

## Chapitre 4

### LES LOIS DES PROBABILITÉS

Le temps est venu d'étudier les lois des probabilités les plus courantes et leurs applications.

#### 4.1 La loi binomiale

Considérons une population dont les individus en proportion  $p$  sont distingués par une caractéristique remarquable : population d'électeurs qui se proposent de voter *Clovis* (parmi les diverses possibilités qui sont offertes), lot de pièces défectueuses (ou non) que l'on se propose d'acquérir (ou non), boules vertes dans une urne, etc...

Dans la population, on tire un individu au hasard, de telle sorte que tous les individus soient équiprobables. Une telle épreuve est *une épreuve de Bernoulli* (avec un seul "i").

On construit une variable aléatoire  $X$  qui prend la valeur 1 si l'individu tiré au hasard présente la caractéristique remarquable évoquée, et la valeur 0 dans le cas contraire.

Une telle variable est une *variable de Bernoulli*.

$$\begin{array}{l} \text{variable : } X = \quad 1 \quad 0 \\ \text{probabilité : } \quad p \quad q = 1 - p \end{array}$$

On utilise les définitions pour calculer l'espérance,  $E(X) = \bar{X}$ , de la variable  $X$  et son écart-type,  $\sigma_X$  :

$$E(X) = \bar{X} = p \quad \text{et} \quad \sigma_X = \sqrt{pq} = \sqrt{p(1-p)}$$

On considère une suite de  $n$  épreuves de Bernoulli successives et indépendantes. On peut ainsi modéliser un sondage pré-électoral\*, ou un échantillon de pièces tirées du lot à acheter.

Un tirage est une succession de 0 et de 1 en nombre  $n$ . On additionne toutes les valeurs des  $n$  variables aléatoires obtenues. On construit ainsi la variable aléatoire  $Z = X_1 + X_2 + \dots + X_n$  où  $X_j$  est la variable aléatoire  $X$  de rang  $j$  (ce n'est pas, ici, l'une des valeurs possible de la variable  $X$ , comme ce fut le cas dans quelques circonstances précédentes).

La valeur obtenue pour  $Z$  est un entier,  $k$ , qui est égal au nombre de 1 obtenus dans le tirage (nombre d'électeurs qui ont l'intention de voter *Clovis*).

La variable  $Z$  suit une *loi binomiale* :

$$\text{proba}[Z = k] = C_n^k p^k (1-p)^{n-k} \quad \text{avec} \quad C_n^k = \frac{n!}{k!(n-k)!} \quad (4.1)$$

---

\*Le plus souvent, pour des raisons d'économie, les instituts de sondage n'utilisent pas cette méthode mais la méthode des quotas qui consiste à interroger un échantillon réduit que l'on croit représentatif.

$C_n^k$  est le nombre de suites différentes de  $n$  nombres comportant  $k$  fois le nombre 1 et  $n - k$  fois le nombre 0, tandis que  $p^k q^{(n-k)} = p^k (1 - p)^{n-k}$  est la probabilité pour que l'une quelconque des suites précédentes soit le résultat des  $n$  épreuves de Bernoulli.

On peut calculer l'espérance, la variance et l'écart-type de la loi binomiale précédente, en utilisant les propriétés concernant la somme de variables indépendantes (cf.3.10) :

$$\boxed{E(Z) = \bar{Z} = n \times E(X) = n p} \tag{4.2}$$

$$\boxed{\mathcal{V}(Z) = n \times \mathcal{V}(X) = np(1 - p)} \quad \boxed{\sigma_Z = \sqrt{n p (1 - p)} = \sqrt{n} \sigma_X} \tag{4.3}$$

La variable  $Z$  est la fréquence absolue des observations  $X = 1$  dans l'échantillonnage au hasard effectué. Introduisons la fréquence relative. Plus précisément posons

$$z = \frac{Z}{n}$$

La valeur observée de  $z$  est la moyenne observée de  $X$  sur l'échantillon. Les relations 3.10, 3.3 et 3.4 donnent l'expression de l'espérance de  $z$  et de son écart type :

$$\boxed{E(z) = \bar{z} = \frac{1}{n} E(Z) = E(X) = p}, \quad \boxed{\sigma_z = \frac{1}{n} \sigma_Z = \frac{\sigma_X}{\sqrt{n}} = \sqrt{\frac{p(1-p)}{n}}}$$

La valeur observée de  $z$  est le résultat d'une campagne de mesures portant sur un échantillon de cardinal  $n$ . En répétant les campagnes de mesures, on obtient des valeurs de  $z$  dispersées autour de la moyenne,  $p$ . L'écart-type  $\sigma_z$ , permet d'apprécier cette dispersion : c'est un ordre de grandeur de l'incertitude sur la mesure de  $p$  lorsqu'on assimile la moyenne observée,  $z$ , à la moyenne réelle, inconnue,  $p$ . Lorsque  $n$  croît,  $\sigma_z$  décroît et par conséquent, l'incertitude décroît. Un grand nombre d'observations permet ainsi d'améliorer la précision du résultat. C'est cette propriété, présentée ici de façon intuitive, que nous allons maintenant démontrer avec plus de rigueur.

### 4.2 L'inégalité de Bienaymé-Tchebychev

#### 4.2.1 L'inégalité de Bienaymé-Tchebychev

Considérons une variable aléatoire  $X$ , susceptible de prendre les valeurs  $X_k$  avec la probabilité  $p_k$ . La moyenne de  $X$  est  $\bar{X} = \sum_k X_k p_k$ , la variance est  $\sigma_X^2 = \sum_k (X_k - \bar{X})^2 p_k$

Donnons nous une valeur  $a$ , positive, arbitraire. On distingue deux régions : la région I correspondant à  $|X - \bar{X}| \leq a$  et la région II correspondant à  $|X - \bar{X}| > a$  (voir la figure 4-1 ci-dessous).

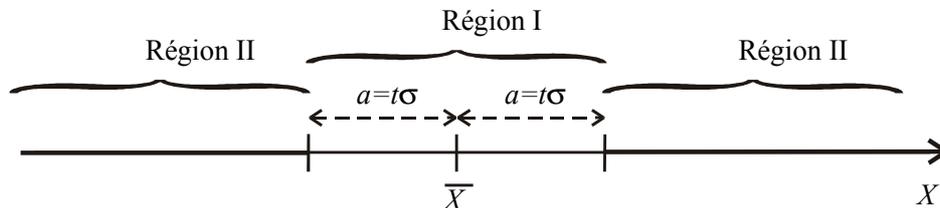


Figure 4-1

Par définition il vient  $\sigma_X^2 = \sum_I (X_k - \bar{X})^2 p_k + \sum_{II} (X_k - \bar{X})^2 p_k \geq \sum_{II} (X_k - \bar{X})^2 p_k$ .

Cependant, dans la région II, il vient  $|X - \bar{X}| > a$  ; on en déduit

$\sigma_X^2 \geq \sum_{\text{II}} (X_k - \bar{X})^2 p_k > a^2 \sum_{\text{II}} p_k$  or  $\sum_{\text{II}} p_k$  est la probabilité de la région II, c'est-à-dire la probabilité de  $|X - \bar{X}| > a$ . On écrit le résultat obtenu sous la forme  $\frac{\sigma_X^2}{a^2} > \sum_{\text{II}} p_k = \text{proba} [|X - \bar{X}| > a]$ . Nous avons vu que  $\sigma_X$  fixe une échelle naturelle de référence pour apprécier si  $|X - \bar{X}|$  est grand ou petit (voir la section 1.6 page 20), on préfère donc poser  $a = t \sigma$  où  $t$  est un nombre sans dimension, positif, arbitraire. L'inégalité précédente s'écrit alors

$$\boxed{\text{proba} [|X - \bar{X}| > t \sigma_X] < \frac{1}{t^2}} \quad (4.4)$$

Cette inégalité est ***l'inégalité de Bienaymé-Tchebychev***<sup>†</sup> dont nous présentons maintenant quelques applications.

[[Jules Irénée Bienaymé (1796-1878) fut inspecteur des finances puis, après la révolution de 1848 professeur à la Sorbonne. Il publia 23 articles scientifiques durant sa vie, la plupart dans d'obscures revues. Ce palmarès lui permettrait aujourd'hui d'être un candidat pas très bien placé à un poste de professeur des Universités, mais à son époque on lisait les articles avant de porter un jugement, les compter n'était pas suffisant. Il est vrai que dans ce temps là, et jusqu'à un passé récent, on tournait 7 fois son porte-plume dans la main avant d'écrire une première page<sup>‡</sup>. Quoique de 25 ans son aîné, Bienaymé était très lié avec Tchebychev dont il traduisit les travaux en Français. Toujours très en avance sur son temps dans le domaine des statistique, Bienaymé est à l'origine de l'inégalité 4.4.]

Pafnuty Lvovich Tchebychev (1821-1894), issu d'une famille aisée de militaires russes, reçut une excellente éducation. Très tôt il apprit le Français à une époque où cette langue était encore magnifiquement écrite et largement répandue dans tous les milieux cultivés, tout particulièrement en Europe. Cette connaissance du Français facilita les relations scientifique de Tchebychev avec de nombreux mathématiciens, parmi eux I. J. Bienaymé dont il utilisa l'inégalité pour exprimer la loi des grands nombres telle que nous l'avons présentée. Depuis, la relation 4.4 est connue comme l'inégalité de Bienaymé-Tchebychev.]]

#### 4.2.2 La loi des grands nombres

Appliquons à la variable  $z$  de la section précédente, l'inégalité de Bienaymé-Tchebychev 4.4 :

$$\text{proba} [|z - \bar{z}| > t \sigma_z] < \frac{1}{t^2}$$

Avec  $\bar{z} = p$  et  $\sigma_z = \sqrt{\frac{p(1-p)}{n}}$  il vient

$$\text{proba} \left[ |z - p| > t \sqrt{\frac{p(1-p)}{n}} \right] < \frac{1}{t^2} \quad (4.5)$$

Remarquons la relation  $p \in [0, 1] \Rightarrow p(1-p) \leq 1/4$ .

<sup>†</sup>L'inégalité de Bienaymé-Tchebychev est un cas particulier de l'inégalité  $\text{proba} [|Z| > t (E[|Z|^r])^{1/r}] < \frac{1}{t^r}$ , valable pour  $t$  positif, sous réserve que  $E[|Z|^r]$  soit défini.

Cette inégalité se démontre comme l'inégalité de Bienaymé-Tchebychev. Dans le cas  $Z = X - \bar{X}$  et  $r = 2$  on obtient l'inégalité de Bienaymé-Tchebychev ; pour  $r = 1$ , c'est l'inégalité de Markov.

<sup>‡</sup>Allusions aux pratiques (trop) fréquentes en ce début de 21<sup>ème</sup> siècle.

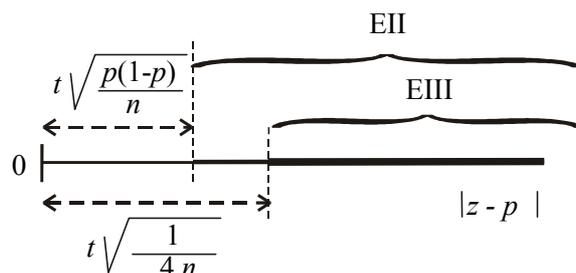


Figure 4-2

La relation  $p(1-p) \leq \frac{1}{4}$  est valide pour  $p \in [0, 1]$ . Nous considérons l'événement EII précédent qui correspond à  $|z-p| > t \sqrt{\frac{p(1-p)}{n}}$ , ainsi que l'événement EIII correspondant à  $|z-p| > t \sqrt{\frac{1}{4n}}$  (cf. figure 4-2). De la relation  $t \sqrt{\frac{p(1-p)}{n}} \leq t \sqrt{\frac{1}{4n}}$ , pour  $t > 0$ , on déduit  $EIII \subseteq EII$ , par conséquent  $proba[EIII] \leq proba[EII]$ . L'inégalité de Bienaymé-Tchebychev s'écrit  $proba[EII] < \frac{1}{t^2}$ , on en déduit donc  $proba[EIII] < \frac{1}{t^2}$ , ce qui s'écrit

$$proba \left[ |z-p| > t \sqrt{\frac{1}{4n}} \right] < \frac{1}{t^2} \quad (4.6)$$

Considérons un exemple. Dans une population donnée, nous cherchons à connaître la proportion d'individus appartenant à un sous-ensemble  $A$  (le sous-ensemble de ceux qui votent *Clovis* par exemple). Soit  $X$  la *variable caractéristique* de  $A$ , c'est à dire la variable qui vaut 1 pour tout individu de  $A$  et 0 pour les autres individus.

1. Nous ne connaissons pas la proportion  $p$  d'individus pour lesquels  $X = 1$ . Pour connaître  $p$ , nous répétons  $n$  fois une épreuve de Bernoulli afin d'extraire de la population étudiée un échantillon au hasard. La proportion observée dans l'échantillon est  $z$ . Pour nous faire souvenir que cette proportion est observée sur un échantillon de cardinal  $n$ , nous changeons de notation :  $z \rightarrow z_n$ . La valeur  $z_n$  est une estimation de  $p$ .
2. En assimilant  $z_n$  et  $p$  nous commettons une erreur  $e \stackrel{\text{déf}}{=} z_n - p$ . Remarquons que cette erreur est inconnue (si nous connaissions  $e$ , nous pourrions obtenir la valeur exacte de  $p = z_n - e$  car  $z_n$  est connu). Nous définissons une erreur acceptable,  $\varepsilon$ . Par exemple  $\varepsilon = 0,01\% = 10^{-4}$ . Notre objectif est de connaître  $p$  avec une erreur inférieure à  $10^{-4}$ . Nous voulons donc éviter d'avoir  $|e| = |z_n - p| > \varepsilon$  (ce qui constitue l'événement  $E_\varepsilon$ ).

Nous ne serons jamais certain d'éviter  $E_\varepsilon$ . Mais nous pouvons rendre cet événement improbable. Pour cela, nous fixons un seuil de probabilité  $\eta$ , tel que toute probabilité inférieure à  $\eta$  est considérée comme négligeable :  $\eta = 10^{-3}$  par exemple. Un événement dont la probabilité est inférieure à  $\eta$  est considéré comme improbable. Son complémentaire de probabilité supérieure à  $1 - \eta$  (supérieur à 0,999 dans l'exemple choisi) est dit "**probable**".

3. Supposons la relation  $\varepsilon \geq t \sqrt{\frac{1}{4n}}$ . Il vient

$$proba[|z_n - p| > \varepsilon] \leq proba \left[ |z_n - p| > t \sqrt{\frac{1}{4n}} \right] < \frac{1}{t^2}$$

Posons en outre  $\eta = \frac{1}{t^2}$ . On en déduit  $n \geq \frac{1}{4\eta\varepsilon^2}$  et

$$\text{proba} [|z_n - p| > \varepsilon] \leq \text{proba} \left[ |z_n - p| > t \sqrt{\frac{1}{4n}} \right] \leq \frac{1}{t^2} = \eta$$

Dans l'exemple choisi,  $n \geq \frac{1}{4\eta\varepsilon^2}$  implique  $n \geq 2,5 \times 10^{10}$

**Ainsi, si  $n$  est assez grand, il est probable que la proportion observée de cas favorables dans l'échantillon est voisine de la proportion de cas favorables dans la population.** On dit que  $z_n$  tend vers  $p$  "en probabilité" lorsque  $n$  tend vers l'infini.

