La planche de Galton

Une introduction à la « loi des grands nombres » en probabilités

Frédéric Élie, juillet 2014

Une façon d'illustrer expérimentalement la loi de probabilité binomiale, et de se convaincre théoriquement de la « loi des grands nombres », consiste à utiliser la planche de Galton (Galton (1822-1911), cousin du naturaliste Charles Darwin, chercha à utiliser les probabilités pour expliquer les lois de l'hérédité dans une population).

SOMMAIRE

1 – Planche de Galton

La planche de Galton est constituée d'une planche sur laquelle des clous sont plantés de façon à former un réseau triangulaire de n étages, et où les nœuds sont les clous disposés en quinconce entre deux rangées consécutives (**figure 1**). Chaque fois que l'on passe d'une rangée à la suivante, le nombre de clous augmente de 1. A la rangée n°1 on a donc un seul clou, à la dernière rangée on a n clous. En dessous de la dernière rangée n, on aménage n+1 compartiments destinés à récupérer à la fin de sa course une bille lâchée depuis le sommet du triangle.

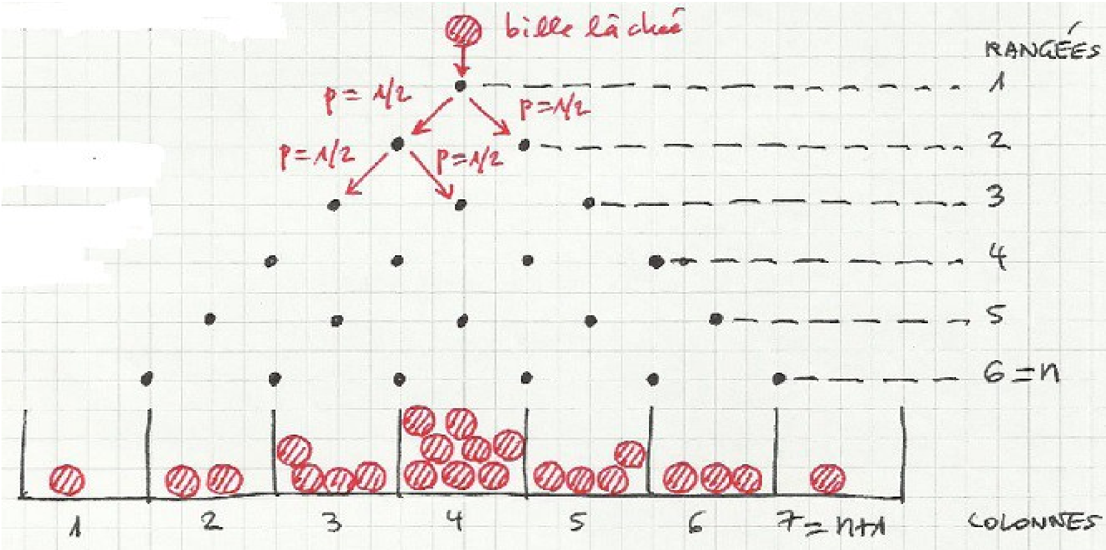

figure 1 – planche de Galton, exemple pour n = 6

On lâche la bille au-dessus du clou de la rangée n°1, située au sommet (la planche, de surface la plus lisse possible, est inclinée pour permettre à la bille de rouler sans quitter la planche).
Lorsque la bille rencontre le clou, elle a une probabilité d'aller à droite notée p, et une probabilité d'aller à gauche notée q = 1 – p (puisque l'on doit avoir p + q = 1). Cette probabilité élémentaire, p, est la même quelle que soit la rangée puisque l'on suppose qu'aucun facteur physique propre à chaque rangée (rugosité, déformations, etc.) n'affecte la descente de la bille.
La bille passe ensuite dans la 2e rangée et rencontre l'un des clous qui y sont présents. La probabilité d'aller à droite est de nouveau p et celle d'aller à gauche q = 1 – p.
Et ainsi de suite jusqu'à la dernière rangée n.

Lorsque la bille rencontre un clou et passe à droite, c'est une expérience aléatoire dont la variable aléatoire est:

$$X = \{\text{bille passe à droite}\}$$

(on aurait tout autant pu choisir X = {bille passe à gauche}).
Le nombre de fois, x, où la bille est allée à droite lors de son parcours du sommet jusqu'à la dernière rangée n est aussi une expérience aléatoire de probabilité notée $P_X(x)$. On a évidemment $0 \leq x \leq n$ ainsi que:

$$P_X(1) = p = \text{probabilité que la bille aille 1 fois à droite}$$
$$P_X(0) = q = 1 - p = \text{probabilité que la bille aille 1 fois à gauche}$$

La rencontre de la bille avec un clou équivaut à un tirage, ou lancer d'une pièce pouvant retomber soit sur face (ce qui correspond à: « la bille va à droite »), soit sur pile (ce qui correspond à: « la bille va à gauche »).
On introduit alors la notion du ***test d'hypothèse nulle H_0***. Il s'exprime par: *« La valeur moyenne d'une grandeur sur un échantillon observé est significativement peu différente d'un résultat théorique prédéfini »*.
Dans l'exemple des tirages d'une pièce tombant sur pile ou face, l'hypothèse H_0 devient: *« Le résultat des n = x + y tirages, qui ont donné x fois pile et y fois face, est significativement peu différent de celui obtenu avec une pièce non pipée »*.
Pour l'expérience de la planche de Galton, qui lui est équivalente, l'hypothèse H_0 s'exprime ainsi: *« Le*

résultat des n tirages, c'est-à-dire des n rencontres de la bille avec les clous, donne x déviations vers la droite (ou si l'on préfère l'événement X s'est produit x fois), et ce résultat est significativement peu différent de celui obtenu avec une planche de Galton où la bille glisse parfaitement, où les clous sont plantés de la même manière, et où la bille est parfaitement sphérique ».

