

## VII. LA LOI NORMALE.

### VII.1. Un exemple simple : la marche aléatoire ou le jeu de pile ou face.

Nous avons déjà remarqué (voir VI.2., exemple b) que l'abscisse  $S$  atteinte après  $2n$  pas par une marche aléatoire était la somme de  $2n$  variables aléatoires  $X_j$ , égales à  $-1$  ou  $+1$  (avec probabilité  $\frac{1}{2}$ ) selon que le  $j^{\text{e}}$  pas est en arrière ou en avant. La loi de la variable aléatoire  $S$  est, comme nous l'avons calculée au chapitre III,

$$\mathcal{P}(S = -2n + 2k) = 2^{-2n} \binom{2n}{k}$$

Sa fonction caractéristique est  $\Phi_S(t) = [\cos(t)]^{2n}$ , ce qui est logique puisque la fonction caractéristique de chacune des  $X_j$  est  $\Phi_{X_j}(t) = \cos(t)$  et que la fonction caractéristique de la somme des  $X_j$ , qui sont indépendantes, est le produit de leurs fonctions caractéristiques.

Nous avons établi au chapitre II l'approximation suivante pour les coefficients binômiaux, valable pour  $n$  grand :

$$\binom{2n}{n+j} \simeq \binom{2n}{n} \exp\left[-\frac{j^2}{n}\right] \simeq \frac{1}{\sqrt{\pi n}} \cdot \exp\left[-\frac{j^2}{n}\right]$$

Cela peut s'interpréter en disant que la probabilité pour que  $S = 2j$  (lorsque  $j$  est un entier entre  $-n$  et  $+n$ ), est  $(1/\sqrt{\pi n}) \exp[-j^2/n]$  (approximativement), ou encore, que *pour  $n$  grand*, la loi de probabilité de la variable aléatoire  $S$  est approximativement gaussienne, avec un écart-type  $\sigma = \sqrt{n/2}$ .

Les valeurs que peut prendre la variable aléatoire  $S$  s'étendent de  $-2n$  à  $+2n$ . Mais les valeurs extrêmes  $\pm 2n$  ont une probabilité prodigieusement petite:  $2^{-2n}$ . La valeur la plus probable est  $S = 0$  qui a une probabilité à peu près égale à  $1/\sqrt{\pi n}$  (ce qui pour  $n$  grand est incomparablement plus grand que  $2^{-2n}$ ) et l'approximation montre que la probabilité décroît autour de ce maximum selon la loi gaussienne

$$\mathcal{P}(S = 2j) = \frac{1}{\sqrt{\pi n}} \exp\left[\frac{-j^2}{n}\right]$$

Comme nous l'avons déjà discuté au chapitre **VI** (exemple **b**), si  $j$  est sensiblement supérieur à l'écart-type  $\sqrt{n/2}$ , cette probabilité devient rapidement très petite, c'est le phénomène connu sous le nom de *loi des grands nombres*. Si on lance 1000 fois une pièce de monnaie (rappelons que la marche aléatoire est aussi un modèle pour les jeux de pile ou face), le calcul ci-dessus montre que le nombre de pile le plus probable est 500 (probabilité  $1/\sqrt{500\pi} \simeq 0.0252$ ; obtenir 519 a une probabilité un peu plus faible :  $1/\sqrt{500\pi} \cdot \exp(-19^2/500) \simeq 0.0122$ . Pour 530 la probabilité tombe à 0.0042, pour 550 à 0.00017, pour 575 à  $3.28 \cdot 10^{-7}$ , et pour 600 à  $5.19 \cdot 10^{-11}$ . Si on lançait la pièce 1 000 000 de fois au lieu de 1 000, on obtiendrait les probabilités correspondantes que voici :

|         |                          |
|---------|--------------------------|
| 500 000 | $7.98 \cdot 10^{-4}$     |
| 519 000 | $2.72 \cdot 10^{-314}$   |
| 530 000 | $1.81 \cdot 10^{-782}$   |
| 575 000 | $1.33 \cdot 10^{-4886}$  |
| 600 000 | $9.4 \cdot 10^{-312693}$ |

On voit que pour des écarts proportionnellement équivalents, les probabilités sont absolument infinitésimales pour un million d'essais, alors qu'elles ne l'étaient pas pour mille essais. Par contre pour des écarts de l'ordre de 500 (donc pour obtenir 500 500 pile sur 1 000 000 de lancers), la probabilité est  $6.21 \cdot 10^{-4}$ , soit seulement 1.28 fois plus petite que la probabilité maximum. On peut conclure de cette petite observation numérique que si on lance une pièce de monnaie mille fois, la probabilité d'obtenir plus de 600 pile et moins de 400 face ou vice-versa (c'est-à-dire d'obtenir des fluctuations supérieures à 20%) sera extrêmement faible, en tous cas inférieure à  $10^{-9}$ . Par contre la probabilité d'obtenir entre 400 et 600 pile sera très proche de 1. Si on lance la pièce 1 000 000 de fois, la probabilité d'obtenir des fluctuations supérieures à 20% sera inférieure à  $10^{-312000}$ , ce qui signifie pratiquement que de telles fluctuations sont *absolument* impossibles. Pour 1 000 000 de lancers, la probabilité d'obtenir des fluctuations de plus de 1% par rapport à 500 000 (c'est-à-dire un nombre de pile ou de face inférieur à 495 000 ou supérieur à 505 000) est de l'ordre de  $10^{-10}$ .

Ce qui détermine le seuil à partir duquel une probabilité qui suit une loi gaussienne devient petite est l'écart-type (ici  $\sqrt{n/2}$ ); cette probabilité devient absolument infinitésimale lorsqu'on s'écarte de la valeur la plus probable de dix fois l'écart-type. Or l'écart-type de la marche aléatoire (ou du jeu de pile ou face) étant de l'ordre de la racine carrée de  $n$ , est, proportionnellement à  $n$  lui-même, d'autant plus petit que  $n$  est plus grand : si on lance  $2n$  fois une pièce de monnaie, la fréquence relative des pile et des face tend vers  $1/2$  quand  $n$  tend vers l'infini, et les fluctuations *en*

valeur relative sont de plus en plus improbables ; si on appelle  $x$  l'amplitude relative ( $x = j/n$ ) de ces fluctuations, la probabilité des fluctuations décroît en  $e^{-nx^2}$ , qui devient pratiquement nul ( $e^{-10} \simeq 4.5 \cdot 10^{-5}$ ) lorsque  $nx^2 > 10$ , et absolument infinitésimal ( $e^{-30} \sim 10^{-13}$ ) lorsque  $nx^2 > 30$ . On peut donc dire que sur  $2n$  lancers de pièce, il est extrêmement improbable et donc pratiquement impossible d'obtenir des écarts supérieurs à  $\sqrt{30/n}$  en valeur relative par rapport à la valeur moyenne.

Il a déjà été dit au chapitre précédent que cette loi des grands nombres transformait – à l'inverse du chaos – le hasard en déterminisme : il suffit pour cela que la précision avec laquelle le résultat est perçu ou mesuré ne soit pas parfaite, que le résultat ne puisse être connu qu'à quelques écart-types près. Les valeurs “les plus probables” deviennent alors certaines.

Il se trouve que cette propriété n'est pas seulement vraie dans le cas de la marche aléatoire ou du jeu de pile ou face, mais qu'elle est vraie pour *n'importe quel* phénomène aléatoire qui se répète un grand nombre de fois, pourvu qu'il y ait — comme c'était le cas pour la pièce de monnaie — indépendance stochastique entre les répétitions. C'est ce que nous allons prouver dans les deux paragraphes suivants.

## VII.2. Une propriété des fonctions caractéristiques.

Soit  $X$  une variable aléatoire de loi  $\{x_k (p_k)\}$ . Sa fonction caractéristique est (voir VI.2)  $\Phi(t) = \sum_k p_k e^{itx_k}$ . Or on dispose pour la fonction  $e^{i\alpha\tau}$  du développement limité suivant :

$$e^{i\alpha} = 1 + i\alpha - \frac{1}{2}\alpha^2 - i\alpha^3 \int_0^1 \frac{(1-\tau)^2}{2} e^{i\alpha\tau} d\tau$$

Si on reporte ce développement dans l'expression de la fonction caractéristique on obtient

$$\begin{aligned} \Phi(t) &= \sum_k p_k e^{itx_k} \\ &= \sum_k p_k \left[ 1 + itx_k - \frac{t^2}{2}x_k^2 - it^3x_k^3 \int_0^1 \frac{(1-\tau)^2}{2} e^{itx_k\tau} d\tau \right] \\ &= \sum p_k + it \sum p_k x_k - \frac{t^2}{2} \sum p_k x_k^2 - it^3 \sum p_k x_k^3 \int_0^1 \frac{(1-\tau)^2}{2} e^{itx_k\tau} d\tau \end{aligned}$$

Cette expression est un développement limité de la fonction caractéristique (pour  $t$  petit). On constate que le coefficient de  $t$  dans ce développement est  $i \sum p_k x_k$ , c'est-à-dire  $i\mathbf{E}(X)$  ; le coefficient de  $t^2$  est  $-\frac{1}{2} \sum p_k x_k^2$ , c'est-à-dire  $-\frac{1}{2}\mathbf{E}(X^2)$ . Si  $\mathbf{E}(X) = 0$ ,  $\mathbf{E}(X^2) = \mathbf{Var}(X)$  : dans ce cas le développement

limité de  $\Phi(t)$  commence par  $1 - \frac{t^2}{2}\mathbf{Var}(X)$ . Cela est vrai pour la fonction caractéristique de *n'importe quelle variable aléatoire*, pourvu que sa moyenne et la moyenne de son carré ne soient pas infinies.

Si on souhaite pousser plus loin le développement limité, aucune difficulté particulière ne s'y oppose. Pour écrire un tel développement, il est commode d'introduire les moments :

$$\begin{aligned} M_1 &= \mathbf{E}(X) & M_2 &= \sum_k p_k x_k^2 \\ M_3 &= \sum_k p_k x_k^3 & M_4 &= \sum_k p_k x_k^4 \\ & \dots \end{aligned}$$

Alors le développement limité de la fonction caractéristique peut s'écrire

$$\Phi(t) = 1 + itM_1 - \frac{t^2}{2}M_2 - i\frac{t^3}{6}M_3 + \frac{t^4}{24}M_4 \quad \dots \quad (VII.1.)$$

On peut observer cela sur les cas particuliers que nous avons déjà étudiés. Par exemple si  $X$  est une variable aléatoire qui prend les valeurs  $+1$  et  $-1$  avec probabilités  $\frac{1}{2}$ , la fonction caractéristique est  $\cos(t)$ ;  $\mathbf{E}(X) = 0$ ,  $\mathbf{Var}(X) = 1$ , et on a bien  $\cos(t) \simeq 1 - \frac{t^2}{2}$ . On pourrait constater qu'il en est de même pour toutes les fonctions caractéristiques de variables aléatoires dont la moyenne et la variance sont finies. Si  $X$  est par exemple le premier retour à zéro d'une variable aléatoire, nous avons vu que (à la limite où le nombre  $n$  de pas tend vers l'infini) la moyenne est infinie; la fonction caractéristique, qui dans ce cas était  $1 - \sqrt{1 - e^{2it}}$  ne possède pas de développement limité à l'ordre 2: en fait elle n'est pas dérivable au point  $t = 0$ .

