# XI. TESTS STATISTIQUES

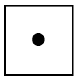
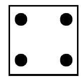
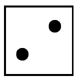
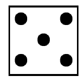
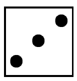
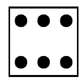
## XI. 1. Les densités $\chi_r^2(t)$

Au chapitre précédent, nous avons vu comment utiliser la loi gaussienne pour distinguer une véritable anomalie d'une simple fluctuation aléatoire. L'exemple que nous avons retenu pour discuter était celui d'une pièce de monnaie qu'on lance un grand nombre $n$ de fois. La loi gaussienne des fluctuations nous dit que la probabilité pour que le nombre de pile ou de face s'écarte de la valeur médiane $n/2$ de plus que, disons, cinq fois l'écart-type $\sigma = \sqrt{n/2}$ est si faible qu'il est pratiquement impossible qu'un tel écart se produise, ce qui permet, si cela se produit, d'affirmer avec un très haut degré de certitude que la pièce était donc mal équilibrée.

Les tests statistiques sont des recettes basées sur ce principe. Tous ne font que transposer le raisonnement précédent, en introduisant un intermédiaire mathématique qui ne change rien quand au fond, mais facilite le calcul. Dans le cas de la pièce de monnaie le recours direct à la loi gaussienne était facile : il suffisait de compter le nombre de pile, puis de comparer son écart par rapport à $n/2$ à l'écart-type $\sigma = \sqrt{n/2}$, et on pouvait alors sans calcul supplémentaire se servir d'une table de la loi gaussienne.

Les choses deviennent plus complexes lorsqu'au lieu d'une pièce de monnaie, on teste un dé. Si on lance le dé 6000 fois, chaque face du dé doit apparaître environ 1000 fois, et c'est alors l'écart par rapport à 1000 qui mesure la fluctuation. Mais il y a *six* fluctuations, une par face. En fait, il n'y en a que cinq indépendantes, car, de même que pour la pièce de monnaie le nombre de pile déterminait automatiquement le nombre de face, pour le dé la connaissance de cinq fluctuations détermine la sixième. Il est bien entendu que chacune de ces six fluctuations obéit à une loi approximativement gaussienne, et qu'on peut donc analyser chacune séparément, avec sa propre loi gaussienne. Mais il est assez rare en Statistique qu'une information aussi détaillée soit utile : il suffit de savoir si le dé est globalement bon ou mauvais et il est inutile de s'encombrer de ce qui arrive à chacune des faces. Le dé sera considéré comme mal équilibré même si un seul des six écarts est trop grand (par exemple supérieur à deux et demi ou trois fois l'écart-type) ; autrement dit : un dé qui est parfaitement équilibré pour les faces $1, 2, 5, 6$, mais mal éqilibré pour les faces 3 et 4, n'est pas meilleur qu'un dé qui est mal équilibré sur toutes les faces. Dans un tel cas tester l'écart pour chacune des six faces séparément est une perte de temps, et on souhaiterait plutôt pouvoir mesurer un écart global, par exemple la somme des carrés des fluctuations. Certes, la somme des valeurs absolues des fluctuations, ou la somme des quatrièmes puissances pourraient convenir également, mais, comme cela a déjà été discuté à propos de la variance (voir **VI.2.**), la somme des carrés est considérée pour diverses raisons comme la plus pratique (bien entendu la somme *algébrique* des fluctuations est par nature toujours nulle et ne mesure donc rien).

Pour que la discussion soit plus concrète, imaginons que le dé, lancé 6000 fois, donne les résultats suivants :

| | | | | |
|---|---|---|---|---|
|  | 1064 fois |  | 1009 fois | |
|  | 987 fois |  | 977 fois | $(XI.1.)$ |
|  | 961 fois |  | 1002 fois | |

Si on veut faire comme nous avons fait pour la pièce de monnaie, il faudra comparer les six résultats aux fluctuations gaussiennes autour de la moyenne 1000 ; dans ce cas il faut déterminer l'écart-type : en l'occurrence il vaut $\sqrt{1000} \simeq 32$. Ici l'écart-type est le même pour chacune des six faces, ce qui est évident puisqu'il y a symétrie (les six faces sont interchangeables). On peut aussi calculer la somme des carrés qui est alors

$$S = 64^2 + 13^2 + 39^2 + 9^2 + 23^2 + 2^2 = 6400 \qquad (XI.2.)$$

Mais plus question de comparer cette valeur à l'écart-type 32. Il y a à cela trois raisons : la première est qu'on fait la somme de six contributions, il faudrait donc comparer à six fois la valeur moyenne attendue pour l'une ; la seconde est qu'on prend les carrés (qui sont beaucoup plus grands), il faudrait donc comparer à six fois le carré de l'écart-type. Mais ce raisonnement serait faux : ce n'est pas à six fois le carré de l'écart-type qu'il faut comparer ce nombre, car (c'est la troisième raison) cette somme de carrés ne suit pas une loi gaussienne. Il se trouve en effet que, si les fluctuations obéissent à une loi gaussienne, ce n'est plus le cas de leurs carrés.

Si on n'a pas du tout réfléchi à la question, on peut être surpris d'apprendre que le carré d'une variable aléatoire n'obéit pas à la même loi que la variable elle-même, mais cela est logique et a déjà été évoqué : par exemple dans le problème des distributions au hasard de cordes sur un cercle (chapitre **I**) la différence entre les trois lois se traduisait mathématiquement

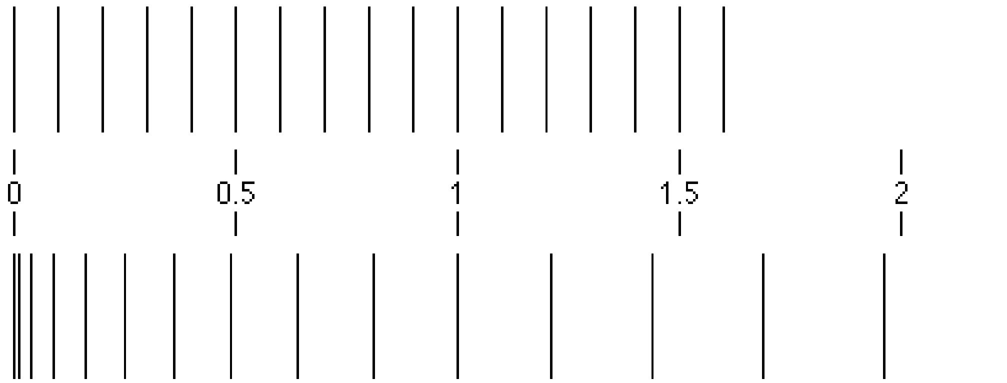



figure 41

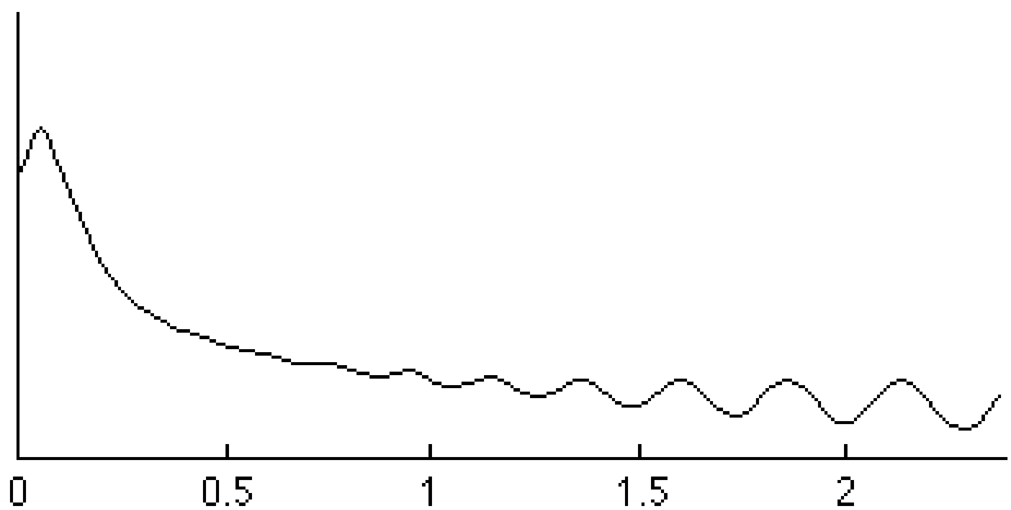



figure 42

par le fait que les variables aléatoires correspondantes sont liées par des relations non linéaires. Sur la figure 41 on peut voir deux graphiques de loi. Celui du haut représente une loi $\{x_j\ ;\ (p_j)\}$ telle que les $p_j$ soient tous égaux à $\frac{1}{17}$ et $x_j = \frac{1}{10}j$ : les valeurs sont donc équidistantes et vont de 0 à 1.6 ; elles sont toutes équiprobables et, comme il y en a 17 en tout, la probabilité de chacune est $\frac{1}{17}$. Le graphique du bas représente la loi du carré : les $x_j$ sont remplacées par les $x_j^2$, mais les $p_j$ ne changent pas : le passage au carré rend plus petit ce qui est inférieur à 1, mais rend plus grand ce qui est supérieur à 1. Ce qui a pour effet que les valeurs inférieures à 1 se trouvent davantage comprimées vers 0, tandis que les valeurs supérieures à 1 sont dilatées. Par conséquent la *densité*, qui était uniforme, se trouve augmentée entre 0 et 1 (d'autant plus qu'on est plus près de 0) et diminuée au-delà de 1. Si on effectue un lissage de cette loi discrète (celle du carré) pour faire apparaître sa densité, on obtient le graphique de la figure 42 ; le lissage de la loi initiale $\{x_j\ ;\ (p_j)\}$ aurait évidemment donné une densité constante entre 0 et 1.6. Ainsi, comme nous en avions déjà eu des exemples ailleurs, ce n'est pas une modification des probabilités qui change la loi, mais une modification dans la répartition des valeurs, consécutive à une transformation non linéaire.

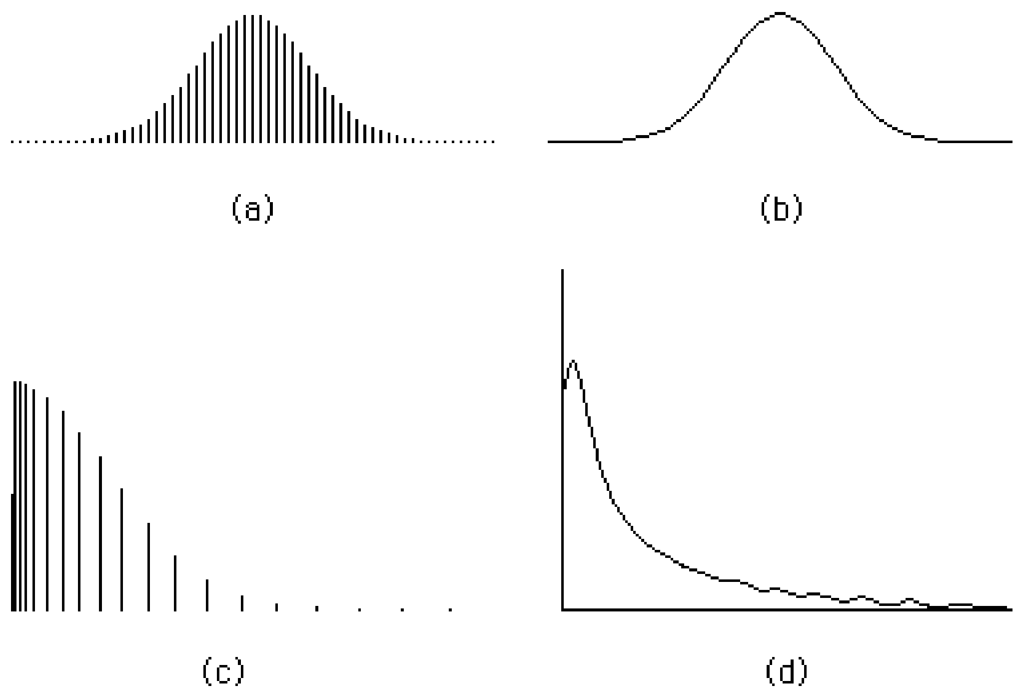



figure 43

On s'attend donc à un effet analogue pour la loi binômiale. Sur la figure 43 on peut voir en $(a)$ la loi binômiale et en $(b)$ la même loi après lissage par convolution ; ces graphiques sont les mêmes que sur les figures 11 (en haut), et 14 ($\varepsilon = 0.42$) du chapitre **VII**. Si $X$ est une variable aléatoire de loi $\{x_j\ ;\ (p_j)\}$, son carré a pour loi $\{x_j^2\ ;\ (p_j)\}$ : c'est-à-dire que les probabilités $p_j$ ne changent pas, mais les valeurs $x_j$ sont remplacées par les carrés. Si parmi les $x_j$ il y en a deux distinctes $x_{j_1}$ et $x_{j_2}$ qui ont le même carré, la probabilité de ce carré sera évidemment $p_{j_1} + p_{j_2}$. Dans le cas de la loi binômiale représentée sur la figure 43 $a$, qui est symétrique, cela se produit pour chacune des $x_j$, sauf 0. Ainsi la probabilité de $+x_j$ est (approximativement) $p_0 \exp(-x_j^2/2)$, et la probabilité de $-x_j$ est aussi

$p_0 \exp(-x_j^2/2)$, de sorte que la probabilité de $x_j^2$ sera $2p_0 \exp(-x_j^2/2)$, à l'exception de $x_0 = 0$ qui est seule à avoir 0 pour carré, et dont la probabilité ne sera donc pas $2p_0$, mais $p_0$. C'est pourquoi on voit en $(c)$ que les bâtons du graphique (dont la longueur est proportionnelle à la probabilité correspondante) sont deux fois plus longs que les bâtons du graphique $(a)$, à l'exception du tout premier, qui se confond presque avec le second. Mais ce n'est pas la seule différence ; car sur le graphique $(a)$ les bâtons sont équidistants (en effet les valeurs $x_j$ de la loi binômiale sont toutes des multiples entiers de $x_1$), et sur le graphique $(c)$ ils ne le sont plus car leurs abscisses sont les carrés des précédentes. Ainsi les valeurs prises par le carré sont plus denses au voisinage de zéro et moins denses loin de l'origine. Si on effectue une convolution sur la loi $(c)$, afin de mieux visualiser la *densité* de la loi, on obtient le graphique $(d)$, qui ne ressemble plus du tout au graphique $(b)$.