Sous l'hypothèse H_0, pour la planche de Galton, la probabilité p d'obtenir une déviation de la bille vers la droite est égale à la probabilité q d'obtenir une déviation vers la gauche:

$$p = q \text{ avec } p + q = 1 \text{ donc } p = q = 1/2$$

Lorsque la bille a parcouru les n étages de clous (n tirages), la probabilité $P_X(x)$ d'obtenir x fois une déviation à droite est donnée par la ***loi binomiale de Jacques Bernoulli***:

$$P_X(x) = \frac{n!}{x!(n-x)!} p^x q^{n-x} \quad (1)$$

Avec p = q = 1/2, (1) se simplifie en:

$$P_X(x) = \frac{1}{2^n} \frac{n!}{x!(n-x)!} \quad (2)$$

Exemple: dans la configuration de la **figure 1** (n = 6), cherchons la probabilité pour que la bille arrive dans le compartiment du milieu, c'est-à-dire le numéro 4.
Pour obtenir cela, il faut que la bille ait dévié 3 fois sur la droite le long de son parcours, on a donc x = 3; (2) donne donc:

$$P_X(3) = \frac{1}{2^6} \frac{6!}{3!(6-3)!} = \frac{1}{2^6} \frac{6 \times 5 \times 4}{3 \times 2} = 0{,}3125$$

Si l'on cherche la probabilité d'une alternative, « la bille dévie x fois à droite <u>ou</u> y fois à droite », elle est égale à la somme des probabilités de chacun de ces deux événements:

$$P_X(X = x \text{ } \underline{ou} \text{ } X = y) = P_X(x) + P_X(y)$$

soit:

$$P_X(X=x \text{ ou } X=y) = \frac{1}{2^n} \frac{n!}{x!(n-x)!} + \frac{1}{2^n} \frac{n!}{y!(n-y)!} = \frac{n!}{2^n} \left[\frac{1}{x!(n-x)!} + \frac{1}{y!(n-y)!} \right] \quad (3)$$

Par exemple, probabilité de dévier 4 fois à droite (donc d'arriver à la case n°5) ou de dévier 6 fois à droite (donc d'arriver à la case n°7), avec x = 4 et y = 6, (3) donne:

$$P(x=4 \text{ ou } y=6) = \frac{6!}{2^6} \left(\frac{1}{4!2!} + \frac{1}{6!0!} \right) = 0{,}25$$

Si x = 0 (aucune déviation à droite, donc la bille termine sa course dans la case n°1), avec toujours y = 6 (la bille arrive à la case n°7), autrement dit la probabilité que la bille arrive à une extrémité, (3) donne:

$$P(x=0 \text{ ou } y=6) = \frac{6!}{2^6} \left(\frac{1}{0!6!} + \frac{1}{6!0!} \right) = \frac{1}{2^5} = 0{,}03125$$

On voit que la probabilité que la bille arrive à l'une des extrémités est beaucoup plus faible que celle de l'exemple précédent où elle arrive aux cases du milieu.

Quel que soit le chemin parcouru, la bille atteint l'une des cases n°1 à 7: il s'agit d'un événement certain donc de probabilité égale à 1. Elle est par ailleurs égale à la somme des probabilités de tous les événements possibles « la bille dévie x fois à droite » avec $0 \leq x \leq n$, par conséquent:

$$P_X(0 \leq x \leq n) = P_X(0) + P_X(1) + ... + P_X(n) = \frac{1}{2^n} \sum_{x=0}^{n} \frac{n!}{x!(n-x)!} = \frac{1}{2^n} \sum_{x=0}^{n} C_n^x$$

où $C_n^x = \frac{n!}{x!(n-x)!}$ est le nombre de combinaisons de n objets pris x à x. Il intervient dans le développement du binôme de Newton:

$$(a+b)^n = \sum_{x=0}^{n} C_n^x a^{n-x} b^x \quad (4)$$

En particulier, si a = b = 1, il vient:

$$2^n = \sum_{x=0}^{n} C_n^x \quad (5)$$

On obtient donc bien:

$$P_X(0 \leq x \leq n) = \frac{2^n}{2^n} = 1 \quad \text{(5bis)}$$

2 – Expression générale de la loi binomiale

On a posé a priori l'équiprobabilité des événements élémentaires: p = q = 1/2. Dans une autre configuration, on peut avoir de manière générale $p \neq q$ avec p + q = 1, une déviation étant plus privilégiée qu'une autre. Dans ce cas, il faut conserver la forme générale (1) de la loi binomiale:

$$P_X(x) = C_n^x p^x q^{n-x} = C_n^x p^x (1-p)^{n-x} \quad (6)$$

La relation (6) exprime la loi binomiale, notée $\boldsymbol{B}(n, p)$, de paramètre p, correspondant au résultat de la somme de n tirages indépendants, chacun des tirages étant un événement élémentaire dont la probabilité est p.