On peut donc dire que toutes les fonctions caractéristiques sont de la forme  $1 + it\mathbf{E}(X) - \frac{t^2}{2}\mathbf{E}(X^2)$  au voisinage de  $t = 0$  (pourvu que  $\mathbf{E}(X)$  et  $\mathbf{E}(X^2)$  soient finies). Une fonction qui n'aurait pas un développement limité de cette forme au voisinage de zéro (qui serait par exemple de la forme  $1 + \frac{t^2}{2}$ ) ne pourrait être la fonction caractéristique d'aucune variable aléatoire<sup>(1)</sup>

---

(1) à moins de considérer des variables aléatoires à valeurs complexes: ainsi une v.a. qui prendrait les valeurs  $\pm i$  avec probabilité  $\frac{1}{2}$  chacune, aurait une fonction caractéristique égale à  $\cosh(t) \simeq 1 + \frac{t^2}{2}$ . Nous ne considérons que des v.a. à valeurs réelles, et s'il y a un jour des valeurs complexes à prendre en compte, on décomposera en partie réelle et partie imaginaire. De toute façon, même en admettant des v.a. complexes, leurs fonctions caractéristiques obéissent à des contraintes strictes (la transformée de Fourier d'une loi de probabilité ne peut pas être n'importe quoi); il existe des ouvrages entiers rien que sur ce sujet: voir par exemple Eugène Lukacs *fonctions caractéristiques* Dunod, Paris, 1964.

En dehors du développement limité au voisinage de zéro, les fonctions caractéristiques ont encore d'autres propriétés caractéristiques; par exemple elles sont toujours bornées par 1 : si  $X$  est une variable aléatoire absolument quelconque, on aura toujours  $|\Phi_X(t)| \leq 1$ . Il est très facile de s'en assurer, il suffit de remarquer que

$$|\Phi_X(t)| = \left| \sum_k p_k e^{itx_k} \right| \leq \sum_k |p_k e^{itx_k}| = \sum_k p_k = 1$$

Cette propriété est donc simplement due au fait que les probabilités  $p_k$  sont des nombres positifs dont la somme vaut 1 (notons qu'on a aussi utilisé le fait que  $|e^{itx_k}| = 1$ , ce qui serait faux si les valeurs  $x_k$  pouvaient être complexes).

Si les valeurs  $x_k$  prises par la variable aléatoire  $X$  sont entières, la fonction caractéristique est périodique (de période  $2\pi$ ). Plus généralement, si les  $x_k$  sont tous des multiples entiers d'un même nombre réel  $a$  (c'est-à-dire qu'ils ne sont pas mutuellement incommensurables), la fonction caractéristique est périodique, de période  $2\pi/a$ . Si par contre les  $x_k$  sont mutuellement incommensurables, la fonction caractéristique n'est pas périodique.

### **VII.3. La somme d'un grand nombre de variables aléatoires indépendantes.**

Nous avons vu au paragraphe précédent que la fonction caractéristique d'une variable aléatoire  $X$  de moyenne nulle est de la forme  $1 - \frac{t^2}{2} \text{Var}(X)$  au voisinage de  $t = 0$ . Maintenant nous allons voir que cette propriété des fonctions caractéristiques a une conséquence capitale : la somme d'un grand nombre de variables aléatoires indépendantes suit une loi approximativement gaussienne, et cela *quelle que soit la loi des variables individuelles*.

En étudiant l'exemple du jeu de pile ou face, nous avons vu que les fluctuations pour  $n$  lancers avaient un écart-type  $\sigma$  de l'ordre de  $\sqrt{n}$ . Cela veut dire que des fluctuations de plusieurs dizaines de fois  $\sigma$  sont pratiquement impossibles, et que donc les fluctuations autour de la valeur moyenne (qui est la plus probable) sont pratiquement toutes de l'ordre de  $\sqrt{n}$ . Ainsi l'amplitude de ces fluctuations tend vers l'infini en même temps que  $n$ , mais plus lentement; si par contre on ne les compte qu'en valeur relative (en les divisant donc par  $n$ ), leur amplitude tend vers zéro. On obtiendrait une valeur-limite de l'écart-type lorsque  $n$  tend vers l'infini en divisant les fluctuations par  $\sqrt{n}$ . C'est pourquoi, en étudiant le cas général, nous ne mesurerons les fluctuations ni en valeur exacte, ni en valeur relative, mais à l'échelle intermédiaire  $\sqrt{n}$ .

Voici donc ce que nous nous proposons d'établir :

## La loi normale

*Théorème*: soient des variables aléatoires  $X_1, X_2, X_3, \dots, X_n$  de moyenne nulle. On considère les fluctuations de la somme (rapportée comme annoncé à l'unité  $\sqrt{n}$ ):

$$S = \frac{X_1 + X_2 + X_3 \cdots + X_n}{\sqrt{n}}$$

Alors la variable aléatoire  $S$  a, pour  $n$  grand, une loi approximativement gaussienne.

Ce théorème mérite deux commentaires :

a) avoir supposé que les variables aléatoires  $X_1, X_2, X_3, \dots, X_n$  sont de moyenne nulle n'est pas une restriction. Nous ne nous intéressons en effet qu'aux *fluctuations* de leur somme autour de la moyenne; si les  $X_j$  ont des moyennes  $\mathbf{E}(X_j) = m_j$  non nulles, leur somme  $T$  a pour moyenne la somme des moyennes, soit  $\mathbf{E}(T) = m = \sum m_j$ . Les fluctuations de  $T$  autour de cette moyenne peuvent s'écrire  $T - m = \sum (X_j - m_j)$  et les fluctuations rapportées à l'unité  $\sqrt{n}$  seront

$$S = \frac{T - m}{\sqrt{n}} = \frac{\sum (X_j - m_j)}{\sqrt{n}}$$

de sorte que tout se ramène à des variables aléatoires de moyenne nulle, à savoir les  $X_j - m_j$ .

b) Le sens de l'expression "une loi approximativement gaussienne" doit être précisé. Dans le cas de la marche aléatoire (voir §VII.1.) ce sens était clair; la loi exacte était donnée par les coefficients du binôme et ceux-ci pouvaient être approchés par une expression de la forme  $e^{-x^2}$ . Mais on savait alors que les valeurs prises par la variable aléatoire étaient entières. Dans le cas général il n'y a plus aucune raison pour que les valeurs soient entières, ni même équidistantes, mais celles-ci sont cependant toujours discrètes; une approximation de cette loi discrète ne peut pas se réduire à une approximation des probabilités de chaque valeur: il faut en outre avoir une approximation des valeurs elles-mêmes. Par ailleurs la fonction  $e^{-x^2}$  est définie sur les nombres réels, c'est-à-dire qu'elle dépend d'une variable qui varie de façon continue; en particulier, dans l'hypothèse où les valeurs prises par la variable aléatoire ne sont pas équidistantes, toute information sur la distance entre les valeurs discrètes a disparu. Dans quel sens peut-on alors dire qu'une loi continue approche une loi discrète?

Cela devra être entendu dans le sens suivant: divisons l'intervalle compris entre la plus petite et la plus grande des valeurs en  $N$  parties égales de longueur  $\varepsilon$ . Il faut que  $\varepsilon$  soit petit, mais pas trop: il doit être assez petit pour que la fonction continue (en l'occurrence  $e^{-x^2}$ ) varie peu sur la distance

$\varepsilon$ , mais il faut aussi qu'il soit grand par rapport à la distance moyenne entre les valeurs prises par la variable aléatoire, afin que chacun des intervalles de longueur  $\varepsilon$  contienne un échantillon statistiquement significatif de valeurs discrètes. Une telle opération s'appelle un échantillonnage. On dira alors qu'une densité continue  $y = f(x)$  approche une loi discrète  $\{x_j, (p_j)\}$  si dans chacun des intervalles d'échantillonnage  $J$  de longueur  $\varepsilon$ , on a

$$\sum_{x_j \in J} p_j \simeq \varepsilon f(x_j)$$

Il serait équivalent de dire ceci : pour tout intervalle  $[a, b[$  dont la longueur est suffisante pour contenir un échantillon statistiquement significatif de valeurs discrètes, on a

$$\sum_{a \leq x_j < b} p_j \simeq \int_a^b f(x) dx$$

Dans l'exemple de la marche aléatoire, la variable aléatoire dont on voulait approcher la loi par une densité continue est le déplacement rapporté à  $\sqrt{n}$ , qui prend des valeurs allant de  $-\sqrt{n}$  à  $+\sqrt{n}$ , avec un pas égal à  $2/\sqrt{n}$ . Dans ce cas le nombre  $\varepsilon$  évoqué ci-dessus, qui représente — en ordre de grandeur — la taille optimale des intervalles d'échantillonnage, doit être sensiblement plus grand que le pas  $2/\sqrt{n}$ , tout en étant petit à l'échelle usuelle, macroscopique (celle où le pas de la marche vaut 1); une valeur de  $\varepsilon$  telle que par exemple  $n^{-1/4}$  serait correcte.

On pourrait penser à la première lecture que toutes ces considérations sur les intervalles qui doivent être “petits, mais pas trop”, qui doivent “contenir un échantillon statistiquement significatif”, etc. sont des simplifications grossières de physicien ou d'ingénieur, mais que les *véritables* mathématiques permettent de dépasser ces vulgarités et offrir des énoncés et des démonstrations “propres” et “rigoureux”, où  $\varepsilon$  pourrait être rendu aussi petit que l'on veut.

