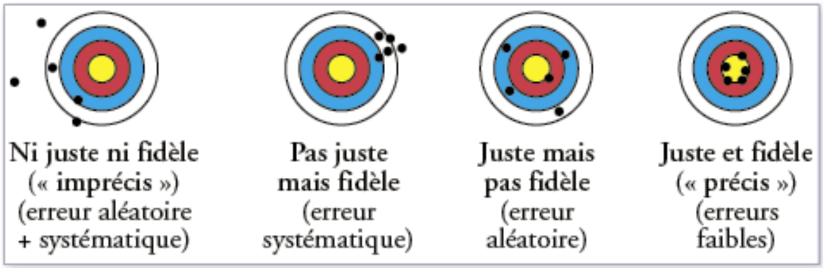
# Analyse de données et exploitation de mesures

## I – Présentation du problème

Pour les sciences expérimentales comme la physique et la chimie, il n'y a pas de mesure exacte. On connait toujours la valeur de la grandeur mesurée avec une certaine précision qui ne peut pas être arbitrairement grande. Il existe en effet plusieurs type d'erreurs :

- les erreurs systématiques, dues aux appareils de mesure (calibrage, tare, dérive en température, ...) ou à des approximations dans le mode opératoire (facteurs extérieurs non pris en compte, mauvaise utilisation d'un appareil, ...)
- les erreurs aléatoires, dues à l'intervention d'un opérateur (erreur de lecture, ...) ou aux appareils (précision, vieillissement, ...)




La précision d'une mesure dépend dépend donc de plusieurs facteurs qu'il faut identifier et essayer d'estimer. La figure ci-dessus présente différents cas de mesures en indiquant la valeur "vraie" au centre de la cible. En pratique, cette valeur est souvent inconnue. Lorsqu'on répète les mesures, il est possible d'identifier et de minimiser les erreurs aléatoires. Pour les erreurs systématiques, il est impossible de les détecter par cette approche. C'est tout le métier de l'expérimentateur de savoir les identifier et les estimer.

**Un résultat sans incertitude associée ne veux rien dire !**

Dans toute la suite, on note le résultat d'une mesure sous la forme :

$$x = x_m \pm \delta x$$

$x_m$ est la meilleur estimation possible de la valeur "vraie" $x_0$ non connue et $\delta x$ est l'incertitude absolue sur la mesure. Par convention, cette incertitude est toujours positive. On peut aussi parler d'incertitude relative $\dfrac{\delta x}{x_0}$. On note alors :

$$x = x_m \left(1 \pm \frac{\delta x}{x_m}\right)$$

Ainsi, le résultat d'une mesure n'est jamais une valeur mais c'est un intervalle des valeurs possibles. Au cours d'une manipulation, deux cas peuvent se présenter :

**Incertitude de type A** si on effectue une série de mesures d'une même grandeur. Les résultats sont répartis au hasard en considérant toutes les mesures indépendantes. Dans ce cas, un traitement statistique des mesures est possible.

**Incertitude de type B** si on ne peut effectuer qu'une seule mesure. Dans ce cas une étude statistique est exclue et seule une évaluation des erreurs et un calcul d'incertitude peut donner un encadrement à la valeur obtenue. Cet encadrement est intéressant pour évaluer la précision de la mesure mais il est à prendre avec beaucoup de précautions car une seule mesure peut toujours être fausse, n'ayant pas de point de comparaison.

Une fois la mesure faite et le calcul d'inertitude effectué, si la valeur mesurée peut-être comparée à une valeur tabulée, théorique, donnée par un constructeur, ... alors il faut analyser la compatibilité des valeurs. La grandeur à déterminer pour cela s'appelle l'écart normalisé (ou z–score). Considérons une grandeur $x_{\text{mes}}$ et l'incertitude–type associée $s_{\text{mes}}$ que l'on souhaite comparer à la grandeur $x_{\text{ref}}$ de "référence" associée elle aussi à une incertitude–type, notée $s_{\text{ref}}$. Le z–score est défini par :

$$Z_{\text{score}} = \frac{|x_{\text{ref}} - x_{\text{mes}}|}{\sqrt{s_{\text{ref}}^2 + s_{\text{mes}}^2}}$$

Si cette grandeur est inférieure à deux, les mesures sont compatibles avec un intervalle de confiance raisonnable.

