

# *Traitement des incertitudes de mesure*

## **I. PROBLEMATIQUE.**

Quand on veut déterminer une grandeur et que l'on effectue plusieurs mesures (simultanées avec des appareillages équivalents, ou répétées dans des conditions semblables avec un seul appareillage), on constate que celles-ci ont donné des valeurs différentes. Cela signifie que la variété des circonstances nous a fourni des valeurs déformées de la grandeur elle-même. Ces constatations permettent d'énoncer ainsi le problème de la mesure :

*il s'agit de saisir une estimation optimale d'une grandeur à laquelle on suppose une existence objective.*

### **Précision des mesures.**

Par définition, une mesure est la comparaison d'une grandeur physique à une grandeur de même espèce prise pour étalon. Pour avoir un sens, le nombre qui exprime une mesure doit être complété par :

- la précision (ou le degré d'incertitude) avec lequel on l'a obtenu
- l'unité choisie

Ex.: vitesse de la lumière dans le vide:  $c = (2.997\,91 \pm 0.000\,01) 10^8 \text{ m s}^{-1}$

## **II. LES INCERTITUDES DE MESURE.**

Chaque mesure physique comporte donc une "incertitude", même si la mesure est faite soigneusement et que tous les appareils fonctionnent et sont bien réglés. Il existe deux types d'incertitudes :

1. *les incertitudes systématiques* (à caractère constant et repérables)
2. *les incertitudes aléatoires ou statistiques* (non prévisibles)

Les sources d'incertitudes sont à chercher :

- a) dans les étalons d'usage qui peuvent s'user, voire se déformer à la longue.

b) dans l'appareillage utilisé, qui peut être très varié selon la précision recherchée. Il possède trois caractéristiques :

- la *sensibilité*, qui peut être définie par la plus petite fraction de la grandeur qu'il permet de déceler,
- la *précision*, définie par la limite supérieure de l'écart entre la valeur obtenue au moyen de l'appareil et la valeur "vraie". Cet écart peut revêtir deux natures: systématique (indications fausses, mais reproductibles) ou accidentelle (variant d'une mesure à l'autre). Généralement les deux écarts coexistent.
- la *fiabilité*, qui peut être définie par l'aptitude de l'appareillage à limiter l'évolution dans le temps de ses caractéristiques internes.

c) dans la méthode utilisée qui a obligatoirement une influence déterminante sur la précision des mesures. Elle peut introduire notamment des incertitudes systématiques que l'on peut compenser ou neutraliser plus ou moins bien.

d) aussi chez l'opérateur dont le rôle joué est complexe. L'imperfection des sens de celui-ci, voire son préjugé, sont aussi une source d'incertitudes qui constitue ce que l'on appelle "l'équation personnelle" (par exemple le retard systématique des chronomètres). Le choix de la méthode ainsi que les progrès de l'automatisation permettent de réduire ces incertitudes.

### *Les incertitudes systématiques*

Les incertitudes systématiques, qui sont à caractère constant et donc repérables, doivent être corrigées. Un exemple simple est la mise à zéro d'un voltmètre. Si le voltmètre n'est pas mis exactement à zéro avant son usage, toutes les valeurs indiquées seront décalées dans un sens. Cependant, même si on met soigneusement le voltmètre à zéro, il subsiste toujours une incertitude.

### *Les incertitudes aléatoires ou statistiques*

Les incertitudes aléatoires peuvent provenir des imprécisions liées à l'instrumentation et/ou à la nature statistique des phénomènes observés.

Dans le premier cas, les "imperfections" consistent en des facteurs non contrôlés au niveau des instruments de mesure ou des conditions de la mesure. Par exemple, il peut y avoir du jeu dans les joints mécaniques d'un micromètre, des variations dans l'épaisseur des traits d'une règle ou l'échelle d'un voltmètre, etc. Ces facteurs vont varier aléatoirement d'une mesure à l'autre et perturberont ainsi le résultat. En conséquence, si on fait une série de mesures répétées, on trouvera une distribution de valeurs. Si on procède à un nombre infini de mesures, la moyenne de cette distribution correspondra à la vraie valeur de la quantité mesurée alors que la dispersion de la distribution sera liée aux imperfections de l'instrument (la précision). Il est évident que plus on contrôlera les différents facteurs de la mesure, plus la mesure deviendra précise et plus la distribution deviendra étroite.