Cette loi est connue comme "le théorème de Jacques Bernoulli" ; c'est la première formulation de la loi de grands nombres.

La loi des grands nombres repose sur l'hypothèse que, dans l'épreuve de Bernoulli, les individus sont choisis au hasard. Mais *a contrario*, ayant défini un mode opératoire pour choisir un individu dans la population de départ, la loi des grands nombres nous permet de vérifier si ce mode opératoire définit une épreuve de Bernoulli. Pour cela, il suffit de mettre en oeuvre ce mode opératoire pour tirer des échantillons d'une population connue et de vérifier que les proportions observées tendent vers les proportions présentes dans la population lorsque le nombre de tirages augmente.

Cette méthode est connue de chacun sans qu'il ne soit besoin d'aucun calcul ni d'aucune inégalité de Bienaymé-Tchebychev. Supposez que vous vous proposiez de jouer à pile ou face. Avant de décider vous observez le comportement de la pièce de monnaie utilisée et vous constatez les résultats suivants :

Pile, face, face, pile, face, face, face, face, pile, face, face, face face pile, pile, face, face, face, pile, pile, face, pile, face, pile, pile, pile, face, face, face, pile, face, face, pile, pile, face, face, face, face, face, face, pile, face, face, pile, face, pile, face, .....

Seriez-vous prêt à jouer, sur *pile*, votre argent votre salaire ou même votre argent de poche ? Non, bien sûr : *face* sort trop souvent, la pièce est sans doute truquée (cf. figure 4-3).

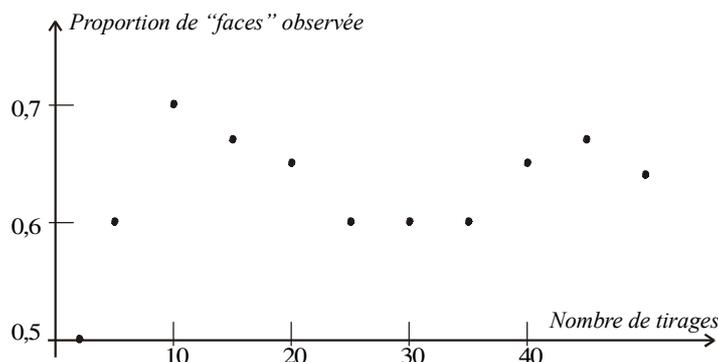


Figure 4-3

Tout ça, c'est-à-dire la loi des grands nombres à la limite où  $n \rightarrow \infty$ , vous le saviez d'abord, comme disent les enfants ; mais l'inégalité de Bienaymé-Tchebychev vous en apprend un peu plus : elle vous permet d'estimer le nombre  $n$ , cardinal de l'échantillon à prélever pour atteindre les objectifs fixés.

Remarquez que  $n$  ne dépend que des objectifs fixés,  $\varepsilon$  et  $\eta$ , et ne dépend pas de la taille de la population. En utilisant la méthode précédente, avec un échantillon de 2000

individus, vous avez la même information que la population soit constituée de 200 000 ou de 2010 individus. Remarquez que la répétition de l'épreuve de Bernoulli suppose que c'est toujours dans la même population que vous choisissez chaque individu. Tout individu sélectionné lors d'un tirage ne doit pas être éliminé ; il fait nécessairement partie du tirage suivant et le cas échéant, il pourra être sélectionné plusieurs fois. Dans ces conditions, il est même possible d'extraire un échantillon de 2000 individus à partir d'une population qui n'en contiendrait que 200. Ce serait absurde mais c'est possible dans le cas d'un tirage non exhaustif.

#### 4.2.3 Quelques exemples

Nous avons utilisé l'inégalité de Bienaymé-Tchebychev sans rien supposer de la proportion  $p$  à déterminer. Seuls, les objectifs (c'est-à-dire  $\varepsilon$  et  $\eta$ ) étaient fixés.

Supposons que l'ordre de grandeur de  $p$  soit connu. L'estimation de  $n$  en est affinée et peut être très inférieure à l'estimation précédente (estimation obtenue dans le cas où  $p(1-p)$  est maximum, c'est-à-dire pour  $p \sim 0,5$ ).

Supposons que nous voulions tester un lot de diodes laser par exemple. La fabrication industrielle nous apprend que la proportion  $p$  de diodes défectueuses est de l'ordre de 6%. Nous mettons en place un perfectionnement de la production qui se traduit par une diminution de  $p$ . C'est la nouvelle valeur de  $p$  que nous voulons estimer. La question se pose de la taille de l'échantillon à prélever (au hasard) pour estimer, "**au seuil de confiance**" de 90%, la valeur de  $p$  avec une erreur qui n'excède pas 1%.

"**au seuil de confiance**" de 90% : Oui ! C'est de cette façon que l'on exprime la condition  $\eta < 0,1 = 10\%$ . Le seuil de confiance c'est le seuil de probabilité,  $\pi_s$ , à partir duquel on peut avoir confiance dans la réalisation de l'événement. La probabilité  $1 - \pi_s$  est donc la probabilité d'un événement dont on a confiance qu'il ne se réalisera pas : c'est ce que nous avons appelé "une probabilité négligeable", on l'appelle aussi le "*risque*".

Au lieu de l'inégalité 4.6, nous utilisons 4.5 avec  $p = 0,06$ . Nous posons

$$\boxed{\frac{1}{t^2} = \eta = 1 - \pi_s} = 0,10 \text{ et } t \sqrt{\frac{p(1-p)}{n}} = \varepsilon = 0,01.$$

$$\text{Avec } p = 0,06, \text{ on trouve } n \gtrsim \frac{p(1-p)}{\varepsilon^2 \eta} = 6000.$$

En utilisant 4.6 ou en supposant  $p \sim 0,5$ , ce qui revient au même, on trouve  $n \gtrsim 25\,000$ . En utilisant notre connaissance de  $p$ , nous avons considérablement diminué le coût du contrôle. Nous verrons que le nombre de contrôles peut encore être considérablement diminué en exploitant notre connaissance de la loi asymptotique suivie par la variable aléatoire que constitue la fréquence observée (voir l'exemple page 75). C'est particulièrement utile dans le cas de contrôles destructifs où l'objet contrôlé est détruit par le contrôle lui-même (certains équipements de sécurité dans des voitures soumises à un choc violent simulant un grave accident, ou les blindages des chars de combats dont on cherche à vérifier la résistance à la pénétration de projectiles de toutes natures).

Deux remarques s'imposent.

1. Il faut contrôler 6 000 diodes laser. Que la production soit de 10 000 ou de 100 000 diodes, c'est toujours 6 000<sup>§</sup>. La seule condition c'est que ces 6 000 diodes soient choisies au hasard sur l'ensemble des diodes produites et que ce ne soient pas 6 000 diodes produites le même jour, même si cette journée est tirée au hasard.
2. Comment constituer l'échantillon ? Lorsque la population n'est pas trop nombreuse, on numérote les individus par ordre de production. Avec une table de nombres au

---

<sup>§</sup>Ce qui représente 60% de la population dans le premier cas et 6% seulement dans le second cas. Nous vous laissons apprécier la pertinence des intuitions de ceux qui dans le présent contexte affirment avec aplomb : "6% est une représentation insuffisante de cette population, 20% ce serait mieux".

hasard (ou un générateur de nombres au hasard) on obtient une liste de numéros d'ordre. On sélectionne alors les individus dont le numéro appartient à la liste.

Si la population est très nombreuse, on la découpe en sous-ensembles de même taille dont certains sont tirés au hasard (toujours grâce aux tables ou aux générateurs de nombres au hasard). Parmi ces sous-ensembles, on sélectionne (au hasard...) les individus qui seront contrôlés. On voit que la difficulté ne vient pas du contrôle de 6000 diodes par an mais de la réalisation d'une équiprobabilité d'accès à chacune d'entre elles.

3. Le contrôle de ces 6 000 diodes permet d'obtenir une nouvelle estimation de  $p$  que l'on note  $p_1$ . C'est l'objet du contrôle, rappelons-le. En utilisant  $p_1$ , on peut calculer  $n_1 = \frac{p_1(1-p_1)}{\eta\varepsilon^2}$ . C'est le cardinal de l'échantillon à étudier pour atteindre les objectifs fixés, compte tenu de la nouvelle estimation de  $p$ . Deux cas sont envisageables selon que  $n_1 \leq n$  ou  $n_1 > n$ .

Pour  $n_1 \leq n$ , on arrête le test et on accepte  $p_1$  comme estimation de  $p$ . En effet, l'ensemble des résultats est cohérent dans la mesure où le résultat du test suggère qu'une étude moins approfondie aurait été suffisante.

Pour  $n_1 > n$ , la situation est différente. Les résultats de l'étude nous disent que si nous acceptons l'estimation  $p_1$ , alors il faut poursuivre l'étude. D'autre part, il est clair que si nous refusons  $p_1$  nous devons poursuivre l'étude car notre objectif est de disposer d'une estimation de  $p$ . Dans tous les cas, nous devons poursuivre en calculant l'estimation  $p_k$  de  $p$  à chaque stade, pour en déduire  $n_k$ . L'étude s'arrête lorsque le cardinal de l'échantillon effectivement étudié est supérieur à l'estimation qui en est faite.

Un tel test où à chaque stade on décide de cesser ou de poursuivre selon la cohérence des résultats est appelé "*un test progressif*".

Les élections démocratiques en Amérique et en Europe relèvent souvent d'une opération de marketing où la publicité joue un rôle plus important que les choix politiques. Cette logique conduit à l'affrontement de deux candidats qui, disposant, de moyens comparables se retrouvent "au coude à coude". Effectuer un sondage d'opinion pour savoir si Clovis l'emportera revient à déterminer la proportion des électeurs qui voteront "Clovis", sachant que cette proportion est voisine de 50%. Les écarts observés dans les élections antérieures étant de l'ordre de 1%, le taux d'abstention étant inconnu, il est raisonnable de rechercher la "certitude" d'une erreur inférieure à 1% = 0,01.

"certitude" : nous devons préciser le sens du mot. Fixons le seuil de confiance à 95%. Pour estimer  $n$  nous posons donc  $p \sim 0,5$  et  $\varepsilon = 0,01$  tandis que  $\eta = 0,05$ . On en déduit  $n = 50\,000$ . Il n'est pas envisageable d'interroger un tel nombre d'électeurs tirés au hasard. Les moyens à mettre en œuvre sont trop importants, et d'un coût excessif. A une telle méthode, que l'on peut qualifier de scientifique, on préfère la méthode qui consiste à interroger un échantillon de quelques centaines d'électeurs constitué sur des bases sociologiques et dont on suppose qu'il est représentatif, avec les succès et les échecs que l'on connaît.

Le choix de  $\varepsilon$  est défini en fonction de l'objectif recherché dans l'étude entreprise. Aucun problème particulier n'est soulevé par le choix de  $\varepsilon$ . Il n'en va pas de même pour le choix de  $\eta$ . Comment choisir un seuil de confiance? Qu'est-ce qu'une probabilité négligeable?

Il n'y a pas de réponse universelle à cette question. Reprenons le jeu de pile ou face de la page 59. La probabilité de face est voisine de 2/3. Pour vous amuser peut-être joueriez vous 1 euro sur *face* en considérant que la probabilité de *pile* (1/3) est

négligeable, mais sans doute pas votre argent du mois et certainement pas votre vie quelle qu'en soit la contre-partie<sup>¶</sup>. A cet exemple vous voyez que  $1/3$  peut être considéré comme une probabilité négligeable ou non selon ce que vous voulez faire.

Pour ce qui concerne les problèmes d'estimation ou les tests d'hypothèses (voir ci-dessus page 60 et ci-après le chapitre 6, page 87) et, de façon générale, pour ce qui concerne les aides à la décision on considère pour probabilité négligeable les valeurs 10%, 5%, 1% selon les espoirs ou les inquiétudes qui hantent le statisticien.

La question du choix du seuil doit être résolue par le statisticien s'il travaille pour son compte mais s'il travaille pour un employeur cette question est fondamentalement du ressort de l'employeur. C'est lui qui investit, qui prend les risques et qui tire les bénéfices de son action, c'est à lui de prendre ses responsabilités. Dans cette optique, le commanditaire d'une étude a dégagé des moyens qui permettent le contrôle de 6000 diodes. L'estimation de la proportion de déchets est alors de 8%. On en déduit  $\eta\varepsilon^2 = 1,2 \times 10^{-5}$ . On trace le graphe de la fonction  $x = 100 \varepsilon \mapsto y = 100 \eta = 12/x^2$ . C'est alors au commanditaire d'interpréter le graphique ci-dessous<sup>||</sup>.

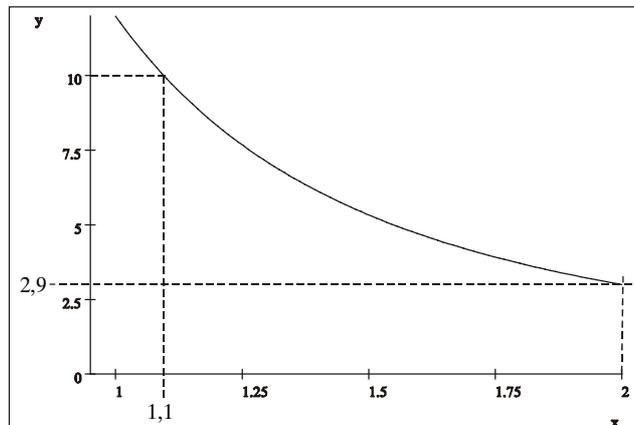


Figure 4-4

Les résultats de l'étude peuvent s'interpréter de diverses manières. A titre d'exemple, nous donnons deux interprétations possibles (en posant  $2,9 \simeq 3$ ).

$$p = 8,0\% \pm 1,1\% \text{ au seuil } 0,9 \quad \text{ou} \quad p = 8,0\% \pm 2\% \text{ au seuil } 0,97$$

où " $p = 8\% \pm 2\%$  au seuil  $0,97$ " signifie " $|p - 8\%| = |p - 0,08| < 2\%$ , événement dont la probabilité est supérieure à  $0,97$ ". Doit-on en conclure que l'amélioration recherchée conduit à une dégradation des résultats (ce qui semble être le cas) ou qu'il faut poursuivre l'étude? C'est au commanditaire d'en décider.

#### 4.2.4 Formulation générale de la loi des grands nombres

Le théorème de Bernoulli que nous avons déduit de l'inégalité de Bienaymé-Tchebychev, apparaît comme un cas particulier d'une loi plus générale : **la loi faible des grands nombres** que nous présentons maintenant. Les démonstrations s'appuient sur les résultats de la section 3.3 et en particulier les paragraphes 3.3.1 et 3.3.4.