La loi de probabilité qui est représentée par le graphique $(c)$ ou, sous forme lissée, par le graphique $(d)$, est appelée *la loi du* $\chi^2$ *à un degré de liberté*. En toute rigueur, la loi du $\chi^2$ à un degré de liberté n'est qu'asymptotiquement la loi du carré d'une variable *binômiale*, c'est-à-dire que la loi discrète$(c)$ a pour limite la loi du $\chi^2$ lorsque le nombre de bâtons tend vers l'infini et que la distance entre eux tend conjointement vers zéro, de même que la loi gaussienne est la limite de la loi binômiale. Les graphiques $(c)$ et $(d)$ ne correspondent donc qu'asymptotiquement à la loi du $\chi^2$ à un degré de liberté.

On peut résumer :

| |
|---|
| La densité normalisée $\chi_1^2(t)$ à un degré de liberté est la densité du carré d'une variable aléatoire de densité gaussienne $\frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$. |

Mais dans $(XI.2.)$ il y avait une somme de carrés. Lorsqu'on additionne des variables aléatoires gaussiennes indépendantes, leur somme est également gaussienne, avec une variance qui est la somme des variances de chacune ; cela résulte d'une propriété très spéciale que les densités gaussiennes possèdent, à savoir que la convolution de plusieurs densités gaussiennes est aussi une densité gaussienne. Les carrés ont pour densité $\chi^2(t)$ et non $e^{-x^2/2}$, donc il n'y a aucune raison pour que leur somme soit aussi une densité $\chi^2(t)$. Nous allons voir ci-dessous lorsque nous calculerons ces fonctions que la densité de la somme de plusieurs carrés (disons $r$ carrés) n'est pas semblable à la densité d'un seul carré, et on l'appelle la densité $\chi_r^2(t)$, ou densité $\chi^2(t)$ à $r$ degrés de liberté.

On pourra donc retenir la définition suivante :

> La densité normalisée $\chi_r^2(t)$ à $r$ degrés de liberté est la densité de la somme des carrés de $r$ variables aléatoires stochastiquement indépendantes, chacune de densité gaussienne $\frac{1}{\sqrt{2\pi}}\exp(-x^2/2)$.

Soit $X$ une variable de loi $e^{-x^2/2}$. On sous-entendra qu'il s'agit de lois asymptotiques et qu'en réalité la loi est discrète. Cela signifie que

$$\mathcal{P}\,(a \leq X < b) \quad \simeq \quad \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{x^2}{2}}\,dx$$

si $b - a$ est suffisamment grand pour contenir un nombre statistiquement significatif de valeurs discrètes. Il suffit d'ailleurs pour caractériser la densité, de connaître la fonction

$$F(x) \quad = \quad \mathcal{P}\,(X < x) \quad \simeq \quad \frac{1}{\sqrt{2\pi}}\int_{-\infty}^x e^{-\frac{t^2}{2}}\,dt \qquad (XI.3.)$$

appelée *fonction de répartition* de $X$ ; en effet, $\mathcal{P}\,(a < X < b) = F(b) - F(a)$. Pour déterminer la densité du carré de $X$, nous partirons donc du fait que la fonction de répartition de $X$ est celle donnée par $(XI.3.)$ et il s'agit de calculer la fonction de répartition $G(x)$ de $X^2$.

Or

$$G(x) \quad = \quad \mathcal{P}\,(X^2 < x) \quad = \quad \mathcal{P}\,(-\sqrt{x} < X < \sqrt{x}) \quad \simeq \quad \frac{1}{\sqrt{2\pi}}\int_{-\sqrt{x}}^{\sqrt{x}} e^{-\frac{t^2}{2}}\,dt$$

Puisque la fonction à intégrer ci-dessus est paire, on a aussi

$$G(x) \quad \simeq \quad \frac{2}{\sqrt{2\pi}}\int_0^{\sqrt{x}} e^{-\frac{t^2}{2}}\,dt$$

Effectuons dans l'intégrale le changement de variable $s = t^2$ ; on obtient

$$G(x) \quad \simeq \quad \frac{1}{\sqrt{2\pi}}\int_0^x s^{-\frac{1}{2}} e^{-\frac{s}{2}}\,ds$$

L'astuce, dans le changement de variable, était de faire en sorte que la borne supérieure de l'intégrale, qui était devenue $\sqrt{x}$, redevienne $x$. Ainsi la probabilité pour que $X^2$ soit inférieur à $x$ est l'intégrale de 0 à $x$ (et non plus de 0 à $\sqrt{x}$) de la densité $\sqrt{1/2\pi}\; s^{-1/2}\, e^{-s/2}$ ; c'est la preuve que la fonction sous l'intégrale est bien la densité cherchée. La densité $\chi_1^2(t)$ sera donc

$$\chi_1^2(t) \quad = \quad \sqrt{\frac{1}{2\pi}}\, t^{-\frac{1}{2}}\, e^{-\frac{s}{2}} \qquad (XI.4.)$$

On voit que cette densité tend vers l'infini comme $1/\sqrt{t}$ quand $t$ tend vers 0. Cette singularité est évidemment un artefact du passage à une limite continue : il ne peut pas y avoir une telle densité infinie pour des lois discrètes et réelles, comme le montrent les figures 42 et 43 ($d$). Le phénomène doit être compris de la façon suivante. Nous avons vu au chapitre **VII** que lorsqu'on dit qu'une loi discrète tend vers une densité continue, cela signifie que c'est la loi échantillonnée ou lissée par convolution qui tend (au sens de la convergence uniforme des fonctions) vers la densité continue ; mais la loi échantillonnée ou lissée est d'autant plus proche de sa limite que les valeurs de la loi discrète sont plus nombreuses et plus proches les unes des autres : ce qui fait qu'une loi discrète "tend" vers une densité continue, est que la distance moyenne entre les valeurs consécutives (le pas de discrétisation) tend vers zéro. Si dans le graphique de la figure 41 ou celui de la figure 43 ($a$) on augmente le nombre de valeurs en même temps qu'on les rapproche les unes des autres (par exemple, au lieu de prendre les 17 valeurs $x_j = \frac{1}{10}j$ pour $j = 0,\, 1,\, \ldots 16$, chacune ayant la probabilité $\frac{1}{17}$, on prend $1,7 \cdot 10^n$ valeurs $x_j = 10^{-n}j$ pour $j = 0,\, 1,\, \ldots 1,7 \cdot 10^n - 1$, chacune ayant la probabilité $1/1,7 \cdot 10^n$, et on fait tendre $n$ vers l'infini), on ne change rien au fait que la loi lissée sera uniforme ; par contre la loi lissée du carré(figure 42) va tendre vers la densité $1/\sqrt{t}$ (bien entendu le paramètre $\varepsilon$ — la longueur de corrélation — du filtre de lissage devra tendre corrélativement vers zéro). Il n'est pas possible, à cause du faible pouvoir séparateur des graphiques, de présenter les dessins qui illustreraient le phénomène ; mais si on le pouvait, on verrait que la courbe de la figure 42 tendrait vers la courbe $y = 1/\sqrt{x}$ : en particulier l'abscisse de son maximum, qui est déjà proche de zéro sur le graphique, s'en rapprocherait de plus en plus, tandis que son ordonnée prendrait des valeurs de plus en plus grandes. Cela se comprend aisément : si les valeurs discrètes se densifient, le fait de les élever au carré produira une densité énorme tout près de zéro (par exemple si la distance entre les valeurs est $\varepsilon$, la distance entre les valeurs du carré au voisinage immédiat de zéro sera de l'ordre de $\varepsilon^2$). Tant que la loi est discrète, cette densité énorme ne peut toutefois pas être infinie, et c'est pourquoi la courbe lissée possède près de zéro un maximum aigu, mais non une singularité. Ce maximum se transforme cependant en singularité à la limite continue.

On peut de la même manière calculer la densité $\chi_r^2(t)$ à $r$ degrés de liberté. Cette fois il s'agit de la *somme* de $r$ carrés indépendants, on a donc $r$ variables aléatoires $X_1,\, X_2,\, X_3,\, \ldots X_r$ stochastiquement indépendantes, et pour chacune on a

$$\mathcal{P}\,(X_j < x) \;\simeq\; \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{t^2}{2}}\, dt$$

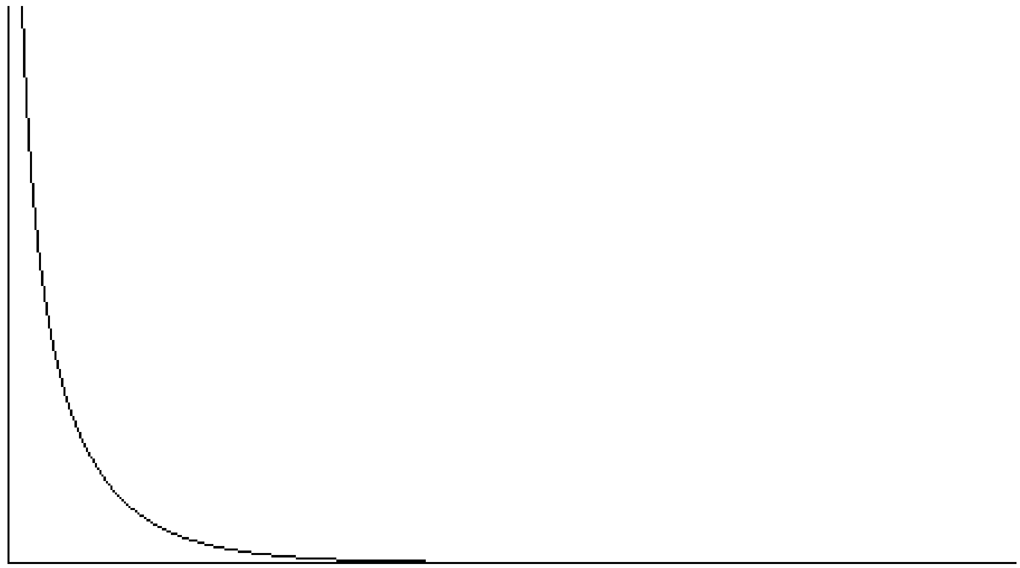
Puisque les $X_j$ sont stochastiquement indépendantes, leur loi conjointe s'obtient par factorisation, de sorte que pour la somme des carrés on aura

$$\begin{aligned}
&\mathcal{P}\,(X_1^2 + X_2^2 + X_3^2 + \cdots X_r^2 < x) \quad \simeq \\
&\quad \simeq \iiint\limits_{t_1^2+t_2^2+t_3^2+\cdots t_r^2<x} \!\!\cdots\! \int \left[\frac{1}{\sqrt{2\pi}}\right]^r e^{-\frac{{t_1}^2}{2}} e^{-\frac{{t_2}^2}{2}} e^{-\frac{{t_3}^2}{2}} \cdots\, e^{-\frac{{t_r}^2}{2}}\; dt_1\, dt_2\, dt_3 \cdots dt_r \\
&\quad = \iiint\limits_{t_1^2+t_2^2+t_3^2+\cdots t_r^2<x} \!\!\cdots\! \int \left[\frac{1}{\sqrt{2\pi}}\right]^r e^{-\frac{{t_1}^2+{t_2}^2+{t_3}^2+\cdots+{t_r}^2}{2}}\; dt_1\, dt_2\, dt_3 \cdots dt_r
\end{aligned}$$

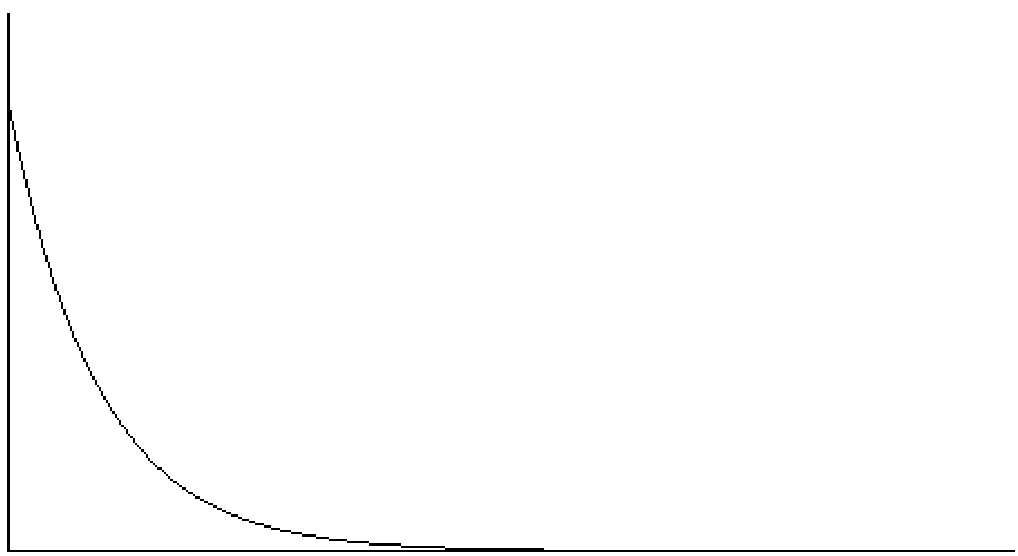
On voit que dans cette intégrale multiple tout est à symétrie sphérique, aussi bien la fonction à intégrer que le domaine d'intégration. On peut donc avantageusement passer en coordonnées sphériques, ce qui donne

$$\mathcal{P}\,(X_1^2 + X_2^2 + X_3^2 + \cdots X_r^2 < x) \;\simeq\; \left[\frac{1}{\sqrt{2\pi}}\right]^r \int_{\rho^2<x} \rho^{r-1} e^{-\frac{\rho^2}{2}}\, d\rho \;\cdot\; \frac{2\,\pi^{\frac{r}{2}}}{\Gamma\!\left(\frac{r}{2}\right)}$$

le facteur $2\,\pi^{r/2}\big/\,\Gamma(r/2)$ provenant de l'intégration des coordonnées sphériques angulaires (c'est l'aire de la sphère unité dans l'espace à $r$ dimensions).

