Journées APMEP 2006
Atelier Asm 10, Michel HENRY

Naissance des probabilités du 13^{e} au 18^{e} siècle
De Huygens à Bernoulli[1]

I - Introduction

En conclusion de son opuscule de 1657, *De ratiociniis in ludo aleae*, Huygens propose 5 problèmes à la sagacité de ses lecteurs. Jacques Bernoulli relèvera le défi, fera une analyse complète du texte de Huygens, résout les problèmes et propose d'appliquer le tout nouveau calcul des probabilités aux affaires civiles, morales et économiques. Il passe les 20 dernières années de sa vie à écrire son traité magistral *Ars Conjectandi* (1713), qui se termine par la démonstration du théorème qui porte son nom, faisant le lien entre probabilité a priori et observation fréquentiste.

Dans cet atelier, j'ai présenté la progression des idées dans la 4ème partie du traité et étudié la démonstration que fait Bernoulli de la loi des grands nombres, donnée en chapitre 5 dans le texte authentique traduit par Norbert Meunier [1].

Cette première démonstration du théorème de Bernoulli par lui-même est intéressante à plus d'un titre [2]. Le travail fin sur les coefficients binomiaux met bien en évidence la propriété de stabilisation des fréquences, source de la loi des grands nombres. Le traitement que fait Bernoulli de la notion de limite en probabilité est aussi très instructif, passant d'un raisonnement erroné inspiré par le calcul infinitésimal à une démonstration rigoureuse en termes de limites, bien avant que cette notion soit clairement dégagée par CAUCHY.
Le caractère constructif de la démonstration de Bernoulli permet de déterminer le nombre optimal d'observations pour estimer une probabilité. C'est sans doute le premier résultat sur l'estimation par intervalle de confiance.
Le résultat peu performant obtenu en fin de compte par Bernoulli est instructif sur les limites de la méthode et appelle les avancées de MOIVRE et de LAPLACE sur l'approximation de la loi binomiale par la loi normale.

Christiaan Huygens (1629-1695)

S'est intéressé aux jeux de hasard et a suscité une reprise des échanges entre Pascal et Fermat en 1656-1657

Publia en 1657: *De Ratiociniis in ludo aleae*

(Du calcul dans les jeux de hasard)

---

[1] Pour les journées de Clermont-Ferrand, j'avais proposé deux ateliers pour illustrer sous la forme de diaporamas quelques étapes essentielles de la naissance du calcul des probabilités. Ce compte rendu reprend les diapositives du deuxième atelier qui constituent la trame de l'article *La démonstration par Jacques Bernoulli de son théorème* [3].

II - Christiaan Huygens (1657) *De Ratiociniis in ludo aleae*

Huygens commence son opuscule par cette affirmation :

« *Quoique dans les jeux de hasard pur les résultats soient incertains, la chance qu'un joueur a de gagner ou de perdre a cependant une valeur déterminée* »…

Puis il énonce le principe suivant :

« *Dans un jeu de hasard, la chance qu'on a de gagner quelque chose a une valeur telle que si on possède cette valeur on peut se procurer la même chance par un jeu équitable…* ».

Huygens définit cette attente dans le cas de l'équiprobabilité, Proposition I :

« *Si j'attends a ou b, dont n'importe lequel pourrait également facilement m'échoir, mon attente doit être* $\frac{a+b}{2}$ »

Il donne ensuite, en proposition III, la définition de l'espérance mathématique :

« *Avoir p chances d'obtenir a et q chances d'obtenir b, les chances étant équivalentes, me vaut* $\frac{pa+qb}{p+q}$ ».

Reprenant le problème des partis, Huygens étudie ensuite diverses situations de jeux de dés, et pour ponctuer son manuel, il propose 5 exercices à la sagacité de ses lecteurs, « *afin que cela leur servît d'exercice et de passe temps* ». Exemple, le 5$^{\text{ème}}$ problème de Huygens :

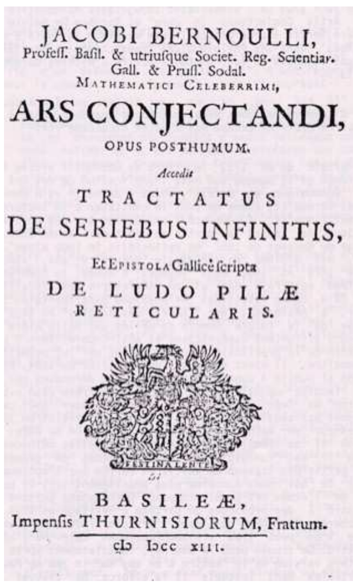
« Ayant pris chacun 12 jetons, A et B jouent avec trois dés à cette condition qu'à chaque coup de 11 points, A doit donner un jeton à B et que B en doit donner un à A à chaque coup de 14 points, et que celui là gagnera qui sera le premier en possession de tous les jetons.