3 – Espérance et variance

La variable x, nombre de déviations à droite, est une variable aléatoire discrète: $0 \leq x \leq n$. Pour chaque x, la probabilité est $P_X(x)$ donnée par (6) dans le cas d'une loi binomiale. La moyenne théorique de x, calculée sur l'ensemble des n épreuves, est appelée ***espérance*** de X et est définie par:

$$E(X) = \sum_{x=0}^{n} x P_X(x) \quad (7)$$

Chaque réalisation x, prise par la variable aléatoire X, a un écart à la moyenne attendue égale à:

x – E(X)

Pour l'ensemble des réalisations x, $0 \leq x \leq n$, la moyenne des écarts est:

$$E(X - E(X)) = \sum_{x=0}^{n} (x - E(X)) P_X(x) \quad (8)$$

(8) est l'espérance de (X-E(X)) c'est-à-dire des écarts. Elle est nulle, en effet:

$$E(X-E(X))=\sum_{x=0}^{n} x P_X(x)-E(X)\sum_{x=0}^{n} P_X(x)$$

Comme $\sum_{x=0}^{n} P_X(x)=1$, et d'après (7), il vient: $E(X-E(X))=E(X)-E(X)\times 1=0$

L'expression (8) ne permet donc pas de caractériser la dispersion des réalisations de X. C'est pourquoi on choisit de caractériser la dispersion par l'espérance du carré des écarts, autrement dit la ***variance*** Var(X), d'autant que cette quantité intervient dans les droites de régression linéaire:

$$Var(X)=\sigma^2(X)=E((X-E(X))^2)=\sum_{x=0}^{n}(x-E(X))^2 P_X(x) \quad (9)$$

En développant (9) on obtient:

$$\mathrm{Var}(X)=\sum_{x=0}^{n}(x^2+E(X)^2-2xE(X))P_X(x)=\sum_{x=0}^{n} x^2 P_X(x)+E(X)^2-2E(X)\sum_{x=0}^{n} x P_X(x)$$

Le premier terme est E(X²), le troisième est: 2E(X)E(X) = 2E(X)².
Finalement:

$$\mathrm{Var}(X)=\sigma^2(X)=E(X^2)-E(X)^2 \quad (10)$$

La racine carrée de la variance, σ, est l'***écart-type***:

$$\sigma(X)=\sqrt{E(X^2)-E(X)^2} \quad (11)$$

Appliquons les définitions (7) et (10) à la planche de Galton, dont la loi de probabilité est la loi binomiale (6) de paramètre p. On obtient:

$$E(X)=\sum_{x=0}^{n} x P_X(x)=\sum_{x=0}^{n} x C_n^x p^x(1-p)^{n-x}=\sum_{x=0}^{n} x \frac{n!}{x!(n-x)!} p^x(1-p)^{n-x}=np$$

$$\mathrm{Var}(X)=\sigma^2(X)=E(X^2)-E(X)^2=\sum_{x=0}^{n} x^2 C_n^x p^x(1-p)^{n-x}-n^2p^2=np(1-p)$$

En conclusion, pour la loi binomiale de paramètre p, on a:

$$\begin{array}{c} E(X)=np \\ \sigma^2(X)=np(1-p) \end{array} \quad (12)$$

Et si p = 1/2, (12) donne:

$$\left.\begin{array}{r} E(X)=\dfrac{n}{2} \\ \sigma^2(X)=\dfrac{n}{4} \end{array}\right| \quad (13)$$

4 – Variables aléatoires continues: fonction de répartition et densité de probabilité

Si x est une variable aléatoire discrète – comme c'est le cas pour l'expérience de la planche de Galton – la probabilité d'observer la réalisation X = x est, par définition:

$$P_X(X=x)=P_X(x)=\lim_{n\to\infty}\frac{N(x)}{n} \quad (14)$$

où N(x) est le nombre de fois que se produit la réalisation X = x, et n le nombre total d'épreuves.

Comme $n=\sum_{x=0}^{n} N(x)$ on a:

$$P_X(0\leq x\leq n)=\sum_{x=0}^{n} P_X(x)=1 \quad (15)$$

qui est la propriété (5bis) déjà vue.

Dans le cas où X est une ***variable aléatoire continue***, la probabilité d'obtenir X = x est par définition nulle, puisque suite à (14), c'est le rapport d'un nombre de réalisations de X = x sur un nombre infini non dénombrable d'épreuves:

$$P_X(X=x)=0 \quad (16)$$

On peut alors seulement utiliser la probabilité d'obtenir une valeur de X dans un intervalle infinitésimal (x,x+dx):

$$P_X(x\leq X\leq x+dx)=dF(x)$$

Par définition, la ***fonction de répartition*** de X est F(x) et est définie par:

$$F(x)=P_X(X<x)=\int_{-\infty}^{x} dF(y) \quad (17)$$

En particulier, la probabilité d'obtenir la variable aléatoire X dans un intervalle fini [a,b] est:

$$P_X(a\leq X\leq b)=F(b)-F(a)$$

On démontre (cf. [1]) les propriétés suivantes de F:

- F(x) est croissante
- F(x) est bornée entre 0 et 1: $\forall\ x\in\mathbb{R},\ 0\leq F(x)\leq 1$
- $1-F(x)=1-P_X(X<x)=P_X(X\geq x)$
- $P_X(a\leq X\leq b)=f(b)-F(a)=1-(P_X(X\leq a)+P_X(X\geq b))$
- $F(-\infty)=\lim_{x\to-\infty} F(x)=0$
- $F(+\infty)=\lim_{x\to+\infty} F(x)=1$
- F(x) est continue si, et seulement si, $\forall\ x\in\mathbb{R},\ P_X(X=x)=0$ qui est la condition (16), appelée encore ***condition de probabilité diffuse***;
- F(x) existe aussi pour les variables aléatoires discrètes. Dans ce cas elle est discontinue et prend l'aspect d'une fonction en escalier.