Considérons une suite de variables numériques aléatoires indépendantes  $X_1, X_2, \dots, X_k, \dots$ . Nous supposons que ces variables sont centrées (c'est-à-dire telles que  $E[X_k] = 0$ ). Soit  $\sigma_k$  l'écart-type de la variable  $X_k$ . Nous supposons la relation  $\frac{1}{n^2} \sum_{k=1}^n \sigma_k^2 \xrightarrow[n \rightarrow \infty]{} 0$ .

<sup>¶</sup>Enfin espérons-le!

<sup>||</sup>S'il en sait assez sur le sujet et s'il n'attend pas de l'expert de prendre les décisions à sa place.

C'est le cas, par exemple, lorsque  $\sigma_k$  est majoré :  $\sigma_k < \sigma$  quel que soit  $k$  (le majorant est ici  $\sigma$ ).

Avec les  $n$  premières variables de cette suite formons la variable  $Z_n = \frac{1}{n} \sum_{k=1}^n X_k$ .

L'espérance de  $Z_n$  est  $E[Z_n] = \frac{1}{n} \sum_{k=1}^n E[X_k] = 0$

Les variables étant indépendantes, la variance de  $Z_n$  est  $(\sigma_{Z_n})^2 = \frac{1}{n^2} \sum_{k=1}^n \sigma_k^2$ .

L'inégalité de Bienaymé-Tchebychev (4.4), appliquée à la variable  $Z_n$  s'écrit alors

$$\text{proba} [ |Z_n| > t \sigma_{Z_n} ] < \frac{1}{t^2}$$

Lorsque  $n$  tend vers l'infini,  $t \sigma_{Z_n}$  tend vers zéro, quel que soit  $t$  choisi aussi grand que l'on veut. Lorsque  $n$  est "assez" grand, il est donc "improbable" que l'on ait  $|Z_n| > \varepsilon$  quel que soit  $\varepsilon$ . Il est donc probable que l'on a  $|Z_n| < \varepsilon$  lorsque  $n$  tend vers l'infini, et cela quel que soit  $\varepsilon$ . Dans ces conditions on dit que  $Z_n$  converge vers zéro *en probabilité*. Dans ces conditions la loi faible\*\* des grands nombres s'écrit

$$Z_n \xrightarrow[n \rightarrow \infty]{pr} 0$$

Dans la pratique, la loi des grands nombres est utilisée le plus souvent pour estimer la moyenne  $\bar{Y}$  d'une variable aléatoire  $Y$ . Par exemple on répète la mesure d'une grandeur physique  $n$  fois. Les résultats obtenus sont les valeurs  $y_1, y_2, \text{etc}, y_n$ . La moyenne observée est  $u_n = \frac{1}{n} (y_1 + y_2 + \dots + y_n)$ . De façon générale, on admet que  $u_n$  est une meilleure estimation de  $\bar{Y}$  que la seule mesure  $y_1$ .

Dans l'expérience décrite, nous disposons de  $n$  épreuves identiques dont les résultats sont les variables indépendantes,  $Y_1, Y_2, \dots, Y_n$ . Les valeurs  $y_1, y_2, \text{etc}$ , prises par ces variables ne sont pas déterminées avant le déroulement des épreuves, seule est connue la loi de probabilité de  $Y_k$  : c'est la loi suivie par la variable  $Y$ , la même pour tous les  $k$ . La valeur  $u_n$  n'est donc pas connue *a priori* ; c'est la valeur d'une variable aléatoire  $U_n = \frac{1}{n} \sum_{k=1}^n Y_k$ .

On pose  $X_k = Y_k - \bar{Y}$ . On vérifie que la variable aléatoire  $X_k$  satisfait les hypothèses de validité de la loi des grands nombres exposées ci-dessus. On en déduit  $Z_n \xrightarrow[n \rightarrow \infty]{pr} 0$ .

Cependant  $Z_n = \frac{1}{n} \sum_{k=1}^n X_k = \frac{1}{n} \sum_{k=1}^n (Y_k - \bar{Y}) = U_n - \bar{Y}$ . La loi des grands nombres s'écrit alors  $U_n - \bar{Y} \xrightarrow[n \rightarrow \infty]{pr} 0$  soit

$$U_n \xrightarrow[n \rightarrow \infty]{pr} \bar{Y}$$

Ainsi, si  $n$  est assez grand, la probabilité est très faible pour que  $u_n$  soit très différent de  $\bar{Y}$ .

De façon intuitive cela tient au fait que la dispersion des valeurs observées d'une variable aléatoire  $Q$ , autour de sa moyenne  $\bar{Q}$  est donnée (en ordre de grandeur) par l'écart-type  $\sigma_Q$ . Dans le cas considéré,  $Y$  et  $U_n$  ont même moyenne,  $\bar{Y}$  mais des écarts types différents :  $\sigma_{U_n} = \frac{\sigma_Y}{\sqrt{n}}$ . La dispersion des valeurs de  $U_n$  tend vers zéro lorsque  $n$  tend vers l'infini, alors que celle des valeurs de  $Y$  est une constante.

\*\*Il existe une "loi forte des grands nombre" correspondant à des hypothèses différentes de celles que nous avons considérées ici.

### 4.3 La loi de Poisson

#### 4.3.1 Présentation de la loi

Soit  $X$  une variable aléatoire. On dit que  $X$  suit une loi de Poisson si elle est susceptible de prendre les valeurs entières  $X_k = k \geq 0$  avec la probabilité  $P_k$  :

$$\boxed{\text{proba } [X = k] \stackrel{\text{def}}{=} P_k = e^{-m} \frac{m^k}{k!}} \quad (4.7)$$

où  $m$  est un paramètre arbitraire, positif et  $0! = 1$  (par définition).

On calcule la moyenne et la variance<sup>††</sup> de  $X$  :

$$\bar{X} = \sum_0^{\infty} k e^{-m} \frac{m^k}{k!} \quad \text{et} \quad \mathcal{V}(X) = \sum_0^{\infty} k^2 e^{-m} \frac{m^k}{k!} - (\bar{X})^2$$

on trouve

$$\boxed{\bar{X} = m = E(X) \quad \text{et} \quad \mathcal{V}(X) = m} \quad (4.8)$$

[[ De la relation  $e^m = \sum_{k=0}^{\infty} \frac{m^k}{k!}$ , valable pour tout  $m$ , on déduit  $\sum_{k=0}^{\infty} P_k = 1$ . Cette même relation permet de démontrer 4.8.

$$\begin{aligned} \bar{X} &= \sum_0^{\infty} k e^{-m} \frac{m^k}{k!} = e^{-m} \times \left( m + \frac{m^2}{1!} + \frac{m^3}{2!} + \dots \right) = e^{-m} \times m \left( 1 + \frac{m}{1!} + \frac{m^2}{2!} + \dots \right) \\ \bar{X} &= e^{-m} \times m e^m = m : \text{c'est ce que nous voulions démontrer.} \end{aligned}$$

$$\begin{aligned} \mathcal{V}(X) + (\bar{X})^2 &= \sum_0^{\infty} k^2 e^{-m} \frac{m^k}{k!} = e^{-m} \left( 1 \times \frac{m}{1!} + 2^2 \frac{m^2}{2!} + 3^2 \frac{m^3}{3!} + \dots \right) \\ \mathcal{V}(X) + (\bar{X})^2 &= e^{-m} \times m \left( 1 + 2 \frac{m}{1!} + 3 \frac{m^2}{2!} + \dots + k \frac{m^{k-1}}{(k-1)!} + \dots \right) \end{aligned}$$

Considérons la fonction  $F(m) = m + \frac{m^2}{1!} + \frac{m^3}{2!} + \dots + \frac{m^k}{(k-1)!} + \dots = m e^m$ .

En dérivant  $F$  par rapport à  $m$ , il vient

$$\frac{dF}{dm} = 1 + 2 \frac{m}{1!} + 3 \frac{m^2}{2!} + \dots + k \frac{m^{k-1}}{(k-1)!} + \dots = e^m + m e^m. \text{ On en déduit}$$

$$\begin{aligned} \mathcal{V}(X) + (\bar{X})^2 &= e^{-m} \times m (e^m + m e^m) = m^2 + m. \text{ En utilisant l'expression } \bar{X} = m \text{ il vient} \\ \mathcal{V}(X) &= m^2 + m - (\bar{X})^2 = m^2 \text{ Q.E.D. }^{\dagger\dagger}] \end{aligned}$$

Dans un schéma d'urne où les boules sont équiprobables,  $P_k$  est la proportion de boules sur lesquelles est écrite la valeur  $X_k = k$  de la variable. Dans ce schéma d'urne, les valeurs de  $P_k$  sont nécessairement des fractions rationnelles. En toute rigueur, un tel schéma ne peut donc pas décrire une loi de Poisson dont les probabilités ne sont pas nécessairement des fractions rationnelles. Toutefois chaque valeur  $P_k$  peut être approchée par une fraction rationnelle avec une précision arbitraire. Pour  $k < K$  on remplace donc chaque  $P_k$  par son approximation,  $\tilde{P}_k$ . On réduit les  $K$  fractions obtenues au même dénominateur,  $D$ . Chacune des fractions  $\tilde{P}_k$  s'écrit alors  $\tilde{P}_k = N_k/D$ . On peut donc imaginer

<sup>††</sup> Vous remarquez l'utilisation ici, d'un vocabulaire issu aussi bien du langage des statistiques (moyenne) que de celui des probabilités (variance). Ce n'est pas très cohérent mais c'est sans ambiguïté.

<sup>‡‡</sup> *Quid erat demonstrandum* : ce qui devait être démontré. A l'époque de Pascal on indiquait ainsi la fin de la démonstration.

une urne qui contienne  $N_k$  boules sur lesquelles est inscrit " $k$ " et  $D - \sum_k N_k$  boules sur lesquelles est inscrit " $k \geq K$ ". Une épreuve de Bernoulli conduit alors "probablement" à la loi de Poisson avec une précision aussi bonne que l'on veut pour les valeurs de  $k < K$  où  $K$  peut être choisi aussi grand que l'on veut. Ainsi, si c'est nécessaire, on peut le plus souvent appuyer les raisonnements sur un schéma d'urne bien qu'en toute rigueur un tel schéma ne puisse convenir. Cette remarque est générale pour la plupart des lois de probabilité que l'on rencontre en pratique.

#### 4.3.2 Exemples de variables poissonniennes

La nature nous fournit naturellement des variables poissonniennes.

Considérons par exemple un événement dont la probabilité d'occurrence pendant la durée infinitésimale  $dt$  est  $dP = \frac{dt}{\tau}$ . Le temps  $\tau$  est appelé "**constante de temps**", tandis que  $1/\tau$  est "**la probabilité d'occurrence par unité de temps**".

Que signifie l'expression "probabilité d'occurrence pendant  $dt$ " ? Quelle est l'épreuve correspondante ? Nul ne cherche à le savoir mais seulement à constater que "tout se passe comme si" pendant chacune des durées successives  $dt$  (infiniment petites) on effectuait une épreuve de Bernoulli en tirant une boule d'une urne où certaines d'entre elles, en proportion  $dP = dt/\tau \ll 1$ , portent la mention "occurrence". Remarquons qu'aucune boule ne porte la mention "deux occurrences" ou qu'une telle mention n'apparaît qu'en proportion négligeable, si bien que la probabilité de tirer une telle boule est très petite et que le phénomène ne se produit pratiquement jamais si  $dt$  est assez petit. Le processus envisagé exclut donc les occurrences simultanées de plusieurs phénomènes. Une telle épreuve virtuelle décrit convenablement de nombreuses situations et permet d'étudier de nombreux problèmes de "*files d'attente*".

L'exemple le plus connu concerne la radioactivité. Un noyau de carbone-14 peut se désintégrer spontanément pour fournir un noyau d'azote\*. De tels phénomènes sont décrits par une probabilité de désintégration par unité de temps indépendante de l'histoire du noyau et de son environnement. Les constantes de temps sont très variables, de quelques fractions de millièmes de secondes, ou moins, à quelques millions d'années, ou plus. Pour le carbone-14, la constante de temps  $\tau$  est voisine de 8270 ans.

Un autre exemple peut être trouvé dans l'occurrence d'un accident sur un tronçon d'autoroute. Dans ce cas, les conditions changent avec l'heure de la journée, le jour de la semaine et bien d'autres circonstances (départ en vacances, présence policière, etc...). On peut donc considérer que  $\tau$  ne reste constant que pendant une durée limitée. Pendant cette durée la loi évoquée est pertinente. A un autre moment la même loi est encore pertinente pendant une durée limitée, mais avec une autre valeur numérique pour  $\tau$ . Dans un tel cas, on dit que "*la constante de temps varie avec le temps*". Cette phrase est curieuse : autocontradictoire mais sans ambiguïté une fois que l'on a expliqué ce qu'elle signifie.

Dans un autre contexte la loi évoquée peut décrire l'arrivée d'une commande ou celle d'un client à un quelconque guichet, la poste ou la caisse d'un super-marché par exemple. Dans la théorie des flux tendus, la commande d'une automobile déclenche le processus de fabrications ; il n'y a pas de stocks (ou alors ils sont sur les routes, acheminés souvent par camions, en petites quantités, ce qui évite l'immobilisation de capitaux privés tandis que le coût<sup>†</sup> du stockage est partiellement payé par la collectivité).

---

"probablement" car il n'est pas complètement exclu de trouver une boule " $k \geq K$ "; toutefois, cette probabilité peut être rendue arbitrairement petite (en théorie!).

\*Formé dans la haute atmosphère, le carbone-14 se trouve en très petite proportion dans la nature où le carbone  $^{12}_6C$  est le plus fréquemment rencontré. La présence de carbone-14 permet cependant de dater certains objets, d'où son intérêt en archéologie.

†Accidents, entretien des routes, pollution, etc...

Il est très important de connaître  $\tau$  car ce paramètre détermine largement les moyens à mettre en oeuvre pour apporter les secours en cas d'accident ou pour limiter l'attente des clients aux guichets ou aux caisses. En fait le but n'est pas de limiter l'attente. A une caisse l'attente doit au contraire être la plus longue possible (permettant l'utilisation d'un minimum de personnel et par conséquent, un coût minimal : c'est l'idéologie dominante aujourd'hui), mais elle ne doit pas être source d'une révolte qui ferait fuir les clients ou provoquerait une colère trop violente. Un compromis est nécessaire.

Les exemples qui précèdent montrent la diversité des raisons qui font de la loi de probabilité étudiée une loi importante. Nous nous proposons de montrer que cette loi est une loi de Poisson. Plus précisément, nous allons montrer que le nombre d'occurrences pendant une durée  $t$ , donnée, est une variable  $X$  qui suit une loi de Poisson, que le phénomène concerne l'occurrence d'une désintégration radioactive, de l'arrivée d'une commande, ou d'un client.