On trouve dans ce cas que la chance de A est à celle de B comme 244 140 625 est à 282 429 536 481 »

Le rapport donné par Huygens est r = $\left(\frac{15}{27}\right)^{12}$ = 8,644 10^{-4}

L'élève Bernoulli résout les exercices et fait des commentaire qui formeront la première partie de son chef d'œuvre: *Ars Conjectandi.*

III – Jacques (Jacobi) Bernoulli, 1654 – 1705 : *Ars Conjectandi* (1713)


JACOBI BERNOULLI,
Profess. Basil. & utriusque Societ. Reg. Scientiar. Gall. & Pruss. Sodal.
MATHEMATICI CELEBERRIMI,
ARS CONJECTANDI,
OPUS POSTHUMUM.
Accedit
TRACTATUS
DE SERIEBUS INFINITIS,
Et EPISTOLA Gallicè scripta
DE LUDO PILÆ
RETICULARIS.

BASILEÆ,
Impensis THURNISIORUM, Fratrum.
 cIɔ Iɔcc XIII.

Dans son introduction de la quatrième partie d'*Ars Conjectandi : De l'usage et l'application de la doctrine précédente aux affaires civiles, morales et économiques,* Jacques Bernoulli s'affirme déterministe :

> « *Tout ce qui bénéficie sous le soleil de l'être ou du devenir, passé, présent ou futur, possède toujours en soi et objectivement une certitude totale.*
> *C'est évident du présent et du passé, ce qui est ou a été ne peut pas ne pas être ou avoir été.*
> *Sur le futur il n'y a pas à discuter ; cependant ce n'est pas par la nécessité de quelque destin qu'il ne peut pas ne pas advenir, mais en raison soit de la prescience soit de la prédétermination divine; car si n'arrivait pas avec certitude tout ce qui est futur, on ne voit pas comment le Créateur suprême pourrait conserver entière la gloire de son omniscience et de son omnipotence.*
> *Quant à dire comment cette certitude de l'avenir peut subsister avec la contingence ou la liberté des causes secondes, que d'autres en disputent ; pour nous, nous ne voulons pas toucher aux points étrangers au but que nous visons.* »

Cette quatrième partie est articulée ainsi :
Chapitre I *: Préliminaires : la certitude, la probabilité, la nécessité, la contingence.*
Bernoulli y donne son point de vue sur la probabilité :

> *« la probabilité est un degré de la certitude et en diffère comme la partie diffère du tout ».*

Chapitre II : *Science et conjecture. L'art de conjecturer. Les arguments des conjectures. Axiomes généraux touchant ces points.*
Il y justifie le titre de son ouvrage :

> « *Conjecturer quelque chose, c'est mesurer sa probabilité : ainsi l'art de conjecturer ou la stochastique se définit pour nous comme l'art de mesurer aussi exactement que possible les probabilités des choses…* ».

Chapitre III : *Les divers espèces d'arguments, et comment estimer leur poids pour supputer les probabilités.*
Bernoulli y donne la première définition de la probabilité :

> « *Posons que le nombre des cas, dans lesquels un argument quelconque peut exister est b ; le nombre de ceux dans lesquels il peut arriver qu'il n'existe pas est c, (…). Or je pose que tous les cas sont également possibles, ou qu'ils peuvent survenir avec une égale facilité ; (…) en sorte qu'un tel argument prouve* $\frac{b}{b+c}$ *de la chose ou de la certitude de la chose* ».

Chapitre IV : *La double manière de chercher les nombres de cas. Ce qu'il faut penser de celui qui est établi par des expériences.*
C'est le grand tournant :

> « *On en est ainsi venu à ce point que pour former selon les règles des conjectures sur n'importe quelle chose, il est seulement requis d'une part que les nombres de cas soient soigneusement déterminés, et d'autre part que soit défini combien les uns peuvent arriver plus facilement que les autres* ».
> « *Mais c'est ici enfin que surgit une difficulté, nous semble-t-il : cela peut se voir à peine dans quelques très rares cas et ne se produit presque pas en dehors des jeux de hasard que leurs premiers inventeurs ont pris soin d'organiser en vue de se ménager l'équité.*
> *(…) Mais qui donc parmi les mortels définira par exemple le nombre de maladies, (…) qui encore recensera les cas innombrables des changements auxquels l'air est soumis chaque jour. (…) Il serait donc absolument d'un insensé de vouloir connaître quelque chose de cette matière* ».