Dans le deuxième cas, c'est le phénomène physique lui-même qui varie. Un exemple est la radioactivité: le nombre de désintégrations subies par une source radioactive dans une période  $\Delta t$  varie aléatoirement selon les lois probabilistes de la mécanique quantique. C'est pourquoi si l'on fait une série de mesures en comptant le nombre de coups d'une source radioactive quelconque pendant une période de temps  $\Delta t$  avec un détecteur idéal, on trouvera toujours une distribution de valeurs.

Contrairement aux incertitudes systématiques, les incertitudes aléatoires peuvent être traitées par la théorie de la statistique.

### III. ENSEMBLES DE MESURES.

Soit un ensemble de mesures  $x_i$  débarrassées des incertitudes systématiques, effectuées un grand nombre de fois. Sur un diagramme, reportons  $x$  en abscisse, puis en ordonnée le nombre  $N$  de fois que l'on a obtenu des valeurs comprises entre  $x_1$  et  $x_2$ ,  $x_2$  et  $x_3$ , etc. On obtient le diagramme des fréquences absolues (ou relatives, c'est-à-dire rapportées au nombre total de mesures), Fig. 1.

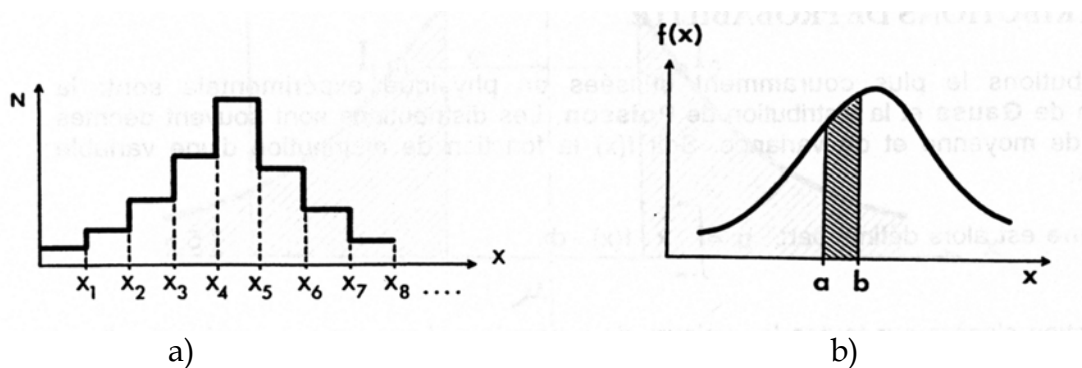


Fig. 1: Diagrammes de fréquence: a) pour un nombre fini de mesures, b) pour un nombre infini de mesures.

Si le nombre des mesures est très grand, on peut envisager une variation pratiquement continue de  $N$ , ce qui conduit à l'introduction d'une fonction  $f(x)$  appelée densité de fréquence ou fonction de distribution qui représente la distribution des mesures (Fig. 1 b)

Ainsi le nombre de mesures dont les valeurs seront comprises entre  $a$  et  $b$  sera donné par:

$$N = \int_a^b f(x) dx \quad (1)$$

Le diagramme des fréquences est caractérisé par :

- la moyenne 
$$m = \frac{1}{M} \cdot \sum_{i=1}^n x_i \cdot f_a(x_i) \quad (2)$$

avec  $f_a$  = fréquence absolue et  $M$  = nombre total de mesures

- la variance 
$$s^2 = \frac{1}{M} \cdot \sum_{i=1}^n (x_i - m)^2 \cdot f_a(x_i) \quad (3)$$

qui, pour un nombre infini de mesures, devient

$$\sigma^2 = \lim_{M \rightarrow \infty} \left[ \frac{1}{M} \cdot \sum_{i=1}^n (x_i - \mu)^2 \cdot f_a(x_i) \right] \quad \text{avec} \quad \mu = \lim_{M \rightarrow \infty} [m] \quad (4)$$

$s^2$  et  $\sigma^2$  sont des paramètres aptes à estimer la largeur de la zone de fréquence maximum.

On désigne par écart-type ou écart-standard la quantité  $\sigma$ .

Pour être complètement fixé au sujet de la valeur des paramètres  $\mu$  et  $\sigma$ , il faudrait pouvoir disposer d'une suite infinie de mesures. Comme cela n'est pas possible, on se contente d'une suite finie de mesures, qui fournissent des estimations  $m$  et  $s$  des paramètres  $\mu$  et  $\sigma$ .

En pratique, on compare les distributions de fréquences observées à des distributions théoriques définies à priori, dans le but de décrire au mieux ce que l'on observe.