Avant d'aborder cette démonstration, montrons comment la pertinence du schéma proposé pour décrire la radioactivité peut être testé expérimentalement grâce à la loi des grands nombres.

A un instant donné, choisi arbitrairement comme l'origine des temps, nous disposons d'une population de noyaux radioactifs en nombre  $N_0$ . A l'instant  $t$  le nombre de noyaux qui ne sont pas désintégrés est  $N(t)$ . Chacun de ces noyaux est susceptible de se désintégrer pendant le temps  $dt$ . Nous pouvons donc considérer que se déroulent simultanément  $N(t)$  expériences de désintégration individuelle, ou encore que nous disposons de  $N(t)$  urnes dont nous tirons de chacune d'elles l'une des boules sur lesquelles est peut-être écrit "occurrence". La loi des grands nombres nous indique que "probablement" la proportion de noyaux qui se désintègrent pendant  $dt$  est voisine de la probabilité  $\frac{dt}{\tau}$  ou encore que le nombre de noyaux qui se désintègrent est "probablement" voisin de  $\delta N = N(t) \times \frac{dt}{\tau}$ . La variation de  $N$  est donc  $dN \simeq -\delta N$ . En négligeant l'erreur commise, on en déduit que  $N(t)$  satisfait l'équation différentielle  $dN = -N(t) \times \frac{dt}{\tau}$ . Avec la condition initiale  $N(0) = N_0$ , il vient

$$N(t) = N_0 e^{-t/\tau}$$

En réalité il faudrait écrire  $N(t) \simeq N_0 e^{-t/\tau}$  mais compte tenu des très grandes quantités de noyaux mis en jeu dans les expériences usuelles, la loi des grands nombres est très bien vérifiée et l'erreur  $|N(t) - N_0 e^{-t/\tau}|$  est, le plus souvent, négligeable.

La loi exponentielle ainsi obtenue est "*probable*", c'est-à-dire qu'elle est "*presque toujours*" satisfaite. Ici encore, la loi des grands nombres étant très bien vérifiée, "*presque toujours*" signifie pratiquement "*toujours*" car la probabilité pour que cette loi soit en défaut est si faible que la proportion d'expériences où la loi exponentielle est infirmée est complètement négligeable (probablement!).

La vérification de la loi de décroissance exponentielle constitue une justification du modèle probabiliste proposé.

#### 4.3.3 Nature poissonnienne des désintégrations radioactives

Nous considérons la loi de probabilité d'une occurrence  $dP = \frac{dt}{\tau}$  où  $\tau$  est une constante. Posons-nous la question suivante : Quelle est le nombre d'occurrences pendant la durée  $t$ ? Avec quelle probabilité? La durée  $t$  étant fixée, le nombre d'occurrences constitue une variable aléatoire  $X$ , dont les valeurs possibles sont les entiers non-négatifs,  $k$ .

Le processus décrit est un processus sans mémoire, on peut donc considérer que la durée  $t$  s'écoule entre 0 et  $t$  et que les occurrences survenues avant la date  $t_0 = 0$  ne comptent pas. La probabilité de  $k$  occurrences entre 0 et  $t$  est notée  $P_k(t)$ . Nous nous proposons de déterminer  $P_k(t)$ .

- Pour  $k = 0$ , on décompose l'intervalle  $(0, t)$  en  $n$  durées successives égales à  $t/n$ . On considère l'événement  $E_0$  : "absence d'occurrence pendant la durée  $t/n$ ". A la limite  $n \rightarrow \infty$ , la durée  $t/n$  est un infiniment petit noté  $dt$ . La probabilité d'une occurrence pendant  $dt$  est  $dt/\tau$ . La probabilité de l'absence d'occurrence est donc  $1 - dt/\tau$  (en effet, rappelez vous! Les seuls cas possibles sont "pas d'occurrence" ou "une occurrence" pendant  $dt$ , choisi assez petit). Les  $n$  épreuves de Bernoulli se déroulent successivement pendant le temps  $t$ , indépendamment les unes des autres. La probabilité pour que l'événement  $E_0$  survienne  $n$  fois est donc  $P_0 = (1 - dt/\tau)^n = \left(1 - \frac{t}{\tau n}\right)^n$ . En fait,  $P_0$  est la limite de cette quantité pour  $n \rightarrow \infty$ . On obtient le résultat cherché

$$P_0(t) = e^{-t/\tau}$$

- Remarquons la relation  $P_0(0) = 1$  : pendant une durée nulle, il est certain qu'aucune occurrence ne se produira. Ce résultat est satisfaisant car il correspond aux observations les plus générales, dans les domaines les plus variés. Il signifie qu'un certain temps est nécessaire pour qu'il se produise quelque chose!

La condition  $P_0(0) = 1$  implique  $P_k(0) = 0$  pour  $k \neq 0$

*Résumons* : La probabilité est nulle pour qu'il arrive quelque chose pendant une durée nulle et la probabilité est 1 pour qu'il n'arrive rien.

- Considérons  $k$  occurrences pendant la durée  $t + dt$ , entre  $t_0 = 0$  et  $t + dt$ . La probabilité en est  $P_k(t + dt)$ . Il y a deux voies indépendantes pour réaliser cet événement :

1. Suivant la première voie, les  $k$  occurrences se sont produites entre 0, et  $t$  (probabilité  $P_k(t)$ ) et aucune occurrence ne s'est produite entre  $t$  et  $t + dt$  (probabilité  $1 - dt/\tau$ ).

Ce processus a pour probabilité  $\pi_1 = P_k(t) \times \left(1 - \frac{dt}{\tau}\right)$

2. Dans la seconde voie, entre 0 et  $t$ , il s'est produit  $k - 1$  occurrences. La probabilité en est  $P_{k-1}(t)$ . Une occurrence s'est alors produite entre  $t$  et  $t + dt$  (probabilité  $dt/\tau$ ).

La probabilité de ce processus est  $\pi_2 = P_{k-1}(t) \times \frac{dt}{\tau}$ .

Il n'y a pas d'autres processus possibles car une seule occurrence au maximum, peut intervenir entre  $t$  et  $t + dt$ .

On en déduit  $P_k(t + dt) = \pi_1 + \pi_2 = P_k(t) - \left(P_k(t) - P_{k-1}(t)\right) \times \frac{dt}{\tau}$ .

Considérons la fonction  $P_k(t)$  et sa dérivée  $\frac{dP_k}{dt} = P'_k(t)$ . Au premier ordre relativement à  $dt$ , il vient  $P_k(t + dt) = P_k(t) + P'_k(t) dt$ . En identifiant les deux expressions obtenues pour  $P_k(t + dt)$ , il vient

$$P'_k(t) = -\frac{1}{\tau} \left( P_k(t) - P_{k-1}(t) \right) \quad (4.9)$$

Compte tenu des conditions initiales, la solution est

$$P_k(t) = \frac{e^{-t/\tau}}{k!} \left(\frac{t}{\tau}\right)^k \text{ pour } k > 0$$

Ainsi pour  $t \neq 0$ , la variable  $X$  (où  $X$  est le nombre d'occurrences) suit une loi de Poisson de moyenne  $m = t/\tau$ .

[[Pour calculer  $P_k(t)$ , nous posons  $P_k(t) = A_k(t) \times e^{-t/\tau}$ . L'équation 4.9 devient  $A'_k(t) = \frac{1}{\tau} A_{k-1}(t)$

Les conditions initiales s'écrivent  $P_k(0) = 0$ , soit  $A_k(0) = 0$ .

$k = 1$  :  $A'_1(t) = \frac{1}{\tau} A_0(t) = \frac{1}{\tau} e^{t/\tau} P_0(t) = \frac{1}{\tau} \Rightarrow A_1(t) = \frac{t}{\tau} + cte$ . Les conditions initiales  $A_1(0) = 0$  impliquent  $A_1(t) = \frac{t}{\tau}$ .

$k = 2$  :  $A'_2(t) = \frac{1}{\tau} A_1(t) = \frac{1}{\tau} \frac{t}{\tau} \Rightarrow A_2(t) = \frac{1}{2} \left(\frac{t}{\tau}\right)^2$ , compte tenu des conditions initiales.

Récurrence

$$\text{Propriété } P_n : A_n(t) = \frac{1}{n!} \left(\frac{t}{\tau}\right)^n.$$

La propriété  $P_1$  est vraie.

$$\text{Hypothèse de récurrence : } P_k \text{ est vraie, c'est-à-dire } A_k(t) = \frac{1}{k!} \left(\frac{t}{\tau}\right)^k.$$

Objectif : démontrer  $P_{k+1}$ .

$A_{k+1}(t)$  satisfait l'équation  $A'_{k+1}(t) = \frac{1}{\tau} A_k(t) = \left(\frac{1}{\tau} \frac{1}{k!}\right) \left(\frac{t}{\tau}\right)^k$ . Compte tenu des conditions initiales, la solution est  $A_{k+1}(t) = \left(\frac{1}{k!}\right) \frac{1}{k+1} \left(\frac{t}{\tau}\right)^{k+1} = \frac{1}{(k+1)!} \left(\frac{t}{\tau}\right)^{k+1}$ . C'est ce que nous voulions démontrer.]]

#### 4.3.4 Approximation poissonnienne de la loi binomiale

On peut étudier le cas  $X = k \neq 0$  comme on a étudié le cas  $k = 0$ . Décomposons l'intervalle  $(0, t)$  en une succession de  $n$  intervalles de durée  $dt = t/n$ . On admet que  $n$  est assez grand pour que, dans chaque intervalle  $dt$ , ne puisse survenir qu'une occurrence au plus. On réalise donc  $X = k$  en choisissant  $k$  intervalles parmi les  $n$  intervalles entre 0 et  $t$ . Ce choix s'effectue de  $C_n^k$  façons différentes. On admet l'apparition d'une occurrence dans chacun de ces  $k$  intervalles et d'aucune occurrence dans les  $n - k$  autres intervalles. La probabilité de réalisation d'un tel événement est  $\left(\frac{t/n}{\tau}\right)^k \left(1 - \frac{t/n}{\tau}\right)^{n-k}$ . On en déduit

une estimation  $\tilde{P}_k$  de la probabilité théorique  $P_k$  :  $P_k \simeq \tilde{P}_k = C_n^k \left(\frac{t/n}{\tau}\right)^k \left(1 - \frac{t/n}{\tau}\right)^{n-k}$ .

La loi de probabilité  $\tilde{P}_k$  est une loi binomiale de moyenne  $p = \frac{t/n}{\tau} \ll 1$  (voir 4.1 et 4.2). Posons  $m = t/\tau$ , c'est la moyenne de la loi de Poisson suivie par la variable  $X$  (voir la démonstration au paragraphe précédent).

Les relations  $p \ll 1$ ,  $n \gg 1$  et  $pn = m$  suggèrent que la loi binomiale puisse être assimilée à une loi de Poisson de même moyenne  $m = np$  pour  $n \gg 1$  et  $p \ll 1$  :

$$e^{-m} \frac{m^k}{k!} = P_k \simeq \tilde{P}_k = C_n^k p^k (1-p)^{n-k} \quad \text{pour } p \ll 1 \text{ et } n \gg 1 \text{ avec } m = np$$

Cette affirmation, quelque peu hâtive, souvent rencontrée dans la littérature scientifique, demande à être nuancée.

Lorsque  $p$  est assez petit ( $p < 0,1$ ), la situation est la suivante. Le rapport  $\tilde{P}_k/P_k$  est voisin de l'unité, sinon  $\tilde{P}_k$  et  $P_k$  sont assez petits pour être négligés devant l'unité.

Dans tous les cas ces deux probabilités  $P_k$  et  $\tilde{P}_k$  peuvent recevoir la même considération. C'est dans ce sens que la loi binomiale peut être remplacée par une loi de Poisson.

La figure 4-5 présente, en fonction de  $k$ , les probabilités  $P_k$  et  $\tilde{P}_k$  ainsi que le rapport  $\tilde{P}_k/P_k$  à un facteur d'échelle près. Nous avons utilisé l'approximation des factoriels par la formule de Stirling (voir page 102).

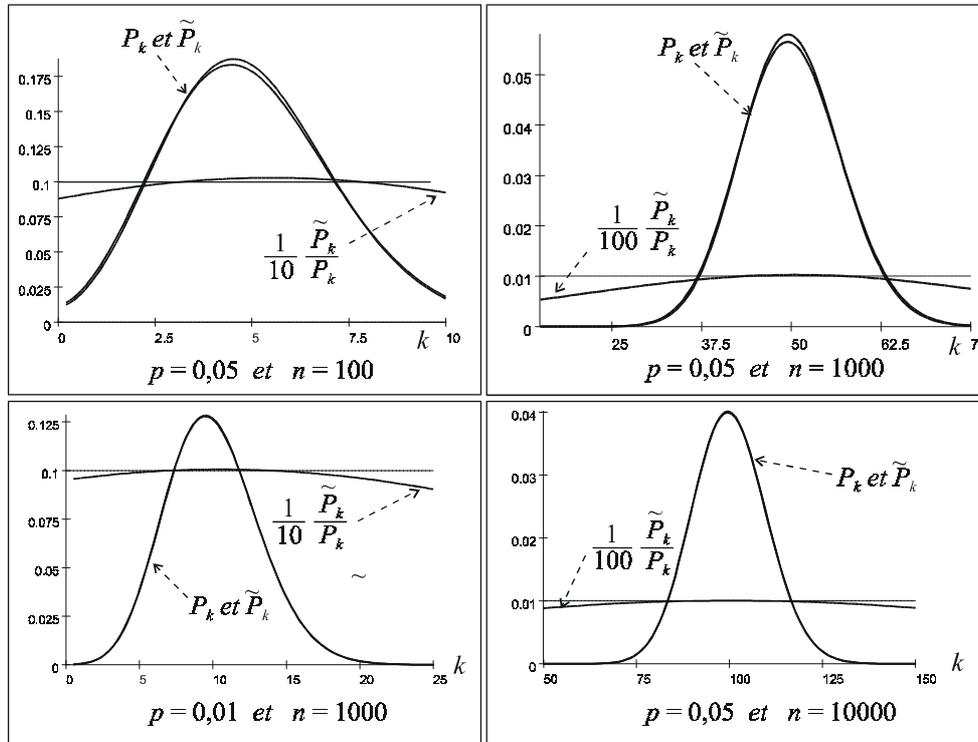


Figure 4-5

On remarquera que les probabilités sont non négligeables au voisinage de la valeur  $np$ , moyenne de la loi de Poisson.

#### 4.4 Un sujet de réflexion en guise de conclusion

Considérons une épreuve dont l'une des issues est un événement de probabilité  $p$ , éventuellement très petite (très inférieure à l'unité). Un théorème dû à Emile Borel affirme que cet événement se produira *certainement* si on répète la même épreuve indéfiniment. Un tel résultat ne se déduit pas de la loi des grands nombres. Celle-ci précise seulement que cet événement se produira *probablement* avec une fréquence voisine de  $p$  si on répète la même épreuve indéfiniment. Il est donc *improbable* que cet événement ne se produise pas. Mais "improbable" ne signifie pas "impossible". Le théorème de Borel, pour sa part, précise qu'il est impossible que cet événement ne se produise pas.