# II – Calcul d'incertitude

## 1 - Approche statistique (type A)

**★ Cas général**

On se place dans le cas où on peut faire plusieurs fois la même mesure dans les mêmes conditions ($n$ fois). En supposant que les diverses observations sont représentées par des variables aléatoires indépendantes (ce qui sera très souvent notre cas), on peut montrer que les mesures se répartissent suivant une loi normale de probabilité (la densité de probabilité de mesure est une Gaussienne). Dans la suite, on suppose que c'est effectivement le cas. On définit alors la valeur moyenne $\overline{x}$ et l'écart type $\sigma(x)$ associés à la série de mesure effectuée :

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

et

$$\sigma(x) = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(\overline{x} - x_i)^2}$$

Dans le cas d'une recherche d'incertitude statistique sur une série de mesures, il est plus précis d'évaluer $\sigma(\overline{x})$ qui correspond à l'écart-type sur la valeur moyenne de l'échantillon. On montre que cet écart type (aussi appelé erreur-type de la moyenne ou incertitude-type sur la valeur de $x$) s'écrit :

$$\sigma(\overline{x}) = s = \frac{\sigma(x)}{\sqrt{n}}$$

**Remarques** :

- La meilleure estimation possible de la valeur cherchée correspond à la moyenne arithmétique des valeurs observées : $x_m = \overline{x}$.
- L'apparition du facteur $n-1$ dans l'expression de $\sigma(x)$ est cohérente avec le fait qu'une seule mesure est insuffisante pour espérer pouvoir l'estimer.
- Plus N est grand, plus $\sigma(\overline{x})$ diminue. Il est donc possible de réduire l'incertitude de type A en augmentant le nombre de mesures réalisées.

On introduit alors la notion d'intervalle de confiance. Il est définit par $[x_m - \delta x; x_m + \delta x]$ avec $\delta x = k \cdot s$ où k est le facteur d'élargissement. À chaque facteur d'élargissement est associée une probabilité pour que la valeur vraie $x_0$ soit dans l'intervalle de confiance. Les correspondances sont données dans des tables, mais on peut retenir que :

- pour $k = 1$, cette probabilité est de 68%
- pour $k = 2$, cette probabilité est de $95,5\%$
- pour $k = 3$, cette probabilité est de $99,7\%$

**★ Cas d'un petit nombre de mesures**

Dans le cas où le nombre $n$ de mesure est trop faible pour que la statistique soit suffisante, il est tout de même possible de généraliser les résultats obtenus, en modifiant le facteur d'élargissement (appelé alors coefficient de student $t_S$) pour obtenir le niveau de confiance souhaité. L'intervalle de confiance est alors donné par :

$$[x_m - t_s \cdot s; x_m + t_s \cdot s]$$

Le tableau suivant donne les valeurs de coefficient de student associés à différents niveaux de confiance, lorsque le nombre de mesures varie.

| n | 90% | 95% | 99% |
|---|---|---|---|
| 15 | 1,75 | 2,13 | 2,95 |
| 20 | 1,72 | 1,09 | 2,85 |
| 25 | 1,71 | 2,06 | 2,8 |
| 30 | 1,7 | 2,04 | 2,75 |
| 40 | 1,68 | 2,02 | 2,7 |
| 60 | 1,67 | 2 | 2,66 |
| 120 | 1,66 | 1,98 | 2,62 |
| ∞ | 1,64 | 1,96 | 2,58 |

## 2 - Incertitude de type B

Il est très souvent impossible de réaliser un grand nombre de fois une mesure donnée. On se place donc maintenant dans le cas d'une mesure unique. La meilleure estimation correspond à la valeur mesurée et il est possible d'exprimer l'incertitude-type suivant le cas qui se présente.

- Pour un appareil de mesure analogique (règle, vernier, ...), l'incertitude-type de lecture est obtenue par :
$$s = \frac{\text{graduation}}{\sqrt{12}}$$
- Pour un appareil numérique, si le constructeur donne l'incertitude sous la forme $\alpha\% + n \cdot UL$ ($UL$ signifie "unité de lecture" et représente la plus petite valeur lisible sur le calibre correspondant, dit souvent "dernier digit"), cela représente l'incertitude-type de mesure.
- Pour un appareil numérique, si le constructeur donne une incertitude $\Delta_C$ sans autre information, l'incertitude-type de mesure est définie par :
$$s = \frac{\Delta_C}{\sqrt{3}}$$
- Les indications données sur la verrerie en chimie correspondent directement aux incertitudes-type de mesure.

## 3 - Propagation des incertitudes

On cherche souvent en TP à mettre en évidence ou vérifier une loi, c'est-à-dire associer aux résultats expérimentaux une fonction $f$. Autrement dit, la grandeur $x$ que l'on cherche à déterminer peut s'exprimer comme une fonction $f(a, b, c)$ de plusieurs grandeurs mesurables $a$, $b$ et $c$. Comment se calcule alors l'incertitude-type sur $x$ ?