#### IV. DISTRIBUTIONS DE PROBABILITE.

Les distributions le plus couramment utilisées en physique expérimentale sont la distribution de **Gauss** et la distribution de **Poisson**. Les distributions sont souvent décrites en termes de moyenne et de variance. Soit  $f(x)$  la fonction de distribution d'une variable aléatoire  $x$ .

La moyenne est alors définie par: 
$$\mu = \int_{-\infty}^{\infty} x f(x) dx \quad (5)$$

où l'intégration s'opère sur toutes les valeurs de  $x$  permises. La moyenne représente d'une certaine façon "le centre de gravité" de la distribution. La condition

$$\int_{-\infty}^{\infty} f(x) dx = 1 \quad \text{exprime la certitude d'avoir une mesure comprise entre } -\infty \text{ et } +\infty.$$

La variance est définie par: 
$$\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx \quad (6)$$

La racine de la variance  $\sigma$  est appelée "écart-type". Ce paramètre donne donc une mesure de la dispersion ou de la largeur de la distribution.

### a) Distribution de Gauss

La distribution Gaussienne (aussi appelée distribution "normale") est la distribution qui décrit pratiquement toutes les incertitudes instrumentales aléatoires (Fig. 2).

Si  $\mu$  est la moyenne et  $\sigma$  l'écart-type, alors la fonction de fréquence vaut:

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right) \quad (7)$$

Et, le changement de variable,  $X = \frac{x - \mu}{\sigma}$  et  $F(X) = \sigma f(x)$  conduit à la courbe normale dite réduite, centrée autour de  $X = 0$ :

$$F(X) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{X^2}{2}\right) \quad (8)$$

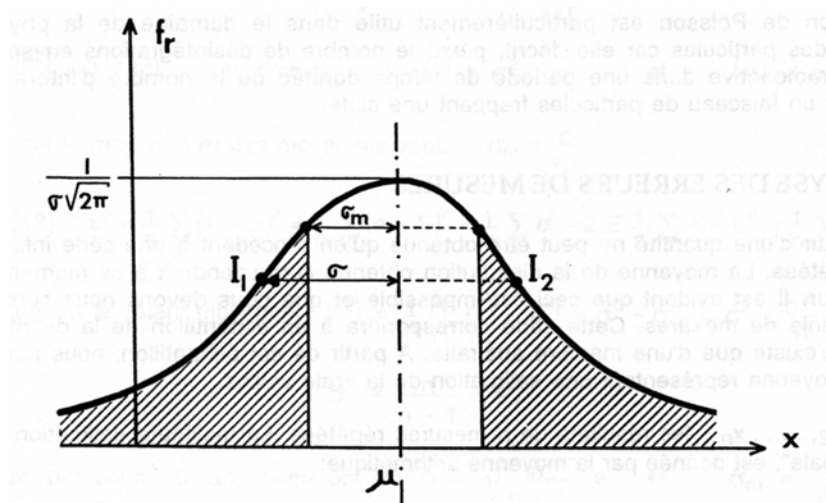


Fig. 2: Distribution de Gauss

Comme on le voit sur la figure,  $\sigma$  est une mesure de la largeur ou dispersion de la distribution. Pour la Gaussienne, l'écart-type correspond à la demi-largeur de la courbe à une hauteur d'à peu près 2/3 du maximum. Il est important de remarquer que la Gaussienne est une distribution très "concentrée". Si on calcule l'aire de la distribution entre  $\mu - \sigma$  et  $\mu + \sigma$ , on trouvera 68,3%. Entre  $\mu - 2\sigma$  et  $\mu + 2\sigma$ , cette valeur s'élève à 95,5% alors qu'entre  $\mu - 3\sigma$  et  $\mu + 3\sigma$ , la surface est de 99,7%.

On définit encore  $\sigma_m$  tel que :

$$\int_{-\infty}^{\mu - \sigma_m} f(x) dx + \int_{\mu + \sigma_m}^{\infty} f(x) dx = 0.5 \quad (9)$$

c'est-à-dire tel que 50% des valeurs sont situées à l'extérieur de l'intervalle  $(\mu - \sigma_m, \mu + \sigma_m)$ .  $\sigma_m$  définit l'écart médian, et il vaut :  $\sigma_m = 0.6745 \sigma$ .