Si je joue au loto chaque semaine, et après moi l'un de mes enfants et ensuite un de ses enfants et que cette tradition se lègue de génération en génération, il est certain qu'un jour, quelqu'un gagnera (C'est le théorème d'Emile Borel (1871-1956) !), sous réserve, toutefois, que la tradition perdure jusqu'à la fin des temps et le loto aussi. Mais, les épreuves successives étant indépendantes, le fait d'avoir joué régulièrement depuis 10 générations, ne nous avance en rien vers le succès : l'infini est grand, très grand, ou comme l'écrit si justement Alphonse Allais "l'éternité c'est long. Surtout vers la fin !"



## Chapitre 5

### LA LOI NORMALE DE GAUSS

Parmi les lois de probabilité, la loi normale joue un rôle tout particulièrement important, rencontrée si fréquemment que la densité de probabilité correspondante est apparue comme "normale"...

Alors que les lois précédentes, loi binomiale et loi de Poisson, concernent des variables aléatoires discrètes, la loi normale concerne le comportement d'une variable réelle,  $X$ , susceptible de prendre une valeur  $x$ , réelle, arbitraire, appartenant à l'intervalle  $(-\infty, +\infty)$ .

#### 5.1 Présentation de la loi normale

La loi normale est décrite par sa densité de probabilité :

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

où  $\sigma$  et  $\mu$  sont des constantes ( $\sigma > 0$ ). On démontre les relations :

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} x e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \mu \quad \text{et} \quad \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \sigma^2 + \mu^2$$

Par conséquent la moyenne est  $\mu$  et la variance  $\sigma^2$  :

$$\overline{X} = E(X) = \mu, \quad \mathcal{V}(X) = \sigma^2$$

La *variable normale, centrée, réduite*  $\left( U = \frac{X - \mu}{\sigma} \right)$ , suit une loi normale, centrée (c'est-à-dire de moyenne nulle) et réduite (c'est-à-dire de variance unité). Pour le démontrer il suffit d'utiliser les relations 3.3, 3.4 et 3.5). Pour déterminer la densité de probabilité de la variable  $U$ , considérons l'événement  $E$ , tel que  $U \in [u, u + du]$ . Cet événement est identique à l'événement  $X = \sigma U + \mu \in [x = \sigma u + \mu, x + dx = \sigma u + \mu + \sigma du]$ . Utilisons les densités de probabilité des variables  $X$  et  $U$  pour exprimer la probabilité de l'événement  $E$  :

$$proba [E] = \pi(u) du = p(x) dx$$

où  $\pi(u)$  est la densité de probabilité de la variable  $U$ . En exprimant  $x$  et  $dx$  en fonction de  $u$  et  $du$ , il vient  $\pi(u) du = p(\sigma u + \mu) \sigma du$ . On en déduit la densité de probabilité d'une variable normale, centrée :

$$\pi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$$

Généralement ce sont les données numériques concernant la loi normale centrée qui fournissent les tables et les logiciels de calcul.

Soit l'intervalle  $[a, b]$  de la variable  $x$ ; la probabilité de l'intervalle  $[a, b]$  est la probabilité de l'événement  $X \in [a, b]$ . Nous posons  $\text{proba}[X \in [a, b]] = P_{[a, b]}$ . L'événement  $X \in [a, b]$  est l'événement  $U \in [u_a, u_b]$  avec  $u_a = \frac{a - \mu}{\sigma}$  et  $u_b = \frac{b - \mu}{\sigma}$ . On pose  $\text{proba}[U \in [u_a, u_b]] = \Pi_{[u_a, u_b]}$ . Il vient

$$P_{[a, b]} = \int_a^b p(x) dx = \int_{u_a}^{u_b} \pi(u) du = \Pi_{[u_a, u_b]}$$

On trouve la fonctions de répartition de la loi normale, centrée, réduite,  $f(u) = \int_{-\infty}^u \pi(t) dt$ , dans la plupart des tables, ce qui permet de calculer  $P_{[a, b]} = f(u_b) - f(u_a)$ . Cependant, c'est la fonction  $\text{erf}(u)$  qui est fournie dans la plupart des logiciels de calcul :

$$\begin{aligned} f(u) &= \int_{-\infty}^u \pi(t) dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^u e^{-\frac{t^2}{2}} dt \\ \text{erf}(u) &= \frac{2}{\sqrt{\pi}} \int_0^u e^{-t^2} dt \Leftrightarrow f(u) = \frac{1}{2} \left( 1 + \text{erf}\left(\frac{u}{\sqrt{2}}\right) \right) \end{aligned}$$

Le symbole "erf" signifie "*error function*". On retrouvera la fonction d'erreur  $\text{erf}(u)$  ou son équivalent,  $f(u)$ , plus loin lorsque nous étudierons erreurs et incertitudes de mesure. Mentionnons l'existence dans la littérature de plusieurs définitions différentes de la fonction d'erreur. C'est une source de difficultés si on n'y prête garde : on rencontre en particulier les définitions  $\text{erf}(u) = \int_0^u e^{-t^2} dt$  et  $\text{erf}(u) = f(u) - \frac{1}{2}$ .

Terminons ce bref résumé avec la propriété suivante : la somme de deux ou plusieurs variables normales indépendantes suit une loi normale et d'autre part, si  $X$  suit une loi normale, alors  $\lambda X$  suit une loi normale (voir la démonstration en annexe page 83).

On en déduit que si  $X$  et  $Y$  suivent des lois normales, il en va de même de toute combinaison linéaire de  $X$  et  $Y$ , par exemple  $X - Y$ .

En particulier, si les variables  $X_k$  suivent la même loi que la variable normale  $X$ , d'espérance  $\mu$  et d'écart-type  $\sigma$ , on démontre que  $\Sigma = \frac{1}{n} (X_1 + X_2 + \dots + X_n)$  suit une loi normale d'espérance  $\mu$  et d'écart type  $\frac{\sigma}{\sqrt{n}}$  (cf le paragraphe 3.3 et plus particulièrement les relations 3.9 et 3.11 pour l'estimation de l'espérance et de l'écart type)

## 5.2 Théorème central limite

### 5.2.1 Approximation normale de la loi binomiale

Considérons une épreuve de Bernoulli dont la variable caractéristique  $X$  prend la valeur  $X = 1$  avec la probabilité  $p$ , et la valeur  $X = 0$  avec la probabilité  $1 - p$ . Cette épreuve est répétée  $n$  fois. Soit  $Z = X_1 + X_2 + \dots + X_n$  la variable aléatoire somme des  $n$  variables aléatoires indépendantes  $X_1, X_2, \dots, X_n$  qui suivent la même loi de probabilité que la variable  $X$  (notez que  $X_k$  représente une variable aléatoire et non la valeur possible d'une variable  $X$ ). La variable  $Z$  suit la loi de probabilité binomiale que nous avons étudié au paragraphe 4.1, page 55. L'espérance de  $Z$  est  $\bar{Z} = np$  et l'écart-type de  $Z$  est  $\sigma_Z = \sqrt{n} \times \sigma$  où  $\sigma = \sqrt{p(1-p)}$  est l'écart-type de la loi suivie par la variable  $X$ .

Définissons la variable  $T = \frac{Z - np}{\sqrt{n}}$ . La variable  $T$  est une variable centrée, d'écart-type  $\sigma_T = \frac{\sigma_Z}{\sqrt{n}} = \sigma$ .

La variable  $Z$  est susceptible de prendre les valeurs entières,  $k \in [0, n]$ . Les valeurs que  $T$  est susceptibles de prendre sont donc les valeurs  $t_k$  :

$$t_k = \frac{k - np}{\sqrt{n}} \Leftrightarrow k = np + t_k \sqrt{n}$$

où  $k$  est un entier non négatif inférieur ou égal à  $n$ .

La probabilité de l'événement  $Z = k$  est donnée par l'expression 4.1. On la note  $P_k$  :

$$proba [Z = k] = P_k = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

Nous supposons la relation  $\frac{t_k}{p(1-p)\sqrt{n}} \ll 1$ . On peut démontrer la relation

$$P_k \simeq N_k \text{ avec } N_k = \frac{1}{\sqrt{2\pi n p q}} e^{-\frac{t_k^2}{2}} \tag{5.1}$$

La démonstration fait usage de la formule de Stirling ; elle est plus fastidieuse que difficile. Nous omettons cette démonstration qui ne présente pas un intérêt considérable.

En réalité, l'approximation  $P_k \simeq N_k$  est de même nature que l'approximation d'une loi binomiale par une loi de Poisson (figure 4-5 page 69).

Lorsque  $n$  tend vers l'infini,  $P_k/N_k$  est voisin de l'unité si  $k$  n'est pas très différent de la moyenne  $\bar{Z} = np$ ; dans le cas contraire,  $P_k$  et  $N_k$  sont tous deux assez petit pour être négligés devant l'unité.

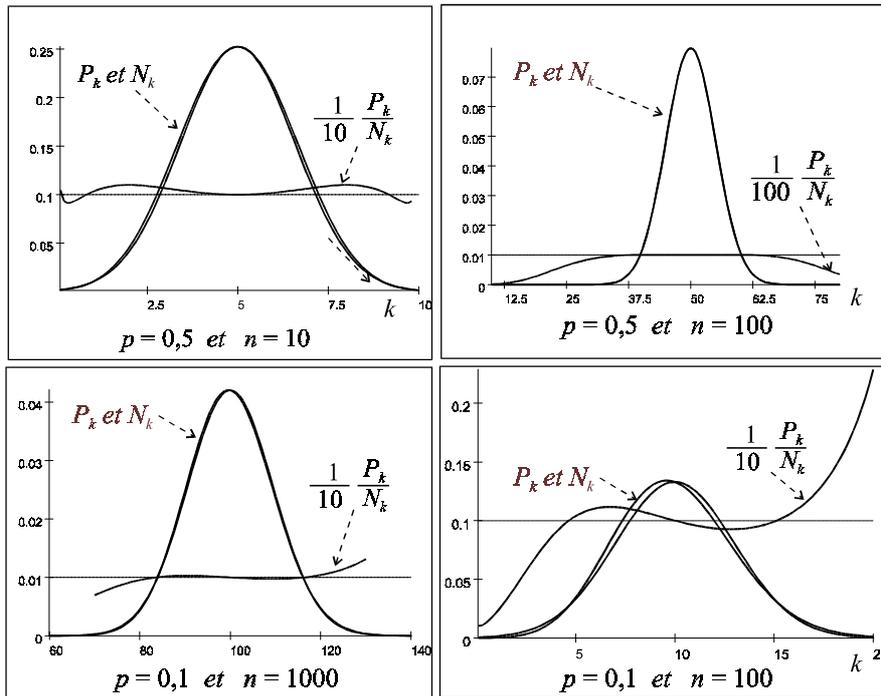


Figure 5-1

L'intervalle entre deux valeurs de  $k$  consécutives est  $k + 1 - k = 1$ , par contre, l'intervalle entre deux valeurs de  $t_k$  consécutives est  $\delta t = t_{k+1} - t_k = \frac{1}{\sqrt{n}}$ . Lorsque  $n$

tend vers  $\infty$ , on peut pratiquement considérer que  $t_k$  est une variable continue,  $t$ , dans la mesure où  $\delta t$  tend vers zéro alors que les intervalles  $\Delta t$  considérés comme significatifs restent constants.

L'exemple concret d'une variable discrète assimilable à une variable continue est donnée à la fin du paragraphe 1.2.3, page 8.

Un autre exemple d'une variable discrète assimilable à une variable continue concerne l'énergie lumineuse. A l'échelle macroscopique un microjoule est une petite quantité d'énergie  $\Delta E$ . Cette quantité élémentaire est cependant l'énergie d'un nombre entier de photons (environ 5000 milliards dans le domaine infra-rouge). C'est l'interprétation de l'effet photoélectrique par Albert Einstein qui nous a appris que l'énergie était répartie en quanta discrets dans une onde électromagnétique.

Supposons donc que nous puissions assimiler  $t_k$  à une variable continue,  $t$ . Posons nous alors la question suivante : "Quelle est la probabilité de l'événement  $T \in [t, t + dt]$ ?" . Posons  $t_{k_1} = t$  et  $t_{k_2} = t + dt$ . Il vient  $k_1 = np + t\sqrt{n}$  et  $k_2 = np + (t + dt)\sqrt{n}$ . On en déduit  $dk = k_2 - k_1 = dt\sqrt{n}$ . En réalité  $dk$  est l'entier le plus proche de cette valeur.

La probabilité de l'événement  $T \in [t, t + dt]$  est la probabilité de l'événement  $Z \in [k_1, k_2]$ , c'est la quantité  $dP = \sum_{k=k_1}^{k_2} N_k = \sum_{k_1}^{k_2} \frac{1}{\sqrt{2\pi n\sigma^2}} e^{-\frac{t_k^2}{2\sigma^2}}$ . Lorsque  $k$  varie de  $k_1$  à  $k_2$ , on considère que  $t_k$  peut être considérée comme une constante :  $t_k \simeq t$ . On en déduit  $dP = \frac{1}{\sqrt{2\pi n\sigma^2}} e^{-\frac{t^2}{2\sigma^2}} \times dk = \frac{1}{\sqrt{2\pi n\sigma^2}} e^{-\frac{t^2}{2\sigma^2}} \times dt\sqrt{n}$ . Ainsi il vient

$$\text{proba } [T \in [t, t + dt]] = dP = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{t^2}{2\sigma^2}} dt \quad (5.2)$$

La variable  $T$  suit donc, ici, une loi normale, centrée, d'écart-type  $\sigma = \sqrt{p(1-p)}$ .

### 5.2.2 Retour sur la loi des grands nombres

Nous pouvons reprendre la démonstration de la loi des grands nombres d'un point de vue différent de celui que nous avons présenté précédemment (paragraphe 4.2.2, page 57).

La fréquence relative de cas favorables observés (cas pour lesquels  $X = 1$ ) est  $\frac{k}{n} = p + \frac{t}{\sqrt{n}}$  où  $t$  suit la loi normale 5.2, de variance  $\sigma^2 = p(1-p)$ . Introduisons la variable  $U = T/\sqrt{n}$  et posons  $u = \frac{t}{\sqrt{n}}$ . Il vient  $\boxed{\frac{k}{n} = p + u}$  où  $\frac{k}{n}$  est la valeur prise par la variable aléatoire  $F = U + p$ .

