la valeur  $k = np (= 2)$ , l'écart relatif n'est jamais inférieur à 6 %.

$k$	$a_k$	$b_k$	binomiale	normale	% écart
0	-1,581	-1,186	0,1074	0,1178	8,9
1	-0,791	-0,395	0,2684	0,2285	17
2	0	0,395	0,302	0,3074	1,7
3	0,791	1,186	0,2013	0,2285	12
4	1,581	1,976	0,0881	0,0938	6,1
5	2,372	2,767	0,0264	0,0212	24
6	3,162	3,558	0,0055	0,0026	108
7	3,953	4,348	0,0008	0,0002	336
8	4,743	5,139	7E-05	7E-06	995
9	5,534	5,929	4E-06	1E-07	###
10	6,325	6,72	1E-07	9E-12	###



Graphique 4

En conclusion, ces quelques exemples montrent (sans démontrer) que, s'il y a bien convergence de la loi binomiale vers la loi normale, cette convergence apparaît d'autant meilleure que la probabilité  $p$  est voisine de 0,5. Par comparaison avec les autres, les graphiques 1 et 4 et les tableaux qui leur sont associés nous fournissent une clé d'explication : ils montrent en effet que, pour  $p = 0,2$ , la distribution binomiale n'est pas symétrique par rapport à la valeur la plus probable,  $np$  (pour toutes les valeurs numériques choisies,  $np$  est entier), au contraire de la distribution normale. Au contraire, pour  $p = 0,5$ , la distribution binomiale, comme la distribution normale, est symétrique par rapport à la valeur  $np$  : on a en effet  $p_k = p_{n-k}$  car

$$\binom{n}{k} = \binom{n}{n-k} \text{ et } p = 1 - p. \text{ Le logiciel, de par ses deux fonctionnalités complémentaires (tableur et grapheur), permet, d'une part de calculer rapidement les distributions de la loi binomiale et de la loi normale « discrétisée », et d'autre part de faire voir ces distributions grâce aux diagrammes en bâtons, ce qui permet de les comparer de façon tant qualitative (grapheur) que quantitative (tableur). Ainsi, même si l'on ne dispose pas des outils théoriques permettant de démontrer la convergence de la loi binomiale } \mathcal{B}(n ; p) \text{ vers la loi normale, on peut s'en persuader grâce à l'outil informatique, et même se faire une idée de l'influence des paramètres } n \text{ et } p \text{ dans la « rapidité »}^{(6)} \text{ de cette convergence}^{(7)}. \text{ En classe, en se restreignant aux deux colonnes de la partie « binomiale », il s'agit d'un type d'activité dans lequel on peut laisser les élèves partir en exploration et éventuellement, si l'expression « courbe en cloche » apparaît, confirmer qu'on « s'en approche » effectivement lorsque } n \text{ « devient grand ». Ce qui peut même permettre, en série L, de revenir sur la notion de plage de normalité rencontrée en première.}$$

taires (tableur et grapheur), permet, d'une part de calculer rapidement les distributions de la loi binomiale et de la loi normale « discrétisée », et d'autre part de faire voir ces distributions grâce aux diagrammes en bâtons, ce qui permet de les comparer de façon tant qualitative (grapheur) que quantitative (tableur). Ainsi, même si l'on ne dispose pas des outils théoriques permettant de démontrer la convergence de la loi binomiale  $\mathcal{B}(n ; p)$  vers la loi normale, on peut s'en persuader grâce à l'outil informatique, et même se faire une idée de l'influence des paramètres  $n$  et  $p$  dans la « rapidité »<sup>(6)</sup> de cette convergence<sup>(7)</sup>. En classe, en se restreignant aux deux colonnes de la partie « binomiale », il s'agit d'un type d'activité dans lequel on peut laisser les élèves partir en exploration et éventuellement, si l'expression « courbe en cloche » apparaît, confirmer qu'on « s'en approche » effectivement lorsque  $n$  « devient grand ». Ce qui peut même permettre, en série L, de revenir sur la notion de plage de normalité rencontrée en première.

(6) Ce mot est ici à prendre dans un sens général, qualitatif, et non dans son sens « technique ».

(7) Le lecteur intéressé pourra se fabriquer ses propres exemples numériques sur la base des tableaux figurant dans le texte. Par exemple, les fonctions d'Excel comportent la probabilité de la loi binomiale (LOI.BINOMIALE) et la fonction de répartition de la loi normale centrée réduite (LOI.NORMAL.STANDARD).