L'allure générale de F(x) est donc celle de la **figure 2**.

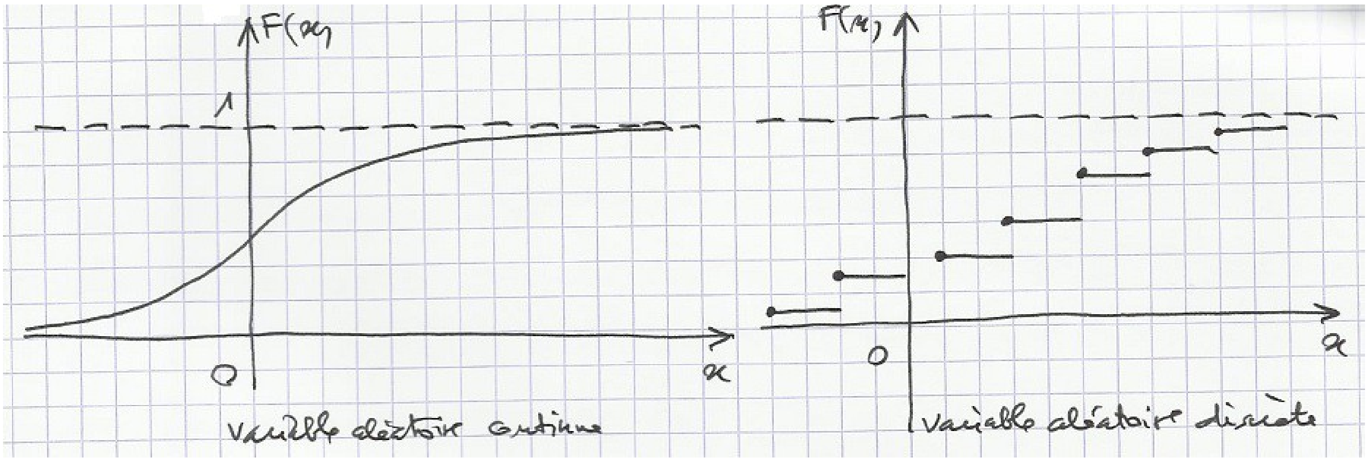

figure 2 – fonction de répartition

Pour une variable aléatoire continue X, la ***densité de probabilité*** ρ(x) est par définition:

$$F(x)=\int_{-\infty}^{x}\rho(y)dy \quad (18)$$

Compte tenu de (17), on serait tenté d'identifier: dF(x) = ρ(x)dx, autrement dit d'assimiler ρ à la dérivée de F. De façon générale, il n'en est rien: cette identification est vraie seulement « presque partout » au sens de la théorie de la mesure en analyse mathématique [1].
Dans la cas de variable aléatoire continue, les expressions (7) et (11) de l'***espérance*** et de la ***variance*** sont remplacées par:

$$E(X)=\int_{-\infty}^{+\infty} y\rho(y)dy \quad (19)$$

$$\mathrm{Var}(X)=\sigma^2(X)=\int_{-\infty}^{+\infty}(y-E(X))^2\rho(y)dy=\int_{-\infty}^{+\infty}y^2\rho(y)dy-E(X)^2 \quad (20)$$

On vérifie, en employant la théorie des distributions, que les définitions (19) et (20) permettent de retrouver (7) et (11) lorsque X est une variable aléatoire discrète.

5 – Moyenne arithmétique; théorème de l'inégalité de Bienaymé-Chebyshev

Dans le cas particulier de la planche de Galton, les variables aléatoires X sont discrètes et prennent leurs valeurs x parmi les réalisations 0, 1, 2, …, n où n est le nombre d'épreuves; c'est pourquoi on mentionne $0\leq x\leq n$, ce qui nous dispensait d'indicer x par x_k, avec $0\leq k\leq n$.
Dans une toute autre expérience, on pourrait avoir x prenant ses valeurs x_k dans un domaine très différent que 0, 1, 2, …, n; c'est pourquoi, de manière générale, il est plus correct et moins source de confusion, d'utiliser dans le cas de variables aléatoires discrètes, une notation indicée pour les réalisations, $X = x_k$.
Ce faisant, les définitions de la probabilité, de l'espérance et de la variance (14), (7) et (10) s'écrivent rigoureusement:

$$P_X(X=x_k)=\lim_{n\rightarrow\infty}\frac{N(x_k)}{n} \quad (21)$$

$$E(X)=\sum_{k=1}^{n}x_k P_X(X=x_k) \quad (22)$$

$$\sigma^2(X)=E((X-E(X))^2)=\sum_{k=1}^{n}(x_k-E(X))^2 P_X(x_k)=E(X^2)-E(X)^2 \quad (23)$$