La probabilité de l'événement  $U \in (u, u + du)$  est la probabilité de l'événement  $T \in (t, t + dt)$  avec  $t = u\sqrt{n}$  et  $dt = du\sqrt{n}$ . Cette probabilité est donnée par l'expression 5.2 :  $dP = \frac{\sqrt{n}}{\sqrt{2\pi\sigma^2}} e^{-\frac{u^2}{2(\sigma/\sqrt{n})^2}} du$ . Représentons la densité de probabilité de la variable  $u$  pour diverses valeurs de  $n$  (figure 5-1)

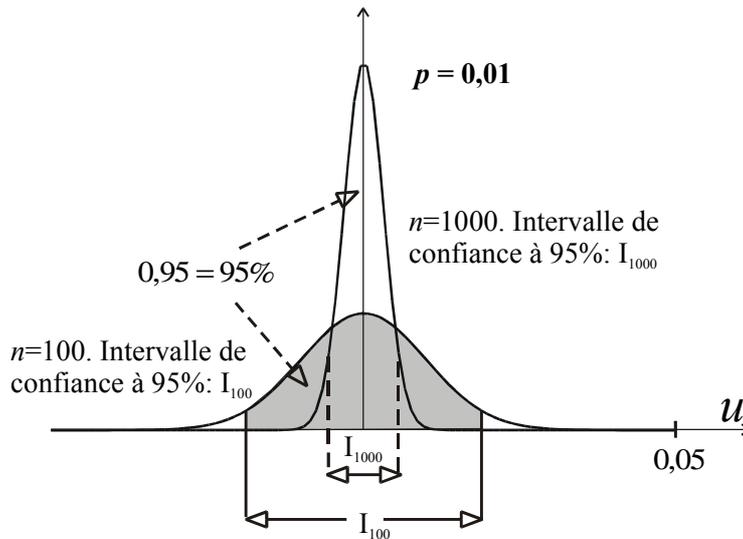


Figure 5-1

Nous constatons que la dispersion diminue lorsque  $n$  croît. L'intervalle  $I_{100}$  est un intervalle\* dont la probabilité est 95% pour  $n = 100$  c'est-à-dire que  $\text{proba}[U \in I_{100}] = 0,95$ .

Lorsque  $n$  passe de la valeur 100 à la valeur 1000, l'intervalle à 95% devient  $I_{1000}$ , d'amplitude très inférieure à  $I_{100}$ . En posant  $\frac{k}{n} = \tilde{p}$  (estimation de  $p$ ) nous commettons une erreur  $u = \frac{k}{n} - p$ . Selon toute vraisemblance (probabilité de 95%), l'erreur diminue ( $\frac{I_{100}}{2} \searrow \frac{I_{1000}}{2} < \frac{I_{100}}{2}$ ) quand  $n$  augmente (de 100 à 1000).

La dispersion devient arbitrairement petite lorsque  $n$  devient arbitrairement grand. Cela signifie qu'ayant choisi un niveau de confiance,  $\pi_s$ , arbitrairement voisin de 1 (par exemple 99,99...9% au lieu de 95%), il existe une valeur de  $n$  assez grande pour que l'intervalle de confiance correspondant,  $I$ , soit arbitrairement petit et par conséquent que l'erreur soit très probablement négligeable. On retrouve ainsi la loi des grands nombres.

**Exemple.** Reprenons l'exemple traité au paragraphe 4.2.3 page 60. Nous estimons  $p \sim 0,06$ . Nous voulons déterminer  $p$  avec une erreur inférieure à  $\varepsilon = 0,01$  au seuil de confiance  $\pi_s = 90\%$ . Le résultat que nous avons obtenu est  $n \geq \frac{p(1-p)}{\varepsilon^2(1-\pi_s)} \sim 6000$ .

Ainsi pour atteindre les objectifs fixés, nous envisageons de contrôler 6000 pièces, issues au hasard de l'ensemble de la production.

En utilisant l'approximation normale de la loi binomiale, l'estimation de  $n$  s'effectue ainsi.

L'intervalle de confiance de la variable  $U$  au seuil  $\pi_s$  est l'intervalle  $(-u_s, u_s)$  tel que

$$\int_{-u_s}^{u_s} \frac{\sqrt{n}}{\sqrt{2\pi\sigma^2}} e^{-\frac{u^2}{2(\sigma/\sqrt{n})^2}} du = \pi_s \text{ avec } \sigma = \sqrt{p(1-p)}$$

\*L'intervalle de confiance au seuil  $\pi_s$  donné est l'intervalle  $[a, b]$  de probabilité  $\pi_s$ , tel que  $b - a$  soit minimal. On démontre que cette relation implique  $p(a) = p(b)$  où  $p(x)$  est la densité de probabilité.

Posons  $u = x\sigma/\sqrt{n}$ , il vient  $\pi_s = \int_{-x_s}^{x_s} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$ . Ainsi  $x_s$  est donné, à partir de  $\pi_s$ , par la lecture des tables de la loi normale centrée, réduite. Ici,  $\pi_s = 0,9 \Rightarrow x_s = 1,645$ .

L'erreur acceptable est  $u_s = x_s\sigma/\sqrt{n} = \varepsilon = 0,01$ . On en déduit  $n = \left(\frac{x_s\sigma}{\varepsilon}\right)^2$ . En utilisant l'estimation de  $p$  pour calculer  $\sigma$ , il vient  $n = \left(\frac{1,645}{0,01}\right)^2 \times 0,06 \times 0,94 \sim 1500$ . En exploitant notre connaissance du comportement asymptotique de la loi binomiale, on réduit considérablement l'estimation du travail de contrôle (de 6000 à 1500 pièces à contrôler)..

### 5.2.3 Théorème centrale limite

La variable  $X$  de l'épreuve de Bernoulli considérée ci-dessus (paragraphe 5.2.1) est susceptible de prendre les valeurs 1 ou 0. La variable  $Z$ , somme de  $n$  variables de Bernoulli, est susceptible de prendre les valeurs entières  $k$ , tandis que la variable  $T$  prend les valeurs  $t_k = (k - np)/\sqrt{np}$ . La variable aléatoire  $Z$  suit une loi binomiale.

Considérons le cas où  $n$  est très grand. *La variable  $Z$  est la somme d'un très grand nombre de variables aléatoires élémentaires indépendantes.* Chacune de ces variables suit la même loi de probabilité que  $X$  et admet donc comme écart-type  $\sigma = \sqrt{p(1-p)}$ , tandis que l'écart-type de la variable  $Z$  est  $\sigma_Z = \sqrt{np(1-p)}$ .

Si  $n$  est assez grand, il vient  $\sigma_Z \gg \sigma$  : *l'écart type de la variable  $Z$  est beaucoup plus grand que l'écart type de chacune des variables élémentaires dont  $Z$  est la somme.*

En assimilant les variables discrètes,  $T$  et  $Z$ , à des variables continues, nous avons montré que, dans ces conditions, celles-ci suivent une loi normale. Ainsi,  *$Z$  est "probablement" une variable normale à la limite où  $n \rightarrow \infty$ .* Cette propriété constitue le théorème de Moivre-Laplace.

La propriété que nous venons d'énoncer se généralise. Nul n'est besoin de supposer que l'épreuve composée est la répétition de la même épreuve de Bernoulli. De façon générale **le théorème central limite s'énonce ainsi :**

**Lemma 1** *soit  $Z = X_1 + X_2 + \dots + X_n$ , une variable aléatoire, somme de  $n$  variables aléatoires **indépendantes**, d'écart-type  $\sigma_1, \sigma_2, \dots, \sigma_n$ . Nous supposons que, pour tout  $j$ , l'écart-type,  $\sigma_Z$ , de la variable  $Z$  satisfait la relation  $\sigma_Z \gg \sigma_j$ . Dans ces conditions, la loi de probabilité suivie par  $Z$  est presque certainement, **une loi normale** à la limite  $n \rightarrow \infty$ .*

C'est cette propriété qui est connue comme le théorème central limite (Nous admettons ce théorème dont on peut donner une démonstration numérique expérimentale sans grandes difficultés.).

L'importance du théorème central limite tient à sa généralité d'une part et à sa validité pour la description des erreurs de mesures.

## 5.3 Incertitudes et erreurs

On distingue deux familles de mesures.

1. Les mesures de la première famille sont celles que décrivent les axiomes de la mécanique quantique. Le résultat de la mesure n'est pas fixé à l'avance, même de façon approximative. Divers résultats sont possibles et seule la probabilité de chacun d'entre eux est déterminée.
2. Les mesures de la seconde famille sont les mesures classiques que l'on rencontre habituellement en physique. Une caractéristique d'un système physique étant bien

définie, on se propose de la mesurer. Par exemple, le nombre d'abeilles dans telle ruche de mon élevage ou la masse du canard que j'ai acheté au marché.

Parmi ces mesures classiques, on en distingue deux types. Celles que l'on ne peut pas répéter et celles que l'on peut répéter à satiété.

2-a) Les mesures "destructives", qui entraînent la "destruction" du système mesuré sont des mesures du premier type. Celles qui demandent l'ouverture de la boîte crânienne et l'auscultation du cerveau d'une souris pour apprécier l'efficacité de tel produit, par exemple ; ou encore celles dont le but est la mesure de la longueur de pénétration de tel type de projectile, dans un blindage.

On peut aussi classer dans cette catégorie les mesures dont la répétition est impossible dans les mêmes conditions, compte-tenu des dispositifs mis en oeuvre (la mesure de la distance Terre-Lune repose sur la mesure du temps de vol aller et retour, d'un signal lumineux. Lorsqu'on répète la mesure, les conditions sont modifiées : les positions, l'indice de l'atmosphère, etc...)

Soit  $g_{vrai}$  la vraie valeur de la grandeur à mesurer et  $g$  la mesure que nous en avons faite. En posant  $g_{vrai} = g$ , nous commettons **une erreur**  $\delta g \stackrel{déf}{=} g - g_{vrai}$ . L'erreur est inconnue.

Si notre seule information est le résultat d'une seule mesure, nous cherchons à estimer un majorant de l'erreur (aussi petit que possible), qui est noté  $\Delta g$ . On écrit alors le résultat sous l'une des trois formes équivalentes<sup>†</sup>

$$\boxed{g_{vrai} = g \pm \Delta g} \Leftrightarrow \boxed{|\delta g| = |g - g_{vrai}| \leq \Delta g} \Leftrightarrow \boxed{g - \Delta g \leq g_{vrai} \leq g + \Delta g} \quad (5.3)$$

$\Delta g$  est l'**incertitude absolue** sur la mesure. *L'incertitude absolue est connue* ; c'est tout au moins le but de l'analyse et du calcul des erreurs.

Remarquons que  $g_{vrai}$ ,  $g$ ,  $\delta g$  et  $\Delta g$  ont mêmes dimensions physiques.

2-b) Les mesures classiques usuelles consistent à déterminer au moyen d'un appareil de mesure (appareil ou appareillage) la grandeur  $g_{vrai}$  inconnue dans des conditions où la mesure peut être répétée. Pour la mesure d'une grandeur, on dispose donc d'un nuage de points (à une dimension). Chacun des points représente le résultat d'une mesure.

Tous les résultats sont différents car il n'est pas possible de reproduire exactement les mêmes conditions physiques : la pression atmosphérique et la température ont été modifiées, la position de l'oeil qui lit le résultat n'est plus exactement la même...

C'est cette catégorie de mesures que nous considérons ici.

De façon générale, les diverses mesures effectuées présentent une certaine dispersion autour de la valeur  $g_{vrai}$ .

- On admet que  $\delta G = G - g_{vrai}$  est une variable aléatoire, dont la valeur,  $\delta g = g - g_{vrai}$ , est la somme de nombreuses perturbations. Chacune d'elles est considérée comme (la valeur d')une variable aléatoire indépendante des autres. Cette hypothèse est généralement bien vérifiée.

[[En effet, supposons que  $\delta g$  soit une fonction des grandeurs physiques indépendantes  $\alpha_1, \alpha_2, \dots, \alpha_n$  (la température, le champ magnétique ambiant, etc...). Il vient  $\delta g = \delta g(\alpha_1, \dots, \alpha_n)$ . Les valeurs  $\alpha_k$  ne reste pas constante d'une mesure à l'autre. On pose  $\alpha_k = \overset{0}{\alpha_k} + \delta\alpha_k$ . La grandeur  $\overset{0}{\alpha_k}$  reste constante, tandis que  $\delta\alpha_k$  fluctue. Le plus souvent  $\delta\alpha_k$  est assez petit pour que l'on puisse assimiler  $g - g_{vrai}$  à un développement limité au premier ordre relativement aux

---

<sup>†</sup>C'est par une convention d'écriture que la première égalité est équivalente aux deux autres.

diverses fluctuations :

$$g - g_{vrai} = \delta g \binom{0}{\alpha} + \sum_{k=1}^n a_k \times \delta \alpha_k \quad \text{avec} \quad a_k = \left( \frac{\partial \delta g}{\partial \alpha_k} \right) \binom{0}{\alpha}$$

où  $\binom{0}{\alpha}$  représente la suite des valeurs  $(\alpha_1, \dots, \alpha_n)$ .

En l'absence d'erreur systématique, il vient  $\delta g \binom{0}{\alpha} = 0$  et  $E[\delta \alpha_k] = 0$ . Dans ces conditions la variable aléatoire  $\delta G$  apparaît comme la somme des variables aléatoires indépendantes  $X_k = a_k \times \delta \alpha_k$  dont l'espérance mathématique est nulle. Les coefficients  $a_k$  sont des constantes, tandis que les  $\delta \alpha_k$  sont les valeurs de variables aléatoires indépendantes, dont l'écart type est  $\sigma_{\alpha_k}$ . La variable  $X_k$  admet donc comme écart type  $\sigma_k = |a_k| \times \sigma_{\alpha_k}$ . L'écart type de la variable  $\delta G$  est donc  $\sigma = \sqrt{\sigma_1^2 + \sigma_2^2 + \dots + \sigma_n^2}$ . Par la suite, on admet la relation  $\sigma_k \ll \sigma$ .]]

• La mesure de  $G$  apparaît comme une épreuve probabiliste dont le résultat est  $g$ . A l'issue de cette épreuve, la variable aléatoire  $\delta G = G - g_{vrai}$  prend la valeur  $\delta g$ . Cette valeur fluctue d'une mesure à l'autre. On admet que l'origine des fluctuations de  $\delta g$  n'est pas principalement due à un seul paramètre mais que de nombreux paramètres interviennent. Cela signifie que l'écart type de  $\delta G$  est beaucoup plus grand que l'écart-type de chacun des paramètres physiques qui influencent le résultat.

Cette situation est généralement satisfaite. Supposons en effet que ce ne soit pas le cas et que  $g$  soit très sensible à la température,  $\theta$ , par exemple, si sensible que les fluctuations de  $\theta$  sont pratiquement les seules fluctuations observées. L'appareil se comporte comme une sorte de thermomètre alors que nous voulons mesurer tout autre chose que  $\theta$ . Pour éviter de telles fluctuations aléatoires, nous stabilisons la température au moyen d'un thermostat. C'est faisable jusqu'au point où les fluctuations de température ne peuvent plus être remarquées parmi l'ensemble des fluctuations des divers paramètres (et, par conséquent, ne peuvent pas être mieux corrigées). Les fluctuations de la température cessent alors d'être prépondérantes. La fluctuation de  $\delta g$ , d'une mesure à l'autre, n'est plus due principalement à la fluctuation de la température.