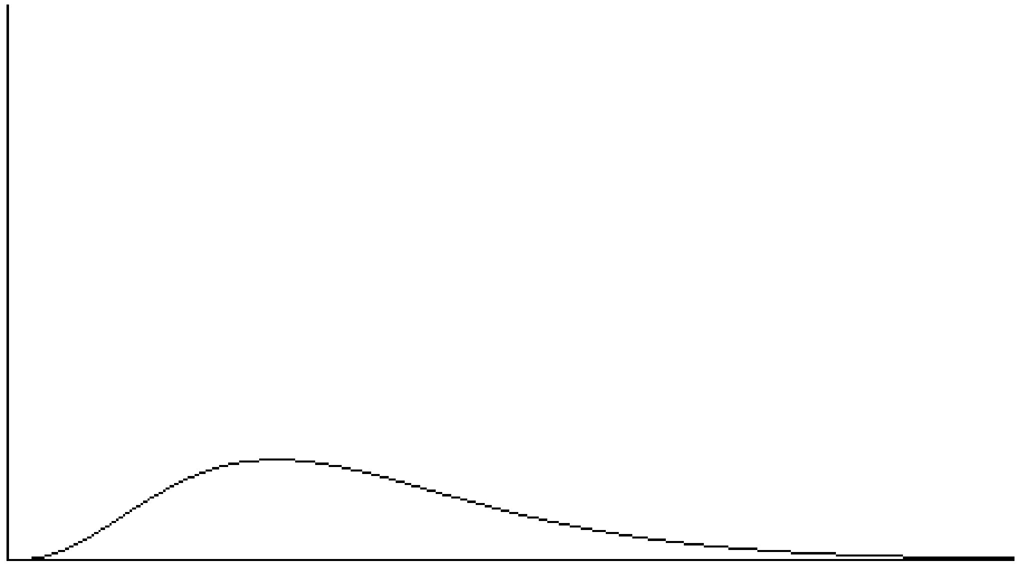


La densité $\chi^2$ à un degré de liberté.

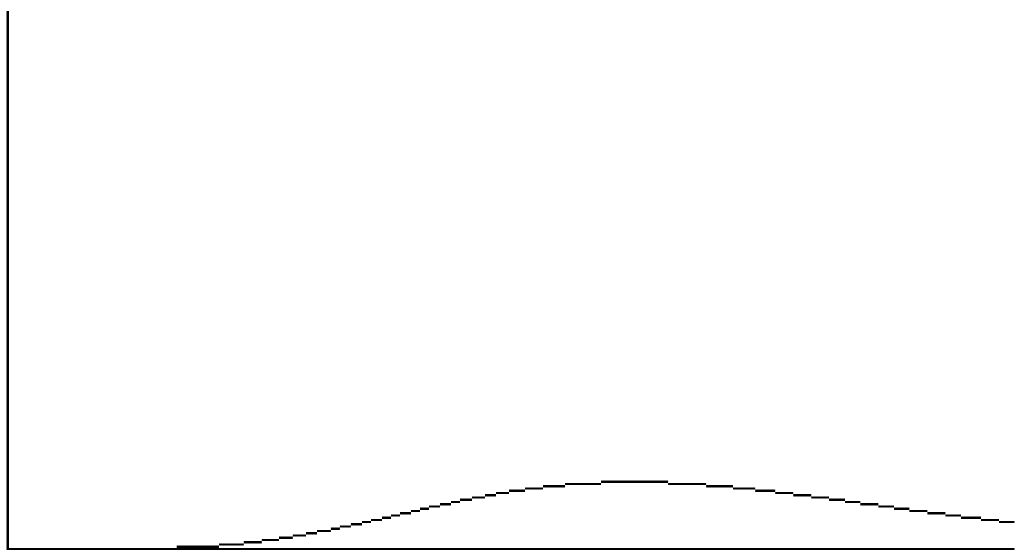


La densité $\chi^2$ à deux degrés de liberté.

figure 44



La densité $\chi^2$ à huit degrés de liberté.



La densité $\chi^2$ à seize degrés de liberté.

figure 45

Voici quatre exemplaires (figures 44 et 45) choisis parmi les densités $\chi^2$ ; les quatre graphiques sont à la même échelle.

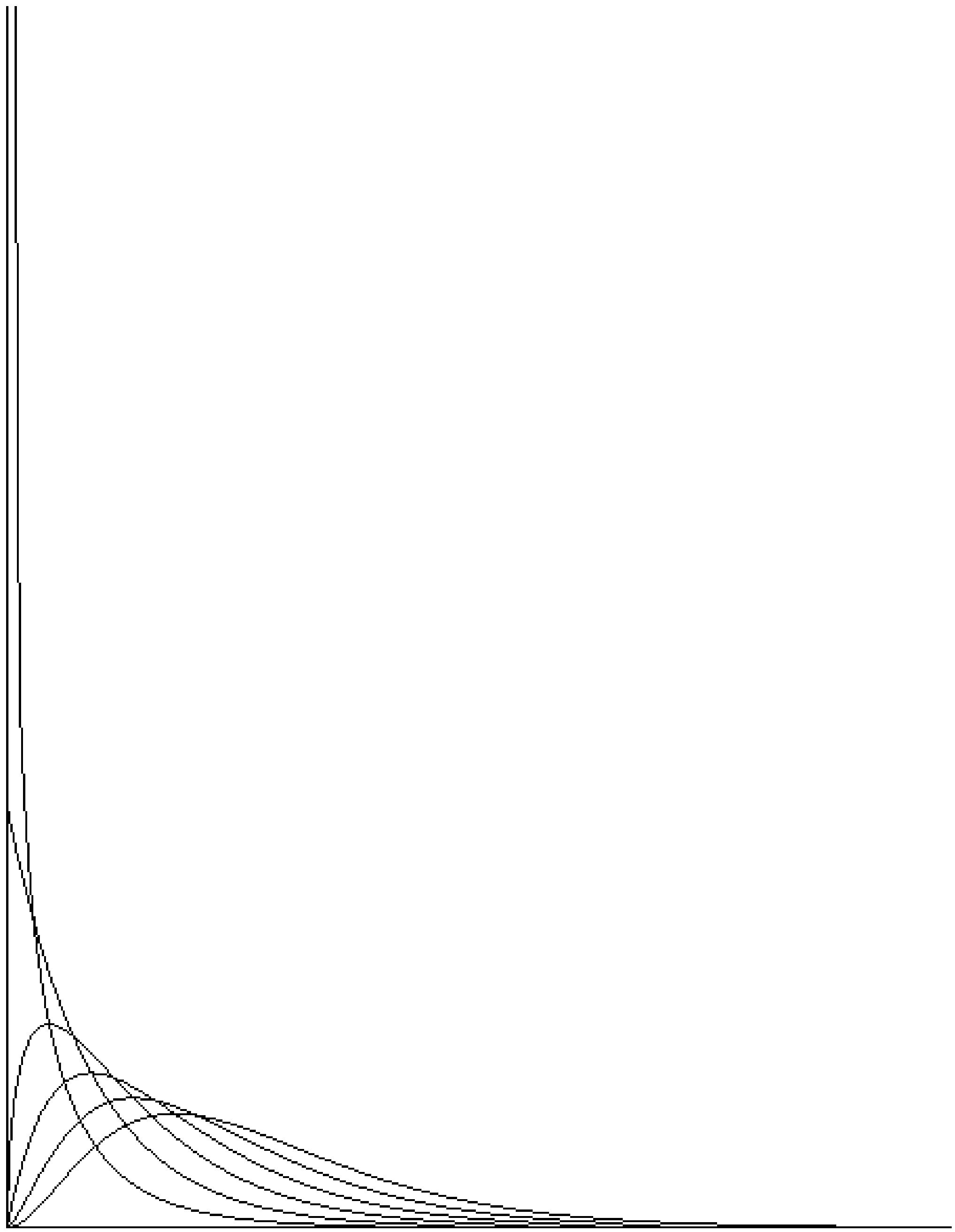


figure 46

Voici, toujours à la même échelle que sur les figures 44 et 45, les graphiques superposés des six premières densités $\chi_r^2$ ($r = 1,\ 2,\ 3,\ 4,\ 5,\ 6$).

À partir de là nous pouvons procéder comme dans le calcul précédent à une dimension : nous faisons le changement de variable $s = \rho^2$, ce qui donne

$$\mathcal{P}\left(X_1^2 + X_2^2 + X_3^2 + \cdots X_r^2 < x\right) \quad \simeq \quad \frac{1}{2^{\frac{r}{2}}\,\Gamma\!\left(\frac{r}{2}\right)} \int_0^x s^{\frac{r}{2}-1} e^{-\frac{s}{2}} ds$$

Sous cette forme on voit apparaître la densité que nous cherchions. On peut donc conclure

$$\chi_r^2(t) \quad = \quad \frac{1}{2^{\frac{r}{2}}\,\Gamma\!\left(\frac{r}{2}\right)}\, t^{\frac{r}{2}-1} e^{-\frac{t}{2}} \qquad (XI.5.)$$

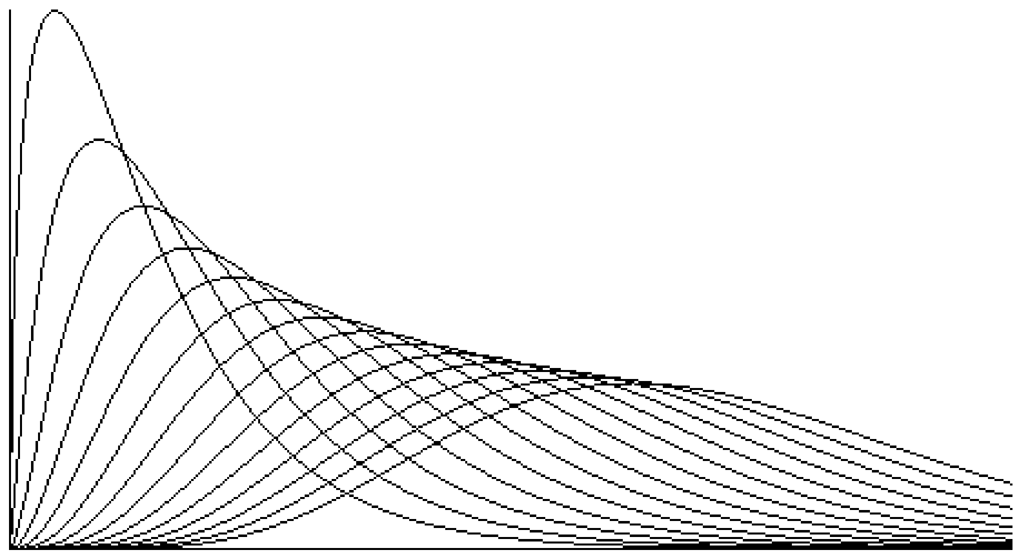


figure 47

Ici sont superposés les graphiques des densités $\chi_r^2$ pour $r = 3$ à $r = 16$. L'échelle des abscisses est la même que sur les figures 44, 45, 46, mais les ordonnées ont été agrandies $2,5$ fois.

## XI. 2. Le test du $\chi^2$

Revenons maintenant à $(XI.2.)$ On avait calculé la somme des carrés des six fluctuations du dé, et il s'agissait de déterminer la loi de probabilité de cette somme de carrés, afin de pouvoir en déduire si la valeur calculée était probable ou pas : si la valeur effectivement observée est très peu probable, on en déduit que le dé est très probablement mal équilibré ; si au contraire la valeur effectivement observée est une valeur probable, on n'a plus aucune raison de conclure que le dé est mauvais. Mais on ne peut plus utiliser la loi gaussienne pour estimer ces probabilités, puisque la somme des carrés ne suit pas une loi gaussienne. Nous allons établir maintenant que la somme des six carrés suit une loi qui *asymptotiquement* a pour densité $\chi_5^2(t)$ (et non $\chi_6^2(t)$). Mathématiquement, cela s'énonce ainsi :

**Théorème :** Soient $X_1$, $X_2$, …, $X_r$ des variables aléatoires gaussiennes normalisées (c'est-à-dire que leurs lois ont une densité égale à $(1/2\pi)\, e^{-x^2/2}$, de moyenne 0 et d'écart-type 1). Si la loi conjointe de ces variables a pour densité $(1/2\pi)^r \exp\{-(x_1^2 + x_2^2 + \cdots x_r^2)/2\}$ et si elles sont liées par la relation $X_1 + X_2 + \cdots + X_r = 0$, alors la somme de leurs carrés a pour densité $\chi^2_{r-1}(t)$.