Avec ces notations, la ***moyenne arithmétique*** d'un ensemble de n valeurs $(x_k)_{1\leq k\leq n}$ est par définition:

$$m=\frac{1}{n}\sum_{k=1}^{n}x_k \quad (24)$$

Remarque très importante: - Pour la planche de Galton, le nombre possible de valeurs de x_k est n+1, et non n, et $0\leq x_k\leq n$. L'espérance (22) devient alors:

$$E(X)=\frac{1}{n+1}\sum_{x=0}^{n}x$$

La somme est celle des n premiers entiers, elle vaut $\frac{1}{2}n(n+1)$ d'où:

$$E(X)=\frac{1}{n+1}\frac{n(n+1)}{2}=\frac{n}{2}$$

Ainsi donc, pour une planche de Galton « non pipée », l'espérance des déviations à droite (ou à gauche) est égale à la moitié du nombre d'étages, ou, ce qui revient au même, du nombre d'épreuves élémentaires. On remarquera que E(X) est égale à (13) comme il se doit, pour une distribution binomiale de paramètre p = 1/2.
Attention: E(X) n'est pas égale à la moyenne arithmétique des réalisations observées m. Comme on l'a déjà annoncé, E(X) est la moyenne _attendue_, tandis que dans l'expérience réelle, certaines valeurs prises par X, x_k, peuvent se répéter ou au contraire, certaines autres ne pas apparaître du tout. On verra que l'on peut assimiler m à E(X) seulement lorsque le nombre d'épreuves n est très grand, ce qui s'exprime par la ***loi faible des grands nombres*** dont la démonstration exploite le ***théorème de l'inégalité de Chebyshev***.

Théorème de l'inégalité de Chebyshev (ou de Bienaymé-Chebyshev)[1]:
Soit un échantillon de valeurs $\{x_k\}$, $0\leq k\leq n$ prises par une variable aléatoire X, dont l'espérance μ = E(X) et la variance $\sigma^2(X)$ sont supposées connues. Le nombre n' de valeurs situées à l'extérieur d'un intervalle [-ε, +ε] centré sur μ vérifie:

$$\frac{n'}{n}=P(|X-\mu|\geq\varepsilon)\leq\frac{\sigma^2}{\varepsilon^2} \quad (25)$$

(ou encore: $P(|X-\mu|\geq\sigma\,\varepsilon)\leq 1/\varepsilon^2$) (25bis).

DEMONSTRATION (heuristique): - On classe les écarts à l'espérance par valeurs croissantes: $|x_k-\mu|$. A partir de l'indice j, les écarts deviennent plus grandes que ε, on a donc $|x_k-\mu|\geq\varepsilon$ pour $j\leq k\leq n$. La variance se décompose alors en deux parties:

$$\sigma^2(X)=\frac{1}{n}\left[\sum_{k=1}^{j}(x_k-\mu)^2+\sum_{k=j+1}^{n}(x_k-\mu)^2\right]$$

Les termes de la première somme, étant par définition plus petits que ε², peuvent être négligés, ε étant choisi arbitrairement petit. Il reste donc:

1 Le théorème a été inventé par le mathématicien russe Pafnouti Chebyshev (1821-1894) et c'est son ami, le mathématicien français Irénée Jules Bienaymé (1796-1876) qui, traduisant toutes les œuvres de Chebyshev, l'a fait connaître en Europe.

$$\sigma^2(X) > \frac{1}{n}\sum_{k=j+1}^{n}(x_k-\mu)^2$$

Or pour $k \geq j+1$ on a par définition $(x_k-\mu)^2 \geq \varepsilon^2$ par conséquent:

$$\sigma^2(X) > \frac{1}{n}\sum_{k=j+1}^{n}\varepsilon^2 = \frac{n-j}{n}\varepsilon^2 = \frac{n'}{n}\varepsilon^2 \quad \text{puisque n' = n – j}$$

Mais n'/n est la probabilité qu'une valeur $|x_k-\mu|$ soit $\geq \varepsilon$ donc:

$$\frac{n'}{n} = P(|x_k-\mu| \geq \varepsilon) < \frac{\sigma^2}{\varepsilon^2} \quad \text{- CQFD}$$

L'inégalité (25) permet d'estimer l'erreur commise lorsque, dans une série de réalisations d'une variable aléatoire, on remplace celle-ci par son espérance μ. Par exemple, (25bis) montre que le résultat diffère de μ de plus de 10σ (ε = 10) avec une probabilité plus petite que 1/ε² = (1/10)² = 1%.
En fait, la connaissance de la densité de probabilité permet d'affiner cette estimation de l'erreur.

6 – Nombre d'épreuves n très élevé: loi faible des grands nombres (théorème de Jacques Bernoulli) et théorème central limite

Lorsqu'on augmente le nombre d'épreuves n, pour une même variable aléatoire X, de réalisations x_k suivant la même loi de probabilité, on s'attend à ce que leur moyenne arithmétique m_n, définie par (24), s'approche de l'espérance μ = E(X), autrement dit que la probabilité d'un écart entre μ et m_n tende vers zéro quand n tend vers l'infini. C'est ce qu'exprime la loi faible des grands nombres, théorème établi par Jacques Bernoulli en 1713.

Théorème: loi faible des grands nombres: - Soit une suite de variables aléatoires (X_k), $1 \leq k \leq n$, indépendantes, de même espérance μ = $E(X_n)$ et de même variance σ². On définit la ***suite de Cesáro***:

$$m_n = \frac{X_1+X_2+...+X_n}{n}$$

Alors (m_n) converge en probabilité vers la variable aléatoire constante égale à μ:

$$\forall\ \varepsilon>0,\ \lim_{n\to\infty} P(|m_n-\mu| \geq \varepsilon) = 0 \qquad (26)$$

DEMONSTRATION: - Il est clair, d'après (24), que $E(m_n)$ = m (la moyenne arithmétique). En appliquant (25) à X = m_n et μ = $E(m_n)$ = m, σ² = $\sigma^2(m_n)$, on a:

$$P(|m_n-\mu| \geq \varepsilon) = P(|m_n-m| \geq \varepsilon) \leq \frac{\sigma^2(m_n)}{\varepsilon^2}$$