Or il n'en est rien. Le fait que la longueur des intervalles d'échantillonnage doive être sensiblement moins petite que  $1/\sqrt{n}$  est une nécessité objective qu'on ne peut pas faire disparaître en introduisant plus de rigueur. Certes, on trouvera des ouvrages plus “mathématiques” que celui-ci sur le Calcul des probabilités, dans lesquels tout cela sera traité soigneusement en termes de limites quand  $n$  tend vers l'infini. Mais si on regarde les démonstrations de plus près, on s'apercevra que cette histoire d'échantillonnage y figure simplement sous une forme déguisée, et si on parvient à y reconnaître le paramètre qui joue le rôle de  $\varepsilon$ , on s'apercevra qu'il tend certes vers zéro, mais de telle sorte que  $\varepsilon \times \sqrt{n}$  tende vers l'infini (ce qui est évidemment équivalent à dire

que  $\varepsilon$  doit être grand par rapport à  $1/\sqrt{n}$ . Le défaut de ce genre d'ouvrage consiste souvent à noyer la propriété importante, qui en l'occurrence est précisément cette affaire d'ordre de grandeur de l'échantillonnage, dans les détails techniques d'une démonstration sophistiquée.

Afin de rendre les choses plus concrètes, le mieux est encore de voir deux exemples extrêmes.

Soit  $f(x)$  une fonction continue de la variable réelle  $x$  sur un intervalle  $[a, b]$ , qui peut représenter une densité de probabilité (elle est donc positive et  $\int f(x)dx = 1$ ). Prenons deux variables aléatoires  $X$  et  $Y$  à valeurs dans  $[a, b]$ , dont l'une ( $X$ ) prend des valeurs équidistantes ( $x_j = j\varepsilon$ ) avec une probabilité variable  $p_j$ , et l'autre ( $Y$ ) des valeurs non équidistantes  $y_j$  mais équiprobables (pour tout  $j$  on a  $\mathcal{P}(Y = y_j) = p$ ). On suppose que pour  $X$  on a  $p_j = f(x_j) = f(j\varepsilon)$  et que pour  $Y$  la distance entre deux valeurs consécutives est  $y_{j+1} - y_j = p/f(y_j)$ . On a représenté graphiquement les deux lois de  $X$  et de  $Y$  sur la figure 11, dans le cas où  $f(x) = e^{-x^2}/\sqrt{\pi}$ .

Si on veut approcher la loi de probabilité de  $X$  par une densité continue, il suffira — comme on l'a fait au §1 pour la marche aléatoire — d'approcher la valeur des  $p_j$ . Mais si on veut approcher la loi de  $Y$ , il faudra tenir compte du fait que, les valeurs de  $Y$  étant plus nombreuses (plus denses) au voisinage de zéro, il y aura une plus forte probabilité pour  $Y$  de prendre une valeur proche de zéro, bien que chaque valeur discrète prise isolément soit équiprobable. L'opération d'échantillonnage évoquée ci-dessus consiste à regrouper les valeurs prises par les variables aléatoires dans des intervalles de longueur égale  $\varepsilon$ ; en choisissant  $\varepsilon$  égal au dixième de la largeur totale des graphiques de la figure 11, on obtient dix intervalles d'échantillonnage, et la probabilité pour que  $X$  ou  $Y$  prenne ses valeurs dans l'un de ces dix intervalles est représentée sur la figure 12. On voit que les deux graphiques sont déjà bien plus ressemblants.

Une autre manière, équivalente à l'échantillonnage, de passer d'une loi discrète à une densité continue est la *convolution*; cette opération consiste à moyenniser les lois de probabilité ou, comme on dit en traitement du signal, à appliquer un filtre passe-bas. Dans ce cas, le paramètre  $\varepsilon$  sera la *longueur de corrélation* de ce filtre passe-bas. Mathématiquement, la convolution est l'opération suivante : étant donnée une loi discrète  $\{x_k(p_k)\}$  et une fonction  $y = \rho(x)$  de la variable réelle  $x$ , on appelle *convolution* de la loi discrète par la fonction  $\rho$  la fonction :

$$g(x) = \sum_k p_k \rho(x - x_k)$$

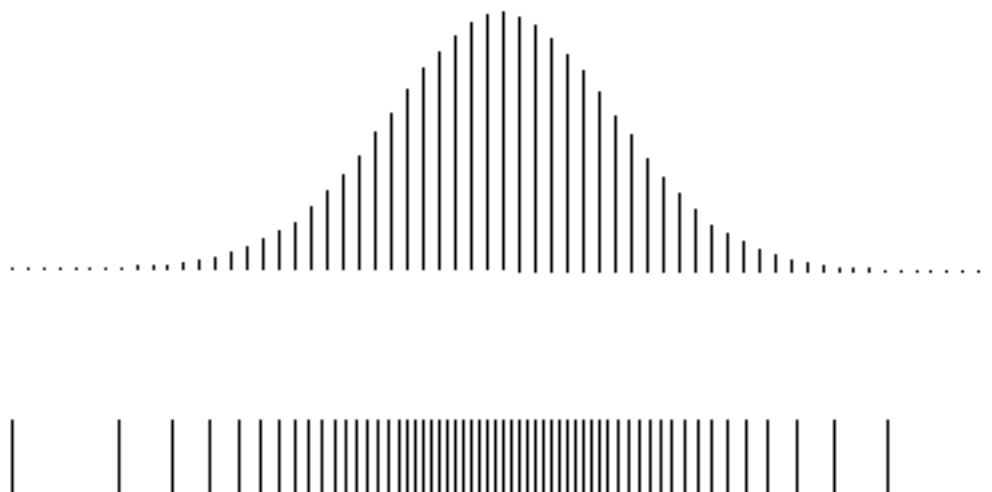


figure 11

On a représenté sur cette figure les graphiques de deux lois de probabilité: la première (graphique du haut) est la loi d'une variable aléatoire qui prend des valeurs équidistantes avec une probabilité variable (en  $e^{-x^2}$ ); la seconde (graphique du bas) est la loi d'une variable aléatoire qui prend des valeurs équiprobables mais non équidistantes; toutefois dans la seconde loi les valeurs ont une densité plus grande au centre qu'au bord, et cette densité (mesurée en nombre de traits par mm.) est exactement proportionnelle à  $e^{-x^2}$ . De façon plus précise, on peut voir que les deux variables aléatoires prennent 59 valeurs chacune; la deuxième prend chaque valeur avec probabilité  $1/59$ ; les deux graphiques sont à la même échelle, c'est-à-dire que la somme des longueurs (variables) des traits du graphique du haut est égale à la somme des longueurs (constantes) des traits du graphique du bas.

Ces deux lois sont représentées approximativement par la *même* densité continue  $e^{-x^2}$ .

La fonction  $\rho$  est le *filtre*; si par exemple

$$\rho(x) = \begin{cases} 0 & \text{si } |x| > \frac{1}{2}\varepsilon \\ \frac{1}{\varepsilon} & \text{si } |x| \leq \frac{1}{2}\varepsilon \end{cases} \quad (VII.2a)$$

alors la convolution est simplement la moyennisation sur des intervalles de largeur  $\varepsilon$ . On peut également prendre des filtres gaussiens d'écart-type  $\varepsilon$ ; dans ce cas on aura

$$\rho(x) = \frac{1}{\sqrt{2\pi} \varepsilon} e^{-\frac{x^2}{2\varepsilon^2}} \quad (VII.2b)$$

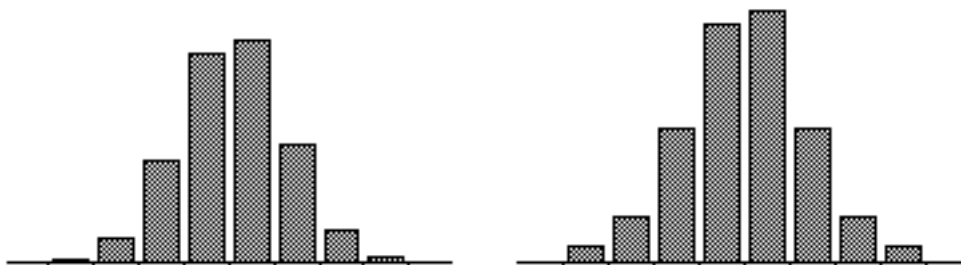


figure 12

On retrouve ici les deux graphiques de la figure 11, *après* échantillonnage (celui de gauche ci-dessus correspond à celui du haut dans la figure 11, et celui de droite à celui du bas).

Un autre filtre fréquemment considéré est le filtre

$$\rho(x) = \frac{\sin(\frac{1}{\varepsilon} \pi x)}{\pi x} \quad (VII.2c)$$

dont l'importance est due principalement au fait que sa transformée de Fourier est la fonction qui vaut 1 dans l'intervalle  $[-\frac{1}{\varepsilon}, +\frac{1}{\varepsilon}]$  et zéro en dehors.

On peut voir le résultat de ces convolutions sur les figures 13 à 16.

On peut remarquer sur ces figures que le lissage de la loi de probabilité discrète par convolution donne bien une courbe approximativement gaussienne, mais seulement lorsque la longueur de corrélation  $\varepsilon$  du filtre se situe dans une certaine plage de valeurs. Si  $\varepsilon$  est trop petit, le *bruit discret* subsiste et la courbe lissée n'est pas proche de la gaussienne. Si  $\varepsilon$  est trop grand, la moyennisation ne porte pas seulement sur les hautes fréquences caractéristiques de la loi discrète, mais aussi sur l'écart-type de la loi elle-même, ce qui aboutit à étaler la loi.

Ainsi, lorsqu'on énonce une phrase telle que “la loi discrète  $X$  est approximativement gaussienne”, il faut comprendre que pour un choix optimal du paramètre  $\varepsilon$  (celui-ci n'étant ni trop grand, ni trop petit), la loi  $X$  *filtrée* (qui, elle, est une densité continue) est proche de la densité continue gaussienne.