Réponse : l'approche fréquentiste :

> *(…) « Mais à la vérité ici s'offre à nous un autre chemin pour obtenir ce que nous cherchons. » … « Ce qu'il n'est pas donné d'obtenir **a priori** l'est du moins **a posteriori**, c'est-à-dire qu'il sera possible de l'extraire en observant l'issue de nombreux exemples semblables, car on doit présumer que, par la suite, chaque fait peut arriver et ne pas arriver dans le même nombre de cas qu'il avait été constaté auparavant, dans un état de choses semblables, qu'il arrivait ou n'arrivait pas ».*

Mais il reste à démontrer que la fréquence observée d'un événement « *issu de nombreux exemples semblables* » est aussi proche que l'on veut de la probabilité de cet événement, supposé réalisé par une multitude de cas équiprobables inaccessibles.

Chapitre V : *Solution du problème précédent* (le Théorème de Bernoulli) :

D'une urne de Bernoulli contenant t boules dont r blanches (fertiles) et s noires (stériles), on tire nt boules avec remises et on compte les boules blanches obtenues (schéma binomial).

Bernoulli formule ainsi son théorème :

> *« Soit le nombre de cas fertiles (…) au nombre de tous dans le rapport* $\frac{r}{t}$ *(…). On peut concevoir des expériences en un nombre tel qu'il soit plus vraisemblable d'autant de fois que l'on veut (…) que le nombre des observations fertiles soit au nombre de toutes les observations dans un rapport ni plus grand que* $\frac{r+1}{t}$*, ni plus petit que* $\frac{r-1}{t}$ *».*

Traduction moderne :

Une même expérience aléatoire est répétée un nombre n de fois suffisamment grand. On s'intéresse à la fréquence F des issues qui réalisent un événement de probabilité p.

On représente cette situation par un schéma binomial (tirages avec remises dans une urne de Bernoulli, contenant t boules dont r blanches) où l'on pose $\varepsilon = 1/t$ et $p = r/t$. Il y a donc une hypothèse d'équiprobabilité quelque part.

Alors, il y a une probabilité aussi voisine de 1 que l'on veut (niveau de confiance) que la fréquence F des issues réalisant un événement donné soit plus proche que tout ε (précision de l'approximation) de la probabilité p de cet événement.

Cette fréquence observée F peut donc être prise pour estimer cette probabilité p et cet énoncé explicite la condition de confiance : $P(F - \varepsilon < p < F + \varepsilon) > 1 - \alpha$.

IV - Principe de la démonstration de Bernoulli

1 – A l'origine, la formule du binôme

On désigne par $p = r/t$ et $q = s/t$ les probabilités respectives du succès (boule blanche) et de l'échec (boule noire). On répète l'épreuve un grand nombre nt de fois.

Le nombre d'échecs dans cette expérience aléatoire est une variable binomiale *N* dont la loi $B(nt, q)$ est donnée par : $P(N = k) = \dfrac{nt.nt-1.nt-2.\ \ldots\ .nt-k+1}{1.2.3.\ \ldots\ .k} p^{nt-k} q^k$, avec les notations de Bernoulli. Il suppose les probabilités binomiales bien connues : « *c'est une chose connue parmi les géomètres* ». Le théorème est démontré à partir de remarques fines sur les coefficients binomiaux.

Il part de la formule du binôme où t = r + s : $(r + s)^{nt} = \sum_{k=0}^{nt} \binom{nt}{k} r^{nt-k} s^k = \sum_{k=0}^{nt} A(k)$

où j'introduis la notation fonctionnelle anachronique A(k) pour condenser les écritures et les formulations. Les probabilités binomiales valent donc $A(k)/t^{nt}$.

Le lemme suivant semble anodin.

Lemme 2 : *cette somme contient* nt + 1 *termes*

2 – Remarques structurantes sur les termes du développement du binôme
Dans les lemmes 1 et 3, Bernoulli étudie les termes A(k), de rang k.
Il distingue trois termes remarquables qu'il note M = A(ns), de rang ns, L = A(ns-n) placé n termes avant M, et Λ = A(ns+n) placé n termes après M.
On peut représenter les termes A(k) de $(r + s)^{nt}$ sur un schéma :

rang à partir du premier :	0	k	ns-n	ns	ns+n	nt
<u>termes du binôme</u> :	r^{nt}	A(k)	L	M	Λ	s^{nt}
rang à partir du dernier :	nt	nt-k	nr+n	nr	nr-n	0