## b) Distribution de Poisson

Contrairement à la distribution de Gauss, la distribution de Poisson est une distribution discrète, c.à.d. une distribution dont les valeurs  $x$  peuvent seulement prendre des valeurs entières. Elle est définie par la fonction de distribution suivante:

$$f(x) = e^{-\mu} \frac{\mu^x}{x!} \quad (10)$$

où  $\mu$  est la moyenne de la distribution. La distribution de Poisson a la propriété que sa variance  $\sigma^2$  est égale à la moyenne  $\mu$ . L'écart-type de la distribution est donc

$$\sigma = \sqrt{\mu} \quad (11)$$

Dans le cas d'un évènement aléatoire  $E$  qui peut se produire un grand nombre de fois durant une longue période d'observation  $t$ , mais qui apparaît rarement pour  $t$  petit,  $f(x)$  est la probabilité pour que  $E$  ait lieu 0, 1, 2, ...,  $x$  fois.

La distribution Poisson est particulièrement utile dans le domaine de la physique nucléaire et des particules car elle décrit, par exemple, le nombre de désintégrations émises par une source radioactive dans une période de temps donnée ou encore le nombre d'interactions produites par un faisceau de particules frappant une cible.

## V. ANALYSE DES INCERTITUDES DE MESURES

La vraie valeur d'une quantité ne peut être obtenue qu'en procédant à une série infinie de mesures répétées. La moyenne de la distribution obtenue correspondrait à ce moment-là à la vraie valeur. Il est évident que ceci est impossible et que nous devons nous contenter d'une série finie de mesures. Cette série correspondra à un échantillon de la distribution "infinie" qui n'existe que d'une manière abstraite. A partir de cet échantillon, nous pouvons calculer la moyenne représentant une estimation de la vraie valeur.

Soient  $x_1, x_2, \dots, x_n$  les résultats de  $n$  mesures répétées. La meilleure estimation de la moyenne "idéale", est donnée par la moyenne arithmétique:

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i \quad (12)$$

De même, on peut estimer la variance ou l'écart-type par:

$$s^2 = \frac{1}{n} \cdot \sum_{i=1}^n (x_i - \bar{x})^2 \quad (13)$$

Remarquons qu'il est important de différencier la moyenne  $\mu$  et la variance  $\sigma$  de la distribution "idéale", c.à.d. celle obtenue dans la limite d'un nombre infini de mesures, de ses "estimateurs", c.à.d. la moyenne  $\bar{x}$  et la variance  $s$  calculées à partir de l'échantillon de  $n$  mesures.

Quelle est maintenant l'incertitude sur l'estimation de la vraie valeur  $\mu$ ? Pour calculer cette quantité, il est important de comprendre que  $\bar{x}$  est aussi une variable aléatoire. Si l'on faisait une deuxième série de  $n$  mesures, on trouverait une moyenne  $\bar{x}$  qui serait probablement différente de la première. Une troisième série donnerait encore un autre résultat, etc. On peut donc prendre un nombre infini d'échantillons de taille  $n$ , calculer la moyenne  $\bar{x}$  de chacun d'eux et regarder la distribution de  $\bar{x}$  ainsi formée.

Soit  $v$  échantillons de  $n$  mesures,  $\bar{x}_i$  la valeur moyenne de l'échantillon  $i$ . La moyenne  $m$  des moyennes vaut simplement:

$$m = \frac{1}{v} \cdot \sum_{i=1}^v \bar{x}_i \quad (14)$$

avec  $m \rightarrow \mu$  pour  $v \rightarrow \infty$

Soit  $e_i = x_i - \mu$ , l'incertitude de la  $i^{\text{ème}}$  mesure. L'incertitude  $E$  sur la moyenne  $\bar{x}$  vaut:

$$E = \bar{x} - \mu = \frac{1}{n} \cdot \sum_{i=1}^n x_i - \mu = \frac{1}{n} \cdot \sum_{i=1}^n (x_i - \mu) = \frac{1}{n} \cdot \sum_{i=1}^n e_i \quad (15)$$

et

$$E^2 = \frac{1}{n^2} \cdot \sum_i e_i^2 + \frac{1}{n^2} \cdot \sum_{i \neq j} e_i \cdot e_j \quad (16)$$

Calculons la moyenne de  $E^2$ . On a :

$$\left\langle \sum_i e_i^2 \right\rangle = n \langle e^2 \rangle \quad \text{et} \quad \left\langle \sum_{i \neq j} e_i \cdot e_j \right\rangle = 0 \quad (\text{car } e_i \text{ et } e_j \text{ indépendants})$$

donc  $\langle E^2 \rangle = \frac{1}{n} \langle e^2 \rangle$  et comme  $\langle E^2 \rangle = \sigma_m^2$  et  $\langle e^2 \rangle = \sigma^2$