Lorsqu'on dit qu'une loi discrète  $X$  est approximativement gaussienne, son graphique (par exemple l'un des deux graphiques de la figure 11) est très différent du graphique de la fonction

$$\frac{1}{\sqrt{2\pi \mathbf{Var}(X)}} \exp\left\{-\frac{x^2}{2 \mathbf{Var}(X)}\right\}$$

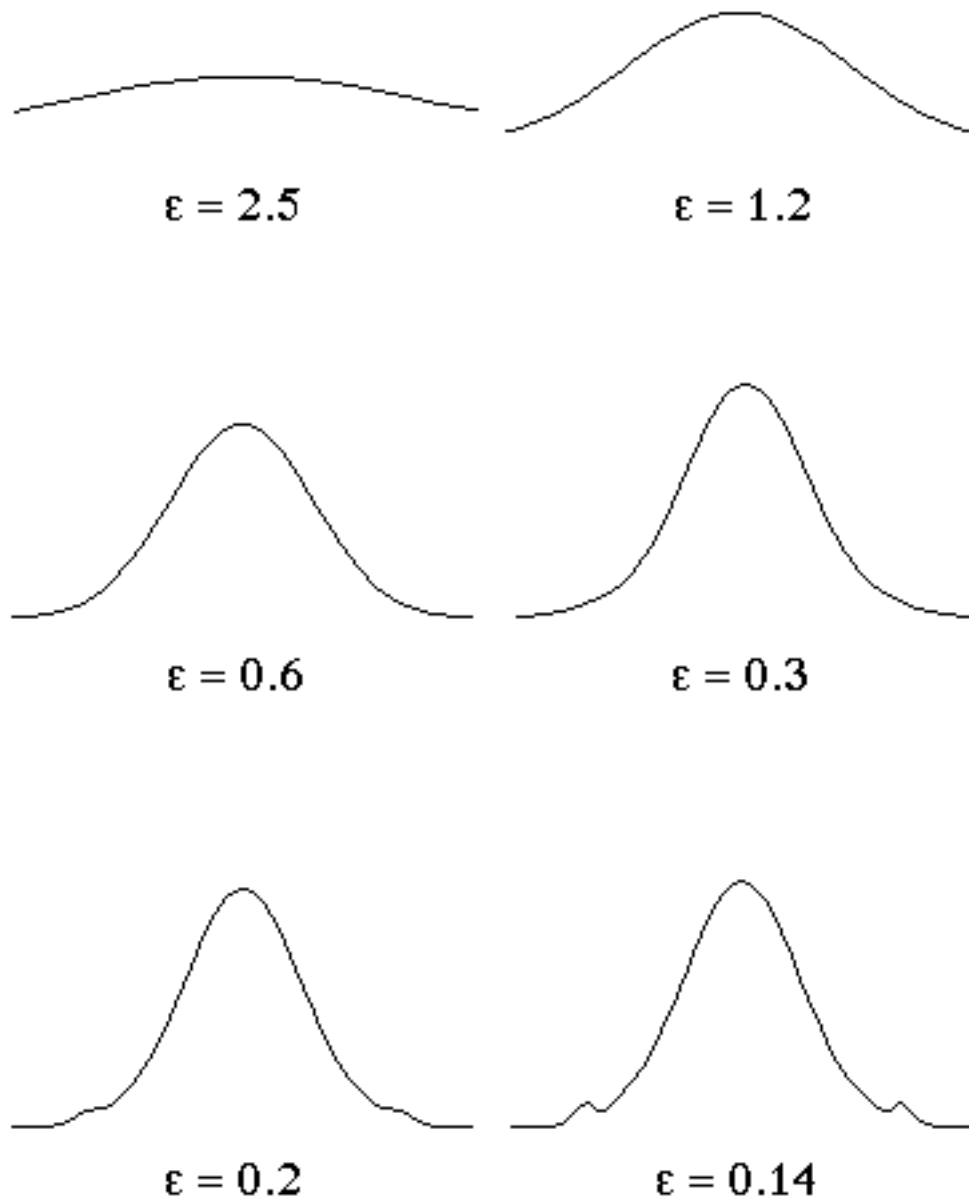


figure 13 (début)

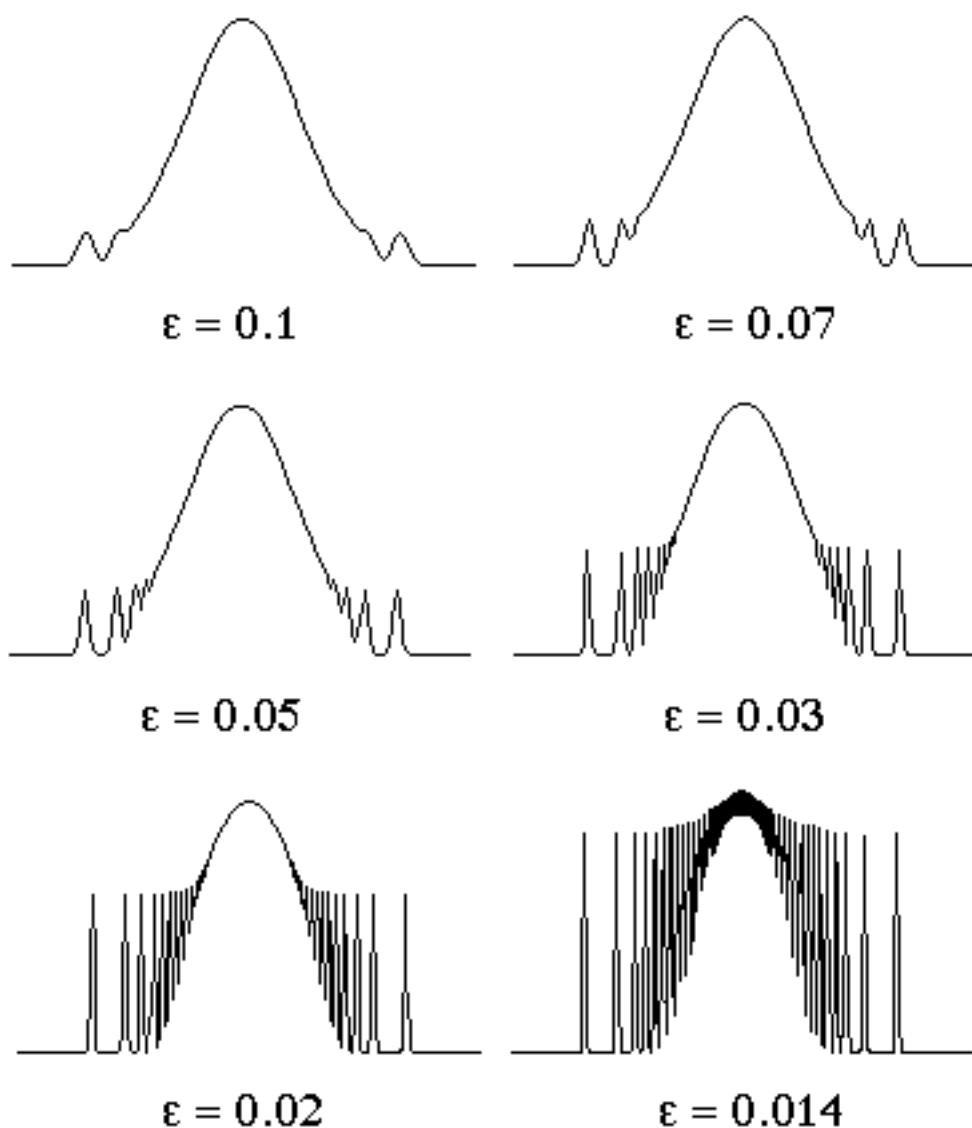


figure 13 (suite)

On peut voir sur ces huit graphiques le résultat de convolutions par un filtre gaussien d'écart-type  $\varepsilon$  variant de 1.2 à 0.014 sur la loi correspondant au graphique du bas dans figure 11. On constate que le filtrage donne une courbe ressemblant à la loi normale en  $e^{-x^2}$  lorsque  $\varepsilon$  est de l'ordre de 0.3 à 0.6. Pour  $\varepsilon$  plus petit, le "bruit" discret n'est pas éliminé, et pour  $\varepsilon$  plus grand, la courbe est trop aplatie (le filtrage élimine alors non seulement les fréquences du bruit discret, mais même celle de la courbe normale).

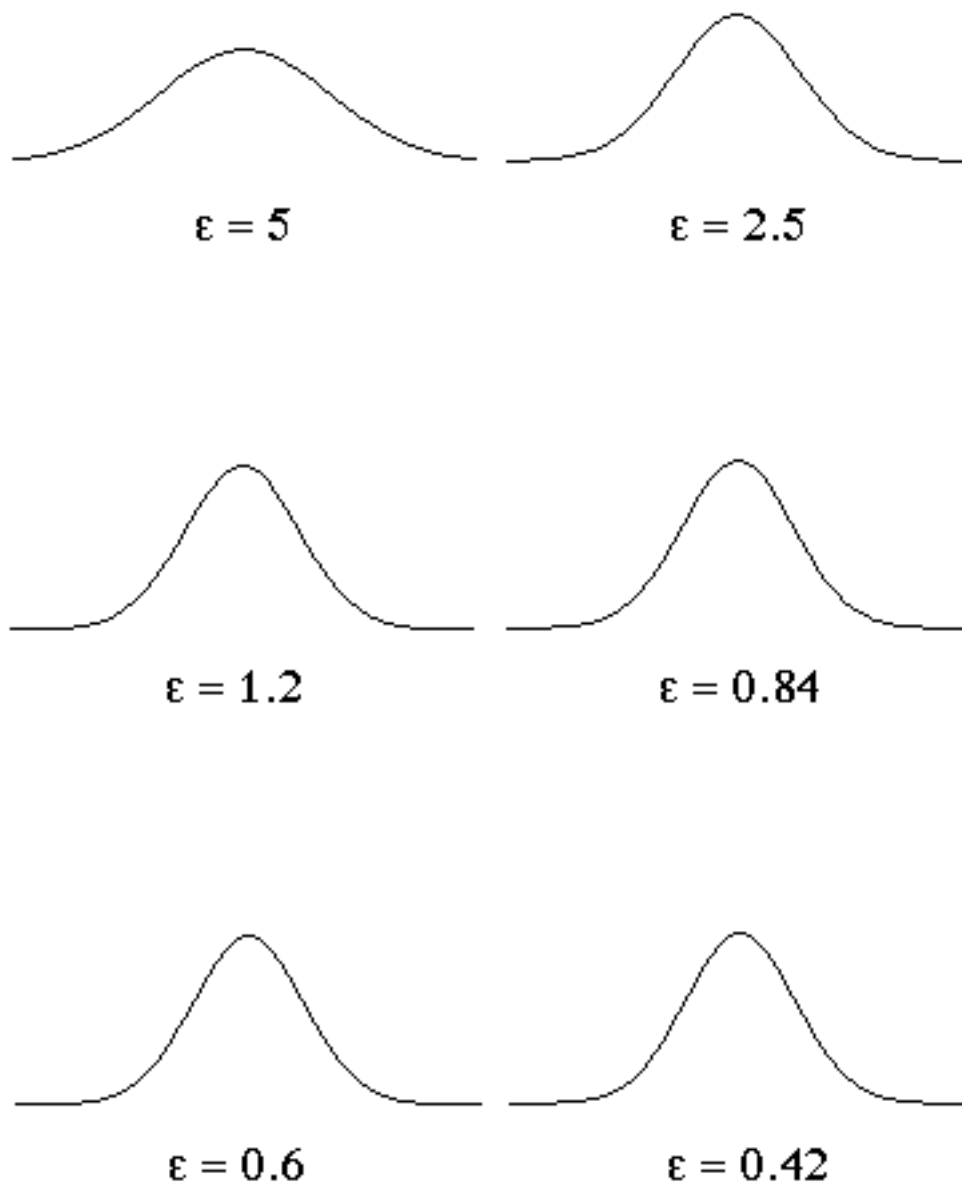


figure 14 (début)