Après la mesure, on ne connait pas $a$, $b$ et $c$ mais $a \pm \delta a$, $b \pm \delta b$ et $c \pm \delta c$ où $\delta a$, $\delta b$ et $\delta c$ sont les erreurs de mesures, associées aux incertitudes-types $s_a$, $s_b$, $s_c$. On considère que les erreurs sont faibles vis à vis des mesures. On calcule donc la valeur de la fonction avec une erreur $\mathrm{d}f$ définie par :

$$\mathrm{d}f = f(a + \delta a, b + \delta b, c + \delta c) - f(a, b, c)$$

Comme les erreurs sont faibles, on peut rapprocher cette erreur de la différentielle de $f$ :

$$\mathrm{d}f = \left.\frac{\partial f}{\partial a}\right|_{(b,c)} \delta a + \left.\frac{\partial f}{\partial b}\right|_{(a,c)} \delta b + \left.\frac{\partial f}{\partial c}\right|_{(a,b)} \delta c$$

Chaque terme de la somme apporte une contribution à l'erreur sur $x$. Si les mesures sont indépendantes, on définit alors l'incertitude-type sur $x$ par :

$$s = \sqrt{\left(\frac{\partial f}{\partial a} s_a\right)^2 + \left(\frac{\partial f}{\partial b} s_b\right)^2 + \left(\frac{\partial f}{\partial c} s_c\right)^2} \tag{1}$$

**Remarque** : Attention, l'incertitude-type n'est en aucun cas donnée en injectant les incertitudes mesurées dans l'expression de la fonction ! !

$$s \neq f(\delta a, \delta b, \delta c)$$

Par exemple, on souhaite déterminer une fréquence en mesurant la période associée. On trouve $T = (0,020 \pm 0,001)$s. La fréquence est donc égale à $f = 50{,}0\,\mathrm{Hz}$ mais l'incertitude sur cette fréquence n'est bien évidemment pas 1000 Hz.

Voyons quelques exemples d'utilisation de la formule (1). Dans la plupart des cas, la fonction à calculer est soit une combinaison linéaire des grandeurs mesurées soit un quotient de ces valeurs.

**Combinaison linéaire.** Dans ce cas, la grandeur $x$ s'écrit $f(a,b,c) = \alpha a + \beta b + \gamma c$. La formule (1) donne alors :

$$s = \sqrt{(\alpha s_a)^2 + (\beta s_b)^2 + (\gamma s_c)^2}$$

**Quotient.** On suppose dans ce cas $x = f(a,b,c) = \dfrac{a \cdot b}{c}$. L'incertitude-type est alors :

$$s = \sqrt{\left(\frac{b}{c} s_a\right)^2 + \left(\frac{a}{c} s_b\right)^2 + \left(-\frac{a \cdot b}{c^2} s_c\right)^2}$$

Cette incertitude-type peut se mettre sous la forme :

$$\frac{s}{x} = \sqrt{\left(\frac{s_a}{a}\right)^2 + \left(\frac{s_b}{b}\right)^2 + \left(-\frac{s_c}{c}\right)^2}$$

Dans le cas d'une grandeur produit, l'incertitude-type relative est donc la somme quadratique des incertitudes-type relatives.

**Remarque** : On retouve ce résultat en utilisant une dérivée logarithmique. En effet, $\ln f = \ln a + \ln b + \ln c$ donc :

$$\mathrm{d}(\ln f) = \frac{\mathrm{d}f}{f} = \frac{\partial \ln f}{\partial a}\delta a + \frac{\partial \ln f}{\partial b}\delta b + \frac{\partial \ln f}{\partial c}\delta c = \frac{\delta a}{a} + \frac{\delta b}{b} + \frac{\delta c}{c}$$

## III – Ajustement de données

Réaliser un ajustement des données est utile lorsque l'on veut vérifier ou proposer une loi physique. Il y a toutefois quelques bonnes pratiques à respecter.

- Si on veut vérifier une loi connue (souvent notre cas), on se ramène au tracé d'une droite. Cela permet de voir très rapidement les écarts à la loi, de réaliser un ajustement plus précis avec un indicateur fiable de l'écart entre le modèle et l'expérience (le coefficient $R^2$) et d'adapter facilement les paramètres de l'ajustement si des écarts sont constatés.
- Si la loi est inconnue à priori, tracer les données brutes et tenter un ajustement linéaire. Si l'ajustement n'est pas correct, repérer si une tendance se dessine de l'écart entre les points et l'ajustement, et procéder à taton.

Dans tous les cas, une fois les points de mesure placés sur le graphique, ne pas oublier d'indiquer les barres d'erreurs. Sans elles, impossible de savoir si les mesures confirment une loi connue ou si le modèle choisi pour l'ajustement est valide.