[----------------------[----------------[]------------------]--------------]

nombre de termes : ns-n n n nr-n

Bernoulli observe que :
Lemme 1 : *Avant* L*, il y a* s – 1 *fois plus de termes qu'entre* L *et* M*, après* Λ *il y a* r – 1 *fois plus de termes qu'entre* M *et* Λ*.*

3 - Remarque (moderne) sur le maximum M
M = A(ns) = $\binom{nt}{ns} r^{nt-ns} s^{ns} = \dfrac{(nt)!}{(nr)!(ns)!}$ est le terme maximal du développement du binôme $(r+s)^{nt} = t^{nt}(p+q)^{nt}$. Donc M/t^{nt} est la valeur maximale de la probabilité binomiale de loi *B*(nt, s/t) d'espérance ns et de variance nrs/t.
L'approximation normale de cette loi, n étant grand, est la loi $N\left(ns, \sqrt{\dfrac{nrs}{t}}\right)$ dont la densité a pour maximum $\dfrac{\sqrt{t}}{\sqrt{2\pi nrs}}$. On en déduit que $M \sim \dfrac{\sqrt{t}}{\sqrt{2\pi nrs}} t^{nt}$ quand n tend vers l'infini.
La formule de Stirling [$n! \sim \left(\dfrac{n}{e}\right)^n \sqrt{2\pi n}$] permet de retrouver directement cette équivalence.

4 - Croissance et décroissance de A(k)
Le **lemme 3** est moins immédiat que les deux premiers, mais sa démonstration est un calcul élémentaire. Il met en évidence le rôle central que joue M : de part et d'autre de M, les termes du binôme vont en décroissant, et cela de plus en plus vite.
Formulation de Bernoulli :

> « M *sera le plus grand de tous, et ceux qui de part et d'autre sont plus proches de lui seront plus grands que d'autres plus éloignés du même côté : mais ce terme* M *sera dans un rapport plus petit avec un terme proche que ce terme avec un terme plus éloigné (pour des intervalles de termes égaux)* ».

Autrement dit : A(k) est une fonction croissante de k de 0 et ns, décroissante de ns à nt.

De plus, pour $0 < k \leq ns$ et j pas trop grand, on a : $1 < \dfrac{A(k)}{A(k-1)} < \dfrac{A(k-j)}{A(k-j-1)}$,

et pour $ns \leq k < nt$ et j pas trop grand, $1 < \dfrac{A(k)}{A(k+1)} < \dfrac{A(k+j)}{A(k+j+1)}$.

En particulier, pour k=ns et j=n, $1<\dfrac{M}{A(ns-1)}<\dfrac{L}{A(ns-n-1)}$ et $1<\dfrac{M}{A(ns+1)}<\dfrac{\Lambda}{A(ns+n+1)}$.

5 - Étude de limites

Les lemmes 4 et 5 étudient les comportements des rapports M/L et M/Λ quand n est grand, préparant la convergence annoncée dans le théorème.

Lemme 4 : Dans le développement de $(r + s)^{nt}$, on peut prendre n assez grand pour que les rapports $\dfrac{M}{L}$ *et* $\dfrac{M}{\Lambda}$ soient aussi grands que l'on veut.

Symboliquement : $\dfrac{M}{L} \xrightarrow[n \to \infty]{} \infty$ et $\dfrac{M}{\Lambda} \xrightarrow[n \to \infty]{} \infty$

Formulation de Bernoulli:

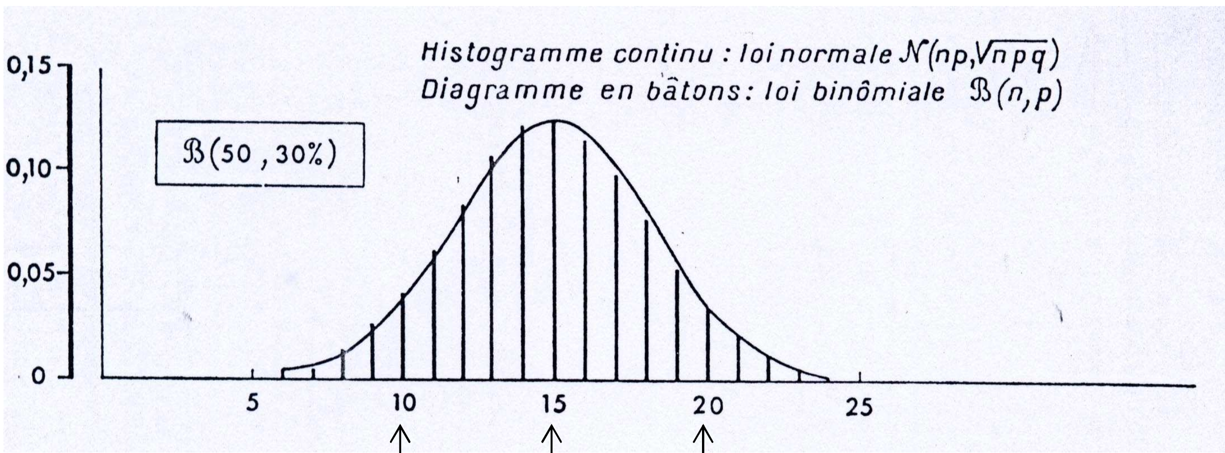
> « *Dans l'élévation à la puissance d'un binôme, dont l'indice est nt, on peut concevoir un nombre n si grand que M le plus grand des termes acquière en comparaison des deux autres L et Λ, distants de M d'un intervalle de n termes à gauche et à droite, un rapport supérieur à la donnée que l'on veut* ».