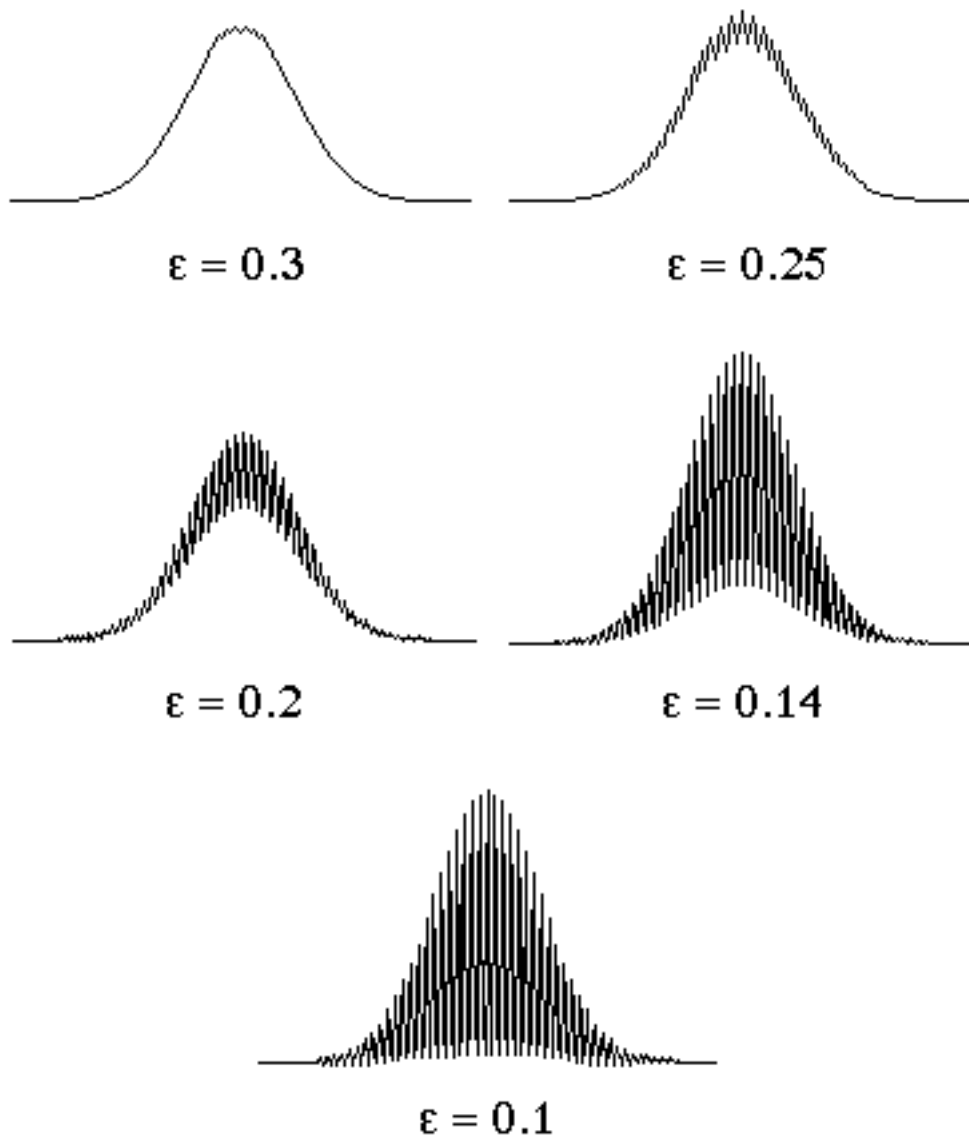


figure 14 (suite)

On a fait la même chose que dans la figure 13 mais pour l'autre loi de probabilité (graphique du haut dans la figure 11). On constate que là aussi, les valeurs de  $\varepsilon$  pour lesquelles le lissage transforme la loi discrète en loi gaussienne sont de l'ordre de  $0.3 - 0.6$ .

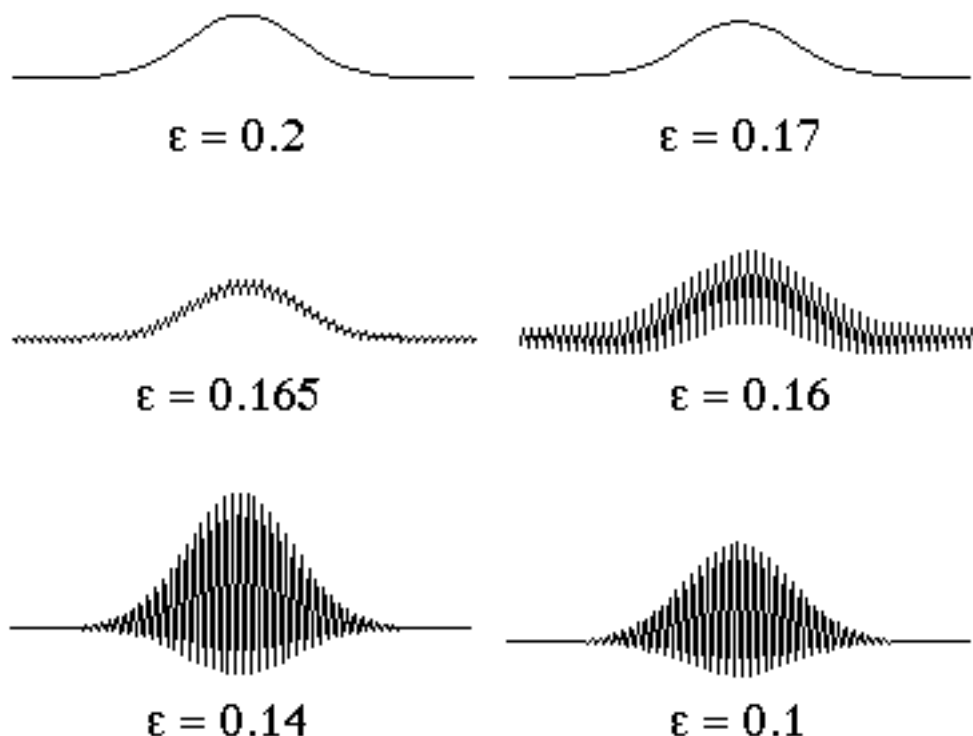


figure 15

Ici on a fait la même chose que dans la figure 14, mais à la place du filtre gaussien on a utilisé pour le lissage le filtre  $\sin(\pi x/\varepsilon)/\pi x$  ; il s'agit donc de la loi de probabilité représentée par le graphique du *haut* dans la figure 11. On retrouve qualitativement les mêmes phénomènes que ceux observés avec le filtre gaussien. On remarquera toutefois que la disparition du bruit discret est beaucoup plus brusque qu'avec le filtre gaussien : elle se produit entre  $\varepsilon = 0.16$ , où le bruit discret est encore quasiment maximum, et  $\varepsilon = 0.17$ , où il a pratiquement disparu. On voit sur la deuxième partie de la figure 14 que dans le cas du filtre gaussien, cette disparition est plus progressive et s'étale de  $\varepsilon = 0.1$  à  $\varepsilon = 0.3$ . Chaque filtre présente des caractéristiques particulières, mais tous ont en commun que la courbe lissée ne correspond à la loi gaussienne limite que lorsque  $\varepsilon$  est de l'ordre de 0.3 à 0.6 (ces chiffres précis n'étant bien sûr valables que pour le cas particulier de l'exemple).

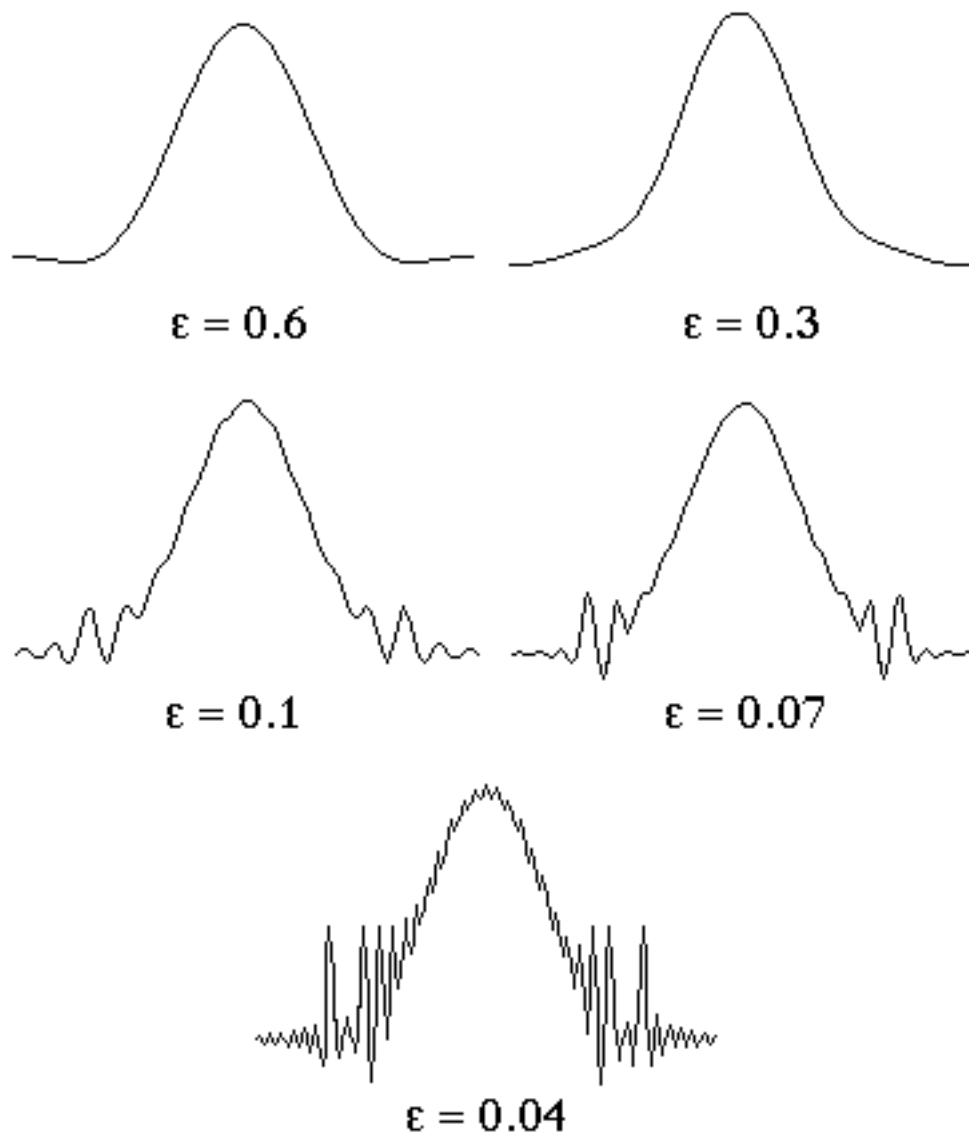


figure 16

Ici on a fait la même chose que dans la figure 13, en remplaçant le filtre gaussien par le filtre  $\sin(1.66x/\epsilon)/x$ ; ou encore: la même chose que dans la figure 15, mais pour la loi de probabilité représentée par le graphique du *bas* dans la figure 11.

mais il lui devient ressemblant si on effectue un échantillonnage (comme le montre la figure 12) ou un filtrage par convolution (comme le montrent les figures 13 à 16).