L'incertitude sur la moyenne  $m$  des moyennes vaut:  $\sigma_m = \frac{\sigma}{\sqrt{n}}$  (17)

mais, de l'éq. (13):

$$s^2 = \frac{1}{n} \cdot \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \cdot \sum_{i=1}^n (e_i - E)^2 = \frac{1}{n} \cdot \sum_{i=1}^n e_i^2 - 2 \cdot E \cdot \frac{1}{n} \cdot \sum_{i=1}^n e_i + E^2 = \frac{1}{n} \cdot \sum_{i=1}^n e_i^2 - E^2$$

et, en prenant la valeur moyenne:  $\langle s^2 \rangle = \left\langle \frac{1}{n} \cdot \sum_{i=1}^n e_i^2 \right\rangle - \langle E^2 \rangle = \sigma^2 - \sigma_m^2 = \sigma^2 - \frac{\sigma^2}{n}$

donc 
$$\sigma^2 = \frac{n}{n-1} \langle s^2 \rangle \quad (18)$$

$\langle s^2 \rangle$  n'est pas connu, on se contente de  $s^2$  comme approximation:

$$\sigma = \sqrt{\frac{n}{n-1}} s \quad \text{et} \quad \sigma_m = \sqrt{\frac{1}{n-1}} s \quad (19)$$

Nous remarquons que dans la limite  $n \rightarrow \infty$ , l'incertitude sur la moyenne tend vers zéro. Notons aussi que dans le cas d'une seule mesure,  $n = 1$ , l'incertitude devient  $\sigma_m = \sigma$ . Ce résultat nous permet d'interpréter  $\sigma$ .

$\sigma$  donne accès à la précision de l'instrument qui est intrinsèque à l'appareil, et qui est déterminée par la largeur de la distribution idéale. La précision de la mesure est donnée par l'incertitude sur la moyenne  $\sigma_m$  qui dépend de la précision de l'instrument et du nombre de mesures.

C'est seulement dans le cas où le nombre de mesures vaut 1 que la précision de la mesure égale la précision de l'instrument.

L'incertitude  $(\mu - \bar{x}) = \sigma_m$  diminue lorsque le nombre de mesures augmente.

L'examen de la courbe  $\mu - \bar{x} = f(n)$  (Fig. 3) montre cependant qu'il n'est guère nécessaire de porter le nombre des mesures au-delà d'une valeur  $n_c = 10$ , vu le ralentissement de la décroissance de la courbe pour  $n \geq 10$ .

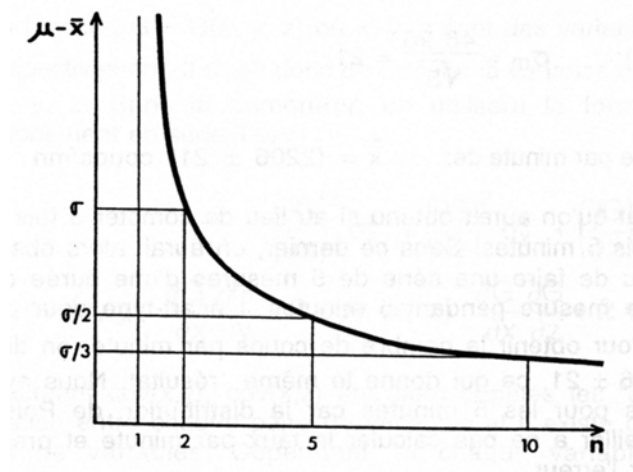


Fig. 3: Variation de l'incertitude sur une moyenne de  $n$  mesures.

## VI. EXEMPLE

Un groupe d'étudiants mesure l'épaisseur d'une feuille métallique à l'aide d'un micromètre dont la précision est inconnue. Ils font 10 mesures et obtiennent les résultats suivants

250 252 249 246 250 253 252 251 249 253  $\mu\text{m}$

Donnez une valeur pour l'épaisseur et son incertitude. Quelle est la précision du micromètre ?



L'épaisseur la plus probable est donnée par la moyenne:  $\bar{x} = \frac{2505}{10} = 250.5 \mu\text{m}$

L'incertitude est obtenue par l'équation (II.19):  $\sigma_m = 0.69 \mu\text{m}$

L'épaisseur de la feuille est alors:  $x = 250.5 \pm 0.7 \mu\text{m}$

Comme nous l'avons déjà vu, la précision du micromètre correspond à l'écart-type, donc précision du micromètre de l'ordre de  $\sigma = 2.2 \mu\text{m}$

## II. 7 REGLE ELEMENTAIRE DU CALCUL DES INCERTITUDES COMPOSEES

On va considérer que les incertitudes individuelles sont en général petites par rapport à la grandeur mesurée.