Lemme 5 : Dans le développement binomial de $(r + s)^{nt}$, pour n assez grand, le rapport de la somme des termes compris entre L et Λ à la somme de tous les autres (ceux qui précèdent L et ceux qui suivent Λ) est plus grand que tout nombre c donné. Traduction symbolique :

$$c < \frac{\displaystyle\sum_{k=ns-n}^{ns+n} A(k)}{(r+s)^{nt} - \displaystyle\sum_{k=ns-n}^{ns+n} A(k)} \xrightarrow[n\to\infty]{} \infty, \text{ ou encore : } \frac{c}{c+1} < \frac{\displaystyle\sum_{k=ns-n}^{ns+n} A(k)}{(r+s)^{nt}} \xrightarrow[n\to\infty]{} 1$$

Ce qui s'écrit : $P(ns-n \le N \le ns+n) = P(q-1/t \le N/nt \le q+1/t) \xrightarrow[n\to\infty]{} 1$

Ces propriétés sont visibles sur le diagramme en bâtons de la loi binomiale, pour n assez grand.



ns–n ns ns+n

$P(ns-n \le N \le ns+n)$

V - Quelques éléments de la démonstration

1 - Démonstration du théorème

Elle découle directement du lemme 5. Bernoulli a énoncé son théorème avec les succès (exp. fertiles) et non les échecs (exp. Stériles). Soit $B_{nt} = nt - N$ le nombre de succès obtenus lorsque l'on répète l'expérience de Bernoulli nt fois. B_{nt} est une variable binomiale.

Soit $F_{nt} = \dfrac{B_{nt}}{nt}$ la fréquence de ces succès.

Avec les notations introduites, où l'indice k désigne le nombre d'échecs, on a :

$P(nt\text{-}ns\text{-}n \le B_{nt} \le nt\text{-}ns\text{+}n) = P(ns-n \le N \le ns+n) = \dfrac{\sum_{k=ns-n}^{ns+n} A(k)}{t^{nt}} \xrightarrow[n\to\infty]{} 1$, ce qui s'écrit pour un c positif quelconque et n assez grand : $P(p - 1/t \le F_{nt} \le p + 1/t) > \dfrac{c}{c+1}$.

En termes modernes, cela signifie que $P(|F_{nt} - p| < 1/t) \xrightarrow[n\to\infty]{} 1$.

On peut choisir le nombre t de boules dans l'urne aussi grand que l'on veut, tel que $1/t < \varepsilon$ donné, ce qui montre que $F_{nt} \xrightarrow[\textit{prob.}]{} p$

2 - Démonstration du lemme 4

Ce lemme consiste en l'étude du comportement des rapports $\dfrac{M}{L}$ et $\dfrac{M}{\Lambda}$ quand n tend vers l'infini. Bernoulli fait d'abord une démonstration intuitive qui ne résiste pas à un contrôle rigoureux. Pressentant la critique, il la reprend dans une *scholie* (remarque critique), de manière bien plus fine. Son étude soigneuse le conduit à une méthode pour déterminer le nombre d'expériences à reproduire pour estimer la probabilité p aussi précisément que l'on veut (à 1/t près), avec une probabilité aussi petite que l'on veut de se tromper (inférieure à 1/c). On a :

$\dfrac{M}{L} = \dfrac{A(ns)}{A(ns-n)} = \dfrac{A(ns)}{A(ns-1)}.\dfrac{A(ns-1)}{A(ns-2)}...\dfrac{A(ns-n+1)}{A(ns-n)}$ (expression analogue pour $\dfrac{M}{\Lambda}$).