D'après ce qui a été vu au §1 ci-dessus, on pourrait dire que si les $r = 6$ fluctuations étaient stochastiquement indépendantes les unes des autres, la somme de leurs carrés aurait pour densité (approximativement) $\chi^2_6(t)$. C'est bien sûr parce que les six fluctuations $X_1$, $X_2$, $X_3$, $X_4$, $X_5$, $X_6$ sont liées par la relation linéaire $X_1 + X_2 + X_3 + X_4 + X_5 + X_6 = 0$ que la dimension baisse ainsi d'une unité. Mais il n'est pas évident sans vérification mathématique que cette liaison linéaire ne modifie pas par exemple l'écart-type. Afin de bien comprendre ce qui se passe, il faut examiner cette affaire de plus près, et pour cela nous allons *démontrer* le théorème. Pour que cette démonstration soit plus imagée, nous gardons le cas particulier $r = 6$ du dé.

Suivre cette démonstration demande un effort : ne la lisez que si vous souhaitez vraiment comprendre *pourquoi* la relation linéaire entre les $X_j$ (qui fait baisser d'une unité le nombre de degrés de libertés) ne change pas le caractère gaussien de la densité.

Si les six fluctuations $X_1$, $X_2$, $X_3$, $X_4$, $X_5$, $X_6$ étaient stochastiquement indépendantes les unes des autres, la densité de leur loi conjointe serait

$$f_6(t_1, t_2, t_3, t_4, t_5, t_6) \;=\; \left[\frac{1}{\sqrt{2\pi}}\right]^r e^{-\frac{{t_1}^2+{t_2}^2+{t_3}^2+\cdots+{t_r}^2}{2}} \qquad (XI.6.)$$

Si elles sont liées par la relation $X_1 + X_2 + X_3 + X_4 + X_5 + X_6 = 0$, cela signifie que le point $M$ de coordonnées $X_1$, $X_2$, $X_3$, $X_4$, $X_5$, $X_6$ est situé sur l'hyperplan $H$ d'équation $X_1 + X_2 + X_3 + X_4 + X_5 + X_6 = 0$. Le fait que les valeurs des $X_j$ soient entières et que leur densité ne soit donnée par $(XI.6.)$ qu'asymptotiquement ne change rien à cela. Supposons qu'on effectue un échantillonnage (ce serait la même chose pour une convolution) des valeurs discrètes. Pour effectuer cet échantillonnage on peut par exemple quadriller l'espace à six dimensions en hypercubes ou pavés dont les arêtes sont parallèles aux six axes $OX_1$, $OX_2$, $OX_3$, $OX_4$, $OX_5$, $OX_6$, mais on peut tout aussi bien le quadriller selon six autres directions mutuellement perpendiculaires, et telles que les cinq premières soient parallèles à l'hyperplan $H$. On choisira bien sûr la taille $\varepsilon$ des pavés de telle sorte qu'ils contiennent chacun un nombre assez grand de points de la distribution discrète, tout en étant suffisamment petit pour conserver un sens à la notion de densité continue. Dans ces conditions il est bien clair que la probabilité pour que le point $M$ de coordonnées $X_1$, $X_2$, $X_3$, $X_4$, $X_5$, $X_6$ soit situé dans un pavé $P$ est approximativement donnée par la densité asymptotique continue :

$$\begin{aligned}\mathcal{P}\,(M \in P) &\simeq \int_P f_6(t_1,\, t_2,\, t_3,\, t_4,\, t_5,\, t_6)\; dt_1\, dt_2\, dt_3\, dt_4\, dt_5\, dt_6 \\ &\simeq \varepsilon^6 \cdot f(M_0) \qquad (XI.7.)\end{aligned}$$

où $M_0$ désigne par exemple le centre du pavé. Or la densité $f(M)$ ne dépend en fait que de $\rho = OM$ (la distance de $M$ à l'origine). Si on impose aux six variables $X_1$, $X_2$, $X_3$, $X_4$, $X_5$, $X_6$ d'avoir une somme nulle, cela signifie qu'on ne considère que les points $M$ situés sur l'hyperplan $H$, ou, ce qui revient au même puisque la densité $f$ n'a pas de discontinuité, les points qui sont distribués dans une bande $H_\varepsilon$ de largeur $\varepsilon$ autour de l'hyperplan (on a supposé que le quadrillage de l'espace en pavés est parallèle à l'hyperplan). Ainsi la densité de la loi conjointe des $X_j$ liées par la relation $\sum X_j = 0$ sera donnée par la répartition des probabilités selon les pavés contigus àl'hyperplan : il s'agit donc simplement des probabilités conditionnelles$\mathcal{P}\,(M \in P \mid M \in H_\varepsilon)$. Or cette probabilitéconditionnelle est, conformément à $(IV.1.)$, donnée par

$$\begin{aligned}\mathcal{P}\,(M \in P \mid M \in H_\varepsilon) \quad &= \quad \frac{\mathcal{P}\,(M \in P)}{\mathcal{P}\,(M \in H_\varepsilon)} \quad \simeq \\ \simeq \quad &\frac{\displaystyle\int_P f_6(s_1,\, s_2,\, s_3,\, s_4,\, s_5,\, s_6)\; ds_1\, ds_2\, ds_3\, ds_4\, ds_5\, ds_6}{\displaystyle\int_{0\le s_6<\varepsilon} f_6(s_1,\, s_2,\, s_3,\, s_4,\, s_5,\, s_6)\; ds_1\, ds_2\, ds_3\, ds_4\, ds_5\, ds_6} \qquad (XI.8.)\end{aligned}$$

En effet, la densité $f(M)$ ne dépendant que de $\rho = OM$ est invariante par changement de repère orthonormé, de sorte que nous pouvons intégrer par rapport aux coordonnées $s_1$, $s_2$, $s_3$, $s_4$, $s_5$, $s_6$ du nouveau repère, dans lequel l'équation de l'hyperplan $H$ est $s_6 = 0$ : la bande de largeur $\varepsilon$ autour de $H$, formée par l'ensemble des pavés contigus à $H$ (qui est le domaine de l'intégrale au dénominateur), est la région de l'espace où $0 \le s_6 < \varepsilon$. La densité $f(M)$ peut être refactorisée dans le nouveau repère. En effet :

$$\begin{aligned}f(M) &= e^{-\frac{{t_1}^2}{2}} \cdot e^{-\frac{{t_2}^2}{2}} \cdot e^{-\frac{{t_3}^2}{2}} \cdot e^{-\frac{{t_4}^2}{2}} \cdot e^{-\frac{{t_5}^2}{2}} \cdot e^{-\frac{{t_6}^2}{2}} \\ &= e^{-\frac{{t_1}^2+{t_2}^2+{t_3}^2+{t_4}^2+{t_5}^2+{t_6}^2}{2}} \\ &= e^{-\frac{{s_1}^2+{s_2}^2+{s_3}^2+{s_4}^2+{s_5}^2+{s_6}^2}{2}} \\ &= e^{-\frac{{s_1}^2}{2}} \cdot e^{-\frac{{s_2}^2}{2}} \cdot e^{-\frac{{s_3}^2}{2}} \cdot e^{-\frac{{s_4}^2}{2}} \cdot e^{-\frac{{s_5}^2}{2}} \cdot e^{-\frac{{s_6}^2}{2}}\end{aligned}$$

Si en utilisant cela on refactorise l'intégrale au dénominateur de $(XI.8.)$ et qu'on remplace celle du numérateur par son expression approchée (pour $\varepsilon$

petit) donnée par $(XI.7.)$, on obtient pour la probabilité conditionnelle

$$\mathcal{P}\,(M \in P \mid M \in H) \ \simeq\ \frac{\varepsilon^6 f_6(s_1,\, s_2,\, s_3,\, s_4,\, s_5,\, 0)}{\varepsilon \int f_6(s_1,\, s_2,\, s_3,\, s_4,\, s_5,\, 0)\, ds_1\, ds_2\, ds_3\, ds_4\, ds_5}$$

Mais $f_6(s_1,\, s_2,\, s_3,\, s_4,\, s_5,\, 0) = f_5(s_1,\, s_2,\, s_3,\, s_4,\, s_5)\,/\sqrt{2\pi}$, de sorte que finalement après simplification

$$\mathcal{P}\,(M \in P \mid M \in H) \ \simeq\ \varepsilon^5 f_5(s_1,\, s_2,\, s_3,\, s_4,\, s_5)$$

On comprend ainsi que

a) la loi conjointe des six variables $X_1$, $X_2$, $X_3$, $X_4$, $X_5$, $X_6$ liées par la relation $X_1 + X_2 + X_3 + X_4 + X_5 + X_6 = 0$ est une loi discrète de valeurs distribuées sur l'espace $\mathrm{R}^6$, mais toutes situées sur l'hyperplan $H$ ;

b) si *à l'intérieur de* $H$ on choisit un repère orthonormé, les valeurs discrètes de cette loi conjointe peuvent être repérées par les *cinq* coordonnées $s_1$, $s_2$, $s_3$, $s_4$, $s_5$, et (après échantillonnage par pavés) elles sont distribuées selon la densité

$$f_5(s_1,\, s_2,\, s_3,\, s_4,\, s_5) \ =\ e^{-\frac{{s_1}^2+{s_2}^2+{s_3}^2+{s_4}^2+{s_5}^2}{2}}$$

Or la propriété caractéristique de la fonction exponentielle (à savoir que l'exponentielle d'une somme est le produit des exponentielles) a pour conséquence que cette densité se factorise, de sorte que tout se passe comme si les valeurs discrètes prises par le vecteur aléatoire $X_1$, $X_2$, $X_3$, $X_4$, $X_5$, $X_6$, qui sont toutes situées sur l'hyperplan $H$, étaient les valeurs discrètes d'un vecteur aléatoire $S_1$, $S_2$, $S_3$, $S_4$, $S_5$ à cinq composantes *stochastiquement indépendantes* (mais dans le nouveau repère).

La démonstration du théorème s'achève ici.

On peut conclure que lorsque six variables aléatoires $X_1$, $X_2$, $X_3$, $X_4$, $X_5$, $X_6$, chacune de densité $e^{-t^2/2}$, sont liées par la relation $X_1 + X_2 + X_3 + X_4 + X_5 + X_6 = 0$, leur loi conjointe est agencée comme suit :

a) la probabilité pour que le vecteur $X_1$, $X_2$, $X_3$, $X_4$, $X_5$, $X_6$ soit en dehors de l'hyperplan est nulle ;

b) à l'intérieur de l'hyperplan elle est approximativement la même que celle de cinq variables aléatoires *stochastiquementindépendantes* et ayant chacune la densité $e^{-t^2/2}$.

Pour démontrer cela, nous avons utilisé, de façon essentielle, deux propriétés de la densité gaussienne : la première est qu'ellene dépend que de la somme des carrés des coordonnées et est donc invariante par changement de repère orthonormé (c'est cette propriété qui entraîne que la trace de $f_6$

sur l'hyperplan est $f_5$) et la seconde est que l'exponentielle d'une somme est le produit des exponentielles (c'est cette propriété qui permet la factorisation de $f_5$ en produit, prouvant ainsi l'indépendance stochastique des cinq nouvelles coordonnées). Le théorème énoncé ci-dessus n'est donc vrai qu'à la limite où les $X_j$ sont gaussiennes; il n'est qu'approximativement vrai pour la loi discrète *exacte*. Cela peut se comprendre à partir d'arguments géométriques simples: supposons pour alléger qu'on ait trois variables $X_1$, $X_2$, $X_3$au lieu de six, avec $X_1+X_2+X_3 = 0$. On peut les exprimer enfonction de deux variables indépendantes $S_1$, $S_2$, par exemple

$$\begin{aligned} X_1 &= S_1 + S_2 \\ X_2 &= S_1 - S_2 \\ X_3 &= -2S_1 \end{aligned}$$

Or si $S_1$ et $S_2$ prennent toutes les valeurs entières possibles, $X_1$ et $X_2$ auront nécessairement la même parité et $X_3$ sera toujours paire, donc on ne couvrira pas ainsi toutes les valeurs que peuvent prendre $X_1$, $X_2$, $X_3$. On pourra essayer toutes les combinaisons pour exprimer $X_1$, $X_2$, $X_3$ en fonction de $S_1$, $S_2$ (de sorte que $X_1$, $X_2$, $X_3$ prennent des valeurs entières et $X_1 + X_2 + X_3 = 0$), il sera impossible de recouvrir exactement les valeurs prises par $X_1$, $X_2$, $X_3$; soit on n'en retrouvera qu'une partie, soit on devra y ajouter des demi-entiers. Cela est dû simplement à ce que, si l'espace $\mathrm{R}^3$ est invariant par rotation, il n'en est plus de même pour une partie *discrète* de $\mathrm{R}^3$. Par contre la propriété est *approximativement* vraie lorsque les lois discrètes sont *approximativement* gaussiennes.