Or: $\sigma^2(m_n) = \sigma^2\left(\frac{X_1+...+X_n}{n}\right) = \sigma^2\left(\frac{X_1}{n}+...+\frac{X_n}{n}\right)$

Les X_k/n étant indépendants, on a l'égalité de Bienaymé:

$$\sigma^2\left(\frac{X_1}{n}+...+\frac{X_n}{n}\right) = \frac{1}{n^2}\sigma^2\left(X_1+...+X_n\right) = \frac{1}{n^2}(\sigma^2(X_1)+...+\sigma^2(X_n)) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

puisque $\sigma^2(X_k)=\sigma^2 \ \forall k$. D'où:

$$P(|m_n-m|\geq\varepsilon)\leq\frac{\sigma^2}{n\varepsilon^2}$$

qui tend vers 0 quand $n\rightarrow\infty$ - CQFD.

NB – Le théorème est qualifié de « faible » car il suppose l'existence des variances $\sigma^2(m_n)$ et leurs égalités. On peut se passer de cette hypothèse moyennant une démonstration plus compliquée, conduisant à la « ***loi forte des grands nombres*** ».

On a vu que la suite (m_n) des moyennes de Cesáro converge vers la variable aléatoire constante m, c'est-à-dire la moyenne que l'on peut assimiler à l'espérance µ lorsque le nombre d'épreuves n est très grand. Le théorème qui suit énonce que, si chaque terme X_n de la suite de Cesáro est une variable aléatoire indépendante des autres et de même loi de probabilité qu'elles, alors la suite converge, en loi, vers une variable aléatoire qui suit la loi normale ou loi gaussienne.
En d'autres termes, la loi de probabilité des variables aléatoires X_n peut être approchée par la loi normale lorsque n devient très grand. Cette loi normale est paramétrée par la variance σ² des variables aléatoires X_n et leur espérance µ (laquelle peut être éliminée par un changement d'échelle appropriée, on parle alors de loi normale centrée).
Le théorème central limite est énoncé ci-après, sans démonstration, celle-ci fait appel aux ***fonctions caractéristiques des lois de probabilité***[2] et au ***théorème de continuité de Paul Lévy*** (1886-1971).

Définition de la loi normale: - Une loi normale (ou gaussienne) de moyenne m, de variance σ², notée **N**(m,σ²), est une loi de probabilité de densité:

$$\rho_{m,\sigma}(x)=\frac{1}{\sqrt{2\pi}\,\sigma}\exp\left[-\frac{(x-m)^2}{2\sigma^2}\right] \quad (27)$$

Elle est dite loi normale centrée si la moyenne est nulle m = 0 ou rendue nulle par un changement approprié des coordonnées x: **N**(0,σ²).

Le changement d'échelle qui permet d'éliminer m conduit à la variable réduite:

$$z=\frac{x-m}{\sigma}$$

et (27) devient alors:

$$\rho_{0,\sigma}(z)=\frac{1}{\sqrt{2\pi}\,\sigma}\exp\left(-\frac{z^2}{2}\right) \quad (28)$$

Théorème central limite [3] : - Soit la suite des moyennes de Cesáro dont les termes sont des variables aléatoires indépendantes X_k, de même loi de probabilité, d'espérance µ et de variance finie σ² ([4]):

$$m_n=\frac{X_1+X_2+...+X_n}{n}$$

---

2 La fonction caractéristique $\Phi_X(\omega)$ d'une variable aléatoire X est la transformée de Fourier de sa densité de probabilité ρ(x):

$$\Phi_X(\omega)=\int_{-\infty}^{+\infty} e^{i\omega x}\rho(x)\,dx$$

L'intérêt des fonctions caractéristiques est qu'elles permettent de calculer la loi de probabilité de sommes et de produits de variables aléatoires.

3 Il a été démontré pour la première fois par Abraham De Moivre (1667-1754) dans le cas d'une loi binomiale.

4 La condition σ² finie est importante: intuitivement, on conçoit bien, en effet, que si une distribution présente un écart à la moyenne infini elle ne peut pas être centrée sur une moyenne selon la loi binomiale. Par exemple, la loi de probabilité de Cauchy présente un σ² non finie donc on ne peut pas lui appliquer le théorème central limite.

Alors la suite de variables aléatoires $\sqrt{n}(m_n - \mu)$ converge en loi vers la loi normale centrée **N**(0,σ²):

$$\forall\ x \in \mathbb{R},\ \lim_{n\to\infty} P\left(\frac{m_n - \mu}{\sigma}\sqrt{n} < x\right) = \boldsymbol{N}(x) = \int_{-\infty}^{x} \exp(-u^2/2)\, du \qquad (29)$$

L'expression (29) signifie, en clair, que les réalisations qui présentent un écart donné à la moyenne suivent une distribution dont la fonction de répartition ***N***(x) a pour densité celle de la loi normale:

$$\boldsymbol{N}(x) = \sqrt{2\pi}\,\sigma \int_{-\infty}^{x} \rho_{0,\sigma}(u)\, du$$

La densité (28) est exprimée sous la forme normalisée:

$$\rho_{0,1}(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \qquad (30)$$

dont la courbe (dite courbe de Gauss) est représentée à la **figure 3**. La fonction de répartition

$$F(z) = P_X(y < z) = \int_{-\infty}^{z} \rho_{0,1}(y)\, dy \qquad (31)$$

est égale à l'aire comprise entre la courbe $\rho_{0,1}$ (z) et l'axe Oz pour le domaine y < z. Pour y > z' la fonction de répartition est F(z') = 1 – P_X (y < z') = $1 - \int_{-\infty}^{z'} \rho_{0,1}(y)\, dy$.