**Les hypothèses précédentes apparaissent donc comme des hypothèses très généralement satisfaites** (ou que l'on cherche à satisfaire) ; **ce sont en outre ces hypothèses qui nous assurent que  $\delta G$  suit une loi normale.**

En l'absence d'erreur systématique, l'espérance mathématique de  $\delta G$  est nulle ; c'est précisément ce que l'on entend par "absence d'erreur systématique". Soit  $\sigma$  l'écart-type de  $\delta G$ . La variable  $U = \delta G / \sigma$  suit une loi normale, centrée (d'espérance nulle), réduite (d'écart-type unité).

### 5.3.1 Incertitude standard, intervalle de confiance

On suppose ici que les fluctuations de  $\delta g$  suivent une loi normale de moyenne nulle et d'écart-type  $\sigma$ . On donne le résultat de la mesure sous la forme

$$g_{vrai} = g \pm \sigma(\text{rms}) \quad (5.4)$$

où  $g$  est le résultat mesuré et  $g_{vrai}$  la vraie valeur de la variable  $G$ .

Une telle expression a la même forme que la première relation 5.3. Son interprétation est cependant très différente.

La relation 5.4 ne signifie pas que  $|g_{vrai} - g|$  est inférieur à  $\sigma$  mais seulement que la probabilité d'un tel résultat est 0,68.

L'écart-type  $\sigma$  est appelé "**incertitude standard**". L'intervalle  $(-\sigma, \sigma)$  est l'intervalle d'erreur standard ou **intervalle de confiance à 68%** ou encore en argot de physicien "barre d'erreur".

Nous avons utilisé l'expression (*rms*) pour désigner l'incertitude standard. Dans la littérature scientifique, on trouve par exemple "une incertitude de  $10^{-8}$  m (*rms*)". Cela

signifie "root mean square" ou encore en français "écart quadratique moyen". L'usage en est fait lorsque l'incertitude est estimée à partir de mesures assez nombreuses pour que, dans le cas considéré,  $10^{-8}$  m apparaisse comme une estimation de l'incertitude standard (écart-type d'une loi pratiquement assimilable à une loi normale).

On pose  $\delta g = u\sigma$  où  $u$  est la valeur de la variable aléatoire  $U$  introduite à la fin du paragraphe précédent. Les valeurs de  $u$  se distribuent suivant une loi normale centrée, réduite. On écrit encore le résultat d'une mesure sous la forme

$$g_{vrai} = g \pm u\sigma \text{ (au seuil de confiance } \pi)$$

Par exemple  $g_{vrai} = g \pm 0,045$  (au seuil de confiance 90%). Cela signifie qu'en répétant les mesures, 9 fois sur 10 on aura  $|g_{vrai} - g| \leq 0,045$ . En d'autres termes, si on assimile la grandeur inconnue  $g_{vrai}$  au résultat connu de la mesure,  $g$ , la probabilité d'une erreur inférieur à 0,045 est 0,9.

Nous posons  $0,045 = u\sigma$ . L'écart-type  $\sigma$  est inconnu mais  $u$  peut être déterminé par la condition que  $\text{proba}[U \in (-u, u)] = 0,90$ . Les tables de la loi normale, centrée, réduite fournissent alors  $u = 1,65$ . On en déduit  $\sigma = \frac{0,045}{1,65} = 2,7 \times 10^{-2}$  :

$$g_{vrai} = g \pm 0,045 \text{ (au seuil de confiance 90\%)} \Leftrightarrow g_{vrai} = g \pm 2,7 \times 10^{-2} \text{ (rms)}$$

Si on veut approcher la certitude et choisir un seuil de confiance de 99%, il convient de poser  $u = 2,6$ ; on écrit alors  $g_{vrai} = g \pm 2,6 \sigma$  (au seuil de 99%). L'intervalle  $[g - 2,6 \sigma, g + 2,6 \sigma]$  est l'intervalle de confiance à 99%. Ici,  $2,6 \sigma = 0,07$ . On est donc pratiquement certain (probabilité 0,99) que l'erreur est inférieure à 0,07.

Le plus souvent on considère l'intervalle de confiance à 95%. Il vient  $g_{vrai} = g \pm \delta$  avec  $\delta = u_{0,95} \times \sigma$  et  $u_{0,95} = 1,65$  (voir la table de la loi normale chapitre 7). L'intervalle  $[g - \delta, g + \delta]$  est communément appelé une "**fourchette**".

*N.B.* La fonction de répartition,  $f(u)$ , de la loi normale, centrée, réduite est donnée dans les tables de tout ouvrage de statistiques. On démontre la relation  $f(u) = \frac{1 + \pi}{2}$ .

Considérons deux mesures différentes<sup>‡</sup> de la même grandeur pour le même seuil de confiance standard de 68% :  $g_{vrai} = 2,75 \pm 0,03$  et  $g_{vrai} = 2,81 \pm 0,02$ . La première mesure implique que "vraisemblablement" on a  $g_{vrai} \leq 2,78$  tandis que la seconde mesure implique le résultat "vraisemblable"  $g_{vrai} \geq 2,79$ . Ces deux résultats sont incompatibles mais "vraisemblable" ne signifie pas "certain" et il n'est pas exclu que les erreurs soient supérieures aux valeurs standard 0,03 et 0,02.

Les valeurs standard sont obtenues pour  $u = 1$ . En posant  $u = 1,5$  par exemple, il vient (au seuil 87% correspondant à  $u = 1,5$ )  $g_{vrai} = 2,75 \pm 0,045$  et  $g_{vrai} = 2,81 \pm 0,03$  : ces résultats ne sont plus incompatibles.

Comment choisir le niveau de confiance  $\pi$  ? Rappelons qu'il n'y a pas de réponse universelle à cette question (cf. page 61). Ici, la probabilité d'une erreur assez grande pour que les résultats soient compatibles, n'est pas si petite que le cas en serait invraisemblable. Il s'en suit que la décision à prendre ne s'impose pas à l'évidence.

Considérons une situation semblable à la situation précédente où nous supposons cependant que les incertitudes standard sont beaucoup plus petites : les résultats des deux mesures au seuil de confiance standard sont maintenant  $g_{vrai} = 2,75 \pm 0,01$  et  $g_{vrai} = 2,81 \pm 0,01$ . Pour que les intervalles de confiance se recouvrent il faut poser

<sup>‡</sup>Ici les procédés pour effectuer les deux mesures sont différents, ce qui explique la différence des incertitudes standard.

$u > 3$ . En effet, en posant  $u = 3$  la première mesure donne  $|g_{vrai} - g| < 2,78$  et la seconde  $|g_{vrai} - g| > 2,78$ . Les résultats ne sont donc pas incompatibles pour  $u > 3$ . Cependant la probabilité d'une erreur si grande que l'on ait  $u > 3$  est faible (elle est donnée par la table de la loi normale et vaut  $2(1 - f(u)) \simeq 2,7 \times 10^{-3}$  environ). Nous avons donc le choix entre admettre que les résultats sont compatibles mais exceptionnels ou que les résultats ne sont pas compatibles et que, par conséquent, l'une des deux mesures au moins a donné un résultat erroné. Cette conclusion est la plus vraisemblable des deux, aussi faudra-t-il refaire les expériences et tenter de comprendre pourquoi l'une d'entre elles donnait un résultat inacceptable.

### 5.3.2 Estimation d'une espérance mathématique : intérêt de la multiplication des mesures

Dans les développements précédents, nous avons supposé connue l'incertitude standard,  $\sigma$ . La question se pose de l'estimation de  $\sigma$ . C'est là qu'intervient la répétabilité des expériences considérées.

Effectuons une campagne de  $n$  mesures de la même grandeur  $G$ . Les observations constituent la série statistique  $\{g_1, g_2, \dots, g_n\}$ .

La somme de cette série est la valeur observée de la variable aléatoire  $\Sigma = G_1 + G_2 + \dots + G_n$  où les  $G_k$  forment un ensemble de  $n$  variables aléatoires indépendantes, qui suivent la même loi de probabilité que  $G$ .

La moyenne de la série,  $\frac{1}{n}(g_1 + g_2 + \dots + g_n) = \bar{g}_n$ , est la valeur de la variable aléatoire  $\frac{\Sigma}{n} = Z$ . Cette valeur fluctue d'une campagne de mesures à l'autre.

Les résultats du paragraphe 3.3 donnent l'espérance,  $E[Z]$  de  $Z$  et son écart type,  $\sigma_Z$  :

$$E[Z] = E[G] \quad \text{et} \quad \sigma_Z = \frac{\sigma}{\sqrt{n}}$$

où  $E[G]$  est l'espérance mathématique de  $G$  et  $\sigma$  l'écart type de  $G$ .

Admettons l'hypothèse d'une distribution normale des erreurs. Soit  $g_1$  le résultat d'une mesure isolée et  $\bar{g}_n$  la moyenne observée lors d'une campagne de  $n$  mesures, il vient :

$$E[G] = g_1 \pm \sigma \quad \text{et} \quad E[G] = \bar{g}_n \pm \frac{\sigma}{\sqrt{n}} \quad (5.5)$$

au seuil de confiance standard de 68%.

**On divise l'incertitude standard par le facteur  $\sqrt{n}$  en remplaçant une mesure unique par la moyenne de  $n$  mesures identiques, indépendantes.**

Remarquons qu'aucune hypothèse particulière n'a été postulée ici (hypothèse de normalité ou autre). En choisissant  $n$  assez grand, on réduit la variance de  $Z$  à une valeur arbitrairement petite. De l'inégalité de Bienaymé-Tchebychev (4.4) on déduit

$$\text{proba} [|Z - E[G]| \leq t\sigma_Z] > 1 - \frac{1}{t^2} \quad (5.6)$$

On choisit  $t$  de telle sorte que  $\eta = \frac{1}{t^2}$  soit arbitrairement voisin de 0. On choisit alors  $n$  de telle sorte que  $\varepsilon = t\sigma_Z = \frac{t\sigma}{\sqrt{n}}$  soit arbitrairement petit. L'estimation de  $E[G]$  est alors arbitrairement bonne.

[[L'inégalité de Bienaymé-Tchebychev s'écrit  $\text{proba} [|Z - E[Z]| > t\sigma_Z] < \frac{1}{t^2}$ .  
Or,  $Z = \frac{1}{n}(G_1 + \dots + G_n) \Rightarrow E[Z] = E[G]$  d'où  $\text{proba} [|Z - E[G]| > t\sigma_Z] < \frac{1}{t^2}$ .

La campagne de  $n$  mesures est une épreuve probabiliste et la réalisation de  $|Z - E[G]| > t\sigma_Z$  constitue un événement de probabilité  $P$ . Le complémentaire de cet événement est l'événement  $|Z - E[G]| \leq t\sigma_Z$ ; sa probabilité est  $1 - P$ . La relation  $P < \frac{1}{t^2}$  implique donc  $1 - P > 1 - \frac{1}{t^2}$ . On en déduit donc la relation 5.6]]

**De façon générale, l'espérance  $E[G]$  d'une variable aléatoire quelconque  $G$ , peut être estimée en prenant la moyenne des valeurs  $g_1, g_2, \dots, g_n$  obtenues lors d'une campagne de  $n$  mesures de  $G$ . La probabilité d'une erreur arbitrairement petite sur l'estimation de  $E[G]$  peut être rendue aussi grande que l'on veut pour  $n \rightarrow \infty$ .**

### 5.3.3 Estimation d'une espérance mathématique : erreur systématique et biais

Considérons une grandeur physique dont la vraie valeur est  $g_{vrai}$ . La mesure de cette grandeur est une variable aléatoire,  $G$  dont la valeur observée varie d'une mesure à l'autre.

Pour déterminer  $E[G]$ , nous effectuons une campagne de  $N$  mesures. La moyenne observée est  $\bar{g}_N$ . En prenant  $N$  aussi grand que nécessaire, nous pouvons atteindre une précision et une confiance arbitrairement grande dans la mesure de  $E[G]$  mais nous ne savons pas si  $E[G]$  est voisin de  $g_{vrai}$  ou non.

L'**erreur systématique** est la quantité  $\Delta_S G = E[G] - g_{vrai}$ . La grandeur  $\Delta_S G$  est encore appelé "**biais**".

L'erreur systématique liée au comportement d'un appareil peut être connue dans une opération d'étalonnage en laboratoire. Une telle opération consiste en effet à mesurer  $G$  pour des valeurs connues de  $g_{vrai}$ . Les mesures pouvant être répétées à l'infini, on connaît  $E[G]$  aussi bien que l'on veut et par conséquent  $\Delta_S G$ .

Lors d'une mesure destinée à déterminer  $G$ , il suffit alors de rajouter  $\Delta_S G$  à la quantité estimée  $E[G]$  que l'on observe. On obtient ainsi une estimation sans biais.

Cependant la procédure indiquée n'est efficace que dans la mesure où l'appareil est **fidèle** (c'est-à-dire que  $g_{vrai}$  étant donné, l'erreur systématique produite est toujours la même).

Ayant vu comment évaluer l'erreur systématique, nous pouvons admettre qu'il en est tenu compte dans le résultat affiché des mesures. Nous supposons donc que les mesures effectuées ne présentent aucune erreur systématique. Cela signifie que  $E[G] = g_{vrai}$ . Dans ces conditions les relations 5.5 deviennent :

$$g_{vrai} = g_1 \pm \sigma \quad \text{et} \quad g_{vrai} = \bar{g}_n \pm \frac{\sigma}{\sqrt{n}} \quad (5.7)$$

où les incertitudes sont les incertitudes standards. Remarquons que  $\sigma$  est, ici, supposé connu. Nous verrons comment l'estimer au paragraphe 5.3.4, ci-dessous.

Pour terminer ce paragraphe, précisons de façon générale ce que l'on entend par "**estimation sans biais**".

Nous sommes fréquemment conduit à estimer des grandeurs à partir des séries statistiques des observations effectuées. Les séries statistiques qui permettent les estimations sont formées par les valeurs observées de variables aléatoires. L'estimation  $\tilde{F}$  d'une grandeur  $F$  est une combinaison des valeurs observées, c'est donc une variable aléatoire. On dit que  $\tilde{F}$  est une estimation sans biais de  $F$  si l'espérance de  $\tilde{F}$  est égale à  $F$ , soit  $E[\tilde{F}] = F$ . La loi des grands nombres nous assure que la multiplication des mesures de  $\tilde{F}$  permet d'estimer  $E[\tilde{F}]$  aussi bien que l'on veut. On peut apprécier l'importance qu'il y a à disposer d'une estimation sans biais des grandeurs que l'on cherche à connaître.

### 5.3.4 Estimation de l'incertitude standard dans l'étalonnage d'un appareil

Nous savons comment estimer  $g_{vrai}$  à partir des résultats de mesures, en supposant connu l'écart-type  $\sigma$  de  $G$ ; mais nous ne savons toujours pas comment estimer  $\sigma$ . Cette quantité est une caractéristique essentielle des appareil de mesure quand ils sont responsables des erreurs de mesure.