Par ailleurs, en y regardant de près, on comprendra également que les $X_j$ doivent avoir *le même écart-type* (ici il valait 1): en effet, si les écarts-types des $X_j$ étaient différents, il n'y aurait plus d'invariance par rotation (toutefois on aurait pu pallier cet inconvénient en choisissant dans $H$ un repère orthogonal, mais non normé).

L'importante propriété que nous venons d'établir, que la loi conjointe de $r$ variables aléatoires gaussiennes $X_1, X_2, X_3, \ldots X_r$ d'écart-type 1 et de somme nulle est également une gaussienne, mais avec un degré de liberté (une dimension) de moins, a donc pour conséquence que la somme de leurs carrés suit une loi dont la densité est $\chi^2_{r-1}$; en effet, la somme des carrés des $X_j$ est alors égale à la somme des carrés de $r - 1$ variables aléatoires, également gaussiennes et d'écart-type 1, mais stochastiquement indépendantes.

Si on veut tester un dé à l'aide de la somme $S$ des carrés des fluctuations ($XI.2.$), il faut donc utiliser la densité $\chi^2_5$; mais comme celle-ci est prévue pour des variables aléatoires d'écart-type 1, il faut encore se ramener à ce cas. En effet, les fluctuations d'un dé qu'on a jeté $6n$ fois ont un écart-type égal à $\sqrt{n}$ ; on se ramènera donc à un écart-type égal à 1 en divisant chaque fluctuation par $\sqrt{n}$. Donc la somme $S$ doit être divisée par 1000 pour qu'on puisse lui appliquer la loi $\chi_5^2$. La probabilité pour que $S$ soit supérieure à un seuil donné $x$ peut donc être calculée approximativement par

$$\mathcal{P}\left(\frac{S}{1000} > x\right) \;\simeq\; \int_x^\infty \chi_5^2(t)\,dt$$

Bien sûr l'intégrale ci-dessus ne se calcule pas par primitives, mais numériquement (il y a pour cela des logiciels ou des tables) ; dans l'exemple pris ici, $S = 6400$, donc $S/1000 = 6,4$ ; si on consulte alors une table on voit que la probabilité pour que $S$ soit supérieur ou égal à $6,4$ est environ $0,27$ ; ces fluctuations n'étant pas "peu probables" on n'est donc pas fondé à conclure que le dé est truqué (ce qui ne signifie pas qu'on est fondé à conclure que le dé n'est pas truqué). Si par contre la somme $S$ des carrés des six fluctuations avait été égale à 15 000, la même table aurait montré qu'avec un dé normal une telle amplitude n'avait qu'une chance sur cent de se produire ; alors on aurait été fondé à conclure que le dé est probablement truqué.

On voit que les opérations à faire sont simples :

a) se fixer un seuil de certitude $\alpha$, par exemple $\alpha = 0.99$ ;

b) calculer la somme $S$ des carrés des fluctuations ;

c) diviser par la variance ($n/6$ si on a lancé le dé $n$ fois) pour obtenir $x = S/(n/6)$ ;

d) chercher sur une table la valeur de $\int_0^x \chi_5^2(t)\,dt$ ; si celle-ci est supérieure à $\alpha$, on a dépassé le seuil, ce qui signifie que la probabilité pour que les fluctuations aient une telle amplitude est inférieure à $1 - \alpha$, et que donc le dé est probablement faussé.

C'est cette série d'opérations qui constitue le *test du* $\chi^2$.

Tout ce qui a été dit depuis le début de ce chapitre avait pour but de montrer que ce test du $\chi^2$ est mathématiquement équivalent à ce qu'aurait donné une analyse séparée de chacune des six fluctuations, et utilisant la loi gaussienne. En effet, on aurait pu également chercher séparément pour chacune des six fluctuations quelle probabilité elle avait de se produire (ou plutôt : quelle était la probabilité pour que la fluctuation ait une amplitude au moins égale à la fluctuation observée), puis accepter ou rejeter, selon le résultat de ces six calculs, l'hypothèse que le dé est faussé. L'avantage d'une étude séparée des six fluctuations est qu'elle permet de savoir laquelle ou lesquelles des faces du dé sont favorisées : si on souhaite disposer de cette information, il ne faut pas faire le test du $\chi^2$, mais l'analyse séparée. Par contre le test du $\chi^2$ est bien plus rapide et est donc préférable dès lors que seul importe le résultat global.

Il reste une dernière remarque à faire. Les densités $\chi_1^2(t)$ et $\chi_2^2(t)$ sont grandes au voisinage de $t = 0$, et deviennent petites pour $t$ grand, comme on peut voir sur la figure 44. Par contre on peut voir sur la figure 45 que pour $r$ grand (par exemple $r = 8$ ou $r = 16$) la densité est petite non seulement pour les grandes valeurs de $t$, mais aussi pour les petites. Cela signifie que lorsqu'il y a beaucoup de fluctuations, c'est-à-dire beaucoup de degrés de liberté, la somme $S$ des carrés de ces fluctuations a non seulement une faible probabilité d'être grande, mais aussi une faible probabilité d'être petite. Or cela se comprend aisément si on y réfléchit. En effet, $S$ est la somme des carrés, donc si $S$ est petit, c'est que *toutes* les fluctuations sont petites; la faible densité au voisinage de zéro signifie donc qu'il est peu probable que toutes les fluctuations soient petites en même temps. De façon plus quantitative: la moyenne des carrés des fluctuations est la variance $V$; il est peu probable que ces carrés deviennent beaucoup plus grands que la variance, mais il est peu probable aussi (s'ils sont nombreux) qu'ils soient *tous en même temps* nettement plus petits que la variance. Cela signifierait en quelque sorte que le hasard a *trop bien* fait les choses. Ainsi pour une pièce de monnaie qu'on lance 2000 fois, le résultat le plus probable est d'avoir 1000 fois pile (en avoir 999 est un tout petit peu moins probable). Mais si on lance 6000 fois un dé, obtenir *exactement* 1000 fois chacune des six faces n'est pas le résultat le plus probable, c'est au contraire improbable, et si au lieu de six faces, il y en avait douze, il serait encore plus improbable que toutes les douze apparaissent exactement 500 fois chacune. En calculant un peu, on peut constater que la densité $\chi_5(t)$ est maximum pour $t = 3,5$; dans l'exemple où on lance 6000 fois un dé, cela signifie que la valeur “la plus probable” de $S/1000$ est $3,5$ et non zéro.

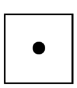
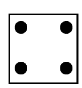
Toutefois lorsqu'on pratique le test du $\chi^2$ pour savoir par exemple si le dé est mal équilibré, on ne peut pas interpréter les faibles valeurs de $S$ de la même façon que les grandes. En effet, si le dé est mal équilibré, il sera encore moins probable d'avoir des valeurs “trop” petites pour $S$; par conséquent on ne pourra conclure que le dé est mal équilibré que si $S$ se situe dans les grandes valeurs improbables. Les valeurs trop petites de $S$ signifient que les résultats étaient *trop* exacts. Imaginons un expérimentateur qui fraude; il annonce comme résultats de mesure des résultats inventés, c'est-à-dire calculés d'après ce qu'il souhaite prouver frauduleusement. Si $V$ est alors la variance des erreurs de mesures (qui est donc un paramètre caractéristique des appareils utilisés), cet expérimentateur devra veiller à inventer également des fluctuations autour des résultats inventés, de telle sorte que la somme des carrés de ces fluctuations, divisée par $V$, corresponde à un maximum de la densité $\chi^2$ (avec le nombre convenable de degrés de liberté). S'il ne prend pas cette précaution, les experts appelés par des collègues soupçonneux ne manqueront pas d'invoquer un test du $\chi^2$ à sa charge. Dans ce cas, le test du $\chi^2$ portera sur les petites valeurs de $S$ et non sur les grandes, et servira à conclure qu'une accumulation de résultats trop exacts est anormale.

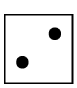
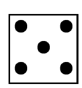
On pourra donc retenir en conclusion que le test du $\chi^2$ peut servir à estimer si des fluctuations sont trop grandes, mais aussi trop petites. Bien entendu l'interprétation du test ne sera pas du tout la même dans les deux cas. En revanche le principe du test est toujours le même ; on teste une hypothèse. Le raisonnement est le suivant : si l'hypothèse est vraie, le résultat observé est improbable, donc on conclut que, l'événement improbable *ayant eu lieu*, c'est l'hypothèse qui est probablement fausse ; ou inversement : si le test conclut que le résultat observé était probable, il n'y a aucune incompatibilité entre l'hypothèse et l'observation. Un tel test permet une quantification précise et objective des seuils de certitude, mais cela ne protège évidemment pas contre la légèreté dans l'interprétation.

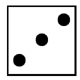
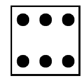
### XI. 3. Comment tester une hypothèse.

Toute la discussion de la section précédente concerne un exemple (le dé). Mais la méthode du test peut s'appliquer de manièrebien plus générale.

Il se trouve que l'hypothèse “le dé est correctement équilibré” se traduit mathématiquement par “les probabilités des six faces sont toutes égales à $\frac{1}{6}$”. Autrement dit, l'hypothèse se traduit mathématiquement par la donnée d'une loi de probabilité. Imaginons qu'après un test négatif, c'est-à-dire un test ayant donné un résultat très peu probable dans l'hypothèse où le dé est correctement équilibré, nous souhaitions trouver la loi de probabilité réelle (et non uniforme) du dé. Le test ayant consisté à lancer 6 000 fois le dé, on peut prendre les fréquences relatives de chaque face comme un première indication. Mettons pour fixer les idées que les résultats obtenus, au lieu d'être ceux de $(XI.1.)$ soient les suivants

 980 fois  993 fois

 812 fois  1133 fois $(XI.9.)$

 1019 fois  1063 fois

Il saute aux yeux que les résultats anormaux sont le deux qui ne sort que 812 fois et le cinq qui sort 1133 fois. À l'aide de la loi gaussienne (cette fois il est préférable de ne pas utiliser le test du $\chi^2$, puisqu'on veut analyser les résultats pour la face deux et la face cinq, et non globalement) on peut voir que la probabilité d'avoir plus de 1100 fois le cinq sur 6000 lancers (dans l'hypothèse d'un dé correctement équilibré) est 0.0008 ; la probabilité d'avoir moins de 850 fois le deux est inférieure à $10^{-4}$. Comme le premier coup d'oeil le laissait pressentir, le test est donc négatif, et on est conduit à conclure que la probabilité du cinq est en réalité supérieure à $\frac{1}{6}$ tandis que la probabilité du deux est inférieure à $\frac{1}{6}$. On peut évaluer ces probabilités par les fréquences observées dans $(XI.9.)$ qui seraient donc

$$\frac{980}{6000}\,, \quad \frac{812}{6000}\,, \quad \frac{1019}{6000}\,, \quad \frac{993}{6000}\,, \quad \frac{1133}{6000}\,, \quad \frac{1063}{6000}\,,$$

mais il serait absurde de les évaluer avec une telle précision puisque nous savons que les fluctuations normales autour deces valeurs sont de l'ordre de $\sqrt{1000}/6000 \simeq 0.005$. Donc le résultat observé pour le un, à savoir 980/6000, ne doit pas être tenu pour significativement différent de 1/6. On peut alors faire la *nouvelle* hypothèse que les probabilités sont $\frac{1}{6}\,,\ 0.13\,,\ \frac{1}{6}\,,\ \frac{1}{6}\,,\ 0.20\,,\ 0.17$. Telles qu'elles sont choisies, ces valeurs donneront évidemment un test positif (les fluctuations autour de ces valeurs sont petites puisqu'elles ont été choisies pour cela). Toutefois il serait prudent de tester la nouvelle hypothèse indépendamment en lançant à nouveau le dé 6000 fois (cela coûte évidemment du temps et de l'argent, mais comme toutes les bonnes choses, la certitude se paie).