Le théorème qui fait l'objet de ce chapitre affirme que la somme d'un grand nombre de variables aléatoires *indépendantes* a une loi approximativement gaussienne, dans le sens qui vient d'être précisé. En fait, la loi d'une telle somme peut être simple (par exemple si chacune des variables aléatoires a pour loi  $\{-1 \ (\frac{1}{2}), +1 \ (\frac{1}{2})\}$ : dans ce cas la loi de la somme correspond au graphique du haut dans la figure 11); mais ce peut être aussi une loi extrêmement compliquée et chaotique (si par exemple les variables aléatoires prennent des valeurs nombreuses et incommensurables entre elles, avec des probabilités respectives également chaotiques: voir la figure 17). Le théorème dit alors que même si la loi discrète *exacte* de la somme est extrêmement compliquée et chaotique, sa *densité*, qu'on peut révéler par moyennisation, sera néanmoins proche d'une densité gaussienne; la partie compliquée de la loi se situe uniquement dans le bruit discret, et après élimination de ce bruit par échantillonnage ou filtrage, les probabilités se répartiront selon une loi gaussienne simple: il se produira pour la loi compliquée et chaotique la même chose que ce qui s'est produit dans les figures 13 et 16.

La démonstration du théorème fait apparaître le mécanisme de ce miracle. Si on effectue la somme de  $n$  variables aléatoires  $X$  stochastiquement indépendantes, toutes de même loi, alors la fonction caractéristique de leur somme  $S = \sum X$  est le produit des fonctions caractéristiques, c'est-à-dire

$$\Phi_S(t) = \Phi_X(t)^n$$

Si on pose  $Z = S/\sqrt{n}$ , alors

$$\Phi_Z(t) = \Phi_S\left(\frac{t}{\sqrt{n}}\right) = \Phi_X\left(\frac{t}{\sqrt{n}}\right)^n$$

En effet,

$$\Phi_Z(t) = \sum_k p_k \exp\left(i \frac{x_k}{\sqrt{n}} t\right) = \sum_k p_k \exp\left(i x_k \frac{t}{\sqrt{n}}\right) = \Phi_S\left(\frac{t}{\sqrt{n}}\right)$$

Il résulte de la section **VII.2.** que  $\Phi_X(t)$ , comme toute fonction caractéristique de variable aléatoire, possède *au voisinage de  $t = 0$*  le développement limité

$$1 - \text{Var}(X) \frac{t^2}{2}$$

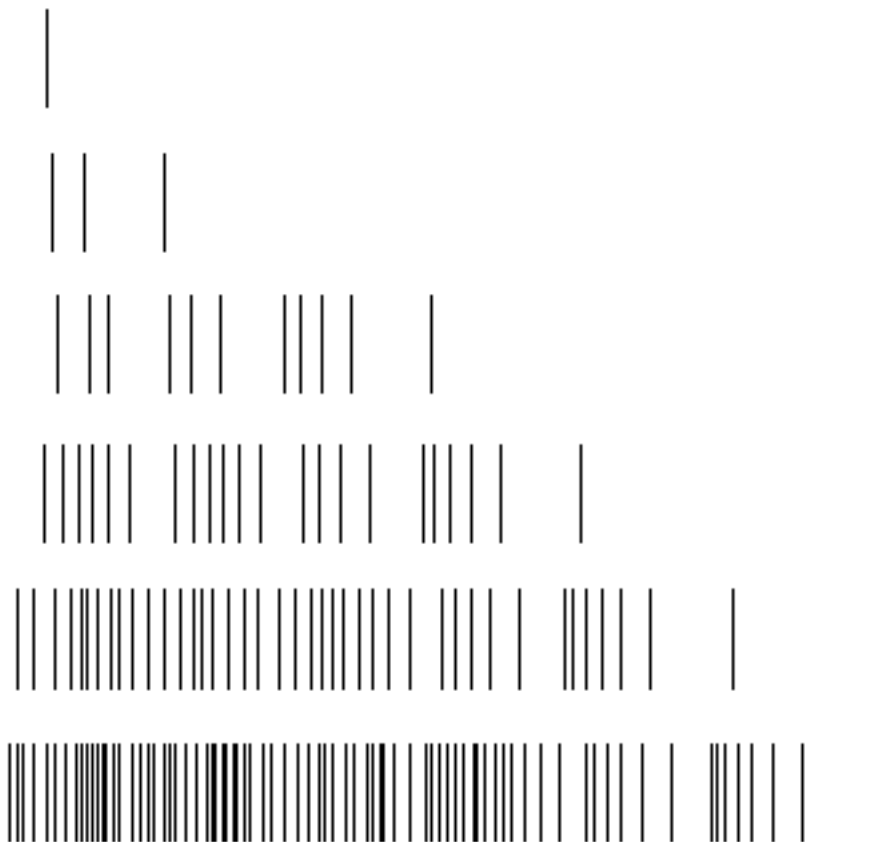


figure 17

Les six graphiques que voici représentent la loi de la somme d'un nombre de plus en plus grand de variables aléatoires. Afin d'économiser la place, on n'a reproduit que la moitié positive du graphique, la moitié négative étant symétrique; ainsi la valeur 0 est à l'extrême gauche du graphique. Les variables aléatoires sont, pour  $j = 1$  à  $j = 6$ ,

$$X_j = \begin{cases} \sqrt{j/3} + \sqrt{2-j/5} + \sqrt{j/7} & \text{avec probabilité } \frac{1}{2} \\ -\sqrt{j/3} - \sqrt{2-j/5} - \sqrt{j/7} & \text{avec probabilité } \frac{1}{2} \end{cases}$$

Ces valeurs bizarres ont été choisies parce qu'elles sont incommensurables, de sorte qu'on ne peut avoir deux fois la même somme avec des termes différents; ainsi toutes les valeurs prises par la somme sont équiprobables et on obtient donc une loi du type de celle du bas dans la figure 12.

Le premier graphique (tout en haut) est celui de la loi de  $X_1$ . Le second celui de la loi de  $X_1 + X_2$ , puis  $X_1 + X_2 + X_3$ , puis  $X_1 + X_2 + X_3 + X_4$ , ...

On devine que, au fur et à mesure que le nombre de termes augmente, la densité se rapproche d'une gaussienne; l'étalement augmente aussi, comme prévu (il est proportionnel à la racine carrée du nombre de termes).

et par conséquent

$$\Phi_Z(t) \simeq \left[1 - \mathbf{Var}(X) \frac{t^2}{2n}\right]^n$$

pourvu que  $t/\sqrt{n}$  soit petit. Or tout est là :  $\Phi_Z$  est la transformée de Fourier de la loi discrète de  $Z$  ; cela signifie que le comportement de  $\Phi_Z(t)$  pour les grandes valeurs de  $t$  est conditionné par la structure fine de la loi de  $Z$  ; en revanche, la loi moyennisée conditionne le comportement de  $\Phi_Z(t)$  pour les petites valeurs de  $t$ . La structure fine de la loi de  $Z$ , qui est le bruit discret, peut être très variable et très compliquée selon les cas, et on ne peut rien dire de général sur le comportement de  $\Phi_Z(t)$  pour les grandes valeurs de  $t$  ; celui-ci peut être arbitrairement compliqué. Par contre, *toutes* les fonctions caractéristiques  $\Phi_X$  ont en commun le développement limité au voisinage de  $t = 0$ , ce qui permet d'affirmer, en ignorant tout de ce qui se passe pour les grandes valeurs de  $t$ , que pour  $t$  petit,

$$\Phi_Z(t) \simeq \left[1 - \mathbf{Var}(X) \frac{t^2}{2n}\right]^n \simeq e^{-\mathbf{Var}(X) \frac{t^2}{2}}$$

Le mécanisme de la formation d'une loi de densité gaussienne par simple accumulation de contributions stochastiquement indépendantes (comme on peut le voir sur la figure 17) se comprend donc qualitativement à partir de cette propriété des fonctions caractéristiques. Il reste à comprendre *quantitativement* le rapport entre la coïncidence des fonctions caractéristiques et celle des densités ; en particulier il va falloir préciser l'expression " $t$  petit" :  $t$  doit être suffisamment petit pour que le développement limité ci-dessus soit valable, mais aussi pour que  $(1 - \mathbf{Var}(X) t^2/2n)^n \simeq e^{-\mathbf{Var}(X) t^2/2}$ . Pour cela il faut que  $\mathbf{Var}(X) t^2/n$  soit petit, c'est-à-dire que  $t$  soit d'un ordre de grandeur petit devant  $\sqrt{n/\mathbf{Var}(X)}$ . Cela précise la longueur de corrélation du filtre : il doit éliminer les fréquences de l'ordre de  $\sqrt{n/\mathbf{Var}(X)}$  ou plus. Pour cette étude quantitative, désignons par  $\{z_k, (p_k)\}$  la loi discrète de la variable aléatoire  $Z$ . Sa fonction caractéristique est alors

$$\Phi_Z(t) = \sum_k p_k e^{-iz_k t}$$

Par ailleurs, la transformée de Fourier de la fonction densité

$$f(x) = \frac{1}{\sqrt{2\pi \mathbf{Var}(X)}} \exp\left\{-\frac{x^2}{2 \mathbf{Var}(X)}\right\}$$

est

$$\phi(t) = \int_{-\infty}^{+\infty} f(x) e^{ixt} dx = \exp\left\{-\mathbf{Var}(X) \frac{t^2}{2}\right\}$$

On constate en comparant  $\phi(t)$  avec le développement limité de  $\Phi_Z(t)$  que les deux expressions sont semblables ; autrement dit, pour  $t$  petit,  $\Phi_Z(t) \simeq \phi(t)$ .