Convention: l'incertitude sur une mesure individuelle  $x$  se note  
soit en valeur absolue  $\Delta x$  (même unité que  $x$ )  
soit en valeur relative  $\frac{\Delta x}{x}$  (en %).

Pour évaluer l'incertitude  $\Delta G$  sur une grandeur  $G$  fonction de plusieurs variables indépendantes  $x, y, z, \dots$ , lorsqu'on connaît les incertitudes individuelles  $\Delta x, \Delta y, \Delta z, \dots$ , on développe  $G$  en série de TAYLOR, et on ne retient que les termes du premier ordre.

$$\Delta G = \frac{\partial G}{\partial x} \cdot \Delta x + \frac{\partial G}{\partial y} \cdot \Delta y + \frac{\partial G}{\partial z} \cdot \Delta z \quad (20)$$

Comme il faut connaître l'incertitude maximum qui entache un résultat, on prend les valeurs absolues:

$$|\Delta G| = \left| \frac{\partial G}{\partial x} \cdot \Delta x \right| + \left| \frac{\partial G}{\partial y} \cdot \Delta y \right| + \left| \frac{\partial G}{\partial z} \cdot \Delta z \right| \quad (21)$$

Cette formule permet en particulier d'établir les règles élémentaires qui régissent le calcul d'incertitude dans des cas simples. Par exemple:

$$\begin{array}{llll} \Delta G = \Delta x + \Delta y & \text{lorsque} & G = x + y & \text{ou} & G = x - y \\ \frac{\Delta G}{G} = \frac{\Delta x}{x} + \frac{\Delta y}{y} & \text{lorsque} & G = x \cdot y & \text{ou} & G = x/y \end{array}$$

## VIII . LOI DE PROPAGATION DES INCERTITUDES

Souvent, il est nécessaire d'utiliser le résultat d'une mesure pour en calculer un autre. Il est évident que l'incertitude sur la mesure va produire aussi une incertitude

au niveau de ce nouveau résultat. On dit que l'incertitude a été *propagée*. Quelle est la valeur de cette incertitude ?

Soit la fonction  $u = f(x, y, z)$  où  $x, y, z$  sont des grandeurs mesurées comportant des incertitudes  $\sigma_x, \sigma_y, \sigma_z$ , respectivement.

Il s'agit donc de calculer la variance de  $u$ ,  $\sigma_u$ , en fonction des variances de  $x, y, z$ .

Pour la mesure n°  $i$ , on a :  $u_i = f(x_i, y_i, z_i)$  et on admettra que la valeur moyenne est donnée par :  $\bar{u} = f(\bar{x}, \bar{y}, \bar{z})$

La variance vaut :

$$\sigma_u^2 = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_i (u_i - \bar{u})^2$$

En développant au premier ordre autour de  $u = \bar{u}, \dots x = \bar{x}, \dots$  on obtient :

$$u_i - \bar{u} \cong \frac{\partial u}{\partial x} (x_i - \bar{x}) + \frac{\partial u}{\partial y} (y_i - \bar{y}) + \frac{\partial u}{\partial z} (z_i - \bar{z}) \quad \text{et}$$

$$\sigma_u^2 = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_i \left[ (x_i - \bar{x})^2 \cdot \left( \frac{\partial u}{\partial x} \right)^2 + (y_i - \bar{y})^2 \cdot \left( \frac{\partial u}{\partial y} \right)^2 + \dots + 2 \cdot (x_i - \bar{x})(y_i - \bar{y}) \cdot \left( \frac{\partial u}{\partial x} \right) \left( \frac{\partial u}{\partial y} \right) + \dots \right]$$

mais  $\lim_{N \rightarrow \infty} \frac{1}{N} \sum_i \left[ (x_i - \bar{x})^2 \right] = \sigma_x^2$  d'où la formule :

$$\begin{aligned} \sigma_u^2 = & \sigma_x^2 \cdot \left( \frac{\partial u}{\partial x} \right)^2 + \sigma_y^2 \cdot \left( \frac{\partial u}{\partial y} \right)^2 + \sigma_z^2 \cdot \left( \frac{\partial u}{\partial z} \right)^2 + \\ & + 2 \cdot \sigma_{xy}^2 \cdot \left( \frac{\partial u}{\partial x} \right) \left( \frac{\partial u}{\partial y} \right) + 2 \cdot \sigma_{xz}^2 \cdot \left( \frac{\partial u}{\partial x} \right) \left( \frac{\partial u}{\partial z} \right) + 2 \cdot \sigma_{yz}^2 \cdot \left( \frac{\partial u}{\partial y} \right) \left( \frac{\partial u}{\partial z} \right) \end{aligned} \quad (22)$$