Les tables, dont un extrait est donné **tableau 1**, donnent F(z) = P_X (y < z) c'est-à-dire l'intégrale (31):

z	F(z)	z	F(z)	z	F(z)
-4,265	0,00001	-1,036	0,15	1,282	0,90
-3,719	0,0001	-0,842	0,20	1,341	0,91
-3,090	0,001	-0,674	0,25	1,405	0,92
-2,576	0,005	-0,524	0,30	1,476	0,93
-2,326	0,01	-0,385	0,35	1,555	0,94
-2,054	0,02	-0,253	0,40	1,645	0,95
-1,960	0,025	-0,126	0,45	1,751	0,96
-1,881	0,03	0	0,50	1,881	0,97
-1,751	0,04	0,126	0,55	1,960	0,975
-1,645	0,05	0,253	0,60	2,054	0,98
-1,555	0,06	0,385	0,65	2,326	0,99
-1,476	0,07	0,524	0,70	2,576	0,995
-1,405	0,08	0,674	0,75	3,090	0,999
-1,341	0,09	0,842	0,80	3,719	0,9999
-1,282	0,10	1,036	0,85	4,265	0,99999

tableau 1 – distribution normale standard $F_{0,1}(z)$

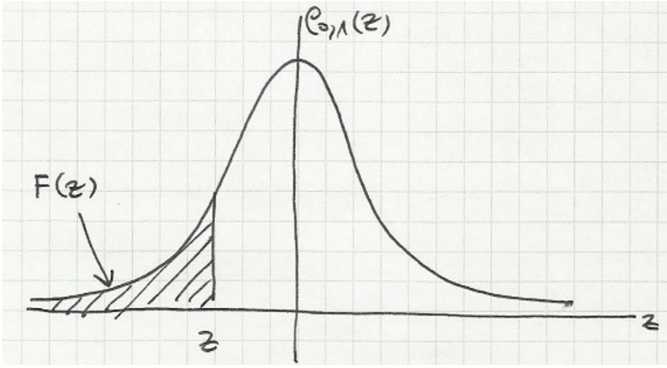

figure 3 – densité de probabilité de la loi normale centrée

Exemple pour la loi binomiale:

Reprenons le cas de la planche de Galton à n = 6 étages (i.e. 6 épreuves pour une même bille). Les relations (12) donnent μ = np = 6/2 = 3 et σ² = np(1-p) = 6/4 = 1,5.
Si x est le nombre de déviations à droite, la variable réduite est:

$$z=\frac{x-m}{\sigma}=\frac{x-3}{\sqrt{1,5}}$$

et la densité « gaussienne » est: $\rho_{0,1}(z)=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{z^2}{2}\right)$ et la fonction de répartition:

$$F_{0,1}(z)=\int_{-\infty}^{z}\rho_{0,1}(z)\,dz$$

Cherchons la probabilité avec laquelle la bille dévie au moins 3 fois à gauche: $x\leq 3$ donc $z\leq 0$. Elle est égale à $F_{0,1}(3)$, d'après le **tableau 1**, on obtient $F_{0,1}(3)$ = 0,50 soit $P(x\leq 3)=0{,}50$ (la bille a autant de chance de dévier à droite qu'à gauche). Rappelons-nous que, d'après la loi binomiale, la probabilité pour que la bille dévie *exactement* 3 fois à droite a été trouvée égale à $P_X(3)$ = 0,3125. Et la probabilité qu'elle dévie *au plus* 3 fois à droite est, d'après (2):

$$P_X(0\leq x\leq 3)=P_X(0)+P_1(1)+P_X(2)+P_X(3)=\frac{6!}{2^6}\left[\frac{1}{0!(6-0)!}+\frac{1}{1!(6-1)!}+\frac{1}{2!(6-2)!}+\frac{1}{3!(6-3)!}\right]=0{,}656$$

valeur à comparer avec $F_{0,1}(x\leq 3)=0{,}50$ dans le cas de la limite gaussienne (n grand).
Pour n = 6 (valeur faible de nombres d'épreuves) la distribution par rapport à la moyenne μ = 3 est éloignée de ce que donne la loi des grands nombres.
Pour n = 10 la moyenne est μ = 5 et la distribution $P(0\leq x\leq 5)$ donne:

$$P(0\leq x\leq 5)=\frac{10!}{2^{10}}\sum_{x=0}^{5}\frac{1}{x!(10-x)!}=0{,}623>0{,}50$$

Pour n = 20, la moyenne est μ = 10 et l'on obtient:

$$P(0\leq x\leq 10)=\frac{20!}{2^{20}}\sum_{x=0}^{10}\frac{1}{x!(20-x)!}=0{,}623>0{,}588$$

valeur encore supérieure à F(μ) mais qui s'en approche.
De façon générale $P(0\leq x\leq\mu)$ tend vers F(μ) = 0,50 de la distribution normale lorsque le nombre d'épreuves n devient très grand.

Pour une planche de Galton à n = 100 étages, on a μ = 50 et σ² = 25. Alors z = (x – μ)/σ = (x – 50)/5. la probabilité pour que la bille dévie au plus 30 fois à droite ($x \leq 30 \rightarrow z \leq -4$) est , d'après le **tableau 1**: $f_{0,1}(-4) = P_X(x \leq 30) \approx 0{,}0001$ elle est donc très faible ! Une bille qui dégringole 100 étages de planche de Galton a une chance de dévier au maximal 30 fois à droite égale à 0,01% ! Un forain qui promettrait un gain important aux joueurs qui parieraient sur cette chance prendrait un risque quasiment nul.