Si avec la nouvelle hypothèse on veut effectuer le second test pour la somme des carrés (c'est-à-dire globalement et non pour une ou plusieurs faces particulières), il faudra tenir compte d'un fait nouveau par rapport à la loi uniforme : les fluctuations $X_j$ n'ont pas toutes le même écart-type ; en effet si les probabilités ne sont pas toutes les six égales à $\frac{1}{6}$, mais ont des valeurs différentes $p_j$, les écarts-types correspondants vaudront $\sqrt{6000\,p_j}$ au lieu de $\sqrt{1000}$. Dans ce cas il serait erroné de calculer la somme des carrés des fluctuations, de la diviser par 1000, puis de la situer par rapport à la répartition $\chi^2_5$, car la densité $\chi^2_5$ est celle de la somme de six fluctuations (de somme nulle) ayant *chacune* un écart-type égal à 1. La loi de la variable aléatoire “somme des carrés” n'a pas pour densité $\chi_5^2$ et il convient de la remplacer par la variable aléatoire

$$S = \Big(\frac{X_1}{\sigma_1}\Big)^2 + \Big(\frac{X_2}{\sigma_2}\Big)^2 + \Big(\frac{X_3}{\sigma_3}\Big)^2 + \Big(\frac{X_4}{\sigma_4}\Big)^2 + \Big(\frac{X_5}{\sigma_5}\Big)^2 + \Big(\frac{X_6}{\sigma_6}\Big)^2$$

qui est la somme des carrés des variables aléatoires $X'_j = X_1/\sigma_1$, qui, elles, ont bien un écart-type égal à 1. On pourrait objecter ici que les variables aléatoires $X'_j$ ont certes chacune un écart-type égal à 1, mais que leur somme n'est pas nulle, car c'est la somme des $X_j$ qui est nulle. Or d'après le théorème énoncé à la section **2**, si six variables aléatoires ont chacune unécart-type égal à 1 et une somme nulle, alors la somme de leurscarrés a une loi de densité $\chi_5^2$. Cela résultait de lapropriété typique que la densité gaussienne à six dimensions, $\exp\{-\frac{1}{2}(x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2 + x_6^2)\}$, est invariante par rotation et que sa trace sur un sous-espace de dimension cinq est une densité gaussienne à cinq dimensions. Mais cela aurait été vrai pour *n'importe quel* sous-espacede dimension cinq, à cause justement de l'invariance par rotation. Ce serait donc tout aussi vrai pour l'hyperplan d'équation$\sigma_1 X'_1 + \sigma_2 X'_2 + \sigma_3 X'_3 + \sigma_4 X'_4 + \sigma_5 X'_5 + \sigma_6 X'_6 = 0$ que pour l'hyperplan d'équation $X'_1 + X'_2 + X'_3 + X'_4 + X'_5 + X'_6 = 0$. Donc malgré la modification, il demeure que la loi de la variable aléatoire $S = \sum {X'}_j^2$ a bien la densité $\chi_5^2$.

Si on refait alors le test avec la nouvelle hypothèse que les six faces ne sont pas équiprobables, mais ont les probabilités $p_1 = \frac{1}{6}$, $p_2 = 0.13$, $p_3 = \frac{1}{6}$, $p_4 = \frac{1}{6}$, $p_5 = 0.20$, $p_6 = 0.17$, et qu'on trouve sur $N$ lancers $N_1$ fois la face 1, $N_2$ fois la face 2, etc. on calculera la grandeur

$$\begin{aligned} S \;=\; & \frac{(N_1 - p_1 N)^2}{p_1 N} + \frac{(N_2 - p_2 N)^2}{p_2 N} + \frac{(N_3 - p_3 N)^2}{p_3 N} + \frac{(N_4 - p_4 N)^2}{p_4 N} + \\ & + \frac{(N_5 - p_5 N)^2}{p_5 N} + \frac{(N_6 - p_6 N)^2}{p_6 N} \qquad (XI.10) \end{aligned}$$

C'est donc sur ce modèle que sont pratiqués les tests d'hypothèse dits “du $\chi^2$”. L'hypothèse est une loi de probabilité qu'on postule a priori soit parce qu'on y est conduit par des considérations théoriques (par exemple que les six faces d'un dé sont équiprobables), soit par une observation préalable, soit par un cahier des charges (si on veut détecter des défauts sur une chaîne de fabrication) ou pour toute autre raison. Ce que nous venons de voir montre que l'hypothèse faite ne comprend pas seulement la loi de probabilité, mais également les écarts-types (ou les variances) des fluctuations. En effet, le calcul de la grandeur $S$ dans $(XI.10)$ fait appel non seulement à la donnée des probabilités $p_j$, mais *aussi* à la donnée des écarts-types $\sigma_j$ des fluctuations. On n'est donc pas en mesure d'appliquer un test du $\chi^2$ si on n'a pas une idée préconçue sur l'écart-type de chaque fluctuation. On peut très bien avoir postulé une loi de probabilité $p_j$ correcte, et cependant avoir un test négatif parce qu'on s'est trompé sur les écarts-types. Les écarts-types retenus dans $(XI.10)$ sont ceux de la loi multinômiale ("sextinômiale" pour être précis), puisque c'est cette loi qui régit les lancers de dés. Donc on ne connaît pas a priori les probabilités $p_j$ car les dés ont un défaut non contrôlé d'équilibrage, mais on est sûr dès le départ que les résultats d'un grand nombre de lancers sont soumis à la loi multinômiale, et les écarts-types sont alors une fonction connue des $p_j$. Il suffit de faire une hypothèse sur les $p_j$ et de calculer ensuite $S$ selon $(XI.10)$.

Toutes les explications précédentes ont été développées à propos d'une expérience de dé qu'on lance six mille fois. Il est bien entendu que cette expérience de pensée n'était ici qu'un support didactique, permettant de mieux comprendre la *signification* du test.

En pratique le test du $\chi^2$ est souvent utilisé dans des situations où la nature gaussienne des fluctuations n'est pas assurée à l'avance avec autant de rigueur. Toutefois l'expérience de pensée du dé reste un *modèle* ou un *idéal*. Il représente l'*esprit des sciences exactes* dans des contextes expérimentaux où cet esprit n'est qu'un idéal inaccessible.

Voici un type d'expérience très simple sur des souris qui est typique des applications possibles de tests statistiques en biologie ou psychologie expérimentale. Quoique de nos jours la sophistication des méthodes statistiques dépasse presque toujours celle de cet exemple scolaire, il n'y a rien de nouveau quant au principe. Nous discuterons cette expérience afin d'en montrer les limites.

La figure 48 montre un dispositif destiné à étudier l'apprentissage (ou toute autre faculté, sens de l'orientation inné, etc.) chez les souris. Sur une arène centrale débouchent $n$ couloirs (ici $n = 8$) absolument identiques, numérotés de 1 à 8 sur la figure. On veut tester si les souris sont capables de reconnaître le numéro en leur offrant par exemple, pendant une période d'apprentissage, dix fois de suite de la nourriture dans le couloir $N°7$. Après cette période d'apprentissage, on veut savoir si les souris se dirigeront préférentiellement du premier coup dans le couloir $N°7$, ou si au contraire, comme la toute première fois, elles choisiront au hasard ("au hasard" voulant dire que les $n$ couloirs sont équiprobables). On peut aussi tester leur sens de l'orientation en plaçant la nourriture dans un couloir dont le numéro change aléatoirement, mais qui reste toujours celui qui va vers le nord. Avec ce procédé, on peut tester toutes sortes de facultés sensorielles.

Après la période d'apprentissage, on place successivement disons deux

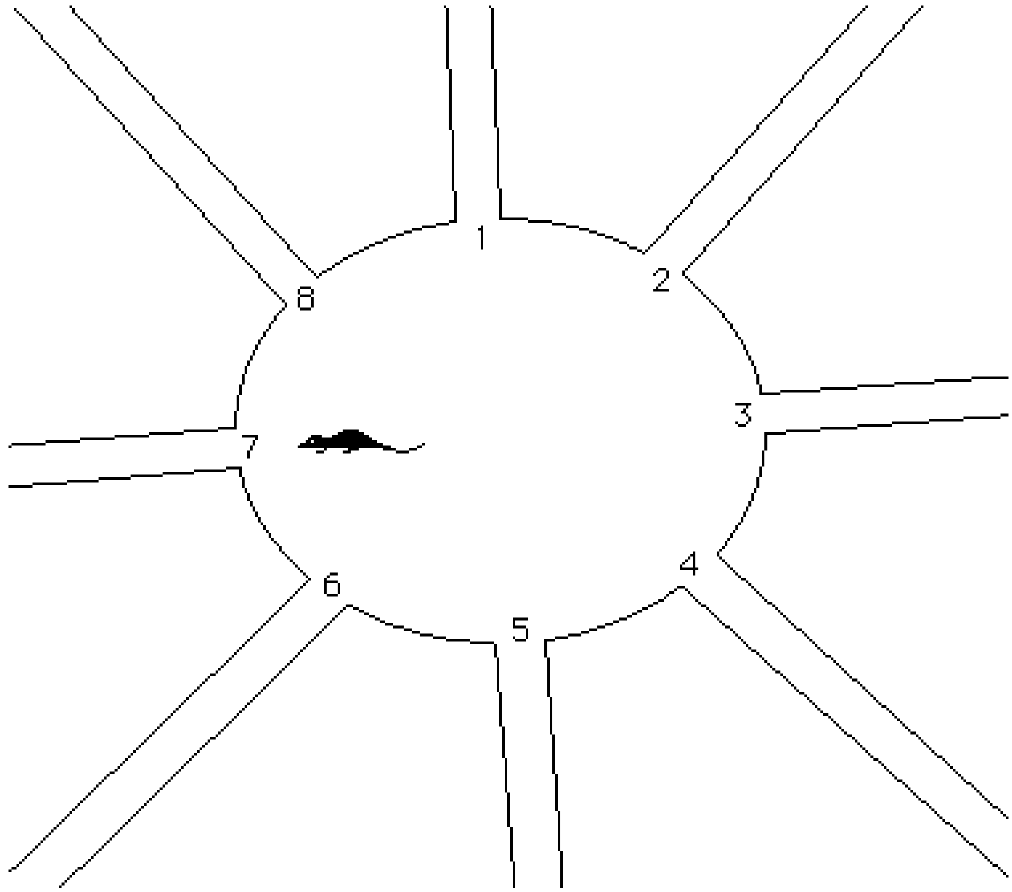



figure 48

cents souris dans l'arène et on répète exactement les conditions qui ont régné durant l'apprentissage ; par exemple, si l'arrivée de nourriture était annoncée par un son, on produira ce même son. Puis on note pour chaque souris le numéro du couloir qu'elle choisit *en premier*. On obtient alors une distribution statistique : pour chacun des huit couloirs, le nombre de souris l'ayant choisi.

Si toutes les deux cents souris (ou presque, par exemple cent quatre-vingt-dix-sept) choisissent le bon tunnel, point n'est besoin de test et la conclusion est claire. Mais en général, dans ce genre d'expériences, les résultats sont loin d'être aussi tranchés. Il faudra alors savoir si une légère préférence pour le bon couloir peut être l'effet du hasard (donc une fluctuation normale), ou bien si, sous l'hypothèse de l'équiprobabilité des huit choix, il est peu vraisemblable que la préférence soit due au hasard.

Il y a une analogie avec le problème longuement discuté ci-dessus du dé. Toutefois des différences sautent aux yeux : avec le dé on savait que les fluctuations seraient gaussiennes et on en connaissait les écarts-types, car on avait la loi binômiale qui résultait des symétries du dé. On ne peut avoir de telles certitudes avec des organismes vivants. Il était raisonnable d'admettre que les lancers successifs du dé étaient stochastiquement indépendants : le dé n'est pas influencé par ce qui s'est passé avant. Mais qui nous prouve que le couloir choisi par la première souris ne garde pas une trace (une odeur par exemple) qui poussera les souris suivantes à emprunter le même couloir ? Si on *veut* tenir compte d'un tel effet, on *peut* améliorer le protocole de l'expérience pour éliminer ce facteur : par exemple permuter aléatoirement les numéros des couloirs, de sorte que la nourriture ne soit pas toujours dans le même couloir, mais toujours au $N°7$. Cependant on n'est jamais à l'abri de facteurs insoupçonnés ; on aura pensé à l'odorat, mais on ne sait pas tout sur la vie des souris.

Une autre différence est aussi le nombre : en répétant six mille fois, les résultats sont bien plus concluants qu'avec deux cents. Mais on peut aussi y remédier en prenant six mille souris. C'est le sens des expériences médico-pharmacologiques à très grande échelle, comme celle qui est citée du journal *The Lancet* au chapitre **X**.

Une différence plus subtile est la suivante : supposons que l'on ait utilisé six mille souris, ou même un million si vous voulez ; le test du $\chi^2$ suppose que, comme pour le dé, les écarts-types des huit fluctuations soient ceux de la loi multinômiale. Cela est nécessairement le cas si l'expérience est *parfaitement reproductible* : quelles que soient les probabilités a priori $p_j$ pour que chaque souris prenne le couloir $N°j$, la répétition de l'expérience conduira forcément à la loi multinômiale et l'expression $XI.10$ (pour huit possibilités au lieu de six) suivra la loi $\chi^2$ à sept degrés de liberté. Mais l'expérience est-elle exactement reproductible ? On retrouve ici la discussion du chapitre **X**. Des êtres vivants aussi évolués que les souris ne pourraient-ils pas être influencés par des détails perceptifs insignifiants pour l'expérimentateur, qui conduiraient à augmenter l'écart-type du quatrième couloir et à reserrer celui du sixième, par exemple ? Imaginons que pendant l'apprentissage le couloir $N°7$ où était disposée la nourriture était trop souvent au nord, l'expérimentateur *n'ayant pas songé* à ce détail. L'expérimentateur croit que seul le numéro distingue les couloirs. Mais essayez de vous mettre à la place d'une de ces souris (elles sont des mammifères comme vous et moi, et ont une certaine forme de pensée) : elles ont faim, et pour manger doivent deviner le bon couloir. Pourquoi penseraient-elles d'abord au numéro, qui pour elles est bien plus abstrait et éloigné de leur mode de perception que par exemple l'orientation (peu importe ici de savoir comment elles arrivent à s'orienter). On observerait alors des effets différents selon que les deux critères, le numéro *et* l'orientation, se conjuguent ou s'opposent. Si les deux critères se conjuguent, les deux catégories de souris, d'une part les intellectuelles qui pensent d'abord au numéro, et d'autre part les instinctives qui pensent d'abord à l'orientation, vont agir de même et "préférer" le couloir $N°7$. Si au contraire les deux critères s'opposent, les souris ayant d'abord pensé à l'orientation choisiront un autre couloir que le $N°7$ et le résultat statistique sera une augmentation de l'écart-type.