Utilisant l'égalité $\dfrac{A(k+1)}{A(k)} = \dfrac{nt-k}{k+1}.\dfrac{s}{r}$ obtenue dans le lemme 3 et remarquant que $nt - ns = nr$, on obtient : $\dfrac{M}{L} = \dfrac{nr+1}{ns}.\dfrac{nr+2}{ns-1}...\dfrac{nr+n}{ns-n+1}.(\dfrac{s}{r})^n$.

Supposant « n *infini* », Bernoulli considère que *« les nombres* 1, 2, 3, *etc. sont négligeables devant* n, [et les facteurs comme] nr+n–1, nr+n–2,... *sont équivalents à* nr+n, *et* ns–n+1, ns–n+2,... *sont équivalents à* ns–n, *et en effectuant la division par* n, ***il s'ensuit*** *que* :

$$\frac{M}{L} = \frac{nr+n}{ns-n}.\frac{nr+n}{ns-n}...\frac{nr+n}{ns-n}.(\frac{s}{r})^n = \left(\frac{rs+s}{rs-r}\right)^n = \left(\frac{r+1}{r}.\frac{s}{s-1}\right)^n$$

et de même : $\dfrac{M}{\Lambda} = \dfrac{ns+n}{nr-n}.\dfrac{ns+n}{nr-n}...\dfrac{ns+n}{nr-n}.(\dfrac{r}{s})^n = \left(\dfrac{rs+r}{rs-s}\right)^n = \left(\dfrac{r}{r-1}.\dfrac{s+1}{s}\right)^n$.

Ce faisant, Bernoulli considère comme *équivalents* à $\dfrac{nr+n}{ns-n}$ des facteurs de la forme $\dfrac{nr+j}{ns-j+1}$ pour j variant de 1 à n, alors que la majoration possible est dans le mauvais sens et ne garantit pas le résultat. De plus, « *le nombre de ces facteurs est* n, *c'est à dire un nombre infini* », remarque qui invalide la démonstration.

Pourtant il conclut que :

> « *les rapports sont des multiples (puissances) infinis des rapports* $\frac{rs+s}{rs-r}$ *et* $\frac{rs+r}{rs-s}$*, aussi, conséquence simple, sont-ils infinis* ».

et il énonce :

> « *le plus grand terme M est envers L et Λ dans un rapport supérieur à tout rapport assignable. C.Q.F.D.* ».

Bernoulli démontre ensuite le lemme 5, puis revient au lemme 4 dans la *scholie* dont l'introduction savoureuse est révélatrice de sa maîtrise de l'infini et implicitement de la notion de limite. **Scholie :**

> « *Il peut être objecté aux 4ème et 5ème lemmes par ceux qui ne sont pas habitués à faire des observations sur l'infini, que même si dans le cas du nombre n infini, les facteurs des quantités qui expriment les rapports* M/L *et* M/Λ , nr+n–1 *etc. et* ns–n+1 *etc. valent autant que* nr+n *et* ns–n*, alors que les nombres* 1, 2, 3 *etc. des facteurs pris un à un s'évanouissent normalement, il peut cependant arriver que, pris tous ensemble ou bien considérés en soi, ils croissent à l'infini (à cause du nombre infini de facteurs) et qu'ainsi ils diminuent infiniment le rapport infiniment multiple du rapport* $\frac{rs+s}{rs-r}$ *ou* $\frac{rs+r}{rs-s}$*, c'est à dire qu'ils le rendent fini. Je ne puis mieux apaiser ce scrupule, que si je montre maintenant un mode d'assignation, en fait un nombre* n *fini, soit une puissance finie du binôme, dans laquelle la somme des termes entre les limites* L *et* Λ *soit à la somme de ceux qui sont en dehors dans un rapport supérieur à un rapport donné ; il va de soi qu'après avoir montré cela l'objection s'écroule nécessairement d'elle-même* ».

Soit un nombre γ aussi grand que l'on veut. Bernoulli montre que l'on peut prendre n assez grand pour que $\frac{M}{L} > \gamma$. On avait : $\frac{M}{L} = \frac{nr+1}{ns}.\frac{nr+2}{ns-1}...\frac{nr+n}{ns-n+1}.(\frac{s}{r})^n = \prod_{j=1}^{n}\frac{nrs+js}{nrs-(j-1)r}$, où chaque facteur est supérieur à 1.