L'expérience suivante montre deux cas de figures pour la planche de Galton: l'une pour 6 étages où l'expérience est répétée 10 fois pour une même bille (on a alors n = 6 X 10 = 60 épreuves), et l'autre, toujours pour 6 étages, où l'expérience est répétée 20 fois (n = 6 X 20 = 120 épreuves). Les histogrammes de la **figure 4** montrent que pour n = 120 la distribution s'approche plus d'une gaussienne que pour n = 60.

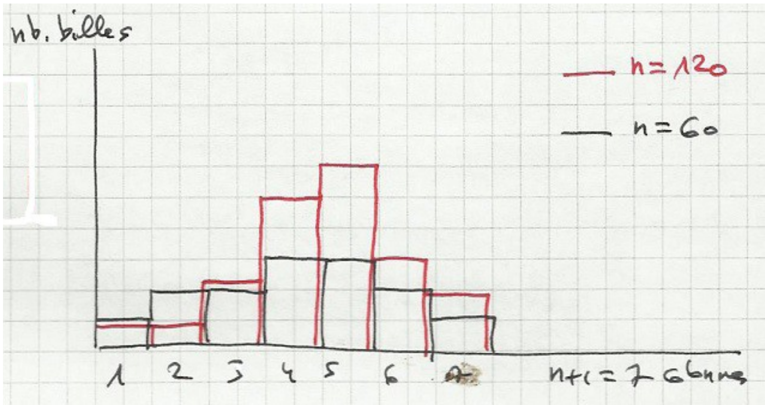

figure 4 – tendance vers une distribution gaussienne pour une planche de Galton à 6 étages

7 – Loi binomiale des événements rares: loi de Poisson

Comme toute loi de probabilité à variance finie, la loi binomiale tend vers la loi normale lorsque l'espérance est grande. Or on avu que pour une loi binomiale:

$$E(X) = \mu = np$$

où p est le paramètre de la loi binomiale, c'est-à-dire la probabilité de chaque épreuve élémentaire (on rappelle que pour la planche de Galton, épreuve de type « pile ou face », on avait p = 1/2). Donc μ est grande si n est grand et np >> 1 (donc p pas trop faible). On a vu à la fin du paragraphe 6 que μ trop petite (inférieure à 10) entraine que la loi binomiale reste encore éloignée de la loi normale.

Que se passe-t-il lorsque, tout en ayant n élevé, p est très faible, autrement dit lorsque l'épreuve élémentaire est rare?

Une telle situation se présente dans l'exemple suivant:
On procède à l'arrosage en pluie, uniformément répartie et régulière, d'une surface parfaitement plane, d'aire S, percée d'un trou de très faible ouverture, d'aire s telle que s/S << 1. La probabilité qu'une goutte entre dans le trou est indépendante de l'entrée des gouttes précédentes et est donc supposée constante et proportionnelle à la surface du trou: p = s/S. Pour l'ensemble des n gouttes qui tombent sur la plaque, la loi de probabilité est binomiale parce qu'elles sont dans deux situations possibles: soit elles entrent dans le trou, probabilité élémentaire p, soit elles n'y entrent pas, probabilité élémentaire 1-p. La probabilité pour que x gouttes parmi les n entrent dans le trou est donc:

$$P_X(x) = C_n^x p^x (1-p)^{n-x}$$

Son espérance est:

$$\mu = E(X) = n\,p = n\frac{s}{S}$$

Comme s/S << 1, on a aussi μ << 1 malgré n éventuellement grand. Dans ca cas on a l'approximation:

$$C_n^x \approx \frac{n^x}{x!}$$

$$(1-p)^{n-x} = \exp(n-x)\ln(1-p) \approx \exp(-np) = \exp(-\mu)$$

d'où:

$$P_X(x) \approx \frac{n^x}{x!} p^x e^{-\mu} = \frac{\mu^x}{x!} e^{-\mu} \quad \text{puisque} \quad n^x p^x = \mu^x$$

L'expression précédente a la forme de la ***loi de probabilité de Poisson*** de paramètre λ:

$$P_\lambda(x) = e^{-\lambda} \frac{\lambda^x}{x!} \qquad (32)$$

et l'on identifie λ = μ.
On montre que la variance d'une loi de probabilité de Poisson est aussi égale au paramètre λ:

$$\sigma^2 = \lambda = E(X) \qquad (33)$$

Le maximum de probabilité est atteint pour $x \approx \lambda = \mu$.
En physique, un exemple de processus aléatoire qui suit la loi de probabilité de Poisson est la ***désintégration radioactive*** [1]. La probabilité pour que x particules soient émises pendant la durée t lors de la désintégration d'un grand nombre n d'atomes est, en effet, égale à:

$$P(x,t) = \frac{(pt)^x}{x!} e^{-pt}$$

où la probabilité du processus élémentaire (émission d'une particule) p est proportionnelle à l'inverse de la période radioactive τ: p ~ 1/τ. Ainsi donc, μ = pt et le processus a une probabilité maximale pour x = μ = pt.

Références

[1] Walter Appel: Mathématiques pour la physique et les physiciens – H&K éditions, 2005

[2] P-A. Baratte, P. Robert: Systèmes de mesure – Dunod, 1986