Si l'expérimentateur pense à randomiser soigneusement l'orientation, il évitera l'artefact correspondant. Mais nos connaissances sur les souris sont beaucoup moins complètes que nos connaissances sur les dés. Combien restera-t-il de facteurs ignorés ? Nul ne le sait. Donc l'hypothèse que les huit écarts-types sont ceux de la loi multinômiale est peut-être systématiquement fausse.

Les mammifères sont plus proches de l'homme que du dé. On peut mieux comprendre les souris en mettant un enfant humain à leur place qu'en mettant un dé à leur place. C'est pourquoi on pourra méditer l'exemple suivant [(1)]. Des élèves du primaire sont confrontés au problème mathématique : "calculer $(10 \times 7)$, puis $(3 \times 7)$ ; en déduire $13 \times 7$." Un des élèves répond respectivement "70, 21, et 0". Le maître *n'avait pas pensé* que l'expression "en déduire" avait deux sens différents en français et conclut que l'élève n'a pas compris. Si l'énoncé précédent figure dans un questionnaire dont les résultats seront traités statistiquement, le mot ambigu aura pour effet d'augmenter l'écart-type des notes (et de baisser leur moyenne). On peut réaliser des expériences consistant à donner des énoncés différents du même problème mathématique ; s'ils sont donnés successivement aux mêmes élèves, ceux-ci ayant en mémoire les précédents énoncés ne répondront pas indépendamment ; si on change la population, on introduit une nouvelle incertitude (les nouveaux élèves, venant peut-être d'un milieu social différent, n'ayant pas forcément la même perception de la langue que les précédents).

Lorsqu'on publie des résultats expérimentaux sur les animaux de laboratoire, les revues spécialisées exigent des tests quantitatifs. La raison en est que même avec les limites indiquées ci-dessus, la présence de statistiques quantifiées dans une publication scientifique permet une comparaison plus objective : même si le test du $\chi^2$ est fallacieux pour une des raisons indiquées, il permet dans une certaine mesure de comparer des résultats obtenus par une équipe japonaise à ceux obtenus par une équipe danoise.

(1) Exemple emprunté à Stella Baruk, mais je ne sais plus dans quel livre.

Une telle comparaison des résultats est relativement sensée, plus parce que les erreurs commises dans la manipulation du test sont partout les mêmes (par exemple, que les “facteurs ignorés” sont les mêmes pour l'équipe japonaise que pour l'équipe danoise), que parce que le test est bien appliqué dans son principe.

### XI.4. Le test de Student.

Le test du $\chi^2$, comme nous avons vu, convient pour tester l'hypothèse que plusieurs possibilités sont équiprobables, ou plus généralement que ces possibilités obéissent à une loi de probabilité supposée (et qu'il s'agit justement de tester), mais à condition de connaître les écarts-types.

Le test dit *de Student* (inventé en fait par le statisticien anglais GOSSET $\sim$ 1900) que nous allons étudier maintenant, est généralement utilisé pour tester des moyennes empiriques, sans connaître l'écart-type.

Afin de bien comprendre l'*idée* du test de Student, imaginons une expérience reproductible modélisée par une variable aléatoire $X$ dont la loi est approximativement gaussienne. Pour fixer les idées on peut imaginer que $X$ représente les fluctuations de la mesure d'une grandeur. Si on mesure un très grand nombre de fois une grandeur (supposée évidemment reproductible), on retient généralement la moyenne empirique des résultats comme étant la vraie valeur ; mais supposons que, par exemple à la suite d'une prédiction théorique, on ait déjà une bonne raison de penser que la valeur de la grandeur soit $m$. La moyenne empirique $M$ obtenue par les mesures ne sera pas forcément égale à $m$, mais comment savoir si la différence est significative et n'est pas une fluctuation due au hasard ? Dans le test du $\chi^2$ on voulait tester si (un dé étant lancé six mille fois, ou des souris étant lâchées huit cents fois) la différence entre les probabilités mesurées et les probabilités supposées est due au hasard ou non. Maintenant nous voudrions tester si la différence entre la moyenne mesurée et la moyenne supposée est due au hasard ou non. On ne peut pas utiliser le test du $\chi^2$ pour cela. Ou plutôt, on pourrait l'utiliser si on connaissait l'écart-type des erreurs de mesure (c'est-à-dire l'écart-type de la variable aléatoire $X$) : il suffirait alors de calculer la somme normalisée des carrés des écarts par rapport à $m$. En effet, une valeur fausse de $m$ rendrait cette somme trop grande et le test du $\chi^2$ serait négatif. Mais pour calculer la somme *normalisée* des carrés des écarts, il faut diviser par l'écart-type ; si celui-ci n'est pas connu, on ne peut pas procéder ainsi.

Dans le cas d'une expérience reproductible avec un nombre plutôt petit de possibilités (six pour les dés, huit pour les souris, etc.) on pouvait déduire l'écart-type de la loi multinômiale (cf. $IX.10$). Mais ici nous avons des erreurs de mesure dont nous ne connaissons pas l'origine ; il est raisonnable de les supposer gaussiennes, mais rien de raisonnable ne permet de deviner leur écart-type. Il faut donc un test indépendant de cet écart-type. C'est le test de Student.

Pour l'analyse mathématique du problème, appelons $m$ la moyenne de $X$ et $\sigma$ son écart-type. Cela signifie qu'en répétant l'expérience on obtient des résultats $x_1, \ x_2, \ldots x_n$ qui ont chaque fois une probabilité a priori $\mathcal{P}\,(X = x_j)$ de se produire, ou, en termes de densité, que la probabilité pour que $x_j$ soit compris entre $x_j - \varepsilon$ et $x_j + \varepsilon$ est égale à

$$\mathcal{P}\,(x_j - \varepsilon < X < x_j + \varepsilon) \;\; \simeq \;\; \frac{1}{\sigma\sqrt{2\pi}} \int_{x_j-\varepsilon}^{x_j+\varepsilon} e^{-\frac{(x-m)^2}{2\sigma^2}} \, dx$$

La répétition à l'identique de l'expérience implique que les résultats successifs $x_j$ sont stochastiquement indépendants et tous distribués selon la loi de $X$ ; autrement dit, les variables aléatoires $(x_j - m)/\sigma$ sont indépendantes et de densité $e^{-x^2/2}$. Par conséquent la somme de leurs carrés, soit

$$S_0 = \sum_{j=1}^{j=n} \frac{(x_j - m)^2}{\sigma^2}$$

obéit à la loi du $\chi^2$ à $n$ degrés de libertés. Mais ici nous ne connaissons pas forcément $\sigma$, qui est l'écart-type *théorique*.

Remarquons à ce propos que, si pour le test du $\chi^2$ la connaissance a priori de l'écart-type est nécessaire, celle de la moyenne théorique en revanche ne l'est pas. Lorsque la moyenne théorique $m$ est inconnue, on peut la remplacer par la moyenne *empirique* $M = (x_1 + x_2 + \cdots + x_n)/n$. Au lieu de calculer la somme $S_0$, on calculera

$$S_1 = \sum_{j=1}^{j=n} \frac{(x_j - M)^2}{\sigma^2}$$

La différence essentielle avec $S_0$ est que les variables aléatoires $x_j - M$ ne sont plus stochastiquement indépendantes, puisque $M$ dépend de chacune d'elles. Toutefois il est clair que la somme des $x_j - M$ est nulle ; on est donc dans la situation du théorème de la section précédente **IX.2** : les variables aléatoires $x_j$ sont indépendantes, mais la somme des $x_j - M$ est nulle. D'après le théorème, la somme $S_1$ obéit à la loi $\chi^2$, mais avec $n - 1$ degrés de liberté. La contrainte que la somme est nulle fait baisser d'une unité le nombre de degrés de liberté. Que cet abaissement ne modifie pas le caractère gaussien de la loi est une propriété spéciale de la loi gaussienne, voir **XI.2**. Toujours est-il que la somme *normalisée* $S_0$ ou $S_1$ des carrés des écarts par rapport à $m$ ou $M$ ne peut être calculée à partir des mesures $x_1,\ x_2, \ldots x_n$ que si on connaît a priori l'écart-type $\sigma$.

Pour faire disparaître le paramètre $\sigma$, l'idée du test de Student consiste alors à considérer la variance empirique [1]

$$V_{\mathrm{emp}} \quad = \quad \frac{1}{n-1}\sum_{j=1}^{j=n}(x_j - M)^2$$

et d'étudier la loi de probabilité (ou plutôt la densité) du quotient

$$S \quad = \quad \frac{1}{\sqrt{V_{\mathrm{emp}}}}\sum_{j=1}^{j=n}(x_j - m)$$

Pour calculer la loi de ce quotient nous commençons par chercher celle de la grandeur $V_{\mathrm{emp}}/\sigma^2$. Chacune des valeurs $x_j$ peut fluctuer autour de sa moyenne $m$ selon une loi gaussienne d'écart-type inconnu $\sigma$. Donc la grandeur $V_{\mathrm{emp}}/\sigma^2$, qui est la somme *normalisée* des carrés des écarts par rapport à $M$, obéit à la loi du $\chi^2$ à $n-1$ degrés de liberté, qui ne fait pas intervenir $\sigma$. D'autre part la grandeur $\frac{1}{\sigma}\sum_j (x_j - m)$, qui est la somme normalisée des écarts par rapport à $m$, obéit à la loi gaussienne centrée et normalisée à $n$ degrés de liberté $\frac{1}{2\pi}\exp\{-(x_1^2 + x_2^2 + \cdots + x_n^2)/2\}$ qui ne fait pas non plus intervenir $\sigma$. Or $S$ est le quotient des deux, dans lequel les $\sigma$ inconnus se simplifient :

$$\begin{aligned} S \quad &= \quad \frac{1}{\sqrt{V_{\mathrm{emp}}}}\sum_{j=1}^{j=n}(x_j - m) \\ &= \quad \frac{1}{\sqrt{V_{\mathrm{emp}}/\sigma^2}} \quad \cdot \quad \frac{\displaystyle\sum_{j=1}^{j=n}(x_j - m)}{\sigma} \end{aligned}$$

On peut donc conclure que la loi de probabilité de $S$ est celle d'un quotient de la forme $Y/\sqrt{Z}$ où $Y$ est une variable aléatoire qui suit la loi gaussienne centrée et normalisée $\frac{1}{2\pi}\, e^{-x^2/2}$ et $Z$ une autre variable aléatoire, indépendante de $Y$, qui suit la loi $\chi^2$ à $n-1$ degrés de libertés. Aucune de ces deux lois ne fait intervenir $\sigma$, donc la loi de $S$ est indifférente à la quantité inconnue $\sigma$. C'est là le principe du test de Student.

---

(1) on notera que dans l'expression de $V_{\mathrm{emp}}$ la somme des carrés des écarts est divisée par $n-1$ et non par $n$. La raison de ce choix est expliquée au chapitre suivant (**XII.2**).

Il ne reste plus qu'à calculer la densité de Student elle-même. Nous procédons comme au chapitre **IX** pour calculer la densité de Cauchy ou au paragraphe **XI.1** pour calculer celle du $\chi^2$. La densité sera en effet donnée par les probabilités $\mathcal{P}\,(a < S < b)$ lorsqu'on les aura exprimées sous la forme d'une intégrale entre les bornes $a$ et $b$. Puisque $S = Y/\sqrt{Z}$ l'événement $\{a < S < b\}$ est identique à l'événement $\{a\sqrt{Z} < Y < b\sqrt{Z}\}$, et on peut écrire

$$\begin{aligned}
\mathcal{P}\,(a < S < b) \quad &\simeq \quad \int\!\!\!\int_{a\sqrt{z}<y<b\sqrt{z}} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{y^2}{2}} \chi^2_{n-1}(z)\, dy\, dz \\
&= \quad \frac{1}{\sqrt{2\pi}}\, \frac{1}{2^{\frac{r}{2}}\Gamma(\frac{r}{2})} \int\!\!\!\int_{a\sqrt{z}<y<b\sqrt{z}} e^{-\frac{y^2}{2}} z^{\frac{r}{2}-1} e^{-\frac{z}{2}}\, dy\, dz
\end{aligned}$$

Nous avons simplement reporté ici l'expression $XI.5$ de la densité $\chi^2_r$ à $r$ degrés de liberté. Dans le présent contexte, $r = n-1$, mais nous oublions l'entier $n$ pour la durée du calcul.