Nous allons voir que le paramètre  $\varepsilon$ , la longueur de corrélation de l'échantillonnage qui est nécessaire pour approcher la loi de  $Z$  par une densité continue, est déterminé par la largeur de l'intervalle dans lequel  $\Phi_Z(t)$  reste proche de  $\phi(t)$ . Cette largeur dépend de  $n$ , mais aussi de la loi de  $X$ . Nous nous proposons maintenant de trouver la valeur optimale de  $\varepsilon$  en fonction de  $n$  et de la loi de  $X$ .

Revenons au développement limité (VII.1.) Afin de bien contrôler le domaine de valeurs de  $t$  dans lequel les approximations sont valables, une bonne méthode consiste à pousser les développements limités à un ordre plus grand que ce qui sera finalement retenu, ce qui permet d'avoir une expression explicite de l'erreur. Dans le cas qui nous intéresse, c'est le développement à l'ordre 2 qui sera finalement retenu, donc nous effectuerons tous les calculs en incluant l'ordre 3. On a alors

$$\Phi_X\left(\frac{t}{\sqrt{n}}\right) \simeq 1 - \frac{t^2}{2n}M_2 - i\frac{t^3}{6n\sqrt{n}}M_3$$

On obtient  $\Phi_Z(t)$  en élevant cela à la puissance  $n$ . Pour avoir un développement limité, prenons les logarithmes. Ici  $\Phi_X(t/\sqrt{n}) = 1 - z$  avec

$$z \simeq \frac{t^2}{2n}M_2 + i\frac{t^3}{6n\sqrt{n}}M_3$$

et  $\ln(1 - z) \simeq -z - \frac{1}{2}z^2$ . La variable  $z$  est complexe, mais comme nous avons affaire à des développements en puissances entières de  $z$ , cela ne pose aucun problème tant que  $|z| < 1$ . Ainsi

$$\begin{aligned} \ln\left\{\Phi_X\left(\frac{t}{\sqrt{n}}\right)\right\} &\simeq -z - \frac{1}{2}z^2 \\ &\simeq -\frac{t^2}{2n}M_2 - i\frac{t^3}{6n\sqrt{n}}M_3 - \frac{1}{2}\left[\frac{t^2}{2n}M_2 + i\frac{t^3}{6n\sqrt{n}}M_3\right]^2 \end{aligned}$$

On voit que les termes correspondant à  $z^2$  sont d'ordre supérieur à  $t^3/n\sqrt{n}$  ; on va donc les négliger. En multipliant par  $n$  cela devient

$$\ln\left\{\Phi_X\left(\frac{t}{\sqrt{n}}\right)^n\right\} \simeq -\frac{t^2}{2}M_2 - i\frac{t^3}{6\sqrt{n}}M_3$$

et en revenant aux exponentielles

$$\Phi_Z(t) = \Phi_X\left(\frac{t}{\sqrt{n}}\right)^n \simeq e^{-\frac{t^2}{2}M_2} \cdot e^{-i\frac{t^3}{6\sqrt{n}}M_3}$$

On voit alors que la condition pour que  $\Phi_Z(t) \simeq \exp(-(t^2/2)M_2)$  est que  $(t^3/6\sqrt{n})M_3$  soit petit, c'est-à-dire que

$$t \ll \sqrt[3]{\frac{6}{|M_3|}} n^{1/6} \quad (VII.3.)$$

Par conséquent l'intervalle de valeurs de  $t$  pour lesquelles  $\Phi_Z(t) \simeq e^{-\frac{t^2}{2}M_2}$  est d'autant plus large que  $M_3$  est plus petit et  $n$  plus grand ; mais sa largeur croît très lentement avec  $n$ . A partir de maintenant on appellera  $\eta$  cette largeur ; elle est caractérisée par l'amplitude qu'on tolère pour la différence entre  $\Phi_Z(t)$  et  $\phi(t)$ . Si on se fixe une différence relative maximum autorisée de 1%, cela signifie que le facteur  $\exp(-i(t^3/6\sqrt{n})M_3)$  doit être compris entre 0.99 et 1.01, ce qui correspond à peu près à la condition  $|(t^3/6\sqrt{n})M_3| \leq 0.01$ , d'où  $\eta = \sqrt[3]{0.06/|M_3|} n^{1/6}$ . Si la marge d'erreur tolérée avait été 1/10 000, on aurait eu  $\eta = \sqrt[3]{0.0006/|M_3|} n^{1/6}$ . Plus généralement, pour une marge d'erreur donnée  $\alpha$ , on aura

$$\eta = \sqrt[3]{6\alpha/|M_3|} n^{1/6} \quad (VII.4.)$$

**Remarque :** Cette estimation de l'erreur a été obtenue en retenant le terme d'ordre 3 et en négligeant les suivants, ce qui est légitime si les suivants sont plus petits que lui ; mais si  $M_3$  était beaucoup plus petit que  $M_4$  ou même nul, ce ne serait plus le cas et il faudrait tenir compte du terme d'ordre 4. Le calcul aurait alors donné :

$$t \ll \sqrt[4]{\frac{24}{|M_4 - 6M_2|}} n^{1/4} \quad (VII.3a.)$$

et par conséquent

$$\eta = \sqrt[4]{24\alpha/|M_4 - 6M_2|} n^{1/4} \quad (VII.4a.)$$

Bien entendu, si en plus de  $M_3 \simeq 0$  on a aussi  $M_4 \simeq 6M_2$ , il faudra pousser les développements limités jusqu'à l'ordre  $t^5$ , et ainsi de suite.

Pour filtrer les lois discrètes et faire apparaître leurs densités, nous avons vu qu'on pouvait procéder par échantillonnage ou par convolution. Nous opterons ici pour la convolution, car les rapports de cette opération avec la transformation de Fourier sont plus commodes. Soit donc  $\rho(x)$  un filtre de convolution, par exemple une des fonctions (VII.3 a, b, c). La convolution de la loi de  $Z$  par le filtre  $\rho$  est alors la fonction  $g(x) = \sum_k p_k \rho(x - x_k)$ . La transformée de Fourier de cette fonction est

$$\int_{-\infty}^{+\infty} g(x) e^{itx} dx = \sum_k p_k \int_{-\infty}^{+\infty} \rho(x - x_k) e^{itx} dx$$

## La loi normale

Avec le changement de variable  $y = x - x_k$  cela devient

$$\begin{aligned} \int_{-\infty}^{+\infty} g(x) e^{itx} dx &= \sum_k p_k e^{itx_k} \int_{-\infty}^{+\infty} \rho(y) e^{ity} dy \\ &= \sum_k p_k e^{itx_k} \hat{\rho}(t) \\ &= \Phi_Z(t) \hat{\rho}(t) \end{aligned}$$

où  $\hat{\rho}(t)$  est la transformée de Fourier de  $\rho(x)$ . Ainsi, la convolution  $g$  de la loi de  $Z$  avec le filtre  $\rho$  a pour transformée de Fourier le produit de  $\Phi_Z$  par  $\hat{\rho}$ . Or, si  $\rho$  est une fonction du type (VII.2c) avec  $\varepsilon = 1/\eta$ , c'est-à-dire si

$$\rho(x) = \frac{\sin(\eta\pi x)}{\pi x}$$

alors  $\hat{\rho}(t)$  est la fonction égale à 0 pour  $|t| \geq \eta$  et égale à 1 pour  $|t| < \eta$ . En convoluant la loi de  $Z$  par  $\rho$ , on multiplie donc  $\Phi_Z(t)$  par 0 lorsque  $|t| \geq \eta$  et par 1 lorsque  $|t| < \eta$ , ce qui revient à tronquer  $\Phi_Z$  de tout ce qui est en dehors de l'intervalle  $[-\eta, +\eta]$ . Cela consiste donc bien à appliquer un filtre passe-bas.

La fonction caractéristique  $\Phi_Z(t)$  n'était proche de  $\phi(t)$  que pour les petites valeurs de  $t$ ; précisément les valeurs comprises entre  $-\eta$  et  $+\eta$ . En dehors de cet intervalle,  $\Phi_Z(t)$  pouvait différer énormément de  $\phi(t)$ . Mais la fonction tronquée, elle, est *partout* proche de  $\phi(t)$ , car en dehors de l'intervalle  $[-\eta, +\eta]$  la fonction tronquée est nulle, et la fonction  $\phi(t)$  est inférieure à  $\exp(-(\eta^2/2)\mathbf{Var}(X))$ , donc pratiquement nulle si  $\eta$  est assez grand.

Il est facile de comparer les fonctions  $f(x)$  et  $g(x)$ ; en effet, on peut inverser l'intégrale de Fourier, de sorte que

$$f(x) - g(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} [\phi(t) - \Phi_Z(t) \hat{\rho}(t)] e^{-itx} dt$$

On en déduit en prenant les modules

$$\begin{aligned} |f(x) - g(x)| &\leq \frac{1}{2\pi} \int_{-\infty}^{+\infty} |\phi(t) - \Phi_Z(t) \hat{\rho}(t)| dt \\ &\leq \frac{1}{2\pi} \int_{-\infty}^{+\infty} |1 - e^{-i\frac{t^3}{6\sqrt{n}}M_3}| \phi(t) dt + \frac{1}{\pi} \int_{\eta}^{+\infty} \phi(t) dt \end{aligned}$$

Le second terme est d'autant plus petit que  $\eta$  est plus grand. Quant au premier, on remarque qu'il est égal au seuil d'erreur  $\alpha$  (cf. VII.4.) qu'on s'est fixé, multiplié par  $\frac{1}{2\pi} \int \phi(t) dt$ , ce qui vaut  $\alpha/\sqrt{2\pi M_2}$

Bien entendu l'erreur *ne peut pas* être rendue arbitrairement petite pour une valeur fixée de  $n$  : on voit que si on prend  $\alpha$  trop petit,  $\eta$  sera trop petit également de sorte que le second terme ci-dessus ne sera pas petit. Le meilleur choix (pour une valeur fixée de  $n$ ) est donc celui qui rend minimum l'expression

$$u(\alpha) = \frac{\alpha}{\sqrt{2\pi M_2}} + \frac{1}{\pi\sqrt{M_2}} \gamma(\eta\sqrt{M_2})$$

avec

$$\gamma(r) = \int_r^{+\infty} e^{-\frac{s^2}{2}} ds$$

et  $\eta$  étant fonction de  $\alpha$  selon (VII.4.) En effet, l'expression  $u(\alpha)$  est la somme de deux termes : le premier croît avec  $\alpha$  et le second décroît avec  $\alpha$ , de sorte qu'il doit y avoir une valeur optimale de  $\alpha$  pour laquelle elle sera minimum.