L'idée consiste à minorer le plus grand nombre m possible de facteurs par $\frac{r+1}{r}$, les autres ne changeant pas, et à obtenir le résultat $\frac{M}{L} > \left(\frac{r+1}{r}\right)^m \geq \gamma$ dès que m est assez grand, c'est-à-dire dès que $m \geq m_L \geq \frac{\log\gamma}{\log(r+1)-\log r}$. C'est possible pour les $j \leq n$ tels que $\frac{ns+r+1}{s+r+1} \leq j$. Il y a $m = \frac{n(r+1)\ +\ s}{s+r+1}$ tels facteurs. Ainsi m peut être aussi grand que l'on veut, r et s étant fixés.

Pour avoir la minoration cherchée, il faut prendre $n \geq n_L = m_L + \frac{m_L s - s}{r+1}$. Idem pour $\frac{M}{\Lambda}$.

On prendra finalement $n \geq \sup(n_L, n_\Lambda)$, d'où le nombre nt d'expériences à réaliser.

3 – Démonstration du lemme 5

Dans le lemme 5, on compare la somme des termes compris entre L et Λ à celle des autres termes du développement de $(r+s)^{nt}$. Pour cela, Bernoulli introduit les notations :

F = A(ns-1) désigne le terme précédant M,

P = A(ns-n-1) le terme précédant L, n termes avant F;

G = A(ns-2) ; Q = A(ns-n-2) ; H = A(ns-3) ; R = A(ns-n-3) ; ...; T = A(ns-2n) :

$$\begin{array}{cccccccccccc} 0 & \ldots & T & \ldots & R & Q & P & L & \ldots & H & G & F \; M \end{array}$$

|________________|____________________|__|__|__|__|____________________|__|__|__|

D'après la deuxième partie du lemme 3, avec j = n, on a : $\frac{M}{F} < \frac{L}{P}$, $\frac{F}{G} < \frac{P}{Q}$, $\frac{G}{H} < \frac{Q}{R}$ etc.

On en déduit : $\frac{M}{L} < \frac{F}{P} < \frac{G}{Q} < \frac{H}{R} < \ldots < \frac{L}{T}$ et par conséquent, par un calcul élémentaire, $\frac{M}{L} < \frac{F+G+H+\ldots+L}{P+Q+R+\ldots+T}$. Ce dernier rapport s'écrit $\frac{\sum_{k=ns-n}^{ns-1} A(k)}{\sum_{k=ns-2n}^{ns-n-1} A(k)}$, supérieur à $\left(\frac{r+1}{r}\right)^m \geq \gamma$ dès que $\frac{M}{L}$ l'est. Le numérateur est la somme des n termes de $(r + s)^{nt}$ qui précèdent M et le dénominateur est la somme des n termes qui précèdent L.

Comme (lemme 3) la fonction A(k) est croissante sur [0, ns], on a pour i = 0, 1,..., s–2, les s–1 inégalités suivantes entre les sommes analogues sommant les A(k) par paquets de n termes :

$$\sum_{k=in}^{(i+1)n-1} A(k) \leq \sum_{k=ns-2n}^{ns-n-1} A(k). \text{ D'où } \sum_{k=0}^{ns-n-1} A(k) \leq (s-1) \sum_{k=ns-2n}^{ns-n-1} A(k).$$

Pour un c donné, aussi grand que l'on veut, posant $\gamma = (s - 1)\, c$, r et s étant fixés, on obtient pour m et n assez grands ($n \geq n_L$ du lemme 4) :

$$\frac{\sum_{k=ns-n}^{ns-1} A(k)}{\sum_{k=0}^{ns-n-1} A(k)} \geq \frac{\sum_{k=ns-n}^{ns-1} A(k)}{(s-1)\sum_{k=ns-2n}^{ns-n-1} A(k)} > \frac{1}{s-1}\left(\frac{r+1}{r}\right)^m \geq \frac{\gamma}{s-1} = c$$

On obtient de même, avec $\delta = (r - 1)\, c$ et $n \geq n_\Lambda$:

$$\frac{\sum_{k=ns+1}^{ns+n} A(k)}{\sum_{k=ns+n+1}^{nt} A(k)} \geq \frac{\sum_{k=ns+1}^{ns+n} A(k)}{(r-1)\sum_{k=ns+n+1}^{ns+2n} A(k)} > \frac{1}{r-1}\left(\frac{s+1}{s}\right)^m \geq \frac{\delta}{r-1} = c$$

D'où, en ajoutant M au numérateur ce qui ne fait que renforcer l'inégalité :

$$\frac{\sum_{k=ns-n}^{ns+n} A(k)}{\sum_{k=0}^{ns-n-1} A(k) + \sum_{k=ns+n+1}^{nt} A(k)} > c$$

Le numérateur est la somme des termes de $(r + s)^{nt}$ compris entre L et Λ, le dénominateur est la somme de tous les autres. Le lemme 5 est ainsi démontré.