Considérons une opération d'étalonnage où la valeur  $g_{vrai}$  de  $G$  est connue. Soit  $g$  le résultat d'une mesure. Posons  $X = (g - g_{vrai})^2$ . Par définition l'espérance  $E[X]$  est égale à la variance de  $G$ , d'où  $\sigma^2 = E[X]$ .

Pour estimer  $E[X]$ , nous effectuons une campagne de  $N$  mesures dont les résultats sont les valeurs des variables aléatoires  $X_1, X_2, \dots, X_N$  qui suivent toutes la même loi de probabilité que  $X$ . La variable aléatoire  $Z = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$  prend la valeur  $z = \frac{1}{n}(x_1 + x_2 + \dots + x_n)$  où  $x_k$  est la valeur de  $X_k$  : c'est-à-dire  $x_k = (g_k - g_{vrai})^2$ . La variable  $Z$  est une estimation sans biais de  $E[X]$  qui peut être aussi bonne que l'on veut en choisissant  $n$  assez grand (voir le paragraphe 5.3.2).

$Z$  est donc une estimation sans biais de la variance,  $\sigma^2$ , de  $G$ . On utilise donc l'estimation suivante de l'écart-type,  $\sigma$ , de  $G$  :

$$\sigma \simeq \sqrt{\frac{1}{n} \sum_k (g_k - g_{vrai})^2}$$

Ce sont ces valeurs de  $\sigma$ , établis par étalonnage, qui caractérisent l'incertitude standard de l'appareil de mesure.

### 5.3.5 Estimation de l'incertitude standard dans une campagne de mesures

Dans le cas général,  $g_{vrai}$  est inconnu et  $\sigma$  doit être déduit des mesures elles-mêmes.

La mesure est répétée  $n$  fois. L'estimation de  $g_{vrai}$  est  $\bar{g}_n = \frac{1}{n}(g_1 + g_2 + \dots + g_n)$ . Il est légitime d'introduire l'écart quadratique moyen de la série statistique des observations :  $y = \frac{1}{n} \sum_k (g_k - \bar{g}_n)^2$ . Remarquons la relation  $y = s_n^2$  où  $s_n$  est l'écart quadratique moyen de la série statistique des observations  $\{g_1, g_2, \dots, g_n\}$ .

La valeur  $s_n^2$  n'est pas une estimation sans biais de  $\sigma^2$ . On utilise l'estimation

$$\sigma \simeq \sqrt{\frac{n}{n-1}} \times s_n \quad \text{avec} \quad s_n^2 = \frac{1}{n} \sum_k (g_k - \bar{g}_n)^2 \quad (5.8)$$

Remarquons que pour  $n \gg 1$ , l'estimation  $\sqrt{\frac{n}{n-1}} \times s_n$  est voisine de  $s_n$ .

[[ La mesure  $n^{\circ}k$  de  $G$  est une variable aléatoire que l'on note  $G_k$ . Soit  $g_k$  le résultat de cette mesure. La quantité  $x_k = g_k - g_{vrai}$  est la valeur qu'a pris la variable aléatoire  $X_k = G_k - g_{vrai}$ .

Les variables  $X_k$  sont indépendantes et suivent la même loi de probabilité. On admet que  $G$  est une estimation sans biais de  $g_{vrai}$ . On en déduit  $E[G] = g_{vrai}$  d'où  $E[X_k] = 0$  et par conséquent  $E[X_k^2] = \sigma^2$  où  $\sigma^2$  est la variance de  $G$ .

On introduit la variable aléatoire  $Y = \frac{1}{n} \sum_k \left( X_k - \frac{X_1 + X_2 + \dots + X_n}{n} \right)^2$ . On vérifie que la valeur de  $Y$  est  $y = \frac{1}{n} \sum_k (g_k - \bar{g}_n)^2 = \frac{1}{n} \sum_k \left( x_k - \frac{x_1 + x_2 + \dots + x_n}{n} \right)^2$ .

On exprime  $Y$  sous la forme

$$Y = \frac{1}{n^3} \sum_k \left( (n-1) X_k - (X_1 + \dots + X_{k-1} + X_{k+1} + \dots + X_n) \right)^2$$

Les variables  $X_i$  et  $X_j$  sont indépendantes, d'espérance mathématique égale à zéro. On en déduit  $E[X_i X_j] = \text{cov}[X_i, X_j] = 0$  et par conséquent :

$$E[Y] = \frac{1}{n^3} \sum_k \left( (n-1)^2 E[X_k^2] + E[X_1^2] + \dots + E[X_{k-1}^2] + E[X_{k+1}^2] + \dots + E[X_n^2] \right)$$

On utilise la relation  $E[X_k^2] = \sigma^2$  :

$$E[Y] = \frac{1}{n^3} \sum_k \left( (n-1)^2 \sigma^2 + (n-1) \sigma^2 \right) = \frac{n-1}{n^3} \sum_k n \sigma^2 = \frac{n-1}{n^2} \sum_k \sigma^2 = \frac{n-1}{n} \sigma^2.$$

La valeur de  $Y$  est  $s_n^2$  on utilise donc l'estimation sans biais  $\sigma^2 = \frac{n}{n-1} s_n^2$  qui conduit à la relation 5.8]]

La campagne de  $n$  mesures indépendantes, conduit au résultat

$$\boxed{g_{vrai} = \bar{g}_n \pm \frac{s_n}{\sqrt{n-1}} (\text{rms})} \quad \text{avec}$$

$$\bar{g}_n = \frac{1}{n} (g_1 + g_2 + \dots + g_n) \quad \text{et} \quad s_n^2 = \frac{1}{n} \sum_k (g_k - \bar{g}_n)^2$$

On constate que la forme du résultat est semblable à celle donnée en 5.7; l'estimation de  $G$  est inchangée mais celle de l'incertitude standard est modifiée.

#### 5.4 Remarques finales

Les résultats d'une campagne de mesures se présentent sous la forme d'un nuage de points à une dimension, assez semblable à une statistique de revenus. Il existe cependant une différence importante. Considérant le cas des mesures, la dispersion du nuage apparaît le plus souvent<sup>§</sup> comme une conséquence de l'imperfection des méthodes physiques employées. Cette dispersion peut être considérablement réduite en améliorant les appareils de mesure. Pour la statistique de revenus, la dispersion des résultats est un élément de description de la population concernée. Ce n'est pas en mesurant les revenus au dixième de centime près que l'on réduira cette dispersion. Il reste néanmoins que les deux nuages de points peuvent être décrits de façons semblables.

Assimilées à des variables continues, la densité de répartition des variables aléatoires les plus diverses présente l'allure d'une courbe en forme de cloche (voir par exemple la figure 1-11 page 14). On réserve cependant l'expression de "**courbe en cloche**" pour la densité de répartition d'une variable normale.

#### Annexe

1- Nous nous proposons de démontrer que la somme de deux variables normales indépendantes suit une loi normale.

Soit  $X$  et  $Y$  deux variables aléatoires dont la densité de probabilité est respectivement  $p(x) = \frac{1}{\sqrt{2\pi} \sigma_X} \exp \left[ -\frac{(x - \bar{X})^2}{2 \sigma_X^2} \right]$  et  $q(y) = \frac{1}{\sqrt{2\pi} \sigma_Y} \exp \left[ -\frac{(y - \bar{Y})^2}{2 \sigma_Y^2} \right]$  où  $\bar{X}$

<sup>§</sup>Ce n'est pas le cas en mécanique quantique où, pour une mesure idéale, la dispersion des résultats est une caractéristique de l'état physique observé et de la grandeur mesurée.

et  $\bar{Y}$  sont les espérances mathématiques de  $X$  et  $Y$ , tandis que  $\sigma_X$  et  $\sigma_Y$  en sont les écarts-types.

La probabilité de  $X \in [x, x + dx]$  est  $p(x)dx$ . La probabilité de  $Y \in [y, y + dy]$  est  $q(y)dy$ . Les variables étant indépendantes, la probabilité des deux événements simultanés est

$$dP = p(x)dx \times q(y)dy = p(x) \times q(y) \times dx dy \quad (\text{A-1})$$

Le résultat d'une mesure de  $X$  et  $Y$  est un couple de valeurs  $(x, y)$  que l'on représente comme les coordonnées d'un point,  $M$ , du plan  $Ox-Oy$ . La probabilité pour que ce point figure dans la surface élémentaire d'aire  $d\mathcal{A}$  est  $dP = f(M) d\mathcal{A}$ . La relation A-1, avec  $d\mathcal{A} = dx dy$  conduit à l'expression

$$dP = f(M) d\mathcal{A} \quad \text{avec} \quad f(M) = p(x) q(y) \quad (\text{A-2})$$

Changeons de coordonnées pour repérer le plan. Posons  $x = \frac{1}{2}(z - u)$  et  $y = \frac{1}{2}(z + u)$  c'est-à-dire  $z = y + x$  et  $u = y - x$ .

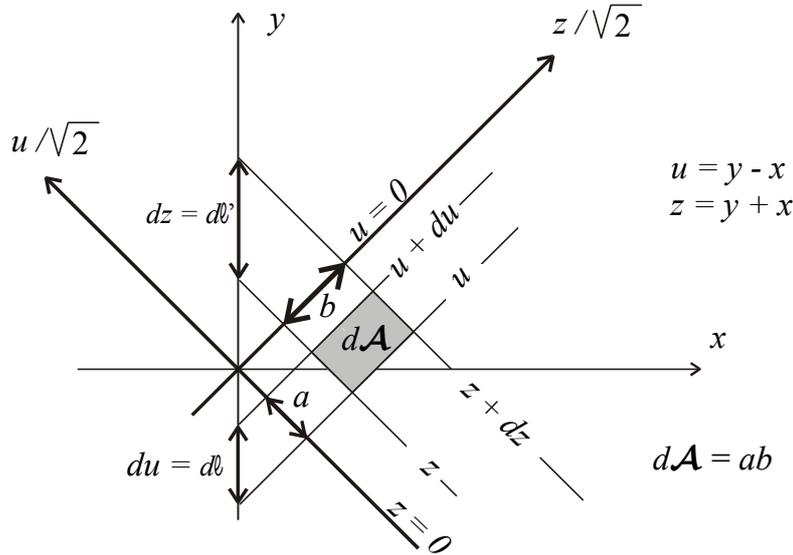


Figure 5-2 :  $d\mathcal{A} = a b = dl dl' / 2 = du dz / 2$

Avec les nouvelles variables, l'élément de surface  $d\mathcal{A}$  s'exprime sous la forme  $d\mathcal{A} = \frac{dz du}{2}$ . La probabilité de l'élément de surface  $d\mathcal{A}$  s'écrit

$$\begin{aligned} dP &= f(M) d\mathcal{A} = p\left(\frac{z-u}{2}\right) q\left(\frac{z+u}{2}\right) \frac{du dz}{2} \\ &= \exp\left[-\frac{\left(\frac{z-u}{2} - \bar{X}\right)^2}{2\sigma_X^2}\right] \exp\left[-\frac{\left(\frac{z+u}{2} - \bar{Y}\right)^2}{2\sigma_Y^2}\right] \frac{du dz}{4\pi\sigma_X\sigma_Y} \end{aligned}$$

Nous écrivons cette expression sous la forme  $dP = \exp[-a^2 z^2 + bz - c^2 u^2 + dzu + eu - f] \times du dz$  avec

$$\begin{aligned}
a^2 = c^2 &= \frac{1}{8\sigma_X^2} + \frac{1}{8\sigma_Y^2} \\
b &= \frac{\bar{X}}{2\sigma_X^2} + \frac{\bar{Y}}{2\sigma_Y^2} \\
d &= \frac{1}{4\sigma_X^2} - \frac{1}{4\sigma_Y^2} \\
e &= \frac{-\bar{X}}{2\sigma_X^2} + \frac{\bar{Y}}{2\sigma_Y^2} \\
f &= \frac{\bar{X}^2}{2\sigma_X^2} + \frac{\bar{Y}^2}{2\sigma_Y^2} + \ln(4\pi\sigma_X\sigma_Y)
\end{aligned}$$

La probabilité de l'intervalle  $(z, z + dz)$  est  $d\Pi = \pi(z) dz$  avec

$$\begin{aligned}
\pi(z) &= \int_{-\infty}^{+\infty} \exp[-a^2z^2 + bz - c^2u^2 + dzu + eu - f] \times du \\
&= \exp[-a^2z^2 + bz - f] \times \int_{-\infty}^{+\infty} \exp[-c^2u^2 + dzu + eu] \times du
\end{aligned}$$

On utilise la relation  $\int_{-\infty}^{+\infty} \exp[-\alpha^2u^2 + \beta u] \times du = e^{\frac{\beta^2}{4\alpha^2}} \frac{\sqrt{\pi}}{|\alpha|}$ . Il vient

$$\pi(z) = \exp[-a^2z^2 + bz - f] \times \exp\left[\left(\frac{dz + e}{2c}\right)^2 + \ln\left[\frac{\sqrt{\pi}}{|\alpha|}\right]\right]$$

On en déduit

$$\pi(z) = \exp[-A^2z^2 + Bz + C] = \Lambda \exp\left[-A^2\left(z - \frac{B}{2A^2}\right)^2\right]$$

où  $\Lambda$  est une constante tandis que  $\frac{1}{A^2} = 2(\sigma_X^2 + \sigma_Y^2)$ . La fonction  $\pi(z)$  est la densité de probabilité d'une loi normale de moyenne  $\frac{B}{2A^2} = \bar{X} + \bar{Y}$  et de variance  $\sigma_Z^2 = \frac{1}{2A^2} = \sigma_X^2 + \sigma_Y^2$ . C'est ce que nous voulions démontrer. Chemin faisant on vérifie que l'espérance d'une somme de variables est la somme des espérances de chacune d'elles et que les variances s'additionnent lorsque les variables sont indépendantes.

**2-** Considérons la variable normale  $X$  dont la densité de probabilité est

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma_X} \exp\left[-\frac{(x - \bar{X})^2}{2\sigma_X^2}\right]$$

Introduisons la variable  $Z = \lambda X + m$  où  $\lambda$  et  $m$  sont des constantes.

Soit  $dP$  la probabilité de l'événement  $Z \in (z, z + dz)$ . C'est la probabilité de l'événement  $X \in (x, x + dx)$ , avec  $x = \frac{1}{\lambda}(z - m)$  et  $dx = \frac{1}{\lambda}dz$ . On en déduit  $dP = \frac{1}{\sqrt{2\pi}\sigma_X} \exp\left[-\frac{(x - \bar{X})^2}{2\sigma_X^2}\right] \times \frac{1}{\lambda}dz = \frac{1}{\sqrt{2\pi}\lambda\sigma_X} \exp\left[-\frac{(z - m - \lambda\bar{X})^2}{2(\lambda\sigma_X)^2}\right] \times dz$

La variable  $Z$  suit donc une loi normale de moyenne  $\bar{Z} = m + \lambda\bar{X}$  et d'écart-type  $\sigma_Z = \lambda\sigma_X$ .