Les facteurs  $\sigma_{xy}^2, \sigma_{xz}^2$ , etc. sont appelés les "covariances". Elles sont liées à la corrélation entre chaque paire de variables et il existe donc autant de covariances que de paires de variables. Cependant, si chaque variable est le résultat d'une mesure indépendante, ce qui est normalement le cas dans les mesures en physique, les covariances sont nulles et la formule se réduit à une somme de carrés, donc :

$$\boxed{\sigma_u^2 = \sigma_x^2 \cdot \left( \frac{\partial u}{\partial x} \right)^2 + \sigma_y^2 \cdot \left( \frac{\partial u}{\partial y} \right)^2 + \sigma_z^2 \cdot \left( \frac{\partial u}{\partial z} \right)^2} \quad (22^*)$$

Exemple: La période  $T$  d'un pendule est donnée par :  $T = 2 \pi \sqrt{\frac{L}{g}}$

où  $L$  est la longueur du pendule et  $g$  la constante de gravité.

Calculer  $g$  et l'incertitude sur  $g$  à partir des résultats de mesures suivantes :

$L = 99.8 \pm 0.3 \text{ cm}$  et  $T = 2.03 \pm 0.05 \text{ s}$ , en supposant  $L$  et  $T$  non corrélés.

Evidemment :

$$g = 4\pi^2 L / T^2 = 9.56 \text{ m/s}^2$$

L'incertitude est obtenue par l'éq. (22\*):  $\sigma_g^2 = \sigma_L^2 \cdot \left(\frac{\partial g}{\partial L}\right)^2 + \sigma_T^2 \cdot \left(\frac{\partial g}{\partial T}\right)^2$

$$\sigma_g^2 = \sigma_L^2 \cdot \left(\frac{4\pi^2}{T^2}\right)^2 + \sigma_T^2 \cdot \left(\frac{-8\pi^2 L}{T^3}\right)^2 = 0.223 \frac{\text{m}^2}{\text{s}^4}$$

donc  $\sigma_g = 0.47 \text{ m/s}^2$ , et donc  $g = 9.56 \pm 0.47 \text{ m/s}^2$

## IX. CORRELATION DE DEUX VARIABLES ALEATOIRES DEPENDANTES ET COURBES DE REGRESSION.

Jusqu'ici on a discuté les problèmes relatifs aux variables aléatoires indépendantes. Considérons maintenant les problèmes de deux variables aléatoires non indépendantes.

La forme limite de la dépendance de deux variables aléatoires rejoint la relation mathématique  $y = f(x)$  et l'on dit que la variable aléatoire dépendante  $y$  est une fonction de la variable aléatoire indépendante  $x$ . En général, on ne connaît pas cette dépendance et pour la découvrir, il faut réunir certaines données qui consistent en observables couplées  $x_i$  et  $y_i$ . Ces observables représentent un "nuage" de points dans le plan  $xy$  (plan de mesure). On se pose le problème fondamental suivant: déterminer la relation mathématique  $y = f(x)$  au moyen du nuage de points  $x_i$  et  $y_i$ . L'image de cette relation dans le plan des mesures définit la **courbe de régression** ou **courbe d'ajustement**  $y = f(x)$  représentée sur la figure 4.

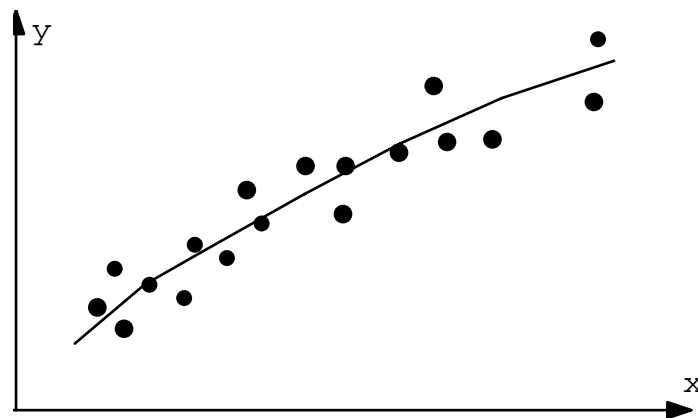


Fig. 4 : Nuage de points  $x_i$  et  $y_i$  et courbe de régression

Une fois connue, cette courbe permet de prédire la valeur probable de la variable dépendante  $y$  en fonction de la variable indépendante  $x$ .