Effectuons maintenant le changement de variable $s = y/\sqrt{z}$, de manière à passer des variables $y, z$ aux variables $s, z$. Le jacobien de la transformation est $\sqrt{z}$, de sorte que notre intégrale devient

$$\begin{aligned}
\mathcal{P}\,(a < S < b) \quad &\simeq \quad \frac{1}{\sqrt{2\pi}}\, \frac{1}{2^{\frac{r}{2}}\Gamma(\frac{r}{2})} \int\!\!\!\int_{a<s<b} e^{-\frac{zs^2}{2}} z^{\frac{r}{2}} e^{-\frac{z}{2}}\, ds\, dz \\
&= \quad \frac{1}{\sqrt{2\pi}}\, \frac{1}{2^{\frac{r}{2}}\Gamma(\frac{r}{2})} \int_a^b \left\{ \int_0^\infty e^{-\frac{s^2+1}{2}z}\, z^{\frac{r-1}{2}}\, dz \right\} ds
\end{aligned}$$

Ceci montre déjà que la densité cherchée sera l'expression entre accolades (multipliée par le coefficient), qui est une intégrale eulérienne connue. En effet, puisque

$$\Gamma(x) \quad = \quad \int_0^\infty e^{-z} z^{x-1} dz$$

on voit que notre densité sera

$$f(s) \quad = \quad \frac{1}{\sqrt{2\pi}}\, \frac{1}{2^{\frac{r}{2}}\Gamma(\frac{r}{2})} \int_0^\infty e^{-\frac{s^2+1}{2}z^2} z^{\frac{r-1}{2}} dz \quad = \quad \frac{1}{\sqrt{2\pi}}\, \frac{1}{2^{\frac{r}{2}}\Gamma(\frac{r}{2})}\, \frac{\Gamma(\frac{r+1}{2})}{(\frac{s^2+1}{2})^{\frac{r+1}{2}}}$$

En regroupant toutes les puissances de 2 éparpillées dans cette dernière expression, on peut simplifier : $\sqrt{2}\cdot 2^{\frac{r}{2}} \cdot 2^{-\frac{r+1}{2}} = 1$ ; d'où

$$f(s) \quad = \quad \frac{1}{\sqrt{\pi}}\, \frac{1}{\Gamma(\frac{r}{2})}\, \frac{\Gamma(\frac{r+1}{2})}{(s^2+1)^{\frac{r+1}{2}}} \qquad (XI.11)$$

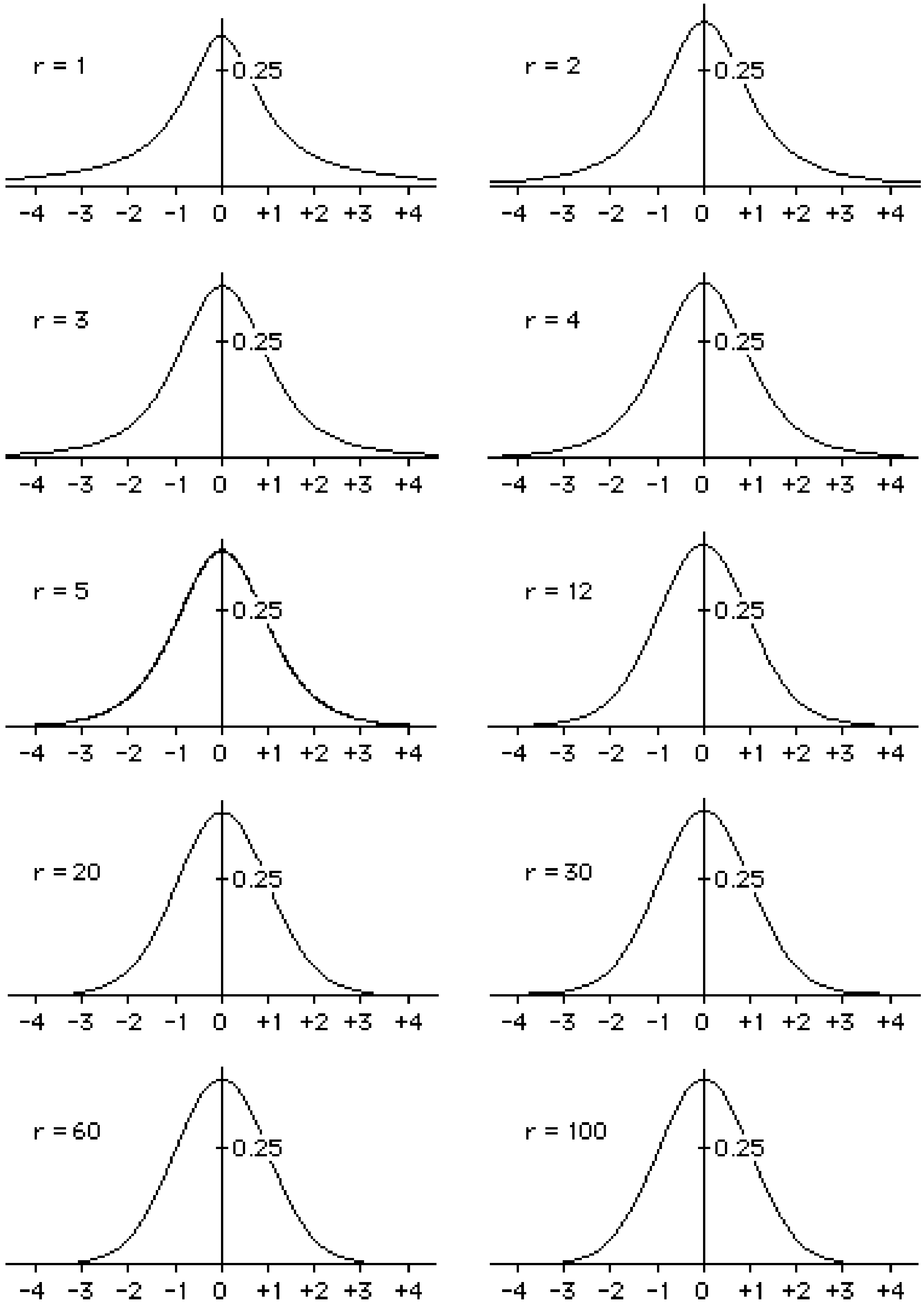



figure 49 : Les graphiques ci-dessus sont ceux des densités de Student à $r$ degrés de liberté pour $r = 1, 2, 3, 4, 5, 12, 20, 30, 60, 100$. Les derniers ne se distinguent pratiquement plus de leur limite gaussienne.

Rappelons que dans ces calculs $r = n-1$. La variable $Z$ est ici la somme des carrés des écarts par rapport à la moyenne théorique $m$ supposée connue. Mais la variance empirique est égale à $Z/(n-1)$ (cf. chapitre suivant $XII.6$). Pour *pratiquer* le test de Student on peut indifféremment considérer $Y/\sqrt{Z}$ ou bien $Y/\sqrt{Z/(n-1)} = \sqrt{n-1} \cdot Y/\sqrt{Z}$. Toutefois la convention communément admise est de considérer $Y/\sqrt{Z/(n-1)}$. La loi de Student (ou plutôt la *densité* de Student) est celle de $Y/\sqrt{Z/(n-1)}$ et non celle de $Y/\sqrt{Z}$. Ce n'est qu'une convention, mais étant donné que tous les logiciels de statistique la respectent, on doit la respecter aussi si on veut employer ces derniers tels qu'ils sont livrés au client. La relation entre les deux densités est simple : c'est la relation entre les densités de deux variables aléatoires proportionnelles : si $f(t)$ est la densité d'une variable aléatoire $X$, et $g(t)$ celle de $\alpha X$, on a bien sûr

$$\mathcal{P}\,(a < X < b) \;\simeq\; \int_a^b f(t)\,dt$$

d'où

$$\begin{aligned}\int_a^b g(t)\,dt \quad &\simeq \quad \mathcal{P}\,(a < \alpha X < b) \quad = \quad \mathcal{P}\,(\frac{a}{\alpha} < X < \frac{b}{\alpha}) \\ &\simeq \quad \int_{a/\alpha}^{b/\alpha} f(t)\,dt \quad = \quad \int_a^b f(\frac{s}{\alpha})\,\frac{1}{\alpha} ds\end{aligned}$$

où la dernière égalité s'obtient par le changement de variable $s = \alpha t$. Pour faire apparaître la densité d'une variable aléatoire $W$, il suffit d'exprimer la probabilité $\mathcal{P}\,(a < W < b)$ sous la forme d'une intégrale dont les bornes sont $a$ et $b$, ce qui est le cas ci-dessus. Par conséquent

$$g(t) \;=\; \frac{1}{\alpha}\, f(\frac{t}{\alpha}) \qquad (XI.12)$$

formule qui permet de calculer très facilement la densité de $\alpha X$ à partir de celle de $X$. Or la densité de Student est celle de $\sqrt{n-1} \cdot Y/\sqrt{Z}$, il suffit donc de prendre $\alpha = \sqrt{n-1} = \sqrt{r}$ dans $XI.12$ : la densité $f(t)$ de $Y/\sqrt{Z}$ étant donnée par $XI.11$, on obtient pour la densité de Student

$$g(t) \;=\; \frac{1}{\sqrt{\pi r}}\,\frac{\Gamma(\frac{r+1}{2})}{\Gamma(\frac{r}{2})}\,\frac{1}{(\frac{t^2}{r}+1)^{\frac{r+1}{2}}} \qquad (XI.13)$$

On l'appelle densité de Student *à $r$ degrés de liberté*. Mais $r$ est égal à $n-1$, car, tout comme pour la densité $\chi^2$, les $n$ variables aléatoires gaussiennes dont la somme des carrés est $Z$ (les $n$ fluctuations autour de $m$) ont une somme nulle, ce qui diminue d'une unité leurs degrés de liberté.

Lorsque $r$ tend vers l'infini, cette densité de Student tend vers une limite. En effet, on sait que le quotient $\Gamma(x+\frac{1}{2})/\Gamma(x)$ est équivalent à $x^{\frac{1}{2}}$ lorsque $x$ tend vers l'infini (ici $x = r/2$) et que $(1+t^2/r)^{-\frac{r+1}{2}}$ tend vers $e^{-\frac{t^2}{2}}$, donc la densité limite sera $(1/\sqrt{2\pi})\, e^{-\frac{t^2}{2}}$, c'est-à-dire la densité gaussienne normalisée. En pratique, pour les exigences de précision courantes, cette limite est déjà atteinte pour $r = 20$ (voir figure 49).

Une fois connue la densité de $S = Y/\sqrt{Z}$ ou de $S' = Y/\sqrt{Z/(n-1)}$ on pourra effectuer un test d'hypothèse : on calcule à partir de la valeur a priori $m$ et des données statistiques la valeur *observée* de la variable aléatoire $S$ ci-dessus. Si elle dépasse un seuil déterminé, on conclura que, *indépendamment de la valeur inconnue de* $\sigma$, les écarts par rapport à la moyenne sont trop forts pour que l'hypothèse d'équiprobabilité puisse être retenue comme plausible. Pour la détermination de ce seuil on procédera comme pour le test du $\chi^2$ : il sera tel que la probabilité de le dépasser soit inférieure à une norme de certitude conventionnelle, par exemple 0.05 ou 0.01.

Ainsi le test du $\chi^2$ doit être utilisé lorsqu'on connaît a priori les écarts-types : on forme alors la somme normalisée des carrés des écarts par rapport à la moyenne empirique. En revanche le test de Student sera utilisé lorsqu'on connaît la moyenne théorique $m$ a priori, et qu'on sait également a priori (par des considérations de symétrie et d'invariance) que les écarts par rapport à $m$ ont tous le même écart-type $\sigma$, mais que ce dernier est inconnu.

Il est essentiel, lorsqu'on recourt à ces tests, de les employer *à bon escient*. Des tests pharmaceutiques effectués en clinique (par exemple l'observation sur un groupe de cinquante personnes recevant un médicament hypotenseur, comme nous en avons discuté au chapitre **X**) ne s'apparentent pas à l'expérience avec les dés, même si on n'en conteste pas la reproductibilité. En effet les fluctuations de la tension dans un groupe (ou au cours du temps chez une même personne) n'ont aucune raison d'obéir à la loi multinômiale ; d'ailleurs, la tension ne prend pas que six ou huit valeurs, c'est un paramètre pratiquement continu. Il est raisonnable de considérer les fluctuations de tension comme gaussiennes, mais on en ignore l'écart-type.

On a établi empiriquement avec une précision assez poussée que la répartition des $Q.I.$ (quotient intellectuel) dans une très grande population est gaussienne, car on a pu disposer de données assez nombreuses. Le $Q.I.$ semble en effet suffisamment passionner les décideurs pour que des expériences à grande échelle puissent trouver un financement. Mais l'écart-type varie beaucoup selon les populations ; il n'y a pas pour le $Q.I.$ un écart-type universel.

Les mécanismes moléculaires de l'hérédité obéissent à une causalité spatio-temporelle simple qui permettent de cerner "le niveau où intervient le hasard pur". Si par déduction a priori on établit une loi multinômiale pour la combinaison de certains gènes, et qu'on souhaite tester sur un échantillon une hypothèse sur les effets observables de cette combinaison, le test du $\chi^2$ prend tout son sens, comme dans l'exemple du dé. En revanche, le métabolisme d'un organisme vivant complexe, ou la psychologie animale, ne sont pas réductibles à une causalité spatio-temporelle simple (plus exactement : même si cette réduction est possible, nous ne savons pas l'effectuer) ; les tests ne reposent alors que sur des extrapolations plausibles. Dans la plupart des cas qui concernent le vivant, l'hypothèse des fluctuations gaussiennes est simplement plausible.

Le choix du test pertinent ne doit donc jamais être laissé au hasard. Lorsque ce choix est dûment motivé, il exprime toujours une hypothèse sur le rôle joué par le hasard dans le phénomène testé. Ces hypothèses devraient — selon le véritable esprit scientifique — être formulées explicitement et discutées, mais c'est rarement le cas dans la pratique.

Bien entendu il existe d'innombrables autres tests statistiques que les deux que nous venons d'étudier, mais ce n'est pas l'ambition de cet ouvrage de présenter un exposé exhaustif de tous les tests. Et bien entendu, on peut rencontrer des situations où *aucun* des deux tests étudiés ne convient ; c'est pourquoi les statisticiens en ont créé d'autres. Mais tout test, aussi sophistiqué qu'il soit, repose sur des hypothèses implicites.