Mais par ailleurs, bien que  $\alpha$  soit petit par nature,  $\eta$  doit être grand (le développement limité de  $\Phi_Z(t)$  doit être correct sur un intervalle assez large, ou, si on préfère, le bruit discret qu'on élimine ne doit comporter réellement que des hautes fréquences, sinon l'idée même d'une approximation de la loi discrète par une densité continue perd son sens). De sorte que le second terme  $\frac{1}{\pi\sqrt{M_2}} \gamma(\eta\sqrt{M_2})$  peut tout simplement être négligé (dès que  $\eta\sqrt{M_2}$  est supérieur à 5 unités<sup>(1)</sup> ce terme n'est déjà plus que de l'ordre de  $10^{-6}$ ). Donc pour que l'erreur  $\alpha$  soit petite tout en respectant la condition  $\eta \geq 5/\sqrt{M_2}$ , il faut avoir

$$1 \gg \alpha = \frac{|M_3|}{6\sqrt{n}} \eta^3 \geq \frac{125 |M_3|}{6 M_2 \sqrt{M_2} \sqrt{n}} \geq \frac{20 |M_3|}{M_2 \sqrt{M_2} \sqrt{n}}$$

ce qui veut dire que

$$n \geq 400 |M_3|^2 / M_2^3 \alpha^2 \quad (\text{VII.5.})$$

Cette inégalité exprime le seuil à partir duquel  $n$  peut être considéré comme assez grand pour que *après filtrage* la loi de  $Z = (X_1 + X_2 + \dots + X_n) / \sqrt{n}$  ne diffère de la densité  $f(x)$  que de  $\alpha$  au plus. On voit que ce seuil est inversement proportionnel à  $\alpha^2$ , ce qui signifie que pour diminuer l'erreur  $\alpha$  d'un facteur deux, il faut prendre  $n$  quatre fois plus grand. Ce seuil est aussi proportionnel à  $|M_3|^2$ , mais n'oublions pas que le calcul précédent

---

<sup>(1)</sup> Ce facteur 5 est évidemment assez arbitraire : il garantit que le terme  $\gamma$  sera négligeable si on tient  $10^{-6}$  pour négligeable ; si on est plus exigeant on pourra prendre 10 au lieu de 5.

n'est valable que si  $|M_3|$  est assez grand; sinon il faut utiliser l'estimation issue de (VII.4a.) qui donne

$$n \geq 25 |M_4 - 6M_2| / M_2^2 \alpha \quad (VII.5a.)$$

dans ce cas le seuil est inversement proportionnel à  $\alpha$  et non plus à  $\alpha^2$ .

Si  $n$  est supérieur à ce seuil, alors la longueur de corrélation  $\varepsilon$  du filtrage est, comme nous l'avons vu, égale à  $1/\eta$ , donc inférieure à  $\sqrt{M_2}/5$ . La signification du paramètre  $\varepsilon$  est que le filtre donne un bon lissage si la distance entre les valeurs discrètes est inférieure à  $\varepsilon$ ; donc la signification du seuil concerne aussi la distance entre les valeurs discrètes de  $Z$ . Lorsqu'on fait la somme des variables  $X_1, X_2, X_3, \dots, X_n$ , on ne réduit pas la distance entre les valeurs, mais on les étale de plus en plus; or  $Z$  est égal à cette somme, *divisée par*  $\sqrt{n}$ ; de sorte que les valeurs de  $Z$  sont de plus en plus serrées au fur et à mesure que  $n$  augmente: en fait, la distance entre les valeurs discrètes de  $Z$  diminue en  $1/\sqrt{n}$ . Il y a donc pour  $n$  un seuil à partir duquel les distances entre deux valeurs discrètes consécutives de  $Z$  deviendront inférieures à  $\varepsilon$ . Comme on peut le voir sur les figures 13 à 16, lorsque  $\varepsilon$  est trop petit le bruit discret n'est pas filtré: c'est ce qui se produit sur ces figures lorsque  $\varepsilon$  est plus petit que la distance entre les valeurs discrètes de la variable aléatoire. Lorsqu'on considère la somme de  $n$  variables aléatoires  $X_j$  stochastiquement indépendantes, ce seuil au delà duquel le filtrage devient correct est donné par les inégalités (VII.5.) ou (VII.5a.)

On peut donc énoncer les choses ainsi: si on moyennise la loi discrète de  $Z$  sur des longueurs de corrélation d'ordre légèrement supérieur à  $\sqrt{\text{Var}(X)}/n$ , on obtient une distribution approximativement gaussienne. La densité de cette distribution continue est alors proche de

$$\frac{1}{\sqrt{2\pi \text{Var}(X)}} \exp\left\{-\frac{z^2}{2 \text{Var}(X)}\right\}$$

Cela revient à dire que si  $\varepsilon$  est la bonne longueur de corrélation, déterminée par (VII.4.) ou (VII.4a.), alors à  $\alpha$  près on a

$$\mathcal{P}\left\{a - \frac{\varepsilon}{2} < \frac{Z}{\sqrt{n}} < a + \frac{\varepsilon}{2}\right\} \simeq \int_{a-\frac{\varepsilon}{2}}^{a+\frac{\varepsilon}{2}} \frac{1}{\sqrt{2\pi \text{Var}(X)}} \exp\left\{-\frac{z^2}{2 \text{Var}(X)}\right\} dz$$

ou plus généralement, pour tous  $a, b$  tels que  $b - a \geq \varepsilon$ :

$$\mathcal{P}\left\{a < \frac{Z}{\sqrt{n}} < b\right\} \simeq \int_a^b \frac{1}{\sqrt{2\pi \text{Var}(X)}} \exp\left\{-\frac{z^2}{2 \text{Var}(X)}\right\} dz \quad (VII.6.)$$

Pour conclure ce chapitre on peut encore faire la remarque suivante : formellement, nous avons établi la propriété que la somme d'un grand nombre de variables aléatoires indépendantes *et de même loi* était approximativement gaussienne. En fait, il n'est pas nécessaire qu'elles aient la même loi (par exemple ce n'est pas le cas dans la figure 17). En effet, si les  $X_j$  ont la même loi, leur fonction caractéristique commune est  $\Phi_X(t)$  et par conséquent celle de  $Z$  sera  $\Phi_X(t/\sqrt{n})^n$ . Si les  $X_j$  ont des lois différentes, leurs fonctions caractéristiques seront différentes, mais pour  $t$  petit (là où le développement limité est valable), elles ne diffèrent que par la valeur de  $\mathbf{Var}(X_j)$ . Si les  $X_j$  sont indépendantes (car cette hypothèse là reste essentielle), la fonction caractéristique de  $Z$  sera le produit

$$\prod_{j=1}^{j=n} \Phi_{X_j}\left(\frac{t}{\sqrt{n}}\right)$$

avec des facteurs non égaux, mais pour  $t$  petit on aura toujours

$$\prod_{j=1}^{j=n} \left(1 - \mathbf{Var}(X_j) \frac{t^2}{2n}\right) \simeq \exp\left\{-A \frac{t^2}{2}\right\}$$

avec  $A = \frac{1}{n} \sum \mathbf{Var}(X_j)$  (moyenne des variances). Il faut toutefois que soient satisfaites les conditions pour que l'approximation soit valable : par exemple si  $\mathbf{Var}(X_j)$  augmente proportionnellement à  $j$ , cela ne marche plus ; on supposera donc que les variances des  $X_j$  restent bornées pendant que  $n$  tend vers l'infini.

La conclusion générale à tirer de ce chapitre est la suivante : lorsqu'un phénomène quelconque subit des fluctuations dues à un grand nombre de perturbations aléatoires, pourvu que ces perturbations agissent indépendamment les unes des autres, les fluctuations suivront une loi approximativement gaussienne. Le fait que, comme nous venons de le voir, le caractère gaussien des fluctuations est universel, permet de postuler a priori que des phénomènes produits par des causes dont on ignore tout, telles que les erreurs de mesure, l'agitation thermique, l'effet d'un médicament, les notes obtenues aux examens, etc. présentent autour de leur position moyenne des fluctuations gaussiennes ; pour pouvoir affirmer cela avec certitude il n'est pas nécessaire de savoir quoi que ce soit sur les causes de ces fluctuations, excepté qu'elles sont nombreuses, qu'elles s'ajoutent les unes aux autres, et qu'elles sont stochastiquement indépendantes les unes des autres.

C'est pourquoi en statistique on fait presque toujours l'hypothèse que les fluctuations autour de la moyenne obéissent à cette loi gaussienne. Les tests statistiques que nous étudierons au chapitre **XI** partent de cette

hypothèse. En effet, en statistique on traite des données brutes, et on ignore généralement tout de l'enchevêtrement des causes. Mais on peut malgré tout avoir la certitude que des paramètres sont intervenus indépendamment les uns des autres : par exemple, si des tireurs à l'arc visent le centre d'une cible, la trajectoire de la flèche est perturbée par le tremblement de leurs muscles, et par la résistance de l'air sur la flèche, qui à grande vitesse accentue fortement les imperceptibles dissymétries de l'empennage ou de la pointe. S'il est absolument impossible, non seulement de calculer ces fluctuations de façon déterministe, mais même de connaître les lois de probabilités des innombrables paramètres qui interviennent, il est cependant déraisonnable de penser que les valeurs aléatoires prises par ces paramètres au cours d'un tir particulier soient influencées par ce qui s'est produit au tir précédent. Sans qu'on puisse rien calculer, de simples principes généraux de causalité peuvent ainsi garantir l'indépendance stochastique des différentes causes de perturbation. En outre, même lorsque rien ne garantit l'indépendance (par exemple si on relève les fluctuations du taux de cholestérol sanguin on a affaire à un organisme vivant dans lequel tout est lié) l'expérience montre presque toujours, si on trace les histogrammes des fluctuations en fonction des écarts par rapport à la moyenne, qu'on obtient au moins qualitativement une répartition gaussienne. A tel point que lorsqu'on obtient une répartition nettement différente, on s'interroge sur les raisons d'une telle anomalie.

Si l'ampleur des fluctuations est *en-deça* de la précision avec laquelle le phénomène est perçu ou mesuré, alors le phénomène apparaît comme déterministe. C'est la raison pour laquelle le monde macroscopique est dominé par le déterminisme, alors que le monde des atomes et des particules est dominé par le hasard.