4 – Épilogue

Bernoulli fait une application numérique. La probabilité p (supposée inconnue) étant fixée, pour une précision ε et un nombre c donnés, il s'agit de déterminer le nombre nt d'expériences à réaliser pour que la fréquence de succès observée F_{nt} soit éloignée de p de

moins de ε avec une probabilité aussi proche de 1 que l'on veut (supérieure au niveau de confiance $\frac{c}{c+1}$).
Bernoulli prend t = 50, r = 30 et s = 20 (p = 3/5), $\varepsilon = 1/t = 0{,}02$ et c = 1 000.
Pour avoir cette précision peu fameuse (elle est en $1/\sqrt{nt}$) de 2 % sur la probabilité à estimer, il calcule qu'à ce niveau de confiance 0,999, il faut réaliser le nombre inaccessible de 25 550 expériences (14 000 au niveau 0,9).
On comprend la déception de Bernoulli qui a laissé *Ars Conjectandi* inachevé, cherchant jusqu'à sa mort en 1705 à améliorer son résultat.
Moivre, exploitant la formule de son ami Stirling (*Doctrine des Chances*, 1718, 1756), obtient des équivalents des probabilités binomiales, plus performants que les majorations de Bernoulli. L'approximation normale qu'il donne permet d'avoir une précision de 3 % au niveau de confiance 0,95 avec seulement 1000 expériences. C'est la situation des sondages habituels (Il suffit de 6 534 expériences avec les données numériques de Bernoulli).

VI – Une démonstration moderne élémentaire du théorème de Bernoulli

Dans un schéma de Bernoulli, soit F_n la fréquence des succès (de probabilité p) en n épreuves. nF_n est une variable binomiale $B(n, p)$, on a $E(nF_n) = np$ et $\mathrm{Var}(nF_n) = np(1-p)$,
d'où $E(F_n) = p$ et $\mathrm{Var}(F_n) = \frac{p(1-p)}{n}$. On note f_i les n+1 valeurs possibles pour F_n [$f_i = i/n$, avec i = 0 … n] et $p_i = P(F_n = f_i)$. [$p_i = C_n^i\, p^i\, (1-p)^{n-i}$].

On a $\mathrm{Var}(F_n) = \sum_{i=0}^{n} (f_i - E(F_n))^2 p_i = \sum_{\{i \mid |f_i - p| < \varepsilon\}} (f_i - p)^2 p_i + \sum_{\{i \mid |f_i - p| \geq \varepsilon\}} (f_i - p)^2 p_i$

d'où $\mathrm{Var}(F_n) = \frac{p(1-p)}{n} \geq \varepsilon^2 \sum_{\{i \mid |f_i - p| \geq \varepsilon\}} p_i = \varepsilon^2 P(|F_n - p| \geq \varepsilon)$.

D'où le théorème de Bernoulli : $P(p \in]F_n - \varepsilon\,;\, F_n + \varepsilon[) \geq 1 - \frac{p(1-p)}{n\varepsilon^2} \xrightarrow[n\to\infty]{} 1$

Cette démonstration établit une convergence moins performante que celle de Bernoulli. Elle s'inspire de celle de l'inégalité de Bienaymé-Tchebychev qui met en œuvre le concept de variance, inconnu de Bernoulli à son époque.

Bibliographie

[1] Bernoulli, J., (1713). *Ars Conjectandi,* 4ème partie, traduction du Latin par Norbert Meunier, IREM de Rouen, 1987.
[2] Meunier, N., (1989). Argumentation et démonstration : à quoi sert la démonstration de la « Loi des grands nombres » de Jacques Bernoulli (1654-1705), *La démonstration mathématique dans l'Histoire*, pp. 81-97, Actes du 7ème colloque inter-IREM Épistémologie et Histoire des Mathématiques, Besançon, 1989.
[3] Henry, M., (2004). La démonstration par Jacques Bernoulli de son théorème, *Histoires de probabilités et de statistiques*, Ellipses, pp. 121-140.
[4] Lanier, D., (1992). Huygens : l'espérance et l'infini, *Histoire d'infini*, pp. 555-577, Actes du 9ème colloque inter-IREM Épistémologie et Histoire des Mathématiques, IREM de Brest, 1992.
[5] Pichard, J.F., (2001). Les probabilités au tournant du XVIIIème siècle, *Autour de la modélisation en probabilités*, Commission inter-IREM Statistique et Probabilités, Presses Universitaires Franc-Comtoises, pp. 13-46.