En général, on recherche une relation linéaire entre les variables  $x$  et  $y$  (échelles log, semi-log,...). Dans le plan  $xy$ , la relation linéaire est représentée par une droite. Deux méthodes s'offrent principalement à nous pour déterminer cette dernière.

### 1) Méthode graphique

Ayant à disposition plusieurs points de la dépendance entre  $x$  et  $y$ , on trace au jugé, dans le plan des mesures  $xy$ , une droite apte à satisfaire le mieux possible une relation linéaire entre les deux variables. La droite ainsi construite est appelée "*droite de régression estimée*". La méthode est simple, mais de nature subjective.

### 2) Méthode des moindres carrés

Ayant rassemblé des points  $(x_i, y_i)$  de significations équivalentes quant aux conditions dans lesquelles les mesures ont été effectuées et quant à la précision de ces dernières, on cherche une droite d'équation

$$y = ax + b \quad (23)$$

rendant minimum la somme  $S$  des segments  $MM_i$  indiqués sur la figure 5.

$$S = \sum_{i=1}^n (y_i - ax_i - b)^2 \quad (24)$$

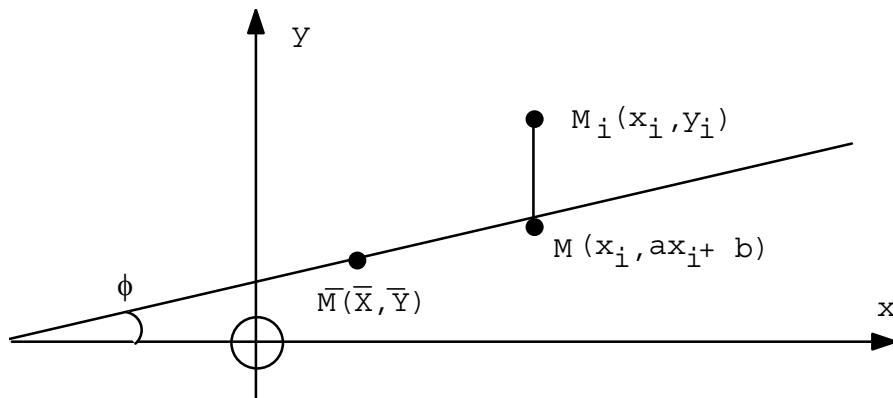


Fig. 5: Droite de régression

Les points stationnaires de  $S$  sont fournis par les deux relations

$$\frac{\partial S}{\partial a} = 0 \quad \text{et} \quad \frac{\partial S}{\partial b} = 0 \quad (25)$$

De la seconde, on tire 
$$\sum_{i=1}^n (y_i - ax_i - b) = 0$$

soit en divisant membre à membre par  $n$  : 
$$\frac{1}{n} \sum_{i=1}^n y_i - a \frac{1}{n} \sum_{i=1}^n x_i - b = 0 \quad (26)$$

$$\frac{\bar{y}}{\quad} \quad \frac{a\bar{x}}{\quad}$$

qui signifie que la droite de régression passe par le centre de gravité  $M(x,y)$  du nuage. De l'équation (26), on obtient:

$$b = \bar{y} - a \cdot \bar{x}$$

Dans ces conditions, la somme  $S$  s'écrit (équ. (24)):

$$S = \sum_{i=1}^n (y_i - \bar{y} - a(x_i - \bar{x}))^2$$

Finalement de  $\frac{\partial S}{\partial a} = 0$ , on tire :

$$a = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (27)$$

Ce qui achève la détermination de la droite d'équation (23).

Incertitudes : Des calculs plus complets permettent de montrer que les incertitudes sur la pente de la droite «  $a$  » et sur l'ordonnée à l'origine «  $b$  » valent :

$$(\Delta a)^2 \cong \frac{\sum_{i=1}^n (y_i - ax_i - b)^2}{(n-2) \cdot \sum_{i=1}^n (x_i - \bar{x})^2} \quad \text{et} \quad (\Delta b)^2 \cong \left( \frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right) \cdot \frac{\sum_{i=1}^n (y_i - ax_i - b)^2}{(n-2)}$$

## X. BIBLIOGRAPHIE

1. P.R. Bevington, Data Reduction and Error Analysis for the Physical Sciences (McGraw-Hill Book Co., New-York (1969)
2. W.R. Leo, Techniques for Nuclear and Particle Physics Experiments (Springer-Verlag, Heidelberg 1987) chap.4
3. S.L. Meyers, Data Analysis for Scientists (John Wiley & Sons, New-York 1976)
4. G. L. Squires, Practical physics. (Cambridge University Press, 3<sup>rd</sup> edition, 1985)